

(19) 日本国特許庁(JP)

(12) 公開特許公報(A)

(11) 特許出願公開番号

特開2004-158005

(P2004-158005A)

(43) 公開日 平成16年6月3日(2004.6.3)

(51) Int. Cl. <sup>7</sup>	F I	テーマコード (参考)
<b>G06F 19/00</b>	G06F 19/00 600	2G045
<b>C12M 1/00</b>	C12M 1/00 A	4B024
<b>C12N 15/09</b>	C12Q 1/34	4B029
<b>C12Q 1/34</b>	C12Q 1/68 A	4B063
<b>C12Q 1/68</b>	GO1N 27/62 K	5B075
	審査請求 未請求 請求項の数 74 O L	(全 94 頁) 最終頁に続く

(21) 出願番号	特願2003-360454 (P2003-360454)	(71) 出願人	500092402
(22) 出願日	平成15年10月21日 (2003.10.21)		シークエノム・インコーポレーテッド
(62) 分割の表示	特願2001-530799 (P2001-530799) の分割		アメリカ合衆国92121カリフォルニア 州 サンディエゴ、ジョン・ホプキンス・ コート3595番
原出願日	平成12年10月13日 (2000.10.13)	(74) 代理人	100062144
(31) 優先権主張番号	60/159, 176		弁理士 青山 稜
(32) 優先日	平成11年10月13日 (1999.10.13)	(74) 代理人	100086405
(33) 優先権主張国	米国 (US)		弁理士 河宮 治
(31) 優先権主張番号	60/217, 658	(74) 代理人	100067035
(32) 優先日	平成12年7月10日 (2000.7.10)		弁理士 岩崎 光隆
(33) 優先権主張国	米国 (US)	(72) 発明者	アンドレアス・ブラウン
(31) 優先権主張番号	60/217, 251		アメリカ合衆国92130カリフォルニア 州サンディエゴ、カーメル・クリーク・ロ ード11237-6番
(32) 優先日	平成12年7月10日 (2000.7.10)		最終頁に続く
(33) 優先権主張国	米国 (US)		

(54) 【発明の名称】 データベースを作成する方法および多型遺伝的マーカーを同定するためのデータベース

## (57) 【要約】

【課題】 DNAのSNPを同定するための唯一の利用可能な方法は、配列決定によるものであり、これは高価で、困難で、そして労働力消費性である。そのため、安価で、簡易で、そして労働力が少なくすむ配列決定法を提供することが目的である。さらには、疾患のゲノム基礎およびそのマーカーを同定するための方法を提供することが目的である。

【解決手段】 健康体ヒトドナーからのゲノムサンプルのデータベースを創出するためのプロセスおよび方法、多型遺伝的マーカーおよび他のマーカーを同定し、そして疾患および状態と関連させるデータベースを使用する方法を提供する。

【選択図】 なし

## 【特許請求の範囲】

## 【請求項 1】

生物学的サンプルが取得される複数の健康体生物を表すデータポイント；およびそれぞれの生物を同定するインデクサーを含むデータベースであって、ここで  
 ( a ) 各データポイントが、生物型および他の同定のための情報を表すデータと、インデクサーによって連関される；かつ  
 ( b ) 該データが蓄積可能である；  
 データベース。

## 【請求項 2】

データポイントが、コンピューター読み取り可能蓄積媒体に蓄積される、請求項 1 のデータベース。 10

## 【請求項 3】

リレーショナルデータベースである、請求項 1 または 2 のデータベース。

## 【請求項 4】

データポイントが、人種、年齢、性、身長、体重、アルコール摂取量、妊娠回数、生児出生回数、菜食主義、肉体的活動の型、住居の状態および/または特定の状態における居住の期間、教育レベル、親の死亡年齢、親の死亡原因、過去のまたは現在の喫煙者、喫煙者としての期間、喫煙頻度、肉親(親、兄妹、子供)における疾患の発生、処方薬の使用および/またはその理由、入院の期間および/または回数、および環境要因にさらされること、  
 からなる群から選択される、1 または 2 以上のパラメータに関する質問に対する回答である、請求項 1 - 3 いずれかのデータベース。 20

## 【請求項 5】

生物が動物、細菌、真菌、原生生物および寄生生物の中から選択されている、請求項 1 - 4 いずれかのデータベース。

## 【請求項 6】

サンプルが体液または組織である、請求項 1 - 5 いずれかのデータベース。

## 【請求項 7】

サンプルが血液、血液画分、細胞、亜細胞性のオルガネラ、細胞、種子、微生物、病原体および生殖組織サンプルから選択される、請求項 1 - 5 いずれかのデータベース。 30

## 【請求項 8】

生物由来の表現型データをさらに含む、請求項 1 - 7 いずれかのデータベース。

## 【請求項 9】

生物から取得した核酸からの遺伝子型のデータをさらに含む、請求項 1 - 8 いずれかのデータベース。

## 【請求項 10】

遺伝子型データが、遺伝的マーカー、非コード領域、マイクロサテライト、RFLP および VNTR を含む、請求項 1 - 9 いずれかのデータベース。

## 【請求項 11】

選択されたパラメータにしたがって請求項 1 - 10 いずれかのデータベースを分類し、  
 選択されたパラメータにマッチするサンプルを同定すること；  
 それぞれの同定されたサンプルから生体高分子を単離すること；  
 所望によりそれぞれ単離された生体高分子をプールすること；  
 所望により生体高分子の量を増幅すること；  
 プールした生体高分子を切断して、そのフラグメントを生成させること；  
 得られるフラグメントのマスペクトルを取得し、対照のマスペクトルと、該マスペクトルを比較し、スペクトル間の相違を同定し、それによって任意の多型を同定すること；  
 を含む集団における多型の発見方法；  
 ここで、 40 50

該対照のマスペクトルは、コレクション中の未分類のサンプルまたは異なるパラメータに従って分類されたサンプルから取得される。

【請求項 1 2】

切断が生体高分子を酵素と接触させることにより行われる、請求項 1 1 の方法。

【請求項 1 3】

酵素がヌクレオチドグリコシラーゼ、ニッカーゼおよび IIS 型制限酵素からなる群から選択される、請求項 1 2 の方法。

【請求項 1 4】

生体高分子が核酸またはタンパク質である、請求項 1 1 の方法。

【請求項 1 5】

質量分析の形式がマトリックス介助レーザー脱離/イオン化、飛行時間型(MALDI-TOF)、電子スプレー(ES)、IR-MALDI、イオンサイクロトロン共鳴(ICR)、フーリエ変換およびそれらの組み合わせの中から選択される、請求項 1 1 の方法。

【請求項 1 6】

選択されたパラメータにしたがって請求項 1 - 1 0 いずれかのデータベースを分類し、選択されたパラメータにマッチするサンプルを同定すること；

それぞれの同定されたサンプルから生体高分子を単離すること；

それぞれ単離された生体高分子をプールすること；

所望により生体高分子の量を増幅すること；および

プールした生体高分子における多型の頻度を決定すること；

を含む多型をパラメータと関連させる方法；

ここで、

対照と比較したときの多型の頻度の変化は、多型と、選択されたパラメータとの相関を示しており；かつ

該対照は、未分類のデータベースまたは異なるパラメータに従って分類されているデータベースより同定されたサンプルから取得した、プールした生体高分子における多型の頻度である。

【請求項 1 7】

多型性の遺伝子座にリンクした任意の経路または遺伝子を同定することをさらに含む請求項 1 6 の方法であって、ここで、

遺伝的マーカーの候補である多型性が同定される、方法。

【請求項 1 8】

頻度が、多型を含むプールした核酸中の核酸分子のパーセンテージである、請求項 1 6 または 1 7 の方法。

【請求項 1 9】

パラメータが、人種、年齢、性、身長、体重、アルコール摂取量、妊娠回数、生児出生回数、肉食主義者、肉体的活動の種類、住居の状態および/または特定の状態における居住の期間、教育レベル、親の死亡年齢、親の死亡原因、過去のまたは現在の喫煙者、喫煙者としての期間、喫煙頻度、肉親(親、兄妹、子供)における疾患の発生、処方薬の使用および/またはその理由、入院の期間および/または回数、および環境要因にさらされること、

からなる群から選択される、請求項 1 1 - 1 8 いずれかの方法。

【請求項 2 0】

パラメータが疾患の発生または肉親のメンバーにおける特定の疾患であり、それによって多型を該疾患と関連させる、請求項 1 9 の方法。

【請求項 2 1】

パラメーターが、年齢、人種または性別であり、かつ、多型性の頻度が、健康体集団における、加齢で、人種で、または性別で決定される、請求項 1 9 の方法。

【請求項 2 2】

10

20

30

40

50

パラメータが年齢であり、それによって多型を死亡(率)、早期罹病(率)、または罹病(率)および早期死亡(率)に対する感受性と相関させる、請求項19の方法。

【請求項23】

切断が生体高分子を酵素と接触させることによりおこなわれる、請求項11-22いずれかの方法。

【請求項24】

酵素がヌクレオチドグリコシラーゼ、ニッカーゼおよびIIS型制限酵素からなる群から選択される、請求項23の方法。

【請求項25】

生体高分子が核酸またはタンパク質である、請求項23の方法。

10

【請求項26】

多型が、

- a)第1のオリゴヌクレオチドを標的核酸とハイブリダイズさせること；
  - b)第2のオリゴヌクレオチドを標的核酸の隣接領域とハイブリダイズさせること；
  - c)ハイブリダイズしたオリゴヌクレオチドをライゲートさせること；および
  - d)標的核酸の存在を指標として、ハイブリダイズした第1のオリゴヌクレオチドを質量分析法により検出すること；
- のステップを含む方法でサンプル中の標的核酸の存在を検出することにより同定される、請求項16-25いずれかの方法。

【請求項27】

20

多型が、

- a)第1のオリゴヌクレオチドを標的核酸とハイブリダイズさせ、第2のオリゴヌクレオチドを標的核酸の隣接領域とハイブリダイズさせること；
  - b)ハイブリダイズした第1および第2のオリゴヌクレオチドを切断酵素と接触させて切断生成物を形成させること；および
  - c)標的核酸の存在を指標として質量分析法により切断生成物を検出すること；
- のステップを含む方法でサンプル中の標的核酸を検出することにより同定される、請求項16-25いずれかの方法。

【請求項28】

核酸の領域を増幅しアプ리콘を作成すること、ここで、得られるアプ리콘には1または2以上の酵素制限部位が含まれ；  
アプ리콘を制限酵素と接触させ、フラグメントを作成すること；  
得られるフラグメントのマスマスペクトルを取得すること、および  
マスマスペクトルのシグナルを解析すること；  
を含む、請求項16-25いずれかの方法；  
ここで、  
該多型はシグナルのパターンから検出される。

30

【請求項29】

質量分析の形式がマトリックス介助レーザー脱離/イオン化、飛行時間型(MALDI-TOF)、電子スプレー(ES)、IR-MALDI、イオンサイクロトロン共鳴(ICR)、フーリエ変換およびそれらの組み合わせの中から選択される、請求項27または28の方法。

40

【請求項30】

プライマーオリゴ塩基伸長法(PROBE法)でサンプル中の標的核酸を同定することにより多型を同定する、請求項16-25いずれかの方法。

【請求項31】

プライマーオリゴ塩基伸長法が、

- a)標的ヌクレオチドを含む核酸分子を取得すること；
- b)所望により核酸分子を固体支持体上に固定し、固定された核酸分子を作成すること；
- c)核酸分子を、標的ヌクレオチドに隣接した部位で、核酸分子に相補的なプライマーオリ

50

ゴヌクレオチドとハイブリダイズさせること；

d)ステップc)の生成物をジデオキシヌクレオシド・トリホスフェートまたは3'-デオキシヌクレオシド・トリホスフェートおよびポリメラーゼを含むコンポジションと接触させ、その結果、標的ヌクレオチドに相補的なジデオキシヌクレオシド・トリホスフェートまたは3'-デオキシヌクレオシド・トリホスフェートのみをプライマー上で伸長させること；および

e)伸長したプライマーを検出し、それにより標的ヌクレオチドを同定すること；を含む、請求項30の方法。

【請求項32】

伸長したプライマーの検出を質量分析法により実施する請求項31の方法であって、ステップd)の生成物をイオン化し、気化させること；および

質量分析法により伸長したプライマーを検出し、それにより標的ヌクレオチドを同定すること；

を含む方法。

【請求項33】

サンプルをチップ上のアレイとして質量分析計に提供し；そして

各サンプルが、マトリックス介助レーザー脱離/イオン化(MALDI)質量分析法で使用する質量分析計におけるレーザーにより発射されたレーザースポットほどの大きさの体積を占める、請求項32の方法。

【請求項34】

反応容器中に含まれる生物学的サンプルに対して、それぞれが処理(procedure)を実行する複数のプロセッシング・ステーションを含むプロセスライン；

プロセッシング・ステーションからプロセッシング・ステーションへと反応容器を運ぶロボットシステム；

プロセスラインのテスト結果を受け取り、テスト結果を自動的にプロセッシングし、反応容器中の生物学的サンプルに関して決定をするデータ解析システム；

各プロセッシング・ステーションでのテストをいつ終了させるかを決定し、それに応じて、反応容器を次のテストステーションへと移動させ、コントロールシステムが停止の指示を受けるまで逐次、反応容器を連続的にプロセッシングするコントロールシステム；および自動化されたプロセスラインによりテストされるサンプルがデータベース内の対象からの

サンプルを含む請求項1-10いずれかのデータベース；

を含む生物学的サンプルのハイスループットプロセッシングシステム。

【請求項35】

プロセッシング・ステーションのうちの1つが質量分析計を含む、請求項34のシステム。

【請求項36】

生物学的サンプルに関するテストデータが1または2以上のシグナルを含むように、データ解析システムが質量分析計からのテストデータを受け取ることによりテスト結果をプロセッシングし、その後、データ解析システムが各シグナルの曲線の下面積を決定し、その結果を正規化し、テストされたサンプル中のコンポーネントの相対量を表す実質的に

【請求項37】

請求項1-10いずれかのデータベース；

自動化されたプロセスライン、ここでサンプルはデータベース内の対象由来のサンプルを含む自動化されたプロセスラインによりテストされる；および

サンプル内の生体高分子の解析のための質量分析計；

を含む、生物学的サンプルのハイスループットプロセッシングのためのシステム。

【請求項38】

生物学的サンプルのハイスループットプロセッシングのための方法であって；

反応容器中に含まれる1または2以上の生物学的サンプルに対してそれぞれが処理を実行

する、複数のプロセッシング・ステーションを有するプロセスラインを含んだ請求項 3 4 のシステムにしたがって反応容器を運ぶこと；  
各プロセッシング・ステーションでのテスト処理をいつ終了させるかを決定し、それに応答して、反応容器を次のプロセッシング・ステーションに移動させること；  
プロセスラインのテスト結果を受け取り、自動的にテスト結果をプロセッシングし、反応容器中の生物学的サンプルに関してデータ解析決定を作成すること；そして  
停止の指示を受けるまで逐次連続的に反応容器を処理すること；  
を含む方法；

ここで、

自動化されたプロセスラインによりテストされるサンプルには、データベース内の患者由来のサンプルが含まれる。 10

【請求項 3 9】

プロセッシング・ステーションのうちの 1 つが質量分析計を含む、請求項 3 8 の方法。

【請求項 4 0】

サンプルが、プライマーオリゴ塩基伸長法 ( P R O B E 法 ) を含む方法により解析される、請求項 3 9 の方法。

【請求項 4 1】

生物学的サンプルに関するテストデータが 1 または 2 以上のシグナルまたはシグナルを表す絶対値を含むように、質量分析計からのテストデータを受け取ることによりテスト結果をプロセッシングし、その後、データ解析システムが各シグナルの曲線の下面積を決定し、その結果を正規化し、テストされたサンプル中のコンポーネントの相対量を表す実質的に定量的な結果を取得すること； 20

をさらに含む請求項 4 0 の方法。

【請求項 4 2】

プライマーオリゴ塩基伸長法が、

- a) 標的ヌクレオチドを含む核酸分子を取得すること；
- b) 所望により核酸分子を固体支持体上に固定し、固定された核酸分子を作成すること；
- c) 核酸分子を、標的ヌクレオチドに隣接した部位で核酸分子に相補的なプライマーオリゴヌクレオチドとハイブリダイズさせること；
- d) ステップ c) の生成物をジデオキシヌクレオシド・トリホスフェートまたは 3' - デオキシヌクレオシド・トリホスフェートおよびポリメラーゼを含むコンポジションと接触させ、その結果、標的ヌクレオチドに相補的なジデオキシヌクレオシドまたは 3' - デオキシヌクレオシド・トリホスフェートのみをプライマー上で伸長させること；および 30
- e) プライマーを検出し、それにより標的ヌクレオチドを同定すること；

を含む、請求項 4 0 の方法。

【請求項 4 3】

伸長したプライマーの検出を質量分析法により実施する、請求項 4 2 の方法であって、ステップ d) の生成物をイオン化し、気化させること；および

質量分析法により伸長したプライマーを検出し、それにより標的ヌクレオチドを同定すること； 40

を含む方法。

【請求項 4 4】

サンプル中の標的核酸が、

- a) 第 1 のオリゴヌクレオチドを標的核酸とハイブリダイズさせること；
- b) 第 2 のオリゴヌクレオチドを標的核酸の隣接領域とハイブリダイズさせること；
- c) 次いでハイブリダイズしたオリゴヌクレオチドをライゲートさせること；および
- c) 標的核酸の存在を指標として、ハイブリダイズした第 1 のオリゴヌクレオチドを質量分析法により検出すること；

のステップを含む方法により検出および / または同定される、請求項 3 9 の方法。

【請求項 4 5】

サンプル中の標的核酸が、

- a) 第1のオリゴヌクレオチドを標的核酸とハイブリダイズさせ、第2のオリゴヌクレオチドを標的核酸の隣接領域とハイブリダイズさせること；
  - b) ハイブリダイズした第1および第2のオリゴヌクレオチドを切断酵素と接触させて切断生成物を形成させること；および
  - c) 標的核酸の存在を指標として質量分析法により切断生成物を検出すること；
- のステップを含む方法により、検出および/または同定される、請求項39の方法。

【請求項46】

評価されるべき健康体の標的集団および遺伝的変異を選択すること；  
 集団のメンバーから取得した生体高分子の複数のサンプルをプールすること；  
 変異を含む生体高分子を質量分析法により決定または検出すること；  
 マススペクトルまたはそのデジタル表示を取得すること；および  
 該集団の変異の頻度を決定すること；  
 を含む遺伝的変異の頻度のハイスループット決定方法。

10

【請求項47】

変異が対立遺伝子変異、翻訳後修飾、核酸修飾、標識、核酸の質量的修飾およびメチル化からなる群から選択され；そして/または  
 関心のある生体高分子の濃度がそれぞれのサンプルにおいて同じである、生体高分子が核酸、タンパク質、多糖類、脂質、低分子有機メタボラリトまたは中間体であり；そして/または  
 該頻度が、ゲノム変異を含む生体高分子の質量に対応する、マススペクトルまたはそのデジタル表示のピークの下の面積を決定することを含む方法を評価することにより決定される；  
 請求項46の方法。

20

【請求項48】

頻度の決定方法が、バックグラウンドが補正されている、全体のマススペクトルの総面積に対するシグナルまたはそのデジタル表示の比率を決定することにより実施される、請求項47の方法。

【請求項49】

(a) 選択されたパラメータにしたがって請求項1 - 10いずれかのデータベースを分類し、選択されたパラメータにマッチするサンプルを同定すること；  
 (b) それぞれの同定されたサンプルから核酸を単離すること；  
 (c) 所望によりそれぞれの単離された核酸をプールすること；  
 (d) 核酸の量を増幅すること；  
 (e) 1本鎖核酸を形成させ、それぞれの1本鎖を別個の反応容器に分けること；  
 (f) それぞれの1本鎖核酸をアダプター核酸と接触させ、アダプター複合体を形成させること；  
 (g) アダプター複合体をヌクレアーゼおよびリガーゼと接触させること；  
 (h) ステップ(g)の生成物を、ライゲートしたアダプターを増幅することができる混合物と接触させ、伸長した生成物を生成させること；  
 (i) ステップ(h)から得られるそれぞれの核酸のマススペクトルを取得し、伸長した生成物に対応するシグナルを同定することにより多型を検出すること；  
 (j) 同じ鎖上の異なる配列にハイブリダイズする別のアダプター核酸とハイブリダイズ可能な核酸を利用して、ステップ(f)~(i)を繰り返すこと；  
 を含む、核酸中の多型をハプロタイピングする方法；  
 ここで、  
 該多型は1より多い伸長生成物を検出することによりハプロタイピングされる。

30

40

【請求項50】

ヌクレアーゼがFen-1である、請求項49の方法。

【請求項51】

50

選択されたパラメータにしたがって請求項 1 - 10 いずれかのデータベースを分類し、  
 選択されたパラメータにマッチするサンプルを同定すること；  
 それぞれの同定されたサンプルから核酸を単離すること；  
 それぞれ単離された核酸をプールすること；  
 所望により核酸の量を増幅すること；  
 核酸を少なくとも 1 つの酵素と接触させ、そのフラグメントを生成させること；  
 得られるフラグメントのマスペクトルを取得すること；  
 を含む、集団中の多型をハプロタイピングする方法；  
 ここで、  
 該多型は、多型に対応するシグナルを検出することにより検出され；そして  
 該多型は、核酸の同じ鎖に位置する多型のマスペクトルから決定することによりハプロ  
 タイピングされる。

10

## 【請求項 5 2】

酵素がニッカーゼである、請求項 4 9 の方法。

## 【請求項 5 3】

ニッカーゼが NY 2 A および NY S 1 からなる群から選択される、請求項 5 2 の方法。

## 【請求項 5 4】

多型を同定すること；

健康体集団における多型の頻度を、年齢によって請求項 1 - 11 いずれかのデータベース  
 を層化することによって決定すること；および  
 加齢による統計的有意性の相違は、該多型性が罹病（率）、早期死亡（率）、または罹病  
 （率）および早期死亡（率）への感受性と相関することを指摘する、年齢群間の多型性頻  
 度の相違の統計的有意性を決定すること；  
 を含む、多型性が罹病（率）、早期死亡（率）、または罹病（率）および早期死亡（率）  
 への感受性と相関するか否か決定する方法。

20

## 【請求項 5 5】

核酸サンプル内のメチル化されたヌクレオチドの検出方法であって、  
 核酸のサンプルを別個の反応容器に分けること；

1 つの反応容器中の核酸を亜硫酸塩と接触させること；

各反応容器中の核酸分子を増幅すること；

30

各反応容器中の核酸分子を切断し、そのフラグメントを生成させること；および

1 つの反応容器から得られるフラグメントのマスペクトル、および別の反応容器から得  
 られるフラグメントの別のマスペクトルを取得すること；

を含む方法；

ここで、

シトシンのメチル化はマスペクトル間のシグナルの相違を同定することにより検出され  
 る。

## 【請求項 5 6】

増幅のステップが d U T P の存在下で行われ；そして

切断のステップがウラシルグリコシラーゼで処理し、該核酸を切断のために有効な条件下  
 でインキュベートすることにより行われる、請求項 5 5 の方法。

40

## 【請求項 5 7】

核酸サンプル内のメチル化ヌクレオチドを検出する方法であって、

任意の非メチル化シトシンを脱アミノ化するために重亜硫酸塩で核酸分子を処理し、ウラ  
 シルを生産する、

処理した分子を増幅する、

増幅した核酸分子をフラグメント化する、

得られたフラグメントのマスペクトルを取得し、シトシンの代わりにウラシルを含むも  
 のを、処理してない核酸分子から得られるスペクトルと、処理した増幅フラグメントから  
 のスペクトルとを比較することによって同定する、

50

ことを含む、方法。

【請求項 58】

フラグメント化が、ウラシルグリコシラーゼで、フォスフェートバックボーンを切断するのに有効な条件下で処理することによって実施される、請求項 57 の方法。

【請求項 59】

請求項 2 - 10 いずれかのデータベースを含むコンピューターシステム。

【請求項 60】

請求項 2 - 10 いずれかのデータベースを含むデータ蓄積媒体。

【請求項 61】

請求項 1 - 10 いずれかのデータベースおよび質量分析計を含む組み合わせ。

10

【請求項 62】

生物学的サンプルを解析するための自動化されたプロセスラインである、請求項 61 の組み合わせ。

【請求項 63】

請求項 1 - 10 いずれかのデータベースおよびインデックスがサンプルが取得された対象を同定するインデックスを付されたサンプルのコレクション；  
を含む組み合わせ。

【請求項 64】

血液、組織、体液、細胞、種子、微生物、病原体および生殖組織のサンプルからなる群から選択される、複数のサンプル；および  
各サンプルのソースおよび/または履歴 (history) を表す、サンプルを含むコンテナのシンボロジー；

20

を含む標的集団由来のサンプルのサブコレクション；

ここで、

標的集団は疾患状態について選択されていない健康体集団であり；

核酸配列情報が該サンプルから由来し；

コレクションは健康体集団からのサンプルを含み；そして

サブコレクションは特異的なパラメータにしたがって該コレクションを分類することにより得られる。

30

【請求項 65】

核酸、胎児組織、タンパク質のサンプルからなる群から選択される、複数のサンプル；  
および

各サンプルのソースおよび/または履歴を表す、サンプルを含むコンテナのシンボロジー；

を含む標的集団由来のサンプルのサブコレクション；

ここで、

該標的集団は疾患状態で選択されていない健康体集団であり；

該コレクションは、健康体集団からのサンプルを含み；そして

該サブコレクションは、特異的なパラメータにしたがってコレクションを分類することにより得られる。

40

【請求項 66】

パラメータが、人種、年齢、性、身長、体重、アルコール摂取量、妊娠回数、生児出生回数、菜食主義者、肉体的活動の型、住居の状態および/または特定の状態における居住の期間、教育レベル、親の死亡年齢、親の死亡原因、過去のまたは現在の喫煙者、喫煙者としての期間、喫煙頻度、肉親 (親、兄妹、子供) における疾患の発生、処方薬の使用および/またはその理由、入院の期間および/または回数、および環境要因にさらされること

からなる群から選択される、請求項 64 または 65 のサブコレクション。

【請求項 67】

50

シンボロジーがバーコードである請求項 6 4 または 6 5 のサブコレクション。

【請求項 6 8】

生物学的サンプルのコンポジションを示すデータセットを作成すること；  
データセットのノイズを除去し、ノイズが除去されたデータを作成すること；  
ノイズが除去されたデータからベースラインを削除し、中間データセット (intermediate data set) を作成すること；  
生物学的サンプルに関して推定されるピークを定義すること；  
推定されるピークを用いて残留ベースライン (residual baseline) を作成すること；  
残留ベースラインを中間データセットから除いて、補正されたデータセットを作成すること；  
残留ベースラインを除去することに応じて、補正されたデータセット中で有り得るピーク (probable peak) を位置決定すること；および  
位置決定した有り得るピークを用いて生物学的サンプルを同定すること；  
を含む、生物学的サンプルの同定方法；  
ここで、作成される生物学的サンプルデータセットは、アッセイ用のフラグメントのセンス鎖およびアンチセンス鎖からのデータを含む。

10

【請求項 6 9】

同定が、  
センス鎖とアンチセンス鎖からのデータを組み合わせること、およびデータを予測されるセンス鎖およびアンチセンス鎖の値と比較して生物学的サンプルを同定することを含む、  
請求項 6 8 に記載の方法。

20

【請求項 7 0】

有り得るピークが、センス鎖のデータ由来であるか、アンチセンス鎖のデータ由来であるかに従い、当該同定が、有り得るピークについてピーク確率を導くことを含む、請求項 6 8 に記載の方法。

【請求項 7 1】

当該同定が、有り得るピークについてピーク確率を導くこと、および有り得るピークの下  
の算出した面積と、データセットにおけるすべてのピークの下  
の算出した予測される平均面積との比率に応じて、対立遺伝子ペナルティーを適用することを含む、請求項 6 8 に  
記載の方法。

30

【請求項 7 2】

生物学的サンプルのコンポジションを示すデータセットを作成すること；  
データセットのノイズを除去し、ノイズが除去されたデータを作成すること；  
ノイズが除去されたデータからベースラインを削除し、中間データセットを作成すること；  
生物学的サンプルに関して推定されるピークを定義すること；  
推定されているピークを用いて残留ベースラインを作成すること；  
残留ベースラインを中間データセットから除いて、補正されたデータセットを作成すること；  
残留ベースラインを除去することに応じて、補正されたデータセット中で有り得るピーク  
を位置決定すること；および  
位置決定した有り得るピークを用いて生物学的サンプルを同定すること；  
を含む、生物学的サンプルの同定方法；  
ここで、  
当該同定には、有り得るピークについてピーク確率を導くこと、および有り得るピークの下  
の算出した面積と、データセットにおけるすべてのピークの下  
の算出される予測される面積との比率に応じて、対立遺伝子ペナルティーを適用することが含まれる。

40

【請求項 7 3】

同定が、適用される対立遺伝子ペナルティーを受け取らなかった有り得るピークからの  
データを比較し、オリゴヌクレオチドの生物学的データにしたがってそれらの質量を決定

50

することを含む、請求項 7 2 に記載の方法。

【請求項 7 4】

対立遺伝子ペナルティーが、予測される面積値に対するピークの下面積の比率が 3 0 % より大きい、有り得るピークには適用されない、請求項 7 2 に記載の方法。

【発明の詳細な説明】

【技術分野】

【0001】

関連出願

以下の出願の優先権の利益をここに請求する：

Andreas Braun, Hubert Koster; Dirk Van den Boonの、米国仮出願番号 6 0 / 2 1 7 6 5 8、7 月 1 0 日出願、名称「データベースを作成する方法および多型遺伝的マーカーを同定するためのデータベース」；Andreas Braun, Hubert Koster, Dirk Van den Boomの、米国仮出願番号 6 0 / 1 5 9 1 7 6、1 9 9 9 年 1 0 月 1 3 日出願、名称「データベースを作成する方法および多型遺伝的マーカーを同定するためのデータベース」；Andreas Braunの、米国仮出願番号 6 0 / 2 1 7 2 5 1、2 0 0 0 年 7 月 1 0 日出願、名称「多型キナーゼアンカータンパク質遺伝子配列、多型キナーゼアンカータンパク質および多型キナーゼアンカータンパク質およびそれらをコードしている核酸を検出する方法」；および Ping Yipの、米国出願番号 0 9 / 6 6 3 9 6 8、2 0 0 0 年 9 月 1 9 日出願、名称「生物学的サンプルを同定するための方法および装置」。

許容される場合、前記の出願および仮出願を引用をもって、その全体をここに含ませる。

【0002】

発明の分野

健康体ヒトドナーからのゲノムサンプルのデータベースを創出するためのプロセスおよび方法。当該データベースを使用して多型遺伝的マーカーおよび他のマーカーを同定し、そして疾患および状態と関連させる方法を提供する。

【背景技術】

【0003】

背景

すべての生物の疾患は、遺伝され、または体の環境ストレス、例えば、ウイルスおよび毒物への応答に起因する、遺伝的コンポーネントを有する。ゲノムリサーチを進める究極的目標は、この情報を使用して、これらの疾患を同定し、処置し、そして有効に治療する新しい方法を開発することである。第一のステップは、個体サンプルのレベルで疾患組織をスクリーニングし、そしてゲノム変化を同定することである。

【0004】

次いで、これらの「疾患」マーカーの同定は、これらの誤りのある遺伝子または多型を検出する診断テストの開発および商業化を活気付ける。医学的および薬学的コミュニティーに対する問題提起は、単一ヌクレオチド多型 (SNP)、マイクロサテライト、タンデムリピート、新規にマッピングされたイントロンおよびエキソンを含む遺伝的マーカーの数を増加させることによって、疾患を同定するだけでなく、また疾患の進行に従い、そして処置に対する生物の応答を予測する、遺伝子型を同定することである。

【0005】

現行の薬学的および生物工学的産業は、疾患を発見し、それからその疾患について、ゲノムの基礎を決定しようと試みる。このアプローチは、時間消費性で、そして高価であり、そして多くの場合、その疾患にどの経路が関係し得るかについて、調査者は推測しなければならない。

【0006】

ゲノミクス

現在、利用可能なゲノム情報を分析するときに使用される 2 種の主要な戦略は、逆遺伝

10

20

30

40

50

学的な力づくの戦略および知識に基づく経路指向的 ( pathway oriented ) フォワードジェネティクス ( forward genetics ) 戦略である。この力づくのアプローチによって、配列情報の大きなデータベースを得るが、医学的または配列情報の他の使用について、わずかな情報しか得られない。ゆえに、この戦略は、価値に疑問のある漠然とした産物をもたらす。知識に基づく戦略によれば、特定の DNA 配列およびその経路の他の産物の医学的使用についての多くの情報を含む小さいデータベースが得られ、そして高価値の具体的な産物が得られる。

#### 【 0 0 0 7 】

##### 多型

多型は血液型の同定によって 1 9 0 1 以来知られている。1 9 5 0 年代に、これらは大集団の遺伝的研究を使用して、タンパク質のレベルで同定された。1 9 8 0 年代および 1 9 9 0 年代に、多くの既知のタンパク質多型が、ゲノム DNA の遺伝子座と相関された。例えば、アポリポタンパク質 E 4 型の対立遺伝子の遺伝子用量 ( gene dose ) が、遅発型ファミリーのアルツハイマー病のリスクと相関され ( 例えば、Corder et al. (1993) Science 261: 921-923 参照 ) ; 血液凝第 V 因子の突然変異が活性化タンパク質 C に対する抵抗性と連関され ( 例えば、Bertina et al. (1994) Nature 369:64-67 参照 ) ; HIV - 1 感染に対する抵抗性が CCR - 5 ケモカインレセプター遺伝子の突然変異体対立遺伝子を有する白人個体で示され ( 例えば、Samson et al. (1996) Nature 382:772-725 参照 ) ; そして抗原提示細胞 ( APC、例えばマクロファージ ) の高頻度突然変異性トラクトが Ashkenzi ユダヤ人バックグラウンドの個体の家族性結腸直腸癌で同定された ( 例えば、Laken et al. (1997) Nature Genet. 17:79-83 参照 ) 。ヒトゲノムの 3 百万より多い多型部位が存在し得る。多くは同定されたが、なお特徴を把握され、またはマッピングされ、またはマーカーと連関されていない。

#### 【 0 0 0 8 】

##### 単一ヌクレオチド多型 ( SNP )

ゲノミクスの焦点の多くは、SNP の同定にあり、それは種々の理由のために重要である。それらは、間接的テスト ( ハプロタイプの連関 ) および直接的テスト ( 機能変異体 ) をもたらす。これらは、最も豊富な、そして安定な遺伝的マーカーである。共通の疾患は、共通する遺伝的な変化によってとくにうまく説明され、そしてヒト集団の天然の変異は、疾患、治療および環境的相互作用の理解の助けとなる。

#### 【 発明の開示 】

#### 【 発明が解決しようとする課題 】

#### 【 0 0 0 9 】

現在、DNA の SNP を同定するための唯一の利用可能な方法は、配列決定によるものであり、これは高価で、困難で、そして労働力消費性である。さらに、ひとたび SNP が発見されると、それが真の多型なのか配列決定のエラーなのか決定するために確認しなければならない。また、次いで、発見された SNP は特定の表現型と連関されるかどうか、評価されなければならない。したがって、疾患についてのゲノム基礎を、およびそのマーカーを同定するための新規パラダイムを開発する必要がある。したがって、疾患のゲノム基礎およびそのマーカーを同定するための方法を提供することがここでの目的である。

#### 【 課題を解決するための手段 】

#### 【 0 0 1 0 】

##### 要約

データベースおよびデータベースを使用する方法を、ここで提供する。このデータベースは、健康であるという基礎のみに基づき選択された集団の対象と連関するパラメータのセットを含む ( すなわち、対象が哺乳動物、例えば、ヒトである場合、彼らは外観上の健康および検出可能な感染症がないことに基づいて選択される ) 。このデータベースは、1 または 2 以上の選択されたパラメータに基づいて分類することができる。

#### 【 0 0 1 1 】

データベースは、好ましくはリレーショナルデータベース ( relational database ) で

あり、それはそれぞれの対象を表すインデックスが、パラメータ（それらはデータ、例えば、年齢、人種、性別、病歴等、そして究極的には遺伝子型情報であり、データベースに入力され、そして蓄積された）を関連づけるのに役立つ。次いで、このデータベースは、これらのパラメータにしたがって分類することができる。最初に、そのパラメータ情報はそれぞれの対象（そこから体組織または体液サンプルを取得する）によって回答される質問票から取得される。それぞれのサンプルについてさらなる情報を取得するように、この情報をデータベースに入力することができ、そして分類パラメータとして役立てることができる。

**【0012】**

健康体個体から取得したデータベースは、多くの使用（例えば、既知の多型を表現型または疾患と関連させること）を有する。このデータベースを使用して、有害である、有益である、そして疾患と関連のある対立遺伝子を同定することができる。

10

**【0013】**

ここでの目的のために、遺伝子型情報を当業者に既知の任意の方法を使用して取得することができるが、好ましくは質量分析法を使用して取得する。

**【0014】**

対象および遺伝子型および他のパラメータ、例えば、年齢、人種、種族および性の既存のデータベースの新規の使用を、またここで提供する。任意のデータベースを、ここでの方法にしたがって分類することができ、そして任意の分類パラメータと統計的に有意な相関を示す対立遺伝子を同定することができる。しかし、ここで提供するデータベースおよびランダムに選択したデータベースが、これらの方法でよりよく機能することが注目され、なぜなら、疾患に基づくデータベースは多くの限界（それらは比較的大きさが小さい点、選択された疾患集団の均質性、およびデータベースの選択のためのマーカーと連関する多型のマスキング効果を含む）をこうむっているからである。ゆえに、ここで提供する健康体のデータベースは、これまで認識され、または開拓されていない有利な効果を提供する。

20

**【0015】**

しかし、ここで提供する方法は、多型の発見および相関のための分類を有する（または有しない）、疾患に基づくデータベースを含む、選択されたデータベースとともに使用することができる。さらに、ここで提供するデータベースは、多型の発見のために典型的に利用される選択されていないデータベースよりもより大きい遺伝的多様性を表し、したがって、これにより、効率的に多型を発見し、そして相関を得ることができる。

30

**【0016】**

ここで提供するデータベースは、同定された多型を取得するために使用し、そしてデータを選択されたパラメータにしたがって分類するときに、それが頻度について変化するかどうかを確認することができる。

**【0017】**

これらの方法のある使用は、既知の遺伝的マーカーの出現に従うことによって、特定のパラメータと選択されたマーカーを関連させることであり、それからこの相関を作成したら、疾患との相関を決定し、または同定する。この使用の例は、p 53 およびリポタンパク質リパーゼ多型である。ここで例示のように、既知のマーカーは、ある種の群、例えば、特定の人種または種族またはある性別と特定の相関を有することが示される。次いで、そのような相関によって、よりよい診断テストおよび処置養生法の開発が可能となる。

40

**【0018】**

これらの方法は、1または2以上の遺伝的マーカーの同定のために有益であり、それらの頻度は、年齢、人種群、性別またはある他の基準の関数（function）として集団内で変化する。これは、以前未知である多型、そして究極的には、疾患の開始（onset）および進行に関係する遺伝子または経路の同定を可能とする。

**【0019】**

ここで提供するデータベースおよび方法によって、とりわけ、その遺伝子基盤によって

50

疾患プロセスのコンポーネント、特にキーコンポーネントの同定が可能となり、そしてまたプロセス、例えば個体の薬物応答の理解が可能となる。ここで提供するデータベースおよび方法は、また病理学的経路の解明に関係する方法で、新規診断アッセイの開発で、新規の可能性のある薬物標的の同定、および新規薬物候補の同定で使用することができる。

【0020】

この方法およびデータベースは、インシリコ (in silico) SNP同定、インビトロ SNP同定/明確化、大集団の遺伝的プロファイル、および生物統計的分析および解釈を含むがこれらに限定されない実験的処理とともに使用することができる。

【0021】

ここで提供するデータベースおよびデータベースの対象からの生物学的サンプル、および好ましくはすべての対象またはデータベース中の複数の対象からの生物学的サンプルを含む組み合わせを、またここで提供する。組織および体液サンプルの収集物をまた提供する。

【0022】

年齢と相関し、多型の同定を含み、そして健康体集団の年齢の増加について多型の頻度を決定する、遺伝的マーカーを決定するための方法を、またここで提供する。

【0023】

遺伝的マーカーが病的状態 (morbidty) に対する感受性、早期死亡 (early mortality)、または病的状態および早期死亡と相関するか否かを決定するための方法であって、多型を同定し、そして健康体集団で年齢の増加にともなう多型の頻度を決定することを含む方法を、またここで提供する。

ここで記載の任意の方法を、マルチプレックスフォーマットで使用することができる。

【0024】

遺伝的情報を正確に同定するための装置およびプロセスを、またここで提供する。遺伝的データから高度に自動化された態様で遺伝的情報を抽出することが、ここでのさらなる目的である。したがって、既知の通常のシステムの欠点を克服するために、生物学的サンプルを同定するための方法および装置を提案する。

