

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第6268184号
(P6268184)

(45) 発行日 平成30年1月24日(2018.1.24)

(24) 登録日 平成30年1月5日(2018.1.5)

(51) Int. Cl.		F I	
G06F	19/18	(2011.01)	G06F 19/18
C12Q	1/68	(2018.01)	C12Q 1/68 A
G06F	19/24	(2011.01)	G06F 19/24

請求項の数 15 (全 17 頁)

(21) 出願番号	特願2015-543547 (P2015-543547)	(73) 特許権者	590000248
(86) (22) 出願日	平成25年11月15日(2013.11.15)		コーニンクレッカ フィリップス エヌ ヴェ
(65) 公表番号	特表2016-504667 (P2016-504667A)		KONINKLIJKE PHILIPS N. V.
(43) 公表日	平成28年2月12日(2016.2.12)		オランダ国 5656 アーエー アイン ドーフエン ハイテック キャンパス 5
(86) 国際出願番号	PCT/IB2013/060150		High Tech Campus 5, NL-5656 AE Eindhoven
(87) 国際公開番号	W02014/080323	(74) 代理人	100107766
(87) 国際公開日	平成26年5月30日(2014.5.30)		弁理士 伊東 忠重
審査請求日	平成28年11月14日(2016.11.14)	(74) 代理人	100070150
(31) 優先権主張番号	61/729, 678		弁理士 伊東 忠彦
(32) 優先日	平成24年11月26日(2012.11.26)		
(33) 優先権主張国	米国 (US)		

最終頁に続く

(54) 【発明の名称】 患患者固有の関連性評価を用いた変異と疾患の関連付けを使用する診断的遺伝子分析

(57) 【特許請求の範囲】

【請求項1】

電子データ処理デバイスによって方法を実行するのに実行可能な命令を格納する非一時的記憶媒体であって、前記方法は、

診断対象の遺伝子データ内において、臨床解析により表現型特性に関連付けられる解析遺伝子変異を識別するステップと、

前記解析遺伝子変異に機能的に関連する多型のセットを識別するステップと、

前記多型のセットについて、前記診断対象の遺伝子データ内で観察される変異のフォアグラウンド分布を計算するステップと、

前記多型のセットについて、前記臨床解析の対象の遺伝子データ内で観察されるバックグラウンド分布を計算するステップと、

前記フォアグラウンド分布と前記バックグラウンド分布を比較する比較メトリックを計算するステップと、

前記比較メトリックに基づいて、前記診断対象に対する前記解析遺伝子変異の関連性を定量化するステップと、

を含み、

前記多型のセットについて、前記診断対象の遺伝子データ内で観察される変異のフォアグラウンド分布を計算するステップは、前記多型のセットについて、前記診断対象の遺伝子データ内で観察される変異を保持するベクトル要素を有する診断対象の特徴ベクトルを計算することを含み、

10

20

前記多型のセットについて、前記臨床解析の対象の遺伝子データ内で観察されるバックグラウンド分布を計算するステップは、前記多型のセットについて、前記臨床解析の各対象について、該対象の遺伝子データ内で観察される変異を保持するベクトル要素を有する対象特徴ベクトルを計算することを含む、
非一時的記憶媒体。

【請求項 2】

前記多型のセットを識別するステップは、
前記解析遺伝子変異が関連付けられる遺伝経路を識別するステップと、
前記識別された遺伝経路に関連付けられる多型を識別するステップと
を含む、請求項 1 に記載の非一時的記憶媒体。

10

【請求項 3】

前記遺伝経路を識別するステップ及び前記識別された遺伝経路に関連付けられる多型を識別するステップは、遺伝経路データベースを参照する、
請求項 2 に記載の非一時的記憶媒体。

【請求項 4】

前記多型のセットを識別するステップは、
1 つ又は複数の臨床解析により前記解析遺伝子変異が関連付けられる表現型特性に関連付けられる変異を有する多型を識別するステップ、
を含む、請求項 1 又は 2 に記載の非一時的記憶媒体。

20

【請求項 5】

前記多型のセットを識別するステップは、
前記解析遺伝子変異を含む表現型遺伝子シグネチャを識別するステップと、
前記識別された表現型遺伝子シグネチャに関連付けられる多型を識別するステップと、
を含む、請求項 1 乃至 3 のいずれかに記載の非一時的記憶媒体。

【請求項 6】

前記フォアグラウンド分布と前記バックグラウンド分布を比較する比較メトリックを計算するステップは、
前記対象特徴ベクトルをクラスタ化して複数のクラスタを規定するステップと、
前記クラスタからの前記対象特徴ベクトルの特異性を測定する異常値メトリックを計算するステップと、
を含む、請求項 1 に記載の非一時的記憶媒体。

30

【請求項 7】

前記フォアグラウンド分布と前記バックグラウンド分布を比較する比較メトリックを計算するステップは、
前記バックグラウンド分布に対する前記フォアグラウンド分布の有意性を示す統計的有意性のメトリックを計算するステップ
を含む、請求項 1 乃至 5 のいずれかに記載の非一時的記憶媒体。

【請求項 8】

前記比較メトリックに基づいて、前記診断対象に対する前記解析遺伝子変異の関連性を定量化するステップは、
より高い値の前記統計的有意性のメトリックが、前記診断対象に対する前記解析遺伝子変異のより低い関連性を示すように、前記統計的有意性のメトリックの逆として関連性を定量化するステップ
を含む、請求項 7 に記載の非一時的記憶媒体。

