

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第5150997号
(P5150997)

(45) 発行日 平成25年2月27日(2013.2.27)

(24) 登録日 平成24年12月14日(2012.12.14)

(51) Int.Cl.		F I	
C 1 2 N 15/09	(2006.01)	C 1 2 N 15/00	Z N A A
C O 7 K 14/47	(2006.01)	C O 7 K 14/47	
C 1 2 N 1/15	(2006.01)	C 1 2 N 1/15	
C 1 2 N 1/19	(2006.01)	C 1 2 N 1/19	
C 1 2 N 1/21	(2006.01)	C 1 2 N 1/21	

請求項の数 37 (全 75 頁) 最終頁に続く

(21) 出願番号	特願2001-571753 (P2001-571753)	(73) 特許権者	502349874
(86) (22) 出願日	平成13年3月27日(2001.3.27)		フォндаシオン・ジャン・ドセーセフ
(65) 公表番号	特表2003-528631 (P2003-528631A)		FONDATION JEAN DAUSSET-CEPH
(43) 公表日	平成15年9月30日(2003.9.30)		フランス国、75010パリ、リュウ・ジュリエット・ドデュ、27
(86) 国際出願番号	PCT/FR2001/000935	(74) 代理人	100081352
(87) 国際公開番号	W02001/072822		弁理士 広瀬 章一
(87) 国際公開日	平成13年10月4日(2001.10.4)	(72) 発明者	ユーゴ、ジャン・ピエール
審査請求日	平成20年3月7日(2008.3.7)		フランス国、75019パリ、リュウ・ドゥ・ティヨンビル、5
(31) 優先権主張番号	00/03832	(72) 発明者	トーマ、ジル
(32) 優先日	平成12年3月27日(2000.3.27)		フランス国、75005パリ、リュウ・ピュフォン、15
(33) 優先権主張国	フランス (FR)		
前置審査			

最終頁に続く

(54) 【発明の名称】 炎症性腸疾患に関与する遺伝子およびその用途

(57) 【特許請求の範囲】

【請求項 1】

下記群の配列から選択される核酸配列を含むことを特徴とする、精製または単離された核酸。

a) 配列番号 1 および配列番号 3 ;

b) a) に定義された配列に対応する相補的配列または RNA 配列。

【請求項 2】

配列番号 1 の配列、またはこの配列に対応する相補的配列もしくは RNA 配列を含むか、あるいはこれからなることを特徴とする、請求項 1 記載の精製または単離された核酸。

【請求項 3】

配列番号 2 のタンパク質をコードすることを特徴とする、精製または単離された核酸。

【請求項 4】

配列番号 2 の配列を含むことを特徴とする、単離されたポリペプチド。

【請求項 5】

配列番号 2 の配列からなることを特徴とする、請求項 4 記載のポリペプチド。

【請求項 6】

請求項 1 ~ 3 のいずれかに記載した核酸、または請求項 4 および 5 のいづ

れかに記載したポリペプチドをコードする核酸を含む、クローニング及び/又は発現ベクター。

【請求項 7】

請求項 6 に記載のベクターで形質転換されたことを特徴とする宿主細胞。

【請求項 8】

請求項 1 ~ 3 のいずれかに記載の核酸配列中の少なくとも 15 の連続したヌクレオチドを含む核酸の、請求項 1 ~ 3 のいずれかに記載の核酸配列を検出及び/又は増幅するための、プローブまたはプライマーとしての使用。

【請求項 9】

請求項 1 ~ 3 のいずれかに記載の核酸配列の、組換えポリペプチドを製造するための使用。

10

【請求項 10】

請求項 7 記載の細胞を、ポリペプチドの発現を可能にする条件下で培養し、そして組換えポリペプチドを回収することを特徴とする、組換えポリペプチドを得る方法。

【請求項 11】

請求項 10 記載の方法を用いて得られることを特徴とする、組換えポリペプチド。

【請求項 12】

請求項 4、5 または 11 のいずれかに記載のポリペプチドを選択的に結合することを特徴とする、モノクローナル抗体またはポリクローナル抗体。

20

【請求項 13】

下記工程を含むことを特徴とする、生物学的試料において請求項 4、5 または 11 のいずれかに記載のポリペプチドを検出する方法。

- a) 生物学的試料を請求項 12 記載の抗体に接触させる、
- b) 抗体と生物学的試料中のポリペプチドとの間に形成される複合体を免疫反応において検出する。

【請求項 14】

下記を含むことを特徴とする、請求項 13 記載の方法を行うための試薬キット。

30

- a) 請求項 12 記載のモノクローナル抗体またはポリクローナル抗体、
- b) 免疫反応に適した媒体を構成する試薬、
- c) 免疫反応の間に形成される抗原 - 抗体複合体を検出するための試薬。

【請求項 15】

配列番号 1 または配列番号 3 に対応する遺伝子の少なくとも 1 つの突然変異及び/又は発現の有害な変化の存在を、患者からの生物学的試料を用いて、該遺伝子に対応する核酸配列の全部または一部を解析することにより決定することを含む、クローン病の診断及び/又は予後の評価のための、IBD 1 遺伝子の解析方法であり、突然変異又は有害な変化が下記からなる群より選ばれる、前記方法：

- ・ 配列番号 1 の 2 1 0 7 番目のヌクレオチド、または配列番号 3 の 1 6 4 6 7 番目のヌクレオチドにおける C から T への変異
- ・ 配列番号 1 の 2 7 2 5 番目のヌクレオチド、または配列番号 3 の 2 7 0 5 9 番目のヌクレオチドにおける G から C への変異
- ・ 配列番号 1 の 3 0 2 2 番目のヌクレオチド、または配列番号 3 の 3 4 2 9 6 番目のヌクレオチドにおける C の挿入。

40

【請求項 16】

請求項 1 ~ 3 のいずれかの項記載の核酸配列を含むことを特徴とする、DNAチップ。

【請求項 17】

請求項 4、5 もしくは 11 のいずれかの項記載のポリペプチド、または

50

請求項 1 2 記載の抗体を含むことを特徴とする、タンパク質チップ。

【請求項 1 8】

下記工程を含むことを特徴とする、生物学的試料中において請求項 1 ~ 3 のいずれかの項記載の核酸を検出及び / 又は解析する方法。

- a) 標識された請求項 1 ~ 3 のいずれかの項記載のポリヌクレオチドを接触させる、
- b) 該ポリヌクレオチドと生物学的試料中の核酸との間に形成されたハイブリッドを検出及び / 又は解析する。

【請求項 1 9】

下記工程を含むことを特徴とする、生物学的試料中に請求項 1 ~ 3 のいずれかの項記載の核酸を検出及び / 又は解析する方法：

- a) 請求項 1 または 2 に記載の核酸から選択される、配列番号 1 または 3 の少なくとも 1 5 の連続したヌクレオチドの断片に対応するプライマーを用いた、該生物学的試料中の核酸の増幅、
- b) 生物学的試料からの増幅した核酸の検出および / または定量。

【請求項 2 0】

2 1 0 7 番目のヌクレオチドにおいて C から T への変異を有する、配列番号 1 を含む精製または単離された核酸。

【請求項 2 1】

2 7 2 5 番目のヌクレオチドにおいて G から C への変異を有する、配列番号 1 を含む精製または単離された核酸。

【請求項 2 2】

3 0 2 2 番目のヌクレオチドにおいて C が挿入された、配列番号 1 を含む精製または単離された核酸。

【請求項 2 3】

7 0 3 番目のアミノ酸が R から W へ置換した、配列番号 2 を含む単離されたポリペプチド。

【請求項 2 4】

9 0 9 番目のアミノ酸が G から R へ置換した、配列番号 2 を含む単離されたポリペプチド。

【請求項 2 5】

1 0 0 8 番目のアミノ酸が L から P へ置換した、配列番号 2 のアミノ酸 1 ~ 1 0 0 8 を含む単離されたポリペプチド。

【請求項 2 6】

下記を含む、ヒト個体において I B D 1 遺伝子中の変異核酸配列の存在または不在を決定する解析方法：

- (a) ヒト個体からの生物学的試料を、1 6 4 6 7 番目のヌクレオチドにおいて C から T への変異を有する配列番号 3、2 7 0 5 9 番目のヌクレオチドにおいて G から C への変異を有する配列番号 3、および 3 4 2 9 6 番目のヌクレオチドにおいて C の挿入を有する配列番号 3 からなる群より選ばれた変異核酸配列に特異的な核酸プローブと接触させること、および
- (b) 該核酸プローブと該変異核酸配列との間のハイブリダイゼーションを検出することにより、該変異核酸配列の存在または不在を決定すること。

【請求項 2 7】

1 6 4 6 7 番目のヌクレオチドにおいて C から T へ変異を有する配列番号 3 の変異核酸配列の存在または不在を決定する、請求項 2 6 記載の方法。

【請求項 2 8】

2 7 0 5 9 番目のヌクレオチドにおいて G から C へ変異を有する配列番号 3 の変異核酸配列の存在または不在を決定する、請求項 2 6 記載の方法。

【請求項 2 9】

3 4 2 9 6 番目のヌクレオチドにおいて C の挿入を有する配列番号 3 の変

10

20

30

40

50

異核酸配列の存在または不在を決定する、請求項 2 6 記載の方法。

【請求項 3 0】

少なくとも 2 種の前記変異核酸配列の存在または不在を決定する、請求項 2 6 記載の方法。

【請求項 3 1】

3 種の前記変異核酸配列の存在または不在を決定する、請求項 2 6 記載の方法。

【請求項 3 2】

前記核酸プローブが検出可能な標識を含む、請求項 2 6 記載の方法。

【請求項 3 3】

前記検出可能な標識が蛍光標識である、請求項 3 2 記載の方法。

【請求項 3 4】

前記生物学的試料がゲノムDNA である、請求項 2 6 記載の方法。

【請求項 3 5】

前記生物学的試料を、前記変異核酸配列に近接する核酸配列に特異的な一対の核酸プライマーと接触させることをさらに含む、請求項 2 6 記載の方法。

【請求項 3 6】

前記核酸プローブが少なくとも 1 5 ヌクレオチドの長さである、請求項 2 6 記載の方法。

【請求項 3 7】

ヒト個体においてクローン病を診断するための、請求項 2 6 ~ 3 6 のいずれかに記載の方法。

【発明の詳細な説明】

【0 0 0 1】

【産業上の利用分野】

本発明は、炎症性及び/もしくは免疫性疾患およびある種の癌、特に原因不明性炎症性腸疾患に關与する遺伝子、並びにこれらの遺伝子にコードされるタンパク質に関する。本発明はまた、炎症性疾患の診断方法にも關する。

【0 0 0 2】

【従来の技術】

原因不明性炎症性腸疾患 (IBD) は、その原因が分からない、消化管の炎症を特徴とする疾患である。炎症の位置および特性により、異なる 2 つの疾病分類学的実体、潰瘍性大腸炎 (UC) およびクローン病 (CD) が区別される。UC は S Wilkes により 1865 年に報告され、一方、限局性回腸炎の最初の患者はクローンにより 1932 年に報告された。実際は、これらの病気はさらにさかのぼることができる。

【0 0 0 3】

IBD は生涯にわたり進行する慢性の疾患であり、西欧諸国においては 1000 人につき約 1 ~ 2 人が罹患し、このことはフランスではこれらの病気の患者が 60,000 人 ~ 100,000 人であることを意味する。これらの疾患は若年層にあらわれ (症例のピークは 30 代である)、寛解期を挟んだ発病を経て進行し、栄養不足、子供における成長の遅滞、骨の脱鈣物質および最後に大腸癌への悪性化などの合併症をしばしば伴う。特別な治療法はない。通常の治療法は抗炎症薬、免疫抑制剤を使用し、また外科的処置を用いる。これらの治療手段はすべて、それら自身がかなり高い医原性疾患の原因である。これらすべての理由により、IBD は大きな公衆衛生上の問題と思われる。

【0 0 0 4】

IBD の病因はこれまで不明である。この病気の発生の長期にわたる増加および一卵性双生児での不完全な一致により示されるように、環境因子がこの病気の発生に關与する。現在知られている環境危険因子は、1) タバコ (その役割は CD において有害であり、UC においては有益である) および 2) 虫垂切除 (UC に対して防御的役割を有する) のみである。

【0 0 0 5】

10

20

30

40

50

これらの疾患をもつ人種集団および家族集団が存在することにより、ずっと以前より遺伝的素因が疑われてきた。事実、IBDはカフカス人の集団において、特に中央ヨーロッパのユダヤ人集団においてよくみられる。家族性の形態はIBD患者の6~20%になる。それらはこの疾患が早期に始まった時、特によくみられる。しかしながら、これらの疾患の遺伝性を確認することを可能にしたのは双生児における研究である。実際、これらの疾患の双生児間の一致率は、二卵性双生児よりも一卵性双生児においてより高く、これはIBD、特にCDに対する遺伝的要素を強力に裏付ける。多分IBDは、いくつかの異なる遺伝子を含み、互いにそして環境因子と相互作用する複合遺伝性疾患であろう。従って、IBDは多因子疾患に分類されうる。

【0006】

IBD-罹病性 (susceptibility、罹患しやすさ) の遺伝子を証明するために2つの主要な方策が開発された。第1は、生理病理学的原因の候補である遺伝子の解析に基づく。こうして、多くの遺伝子がIBDにとって重要である可能性があるとして提案された。それらは炎症および免疫応答においてある役割を有する遺伝子であることが多い。HLA、TAP、TNFおよびMICA遺伝子、Tリンパ球受容体、ICAM1、インタロイキン1、CCR5等が挙げられる。その他の遺伝子は、GAI2、モチリン、MRAMP、HMLH1などの種々の機能に關与する。実際、検討された各種候補遺伝子はいずれも、IBDの発生に何らかの役割を果たすことはこれまで明確には証明されていない。

【0007】

高度に多型性の遺伝的マーカーを用いた最近のヒトゲノム地図の進展により、遺伝学者が全ゲノムにわたる非標的化方法を開発することが可能となった。この方法は、逆遺伝学またはポジショナルクローニングと称され、病気に關与する遺伝子に關する仮説を立てずに、ゲノムの系統的スクリーニングによりそれらを発見しようとする。複合遺伝性疾患に最もよく使用される方法は、同じ家族内の罹患した個人の伝わり方による同一性を検討することに基づく。ゲノムに均一に(10cM毎に)分布する非常に多数(300~400)の多型マーカーについてこの値を算出する。患者間に過剰の同一性がある場合は、試験したマーカーはその疾患の罹病性 (susceptibility) に關する遺伝子を含むと考えられる領域を示す。複合遺伝性疾患の場合は、遺伝的素因の基礎となるモデル(遺伝子の数およびそれらの各々の相対的重要性)が未知であるので、使用する統計的方法を調整しなければならないであろう。

【0008】

本発明はIBDおよびその他の炎症性疾患に關与する遺伝子の核酸配列の証明、およびこれらの核酸配列の使用に關する。

本発明に關連して、本発明者等による予備検討によりCD罹病性の遺伝子の位置決定が既に可能となった。具体的には、本発明者等(Hugot et al., 1996)は、CD罹病性の遺伝子は第16染色体の動原体周囲(pericentromeric)の領域に位置する(図1)ことを示した。これは、ポジショナルクローニングにより位置決定され、文献(LanderおよびKruglyak, 1995)で提案された厳格な基準を満足するような、複合遺伝性疾患に対する罹病性の最初の遺伝子であった。この遺伝子はIBD1(inflammatory bowel disease 1、炎症性腸疾患1)と命名された。それ以来、別の著者等によりその他の位置、特に第12、1、3、6および7染色体上の位置が提案された(Satsangi et al., 1996; Cho et al., 1998)。それらは位置決定されたが、現在までこれらのIBD罹病性遺伝子のいずれも同定することはできなかった。

【0009】

数人の著者(Rioux et al., 1998)はこの位置を複製することができなかった。しかし、これは、遺伝的不均一性がありうる複合遺伝性疾患の場合には意外なことではない。

【0010】

同じポジショナルクローニングのアプローチにより、いくつかの免疫性および炎症性疾患、例えば硬直性脊椎炎、ブラウ症候群(Blau's syndrome)、乾癬など、については第16染色体上の位置が提案されている(Becker et al., 1998; Tromp et al., 1996)。そこで、

10

20

30

40

50

これらの疾患はすべて第16染色体上に位置する同じ遺伝子（または同じ遺伝子群）を共有しているかもしれない。

【0011】

事実、最大の遺伝子連鎖試験は、2cMしか離れていないD16S409 またはD16S411 の領域における同じポジションに常に位置している。この結果は、非パラメトリック連鎖解析を用いるアプローチにより遺伝子位置決定に帰属されうる、かなりの大きさの信頼区間（通常20cMより大）と矛盾する。

【0012】

本発明者等による検討において使用した統計試験の比較により、家系による完全な同一性（Tz2）に基づく試験は、家系による同一性の平均（Tz）に基づくものより良好であることが示される（図1）。かかる違いはIBD1の劣性効果により説明できる。

10

【0013】

第16染色体の動原体周囲の領域にあることが知られるいくつかの遺伝子、例えばインターロイキン4受容体、CD19、CD43またはCD11、はCDに対する可能性の高い候補であるようである。しかし、予備試験の結果は、これらの遺伝子がCDに関与していることを有利に示すものではない。

【0014】

特に、本発明はIBD1遺伝子の配列だけでなく、以下の実施例での報告にあるように実証された、IBDの近傍に位置するためにIBD1proxと称される別の遺伝子の部分的配列にも関する。従って、そのcDNA配列がそれぞれ配列番号1および配列番号4に対応するこれらの遺伝子は、多くの炎症性及び/又は免疫性疾患および癌に関与している可能性がある。

20

【0015】

IBD1およびIBD1prox遺伝子により発現されるペプチド配列は、それぞれ配列番号2および配列番号5により表される。これらの遺伝子のゲノム配列はそれぞれ配列番号3および配列番号6により表される。

【0016】

従って、本発明の主題は、下記群の配列から選択される核酸配列を含むことを特徴とする、精製または単離された核酸である。

- a) 配列番号1、配列番号3、配列番号4および配列番号6；
- b) 配列番号1、配列番号3、配列番号4および配列番号6から選択される配列の少なくとも15の連続したヌクレオチドの断片の配列；
- c) 最適の整列化後、a)またはb)に規定された配列と少なくとも80%の同一性割合を有する核酸配列；
- d) 高ストリンジェントな条件下でa)またはb)に規定された核酸配列とハイブリダイズする核酸配列；
- e) a)、b)、c)またはd)に規定された配列に対応する相補的配列またはRNA配列。

30

【0017】

c)で規定された本発明の核酸配列は、最適の整列化後、上記a)またはb)に規定された配列と少なくとも80%、好ましくは90%、最も好ましくは98%の同一性割合を有する。

【0018】

本明細書において区別せずに用いられる、「核酸」、「核酸配列」、「ポリヌクレオチド」、「オリゴヌクレオチド」、「ポリヌクレオチド配列」および「ヌクレオチド配列」なる用語は、改変されていなくてもよい、一連のヌクレオチドを意味し、異常なヌクレオチドを含んでいなくてもよく、そして二本鎖DNA D、一本鎖DNAおよび該DNAの転写産物に対応してもよい、核酸の断片または領域を規定するのを可能にする。従って、本発明の核酸配列はPNA（ペプチド核酸）なども含む。

40

【0019】

本発明は天然の染色体環境における、すなわち天然の状態のヌクレオチド配列には関しないことは理解されるべきである。それらは、単離及び/又は精製された、すなわち、例えば複製により直接または間接に採取され、それらの環境は少なくとも部分的に改変された

50

ものであるような、配列である。従って、化学合成により得られる核酸も含まれる。

【0020】

本発明の目的にとって、2つの核酸またはアミノ酸配列の間の「同一性割合」なる用語は、最良の整列化後に得られる、比較すべき2つの配列間で同じであるヌクレオチドまたはアミノ酸残基の割合を意味し、この割合は純粋に統計的であり、2つの配列間の違いはランダムにその全長にわたり分布する。「最良の整列化」または「最適の整列化」は、以下のようにして決定する同一性割合が最も高い整列化を意味する。2つの核酸またはアミノ酸配列間の配列比較は、それらを最適に整列させた後これらの配列を比較することにより慣用の方法で行われ、この比較は配列類似性を有する局部的領域を同定し比較するために、セグメントによりまたは「比較ウインドウ」により行われる。比較のための配列の最適整列化は、手作業の他、Smith およびWaterman (1981) の局所相同性アルゴリズム、Neddelmann およびWunsch (1970) の局所相同性アルゴリズム、Pearson およびLipman (1988) の類似性調査法、これらのアルゴリズムを用いたコンピュータープログラム (Wisconsin Genetics Software Package, Genetics Computer Group, 575 Science Dr., Madison, WI のGAP, BESTFIT, BLAST P, BLAST N, FASTA およびTFASTA) により行うことができる。最適の整列化を得るには、BLOSUM 62 マトリックスを有するBLAST が好ましく使用される。PAM またはPAM250マトリックスも使用できる。

10

【0021】

2つの核酸またはアミノ酸配列間の同一性割合は、最適に整列化したこれらの2つの配列を比較することにより決定され、比較されるべき核酸またはアミノ酸配列は、これら2つの配列間の最適整列化のための参照配列に対して付加または欠失を多分含む。同一性割合は、ヌクレオチドまたはアミノ酸残基が2つの配列間で同一である同一位置の数を決定し、この同一位置の数を比較する全位置数で割り、そして得られた数を100倍して、これら2つの配列間の同一性割合を得ることにより算出する。

20

【0022】

「最適の整列化後に参照配列と少なくとも80%、好ましくは90%、より好ましくは98%の同一性割合を有する核酸配列」なる表現は、参照核酸配列と比較した場合にある種の改変、例えば特に欠失、短縮、伸長、キメラ的融合及び/又は置換(特に点での)を有する核酸配列であり、その核酸配列が最適の整列化後に参照核酸配列と少なくとも80%、好ましくは90%、より好ましくは98%の同一性割合を有する核酸配列を意味する。それらは好ましくは、その相補的配列が本発明の配列番号1または配列番号4の配列に特異的にハイブリダイズしうる配列である。好ましくは、特異的または高ストリンジェントなハイブリダイゼーション条件は、2つの配列の一方と、他方に相補的な配列との間で、最適の整列化後に、少なくとも80%、好ましくは90%、より好ましくは98%の同一性を確実にするようなものであろう。

30

【0023】

高ストリンジェントな条件下でのハイブリダイゼーションとは、温度およびイオン強度の条件が、2つの相補的 DNA断片間のハイブリダイゼーションが維持されることを可能にするように選択されることを意味する。実例として、上記したポリヌクレオチド断片を規定する目的の、ハイブリダイゼーション工程のための高ストリンジェントな条件は有利には以下の通りである。

40

【0024】

DNA-DNA またはDNA-RNA ハイブリダイゼーションは2つの工程において行われる：(1) 5 × SSC (1 × SSC は0.15M NaCl + 0.015Mクエン酸ナトリウムの溶液に相当する)、50%ホルムアミド、7%トデシル硫酸ナトリウム (SDS)、10 × デンハート溶液、5%デキストラン硫酸および1%サケ精子 DNAを含むリン酸塩緩衝液 (20 mM, pH 7.5) 中42 °C で3時間の予備ハイブリダイゼーション、(2) プロブの長さに応じた温度 (すなわち、プロブ > 100 ヌクレオチド長に対して42 °C) での20時間のハイブリダイゼーション自体、次いで2 × SSC + 2% SDS 中20 °C で20分間の洗浄2回、および 0.1 × SSC + 0.1% SDS 中20 °C で20分間の洗浄1回。最後の洗浄を 0.1 × SSC + 0.1% SDS 中、プロブ > 100 ヌクレオチド長

50

に対して60 で30分間行う。規定の長さのポリヌクレオチドに対する上記高ストリンジェントなハイブリダイゼーション条件は、Sambrook et al., 1989 の教示によって、より長いまたは短いオリゴヌクレオチドに対して当業者により調整できる。

【0025】

最適の整列化後本発明の配列と少なくとも80%、好ましくは90%、より好ましくは98%の同一性割合を有する核酸配列の中で、配列番号1または配列番号4の変異核酸配列、またはその断片、すなわち対立遺伝子性変異体、すなわち配列番号1または配列番号4の配列の個々の変異、に対応するすべての核酸配列もまた好ましい。これらの自然突然変異配列は哺乳動物、特にヒトに存在する多型、特に病的状態を引き起こすかもしれない多型に相当する。好ましくは本発明は、突然変異により配列番号1または配列番号4の正常配列によりコードされるポリペプチドまたはその断片のアミノ酸配列の改変を生じている変異核酸配列に関する。

10

【0026】

「変異核酸配列」なる表現はまた、そのcDNAが配列番号1または配列番号4の配列を有するゲノム核酸配列のスプライス部位の突然変異及び/又は変異から生じる任意のRNAまたはcDNAも意味する。

【0027】

本発明は好ましくは、配列番号1もしくは配列番号4の配列、その相補的配列、または配列番号1もしくは配列番号4に対応するRNA配列のいずれかを含むかまたはそれからなることを特徴とする、精製または単離された本発明の核酸配列に関する。

20

【0028】

本発明核酸の配列を含むことを特徴とするプローブまたはプライマーもまた本発明の一部である。

従って、本発明はまた、特に、変異核酸配列を実証または区別すること、あるいは、そのcDNAが配列番号1または配列番号4で表される遺伝子のゲノム配列を、特にPCR法もしくは関連方法などの増幅方法を用いて同定することを可能にする、本発明のプライマーまたはプローブに関する。

【0029】

本発明はまた、本発明の核酸配列の、ある核酸配列を検出、同定、解析または増幅するためのプローブまたはプライマーとしての使用に関する。

30

本発明によれば、ある核酸配列を検出、同定、解析または増幅するための方法においてプローブまたはプライマーとして使用できるポリヌクレオチドは長さが最小15塩基、好ましくは20塩基、またはさらに好ましくは25~30塩基である。

【0030】

本発明のプローブおよびプライマーは、検出可能な、及び/又は定量可能なシグナルを得るために、当業者に周知の方法を用いて放射性または非放射性化合物で直接または間接的に標識してもよい。

【0031】

標識されていない本発明のポリヌクレオチド配列をプローブまたはプライマーとして直接使用することもできる。

40

一般的には多くの用途に使用できる配列を得るために配列を標識する。本発明のプライマーまたはプローブは放射性要素または非放射性分子で標識する。

【0032】

使用される放射性同位体には、 ^{32}P 、 ^{33}P 、 ^{35}S 、 ^3H または ^{125}I が挙げられる。非放射性の物質は、ビオチン、アビジン、ストレプトアビジンまたはジオキシゲニン (dioxygenin) などのリガンド、ハプテン、染料および発光試薬 (放射発光、化学発光、生物発光、蛍光または燐光試薬など) から選ばれる。

【0033】

従って、本発明のポリヌクレオチドは、特にPCR (ポリメラーゼ連鎖反応) 手法 (Rolf s et al., 1991) を用いる方法においてプライマー及び/又はプローブとして使用するこ

50

とができる。この手法では、増幅すべき断片の境を設けるオリゴヌクレオチドプライマー対を選ぶことが必要である。例えば、米国特許第4,683,202号に記載の手法が例示される。増幅断片は、例えば、アガロースもしくはポリアクリルアミドゲル電気泳動の後、またはゲル濾過やイオン交換クロマトグラフィーなどのクロマトグラフィー手法の後に同定し、次いで配列決定できる。増幅の特異性は、プライマーとして本発明ポリヌクレオチドの核酸配列を用い、マトリックスとしてこれらの配列を含むプラスミドあるいは誘導された増幅産物を用いて調節できる。増幅されたヌクレオチド断片は、生物学的試料中の、この増幅ヌクレオチド断片の配列に相補的な配列の標的核酸の存在を実証するために、ハイブリダイゼーション反応における試薬として使用できる。

【0034】

本発明はまた、本発明のプライマーを用いた増幅により得ることができる核酸にも関する。

標的核酸を増幅するその他の手法を、本発明のヌクレオチド配列を有するプライマー対を用いるPCR法(PCR変法)の代わりとして有利に用いることができる。「PCR変法」なる用語は、核酸配列の直接または間接的再生産を用いる、あるいは標識系が増幅されたすべての方法を意味する。これらの方法は当然既知である。一般的に、それらはDNAのポリメラーゼによる増幅を含み、もとの試料がRNAである場合は逆転写を予め実施しておくべきである。かかる増幅には現在、非常に多くの方法が存在し、例えば、SDA (strand displacement amplification、鎖置換増幅)法(Walker et al., 1992)、Knoh et al., (1989)により既報のTAS (transcription-based amplification system、転写に基づく増幅系)法、Guatelli et al. (1990)により既報の3SR (self-sustained sequence replication、自立配列複製)法、Kievitis et al. (1991)により既報のNASBA (nucleic acid sequence based amplification、核酸配列に基づく増幅)法、TMA (transcription mediated amplification、転写媒介増幅)法、Landegren et al. (1988)により既報の(ligase chain reaction、リガーゼ連鎖反応)法、Segev (1992)により既報のRCR (repair chain reaction、修復連鎖反応)法、Duck et al. (1990)により既報のCPR (cycling probe reaction、循環プローブ反応)法およびMiele et al. (1983)により既報のQ-ベータ-レプリカーゼ増幅(Q-beta-replicase amplification)法がある。これらの手法のいくつかはこれまで改善されている。

【0035】

検出すべき標的ポリヌクレオチドがmRNAである場合は、生物学的試料中に含まれるmRNAからcDNAを得るために、本発明のプライマーを用いて増幅反応を行うか、または本発明のプローブを用いる検出法を実施する前に、逆転写酵素型の酵素が有利に使用される。次いで、得られるcDNAは、本発明による増幅または検出方法に用いるプライマーまたはプローブの標的として働く。

【0036】

プローブハイブリダイゼーション手法は多くのやり方で実施できる(Matthews et al., 1988)。最も一般的な方法は、各種の組織の細胞、または培養中の細胞から抽出した核酸を支持体(ニトロセルロース、ナイロンまたはポリスチレンなど)に固定し、固定した標的核酸を規定の条件下でプローブと共にインキュベートすることからなる。ハイブリダイゼーション後、過剰のプローブを除去し、形成された雑種分子を適宜方法(プローブに結合した放射活性、蛍光または酵素活性の測定)を用いて検出する。

【0037】

本発明の核酸プローブの別の態様によれば、これは捕捉プローブ(capture probe)として使用できる。この場合、「捕捉プローブ」と称されるプローブは、支持体上に固定され、特異的ハイブリダイゼーションにより、試験すべき生物学的試料から得られる標的核酸を捕捉するのに使用され、次いで標的核酸を、容易に検出する要素で標識した「検出プローブ」と称される第2のプローブを用いて検出する。