【0025】

簡単には、生物学的サンプルを同定するための方法およびシステムは、生物学的サンプルのコンポジションを示すデータセットを作成する。特定の例では、そのデータセットは、質量分析計から受けとられるDNAスペクトロメトリデータである。このデータセットは、ノイズが除去され、そしてベースラインが削除されている。生物学的サンプルのあり得るコンポジションが既知であり得るから、予測ピークエリア (area) を決定し得る。予測ピークエリアを使用し、残留ベースラインを作成し、データセットをさらに補正する。次いで、有り得るピーク (probable peak) が補正されたデータセットでは同定可能であり、それを、生物学的サンプルのコンポジションを同定するために使用する。開示の例では、統計的方法を使用して、有り得るピークが実際のピークである、または実際のピークではない、またはそのデータがあまりに決定的でないため利用 (call) できない、確率を決定する。

【0026】

生物学的サンプルを同定するための方法およびシステムは、コンポジションコール (composition call(s)) を高度に自動化された態様で正確に作成することが、有利である。そのような態様では、例えば、完全なSNPプロファイル情報が効率的に収集され得る。収集されたデータが高度に正確な結果で分析されることがより重要である。例えば、特定のコンポジションがコールされるとき、結果は高い信頼性で信頼し得る。そのような信頼性を、使用の確固たる (robust) コンピュータ的プロセスによって提供する。

【0027】

図面の説明

図1は、例示的なサンプルバンクを示す。パネル1は、性別および人種の関数としてのサンプルを示す。パネル2は、年齢の関数として白人を示す。パネル3は、年齢の関数と

してヒスパニックを示す。

図 2 A および 2 C は、リポタンパク質リパーゼ遺伝子の 2 9 1 S 対立遺伝子の年齢および性別分布を示し、全部で 4 3 6 人の男性および 5 8 9 人の女性を調査した。図 2 B は、4 3 6 人の男性について年齢の分布を示す。

【 0 0 2 8 】

図 3 は、集団に基づくサンプルバンキングのための例示的な質問票である。

図 4 は、血液サンプルコンポーネントのプロセッシングおよびトラッキングを示す。

図 5 は、「病気」対立遺伝子および「健康体」対立遺伝子の対立遺伝子頻度を年齢の関数として示す。健康体対立遺伝子の相対的頻度が年齢の増加とともに集団中で増加することが注目される。

10

図 6 は、A p o E 遺伝子型の年齢依存性分布を示す ( Schaechter et al. (1994) Nature Genetics 6:29-32 参照 ) 。

【 0 0 2 9 】

図 7 A - D は、データベース中の白人集団中の p 5 3 ( 腫瘍サプレッサー ) コドン 7 2 の年齢依存性および遺伝子型頻度を示す。\* R 7 2 および \* P 7 2 はデータベース集団中の対立遺伝子の頻度を表す。R 7 2、R 7 2 P、および P 7 2 は集団中の個体の遺伝子型を表す。ホモ接合性 P 7 2 対立遺伝子の頻度は、年齢とともに 6 . 7 % から 3 . 7 % に低下する。

図 8 は、年齢の関数として、p 2 1 S 3 1 R 対立遺伝子の対立遺伝子および遺伝子型頻度を示す。

20

図 9 は、個体サンプルに対してプールされた F V I I 対立遺伝子 3 5 3 Q の頻度を示す。

【 0 0 3 0 】

図 1 0 は、個体サンプルに対してプールされた、C E T P ( コレステロールエステル輸送タンパク質 ) 対立遺伝子の頻度を示す。

図 1 1 は、個体サンプルに対してプールされた、プラスミノゲンアクチベーターインヒビター - 1 ( P A I - 1 ) 5 G の頻度を示す。

図 1 2 は、P A I - 1 対立遺伝子のサンプルのマススペクトルおよび人種分布を示す。

図 1 3 は、C E T P 4 0 5 対立遺伝子のサンプルのマススペクトルおよび人種分布を示す。

30

【 0 0 3 1 】

図 1 4 は、第 V I I 因子 3 5 3 対立遺伝子のサンプルのマススペクトルおよび人種分布を示す。

図 1 5 は、プールされた D N A サンプルを使用する、P A I - 1、C E T P および第 V I I 因子の人種分布を示す。

図 1 6 は、p 5 3 - R b 経路およびその経路の種々の因子の間の関連を示す。

図 1 7 は、ここで記載したプロセスおよびデータベースを提供する、コンピューター構築性のブロックダイアグラムであり、これは、ここで提供するデータベースを蓄積し、そして分類するための、そしてここで提供する方法を実行する、典型的なコンピューターシステムを示す。

40

【 0 0 3 2 】

図 1 8 は、多型遺伝的マーカーを同定するためのデータベースへのアクセスを維持し、そして提供するための、図 1 7 に示されるコンピューターを使用して実行される、プロセッシングステップを示すフローダイアグラムを示す。

図 1 9 は、A K A P 1 0 - 1 遺伝子座について、年齢および性別で層化された、白人集団での対立遺伝子および遺伝子型分布を示すヒストグラムである。明緑色のバーは、4 0 歳より若齢の個体での頻度を示す。暗緑色のバーは、6 0 歳より高齢の個体での頻度を示す。

図 2 0 は、A K A P 1 0 - 5 遺伝子座について、年齢および性別で層化された、白人集団の対立遺伝子および遺伝子型分布を示すヒストグラムである。明緑色のバーは、4 0 歳

50

より若齢の個体の頻度を示す。暗緑色のバーは、60歳より高齢の個体の頻度を示す。

【0033】

図21は、h - m s r - A 遺伝子座について、年齢および性別で層化された白人集団の、対立遺伝子および遺伝子型分布を示すヒストグラムである。明緑色のバーは、40歳より若齢の個体での頻度を示す；暗緑色のバーは、60歳より高齢の個体での頻度を示す。

図22 A - D は、健康体のデータベースのために使用する、サンプルデータ収集質問票である。

【0034】

図23は、アッセイフラグメントからのセンス鎖およびアンチセンス鎖の遺伝子型決定をおこなうときに、図24のコンピューティング装置によって実行されるプロセッシングを示すフローチャートである。 10

図24は、本発明に従うシステムを示すブロックダイアグラムである。

図25は、本発明に従う、生物学的サンプルを同定する方法のフローチャートである。

図26は、質量分析計からのデータをグラフに表したものである。

図27は、質量分析法データのウェーブレット変換のダイアグラムである。

【0035】

図28は、ウェーブレットステージ0 hi データをグラフに表したものである。

図29は、ステージ0 ノイズプロファイルをグラフに表したものである。

図30は、ステージノイズ標準偏差の作成をグラフに表したものである。

図31は、データステージのスレシヨルド ( threshold ) の適用をグラフに表したものである。 20

図32は、希薄データセットをグラフに表したものである。

図33は、シグナルをシフトさせる式である。

図34は、ノイズを除去し、そしてシフトしたシグナルの、ウェーブレット変換をグラフに表したものである。

【0036】

図35は、ノイズを除去し、そしてシフトしたシグナルをグラフに表したものである。

図36は、ピークセクションの除去をグラフに表したものである。

図37は、ピークフリーシグナルの作成をグラフに表したものである。

図38は、ベースライン補正を作成する方法のブロックダイアグラムである。 30

図39は、ベースラインおよびシグナルをグラフに表したものである。

図40は、ベースラインを除去したシグナルをグラフに表したものである。

図41は、圧縮したデータを示す表である。

図42は、データを圧縮する方法のフローチャートである。

図43は、マスシフティングをグラフに表したものである。

【0037】

図44は、ピーク幅の決定をグラフに表したものである。

図45は、ピークの除去をグラフに表したものである。

図46は、ピークを除去したシグナルをグラフに表したものである。

図47は、残留ベースラインをグラフに表したものである。 40

図48は、残留ベースラインが除去されたシグナルをグラフに表したものである。

【0038】

図49は、ピーク高の決定をグラフに表したものである。

図50は、各ピークについて、シグナル - 対 - ノイズ決定をグラフに表したものである。

図51は、各ピークについて、残留エラーの決定をグラフに表したものである。

図52は、ピーク確率をグラフに表したものである。

図53は、ピーク確率に対する対立遺伝子比率の適用をグラフに表したものである。

【0039】

図54は、ピーク確率の決定をグラフに表したものである。 50

図 5 5 は、遺伝子型のコールをグラフに表したものである。

図 5 6 は、遺伝子型をコールするための統計的処理を示すフロチャートである。

図 5 7 は、標準を欠く遺伝子型決定をおこなうとき、図 1 のコンピューティング装置によって実行されるプロセッシングを示すフローチャートである。そして

図 5 8 は、標準を欠く遺伝子型プロセッシングのための、ピーク確率に対する対立遺伝子比率の適用をグラフに表したものである。

【 0 0 4 0 】

詳細な説明

特記しない限り、ここで使用するすべての技術的および科学的用語は、本発明の属する業界の当業者に一般に理解されるのと同じ意味を有する。本開示すべてにおいて、ここで引用される、すべての特許、出願、公開された出願および他の刊行物および GenBank および他のデータベースからの配列は、引用によってそのすべてが含まれる。

【 0 0 4 1 】

ここで使用するように、生体高分子は、核酸、タンパク質、多糖類、脂質および他の大分子を含むがこれらに限定されない。核酸は、DNA、RNA、およびそのフラグメントを含む。核酸はゲノム DNA、RNA、ミトコンドリア核酸、クロロプラスト核酸および別の遺伝的物質を有する他のオルガネラに由来し得る。

【 0 0 4 2 】

ここで使用するように、病的状態（率）（morbidity）は、状態、例えば、疾患または障害であって、生物、例えば、動物の健康および幸福を妥協する状態を言う。病的状態感受性または病的状態関連性遺伝子は、例えば、ヌクレオチド配列の変化によって変化されるとき、特異的な疾患臨床的表現型の発現を促進する（facilitate）遺伝子である。したがって、病的状態感受性遺伝子は、変化されるとき、生物が特定の疾患を発生させる、尤度または一般的なリスクを増加させるポテンシャルを有する。

【 0 0 4 3 】

ここで使用するように、死亡（率）（mortality）は、生物、特に動物が、予測される寿命全部を生存しない統計的尤度を意味する。

ゆえに、高い死亡（率）と関連する、特徴（trait）またはマーカー、例えば、多型は、集団のより若齢のセグメントよりもより高齢なもので、より低い頻度で観察される。

【 0 0 4 4 】

ここで使用するように、多型、例えば遺伝的変異は、集団中のゲノムの遺伝子の配列の変異、例えば対立遺伝子変異および発生し、または観察される他の変異を意味する。したがって、多型は、集団中の、2種またはより多い遺伝的に決定される別の配列または対立遺伝子の存在を意味する。これらの相違は、ゲノムのコードおよび非コード部分に存在することができ、そして例えば転写、プロセッシング、翻訳、輸送、タンパク質プロセッシング、トラフィック、DNA合成、発現されたタンパク質、他の遺伝子産物または生物学的経路のまたはポスト翻訳修飾の産物および集団のメンバー中に明示される任意の他の相違を含む、核酸配列、遺伝子発現での相違として明示され、または検出されることができる。単一ヌクレオチド多型（SNP）は、単一塩基変化、例えば、塩基の挿入、欠失または変化の結果として生じる多型を意味する。

【 0 0 4 5 】

多型マーカーまたは部位は、分岐が存在する遺伝子座である。そのような部位は、一塩基対（SNP）と同じくらい小さくてよい。多型マーカーは、制限フラグメント長多型、種々の数のタンデムリピート（VNTRの）、過剰可変領域、ミニサテライト、ジヌクレオチドリピート、トリヌクレオチドリピート、テトラヌクレオチドリピート、および他の反復パターン、単純配列リピートおよび挿入エレメント、例えばAluを含むがこれらに限定されない。多型形態は、また遺伝子について異なるメンデル的対立遺伝子として明示される。多型は、タンパク質の相違、タンパク質修飾、RNA発現修飾、DNAおよびRNAメチル化、遺伝子発現およびDNA複製を変化させる調節ファクター、およびゲノム核酸またはオルガネラ核酸の任意の他の明示または変化によって観察され得る。

10

20

30

40

50

## 【0046】

ここで使用するように、健康体集団は、疾患フリーである、動物、バクテリア、ウイルス、寄生生物、植物、ユーバクテリア、およびその他を含むがこれらに限定されない生物の集団を意味する。疾患フリーの概念は、選択される生物の関数 (function) である。例えば、哺乳動物について、それは、任意の疾患状態を明示しない対象を意味する。実際上、ヒトの場合に、健康体対象は、一般集団での最終的な使用のために、血液を献血する血液バンクの基準を通過するヒトドナーとして定義される。これらの基準は以下の通りである：

検出可能なウイルス性、バクテリア性、マイコプラズマ性、および寄生生物性感染症フリーで；貧血性でなく；それから質問票に関連する履歴 (history) に基づいてさらに選択される (図3参照)。

## 【0047】

したがって、健康体集団は、血液バンク基準にしたがって血液を献血するのに十分健康であり、任意の疾患状態についてさらに選択されていない、偏りのない集団を表す。典型的には、そのような個体は、いずれの明示も有しない。植物について、例えば、それは植物に関連する病害病理を明示しない植物集団である。バクテリアについて、それは環境ストレス、例えば選択物質、熱および他の病原なく複製しているバクテリア集団である。

## 【0048】

ここで使用するように、健康体のデータベース (または健康な患者データベース) は、任意の特定の疾患について予め選択されていない対象のプロファイルのデータベースを意味する。ゆえに、予め決定された基準にしたがって、健康であるように、データベースのデータのソースとして役立つ対象が選択される。特定の疾患または他の形質 (characteristic) を有する対象について予め選択された、他のそのようなデータベースと異なり、ここで提供するデータベースのための対象は、そのように選択されていない。

## 【0049】

また、対象が疾患または他の状態を明示しているならば、発見され、または特徴付けられた任意の多型が、無関係の疾患または状態に関連付けられるべきである。好ましい実施態様では、対象がヒトである場合、健康な対象は、疾患症候を明示せず、そして基準、例えば血液ドナーのために血液バンクによってセットされたものに適合する。

## 【0050】

したがって、データベースのための対象は、動物、植物、バクテリア、ウイルス、寄生生物および核酸を有する任意の他の生物または存在を含むがこれらに限定されない、任意の生物の集団である。好ましい対象には、哺乳動物、好ましいが必然的ではなくヒトが含まれる。そのようなデータベースは、集団の多様性を捕捉し、こうしてまれな多型の発見をすることができる。

## 【0051】

ここで使用するように、プロファイルは、すべての、年齢、性別、人種、疾患履歴、家族履歴、表現型的な形質、例えば、身長、体重および他の表れるパラメータを必須でなく含むがこれらに限定されないが、これらに関連する情報を意味する。サンプル収集情報フォームは、図22に示され、それはプロファイルの意図を例示している。

## 【0052】

ここで使用するように、疾患状態は、状態または異常または障害であって、遺伝され、または環境ストレス、例えば、毒物、細菌、真菌およびウイルス性感染から生じ得るものである。

## 【0053】

ここで使用するように、非選択的对象のセットは、共通の疾患または他の形質を有する予め選択されない対象を意味する。それらは、ここで定義されるように健康であるべきであるように選択することができる。

## 【0054】

ここで使用するように、表現型は、生物の任意の区別可能な特徴を含むパラメータのセ

10

20

30

40

50

ットを意味する。表現型は、身体的特徴であることができ、そして例えば、対象が動物である場合、精神的特徴、例えば感情的特徴であることができる。幾つかの表現型は、質問票によって導いた観察によって（例えば、図3および22参照）または先行する医学的および他の記録を参照することによって決定することができる。ここでの目的のために、表現型は、データベースが分類（sort）されることができる周辺のパラメータである。

**【0055】**

ここで使用するように、パラメータはデータベースを分類するための基礎として役立つ任意の入力データである。これらのパラメータは、表現型特徴、病歴、家族履歴および対象から引き出され、または対象について観察される任意の他のそのような情報を含む。パラメータは、対象、対象によって経験されるある種の履歴または現在の環境的または社会的影響、または対象に関連する者に及ぼす状態または環境的影響を説明し得る。パラメータは、ここで記載され、そして当業者に既知の任意のものを含むがこれらに限定されない。

10

**【0056】**

ここで使用するように、ハプロタイプは、単一のDNA鎖に位置する2またはより多い多型を意味する。ゆえに、ハプロタイピングは、単一のDNA鎖上の2またはそれより多い多型の同定を意味する。ハプロタイプは、表現型を示すことができる。幾つかの障害については、単一の多型は、特徴を表すために十分で有り得る。その他の場合には、複数のもの（すなわちハプロタイプ）が必要とされ得る。ハプロタイピングを、核酸を単離し、そして鎖を分離することによって行うことができる。加えて、酵素、例えば、各鎖から異なるサイズのフラグメントを生成する、ある種のヌクレアーゼを使用するとき、鎖分離はハプロタイピングに必要ではない。

20

**【0057】**

ここで使用する、使用するように、マススペクトルまたは質量分析法的分析の参照を有するパターンは、形質分布およびシグナルの数（そのようなピークまたはそのデジタル表示）を意味する。

**【0058】**

ここで使用するように、マススペクトルのコンテキストおよびその分析におけるシグナルは、特定の質量を有する分子の数または相対数である、出力データを意味する。シグナルは、「ピーク」およびそのデジタル表示を含む。

30

**【0059】**

ここで使用するように、アダプターは、Fenリガーゼを使用するハプロタイピングを引用して使用するときに、所望の多型に特異的にハイブリダイズする核酸を意味する。アダプターは、部分的に二本鎖であることができる。アダプターがその標的にハイブリダイズするとき、アダプター複合体が形成される。

**【0060】**

ここで使用するように、標的核酸は、サンプルの所望の任意の核酸を意味する。それは1または2以上のヌクレオチドを含むことができる。

**【0061】**

ここで使用するように、標準を欠く分析は、内部標準に基づく決定を意味する。例えば、多型の頻度は、単一のマススペクトル内のシグナルを比較することによってここに決定することができる。

40

**【0062】**

ここで使用するように、増幅は、生体高分子、特に核酸の量を増加させる手段を意味する。選択される5'および3'プライマーに基づき、増幅は、そしてまた分析の対象であるゲノムの領域を制限し、そして定義するように役立つ。増幅は、当業界で既知の任意の手段によってなされることができ、ポリメラーゼ連鎖反応（PCR）等の使用を含む。多型の頻度を決定することが要求されるとき、増幅、例えばPCRは、定量的にされなければならない。

**【0063】**

50

ここで使用するように、切断は、生体高分子の非特異的および特異的フラグメント化を意味する。

ここで使用するように、マルチプレックシングは、1より多い多型の同時的な検出を意味する。マルチプレックシング反応を実施するための方法は、質量分析法と組み合わせて、実際的に既知である（例えば米国特許番号6043031、5547835および国際PCT出願番号97/37041参照）。

【0064】

ここで使用するように、質量分析法の引用は、当業者に既知の、任意の適当な質量分析法形式を含む。そのような形式は、マトリックス介助レーザー脱離/イオン化飛行時間型（MALDI-TOF）、エレクトロスプレー（ES）、IR-MALDI（例えば公開された国際PCT出願番号99/57318および米国特許番号5118937参照）、イオンサイクロトロン共鳴（ICR）、フーリエ変換およびそれらの組み合わせを含むがこれらに限定されない。MALDI、特にUVおよびIRが好ましい形式に含まれる。

10

【0065】

ここで使用するように、マススペクトルは、質量分析法によってグラフ的に、または数値的にコード化された生体高分子またはそのフラグメントを分析することから取得されたデータを表したものを意味する。

【0066】

ここで使用するように、血液コンポーネントは、血液から分離されるコンポーネントを意味し、そして赤血球細胞および血小板、血液凝固因子、血漿、酵素、プラスミノゲン、免疫グロブリンを含むがこれらに限定されない。細胞性血液コンポーネントは、血液のコンポーネント、例えば、細胞である赤血球細胞である。血液タンパク質は血液中に通常見出されるタンパク質である。そのようなタンパク質の例は、血液第VIIおよび第VII因子である。そのようなタンパク質およびコンポーネントは、当業者に周知である。

20

【0067】

ここで使用するように、血漿を、当業者に既知の任意の方法によって分離することができる。例えば、それは赤色細胞にペレット化する力で血液を遠心分離することによって調製することができ、そして赤色細胞およびパフィーコート（buffy coat）（それは白血球を含み、さらにそれは血漿である）の間の界面を形成する。例えば、典型的な血小板濃縮物は少なくとも10%の血漿を含む。

30

【0068】

血液は、そのコンポーネント（血漿、血小板および赤血球細胞を含むがこれらに限定されない）に、当業者に既知の任意の方法によって分離することができる。例えば、血液は、十分な時間、そして十分な加速度で遠心分離し、赤血球細胞を含むペレットを形成することができる。白血球は、ペレットおよびパフィーコート領域の上清の界面で、おもに収集する。血漿、血小板、および他の血液コンポーネントを含む上清は、次いで除去され、そしてより高い加速度で遠心分離され、それによって血小板ペレットが得られ得る。

【0069】

ここで使用するように、p53は、DNA損傷を評価し、そして細胞成長、DNA修復およびアポトーシスを制御する、転写因子調節遺伝子として働く、細胞周期コントロールタンパク質である。p53突然変異は、種々の頻度を有するすべての異なる型の白血球を含む、広く種々の異なる癌に見出される。通常のp53機能の喪失は、ゲノムの不安定性および宿主細胞のコントロールされない成長を生じる。

40

【0070】

ここで使用するように、p21は、通常細胞のG1相アレストと関連する、サイクリン依存性キナーゼインヒビターである。発現トリガーアポトーシスまたはプログラムされた細胞死は、またWilmの腫瘍、小児性腎臓癌と関連された。

【0071】

ここで使用するように、第VII因子は独特な血液凝固カスケードと関係するセリンプロテアーゼである。この因子は、トロンビンによって活性化され、そして組織因子（第I

50

II因子)と、第X因子の第Xa因子へのプロセッシングにおいて働く。証拠は、遺伝子の多型と、心筋梗塞を含む、虚血性心臓血管疾患の高リスクを生じ得るVII因子活性の増加の間の連関を支持した。

#### 【0072】

ここで使用するように、リレーショナルデータベースは、データの裂およびコラムを含む、例えば2次元の表、またはより高次元のマトリックスのような、マトリックスを表す形態で情報を蓄積する。例えば、1つの実施態様では、リレーショナルデータベースは、それぞれがパラメータを有する別々の表を有する。表は、記録された数と連結され、それはまたインデックスとして作用する。データベースは、表のデータを使用することによってサーチし、または分類することができ、そして任意の適当な蓄積媒体、例えばフロッピー(登録商標)ディスク、CD-ROMディスク、ハードドライブまたは他の適当な媒体に蓄積される。

10

#### 【0073】

ここで使用するように、バーコードは、任意の望まれる大きさおよび形状の、所望により読み取り可能なマークの任意のアレイを意味し、それらは、必ずしも必須ではないのだが、好ましくは、1または2以上のコラムおよび1または2以上の列の参照コンテキストまたはフレームに編成される。ここでの目的のために、バーコードは、必ずしも「バー」である必要はないがドット、文字または任意のシンボル(複数含む)を含み得る、任意のシンボロジーを意味する。

#### 【0074】

ここで使用するように、シンボロジーは、サンプルに連結される、識別子(identifier)コードまたはシンボル、例えば、バーコードを意味する。インデックスは、それぞれのそのようなシンボロジーに参照符を付する。シンボロジーは、使用者によって知られ、または設計される任意のコードである。シンボルは、データベースに蓄積される情報と連関される。例えば、各サンプルを、コード化されたシンボロジーで独特に同定することができる。パラメータ、例えば、質問に対する回答および、サンプルの分析により取得される次の遺伝型および他の情報をデータベースに含ませ、そしてシンボロジーと連関させる。データベースを、任意の適当な記録媒体、例えば、ハードドライブ、フロッピー(登録商標)ディスク、テープ、CD-ROM、DVDディスクおよび任意の他の適当な媒体に蓄積する。

20

30

#### 【0075】

データベース

ヒトの遺伝子型決定は、現在は、疾患組織のサンプルを提供する、病院、組織バンクおよびリサーチ機関との共同に依存している。このアプローチは、疾患の開始および/または進行を、多型または他の遺伝的マーカーの存在と相関させることができるという概念に基づいている。このアプローチは、疾患が特異的なマーカーの存在および特異的なマーカーの不存在と相関しているべきものである、とは考えない。

#### 【0076】

マーカーの出現および消失の同定およびスコア化は、これらのマーカーが健康体対象のバックグラウンドで測定され、ここで、疾患の開始が多型の存在の変化をマスクしないときのみ、可能であることがここで示される。疾患集団からの情報のデータベースは、サンプルの大きさが小さい点、選択の偏り、および不均質性をこうむっている。健康体集団からここで提供したデータベースは、サンプルバンドが大きい点、簡単な選択方法および不均質性を希釈することによってこれらの問題を解決する。

40

#### 【0077】

非選択性、特に健康な対象と連関した、パラメータの第1のデータベースをここで提供する。また、データベースと、各対象から取得されるインデックス化されたサンプルの組み合わせも提供する。さらに、第1のデータベースから作成されるデータベースを提供する。これらは、本来のパラメータ情報に加えて、例えば、サンプルに由来する、ゲノム配列情報を含むがこれらに限定されない遺伝子型情報を含む。

50

## 【0078】

ここで健康体データベースと命名したデータベースは、特定の疾患について予め選択された対象から取得されないため、そのように命名する。ゆえに、個々のメンバーは疾患を有しているかもしれないが、個体の収集は特定の疾患を有するように選択されているわけではない。

## 【0079】

パラメータを取得する対象は、好ましくはすべての集団にわたり、ランダムに選択されている対象のセット、または疾患フリーまたは健康であるように予め選択される、対象のセットのいずれかを含む。結果として、データベースは、任意の予め選択された表現型、遺伝子型、疾患または他の形質を表すようには選択されない。典型的には、データベースが調製される対象の数は、ここで提供する方法に使用するとき、統計的に有意な結果が得られるように選択される。

10

## 【0080】

好ましくは、対象の数は、100より多く、より好ましくは200より多く、なおより好ましくは、1000より多い。正確な数は、データベースを分類するために使用するべきパラメータ（複数含む）の頻度に基づいて経験的に決定することができる。一般に、集団は、少なくとも50、少なくとも100、少なくとも200、少なくとも500、少なくとも1000、少なくとも5000、または少なくとも10000またはより多い対象を有することができる。

## 【0081】

対象の収集物を同定し、各対象についての情報を記録し、そしてデータベースとして各対象と関連させる。各対象と関連している情報は、対象の履歴的形質に関連する情報、表現型的形質およびまた遺伝型形質、医学的形質および任意の他の特徴および決定することのできる対象についての形質を含むがこれらに限定されない。この情報は、データベースを分類するための基礎として役立つ。

20

## 【0082】

例示的实施態様では、対象は、哺乳動物、例えば、ヒトであり、そして情報は、1または2以上のパラメータ、例えば年齢、性別、病歴、人種および任意の他の要因と関連する情報である。そのような情報は、動物が例えばヒトであるとき、質問票によって、そして個体についての観察、例えば髪の色、目の色、および他の形質によって取得することができる。遺伝子伝型情報は、対象からの組織または他の体および体液サンプルから取得することができる。

30

## 【0083】

健康体ゲノムデータベースは、血液サンプルのライブラリーから健康体個体からのプロファイルおよび多型を含むことができ、ここでライブラリーの各サンプルは、個体および分離血液または他の組織サンプルである。データベースの各サンプルは、ドナーの性別、年齢、人種群、および病歴についてプロファイル化される。

## 【0084】

データベースは、最初に対象の健康体集団を同定し、そして各対象についてデータベースのパラメータを分類するように役立つ情報を取得することによって作成する。この情報を好ましくは蓄積媒体、例えば、コンピューターのメモリーに入力する。

40

## 【0085】

データベースを作成するために使用する集団の各対象について取得された情報を、コンピューターメモリーまたは他の適当な蓄積媒体に蓄積する。この情報を各対象と関連された識別子 (identifier) と連結させる。ゆえに、データベースは、対象を、例えば、バーコードの表すデータポイントによって同定し、それから個体について、すべての情報、例えば、質問票からの情報をデータポイントと関連させる。情報が収集されるように、データベースは作成される。

## 【0086】

したがって、例えば、プロファイル情報、例えば、質問票から取得される対象履歴を、

50

データベースに収集する。得られたデータベースは、標準的ソフトウェアを使用し、例えば、年齢、性別および/または人種によって、望まれるように分類することができる。サンプルが取得されるべき対象のための例示的質問票を、図 2 2 A - D に示す。各質問票を、好ましくはバーコード、特にデータベースへの入力のため機械読み取り可能バーコードによって同定する。対象がデータを提供し、そして健康であると思われる後（すなわち、献血のための基準に適合する）、質問票のデータをデータベースに入力し、そしてバーコードと関連させる。組織、細胞または血液サンプルを対象から取得する。

【 0 0 8 7 】

図 4 は、血液サンプルコンポーネントのプロセッシングおよび追跡（tracking）を例示する。各コンポーネントをバーコードで追跡し、日付を付し、データベースに入力し、そして対象および対象のプロファイルと関連させる。典型的には、全血を遠心分離し、血漿、赤血球細胞（ペレット性である）および間の層である Buffy コートに見出される白血球を生成する。種々のサンプルを取得し、そしてバーコードでコード化し、そして必要な使用のために蓄積する。

10

【 0 0 8 8 】

サンプルを対象から収集する。サンプルは、組織、細胞、および流体、例えば核酸、血液、血漿、羊水、滑膜液、尿、唾液、水性体液、汗、精液サンプルおよび脳髄液を含むがこれらに限定されない。サンプルの特定のセットは集団の生物に依存することが理解される。

【 0 0 8 9 】

ひとたびサンプルを取得すると、収集物を好ましい実施態様では貯蔵することができ、各サンプルを識別子、特に機械読み取り可能コード、例えばバーコードでインデックスを付する。分析のために、サンプルまたはサンプルのコンポーネント、特に生体高分子および小分子、例えば核酸および/またはタンパク質およびメタボライトを単離する。

20

【 0 0 9 0 】

サンプルの分析後、この情報を蓄積媒体のメモリーにおいてデータベースに入力し、そして各対象と関連させる。この情報は、遺伝子型情報を含むがこれらに限定されない。特に、多型を示す核酸配列情報および他の情報、例えば、PCR フラグメントの質量、ペプチドフラグメント配列または質量、生体高分子および小分子のスペクトルおよび遺伝子、遺伝子産物または集団内の多型の存在が推論されることができ他のマーカーの構造または機能の他の印。

30

【 0 0 9 1 】

例示の実施態様では、データベースは、血液サンプルの収集物に由来することができる。例えば、図 1（また図 1 0 参照）は、5 0 0 0 個体サンプルを超える収集物の状態を示す。サンプルを、S O P（標準的取り扱い処理）ガイドラインにしたがって実験室でプロセッシングした。任意の標準的血液プロセッシングプロトコルを使用し得る。

【 0 0 9 2 】

ここで記載の例示的データベースのために、以下の基準を使用して対象を選択した：  
感染性物質についてはテストはしない。

年齢：少なくとも 1 7 歳

体重：最小で 1 1 0 ポンド

永久に不適格：

肝炎の履歴（1 1 歳より後）

白血病リンパ腫

ヒト免疫不全ウイルス（H I V）、A I D S

慢性腎臓疾患

40

【 0 0 9 3 】

一時的不適格：

妊娠 - 出産、流産、中絶後 6 週まで

大きな外科手術または点滴 - 1 年間

50

単球増加症 - 完全な回復まで

全血献血のまえ - 8 週間

注射による抗生物質、1 週間；口により、4 8 時間、

皮膚複合感染のための抗生物質を除く；

年繰り延べ

もし除去されないなら、内部癌および皮膚癌は、なお、そして再発はない

これらは献血について血液バンク基準に対応し、そしてヒト健康体データベースについてここで定義されるような健康体集団を表す。

【0094】

データベースの構造

当業者に既知の任意の適当なデータベース構造およびフォーマットを使用し得る。例えば、リレーショナルデータベースは好ましいフォーマットであり、そこではデータが各対象を同定するインデクサー (indexer) によって連結されたパラメータのマトリックスまたは表として蓄積される。データベースを分類することを含む、調製および操作のためのソフトウェアを、商業的に利用可能であるソフトウェア、例えば、マイクロソフトアクセスから容易に開発し、または適合させることができる。

【0095】

質コントロール

質コントロール処理を実行することができる。例えば、サンプルの収集の後、バンクの収集物の質を評価することができる。例えば、混合したサンプルを、既知のマーカ、例えば性別についてのテストによってチェックすることができる。サンプルの人種による分離の後、サンプルを特定の人種、例えば、HLA DQA 1 群特異的コンポーネントと関連するマーカについてランダムにテストし、サンプルが適当に人種群によって分類されたかどうかを評価する。例示的サンプルバンクを図 4 に示す。

【0096】

データベースのための遺伝子型データおよび他のパラメータの取得

情報および履歴パラメータのデータベースへの入力の後、各対象から取得されるサンプルからの材料を分析する。分析する材料は、タンパク質、メタボライト、核酸、脂質および任意の他の望まれる材料の構成物を含む。例えば、核酸、例えば、ゲノム DNA を配列決定によって分析することができる。

【0097】

配列決定を当業者に任意の既知の方法を使用して、行うことができる。例えば、多型が同定され、または既知であり、そしてデータベースの対象のその頻度または存在を評価することが望まれるなら、各サンプルからの所望の領域を、例えば、PCR または制限フラグメント、ハイブリダイゼーションまたは当業者に既知の他の適当な方法によって単離することができる、そして配列決定する。ここでの目的のために、配列決定分析は、好ましくは質量分析法 (例えば、米国特許番号 5 5 4 7 8 3 5、5 6 2 2 8 2 4、5 8 5 1 7 6 5、および 5 9 2 8 9 0 6 参照) を使用して実施する。核酸は、ハイブリダイゼーションによって、また配列決定することができ (例えば米国特許番号 5 5 0 3 9 8 0、5 6 3 1 1 3 4、5 7 9 5 7 1 4 参照)、これらは質量分析法による分析を含む (米国出願番号 0 8 / 4 1 9 9 9 4 および 0 9 / 3 9 5 4 0 9 参照)。

【0098】

他の検出方法では、対立遺伝子変異体を同定する前に、まず増幅することが必要である。増幅は、当業界で既知の方法にしたがって、例えば PCR および / または LCR によって行うことができる。ある実施態様では、細胞のゲノム DNA を 2 種の PCR プライマーに曝露し、そして望まれる量の増幅された DNA を生成するのに十分な何回かのサイクルの増幅に付する。好ましい実施態様では、プライマーは、1 5 0 ないし 3 5 0 塩基対離れて位置する。

【0099】

別の増幅方法は、下記のものを含む：

10

20

30

40

50

自己持続性配列複製 (Guatelli, J. C. et al., 1990, Proc. Natl. Acad. Sci. U.S.A. 87:1874-1878)、転写的増幅系 (Kwoh, D. Y. et al., 1989, Proc. Natl. Acad. Sci. U.S.A. 86: 1173-1177)、Q-Beta Replicase (Lizardi, P. M. et al., 1988, Bio/Technology 6:1197)、または任意の他の核酸増幅方法、次いで当業界で周知の方法を使用する増幅された分子の検出。これらの検出スキームは、そのような分子が非常に少量で存在するならば、核酸分子の検出のために特に有用である。

#### 【0100】

核酸は、また検出方法およびプロトコル、特に質量分析法によるものによって分析することができる (例えば、米国特許番号 5605798、6043031、共係属している、米国出願番号 08/744481、米国出願番号 08/990851、国際 PCT 出願番号 WO99/31273、国際 PCT 出願番号 WO98/20019 参照)。これらの方法は、自動化することができる (例えば、共係属している、米国出願番号 09/285481 および公開された国際 PCT 出願番号 PCT/US00/08111 であって自動化されたプロセスラインを記載するもの参照)。

10

#### 【0101】

ここでの分析の方法の好ましいものは、検出のため質量分析法でのプライマーオリゴ塩基伸長 (PROBE) 反応に関係するものを含む (ここで記載され、そして他に、例えば、米国特許番号 6043031 参照; また米国出願番号 09/287681、09/287682、09/287141 および 09/287679、共係属している米国出願番号 08/744481、国際 PCT 出願番号 WO98/20019 として公開され、そして米国出願番号 08/744481、08/744590、08/746036、08/746055、08/786988、08/787639、08/933792、08/746055、08/786988 および 08/787639 に基づく、国際 PCT 出願番号 PCT/US97/20444 参照; また米国出願番号 09/074936、米国特許番号 6024925、および米国出願番号 08/746055 および 08/786988 および公開された国際 PCT 出願番号 WO98/20020 参照)。

20

#### 【0102】

分析を実施するための好ましいフォーマットは、固体支持体、例えば、シリコンまたはシリコンコート基質に生体高分子が連結されているチップに基づくフォーマット、好ましくはアレイの形態である。より好ましくは、分析を、質量分析法、特に MALDI を使用して実施するとき、小さいナノリットルレベルの体積のサンプルを負荷し、こうして得られるスポットはレーザースポットのサイズくらい、またはより小さい。これは達成されるとき、質量分析分析から得られる結果は定量的であることが見出される。得られる質量スペクトルのシグナルの面積は濃度に比例的である (バックグラウンドについて標準化され、そして補正されるとき)。

30

#### 【0103】

そのようなチップを調製し、そして使用する方法は、米国特許番号 6024925、共係属している、米国出願番号 08/786988、09/364774、09/371150 および 09/297575 に記載され; また米国出願番号 PCT/US97/20195 であって WO98/20020 として公開されているもの参照。これらの分析を実施するためのチップおよびキットは、SEQUENOM から商標 MassARRAY で商業的に入手可能である。MassArray は、結果をより迅速に送達するために、ミニチュア化されたアレイおよび MALDI-TOF (マトリックス介助レーザー脱離イオン化 - 飛行時間型) 質量分析法と結合された酵素的プライマー伸長反応の忠実性による。それは、タグのない遺伝的変異体と関連した DNA フラグメントのサイズの単一塩基変化を正確に区別する。

40

#### 【0104】

ここで提供する方法は、対立遺伝子の定量的な決定を可能とする。マススペクトルのシグナルの面積は、定量的な決定のために使用することができる。頻度は、すべてのスペクトルの総面積に対するシグナルの比率 (ratio) から決定され、バックグラウンドについ

50

て補正される。これは前記出願に記載されたような P R O B E 技法により可能であり、引用によりここに含ませる。

【0105】

核酸を分析するさらなる方法は、ポリメラーゼ連鎖反応 ( P C R )、リガーゼ連鎖反応 ( L C R )、ミニ P C R、ローリングサークル増幅、自己触媒的方法、例えば、Q レプリカーゼ、T A S、3 S R、および当業者に既知である任意の他の適当な方法を含む、増幅に基づく方法を含む。

【0106】

多型の分析および同定および検出のための他の方法は、対立遺伝子特異的プローブ、サザン分析、および他のそのような分析を含むがこれらに限定されない。

10

【0107】

以下に記載の方法は、質量分析法がフラグメント混合物を分析するために使用されるとき、所望の増幅され、または非増幅性ヌクレオチド配列をフラグメント化し、それによって質量シグナルのセットを生成する方法を提供する。増幅されたフラグメントは、標準的ポリメラーゼ連鎖反応 ( U S 4 6 8 3 1 9 5 および 4 6 8 3 2 0 2 ) によってもたらされる。フラグメント化方法は、D N A の一本鎖または二本鎖を切断する酵素および D N A をライゲートする酵素の使用に関係する。切断酵素はグリコシラーゼ、ニックターゼ、および部位特異的および部位非特異的ヌクレアーゼであることができ、グリコシラーゼ、ニックターゼ、および部位特異的ヌクレアーゼが最も好ましい。

【0108】

20

グリコシラーゼフラグメント化方法

D N A グリコシラーゼはある種のタイプのヌクレオ塩基を所望の D N A フラグメントから特異的に除去する。これらの酵素は、それによって脱塩基 (abasic) 部位を生成する事ができ、それらは、脱塩基部位で特異的にさらされたフォスフェート骨格を切断し、そして配列を示すヌクレオ塩基特異的フラグメントのセットを生成する、さらなる切断酵素によって、またはアルカリ溶液およびまたは熱のような化学的手段によって認識することができる。D N A グリコシラーゼおよびその標的化されたヌクレオチドのある組み合わせの使用は、任意の所望の標的領域の塩基特異的シグナチュアパターンを生成するために十分である。

【0109】

30

多くの D N A グリコシラーゼが既知であり、酵素の修飾ヌクレオチドまたはヌクレオチドアナログ標的に対応し、例えば、D N A グリコシラーゼはウラシル D N A グリコシラーゼ ( U D G )、3 - メチルアデニン D N A グリコシラーゼ、3 - メチルアデニン D N A グリコシラーゼ I I、ピリミジンヒドレート D N A グリコシラーゼ、F a P y D N A グリコシラーゼ、チミジンミスマッチ D N A グリコシラーゼ、ヒポキサンチン D N A グリコシラーゼ、5 - ヒドロキシメチルウラシル D N A グリコシラーゼ ( H m U D G )、5 - ヒドロキシメチルシトシン D N A グリコシラーゼ、または 1 , N 6 - エテノアデニン D N A グリコシラーゼであることができる ( 例えば米国特許番号 5 5 3 6 6 4 9、5 8 8 8 7 9 5、5 9 5 2 1 7 6 および 6 0 9 9 5 5 3、国際 P C T 出願番号 W O 9 7 / 0 3 2 1 0、W O 9 9 / 5 4 5 0 1 参照 ; また Eftedal et al. (1993) Nucleic Acids Res 21:2095-2101, Bjelland and Seeberg (1987) Nucleic Acids Res. 15:2787-2801, Sapparbaev et al. (1995) Nucleic Acids Res. 23:3750-3755, Bessho (1999) Nucleic Acids Res. 27:979-983 参照 )。好ましいグリコシラーゼはウラシル D N A グリコシラーゼ ( U D G ) である。