40

【請求項 9】

前記の解析遺伝子変異を識別するステップと、前記の多型のセットを識別するステップと、前記のフォアグラウンド分布を計算するステップと、前記のバックグラウンド分布を計算するステップと、前記の比較メトリックを計算するステップと、前記の解析遺伝子変異の関連性を定量化するステップとを繰り返し実行して、前記診断対象の遺伝子データにおける解析遺伝子変異のセットの関連性を識別して定量化するステップと、

50

関連性によってランク付けされた前記解析遺伝子変異を表示するステップと、
を更に含む、請求項 1 乃至 8 のいずれかに記載の非一時的記憶媒体。

【請求項 10】

請求項 1 乃至 9 のいずれかに記載される非一時的記憶媒体と、
前記非一時的記憶媒体上に格納される命令を実行するように構成される電子データ処理
デバイスと、
を備える、装置。

【請求項 11】

診断対象の遺伝子データ内において観察される、臨床解析により表現型特性に関連付け
られる解析遺伝子変異に機能的に関連する多型のセットを識別するステップと、
前記多型のセットについて、前記診断対象の遺伝子データ内で観察される変異のフォア
グラウンド分布を計算するステップと、
前記多型のセットについて、前記臨床解析の対象の遺伝子データ内で観察されるバック
グラウンド分布を計算するステップと、
前記フォアグラウンド分布と前記バックグラウンド分布を比較する比較メトリックを計
算するステップと、
前記比較メトリックに基づいて、前記診断対象に対する前記解析遺伝子変異の関連性を
定量化するステップと、

を含み、

前記の多型のセットを識別するステップと、前記のフォアグラウンド分布を計算するス
テップと、前記のバックグラウンド分布を計算するステップと、前記の比較メトリックを
計算するステップは、電子データ処理デバイスによって実行され、

前記多型のセットについて、前記診断対象の遺伝子データ内で観察される変異のフォア
グラウンド分布を計算するステップは、前記多型のセットについて、前記診断対象の遺伝
子データ内で観察される変異を保持するベクトル要素を有する診断対象の特徴ベクトルを
計算することを含み、

前記多型のセットについて、前記臨床解析の対象の遺伝子データ内で観察されるバック
グラウンド分布を計算するステップは、前記多型のセットについて、前記臨床解析の各対
象について、該対象の遺伝子データ内で観察される変異を保持するベクトル要素を有する
対象特徴ベクトルを計算することを含む、方法。

【請求項 12】

前記診断対象の遺伝子データ内において前記解析遺伝子変異を識別するステップを更に
含む、

前記解析遺伝子変異は、遺伝子配列データとマイクロアレイデータのうちの少なくとも
一方を含む、

請求項 11 に記載の方法。

【請求項 13】

前記多型のセットを識別するステップは、

前記解析遺伝子変異が関連付けられる遺伝経路を識別するステップと、

前記識別された遺伝経路に関連付けられる多型を識別するステップと

を含む、請求項 11 又は 12 に記載の方法。

【請求項 14】

前記多型のセットを識別するステップは、

前記解析遺伝子変異が関連付けられる前記表現型特性に 1 つ又は複数の臨床解析により
関連付けられる変異を有する多型を識別するステップ、

を含む、請求項 11 乃至 13 のいずれかに記載の方法。

【請求項 15】

前記多型のセットを識別するステップは、

前記解析遺伝子変異を含む表現型遺伝子シグネチャを識別するステップと、

前記識別された表現型遺伝子シグネチャに関連付けられる多型を識別するステップと、

を含む、請求項 1 1 乃至 1 4 のいずれかに記載の方法。

【発明の詳細な説明】

【技術分野】

【0001】

以下の説明は、遺伝子分析の分野、医療分野及びその用途に関し、医療分野は腫瘍学の分野、獣医学の分野等を含む。

【背景技術】

【0002】

ゲノムワイド関連 (GWA : genome-wide association) 解析は、異なる個体のゲノムにおける変異を検査して、異なる個体で見つかった変異と、疾患の罹患性のようなその表現型との間の関係を識別する。次いで、異なる変異が、疾患のような異なる形質に関連付けられる。そのような臨床解析は、後に患者において検出されて、関連する疾患 (又は他の表現型特性) のインジケータとして機能し得る、変異を識別する。

【0003】

ヒトにおける GWA 解析は、特定の遺伝子と、加齢性黄斑変性症として知られる眼の疾患や糖尿病のような疾患との関連の発見につながった。典型的に、これらの解析では、何百、何千という個体が、そのゲノムに関して何十万という変異について検査される。変異は通常、一塩基多型、すなわち SNP (single-nucleotide polymorphism) である。1200 超の GWA 解析で、200 超の疾患及び形質について検査し、4000 近くの SNP 関連を見つけた。

【0004】

特許文献 1 (国際公開第 01 / 79540 号パンフレット) は、遺伝子型情報を、表現型プロファイルに相関させることにより、薬物耐性を測定するための方法に関する。特定の遺伝子配列における変異パターンが識別され、変異は、特定の治療に対する耐性と関連付けられる。遺伝子型のデータベースが、同じ効果を有する類似の意遠位パターンについて検索される。