【0038】

従って、有利な核酸断片には、特にアンチセンスオリゴヌクレオチド、すなわち、その構

10

20

30

40

50

造が、標的配列とのハイブリダイゼーションにより対応する産物の発現の阻害を確実にするようなオリゴヌクレオチドが挙げられる。また、対応する産物の発現の調節に關与するタンパク質と相互作用することにより、この発現の阻害または活性化を生じるセンスオリゴヌクレオチドも例示される。

【0039】

いずれの場合（センスおよびアンチセンス）も本発明のオリゴヌクレオチドをin vitroおよびin vivoで使用できる。

本発明はまた、以下から選択されるポリペプチドを含むことを特徴とする、単離されたポリペプチドに關する。

- a) 配列番号2または配列番号5の配列を有するポリペプチド；
- b) a)で定義した配列を有するポリペプチドの変異ポリペプチド；
- c) a)のポリペプチドと少なくとも80%の同一性を有する、a)またはb)で定義したポリペプチドに相同的なポリペプチド；
- d) a)、b)またはc)に定義したポリペプチドの少なくとも15の連続するアミノ酸の断片；
- e) a)、b)またはc)に定義したポリペプチドの生物学的に活性な断片。

10

【0040】

本発明の目的にとって、「ポリペプチド」なる用語はタンパク質またはペプチドを意味する。

「生物学的に活性な断片」なる表現は、それが由来するペプチド断片と、同じ生物活性、好ましくは同じ桁の範囲内（10倍の範囲内）の活性を有する断片を意味する。そこで、実施例はIBD1タンパク質（配列番号2）がアポトーシス現象における役割を有する可能性があることを示す。よって、IBD1タンパク質の生物学的に活性な断片は、配列番号2に由来し、かつアポトーシスにおける役割を有するポリペプチドからなる。以下の実施例は、IBD1およびIBD1proxタンパク質のペプチドドメインの機能として、これらのタンパク質に關する生物学的機能を提案し、従って、当分野の技術者が生物学的に活性な断片を同定することを可能にする。

20

【0041】

好ましくは、本発明のポリペプチドは配列番号2の配列（IBD1遺伝子によりコードされるタンパク質に相当）、もしくは配列番号5の配列（IBD1proxによりコードされるタンパク質に相当）、または最適の整列化後に配列番号2もしくは配列番号5と少なくとも80%の同一性を有する配列からなるポリペプチドである。

30

【0042】

このポリペプチドの配列は、最適の整列化後に配列番号2または配列番号5の配列と少なくとも80%、好ましくは90%、より好ましくは98%の同一性割合を有する。

【0043】

「そのアミノ酸配列が、最適の整列化後に参照配列と少なくとも80%、好ましくは90%、より好ましくは98%の同一性割合を有するポリペプチド」なる表現は、参照ポリペプチドと比較してある種の改変、例えば、特に1または2以上の欠失及び/又は短縮、伸長、キメラ的融合及び/又は1または2以上の置換を有するポリペプチドを意味する。

【0044】

そのアミノ酸配列が、最適の整列化後に本発明の配列番号2または配列番号5の配列、あるいはその断片と、少なくとも80%、好ましくは90%、より好ましくは98%の同一性割合を有するポリペプチドの中で好ましいのは、上で規定した変異核酸配列によりコードされる変異ポリペプチド、特に、そのアミノ酸配列が配列番号2または配列番号5の配列あるいはその断片と比べて少なくとも1つのアミノ酸残基の短縮、欠失、置換及び/又は付加に相当する少なくとも1つの突然変異を有するポリペプチド、より好ましくは病的状態に關連する突然変異を有する変異ポリペプチドである。

40

【0045】

本発明はまた、本発明のポリペプチドをコードする核酸を含むクローニングベクター及び/又は発現ベクターに關する。かかるベクターは、発現および、場合により宿主細胞中で

50

のポリペプチドの分泌に必要とされる要素も含んでいてよい。かかる宿主細胞もまた本発明の主題である。

【0046】

本発明のプロモーター及び/又は調節配列を含むことを特徴とするベクターもまた本発明の一部である。

このベクターは好ましくは、プロモーター、翻訳開始および終止シグナル、並びに転写の調節に適切な領域も含む。それらは細胞中で安定的に保持できなければならない、場合により翻訳されたタンパク質の分泌を指定する特定のシグナルを含んでもよい。

【0047】

これらの各種調節シグナルは、使用される細胞性宿主に関連して選択される。この趣旨で、本発明の核酸配列を、選択された宿主中で自律的に複製するベクターに挿入しても、または選択された宿主に組み込まれるベクターに挿入してもよい。

【0048】

自律的に複製する系の中で好ましいのは、宿主細胞に応じて、プラスミドまたはウイルス型の系が使用され、ウイルスベクターとしては特にアデノウイルス (Perricaudet et al., 1992)、レトロウイルス、レンチウイルス、ポックスウイルスまたはヘルペスウイルス (Epstein et al., 1992) が可能である。当分野の技術者はこれらの系のそれぞれについて使用できる技術を知っている。

【0049】

宿主細胞の染色体への配列の組み込みが望ましい場合、例えば、プラスミドまたはウイルス型の系が使用でき、かかるウイルスは例えばレトロウイルス (Temin, 1986) または AAV (Carter, 1993) である。

【0050】

非ウイルスベクターの中で好ましいのは、VICAL 社により開発された手法による、裸の DNA または裸の RNA などの裸のポリヌクレオチド、細菌の人工染色体 (BAC)、酵母での発現のための酵母の人工染色体 (YAC)、マウス細胞での発現のためのマウスの人工染色体 (MAC)、好ましくはヒト細胞での発現のためのヒトの人工染色体 (HAC) である。

【0051】

かかるベクターは当分野の技術者により通常使用される方法により製造され、それより得られるクローンは、例えばリポフェクション、電気穿孔法、熱ショック、膜の化学的透過性化後の形質転換または細胞融合などの標準的方法を用いて適宜宿主に導入することができる。

【0052】

本発明はまた、本発明のベクターで形質転換された宿主細胞、特に真核細胞および原核細胞、並びにこの本発明の形質転換された細胞の1種を含む、ヒト以外のトランスジェニック動物、好ましくは哺乳動物をも含む。これらの動物は、炎症性及び/又は免疫性疾患、特に消化管の炎症性疾患の病因の研究、あるいは癌の研究のためのモデルとして使用できる。

【0053】

本発明の目的のために使用できる細胞には、細菌細胞 (Olins and Lee, 1993)、酵母細胞 (Buckholz, 1993) および動物細胞、特に哺乳動物細胞 (Edwards and Aruffo, 1993)、殊にチャニーズハムスター卵巣 (CHO) 細胞が挙げられる。また、例えばバキュロウイルス (Luckow, 1993) を用いる方法を使用できる昆虫細胞も挙げられる。本発明のタンパク質の発現のための好ましい細胞性宿主は COS 細胞からなる。

【0054】

本発明の動物の中では、本発明のポリペプチドを発現する齧歯類、特にマウス、ラットまたはウサギなどの動物が好ましい。

本発明の哺乳動物の中で好ましいのは、マウス、ラットまたはウサギなどの動物であり、これらの動物は配列番号2もしくは配列番号5の配列を有するタンパク質をコードする遺伝子、またはその配列がこれらの動物における相同 (homologous) 遺伝子によりコードさ

10

20

30

40

50

れるものが機能しないか、ノックアウトされているか、あるいは少なくとも1つの変異を含むことを特徴とする動物である。

【0055】

これらのトランスジェニック動物は、例えば、胚性幹細胞での相同的組換え、これらの幹細胞の胚への移行、生殖系列で作製されたキメラの選択およびこのキメラの成育により得られる。

【0056】

本発明のトランスジェニック動物は、本発明のタンパク質をコードする遺伝子、またはその相同遺伝子を過発現しても、また変異が導入された該遺伝子を発現してもよい。これらのトランスジェニック動物、特にマウスは、例えば、強力で偏在する、もしくは組織型に

10

【0057】

あるいは本発明のトランスジェニック動物は、LOXP/CREリコンビナーゼ系 (Rohlmann et al, 1996) またはこの遺伝子の発現を不活化するためのその他の任意の系を用いた不活化により、配列番号2もしくは配列番号5の配列のポリペプチドの1つをコードする遺伝子、またはその同種遺伝子を欠損させてもよい。

【0058】

本発明の細胞および哺乳動物は、以下に記載する、本発明のポリペプチドを産生する方法に使用することができ、そしてまた、解析用のモデルとして使用してもよい。

20

【0059】

上記形質転換された細胞または哺乳動物はまた、本発明のポリペプチドと、本発明のポリペプチドの活性に直接もしくは間接的に関与する化合物もしくはタンパク質化合物との間の相互作用を研究するためのモデルとして使用でき、これは関与する各種機構および相互作用を研究するためである。

【0060】

それらは特に、補助因子もしくは阻害剤、特に競合阻害剤として、本発明のポリペプチド、特に本発明の配列番号2もしくは配列番号5の配列を有するタンパク質またはその変異体と相互作用する産物、あるいは本発明のポリペプチドの活性に関してアゴニストもしくは

30

【0061】

本発明はまた、本発明のポリペプチドと直接または間接的に相互作用でき、及び/又はこれらのポリペプチドの発現もしくは活性を調節しうる化学的もしくは生化学的化合物をスクリーニングするための、本発明の細胞、哺乳動物またはポリペプチドの使用に関する。

【0062】

同様に、本発明は、in vitroもしくはin vivo で本発明の核酸と相互作用しうる化合物を、本発明の核酸、細胞もしくは哺乳動物を用いてスクリーニングし、候補化合物および本発明の核酸との複合体の形成を検出する方法にも関する。

40

【0063】

こうして選択された化合物もまた本発明の主題である。

本発明はまた、組換えポリペプチドの合成のための、本発明の核酸の使用にも関する。

【0064】

それ自体本発明に含まれる、組換え体の形態の本発明のポリペプチドを製造する方法は、形質転換細胞、特に本発明の細胞または哺乳動物を、本発明の核酸配列によってコードされる組換えポリペプチドの発現を可能にする条件下で培養し、この組換えポリペプチドを回収することを特徴とする。

【0065】

この製造方法を用いて得ることができる組換えポリペプチドもまた本発明の一部である。

50

上記のようにして得られた組換えポリペプチドは、グリコシル化形態でも非グリコシル化形態でもよく、そして天然の3次構造を有しても有していなくてもよい。

【0066】

組換えポリペプチドの配列はまた、その溶解性、特に水性溶媒中での溶解性を改善するために改変されていてもよい。

例えば、疎水性ドメインの欠如や疎水性アミノ酸の親水性アミノ酸での置換などの、かかる改変は当分野の技術者には既知である。

【0067】

これらのポリペプチドは、当分野の技術者に既知の組換えポリペプチドの製造のための技術により、上記核酸配列を用いて製造できる。この場合、使用する核酸配列は、細胞性宿主におけるその発現を可能にするシグナルの制御下に置かれる。

10

【0068】

組換えポリペプチドの製造のための有効な系は、本発明のベクターおよび宿主細胞を有することが必要である。

これらの細胞は、上記ベクターに挿入されたヌクレオチド配列を宿主細胞に導入し、次いでこの細胞を、トランスフェクトされたヌクレオチド配列の複製及び/又は発現を可能にする条件下で培養することにより得ることができる。

【0069】

組換えポリペプチドを精製するのに使用される方法は、当分野の技術者に既知である。組換えポリペプチドは、細胞溶解液またはその抽出物から、あるいは培地上清から、分画、クロマトグラフィー法、特異的モノクローナルもしくはポリクローナル抗体を用いた免疫アフィニティー法などを単独あるいは組み合わせて使用する方法により精製することができる。

20

【0070】

本発明のポリペプチドはまた、多数の既知のペプチド合成方法の1つを用いた化学合成、例えば固相を用いる方法（特に、Stewart et al., 1984）または部分的固相を用いる方法、断片縮合、または溶液中の通常の合成により得ることもできる。

【0071】

化学合成により得られ、対応する非天然のアミノ酸を含んでいてよいポリペプチドもまた本発明に含まれる。

30

本発明のポリペプチドを特異的に認識しうることを特徴とする、モノクローナルもしくはポリクローナル抗体、もしくはその断片、キメラ抗体または免疫複合体 (immunoconjugate) は本発明の一部である。

【0072】

特異的ポリクローナル抗体は、本発明のポリペプチド、特に遺伝的組換えまたは通常のコラーゲンによるペプチド合成により製造されたポリペプチドで免疫された動物の血清から得ることができる。

【0073】

本発明のある種のポリペプチド、変異体またはその免疫原性断片を特異的に認識する抗体の利点が特に言及される。

40

配列番号2または配列番号5の配列のポリペプチドを認識しうることを特徴とする、モノクローナルもしくはポリクローナル抗体、もしくはその断片、キメラ抗体または免疫複合体が特に好ましい。

【0074】

特異的モノクローナル抗体は、KohlerおよびMilstein (1975) により既報のハイブリドーマ培養の慣用方法により得ることができる。

本発明の抗体は、例えばキメラ抗体、ヒト化抗体またはFab もしくはF(ab')₂断片である。それらはまた、検出可能な及び/又は定量可能なシグナルを得るために、免疫複合体または標識抗体の形態であってもよい。

【0075】

50

本発明はまた、本発明の抗体を用いることを特徴とする、本発明のポリペプチドを検出及び/又は精製する方法に関する。

本発明は、本発明方法を用いて得られることを特徴とする精製ポリペプチドも含む。

【0076】

さらに、本発明の抗体、特にモノクローナル抗体は、ポリペプチドの精製のための使用に加え、生物学的試料中のこれらのポリペプチドの検出にも使用できる。

【0077】

従って、それらは、本発明ポリペプチド、特に配列番号2もしくは配列番号5の配列のポリペプチドまたはその変異体の特定組織部分上での発現の、例えば、免疫蛍光法、金標識及び/又は酵素的免疫複合体を用いる免疫細胞化学的または免疫組織化学的解析のための手段を構成する。

10

【0078】

それらは特に、生物学的試料または組織中のこれらのポリペプチドの異常な発現を証明することを可能にする。

さらに一般的には、本発明の抗体は、正常なまたは突然変異した本発明ポリペプチドの発現を観察しなければならない任意の状況において有利に使用できる。

【0079】

従って、生物学的試料を本発明の抗体に接触させる工程、および形成される抗原-抗体複合体を証明する工程を含む、生物学的試料中の本発明のポリペプチド検出方法も本発明の主題であり、またかかる方法を実施するためのキットもそうである。かかるキットは特に以下を含む。

20

- a) 本発明のモノクローナルもしくはポリクローナル抗体、
- b) 場合により、免疫反応に適した媒体を構成するための試薬、
- c) 免疫反応の間に生成する抗原-抗体複合体を検出する試薬。

【0080】

本発明の抗体はまた、IBD1遺伝子またはIBD1prox遺伝子の異常な発現がみられる場合の、ヒトにおける炎症性及び/もしくは免疫性疾患、または癌の治療に使用できる。異常な発現とは過発現または変異したタンパク質の発現を意味する。

【0081】

これらの抗体はヒト血清から直接得ても、本発明のポリペプチドで免疫した動物から得て、次いで「ヒト化」してもよく、そのままで、あるいは上記疾患の治療のための医薬品の製造において使用してもよい。

30

【0082】

本発明の核酸配列、ポリペプチドまたは抗体を用いることを特徴とする、対立遺伝子変動性、突然変異、欠失、ヘテロ接合性の消失、または本発明遺伝子の任意の遺伝的異常を決定する方法もまた本発明の一部である。

【0083】

本発明は事実、炎症性及び/もしくは免疫性疾患、特にIBDに關与するIBD1およびIBD1prox遺伝子の配列を提供する。本発明の方法の1つは、これらの炎症性及び/もしくは免疫性疾患の1つに対応する表現型と関連する、これらの核酸またはポリペプチド配列における突然変異を特定することである。

40

【0084】

これらの突然変異は本発明の核酸および配列(ゲノムDNA、RNAまたはcDNA)の解析により直接検出できるだけでなく、本発明のポリペプチドを介しても可能である。特に、突然変異を有するエピトープを認識する本発明の抗体を使用することにより、「健全な」タンパク質と「病状と関連する」タンパク質との間の区別をすることができる。

【0085】

従って、以下の実施例により実証されるように、各種炎症性及び/又は免疫性のヒト疾患におけるIBD1遺伝子の検討により、この遺伝子の配列の変異がクローン病、潰瘍性大腸炎およびブラウ(Blau)症候群に存在することが示される。これらの配列の変異は推定され

50

るタンパク質配列においてかなりの変化をもたらす。事実、それらは重要な機能性ドメイン中のタンパク質の非常に保存された部位上に位置するか、短縮されたタンパク質の合成を生じる。従って、これらの有害な変化はタンパク質の機能の変化をもたらす、そしてこれらの疾患の発生において原因となる影響を及ぼす可能性が極めて高い。

【0086】

これらの突然変異がみられる様々な疾患は、IBD1遺伝子が多くの炎症性及び/又は免疫性疾患において重要である可能性があることを示唆する。この結果は、第16染色体の動原体周囲領域が、各種ヒト疾患、例えば強直性脊椎炎や乾癬性関節炎、に対する罹病性に関与する遺伝子を含むとして報告された事実と比べるべきである。従って、IBD1遺伝子は多くの炎症性及び/又は免疫性疾患において重要役割をもつことが考えられる。

10

【0087】

特に、IBD1は肉芽腫性の炎症性疾患に関与しうる。ブラウ症候群およびCDは事実、この系統の一部である疾患である。従って、IBD1遺伝子における変化は同じ系統のその他の疾患(サルコイド症、ベーチェット病など)に関しても見出されるであろう。

【0088】

さらに、アポトーシスに至る細胞経路におけるIBD1の関与は、その発癌性の役割の可能性の問題も提起する。実際、IBD1の調節異常は癌への素因となることが予想される。この仮説は、大腸癌の素因が炎症性腸疾患に存在するという事実により裏付けられる。IBD1は癌へのこの罹病性を一部説明でき、新しい発癌性経路を明かにしうる。

【0089】

IBD1遺伝子で観察しうる突然変異の正確な記述により、この役割が実証される炎症性または免疫性疾患の分子診断(molecular diagnosis)の基礎をきずくことが可能となる。この遺伝子での突然変異の検索に基づくかかるアプローチは、これらの疾患の診断への寄与、そして多分、侵襲性であり高価なある種の追加検査の量を減らすことを可能にするであろう。本発明はIBD1における突然変異の検索に基づくかかる分子診断の基礎をつくるものである。

20

【0090】

炎症性疾患の分子診断はまた、これらの疾患の疾病分類学的分類を改善し、特定の疾患のサブグループをその臨床的特性、病気の進行性またはある種の処置に対する応答性によりより明確に規定することを可能にする。例えば、存在する突然変異を除去すると、10%以上炎症性腸疾患を発症する現在原因不明の型の大腸炎を分類することが可能となるかもしれない。かかるアプローチはそれぞれの患者に適した初期処置を提案することを可能にするであろう。一般的に、かかるアプローチは、最後には、治療用および予防用の手段を含む、各病気の遺伝的領域に応じたその病気の個別的処置を決定しうることを望める。

30

【0091】

特に、患者からの生物学的試料を用い、遺伝子に相当する核酸配列の全部または一部を解析することにより、配列番号1または配列番号4に相当する遺伝子の少なくとも1つの突然変異の存在及び/又は発現の有害な変化が決定されることを特徴とする、炎症性疾患または癌の診断及び/又は予後の評価の方法が好ましい。

【0092】

この診断及び/又は予後の評価方法は、予防的に(炎症性疾患または癌の素因の研究)、あるいは患者の臨床状態を確立及び/又は確認するのに役立たせるために用いることができる。

40

【0093】

好ましくは、炎症性疾患は消化管の炎症性疾患であり、そして癌は消化管(小腸および大腸)の癌である。

本発明の教示により、消化管の炎症性疾患と連鎖不平衡を示し、従ってこの疾患と関連する突然変異を決定することが可能となる。

【0094】

解析は、遺伝子の全部または一部を配列決定することにより、あるいは当分野の技術者に

50

既知のその他の方法により行える。PCR法に基づく方法、例えば点変異を検出可能にするPCR-SSCPが特に使用できる。

【0095】

解析はまた、配列番号1、3、4または6の配列のものに対応する、本発明のプローブをDNAチップに連結し、これらのマイクロプレート上でのハイブリダイゼーションにより行ってもよい。本発明の配列を含むDNAチップもまた本発明の主題の1つである。

【0096】

同様に、本発明のアミノ酸配列を含むタンパク質チップもまた本発明の主題である。かかるタンパク質チップは本発明のポリペプチドと、その他のタンパク質または化合物との間の相互作用を研究することを可能にし、従って、本発明のポリペプチドと相互作用する化合物をスクリーニングするのに有用である。本発明のタンパク質チップはまた、患者の血清中の本発明のポリペプチドに対する抗体の存在を検出するのににも使用できる。本発明の抗体を含むタンパク質チップも使用できる。

10

【0097】

当分野の技術者はまた、遺伝子の発現の有害な変化を研究するための方法を、例えばmRNA（特にノーザンブロットングまたはRT-PCR実験により）を本発明のプローブを用いて検討すること、あるいは特に本発明の抗体を用いたウェスタンブロットングにより、発現されたタンパク質を検討することにより行うことができる。

【0098】

試験される遺伝子は、好ましくは配列番号1の配列をもつ遺伝子であり、罹病性の予測を意図する炎症性疾患は消化管の疾患、特にクローン病または潰瘍性大腸炎である。癌の検出を意図する場合は、大腸癌が好ましい。

20

【0099】

本発明はまた、以下の工程を含む、検出可能な表現型に関連するIBD1遺伝子の対立遺伝子を得る方法にも関する。

- a) 該検出可能な表現型を発現している個体から核酸試料を得る、
- b) 該核酸試料を、IBD1タンパク質をコードする核酸を特異的に検出しうる試薬に接触させる、
- c) IBD1タンパク質をコードする該核酸を単離する。

【0100】

かかる方法の次に、IBD1タンパク質をコードする核酸の全部または一部を配列決定する工程を行ってもよく、これは炎症性疾患または癌への罹病性を予測することを可能にする。

30

【0101】

IBD1タンパク質をコードする核酸を特異的に検出しうる試薬は、有利には、本発明のオリゴヌクレオチドプローブであり、これは改変されてもされていなくてもよいDNA、RNAまたはPNAからなっていてよい。この改変は、放射性または蛍光標識を含むか、あるいは塩基間の結合における改変によるものでもよい（例えば、チオリン酸エステルまたはホスホン酸メチル）。当分野の技術者は特定のDNA配列を単離するためのプロトコルは承知している。上記方法の工程b)は、上記増幅工程であってもよい。

【0102】

本発明はまた、本発明のプローブを生物学的試料と接触させる工程、および、前記ポリヌクレオチドと生物学的試料の核酸との間に形成された雑種を検出及び/又は解析する工程を含む、本発明の核酸を検出及び/又は解析する方法にも関する。

40

【0103】

当分野の技術者はかかる方法を行うことができ、特に以下を含む試薬キットを使用できる。

- a) プローブとして使用される、本発明のポリヌクレオチド、
- b) 該プローブと生物学的試料の核酸との間のハイブリダイゼーション反応を行うのに必要な試薬、
- c) 該プローブと生物学的試料の核酸との間に形成される雑種を検出及び/又は分析するの

50

に必要な試薬。

【0104】

これもまた本発明の主題である。

上記キットは得られる結果の質を確実にするために、陽性または陰性対照を含んでいてもよい。

【0105】

しかし、本発明の核酸を検出及び/又は解析するためには、当分野の技術者は本発明の配列から選択されるプライマーを用いた増幅工程を行ってもよい。

最後に、本発明は、本発明の核酸、ポリペプチド、ベクター、細胞もしくは抗体から選択される化合物、または本発明のスクリーニング方法を用いて得られる、医薬品としての化合物、特に、配列番号1または配列番号4に相当する遺伝子の少なくとも1つの突然変異の存在と関連する炎症性及び/もしくは免疫性疾患または癌、特に消化管の炎症性疾患、殊にクローン病または潰瘍性大腸炎、の予防及び/又は治療のための医薬品としての化合物にも関する。

10

【0106】

以下の実施例は本発明の利点をより明瞭に理解することを可能にするものであり、本発明の範囲を限定するものでないことは当然である。

【0107】

【実施例】

【0108】

【実施例1】

IBD1の微細位置決定

IBD1遺伝子の同定への最初の工程は、D16S409 とD16S419 の間に位置するマーカーD16S411 (Hugot et al., 1996 および図1) をまず中心にして、対象の遺伝子領域のサイズを狭めることである。一群の接近したマーカー(高解像度遺伝子地図)を、遺伝子領域をより明確に特定するために使用し、これにより遺伝的連鎖解析が完全になり、病気に関する遺伝的連鎖不平衡を検索できる。

20

【0109】

検討はCDに罹患した少なくとも2人の親族を含む78家族 - これは119の罹患した組に対応する - に関するものであった。UCに罹患した患者を含む家族はこの検討から除いた。

30

【0110】

26のマイクロサテライト型の遺伝的多型マーカーを検討した。これらのマーカーは一緒になって、対象遺伝子領域でマーカー間の平均距離が1cM 程度の高解像度の地図を作成した。検討したマーカーの特性を表1に示す。

【0111】

【表1】

I B D I の微細位置決定に用いた
マイクロサテライト型の多型マーカー

多型マーカーの名称	累積距離 (cM)	P C Rプライマー
D16S3120 (AFM326vc5)	0	配列番号 7 配列番号 8
D16S298 (AFMa189wg5)	2.9	配列番号 9 配列番号 10
D16S299	3.4	配列番号 11 配列番号 12
SPN	3.9	配列番号 13 配列番号 14
D16S383	4.3	配列番号 15 配列番号 16
D16S753 (GGAA3G05)	4.9	配列番号 17 配列番号 18
D16S3044 (AFMa222za9)	5.8	配列番号 19 配列番号 20
D16S409 (AFM161xa1)	5.8	配列番号 21 配列番号 22
D16S3105 (AFMb341zc5)	6.1	配列番号 23 配列番号 24
D16S261 (MFD24)	6.8	配列番号 25 配列番号 26
D16S540 (GATA7B02)	6.9	配列番号 27 配列番号 28
D16S3080 (AFMb068zb9)	7	配列番号 29 配列番号 30
D16S517 (AFMa132we9)	7	配列番号 31 配列番号 32
D16S411 (AFM186xa3)	8	配列番号 33 配列番号 34
D16S3035 (AFMa189wg5)	10.4	配列番号 35 配列番号 36
D16S3136 (AFMa061xe5)	10.4	配列番号 37 配列番号 38
D16S541 (GATA7E02)	11.4	配列番号 39 配列番号 40
D16S3117 (AFM288wb1)	11.5	配列番号 41 配列番号 42
D16S416 (AFM210yg3)	12.4	配列番号 43 配列番号 44
D16S770 (GGAA20G02)	13.2	配列番号 45 配列番号 46
D16S2623 (GATA81B12)	15	配列番号 47 配列番号 48
D16S390	16.5	配列番号 49 配列番号 50
D16S419 (AFM225zf2)	20.4	配列番号 51 配列番号 52
D16S771 (GGAA23C09)	21.8	配列番号 53 配列番号 54
D16S408 (AFM137xf8)	25.6	配列番号 55 配列番号 56
D16S508 (AFM304xf1)	38.4	配列番号 57 配列番号 58

10

20

30

40

【 0 1 1 2 】

各マーカーは国際命名法と、その大部分は発見した研究所により提案された名称により列

50

拳されている。マーカーは染色体上の順序によっている (16p から16q)。マーカー間の遺伝的距離 (Crimapプログラムを用いて実験データから算出した、Kosambi センチモルガン単位) を第 2 欄に示す。1 番目の多型マーカーを参照点としてランダムにとる。ポリメラーゼ連鎖反応 (PCR) に用いたオリゴヌクレオチドを第 3 欄に示す。

【 0 1 1 3 】

これらのマイクロサテライトマーカーの遺伝子型判定 (genotyping) は蛍光プライマーを用いた自動シーケンサー方法に基づいて行った。簡単に説明すると、増幅後、蛍光ポリメラーゼ連鎖反応 (PCR) 産物を、製造者 (Perkin Elmer) の推奨に従い自動シーケンサーのポリアクリルアミドゲル上に載せた。各個人の対立遺伝子の大きさをGenescan^R および Genotyper^R ソフトウェアを用いて推定した。次いで、データを家系の表現型および遺伝的データを含む統合コンピューターに保存した。次いでそれらを遺伝的連鎖解析に用いた。

10

【 0 1 1 4 】

遺伝子型判定操作の間いくつかの品質管理を行った。

- 遺伝子型判定用データの独立した 2 回の読み、
- 各電気泳動での移行の内部対照として標準 DNA の使用、
- 観察された各対立遺伝子の大きさの範囲の調節、
- メンデル遺伝誤差の検索、
- マーカー間の遺伝的距離の算出 (CRIMAPプログラム) およびこの距離と文献記載のデータとの比較、
- 近いマーカー間の組換えが見られたマーカーの追加の型判定。

20

【 0 1 1 5 】

遺伝子型判定用データを複数点での非パラメトリック遺伝的連鎖法 (GENEHUNTERプログラム、バージョン1.3) により解析した。マーカーシステムの有益さは検討した領域について 80% より大きかった。マーカー-D16S541、D16S3117、D16S770 およびD16S416 では試験最大値 (NPL = 3.33 ; P = 0.0004) が得られた (図 2)。

【 0 1 1 6 】

これらの26の多型マーカーの型判定用データはまた、伝達不平衡 (transmission disequilibrium) を検索するためにも解析された。CDに罹患した 1 人または 2 人以上の患者のいる 108 家族および76の家族 2 つのグループを検討した。伝達不平衡の統計的試験はSpielman et al, (1993) により報告されている。本検討では、1 家族につき 1 人の患者を考慮に入れ、p 値は、検討した各マーカーで試験した対立遺伝子の数により補正した。

30

【 0 1 1 7 】

伝達不平衡はマーカー-D16S3136の対立遺伝子 4 および 5 (それぞれ、サイズは205 および 207 塩基対) で見られた (それぞれ p = 0.05 および p = 0.01)。