40

【0110】

例えばウラシルは、通常の D N A 前駆体ヌクレオチド ( 例えば、d C T P、d A T P、および d G T P ) および d U T P の存在下で、D N A を増幅することによって増幅 D N A 分子に組み込むことができる。増幅された産物が U D G で処理されるとき、ウラシル残基が切断される。U D G 反応からの産物の次の化学的処理によって、フォスフェート骨格の切断およびヌクレオ塩基特異的フラグメントの生成を得る。さらに、グリコシラーゼ処理の前に、増幅産物の相補鎖の分離により、生成されるべきフラグメント化の相補的パター

50

ンを得る。

【0111】

したがって、dUTPおよびウラシルDNAグリコシラーゼの使用によって、相補鎖のT特異的フラグメントの生成が可能となり、したがってこれは、所望の配列内のTおよびA位置の情報を提供する。これと同様に、両方の（相補的）鎖（すなわちC特異的グリコシラーゼ）のC特異的反応は、もし両方の増幅鎖のフラグメント化パターンが別個に分析されるならば、所望の配列内のCおよびG位置の情報をもたらす。したがって、グリコシラーゼ法および質量分析法によって、A、C、GおよびT特異的フラグメント化パターンの全シリーズが分析されることができる。

【0112】

ニッカーゼフラグメント化法

DNAニッカーゼ、またはDNAアーゼを使用して、DNA二重らせんの1の鎖を認識し、そして切断することができる。多くのニッカーゼが既知である。これらのうちに、例えば、ニッカーゼNY2AニッカーゼおよびNYS1ニッカーゼ（Megabase）があり、以下の切断部位を有する；

NY2A：5'・・・R AG・・・3'

3'・・・Y TC・・・5'、ここで、R = AまたはGおよびY = CまたはT

NYS1：5'・・・CC[A/G/T]・・・3'

3'・・・GG[T/C/A]・・・5'。

【0113】

Fenリガーゼフラグメント化法

Fenリガーゼ法は、2種の酵素：Fen-1酵素およびリガーゼに関係する。Fen-1酵素は“フラップ”エンドヌクレアーゼとして知られる部位特異的ヌクレアーゼである（米国特許5843669、5874283および6090606）。この酵素は、標的DNA鎖にハイブリダイズされる2種のオリゴヌクレオチドのオーバーラップによって創出されるDNA“フラップ”を認識し、そして切断する。この切断は高度に特異的であり、そして単一塩基対突然変異を認識することができ、これにより所望のあるSNPの個々のヘテロ接合性から単一ホモログの検出、それからフラグメント内に存在する他のSNPのそのホモログの遺伝子型決定を可能とする。Fen-1酵素は、Fen-1様ヌクレアーゼ、例えばヒト、ハツカネズミ、およびツメガエル属XPG酵素および酵母RAD2ヌクレアーゼまたは例えば、*M. jannaschii*、*P. furiosus*、および*P. woesei*からのFen-1エンドヌクレアーゼであることができる。好ましい酵素には、Fen-1酵素が含まれる。

【0114】

リガーゼ酵素は、2つの二本鎖核酸フラグメントの間のホスホジエステル結合を形成する。リガーゼは、DNAリガーゼIまたはDNAリガーゼIIIであることができる（例えば、米国特許番号5506137、5700672、5858705および5976806参照；またWaga. et al. (1994) *J. Biol. Chem.* 269: 10923-10934、Li et al. (1994) *Nucleic Acids Res.* 22:632-638、Arrand et al. (1986) *J. Biol. Chem.* 261: 9079-9082、Lehman (1974) *Science* 186: 790-797、Higgins and Cozzarelli (1979) *Methods Enzymol.* 68:50-71、Lasko et al. (1990) *Mutation Res.* 236:277-287、およびLindahl and Barnes (1992) *Ann. Rev. Biochem.* 61:251-281参照)。

【0115】

熱安定性リガーゼ（Epicenter Technologies）は、ここでの使用のための好ましいリガーゼに含まれ、ここで“熱安定性”は、DNAの2つの鎖を分離するために必要である温度への曝露の後、なお活性を保持しているリガーゼを意味する。

【0116】

タイプIIS酵素フラグメント化法

制限酵素は、特定の認識配列内、またはそれに隣接する特異的部位で二本鎖DNAに特異的に結合し、そして切断する。これらの酵素を、当業者に知られているように2種の群

10

20

30

40

50

に分類する（例えばタイプ I、II、および III）。タイプ I およびタイプ III 酵素の特性のために、これらは分子生物学的応用に広く使用されていない。したがって、本発明のためにタイプ II 制限酵素が好ましい。当業界で既知である何千もの制限酵素のうち、179種の異なるタイプ II 特異性がある。179種の独特のタイプ II 制限エンドヌクレアーゼのうち、31種が4塩基認識配列を有し、11種が5塩基認識配列を有し、127種が6塩基認識配列を有し、そして10種が6塩基より大きい認識配列を有する（米国特許5604098）。カテゴリータイプ II のうち、タイプ IIS が好ましい。

#### 【0117】

タイプ IIS 酵素は、Alw XI、Bbv I、Bce 83、Bpm I、Bsg I、Bsm AI、Bsm FI、Bsa I、Bcc I、Bcg I、Ear I、Eco 571、Esp 31、Fau I、Fok I、Gsu I、Hga I、Mme I、Mbo II、Sap I、等であることができる。好ましいタイプ IIS 酵素は Fok I である。

Fok I 酵素エンドヌクレアーゼは、例示的な周知の、タイプ IIS クラスの特徴を把握されたメンバーである（例えば、米国特許5714330、5604098、5436150、6054276 および 5871911 参照；また Szybalski et al. (1991) Gene 100:13-26、Wilson and Murray (1991) Ann. Rev. Genet. 25:585-627、Sugisaki et al. (1981) Gene 16:73-78、Podhajski and Szalski (1985) Gene 40:175-182 参照。Fok I は配列 5' GGATG - 3' を認識し、そしてそれにより DNA を切断する。タイプ IIS 制限部位は、DNA 標的を増幅するために使用されるプライマーにタイプ IIS 制限部位を組み込むことによって、DNA 標的に導入することができる。Fok I での消化によって生成したフラグメントは、部位特異的であり、そして質量分析法、例えば、MALDI-TOF 質量分析法、ESI-TOF 質量分析法、および当業者に周知である任意の他のタイプの質量分析法によって分析することができる。

#### 【0118】

かつて、多型が年齢のようなパラメータと相関していることが見出された。対立遺伝子脱落のための誤りの結果の可能性を、ゲノムの隣接領域で比較 PCR を行うことによって試験することができる。

#### 【0119】

分析

データベースの使用では、対立遺伝子頻度を集団にわたり、集団の各サンプルを個々に分析し、各個体サンプルの所望の対立遺伝子またはマーカーの存在または不存在を決定し、それから集団のマーカーの頻度を決定することによって決定することができる。次いで、このデータベースを分類し（層化し）、標準的統計的分析を使用して、対立遺伝子および選択されたパラメータの間の任意の相関を同定することができる。もし相関、例えば、年齢と特定のマーカーの減少または性別または他のパラメータとの相関が観察されるならば、そのときはそのマーカーがさらなる研究、例えば、関係する遺伝子または経路を同定するための遺伝的マッピングのための候補である。

#### 【0120】

次いで、例えば、マーカーは疾患と相関づけられ得る。ハプロタイピングを、また実施することができる。遺伝的マッピングを標準的方法を使用して実施することができ、そしてまた他のデータベースの使用、例えば、障害と関連されるべき予め決定されたデータベースを要し得る。

例示的な分析を実施し、そしてこれらを図に示し、そしてここで考察する。

#### 【0121】

サンプルをプールする

ここで提供するデータベース、またはそのような情報の任意の他のデータベースを使用して、それぞれのサンプルを別々に試験することによって取得された実質的に同じ頻度を、サンプルを、例えば、10、20、50、100、200、500、1000 または任意の他の数のバッチでプールすることによって取得することができることが見出される。

正確な数を必要ならば経験的に決定し、そして3より小さいことができる。

【0122】

ある実施態様では、遺伝子型のおよび他のマーカーの頻度を、サンプルをプールすることによって取得することができる。この標的集団を実施するために、そして評価すべき遺伝的変異を選択し、複数の生体高分子のサンプルを集団のメンバーから取得し、そしてマーカーまたは遺伝子型を推論することができる生体高分子を決定し、または検出する。プール中のテストするサンプルを個々に比較し、そしてその分類した結果を図9に示し、第VII因子対立遺伝子353Qの頻度を示す。図10は、プール性対個体サンプルにおける、CETP対立遺伝子CETPの頻度を示す。図15は、データを取得するためにプールしたDNAサンプルを使用する、データベース中の種々の人種の群のなかの人種の多様性を示す。図12-14はこれらのサンプルについてマススペクトルを示す。

10

【0123】

テストサンプルをプールすることは、ここで提供する健康体のデータベースに適用できるだけでなく、また疾患性集団に由来する典型的なデータベースを含む、対象および遺伝子型の情報の任意のデータベースへの入力のためのデータを集めるときに使用できる。ここで実証されることは、達成した結果は、もし各サンプルを別に分析するならば達成されるのと統計的に同じ結果であるという、発見である。方法、例えばここで提供する質量分析法によってプールしたサンプルの分析によって、そのようなデータの解決および結果の数量化が可能となる。

【0124】

第VII因子についてR53Q酸多型を評価する。図9では、“個体”データは92個体の反応物で観察された対立遺伝子頻度を表す。プールしたデータは、単一のプローブの反応にプールした同じ92個体の対立遺伝子頻度を表す。個体のドナーのサンプル中のDNAの濃度は、250ナノグラムである。プールしたサンプルのDNAの総濃度は、また250ナノグラムであり、ここですべての個々のDNAの濃度は、2.7ナノグラムである。

20

【0125】

スペクトルの質または検出するサンプルの量を定量する能力の何らの変化なく、2.7ナノグラムから0.27ナノグラムに、プールしたサンプルの個体のDNA濃度を減少させることが可能であることがまた示された。ゆえに、低濃度のサンプルを、このプールする方法で使用し得る。

30

【0126】

データベースおよびそれによって同定したマーカーの使用

ゲノムの首尾よい使用は、科学的な仮説（すなわち一般的な遺伝的変異、例えば、SNP）、研究計画（すなわち複合的な障害）、サンプルおよび技術、例えばチップに基づく質量分析分析（例えば、米国特許番号5605798、米国特許番号5777324、米国特許番号6043031、共係属している米国出願番号08/744481、米国出願番号08/990851、国際PCT出願番号WO98/20019、共係属している米国出願番号09/285481であって分析のための自動化されたプロセスラインを記載しているもの参照；また例えば、米国出願番号08/617256、09/287681、09/287682、09/287141および09/287679、共係属している米国出願番号08/744481、米国出願番号08/744481、08/744590、08/746036、08/746055、08/786988、08/787639、08/933792、08/746055、09/266409、08/786988および08/787639に基づき、国際PCT出願番号WO98/20019として公開された、国際PCT出願番号PCT/US97/20444参照；また米国出願番号09/074936参照）を要求する。これらの態様のすべては、ここで提供するデータベースおよび収集中のサンプルと結合させて使用することができる。

40

【0127】

それによって同定したデータベースおよびマーカーを、例えば、以前同定されていない

50

、または未知の遺伝的マーカーの同定のために、そして既知のマーカーの新規な使用を同定するために、使用することができる。同定されるマーカーとして、これらをデータベースに入力し、さらなる相関を決定し得るパラメータを分類するように使用するため、データベースに入力し得る。

【0128】

予め同定されていない未知の遺伝的マーカー

健康体のデータベースのサンプルを使用して、またデータベースの集団の多型を調査するときに、任意のマッピング、配列決定、増幅および他の方法論を使用し、新規の多型および遺伝的マーカーを同定することができる。次いで、こうして同定した多型を、各サンプルについてデータベースに入力することができ、そしてその多型を分類用パラメータとして使用して、データベースを分類（層化）することができ、例えば同定したマーカーの頻度の年齢相関性変化を表す、任意のパターンおよび相関を同定することができる。相関を同定するならば、マーカーの遺伝子座をマッピングし、そしてその機能または効果を評価し、または演繹することができる。

10

【0129】

こうして、ここにデータベースは以下のための手段を提供する：

集団で、年齢の増加にともなうマーカーの発生または消失を比較することによって、遺伝子的要因の有意に異なる対立遺伝子頻度を同定し、それからそのマーカーと疾患または生化学的経路を連関させること；

【0130】

男性を女性集団と比較し、または他の選択された層化された集団を比較することによって、疾患を生じる遺伝的要因の有意に異なる対立遺伝子頻度を同定し、そしてそのマーカーを疾患または生化学的経路と連関させること；

20

【0131】

異なる人種群を比較することによって、疾患を生じる遺伝的要因の有意に異なる対立遺伝子変異を同定し、そしてそのマーカーをその人種群に高頻度で存在すると知られている、疾患または生化学的経路と連関させること；

【0132】

年齢、性別、および人種起源にしたがって層化された、全般に任意交配された集団中の遺伝子の潜在的に機能的な変異体をプロファイルし、そしてそれによって調査した集団の身体的状態への変異体遺伝子の寄与を実証すること；

30

【0133】

年齢、性別、および人種起源にしたがって層化された、全般に任意交配された集団内で遺伝子不平衡分析を実施することによって機能的に関連のある遺伝子変異体を同定し、そしてそれによって調査した集団の身体的状態へのそれらの寄与を実証すること；

【0134】

年齢、性別、および人種起源にしたがって層化された、全般に任意交配された集団内で連鎖不平衡分析を実施することによって、潜在的に機能的な染色体または染色体の部分の変異体を同定し、そしてそれによって調査した集団の身体的状態に対するそれらの寄与を実証すること。

40

【0135】

同定したマーカーおよび既知のマーカーの使用

このデータベースを、また既知マーカーと組み合わせて使用し、任意の相関を同定するために分類し得る。例えば、データベースは、以下のために使用することができる；

医学的に関連のある多型マーカーの存在を、決定および評価すること；

【0136】

医学的に関連のある遺伝的要因の診断的な特異性を決定および評価すること；

医学的に関連のある遺伝的要因の肯定的な予測値を決定および評価すること；

原因となる遺伝的要因に関して、全体の集団内で、例えば、糖尿病、高血圧、自己免疫性疾患、アテローム性動脈硬化症、癌および他の疾患を含むがこれらに限定されない、複

50

合的な疾患の発生を決定し、および評価すること；

【0137】

予防的な疾患処置のための適当な戦略を描くこと；

原発性疾患介入のための適当なタイムラインを描くこと；

これらの一般的な利用可能性に関して単離した集団で同定された関連のある遺伝的要因を医学的に確認すること；

これらの一般的な利用可能性に関して単離された集団で同定されたすべての潜在的な標的構造を含む疾患経路を確認すること；および

これらの一般的な利用可能性に関して単離された集団内で同定された適当な薬物標的を確認すること。

10

【0138】

多型が連鎖し得る疾患および障害には、先天性の代謝の誤りに連鎖したもの、後天性の代謝的障害、介在性代謝、発癌経路、血液凝固経路、およびDNA合成および修復経路DNA修復/複製/転写因子および活性、例えば発癌、加齢および血液凝固に関係する遺伝子および血栓に関連する関連性生化学的経路、塞栓、発作、心筋梗塞、脈管形成および発癌を含み得る。

【0139】

例えば、多くの疾患が、中間代謝の酵素の不足または欠損によって引き起こされ、または関係し（例えば、以下の表1および2参照）、それは酵素基質を摂取すると、臓器および組織（特に、幼児の脳および他の臓器に発生し、精神的遅滞および他の発育障害を生じる）を損傷させる有害なメタボライトの蓄積を生じる。

20

【0140】

非常に興味ある障害のマーカーおよび遺伝子の同定  
モデル系

幾つかの遺伝子系、例えば、p21、p53およびリポタンパク質リパーゼ多型（N291S）を選択した。p53遺伝子は、多様な腫瘍タイプで突然変異している腫瘍抑制遺伝子である。ある共通する対立遺伝子変異体がコドン72に存在する。p53遺伝子で同定された多型、すなわちR72P対立遺伝子は、当該遺伝子のコドン72でアルギニンからプロリンへのアミノ酸の置換を生じる。

【0141】

疾患性集団を使用して、米国のアフリカ系アメリカ人および白人のなかのこれらの対立遺伝子の対立遺伝子分布の人種間相違があることが示された。この結果はここに、この発見を支持し、そしてまた健康体のデータベースから取得された結果が意義深いことを実証している（図7B参照）。

30

【0142】

291Sの対立遺伝子は、高密度リポタンパク質コレステロール（HDL-C）（これはアテローム性動脈硬化症、特に、心筋梗塞の男性の高いリスクと関連する）を生じる（Reymer et al. (1995) Nature Genetics 10:28-34参照）。

【0143】

両方の遺伝的多型が白人集団に基づくサンプルバンクの一部の中にプロファイルされた。リポタンパク質リパーゼ遺伝子に位置する多型について、1025人の非選択性個体（436人の男性および589人の女性）のすべてをテストした。ゲノムDNAを当該個体から取得した血液サンプルから単離した。

40

【0144】

実施例および図面に示すように、約5000の対象を含む、例示的データベースは、質問票に対する回答（図3参照）、そして遺伝子型情報を層化する。特定の既知の対立遺伝子を選択し、そして質量分析法を使用してマーカーについてサンプルをテストし、特定のPROBE（実施例参照）を使用して、各サンプル中の多型を同定した。データベース中の集団を、種々のパラメータにしたがって分類し、そして相関を観察した。例えば、図2A-Cは、データベース中の白人集団のリポタンパク質リパーゼ遺伝子についての、年齢

50

および性別によるデータの分類を示している。結果は、男性について年齢にともなってこの対立遺伝子の頻度は減少するが、女性ではそのような減少がないことを示している。データベースに対してテストした他の対立遺伝子は、p 53、p 21および第VII因子を含む。年齢によって分類したときの結果を図に示す。

【0145】

これらの例は、一般的な集団内の疾患を生じる遺伝的要因の変化した頻度の影響を実証している。これらの結果の科学的な解釈によって、多型性の遺伝的変化の医学的な関連性の予測が可能となる。さらに、これらの一貫性、診断的な特異性、肯定的な予測値、疾患の開始、予防戦略の最も適当な開始、および任意交配された集団について単離された集団で同定された遺伝的変化の一般的な利用可能性について結論を引き出すことができる。

10

【0146】

したがって、人種的に同質である、年齢および性別で層化された集団に基づくサンプルバンクは、それらの潜在的な医学的有用性に関して遺伝子要因の迅速な同定および確認のための適当なツールである。

【0147】

データベースシステムを創出し、分類し、そしてプロセッシングするための例示的コンピュータシステム

データベースを含む、コンピュータを含むシステムをここで提供する。コンピュータおよびデータベースを、例えば、APLシステム（共係属している米国出願番号09/285481参照）と組み合わせて使用することができる。これは、生体高分子、特に核酸を分析するための自動化されたシステムである。APLシステム由来の結果をデータベースに入力することができる。

20

【0148】

任意の適当なコンピュータシステムを使用し得る。コンピュータシステムをサンプル分析、例えば、ここで記載の自動化プロセスラインのためにシステムに組み込み得る（例えば、共係属している米国出願番号09/285481参照）。

【0149】

図17は、ここで記載のデータベースを提供し、そしてプロセッシングするための、構築されたコンピュータのブロック図である。データベースを維持し、そして方法および処理を実施するプロセッシングを、同様の構築を有するすべての複数コンピュータ上で実施し得、または単一の、総合的なコンピュータによって実施し得る。例えば、データをデータベースに加えるコンピュータを、データベースを蓄積するコンピュータから分離し得、またはそれによって総合化し得る。別の編成では、プロセッシングを実施するコンピュータは、図17に例示されるような構築を有し得る。

30

【0150】

図17は、前記のデータベースを維持する例示的なコンピュータ1700のブロック図であり、そしてこの方法および処理を実行する。各コンピュータ1700は、中央プロセッサユニット（CPU）1702、例えば、“ペンティアム”マイクロプロセッサの制御下で作動し、そしてインテルコーポレーション（Santa Clara, California, USA）から入手可能である総合的サーキットチップと関連している。コンピュータ使用者は、コマンドおよびデータをキーボードおよびディスプレイマウス1704から入力することができる、そしてディスプレイ1706で入力およびコンピュータ出力を見ることができ、ディスプレイは、典型的には、ビデオモニターまたはフラットパネルディスプレイ装置である。

40

【0151】

コンピュータ1700はまた、直接的アクセス蓄積装置（DASD）1707、例えば、固定ハードディスクドライブをも含む。メモリー1708は、典型的には、揮発性セミアダクチャーランダムアクセスメモリー（RAM）を含む。各コンピュータは好ましくは、プログラム製品リーダーがデータを読むことができる（そしてそれが所望によりデータを書き込むことができる）、プログラム製品蓄積装置1712を許容する、プログラ

50

ム製品リーダー 1710 を含む。

【0152】

プログラム製品リーダーは、例えば、ディスクドライブを含むことができ、そしてプログラム製品蓄積装置は取り外し可能蓄積媒体、例えば磁気フロッピー(登録商標)ディスク、光学的CD-ROMディスク、CD-Rディスク、CD-RWディスク、またはDVDデータディスクを含むことができる。要すれば、コンピューターは相互に、ネットワーク1713上の他の接続されたコンピューターと通信することができるように、接続することができる。各コンピューター1700は、ネットワークとコンピューターの間の接続1716上の通信を可能とする、ネットワークインターフェース1714を経由して、ネットワーク1713上の接続された他のコンピューターと通信することができる。

10

【0153】

コンピューター1700は通常のコンピューター構築物にしたがう、メモリー1708に一時的に蓄積される、プログラミングステップの制御下で作動する。プログラミングステップをCPU1702によって実行するとき、付属システムコンポーネントはそのそれぞれの機能を実行する。こうして、プログラミングステップは前記のようなシステムの機能性を履行する。プログラミングステップを、プログラム製品リーダー1712を経由して、またはネットワーク接続1716を経由して、DASD1707から受け取ることができる。

【0154】

蓄積ドライブ1710は、プログラム製品を受け取り、そこに記録されたプログラミングステップを読み、そしてCPU1702による履行のためにメモリー1708へプログラミングステップを送達することができる。前記のように、プログラム製品蓄積装置1710は、磁気フロッピー(登録商標)ディスクおよびCD-ROM蓄積ディスクを含む、記録されたコンピューター読み取り可能指令を有する、複数の取り外し可能媒体のいずれかを含むことができる。他の適当なプログラム製品蓄積装置は、磁気テープ、およびセミコンダクターメモリーチップを含むことができる。この方法で、作動のために必要なプロセッシングステップをプログラム製品上に埋め込むことができる。

20

【0155】

あるいは、プログラムステップを、ネットワーク1713上で作動しているメモリー1708へと受け取ることができる。このネットワーク方法では、ネットワーク通信がさらなる説明なくして当業者によって理解される周知の方法によってネットワーク接続1716上に確立された後に、ネットワークインターフェース1714を経由して、コンピューターがプログラムステップを含むデータをメモリー1708に受け取る。プログラムステップを次いで、ガメントデータベースシステムのプロセッシングを履行するために、CPU1702によって実行する。

30

【0156】

システムのすべてのコンピューターが好ましくは図17に示すのと同様の構築物を有し、したがって図17のコンピューター1700に関して記載された詳細がシステム1700のすべてのコンピューターにあてはまるのが理解されるべきである。このことは、ネットワーク1713に接続されて示される、複数のコンピューター1700によって指摘される。コンピューター1700のいずれかは、それらが他のコンピューターと通信することができるが、そしてここで記載の機能性を支持することができる限り、別の構築物を有する。

40

【0157】

図18は、図17に例示したコンピューターを使用し、データベース、例えば、多型遺伝的マーカーを同定するためのデータベースを維持し、そしてアクセスを提供して実施されるプロセッシングステップを例示する、フロー図である。特に、データベースに含まれる情報を、図17に例示されるのと同様の構築を有するコンピューターに蓄積する。図18に指摘のように、データベースを維持するための第1のステップは、集団の健康なメンバーを同定することである。前記の様に、集団メンバーは健康であることのみを基礎に選

50

択される対象であり、そしてここで、対象は哺乳動物、例えば、ヒトであり、彼らは／これらは好ましくは外観上の健康および検出可能な感染症の不存在に基づいて選択される。同定するステップは、1802の数を付したフロー図ボックスによって表す。

【0158】

1804の数を付したフロー図ボックスによって表される次のステップは、同定される集団のメンバーに関連する、同定のための、そして履歴的情報およびデータを取得することである。情報およびデータは、それぞれの集団メンバー、例えば、メンバーの年齢、人種、性別、医学的履歴、および最終的には遺伝子型情報についてのパラメータを含む。最初に、パラメータ情報を、体組織または体液サンプルがまた取得される、各メンバーによって回答された質問票から取得する。これらのパラメータをコンピューターのデータベースに入力し、そして蓄積するステップを、1806の数を付したフロー図ボックスによって表す。各集団についてのさらなる情報として、メンバーおよび対応するサンプルを取得し、この情報をデータベースに入力し、そして蓄積パラメータとして役立てることができる。

10

【0159】

1808の数を付したフロー図ボックスによって表した、次のステップでは、メンバーのパラメータをインデクサーと関連させる。このステップを、例えば、新規のデータ記録がリレーショナルデータベース構造にしたがって蓄積され、そしてその構造にしたがって他の記録と自動的に連結されるとき、データベース蓄積操作の一部として実行し得る。ステップ1806はまた、データベース入力が入力サーチにしたがってサーチされ、またはデータの寄与を決定するためのキー値をインデックス化する、通常の方法または修復プロセスの一部として履行し得る。

20

【0160】

例えば、そのようなサーチおよび分類技術を使用し、既知の遺伝的マーカーの存在にしたがい、それからそれらが企図していた、疾患との相関があるかどうか決定し得る。この使用の例は、p53およびリポタンパク質リパーゼ多型の頻度を評価するためのものである。

【0161】

そのようなデータベースのサーチはまた、集団ないの頻度が年齢、年齢、人種群、性別、またはある他の基準の関数として変化する、1または2以上の遺伝的マーカーを同定するために有益であってよい。これは以前未知の多型の同定および、最終的には疾患の開始および進行と関係する遺伝子または経路の同定をもたらすことができる。

30

【0162】

さらに、データベースは、同定された多型を取得するために使用することができ、そして選択されたパラメータにしたがってデータを分類するとき、頻度について変化するかどうか確認することができる。

【0163】

この方法では、ここで提供するデータベースおよび方法が、とりわけ、コンポーネント、特に疾患プロセスのキーコンポーネントの同定が、その遺伝的基盤の理解、およびまたプロセス、例えば個体の薬物応答の理解によって可能となる。ここで提供するデータベースおよび方法をまた、病理学的経路の解明に關係する方法で、新規な診断アッセイの開発で、新規の可能性のある薬物標的の同定、および新規薬物候補の同定で、使用することができる。

40

【0164】

病的状態および／または早期死亡関連性多型

何らの特定の疾患について選択されていない、健康な血液ドナーの集団によって提供された情報を含むデータベースを使用して、存在し、年齢にともない頻度が減少する多型および対立遺伝子を同定することができる。これらは病的状態感受性マーカーおよび遺伝子を表し得る。

【0165】

50

ゲノムの多型は遺伝子機能、タンパク質機能またはゲノム不安定性の変化につながるることができる。臨床的関連性/有用性を有するこれらの多型を同定することが、全世界の科学的努力の目標である。そのような多型の発見が、疾患を治療するための新規薬物化合物の同定および開発に基本的な影響を有することを期待することができる。しかし、種々の多型を同定するための戦略は、煩雑であり、そして非常に多数の患者、および疾患連関を示すためのコントロールコホートの利用可能性に依存する。特に、任意の疾患(病的状態感受性遺伝子)に罹患している集団の一般的リスクを生じる遺伝子により、これらのケース/コントロール研究から完全に免れる。

#### 【0166】

広範な異なる疾患に存在する、病的状態感受性遺伝子を同定するためのスクリーニング戦略をここに記載する。病的状態感受性遺伝子の定義は、多くの異なる細胞型または組織に発現されている遺伝子(ハウスキーピング遺伝子)であり、そしてその変化した機能は、この障害に特異的な経路と関係する、疾患特異的感受性遺伝子によって生じた臨床的表現型の発現を促進することができる。換言すると、これらの病的状態感受性遺伝子は、この疾患についてのその遺伝子的な構成(make-up)にしたがって、ヒトに明瞭な疾患を発生させる素因をつくる。

#### 【0167】

病的状態感受性遺伝子の候補は、転写、翻訳、熱ショックタンパク質、タンパク質トラフィック(protein trafficking)、DNA修復、細胞内構造(例えば、ミトコンドリア、ペルオキシソームおよび他の細胞内ミクロボディ)についての組み立てシステム、レセプターシグナルカスケード、免疫等に関係する経路の根本のレベルで見出すことができる。これらの経路は、細胞内レベルで、並びに全生物について生命の質を制御する。

#### 【0168】

これらの経路についてのタンパク質をコードする遺伝子に位置する突然変異/多型は、細胞の適応性を減少させ、そして生物をより感受性にさせ、疾患特異的感受性遺伝子の作用によって生じた、臨床的な表現型を発現させることができる。したがって、これらの病的状態感受性遺伝子は、すべてでないにしても、広範な種々の複合的な疾患に潜在的に関係することができる。疾患特異的感受性遺伝子は、グルコース、脂質、ホルモン代謝、等のような疾患特異的な経路として考えることができる経路と関係する。

#### 【0169】

例示した方法は、とりわけ、病的状態および/または死亡に対するヒトの全般的な感受性と関係する、遺伝子および/または遺伝子産物の同定;ヒトの疾患の遺伝子的基盤を解明するための研究におけるこれらの遺伝子および/または遺伝子産物の使用;疾患特異的感受性遺伝子なく、またはそれとともに組みあわせによる統計的分析におけるこれらの遺伝子および/または遺伝子産物の使用;疾患感受性遺伝子の一貫性を予測するためのこれらの遺伝子および/または遺伝子産物の使用;素因および/または急性医学的診断におけるこれらの遺伝子および/または遺伝子産物の使用、および疾患を治療するためおよび/またはヒトの寿命を延長するための薬物を開発するためのこれらの遺伝子および/または遺伝子産物の使用を可能とする。

#### 【0170】

##### スクリーニングプロセス

年齢、性別および人種等によって層化された健康体集団は、遺伝子に関連する病的状態についての、大変効率的で、普遍的なスクリーニング手段である。高齢な集団と比較したときの若齢の集団中の対立遺伝子頻度の変化は、推定的な病的状態感受性遺伝子を示すことが期待される。この健康体集団ベースの個々のサンプルを、スループットをさらに増加させるためにプールすることができる。原則的な実験の証明において、若齢および高齢な白人女性および男性のプールを適用して、多くの異なる遺伝子に位置する、400より多いランダムに選択された単一ヌクレオチド多型をスクリーニングした。

#### 【0171】

対立遺伝子的な相違が、両方または一方のみの性について、若齢および高齢者の間で8

10

20

30

40

50

%より大きいならば、候補多型が同定された。当初の結果を、少なくとも1つの独立的な次の実験でもう一度アッセイした。反復実験は、不安定な生化学的反応をみとめるために必要であり、そしてそれは約2 - 3%の頻度で存在し、そして年齢関連性の対立遺伝子頻度の相違に似ていることができる。

#### 【0172】

平均の頻度の相違および標準偏差を、当初の結果の良好な再現性があった後に算出する。最終的な対立遺伝子頻度を次いで、白人CEPHサンプルプールの引用集団と比較する。結果は、若齢白人集団の対立遺伝子頻度に類似することを示すものとする。次いで、遺伝子型情報を含む、候補の正確な対立遺伝子頻度を、すべての個々のサンプルを分析することによって取得した。この処理は、時間およびコストに関して単純である。それは莫大な数のSNPのスクリーニングを可能とする。したがって、年齢に対して高度に有意な連関を有する幾つかのマーカーを同定し、そして以下に記載する。

10

#### 【0173】

一般に、統計的に有意な結果を得るために、スクリーニングされるべき、層化した集団の少なくとも5の個体が必要である。年齢層化集団について、対立遺伝子の頻度を決定する。次いで、カイ二乗分析を対立遺伝子頻度に基づいて実施し、年齢の群の間の相違が統計的に有意かどうか決定する。0.1より小さいp値は、統計的に有意な相違であることを示すと考えられる。より好ましくはp値は、0.05より小さいものとする。

#### 【0174】

##### 臨床試験

集団内の頻度が年齢とともに減少するマーカーの同定によって、またよりよく設計され、そしてバランスが取れている臨床試験を得ることができる。現在では、臨床試験が研究で重要な指標(endpoint)としてマーカーを利用し、そしてそのマーカーが年齢とともに消失するならば、そのときは研究の結果は不正確かもしれない。ここで提供した方法を使用することによって、マーカーが年齢とともに頻度が減少するかどうか確認することができる。研究を計画するとき、この情報を考慮し、そして管理する(control)ことができる。例えば、年齢に独立的なマーカーは、適当な場合に置換し得る。

20

#### 【実施例】

#### 【0175】

以下の実施例は、解説目的のみのために含まれ、本発明の範囲の制限を目的とするものではない。

30

#### 【0176】

##### 実施例1

この実施例は、年齢に伴う既知遺伝的マーカーの対立遺伝子頻度の分布を測定するための任意の特定疾患で選択されているわけではない健康体血液ドナー集団により提供される情報を含むデータベースの使用および当該データベースの白人部分集団における性別により提供される情報を含むデータベースの使用について記載する。この実施例に記載の結果により、疾患関連遺伝的マーカーまたは多型は、年齢、性別および人種のようなパラメータまたはパラメータ群による健康体データベースの分類より同定され得る。

#### 【0177】

##### データベースの作成

血液は、提供される血液用の血液バンク基準を満たしたヒト対象から静脈穿刺により得た。血液サンプルは、pH8.0のEDTAで保存し、ラベルした。各ドナーから、年齢、性別、人種、病歴および家族の病歴のような情報を得た。各サンプルを、情報の同定を提示するバーコードでラベルした。データベースは、商業的に利用可能なソフトウェア、例えばMicrosoft Accessを用いコンピューター保存媒体のメモリ中に入れられる、各ドナーについての、相当する対象の識別子(identifier)および情報を入力することにより、作成した。

40

#### 【0178】

##### モデル遺伝的マーカー

50

幾つかのレベルで疾患と関連すると知られている多型の頻度は、データベースで提示される対象の部分集団において測定された。これら既知多型は、p 2 1、p 5 3およびリポタンパク質リパーゼ遺伝子中で生じる。特に、アミノ酸コドン 2 9 1 でセリンとアスパラギンとの置換を生じるリポタンパク質リパーゼ遺伝子の N 2 9 1 S 多型 (N 2 9 1 S) は、男性のアテローム性動脈硬化および特に心筋梗塞の危険性の増加と関連する、高密度リポタンパク質コレステロール (H D L - C) のレベルを低下することにつながる (Reymer et al. (1995) Nature Genetics 10: 28-34)。

#### 【 0 1 7 9 】

p 5 3 遺伝子は、D N A 損傷を評価する細胞周期制御タンパク質をコードし、細胞増殖、D N A 修復およびアポトーシス (プログラム化された細胞死) を制御する転写因子制御遺伝子として作用する。p 5 3 遺伝子中の突然変異は、種々の頻度を有する、種々の型の白血病を含む広く多種の癌で見つかった。通常 p 5 3 機能の喪失は、非制御細胞増殖をゲノムの不安定とする。p 5 3 遺伝子中で同定された多型、すなわち、R 7 2 P 対立遺伝子は、当該遺伝子のアミノ酸コドン 7 2 でプロリンとアルギニンとの置換を生じる。

10

#### 【 0 1 8 0 】

p 2 1 遺伝子は、通常細胞の G 1 期進行停止を伴うサイクリン依存キナーゼインヒビターをコードする。p 2 1 遺伝子の発現はアポトーシスを誘発する。p 2 1 遺伝子の多型は、Wilms's 腫瘍、小児科腎臓癌と関連する。p 2 1 遺伝子の 1 つの多型、S 3 1 R 多型は、アミノ酸コドン 3 1 においてアルギニンとセリンの置換を生じる。

#### 【 0 1 8 1 】

データベース分析

特定パラメーターによる対象の分類

遺伝的多型は、サンプルバンクの白人部分集団のセグメント内でプロファイルした。p 5 3 のプロファイルの場合、年齢 1 8 - 5 9 歳の合計 1 2 7 7 人の白人対象および年齢 6 0 - 7 9 歳の合計 4 5 7 人の白人対象の血液から単離したゲノム D N A を分析した。p 2 1 のプロファイルの場合、年齢 1 8 - 4 9 歳の合計 9 1 0 白人対象および年齢 5 0 - 7 9 歳の合計 8 2 4 白人対象の血液から単離したゲノム D N A を分析した。リポタンパク質リパーゼ遺伝子のプロファイルの場合、年齢 6 0 歳未満の合計 1 4 6 4 人の白人女性および合計 1 4 7 0 人の白人男性ならびに 6 0 歳を超える合計 4 7 8 人の白人女性および合計 5 6 0 人の白人男性を分析した。

20

30

#### 【 0 1 8 2 】

ゲノム D N A の単離および分析

ゲノム D N A は、個体から得た血液サンプルから単離した。各個体からの全血 1 0 m l を 2 0 0 0 × g で遠心分離した。 Buffy coat (buffy coat) 1 m l を、1 5 5 m M N H<sub>4</sub> C l、1 0 m M K H C O<sub>3</sub>、および 0 . 1 m M N a<sub>2</sub> E D T A の 9 m l に加え、室温で 1 0 分間インキュベーションし、1 0 分間 2 0 0 0 × g で遠心分離した。上清を取り除き、白色の細胞ペレットを、1 5 5 m M N H<sub>4</sub> C l、1 0 m M K H C O<sub>3</sub> および 0 . 1 m M N a<sub>2</sub> E D T A 中で洗浄し、5 0 m M T r i s、5 m M E D T A および 1 % S D S の 4 . 5 m l 中に再懸濁した。タンパク質は、6 m M 酢酸アンモニウム、p H 7 . 3 により細胞ライゼートより沈殿させ、次いで、3 0 0 0 × g の遠心分離により分離した。当該核酸を、同体積の 1 0 0 % イソプロパノールの添加および 2 0 0 0 × g の遠心分離により上清から取り除いた。当該乾燥核酸ペレットを、1 0 m M T r i s、p H 7 . 6 および 1 m M N a<sub>2</sub> E D T A 中で水和し、4 ° で保存した。

40

#### 【 0 1 8 3 】

既知遺伝的マーカーの存在または非存在を決定するゲノム D N A のアッセイは、B i o m a s s P R O B E (商標) 検出方法 (プライマーオリゴ塩基伸張) 反応を用い開発した。当該方法は、単一の検出プライマーを用い、その後オリゴヌクレオチド伸張ステップを行い、質量分析法、特に M A L D I - T O F 質量分析法により容易に分解し得る産物を得た。当該産物は、多型の存在または非存在に依存して長さが異なる。この方法では、検出プライマーは、可変ヌクレオチドまたはヌクレオチド配列の部位に隣接してアニーリングし、

50

当該プライマーは、1以上のジデオキシNTP、所望により1以上のデオキシNTPの存在下DNAポリメラーゼを用い伸張する。当該得られた産物を、MALDI-TOF質量分析法により、分解する。MALDI-TOF質量分析法により測定するような当該産物の質量は、可変部位に存在するヌクレオチドの決定を可能とする。

#### 【0184】

最初に、各白人ゲノムDNAサンプルを、p21(S31R対立遺伝子)、p53(R72P対立遺伝子)およびポタンパク質リパーゼ(N291S対立遺伝子)の遺伝子の多型部位の5'および3'部位に相当するプライマーを用い核酸増幅した。各プライマー対の1プライマーを、固体支持体に増幅産物を固定化し得るようにビオチニル化した。特に、p21、p53およびポタンパク質リパーゼの遺伝子の関連セグメントの増幅に用いるポリメラーゼ連鎖反応プライマーを以下に示す：p21の遺伝子増幅のためのUS4p21c31-2F(配列番号9)およびUS5p21-2R(配列番号10)；p53の遺伝子増幅のためのUS4-p53-ex4-F(p53-ex4US4(配列番号2)としても示される)およびUS5-p53/2-4R(US5p53/4R(配列番号3)としても示される)；およびポタンパク質リパーゼの遺伝子増幅のためのUS4-LPL-F2(配列番号16)およびUS5-LPL-R2(配列番号17)。

10

#### 【0185】

各DNA配列の増幅は、標準的なプロトコールにより行った。例えば、プライマーは濃度8pmolで用い得る。当該反応混合物(例えば、全体積50μl)は、10×緩衝液およびdTNPを含むTaq-ポリメラーゼを含み得る。ポリメラーゼ連鎖反応増幅のサイクリング条件は、典型的には、最初に95 5分間、その後、94 1分間、53 45秒、および72 30秒を40サイクル、最終伸張として72 5分間、行い得る。増幅産物は、製造者説明に従い、Qiagen PCR精製キット(No.28106)を用い精製し得る。カラムから精製した産物の溶出は、TE緩衝液(10mM Tris、1mM EDTA、pH7.5)50μl中で行うことができる。