【先行技術文献】

【特許文献】

【0005】

【特許文献 1】国際公開第 01 / 79540 号パンフレット

【発明の概要】

【発明が解決しようとする課題】

【0006】

以下では、上記の制限及び他の事項を克服する、改善された装置及び方法を検討する。

【課題を解決するための手段】

【0007】

一態様によると、非一時的記憶媒体は、電子データ処理デバイスによって方法を実行するのに実行可能な命令を格納しており、上記の方法は：診断対象の遺伝子データ内において、臨床解析により表現型特性に関連付けられる解析遺伝子変異 (study genetic variant) を識別するステップと；解析遺伝子変異に機能的に関連する多型のセットを識別するステップと；多型のセットについて、診断対象の遺伝子データ内で観察される変異のフォアグラウンド分布を計算するステップと；多型のセットについて、臨床解析の対象の遺伝子データ内で観察されるバックグラウンド分布を計算するステップと；フォアグラウンド分布とバックグラウンド分布を比較する比較メトリックを計算するステップと；比較メトリックに基づいて、診断対象に対する解析遺伝子変異の関連性を定量化するステップとを含む。多型のセットについて、診断対象の遺伝子データ内で観察される変異のフォアグラウンド分布を計算するステップは、多型のセットについて、診断対象の遺伝子データ内で観察される変異を保持するベクトル要素を有する診断対象の特徴ベクトルを計算することを含み、多型のセットについて、臨床解析の対象の遺伝子データ内で観察されるバックグラウンド分布を計算するステップは、多型のセットについて、臨床解析の各対象につい

10

20

30

40

50

て、該対象の遺伝子データ内で観察される変異を保持するベクトル要素を有する対象特徴ベクトルを計算することを含む。

【0008】

別の態様によると、装置は、直ぐ上の段落に記載される非一時的記憶媒体と；この非一時的記憶媒体上に格納される命令を実行するように構成される電子データ処理デバイスとを備える。

【0009】

別の態様によると、方法は、診断対象の遺伝子データ内において観察される、臨床解析により表現型特性に関連付けられる解析遺伝子変異に機能的に関連する多型のセットを識別するステップと；多型のセットについて、診断対象の遺伝子データ内で観察される変異のフォアグラウンド分布を計算するステップと；多型のセットについて、臨床解析の対象の遺伝子データ内で観察されるバックグラウンド分布を計算するステップと；フォアグラウンド分布とバックグラウンド分布を比較する比較メトリックを計算するステップと、比較メトリックに基づいて、診断対象に対する前記解析遺伝子変異の関連性を定量化するステップとを含む。多型のセットについて、診断対象の遺伝子データ内で観察される変異のフォアグラウンド分布を計算するステップは、多型のセットについて、診断対象の遺伝子データ内で観察される変異を保持するベクトル要素を有する診断対象の特徴ベクトルを計算することを含み、多型のセットについて、臨床解析の対象の遺伝子データ内で観察されるバックグラウンド分布を計算するステップは、多型のセットについて、臨床解析の各対象について、該対象の遺伝子データ内で観察される変異を保持するベクトル要素を有する対象特徴ベクトルを計算することを含む。

【0010】

上記の多型のセットを識別するステップと、フォアグラウンド分布を計算するステップと、バックグラウンド分布を計算するステップと、比較メトリックを計算するステップは適切に、電子データ処理デバイスによって実行される。

【0011】

更に別の態様によると、上記方法において、フォアグラウンド分布とバックグラウンド分布を比較する比較メトリックを計算するステップは、対象特徴ベクトルをクラスタ化して複数のクラスタを規定するステップと、クラスタからの対象特徴ベクトルの特異性を測定する異常値メトリックを計算するステップとを含む。

【0012】

更に別の態様によると、上記方法は、解析遺伝子変異を識別するステップと、多型のセットを識別するステップと、フォアグラウンド分布を計算するステップと、バックグラウンド分布を計算するステップと、前較メトリックを計算するステップとを、診断対象の遺伝子データ内において観察される解析遺伝子変異のセットであって、1つ又は複数の表現型特性を有する1つ又は複数の臨床解析によって関連付けられる解析遺伝子変異のセットの各遺伝子変異について繰り返すステップと、解析遺伝子変異のセットを、診断対象に対する各解析遺伝子変異の関連性の指示とともに表示するステップとを更に含み、であって、関連性の指示は、比較メトリックに基づいて計算される。

【0013】

更に別の態様によると、多型のセットを識別するステップは、一塩基多型(SNP)、配列多型、挿入、削除、コピー数多型、構造的多型及びメチル化多型からなるグループから多型を識別する。

【0014】

更に別の態様によると、多型のセットを識別するステップは、一塩基多型(SNP)のみを識別することに制限される。

【0015】

1つの利点は、疾患又は別の表現型特性に相関させるよう、一般的に知られる遺伝子変異について患者固有の関連性評価を提供することにある。

【0016】

別の利点は、遺伝子マーカを使用して、より正確な臨床検査を行うことにある。

【0017】

様々な追加の利点及び効果は、以下の詳細な説明を読むことにより、当業者に明らかになるであろう。

【図面の簡単な説明】

【0018】

本発明は、様々なコンポーネントとコンポーネントの様々な配置及び様々な処理動作と処理動作の様々な配置の形をとってよい。図面は、好ましい実施形態を説明するためだけのものであり、本発明を限定するものとして解釈されるべきではない。