マーカー-D16S3136とCDとの間の関連を示唆するこれらの結果から、D16S3136を中心とする遺伝子領域の物理的地図を作成し、そしてこの多型部位を含む大きいゲノム DNAセグメント (BAC) の配列を確立することができた。次いで、D16S3136の領域でより多数の多型マーカーを同定および解析することができ、またこの領域に存在する転写配列の決定および検討も可能となった。

40

【 0 1 1 8 】

【 実施例 2 】

IBD1領域の物理的マッピング

マーカー-D16S3136、D16S3117、D16S770 およびD16S416 を中心としたゲノム DNA断片のコンティグ (contig) をJean Dausset財団/CEPH のヒトゲノム DNAライブラリーから作製した。染色体 DNAセグメントを、微細遺伝子マッピングに用いたある種の多型マーカー (D16S411、D16S416、D16S541、D16S770、D16S2623、D16S3035、D16S3117およびD16S3136) に基づいて同定した。マーカー配列を含むクローンを探すために、各マーカーについて、細菌人工染色体 (BAC) ライブラリーを PCR法でスクリーニングした。試験される配列が BACクローンに存在するか否かにより、Segmapソフトウェア、バージョン3.35を用いてク

50

ローンを互いの間で統合することができた。

【0119】

当分野の技術者に既知の方法 (Rouquier et al., 1994; Kim et al., 1996; Asakawa et al., 1997) により、対象の遺伝子領域をカバーする連続的統合 (コンティグ) を BAC について確立することができた。これを行うために、同定した BAC の末端を配列決定し、次いでこれらの新しい配列データを使用して、繰り返し BAC ライブラリーをスクリーニングした。次いで、各スクリーニングにおいて、重複したクローンの連続体が得られるまで、1 つずつ BAC コンティグが進んだ。コンティグに寄与する各 BAC の大きさは、パルスフィールドアガロースゲル上の移行プロフィールから推定された。

【0120】

このようにして、コンティグの各点で平均 5.5 BAC の重複を有する、101 BAC を含み、2.5 Mb より長い全距離にわたって伸びる BAC コンティグを作製した。BAC の平均サイズは 136 kb である。

【0121】

【実施例 3】

BAC hb87b10 の配列決定

大きさ 163761bp の、多型マーカー D16S3136 を含むこのコンティグの BAC (hb87b10 と称する) を「ショットガン」法により配列決定した。簡単に説明すると、BAC DNA を超音波処理により断片化した。こうして得られた DNA 断片をアガロースゲル電気泳動にかけ、1.5 kb より大きいサイズの断片を溶出して分析した。次いで、これらの断片を m13 ファージ中にクローン化し、それ自身を電気穿孔により、コンピテントにされた細菌へ導入した。培養後、クローンの DNA を回収し、自動シーケンサーで、m13 ベクターの蛍光プライマーを用いた自動配列決定方法により配列決定した。

【0122】

平均サイズ 600 bp の 1526 の異なる配列が作製され、これを Polyphredphrap^R ソフトウェアを用いて互いに統合して、全 BAC をカバーする配列コンティグが得られた。こうして得られた配列は、平均 5.5 ゲノム当量の重複を有していた。m13 クローンライブラリー中に現れない稀少 ($n = 5$) 配列ギャップを、これらのギャップの一方の側で、特異的 PCR プライマーを作製し、健全な個体のゲノム DNA 由来の PCR 産物を解析することにより充填した。

【0123】

公開の遺伝子データベース (Genbank) で入手しうる配列を有する相同の配列を探した。この 163kb の領域では既知の遺伝子は同定されなかった。いくつかの EST を位置決定し、これは未知の遺伝子がこの配列中に含まれることを示唆した。公開の遺伝子データベース (Genbank, GDB, Unigene, dbEST) 由来のこれらの EST は次の参照番号を有していた: A1167910, A1011720, Rn24957, Mm30219, hs132289, AA236306, hs87296, AA055131, hs151708, AA417809, AA417810, hs61309, hs116424, HUMGS01037, AA835524, hs105242, SHGC17274, hs146128, hs122983, hs87280 および hs135201。GRAIL コンピュータープログラムを用いた推定エキソンの検索により、いくつかのありうるエキソン、ポリアデニル化部位およびプロモーター配列の同定を行うことができた。

【0124】

【実施例 4】

伝達不平衡の検討

12 の二対立遺伝子性 (biallelic) 多型マーカー (SNP) を約 250kb の範囲の、BAC hb87b10 を中心とする領域において同定した。これらの多型は 10 人程度の CD に罹患した無関係の患者の配列を解析することより得られた。配列決定は、BAC に位置する既知の EST またはその領域で主に行った。GRAIL コンピュータープログラムにより予測される推定エキソンも解析した。このように同定された多型マーカーの特性を表 2 に示す。

【0125】

【表 2】

I B D I の領域で検討した
二対立遺伝子性多型マーカーの特性

I	II	III	IV	V	VI
1	KIAA0849ex9	AS-PCR		配列番号88～90	116
2	hb27G11F	PCR-RFLP	BsrI	配列番号86、87	185 116 69
3	Ctg22Ex1	PCR-RFLP	RsaI	配列番号84、85	381 313 69
4	SNP1	AS-PCR		配列番号81～83	410
5	ctg2931-3ac/ola	LO		配列番号78～80	51 49
6	ctg2931-5ag/ola	LO		配列番号75～77	44 42
7	SNP3-2931	AS-PCR		配列番号72～74	245
8	Ctg25Ex1	PCR-RFLP	BstEII	配列番号70、71	207 122 85
9	CTG35ExA	AS-PCR		配列番号67～69	333
10	ctg35ExC	AS-PCR		配列番号64～66	198
11	D16S3136			配列番号37、38	
12	hb133D1f	PCR-RFLP	TaqI	配列番号62、63	369 295 74
13	D16S3035			配列番号35、36	
14	ADCY7int7	AS-PCR		配列番号59～61	140

【 0 1 2 6 】

AS-PCR: 対立遺伝子特異的PCR ; LO: オリゴヌクレオチドの連結

この検討で新たに報告された12の二対立遺伝子性多型マーカーをこの表に挙げる。そのそれぞれについて以下のことが示される。

- 遺伝子座 (欄 I)
- 名称 (欄 II)
- 使用した遺伝子型判定法 (欄 III)
- 使用可能な制限酵素 (欄 IV)
- ポリメラーゼ連鎖反応または連結に使用されるオリゴヌクレオチドプライマー (欄 V)
- 型判定の間に予測された産物の大きさ (欄 VI)

CDに罹患した1人またはそれ以上の患者を含む199家族を、これらの12の多型マーカーについて、またBAC hb87b10に位置するマーカーD16S3035およびD16S3136についても型判定した。UCに罹患した病人を含む家族は考慮に入れなかった。検討された多型の型判定のための方法は、多型の種類により異なり、以下の方法を用いる。

- 多型が酵素による切断部位上にある場合にはPCR-RFLP法 (増幅、次いでPCR産物の酵素的切断)
- 多型部位に特異的なプライマーを用いる PCR法: 各対立遺伝子に特異的なプライマーを

10

20

30

40

50

用いる2つの対立遺伝子の差別的 (differential) 増幅

- オリゴライゲーション (oligoligation) 試験: 各対立遺伝子に特異的なオリゴヌクレオチドを用いる差別的ライゲーション、次いでポリアクリルアミドゲル電気泳動
次いで、型判定用データを伝達不平衡試験を用いて解析した (GENEHUNTERソフトウェア、バージョン2のTDTコンピュータープログラム)。数人の罹患した親族を含む家族について、1人の患者のみを解析のために考慮した。事実、もし数人の血縁関係のある患者を考慮すると、統計計算においてデータの非独立性の問題を生じ、試験値の膨張を生じ得る。解析に用いた患者は、自動無作為化操作を用いて各家族内でロットで引き抜かれた。この無作為化により、得られる統計試験の値は検討した家族のグループからのただ1つの可能なサンプルを表した。解析をこの1つの可能なサンプルに限定しないように、かつ得られる結果の有効性をより明瞭に理解するために、各試験において約100の無作為サンプルを得て解析した。

10

【0127】

マーカーは別々に検討し、次いでその染色体セグメント上での順序により分類した (K1AA0849ex9(遺伝子座1)、hb27G11F(遺伝子座2)、Ctg22Ex1(遺伝子座3)、SNP1(遺伝子座4)、ctg2931-3ac/ola(遺伝子座5)、ctg2931-5ag/ola(遺伝子座6)、SNP3-2931(遺伝子座7)、Ctg25Ex1(遺伝子座8)、CTG35ExA(遺伝子座9)、ctg35ExC(遺伝子座10)、d16s3136(遺伝子座11)、hb133D1f(遺伝子座12)、D16S3035(遺伝子座13)、ADCY7int7(遺伝子座14)(表2)。従って、2、3および4の連続したマーカーを含むハプロタイプも

20

【0128】

試験した各サンプルにつき、少なくとも10の親染色体に保持される遺伝子型 (またはハプロタイプ) のみが考慮された。従って、平均して250の異なる試験を各サンプルにつき行った。次いで、各有意な閾値に対して陽性であると期待される試験数を推定することおよびこの分布を観察される分布と比較することができた。健常人については、試験の分布は無作為基準で期待されるものと違いはない ($\chi^2 = 2.85$, $ddl = 4$, $p = 0.58$)。一方、患者の場合は陽性試験が過剰であり、これは検討領域中の伝達不平衡を反映している。

【0129】

別々に考慮された各多型マーカーについての、または最も強い伝達不平衡を示すハプロタイプについての伝達不平衡試験の結果は、次のマーカーおよび疾患が連鎖不平衡であることを示した: Ctg22Ex1(遺伝子座3)、SNP1(遺伝子座4)、ctg2931-5ag/ola(遺伝子座6)、SNP3-2931(遺伝子座7)、Ctg25Ex1(遺伝子座8)、ctg35ExC(遺伝子座10)。これらのマーカーは約50kbの範囲にわたる (hb87b10の配列上の74736 ~ 124285位)。

30

【0130】

クローン病と最も強く関連しているハプロタイプもそれ自身この領域にわたる。従って、大部分の無作為サンプルにとって、伝達試験 (transmission test) は、以下のマーカーを結合しているハプロタイプに対して陽性 ($p < 0.01$) であった。

- 遺伝子座5-6、遺伝子6-7、遺伝子座7-8、遺伝子座8-9、遺伝子座9-10、
遺伝子10-11

40

- 遺伝子座5-6-7、遺伝子6-7-8、遺伝子座7-8-9、遺伝子座8-9-10、
遺伝子9-10-11

- 遺伝子座5-6-7-8、遺伝子6-7-8-9、遺伝子座7-8-9-10

罹病率の最もリスクが高いハプロタイプは遺伝子座7~10により規定される。これはハプロタイプ1-2-1-2(表2)である。

【0131】

試験したマーカーは予想通り、互いに連鎖不平衡である。

より最近では、2000年6月に発行された (Martin et al., 2000) Pedigree Disequilibrium (家系不平衡) 試験が、TDTコンピュータープログラムを用いて得られた結果の意味をより明確に理解するために用いられた。この新しい統計学は、実際、ある家族における患

50

者および健常人の両方から得られる情報をすべて利用することを可能にし、各家族に対して統計全体における各親族の重要性と釣り合うことを可能にする。PDT試験に対応し、クローン病に罹患した1人またはそれ以上の親族をもつ235の家族の拡大されたグループから得られたp値は表3に示される。この新しい解析法により、BAC hb87b10の領域は実際クローン病に関連することが確認される。

【0132】

【表3】

クローン病に罹患した235家族の
PDT試験の結果 (NS: 有意性なし)

遺伝子座	PDT試験のp値
K1AA0849ex9	NS
hb27g11f	0.05
ctg22ex1	0.01
SNP1	0.001
ctg2931-3ac/ola	NS
ctg2931-5ag/ola	0.0001
SNP3-2931	0.0001
ctg25ex1	0.0006
ctg35exA	NS
ctg35exC	0.00002
D16S3136	NS
hb133d1f	NS
D16S3035	NS

【0133】

【実施例5】

IBD1遺伝子の同定

BAC hb87b10に存在する公開されたESTグループ (Unigene 参照番号: Hs135201, Hs87280, Hs122983, Hs146128, Hs105242, Hs116424, Hs61309, Hs151708, Hs87296 およびHs132289) をより完全な相補的DNA (cDNA)の配列の検索のために検討した。IBD1proxについては、公開のライブラリーで入手可能なクローンを配列決定し、配列を互いに統合した。IBD1については、末梢血の相補的DNAライブラリー (Stratageneヒト血cDNA zapexpress ref 938202)を製造者の提案する方法によって既知のESTから作製したPCR産物でスクリーニングした。次に、こうして同定されたcDNAを、示されたcDNAが得られるまで、cDNAライブラリーなどをさらにスクリーニングするために使用した。

【0134】

EST hs135201 (UniGene)により、利用しうる遺伝子データベース (Genbank)にないcDNAの同定が可能となった。従って、これは新規なヒト遺伝子に対応する。cDNAおよびゲノムDNAの配列の比較により、この遺伝子は11のエキソンと10のイントロンからなることが分かった。同定したcDNAの5'に位置する、追加のエキソンがGraiIプログラムでの配列解析により予測される。これらのエキソンはCARD4/NOD1遺伝子の第1のエキソンと非常に相同性である。同定されたエキソンすべてと推定の追加のエキソンを考慮すると、この新規な遺

伝子はCARD4/NOD1と非常に近いゲノム構造を有するようである。さらに、転写開始部位は第1の推定エキソンの上流に現れる。これらのすべての理由により、推定エキソンはこの新規遺伝子に寄与していると考えられた。従って、添付書類（配列番号1）に示したcDNAは、同定した配列全部に加え、コンピューターモデリングにより予測された配列を含み、この相補的DNAは予測されたコーディング配列の最初のATGコドンでランダムに開始する。従って、これに基づき、この遺伝子は12のエキソンと11のイントロンを含むであろう。この遺伝子のイントロン-エキソン構造は配列番号3に報告されている。

【0135】

ヌクレオチド配列から推定されるタンパク質配列は1041のアミノ酸を含む(配列番号2)。この配列は生物学データベース(Genpept, pir, swissprot)のいずれにも見出されない。

10

【0136】

これまで、上記推定エキソンを確認することはできなかった。従って、IBD1遺伝子は事実上、11のエキソンと10のイントロンのみを含み、1013のアミノ酸(即ち、最初に決定されたよりも28アミノ酸少ない)のタンパク質をコードする。

【0137】

推定タンパク質配列の検討により、この遺伝子は3つの異なる機能性ドメインを含むことが示される(図3)。

- アポトーシスおよびNFkappa B経路の活性化を調節するタンパク質間の相互作用に関与することが知られているCARDドメイン(Caspase Recruitment Domain、カスパーゼ加入ドメイン)。このCARDドメインにより、この新規なタンパク質をCARDタンパク質ファミリーに分類することが可能となる。このファミリーの最も知られたメンバーはCED4, APAF1 およびRICKである。

20

- ATP-認識部位およびマグネシウム結合部位を含むNBDドメイン(Nucleotide-Binding Domain、ヌクレオチド結合ドメイン)。従って、このタンパク質は多分キナーゼ活性を有するはずである。

- その他の既報のタンパク質ドメインとの類似性により、タンパク質間の相互作用に關与すると推測されるLRRドメイン(Leucine-Rich Domain、ロイシンに富むドメイン)。

【0138】

さらに、このタンパク質のLRRドメインにより、このタンパク質が細胞内シグナリングに關与し、植物にも動物にも存在するタンパク質のファミリーに入ることが可能となった。

30

【0139】

この新規な遺伝子と、これまで同定された公開のデータベースで利用可能な遺伝子とを比較すると、この遺伝子がCARD4/NOD1(Bertin et al., 1999; Inohara et al., 1999)と非常に相同性であることが分かる。この相同性は相補的DNAの配列、遺伝子のイントロン-エキソン構造およびタンパク質配列に関する。2つの相補的DNAの配列同一性は58%である。類似性はイントロン-エキソン構造のレベルでもみられる。タンパク質レベルでの配列の相同性は40%のオーダーである。

【0140】

この新規遺伝子とCARD4/NOD1との間の類似性は、IBD1タンパク質は、CARD4/NOD1と同じく、アポトーシスおよびNF-kappa B(Bertin et al., 1999; Inohara et al., 1999)活性化の調節に關与していることを示唆する。細胞のアポトーシスおよびNF-kappa Bの活性化の調節は免疫反応において必須の細胞内シグナリング経路である。具体的には、これらのシグナル伝達経路は、細胞-細胞相互作用および炎症の各種メディエーター(サイトカイン)に対する細胞応答に關与するTNF(Tumor Necrosis Factor、腫瘍壊死因子)受容体ファミリーのタンパク質のエフェクター経路である。従って、この新規遺伝子は一般的に炎症性反応において重要である可能性がある。

40

【0141】

いくつかの立証が、クローン病におけるNF-kbの細菌誘導性の脱制御(deregulation)を支持している。まず、マウスにおけるIBDの自然発生的罹病性は、LRRドメインを介して

50

LPSに結合しており (Politorak et al., 1998 および Sundberg et al., 1994)、NF- κ B ファミリーのアクチベーターの一員であることが知られている分子、Tlr4における突然変異と関連していた。第2に、抗生物質での治療はCDに罹患した患者において一時的改善を生じるが、これは腸内細菌がクローン病の病因としての役割を果たしているという仮説 (Mc Kay, 1999)を支持する。第3に、NF- κ B は炎症性腸疾患において中枢的役割を果たし、クローン病では基底膜単核細胞において活性化される (Schreiber et al., 1998)。第4に、クローン病の治療はNF- κ B 阻害剤として知られているスルファサラジンおよびグルココルチコイドの使用に基づく (Auphan et al., 1995 および Wahl et al., 1998)。

【0142】

さらにより最近、IBD1候補遺伝子がCED4/APAF1スーパーファミリーの一員であるNOD2に非常に似たタンパク質コードすることが示された (Ogura et al., 2000)。IBD1およびNOD2のヌクレオチドおよびタンパク質配列は実際、2つの既報の配列の開始の小部分で異なるのみである。さらに、Nod2およびIBD1の組織発現は重なり得る。従って、これらの2つの遺伝子 (タンパク質) は同一であると考えられる。Nod2のLRRドメインは細菌のリポ多糖類 (LPS)に対する結合活性を有し (Inohara et al., 2000)、そしてその欠失はNF κ B経路を刺激することが実証された。この結果は本発明のデータを確認するものである。

10

【0143】

次いで、IBD1の組織発現をノーザンブロットングにより検討した。ほとんどのヒト組織において4.5 kbの転写物がみられた。転写物の大きさはcDNAにより予測された大きさと一致する。この4.5 kbの添加物は小腸および大腸では極めて少ないようである。一方、これは白血球細胞では非常に強く発現される。これは、クローン病が循環している免疫細胞と関連する疾患である可能性があることを示唆する、移植での臨床データと合致する。事実、腸移植はクローン病における移植体での再発を防止しないが、一方骨髄移植はこの病気の進行に対しよい影響を与える。

20

【0144】

また、データは選択的スプライシングに注意を向けさせる。これは炎症性疾患の進行においてある役割を果たしうる突然変異の発生の可能性において、重要な要素であると判明するかもしれない。

【0145】

IBD1遺伝子のプロモーターは現在正確には同定されていない。しかし、非常に多くの遺伝子との類似性により、このプロモーターは少なくとも一部、この遺伝子のすぐ上流、その5'部分に存在すると考えるのが合理的である。この遺伝子領域は、EST (HUMGS01037, AA835524, hs.105242, SHGC17274, hs.146128, hs.122983, hs.87280) の存在により証明されるように、転写される配列を含む。これらの配列を含むATCCクローンを実験室で配列決定および解析し、選択的スプライシングを伴うかもしれないエキソンおよびイントロンの構成を実証できた。これらのデータは別の遺伝子 (IBD1の近傍にあることからIBD1proxと命名された) の存在を示唆する。IBD1proxの相補的DNAの部分的配列を記載する (配列番号4)。これは配列番号6上のイントロン - エキソン構造である。

30

【0146】

IBD1proxに対応するcDNAの翻訳により、ホメオボックスを含むタンパク質が得られる。しかし、この遺伝子のいくつかのcDNAの解析では選択的スプライシングの存在が示唆される。可能な選択的スプライシングの1つによれば、IBD1proxはアノニマス EST HUMGS01037に対応し、そのRNAは未分化系列よりも分化した白血球系列においてより強く発現される。

40

【0147】

従って、この遺伝子が炎症および細胞分化においてある役割を有する可能性がある。よって、それ自身もまたIBDの罹病性に対するよい候補であると考えられる。CDと、IBD1proxのコーディング配列上に位置する多型ctg35ExCとの間の関連は、この多型がタンパク質レベルで何らの配列変化も生じないにもかかわらず、この仮説を支持する。

【0148】

50

最後に、より最近、クローン病に罹患し、IBD1遺伝子に何の突然変異も含まない家族での遺伝的連鎖の存在それ自身も、IBD1proxがIBD1に加え、この病気の遺伝的素因においてある役割を有していることを示唆している。

【0149】

IBD1とIBD1proxとの間の機能的関係は現在確立されていない。しかし、この2つの遺伝子がかなり近接していることはそれらの間の相互作用を反映するかもしれない。この場合、これらの遺伝子の「頭-尾」の位置関係は、それらが共通または互いに依存する調節方法を有するかもしれないことを示唆する。

【0150】

【実施例6】

炎症性疾患におけるIBD1遺伝子突然変異の同定

炎症性疾患におけるIBD1の役割を確認するために、この遺伝子のコーディング配列およびイントロン-エキソン結合部を、無関係の70人、即ち、CDに罹患した50人の患者、UCに罹患した10人の患者、プラウ症候群に罹患した1人の患者および健常人9人において、エキソン2からエキソン12まで配列決定した。検討した患者はほとんどが家族性型のこの病気であり、伝達不平衡研究により明らかになった罹病性ハプロタイプの保持者であることが多い。

【0151】

こうして、この血縁関係のない70人のグループにおいて24の変異配列を同定した(表3)

。

報告される突然変異の命名は、1041のアミノ酸を含むタンパク質の最初の配列に関する。より最近提案された命名法は、最初の配列から28のアミノ酸を除くことによって容易に推定でき、従って、1013アミノ酸を含むタンパク質に対応する(実施例5参照)。

【0152】

【表4】

10

20

I B D I 遺伝子に見られた突然変異

エキソン	変異 ヌクレオチド	変異 タンパク質	クローン 病	潰瘍性 大腸炎	健全対照
1	試験せず				
2	G417A	サイレント			
2	C537G	サイレント			
3	無し				
4	T805C	S269P	48/100	6/20	3/18
4	A869G	N290S	0	0	1/18
4	C905T	A302V	1/100	0	0
4	C1283T	P428L	1/100	0	0
4	C1284A	サイレント			
4	C1287T	サイレント			
4	T1380C	サイレント			
4	T1764G	サイレント			
4	G1837A	A613T	1/100	0	0
4	C2107T	R703W	10/10	1/20	1/18
4	C2110T	R704C	4/10	1/20	0
5	G2365A	R792Q	1/100	0	0
5	G2370A	V794M	0	1/20	0
5	G2530A	E844K	1/10	0	0
6	A2558G	N853S	1/100	0	0
6	A2590G	M864V	1/100	0	0
7	無し				
8	G2725C	G909R	7/100	0	0
8	C2756A	A919D	1/100	0	0
9	G2866A	V956I	2/100	1/20	3/18
10	C2928T	サイレント			
11	3022insC	終結	20/100	0	0
12	無し				

【 0 1 5 3 】

各エキソンに見られるサイレント突然変異以外の突然変異を記載する。それらはペプチド鎖における変異により示されている。検討した各突然変異および各表現型について、突然変異が見られた回数を試験した染色体の数に関連させて示す。

10

20

30

40

50

【 0 1 5 4 】

機能的な変異配列はエキソン 1 ~ 3 (タンパク質のCARDドメインに対応) においては同定されなかった。エキソン 7 および 12 もまた配列の変異を示さなかった。ある種の変異は、既に同定され、伝達不平衡検討で型判定された多型に相当する、すなわち、

- Snp3-2931: ヌクレオチド変異 T805C、タンパク質変異 S269P
- ctg2931-5ag/ola: ヌクレオチド変異 T1380C (サイレント)
- ctg2931-5ac/ola: ヌクレオチド変異 T1746G (サイレント)
- SNP1: ヌクレオチド変異 C2107T、タンパク質変異 R703W

いくつかの変異配列はサイレントであり (G417A, C537G, C1284A, C1287T, T1380C, T1764G および C2928T)、タンパク質配列の変化は何らもたらさない。それらはここではこれ以上検討しなかった。

10

【 0 1 5 5 】

16の非サイレント配列変異については、タンパク質配列変異が43/50 CDにおいて見られ、健常者対照では5/9 であり、また6/10 UC であった。1 または 2 以上の配列変異の存在は、CD表現型と関連するようであった。いくつかの配列変異はCDに罹患した同じ個人においてしばしば存在し、これはCDに対する遺伝子の時による劣性効果を示唆する。これに反し、UCに罹患した患者または健常人の対照では複合ヘテロ接合体またはホモ接合体は見られなかった。

【 0 1 5 6 】

いくつかの非サイレント変異がUCまたはCDに罹患した患者および健常人の両方に存在した。それらはエキソン 2、4 および 9 に位置する変異S269P, N290S, R703W およびV956I であった。従って、これらの配列変異に対して可能な機能的役割を選択する前にさらなる情報が必要なようである。

20

【 0 1 5 7 】

V956I は保存的 (conservative) 配列変異 (脂肪族アミノ酸) である。

配列変異S269P はヌクレオチド結合ドメインの最初におけるアミノ酸の種類の変化 (イムノ酸 (immuno acid)へヒドロキシル化) に相当する。この配列変異およびCDは伝達不平衡にある。事実それは多型Snp3である (上記参照)。

【 0 1 5 8 】

R703W はアミノ酸の種類の変更をもたらす (塩基性の代わりに芳香族)。この変更は、IB D1とCARD4/NOD1の間の保存領域である、NBDおよび LRRドメインの間の中間領域に生じる。これは従って、この多型には機能的役割が推測される。この配列変異 (多型部位Snp1に対応) はCDに罹患した患者にランダムな場合よりも多い頻度で伝達され (上記参照)、このことはこの多型がCDと関連していることを確かにする。健常人にこの突然変異体が存在することは、慢性炎症性腸疾患などの複合遺伝病について予測されるように突然変異の不完全な浸透を反映する。

30

【 0 1 5 9 】

R703W のすぐ隣に位置する変異R704C はCDおよびUCの両方において同定できた。それはまた、それ自身、同じタンパク質領域でのタンパク質の非保存的 (nonconservative) 変異 (塩基性アミノ酸の代わりに硫黄含有アミノ酸) に対応し、このことはR703W に対すると同じ程度に重要なR704C に対する機能的効果を示唆する。

40

【 0 1 6 0 】

その他の配列変異はCD、UCまたはブラウ症候群に特異的である

逆に、いくつかの配列変異は稀であり、1 人または数人の患者に存在する (A613T, R704C, E844K, N853S, M864V, A919D)。それらは常に、ロンシンに富む領域の、これらの領域内で重要な位置にタンパク質の非保存的変更を生じる変異である。これらの各種要素は、これらの変異が機能的役割を有することを示唆する。

【 0 1 6 1 】

配列変異 (G909R および L1008P *) は極めて多くのクローン病で見出され (それぞれ7/50および16/50)、一方それらは対照またはUCに罹患した人において検出されない。

50

【0162】

コドン1008のグアノシンの欠失/挿入は、最後の LRRのヘリックスの3番目のロイシンをプロリン（終止コドンが続く）に変化させる結果となる（L1008P^{*}）。従って、この配列変異はタンパク質の重要な改変を生じさせる：タンパク質の大きさの減少（短縮 LRRドメインを有するタンパク質）および高度に保存されたアミノ酸（ロイシン）の変化。この配列変化は、この突然変異を有する16家族での伝達不平衡の検討（ $p = 0.008$ ）により立証されたように、CDと関連している。

【0163】

突然変異G909Rは6番目のLRRモチーフの最後のアミノ酸に生じる。それは脂肪族アミノ酸を塩基性アミノ酸で置換するものである。この変異は、ロイシンに富むモチーフ（IBD1およびNOD1/CARD4の両方について）の末端位置におけるアミノ酸の通常中性または極性の性質、およびIBD1およびNOD1/CARD4タンパク質のアミノ酸の保存された性質を考慮すると重要である可能性がある。

10

【0164】

ブラウ症候群では、検討した家族の患者（ $n = 2$ ）は、エキソン4に位置し、タンパク質のNBDドメインに対応する特定の配列変異（L470F）を保持していた。この系列では、この配列変異はブラウ症候群に特異的であった。

【0165】

UCでは、健常人に見出されなかったいくつかの配列変異も同定された。突然変異を有する病人の割合は、IBD1とUC間のあまり強く確立されていない連鎖と、後者の疾患のおそらく低い遺伝性を考慮して予測されるように、CDに関するよりも小さい。配列変異はCDに対しておよびUCに対して共通である（R703W, R704C）。逆に、その他はUCに対して特異的であるようであった（V794M）。この観察により、CDとUCは、少なくとも一部同じ遺伝的素因を共有する疾患であることが確認できる。これはIBDに対する疾病分類学の基礎を築く。

20

【0166】

このように、IBD1遺伝子の配列変異の検討により、かなり可能性のある機能的効果（例えば、短縮タンパク質）を有し、クローン病、UCおよびブラウ症候群と関連するいくつかの変異を同定することが可能となった。

【0167】

この遺伝子のプロモーターは現在決定されていない。しかし、多分、この遺伝子上流の5'領域に位置するであろう。この仮説によれば、この領域に見出された配列変異は機能的効果を有するかもしれない。これは、CDとある種の多型遺伝子座（ctg35ExCまたはCtg25Ex1など）との間の非常に強力な関連を説明できる。

30

【0168】

従って、本発明はヒトにおいてCARDドメインを含有する遺伝子ファミリーにおける突然変異を最初に記載するものである。各種の炎症性疾患におけるこれらの突然変異の頻度は、IBD1遺伝子が正常および病的な炎症過程において必須の役割を果たすことを示す。本発明は、正常および病的な炎症過程の生理病理学分野での理解および研究の新規な方法を提供する。その結果、IBD1により制御されるエフェクター経路を調節し、炎症性疾患の治療および一般的炎症過程の調節に有用な新規な薬剤分子の開発を考えることが可能になる。