20

#### 【0186】

精製増幅産物を、ビオチン-アビジン連結を介し、ストレプトアビジン被覆ビーズに固定化し、二本鎖DNAを変性させた。次いで、検出プライマーを、例えば、以下のような条件を用い、固定化DNAにアニーリングさせた：50 の50μlアニーリング緩衝液(20mM Tris、10mM KCl、10mM (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>、2mM MgSO<sub>2</sub>、1% Triton X-100、pH8)で10分間、その後、ビーズを洗浄緩衝液(40mM Tris、1mM EDTA、50mM NaCl、0.1% Tween 20、pH8.8)200μlで3回洗浄し、TE緩衝液200μlで1回洗浄した。

30

#### 【0187】

PROBE伸張反応は、例えば、USBからのDNAシーケンシングキット(No.70770)の幾つかの内容物およびPharmaciaのdNTPまたはddNTPを用いることにより、行った。典型的なプロトコールには、水21μl、シークエナーゼ緩衝液6μl、10mM DTT溶液3μl、0.5mMの3つのdNTP4.5μl、2mMの非含有(missing)の1つのddNTP4.5μl、グリセロール酵素希釈緩衝液5.5μl、シークエナーゼ2.00.25μl、およびピロホスファターゼ0.25μlを含む、全反応物体積45μlが含まれ得る。次いで、当該反応物を氷上でピペティングし、15分間室温でそして5分間37 でインキュベーションし得る。当該ビーズを洗浄緩衝液200μlで3回洗浄し、70mM NH<sub>4</sub>クエン酸溶液60μlで1回洗浄した。

40

#### 【0188】

当該DNAを変性し、固定化鋳型から伸張プライマーを開放させた。それぞれの得られた伸張産物を、マトリックスとして3-ヒドロキシピコリン酸(3-HPA)およびUVレーザーを用いMALDI-TOF質量分析法により別々に分析した。

#### 【0189】

特に、PROBE反応に用いるプライマーは以下のように示される：p21多型部位のPROBE分析用のP21/31-3(配列番号12)；p53多型部位のPROBE分析

50

用の P 5 3 / 7 2 (配列番号 4) ; およびリポタンパク質リパーゼ遺伝子多型部位の P R O B 分析用の L P L - 2 。 p 2 1 多型部位の P R O B E 分析において、伸張反応は、ジデオキシ C を用いた。“野生型”対立遺伝子鋳型(この場合、コドン 3 1 はセリンをコードする)で起こる反応からおよび多型 S 3 1 R 対立遺伝子鋳型(この場合、コドン 3 1 はアルギニンをコードする)で起こる反応から生ずる産物は、以下のように示し、それぞれ、P 2 1 / 3 1 - 3 S e r (w t )(配列番号 1 3 ) および P 2 1 / 3 1 - 3 A r g (配列番号 1 4 ) のように設計した。M A L D I - T O F 質量分析法で測定できるような各産物の質量がまた、提供される(野生型産物の場合、4 9 0 0 . 2 D a および多型産物の場合、5 2 1 3 . 4 D a ) 。

## 【 0 1 9 0 】

p 5 3 多型部位の P R O B E 分析において、当該伸張反応はジデオキシ C を用いた。“野生型”対立遺伝子鋳型(この場合、コドン 7 2 はアルギニンをコードする)で起こる反応からおよび多型 R 7 2 P 対立遺伝子鋳型(この場合、コドン 7 2 はプロリンをコードする)で起こる反応から生ずる産物を、以下のように示し、それぞれ、C o d 7 2 G A r g (w t ) および C o d 7 2 C P r o のように設計した。M A L D I - T O F 質量分析法により測定され得るような各産物の質量がまた、提供される(野生型産物の場合、5 7 3 4 . 8 D a および多型産物の場合、5 4 0 5 . 6 D a ) 。

## 【 0 1 9 1 】

リポタンパク質リパーゼ遺伝子多型部位の P R O B E 分析において、当該伸張反応は d A および d d T の混合物を用いた。“野生型”対立遺伝子鋳型(この場合、コドン 2 9 1 はアスパラギンをコードする)で起こる反応からおよび多型 N 2 9 1 S 対立遺伝子鋳型(この場合、コドン 2 9 1 はセリンをコードする)で起こる反応から生ずる産物は、以下のように示し、それぞれ、2 9 1 A s n および 2 9 1 S e r のように設計した。M A L D I - T O F 質量分析法により測定され得るような各産物の質量がまた、提供される(野生型産物の場合、6 4 3 8 . 2 D a および多型産物の場合、6 7 5 8 . 4 D a ) 。

## 【 0 1 9 2 】

P 5 3 - 1 ( R 7 2 P )

P C R 産物の長さ : 4 0 7 b p (配列番号 1 )

## 【 化 1 】

```

                                US4-p53-ex4-F
                                ctg aggacctggt cctctgactg
ctcttttcac ccatctacaag tcccccttgc cgteccaagc aatggatgat ttgatgctgt
cccggacga tattgaacaa tggttcactg aagaccagg tccagatgaa gctcccagaa
P53/72          72R
tgccagaagg tgctccccgc gtggcccctg caccagcagc tctacaccg ggggcccctg
c          72P
caccagcccc ctctggccc ctgtcatctt ctgtcccttc ccagaaaacc taccagggca
gctaeggttt ccgtctgggc ttcttgcaatt ctgggacagc caagtctgtg acttgcacgg
tcagttgcc tgaggggctg gcttccatga gacttcaa
                                US5-p53/2-4R

```

プライマー (配列番号 2 - 4 )

## 【 化 2 】

```

p53-ex4FUS4 ccc agt cac gac gtt gta aaa cgc tga gga cct ggt cct ctg ac
US5P53/4R   agc gga taa caa ttt cac aca ggt tga agt ctg atg gaa gcc
P53/72      gcc aga ggc tgc tcc cc

```

質量

10

20

30

40

【表 1】

対立遺伝子	終結産物 : ddC	配列#	長さ	質量
P53/72	gccagaggctgctcccc	5	17	5132.4
Cod72 G Arg (wt)	gccagaggctgctccccgc	6	19	5734.8
Cod72 C Pro	gccagaggctgctcccc	7	18	5405.6

ビオチン化 U S 5 プライマーを P C R 増幅に用いる。

【 0 1 9 3 】

L P L - 1 ( N 2 9 1 S )

リポタンパク質リパーゼ遺伝子のコドン 2 9 1 のアスパラギンをセリンにアミノ酸置換する。

【 0 1 9 4 】

P C R 産物の長さ : 2 5 1 b p ( 配列番号 1 5 )

U S 4 - L P L - F 2 ( 配列番号 1 6 )

【 化 3 】

```

gcgctccatt catctcttca tgcactctct gttgaatgaa gaaaatccaa gtaaggccta
cagggtgcagt tccaaggaag cctttgagaa agggctctgc ttgagttgta gaaagaaccg
                LPL-2                291N
ctgcaacaat ctgggctatg agatcaataa agtcagagcc aaaagaagca gcaaaatgta
                g 291S
cctgaagact cgttctcaga tgccc
                US4-LPL-R2

```

プライマー ( 配列番号 1 6 - 1 8 )

【 化 4 】

US4-LPL-F2 ccc agt cac gac gtt gta aaa cgg cgc tcc att cat ctc ttc

US5-LPL-R2 agc gga taa caa ttt cac aca ggg ggc atc tga gaa cga gtc

LPL-2 caa tct ggg cta tga gat ca

質量

【表 2】

対立遺伝子	終結産物 : ddA, ddT	配列#	長さ	質量
LPL-2	caatctgggctatgagatca	19	20	6141
291 Asn	caatctgggctatgagatcaa	20	21	6438.2
291 Ser	caatctgggctatgagatcagt	21	22	6758.4

ビオチン化 U S 5 プライマーを P C R 増幅に用いる。

【 0 1 9 5 】

P 2 1 - 1 ( S 3 1 R )

腫瘍サプレッサー遺伝子 p 2 1 のコドン 3 1 のセリンをアルギニンにアミノ酸置換する。産物の長さ : 2 0 7 b p ( 配列番号 8 )

【 化 5 】

US4p21c31-2F

```

                gtcc gtcagaaccc atgcggcagc
                p21/31-3 31S
aaggcctgcc gccgcctctt cggcccagtg gacagcgaqc agctgagccg cgactgtgat
                a 31R
gcgctaattg cgggctgcat ccaggaggcc cgtgagcgat ggaacttcga ctttgtcacc
gagacaccac tggaggg
                US5p21-2R

```

10

20

30

40

50

プライマー (配列番号 9 - 1 1 )

【化 6】

US4p21c31-2F ccc agt cac gac gtt gta aaa cgg tcc gtc aga acc cat gcg g

US5p21-2R agc gga taa caa ttt cac aca ggc tcc agt ggt gtc tcg gtg ac

P21/31-3 cag cga gca gct gag

質量

【表 3】

対立遺伝子	終結産物 : ddC	配列#	長さ	質量
p21/31-3	cagcgagcagctgag	12	15	4627
P21/31-3 Ser (wt)	cagcgagcagctgagc	13	16	4900.2
P21/31-3 Arg	cagcgagcagctgagac	14	17	5213.4

ビオチン化 U S 5 プライマーを P C R 増幅に用いる。

【0 1 9 6】

各白人対象 DNA サンプルを、MALDI - TOF 質量分析法により個々に分析し、多型部位のヌクレオチドの同一性を測定した。各アッセイの遺伝子型結果をデータベースに入力できる。次いで、結果を、年齢および / または性別により分類し、年齢および / または性別の対立遺伝子頻度の分布を測定した。各ケースの結果のヒストグラムを示す図中で示すように、p 2 1、p 5 3 およびリポタンパク質リパーゼ遺伝子の多型の遺伝的マーカーの対立遺伝子頻度の差異分布があった。

【0 1 9 7】

図 8 は、p 2 1 遺伝的マーカーアッセイの結果から、年齢 (年齢 1 8 - 4 9 歳を年齢 5 0 - 7 9 歳と比較した) の白人のヘテロ接合遺伝子型 (S 3 1 R) の頻度に統計的に有意な減少 (1 3 . 3 % から 9 . 2 %) が見られることを示す。2 つの年代群のホモ接合 (S 3 1 および R 3 1) 遺伝子型の頻度もまた、2 つの年代群における S 3 1 および R 3 1 対立遺伝子 (図中、それぞれ \* S 3 1 および \* R 3 1 として命名した) の全体的頻度であるため、示される。

【0 1 9 8】

図 7 A - C は、p 5 3 遺伝的マーカーアッセイの結果を示し、年齢 (年齢 1 8 - 5 9 歳を年齢 6 0 - 7 9 歳と比較した) の白人のホモ接合遺伝子型 (P 7 2) の頻度に満足できる有意な減少 (6 . 7 % から 3 . 7 %) を示す。2 つの年代群のホモ接合 “野生型” 遺伝子型 (R 7 2) およびヘテロ接合遺伝子型 (R 7 2 P) の頻度もまた、2 つの年代群における R 7 2 および P 7 2 対立遺伝子 (図中、それぞれ \* R 7 2 および \* P 7 2 として設計した) の全体的頻度であるため、示される。これらの結果は、p 5 3 が第二のタンパク質、細胞周期を通して細胞を動かす必要のあるサイクリン依存キナーゼ (C D K) を阻害する p 2 1 の発現を制御する (何れかの遺伝子の変異が細胞周期を崩壊させ、細胞分裂の増加を生ずる) ため、対立遺伝子が良性ではないという観察からなる。

【0 1 9 9】

図 2 C は、リポタンパク質リパーゼ遺伝子遺伝的マーカーアッセイの結果は、年齢 (また Reymer et al. (1995) Nature Genetics 10: 28-34 参照) の白人男性の多型対立遺伝子 (S 2 9 1) の頻度に統計的に有意な減少 (1 . 9 7 % から 0 . 5 4 %) が見られることを示す。別の年代群の白人女性のこの対立遺伝子頻度もまた示される。

【0 2 0 0】

実施例 2

この実施例は、因子 V I I 遺伝子の多型対立遺伝子 (3 5 3 Q 対立遺伝子) の存在または非存在を評価する複数対象の個々サンプルおよびプールしたサンプルとして多くの対象の DNA サンプルを分析し、対象の群中の対立遺伝子頻度を決定する MALDI - TOF 質量分析法の使用について記載する。この研究結果は、本質的に、同じ対立遺伝子頻度は、

10

20

30

40

50

別々に各サンプルを分析するようにプールしたDNAサンプルを分析することにより得られ、それにより、核酸の分析におけるMALDI-TOF質量分析法の量的性質が説明されることを示す。

#### 【0201】

##### 因子VII

因子VIIは、外来的血液凝固カスケード中に含まれるセリンプロテアーゼである。この因子はトロンビンにより活性化され、因子Xから因子Xaへのプロセッシングにおいて組織因子(因子III)と共に働く。因子VII遺伝子中の多型間の関連および心筋梗塞を含む虚血性心血管疾患の危険性を増大し得る因子VII活性の増加をサポートする証拠がある。本研究で調査する多型は、R353Qである(すなわち、因子VII遺伝子のコドン353においてグルタミン酸残基からアルギニン残基への置換)(表5参照)。

10

#### 【0202】

因子VII遺伝子の353Q対立遺伝子の存在または非存在に関するDNAサンプルの分析

ゲノムDNAは、群あたり92対象の複数の群に分けた大多数の対象から得られた別々の血液サンプルから単離された。ゲノムDNAの各サンプルは、実施例1に記載するようなBiomassPROBE(商標)アッセイを用い分析し、因子VII遺伝子の353Q多型の存在または非存在を決定した。

#### 【0203】

最初に、各サンプル由来のDNAを、以下に示すように、F7-353FUS4(配列番号24)およびF7-353RUS5(配列番号26)を用い、例えば、実施例1に記載するような標準的な条件を用いポリメラーゼ連鎖反応において増幅した。プライマーの1つをビオチニル化し、増幅産物を固体支持体に固定化し得た。精製増幅産物を、ストレプトアビジン被覆ビーズへのビオチン-アビジン連結を介して固定化し、二本鎖DNAを変性した。次いで、検出プライマーを、例えば、実施例1に記載のような条件を用い固定化DNAにアニーリングした。検出プライマーは、以下のようにF7-353-P(配列番号27)として示す。PROBE伸張反応は、例えば、実施例1に示すような条件を用い行った。当該反応は、ddGを用い行った。

20

#### 【0204】

DNAを変性し、固定化鋳型から伸張プライマーを開放した。それぞれの得られた伸張産物は、MALDI-TOF質量分析法により別々に分析した。3-ヒドロキシピコリン酸(3-HPA)のようなマトリックスおよびUVレーザーをMALDI-TOF質量分析法分析に用いた。“野生型”対立遺伝子鋳型(この場合、コドン353はアルギニンをコードする)で起こる反応からおよび多型353Q対立遺伝子鋳型(この場合、コドン353はグルタミン酸をコードする)で起こる反応から生ずる産物は、以下のように示し、それぞれ、353CGGおよび353CAGのように設計した。MALDI-TOF質量分析法により測定され得るような各産物の質量がまた、提供される(すなわち、野生型産物の場合、5646.8Daおよび多型産物の場合、5960Da)。

30

#### 【0205】

各DNAサンプルのPROBE反応のMALDI-TOF質量分析法分析は、各サンプルにおいて最初に別々に行われる(分析あたりのDNAの合計濃度250ng)。92対象の群における353Q多型の対立遺伝子頻度は、検出される多くの各対象に基づき算出された。

40

#### 【0206】

次に、92対象のサンプルをプールし(任意の個々DNAの濃度が2.7ngである、DNAの合計濃度250ng)、DNAのプールをMALDI-TOF質量分析法分析した。生じるスペクトルにおける353Q多型PROBE伸張産物の質量に相当するシグナルの下の領域が、存在するDNA量を定量するため、組込まれた。合計DNAに対する、この量の割合が用いられ、対象の群における353Q多型の対立遺伝子頻度が測定された。この型の個々サンプル対プールしたサンプルの分析を、92種サンプルの多種の群につ

50

いて繰り返した。

【0207】

92の各群の92種のサンプルの各MALDI-TOF質量分析法分析に基づき算出される頻度を、図9において、92サンプルからのDNAのプールのMALDI-TOF質量分析法分析に基づき算出したものと比較する。これらの比較を、図中、棒グラフで“対”として示し、各対は、別々の“プール”番号、例えば、P1、P16、P2などとして標識されている。そのため、例えば、P1の場合、92サンプルそれぞれの別々の分析により算出される多型の対立遺伝子頻度は、11.41%であり、92DNAサンプル全てのプールの分析により算出される頻度は、12.09%であった。

【0208】

別々のDNAサンプルを個々に分析することによりおよびDNAサンプルのプールにより算出される頻度の類似性から、MALDI-TOF質量分析法の量的性質を通して、プールしたサンプルを分析し、正確な頻度測定を得ることが可能となることが証明される。プールしたDNAサンプルの分析能は、本明細書に記載されたような非選択性の健康体データベースの使用を含む時間および費用を有意に減少する。スペクトルの特質または検出されたサンプル量を図る能力を全く変化せずにプール混合物中の個々サンプルのDNA濃度を2.7ngから0.27ngに減少させることが可能となることもまた示される。

【0209】

因子VIIR353QPROBEアッセイ

コード353 CGG > CAG (Arg > Gln)、エキソン9G > AのためのPROBEアッセイ

【0210】

PCRフラグメント：134bp (USタグを含む；配列番号22および23)

対立遺伝子頻度：ヨーロッパ人約0.1、日本人/中国人約0.03 - 0.05 (Thromb. Haemost. 1995, 73: 617-22; Diabetologia 1998, 41: 760-6)；

【化7】

F7-353FUS4>

1201 GTGCCGGCTA CTCGGATGGC AGCAAGGACT CCTGCAAGGG GGACAGTGGA GGCCACATG

F7-353-P> A <F7-353RUS5

1261 CCACCCACTA CCGGGGCACG TGGTACCTGA CGGGCATCGT CAGCTGGGGC CAGGGCTGCC

プライマー (配列番号24-26)

		Tm <sup>9a</sup>
F7-353FUS4	CCC AGT CAC GAC GTT GTA AAA CGA TGG CAG CAA GGA CTC CTG	64°C
F7-353-P	CAC ATG CCA CCC ACT ACC	
F7-353RUS5	AGC GGA TAA CAA TTT CAC ACA GGT GAC GAT GCC CGT CAG GTA C	64°C

質量

【表4】

対立遺伝子	終結産物 : ddG	配列#	長さ	質量
F7-353-P	atgccaccactacc	27	18	5333.6
353 CGG	cacatgccaccactaccg	28	19	5646.8
353 CAG	cacatgccaccactaccag	29	20	5960
US5-bio bio-	agcggataacaatttcacacagg	30	23	7648.6

【0211】

結論

上記実施例から、一般的な群の遺伝的因子に起因する疾患の頻度を変化させる効果が証明される。これらの結果から、多型遺伝的変化の医学的関連性が予測され得ると解釈される。加えて、結論が、浸透度、診断特異性、陽性適中率、疾患の攻撃、予防ストラテジーの最も適する攻撃、および単離された群を任意交配した群と同定する遺伝的変化の一般的適用可能性を導く。そのため、人種的に同種である、年齢および性別で層化した群に基づ

10

20

30

40

50

くサンプルバンクは、可能性ある医療的有用性に関し、遺伝的因子の迅速な同定および確認のための適当なツールとなる。

【0212】

実施例3

病的状態および死亡のマーカー

サンプルバンドおよび初期スクリーニング

健康体サンプルは、San Bernardino, CAの血液バンクを通じて取得した。ドナーに、血液回収の前に同意書にサインしてもらい、血液をヒトの加齢に関する遺伝的研究に使用することに同意してもらった。すべてのサンプルをアノニマイズ(anonymize)した。サンプルのトラッキングバック(tracking back)は可能ではない。

10

【0213】

健康体ドナー群の血液サンプルからのDNAの単離

血液を静脈穿刺により取得し、1 mM EDTA pH 8.0で保存する。各ドナーからの全血10 mlを2000 x gで遠心分離した。パフィーコート(buffy coat)1 mlを、155 mM NH<sub>4</sub>Cl、10 mM KHCO<sub>3</sub>、および0.1 mM Na<sub>2</sub>EDTAの9 mlに加え、室温で10分間インキュベーションし、10分間2000 x gで遠心分離した。上清を取り除き、白色の細胞ペレットを、155 mM NH<sub>4</sub>Cl、10 mM KHCO<sub>3</sub> および0.1 mM Na<sub>2</sub>EDTA中で洗浄し、50 mM Tris、5 mM EDTA および1% SDSの4.5 ml中に再懸濁した。タンパク質は、6 M 酢酸アンモニウム、pH 7.3により細胞ライゼートより沈殿させ、次いで、3000 x gの遠心分離により核酸から分離した。当該核酸を、同体積の100%イソプロパノールの添加および2000 x gの遠心分離により上清から取り除いた。当該乾燥核酸ペレットを、10 mM Tris、pH 7.6 および1 mM Na<sub>2</sub>EDTA中で水和し、4 で保存した。

20

【0214】

この研究では、サンプルを表1に示すようにプールした。血液ドナーの両親は白人起源であった。

表1

【表5】

プールID	性別	年齢-範囲	個体数
SP1	女性	18-39 歳	276
SP2	男性	18-39 歳	276
SP3	女性	60-69 歳	184
SP4	男性	60-79 歳	368

30

【0215】

400を超えるSNPを4つ全てのプールを用い試験した。試験を行った後、34アッセイを少なくとも一度、再アッセイするよう選択した。最終的に、10アッセイにより、幾らかの割合の対立遺伝子頻度に繰り返しの相違が見られ、それによって、各サンプルを用い試験される基準を満たした。平均対立遺伝子頻度および標準偏差を表2に表とする。

40

表2

【表 6】

アッセイID	SP1	SP1-STD	SP2	SP2-STD	SP3	SP3-STD	SP4	SP4-STD
47861	0.457	0.028	0.433	0.042	0.384	0.034	0.380	0.015
47751	0.276	0.007	0.403	0.006	0.428	0.052	0.400	0.097
48319	0.676	0.013	0.627	0.018	0.755	0.009	0.686	0.034
48070	0.581	0.034	0.617	0.045	0.561	n.a.	0.539	0.032
49807	0.504	0.034	0.422	0.020	0.477	0.030	0.556	0.005
49534	0.537	0.017	0.503	n.a.	0.623	0.023	0.535	0.009
49733	0.560	0.006	0.527	0.059	0.546	0.032	0.436	0.016
49947	0.754	0.008	0.763	0.047	0.736	0.052	0.689	0.025
50128	0.401	0.022	0.363	0.001	0.294	0.059	0.345	0.013
63306	0.697	0.012	0.674	0.013	0.712	0.017	0.719	0.005

10

20

10のうち7の可能性のある病的状態のマーカーが十分に分析された。これらSNPが位置する遺伝子に関する更なる情報を、Genbankのような公共のデータベースを通じて集めた。

【0216】

AKAPS

候補の病的状態および死亡のマーカーには、シグナルトランスダクションに含まれる遺伝子のようなハウスキーピング遺伝子が含まれる。その遺伝子には、タンパク質リン酸化を含むシグナルトランスダクション経路に関係するAキナーゼアンカータンパク質(AKAP)遺伝子がある。タンパク質リン酸化は、酵素制御、および真核細胞の細胞膜を通過する細胞外シグナルのトランスダクションに重要な機構である。酵素、膜レセプター、イオンチャンネルおよび転写因子を含む広範の種類の細胞性基質は、細胞と相互作用する細胞外シグナルに応答してリン酸化され得る。ホルモンおよび神経伝達物質に応答する細胞性タンパク質のリン酸化において鍵となる酵素は、サイクリックAMP(cAMP)依存タンパク質キナーゼである(PKA)。そのため、cAMPによる活性化において、PKAは、その細胞外シグナルに対する種々の細胞性応答を仲介する。PKAアイソザイムのアレイは、哺乳類細胞中で発現する。PKAは、通常、制御(R)サブユニットダイマーおよび2つの触媒(C)サブユニットを含む不活性テトラマーとして存在する。3つのCサブユニット(C<sub>1</sub>、C<sub>2</sub>およびC<sub>3</sub>)および4つのRサブユニット(RI<sub>1</sub>、RI<sub>2</sub>、RII<sub>1</sub>およびRII<sub>2</sub>)をコードする遺伝子を同定した(Takio et al. (1982) Proc. Natl. Acad. Sci. U.S.A. 79: 2544-2548; Lee et al. (1983) Proc. Natl. Acad. Sci. U.S.A. 80: 3608-3612; Jahnsen et al. (1996) J. Biol. Chem. 261: 12352-12361; Clegg et al. (1988) Proc. Natl. Acad. Sci. U.S.A. 85: 3703-3707;およびScott (1991) Pharmacol. Ther. 50: 123-145参照)。型I(RI<sub>1</sub>)および型II(RII<sub>1</sub>)サブユニットは、偏在的に分布し、この場合、RI<sub>1</sub>およびRII<sub>1</sub>は主に脳に存在する(例えば、Miki and Eddy (1999) J. Biol. Chem. 274: 29057-29062参照)。型I PKAホロ酵素(RI<sub>1</sub>およびRII<sub>1</sub>)は、細胞質中に主に存在し、一方、多数の型II PKA(RII<sub>2</sub>およびRII<sub>1</sub>)は細胞性構造および細胞小器官に関係する(Scott (1991) Pharmacol. Ther. 50: 123-145)。多くのホルモンおよび他のシグナルがレセプターを介して作用し、PKAのRサブユニットに結合し、放出し、タンパク質をリン酸化するCサブユニットを活性化するcAMPを生ずる。タンパク質キナーゼおよびその基質は広く細胞全体に分布するため、異なる

30

40

50

るシグナルに対するタンパク質キナーゼ仲介応答を局在化させる細胞中に存在する機構がある。その機構の1つには、特定の細胞小器官または細胞骨格コンパートメントにきわめて接近してPKAを位置させ、それによりより特異的なPKA相互作用および局在化応答を提供するAキナーゼアンカータンパク質(AKAP)と呼ばれる、アンカータンパク質の付随を介するPKAの亜細胞性ターゲティングが含まれる(例えば、Scott et al. (1990) *J. Biol. Chem.* 265: 21561-21566; Bregman et al. (1991) *J. Biol. Chem.* 266: 7207-7213; およびMiki and Eddy (1999) *J. Biol. Chem.* 274: 29057-29062参照)。アンカーは好ましい基質に近づけてキナーゼを位置させるばかりでなく、所望により第二メッセンジャーcAMPの揺らぎ(fluctuation)に応答し得る部位にPKAホロ酵素を位置させる(Mochly-Rosen (1995) *Science* 268: 247-251; Faux and Scott (1996) *Trends Biochem. Sci.* 21: 312-315; Hubbard and Cohen (1993) *Trends Biochem. Sci.* 18: 172-177参照)。

10

## 【0217】

75%までの型II PKAは、制御サブユニット(RII)によるAKAPとの会合を介して様々な細胞内部位に局在する(例えば、Hausken et al. (1996) *J. Biol. Chem.* 271: 29016-29022参照)。PKAのRIIサブユニットは、ナノモラーの親和性でAKAPと結合し(Carr et al. (1992) *J. Biol. Chem.* 267: 13376-13382)、多くのAKAP-RII複合体が、細胞抽出物から単離された。PKAのRIサブユニットは、マイクロモラーのみの親和性でAKAPに結合する(Burton et al. (1997) *Proc. Natl. Acad. Sci. U.S.A.* 94: 11067-11072)。PKA RIIサブユニットによるAKAPへの結合の証明は報告されており(Miki and Eddy (1998) *J. Biol. Chem.* 273: 34384-34390)、それは、RII 特異的およびRII / RII 二重特異性PKAアンカードメインは、FSC1 / AKAP82で同定された。PKAの型Iおよび型II制御サブユニットと相互作用するD-AKAP1およびD-AKAP2と呼ばれる、更なる二重特異的AKAPもまた報告されている(Huang et al. (1997) *J. Biol. Chem.* 272: 8057-8064; Huang et al. (1997) *Proc. Natl. Acad. Sci. U.S.A.* 94: 11184-11189)。

20

## 【0218】

20を超えるAKAPは、種々の組織および種で報告されている。AKAPをコードする相補DNA(cDNA)は、Caenorhabditis elegansおよびDrosophilaからヒトまでの範囲の多様な種から単離された(例えば、Colledge and Scott (1999) *Trends Cell Biol.* 9: 216-221参照)。PKAのRIIサブユニットとの会合を仲介するAKAP内の領域を同定した。約10 - 18アミノ酸残基のこれらの領域は、実質的に、第一の配列で変化するが、第二の構造予測から、ヘリックスの面に沿って並ぶ疎水性残基および他のものに沿った荷電性残基と共に両親媒性ヘリックスを形成するようであることが示されている(Carr et al. (1991) *J. Biol. Chem.* 266: 14188-14192; Carr et al. (1992) *J. Biol. Chem.* 267: 13376-13382)。長い脂肪族側鎖を有する疎水性アミノ酸、例えば、バリン、ロイシンまたはイソロイシンは、RIIサブユニットへの結合に関連し得る(Glantz et al. (1993) *J. Biol. Chem.* 268: 12796-12804)。

30

## 【0219】

多くのAKAPがまた、他のシグナリング酵素を含む複数のタンパク質に結合する能力を有する。例えば、AKAP79は、PKA、タンパク質キナーゼC(PKC)、およびプロテインホスファターゼカルシニユリン(PP2B)に結合する(Coghlan et al. (1995) *Science* 267: 108-112およびKlauck et al. (1996) *Science* 271: 1589-1592)。そのため、ニューロン性シナプス後膜に対するAKAP79のターゲティングにより、単一複合体における逆触媒活性を有する酵素が同時に得られる。

40

## 【0220】

そのため、AKAPは、cAMP仲介応答の選択性および強度を増大する可能性ある制御機構としての役割をする。そのため、細胞の基礎となる機能においてこれらのタンパク質が担う重要な役割を完全に理解するため、AKAPの構造的および機能的な性質を明らかにし解明する必要がある。

50

## 【0221】

## AKAP10

ヒトAKAP10 cDNA(D-AKAP2とも称する)の配列はGenBankデータベースにおいて、受け入れ番号AF037439(配列番号31)およびNM007202として利用可能である。AKAP10遺伝子はクロモソーム17上に位置する。

## 【0222】

マウスD-AKAP2 cDNAの配列もまた、GenBankデータベース(受け入れ番号AF021833)において利用可能である。マウスD-AKAP2タンパク質は、Gサブユニットと相互作用しGTPase活性化タンパク質様活性を有するタンパク質の特性であるアミノ末端近辺のRGSドメインを含む(Huang et al. (1997) Proc. Natl. Acad. Sci. U.S.A. 94: 11184-11189)。ヒトAKAP10タンパク質はまた、RGSドメインと相同な配列を有する。マウスD-AKAP2タンパク質のカルボキシ末端40残基は、PKAの制御サブユニットとの相互作用に応答する。この配列は、マウスD-AKAP2とヒトAKAP10タンパク質との間で完全に保存されている。

## 【0223】

ヒトAKAP10遺伝子の多型および多型AKAP10タンパク質

遺伝子発現、制御、タンパク質構造および/またはタンパク質機能を変化させる、AKAP遺伝子の多型は、遺伝子および/またはタンパク質機能を変えない多型よりも、酵素(特にPKA)活性の制御、シグナルの細胞性形質導入およびその応答においてならびに細胞の基礎機能において、より重要な効果を有するようである。本明細書で提供される多型AKAPに含まれるものは、番号646位に異なるアミノ酸残基を含むヒトAKAP10タンパク質である。

## 【0224】

ヒトAKAP10タンパク質のアミノ酸646は、PKAのRサブユニットの結合に係するセグメント内のタンパク質のカルボキシ末端領域に位置する。このセグメントは、カルボキシ末端40アミノ酸を含む。

## 【0225】

ヒトAKAP10タンパク質の646位に関し報告されているアミノ酸残基はイソロイシンである。本明細書で提供される多型ヒトAKAP10タンパク質は、アミノ酸配列であるが、当該タンパク質の646位アミノ酸のイソロイシン以外の残基を含む。本明細書で提供される多型ヒトAKAP10タンパク質の特定の実施態様では、646位のアミノ酸は、バリン、ロイシンまたはフェニルアラニン残基である。

## 【0226】

ヒトAKAP10コーディング配列のヌクレオチド2073位におけるAからGへの変化  
本明細書で記載の通り、コーディング配列の2073位の特異的多型を含み、それにより、646位でバリンをコードするヒトAKAP10遺伝子の対立遺伝子は、ヒト集団のより若齢およびより高齢のセグメント由来のDNAサンプルにおいて多様な頻度で検出された。この対立遺伝子において、AKAP10遺伝子コーディング配列の2073位のAは、AからGに変化させ、アミノ酸646のコードをイソロイシンのコードのATTからバリンのコードのGTTへと変化させて変化した配列を生じさせた。

## 【0227】

病的状態マーカー1: ヒトタンパク質キナーゼAアンカータンパク質(AKAP10-1)健康体ドナー集団におけるAKAP10-1のPCR増幅およびBiomass PROBEアッセイ検出

AKAP10のドナー集団のPCR増幅

PCRプライマーは、OPERONにより、ホスホラミダイト化学を用い合成された。AKAP10標的配列の増幅は、PCR反応物50μl中に、プールしたヒトゲノムDNA100ng-1μgを有する単一のPCR反応物50μl中で行った。プールしたサンプル内の各DNA濃度は、最終濃度1-25ngの範囲の等濃度で存在した。各反応物は、1×PCR緩衝液(Qiagen, Valencia, CA)、200μM dNTP、1U Hotstar Ta

qポリメラーゼ(Qiagen, Valencia, CA)、4 mM MgCl<sub>2</sub>、ならびにユニバーサルプライマー配列および標的特異的配列5'-TCTCAATCATGTGCATTGAGG-3'(配列番号45)を含む25 pmol フォワードプライマー、および2 pmol リバースプライマー5'-AGCGGATAACAATTTTCACACAGGGATCACACAGCCATCAGCAG-3'(配列番号46)、ならびにPCRアンプリコンの5'末端に相補的なビオチニル化ユニバーサルプライマー10 pmol 5'-AGCGGATAACAATTTTCACACAGG-3'(配列番号47)を含む。特異的フォワードおよびリバースプライマーを用いる最初のラウンドの増幅の後、次いで、5'ビオチニル化ユニバーサルプライマーをハイブリダイズさせ、リバースプライマーとして作用させ、それにより、3'ビオチン捕捉部分を当該分子中に導入した。当該増幅プロトコールは、5'-ビオチニル化二本鎖DNAアンプリコンを生じ、遺伝子型決定(genotyping)に用いる5'ビオチン標識各フォワードプライマーの必要性を除くことによるハイスループット遺伝子型決定の費用を劇的に減少する。サーマルサイクリングは、0.2 mLチューブまたは96ウェルプレート中で、MJ Research Thermal Cycler(算出温度)を用い、以下のパラメーターで行った：94

10

20

30

40

50

#### 【0228】

##### DNAの固定化

PCR反応物50 μlを、事前に3回洗浄し1 M NH<sub>4</sub>Cl、0.06 M NH<sub>4</sub>OH中に再懸濁したストレプトアビジン被覆磁性ビーズ25 μlに加えた。PCRアンプリコンは、室温で15分間でビーズに結合し得る。次いで、当該ビーズを、磁石を用い回収し、非結合DNAを含む上清を取り除いた。非結合鎖は、100 mM NaOH中でインキュベーションし、10 mM Tris pH 8.0でビーズを3回洗浄することにより二本鎖アンプリコンから解放した。

#### 【0229】

A K A P 1 0 - 1のドナー集団のB i o m a s s P R O B Eアッセイ分析(クローン48319)

B i o m a s s P R O B Eアッセイ法を用いる遺伝子型決定は、26 mM Tris-HCl pH 9.5、6.5 mM MgCl<sub>2</sub>および50 mM各d T T Pおよび50 mM各d d C T P、d d A T P、d d G T P、熱安定性DNAポリメラーゼ2.5 U (Amersham)ならびに鑄型特異的オリゴヌクレオチドP R O B Eプライマー5'-CTGGCGCCCCACGTGGTCAA-3'(配列番号48)(Operon)にDNA被覆磁性ビーズを再懸濁することにより行った。プライマー伸張は、オリゴヌクレオチドプライマーハイブリダイゼーションおよび伸張の3サイクルで生じる。当該伸張産物は、50 mM NH<sub>4</sub>Clで鑄型から変性させ、各サンプル150 nLを、H3PAマトリックス物質150 nLで事前負荷したシリコンチップへ移した後、分析した。当該サンプル物質を結晶化し、MALDI-TOF (Bruker, PerSeptive)で分析した。A K A P 1 0中に存在するSNPは、A K A P 1 0遺伝子(GenBank受け入れ番号AC005730(配列番号36)のゲノムクローンの配列のヌクレオチド番号156277においてTをCに変化する。配列番号35は、ヒトA K A P 1 0遺伝子のゲノムヌクレオチド配列を含むヒトクロモソーム17のヌクレオチド配列を示し、配列番号36は、ヒトA K A P 1 0対立遺伝子のゲノムヌクレオチド配列を含むヒトクロモソーム17のヌクレオチド配列を示す。B i o M a s sプローブ反応に使用したプライマーの質量は、5500.6ダルトンであった。SNPの存在下、当該プライマーは、質量5773.8のd d Cの添加により伸張する。野生型遺伝子は、d Tおよびd d Gをプライマーに添加し、質量6101ダルトンを有する伸張産物を産生する。

#### 【0230】

SNPの頻度は、年齢で選択した健康個体の集団で測定した。年齢18-39歳の五百五十二(552)個体(276女性、276男性)および年齢60-79の552個体(年齢60-69では184女性、年齢60-79では368男性)を、A K A P 1 0の非翻訳3'領域中に局在する多型の存在に関し試験した。増加する年齢の群にともなうこの多型の頻度における相違が、健康個体で観察された。統計学的な分析により、“より若齢”および“より高齢”の集団の間の対立遺伝子に関する対立遺伝子頻度の相違の有意なレ

ベルは、 $p = 0.0009$  および遺伝子型の有意なレベルは  $p = 0.003$  であることが示された。年齢群の間の相違は、有意である。すべての集団対立遺伝子有意性は、 $p = 0.0009$  であり、遺伝子型の有意性は、 $p = 0.003$  である。

#### 【0231】

最善の有意性を生ずるこのマーカーは、年齢分類集団における対立遺伝子および遺伝子型頻度に関する結果である。図19は、両性および集団全体における対立遺伝子および遺伝子型頻度を示す。対立遺伝子に関する後者の有意性は、 $p = 0.0009$  であり、遺伝子型の有意性は、 $p = 0.003$  であった。若齢および高齢の集団は、ハーディ-ヴァインベルグ平衡であった。ある特定の遺伝子型の好ましい変化は、見られなかった。

#### 【0232】

多型は、ヒトプロテインキナーゼAアンカータンパク質(AKAP10)をコードする遺伝子の非翻訳3'-領域中に局在する。当該遺伝子は、クロモソーム17に局在する。その構造には、15エクソンおよび14介在配列(イントロン)を含む。コードタンパク質は、cAMP依存プロテインキナーゼの垂細胞局在に应答し、それによって、G-プロテイン仲介レセプターシグナリング経路において鍵の役割をする(Huang et al. (1007) PNAS 94: 11184-11189)。その局在は、コーディング領域の外側であるため、この多型は、アミノ酸置換およびその後のタンパク質の機能変化を誘引し得る他の非同義性多型と共に、ほぼ連鎖不平衡(LD)であるようである。この遺伝子に関する異なるGenBankデータベースの配列比較により、さらに6つの可能性ある多型が示され、その2つ各アミノ酸の変化が考えられる(表3参照)。

#### 【表7】

表 3

エクソン	コドン	ヌクレオチド	アミノ酸
3	100	GCT>GCC	Ala>Ala
4	177	AGT>GTG	Met>Val
8	424	GGG>GGC	Gly>Gly
10	524	CCG>CTG	Pro>Leu
12	591	GTG>GTC	Val>Val
12	599	CGC>CGA	Arg>Arg

#### 【0233】

病的状態マーカー2：ヒトプロテインキナーゼAアンカータンパク質(AKAP10-5) AKAP10-5対立遺伝子(配列番号33)の発見

ゲノムDNAは、AKAP10-1遺伝子座で遺伝子型CCを有する十七(17)個体および単一のヘテロ接合性個体(CT)(上記のように)の血液(上記のように)から単離された。C末端PKA結合ドメインをコードするAKAP10-1遺伝子中の標的配列は、ポリメラーゼ連鎖反応を用い増幅した。PCRプライマーは、OPERONにより、ホスホラミダイト化学を用い合成した。AKAP10-1標的配列の増幅は、ヒトゲノムDNA鋳型25ngを有する各PCR反応物50μlで行った。各反応物は、1×PCR緩衝液(Qiagen, Valencia, CA)、200μM dNTP、1U Hotstar Taqポリメラーゼ(Qiagen, Valencia, CA)、4mM MgCl<sub>2</sub>、ユニバーサルプライマー配列および標的的特異的配列5'-TCCCAAAGTGCTGGAATTAC-3'(配列番号53)を含む25pmolのフォワードプライマー(E x 1 3 F)、および2pmolリバープライマー(E x 1 4 R)5'-GTCCAATATATGCAACAGTTG-3'(配列番号54)を含む。サーマルサイクリングは、0.2mLチューブまたは96ウェルプレート中で、MJ Research Thermal Cycler(MJ Research, Waltham, MA)(算出温度)を用い、以下のサイクリングパラメーターで行った：94 5分間；45サイクル：94 20秒、56 30秒、72 60秒；72 3分間。増幅後、当該アンプ