【図1】患者の遺伝子データセットを分析して、患者の疾患及び別の表現型特性を検出するためのシステムを図形的に示す図である。

10

【図2】図1のシステムのうちの選択されたコンポーネントの詳細な描写を図形的に示す図である。

【図3】図1のシステムのうちの選択されたコンポーネントの詳細な描写を図形的に示す図である。

【図4】図1のシステムのうちの選択されたコンポーネントの詳細な描写を図形的に示す図である。

【図5】図1のシステムのうちの選択されたコンポーネントの詳細な描写を図形的に示す図である。

【図6】図1のシステムのうちの選択されたコンポーネントの詳細な描写を図形的に示す図である。

20

【発明を実施するための形態】

【0019】

遺伝的評価は典型的に、患者の遺伝子データを取得することと、この遺伝子データを、臨床解析によって示されるような疾患の関連性（又は他の表現型の関連性）を有する遺伝子変異と比較することを含む。説明上の例として、取得した患者の遺伝子データは、全ゲノム配列（WGS: whole genome sequence）を含むことがあり（典型的には核DNA含量の少なくとも90%、より典型的には、核DNA含量の95%以上を含むと考えられ、任意選択によりRNA又は他の遺伝子配列決定データ（genetic sequencing data）も含むと考えられる）、遺伝子配列決定データは、WGSより少なく、マイクロアレイを使用して取得される、例えば何十万の一塩基多型（SNP）測定値を有する遺伝子データ等（潜在的には、遺伝子発現の形の遺伝子データ、すなわちプロテオミックの情報を含む）を備える。例えば高いスループットの遺伝子配列決定処理又はマイクロアレイ処理によって作成されるような、大規模な患者の遺伝子データセットについて、何千もの遺伝子変異が患者の遺伝子データ内で観察されることがある。典型的には、これらの変異の何十、何百という変異は、何らかのGWA解析において、関連する疾患となると既に判断されている。

30

【0020】

しかしながら、特定の個体で観察される遺伝子変異に関連付けられるリスクは、臨床解析の結果のみに基づいて、その個人について容易に定量化することができない。この理由の1つは、GWA解析によって識別された疾患（又は他の表現型特性）と変異の関連性が、臨床解析が行われる解析対象の母集団のコンテキストにおいて識別されたことである。例えばGWA解析が主に中央ヨーロッパ系の母集団において行われた場合、SNPの関連付けは、中央ヨーロッパ系の患者についてのみ関係する可能性がある。アフリカやアジア系の人々のような他の母集団は、SNPに関連する疾患が属するのと同じ経路の他の遺伝子に補正的変異（compensatory variation）を有する可能性があり、これにより、疾患の関連付けの有意性は低下する。同様に、GWA解析は、単一の性別の対象又は特定の年齢階層内等に制限されることがある。解析対象の母集団におけるこれらの制約が、実際に変異と疾患の関連付けに影響を与えるかどうかに関わらず、特定の患者は容易に確認されない。例えば所与の臨床解析が、中央ヨーロッパ系の母集団に対して実行されたとしても

40

50

、その結果として得られる変異の関連性は、それでもなお、他の民族の人々と関連すること、あるいは、そのような母集団と全く関連しないか、その関連は他の民族の人々に対する関連よりも弱いことがある。

【 0 0 2 1 】

典型的に、解析結果の公表は、その解析が行われた母集団のプールに関する人口統計情報を含む。これに基づいて、患者の医師は、臨床解析結果その患者に適用可能であるかどうかを判断することができ、したがって、その患者に検査をオーダーするかどうかを決めることができる。しかしながら、通常は、医師がこの決定を行うための理にかなった基準はない。一見小さな人口的な差異が、実際には、患者に対して適用可能でない検査となることがあり、一方、一見顕著な人口的な差異が、実際には患者に対する適用性にほとんど又は全く影響を与えないこともある。さらに、患者に対する検査の適用可能性に実際に影響する患者と解析の母集団との差異は、解析結果の公表には記録されないことがある。

10

【 0 0 2 2 】

したがって、患者が、解析対象の母集団のプールから逸脱するように見えるとき、医師には、異なる選択肢が残される：すなわち、誤った結果を得るリスクの下で、解析結果を患者に適用するか；あるいは、証拠となる医療情報を取得しないというリスクの下で、解析結果を患者に適用しない、という選択肢が残される。

【 0 0 2 3 】

これらの問題についての解決策が本明細書において提示される。これらのアプローチでは、患者が、（その臨床解析において使用される母集団において、その関連性が評価されている）解析変異に機能的に関連する類似の変異分布を有する場合、解析変異と疾患の関連付けは、その患者において顕著なものとなる可能性が高く、したがって、その患者に対して関連する医療情報を提供する。

20

【 0 0 2 4 】

本明細書で使用される用語に関する一部の解説は次の通りである。

【 0 0 2 5 】

「診断対象」という用語は、本明細書において、患者、獣医学的对象、法医学的对象（例えば検視解剖を受ける死体）、考古学的対象（例えばミイラ又は古い人骨）等を示すのに使用され、これらの対象者から（又は対象物から）診断対象の遺伝子データが取得され、その診断対象における疾患又は他の表現型特性を識別する目的で、臨床解析の結果と比較される。この識別処理は、臨床解析により疾患又は他の表現型特性と関連付けられる遺伝子変異を診断対象において観察し、次いで、その関連付けについて、診断対象に対する実際の関連性を、本明細書で開示されるアプローチを使用して評価することにより行われる。例示の実施形態では、診断対象は患者であり、そのようなものとして言及されるが、例示の実施形態は、他のタイプの診断対象（例えば獣医学的对象、法医学的对象等）に容易に適用される。臨床解析の対象は、典型的に、診断対象と同じ「タイプ」である。したがって、患者、人間の死体等に適用される臨床解析は、人間の母集団を用い、一方、イヌという獣医学的对象に適用される臨床解析は、イヌを対象とする母集団を用いる。考古学的診断対象の場合、適用される臨床解析の母集団のプールは、現代生活の母集団を備えてもよく、あるいは（他の）考古学的遺物の母集団を備えてもよい。