40

【0169】

【実施例7】

クローン病の罹病性の生物学的診断のための基礎

より最近、クローン病に罹患した457人の独立した患者、潰瘍性大腸炎に罹患した159人の独立した患者および対照となる健常な137人を、突然変異の検索において検討した。この検討により、これまで報告された突然変異を確認し、そして図4に報告される追加の突然変異を同定することができた。次いで、主要な突然変異を、クローン病に罹患した235家族において遺伝子型の判定を行った。このより最近の研究は、参照として、より短いタンパク質配列（1013アミノ酸、実施例5参照）を用いて報告されるが、従来の突然変異の命名は、アミノ酸の位置を示す数に28足すことによりこれから容易に推定される。

50

【 0 1 7 0 】

5つの最もよくみられる突然変異の中で、保存的突然変異V981I (以前はV956I)は炎症性腸疾患の1つまたはその他とはあまり関連しておらず、従って、この疾患に重要な役割を果たしているようにはみえない。

【 0 1 7 1 】

変異S241P(以前はS269P)はその他の主要な突然変異と連鎖不平衡にあり、それ自身は、炎症性腸疾患の罹病性において重要な役割を果たしているようではない (データは示さず)。

【 0 1 7 2 】

逆に、その他の3つの突然変異、R675W(以前はR703W)、G881R(以前はG909R)および980fs(以前はL1008P*)はクローン病と顕著に関連するが、潰瘍性大腸炎とは関連しない(下記参照)。LRR中の、またはその中程度近傍での、3つのよくみられる突然変異の位置は、多分、変異したタンパク質によるNFkBの負の調節における欠損を介する、このタンパク質ドメインが関与する作用機構を非常に有利に説明する。その他の突然変異はもっと稀である(図4)。これらの累積突然変異はクローン病に罹患した人では17%存在するのに対し、健常人および潰瘍性大腸炎に罹患した人ではそれぞれ4%および5%である。非常に多数の稀少突然変異もLRRに存在する。

10

【 0 1 7 3 】

クローン病において最もよくある3つの多型の家族内検討により、この3つはすべてこの疾患に関与していることが示される(表5)。予想通り、非常に有害であると考えられる突然変異については、最も強く関連している多型は短縮性の突然変異である。これら3つの突然変異の1つより多くを有する染色体を235家族において同定することができなかったため、これらの3つの多型は独立して、クローン病と関連している。これらの関連の独立性は、IBD1遺伝子はクローン病の遺伝的素因に明らかに関与しているという仮説をかなり支持する。

20

【 0 1 7 4 】

【表5】

クローン病に罹患した235 家族のIBD1の3つのよくみられる多型の検討

30

突然変異	PDT 試験のp値
R675W	0.001
G881R	0.003
980fs	0.000006

【 0 1 7 5 】

患者 - 対照検討により、この関連が確認される(表6)。それらは、クローン病に最もよくある突然変異は潰瘍性大腸炎ではあまりみられないことを示す。

40

【 0 1 7 6 】

【表6】

炎症性腸疾患におけるIBD1の3つの
よくみられる多型の患者－対照研究

突然変異	検討 染色体数	危険対立 遺伝子R675W の頻度	危険対立 遺伝子G881R の頻度	危険対立 遺伝子980fs の頻度	合計 危険対立 遺伝子
健全対照	206	0.04	0.01	0.02	0.07
潰瘍性大腸炎	318	0.03	0.00	0.01	0.05
クローン病	936	0.11	0.06	0.12	0.29

10

【0177】

これらの突然変異の用量 - 効果の検討により、ホモ接合体または複合ヘテロ接合体の状態
で突然変異を有する人は、これらの突然変異をもたないかヘテロ接合体の状態である人よ
りこの疾患を発症する危険が非常に大きいことが示される（表7）。

【0178】

【表7】

分 布	遺 伝 子 型			
	変異なし	単純ヘテロ接合体	ホモ接合体	複合ヘテロ接合体
健全	88	15	0	0
潰瘍性大腸炎	145	13	1	0
クローン病	267	133	28	40
帰属可能な CD危険率： 相対危険率 絶対危険率	1 0.0007	3 0.002	38 0.03	44 0.03

20

30

【0179】

集団全体では、0.001 というクローン病の危険率を参照として採用し、突然変異がHardy-
Weinberg平衡にあることが推定された。

上記検討により、以前の予備的データが確認され、IBD1変異を検討することによる、クロ
ーン病の生物学的診断の詳細な基礎が提供される。事実、この検討は、

1) 混血のカフカス人集団においては0.001 より大きい頻度の突然変異を明かにし、
2) 観察された突然変異の頻度を決定し、クローン病に関連する3つの主要突然変異を決定
することを可能にする。従って、この検討により、疾患性の変異の検索のために遺伝子を
検討するための方策を決定できる：即ち、まず3つの主要な突然変異を型判定し、第2に
最後の7エキソンの突然変異を検索し、第3に、その他の配列変異を検索する。
3) これらの突然変異をその位置および性質を指摘することにより検索するための実用的様
式を明かにする。事実、次いで、当分野の技術者がその個人的専門技術により型判定およ
び配列決定方法を発展させることは容易である。特に、PCR、次いで酵素的切断および電
気泳動、dHPCL、DGGEまたはSSCPによる移行プロフィールの検討、オリゴライゲーション
、微細配列決定などにより、3つの主要な突然変異の遺伝子型を判定する可能性が挙げら
れる。

40

4) この拡大し変化した集団では同じ染色体上に見られない最もよくある突然変異の独立性
を実証する。この情報により、2倍量の遺伝子内変異の保持者として複合ヘテロ接合体（
2つの突然変異を有する）である人を確実に分類できる。

50

5) 大部分の突然変異は潰瘍性大腸炎のリスクに全くあるいは小さな影響を及ぼすだけであることを実証する。この結果から、この2つの疾患の間の差別的診断において臨床医科の助けとなることを予想できる。実際、約10%の患者において炎症性腸疾患は生物学的、放射線医学的および内視鏡的検査にもかかわらず、未分類のままである。

6) 最もよくみられる遺伝子型に対して病気の相対的および絶対的リスクを明かにする。この結果は、リスクのある集団、特に患者の親族において予防的監視および介入のアプローチにおいて有用である可能性のある、予測的診断の基礎を築く。

7) IBD1遺伝子の用量 - 効果の存在を実証し、クローン病の遺伝的素因が一部劣性であることを確認する。従って、遺伝相談および家族内前臨床診断のための基礎を築くことができる。

10

【0180】

最後に、NBDドメインの追加の突然変異がブラウ症候群を有する第2の家族において単離されたことに注意すべきである。2つの異なる家族での2つの事象が稀であることは、この遺伝子がブラウ症候群および一般的肉芽腫性疾患に関与していることを確認するのに十分である。

【0181】

これらのデータはすべて、日常的实施において実施者にとって直接応用でき、有用である診断手段を提供する。

IBD1のプロモーター領域に位置し、その部分的配列が本発明に開示されているIBD1prox遺伝子もまた、免疫系の成熟細胞の差別的発現により示唆されるように、それ自身、細胞アポトーシスおよび炎症性過程の調節において重要な役割を有しているかもしれない。ここで報告された、多型マーカーctg35ExC（この遺伝子の転写される領域に位置する）とクローン病との間の強い関連もこの仮説を非常に強く支持する。

20

【0182】

炎症性腸疾患は、これまで罹病性遺伝子が確実に同定されていなかった複合遺伝病である。本発明は、ポジショナルクローニング（または逆遺伝学）のアプローチを用いて、クローン病の罹病性に対する最初の遺伝子を同定することを可能にした。これは、複合遺伝病に対するかかるアプローチを用いて得られた最初の遺伝子位置決定であり、少なくとも複合遺伝病のある種の患者でその有用性および実施可能性を実証する。

【0183】

本発明はまた、配列番号2および配列番号5から選択されたタンパク質の少なくとも200のアミノ酸の連続断片を含むポリペプチドをコードすることを特徴とする、精製または単離された核酸に関する。

30

【図面の簡単な説明】

【図1】 第16染色体の動原体周囲領域におけるクローン病のための非パラメトリックの遺伝的連鎖試験 (Hugot et al., 1996による)。第16染色体の動原体周囲領域のマーカーの家系による同一性に基づく複数点での連鎖解析。マーカー間の遺伝的距離はCRIMAPプログラムを用いて推定された。ロッドスコア (MAPMAKER/SIBS) を左側の図に示す。2つの仮確率試験を開発し、右側の図に示した。1番目 (Tz) は平均の試験に類似している。2番目 (Tz2) は2つの対立遺伝子を共有する罹患した組の割合の試験に類似する。

40

【図2】 複数点の非パラメトリックの遺伝的連鎖解析。クローン病に罹患した数人の親族をもつ78家族を、第16染色体の動原体周囲領域の26の多型マーカーについて遺伝子型決定した。各マーカーの位置は矢で表した。マーカーの順序およびそれらの間の距離はCrimapソフトウェアを用いた実験データの解析による。曲線の下は発表された最初の研究 (Hugot et al., 1996) に用いられたマーカー SPN、D16S409 およびD16S416 を示す。図の上部にある矢は、遺伝的連鎖試験の最大に位置するマーカーD16S3136、D16S541、D16S3117、D16S416 およびD16S770 に相当する。型判定用データはGenehunterソフトウェア、バージョン1.3の複数点の非パラメトリック分析プログラムを用いて解析した。最大 NLSスコアは3.33 ($p = 0.0004$) であった。

【図3】 IBD1によりコードされるタンパク質の図式的表示。IBD1によりコードされるタ

50

ンパク質は水平に表示する。構成する各種ドメインは各ドメインの開始と終止に対応するアミノ酸参照番号を付した図の上に表示される。タンパク質はCARDドメイン、ヌクレオチド結合ドメイン (NBD) およびロイシンに富むモチーフ (LRR) からなる。

【図4】 CDに関連する3つの変異体におけるIBD1/NOD2 タンパク質の図式的表示。

A: IBD1候補遺伝子のcDNA配列から推定して作製した翻訳物はNOD2 (Ogura et al., 2000) のものと同じである。このポリペプチドは2つのCARDドメイン (Caspase Recruitment Domain)、ヌクレオチド結合ドメイン (NBD)、27アミノ酸の10回の繰り返しおよびロイシンに富むモチーフ (LRR) を含む。NBDのモチーフA (Pループ) のATP/GTP結合部位の共通配列は黒丸で示す。CDと関連する3つの主要変異体によりコードされる配列の変化はSNP 8 (R675W)、SNP 12 (G881R) およびSNP 13 (フレームシフト 980) である。このフレームシフトは980位でロイシンコドンプロリンコドンに変化させ、直ちに終結コドンが続く。

B: 457人のCD患者、159人のUC患者および罹患しておらず血縁関係のない103人におけるNOD2の稀少ミスセンス変異体。稀少ミスセンス変異体の位置は3つのグループに対して示される。左の目盛りは検討中のグループで同定された変異体の数を示し、右の目盛りは突然変異の頻度を計る。多型V928Iの対立遺伝子頻度は3つのグループであまり異ならず (0.92 : 0.08)、対応する遺伝子型はHardy-Weinberg平衡にある。

参考文献

- Auphan et al. (1995), *Science* 270, 286-90.
- Asakawa et al. (1997), *Gene*, 191, 69.
- Becker et al. (1998), *Proc. Natl. Acad. Sci. USA*, 95, 9979.
- Bertin et al. (1999), *J. Biol. Chem.*, 274, 12955.
- Buckholz (1993), *Curr. Op. Biotechnology* 4, 538.
- Carter, (1993), *Curr. Op. Biotechnology* 3, 533.
- Cho et al. (1998), *Proc. Natl. Acad. Sci. USA*, 95, 7502.
- Duck et al. (1990), *Biotechniques*, 9, 142.
- Edwards and Aruffo (1993), *Curr. Op. Biotechnology*, 4, 558.
- Epstein (1992), *Medecine/Sciences*, 8, 902.
- Guatelli et al. (1990), *Proc. Natl. Acad. Sci. USA* 87: 1874.
- Hugot et al. (1996), *Nature*, 379, 821.
- Inohara et al. (1999), *J. Biol. Chem.*, 274, 14560.
- Inohara et al. (2000), *J. Biol. Chem.*
- Kievitis et al. (1991), *J. Virol. Methods*, 35, 273.
- Kim et al. (1996), *Genomics*, 34, 213.
- Kohler and Milstein (1975), *Nature*, 256, 495.
- Kwoh et al. (1989), *Proc. Natl. Acad. Sci. USA*, 86, 1173.
- Landegren et al. (1988), *Science* 241, 1077.
- Lander and Kruglyak (1995), *Nat. Genet.*, 11, 241.
- Luckow (1993), *Curr. Op. Biotechnology* 4, 564.
- Martin et al. (2000), *Am. J. Hum. Genet.* 67: 146-54.
- Matthews et al. (1988), *Anal. Biochem.*, 169, 1-25.
- Mckay (1999), *Gastroenterol.* 13, 509-516.
- Miele et al. (1983), *J. Mol. Biol.*, 171, 281.
- Neddleman and Wunsch (1970), *J. Mol. Biol.* 48: 443.
- Ogura et al. (2000), *J. Biol. Chem.*
- Olins and Lee (1993), *Curr. Op. Biotechnology* 4: 520.
- Perricaudet et al. (1992), *La Recherche* 23: 471.
- Pearson and Lipman (1988), *Proc. Natl. Acad. Sci. USA*, 85: 2444.
- Poltorak et al. (1998), *Sciences* 282, 2085-8.
- Rioux et al. (1998), *Gastroenterology*, 115: 1062.
- Rohlmann et al. (1996), *Nature Biotech.* 14: 1562.

- Rolfs, A. et al. (1991), Berlin: Springer-Verlag.
- Rouquier et al. (1994), *Anal. Biochem.* 217, 205.
- Sambrook et al. (1989), *Molecular cloning: a laboratory manual*, 2nd ed. Cold Spring Harbor Lab., Cold Spring Harbor, New York.
- Satsangi et al. (1996), *Nat. Genet.*, 14: 199.
- Schreiber et al. (1998), *Gut* 42, 477-84.
- Segev (1992), Kessler C. Springer Verlag, Berlin, New York, 197-205.
- Smith and Waterman (1981) *Ad. App. Math.* 2: 482.
- Steward and Yound (1984), *Solid phase peptides synthesis*, Pierce Chem. Company, Rockford, 111, 2nd ed. (1984).
- Spielman et al. (1993), *Am. J. Hum. Genet.*, 52, 506.
- Sundberg et al. (1994), *Gastroenterology*, 107, 1726-35.
- Temin (1986), *Retrovirus vectors for gene transfer*. In Kucherlapati R., ed. *Gene Transfer*, New York, Plenum Press, 149-187.
- Tromp et al. (1996), *Am. J. Hum. Genet.*, 59: 1097.
- Wahl et al. (1998), *B. J. Clin. Invest.* 101, 1163-74.
- Walker (1992), *Nucleic Acids Res.* 20: 1691.

【配列表】

LISTE DE SÉQUENCES

<110>	Fondation Jean Dausset - CEPH	
<120>	Gènes impliqués dans les maladies inflammatoires de l'intestin et leur utilisation	
<130>	D18702	
<160>	90	
<170>	PatentIn Ver. 2.1	
<210>	1	
<211>	4322	10
<212>	ADN	
<213>	Homo sapiens	
<220>		
<221>	CDS	
<222>	(1)..(3123)	
<400>	1	
atg gag aag aga agg ggt cta acc att gag tgc tgg ggc ccc caa agt	48	
Met Glu Lys Arg Arg Gly Leu Thr Ile Glu Cys Trp Gly Pro Gln Ser		
1 5 10 15		
ccc tca ctg acc ttg ttc tcc tcc cca ggt tgt gaa atg tgc tgg cag	96	
Pro Ser Leu Thr Leu Phe Ser Ser Pro Gly Cys Glu Met Cys Ser Gln		
20 25 30		20
gag gct ttt cag gca cag agg agc cag ctg gtc gag ctg ctg gtc tca	144	
Glu Ala Phe Gln Ala Gln Arg Ser Gln Leu Val Glu Leu Leu Val Ser		
35 40 45		
ggg tcc ctg gaa ggc ttc gag agt gtc ctg gac tgg ctg ctg tcc tgg	192	
Gly Ser Leu Glu Gly Phe Glu Ser Val Leu Asp Trp Leu Leu Ser Trp		
50 55 60		
gag gtc ctc tcc tgg gag gac tac gag ggc ttc cac ctc ctg ggc cag	240	
Glu Val Leu Ser Trp Glu Asp Tyr Glu Gly Phe His Leu Leu Gly Gln		
65 70 75 80		
cct ctc tcc cac ttg gcc agg cgc ctt ctg gac acc gtc tgg aat aag	288	
Pro Leu Ser His Leu Ala Arg Arg Leu Leu Asp Thr Val Trp Asn Lys		
85 90 95		30
ggt act tgg gcc tgt cag aag ctc atc gcg gct gcc caa gaa gcc cag	336	
Gly Thr Trp Ala Cys Gln Lys Leu Ile Ala Ala Ala Gln Glu Ala Gln		
100 105 110		
gcc gac agc cag tcc ccc aag ctg cat ggc tgc tgg gac ccc cac tgg	384	
Ala Asp Ser Gln Ser Pro Lys Leu His Gly Cys Trp Asp Pro His Ser		
115 120 125		
ctc cac cca gcc cga gac ctg cag agt cac cgg cca gcc att gtc agg	432	
Leu His Pro Ala Arg Asp Leu Gln Ser His Arg Pro Ala Ile Val Arg		
130 135 140		
agg ctc cac agc cat gtg gag aac atg ctg gac ctg gca tgg gag cgg	480	
Arg Leu His Ser His Val Glu Asn Met Leu Asp Leu Ala Trp Glu Arg		
		40

145	150	155	160		
ggt ttc gtc agc cag tat gaa tgt gat gaa atc agg ttg ccg atc ttc				528	
Gly Phe Val Ser Gln Tyr Glu Cys Asp Glu Ile Arg Leu Pro Ile Phe	165	170	175		
aca ccg tcc cag agg gca aga agg ctg ctt gat ctt gcc acg gtg aaa				576	
Thr Pro Ser Gln Arg Ala Arg Arg Leu Leu Asp Leu Ala Thr Val Lys	180	185	190		
gcg aat gga ttg gct gcc ttc ctt cta caa cat gtt cag gaa tta cca				624	
Ala Asn Gly Leu Ala Ala Phe Leu Leu Gln His Val Gln Glu Leu Pro	195	200	205		
gtc cca ttg gcc ctg cct ttg gaa gct gcc aca tgc aag aag tat atg				672	10
Val Pro Leu Ala Leu Pro Leu Glu Ala Ala Thr Cys Lys Lys Tyr Met	210	215	220		
gcc aag ctg agg acc acg gtg tct gct cag tct cgc ttc ctc agt acc				720	
Ala Lys Leu Arg Thr Thr Val Ser Ala Gln Ser Arg Phe Leu Ser Thr	225	230	235	240	
tat gat gga gca gag acg ctc tgc ctg gag gac ata tac aca gag aat				768	
Tyr Asp Gly Ala Glu Thr Leu Cys Leu Glu Asp Ile Tyr Thr Glu Asn	245	250	255		
gtc ctg gag gtc tgg gca gat gtg ggc atg gct gga tcc ccg cag aag				816	
Val Leu Glu Val Trp Ala Asp Val Gly Met Ala Gly Ser Pro Gln Lys	260	265	270		
agc cca gcc acc ctg ggc ctg gag gag ctc ttc agc acc cct ggc cac				864	20
Ser Pro Ala Thr Leu Gly Leu Glu Glu Leu Phe Ser Thr Pro Gly His	275	280	285		
ctc aat gac gat gcg gac act gtg ctg gtg gtg ggt gag gcg ggc agt				912	
Leu Asn Asp Asp Ala Asp Thr Val Leu Val Val Gly Thr Glu Ala Gly Ser	290	295	300		
ggc aag agc acg ctc ctg cag cgg ctg cac ttg ctg tgg gct gca ggg				960	
Gly Lys Ser Thr Leu Leu Gln Arg Leu His Leu Leu Trp Ala Ala Gly	305	310	315	320	
caa gac ttc cag gaa ttt ctc ttt gtc ttc cca ttc agc tgc cgg cag				1008	
Gln Asp Phe Gln Glu Phe Leu Phe Val Phe Pro Phe Ser Cys Arg Gln	325	330	335		
ctg cag tgc atg gcc aaa cca ctc tct gtg cgg act cta ctc ttt gag				1056	30
Leu Gln Cys Met Ala Lys Pro Leu Ser Val Arg Thr Leu Leu Phe Glu	340	345	350		
cac tgc tgt tgg cct gat gtt ggt caa gaa gac atc ttc cag tta ctc				1104	
His Cys Cys Trp Pro Asp Val Gly Gln Glu Asp Ile Phe Gln Leu Leu	355	360	365		
ctt gac cac cct gac cgt gtc ctg tta acc ttt gat gcc ttt gac gag				1152	
Leu Asp His Pro Asp Arg Val Leu Leu Thr Phe Asp Gly Phe Asp Glu	370	375	380		
ttc aag ttc agg ttc acg gat cgt gaa cgc cac tgc tcc ccg acc gac				1200	
Phe Lys Phe Arg Phe Thr Asp Arg Glu Arg His Cys Ser Pro Thr Asp	385	390	395	400	40

ccc acc tct gtc cag acc ctg ctc ttc aac ctt ctg cag ggc aac ctg 1248
Pro Thr Ser Val Gln Thr Leu Leu Phe Asn Leu Leu Gln Gly Asn Leu
405 410 415

ctg aag aat gcc cgc aag gtg gtg acc agc cgt ccg gcc gct gtg tcg 1296
Leu Lys Asn Ala Arg Lys Val Val Thr Ser Arg Pro Ala Ala Val Ser
420 425 430

gcg ttc ctc agg aag tac atc cgc acc gag ttc aac ctc aag ggc ttc 1344
Ala Phe Leu Arg Lys Tyr Ile Arg Thr Glu Phe Asn Leu Lys Gly Phe
435 440 445

tct gaa cag ggc atc gag ctg tac ctg agg aag cgt cat cat gag ccc 1392
Ser Glu Gln Gly Ile Glu Leu Tyr Leu Arg Lys Arg His His Glu Pro
450 455 460

ggg gtg gcg gac cgc ctc atc cgc ctg ctc caa gag acc tca gcc ctg 1440
Gly Val Ala Asp Arg Leu Ile Arg Leu Leu Gln Glu Thr Ser Ala Leu
465 470 475 480

cac ggt ttg tgc cac ctg cct gtc ttc tca tgg atg gtg tcc aaa tgc 1488
His Gly Leu Cys His Leu Pro Val Phe Ser Trp Met Val Ser Lys Cys
485 490 495

cac cag gaa ctg ttg ctg cag gag ggg ggg tcc cca aag acc act aca 1536
His Gln Glu Leu Leu Leu Gln Glu Gly Ser Pro Lys Thr Thr Thr
500 505 510

gat atg tac ctg ctg att ctg cag cat ttt ctg ctg cat gcc acc ccc 1584
Asp Met Tyr Leu Leu Ile Leu Gln His Phe Leu Leu His Ala Thr Pro
515 520 525

cca gac tca gct tcc caa ggt ctg gga ccc agt ctt ctt cgg ggc cgc 1632
Pro Asp Ser Ala Ser Gln Gly Leu Gly Pro Ser Leu Leu Arg Gly Arg
530 535 540

ctc ccc acc ctc ctg cac ctg ggc aga ctg gct ctg tgg ggc ctg ggc 1680
Leu Pro Thr Leu Leu His Leu Gly Arg Leu Ala Leu Trp Gly Leu Gly
545 550 555 560

atg tgc tgc tac gtg ttc tca gcc cag cag ctc cag gca gca cag gtc 1728
Met Cys Cys Tyr Val Phe Ser Ala Gln Gln Leu Gln Ala Ala Gln Val
565 570 575

agc cct gat gac att tct ctt ggc ttc ctg gtg cgt gcc aaa ggt gtc 1776
Ser Pro Asp Asp Ile Ser Leu Gly Phe Leu Val Arg Ala Lys Gly Val
580 585 590

gtg cca ggg agt acg gcg ccc ctg gaa ttc ctt cac atc act ttc cag 1824
Val Pro Gly Ser Thr Ala Pro Leu Glu Phe Leu His Ile Thr Phe Gln
595 600 605

tgc ttc ttt gcc gcg ttc tac ctg gca ctc agt gct gat gtg cca cca 1872
Cys Phe Phe Ala Ala Phe Tyr Leu Ala Leu Ser Ala Asp Val Pro Pro
610 615 620

gct ttg ctc aga cac ctc ttc aat tgt ggc agg cca ggc aac tca cca 1920
Ala Leu Leu Arg His Leu Phe Asn Cys Gly Arg Pro Gly Asn Ser Pro
625 630 635 640

10

20

30

atg gcc agg ctc	ctg ccc acg atg tgc atc cag gcc tcg gag gga aag	1968	
Met Ala Arg Leu	Leu Pro Thr Met Cys Ile Gln Ala Ser Glu Gly Lys		
	645 650 655		
gac agc agc gtg	gca gct ttg ctg cag aag gcc gag ccg cac aac ctt	2016	
Asp Ser Ser Val	Ala Ala Leu Leu Gln Lys Ala Glu Pro His Asn Leu		
	660 665 670		
cag atc aca gca	gcc ttc ctg gca ggg ctg ttg tcc cgg gag cac tgg	2064	
Gln Ile Thr Ala	Ala Phe Leu Ala Gly Leu Leu Ser Arg Glu His Trp		
	675 680 685		
ggc ctg ctg gct	gag tgc cag aca tct gag aag gcc ctg ctc cgg cgc	2112	10
Gly Leu Leu Ala	Glu Cys Gln Thr Ser Glu Lys Ala Leu Leu Arg Arg		
	690 695 700		
cag gcc tgt gcc	cgc tgg tgt ctg gcc cgc agc ctc cgc aag cac ttc	2160	
Gln Ala Cys Ala	Arg Trp Cys Leu Ala Arg Ser Leu Arg Lys His Phe		
	705 710 715 720		
cac tcc atc ccg	cca gct gca ccg ggt gag gcc aag agc gtg cat gcc	2208	
His Ser Ile Pro	Pro Ala Ala Pro Gly Glu Ala Lys Ser Val His Ala		
	725 730 735		
atg ccc ggg ttc	atc tgg ctc atc cgg agc ctg tac gag atg cag gag	2256	
Met Pro Gly Phe	Ile Trp Leu Ile Arg Ser Leu Tyr Glu Met Gln Glu		
	740 745 750		
gag cgg ctg gct	cgg aag gct gca cgt ggc ctg aat gtt ggg cac ctc	2304	20
Glu Arg Leu Ala	Arg Lys Ala Ala Arg Gly Leu Asn Val Gly His Leu		
	755 760 765		
aag ttg aca ttt	tgc agt gtg ggc ccc act gag tgt gct gcc ctg gcc	2352	
Lys Leu Thr Phe	Cys Ser Val Gly Pro Thr Glu Cys Ala Ala Leu Ala		
	770 775 780		
ttt gtg ctg cag	cac ctt cgg cgg ccc gtg gcc ctg cag ctg gac tac	2400	
Phe Val Leu Gln	His Leu Arg Arg Pro Val Ala Leu Gln Leu Asp Tyr		
	785 790 795 800		
aac tct gtg ggt	gac att ggc gtg gag cag ctg ctg cct tgc ctt ggt	2448	
Asn Ser Val Gly	Asp Ile Gly Val Glu Gln Leu Leu Pro Cys Leu Gly		
	805 810 815		
gtc tgc aag gct	ctg tat ttg cgc gat aac aat atc tca gac cga ggc	2496	30
Val Cys Lys Ala	Leu Tyr Leu Arg Asp Asn Asn Ile Ser Asp Arg Gly		
	820 825 830		
atc tgc aag ctc	att gaa tgt gct ctt cac tgc gag caa ttg cag aag	2544	
Ile Cys Lys Leu	Ile Glu Cys Ala Leu His Cys Glu Gln Leu Gln Lys		
	835 840 845		
tta gct cta ttc	aac aac aaa ttg act gac ggc tgt gca cac tcc atg	2592	
Leu Ala Leu Phe	Asn Asn Lys Leu Thr Asp Gly Cys Ala His Ser Met		
	850 855 860		
gct aag ctc ctt	gca tgc agg cag aac ttc ttg gca ttg agg ctg ggg	2640	
Ala Lys Leu Leu	Ala Cys Arg Gln Asn Phe Leu Ala Leu Arg Leu Gly		
	865 870 875 880		
aat aac tac atc	act gcc gcg gga gcc caa gtg ctg gcc gag ggg ctc	2688	40