リコンは、クロマトグラフィーを用い精製した (Mo Bio Laboratories (Solana Beach, CA))。

#### 【0234】

標的領域を示す18アンプリコンの配列は、PCRアンプリコン25nmol、3.2 μM DNAシーケンシングプライマー5'-CCCACAGCAGTTAATCCTTC-3'(配列番号55)、および鎖終結ローダミン標識2', 3'ジデオキシヌクレオチド(PE Biosystems, Foster City, CA)を伴う標準的サンガーサイクルシーケンシング法を用い、以下のサイクリングパラメーターに従い決定した：96 15秒；25サイクル：55 15秒、60 4分間。当該シーケンシング産物は、0.3M NaOAcおよびエタノールにより沈殿した。当該沈殿を遠心分離し、乾燥させた。当該ペレットを脱イオン化ホルムアミド中に再懸濁し、5%ポリアクリルアミドゲルで分離した。当該配列は、“Sequencher”ソフトウェアを用い決定した(Gene Codes, Ann Arbor, MI)。

10

#### 【0235】

アンプリコンのAKAP10-1SNPについてホモ接合性である、17すべてのアンプリコンの配列は、AとGが置換えられたヌクレオチド152171位(AKAP10ゲノムクローン(配列番号35)についてGenBank受け入れ番号AC005730)の多型を示す。このSNPはまた、野生型AKAP10(GenBank受け入れ番号AF037439)(配列番号31)のcDNAクローンのヌクレオチド2073位に位置するように設計できる。ヒトAKAP10タンパク質のアミノ酸配列は、配列番号32として提供する。この単一のヌクレオチド多型は、AKAP10-5(配列番号33)として命名され、ヒトAKAP10タンパク質(配列番号32)のアミノ酸配列のアミノ酸646位でバリンとイソロイシンの置換を生じる。

20

#### 【0236】

健康体ドナー集団におけるAKAP10-5のPCR増幅およびBiomassPROBEアッセイ検出

年齢により層化される健康体集団は、高齢の集団と比べ若齢の集団において対立遺伝子頻度の変化を検出し得ることにより、病的状態随伴遺伝子用の非常に有効でユニバーサルなスクリーニングツールである。この健康体集団ベースの各サンプルは、更にスループットを増加させるためプールすることができる。

#### 【0237】

健康体サンプルは、San Bernardino, CAの血液バンクを通じて得た。血液ドナーの両親は、白人起源であった。事実上、ヒトであるとき、健康対象体は、一般的集団における結果としての使用のため血液が提供される血液バンクの基準を有するヒトドナーとして定義される。これらの基準は、以下の通りである：検出可能なウイルス、細菌、マイコプラズマ、および寄生体感染がないこと；貧血症でないこと；および次いで更に病歴に関する質問表に基づき選択する(図3参照)。そのため、健康体集団は、血液バンク基準に従い血液が提供された十分に健康の先入観のない集団を示し、更に任意の疾患の状態に関して選択されることはない。典型的にその個体は、任意の薬物療法を受けていない。

30

#### 【0238】

PCRプライマーは、OPERONによりホスホラミダイト化学を用いて合成した。AKAP10標的配列の増幅は、PCR反応物50 μlにおいて、プールしたヒトゲノムDNA100ng - 1 μgを伴う単一のPCR反応物50 μl中で行った。プールしたサンプル内の各DNA濃度は、最終濃度1 - 25 ngの範囲の等濃度で存在した。各反応物は、1 × PCR緩衝液(Qiagen, Valencia, CA)、200 μM dNTP、1 U Hotstar Taqポリメラーゼ(Qiagen, Valencia, CA)、4 mM MgCl<sub>2</sub>、ならびにユニバーサルプライマー配列および標的的特異的配列5'-AGCGGATAACAATTTTCACACAGGGAGCTAGCTTGGGAAGATTGC-3'(配列番号41)を含む25 pmolのフォワードプライマー、および2 pmolリバープライマー5'-GTCCAATATATGCAAACAGTTG-3'(配列番号54)、ならびにPCRアンプリコンBIOの5'末端に相補的なピオチニル化ユニバーサルプライマー10 pmol：5'-AGCGGATAACAATTTTCACACAGG-3'(配列番号43)を含む。特異的フォワードおよびリバープライマー

40

50

ライマーを用いる標的の最初のラウンドの増幅の後、次いで、5'ビオチニル化ユニバーサルプライマーをハイブリダイズさせ、フォワードプライマーとして作用させ、それにより、5'ビオチン捕捉部分を当該分子中に導入した。当該増幅プロトコールは、5'-ビオチニル化二本鎖DNAアンプリコン中で生じ、遺伝子型決定に用いる5'ビオチン標識各フォワードプライマーの必要性を除くことによるハイスル-プット遺伝子型決定の費用を劇的に減少する。

【0239】

サーマルサイクリングは、0.2 mLチューブまたは96ウェルプレート中で、MJ Research Thermal Cycler(算出温度)を用い、以下のパラメーターで行った：94 5分間；45サイクル：94 20秒、56 30秒、72 60秒；72 3分間。

10

【0240】

DNAの固定化

PCR反応物50 μlを、事前に3回洗浄し1 M NH<sub>4</sub>Cl、0.06 M NH<sub>4</sub>OH中に再懸濁したストレプトアビジン被覆磁性ビーズ(Dynal, Oslo, Norway)25 μlに加えた。二本鎖PCRアンプリコンの一本鎖の5'末端は、室温で15分間でビーズに結合し得る。次いで、当該ビーズを、磁石を用い回収し、非結合DNAを含む上清を取り除いた。ハイブリダイズしているが非結合の鎖は、100 mM NaOH中でインキュベーションし、10 mM Tris pH 8.0でビーズを3回洗浄することにより二本鎖アンプリコンから解放した。

【0241】

Biomass PROBE (商標)アッセイを用いるAKAP10-5の検出

AKAP10-5(配列番号33)のドナー集団のプライマー伸張分析(米国特許番号6,043,031)のBiomass PROBE(商標)アッセイを用いた。これらの方法を用いる遺伝子型決定は、26 mM Tris-HCl pH 9.5、6.5 mM MgCl<sub>2</sub>、50 mM dTTPおよび50 mM各ddCTP、ddATP、ddGTP、2.5 U 熱安定性DNAポリメラーゼ(Amersham)ならびに鋳型特異的オリゴヌクレオチドPROBEプライマー5'-ACTGAGCCTG CTGCATAA-3'(配列番号44)(Operon)にDNA被覆磁性ビーズを再懸濁することにより行った。プライマー伸張は、ハイブリダイゼーションおよび伸張を伴うオリゴヌクレオチドプライマーの3サイクルで生じる。当該伸張産物は、50 mM NH<sub>4</sub>Clで鋳型から変性させ、各サンプル150 nLを、H3PAマトリックス物質150 nLで事前負荷したシリコンチップへ移した後、分析した。当該サンプル物質を結晶化し、MALDI-TOF(Bruker, PerSeptive)で分析した。当該プライマーは、質量5483.6ダルトンを有する。SNPはプライマーにddCを付加し、質量5756.8ダルトンの伸張産物を得た。当該野生型は、プライマーへTおよびddGを付加し、質量6101ダルトンを得る。

20

30

【0242】

SNPの頻度は、年齢で選択した健康個体の集団で測定した。年齢40歳未満の七百十三(713)個体(360女性、353男性)および60歳を超える年齢の703個体(322女性、381男性)を、SNP、AKAP10-5(配列番号33)の存在下、試験した。結果を以下の表1に示す。

40

【表 8】

			<40	>60	デルタG 対立遺伝子
女性	対立遺伝子	*G	38.6	34.6	4.0
		*A	61.4	65.4	
	遺伝子型	G	13.9	11.8	2.1
		GA	49.4	45.7	
		A	36.7	42.5	

10

【表 9】

男性	対立遺伝子	*G	41.4	37.0	4.4
		*A	58.6	63.0	
	遺伝子型	G	18.4	10.8	7.7
		GA	45.9	52.5	
		A	35.7	36.7	
全	対立遺伝子	*G	40.0	35.9	4.1
		*A	60.0	64.1	
	遺伝子型	G	16.1	11.2	4.9
		GA	47.7	49.4	
		A	36.2	39.4	

20

30

## 【0243】

図 20 は、年齢および性別で分類された白人集団における対立遺伝子および遺伝子型の分布の結果を描写的に示す。

## 【0244】

病的状態マーカー 3 : ヒトメチオニンスルホキシドレダクターゼ A (msrA)

両性および集団全体におけるこのマーカーの年齢関連の対立遺伝子および遺伝子型の頻度を図 21 に示す。高齢の男性集団におけるホモ接合 CC 遺伝子型の減少は、高度に有意である。

40

## 【0245】

メチオニンスルホキシドレダクターゼ A (#63306)

ヒト健康体ドナー集団におけるヒトメチオニンスルホキシドレダクターゼ A (h-msr-A) の PCR 増幅および Biomas s P R O B E アッセイ検出

h-msr-A のドナー集団の PCR 増幅

PCR プライマーは、OPERON により、ホスホラミダイト化学を用い合成した。AKAP10 標的配列の増幅は、PCR 反応物 50  $\mu$ l 中に、プールしたヒトゲノム DNA 鋳型 100 ng - 1  $\mu$ g を有する単一の PCR 反応物 50  $\mu$ l 中で行った。プールしたサンプル内の各 DNA 濃度は、最終濃度 1 - 25 ng の範囲の等濃度で存在した。各反応物は、1  $\times$  PCR 緩衝液 (Qiagen, Valencia, CA)、200  $\mu$ M dNTP、1 U Hotstar T

50

a q ポリメラーゼ (Qiagen, Valencia, CA)、4 mM MgCl<sub>2</sub>、ならびにユニバーサルプライマー配列および標的特異的配列 5'-TTTCTCTGCACAGAGAGGGC-3' (配列番号 49) を含む 25 pmol のフォワードプライマー、および 2 pmol リバースプライマー 5'-AGCGGATACAATTTTCACACAGGGCTGAAATCCTTCGCTTTACC-3' (配列番号 50)、ならびに PCR アンプリコンの 5' 末端に相補的なビオチニル化ユニバーサルプライマー 10 pmol 5'-AGCGGATAACAATTTTCACACAGG-3' (配列番号 51) を含む。特異的フォワードおよびリバースプライマーを用いる標的の最初のラウンドの増幅の後、次いで、5' ビオチニル化ユニバーサルプライマーをハイブリダイズさせ、リバースプライマーとして作用させ、それにより、3' ビオチン捕捉部分を当該分子中に導入した。当該増幅プロトコールは、5' - ビオチニル化二本鎖 DNA アンプリコンを生じ、遺伝子型決定に用いる 5' ビオチン標識各フォワードプライマーの必要性を除くことによってハイスル - プット遺伝子型決定の費用を劇的に減少する。サーマルサイクリングは、0.2 mL チューブまたは 96 ウェルプレート中で、MJ Research Thermal Cycler (算出温度) を用い、以下のサイクリングパラメーターで行った：94 5 分間；45 サイクル：94 20 秒、56 30 秒、72 60 秒；72 3 分間。

10

## 【0246】

## DNA の固定化

PCR 反応物 50 μl を、事前に 3 回洗浄し 1 M NH<sub>4</sub>Cl、0.06 M NH<sub>4</sub>OH 中に再懸濁したストレプトアビジン被覆磁性ビーズ (DynaI) 25 μl に加えた。PCR アンプリコンは、室温で 15 分間でビーズに結合し得る。次いで、当該ビーズを、磁石を用い回収し、非結合 DNA を含む上清を取り除いた。非結合鎖は、100 mM NaOH 中でインキュベーションし、10 mM Tris pH 8.0 でビーズを 3 回洗浄することにより二本鎖アンプリコンから解放した。

20

## 【0247】

## h - m s r - A のドナー集団の Biomas s P R O B E アッセイ分析

Biomas s P R O B E アッセイ法を用いる遺伝子型決定は、26 mM Tris - HCl pH 9.5、6.5 mM MgCl<sub>2</sub>、50 mM dTTP および 50 mM 各 ddCTP、ddATP、ddGTP、2.5 U 熱安定性 DNA ポリメラーゼ (Amersham) ならびに 20 pmol の鑄型特異的オリゴヌクレオチド P R O B E プライマー 5'-CTGAAAAGG GAGAGAAAG-3' (Operon) (配列番号 52) に DNA 被覆磁性ビーズを再懸濁することにより行 った。プライマー伸張は、ハイブリダイゼーションおよび伸張を伴うオリゴヌクレオチド プライマーの 3 サイクルで生じる。当該伸張産物は、50 mM NH<sub>4</sub>Cl で鑄型から変 性させ、各サンプル 150 nl を、H3PA マトリックス物質 150 nl で事前負荷した シリコンチップへ移した後、分析した。当該サンプル物質を結晶化し、MALDI - T O F (Bruker, PerSeptive) で分析した。SNP は、2 つの E S T の配列における T から C へ の変化として示される。野生型は、野生型ヒト m s r A 遺伝子 (配列番号 39) の一部であ る E S T のヌクレオチド配列を示す、GenBank 受け入れ番号 AW195104 の 1 2 8 位に T を有することにより示される。SNP は、ヒト m s r A 遺伝子 (配列番号 40) の対立遺伝子の一部である E S T のヌクレオチド配列を示す、GenBank 受け入れ番 号 AW874187 の 129 位の C として示される。

30

40

## 【0248】

ゲノム配列において、SNP は、A から G への変化として示される。Biomas s プ ロープ反応で使用するプライマーは、質量 5654.8 ダルトンを有していた。SNP の 存在下、当該プライマーは、ddC の組込みにより伸張し、質量 5928 を有する。野生 型の存在下、当該プライマーは、dT および DDC の添加により伸張し、質量 6232. 1 ダルトンを生ずる。

## 【0249】

SNP の頻度は、年齢で選択した健康個体の集団で測定した。年齢 18 - 39 歳の五百 五十二 (552) 個体 (276 女性、276 男性) および年齢 60 - 79 の 552 個体 (年齢 60 - 69 では 184 女性、年齢 60 - 79 では 368 男性) を、h - m s r - A の非翻

50

訳 3' 領域中に局在する多型の存在に関し試験した。

【0250】

健康体個体間の男性年齢群の遺伝子型の相違が有意であった。男性集団の場合、対立遺伝子有意性は、 $p = 0.0009$  および遺伝子型有意性は  $p = 0.003$  である。両性および集団全体における、このマーカーの年齢関連の対立遺伝子および遺伝子型の頻度を図 21 に示す。高齢の男性集団におけるホモ接合 CC 遺伝子型の減少は、非常に有意である。

【0251】

多型は、ヒトメチオニンスルホキシドレダクターゼ A (h-msr-A) をコードする遺伝子の非翻訳 3' - 領域中に局在する。当該正確な局在は、ストップコドン (TAA) の 451 塩基対下流である。この SNP は、コーディングまたはプロモーターの領域よりも上流の他の多型と共に連鎖不平衡 (LD) となるようであり；そのため、直接病的状態の原因とはならない。酵素メチオニンスルホキシドレダクターゼは、複数の生物学的機能を示すことが提唱されている。酸化性タンパク質損傷の修復ばかりでなく、生物学的機能の活性化または不活性化によるタンパク質の制御においてまた重要な役割をする (Moskovitz et al. (1990) PNAS 95: 14071-14075)。その活性が、アルツハイマー患者の脳組織において有意に減少することもまた示された (Gabbita et al. (1999) J. Neurochem 73: 1660-1666)。反応性酸素種の代謝に含まれるタンパク質が疾患に関係すると、科学的に考えられる。

【0252】

結論

健康体集団の使用は、病的状態マーカーの同定を提供する。G-プロテイン結合シグナリング形質導入経路にまたは酸化ストレスの緩和に含まれるタンパク質の同定は、説得力のある結果であると考えられる。ヒトプロテインキナーゼ A アンカータンパク質をコードする遺伝子において *in silico* で既に同定されている他の可能性ある多型の更なる確認および検証は、病的状態に強力に付随し、この遺伝子産物は、適当な医薬的または診断的標的であることを証明する。

【0253】

実施例 4

MALDI-TOF 質量分析法分析

以下に挙げた酵素アッセイの産物のすべてを MALDI-TOF 質量分析法により分析した。水で 2.5 倍に希釈された 1:1 水:アセトニトリル中の 10:1 3-ヒドロキシピコリン酸:クエン酸アンモニウムを含む希釈マトリックス溶液 ( $0.15 \mu\text{L}$ ) は、SpectroChip (Sequenom, Inc.) にピペティングし、結晶化することができた。次いで、サンプル  $0.15 \mu\text{L}$  を加えた。陽性イオンモードで操作する、直線状の PerSeptive Voyager DE マススペクトロメーターまたは Bruker Biflex MALDI-TOF マススペクトロメーターを測定に使用した。当該サンプルプレートを各 UV レーザーショット (合計約 250 レーザーショット) 後は  $400 \text{ nm}$  の  $18.2 \text{ kV}$  で維持し、次いで、標的ボルト数を  $20 \text{ kV}$  に上昇させた。オリジナルのスペクトルは、 $500 \text{ MHz}$  でデジタル化した。

【0254】

実施例 5

サンプルコンディショニング

以下の実施例に示す場合、酵素消化物の産物は、ZipTip (Millipore, Bedford, MA) で精製した。ZipTip は、 $50\%$  アセトニトリル  $10 \mu\text{L}$  で事前に濡らしておき、 $0.1 \text{ M TEAA}$   $10 \mu\text{L}$  で 4 回平衡化した。オリゴヌクレオチドフラグメントを、連続的アスピレーションおよび各サンプルを ZipTip に分配することにより、C18 に結合させた。各消化オリゴヌクレオチドを、 $0.1 \text{ M TEAA}$   $10 \mu\text{L}$  で洗浄し、その後、 $\text{H}_2\text{O}$   $10 \mu\text{L}$  で 4 回洗浄ステップを行うことによりコンディショニングした。DNA フラグメントは、 $50\%$  アセトニトリル  $7 \mu\text{L}$  を用い ZipTip から溶出させた。

## 【0255】

当該サンプルのコンディションの任意の方法を用い得る。一般的にピーク分解の増大に使用するコンディショニングの方法が、既知である(例えば、国際特許出願番号W098/20019参照)。

## 【0256】

## 実施例 6

## DNAグリコシラーゼ仲介配列分析

DNAグリコシラーゼは、特異的核酸塩基がDNA中に存在する各位置においてDNAを修飾し、それにより、脱塩基部位を生ずる。他の酵素、化学、または熱によるその後の反応において、脱塩基部位のリン酸バックボーンが切断され得る。

10

## 【0257】

以下の手順に従い利用されるグリコシラーゼは、ウラシル-DNAグリコシラーゼ(UDG)であった。ウラシル塩基は、ウラシルの存在下DNA標的配列を増幅することによりチミン塩基が通常占める各位置においてDNAフラグメント中に組込まれた。各ウラシル置換DNAアンプリコンは、アンプリコン中の各ウラシル塩基を切断するUDGと共にインキュベーションし、次いでDNAフラグメントを生ずる各脱塩基部位でバックボーン切断に効果的なコンディションにした。DNAフラグメントをMALDI-TOF質量分析法分析した。次いで、標的DNAの遺伝的変異性をマスペクトルを分析することにより評価した。

## 【0258】

本明細書に記載のように、ヌクレオチド類似体または修飾ヌクレオチドに特異的なグリコシラーゼが以下の手順でUDGで置換され得る。下記のグリコシラーゼ法は、リン酸バックボーン切断およびMALDIと共に用いられ、SNPスクランニング、細菌型別化、メチル化分析、マイクロサテライト分析、遺伝子型決定、およびヌクレオチド配列決定および再配列決定の目的でDNAフラグメントを分析し得る。

20

## 【0259】

## A. 遺伝子型決定

グリコシラーゼ手順を用い、UCP-2(非結合タンパク質2)をコードするDNA配列を遺伝子型決定した。UCP-2の配列は、受け例番号AF096289でGenBankに登録されている。以下の手順で遺伝子型決定した配列の変化は、ヌクレオチド4790位におけるシトシン(C-対立遺伝子)からチミン(T-対立遺伝子)への変化であり、その結果、UCP-2ポリペプチドの55位においてアラニンからバリンへの変異を生ずる。

30

## 【0260】

DNAは、配列5'-TGCTTATCCCTGTAGCTACCCTGTCTTGGCCTTGCAGATCCAA-3'(配列番号91)を有する5pmolビオチニル化プライマー、配列5'-AGCGGATAACAATTTACACAGGCCATCACACCGCGGTACTG-3'(配列番号92)を有する15pmol非ビオチニル化プライマー、200μM dATP、200μM dCTP、200μM dGTP、600μM dUTP(完全にdTTPに置換され得る)、1.5mMから3mM MgCl<sub>2</sub>、1U Hotstar Taqポリメラーゼ、およびCEPH DNA 25ngを含む反応体積50μLにPCR手順を用い増幅した。増幅は、アニーリング温度56で45サイクル行った。

40

## 【0261】

次いで、増幅産物は、事前に洗浄したDynabeads5μLを伴う増幅反応物50μLを室温で20分間インキュベーションすることにより、固体支持体に固定化した。上清を取り除き、ビーズを0.1M NaOH50μLと共に室温5分間インキュベーションし、一本鎖DNAをビーズに連結させた方法で二本鎖PCR産物を変性させた。次いで、当該ビーズを10mM TrisHCl(pH8)50μLで3回洗浄することにより中性化した。当該ビーズを60mM TrisHCl/1mM EDTA(pH7.9)溶液中で再懸濁し、1U ウラシルDNAグリコシラーゼを37で45分間、当該溶液に加え、ビーズに連結する一本鎖DNA中に存在するウラシルヌクレオチドを取り除いた。次いで、当該

50

ビーズを、10 mM Tris HCl (pH 8) 25  $\mu$ L で2回洗浄し、水10  $\mu$ L で1回洗浄した。次いで、ビオチン化鎖を、60 で10分間、2 M NH<sub>4</sub> OH 12  $\mu$ L でビーズから溶出した。DNAのバックボーンを、95 10分間、サンプルをインキュベーションすることにより(密閉のふたと共に)、切断し、アンモニアを、サンプルを80 で11分間インキュベーションすることによりサンプルから蒸発させた。

#### 【0262】

次いで、切断フラグメントを、実施例4に記載のようにMALDI-TOF質量分析法より分析した。T対立遺伝子は、3254ダルトンの特有フラグメントを生じた。C対立遺伝子は、4788ダルトンの特有フラグメントを生じた。これらのフラグメントは、マススペクトルで識別可能であった。そのため、上記同定手順は、UCP-2中のC対立遺伝子およびT対立遺伝子にヘテロ接合の個体の遺伝子型決定に成功的に利用できた。

10

#### 【0263】

B. プールしたDNAサンプルを用いるグリコシラーゼ分析

グリコシラーゼアッセイを、プールしたサンプルを用い構成し、UCP-2遺伝子配座の遺伝的多様性を検出した。既知遺伝子型のDNAを11個体からプールし、固定濃度の5 ng /  $\mu$ L に希釈した。実施例3Aで提供された手順は、配列5'-CCCAGTCACGACGTTGTAA AACGTCTTGGCCTTGCAGATCCAAG-3' (配列番号93)を有するフォワードプライマー2 pmol および配列5'-AGCGGATAACAATTTACACAGGCCATCACACCGCGGTAAGT-3' (配列番号94)を有するリバースプライマー15 pmol を用いることができた。加えて、配列5'bioCCCAGTCACGACGTTGTAAAACG-3' (配列番号97)を有するビオチニル化プライマー5 pmol を、約2サイクル後にPCR反応物に導入し得る。当該フラグメントは、MALDI-TOF質量分析法(実施例4)で分析した。実施例3Aで測定したように、3254ダルトンの特有のフラグメントを生じるT対立遺伝子は、マススペクトルで、4788ダルトンの特有のフラグメントを生じるC対立遺伝子と区別し得る。プールしたサンプルの対立遺伝子頻度は、対立遺伝子フラグメントに相当する各シグナルの下の領域を積算することにより定量された。積算は、当業者に既知の等式を用い手計算で行った。11サンプルのプールにおいて、この手順は、40.9%の個体がT対立遺伝子を保有し、59.09%の個体がC対立遺伝子を保有することを提唱した。

20

#### 【0264】

C. グリコシラーゼ仲介マイクロサテライト分析

グリコシラーゼ手順を用い、Bradykinin Receptor 2(BKR-2)配列のマイクロサテライトを同定した。BKR-2の配列は、受け入れ番号X86173でGenBankに登録されている。BKR-2は、プロモーター領域にCからTへの変化のSNPを含み、そして繰り返し単位中にGからTへの変化のSNPを含む。実施例3Aで提供された当該手順を用い、プロモーター領域中のSNP、マイクロサテライト繰り返し単位領域中のSNP、およびBKR-2のマイクロサテライト領域の繰り返し単位の数を同定した。特に配列5'-CTCCAGCTGGGCAGGAGTGC-3' (配列番号95)を有するフォワードPCRプライマーおよび配列5'-CACTTCAGTCGCTCCCT-3' (配列番号96)を有するリバースプライマーを用い、ウラシル存在下、BKR-2 DNAを増幅した。当該アンプリコンは、UDGによりフラグメント化し、その後、バックボーン切断した。当該切断フラグメントは実施例4に記載のようにMALDI-TOF質量分析法により分析した。

30

40

#### 【0265】

CからTへの変化を有するBKR-2プロモーター領域中のSNPに関し、C対立遺伝子は、質量7342.4ダルトンを有する特有のフラグメントを生じ、T対立遺伝子は、質量7053.2ダルトンを有する特有のフラグメントを生じた。これらフラグメントは、マススペクトルで区別できた。そのため、上記特定した手順は、BKR-2のプロモーター領域中のC対立遺伝子およびT対立遺伝子についてヘテロ接合の個体を遺伝子型決定するため成功的に利用できた。

#### 【0266】

GからTへの変化を有するBKR-2の繰り返し領域中のSNPに関し、T対立遺伝子

50

は、質量 1784 ダルトンを有する特有のフラグメントを生じ、それは、容易にマスペクトルで検出できた。それゆえ、T 対立遺伝子の存在は、BKR - 2 の繰り返し領域中の G から T への配列変化を示唆した。

【0267】

加えて、繰り返し領域の数は、BKR - 2 中に、2 つの繰り返し配列を有する個体と 3 つの繰り返し配列を有する個体とを区別できた。これら個体の DNA は、各繰り返し配列が SNP 遺伝子配座で G を含んでいるため、繰り返し配列中の G から T への配列変化を保持していなかった。繰り返し領域の数は、質量 2771.6 ダルトンを有する特有の DNA フラグメントに相当するシグナルの下の面積を算出することにより、個々のサンプルで決定された。2 つの繰り返し領域を有する個体から生ずるスペクトルにおけるこのシグナルは、3 つの繰り返し領域を有する個体から生ずるスペクトル中の同じシグナルの下の面積の 33% 未満の領域である、領域を有していた。そのため、上記考察の手順を用い、BKR - 2 中に存在する繰り返し領域の数について個体の遺伝子型決定することができた。

【0268】

D. グリコシラーゼ消化と合わせたバイサルファイト

ゲノム DNA のバイサルファイト処置を用い、DNA 内のメチル化シトシン残基の位置を分析することができる。バイサルファイトによる核酸処置により、シトシン残基をウラシル残基にジアミノ化し、その一方、メチル化シトシンは修飾しないままである。そのため、バイサルファイトで処置していないゲノム DNA から生ずる PCR 産物の配列を、バイサルファイトで処置するゲノム DNA から生じた PCR 産物の配列と比較することにより、核酸中およびシトシンがメチル化されている位置におけるメチル化の程度を推測することができる。

【0269】

ゲノム DNA (2  $\mu$ g) を、制限酵素 1  $\mu$ L で 37  $^{\circ}$ C 2 時間インキュベーションすることにより消化した。3 M NaOH のアリコートを追加し、消化溶液中、最終濃度 0.3 M NaOH を得た。当該反応物を 37  $^{\circ}$ C 15 分間インキュベーションし、その後、5.35 M 尿素、4.44 M バィサルファイトおよび 10 mM ヒドロキノリンで処置し、この場合、ヒドロキノリンの最終濃度は 0.5 mM であった。

【0270】

バイサルファイト (サンプル A) で処置したサンプルを、バイサルファイト処置 (サンプル B) していない同じ消化サンプルと比較した。サンプル A を上記のようにバイサルファイトで処置した後、サンプル A およびサンプル B を標準的な PCR 手順により増幅した。当該 PCR 手順には、各サンプルをミネラルオイルで覆い、当該サンプルをサーマルサイクリング (55  $^{\circ}$ C 15 分間、その後、95  $^{\circ}$ C 30 秒間の 20 サイクル) するステップを含む。当該 PCR 反応物には、4 つのヌクレオチド塩基、C、A、G、および U を含む。当該ミネラルオイルを各サンプルから取り除き、PCR 産物をガラスミルクで精製した。ヨウ化ナトリウム (3 体積) およびガラスミルク (5  $\mu$ L) をサンプル A およびサンプル B に添加した。次いで、当該サンプルを氷上に 8 分間置き、冷緩衝液 420  $\mu$ L で洗浄し、10 秒間遠心分離し、上清画分を取り除いた。この方法を 2 回繰り返し、次いで、水 25  $\mu$ L を添加した。サンプルを 37  $^{\circ}$ C 5 分間インキュベーションし、20 秒間遠心分離し、上清画分を回収し、そしてこのインキュベーション / 遠心分離 / 上清画分回収の手順を繰り返した。次いで、0.1 M NaOH 50  $\mu$ L を当該サンプルに添加し、DNA を変性させた。当該サンプルを室温 5 分間インキュベーションし、10 mM Tris HCl (pH 8) 50  $\mu$ L で 3 回洗浄し、そして 60 mM Tris HCl / 1 mM EDTA、pH 7.9 10  $\mu$ L 中に再懸濁した。

【0271】

次いで、サンプル A およびサンプル B からの PCR 産物の配列を 2 UUDG (MBI Fermentas) で処理し、次いで、上記のようにバックボーンを切断した。各サンプル A およびサンプル B から得られたフラグメントは実施例 4 に記載のように MALDI - TOF 質量分析法により分析した。サンプル A は、サンプル B から得られるフラグメントの数よりも多

いフラグメントの数が生じ、これは、核酸が少なくとも1つのメチル化シトシン部分を保持することを示唆する。

【0272】

実施例7

Fenリガーゼ仲介ハプロタイピング

ハプロタイピング手順により、個体の2つの相同性クロモソームの1つからフラグメントを選択し得、そしてそのフラグメントにおける連鎖したSNPを遺伝子型決定し得る。ハプロタイプの直接決定により、情報コンテンツを増加し、任意の連鎖疾患遺伝子の診断を改善し、またはこれら疾患の関連を同定することができる。従前の研究では、ハプロタイプは、典型的には、系統分析(系統が利用可能である場合)を介して、困難で信頼性のない対立遺伝子特異的PCRを介して、または当分野に既知の単一分子希釈を介して、再構成された。

10

【0273】

ハプロタイピング手順を用い、DNAサンプル中の一本鎖に存在する、SNP1およびSNP2と称せられる2つのSNPの存在を決定した。ハプロタイピング手順を、Fen-1を利用するこのアッセイに使用し、DNA“フラップ(flap)”を切断する部位特異的“フラップ”エンドヌクレアーゼは、標的DNA鎖にハイブリダイズする2つのオリゴヌクレオチドのオーバーラップにより作成した。この実施例の当該2つのオーバーラッピングオリゴヌクレオチドは、短アームおよび長アーム対立遺伝子特異的アダプターであった。標的DNAは、変性しSNP1およびSNP2を含む増幅核酸であった。

20

【0274】

短アームアダプターには、標的DNAには見られない特有の配列を含む。短アームアダプターの3'末端ヌクレオチドは、SNP1対立遺伝子の1つと同一であった。更に、長アームアダプターには、2つの領域：短アームに相補的な3'領域およびSNPに隣接する目的のフラグメントに相補的な5'遺伝子特異的領域を含む。アダプターとその相同体の1つとの間にマッチがあれば、Fen酵素は、認識し、オーバーラッピングフラップを切断する。次いで、当該アダプターの短アームを標的フラグメントの残り(SNP部位を除く)にライゲーションした。このライゲーションしたフラグメントを、ライゲーションした相同体のみを増幅する第二のPCR反応のフォワードプライマーとして用いた。次いで、第二のPCR産物(PCR2)を質量分析法で分析した。アダプターと標的DNAとの間にマッチがなければ、Fen-1によるオーバーラップまたは切断はなく、そのため、目的のPCR2は存在しない。

30

【0275】

目的配列中に1を超えるSNPが存在するならば、第二のSNP(SNP2)は、SNP2に特異的なアダプターを用いることにより、および第一のSNPを含むPCR2産物にアダプターをハイブリダイズさせることにより、発見できた。Fenリガーゼおよび増幅手順は、第一のSNPを含むPCR2産物で繰り返した。増幅産物が第二のSNPを生ずるならば、SNP1およびSNP2は同じフラグメントにあった。

【0276】

SNPが知られていないならば、4つの対立遺伝子特異的アダプター(例えば、C、G、AおよびT)を用い標的DNAとハイブリダイズさせることができる。次いで、基質を、増幅を含むFenリガーゼプロトコールで処理する。PCR2産物を本明細書に記載のようにPROBEで分析し、どのアダプターがDNA標的にハイブリダイズするかを決定し、それにより、当該配列中のSNPを同定し得る。

40

【0277】

Fenリガーゼアッセイを用い、因子VII中に存在する2つのSNPを検出した。これらのSNPは、お互いから814塩基対離れて位置している。SNP1は、8401位(CからTへ)に位置し、SNP2は、9215位(GからAへ)に位置する(配列番号#)

【0278】

A. 最初の増幅ステップ

50

PCR産物(PCR1)は、SNPの5'末端から僅かに離れたSNP1における既知ヘテロ接合体として生じた。特に、PCR反応物10 $\mu$ Lは、1.5mM MgCl<sub>2</sub>、200 $\mu$ M各dNTP、0.5U Hotstarポリメラーゼ、配列5'-GCGCTCCTGTCCGGTGCCA-3'(配列番号56)を有するフォワードプライマー0.1 $\mu$ M、配列5'-GCCTGACTGGTGGGGCC-3'(配列番号57)を有するリバースプライマー0.1 $\mu$ M、およびゲノムDNA1ngを混合することにより行った。アニーリング温度は58 $^{\circ}$ Cであり、増幅処理により、長さが861であるフラグメントを得た。

#### 【0279】

PCR1反応混合物を半分に分け、1.0 $\mu$ L SAPおよび0.1 $\mu$ L エキソン1を含むエキソヌクレアーゼ1/SAP混合物(0.22 $\mu$ L混合物/5 $\mu$ L PCR1反応物)で処理した。エキソヌクレアーゼ処理は、37 $^{\circ}$ C 30分間行い、次いで、85 $^{\circ}$ C 20分間行いDNAを変性させた。

#### 【0280】

##### B. アダプターオリゴヌクレオチド

アダプターあたり、1つの長および1つの短オリゴヌクレオチドを含む対立遺伝子特異的アダプター(CおよびT)の溶液を調製した。各アダプター(10 $\mu$ M)の長アームおよび短アームオリゴヌクレオチドを、1:1の割合で混合し、95 $^{\circ}$ C 30秒間加熱した。当該温度を、アニーリングのため2 $^{\circ}$ Cの誤差で、37 $^{\circ}$ Cに下げた。Cアダプターは、5'-CATGCATGCACGGTC-3'(配列番号58)の短アーム配列および5'-CAGAGAGTACCCCTCGACCGTGCATGCATG-3'(配列番号59)の長アーム配列を有する。これゆえ、アダプターの長アームは30bp(15bp遺伝子特異的)であり、短アームは15bpであった。Tアダプターは、5'-CATGCATGCACGGTT-3'(配列番号60)の短アーム配列および5'-GTACGTACGTGCCAACTCCCATGAGAGAC-3'(配列番号61)の長アーム配列を有した。当該アダプターはまた、3から10ヌクレオチド(配列番号118)を含むループにより短および長アームが分離されるヘアピン構造を有し得る。

#### 【0281】

##### C. Fenリガーゼ反応

2つのチューブに(サンプルあたりの各対立遺伝子特異的アダプターとして1チューブ)、10mM 16%PEG/50mM MOPS 3.5 $\mu$ l、25mM MgCl<sub>2</sub> 1.2 $\mu$ l、10x Ampligase緩衝液1.5 $\mu$ l、および2.5 $\mu$ l PCR1を含む溶液(溶液A)を入れた。溶液Aを含む各チューブは、95 $^{\circ}$ C 5分間インキュベーションし、PCR1産物を変性させた。Ampligase(熱安定性リガーゼ、Epicentre Technologies)1.65 $\mu$ l、200ng/ $\mu$ l MFEN(Methanococcus jannaschii由来)1.65 $\mu$ lおよび対立遺伝子特異的アダプター(CまたはT)3.0 $\mu$ lを含む第二の溶液(溶液B)を調製した。溶液Bの種々の変化物、種々の対立遺伝子特異的アダプターを含む各変化物を作成した。溶液Bを95 $^{\circ}$ Cで溶液Aに添加し、55 $^{\circ}$ C 3時間インキュベーションした。全反応物体積をアダプター特異的反応物あたり15.0 $\mu$ lとした。二対立遺伝子システム(bi-allelic system)では、2x15.0 $\mu$ l反応物を必要とした。

#### 【0282】

次いで、各チューブにおけるFenリガーゼ反応物を、10mM EDTA 8.0 $\mu$ lを加えることにより不活性化した。次いで、エキソIII/緩衝液(70%/30%)溶液1.0 $\mu$ lを各サンプルに加え、37 $^{\circ}$ C 30分間、70 $^{\circ}$ C 20分間(エキソIIIを不活性化)、および95 $^{\circ}$ C 5分間(サンプルを変性させ、使用していないアダプターを鋳型から分離した)インキュベーションした。当該サンプルを氷スラリーで冷却し、長さ100塩基対未満のすべてのフラグメントを取り除くUltraClean PCR Clean-up(MoBio)スピンカラムで精製した。当該フラグメントはH<sub>2</sub>O 50 $\mu$ lで溶出した。

#### 【0283】

##### D. 第二の増幅ステップ

第二の増幅反応(PCR2)を、各サンプルチューブにおいて、フォワードプライマーとして短アームアダプター(CまたはT)配列を用い(SNP1部位を除く)、行った。ライゲ

ーション相同体のみを増幅した。標準的PCR反応物は、1×緩衝液(最終濃度)、1.5 mM 最終濃度MgCl<sub>2</sub>、200 μM最終濃度dNTP、0.5 U Hotstar ポリメラーゼ、0.1 μM 最終濃度フォワードプライマー5'-CATGCATGCACGGT-3'(配列番号62)、0.1 μM 最終濃度リバースプライマー5'-GCCTGACTGGTGGGGCCC-3'(配列番号63)、および精製FENリガーゼ反応溶液1.0 μlを含む全体積10.0 μlで行った。アニーリング温度は、58 °Cであった。PCR2産物を実施例4で示したようにMALDI-TOF質量分析法で分析した。Fen SNP1のマススペクトルは、質量6084.08ダルトンであり、それは、C対立遺伝子を示している。

#### 【0284】

##### E. 付加SNPの遺伝子型決定

第二のSNP(SNP2)は、SNP2に特異的なアダプターを用いることにより、および第一のSNPを含むPCR2産物にアダプターをハイブリダイズさせることにより、発見できる。Fenリガーゼおよび増幅手順は、第一のSNPを含むPCR2産物で繰り返す。増幅産物が第二のSNPを生ずるならば、SNP1およびSNP2は同じフラグメントにある。T対立遺伝子を示すSNP2のマススペクトルは質量6359.88ダルトンであった。

#### 【0285】

このアッセイはまた、プールしたDNAで行い、本明細書に記載するようにハプロタイプ頻度を生ずることが可能である。Fenリガーゼアッセイを用い、本明細書に記載のようにマルチプレックスを分析できる。

#### 【0286】

##### 実施例8

##### ニッカーゼ仲介配列分析

DNAニッカーゼまたはDNaseを用い、DNA二重らせんの一本鎖を認識し切断した。使用した2つのニッカーゼはNY2AニッカーゼおよびNYS1ニッカーゼ(Megabase)であり、それらは以下の部位でDNAを切断する：

#### 【化8】

NY2A: 5'...RAG...3'

3'...Y↓TC...5' ここでR = AまたはG, Y = CまたはT

NYS1: 5'...↓CC[A/G/T]...3'

3'... GG[T/C/A]...5'.