30

40

【 0 0 2 6 】

本明細書で使用されるとき、「多型 (polymorphism)」という用語は、2つ又はそれ以上の異なる潜在的な「変異」を有するサブ配列（その「位置」）を指示するのに使用される。例えば一塩基多型 (SNP) は、染色体 x の位置 y によって指定されるが、変異は、染色体 x の位置 y におけるヌクレオチド「A」（アデニン）として指定されてよい。多型は、母集団において観察される2つ又はそれ以上の変異を有する遺伝子サブ配列である。典型的に、WGA解析は、遺伝子サブ配列が、母集団内で生じる2つ又はそれ以上の異なる潜在的なヌクレオチド（変異）を有するゲノム内の単一のヌクレオチドの位置である、一塩基多型 (SNP) を識別する。例えば所与のSNPは通常、ヌクレオチド「A」（アデニン）を有するが、一部の個体はヌクレオチド「G」（グアニン）を有するか、まれに

50

ヌクレオチド「C」(シトシン)を有することがある。この例示的なSNPは、母集団内で生じる3つの変異(「A」、「G」及び「C」)を有する。(「ヌクレオチド」という用語と「塩基(base)」という用語は、本明細書では交換可能に使用されることに留意されたい。)

SNPは、臨床解析において最も一般的に特徴付けられる多型であるが、多型は、他の構成を有することもできる。多型の他の(非網羅的な)例には、挿入(遺伝子配列は、挿入される追加のヌクレオチドを有することがある)、削除(遺伝子配列は、省略されるヌクレオチドを有することがある)、コピー数多型(遺伝子配列は、1回又は複数回複製されることがある)、構造的多型(遺伝子配列は、2つ又はそれ以上の可能性ある二次的構造の構成を有する)、メチル化多型(遺伝子配列は、メチル化ヌクレオチドを含むことがある)等が含まれる。本明細書で使用されるとき、多型は、配列多型(sequence polymorphism)、すなわち1つより多くの配列が生じることができる位置も含む。例えばジヌクレオチドの形の変異が、ある場所に存在することができ、あるいは1つの遺伝子が1つ又は複数のコピーとしてゲノム内に存在することができる。本明細書で使用されるとき、「遺伝子多型」又は単に「多型」という用語は、配列(実際にはゲノム内の「位置」)を指し、一方「遺伝子変異」又は単に「変異」という用語は、多型の特定の値(例えばSNPの特有のヌクレオチド、又は配列多型の場合は特有の配列)を指す。このため、例えばSNPは、染色体Cにおけるその位置xによって指定されることが可能であり、前述の例では、潜在的な変異「A」、「C」又は「G」を有することがある。変異は、したがって、その位置と、その位置の値とによって十分に指定される(例えば変異は、染色体Cにおける位置xでヌクレオチド「A」を有するものとして指定され得る)。

【0027】

図1に関連して、診断対象4(すなわち、これらの具体例では患者)の疾患又は他の表現型特性を診断する(すなわち、識別する)ためのシステムを説明する。患者4は、サンプル抽出研究室6内でサンプル抽出プロシージャを受けており、口腔スワブ、生体標本又は他の組織サンプル10(図1では、小瓶によって図形的に示されているが、スライド型又は他の適切な組織サンプルコンテナ若しくはサポートによって適切に行われてよい)を抽出する。組織サンプル10は、遺伝子データ取得装置12によって処理されて、診断対象の遺伝子データセット14(すなわち、図示される例では、患者の遺伝子データセット)を生成する。

【0028】

一部の実施形態において、遺伝子データ取得装置12は、配列決定読取値(sequencing read)を生成するシーケンサと、適切なデータ処理ハードウェア(例えば適切にプログラムされたコンピュータ)を含む遺伝子配列決定システムであり、上記の配列決定読取値を前処理して整列させ(任意選択により、基準配列にマッピングして)、全ゲノム配列(WGS)又はこのゲノムの一部を生成する。患者の遺伝子データセット14を生成するデータ処理は、注釈付けの動作を更に含んでよく、例えば既知の変異にラベル付けをすることを含んでよい。シーケンサ装置は、次世代配列決定(NGS:next generation sequencing)装置や、サンガー配列決定機関のようなより従来の配列決定装置であってよい。シーケンサ装置は、一部の実施形態において、米国のカリフォルニア州サンディエゴのIllumina社、米国のマサチューセッツ州ケンブリッジのKnome社、米国のコネチカット州ギルフォードのIon Torrent社又は他のNGSベンダから入手可能な市販の配列決定装置であってよいが、市販されていないシーケンサやカスタムビルド型のシーケンサも考えられる。

【0029】

あるいはまた、遺伝子データ取得装置12は、マイクロアレイ処理ハードウェアを含んでもよく、このマイクロアレイ処理ハードウェアによって、組織サンプル10を処理してマイクロアレイデータを生成する。従来の構成では、マイクロアレイは、多数のセル(より多くの又はより少ない数のセルも考えられるが、例えば一部の実施形態では600,000個のセル)を含み、その各セルが、特定のたんぱく質の発現量のような、特定の遺伝

10

20

30

40

50

子マーカを測定するための付着物や化学物質等を含む。ここで、本明細書で使用されるとき、「遺伝子データ」という用語は、遺伝子配列決定データに加え、プロテオミクデータ (proteomic data) (例えば遺伝子の発現量) やメチル化データ等を包含するよう広く意図されていることに留意されたい。患者の遺伝子データセット 14 についても、異なるソースからの遺伝子データの組合せ、例えば配列決定装置によって生成される配列決定データと、マイクロアレイ分析からのたんぱく質発現量のような異なるソースからの遺伝子データの組合せも含むことが考えられる。さらに、遺伝子データ取得装置 12 の動作が、いくらかの手動の動作、例えば適切に訓練された人によって行われるサンプルの準備とロード動作を要することがあることも理解されよう。