Asn Asn Tyr Ile Thr Ala Ala Gly Ala Gln Val Leu Ala Glu Gly Leu
885 890 895

cga ggc aac acc tcc ttg cag ttc ctg gga ttc tgg ggc aac aga gtg 2736
Arg Gly Asn Thr Ser Leu Gln Phe Leu Gly Phe Trp Gly Asn Arg Val
900 905 910

ggt gac gag ggg gcc cag gcc ctg gct gaa gcc ttg ggt gat cac cag 2784
Gly Asp Glu Gly Ala Gln Ala Leu Ala Glu Ala Leu Gly Asp His Gln
915 920 925

agc ttg agg tgg ctc agc ctg gtg ggg aac aac att ggc agt gtg ggt 2832
Ser Leu Arg Trp Leu Ser Leu Val Gly Asn Asn Ile Gly Ser Val Gly
930 935 940

gcc caa gcc ttg gca ctg atg ctg gca aag aac gtc atg cta gaa gaa 2880
Ala Gln Ala Leu Ala Leu Met Leu Ala Lys Asn Val Met Leu Glu Glu
945 950 955 960

ctc tgc ctg gag gag aac cat ctc cag gat gaa ggt gta tgt tct ctc 2928
Leu Cys Leu Glu Glu Asn His Leu Gln Asp Glu Gly Val Cys Ser Leu
965 970 975

gca gaa gga ctg aag aaa aat tca agt ttg aaa atc ctg aag ttg tcc 2976
Ala Glu Gly Leu Lys Lys Asn Ser Ser Leu Lys Ile Leu Lys Leu Ser
980 985 990

aat aac tgc atc acc tac cta ggg gca gaa gcc ctc ctg cag gcc ctt 3024
Asn Asn Cys Ile Thr Tyr Leu Gly Ala Glu Ala Leu Leu Gln Ala Leu
995 1000 1005

gaa agg aat gac acc atc ctg gaa gtc tgg ctc cga ggg aac act ttc 3072
Glu Arg Asn Asp Thr Ile Leu Glu Val Trp Leu Arg Gly Asn Thr Phe
1010 1015 1020

tct cta gag gag gtt gac aag ctc ggc tgc agg gac acc aga ctc ttg 3120
Ser Leu Glu Glu Val Asp Lys Leu Gly Cys Arg Asp Thr Arg Leu Leu
1025 1030 1035 1040

ctt tgaagtctcc gggaggatgt tcgtctcagt ttgtttgtga caggctgtga 3173
Leu

gtttggggccc cagaggctgg gtgacatgtg ttggcagcct ctccaaaatg agccctgtcc 3233

tgccctaaggc tgaacttggt ttctgggaac accataggtc acctttatcc tggcagagga 3293

gggagcatca gtgccctcca ggatagactt ttcccaagcc tacttttgcc attgacttct 3353

tcccaagatt caatcccagg atgtacaagg acagcccccc tccatagtat gggactggcc 3413

tctgctgac ccccagget tccgtgtggg tcagtggggc ccatggatgt gcttggtaac 3473

tgagtgcctt ttggtggaga ggccggccc acataattca ggaagcagct ttccccatgt 3533

ctcgactcat ccatccaggc cattcccgt ctctggttcc tcccctctc ctggactcct 3593

gcacacgctc ctccctctga ggctgaaatt cagaatatta gtgacctcag ctttgatatt 3653

tcacttacag ccccccaac cctggcacc aggggtggaa gggctacacc tttagcctgcc 3713

ctcctttccg gtgtttaaga catttttga aggggacacg tgacagccgt ttgttcccc 3773

10

20

30

40

agacattcta ggtttgcaag aaaaaatga ccacactcca gctgggatca catgtggact 3833
 tttatttcca gtgaaatcag ttactcttca gttaagcctt tggaacacagc tcgactttaa 3893
 aaagctccaa atgcagcttt aaaaaattaa tctggggccag aatttcaaac ggccctcacta 3953
 ggcttctggt tgatgcoctgt gaactgaact ctgacaacag acttctgaaa tagaccaca 4013
 agaggcagtt ccatttcatt tgtgccagaa tgcttttagga tgtacagtta tggattgaaa 4073
 gtttacagga aaaaaaatta ggccgttcct tcaaagcaaa tgtcttctctg gattattcaa 4133
 aatgatgtat gttgaagcct ttgtaaattg tcagatgctg tgcaaatggt attattttaa 4193
 acattatgat gtgtgaaaac tggttaatat ttataggtca ctttgtttta ctgtcttaag 4253
 tttatactct tatagacaac atggccgtga actttatgct gtaaataatc agaggggaat 4313
 aaactggtg 4322

10

<210> 2
 <211> 1041
 <212> PRT
 <213> Homo sapiens

<400> 2
 Met Glu Lys Arg Arg Gly Leu Thr Ile Glu Cys Trp Gly Pro Gln Ser
 1 5 10 15
 Pro Ser Leu Thr Leu Phe Ser Ser Pro Gly Cys Glu Met Cys Ser Gln
 20 25 30
 Glu Ala Phe Gln Ala Gln Arg Ser Gln Leu Val Glu Leu Leu Val Ser
 35 40 45
 Gly Ser Leu Glu Gly Phe Glu Ser Val Leu Asp Trp Leu Leu Ser Trp
 50 55 60
 Glu Val Leu Ser Trp Glu Asp Tyr Glu Gly Phe His Leu Leu Gly Gln
 65 70 75 80
 Pro Leu Ser His Leu Ala Arg Arg Leu Leu Asp Thr Val Trp Asn Lys
 85 90 95
 Gly Thr Trp Ala Cys Gln Lys Leu Ile Ala Ala Ala Gln Glu Ala Gln
 100 105 110
 Ala Asp Ser Gln Ser Pro Lys Leu His Gly Cys Trp Asp Pro His Ser
 115 120 125
 Leu His Pro Ala Arg Asp Leu Gln Ser His Arg Pro Ala Ile Val Arg
 130 135 140
 Arg Leu His Ser His Val Glu Asn Met Leu Asp Leu Ala Trp Glu Arg
 145 150 155 160
 Gly Phe Val Ser Gln Tyr Glu Cys Asp Glu Ile Arg Leu Pro Ile Phe
 165 170 175

20

30

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

10

20

30

40

500			505			510										
Asp	Met	Tyr	Leu	Leu	Ile	Leu	Gln	His	Phe	Leu	Leu	His	Ala	Thr	Pro	
	515						520					525				
Pro	Asp	Ser	Ala	Ser	Gln	Gly	Leu	Gly	Pro	Ser	Leu	Leu	Arg	Gly	Arg	
	530					535					540					
Leu	Pro	Thr	Leu	Leu	His	Leu	Gly	Arg	Leu	Ala	Leu	Trp	Gly	Leu	Gly	
545					550					555					560	
Met	Cys	Cys	Tyr	Val	Phe	Ser	Ala	Gln	Gln	Leu	Gln	Ala	Ala	Gln	Val	
			565						570					575		
Ser	Pro	Asp	Asp	Ile	Ser	Leu	Gly	Phe	Leu	Val	Arg	Ala	Lys	Gly	Val	10
		580						585					590			
Val	Pro	Gly	Ser	Thr	Ala	Pro	Leu	Glu	Phe	Leu	His	Ile	Thr	Phe	Gln	
		595					600					605				
Cys	Phe	Phe	Ala	Ala	Phe	Tyr	Leu	Ala	Leu	Ser	Ala	Asp	Val	Pro	Pro	
	610					615					620					
Ala	Leu	Leu	Arg	His	Leu	Phe	Asn	Cys	Gly	Arg	Pro	Gly	Asn	Ser	Pro	
625					630					635					640	
Met	Ala	Arg	Leu	Leu	Pro	Thr	Met	Cys	Ile	Gln	Ala	Ser	Glu	Gly	Lys	
			645						650					655		
Asp	Ser	Ser	Val	Ala	Ala	Leu	Leu	Gln	Lys	Ala	Glu	Pro	His	Asn	Leu	20
			660					665					670			
Gln	Ile	Thr	Ala	Ala	Phe	Leu	Ala	Gly	Leu	Leu	Ser	Arg	Glu	His	Trp	
		675					680						685			
Gly	Leu	Leu	Ala	Glu	Cys	Gln	Thr	Ser	Glu	Lys	Ala	Leu	Leu	Arg	Arg	
	690					695					700					
Gln	Ala	Cys	Ala	Arg	Trp	Cys	Leu	Ala	Arg	Ser	Leu	Arg	Lys	His	Phe	
705					710					715					720	
His	Ser	Ile	Pro	Pro	Ala	Ala	Pro	Gly	Glu	Ala	Lys	Ser	Val	His	Ala	
			725					730						735		
Met	Pro	Gly	Phe	Ile	Trp	Leu	Ile	Arg	Ser	Leu	Tyr	Glu	Met	Gln	Glu	30
			740					745					750			
Glu	Arg	Leu	Ala	Arg	Lys	Ala	Ala	Arg	Gly	Leu	Asn	Val	Gly	His	Leu	
		755					760						765			
Lys	Leu	Thr	Phe	Cys	Ser	Val	Gly	Pro	Thr	Glu	Cys	Ala	Ala	Leu	Ala	
	770					775					780					
Phe	Val	Leu	Gln	His	Leu	Arg	Arg	Pro	Val	Ala	Leu	Gln	Leu	Asp	Tyr	
785					790					795					800	
Asn	Ser	Val	Gly	Asp	Ile	Gly	Val	Glu	Gln	Leu	Leu	Pro	Cys	Leu	Gly	
			805					810						815		
Val	Cys	Lys	Ala	Leu	Tyr	Leu	Arg	Asp	Asn	Asn	Ile	Ser	Asp	Arg	Gly	40
			820					825					830			

Ile Cys Lys Leu Ile Glu Cys Ala Leu His Cys Glu Gln Leu Gln Lys
835 840 845

Leu Ala Leu Phe Asn Asn Lys Leu Thr Asp Gly Cys Ala His Ser Met
850 855 860

Ala Lys Leu Leu Ala Cys Arg Gln Asn Phe Leu Ala Leu Arg Leu Gly
865 870 875 880

Asn Asn Tyr Ile Thr Ala Ala Gly Ala Gln Val Leu Ala Glu Gly Leu
885 890 895

Arg Gly Asn Thr Ser Leu Gln Phe Leu Gly Phe Trp Gly Asn Arg Val
900 905 910

Gly Asp Glu Gly Ala Gln Ala Leu Ala Glu Ala Leu Gly Asp His Gln
915 920 925

Ser Leu Arg Trp Leu Ser Leu Val Gly Asn Asn Ile Gly Ser Val Gly
930 935 940

Ala Gln Ala Leu Ala Leu Met Leu Ala Lys Asn Val Met Leu Glu Glu
945 950 955 960

Leu Cys Leu Glu Glu Asn His Leu Gln Asp Glu Gly Val Cys Ser Leu
965 970 975

Ala Glu Gly Leu Lys Lys Asn Ser Ser Leu Lys Ile Leu Lys Leu Ser
980 985 990

Asn Asn Cys Ile Thr Tyr Leu Gly Ala Glu Ala Leu Leu Gln Ala Leu
995 1000 1005

Glu Arg Asn Asp Thr Ile Leu Glu Val Trp Leu Arg Gly Asn Thr Phe
1010 1015 1020

Ser Leu Glu Glu Val Asp Lys Leu Gly Cys Arg Asp Thr Arg Leu Leu
1025 1030 1035 1040

Leu

10

20

<210> 3
<211> 37443
<212> ADN
<213> Homo sapiens

30

<220>
<221> exon
<222> (63)..(106)

<220>
<221> exon
<222> (3908)..(4406)

<220>
<221> exon
<222> (12307)..(12412)

40

<220>
 <221> exon
 <222> (15010)..(16825)

<220>
 <221> exon
 <222> (21017)..(21100)

<220>
 <221> exon
 <222> (21321)..(21404)

<220>
 <221> exon
 <222> (24355)..(24438)

<220>
 <221> exon
 <222> (27052)..(27135)

<220>
 <221> exon
 <222> (27730)..(27813)

<220>
 <221> exon
 <222> (29917)..(30000)

<220>
 <221> exon
 <222> (34244)..(34327)

<220>
 <221> exon
 <222> (36123)..(37443)

10

20

<400> 3
 tcaccatata actggtatth aaagccacaa gagcagggtg gctcatctag ggatggagtg 60
 atatggagaa gagaaggggt ctaaccattg agtgctgggg cccccagtgtag taggaaccag 120
 ccaagaagac agaaagagtg aaaatcagag agttgggggt tcctggagga aatgaagaaa 180
 atgccccaaa gaggaaggag ggaacaaaata tgaccaatgc ccctggcaga gcaagcaggc 240
 tgagggctga ggattgagca atgggaggtc actggtgaca gtttactgg agctggatgg 300
 ggaactagag ggaatgggag gggatgggag gacttgggga cagcagtaca ggcaacagac 360
 aagggggcct gctgtaaagg gagcagataa atgggattgg agccaaatga agaaggggag 420
 tgtcaagaga gtgctttact ttacaatgg agaattagag tgcattgtgc actgggtggg 480
 ggatttgatc tcttagggag agaacagtg tagggaggga gaatgcagga tagctggggg 540
 aggggtgggg gcttggcccc agcagagact caggacactt gggaagtga gcttccctgg 600
 gcttcccctc ctctcctgtc tgcaaggggt cagtgggctg agatttcagc acttaagcaa 660
 agcatttgct cttggcccca gagaaaccgg gctggctgtg gtctcaggaa ggaaggaggt 720
 gtccaggctc aggcctgggc ctgggtttca gggagggccc acgtgggtca ccccttgacc 780
 ctctctttca gcaaggaagt gatcctttct ctacatgggc ctcaccttgg ggaggacaat 840
 ggtgtctttg aagtgtgagt aactgaagta gagatcaaaa ggcaatgcag atagactgac 900
 agatttcgcc tgaagagggg aagcccagcc aggttaataa ggagtaagag gaaggatgtt 960
 aaggacaatt ttaggaaaca gataatgagt gaatattttt tctctctctt tcccaattta 1020
 aactgaagca ggagaaactg aagctagaca taatgattaa ctcccaagc tggtagctt 1080
 cctgagctgg ttagtgagaa cagcactaag gccaggttct cctcccaga tgtttaagat 1140
 gagacaggac aatgcctgct cagagacagg gcctggctga attggccctc aggattctct 1200
 ctgctctgag gtttctggaa gaaggccagg gcagaggtgt ggtgatgtag ctgctgggag 1260
 gacagagctc caggtcacgt ggcttgggcg ggcctcccct tcctgggtgc cacagaagcc 1320
 caacgtcact agctggggtg tgtatggctc acacgtaggc caggctgccc taggcttgg 1380

30

40

```

gtgcaagggga ggggccccta cttacttgtg gcctgtcccc tcgtgaatgt gtctcatgtc 1440
cccagttggg tttttcagtg agggcatgg tctccaggat gcacaaggct ttgtgcca 1500
atbtgcttga atgtcctagl tctggaagge tggttggcca actctggcct ccggcttttc 1560
ctttgggaat tccccttga ggtggggttg gttagacagat ccaggctcac cagtcctgtg 1620
ccactgggct tttggcattc tgcacaagge ctaccocgag atgccatgcc tgctcccca 1680
gcctaattggg ctttgatggg ggaagagggg ggttcagcct ctacagatga ggaggaaaga 1740
gcaagtgtcc tctcgggaca ttctccgggt aagaggagca ggcattgtcc cgtcccagct 1800
tgatcctcag ccttctttca tccctggccg cgacatgctc ccaggcctgg ggtcagatgg 1860
ggagtgtgta cttctgtttc gggctgtttt ctggggagaa tgggtcggcg ggtttttttc 1920
cccaggacct gggcagggtc aatggtgggg gccgctgtcg catccttggc tgggttttcc 1980
acagctgaga accactccag ggccaagccc agagcttatt ctaccctttt ttgtcctctc 2040
ttcccctgtc ctggccacc ccacctctt ggctcctctg cttagatgtg ggcacaagga 2100
ggagaactcc ttggcctgag agaactacct tagatcctgg ctccagtggt cctctgcagg 2160
ggggtacacc ctctctccca agcagccaga cacacaagta acctcattgc ctcagtttcc 2220
ccatctgacc agcacagggc cccctgtgcc ccagcagcgt tctgagagat tggagcttcc 2280
tccttttgtc taccttggct accgtatgag gacggataca gagtgttccc cccaccccca 2340
gcccagggga tatlgtattc atgaacattc cctcagtgtc tttgtggggg acaatgtctg 2400
gccaggctca gggatgccag gacgagtaag acccaggctc ccacgtggcc caggcaggga 2460
gagagacaca taaacaacca tcaggaaaga ggtaaaatcc ccaggcactc tggcatctgc 2520
tcccctgagt gtcctgggat gtcctgatt tataaaaaga agctgacggc cctctttgtt 2580
gtccatggct acaccttct cttcggggca ctgacgagc ccttgtccac 2640
agaccccatg acaatcgag aactgacat gctgagagat ttcttggct gctcagggac 2700
cctgccaggg ctggaagctc ctggagggtc acttgccttc aaattcccag aacgcacagc 2760
aggctcactga tgatagcagt ggcagcagtc tgtgcacggt ggtttcaggg gcgtgggagg 2820
gaggtgaggg ccctagggca agtgtgtgtg ggaagtgtg atgggggaca aggcaccaga 2880
acgctcggaa acaacttagt ttgcaccgta attttctact tgccttagga caggacctt 2940
agagcaatat tctgagtcta cccctggag tagcagtgtg caaaacacac agcacgggct 3000
tggggccccc gtggggaacc caaatgtaag agttagagac atgcatcccg gagtcataca 3060
tggctcgtgt tgaaatctct actctgcctg tctagctgtg acacatcgta caaatcactt 3120
agcttctgtg tgcctcagtg tcttctctg tagaatgggt agatcatagg cactacttca 3180
gagtggtggg gagggttcag tgaattcctg caggagagca cttagaatgg cacttgggtg 3240
gtagtattat cttaattaat attagccgtt actgaaactg ctgtagcctg aatccagcca 3300
gcatgaaaga gcccctctca cctgtctctg aagagaatga attccctgat tglttggaag 3360
atctctctct ctctctctgt cttttttttt ttttttttag aaacggtctt gctctcttgc 3420
ccaggctgga gcgcaatggg gccatcttgg ctcaactgcaa cctctgcctc ccgggttcaa 3480
gtgatctccc tgtctcagcc tctgagtag ctgggattac aggcctctgc caccacgctc 3540
ggctaatttt tgtattttta gtagagacag cgtttcaccg tgbtggccgg gctggcttag 3600
cgctcctgat ctcaagtgac ctltgggagat ctcttgcctc taatattacc tcaagccttt 3660
ttaaacgltt taagccggag accaagcatg gatatgggag ttaggggctc tgatttaatt 3720
cttggttgct tcaaacctctg tggaaacctg aggtgtttct tgccctctct ggttctcaat 3780
tttccacatc atatgggtgg gagcttggat tgggtaatgt ctgaggctag aacctaggcc 3840
aactcggggt ctgctggggc tgaactgccc tggccttccc tgaccacctc gcatctggct 3900
tctggagaag tccctcactg acctgttct cctcccagg ttgtgaaatg tgctcagag 3960
aggcttttca ggcacagagc agccagctgg tccagctgct ggtctcaggg tccctggaag 4020
gcttcgagag tgtcctggag tggctgctgt cctgggaggt cctctcctgg gaggactacg 4080
agggcttcca cctcctgggc cagcctctct cccacttggc caggcgcctt ctggacaccg 4140
tctggaataa gggtaacttg gcctgtcaga agctcatcgc ggctgcccaa gaagccagg 4200
ccgacagcca gtcccccaag ctgcatggct gctgggacc ccactcgtc caccagccc 4260
gagacctgca gagtccagg ccagccatg tcaggaggct ccacagccat gtggagaaca 4320
tgctggacct ggcatgggag cggggtttct tcagccagta tgaatgtgat gaaatcaggt 4380
tgccgatctt cacaccgtcc cagaggggtga ggcactcctg gtgtgcatca cagagtctc 4440
aggaaagggg tgcttagtca ccaagactga tttgtctca tgaagttagc ctgtggggta 4500
acttggctcc tgggatttcc cctaaaaagg tagccaggca ggtaaaattt gctcttgact 4560
cttggcagga aacatacaac tcttcttctt tcttcttctt tcttcttctt cactctgta 4620
ccctggctag aatgtagtggt cacaatcata gctcactgta gccttgaatt cctgcgctca 4680
agtgatcttc tggccttaga gtagctggga ctacggctgc tgtaccacca tgaacagcta 4740
atttttttt tttcttttag agatgggggt ttgctatggt gccaggctg gtctccagct 4800
cctggcttta agcaatctc ccgcttggc ctcccaact gttgggattg caggcatgag 4860
ccacttggc tggccaacag aacacttctg ccgagaggaa gtgtgtggtg gccaggaaact 4920
cagattctg agccagaatg gtgcaggctc aaggtaacc ctgtgtgatc tcaggcttcc 4980
ctatggagcc tctccagcct cagtctcctc tgttctcagt tctcactca .caaaaacatg 5040

```

10

20

30

40

```

ttaatagtca aatggtgcct atcctataag gctcctggga ggattcagtg agttaatttg 5100
agtaatgctt aggatagtgt ctattaccac tggctgctat ttattatbbc tgttatgagt 5160
gatactctgt actgtgtacac ttttatttct gtctgtttta aattaacagc acaacagacc 5220
ataaacctgc agtatattga atttatttta taattaacat agcatattat aaactaatat 5280
agcttaaatg tttatgtagg atttctgaca tgaaatgca ttagatcata gatgttcaga 5340
gttggtatat aacagcccct gagaatgtag taactcagca gagaccagaa ggtcagagaa 5400
atgaccactg agtattttly aaactctttt gttttcttcc aaatagtgat tcttagggct 5460
cctgagagggc agatggaaca atcattaaca ttccacttta taaatcggga agttgagacc 5520
aaggaaagta gtttgaataa gctcacagta gttaatgagg gggccagtg tggaccaatt 5580
ggccagcact ggtcattgac ttatcatcc atcattcatt taltcagcca gaatctatta 5640
ggtgcttcat acatatttgc ttaaagtttg ttgtgttcat agagctttgc acacggtagg 5700
tactccataa acatttgttg atgaataag tgagttactg aatgaatgat tgaattagaa 5760
tgacactgca ttagttaaaa gggctggggt ggggaacatt ttagtltttg tttttgtctg 5820
ttttccaaaa atgtatgtgt tgttccatg agtctggata accctagatt gagattgatg 5880
acataaataa atttgtcttc aaggctgcac taaagctggc tcacatggct aggtatttac 5940
agagcagaag tgggtgcagtc ctctctgatt agttgcacgt acagaagaca taltcgttat 6000
tggactgacc ttagtttctc ttataatttg ttaggggaa tgaatcagcc catctgagaa 6060
gttacaagat tgtgtcttgt catctttaaa agttcagcaa tgtgatgtgg tacagatggt 6120
ctgagggggt tggagaagggt agcctagatc cctagggccc agagaagaca gbatgtgaa 6180
agaggaagta catggatttg tgaagaaaag aaatgggata actcatgggt caaagaagaa 6240
aatctatggg aaatcagaaa atattcagaa ccatacaata atgagaatat talttatcaa 6300
ataggaag aaggaaggca tttgtttatt talttgttta tttatltatt tgagatgggg 6420
gtctcactgt gtcacccagg ctgctggagt gcagtagcac gatcataaat cactgaagtc 6480
tcgaacttct gggctgaagt gatcctcccg cctcagcctt ccaagtaggt gggacacag 6540
ctagcaccac ctaaccaggc taattttttt ttgttagaca cagggtcttg ctatgtttag 6600
gtctcaact cctgggctca agtaatcctc ctccctcgge ttcccaaagt gctgggatta 6660
caggcatgag ccaactgccc catctaaggc tgaatlttaa tgagctaaga attcatctta 6720
agaaagggct aaatagacag caaaagcaaa cattgaaggt tgggactgag ctgagtgggt 6780
agcagggatg ggagacaaca gatctgagga gagcaggaga ttttgaagg attgcactgc 6840
ctgaggttta agcctttaga atccagctct ctctgagctc cctttgagct ctgacattct 6900
gtgactctga tttggtggcc ttcccttagt ggcttactg atttcatttg gatggtgctt 6960
gtggtatata caaccaacat gtcttcccaa atggcctttt aatttctctat aaagaagtag 7020
ttgtcattga ttgcaggtta gggacagaaa atgctgtgga atgaaacaaa atgcaagtta 7080
aagaactaaa ttccaaaaat acccattgct actattgact gagtgaattc ctactgttg 7140
ccagacactg taccagctcc attccctgta ttgttttatt taagcctcac aegggtatag 7200
tgtgactaca ctgtttctta acaatgaaga aactgcccac atcgccctc tgggaagcgg 7260
cccagctaga atttgaaatcc aggcctgttt tectccagag ctgtgtctat tctctgtctg 7320
tcataaaatg tgggggcttt gtgtgtgtaa cttgtctcagt tgggcatagc agttgttagg 7380
aaaactgagg ctggtaaacac cagctgtaat accagctgct cgtctgactc atgcaactgt 7440
taaagtgtat agggctgagg tgtcagactg agctctgaa tgcctgatc ctataacaat 7500
attaacttaa acatttttta aattgggaaa tgcaccatgc atacagaaga gtgtgtatat 7560
ttcatatgta tagtgtaaac tgttcccatc acccaggtta aaaaacagga tgttgccagt 7620
acctggggcc ttcttttaact gcaactgcta gaggtaaaca ctggcttgac ttttgtgtaa 7680
atcatctctt tgcctttctt taatgtttta gcatctttta aaataaatcc ccaataaatg 7740
tattgttcta ttttgaaaaa ctgagtagca agccaaaaat agctgtgtaa agaaaggta 7800
cttaaattag gctgggtgca gtggctcaag cctttaaacc cagtactttg ggaggctgag 7860
gcagggtgat cacaaggta gtagatcgag accatcctgg ccaacatgga gaaaccccgt 7920
ctctactaaa aatacaaaaa attagccaag aatagtgcca tgtgctgta gtcccagcta 7980
ctcgggagggc tgaggcagga gaatcgcttg aacccgggag gcagatgttg cagtgagctg 8040
agatcgcacg gcttgaaccc' gggaggcaga ggttgacgtg agccaagatc gcaccactgc 8100
actctagcct ggtcacaga gcaagactct gtctcaaaaa aaaaaaaaaa aaaaagaag 8160
gttactattg ccttttctta gatgaaggtt ccaaggcag ggaagctaa gtggagtctc 8220
agggacttgg tctggctttt ccttccctgg gaattataa ggacctctc tgggaagtca 8280
gtcggcaatg ccatgaatga gtctggggaa atattgggct cattgcaact ggagggtctg 8340
gtaggactga tbtgaattag gtgtgtgtc cggaggaaaa tggccagagg aagtgggctg 8400
ctttgtacag tcagtgttaa agltgccaaa ggctattata gctcacagga atgggccaag 8460
gctaaacact cctgtggagt gaaatgaatg tectcagctg actgaggcag cgggagtga 8520
gaagaaacga tattagttca tgytgaagac aagtcaata tagataaagg ttagggtcag 8580
gcttgccctgg acatctagga gataactgcc ctcaacttgt ttgaaacttg agtcaactgc 8640
ccattttggt tgaactggtg gccatctact tatagtatac agccatcaac ctgagatbbc 8700

```

10

20

30

40

```

cctacatggt cttcctgcct tggctcctg taccctgaat cctatggcct cttcttcct 8760
ggtttactac attttgctag accgtatcct ccagtcgaat ccttagaatg aatgtatgaa 8820
agttaaaatt tctgaggtct cacatgtctt aaagttccct cactctggat tgatagtgtg 8880
gctgggtata aaattctggg ctggccatca ttttccttca gaattttgat tgcattattc 8940
cattatcctc tcttttcaat attgcttcta agaattccaa aacctttttt tttttttctt 9000
tttgagacag tgtctcactc tgtcaccacg gctggaatgc agtagtgtga tctcagctca 9060
ctgcaacctc caectcctgg gttaagcga ttcttcttcc tcagcctcct gaggagctgg 9120
gattacaggg acccaccacc acacccttta gtagagatgg gtttttgcta tegtggccag 9180
gctggtcttg aacttctgac tttagggtgat ctgcctactt cggcctccca aagtgtctgg 9240
attaagggcg tgagccacca caccagcctt ccaaaaacct tttaaaactc tttctggaag 9300
cttttaaaat tttcttttag tcccagaat tttaaaattt caattatgtg ccttggtgtt 9360
cttccattat atlagtcacc caagaggtag tttcaatctg gaaacttctc tatgttttgg 9420
gaaatgttct tgattagtct actctgtttt tcttctctctc ctttttatct cttctctttt 9480
catgaaacta ctattaattc aatgttagaa ttccttgact gatcatttaa tttcttcta 9540
ttttccatct ctgtgtcttt ttgtctact tttctatgat agtcacagct ctatctttaa 9600
actcttgagt ttttcatttt tgatgtcatg attttaattt gcaagaggtg ggtttgactg 9660
attctttttt gtagtatctt actctgtttt tatggatgca acatcttctt tgacttaagg 9720
atcataagat aggtgggttc tttgtttggt tgtttgactg tttttcaacc tatgtaaact 9780
ttttctacaa gtttcttctc ctttcccccc tttttggctt ctatctccca cattagatgc 9840
tttctctggg tctcatgata tctttggttt tcttctctca gattgacagg taggacttta 9900
aaacttgttg agcatgctgg tgaacttgtt ctaccatgaa tttcactgta gatattttgg 9960
agattgacag attttgctc ttagatctc accctctggg ttgatcaagt tatctgagta 10020
caccacagac cttttgctg gggataaacc agaaatctgt ttcagaaacc actttgattc 10080
agtcttccct gtttttagtca tttccttcag tcccgagggt ccgtcatgct gatcattcca 10140
gagcccttta cagatcctag ggtacacact gcatggtttt caactttctt gttttgggg 10200
taagatttgg ctttcaggag tctcctcagt ccgttactat tcattcaatc agcaagtctt 10260
tgagcacctg attttgctca gacattcttc taggtgttag ggatacctca gtgaacaaaa 10320
cagacaaaaa tctttgtctt ggaatacac acactccagt caggggagag ggacaataag 10380
ccaaggaag gaaattacag cgtgtgctag aagggtataa gtgctgtaga aagtaagtaa 10440
agtgggtttg ggagttgaga gtttgggaag gggataaatg atggcaatbg taaatagagt 10500
agtcagagtt ctcatctaga aggtgaaatt caagtaaaga cttgaaggag gacaggggat 10560
tagccacatg gatgctagg ggaaggcttc caagctgaga ggacagccag agccaaggcc 10620
cagaggcagg agcatacctg gtatgttttag gaaacaggag gccaggatgc tgagtggagt 10680
aagagggggc atgaaaggag aaacttgggt ccacgtggtt ctagacaggt atttttgtct 10740
gttttggggc ctgaaagtta ctattggact tggactctta ctctgaggaa atagggacgc 10800
tatggggacg tttgtacagg agcaatgtga cctgagtttt gtttgtaag gatagactc 10860
tggctgtggc attaaggcta ggctgtgggg gcaggaaacag aagcaggggg accagttttg 10920
cagcctgtgc agctttccag ataagcaggg atttgtgctt ggaggaggat ggtatagagg 10980
aggtgacaag aaatgactct atgtctgtga tgtagatatt ggccacagat gccatttgag 11040
cactagagac ctggctggtc cacatggagt ttccataagc acataataca cattagattt 11100
caaaagctta atatgaaaaa aaaaatttaa cgggccocgg gaattttttt cttttttttt 11160
ttttttgaga cccagtcctg ctctgtcacc caggctggag tgcagtgtgt tgatctcggc 11220
tcaactgcaac ctccgctctc caggttcaag tgaattctct gcctcagcct cctgagbacc 11280
tgggactaca ggcacctgcc accacgcctg gctaattttt tgtattttta gtatgtatgg 11340
gttttcacca tgtgtccag gctggtctgg aactccggac cttaggggat ctaccocctt 11400
tggcctccca aattgctggg attacaggca tgagccacca tctctgcta 11460
ttttctacat ggattacatg ttgaaatggt aatgttttgg ctattgtgga ttaaatagaa 11520
tatatgatta aagttgatct catctatttc ttttaacttt aaaaaatag tctgttagag 11580
gatttgaaat tccacatgct gcttgcattt gtgacctgca tttcatttct gtggaacagt 11640
gocctttttg ggacatgctt tgaagggtga gtcaacagga tttggcagat tacagacgag 11700
aggcttcaag ggtgactcca agacttcggg gcagagcacc tggaaagaaag ggttaatat 11760
tagccaagat gaggaggctt gtoggtttgg caggtgcabg ggcaggttag gagttagtt 11820
ttgaatatgt tggaggtgtt tatgaaactt ttaagtggag atggaaaaata ggcagttgga 11880
tgtgcaagtc caggttccag ggagacagtt caggctggag abgaagatgt gggagtctga 11940
ggagagattg tattcaataa ttcaatccat gagacttgat gaaatcactt ctctccaaa 12000
tgatttacag cctgcagaat cabtttccct atctttgtag gtttatgtct tcattttgtt 12060
tcatttattt ttcagttatt cactgtttta gtgagttttg agtaggagcc agattggatg 12120
catgcttca attcaccatc caacactgta ttaactactt gaaactcatg tggttgttctg 12180
gtttgttttt tgacctttta tcttggatgg aagagagatg cttatgaagt tgcagtaatc 12240
agtaagcctt cccacattgc tccatcagcc ttctggaag aataatgtct tctgccttc 12300
ctgtaggcaa gaaggctgct tgatcttggc acgggtgaaag cgaatggatt ggctgccttc 12360