#### 【0287】

##### A. ニッカーゼ消化

Tris-HCl(10 mM)、KCl(10 mM、pH 8.3)、酢酸マグネシウム(25 mM)、BSA(1 mg/mL)、および6 U Cvi NY2AまたはCvi NYS1ニッカーゼ(Megabase Research)を、標準的ホスホラミダイト化学を用い合成される配列5'-CGCAGGGTTTCCTCGTCGCACTGGGCATGTG-3'(配列番号90、Operon, Alameda, CA)を有する二本鎖オリゴヌクレオチド鑄型25 pmolに加えた。全体積20 μLを用い、当該反応混合物を37 °C 5時間インキュベーションし、当該消化産物を実施例5に記載するようにZipTip(Millipore, Bedford, MA)を用い精製した。当該サンプルを、実施例1に記載するようにMALDI-TOF質量分析法で分析した。ニッカーゼCvi NY2Aは質量4049.76ダルトン、5473.14ダルトンおよび9540.71ダルトンの3つのフラグメントを生じた。Cvi NYS1ニッカーゼは質量2063.18ダルトン、3056.48ダルトン、6492.81ダルトンおよび7450.14ダルトンのフラグメントを生じた。

#### 【0288】

##### B. プールしたサンプルのニッカーゼ消化

DQA(HLA Class II-DQ Alpha、予想フラグメントサイズ = 225 bp)は、100の健

10

20

30

40

50

康個体のゲノムDNAから増幅した。DQAは、標準的PCR化学を用い、10mM Tris-HCl、10mM KCl (pH 8.3)、2.5mM MgCl<sub>2</sub>、200μM 各dNTP、配列5'-GTGCTGCAGGTGTAACTTGTACCAG-3' (配列番号64)を有するフォワードプライマー10pmol、配列5'-CACGGATCCGGTAGCAGCGGTAGAGTTG-3' (配列番号65)を有するリバースプライマー10pmol、1U DNAポリメラーゼ(Stoffel fragment, Perkin Elmer)、およびヒトゲノムDNA(2ng DNA/個体)200ngを含む全体積50μlを有する反応物中で増幅した。当該鋳型を94 5分間変性させた。サーマルサイクリングは、94 20秒、56 30秒、72 1分間の45サイクルおよび72 3分間の最終伸張を含むタッチダウンプログラムで続けた。粗PCR産物をその後のニッカーゼ反応に用いた。

10

## 【0289】

非精製PCR産物をニッカーゼ消化した。Tris-HCl(10mM)、KCl(10mM、pH 8.3)、酢酸マグネシウム(25mM)、BSA(1mg/mL)、および5U Cvi NY2AまたはCvi NYS1ニッカーゼ(Megabase Research)を、全反応物体積20μLを有する増幅鋳型25pmolに加えた。次いで、当該反応混合物を37 5時間インキュベーションした。当該消化産物を実施例5に記載するようにZipTip(Millipore, Bedford, MA)を用い精製した。当該サンプルを、実施例4に記載するようにMALDI-TOF質量分析法で分析した。また、このアッセイを用い、本明細書に記載のように、マルチプレックシングし、標準的遺伝子型決定を行い得る。

20

## 【0290】

ニッカーゼマスマスペクトルを単純化するため、2つの相補鎖は、捕捉プローブとして一本鎖非消化PCR産物を用いた消化の後、分離することができる。このプローブ(以下の実施例8Cで示した調製物)を、200mM クエン酸ナトリウムおよび1%プロッキング剤(Boehringer Mannheim)を含むハイブリダイゼーション緩衝液中のニッカーゼフラグメントにハイブリダイズさせることができる。当該反応物を、thermal cycler(PTC-200 DNA engine, MJ Research, Waltham, MA)を用い95 5分間加熱し、30分間を超えて室温で冷却する。捕捉プローブニッカーゼフラグメントをストレプトアビジン被覆磁性ビーズ140μg上に固定化した。その後、当該ビーズを70mM クエン酸アンモニウムで3回洗浄する。捕捉一本鎖ニッカーゼフラグメントを、50mM 水酸化アンモニウム 5μL中で80 5分間加熱することにより、溶出する。

30

## 【0291】

## C. 捕捉プローブの調製

捕捉プローブは、PCR法を介し、GeneAmp 1×PCR緩衝液II、10mM Tris-HCl、pH 8.3、50mM KCl、2mM MgCl<sub>2</sub>、0.2mM dNTP混合物、各プライマー(フォワードプライマー5'-ACTGGGCATGTGGAGACAG-3' (配列番号66))10pmolおよびビオチニル化リバースプライマー bio5'-GCACTTTCTTGCCATGAG-3' (配列番号67)、2U AmpliTaq Gold、およびヒトゲノムDNA 200ngを含む全体積50μL中でヒト グロビン遺伝子を増幅することにより、調製する。当該鋳型は94 8分間で変性させる。サーマルサイクリングは、94 20秒、64 30秒、72 1分間の11サイクルおよび72 5分間の最終伸張を含むタッチダウンプログラムで続けた。当該アンプリコンは、UltraClean(商標) PCR clean-upキット(MO Bio Laboratories, Solano Beach, CA)を用い精製する。

40

## 【0292】

## 実施例9

## マルチプレックスIIS型SNPアッセイ

IIS型アッセイを用い、既知SNPを有するヒト遺伝子配列を同定した。このアッセイ中で使用したIIS型酵素は、標的DNAを二本鎖切断するフォークIであった。当該アッセイは、増幅のステップおよびアンプリコンのフォークI処理を含む。増幅ステップでは、フォークI認識配列がアンプリコンの5'および3'末端に組込まれるため、設計遺伝子標的の各PCR産物が100塩基未満となるように当該プライマーを設計した。そ

50

のため、フォーク I で切断されるフラグメントは、目的の S N P を含む中央フラグメントを含んだ。

#### 【 0 2 9 3 】

既知 S N P を有する 1 0 のヒト遺伝子標的は、このアッセイにより分析された。1 0 の遺伝子標的、および標的領域の増幅に用いるプライマーの配列を表 5 に示す。1 0 の標的は、リポタンパク質リパーゼ、プロトロンビン、因子 V、コレステロールエステル転移タンパク質 (C E T P)、因子 V I I、因子 X I I I、H L A - H エキソン 2、H L A - H エキソン 4、メチレンテトラヒドロフォレートレダクターゼ (M T H R) および P 5 3 エキソン 4 コドン 7 2 であった。

#### 【 0 2 9 4 】

1 0 の遺伝子配列の増幅は、5 P C R 反応チューブ中にヒトゲノム D N A 鋳型 2 0 n g を有する単一の 5 0 μ L 体積 P C R 反応物中で行った。各反応バイアルは、1 × P C R 緩衝液 (Qiagen)、2 0 0 μ M d N T P、1 U Hotstar T a q ポリメラーゼ (Qiagen)、4 m M M g C l<sub>2</sub>、および各プライマー 1 0 p m o l を含んでいた。5' TCAGTCACGACGTT3' (配列番号 6 8) の配列を有する U S 8、および 5' CGGATAACAATTTTC3' (配列番号 6 9) の配列を有する U S 9 を、それぞれフォワードおよびリバースプライマーとして使用した。更に、当該プライマーは、フォーク I 認識部位がアンプリコンの 5' および 3' 末端に組込まれるように、設計した。サーマルサイクリングは、0.2 mL チューブまたは 9 6 ウェルプレート中で、MJ Research Thermal Cycler (算出温度) を用い、以下のサイクリングパラメーターで行った：9 4 5 分間；4 5 サイクル：9 4 2 0 秒、5 6 2 0 秒、7 2 6 0 秒；および 7 2 3 分間。

#### 【 0 2 9 5 】

P C R 後、当該サンプルを、0.2 U エキソヌクレアーゼ I (Amersham Pharmacia) および S アルカリホスファターゼ (Amersham Pharmacia) で処理し、組込まれていないプライマーおよび d N T P を取り除いた。典型的に、0.2 U エキソヌクレアーゼ I および S A P を P C R サンプル 5 μ L に加えた。次いで、当該サンプルを 3 7 1 5 分間インキュベーションした。次いで、エキソヌクレアーゼ I および S A P を、8 5 までの温度で 1 5 分間サンプルを加熱することにより、不活性化した。フォーク I 消化は、2 U フォーク I (New England Biolab) を P C R サンプル 5 μ L に加えることにより、および 3 7 3 0 分間インキュベーションすることにより、行った。フォーク I 制限酵素部位は、アンプリコンの両サイドに位置するため、5' および 3' カットオフフラグメントは S N P を含む中央フラグメントよりも大きい質量を有する。次いで、当該サンプルを陰イオン交換により精製し、実施例 4 に記載のように M A L D I - T O F 質量分析法により分析した。このマルチプレクシング実験からの遺伝子フラグメントの質量を表 6 に挙げる。これら遺伝子フラグメントはマススペクトルで分解し、それにより、この遺伝子中の配列変化をマルチプレクシング分析し得る。

表 5

マルチプレクシング I I S 型アッセイの遺伝子

10

20

30

【表10】

遺伝子	配列	配列番号	プライマー	配列番号
リポタンパク質リパーゼ (Asn291Ser)	cctttgagaa agggctctgc ttgagttgta gaaagaaccg ctgcaacaat <u>ctgggctatg agatcaia&gt;gitea apcagagcc</u> <u>aaaagaagca gcaaaatgta</u>	98-99	5' caatttcacgctggatgcaatct gggctatgagatc 3'	70
			5' caatttcacacagcggatgcttct ttggctctgact 3'	71
プロトロンビン	26731 gaattatmtt tgtgttcta aaactatggt <u>tccaataaa agtgactctc</u> 26781 <u>agcig&gt;alagcctc aatgctccca</u> <u>gtgctattca tggcagctc tctgggctca</u>	100-101	5' tcagtcacgacgttggatgcca <u>taaaatgactctcagc 3'</u>	72
			5' cggataacaatttcggatgcaact <u>gggagcattgagcc 3'</u>	73
因子V (Arg506Gln)	taataggact acttctaate tgtaagagca <u>gatccctgga caggcig&gt;alagga</u>  <u>atcacagtat tttgtccttg aagtaacctt tcag</u>	102-103	5' tcagtcacgacgttggatgagca <u>gatccctggacagcc 3'</u>	74
			5' cggataacaatttcggatgagca <u>aastacctgtattcc 3'</u>	75
コレステロールエステル転移タンパク質 (CETP) (I405V)	1261 ctaccatgg gcatttgatt <u>gcagagcage</u> <u>tccagatccig&gt;a) tccagagctt</u> 1311 <u>cctgcantca atgatcaccg ctgtggccat</u> cctgaggtc atgtctgta	104-105	5' tcagtcacgacgttggatgagca <u>gcagctccagtc 3'</u>	76
			5' <u>cagcggatgatcattggatgagc</u> <u>aagctctgg 3'</u>	77
因子VII (R353Q)	1221 agcaaggact cctgcaaggg ggacaptgga <u>ggccacatg ccaccacta</u> 1271 <u>gcia&gt;gigggcacg tggtaacctga</u> <u>cgggcatcgt cagctggggc cagggctcgg</u>	106-107	5' tcagtcacgacgttggatgcca <u>catgccaccactac 3'</u>	78
			5' cggataacaatttcggatgccc <u>tcaggtaccacg 3'</u>	79
因子XIII (V34L)	111 caataactct aatgcagcgg aagatgacct <u>gcccacagtg gagcttcagg</u> 161 <u>gcig&gt;tgggtgcc cggggcgtc</u> <u>aacctgcaag gtatgagcat accccccttc</u>	108-109	5' tcagtcacgacgttggatgcca <u>cagtgagcttcag 3'</u>	80
			5' <u>gctcataccttcagagatgacg</u> 3'	81
HLA-IIキリン2 (His83Asp)	361 ttgaagcttt gggctacgtg <u>gatgaccagc</u> <u>tggtcgtgtt ctatgat(c&gt;g)at</u> 411 <u>gagatcgcg gtgtgagcc ccgaactcca</u> <u>tgggttcca gtagaatttc</u>	110-111	5' tcagtcacgacgttggatgcca <u>gctattcgtgttc 3'</u>	82
			5' tacctgaggttcgggagatgaca <u>cgccgactctc 3'</u>	83
HLA-IIキリン4 (Cys282Tyr)	1021 ggataacctt ggtgtacc cctggggag <u>agcagagata tacgtig&gt;alccag</u> 1071 <u>gtgggcacc caggcctgga tcagccctc</u> attgtatct gggagccctc	112-113	5' tcagtcacgacgttggatgggg <u>ggagcagagatatacgt 3'</u>	84
			5' <u>gagggactatccagatgggt</u> <u>gctccac 3'</u>	85

10

20

30

【表 1 1】

遺伝子	配列	配列番号	プライマー	配列番号
メチレンテトラヒドロフォレートレダクターゼ (MTHFR) (Ala222Val)	761 <u>tgaaacactt gaagga gaag gtatctcgg</u> <u>gag(c&gt;t)cgattt catcaicacg</u>	114-115	5' <u>tcagtcacgacgttggatggaga</u> <u>agagcagagatacgt 3'</u>	86
	811 <u>cagctttct ttgaggctga cacattcttc</u>		5' <u>gagggctgatccaggatgggt</u> <u>gctcac 3'</u>	87
P53 エキソン 4 コドン 72 (Arg72Pro)	12101 <u>tccagatgaa gctcccaga</u> <u>tgccagagge tgcctccc(g&gt;c)g gtggccctg</u>	116-117	5' <u>gatgagctcccagatgccag</u> <u>agac 3'</u>	88
	12151 <u>caccagcagc tctacaccg</u> <u>gagccctg</u>		5' <u>gcccggatagagatgctgctg</u> <u>gtgc 3'</u>	89

10

表 6

I I S アッセイによる 1 0 種の S N P の分類分けのための中央フラグメントの質量

【表 1 2】

遺伝子	LPL( <sup>rs</sup> 2915 <sup>rs</sup> )		プロトロンビン		FV( <sup>rs</sup> 506 <sup>rs</sup> )		CETP( <sup>rs</sup> 405 <sup>rs</sup> )		FVII( <sup>rs</sup> 353 <sup>rs</sup> )		FXIII( <sup>rs</sup> 34)	
	A	G	G	A	G	A	G	A	G	A	G	T
遺伝子型												
+鎖質量 (Da)	6213	6229	5845	5829	5677	5661	3388	3372	6128	6112	5058	5033
-鎖質量 (Da)	6129	6114	5949	5964	5472	5487	3437	3452	6174	6189	4816	4940

遺伝子	H1ah2		H1ah4		MTHR( <sup>rs</sup> 222 <sup>rs</sup> )		P53T( <sup>rs</sup> 72 <sup>rs</sup> )	
	C	G	G	A	C	T	G	C
遺伝子型								
+鎖質量 (Da)	5889	5929	4392	4376	4400	4415	4586	4546
-鎖質量 (Da)	5836	5796	4319	4334	4368	4352	4724	4764

10

20

30

【0296】

実施例 10

健康体データベースの層化のための親病歴パラメーターの典型的使用

健康体データベースを用い、年齢と対立遺伝子、特にホモ接合遺伝子型との間に強力な関係が見られることが判明した特定対立遺伝子(SNP)と疾患状態とを関連させることができる。当該方法は、年齢に依存する関連性の同定に用いられる同じ健康体データベースを用いることを含むが、層化は、両親が患っている通常疾患(ドナーの家族性の病歴)に関するドナーにより提供される情報に基づく。ドナーが両親の健康状態に関し得られる3つの可能性ある答えがある:全く影響はない、一方の影響を受ける、または両方の影響を受ける。ドナーの両親が臨床的疾患の表現型が見られるぐらいにまで加齢していなければならぬため、疾患に依存して、特定の最小年齢を超えるドナーのみを用いる。これらの各群の遺伝子型頻度が決定され、互いに比較される。ドナーのマーカーが疾患と関連すれば、ヘテロ接合遺伝子型の頻度は増大する。ホモ接合遺伝子型の頻度は、健康体集団において表示不足(underrepresented)が有意であるため、増大しない。

40

【0297】

50

## 実施例 1 1

## 生物学的サンプルを同定する方法および装置

## 詳細

本発明に従い、生物学的サンプルを同定する方法および装置を提供する。ここで、図 2 4 について言及すると、生物学的サンプルを同定する装置 1 0 を開示する。生物学的サンプルを同定する装置 1 0 は、一般的に、コンピューティング装置 2 0 に伝達するマススペクトロメーター 1 5 を含む。好ましい実施態様では、マススペクトロメーターは、Bruker -Franzen Analytik GmbHにより製造されたMALDI - TOFマススペクトロメーターであり得る；しかし、他のマススペクトロメーターが置換され得ることが認識される。コンピューティング装置 2 0 は、好ましくは汎用コンピューティング装置である。しかし、コンピューティング装置は、他に形成され得、例えば、マススペクトロメーターに組込まれ得るか、または巨大なネットワークシステム中のコンピューターの一部であり得ることが認識される。

10

20

30

40

50

## 【0298】

生物学的サンプルを同定するための装置 1 0 は、マススペクトロメーター 1 5 の受入エリア 3 1 にサンプルプレート 2 9 を送達するよう形成したロボットアーム 2 7 を有するロボット 2 5 を有する自動化同定システムとして操作し得る。その形式では、同定されるサンプルをプレート 2 9 に置き、それは自動的にマススペクトロメーター 1 5 中へと受け取られる。次いで、生物学的サンプルをマススペクトロメーター中で処理し、生物学的サンプル中でDNAフラグメントの質量を示すデータを得る。このデータをコンピューティング装置 2 0 に直接送るか、またはマススペクトロメーター内で行われるプレプロセッシングまたはフィルタリングをし得る。好ましい実施態様では、マススペクトロメーター 1 5 は、未プロセッシングおよび未フィルタリングの質量分析法データをコンピューティング装置 2 0 に伝達する。しかし、コンピューティング装置での分析を、マススペクトロメーター内で行われるプレプロセッシングまたはフィルタリングに対応するように調節し得ることが認識される。

## 【0299】

ここで、図 2 5 について言及すると、生物学的サンプルを同定するための一般的方法 3 5 を示す。方法 3 5 では、データは、ブロック 4 0 中の試験装置からコンピューティング装置中へ受け取られる。好ましくは、当該データは、未処理、未プロセスおよび未フィルターの形態で受け取られるが、他に、適用される幾つかの形態のフィルタリングまたはプロセッシングであり得る。好ましい実施対応の試験装置は、上記のようなマススペクトロメーターである。しかし、他の試験装置がマススペクトロメーターと置換されることが認識される。

## 【0300】

試験装置により生ずるデータ、および特にマススペクトロメーターは、生物学的サンプルの同定を示す情報を含む。より特に、当該データは、生物学的サンプルのDNA組成物を示す。典型的に、DNA増幅技術から得られたDNAサンプルから集められた質量分析法データは、例えば、典型的なタンパク質サンプル由来のものよりもノイズが多い。これは、タンパク質サンプルがより容易により多量に調製されること、およびタンパク質サンプルはDNAサンプルと比較してより容易にイオン化することに一部起因する。従って、通常のマススペクトロメーターデータ分析技術は、一般的に、生物学的サンプルのDNA分析に効果的ではない。DNA組成物データがより容易に識別できるように分析能を改善するため、好ましい実施態様は、DNA質量分析法データを分析するためのウェーブレット技術を用いる。ウェーブレットは、シグナルプロセッシング、数値解析、および数学的モデリングのための分析ツールである。ウェーブレット技術は、データセットに適用される基本的な拡張機能を提供する。ウェーブレット分解を用い、当該データセットを同時に時間および頻度の範囲内で分析し得る。ウェーブレット変換は、MALDI - TOF DNAデータのような複雑な時間(質量)および頻度の領域を示すデータの分析における選択の技術である。本明細書で述べるようなウェーブレット変換は、通常のフーリエ分析技術

との比較として優れたノイズ除去性を有する。ウェーブレット変換は、DNAサンプルの固有のノイズMALDI-TOFスペクトルのインタープリティングに特に効果的であることが証明された。ウェーブレットの使用では、“小さな波”または“スケーリング関数”を用いステージにデータセットを変換し、各ステージはデータセット中の頻度コンポーネントを示す。ウェーブレット変換を用い、質量分析法データをプロセスし、フィルターし、および生物学的サンプルのDNA組成物の同定に有用な有意な識別で分析し得る。

#### 【0301】

再び、図25について言及すると、ブロック40で受け取るデータは、ブロック45でノイズ除去される。次いで、ノイズが除去されたデータは、ブロック50で適用されるベースライン補正を有する。ベースライン補正は、試験装置、特にマスマススペクトロメーター装置から来るデータとして一般的に必要であり、一般的に、指数的減衰手法で配列されるデータを有する。この一般的な指数的減衰配列は、生物学的サンプルの組成物には起因しないが、試験装置の物理的性質および特性の結果であり、他の化学物質には、DNAサンプル調製物が含まれる。従って、ベースライン補正は、実質的にデータを校正し、試験システムに帰するデータ、およびサンプル調製物特性のコンポーネントを取り除く。

#### 【0302】

ブロック45でノイズ除去し、ブロック50でベースライン補正した後、シグナルは残存し、生物学的サンプルの組成物を一般的に示す。しかし、生物学的サンプルのDNA組成物の分析に必要とされる通常以上の識別のため、当該組成物は、ノイズ除去化および校正化シグナルからは容易には明白とはならない。例えば、シグナルはピーク領域を含み得るが、これらの“推定”ピークが実際にDNA組成物を表しているのかどうか、または推定ピークが合成的もしくは化学的な異常型の結果であるのかどうか、まだ明らかとはならない。更に、生物学的サンプルの組成物の任意のコール(call)は、臨床的または治療的目的としては受け入れがたいエラーの可能性を有し得る。そのような危険な状態では、サンプルの任意のコールまたは同定が正確であるという、高い程度の確実性が必要である。そのため、更なるデータプロセッシングおよびインタープリテーションが、サンプルを正確におよび確実に同定する前に必要となる。

#### 【0303】

各質量分析法試験から得られたデータの量は、典型的には数千のデータポイントであり、自動システムがセットされて時間あたり数百または更には数千の試験を行うため、得られた質量分析法データの量は膨大となる。質量分析法データの効率的伝達および保存を促進するため、ブロック55は、ノイズ除去化およびベースライン補正データが圧縮されることを示す。

#### 【0304】

好ましい実施態様では、生物学的サンプルを選択し、制限された範囲の可能な組成物のみを有するよう処理をする。従って、そのため、組成物を示すピークがどこに位置するかが、存在するならば、判る。これら予想されるピークの位置を知る利点があるため、ブロック60において、方法35は、プロセスしたシグナルにおける推定ピークを予想ピークの位置にマッチさせる。その方法では、生物学的サンプルの組成物を示す実際のピークであるデータ内の各推定ピークの確率を決定することができる。各ピークの確率をブロック60で一旦決定すると、次いで、ブロック65において、方法35は、生物学的サンプルの組成物を統計学的に決定し、遺伝子型をコールする程度に確実性が高いならば決定する。

#### 【0305】

再びブロック40について言及すると、データは、試験装置、好ましくはマスマススペクトロメーターから受け取られる。特定の説明では、図26は、マスマススペクトロメーターからのデータの例を示す。当該マスマススペクトロメーターデータ70は、x軸71およびy軸72に沿って分布するデータポイントを一般に含んでいる。x軸71は、決定された粒子の質量を示し、その一方、y軸72は、当該粒子の数値濃度を示す。図26に示すように、質量分析法データ70は、x軸71のより重い端(heavier end)74のデータの方向へ

10

20

30

40

50

一般的に指数的に減衰する x 軸 7 3 の左端のデータに伴って、一般に指数的に減衰する。しかし、当該データの一般的な指数的表示は、生物学的サンプルの組成物を示さないが、合成エラーおよび特性をより反映させる。更に、上記および図 2 6 に示すように、相当なノイズが質量分析法 DNA データ 7 0 に存在する。

#### 【0306】

再びブロック 4 5 について言及すると、ブロック 4 0 で受けた未処理データをノイズ除去化する場合、ノイズ除去化プロセスをより詳細に記載する。図 2 5 に示したように、ノイズ除去化プロセスは、一般に、1) 未処理データにおいてウェーブレット変換を行い、未処理データをウェーブレットステージ係数に分解すること、2) ウェーブレット係数の最も高いステージからノイズプロファイルを作成すること、および 3) ウェーブレット変換における他のステージにスケール化ノイズプロファイルを適用することを必要とする。ノイズ除去プロセスの各ステップを更に以下に記載する。

#### 【0307】

ここで図 2 7 について言及すると、未処理質量分析法データのウェーブレット変換を一般的にダイアグラム化する。ウェーブレット変換技術を用い、その後、質量分析法データ 7 0 をステージに変換する。各ステージステップでは、当該データを High ステージおよび Low ステージにおいて示し、Low ステージは次のその後のステージへの入力として作用する。例えば、質量分析法データ 7 0 を、ステージ 0 ハイデータ 8 2 およびステージ 0 ローデータ 8 3 に変換する。次いで、ステージ 0 ローデータ 8 3 を次のレベルの変換への入力として用い、ステージ 1 ハイデータ 8 4 およびステージ 1 ローデータ 8 5 を得る。同様の方法で、ステージ 1 ローデータ 8 5 を、ステージ 2 ハイデータ 8 6 およびステージ 2 ローデータ 8 7 へ変換する入力として使用する。変換は、より有用な情報が更なるウェーブレット変換により得ることができなくなるまで、続ける。例えば、好ましい実施態様では、24 ポイントウェーブレットを用いる。より特に、Daubechies 2 4 として通常言及されるウェーブレットを用い、未処理データを分解する。しかし、他のウェーブレットがウェーブレット変換に使用できることが認識される。ウェーブレット変換の各ステージは前のステージのデータポイントの半分となるため、ウェーブレット変換は、ステージ n ローデータ 8 9 が約 5 0 ポイントとなるまで続けることができる。従って、ステージ n ハイ 8 8 は、約 1 0 0 データポイントを含む。好ましいウェーブレットは 2 4 ポイント長であるため、小さなデータまたは情報が、約 5 0 ポイントのデータセットにおいてウェーブレット変換続けることにより得ることができる。

#### 【0308】

図 2 8 は、ステージ 0 ハイデータ 9 5 の例を示す。ステージ 0 ハイデータ 9 5 は質量分析法データにおいて最も高い頻度を一般的に示すため、ステージ 0 ハイデータ 9 5 は質量分析法データにおける高頻度ノイズの量に密接に関連する。図 2 9 において、指数フィッティング式 (exponential fitting formula) をステージ 0 ハイデータ 9 5 に適用し、ステージ 0 ノイズプロファイル 9 7 を得る。特に、指数フィッティング式は、式  $A_0 + A_1 E^{-A_2 m}$  である。他の指数フィッティング式または他の型のカーブフィットも用い得ることが認識される。

#### 【0309】

ここで図 3 0 について言及すると、他の High ステージのノイズプロファイルを決定する。各ステージの後のデータポイントは、各ステージのノイズのレベルを示すようであるため、各ステージの後のデータポイントのみを用い、特定ステージのノイズコンテンツを示す標準偏差図を得る。より特に、各残りステージのノイズプロファイルの作成において、各ステージの最後 5 % のデータポイントのみを分析し、標準偏差数を決定する。他に多くのポイントまたは他の方法を用い、その標準偏差図を得ることが認識される。

#### 【0310】

各ステージの標準偏差数をステージ 0 ノイズプロファイル (指数曲線) 9 7 と共に用い、各ステージのスケール化ノイズプロファイルを得る。例えば、図 3 0 は、ステージ 1 ハイデータ 9 8 がエリア 9 9 により示される最後 5 % のデータポイントと共にステージ 1 ハイ

10

20

30

40

50

データ103を有することが示される。エリア99のポイントを評価し、ステージ1ハイデータ103のノイズコンテンツを示す標準偏差数を決定する。次いで、標準偏差数をステージ0ノイズプロファイル97と共に用いステージ1ノイズプロファイルを得る。

#### 【0311】

同様の方法で、ステージ2ハイ100は、エリア101により示される最後5%のポイントと共にステージ2ハイデータ104を有する。次いで、エリア101のデータポイントを用い標準偏差数を算出し、次いで、それを用い、ステージ0ノイズプロファイル97をスケール化し、ステージ2データのノイズプロファイルを得る。この同じプロセスを、ステージnハイ105により示されるように各ステージハイデータについて続ける。ステージnハイ105に関し、ステージnハイデータ108は、エリア106により示される最後5%のデータポイントを有する。エリア106のデータポイントを用い、ステージnの標準偏差数を決定する。次いで、ステージn標準偏差数をステージ0ノイズプロファイル97と共に用い、ステージnのノイズプロファイルを得る。従って、各ハイデータステージはノイズプロファイルを有する。

10

#### 【0312】

図31は、どのようにノイズプロファイルを各ステージのデータに適用するかを示す。一般にノイズプロファイルを用い、各ステージのデータに適用するスレシヨルド(threshold)を得る。ノイズプロファイルを既にスケール化し、各ステージのノイズコンテンツを調製するため、スレシヨルドの算出により、除かれるノイズの量を調節する調製が可能となる。スレシヨルド未満のウェーブレット係数は無視でき、その一方、スレシヨルドの上のもの

20

#### 【0313】

ウェーブレット変換の特性のため、ステージ0および1のような低いステージは、ステージ2またはステージnのような後のステージよりもよりノイズコンテンツを有している。実際、ステージnローデータは、殆どノイズを有してないようである。そのため、好ましい実施態様では、ノイズプロファイルを、より低いステージではより積極的に適用し、より後のステージではあまり積極的に適用しない。例えば、図31は、ステージ0ハイスレシヨルドがステージ0ノイズプロファイルを4つの因子に掛けることにより決定されることを示す。その方法で、ステージ0ハイデータ95内のデータポイントの有意な数はスレシヨルド未満となり、それゆえ、除かれる。ステージ1ハイスレシヨルド112は、ステージ1ハイデータのノイズプロファイルの2倍のセットであり、ステージ2ハイスレシヨルド114は、ステージ2ハイのノイズプロファイルと等しいセットである。そのため、このゲノム経過の後、ステージnハイスレシヨルド116は、 $(1/2^{n-2})$ に等しい因子による各ステージnハイのノイズプロファイルのスケールリングにより決定される。他の因子が各ステージのノイズプロファイルのスケールに適用されることが認識される。例えば、ノイズプロファイルを多かれ少なかれ積極的にスケールし、特異的組織特性またはサンプル組成物に適應させる。上記したように、ステージnローデータ118は僅かにノイズコンテンツを有するか全く有していないと想定すると、ステージnローデータは、適用されるノイズプロファイルを有してはいない。スケール化ノイズプロファイルを各ハイデータステージに適用した後、質量分析法データ70はノイズ除去され、容易に更にプロセッシングされる。ノイズ除去化シグナルのウェーブレット変換は、図31に示すように希薄データセット120を生ずる。

30

40

#### 【0314】

図25について再び言及すると、ブロック40で受ける質量分析法データをブロック45でノイズ除去し、ベースライン補正のためブロック50に移す。ベースライン補正を行う前に、ウェーブレット変換手順により導入されるアーテファクト(artifact)を好ましくは取り除く。ウェーブレット変換は、ウェーブレットのポイントを開始点として用いることに僅かに依存する変化を生ずる。例えば、好ましい実施態様は、24ポイントのDaubechies24ウェーブレットを用いる。ウェーブレットの0ポイントの変換の開始により、僅

50

かに異なる結果が、ウェーブレットのポイント1または2で開始する場合よりも得られる。そのため、ノイズ除去データは、各利用可能な開始点を用い変換し、その結果を平均し最終的なノイズ除去およびシフト化シグナルを決定する。例えば、図33は、ウェーブレット係数を24種の時間に適用し、次いで、その結果を平均し最終データセットを得る。他の技術を用いウェーブレットシフティングにより導入される僅かなエラーを適応させる。

#### 【0315】

式125は、図33で一般的に示される。シグナルを、一旦、ノイズ除去およびシフトすると、ノイズ除去およびシフト化シグナル130を図58に示されるように生ずる。図34は、ノイズ除去およびシフト化シグナル130からのウェーブレット係数135データセットの例を示す。

10

#### 【0316】

図36は、推定ピークエリア145、147および149は、ノイズ除去およびシフト化シグナル150に位置することを示している。推定ピークエリアを、シグナル150に沿って移動平均すること、および移動平均に関連するスレシヨルドを超えるシグナル150のセクションを同定することにより組織的に同定される。他の方法を用い、シグナル150における推定ピークエリアを同定し得る。

#### 【0317】

推定ピークエリア145、147および149をシグナル150から取り除き、図37に示すようにピークフリーシグナル155を作成する。ピークフリーシグナル155を更に分析し、残存最小値157を同定し、残存最小値157を接続し、ピークフリーシグナル155を得る。

20

#### 【0318】

図38は、ピークフリーシグナル155を用い図39に示すようにベースライン170を得るプロセスを示す。ブロック162に示すように、ウェーブレット変換をピークフリーシグナル155で行う。ウェーブレット変換からの全ステージを、nLowステージを除くブロック164で排除する。nLowステージは、一般的に、ピークフリーシグナル155の最も低い頻度コンポーネントを示し、それにより、一般的に、システム指数特性を示す。ブロック166は、シグナルはnロー係数から再構成され、ベースラインシグナル170をブロック168で得る。

30

#### 【0319】

図39は、校正ベースライン170に隣接して位置するノイズ除去およびシフト化データシグナル172を示す。ベースライン補正170は、ノイズ除去およびシフト化シグナル172から差し引かれ、図40に示すように適用されたベースライン補正を有するシグナル175を得る。そのノイズ除去、シフト化および校正化シグナルは、殆どの同定目的に重要であるが、シグナル175中の推定ピークは、生物学的サンプルのDNA組成物をコールする有意な正確性または確実性を伴って同定できない。

#### 【0320】

再び図25について言及すると、ベースライン補正50のデータをブロック55で圧縮し、好ましい実施態様に用いる圧縮技術を図41に詳述する。図41では、ベースライン補正化データ内のデータは、連関データ値184を有するx軸ポイント183と共にアレイフォーマット182に存在する。x軸は、ゼロではないウェーブレット係数による指標となり、連関する値は、ウェーブレット係数の値である。表182の解説データ例では、最大値184は、1000となることを示す。質量分析法データに特に利点の圧縮技術を示すが、他の圧縮技術も使用し得ることが認識される。好ましくないが、当該データはまた、圧縮なしに保存され得る。

40

#### 【0321】

好ましい実施態様によるデータ圧縮において、中間フォーマット186を得る。中間フォーマット186は、一般的に、自然数部分188および少数部分190を有する実数を含む。自然数部分は、x軸ポイント183であり、その一方、少数部分は、最大値で割ら

50

れる値データ184である。例えば、データ182では、データ値“25”は、x軸ポイント“100”で示される。このデータポイントの中間値は“100.025”となり得る。

#### 【0322】

中間圧縮データ186から、最終圧縮データ195を得る。中間データファイルの最初のポイントは、圧縮データの開始点となる。その後、圧縮データ195の各データポイントは、以下のように算出される：自然数部分(小数点の左)は、現在の自然数と過去の自然数との間の違いにより置換えられる。残り部分(小数点の右)は未処理のままである。例えば、圧縮データ195の開始点は、“100.025”である中間データポイントと同じであると示されている。最初の中間データポイント“100.025”と第二の中間データポイント“150.220”との間は、“50.220”である。そのため、“50.220”は圧縮データ195の第二のポイントとなる。同様の方法で、第二の中間データポイントは、“150.220”であり、第三の中間データポイントは“500.0001”である。そのため、第三の圧縮データは、“350.000”となる。決定された圧縮データポイントの算出は、データポイントの全アレイが実数の単一アレイに変換されるまで続ける。

10

#### 【0323】

図42は、一般的に、質量分析法データを圧縮する方法を記載し、それは、ブロック201のデータファイルがブロック202で係数のアレイとして表されることを示す。当該データ開始点および最大は、ブロック203に示すように決定され、中間実数は、上記のようにブロック204で算出される。得られた中間データポイントを用い、圧縮データをブロック205で得る。記載された圧縮方法は、質量分析法装置からのプロセス化データセットのような圧縮データセットに高い利点および効果がある。当該方法は、多数使用し、x軸データにおいて部分的なギャップを有するようにプロセスする、質量分析法データのようなデータに特に有用である。従って、プロセス化質量分析法データのx-yデータアレイは、10xまたはそれを超える効果的な圧縮率で保存され得る。圧縮技術が質量分析法データに適用されるが、当該方法は、他のデータセットに有利に適用されることが認識される。

20

#### 【0324】

再び図25について言及すると、ピークの高さは、ブロック60で決定される。ピークの高さを決定する最初のステップを図43に示し、この場合、シグナル210は、左または右にシフトし、それは予測ピークの位置に相当する。質量分析法データを生ずる前に生物学的サンプルにおいて可能性ある組成物のセットが判るため、予測ピークの可能な位置は既に判る。これらのあり得るピークを、予測ピーク212、214および216のような予測ピークとして称する。校正または試験装置データ中の他のエラーのため、全シグナルは、実際の位置から左または右にシフトし得、そのため、推定ピーク218、222および224のようなシグナルに位置する推定ピークを予測ピーク212、214および216とそれぞれ比較し得る。次いで、全シグナルを、推定ピークが予測ピークにより近づいて位置するようにシフトする。

30

#### 【0325】

推定ピークが一旦シフトし、予測ピークとマッチすると、最も強い推定ピークが図44で同定される。好ましい実施態様では、最も強いピークを、ピークの全体の高さおよびピークの真下のエリアの分析の組合せとして算出する。例えば、幅の広いピークでなく適度な高さのピークは、極端に幅が狭く非常に高いよりも強力となり得る。推定ピーク225のような同定される最も強力な推定ピークを用い、ガウス228カーブをピーク225にフィットさせる。一旦、ガウスをフィットさせると、ガウスの幅(W)を測定し、将来的に算出のためのピーク幅として使用する。

40

#### 【0326】

一般に上記のような取り組みとして、ノイズ除去化、シフト化、およびベースライン補正化シグナルは、生物学的サンプルのDNA組成物を確実にコールするほどには充分には

50

プロセスされない。例えば、ベースラインは、一般に取り除かれるが、なお、現在の残留ベースライン効果が存在する。そのため、これら残留ベースライン効果を取り除き、同定の正確性および確実性を増大させる。

#### 【0327】

残留ベースライン効果を取り除くため、図45は、推定ピーク218、222および224をベースライン補正シグナルから取り除くことを示す。当該ピークを、推定ピーク218、222および224、それぞれのセンターライン230、232、および234を同定することにより、および同定センターラインの左および右にエリアを取り除くことにより、取り除く。各推定ピークの場合、ガウスの2倍の幅(W)に等しいエリアをセンターラインの左から取り除き、その一方、50ダルトンに等しいエリアをセンターラインの右から取り除く。50ダルトンを示すエリアを適合させ、実際のピークに付随し得る塩付加物の効果を十分に適当に取り除くことが発見された。その付加物は、実際のピークの右に現れ、マススペクトルの取得を含む化学からの自然効果を有する。50ダルトン緩衝液を選択するけれども、他の範囲または方法が使用され付加物の効果を減少させるか取り除き得ることが認識される。

10

#### 【0328】

当該ピークが取り除かれ、シグナル245の作成に関連する最小247により図46で示されるように位置する最小247が残存する。四次多項式をシグナル245に適用し、図47に示すように残留ベースライン250を得る。残留ベースライン250を、シグナル225から差し引き、図48に示すように最終シグナル255を得る。残留ベースラインは、シグナル245への四次フィットの結果であるが、他の技術を用い残留ベースラインをスムーズにするか、フィットさせ得る。

20

#### 【0329】

図49に示すようにピークの高さを決定するため、ガウス266、268および270のようなガウスを、それぞれピーク260、262、および264のような各ピークにフィットさせる。従って、ガウスの高さは、高さ272、274、および276として決定する。一旦、各ガウスピークの高さを決定すると、次いで、生物学的化合物35を同定する方法は、図25に示すように遺伝子型決定フェーズ65へと移ることができる。

#### 【0330】

各推定ピークが実際のピークであるという確実性の現れ(indication)は、各推定ピークについてノイズ対シグナル比率(signal to noise ratio)を算出することにより認識し得る。従って、ノイズに対する強力なシグナルの比を有する推定ピークは、ノイズに対するより低いシグナルの比を有する推定ピークよりもより実際のピークとなるようである。上記および図50に示したように、高さ272、274、および276のような各ピークの高さを各ピークについて決定し、この場合、高さが各ピークのシグナル強度インディケータとなる。ノイズプロファイル97のようなノイズプロファイルを、同定ピークを通してノイズプロファイル280に外挿する。各ピークのセンターラインにおいて、ノイズ値をノイズ値282、283、および284のように決定する。得られるシグナル値およびノイズ値を用い、ノイズ対シグナル比率を各ピークで算出することができる。例えば、図50の最初のピークについてのノイズ対シグナル比率は、ノイズ値282で割ったシグナル値272として算出し、同様の方法で、図50の中央ピークのノイズ対シグナル比率は、ノイズ値283で割ったシグナル274として決定し得る。

30

40

#### 【0331】

ノイズ対シグナル比率は、実際ピークの存在の有用な一般的なインディケータであるが、更なるプロセッシングにより、サンプルが同定され得る確実性を増大することが発見された。例えば、好ましい実施態様において各ピークに関するノイズ対シグナル比率は、好ましくは、ガウスと各推定ピークとの間でのフィットの長所により調節される。それは、通常の分布に一般的に従う方法でサンプル物質を検出するマススペクトロメーターの特性である。従って、少ない通常分布を有するシグナルよりも、よりよい確実性が、ガウス形を有する推定シグナルを付随する。非ガウス形を有することから生ずるエラーを“残留

50

エラー”と称する。

【0332】

図51について言及すると、残留エラーは、データシグナルにおけるガウス293と推定ピーク290との間で二乗平均を算出することにより、算出する。当該算出は、ガウスのセンターラインの何れかの側の1つの幅内のデータで行う。残留エラーは、：