【0030】

診断分析を実行するために、患者の遺伝子データセット 14 が臨床解析 20 の結果と比較される。これらの臨床解析 20 の結果は、臨床解析 20 により疾患又は他の表現型特性に関連付けられた解析遺伝子変異 22 を含む。臨床解析 20 の結果は、「生データ」、すなわち解析対象について取得した遺伝子データセット 24 も含む。典型的に、臨床解析 20 は、論文審査のある医療雑誌に公表された、公表済みの解析であり、遺伝子データセット 24 は、そのような雑誌により、あるいは医師会や雑誌に関連する他のエンティティにより管理される、公衆にアクセス可能なデータリポジトリにおいて取得可能である。あるいは、臨床解析 20 は、未公表の研究であってもよく、例えば商業的な遺伝子検査サービスにおいて使用するための変異と疾患の関連付けデータを生成するために医療サービス会社によって行われる独自の研究であってもよい。この場合において、対象の遺伝子データ

【0031】

診断分析は、図 1 に図形的に示される動作 32、34、36 を実行するコンピュータ 30 又は他の電子データ処理デバイスによって適切に実行される。動作 32 において、患者の遺伝子データセット 14 を検索して、臨床解析 20 によって疾患 (あるいは、より一般的には表現型特性) と関連性を有するものとして識別される、解析遺伝子変異 22 のいずれかが、患者の遺伝子データセット 14 において観察されるかどうかを判断する。変異の識別動作 32 は、患者の遺伝子データセット 14 の変異の注釈又は他の注釈を、そのような注釈が利用可能である場合には適切に参照する。あるいはまた、サブ配列パターンマッ

【0032】

観察動作 32 の出力は、従来的に、遺伝子検査の出力とみなされる。すなわち、解析変異の観察値は、患者が、関連する疾患 (又は表現型特性) を有することの証拠とみなされる。しかしながら、これは、臨床解析 20 によって決定される変異と疾患の関連付けが、特定の患者 4 について有効であると想定するが、この変異と疾患の関連付けは、その患者が、異なる人口層の患者であるか、あるいはそうでなくとも解析対象の母集団と「著しく」異なる場合は有効でないことがある。このような場合において、変異と疾患の関連付けは、患者 4 にとって有効でないか、あるいは、臨床解析 20 の結果に基づいて想定されるものよりもより狭い範囲で有効なものとなることがある。

【0033】

したがって、図1の診断システムでは、更なる動作34、36が、診断を受けている特定の患者4に対する、変異と疾患の関連付けの関連性を評価するのに実行される。使用されるアプローチは、(1)観察される解析変異に機能的に関連する多型のセットについて、患者の遺伝子データセット14内で観察される変異の「フォアグラウンド」の分布を識別し(動作34)、(2)このフォアグラウンドの分布を、同じ多型のセットについて、対象の遺伝子データセット24内で観察される変異の「バックグラウンド」の分布と比較する(動作36)ことである。この分析は、患者(フォアグラウンド)において機能的に関連する多型のセット内の各多型をとり、「バックグラウンド」の母集団におけるその発現の可能性を計算することによって、合致の近さが決定される。この場合、バックグラウンドの母集団は、多型の臨床的関連性が確立された母集団である。バックグラウンドのセットにおける多型の対立遺伝子の頻度を所与とすると、そのバックグラウンドのセットにおいて患者の多型それぞれの確率を評価することができ、バックグラウンドのセットにおいて患者の多型を有する個体を見つける同時確率(joint probability)を集散的に確立することができる。この確率は、標準的な方法を使用して(例えば線形関数又は同様の単調関数の使用を通じて)近接性スコアに変換される。この同時確率は、解析下にある多型のセット全体について、検討されている多型の独立を仮定するネイティブのベイズアプローチにより、あるいはハプロタイプ関係(haplotype relationship)をキャプチャするベイズネットワークモデルを通して計算される。動作36において、フォアグラウンド分布とバックグラウンド分布の近い合致は、患者4が臨床解析20の対象プールとの近く合致することを示すこととなり、したがって、解析20によって決定される解析変異の疾患との関連付けは、患者4と関連するとみなされる。一方、フォアグラウンド分布とバックグラウンド分布との間の実質的な不一致は、(その差異が、解析変異に機能的に関連する多型に関連するので)患者4が臨床解析20の対象プールとは著しく異なることを示すこととなり、したがって、解析20によって決定される解析変異の疾患との関連付けは、患者4に関連するとみなされない。より一般的に、動作36は、典型的に、フォアグラウンドの分布とバックグラウンドの分布を比較する定量的類似性(又は距離)の指標(一般的に、「比較メトリック」という用語に含まれる)を用い、この比較は関連性の定量化を提供する。この定量化は、(比較メトリックを閾値処理することによって「関連する又は関連しない」という二成分の決定を行うことはできるが、必ずしも二値ではない。動作38において、観察される解析変異(1つ又は複数)を、観察される遺伝子変異それぞれの患者4に対する関連性の指示とともに、列挙することができる。