```

10

20

30

40

cttctacaac	atgttcagga	attaccagtc	ccattggccc	tgcctttgga	aggtaggtgt	12420
atgttctcag	ttaatcagaa	aggggaagggc	agtcagtgca	gatccatggt	taagagcaga	12480
acacacctcg	gttaacatcc	catatgctgg	cagtatagcc	tccctatgac	tcaatttccct	12540
tgttttaagg	ctagcaccac	cccgtctcat	tgggattttg	ggagcattaa	aaggacaaaa	12600
gcggtaatg	ttagctatta	gctttcatta	tctcccacac	agtatactga	caattgggct	12660
accatatatt	gagggctaac	taaaggtggt	acttaccatc	caaactctca	tbatctgtac	12720
cgaaaagata	tggacacatg	ttttgagtta	gggctggtat	ctcttgatct	ctgaaattta	12780
gcagctcaca	atgggaaact	caagaaccaa	gtggatctag	agactctggt	atccctcagt	12840
gccacgggtc	accacccaaa	ctcaggaaca	ggaggggctt	ggaccgcacc	acttgaacat	12900
accagggcatc	ctgccagggtg	cittatggac	aatgtctacc	ctttgcaaca	accctgagaa	12960
gtaggtgggtg	tttttttcca	ccttatagat	gtggaactg	ggcagggagg	ttaagtgcg	13020
agggagggga	agatgggtct	gattgtaaat	tgtccccacc	tacactttct	cttttcttgg	13080
gagaagaaat	gtcagttgta	aagagagagt	gcaagcctgg	cactctttag	ggcttggtcc	13140
tacaccactg	tagggaaagc	tcattggcac	tgaagcccc	tgagctgtgt	gtggctgctg	13200
cagatgggtc	tatcacctcg	gactgtgtcc	tctgggcagc	aagcaagcct	gtgggcgggg	13260
tggctggaag	tctgtgctg	gcactcgcga	gtgcaccgtc	tcattgaaga	acaggatcta	13320
aaatcagtg	ccccacagca	gggtgcgog	cacggagtgc	aggccctggt	ttggcccttg	13380
ggtgaggttt	gctgttgaca	tcataagca	cagctagtca	ctgtaagacc	aggccagggt	13440
gcaagattcc	ccacacttct	aaaggtgaca	atbgtgtat	ttatctctct	ataaaatgac	13500
attttttttt	tctggagaat	tttagtatca	ttggtgatga	ctggaaaacc	tgcatcagaa	13560
atcaggtcgg	aagaggaaga	tatatatctg	atatgtactg	gagaggaaga	tatctatctt	13620
atggtcctaag	ttcagggatc	ctggtatatt	cagagggcagc	aaagctcagc	aataatcatc	13680
aactctggga	acagaggtga	cataaacaca	ggcgtcccc	tttbtgtgac	tgcagatagt	13740
catcagtgag	ctcagagctc	tatgaaaatt	acttgctagt	ttttgggttg	aaaatagtgg	13800
gccagtggtt	ggttgggggc	agtgaggctg	tgatggcggg	ggaccatgcc	aagctcctac	13860
cagcctggga	cgctaaacca	gcacttcccc	atttctgaa	aggggaacta	aactctgaca	13920
caggaaatgg	tttgcttgca	ttactttcag	gatgagaaag	gaagagcact	ggccttccaa	13980
acacaccccg	tgcataaaaa	ctctccctgc	atggggtgca	tggggaggat	ggggaagtgg	14040
aggcaggatc	acagactcct	gttcgagtg	tcagctgggg	caccccggtg	accccgaggc	14100
cttcccttgc	taggtccacc	cagatcaatc	aggatcatct	ccccatctcg	aagtttaact	14160
ttatcacatc	tcagagttcc	ttttgccaag	taaggtaaca	tattcacagg	ttctgagaat	14220
ccggacatgg	acatctttga	gggtctatgg	ttgtgcttac	tatatccatg	aataaatgag	14280
ataataagca	ccattttttg	agagtttgcc	atgtcagata	ttcttttaaa	ctgtatttta	14340
tctcgtctcc	tccgaaaaaa	atccttccag	gtgtatatbg	tccccattht	tacagatgag	14400
agaactgagg	cccagaaagg	ctaaatggct	tgcccaagtg	tatggtggac	ccaggttttc	14460
aaactcaggt	tcagagcgtg	tcagagactg	ggctcctgag	cccttaagcc	ctttgttccc	14520
ctttagaaaa	agtcacctga	ggctgagtg	tgaagggatt	tatccaaagc	cacccggcca	14580
ctatggcagg	acagatatca	gaatacaggt	cttccgatcc	cagccagag	ccccttccc	14640
tcattctagaa	ctcctcctgg	tgctcagtaat	gataacggca	gtcactgatg	tcttttgagc	14700
acttactttg	tgttgagcac	ttacactgtg	ctaagcactt	gacataggtc	atcttagttg	14760
atccgtgtaa	aactctgtga	ggtagtgacc	aacatttctc	ccaccttaca	gaggtggaaa	14820
ctgaggggta	ggaagtthcc	ttgactgtcc	tcaaagtgca	cagcttgtga	atggaggagc	14880
caggatgggc	gcccgtctgc	tctcctatcc	cttcagttat	gtcagcgtcc	cccgcagcag	14940
ccpattgtct	ggttaggtcc	cytcttcacc	atggtgccac	cttcatctgc	ctcttcttct	15000
gccttccagc	tgccacatgc	aagaagtata	tggccaagct	gaggaccacg	gtgtctgctc	15060
agtctogctt	cctcagtaac	tatgatggag	cagagacgct	ctgcctggag	gacatataca	15120
cagagaatgt	cctggaggtc	tgggcagatg	tgggcatggc	tggatcccc	cagaagagcc	15180
cagccacctt	gggcctggag	gagctcttca	gcacccctgg	ccacctcaat	gacgatgctg	15240
acactgtgct	ggtgggtggg	gaggcgggca	gtggcaagag	cacgctcctg	cagcggctgc	15300
actttgctgtg	ggctgcagg	caagacttcc	aggaatttct	ctttgtcttc	ccattcagct	15360
gcccggcagct	gcagtgcatg	gccaaaccac	tctctgtgcg	gactctactc	tttgagcaat	15420
gctgttggcc	tgatgttgg	caagaagaca	tcttccagtt	actccttgac	caccttgacc	15480
gtgtcctggt	aacctttgat	ggctttgacg	agttcaagtt	caggtttcacg	gatcgtgaac	15540
gccactgctc	cccgaccgac	cccacctctg	tccagacctt	gctcttcaac	cttctgcagg	15600
gcaacctgct	gaagaatgcc	cgcaaggtgg	tgaccagccg	tccggcggct	gtgtcggctg	15660
tcctcaggaa	gtacatccgc	accgagttca	acctcaaggg	cttctctgaa	cagggcatcg	15720
agctgtacct	gaggaagcgt	catcatgagc	ccgggtggc	ggaccgcctc	atccgcctgc	15780
tccaagagac	ctcagccctg	cacggtttgt	gccacctgcc	tgtcttctca	tggatgggtg	15840
ccaaatgcca	ccaggaactg	ttgctgcagg	agggggggctc	cccaaagacc	actacagata	15900
tgtacctgct	gattctgcag	cattttctgc	tgcatgccac	ccccccagac	tcagcttccc	15960
aaggctctggg	accagctctt	cttcggggcc	gcctccccac	cctcctgcac	ctgggcagac	16020

10

20

30

40

tggctctgtg gggcctgggc atgtgctgct acgtgttctc agcccagcag ctccaggcag 16080
 cacaggtcag ccctgactgac atttctcttg gcttccctgg gcttgccaaa ggtgtcgtgc 16140
 caggagtagc agcgccctcg gaattccctc acatcacttt ccagtgtctc tttgcccgt 16200
 tctacctggc actcagtgct gatgtgccac cagctttgct cagacacctc ttcaattgtg 16260
 gcaggccagg caactacca atggccaggc tccctgccac gatgtgcac caggcctcgg 16320
 agggaaagga cagcagcgtg gcagctttgc tgcagaaggc cgagccgcac aacctcaga 16380
 tcacagcagc cttcctggca gggctgttgt cccgggagca ctggggcctg ctggctgagt 16440
 gccagacatc tgagaaggcc ctgctctggc gccaggcctg tgcccgtgg tgtctggccc 16500
 gcagcctccg caagcacttc cactccatcc cgccagctgc accgggtgag gccaaagagc 16560
 tgcatgccat gccggggttc atctggctca tccggagcct gtacgagatg caggaggagc 16620
 ggctggctcg gaaggctgca cgtggcctga atgttgggca cctcaagtgt acattttgca 16680
 gtgtgggccc cactgagtgt gctgcccctg cctttgtgct gcagcacctt cggcggccc 16740
 tggccctgca gctggactac aactctgtgg gtgacattgg cgtggagcag ctgctgcctt 16800
 gccttgggtg ctgcaaggct ctgtagttag tgttactggg cattgctggt caggtatggg 16860
 ggagaccat caaggctaag tgtgggagca ccgagctggg ctctagaagt ctgggcccag 16920
 cttgcctct gccaccctgc tttgcaacac tgcccagatc ccttcccttc tgggccttaa 16980
 tttcaatatg tgatgatgac agccacactt tattgactgg cctatgtgct gggctcgggt 17040
 ctatgctttc cggaatgacc tcactaatac tctacaacca cctgggggg taggcaggaa 17100
 tgttattatc tccattatcc ttgacttag gctcagagaa gtgaagtaac ttgtccagga 17160
 aatggcagag ctgggggtca caaattgcat cattctgatt acaggttttc tgccctcccac 17220
 cagtctatgg atacacttca gaggtccct gaaaacctg aggtcacttg cagaagttt 17280
 tgtgtagtat gctgcccgtat caggaacaac accaaatcag aggtgacttg tgcccacatc 17340
 gagactttaa cagcccaacc ttcaggacc ccagaaataga aagtggctgc 17400
 agggttacaa ctactgttgg atctctgagg tagcacagtg tccaaacagg atttcagcac 17460
 taccggtatt gcttagagcc ccagccaaag atgtgagggt ttgcccttg gagaatctgt 17520
 gccctgaac tggggggcct cttccacat cttgggggca ggcaaggcca gagggtgtgc 17580
 ctaggcctgc ggatcagcat gcgacagatt ccccaacatc cttccagctt gaaaggggat 17640
 tgccctgctt ctatttagaa cctataggaa agcagaagtt ctgattgaa gttaaaattg 17700
 attcccagcc tccaggggct ttgggctaca cctggatgac ctttaattgac cctaaagcat 17760
 ggacaaacca cttcctgaga gtattagat ggtatacatc ttctctgggg gcaagcaac 17820
 aagatttatt tttcatcatg gaccaaacac atggaatccc actagaaact gtgtagttaa 17880
 tttgttaac cttgacatag ggaccatggt ctttaggtta aagcataata acaacataat 17940
 acataacata tatagcgaat atatatatgt attatatgca atgaatgtaa atatgattat 18000
 acccatcatg gtcttgagg aaacagatga cacacttaaa atgggtgttt tgaggagagt 18060
 ttgaaaaaca gattgtttac aagccatggg caggagttag gaagagttag agggttgggt 18120
 caggggctg gggttagtaa cagctggggg agggtagact tgaaggggga aggggaggga 18180
 gactaattag ctgggggaa ggtatggaga cggctgcctg agcttctgca aagtggaaga 18240
 atactgcttg gccctaactc ctacccccaa ctcttgcctg tggccagcgc cttcccagg 18300
 ctggacccat cagggaggcc gagtgggctg tctgctggag tagtcccag gcatcagcct 18360
 ccaggagcc agggcgggt agagaagggg gagagtggat ctggccaggc aaatggaaaa 18420
 cagccagcac caaactctat ttcccaggca gggaggatca tgatactttg agtgggaatt 18480
 tggaaacctg tctgttggag caatttccct gatagaaata agaattgtgca ttttctggg 18540
 tagtagactc agtttttacc ccaagaggcc aggcactact ggcctgtgtg atcctcatag 18600
 gccagtccat ctctggaatt cttgaatgga tcatccatcc ttgattaggg atgtccccgt 18660
 gattaccagg gtgtgcagaa gggctctggg aaacctgtgg gtctgtctct gtgttcagag 18720
 aaagggtgag gtggcctggg tctagctcat ggtgctcaga ctgtgtgtgt taaaggcact 18780
 cgtggcaatg cagattcctg ggcctgctc tagtgattcc cattcagtag gtttgggggt 18840
 gggcccagga aatctatatt tttcacagac acccctgggt attctgatac aagtggctct 18900
 gcctgggag aactactggt ctgcagcaac cagcttgggt ttccattagc aatfactgtc 18960
 cttgagcgag ttttactgct cttcacctta cacacactaa aactgccaa gccgtagggg 19020
 aggggaaagca accatgaggt tgctgtgagt gcactgtgtg tgtgtgtgtg tgtgtgtgtg 19080
 tgtgtgtgtg tgtatgagag agagagagag attgagaaag agaggaaggg aggaagggg 19140
 agggcacagg ctctctccc acagtgccaa cctgctctc tcccacttga agcgtttcca 19200
 tgccaactga aatctcagc ctctaggaaa cctatatac acagtcccc tatataggtt 19260
 tcttttagact ctgctctct cagactctag agtgatggct ttaaaagtgt tatgttacc 19320
 acagagagag agcacgcacc accatgtaaa catggaacct aagtttcaca aaatgacttc 19380
 gctttatgaa ctctgagaca ctctgctctc ttctgttctg ttctatttcc attttagaaa 19440
 tgctgctcag gacctcaaa atgatttgca tgacctgcaa cctgcagtct gaaaaatcac 19500
 tgcactacag aagtggccat aagaggccct gagggagaag ctgcacaatg tcatggttaa 19560
 gagtggggt tggagccaa cgcctaggg tcaaagcctt tatgtgccgt acaaccttg 19620
 caaagtcact tgcctgtct gtgcctcagt ttctttctca cgaatgctca taataatggt 19680

10

20

30

40

```

tccoatttca ctggcttgtt gtgaggatga aatagtgtaa ttattgagaa gtgtaaggg 19740
tagtgatcag tgctagcgat catgattcta ggtgactttt actgtgtacc ggggtctcac 19800
aaggctttat gtgcacagcc ttggtaggct gataatacta ttgttccctc tttttttttt 19860
ttggaaacgg agtctctgtt tgttgcccag gctgggggta cagtggcaca atctcggctc 19920
atgcaatctc tgccctcccg gttaacgcca ttctcctgcc tcagcctccc aagtagctgg 19980
gactacaggg gcctgcccac acgcccggct aatltttttg tatttttggg agcgacaggg 20040
ttbcaactgt ttaaccagga tggctctgat ctctgacct cgtgatccgc ccgctcggc 20100
ctcccaaagt gctgggatta caggegtgag ccaccgtgcc cggcctgttc cctcttttat 20160
agatgaagag accagcaaat aactagtaag tcgctgatca ggatcacaat atccagctga 20220
ggcactccag agcctgagct gtttaaccatt cagtcagggc ctcccaagtt tgcctaaaga 20280
taaagaatca tgtgacagct tgttaaaata tacagattcc tggggccccc cccgcagata 20340
cttgattgcc agctccaggg tatgggcctg agaactctgtc ttttagggaa gctttcagat 20400
gatgtttgta tcaggtgagt tttgggaatg gtgcccacag aggagtggca gacagggctt 20460
gctcggcagg gactagcctg ttggagtggg gccattgggg ttaaggactg ggcagcaggg 20520
cctcactaac cacagcctat atgctctgtt ctgaagtttt ggccactctc atccagctgg 20580
tctactgtct gctgacctag atgatggtaa attgtcccca ggggtagcct gctagttca 20640
ggctgcaact ttccgcatata tcagctcctt tccaccatca tcccccttgt gaggctgctg 20700
tgattatcat gttecltttg cacagatgga aacattgctt caaattagct ctgtcatttc 20760
ctaaggattc cagggttctt tagtaggggg tctggatcct acgtcctggg ccatcccat 20820
catagtgcac cacgtcacct cctggccag ggaccgtggg gtctccactt ttttggggtg 20880
ctccatctat gcagggtttc ctggaagcac agatgctggc acttcagggg tgaatgaaag 20940
tctttttggg ggatttgtag attttttct tgtcttacta gctccatttt caaatgtatt 21000
tattttgtct ctttagtttg cgcgataaca atatctcaga cagaggcacc tgcaagctca 21060
ttgaatgtgc tcttcaactgc gagcaattgc agaagttagc gtaagtccagc ctgggctgtg 21120
gacaatgggc tccaagtgcc ctggctcac cccaggtcgt gcagcctggg aagctgtgag 21180
tgatgggctg gggcaggggc tgtttgcatg atggggggtg caggtgatcc ctgcccagag 21240
gggaagggca accctgggat ttgggtgctca ctgtccaatg tgctttgctt ctgtgtctcc 21300
tctcttcttg aactgaacag tctattcaac aacaaattga ctgacggctg tgcacactcc 21360
atggctaagc tccctgcatg cagggcagaac ttcttggcat tgaggtgagc ccaggttttc 21420
cttattccct ggaaactatt ttttgcccca ttctctgagtc agtctgatct ggtcttggcc 21480
tggcactgcc cacactggct cctgacctcc tgattgaatg cagggacagt gtctcatttt 21540
aagcaggggt tctctaagtc tctgatctcc ccagtaaaact ctggactagc tctgctgagg 21600
acttctgttc ttttgacctt tagcccgtag ggcaagaaag cttttctagg cccctttcct 21660
ttctctgttc taagagtgtc acagcttctt ggggttactg agttccacga tgcatttga 21720
gctcgtctct gtgggggagg catacacagt tacttgccac cccagctgtg gcagcgagtt 21780
gctgcaacac tcccaggagg tcccttcacc actcagagca tgcaaggttt gcagtccatc 21840
tggttctgca tttctgctac tccagtgctc cccagtttca acaggagtct ctctctctcc 21900
tacctgatgc ctttaaattg cccctctagc tggccgctgg gttggcctgg ctctctctc 21960
cttctctctc tctcagatat tcttgctctc tgtgatttgt gaggcagtaa aaaaagacaa 22020
agtaaagaat tgcttccatc tattctttta cctcttgggc tgggtttgtg gatgggagcc 22080
gccattttaa aatggcgggc cacatagctc agtctcggca agggctactg agatcagaac 22140
cacaggtgcc aatttgtaca aaggactcag tctctgctacc actgctgat cctcagact 22200
cacaagcctg gaataggctg tggccagacc tggctggccc atccctgaga agggtgctag 22260
tttcagaaat ggaggtgag tttgtggcca acacagtagt cctccggtat gtgcaggaga 22320
gatgttctaa gacccagtg gatgcctgaa accatggaga gtatcaagcc ctacacatac 22380
catgcttttc ccaataccta cacacctgca ataaagtgtg gtttataaat taggctcagt 22440
aagagagtaa tagcaactca taataaaata gaacaattat aacaatcaat atactataat 22500
aacactatgt gaatgtggac tctctccatc tccctcaaaa tatcttcttg tactgtactc 22560
accctctctc ttgggaagat gtgtggtggt aaaatgcctg tgtgatggga ggaagtgagg 22620
tggatgacgc atgcagcact gtgctctagc gctgggctgc tgttgacctg accacacttc 22680
agaaggagaa tcatctgctc ccagagatcc ctaatctttg agcaacaatg aggtcggcag 22740
ctggatgtca ggagcagacg atcttgatga ttaccaaag ggagcgtata gagcgtggat 22800
gcgctggacg ggggctgat tcacgtcctg ggtgggatgg agctggatgg cacgtgatca 22860
gaatagcatg caatttaaaa tgtatgaatt gtttatctct agaattttcc atttaaatatt 22920
tttgactgac agltgatttc agataactga aaccatagaa ggcgaagctg cggataagca 22980
ggggcaggg attaccgtat atcattgtaa tagagagcac aggtctgga gccagactgc 23040
ccgaggtttg aaccctcatt agctgcgtga cctcaggtca gcccaatgtc tgtgtgcctc 23100
cgtttccctc tctgtagaat ggaggtaata accctggcta cctcagggc tgtagtgatg 23160
agcaagcaag ttaatccaca tgaagggctg caccgtctgg caggggcttt atatagtaag 23220
cgagtggctg aaagatgatg ggtaaatcac acaagcactc agcttgtttc tcttatgtg 23280
agtccggctc tccaagcagg gattcaatgt gccaccatt tattggggaa aagtcctaaa 23340

```

10

20

30

40

```

aggggaagtg ggaaggggag ctgggggagg ctgggaggtg tgtccctgag tgaaggagag 23400
agggaaaggaa ggaaggttga gactggggcac cttggacttc agtgcagtcc taagacatct 23460
gggcaaggot gatgaggagt tcttgaacca aattcaccag gcaggggagc ctgatgtctc 23520
aggcaggggc tggcaagtgc agatgcgagg atgttagatt ttggagcaca gcagctgggg 23580
cccttggcta cctccaagga gctgaggctg gagacctgaa aggcgagttc tccttagctgc 23640
cacacccttt ctccaaggat acaataatat ctgccttata ggattgttgt gagctgagtg 23700
gcttgacgtt ccttgaaaga atgaaagcgt atagttatcc caggaagcct aggggtgcag 23760
gtgagagctc tggggcttct ccgaagctct ccgaggtgtc tggattcagt tgcagcagga 23820
gccttctctg ctgggatctt cccccacccc tagccttggc cctccctctc tccttctctt 23880
ctggaaggct cagtgggccc caccctccc tccagccacc tggacctgcc cagcgtctt 23940
gtgcaacagg taagcctac ctgtagcaac aacagatctg ggaaggctgc agagggcacg 24000
atgggggtct gatcgagggc ggctgagacc agagggaaag gtgtgacct gatcaccct 24060
cgctgtcccg gggaaaccac ctcccaggac agctgcctac tgtgctctcc gcctggaatt 24120
gtcacactgc tgtgcaaca gctcccgtc gcccttctc ctttctggg ggaaaaatgaa 24180
gttgtgggag ccgctgagta aactagacct agcagcgagg gcacctgatg tggctgtgc 24240
ctcccgggca ggtcttcaat gcttctctcc tgtgttctcc tggccagggc acagacggcc 24300
ctcctttctc gcctgcccgt gtgtctctctc agcctctctc gtcttccctt ccaggctggg 24360
gaataactac atcactgccc gggagccca agtgcctgac gagggctcc gaggcaacac 24420
ctcctgacag ttcctggggt aggttggatt actcttcagc caggaggccc cagaggcagc 24540
gggacttttg aggatttagg ggcaggtgaa actcttcagc caggaggccc tgcaccttc 24600
ccagctccag tggggaggac aagccaggga gagagtgggc ggcccttgac tgcaccttc 24660
atccttggtc tatgctgac aaacaggaag tttgggatgt tggggctagg ggaggacagt 24660
gcccacgagc tgggtgacagg aagccctctg atcctcaggg gccctgagg ctgtacttta 24720
gctgcatatt aaaaccacct ggaagcttct aaacactatt gccaggcctc ccacccaga 24780
ctgatgaaat gcaaatatct agtgcaagg cccaggtatc aggagtttta aaaagcttcc 24840
caggggatgt acagccaggg gtgaggacc ctgacctaa aaagagaagg aatbggggaa 24900
ggataggaag gcaccagga taagaggggc tgtgctaggt ccctcggagc tcttctctcc 24960
tgtaggacca tgcctaggcc tgcaggag gggagtacc caacctgcag cccaggggtg 25020
ggcttctctc gtttgctagg caccaggct tgcacctgtg ctgttccag cagcctctct 25080
cctatcctgt catgcccag tbtgaactgg agtccatttg acaagaactg ggagttttag 25140
aacctgggac tbtggaaga gagaataacc ttaggcccta ggtgttccag cccatttcc 25200
aggagggcaa gttgccccca agctcagttt tttgtttgt tttgtttgt ttgagatgta 25260
gtctcactct gttgcccagg ctgaggtgca gtggcacgat cttgctcac tgcacacctc 25320
gcctccttgg tcaagcgat tcacctgcct cagcttctca agtagctggg attataggca 25380
ccccaccaca cgcacagcta atttttgtat ttttagtaga gacagggttt caccatgttg 25440
gcocggctg tcttgaactc ctgatctcag atgatccgcc cgcctcggcc tcccaaagtg 25500
ctgggattac aggtgtgagc caccgcaacc ggcccccaag ctcagtttga gccacaaatg 25560
ggactatgtt gctctagaaa tcaacatctt tccacactg cattagtagc aacagagtct 25620
agaacaaagg aggcacagc cccactgaac tctctctctc ttgaggtcac atctgccaca 25680
tcaggggtat ttacctctt caacacatat ttattagggc acctgtctgg gccaggcgtt 25740
gtgctaaaac ccccaaacgc tgcataatga tacaagtgt tctgtaactt gcttggtttt 25800
tttttttgt tgtttgtttg ttttgtttgt ttttgtttgt tgttttttt tgcctcggca 25860
tatattatag gaattttttt aggtcattat gacctcttta tttacttaat tatctattta 25920
tttattttac taatattttc agaaagggtc tcaactctgt acccaggctg gagtgcagtg 25980
gttgcaatca tagctcattg tagcctttaa ctctgagct caagtgatct tcctacctcg 26040
gcctcctgag tagctgggac tacaggcaca agccacctg cctggccgat atttttatgt 26100
ttttagagag cgggtgtctc ctatgttggc caggctggtc tcaaaactct gggetcaggt 26160
gatcctccct cctttgcctc ccaaagtatt gggattacac aagtgcagca ccttgcctag 26220
cctgacctca tttttcaaag agctgcagag tgttacataa tgtatttaac tggtcacttt 26280
ttgatgacta ttaagtgttt ttcaggtttt ttgttattac agtgtcatal ccttggggca 26340
cagagcagtg ctggcacata gccagagctc aatcgataca tacctaata atgaaagtac 26400
agtggacatc ctaattcagc cattctttgc taacttgtgt acatacctgt ccaggytagg 26460
tccctagaat acagtcaata agtcagaagg tgtgagttgg gatctacct ttggaaaggg 26520
atgttttcaa actacagtga gtcagaggag gatggcccag aagctggggg agttgaaagt 26580
gatggcgtga aggaattagg ggtgttagga agaagcagga gataaagagc tagcttgcag 26640
aagaagtgtt agacttgtta tgggcaggta ctggagggtg gctaaggact tgtgggtggc 26700
agttaccagg aagcgtatct gaactaagtg tcagaaaaag tgtcacaact gtaattact 26760
cttgtcagtg agttcctgtc cttaagggtt agggctgggt agccctctac tattctctaa 26820
gtctgtaatg taaagccact gaaaactctt gggttaagtt tggccatccc acccaaaaga 26880
tggagcgagg tccactttgc tgggaccagg agccccagtg aggccactct gggatlagt 26940
ggtcctgccc ctctggctgg gactgcagag ggaggaggac tgttagttca tgtctagaac 27000