【数1】

$$\sqrt{(G-R)^2/N}$$

として算出される。この場合、Gはガウスシグナル値であり、Rは推定ピーク値であり、そしてNは-Wから+Wまでのポイントの数である。算出残留エラーを用い、下記のような、調節されたノイズ対シグナル比率(adjusted signal to noise ratio)を得る。 10

【0333】

調節されたシグナルノイズ比率は、式

【化9】

$$(S/N) * \text{EXP}^{(-1/R)}$$

を用いる各推定ピークについて算出する。この場合、S/Nはノイズ対シグナル比率であり、Rは上記で決定した残留エラーである。好ましい実施態様で、各ピークの残留エラーを用いる調節されたノイズ対シグナル比率を算出するが、他の技術を用い、ガウスと実際のシグナルとの間でのフィットの長所が説明されることが認識される。

【0334】

図52について言及すると、推定ピークが実際のピークとなる確率を決定する。ピークの確率の決定において、調節されたノイズ対シグナル比率がx軸であり、確率がy軸である、確率プロファイル300を得る。確率は、確率0%と、1として示される確率100%との範囲にあることが必要である。一般的に、調節されたノイズ対シグナル比率が高くなればなるほど、推定ピークが実際のピークである確実性が高くなる。 20

【0335】

調節されたノイズ対シグナル比率の幾つかの標的値において、確率が100%であるとは、推定ピークが実際のピークであり、確実に生物学的サンプルのDNA組成物同定に用い得ることであると判る。しかし、確率が100%であるとみなせる場合、調節されたノイズ対シグナル比率の標的値は、適用特異的基準に従うセットである可変パラメーターとなる。例えば、標的のノイズ対シグナル比率は、試行実験、サンプル特性およびシステム全体の許容可能エラー耐性に依りて、調節される。より特異的に、エラーを許容できない保存アプローチを必要とする状態では、標的の調節されたノイズ対シグナル比率がセットされ得、例えば実施例10、より高くなり得る。従って、100%確率が、調節されたノイズ対シグナル比率が10またはそれを超えないならば、ピークに割り当てられない。 30

【0336】

他の状況では、サンプルデータが明白であるか、またはエラーの危険性を減少し得るため、より積極的なアプローチを取り得る。その状況で、当該システムをセットすると、5またはそれ以上の標的のノイズ対シグナル比率が100%の確率とみなされ得る。もちろん、エラーの極度の危険性がみなされるとき、中間のノイズ対シグナル比率標的図を、7のように選択する。標的の調節されたノイズ対シグナル比率を、一旦当該方法にセットすると、次いで、任意の調節されたノイズ対シグナル比率に関し、確率により、推定ピークが実際のピークであると決定され得る。 40

【0337】

同定試験、特にDNA増幅により調製されるサンプルの質量分析法試験の実行を含む化学により、最も高いピークのシグナル強度と二番目(または三番目など)に高いピークのシグナル強度の間の対立遺伝子比率は、予想される比率内に含まれる。対立遺伝子比率が通常のガイドラインの外側に含まれるならば、好ましい実施態様は、当該確率に対し対立遺伝子比のペナルティーを強要することとなる。例えば、図53は対立遺伝子ペナルティー315を示し、それは、最も高いピークのシグナル強度で割った二番目に高いピークのシ 50

グナル強度の間の割合である x 軸を有する。y 軸 3 1 9 は、決定された対立遺伝子比に依存した 0 から 1 の間のペナルティーとみなせる。好ましい実施態様では、30%を超える対立遺伝子比率が予測範囲内に含まれ、それにより、ペナルティーは適用されないとみなせる。10%と30%との間の比において、ペナルティーは、10%未満の対立遺伝子比率において二番目に高いピークが現実でないこととみなせるまで、直線的に増大する。10%と30%との間の対立遺伝子比率の場合、対立遺伝子ペナルティーチャート 3 1 5 を用いペナルティー 3 1 9 を決定し、図 5 2 で決定したピーク確率によりマルチプレクシングし、最終ピークの確率を決定する。好ましい実施態様を、対立遺伝子比率ペナルティーに組み込み、可能な化学エラーを考慮するが、他の技術を使用し得ることが認識される。同様の処置を他のピークに適用する。

10

## 【0338】

決定される各ピークのピーク確率を用い、種々組成物コンポーネントの統計学的確率を決定し得る。実施例のように、2つのピーク、ピーク G、ピーク C および GG、CC および GC の組合せのそれぞれ 3 つの可能な組合せの確率を決定するためである。図 5 4 は、最もあり得るピーク 3 2 5 を測定すると、最終ピーク確率 90% を有していたことを示す。ピーク 3 2 5 は、生物学的サンプル中に G コンポーネントを示すように位置する。従って、G が生物学的サンプル中に存在する確率 90% が維持され得る。図 5 4 に示す例においても、二番目に高い確率はピーク 3 3 0 であり、20%のピーク確率を有する。ピーク 3 3 0 は C 組成物を付随する位置にある。従って、C が生物学的サンプル中に存在する確率 20% が維持され得る。

20

## 【0339】

開始点として G 存在の確率 (90%) および C 存在の確率 (20%) により、G および C 存在の組合せの確率を算出し得る。例えば、図 5 4 は、GG 存在の確率 3 2 9 が 72% として算出されることを示す。これは、GG の確率が、C が存在しない確率 (100% - 20%) を掛けた G 存在の確率 (90%) に等しくなると、算出される。G 存在の確率が 90% であり、C が存在しない確率が 80% ならば、GG の確率は 72% である。

## 【0340】

同様の方法で、CC が存在する確率は、G が存在しない確率 (100% - 90%) を掛けた C 存在の確率 (20%) に等しくなる。図 5 4 に示すように、C 存在の確率が 20% である一方、G が存在しない確率は 10% であり、そのため、CC の確率は 2% のみである。最終的に、GC 存在の確率は、C 存在の確率 (20%) を掛けた G 存在の確率 (90%) に等しくなる。G 存在の確率が 90% であり、C 存在の確率が 20% であるならば、GC 存在の確率は 18% である。略式の形式では、次いで、生物学的サンプルの組成物の確率は、

30

GG の確率 : 72%

GC の確率 : 18%、および

CC の確率 : 2%

である。

## 【0341】

可能性ある組合せのそれぞれの確率を一旦決定すると、図 5 5 は、遺伝子型をコールする重要な確実性が存在するかどうかの決定に使用される。図 5 5 は、最も高い組合せ確率と二番目に高い組合せ確率との比率である x 軸 3 3 7 を有するコールチャート 3 3 5 を示す。y 軸 3 3 9 は、当該比率が、遺伝子型のコールを正しいとするほどに十分に高いかどうかを単に示す。当該比率の値は M 3 4 0 で示し得る。M 値は、試行データ、サンプル組成物およびエラー許容能に依存するセットである。例えば、M 値は、値 4 のように比較的高いセットとなり得、そのため、最も高い確率は、確実性により遺伝子型のコールを確立する前に二番目に高い確率よりも少なくとも約 4 倍大きくなければならない。しかし、特定レベルのエラーが許容されるならば、M 値は 3 のようなより積極的な値のセットであり得、そのため、最も高い確率と二番目に高い確率との間の比率は、3 のみの比またはそれ以上となることが必要となる。もちろん、適度の危険性を許容できるとき、適度の値が M に関し選択され得る。GG の確率が 72% であり、GC の確率が 18% である図 5 4 の例

40

50

を用い、72%と18%との間の比率は、4.0であり、そのため、Mのセットは3、3.5、または4の何れかとなり、そのシステムはGGとして遺伝子型をコールし得る。好ましい実施態様が、遺伝子型が確実にコールするかどうかを決定する2つの最も高いピーク確率間の比を用いるが、他の方法が置換され得ることが認識される。上記技術が、確率の算出および2を超えるピークの組合せを含む遺伝子型(またはより一般的なDNAパターン)の決定に用いられ得る。

#### 【0342】

ここで図56について言及すると、フローチャートを示しており、上記の遺伝子型の統計学的コールのプロセスを一般的に定義している。図56では、ブロック402は、各ピークの高さを決定し、ブロック404におけるノイズプロファイルが各ピークについて外挿されることを示す。当該シグナルを、ブロック406の各ピークの高さから決定し、各ピークのノイズを、ブロック408のノイズプロファイルを用い決定する。ブロック410では、ノイズ対シグナル比率を各ピークについて算出する。非ガウスピーク形を説明するため、残留エラーをブロック412で決定し、調節されたノイズ対シグナル比率をブロック404で算出する。ブロック416は、確率プロファイルを作成すること、この場合、存在する各ピークの確率がブロック418に見られることを示す。対立遺伝子ペナルティをブロック420に適用し得、この場合、対立遺伝子ペナルティを、ブロック442の調節されたピーク確率に適用する。コンポーネントの各組み合わせの確率をブロック424で算出し、この場合、2つの最も高い確率の間の比をブロック426で決定する。次いで、確率の比率がスレシールド値を超えるならば、次いで、遺伝子型をブロック428でコールする。

#### 【0343】

本発明の他の実施態様では、コンピューティング装置20(図24)は、推定SNPを含むデータピークを同定することにより“標準”遺伝子型決定をサポートする。例えば、標準遺伝子型決定を用い、この場合、サンプルについての情報が十分に知られておらず、予想ピークの位置を決定し、それに対し、上記のような対立遺伝子ペナルティが確実性をもって算出される。これにより、標的DNA分子をフラグメント化する任意のアッセイにより生ずるデータから推定SNPが含まれるピークの同定に関しコンピューティング装置を用い得る。その標準遺伝子型決定の場合、データスペクトル中の他のピークの典型的エリアから有意に誘導されるデータ曲線の下エリアに付随するピークが同定され、その相当する質量(x軸に沿った位置)を決定する。

#### 【0344】

より特に、データ中の他のピークの平均エリアから有意に誘導されるピークを同定し、データピーク間の予想対立遺伝子比率を、データピークの下エリアの比率について定義する。理論的には、各遺伝的配座が同じモル濃度のアナライトを有する場合、各相当するピークの下エリアは同じであり、そのため、任意の2つのピーク間に1.0比率のピークエリアを生ずる。本発明により、データ中の他のピークに比例してより小さくなる比を有するピークは、ピークとして認識されない。より特に、ピークエリアの見た目の値に比例して30%未満となるエリア比率を有するピークを対立遺伝子ペナルティとする。残存ピークの質量(当該データのx軸に沿った位置)をオリゴヌクレオチド標準に基づき決定する。

#### 【0345】

図57は、標準遺伝子型決定を行うとき、コンピューティング装置20(図24)によるプロセッシングのフローダイアグラム表示を示す。フローダイアグラムボックス番号502により示される、最初の操作において、コンピューティング装置は、マススペクトロメーターからデータを受ける。次に、データサンプル中の各推定ピークの高さを、ブロック504により示されるように、決定した。マススペクトロメーターデータ内の各ピークの高さを決定した後、ノイズ除去プロセス505を行い、ノイズプロファイルの外挿を開始し(ブロック506)、その後、各ピークのノイズを発見し(ブロック508)、そして各データサンプル(ブロック510)のノイズ対シグナル比率を算出する(ブロック510)。各

これらの操作は、図 2 5 のノイズ除去操作 4 5 のため上記により行い得る。他の適当なノイズ除去操作が当業者により思いつく。

#### 【0346】

次の操作は、各データポイントに関連する残留エラーを見つけることである。これは、図 5 7 のブロック 5 1 2 により表されている。次のステップ、ブロック 5 1 4 は、各同定ピークの調節されたノイズ対シグナル比率を算出することを含む。確率プロファイルは、次(ブロック 5 1 6)で作成され、その後、ブロック 5 1 8 でピーク確率を決定する。好ましい実施態様では、ブロック 5 0 2 からブロック 5 1 8 までを含む図 5 7 のノイズ除去操作は、図 5 6 と関連する相当する、ブロック 4 0 2 からブロック 4 1 8、それぞれのための上記操作を含む。

10

#### 【0347】

標準遺伝子型決定プロセッシングのための次の動作で、ブロック 5 2 4 により示される各ピークの対立遺伝子ペナルティーを決定する。上記のように、図 5 7 の標準遺伝子決定プロセッシングは、ピークの下エリアを比較することにより対立遺伝子ペナルティーを決定する。そのため、上記図 5 3 のような対立遺伝子ペナルティーを決定する比較シグナル強度比以外では、標準プロセッシングは、各同定ピークの下エリアを決定し、これらエリアの比率を比較する。各ピークの下エリアの決定は、実験データの曲線の下エリアを計算するための通常多くの分析技術を用い計算し得る。

#### 【0348】

そのため、対立遺伝子ペナルティーを、0.30(30%)を超える予想平均エリア値に比例するピークエリアを有するピークに割り当てられるペナルティーはないことを示す図 5 8 により割り当てる。対立遺伝子ペナルティーを、図 5 2 に記載のようなプロセスにより決定され得る、ピーク確率値に適用する。ピークを 30%未満の比とする対立遺伝子ペナルティーより、ピークが更なる測定およびプロセッシングから取り除かれるということが、図 5 8 から明白となる。しかし、他のペナルティースキームを、当業者が決定するように、プロセスされるデータに関する知識により、課す。

20

#### 【0349】

対立遺伝子ペナルティーを決定し適用した後、標準遺伝子型プロセッシングは、残存推定ピークとオリゴヌクレオチド標準の位置を比較し、ブロック 5 2 4 のプロセッシングにおいて相当する質量を決定する。標準遺伝子型データの場合、図 3 3 のブロック 4 2 4、4 2 6、および 4 2 8 に相当する操作を行うよりもむしろ、ブロック 5 2 4 のプロセッシングを行い、質量および遺伝子型を決定する。その比較を行い、質量を測定する技術は当業者に既知である。

30

#### 【0350】

他の実施態様では、コンピューティング装置 2 0 (図 2 4)により、アッセイにおいて生ずるフラグメントのセンスおよびアンチセンス鎖の質量(当該データのx軸に沿った位置)の検出および決定が可能となる。望ましいならば、コンピューティング装置はまた、標準遺伝子型プロセッシングのための上記した同様技術を用い、各センスおよびアンチセンス鎖の量(各ピークの下エリア)を検出し、決定し得る。次いで、各型の鎖を生ずるデータを合わせ、データ冗長を達成し、それにより、決定化遺伝子型の確実性レベルを増大する。この技術により、他の診断方法からデータ中にしばしば観察されるプライマーピークを事前に除去し、それにより、より高いレベルのマルチプレクシングが可能となる。加えて、定量化を、プールした実験で用いるとき、測定ピークエリアの比率は、データ冗長のため、ピーク同定技術よりもより確実に算出される。

40

#### 【0351】

図 2 3 は、センスおよびアンチセンスプロセッシングを行うコンピューティング装置 2 0 により実行するプロセッシングを示すフローダイアグラムである。フローダイアグラムボックス番号 6 0 2 により示される最初の操作で、コンピューティング装置は、マススペクトロメーターからデータを受け取る。このデータは、アッセイフラグメントのセンス鎖およびアンチセンス鎖のデータを含む。次に、データサンプル中の各推定ピークの高さを

50

ブロック 604 で示すように決定する。マススペクトロメーターデータ中の各ピークの高さを決定した後、ノイズ除去プロセス 605 を行い、ノイズプロファイルに外挿する操作(ブロック 606)を開始し、その後、各ピークのノイズを発見し(ブロック 608)、そして各データサンプルのノイズ対シグナル比率を算出する(ブロック 610)。各これらの操作は、図 25 のノイズ除去操作 45 のため上記により行い得る。他の適当なノイズ除去操作が当業者により思いつく。次の操作は、各データポイントに付随する残留エラーを発見することである。これは、図 36 のブロック 612 により示す。

#### 【0352】

センス鎖およびアンチセンス鎖のデータのための残留エラーを行った後、遺伝子型を同定するプロセッシングは、センス鎖で行い、また、アンチセンス鎖で行う。そのため、図 23 は、プロセッシングにはセンス鎖プロセッシング(ブロック 630)およびアンチセンス鎖プロセッシング(ブロック 640)が含まれることを示す。各ブロック 630、640 には、図 56 のブロック 414 から 426 と関連する上記するような、ノイズ対シグナル比率を調節すること、確率プロファイルを作成すること、対立遺伝子ペナルティーを決定すること、対立遺伝子ペナルティーによりピーク確率を調節すること、遺伝子型決定の確率を算出すること、および遺伝子型確率比を試験すること、に相当するプロセッシングが含まれる。各ブロック 630、640 のプロセッシングには、望ましいならば、図 57 と関連する上記のような標準プロセッシング操作が含まれる。標準プロセッシングは、図 56 のプロセッシング操作に代えてまたはそれに加えて、含まれ得る。

#### 【0353】

遺伝子型確率プロセッシングを行った後、センス鎖およびアンチセンス鎖プロセッシングからのデータを合わせ、予測データベース値と比較すると、センス鎖とアンチセンス鎖との間のデータ冗長の利点を得る。当業者ならば、技術により、アッセイフラグメントのセンス鎖とアンチセンス鎖との間の既知のデータ冗長が利点となることを理解し得る。このプロセッシングをブロック 650 に示している。二本鎖からのデータをプロセッシングのために合わせた後、遺伝子型プロセッシングを行い(ブロック 660)、遺伝子型を同定する。

#### 【0354】

修飾が当業者に明白であるため、本発明が添付の請求の範囲にのみによって限定されることを意図する。

#### 【図面の簡単な説明】

#### 【0355】

【図 1 - A】図 1 は、例示的なサンプルバンクを示す。図 1 - A は、性別および人種の関数としてのサンプルを示す。

【図 1 - B】図 1 は、例示的なサンプルバンクを示す。図 1 - B は、年齢の関数として白人を示す。

【図 1 - C】図 1 は、例示的なサンプルバンクを示す。図 1 - C は、年齢の関数としてヒスパニックを示す。

【図 2 - A】図 2 - A は、リポタンパク質リパーゼ遺伝子の 291S 対立遺伝子の年齢および性別分布を示し、全部で 436 人の男性および 589 人の女性を調査した。

【図 2 - B】図 2 - B は、436 人の男性について年齢の分布を示す。

【図 2 - C】図 2 - C は、リポタンパク質リパーゼ遺伝子の 291S 対立遺伝子の年齢および性別分布を示し、全部で 436 人の男性および 589 人の女性を調査した。

【図 3】図 3 は、集団に基づくサンプルバンキングのための例示的な質問票である。

【図 4】図 4 は、血液サンプルコンポーネントのプロセッシングおよびトラッキングを示す。

【図 5】図 5 は、「病気」対立遺伝子および「健康体」対立遺伝子の対立遺伝子頻度を年齢の関数として示す。健康体対立遺伝子の相対的頻度が年齢の増加とともに集団中で増加することが注目される。

【図 6】図 6 は、A p o E 遺伝子型の年齢依存性分布を示す (Schaechter et al. (1994))

10

20

30

40

50

Nature Genetics 6:29-32参照)。

【図7-A】図7-Aは、データベース中の白人集団中のp53(腫瘍抑制因子)コードン72の年齢依存性および遺伝子型頻度を示す。\*R72および\*P72はデータベース集団中の対立遺伝子の頻度を表す。R72、R72P、およびP72は集団中の個体の遺伝子型を表す。ホモ接合体P72対立遺伝子の頻度は、年齢とともに6.7%から3.7%に低下する。

【図7-B】図7-Bは、データベース中の白人集団中のp53(腫瘍抑制因子)コードン72の年齢依存性および遺伝子型頻度を示す。\*R72および\*P72はデータベース集団中の対立遺伝子の頻度を表す。R72、R72P、およびP72は集団中の個体の遺伝子型を表す。ホモ接合体P72対立遺伝子の頻度は、年齢とともに6.7%から3.7%に低下する。 10

【図7-C】図7-Cは、データベース中の白人集団中のp53(腫瘍抑制因子)コードン72の年齢依存性および遺伝子型頻度を示す。\*R72および\*P72はデータベース集団中の対立遺伝子の頻度を表す。R72、R72P、およびP72は集団中の個体の遺伝子型を表す。ホモ接合体P72対立遺伝子の頻度は、年齢とともに6.7%から3.7%に低下する。

【図7-D】図7-Dは、データベース中の白人集団中のp53(腫瘍抑制因子)コードン72の年齢依存性および遺伝子型頻度を示す。\*R72および\*P72はデータベース集団中の対立遺伝子の頻度を表す。R72、R72P、およびP72は集団中の個体の遺伝子型を表す。ホモ接合体P72対立遺伝子の頻度は、年齢とともに6.7%から3.7%に低下する。 20

【図8】図8は、年齢の関数として、p21 S31R対立遺伝子の対立遺伝子および遺伝子型頻度を示す。

【図9】図9は、個体サンプルに対してプールされたFVII対立遺伝子353Qの頻度を示す。

【図10】図10は、個体サンプルに対してプールされた、CETP(コレステロールエステル輸送タンパク質)対立遺伝子の頻度を示す。

【図11】図11は、個体サンプルに対してプールされた、プラスミノーゲンアクチベーターインヒビター-1(PAI-1)5Gの頻度を示す。

【図12】図12は、PAI-1対立遺伝子のサンプルのマススペクトルおよび人種分布を示す。 30

【図13】図13は、CETP405対立遺伝子のサンプルのマススペクトルおよび人種分布を示す。

【図14】図14は、第VII因子353対立遺伝子のサンプルのマススペクトルおよび人種分布を示す。

【図15】図15は、プールされたDNAサンプルを使用する、PAI-1、CETPおよび第VII因子の人種分布を示す。

【図16】図16は、p53-Rb経路およびその経路の種々の因子の間の関連を示す。

【図17】図17は、ここで記載したプロセスおよびデータベースを提供する、コンピューター構築性のブロックダイアグラムであり、これは、ここで提供するデータベースを蓄積し、そして分類するための、そしてここで提供する方法を実行する、典型的なコンピューターシステムを示す。 40

【図18】図18は、多型遺伝的マーカーを同定するためのデータベースへのアクセスを維持し、そして提供するための、図17に示されるコンピューターを使用して実行される、プロセッシングステップを示すフローダイアグラムを示す。

【図19】図19は、AKAP10-1遺伝子座について、年齢および性別で層化された、白人集団での対立遺伝子および遺伝子型分布を示すヒストグラムである。明緑色のバーは、40歳より若齢の個体での頻度を示す。暗緑色のバーは、60歳より高齢の個体での頻度を示す。

【図20】図20は、AKAP10-5遺伝子座について、年齢および性別で層化された 50

、白人集団の対立遺伝子および遺伝子型分布を示すヒストグラムである。明緑色のバーは、40歳より若齢の個体の頻度を示す。暗緑色のバーは、60歳より高齢の個体の頻度を示す。

【図21】図21は、h - m s r - A 遺伝子座について、年齢および性別で層化された白人集団の、対立遺伝子および遺伝子型分布を示すヒストグラムである。明緑色のバーは、40歳より若齢の個体での頻度を示す；暗緑色のバーは、60歳より高齢の個体での頻度を示す。

【図22 - A】図22 - Aは、健康体のデータベースのために使用する、サンプルデータ収集質問票である。

【図22 - B】図22 - Bは、健康体のデータベースのために使用する、サンプルデータ収集質問票である。 10

【図22 - C】図22 - Cは、健康体のデータベースのために使用する、サンプルデータ収集質問票である。

【図22 - D】図22 - Dは、健康体のデータベースのために使用する、サンプルデータ収集質問票である。

【図23】図23は、アクセイフラグメントからのセンス鎖およびアンチセンス鎖の遺伝子型決定をおこなうときに、図24のコンピューティング装置によって実行されるプロセッシングを示すフローチャートである。

【図24】図24は、本発明に従うシステムを示すブロックダイヤグラムである。

【図25】図25は、本発明に従う、生物学的サンプルを同定する方法のフローチャートである。 20

【図26】図26は、質量分析計からのデータをグラフに表したものである。

【図27】図27は、質量分析法データのウェーブレット変換のダイヤグラムである。

【図28】図28は、ウェーブレットステージ0 h i データをグラフに表したものである。

【図29】図29は、ステージ0 ノイズプロファイルをグラフに表したものである。

【図30】図30は、ステージノイズ標準偏差の作成をグラフに表したものである。

【図31】図31は、データステージのスレシヨルド (threshold) の適用をグラフに表したものである。

【図32】図32は、希薄データセットをグラフに表したものである。 30

【図33】図33は、シグナルをシフトさせる式である。

【図34】図34は、ノイズを除去し、そしてシフトしたシグナルの、ウェーブレット変換をグラフに表したものである。

【図35】図35は、ノイズを除去し、そしてシフトしたシグナルをグラフに表したものである。

【図36】図36は、ピークセクションの除去をグラフに表したものである。

【図37】図37は、ピークフリーシグナルの作成をグラフに表したものである。

【図38】図38は、ベースライン補正を作成する方法のブロックダイヤグラムである。

【図39】図39は、ベースラインおよびシグナルをグラフに表したものである。

【図40】図40は、ベースラインを除去したシグナルをグラフに表したものである。 40

【図41】図41は、圧縮したデータを示す表である。

【図42】図42は、データを圧縮する方法のフローチャートである。

【図43】図43は、マスシフティングをグラフに表したものである。

【図44】図44は、ピーク幅の決定をグラフに表したものである。

【図45】図45は、ピークの除去をグラフに表したものである。

【図46】図46は、ピークを除去したシグナルをグラフに表したものである。

【図47】図47は、残留ベースラインをグラフに表したものである。

【図48】図48は、残留ベースラインが除去されたシグナルをグラフに表したものである。

【図49】図49は、ピーク高の決定をグラフに表したものである。 50

【図50】図50は、各ピークについて、シグナル - 対 - ノイズ決定をグラフに表したものである。

【図51】図51は、各ピークについて、残留エラーの決定をグラフに表したものである。

【図52】図52は、ピーク確率をグラフに表したものである。

【図53】図53は、ピーク確率に対する対立遺伝子比率の適用をグラフに表したものである。

【図54】図54は、ピーク確率の決定をグラフに表したものである。

【図55】図55は、遺伝子型のコールをグラフに表したものである。

【図56】図56は、遺伝子型をコールするための統計的処理を示すフローチャートである。

【図57】図57は、標準を欠く遺伝子型決定をおこなうとき、図1のコンピューティング装置によって実行されるプロセッシングを示すフローチャートである。

【図58】図58は、標準を欠く遺伝子型プロセッシングのための、ピーク確率に対する対立遺伝子比率の適用をグラフに表したものである。

【図1 - A】

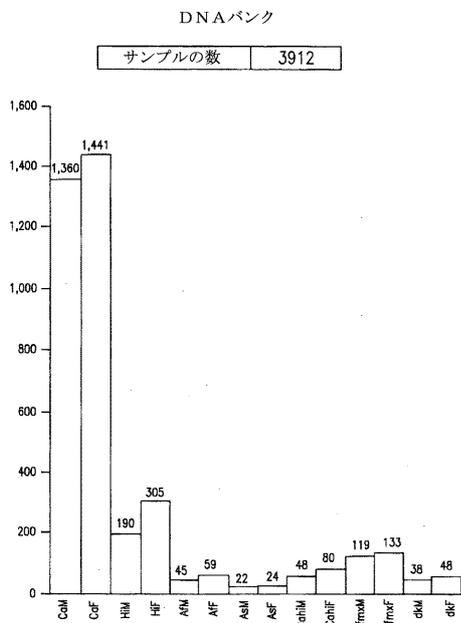


FIG. 1A

【図1 - B】

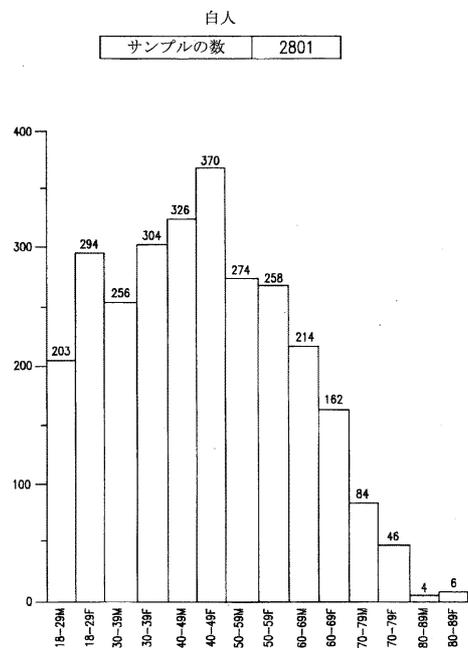
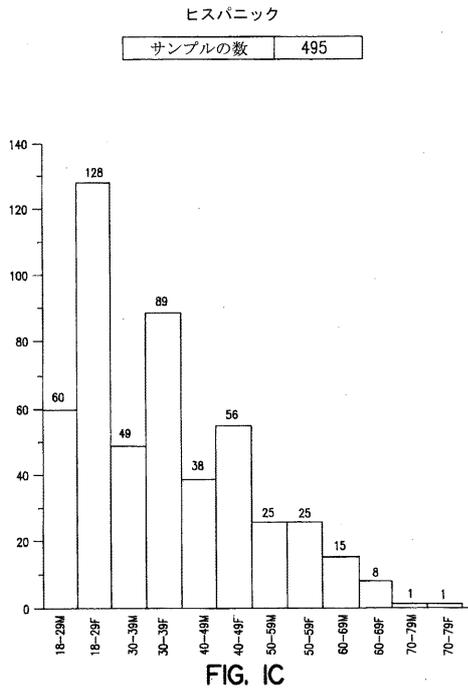
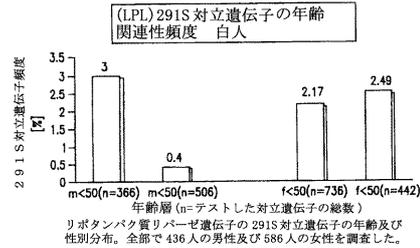


FIG. 1B

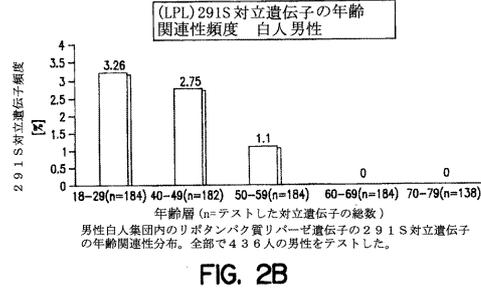
【 図 1 - C 】



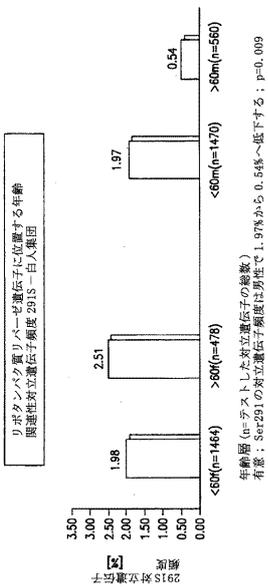
【 図 2 - A 】



【 図 2 - B 】



【 図 2 - C 】



【 図 3 】

集団に基づくサンプルシフトのための質問票

データ収集フォーム

収集情報

同意書にサインしたか Yes No

収集の日付 (MM/DD/YY) \_\_\_/\_\_\_/98

サンプル収集の時間 (24時間時計形式の最近時) \_\_\_\_\_

データ収集者のイニシャル \_\_\_\_\_ 収集コーディネーター \_\_\_\_\_

(記入してはならない(データ入力のための)ラベル) 無傷 \_\_\_\_\_ 喪失 \_\_\_\_\_ 破壊 \_\_\_\_\_

ドナー情報

性別:  男性  女性

どの国 (state)に住んでいるか? \_\_\_\_\_ (MM/YY) / \_\_\_\_年

そこにどのくらいの期間住んでいるか? \_\_\_\_\_ 年

第8年より低  学校で修了最高程度は何ですか? \_\_\_\_\_

ある種の大学第2年  第8、第9、第10または第11年  高等学校卒業または相当

大学卒業第4年  卒業後教育または程度

あなた知っている範囲であった人種起源は何か

父 母

白人 (もしわかれば以下の特定の地理的領域をチェックせよ)

北ヨーロッパ (イギリス、フランス、ドイツ、イタリア、ポーランド、オランダ、デンマーク、スウェーデン、フィンランド)

南ヨーロッパ (ポルトガル、スペイン)

東ヨーロッパ (ロシア、ウクライナ、ポーランド、チェコ、スロバキア)

中央ヨーロッパ (ドイツ、オーストリア、ハンガリー、スイス、フランス、他のヨーロッパ国)

アフリカ系アメリカ人

ヒスパニック(もしわかれば特定の地理的領域をチェックせよ)

メキシコ

中央アメリカ、南アメリカ人

キューバ、プエルトリコ、他のカリブ人

アジア人(もしわかれば以下の特定の地理的領域をチェックせよ)

日本人

中国人

韓国人

ベトナム人

他のアジア人

その他 \_\_\_\_\_

わからない

健康の情報 : あなたまたは両親が以下のものを有しているか? (両親、兄弟、姉妹、またはあなたの子供) 当てはまるものを全てチェックせよ

疾患 : \_\_\_\_\_ あなた 母 父 姉妹 兄弟 子供

心臓疾患発作またはアテロ硬化性動脈硬化症

癌(もしわかればタイプを特定せよ)

アルツハイマー病または痴呆

糖尿病発症または自己免疫性疾患

腎臓系疾患発症または腎臓機能低下

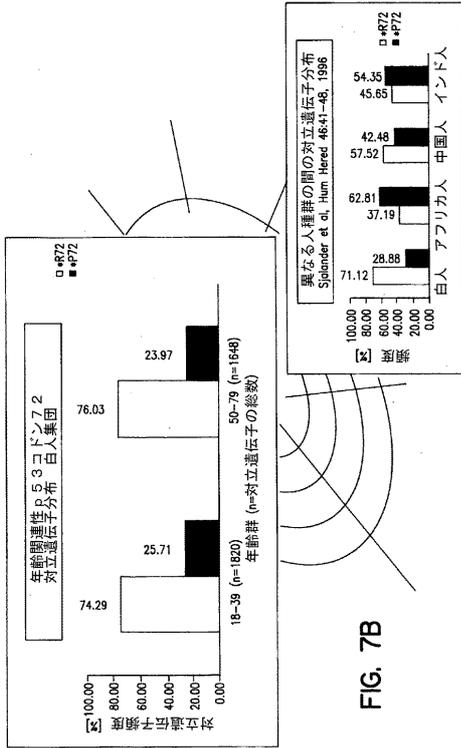
その他(特定せよ)

更なる詳細な健康の情報をあれば記入:

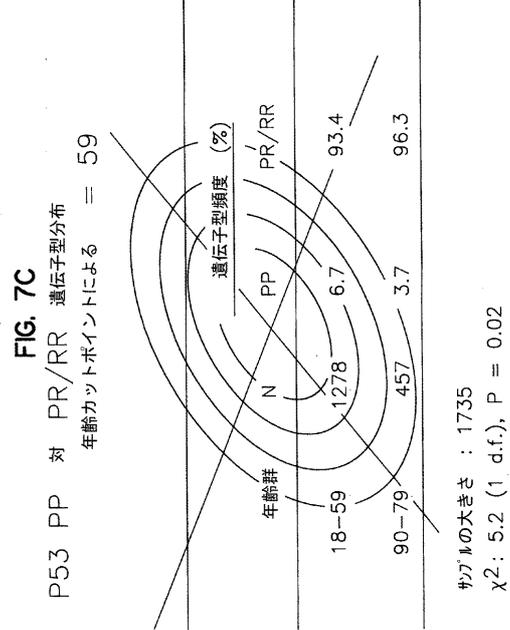
FIG. 3



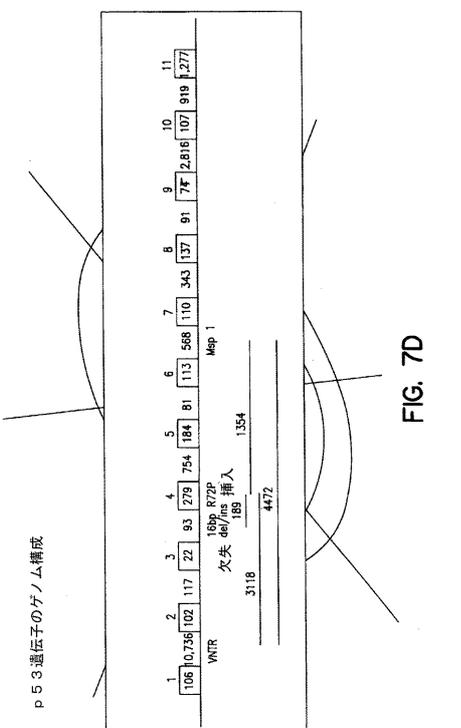
【 図 7 - B 】



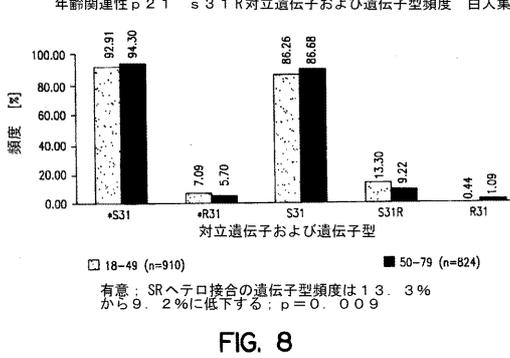
【 図 7 - C 】



【 図 7 - D 】



【 図 8 】



【 図 9 】

プール性対個体サンプルにおける対立遺伝子頻度  
F V I T 対立遺伝子 3 5 3 Q

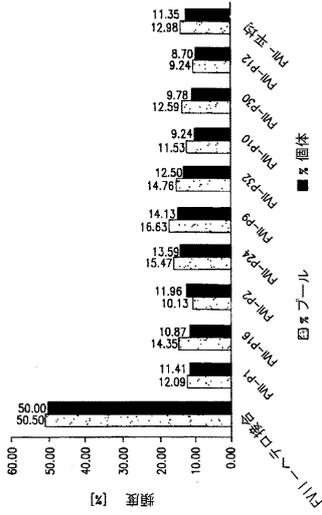


FIG. 9

【 図 1 0 】

プール性対個体サンプルにおける対立遺伝子頻度  
C E T P 対立遺伝子 4 0 5 V

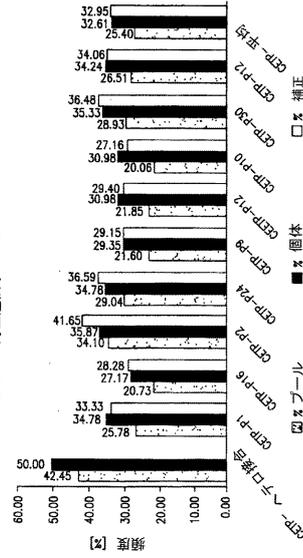


FIG. 10

【 図 1 1 】

プール性対個体サンプルにおける対立遺伝子頻度  
対立遺伝子 P A N 1

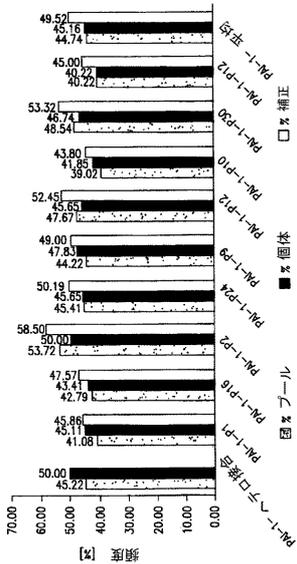


FIG. 11

【 図 1 2 】

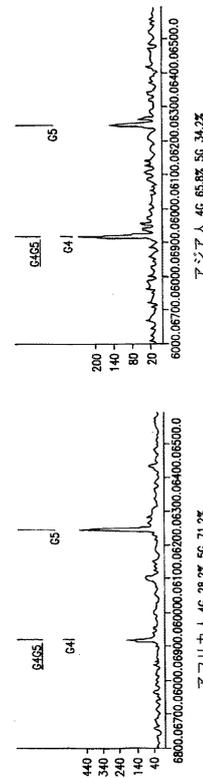


FIG. 12A

FIG. 12B

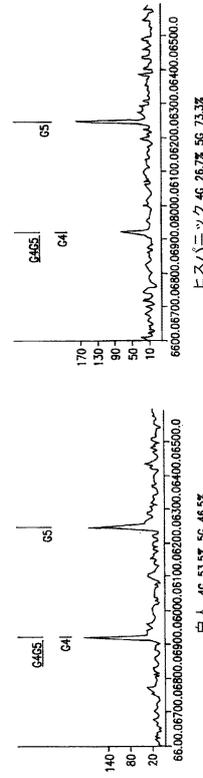
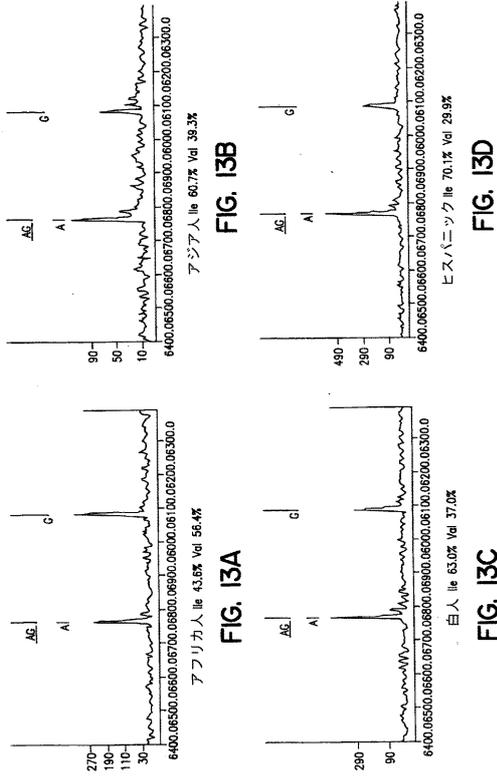