【0034】

動作34は、観察された解析変異に機能的に関連する多型のセットを入力として必要とする。動作36は、(動作34からの)フォアグラウンド分布とバックグラウンド分布を入力として必要とする。任意選択により、機能的に関連する多型のセットと、バックグラウンド分布との双方が、患者4についての診断動作32、34、36、38を実行する前に各解析変異について決定され得る。これらの情報は患者4とは独立であるので、各解析変異についての機能的に関連する多型のセットとバックグラウンド分布のそのような「オフライン」の決定は、計算効率を改善することができる。したがって、図1の診断システムにおいて、別個のコンピュータ40又は他の電子データ処理デバイスが、解析変異と機能的に関連する多型のセット44を決定する動作42を実行し、多型のセット44について、対象の遺伝子データセットにおいて観察される変異のバックグラウンド分布48を決定する動作46を実行する。一般的に、各解析変異は、機能的に関連する多型の異なるセットを有することがあるので、機能的に関連する多型のセット44及び変異のバックグラウンド分布48は、各解析変異について計算される。一般に、多型のセット44は、例えばSNPや配列多型(例えば遺伝子コピー、マルチヌクレオチド変異を含む位置等)のような任意のタイプの多型を含んでよい。しかしながら、計算の効率性のため、一部の実施形態では、多型のセット44はSNPに限定することが考えられる。

【0035】

単なる検討として、動作42、46のオフラインの実行、すなわち、各解析変異につい

10

20

30

40

50

ての機能的に関連する多型のセット44とバックグラウンド分布48とを、患者診断動作32、34、35の開始の前に予め決定することは、説明される商業的な医療サービスプロバイダの場合のように、患者についての大きなスループットが予測されるとき、計算的に効率的である。しかしながら、動作42、46が患者の診断の時に実行されることも考えられる。同様に、別個のコンピュータ30、40が患者の診断動作32、34、36と、患者に特有ではない事前計算動作42、46を実行するように示されているが、これらの動作の双方のセットについて同じコンピュータを用いることも考えられる。(言い換えると、一部の実施形態では、2つの図示されるコンピュータ30、40が、全ての動作32、34、36、42、46を実行する単一のコンピュータであってよい。)

図2~図4に関連して、様々なアプローチを使用して、機能的に関連する多型のセット44を識別する動作42を実行することができる。

【0036】

図2に関連して、1つのアプローチでは、動作42は、解析遺伝子変異が関連付けられる遺伝経路を識別し、識別された遺伝経路に関連付けられる多型を識別する。このために、臨床的関連性を有する各解析変異50について、動作52において、その解析変異を含む遺伝子が識別される。遺伝子識別は、例えば注釈付けされた基準遺伝子データベース54を参照することによって実行することができる。一部の実施形態において、遺伝子配列決定動作は、患者の遺伝子データセット14に、遺伝子識別情報で注釈を付けることを含み、この場合、動作52は適切に省略される(これらの実施形態では、この動作は、取得装置12によって実行されるポスト取得データ処理動作に包含される)。動作56において、解析変異を含む遺伝子によって影響を受ける遺伝経路(複数可)が識別される。動作56は、経路データベース58に適切にアクセスして、この識別処理を行う。例えば経路データベース58は、<http://www.genome.jp/kegg/>(最終アクセス2012年5月4日)から入手可能な、K e g g (Kyoto Encyclopedia of Genes and Genomes) 経路データベースとすることができる。動作60において、解析変異を含む遺伝子によって影響を受ける遺伝経路(複数可)に関連付けられる全ての遺伝子が識別される。動作60は、経路データベース58に再びアクセスして、この識別処理を行うことができる。動作62において、解析変異を含む遺伝子によって影響を受ける遺伝経路(複数可)に関連付けられる遺伝子内の全ての多型が識別され、これらが、その解析変異に機能的に関連するとみられる多型のセット44を形成する(又はそのセット内に含まれる)。動作50、52、56、60、62は、臨床解析20によって識別される解析変異のセット22の各解析変異について繰り返し行われる。

【0037】

図3に関連して、別のアプローチでは、動作42は、各解析変異70について、その解析変異と同じ臨床的関連性を有する他の複数の遺伝子変異を識別する、識別動作72を実行する。同じ臨床的関連性を有するこれらの他の遺伝子変異を、同じ解析20で識別することができる、あるいは任意選択により他の臨床解析によって識別することもできる。例えば解析遺伝子変異が、解析20によって眼球の黄斑縮退(ocular macular degeneracy)に関連付けられる場合、次いで、動作72は、眼球の黄斑縮退に関連付けられるものとして識別された他の遺伝子変異を識別する。動作72によって識別されるこれらの変異に対応する多型は、次いで、その解析変異に機能的に関連するとみられる多型のセット44を形成する(又はそのセット内に含まれる)。動作70、72は、臨床解析20によって識別される解析変異のセット22の各解析変異について繰り返し行われる。

【0038】

図4に関連して、別のアプローチでは、動作42は、各解析変異80について、その解析変異を含む疾患シグネチャを識別する、識別動作82を実行する。識別動作82では適切に、疾患シグネチャデータベース84又は疾患シグネチャを規定する他の文献を参照する。解析変異を含む疾患シグネチャの他の変異に対応する多型は、次いで、その解析変異に機能的に関連するとみられる多型のセット44を形成する(又はそのセット内に含まれる)。動作80、82は、臨床解析20によって識別される解析変異のセット22の各解

10

20

30

40

50

析変異について繰り返し行われる。

【0039】

図2～図4に関連して説明した動作42を実行するための具体的な例の適切なアプローチは単なる例示であり、他のアプローチを使用して、解析変異に機能的に関連するとみられる多型のセット44を識別することができる。さらに、図2～図4のアプローチを様々な組み合わせることができる。例えば動作42は、図2の経路分析アプローチを用いて、幾つかの機能的に関連する多型を識別することができ、図4のように解析変異を含む既知の疾患シグネチャを参照して、機能的に関連する更なる多型を識別することができ、そして、機能的に関連する多型の最終的なセット44は、それぞれ図2のプロセスと図4のプロセスという2つのプロセスの出力の組合せである。