```

10

20

30

40

```

acatatcagg tactcactga cactgtctgt tgactctttt gcccttttca gattctgggg 27060
caacagagtg ggtgacgagg gggcccaggc cctggctgaa gccttgggtg atcaccagag 27120
cttgaggtag ctccaggtaag cttcagagtc tatcctgcag tttctttggg gagatcaggg 27180
gaagaggagg gacctggggc cagttctgaa ggtccttgaa ctttatttct accccacaat 27240
gttaggcaat ggagtaagga aaaaagacca ttggatttca agagaggaca cttgagtcct 27300
tctgggtgac ttggaaatgt cccttgtcct ctcagggttt tgatacagta tctgtaaatt 27360
gaagatattg gctgggatca ggtacatfff atcttaaggg ccaattccaa tccattggta 27420
gtgggtgcc agtgcaccac attaaaaga atctaaaggc tgcaacctgg cttaaagaag 27480
agcactataa tcaattagtg atgtctaaaa aagctaaaaa aaaaaaaaaa gagcactgca 27540
ttcaattagt gatgtctaaa aagggtagaa aaaaaaaaaa aaagaaaaaa gaaagagcac 27600
cgcaatcaat tagtgatgbc tgaaatggag cagaccagga gagcaccacg aattttgccc 27660
tccataggtt agctcatctc tgaggctctt ccctgctctg acatactttt gttccatgat 27720
tacctccagc ctggtgggga acaacattgg cagtgtgggt gcccaagcct tggcactgat 27780
gctggcaaaq aacgtcatgc tagaagaact ctggtgagtt tgggggattc tctgctctgg 27840
ggaagtggat cacaatctct gttgatcccc tggcctcabc cataggagcg gttgtgtgga 27900
cagacaaaag tggatgattg agtgattgac tgattgattg attgtgtttg tctttatatg 27960
tactgagtgq tatgaagctt atagagcctg gtatgtacat gctaaatfff ttatttaata 28020
aaatataatg gtttctggtt ttggtgactg cctccacatg gcataagttg taagagcaca 28080
gactctgtaa tcaagcaggg cgtgatctta ggcaagthaa ataacaatff cagaatctca 28140
agtttcatgt ctgtaaaaatg agggtaagaa tacttccaac cataaaggat ttttgcaaga 28200
attagataaa gtagtgccctg tgaagacctt aatatagtgct ctggcatatt tgtaagtgtc 28260
ccataaatgt taattagaa taatggcagg gtbactacta ctatctactg tgctgctgct 28320
gctgctgctg tacaactac tatagtaact tgactactac tactaataaa gttttgttat 28380
ttbaaagtga ttttgagttc ctaggagcac tgggtattca agtcttaggt cattttggaa 28440
ggtgtaatgg agttttgata gttgaaagag gaaccatgaa tcatgcttat actgttgacc 28500
tgaagcagat tctaagtttc tcatccttta gatgccacta gtatagtttt ctgacatggt 28560
ctgggcagct tcagattatg tcagggagat aaaactactg atgtttgatt ttcccgggaa 28620
gcagaaagcc actgcaacat atgggcattg ccataaacag attttatgga tggaccttg 28680
ctgttgcaag gcttactagc tctactcaag tatgattgat tctatcctga ctggatfff 28740
ccacttgcaa tttcttagta gaggagaacc ttgttatgag agcatcagtt atgattactg 28800
ttaaagaaa aactttaggg aaattaaatt tagcagaact ggtttgaaca tacagcaatt 28860
tatgaattgg gcagacttca gaactgggag tgcctccacc agcaaggtag gcaagcagta 28920
tctatagaca ggaagggaa gtgatgtaca aacagcttg atgtgtgca gctgggcatt 28980
tgctttatat gggcatggtg tgatgaggca tttctcttat atggatatag actgatcagc 29040
tggtagactg tgactgactg aagcctggct gctgtgattg gctaagactt agctgtttgt 29100
tbaagagata tghtgttagg ttgcagtttg ctacatagga actcaaagta cagaggcaqt 29160
ctcaggccaa attttagtta actatatggt aagctgcagg tgacagaata cctccatcta 29220
tagaggttta aacaagggaa gggtttattt ttctctgtat aggcagctgg atgtaggcag 29280
tgtagggttt gtacagtggc tacaagaggc caggaggggt ctcagctctg tctcattctc 29340
ttctgtttcc atcactctta gcctgtaact tcattoacat ggttggttgt ctcatgatca 29400
caggatggct gctccaggtg cagcactact tctgtattcc cggattcogat ctataatacc 29460
aggaaagcca tctgggttct ctcccttaaa aagcattcct ggaagcccca cctgtcagct 29520
tccccttatg tatcaacat gtgtatgtca cttgaccaac ccacttgat gttgtttgac 29580
cagccctggc tgcaatggag agtgggaaat acagtffff caccaagtgc atggctgtcc 29640
aaatgaaatg agacttccat taataaggaa gaaaggaaag atggagatca ggaagctggg 29700
ggatcaggga acttattaca ttgagagccc ttggagtga tctcttgca aatatgtccc 29760
tggaaattgag aatccccaca acgtctttat ctgttctttc tttatccatg agtttgggtt 29820
ttcagatggt ggatttccta tatggggggc atgtgagttc atcatcttcc ataataatg 29880
ttgtatcaac tggatfffft ctcttctctc caccagcctg gaggagaacc atctccagga 29940
tgaaggtgta tghttctctg cagaaggact gaagaaaaat tcaagtttga aaatcctgaa 30000
gtaagggaacc cataagcagg aaacaggaca ataattgctg gcctttggaa ggggcatttc 30060
tgabtaagat ctgggcoct ctccgctggg ctaactcatg tgagggtggc tggtagaaca 30120
gcttgccctg gtctaggtgg acaaggattc cagtgcagat tgtttatctg ggaggtggtc 30180
ccagtaaatt ctgataggag agtgggtgaag tgagatgggg aagtgaaggt aaccaataaa 30240
ggggagttat caagccagtt atcaatgagg gaaattggag ctcaagtactc tggggcactc 30300
ctggagccag tgcagaacac acatggctac ctacccaacc aatgggcaag aaagccatgg 30360
catttatcca ccaaccctct gtccctccta tghtgatgtg cgctcatggg gcactgattc 30420
tccagcactt ccagctcacc ctacccagc tgaacatgct tctgggtca ggagaatggc 30480
ctcaggcaga gagtggcagg tcttctctgc aagcagtgcc tggggaggtg atgtgatggg 30540
gagtaactgt gcctcctoca gtggctgact cagtggcttg ggacttgtgc cacaagaga 30600
tggacagctc aggtgaacat gaaccacct agtgaccatc atgggtttgt cagggtgctc 30660

```

10

20

30

40

```

tctgaggctg atgccaaaat tcttatttca agtagacctc aggaacccca tcagatggct 30720
ccttttgctg gaggaaaatg gcatctgcct aggcaaatgt ggtccttaga aaacgcttgc 30780
cttttagagac agacagacag acagctgcct ctgtgagtgc cagctttgct gccaggctgc 30840
tcccactctg ggcgacactc atttgtgttg ctttcacaag ctaggaagt tccaaatatt 30900
tggagaaaac acttccacta attatttggg tggaaatggg ctgggaagt ggggtgaagc 30960
ccggatgtgt ctgagccaga tgccagcttt gcactgaagg tggcctttg ggaataccaa 31020
gcccattatc aaccaggtgt ggatalggca ggtttgtctt cctccttgt cacagcctta 31080
ctcccactga ctcccattga tgccaggcaa tgaggctggg gttggctcca tgccaccctg 31140
tcactcagct tatttttcag catcctaaac tatatcatcc cccacaaaaa ttgaacttct 31200
gatatacttt ttataaaaaa gagaatgcc tacatctttc tttccagga ttagtttctg 31260
ccaagagttg gttgagagcc caggcttctg ggggtcagtg gctcacacct gtaatcccag 31320
cactttggga ggctgaggcg ggtggatcac ctgaggctggg gagttccata ccagcctgac 31380
caacatggag aaaccccatc tctactaaaa atacaaaatt agccgggctt ggtggcatac 31440
acctgtaatc ccatctactc aggaagctga ggcaggagaa tcacttgaac ctgggaggtg 31500
gaggttgcca tgagccaaga tcacaccatt gcaccctaga ctggacaaga gagaacttc 31560
catctcaaaa aaaaaaaaaa ggatgagaaa aataataatt taaaaaaaaa agtccaggct 31620
ctggaaccag acagcctggg tcttaccctt gctccaccat taccagccag ttcttcttgg 31680
atgagtgcct cagltgcctc aagtgtaaat ggagataatg cctggacctt cattaaggc 31740
catgagcatt cactgagaga atgtagctaa caaaagttag ttgtaggctg gagcaaaagt 31800
aattgtggtt tcagaccatg aactttaaatt tattataact aggctaaaaa acatctttat 31860
taatcaaaat aggaaccatt aaaaatcaaca cttttttgcc aataagaaat aagtttgttt 31920
atlcctgtag cataaaaatt catgcttcgg gattcaacaa actctggaa agcattttct 31980
gcactcctct ggttgtggaa gcatttttcc tgcagaaagt tgtcaagatt cttgaagaaa 32040
tggtagtccg ttggctagag gtcaggtaaa tatggcggat gaggcaaac ttcatagtcc 32100
aattcattca acttttgaag ctttggttgt gtgacatgca gtcgggtgtt tgtcgggag 32160
aattggacc cttctgttga cgaatgcccg ttgacaggtt tgcagtttc agtgcactc 32220
attgacttgc cgagcactct tctcatatgt aatgglttcg cagggattca gaaagctgta 32280
ggggatcaga ctagcagcag accaccagt accatgacct ttttttttgg gtgcgaattt 32340
gcctttggga agtgctttgg agcttcttct cggtcacaacc actgagctag tcatbgccag 32400
ttgtataaaa tccacttttc atgcacgct acaatcagat caagaaatgg ttctgtgtt 32460
ttgtgtagaa taagagaaga tgacacttca aaatgacgat ttcttgggtt ttcaactcagc 32520
tcactgagca cacacttatc gaggtttttc acctttccaa ttgtctcaa atgctgaatg 32580
accatggaat ggtcagatgtt gagttctcaa gtagttgtaa gaaaatcagc ttgatgatt 32640
gctctcaatt ggtcattgtc agcttctgat ggcctgccag tacactctc atctcaagg 32700
ctcttatctc cttcgcaaaa ctcttgaac caccactgca ctatacgtta gttagcagtt 32760
cctgggccc aatggctgct galgttgtga gttgtctccg ctgctttaca accatltttg 32820
aattcaaaata agaaaattgc ttgaatttgc tttttgtcta acatcatttt catagtctaa 32880
aataaatata aaataaacag aaagtattaa gtcattagca aaaaatcata aagtgagaat 32940
tgtgcattaa aatgatgtat agcataacca cttttattta agaatgtatt ccaatatcaa 33000
atggcaaat tcacaatgca aaaaactgca attacttttg caccaatcta atagaagttc 33060
aataaatact gccaattaca attggcattg ccttagggtc aacttgtaag acattcctga 33120
aattgtggga aagggggagg acctggagt gacattattg gaaggcaag ctgtaaccua 33180
aagagcaacc tgggaaacac atgactcctc tgttctgtc cctggccta tctgtctcc 33240
cctcctgtt gtcagctacc tcatatgttc tctaactct ctatcaatta agtactgtat 33300
cccctgaaa atagaaatat tactgctcat tggttatttt ctatcaatta agtactgtat 33360
tagtccgttt tcatgctgat gataaagata tacccaagac tgggcacttt atgaaagaaa 33420
gagttttatt gaacttacag ttccacgtgg ctggggagg ctcaaatca tggctgaagg 33480
tgaaggcac atctcacatg gcagcagaca ggagaagagg gcttgttcag gaaactccc 33540
ctttttaaaa ccatcagata tcatgaaact tatttactgt aatgagaaca ggtgggatt 33600
caattacctt ccactgggtc gctcccacaa cacgtgggaa ttcaagagat ttgggtggg 33660
acacagccaa accatatcaa gtactgtgca agtgttttag gcatgcagag agtgggtgg 33720
cttccagca agcagagtgt ggggaggtaa tgggggactg gtggctgact baatggcca 33780
ggaccatgc cacaaaggaga tggatggtgg atgtgaatag gagcctgctt acaccatca 33840
caatttagat tcttatgctc gatggcacgg gtactctttt aggccattt taccaatgag 33900
gagattggga ctaatttctt cgagatcaaa aaagaagtgg tgtagggtgg atttaaacc 33960
aggatgtcta gcactaaaat gcaggctactt aaccactatc ctaagggagt ggctactta 34020
tttgataaac tcatctagtg aatggaagag agacggttac atttcactga tggactgag 34080
cctttgttga tgagctcatt gggaaatctca gacatgagca ggatgtgtct aagggacagg 34140
tgggcttcag tagactgctt aactcctgca gctcttttaa ctggacagtt tcaagaggaa 34200
aaccaagaat ccttgaagct caccattgta tcttcttttc caggttgtcc aataactgca 34260
tcacctacct aggggcagaa gcctcctgc aggccttga aaggaatgac accatcctgg 34320

```

10

20

30

40

```

aagtctggta aggccccctgg gcaggcctgt ttttagctctc cgaacctcag tttttctatc 34380
tgtaaaatgg ggtgacggga gagaggaatg gcagaatfff gaggatccct tctgattctg 34440
acattcagtg agaatgattc tgcatgtgaa ggcactgatt ctctgtctaa gaagaagtc 34500
tttacctctt taagttaggga gcaatgattt catttttaaa ccttgactat ttattcagca 34560
acttctctgc tctatgagat agtgtaggaa tggggatgtg gttgaagaat gaaaagaaaa 34620
gtcagctccc gccctcctag aaattgcatc tgccttcaca ggtcaaggat attggatcag 34680
accbtctcgg gttctgaaatg gagattacac aggttaggag caggttgcac agtgtttcca 34740
attctctata attaaagoca tagactttca tgtattgaaa aaagcaagaa ttgcattctt 34800
gacagattct ttcattgcct taaaaagaat gactagcctt gggagtctgg gcagctgggt 34860
ccagtggtgt agactttctc tctgctgagc cacagcttca aagatttgtc cttcttgttt 34920
ccagggatct atttctcaga caataagtaa aggccttccc tggcctaagt tgctgtaagt 34980
gaatgctact atatatgttc caggcactgg gctagagact aatattttaa agccaggaaa 35040
tttctctatg aaaatctata tctcagggtt ttctcaaaaag agctgggaac tctggatgcc 35100
cattcatgat tccagtagtt aaccagagta caagaagggc tgagtcttct cagatgggca 35160
aaccactctt ggctgactgc agatccacca agcctattgt cttagaccag gaccctttgg 35220
caactcattc ccataagcct gtgacccttg ctttaaatat gcaggccttg tcttctctca 35280
aaaagcacat caaggctgca gcgaatgcag atatcaaatg atgaagttaa aacaaaagc 35340
tttgctgggc gtggcagctc aacactgtaa tcctagcact ttgggaggct gaggcaggag 35400
gatcacttta ggccagaggt tcaacaccag accttgtctc tcaaaaaata aaaaattcag 35460
ctgggtgctg tgtagttcct agcccacttg gaggctggga tgggaaggatc cctgaacc 35520
aggagtcaa ggctgcagtg ggccatgatt gcatcactgc acaggcgaca gaattagatc 35580
ccatctctta aaaaaataaa aaatttaaa gtgacttcaa aaatctatgc tgbatggag 35640
agatTTTTTc ttctgtatga ttgtgatagc tctgtggcct atgacgtcat caggttctgg 35700
gcaaagtgtg ggttttctgt ttctttgttt ttgaaacctg tgcacagtcc taagaaacat 35760
cacattcttg gtctctggca ccagccaaca tgaggtgagg gcaccagggt ttgctcattg 35820
cattcttgac agattctctt attgccttaa aaagaatcac tggccttggg gagtctgtgg 35880
ctggctgggt gcagtgttgt ggactctctc tgcagagtca tggagccttg ttcagaatgc 35940
ttctgtgact tttgaagtct gccctgggtg gccaaggtta aaaaacagccc tgacttcccc gcaagaaaca 36000
ctgcagctgg gccagagagt cagcccattc caggcatggg tttaaaaagt ggaggctttt 36060
gtttgaagc cctgctctaa ttttgtctc actcaaacct ctgttcaact gatctgcttt 36120
aggctccgag ggaacacttt ctctctagag gaggttgaca agctcggctg cagggaacc 36180
agactcttgc tttgaagtct ccgggaggat gttcgtctca gtttgtttgt gaggcctg 36240
tgagtttggg ccccagaggc tgggtgacat gtgttggcag cctcttcaaa atgagccctg 36300
tctgcctaa ggctgaactt gttttctggg aacaccatag gtcaccttta ttctggcaga 36360
ggagggagca tcagtgcctt ccaggataga cttttcccaa gcctactttt gccattgact 36420
tcttcccaat attcaatccc aggatgtaca aggacagccc ctctccata gtatgggact 36480
ggcctctgct gatcctccca ggcttccgtg tgggtcagtg gggcccatgg atgtgcttgt 36540
taactgagtg ccttttggtg gagaggcccg gcctctcaca aaagaccctt taccactgct 36600
ctgatgaaga ggagtacaca gaacacataa ttcaggaagc agctttcccc atgtctcgac 36660
tcatccatcc aggccattcc ccgtctctgg ttctctcccc cctcctggac tctgtcacac 36720
gctccttctc ctgaggctga aattcagaat attagtgacc tcagcttga tatttcaact 36780
acagcaccct caaccctggc acccagggtg ggaagggcta caccttagcc tgcctcctt 36840
tccggtgttt aagacatttt tgaaggggga cacgtgacag ccgtttgttc cccaagacat 36900
tctaggtttg caagaaaaat atgaccacac tccagctggg atcacatgtg gacttttatt 36960
tccagtgaat tcagttactc ttcagttaag cctttggaaa cagctcgact ttaaaaagct 37020
ccaaatgcag ctttaaaaaa ttaatctggg ccagaatctc aaacggcctc actaggcttc 37080
tggttgatgc ctgtgaactg aactctgaca acagacttct gaaatagacc cacaagaggc 37140
agttccattt catttgtgcc agaatgcttt aggatgtaca gttatggatt gaaagtttac 37200
aggaaaaaaa attaggccgt tccttcaaag caaatgtctt cctggattat tcaaaatgat 37260
gtatgttgaa gcctttgtaa attgtcagat gctgtgcaaa tgtttatttt ttaaacatta 37320
tgatgtgtga aaactggtta atatttatag gtcactttgt ttactgtct taagtttata 37380
ctcttataga caacatggcc gtgaacttta tgctgtaaat aatcagaggg gaataaactg 37440
ttg 37443

```

10

20

30

```

<210> 4
<211> 1315
<212> ADN
<213> Homo sapiens

```

<220>

40

<221> CDS

<222> (117)..(1118)

<400> 4

cgatcagaag caggtcacac agcctgtttc ctgttttcaa acggggaact tagaaagtgg 60

cagcccctcg gcttgctgcc ggagctgaga accaagagct cgaaggggcc atatga cac 119

His
1

tcc tcc cgg acc cct gga cac aca cag ccc tgg aga ctg gag cct tgg 167
Ser Ser Arg Thr Pro Gly His Thr Gln Pro Trp Arg Leu Glu Pro Trp
5 10 15

agc atg gca agt cca gag cac cct ggg agc cct ggc tgc atg gga ccc 215
Ser Met Ala Ser Pro Glu His Pro Gly Ser Pro Gly Cys Met Gly Pro
20 25 30

10

ata acc cag tgc acg gca agg acc cag cag gaa gca cca gcc act ggc 263
Ile Thr Gln Cys Thr Ala Arg Thr Gln Gln Glu Ala Pro Ala Thr Gly
35 40 45

ccc gac ctc cgg cac cca gga cct gac ggg cac tta gac aca cac agt 311
Pro Asp Leu Pro His Pro Gly Pro Asp Gly His Leu Asp Thr His Ser
50 55 60 65

ggc ctg agc tcc aac tcc agc atg acc acg cgg gag ott cag cag tac 359
Gly Leu Ser Ser Asn Ser Ser Met Thr Thr Arg Glu Leu Gln Gln Tyr
70 75 80

20

tgg cag aac cag aaa tgc cgc tgg aag cac gtc aaa ctg ctc ttt gag 407
Trp Gln Asn Gln Lys Cys Arg Trp Lys His Val Lys Leu Leu Phe Glu
85 90 95

att gct tca gct cgc atc gag gag aga aaa gtc tct aag ttt gtg gtg 455
Ile Ala Ser Ala Arg Ile Glu Glu Arg Lys Val Ser Lys Phe Val Val
100 105 110

tac caa atc atc gtc atc cag act ggg agc ttt gac aac aac aag gcc 503
Tyr Gln Ile Ile Val Ile Gln Thr Gly Ser Phe Asp Asn Asn Lys Ala
115 120 125

gtc ctg gaa cgg cgc tat tcc gac ttc gcg aag ctc cag aaa gcg ctg 551
Val Leu Glu Arg Arg Tyr Ser Asp Phe Ala Lys Leu Gln Lys Ala Leu
130 135 140 145

30

ctg aag acg ttc agg gag gag atc gaa gac gtg gag ttt ccc agg aag 599
Leu Lys Thr Phe Arg Glu Glu Ile Glu Asp Val Glu Phe Pro Arg Lys
150 155 160

cac ctg act ggg aac ttc gct gag gag atg atc tgt gag cgt cgg cgc 647
His Leu Thr Gly Asn Phe Ala Glu Glu Met Ile Cys Glu Arg Arg Arg
165 170 175

gcc ctg cag gag tac ctg ggc ctg ctc tac gcc atc cgc tgc gtg cgc 695
Ala Leu Gln Glu Tyr Leu Gly Leu Leu Tyr Ala Ile Arg Cys Val Arg
180 185 190

cgc tcc cgg gag ttc ctg gac ttc ctc acg cgg ccg gag ctg cgc gag 743
Arg Ser Arg Glu Phe Leu Asp Phe Leu Thr Arg Pro Glu Leu Arg Glu
195 200 205

40

gct ttc ggc tgc ctg cgg gcc ggc cag tac ccg cgc gcc ctg gag ctg 791
 Ala Phe Gly Cys Leu Arg Ala Gly Gln Tyr Pro Arg Ala Leu Glu Leu
 210 215 220 225

ctg ctg cgc gtg ctg ccg ctg cag gag aag ctc acc gcc cac tgc cct 839
 Leu Leu Arg Val Leu Pro Leu Gln Glu Lys Leu Thr Ala His Cys Pro
 230 235 240

gcg gcc gcc gtc ccg gcc ctg tgc gcc gtg ctg ctg tgc cac cgc gac 887
 Ala Ala Ala Val Pro Ala Leu Cys Ala Val Leu Leu Cys His Arg Asp
 245 250 255

ctc gac cgc ccc gcc gag gcc ttc cgc gcc gga gag agg gcc ctg cag 935
 Leu Asp Arg Pro Ala Glu Ala Phe Ala Ala Gly Glu Arg Ala Leu Gln
 260 265 270

cgc ctg cag gcc cgg gag ggc cat cgc tac tat gcg cct ctg ctg gac 983
 Arg Leu Gln Ala Arg Glu Gly His Arg Tyr Tyr Ala Pro Leu Leu Asp
 275 280 285

gcc atg gtc cgc ctg gcc tac gcg ctg gcc aag gac ttc gtg act ctg 1031
 Ala Met Val Arg Leu Ala Tyr Ala Leu Gly Lys Asp Phe Val Thr Leu
 290 295 300 305

cag gag agg ctg gag gag agc cag ctc cgg agg ccc acg ccc cga gcc 1079
 Gln Glu Arg Leu Glu Glu Ser Gln Leu Arg Arg Pro Thr Pro Arg Gly
 310 315 320

atc acc ctg aag gag ctc act gtg cga gaa tac ctg cac tgagccggcc 1128
 Ile Thr Leu Lys Glu Leu Thr Val Arg Glu Tyr Leu His
 325 330

tgggaccccg cagggacgct ggagatttgg ggtcaccatg gctcacagtg ggctgtttgg 1188

ggttcttttt ttttatTTTT ccttttcttt tttgttattt gagacagtct tgctctgtca 1248

cccagactga agtgacgtgg ctcaattatg tctcactgca gcctcaaact cctgggcaca 1308

agcaatc 1315

<210> 5
 <211> 334
 <212> PRT
 <213> Homo sapiens

<400> 5
 His Ser Ser Arg Thr Pro Gly His Thr Gln Pro Trp Arg Leu Glu Pro
 1 5 10 15

Trp Ser Met Ala Ser Pro Glu His Pro Gly Ser Pro Gly Cys Met Gly
 20 25 30

Pro Ile Thr Gln Cys Thr Ala Arg Thr Gln Gln Glu Ala Pro Ala Thr
 35 40 45

Gly Pro Asp Leu Pro His Pro Gly Pro Asp Gly His Leu Asp Thr His
 50 55 60

Ser Gly Leu Ser Ser Asn Ser Ser Met Thr Thr Arg Glu Leu Gln Gln

10

20

30

40

<220>
<221> exon
<222> (3812)..(3950)

<220>
<221> exon
<222> (5426)..(5577)

<220>
<221> exon
<222> (7273)..(8135)

<400> 6

cgatcagaag	caggtcacac	agcctgtttc	ctgttttcaa	acggggaact	tagaaagtgg	60	
cagccccctcg	gcttgtcgcc	ggagctgaga	accaagagct	cgaaggggcc	atagacact	120	
ccctcccgac	ccctggacac	acacagccct	ggagactgga	ggtcagtatt	tgatcccaag	180	10
ctcagctgtc	ctctgcctgc	tgtggcctga	gtccccttct	cctggggccc	tgctggcac	240	
ctgctggggg	caggggtggga	gggggaagag	ttagtacag	cogctgtgtc	tggagctctc	300	
cttagcacac	tgaggcagag	gaaggacag	ctcctggacc	ttccatcac	tccattcctt	360	
ttgaaatgct	aggcgttgt	acaacccatc	ttgggcctgg	agaataagtc	accacacctg	420	
tgttttctca	aagaacagtg	tcaggaacc	cctgcctcag	cacagcctta	gaggactcat	480	
ggaaaaatgca	gaatccaggc	ctgttcaatg	gcaccttctt	atgttagcag	ccaggaaacc	540	
tgctcttggg	caagccccctg	ggatcccacc	cccaccccac	caggggattc	ttacacacac	600	
tgggttggga	gcccctggct	ttggcaaggg	ttctcaggtg	agcgtccagt	tgttggaggg	660	
taccacacct	ttcccacaaga	gaggcagcca	cacatccaac	atcctgggat	ctctgtctcc	720	
cagcgtgggc	catgtgcttt	atttcacccc	ctagaggctc	atcccccatg	aaaagtcctc	780	
cgcaggccct	cagaaagata	gtgtggcctc	tgtgtgcca	gcagaagaag	gactggactt	840	
ggcagtcagc	tcttggagag	gggttgggta	ggacacctgg	ggacaggagg	aggagaatga	900	
ctgtctgtgc	acacacggct	ggaaggtaca	ggaggctggg	aaactgtctt	gtcccctggg	960	
ccaactacag	gccccaggg	caacagcaac	aacactttta	gtattttgtt	ataaagtcaa	1020	
gaaatctttg	ctacagaggg	tgaggagagg	gaaggaaagg	gcatggaac	cgtctatgtg	1080	20
getatcccca	gagagctttt	agagtacag	gattgctttc	ccatttcaca	gatgaggaaa	1140	
ctgaggcctg	gagagggatg	ggaagctacc	caaggcccca	tgatacaccc	agtgcacaac	1200	
tctttccttc	ccctcctctt	ttaaatgggt	gattcccatt	gaaacctgta	agagacaacc	1260	
ataagggagc	tgactgtggc	tgctgaattt	gattttattc	taaggcctgg	ttttataatc	1320	
agctttctca	gtctttactg	gagtgtaacg	cogaggcatc	atctctaggg	tcttacaggg	1380	
tctctgggcc	aatagtggcc	tgcttctgac	ctggagccag	ctgcctggtc	atgaaagcag	1440	
atctgcaaag	gctggggccc	ctgaggccaa	ggccactcgc	catcacccat	tttacagaag	1500	
tgctgagcat	aggagtggcc	tgggccccca	agaatcccag	ccaccaagaa	tcaogtaaac	1560	
catccactgt	ctcacttagg	caccagtcag	aatgtaggga	acccaccctt	agtcatecat	1620	
catcttatca	acaggacggg	gcttgtagcc	acatttatca	ggtagggaaa	ctgaagccta	1680	
gagatattaa	agcacttgct	taaggacaca	cggttggtca	ggatggaagg	cgatgtctcc	1740	
tgactccctg	acaggcacia	gagacaagcg	agaggtgccc	gtgacggcat	gctcaagaac	1800	
gtgcagccct	gggcccagcca	ggcccctgct	ccgtgcctct	gtttgcccat	ctgtaaaagg	1860	
tgaggttgga	tcgagggctc	ctgagggccg	cccactggat	ggctgtgcag	agccaaacgg	1920	
agaaggcccc	agggttctct	tcaccggaca	cagcaagcac	ttcccctga	agtgcaggct	1980	30
ccaggcccca	gctgacctcc	cctctcccag	gccagcggct	ctcacccctg	gagcaaggga	2040	
cagggcctgg	ctgtgctcag	ggacatgcat	gactcccgcc	cccatctgtg	ctcagggggt	2100	
gccagggagg	cactggctct	atctttctct	aggccgtagt	cagcccaggg	gttcagacca	2160	
agagcccaga	atccaacaga	tcagagttca	agtcccagct	ctacctctat	gttccactgg	2220	
cagcttctct	aggtcatttg	caccttctct	gtcttgaatt	tccatgccta	accagtatac	2280	
cagctactcc	ctcagccga	tctaattggt	taattgtccc	ttctcttaag	ttgtctcaa	2340	
catttgtaat	tctattccaa	tccaccttaa	tttagtcatt	tatttcacaa	atatttctgg	2400	
aaacatctag	cacttaacag	acactaaaag	cgggggtact	acacagtccc	tgggatggac	2460	
agggccctga	gctgaggett	cagagtctgc	ctgactgaat	cctcacccca	gccttgtgaa	2520	
cgtgggttct	gttatttatcc	ccaatttata	ggaaacagaa	gcacagagaa	gttgagtcac	2580	
ttgccagcta	ccaggtcctc	ccttccactt	atccgggtca	cagacagagt	tattatgtaa	2640	
accagatccc	agctgcctgt	tctccctccc	tgagtaaggt	ggagagaatt	ctgaagtcag	2700	
ccagcctggg	gtctgtatcc	tgccaccac	tcaccagctc	ctcatctttg	gcaactctaa	2760	
gtctcagttc	ccttatcata	aaaggagat	gtaaacagtc	ctgagtgacg	acagtgttca	2820	
ggttagtgca	agagtgtgtg	ctgggtgtga	agtgcacagc	cagcacgtca	caagcactgg	2880	40

```

agacaaattc agctttgctt gttgpcgaca ctcaccagct gcgtgacttt agacctcagt 2940
tttctcatct gttatgtggt ggtaaatgata gacttttgtg agcattaaac tagattaggg 3000
gctatggaga acctagatgg gttatgaagtg ggtataataa gctatcagtt aattttgctg 3060
atagatagat tattgattga ttgatcgata gaagattcat accagtatct acctgctctg 3120
aacactgacc tttctttttt tcttttttgag atggctctgt tctgtcacc agactggagt 3180
gcagtggcat catcatagct cactgcagcc tcagctctct gggcttaagg gatcctcctg 3240
ttctcagctc ccaagtagct gggaccacag gcgtgcatcc tggataattt ttttttattt 3300
ttcttagaga cggggtctca ctacattggc caggctggtc tcaaattcct gggctcaagt 3360
gatccttcta acccagcctc ccaaagcgct gggattacag gcatgagtgg ccatgttcaa 3420
cttgaacact gagaacttcat tcgcatgtgt aacataaaac tgagtatcta gacaagccag 3480
catctttctt tcaagtaatc actaaagcca atacttttac ttgaaatcat ctcatttaaa 3540
actctgagca atacgtaagg atcacctcaa taacatattg atcatcgcaa taggtgaagg 3600
gtcttctctg ccttggagta acctgccag caaaggggca gaccagatt tgggatctgg 3660
cagctgggag agtggggaag gttgagccgt ggggcccttg tcattccctc tgctgccag 3720
gagggggcat gacacagctc ctaggcacc caggagccac cgggaacccc aactggagtg 3780
ggtcctcact gttctctttt tcctctggca gccttggagc atggcaagt cagagcacc 3840
tgggagcctt gctgcatgg gaccataaac ccagtgcaag gcaaggacc agcaggaagc 3900
accagccact ggcocccgacc tcccgcaccc aggacctgac gggcacttag gtgggcttga 3960
ggcttgagac tcggtctggg ggagagggtc gaagacattc aaagtacaaa tbtgggtcac 4020
tttgggggat gcagcaagag gcccgggcag ctcttgtaac ttgggttate ccaaacaga 4080
cactgagaca cagatctagi gcaagctgtt tatccgggag acggtcctag gagtcatggc 4140
agccgagtgg cagtctgga aaagtgggag aggccagggc aggcacatcag tgaacagata 4200
ggcacggtag gtggctgaag ctcaacccca gcgggggtct tctgggagac cctggaacat 4260
atctctgggt tgcctatcc taggggtgag gaagccgggc tgttatctac cagtcctgcc 4320
ctgcatagga gaagggacgc tcctgggctt gctgctatgg ccctagaaa cctcagga 4380
agccagtgcc agttctgga ccaagagggc ccaagagggc acggtccagc ctggggcatg 4440
gacagcatct gctgtagtgc catctcctgg aacagatctt ttcttaccag ccttcgagat 4500
gcctatttca atacctgctc tgttctggc cctatgcagg gcaactggaga aacagaaaca 4560
ggaagaaatc aaacactgca ctagtctctga ggtttgtag agaaacagat cagtggagaa 4620
cagttacacg tgccacgaga aataaataaa taaaatgaaa aacctgtagg aacaaggtgg 4680
gaagctctta ctctaattgc aagggcatt tgcagtgatg tgggggctgg gtcttgaagg 4740
gtagactgga aaagggctgg gaccatgccc ctttgcaata aaatgcacaa ttatttgtgc 4800
ttcttaagaa cctcagagtg gcgcagggct caagtggggt taaagaaaca ctgtgtctc 4860
tttccaggcg tggaaataga ggggtggatg caaggcagag cagtgcacgt ccgagaagag 4920
cccgcatgtt gggcagttag atgagaaggt taggaagggc cagcccgctg aggtggaac 4980
ataacatcct cctcactgcc tcccctgccc actgatgtgt gctcaaggag tctgggcaac 5040
agtcacgaag tcagggctgc agggagcaca gaaacacaca agccaccgtc tctgcttgtc 5100
cagagcaggg atttcacctt ggccaatcta cagaccagaa gtggacgatg caaagtgcc 5160
gcaccgcatc ccaaagcttt gaaaccattt ggggtgatg ggctatttgg gattgtcgtt 5220
ggttaggtgg attctgccc gctgggcaca gaggctctgtc tgatgcccc attgggcta 5280
taaattggcg ggtgggagag agggatattc aatactcttc aggagtctg atatgccatc 5340
tcagatagac ccagccatct ccccaagccc atgcctcggg agtgcactga cagggtgcag 5400
atccttaagg gtgttgcct tcagacaca cacagtggcc tgagctcca ctccagcatg 5460
accagcggg agcttcagca gtactggcag aaccagaaat gccgctggaa gcacgtcaa 5520
ctgctctttg agatcgctc agctcgcatc gaggagagaa agtctctaa gtttgggta 5580
agcagagatt gggaaatggt ggagcctctt tcaactctgt tcttctctgg cctgaataa 5640
gtctttaga gcctcaggtt tcccactat gaaatgggtc aacacactaa ctacagctt 5700
tcttctggg aaaaaggcca aagagcaaga tttcaggctc agcaccgtct agggctctgt 5760
aggattcgaa ccatataagt catatttctt ggtccaaga aggaaatagc ccagtttaat 5820
cccactctat caggtgtcag tcacctgtgt ccttcttca ccaatttgc catatcactg 5880
tatctgttct aattattatt acttattttt ttctttaa at tggatcactt ttaaaaaa 5940
tgaagcaca ttatttcaaa gagaatacc ttaaattgaa aaccaatct acatggcaca 6000
aagcaaaagt aacatactag aaaagtcgat acaaggaaag tcaatacaag gaaagctatg 6060
tgctgttatt aaattctagc tggttactgt ggcttcggga aagccctgtg cctgggagct 6120
gctcctctcc ctgttagaat ggaattttag ctgtgttaa gggatgttaa agattgccta 6180
agagccacac ttcacctctc tccttcaact acctgggacc gggataaata acatagctac 6240
cactgaatgc caatggcatg ccgggcacag ctccatgtgg tttcagtgc ttaactcatt 6300
taatcctcac tgggtgaggt aggcactatg cctatccttg ttttatgaat gagaaaagt 6360
agactcggag aggttaaat actcatctaa aaccacacag ctagaccatg gtagggctat 6420
aattacaacc catgcaatct ggctctggag tcagatgat gggttataat tgcccttaat 6480
atataattgc ccgtaatcag gattctcttg aaagatgatt gaaaaggatt gattttctta 6540