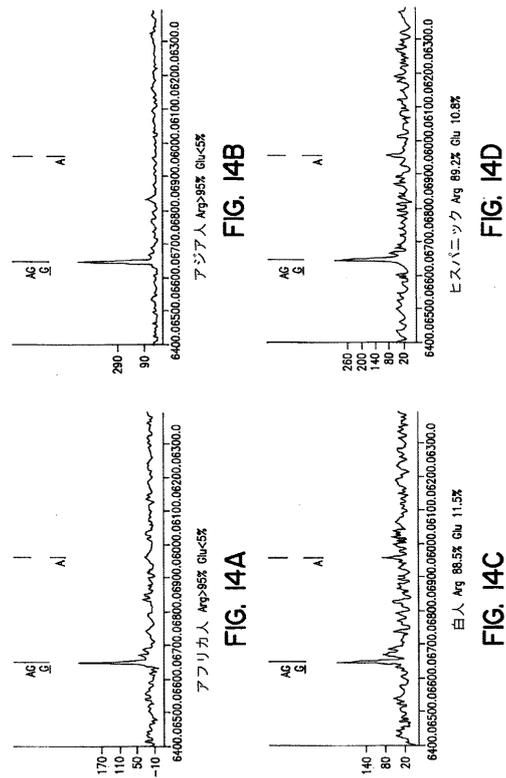
FIG. 12C

FIG. 12D

【 図 1 3 】



【 図 1 4 】



【 図 1 5 】

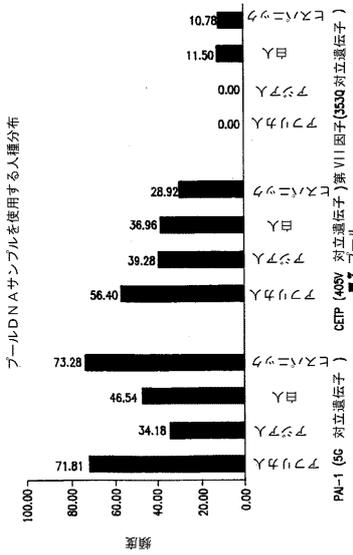


FIG. 15

【 図 1 6 】

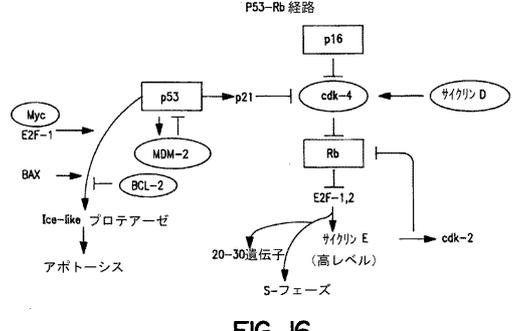


FIG. 16

【 図 1 7 】

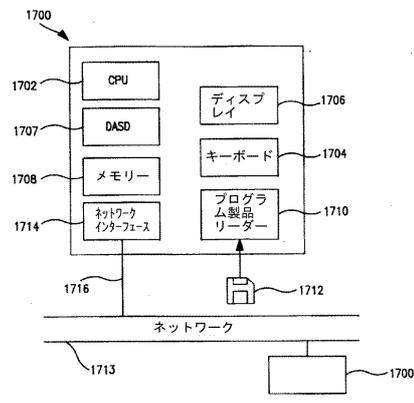


FIG. 17

【 図 18 】

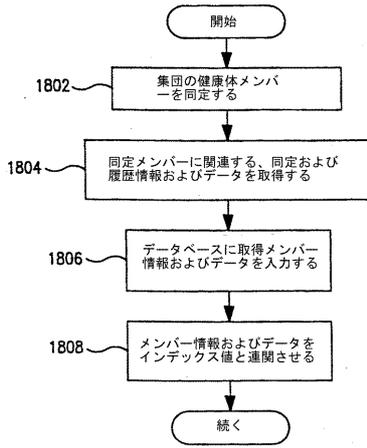


FIG. 18

【 図 19 】

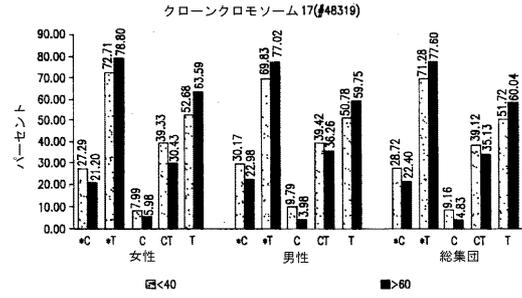


FIG. 19

【 図 20 】

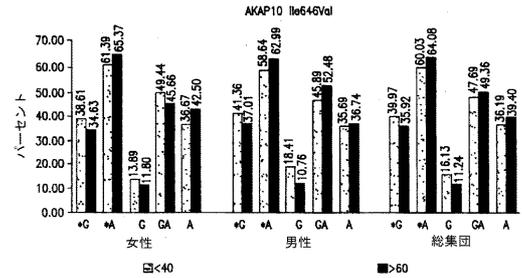


FIG. 20

【 図 21 】

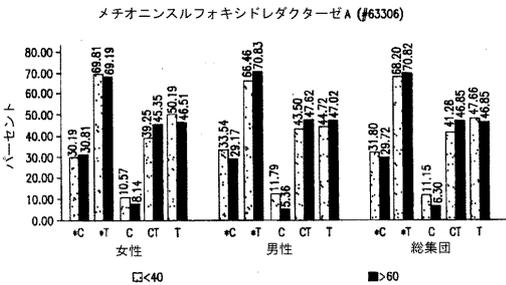


FIG. 21

【 図 22 - A 】

収集情報

同意書にサインしたか  
 Yes  No

Month Year  
 JAN FEB MAR APR MAY JUN JUL AUG SEP OCT NOV DEC

収集の時間 (24時間時計形式で最近時)  
 00 05 10 15 20 25 30 35 40 45 50 55

イニシャル  
 姓: [ ]  
 名: [ ]

データ収集者のイニシャル  
 姓: [ ]  
 名: [ ]

記入してはならない  
 データ入力のための  
 無効  喪失  破壊

ドナー情報

誕生日  
 月 Year  
 JAN FEB MAR APR MAY JUN JUL AUG SEP OCT NOV DEC

性別  
 男性  女性

身長 (cm)  
 [ ]

体重 (kg)  
 [ ]

定期的な身体的活動をしているか?  
 Yes  No

ランニング  
 Yes  No

サイクリング  
 Yes  No

その他の  
 Yes  No

ペジタリ アンカ?  Yes  No

何回妊娠したか? [ ]

何回出産したか? [ ]

わかる範囲であなたの人種は?  
 白人 (わかれば以下の特定の地理的領域をマークせよ)  
 北ヨーロッパ (オランダ、ドイツ、フランス、イギリス、イタリア、オーストリア、デンマーク)  
 南ヨーロッパ (ポルトガル、スペイン、ギリシャ、トルコ、イタリア)  
 中央ヨーロッパ (オーストリア、ドイツ、ポーランド、チェコ、スロバキア、ハンガリー)  
 アフリカ系アメリカ人

どの国に住んでいるか?  
 北アメリカ (わかれば以下の特定の地理的領域をマークせよ)  
 中央アメリカ、南アメリカ  
 キューバ、プエルトリコ、他のカリブ人

人種  
 白人 (わかれば以下の特定の地理的領域をマークせよ)  
 中国人  
 韓国人  
 ヒスパニック  
 フィリピン人  
 ネイティブアメリカ人  
 その他  
 わからない

学校で修了した最高学位は?  
 第8、第9、第10、または第11年  
 同等  
 ある種の大学、2年  
 卒業後教育またはその程度

母は死んだか? 母の死の原因: 父は死んだか? 父の死の原因:  
 Yes  No  Yes  No

心臓疾患  
 <20  20-29  30-39  40-49  50-59  60-69  70-79  80-89  >90

心臓疾患  
 <20  20-29  30-39  40-49  50-59  60-69  70-79  80-89  >90

心臓疾患  
 <20  20-29  30-39  40-49  50-59  60-69  70-79  80-89  >90

FIG. 22A

【 図 22 - B 】

喫煙したことがあるか?  Yes  No 過去5年間で1日に6日より多く入院したか?  Yes  No

Yesの場合、どのくらいの期間か?  年  月  日

Yesの場合、何回か?  1回  2回  3回  4回  5回  6回  7回  8回  9回  10回  11回  12回  13回  14回  15回  16回  17回  18回  19回  20回  21回  22回  23回  24回  25回  26回  27回  28回  29回  30回  31回  32回  33回  34回  35回  36回  37回  38回  39回  40回  41回  42回  43回  44回  45回  46回  47回  48回  49回  50回  51回  52回  53回  54回  55回  56回  57回  58回  59回  60回  61回  62回  63回  64回  65回  66回  67回  68回  69回  70回  71回  72回  73回  74回  75回  76回  77回  78回  79回  80回  81回  82回  83回  84回  85回  86回  87回  88回  89回  90回  91回  92回  93回  94回  95回  96回  97回  98回  99回  100回

それぞれの入院について (もし同じならば) どのくらいの期間入院し、どのような理由か?  急性性障害、感染症及び血栓を含む  慢性性障害  その他

あなたは両親 (両親、兄弟、姉妹、または子供) が以下のものを有したか? 当てはまるものを全てにマークせよ

疾患	あなた	母	父	姉妹	兄弟	子供
心臓疾患、アテローム性動脈硬化症を含む	<input type="checkbox"/>					
糖尿病	<input type="checkbox"/>					
高血圧	<input type="checkbox"/>					
脂質異常症	<input type="checkbox"/>					
慢性腎臓病	<input type="checkbox"/>					
慢性肺疾患	<input type="checkbox"/>					
慢性骨格系疾患	<input type="checkbox"/>					
慢性消化器系疾患	<input type="checkbox"/>					
慢性泌尿器系疾患	<input type="checkbox"/>					
慢性皮膚疾患	<input type="checkbox"/>					
慢性神経系疾患	<input type="checkbox"/>					
慢性精神疾患	<input type="checkbox"/>					
慢性免疫系疾患	<input type="checkbox"/>					
慢性血液系疾患	<input type="checkbox"/>					
慢性眼疾患	<input type="checkbox"/>					
慢性耳疾患	<input type="checkbox"/>					
慢性鼻疾患	<input type="checkbox"/>					
慢性歯疾患	<input type="checkbox"/>					
慢性口腔疾患	<input type="checkbox"/>					
慢性皮膚疾患	<input type="checkbox"/>					
慢性免疫系疾患	<input type="checkbox"/>					
慢性血液系疾患	<input type="checkbox"/>					
慢性眼疾患	<input type="checkbox"/>					
慢性耳疾患	<input type="checkbox"/>					
慢性鼻疾患	<input type="checkbox"/>					
慢性歯疾患	<input type="checkbox"/>					
慢性口腔疾患	<input type="checkbox"/>					
慢性皮膚疾患	<input type="checkbox"/>					
慢性免疫系疾患	<input type="checkbox"/>					
慢性血液系疾患	<input type="checkbox"/>					
慢性眼疾患	<input type="checkbox"/>					
慢性耳疾患	<input type="checkbox"/>					
慢性鼻疾患	<input type="checkbox"/>					
慢性歯疾患	<input type="checkbox"/>					
慢性口腔疾患	<input type="checkbox"/>					
慢性皮膚疾患	<input type="checkbox"/>					
慢性免疫系疾患	<input type="checkbox"/>					
慢性血液系疾患	<input type="checkbox"/>					
慢性眼疾患	<input type="checkbox"/>					
慢性耳疾患	<input type="checkbox"/>					
慢性鼻疾患	<input type="checkbox"/>					
慢性歯疾患	<input type="checkbox"/>					
慢性口腔疾患	<input type="checkbox"/>					
慢性皮膚疾患	<input type="checkbox"/>					
慢性免疫系疾患	<input type="checkbox"/>					
慢性血液系疾患	<input type="checkbox"/>					
慢性眼疾患	<input type="checkbox"/>					
慢性耳疾患	<input type="checkbox"/>					
慢性鼻疾患	<input type="checkbox"/>					
慢性歯疾患	<input type="checkbox"/>					
慢性口腔疾患	<input type="checkbox"/>					
慢性皮膚疾患	<input type="checkbox"/>					
慢性免疫系疾患	<input type="checkbox"/>					
慢性血液系疾患	<input type="checkbox"/>					
慢性眼疾患	<input type="checkbox"/>					
慢性耳疾患	<input type="checkbox"/>					
慢性鼻疾患	<input type="checkbox"/>					
慢性歯疾患	<input type="checkbox"/>					
慢性口腔疾患	<input type="checkbox"/>					
慢性皮膚疾患	<input type="checkbox"/>					
慢性免疫系疾患	<input type="checkbox"/>					
慢性血液系疾患	<input type="checkbox"/>					
慢性眼疾患	<input type="checkbox"/>					
慢性耳疾患	<input type="checkbox"/>					
慢性鼻疾患	<input type="checkbox"/>					
慢性歯疾患	<input type="checkbox"/>					
慢性口腔疾患	<input type="checkbox"/>					
慢性皮膚疾患	<input type="checkbox"/>					
慢性免疫系疾患	<input type="checkbox"/>					
慢性血液系疾患	<input type="checkbox"/>					
慢性眼疾患	<input type="checkbox"/>					
慢性耳疾患	<input type="checkbox"/>					
慢性鼻疾患	<input type="checkbox"/>					
慢性歯疾患	<input type="checkbox"/>					
慢性口腔疾患	<input type="checkbox"/>					
慢性皮膚疾患	<input type="checkbox"/>					
慢性免疫系疾患	<input type="checkbox"/>					
慢性血液系疾患	<input type="checkbox"/>					
慢性眼疾患	<input type="checkbox"/>					
慢性耳疾患	<input type="checkbox"/>					
慢性鼻疾患	<input type="checkbox"/>					
慢性歯疾患	<input type="checkbox"/>					
慢性口腔疾患	<input type="checkbox"/>					
慢性皮膚疾患	<input type="checkbox"/>					
慢性免疫系疾患	<input type="checkbox"/>					
慢性血液系疾患	<input type="checkbox"/>					
慢性眼疾患	<input type="checkbox"/>					
慢性耳疾患	<input type="checkbox"/>					
慢性鼻疾患	<input type="checkbox"/>					
慢性歯疾患	<input type="checkbox"/>					
慢性口腔疾患	<input type="checkbox"/>					
慢性皮膚疾患	<input type="checkbox"/>					
慢性免疫系疾患	<input type="checkbox"/>					
慢性血液系疾患	<input type="checkbox"/>					
慢性眼疾患	<input type="checkbox"/>					
慢性耳疾患	<input type="checkbox"/>					
慢性鼻疾患	<input type="checkbox"/>					
慢性歯疾患	<input type="checkbox"/>					
慢性口腔疾患	<input type="checkbox"/>					
慢性皮膚疾患	<input type="checkbox"/>					
慢性免疫系疾患	<input type="checkbox"/>					
慢性血液系疾患	<input type="checkbox"/>					
慢性眼疾患	<input type="checkbox"/>					
慢性耳疾患	<input type="checkbox"/>					
慢性鼻疾患	<input type="checkbox"/>					
慢性歯疾患	<input type="checkbox"/>					
慢性口腔疾患	<input type="checkbox"/>					
慢性皮膚疾患	<input type="checkbox"/>					
慢性免疫系疾患	<input type="checkbox"/>					
慢性血液系疾患	<input type="checkbox"/>					
慢性眼疾患	<input type="checkbox"/>					
慢性耳疾患	<input type="checkbox"/>					
慢性鼻疾患	<input type="checkbox"/>					
慢性歯疾患	<input type="checkbox"/>					
慢性口腔疾患	<input type="checkbox"/>					
慢性皮膚疾患	<input type="checkbox"/>					
慢性免疫系疾患	<input type="checkbox"/>					
慢性血液系疾患	<input type="checkbox"/>					
慢性眼疾患	<input type="checkbox"/>					
慢性耳疾患	<input type="checkbox"/>					
慢性鼻疾患	<input type="checkbox"/>					
慢性歯疾患	<input type="checkbox"/>					
慢性口腔疾患	<input type="checkbox"/>					
慢性皮膚疾患	<input type="checkbox"/>					
慢性免疫系疾患	<input type="checkbox"/>					
慢性血液系疾患	<input type="checkbox"/>					
慢性眼疾患	<input type="checkbox"/>					
慢性耳疾患	<input type="checkbox"/>					
慢性鼻疾患	<input type="checkbox"/>					
慢性歯疾患	<input type="checkbox"/>					
慢性口腔疾患	<input type="checkbox"/>					
慢性皮膚疾患	<input type="checkbox"/>					
慢性免疫系疾患	<input type="checkbox"/>					
慢性血液系疾患	<input type="checkbox"/>					
慢性眼疾患	<input type="checkbox"/>					
慢性耳疾患	<input type="checkbox"/>					
慢性鼻疾患	<input type="checkbox"/>					
慢性歯疾患	<input type="checkbox"/>					
慢性口腔疾患	<input type="checkbox"/>					
慢性皮膚疾患	<input type="checkbox"/>					
慢性免疫系疾患	<input type="checkbox"/>					
慢性血液系疾患	<input type="checkbox"/>					
慢性眼疾患	<input type="checkbox"/>					
慢性耳疾患	<input type="checkbox"/>					
慢性鼻疾患	<input type="checkbox"/>					
慢性歯疾患	<input type="checkbox"/>					
慢性口腔疾患	<input type="checkbox"/>					
慢性皮膚疾患	<input type="checkbox"/>					
慢性免疫系疾患	<input type="checkbox"/>					
慢性血液系疾患	<input type="checkbox"/>					
慢性眼疾患	<input type="checkbox"/>					
慢性耳疾患	<input type="checkbox"/>					
慢性鼻疾患	<input type="checkbox"/>					
慢性歯疾患	<input type="checkbox"/>					
慢性口腔疾患	<input type="checkbox"/>					
慢性皮膚疾患	<input type="checkbox"/>					
慢性免疫系疾患	<input type="checkbox"/>					
慢性血液系疾患	<input type="checkbox"/>					
慢性眼疾患	<input type="checkbox"/>					
慢性耳疾患	<input type="checkbox"/>					
慢性鼻疾患	<input type="checkbox"/>					
慢性歯疾患	<input type="checkbox"/>					
慢性口腔疾患	<input type="checkbox"/>					

【図24】

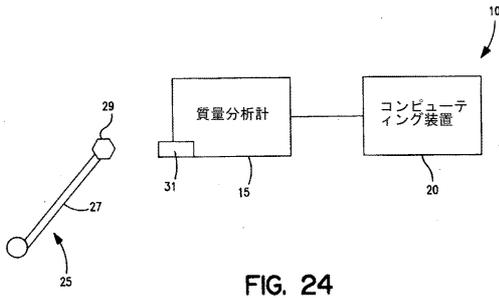


FIG. 24

【図25】

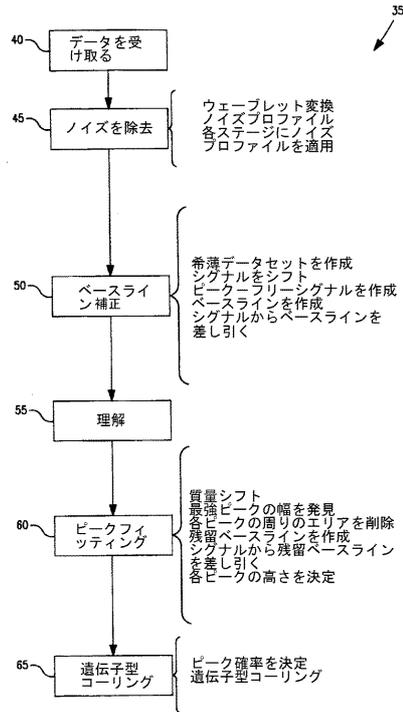


FIG. 25

【図26】

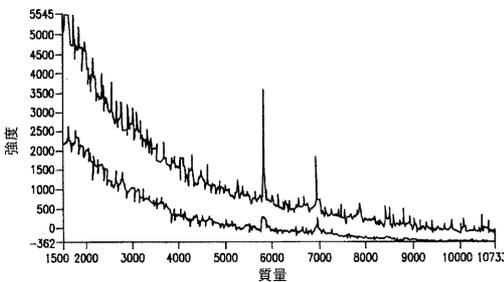


FIG. 26

【図29】

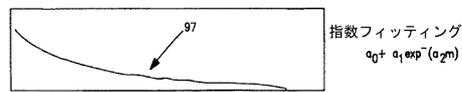


FIG. 29

【図30】

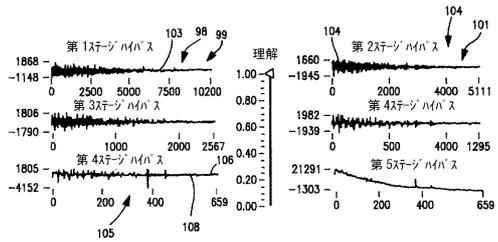


FIG. 30

【図27】

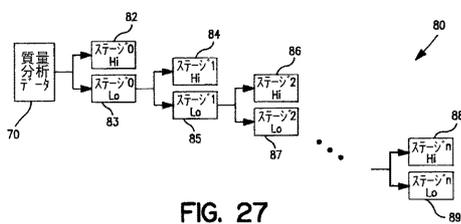


FIG. 27

【図28】

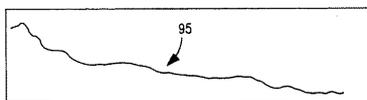


FIG. 28

【図 3 1】

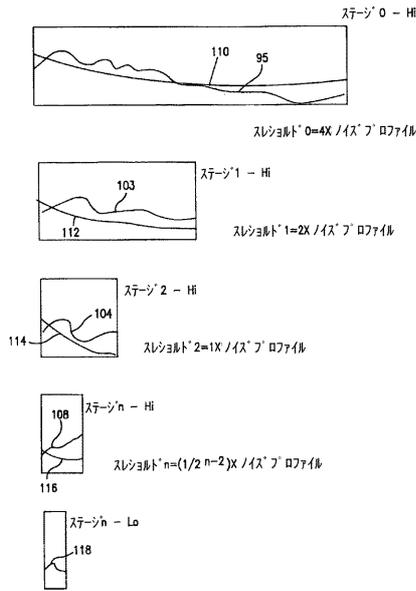


FIG. 31

【図 3 2】

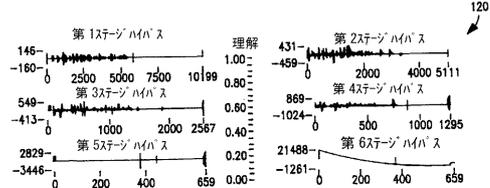


FIG. 32

【図 3 3】

$$\text{シグナル}(t) = \frac{\text{開始}0(t) + \text{開始}1(t) + \text{開始}2(t) + \dots + \text{開始}23(t)}{24}$$

開始点による変異を説明  
するためのシフトシグナル

FIG. 33

【図 3 4】

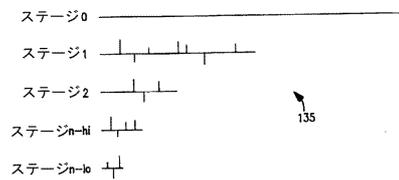


FIG. 34

【図 3 5】

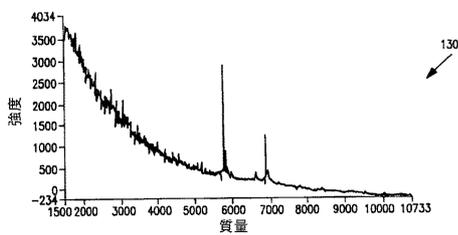


FIG. 35

【図 3 7】

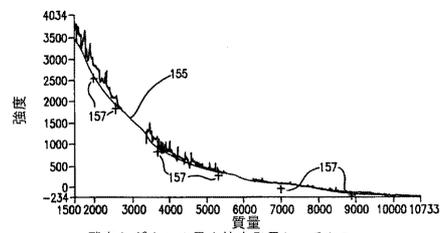


FIG. 37 残存シグナルの最小値を発見し、そしてピークフリーシグナルを形成するために接続する

【図 3 6】

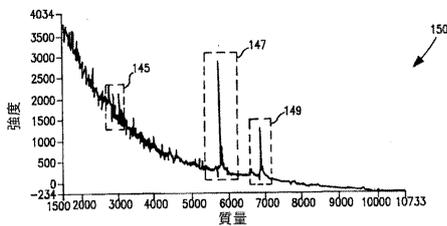


FIG. 13- 移動平均を取得し、スレシヨルトを超えるセクションを取り除く

FIG. 36

【図 3 8】

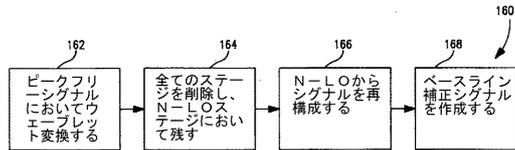


FIG. 38 ベースライン補正を作成する

【図 3 9】

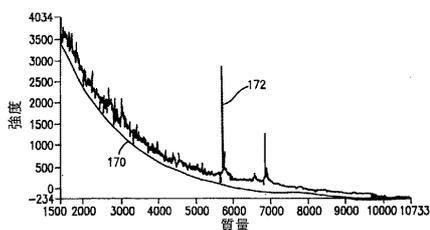


FIG. 39

【 図 4 0 】

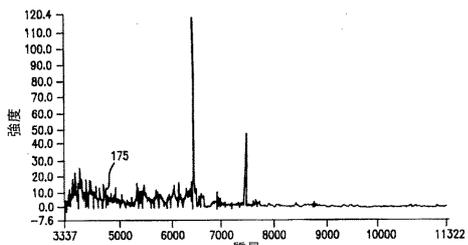


FIG. 40

【 図 4 1 】

NON-0 係数	係数	中間値	比較値
100	25	100.025	100.025
150	220	150.220	50.220
500	.1	500.0001	350.0001
10,050	800	10,050.8	9550.8
10,075	890	10,075.89	25.89
11,125	910	11,125.91	150.91
12,100	1000 (最大)	12,100.99999	975.99999
13,250	940	13,250.94	1150.94

FIG. 41

【 図 4 2 】

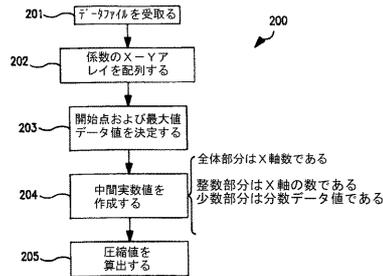


FIG. 42

【 図 4 3 】

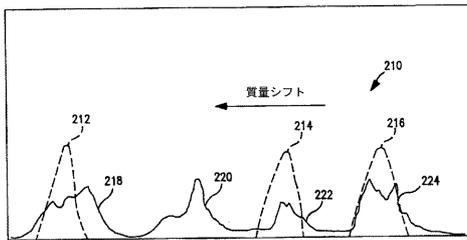


FIG. 43

【 図 4 4 】

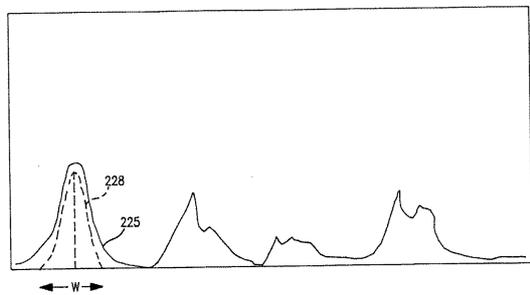


FIG. 44

【 図 4 6 】

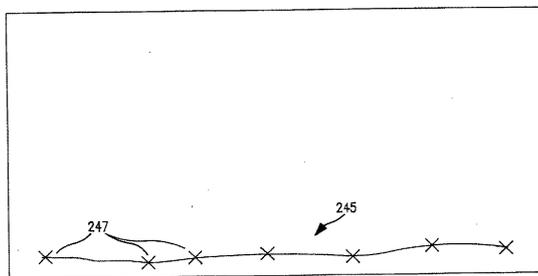


FIG. 46

【 図 4 5 】

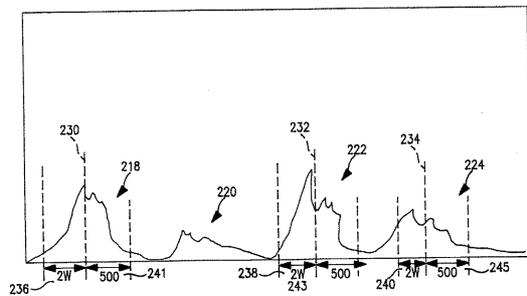


FIG. 45

【 図 4 7 】

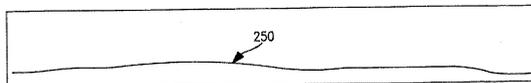


FIG. 47

【 図 4 8 】

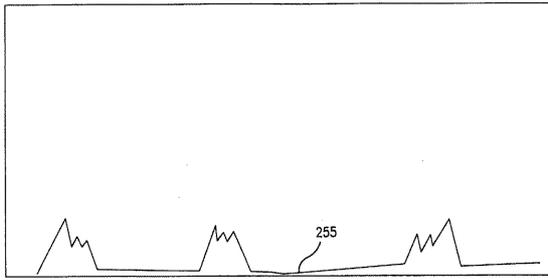


FIG. 48

【 図 5 0 】

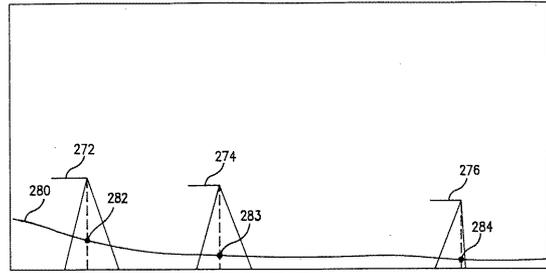


FIG. 50

【 図 4 9 】

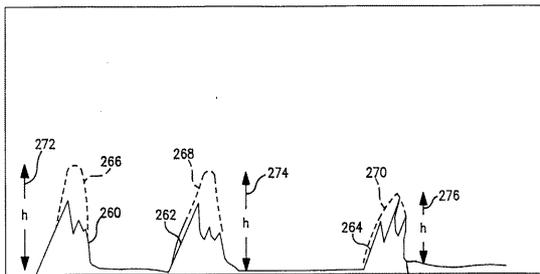


FIG. 49

【 図 5 1 】

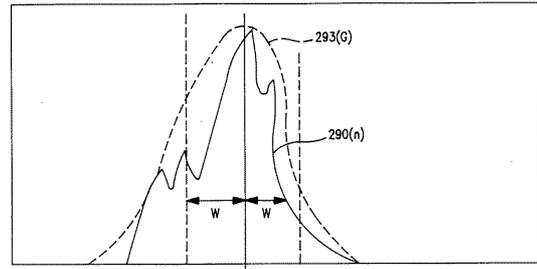


FIG. 51

【 図 5 2 】

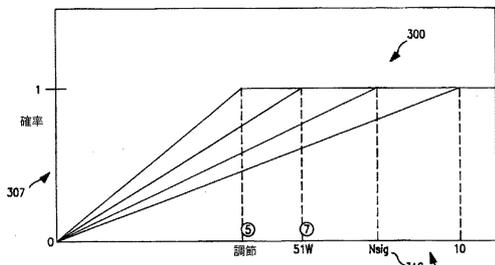


FIG. 52

【 図 5 4 】

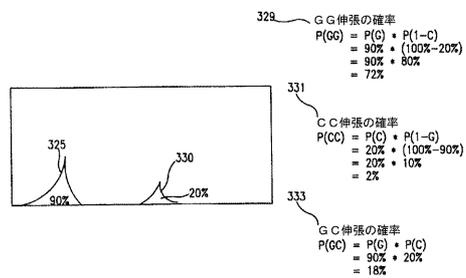


FIG. 54

【 図 5 3 】

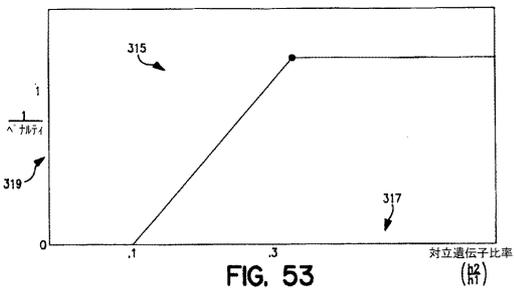


FIG. 53

【 図 5 5 】

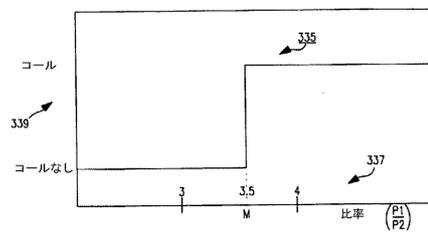


FIG. 55

【 図 5 6 】

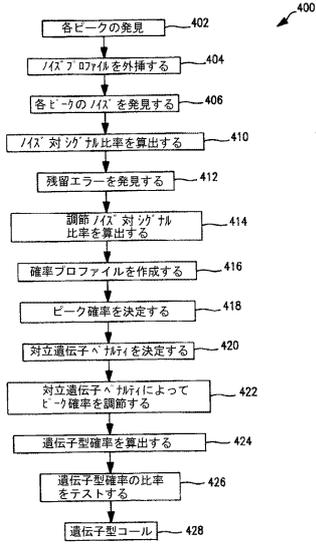


FIG. 56

【 図 5 7 】

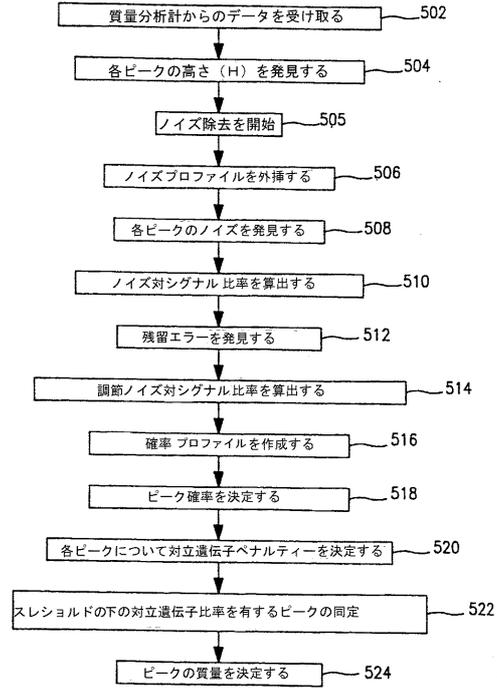


FIG. 57

【 図 5 8 】

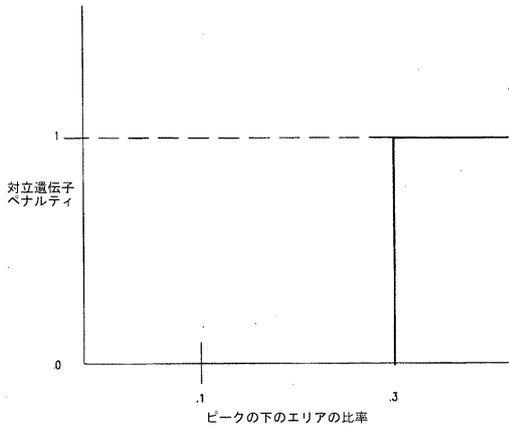


FIG. 58

【配列表】

2004158005000001.app

## フロントページの続き

(51) Int.Cl. <sup>7</sup>	F I	テーマコード(参考)
G 0 1 N 27/62	G 0 1 N 27/62	V
G 0 1 N 27/64	G 0 1 N 27/64	B
G 0 1 N 33/483	G 0 1 N 33/483	Z
G 0 1 N 33/50	G 0 1 N 33/50	P
G 0 6 F 17/30	G 0 6 F 17/30	1 7 0 F
	C 1 2 N 15/00	Z N A A

(31)優先権主張番号 09/663,968

(32)優先日 平成12年9月19日(2000.9.19)

(33)優先権主張国 米国(US)

(特許庁注：以下のものは登録商標)

1. ペンティアム

(72)発明者 フーベルト・ケスター

スイス、ツェーハー - 6 9 0 0 ルガーノ - カッサラーテ、ピア・デッレ・スクオーレ1番

(72)発明者 ディルク・バン・デン・ボーム

ドイツ連邦共和国デー - 2 0 2 5 3 ハンブルク、エッペンドルファー・ベーク205デー番

(72)発明者 イップ・ピン

アメリカ合衆国9 2 1 1 6 カリフォルニア州サンディエゴ、コプリー・アベニュー3 6 4 1番

(72)発明者 チャーリー・ロディ

アメリカ合衆国9 2 0 1 4 カリフォルニア州デル・マー、レクエルド・ドライブ1 3 8 2 3番

(72)発明者 ヘ・リヤン

アメリカ合衆国9 2 1 2 8 カリフォルニア州サンディエゴ、クリーク・ブリッジ・プレイス1 0 9  
4 8番

(72)発明者 ノーマン・チウ

アメリカ合衆国9 2 1 2 6 カリフォルニア州サンディエゴ、カミニト・アルバレス1 1 2 8番

(72)発明者 クリスティアン・ユリンケ

ドイツ連邦共和国デー - 2 0 2 5 5 ハンブルク、ロンベルクシュトラッセ2 2番

Fターム(参考) 2G045 DA13 DA36 FA40 FB01 JA01

4B024 AA11 CA02 CA09 HA14 HA19

4B029 AA07 AA21 AA23 BB20 CC03 FA02

4B063 QA12 QA17 QA18 QA19 QQ02 QQ03 QQ42 QQ52 QR10 QR55

QR62 QR82 QS25 QS34 QS39

5B075 ND03 UU19

专利名称(译)	创建用于鉴定多态性遗传标记的数据库和数据库的方法		
公开(公告)号	<a href="#">JP2004158005A</a>	公开(公告)日	2004-06-03
申请号	JP2003360454	申请日	2003-10-21
[标]申请(专利权)人(译)	每次海阔喃股份有限公司忘川		
申请(专利权)人(译)	每次Shikuenomu酒店股份有限公司的Rete		
[标]发明人	アンドレアスブラウン フーベルトケスター デイルクバンデンポーム イップピン チャーリーロディ ヘリヤン ノーマンチウ クリステリアンユリンケ		
发明人	アンドレアス・ブラウン フーベルト・ケスター デイルク・バン・デン・ポーム イップ・ピン チャーリー・ロディ ヘ・リヤン ノーマン・チウ クリステリアン・ユリンケ		
IPC分类号	G01N27/62 C12M1/00 C12N15/09 C12Q1/00 C12Q1/34 C12Q1/44 C12Q1/48 C12Q1/68 G01N27/64 G01N33/48 G01N33/483 G01N33/50 G01N33/53 G01N33/566 G06F17/30 G06F19/18 G06F19/22 G06F19/28 G06Q50/00 G06F19/00		
CPC分类号	C12Q1/6827 G16B20/00 G16B30/00 G16B50/00 C12Q2563/167 C12Q2523/125		
FI分类号	G06F19/00.600 C12M1/00.A C12Q1/34 C12Q1/68.A G01N27/62.K G01N27/62.V G01N27/64.B G01N33/483.Z G01N33/50.P G06F17/30.170.F C12N15/00.ZNA.A C12N15/00.A C12N15/00.AZN.A C12Q1/6827.C C12Q1/6827.Z G06F16/20 G06F16/28 G06F19/00.618 G06F19/00.628 G06F19/18 G06F19/28 G16B20/00 G16B50/00 G16Z99/00		
F-TERM分类号	2G045/DA13 2G045/DA36 2G045/FA40 2G045/FB01 2G045/JA01 4B024/AA11 4B024/CA02 4B024 /CA09 4B024/HA14 4B024/HA19 4B029/AA07 4B029/AA21 4B029/AA23 4B029/BB20 4B029/CC03 4B029/FA02 4B063/QA12 4B063/QA17 4B063/QA18 4B063/QA19 4B063/QQ02 4B063/QQ03 4B063 /QQ42 4B063/QQ52 4B063/QR10 4B063/QR55 4B063/QR62 4B063/QR82 4B063/QS25 4B063/QS34 4B063/QS39 5B075/ND03 5B075/UU19 2G041/CA01 2G041/DA04 2G041/DA05 2G041/FA10 2G041 /FA11 2G041/FA12 2G041/FA13 2G041/GA05 2G041/GA06 2G041/JA02 2G041/LA06 2G041/LA07 5L049/DD06		
优先权	60/159176 1999-10-13 US 60/217658 2000-07-10 US 60/217251 2000-07-10 US 09/663968 2000-09-19 US		
外部链接	<a href="#">Espacenet</a>		

#### 摘要(译)

鉴定DNA中SNP的唯一可用方法是测序，这种方法昂贵，困难且费力。因此，目的是提供一种廉价，简单且省力的测序方法。此外，一个目的是提供一种用于鉴定疾病的基因组基础及其标志物的方法。提供了用于创建来自健康人类供体的基因组样品的数据

库，识别多态性遗传标记和其他标记以及使用数据库与疾病和状况相关的方法和。 [选择图]无

対立遺伝子	終結産物 : ddA, ddT	配列#	長さ	頻度
LPL-2	caatctgggctatgagatca	19	20	6141
291 Asn	caatctgggctatgagatcaa	20	21	6438.2
291 Ser	caatctgggctatgagatcagt	21	22	6758.4