```

10

20

30

40

```

ccatataacg gcatcaccag tgtacctaaa tgatgttata ttgtacgtaa aactaattcc 6600
caagtgtgaa acatttggaa aacacagcat ctcagttcag aaaacagagg cccagtttta 6660
gcaagtaaaag ccaagagggga cccagcagc ctgcagggca ggacctctg cctttctcc 6720
tcccagatgt ccccaccttg ctgtgttgtt gtccagggt tgactcagct gatgccaata 6780
gcaatttaaa acagaattgg gccaggtgca gtggctcatg cctgtaatcc cagcactttg 6840
ggaggccccag gtagggaggat cgcttgagcc caggagttgg agaccagcct gggcaacaca 6900
gccagacccc atcttttaaa aagaatcaaa aaatctgcca ggtagtgggt gtgctgtag 6960
tcccagctac tcaggaggct caggtyggca ggtcaattga gccataagt tcaaggttgc 7020
agtggaggat gatcgatca ctgtactcca gcctgggtaa cagtgcgaga ccctgtctct 7080
aaaaataaat aaataaataa ataaataaat aaataaacia acaacaacaa aaacaacaa 7140
tcaattgcat ataaggatcg ccgttttca gggcatgctt tacaccggcc tggttaactt 7200
tactctgggt gtgctccgtc cgcgcagcc cccgcggga ggtggccaca gctctctctg 7260
gttgcccccct aggtgtacca aatcatcgtc atccagactg ggagctttga caacaacaag 7320
gocgtcctgg aacggcgcta ttccgacttc gcgaagctcc agaaagcgtc gctgaagacg 7380
ttcagggagg agatcgaaga cgtggagttt cccaggaagc acctgactgg gaacttcgct 7440
gaggagatga tctgtgagcg tcggcgcgcc ctgcaggagt acctgggctt gctctacgcc 7500
atccgctgcg tgcgcccgtc ccgggagttc ctggacttcc tcacgcccgc ggagctgccc 7560
gaggttttcg gctgctgccc ggccggccag taccgcccgc ccctggagct gctgctgccc 7620
gtgctgcccg tgcaggagaa gctcaccgcc cactgccctg cggcccgcgt cccggcccctg 7680
tgcgcccgtc tgcgtgtcca ccgcaacctc gaccgcccgc ccgagycctt cgcggcccga 7740
gagagggccc tgcagcgcct gcaggcccgg gagggccatc gctactatgc gcctctgctg 7800
gacgcccattg tccgcccgtc ctacgcgctg ggcaaggact tcgtgactct gcaggagagg 7860
ctggaggaga gccagctccg gaggcccacg ccccgaggca tcacctgaa ggagctcact 7920
gtgagagaat acctgcactg agcccgcctg ggaccccgca ggagcgtgg agatttgggg 7980
tcaccatggc tcacagtggg ctgtttgggg ttcttttttt ttatttttcc ttttcttttt 8040
tgttatttga gacagtcttg ctctgtcacc cagactgaag tgcaagtggc caattatgtc 8100
tcaactgcag ctcaaacctc tgggcacaa gcaatc 8135

```

10

```

<210> 7
<211> 16
<212> ADN
<213> Homo sapiens

```

20

```

<400> 7
ctgggtgcca ttgctc 16

```

```

<210> 8
<211> 16
<212> ADN
<213> Homo sapiens

```

```

<400> 8
ccaggcccca tgacag 16

```

30

```

<210> 9
<211> 25
<212> ADN
<213> Homo sapiens

```

```

<400> 9
tgggtcccggc ccaatcccaa tgctt 25

```

```

<210> 10
<211> 28
<212> ADN
<213> Homo sapiens

```

<400> 10 ttcctcatgt ataaattggg tgtggcca	28	
<210> 11 <211> 25 <212> ADN <213> Homo sapiens		
<400> 11 acagagtgag gaccccatct ctatc	25	
<210> 12 <211> 25 <212> ADN <213> Homo sapiens		10
<400> 12 tccaactgct gggattacag gcaca	25	
<210> 13 <211> 22 <212> ADN <213> Homo sapiens		
<400> 13 agtccccgag accagggcaa ac	22	
<210> 14 <211> 23 <212> ADN <213> Homo sapiens		20
<400> 14 tccatttctg cagtacacat gca	23	
<210> 15 <211> 20 <212> ADN <213> Homo sapiens		
<400> 15 ctctcccat agaagcatc	20	30
<210> 16 <211> 20 <212> ADN <213> Homo sapiens		
<400> 16 ggatagagac gttctcttaa	20	
<210> 17 <211> 20 <212> ADN		40

<213> Homo sapiens		
<400> 17 caggctgaat gacagaacaa	20	
<210> 18 <211> 20 <212> ADN <213> Homo sapiens		
<400> 18 attgaaaaca actccgtcca	20	
<210> 19 <211> 25 <212> ADN <213> Homo sapiens		10
<400> 19 atactcactt ttagacagtt caggg	25	
<210> 20 <211> 21 <212> ADN <213> Homo sapiens		
<400> 20 ggctcagttc ctaaccagtt c	21	20
<210> 21 <211> 20 <212> ADN <213> Homo sapiens		
<400> 21 agtcagtctg tccagaggtg	20	
<210> 22 <211> 20 <212> ADN <213> Homo sapiens		30
<400> 22 tgaatcttac atcccatccc	20	
<210> 23 <211> 17 <212> ADN <213> Homo sapiens		
<400> 23 gatcttccca aagcgcc	17	
<210> 24		40

<211> 17		
<212> ADN		
<213> Homo sapiens		
<400> 24		
tcccgtcagc caagcta	17	
<210> 25		
<211> 20		
<212> ADN		
<213> Homo sapiens		
<400> 25		
aagcttgtat ctttctcagg	20	10
<210> 26		
<211> 20		
<212> ADN		
<213> Homo sapiens		
<400> 26		
atctaccttg gctgtcattg	20	
<210> 27		
<211> 20		
<212> ADN		
<213> Homo sapiens		20
<400> 27		
cctccataat catgtgagcc	20	
<210> 28		
<211> 20		
<212> ADN		
<213> Homo sapiens		
<400> 28		
aatctcccca actcaagacc	20	
<210> 29		
<211> 20		30
<212> ADN		
<213> Homo sapiens		
<400> 29		
ggatgcctgc tctaaatacc	20	
<210> 30		
<211> 19		
<212> ADN		
<213> Homo sapiens		
<400> 30		
cccaggggtc aaacttaat	19	

<210> 31		
<211> 21		
<212> ADN		
<213> Homo sapiens		
<400> 31		
ggtttgaaag tatchccagg g	21	
<210> 32		
<211> 21		
<212> ADN		
<213> Homo sapiens		
<400> 32		
ggtttgaaag tatchccagg g	21	10
<210> 33		
<211> 20		
<212> ADN		
<213> Homo sapiens		
<400> 33		
gtgcatgtgt tcgtatcaac	20	
<210> 34		
<211> 20		
<212> ADN		
<213> Homo sapiens		20
<400> 34		
tcatchccaa aggagtttct	20	
<210> 35		
<211> 18		
<212> ADN		
<213> Homo sapiens		
<400> 35		
aaagccaacc ttgettca	18	
<210> 36		
<211> 20		
<212> ADN		
<213> Homo sapiens		30
<400> 36		
tcttgaaac aggtaagtc	20	
<210> 37		
<211> 18		
<212> ADN		
<213> Homo sapiens		
<400> 37		

attgccctca agaacagc	18	
<210> 38		
<211> 17		
<212> ADN		
<213> Homo sapiens		
<400> 38		
gtgctatgcc atcccag	17	
<210> 39		
<211> 20		
<212> ADN		10
<213> Homo sapiens		
<400> 39		
ccacaccagc gtttttctaa	20	
<210> 40		
<211> 24		
<212> ADN		
<213> Homo sapiens		
<400> 40		
cacactttac acacacctat accc	24	
<210> 41		20
<211> 22		
<212> ADN		
<213> Homo sapiens		
<400> 41		
aagccatatt aggtctgtcc at	22	
<210> 42		
<211> 19		
<212> ADN		
<213> Homo sapiens		
<400> 42		
gcttgggtta aatgcgtgt	19	30
<210> 43		
<211> 20		
<212> ADN		
<213> Homo sapiens		
<400> 43		
agcagtttgg gtaaacattg	20	
<210> 44		
<211> 20		
<212> ADN		
<213> Homo sapiens		40

<400> 44 aaatatgcct tctggaggtg	20	
<210> 45 <211> 20 <212> ADN <213> Homo sapiens		
<400> 45 ggaggatcag gggagtttat	20	
<210> 46 <211> 24 <212> ADN <213> Homo sapiens		10
<400> 46 caaagtaaat gaatgtctac tgcc	24	
<210> 47 <211> 23 <212> ADN <213> Homo sapiens		
<400> 47 ccaactctgt agtttcaaag agc	23	
<210> 48 <211> 20 <212> ADN <213> Homo sapiens		20
<400> 48 tcacagccta cttgcttggt	20	
<210> 49 <211> 25 <212> ADN <213> Homo sapiens		
<400> 49 gacagcctca aatgaaatat aacac	25	30
<210> 50 <211> 25 <212> ADN <213> Homo sapiens		
<400> 50 gctctcagct agggtagttg tttat	25	
<210> 51 <211> 25		

<212> ADN		
<213> Homo sapiens		
<400> 51		
atttttaagg aatgtaaagn acaca	25	
<210> 52		
<211> 20		
<212> ADN		
<213> Homo sapiens		
<400> 52		
gaccaggagt cagtaaaagg	20	10
<210> 53		
<211> 20		
<212> ADN		
<213> Homo sapiens		
<400> 53		
gtccaaaaca ccaccctcta	20	
<210> 54		
<211> 24		
<212> ADN		
<213> Homo sapiens		
<400> 54		
gaagtagatc agtcatcttg ctgc	24	20
<210> 55		
<211> 19		
<212> ADN		
<213> Homo sapiens		
<400> 55		
tcctctgggg gattcactc	19	
<210> 56		
<211> 20		
<212> ADN		
<213> Homo sapiens		30
<400> 56		
gggacatcac caagcacaag	20	
<210> 57		
<211> 25		
<212> ADN		
<213> Homo sapiens		
<400> 57		
caggaaaata aatctaacac acata	25	

<210> 58		
<211> 20		
<212> ADN		
<213> Homo sapiens		
<400> 58		
cctgtgggca ctgataaata	20	
<210> 59		
<211> 19		
<212> ADN		
<213> Homo sapiens		
<400> 59		10
cccagcccc atctcaccg	19	
<210> 60		
<211> 19		
<212> ADN		
<213> Homo sapiens		
<400> 60		
cccagcccc atctcacca	19	
<210> 61		
<211> 19		
<212> ADN		
<213> Homo sapiens		20
<400> 61		
ctgctggagga ggctgctgg	19	
<210> 62		
<211> 19		
<212> ADN		
<213> Homo sapiens		
<400> 62		
tcactcccac caccctttc	19	
<210> 63		30
<211> 20		
<212> ADN		
<213> Homo sapiens		
<400> 63		
agaagtttag tgtggcgtgg	20	
<210> 64		
<211> 17		
<212> ADN		
<213> Homo sapiens		
<400> 64		
gccatctccc caagccc	17	40

<210> 65
 <211> 18
 <212> ADN
 <213> Homo sapiens

<400> 65
 tcgatgcgag ctgaagcg 18

<210> 66
 <211> 18
 <212> ADN
 <213> Homo sapiens

<400> 66 10
 tcgatgcgag ctgaagca 18

<210> 67
 <211> 20
 <212> ADN
 <213> Homo sapiens

<400> 67
 tgaatgttaa agggctctgg 20

<210> 68
 <211> 19
 <212> ADN
 <213> Homo sapiens

<400> 68 20
 ttggttctca gctccggcg 19

<210> 69
 <211> 19
 <212> ADN
 <213> Homo sapiens

<400> 69
 ttggttctca gctccggca 19

<210> 70
 <211> 19
 <212> ADN
 <213> Homo sapiens

<400> 70 30
 agaaaccggg ctggctgtg 19

<210> 71
 <211> 21
 <212> ADN
 <213> Homo sapiens

<400> 71 gcattgcctt ttgatctcta c	21	
<210> 72 <211> 18 <212> ADN <213> Homo sapiens		
<400> 72 tgggctcttc tgcgggga	18	
<210> 73 <211> 18 <212> ADN <213> Homo sapiens		10
<400> 73 tgggctcttc tgcggggg	18	
<210> 74 <211> 20 <212> ADN <213> Homo sapiens		
<400> 74 tgcctcttct tctgccttcc	20	20
<210> 75 <211> 22 <212> ADN <213> Homo sapiens		
<400> 75 cgagctgtac ctgaggaagc gt	22	
<210> 76 <211> 24 <212> ADN <213> Homo sapiens		
<400> 76 cctgagctgt acctgaggaa gcgc	24	30
<210> 77 <211> 20 <212> ADN <213> Homo sapiens		
<400> 77 catcatgagc ccgggtggc	20	
<210> 78 <211> 23 <212> ADN		40

<213> Homo sapiens		
<400> 78 tttctcttgg cttcctggtg cgt	23	
<210> 79 <211> 25 <212> ADN <213> Homo sapiens		
<400> 79 accttctctt ggcttctctg tgcgg	25	
<210> 80 <211> 26 <212> ADN <213> Homo sapiens		10
<400> 80 gccaaagggtg tcgtgccagg gctcca	26	
<210> 81 <211> 20 <212> ADN <213> Homo sapiens		
<400> 81 atctgagaag gccctgctct	20	20
<210> 82 <211> 20 <212> ADN <213> Homo sapiens		
<400> 82 atctgagaag gccctgctcc	20	
<210> 83 <211> 19 <212> ADN <213> Homo sapiens		30
<400> 83 cccacactta gccttgatg	19	
<210> 84 <211> 19 <212> ADN <213> Homo sapiens		
<400> 84 atgagtttagc ccagcggag	19	
<210> 85		40

<211> 19		
<212> ADN		
<213> Homo sapiens		
<400> 85		
attgagagcc cttggagtg	19	
<210> 86		
<211> 19		
<212> ADN		
<213> Homo sapiens		
<400> 86		
tgatttcgta agacaagtg	19	10
<210> 87		
<211> 20		
<212> ADN		
<213> Homo sapiens		
<400> 87		
agcaaattct aggagttatg	20	
<210> 88		
<211> 19		
<212> ADN		
<213> Homo sapiens		20
<400> 88		
agctgagatg tccggatcg	19	
<210> 89		
<211> 18		
<212> ADN		
<213> Homo sapiens		
<400> 89		
agctgagatt ccggatca	18	
<210> 90		
<211> 20		30
<212> ADN		
<213> Homo sapiens		
<400> 90		
gtcctcttaa cttcccttcc	20	

<211> 19		
<212> ADN		
<213> Homo sapiens		
<400> 85		
attgagagcc cttggagtg	19	
<210> 86		
<211> 19		
<212> ADN		
<213> Homo sapiens		
<400> 86		
tgatttcgta agacaagtg	19	10
<210> 87		
<211> 20		
<212> ADN		
<213> Homo sapiens		
<400> 87		
agcaaattct aggagttatg	20	
<210> 88		
<211> 19		
<212> ADN		
<213> Homo sapiens		20
<400> 88		
agctgagatg tccggatcg	19	
<210> 89		
<211> 18		
<212> ADN		
<213> Homo sapiens		
<400> 89		
agctgagatt ccggatca	18	
<210> 90		
<211> 20		30
<212> ADN		
<213> Homo sapiens		
<400> 90		
gtcctcttaa cttcccttcc	20	

【 図 1 】

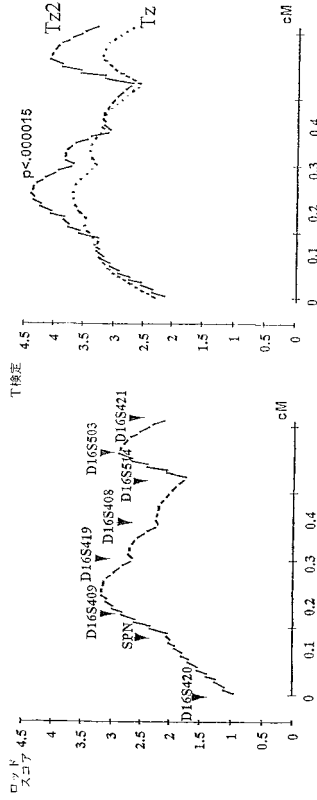


FIG.1

【 図 2 】

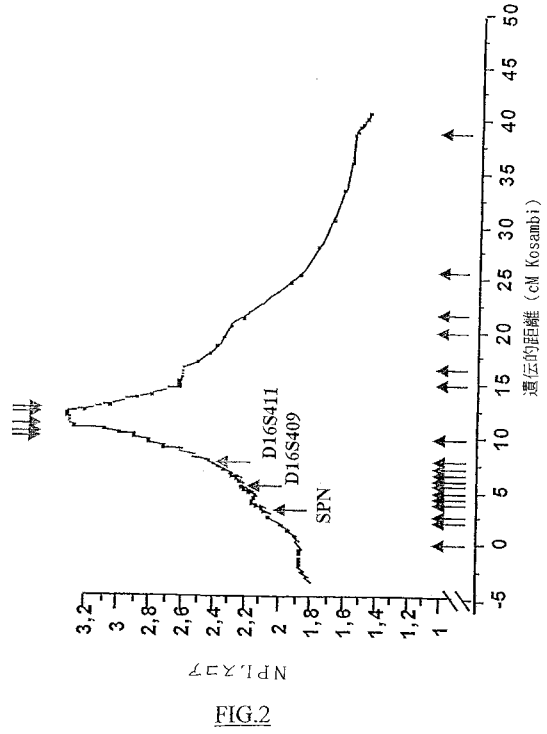


FIG.2

【 図 3 】

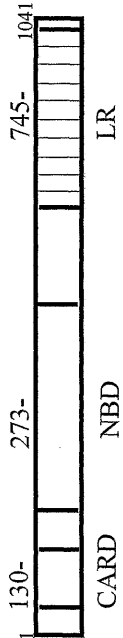


FIG.3

【 図 4 】

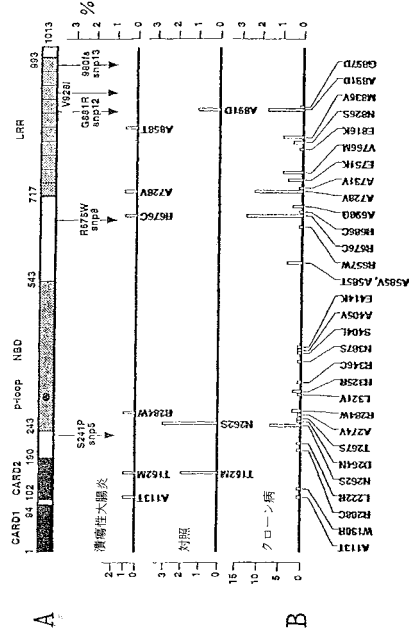


FIG.4

フロントページの続き

(51)Int.Cl.		F I	
<i>C 1 2 N</i>	<i>5/10</i>	<i>(2006.01)</i>	<i>C 1 2 N</i> 5/00 1 0 1
<i>C 1 2 P</i>	<i>21/02</i>	<i>(2006.01)</i>	<i>C 1 2 P</i> 21/02 C
<i>C 0 7 K</i>	<i>16/18</i>	<i>(2006.01)</i>	<i>C 0 7 K</i> 16/18
<i>C 1 2 Q</i>	<i>1/68</i>	<i>(2006.01)</i>	<i>C 1 2 Q</i> 1/68 A
<i>C 1 2 M</i>	<i>1/00</i>	<i>(2006.01)</i>	<i>C 1 2 N</i> 15/00 F
<i>G 0 1 N</i>	<i>33/15</i>	<i>(2006.01)</i>	<i>C 1 2 M</i> 1/00 A
<i>G 0 1 N</i>	<i>33/50</i>	<i>(2006.01)</i>	<i>G 0 1 N</i> 33/15 Z
<i>G 0 1 N</i>	<i>33/53</i>	<i>(2006.01)</i>	<i>G 0 1 N</i> 33/50 Z
<i>G 0 1 N</i>	<i>37/00</i>	<i>(2006.01)</i>	<i>G 0 1 N</i> 33/53 D
<i>C 1 2 P</i>	<i>21/08</i>	<i>(2006.01)</i>	<i>G 0 1 N</i> 33/53 M
			<i>G 0 1 N</i> 37/00 1 0 2
			<i>C 1 2 P</i> 21/08

- (72)発明者 ザウアリ、モハムド
フランス国、9 2 2 2 0 バニュー、リユー・ベルトレ・アルブレヒト、4
- (72)発明者 ルサージュ、シュザンヌ
フランス国、7 8 7 0 0 コンフラン - サン - トノリーヌ、アレ・ドゥ・ラ・ロカド、2
- (72)発明者 シャメラール、マティアス
フランス国、3 7 3 0 0 ジュ - レ - トゥール、リユー・デ・エキュレイユ、3

審査官 神谷 昌男

- (56)参考文献 特表2 0 0 4 - 5 0 7 2 1 5 (J P , A)
J.Biol.Chem.,Feb.2001,Vol.276,No.7,p.4812-4818
Nature,May 2001,Vol.411,p.599-603

- (58)調査した分野(Int.Cl. , D B 名)
C12N15/00-15/90
GenBank/EMBL/DDBJ/GeneSeq
UniProt/GeneSeq

专利名称(译)	参与炎症肠病的基因及其用途		
公开(公告)号	JP5150997B2	公开(公告)日	2013-02-27
申请号	JP2001571753	申请日	2001-03-27
[标]申请(专利权)人(译)	FOND让·多塞CEPH		
申请(专利权)人(译)	方达西昂让德晒 - 约瑟夫		
当前申请(专利权)人(译)	方达西昂让德晒 - 约瑟夫		
[标]发明人	ユーゴジャンピエール トーマジル ザウアリモハムド ルサージュシュザンヌ シャメラールマティアス		
发明人	ユーゴ、ジャン-ピエール トーマ、ジル ザウアリ、モハムド ルサージュ、シュザンヌ シャメラール、マティアス		
IPC分类号	C12N15/09 C07K14/47 C12N1/15 C12N1/19 C12N1/21 C12N5/10 C12P21/02 C07K16/18 C12Q1/68 C12M1/00 G01N33/15 G01N33/50 G01N33/53 G01N37/00 C12P21/08 A01K67/027 A61K31/7088 A61K35/12 A61K35/64 A61K35/74 A61K35/76 A61K36/06 A61K38/00 A61K39/395 A61K48/00 A61P1 /00 A61P1/04 A61P29/00 A61P35/00 A61P37/00 C12N15/12 C12Q1/6883 C12Q1/6886		
CPC分类号	A61K38/00 A61P1/00 A61P1/04 A61P29/00 C07K14/47 C07K14/4747 C12Q1/6883 C12Q1/6886 C12Q2600/156 C12Q2600/158		
FI分类号	C12N15/00.ZNA.A C07K14/47 C12N1/15 C12N1/19 C12N1/21 C12N5/00.101 C12P21/02.C C07K16 /18 C12Q1/68.A C12N15/00.F C12M1/00.A G01N33/15.Z G01N33/50.Z G01N33/53.D G01N33/53.M G01N37/00.102 C12P21/08		
审查员(译)	神谷正夫		
优先权	2000003832 2000-03-27 FR		
其他公开文献	JP2003528631A		
外部链接	Espacenet		

摘要(译)

本发明涉及涉及炎症和/或免疫疾病和一些癌症，特别是肠源性炎症性疾病的基因，以及由所述基因编码的蛋白质。本发明还涉及诊断炎症性疾病的方法。

I B D I の微細位置決定に用いたカー		
多型マーカーの名称	累積距離 (cM)	T C R マーカー
D16S3120	0	配列番号 7
(AFM328ve5)		配列番号 8
D16S298	2.9	配列番号 9
(AFMa189wg5)		配列番号 10
D16S299	3.4	配列番号 11
SPN	3.9	配列番号 12
		配列番号 13
D16S388	4.3	配列番号 14
		配列番号 15
D16S753	4.9	配列番号 16
(GGA3605)		配列番号 17
D16S3044	5.8	配列番号 18
(AFMa222zg9)		配列番号 19
D16S409	5.8	配列番号 20
(AFM151xa1)		配列番号 21
D16S3105	6.1	配列番号 22
(AFMb341ze5)		配列番号 23
D16S261	6.8	配列番号 24
(MPD24)		配列番号 25
D16S340	6.9	配列番号 26
(GATA7B02)		配列番号 27
D16S3090	7	配列番号 28
(AFMb068zb9)		配列番号 29
D16S517	7	配列番号 30
(AFMa132we9)		配列番号 31
D16S411	8	配列番号 32
(AFM156xa3)		配列番号 33
D16S3035	10.4	配列番号 34
(AFMa189wg5)		配列番号 35
D16S3136	10.4	配列番号 36
(AFM061xe5)		配列番号 37
D16S541	11.4	配列番号 38
(GATA7B02)		配列番号 39
D16S3117	11.5	配列番号 40
(AFM288wb1)		配列番号 41
D16S416	12.4	配列番号 42
(AFM210yg3)		配列番号 43
D16S770	12.4	配列番号 44
(GGA20G02)		配列番号 45
D16S263	13.2	配列番号 46
(GATA81B12)		配列番号 47
D16S390	15	配列番号 48
		配列番号 49
D16S419	16.5	配列番号 50
(AFM225zf2)		配列番号 51
D16S771	20.4	配列番号 52
(GGA23CG9)		配列番号 53
D16S408	21.8	配列番号 54
(AFM157xf8)		配列番号 55
D16S508	25.6	配列番号 56
(AFM304xf1)		配列番号 57
	38.4	配列番号 58