10

【0040】

図1を再び参照し、更に図5も参照すると、解析対象の母集団において機能的に関連する変異のバックグラウンド分布48を生成する動作46は、様々なアプローチを用いることができる。図5の具体的な例では、各解析変異90について、(動作42によって、例えば図2～図4の方法のうちの1つ又は複数を使用して生成された)その解析変異に機能的に関連する多型のセット44が取り出される。動作92では、各解析対象について、対応する遺伝子データセットが、解析対象についての1組の遺伝子データセット24から取り出され、多型のセット44について、対象内で観察される変異を保持しているベクトル要素を有する、特徴ベクトルが生成される。

【0041】

様々なタイプのフォーマット化を、ベクトル要素の値に使用することができる。SNPの場合、その値を実際のヌクレオチド(又はヌクレオチドと等価な数字、例えば「0」=アデニン、「1」=シトシン、「2」=グアニン、「3」=チミン)として保持してよい。あるいは、SNP値を、バイナリ値として保持してもよく、例えばSNPが通常の基本値を有する場合には「0」とし、SNPが通常の基本値以外の基本値を有する場合は「1」として保持してもよい。他のエンコードを使用して、多型の様々な他のタイプの値を保持することができる。全ての解析対象の特徴ベクトルが、同じフォーマットの同じベクトル要素を有するべきである。例えばベクトル要素7が、ある解析対象について、xというSNP値を保持する場合、その後、同じベクトル要素7は、全ての他の解析対象についてSNP値xを保持するべきである。

20

【0042】

一部の対象の遺伝子データセットが不十分であり、全ての対象について多型のセット44の必ずしも全ての多型を決定することができない場合、デフォルトの「充填(filler)」値を使用して、値が利用可能でないことを示すことができ、あるいは最も確からしい値(例えば臨床解析22の母集団において最も頻繁に生じている値)によりその値を充填することができる。

30

【0043】

動作92の出力は、臨床解析22の対象についての特徴ベクトルのセットである。例えば臨床解析22がN個の対象を含んでいた場合、動作92は、対応するN個の特徴ベクトルを出力する。多型のセット44のうちの全ての多型が、多型毎に1つのベクトル要素を有する特徴ベクトル内にエンコードされる場合、多型のセット44がR個の多型を含む場合には、特徴ベクトルはR次元を有する。バックグラウンド分布48は、したがって、R次元のN個の特徴ベクトルのセットである。一部の実施形態において、これはN×Rの行列として(あるいはR×Nの行列として)保持される。

40

【0044】

図1を再び参照し、更に図6も参照して、患者関連性評価の動作34、36の例示的な実施形態を説明する。患者関連性評価は、臨床解析20に従う疾患(又は表現型特性の)関連付けを有する患者の遺伝子データセット14内で観察される各変異100について実行される。解析変異に関連する多型のセット44が取り出され、動作102において、患者について患者特徴ベクトルが生成される。動作102は、変異が患者の遺伝子データセ

50

ット14から取得されることを除いて、動作92(図5を参照されたい)と同じ方法で適切に実行される。患者特徴ベクトルは、解析対象について動作92で生成される特徴ベクトルと同じフォーマットで同じベクトル要素を有するべきである。例えばベクトル要素7が、解析対象について特徴ベクトル内にSNP値xを保持する場合、その後、患者特徴ベクトルの同じベクトル要素7は、その患者についてSNP値xを保持するべきである。結果として得られる患者特徴ベクトルは、その患者において、機能的に関連する変異のフォアグラウンド分布104として適切に機能する。

【0045】

動作106において、フォアグラウンド分布104とバックグラウンド分布48を比較する比較メトリックが計算される。直接的なアプローチでは、患者特徴ベクトルと各解析対象特徴ベクトルとの間の距離(例えばユークリッド距離)を計算してその距離を合計し、任意選択により対象特徴ベクトルの数(N)によって正規化して、比較メトリックを生成する。

10

【0046】

別のアプローチでは、バックグラウンド分布に対するフォアグラウンド分布の有意性を示す統計的有意性メトリック(statistical significance metric)が計算される。例えば統計的有意性メトリックは、フィッシャー直接検定、ks検定又はウィルコクソンの符号順位検定とすることができる。統計的有意性メトリックによって計算されるp値は、適切な比較メトリックを提供する。

【0047】

20

距離メトリック又は統計的有意性メトリックのいずれかを使用して、距離又はpメトリックのより低い値は、バックグラウンド分布に対するフォアグラウンド分布のより近い類似性を示す。より近い類似性は、統計的意味では、患者への解析変異の関連性を示すと予想されるパラメータ(解析変異に関連する多型44)に関して、解析対象により類似している患者に対応する。したがって、距離又はpメトリックのより低い値は、患者への解析変異の高い関連性を示し、距離又はpメトリックのより高い値は、患者への解析変異の低い関連性を示す。

【0048】

他の比較メトリックも考えられる。例えば別のアプローチでは、対象の特徴ベクトルをクラスタ化して複数のクラスタを規定し、次いでクラスタからの患者の特徴ベクトルの特異性を測定する異常値メトリックを計算する。この場合、より高い異常値メトリックは、クラスタからのより大きい特異性を示し、これは患者への解析変異の関連性が低いことに対応する。

30

【0049】

動作108において、患者で観察される遺伝子変異のリストが、臨床的関連性及び患者との関連性の指示とともに提示される(例えばディスプレイデバイス上で表示されるか、かつ/又はプリンタ若しくは他のマーキングエンジンを使用して紙に印刷される)。この出力は様々な形式をとることができる。1つのアプローチでは、変異は、最も関連する変異がリストのトップにあるように、関連性に従ってランク付けされる。そのようなリストは、任意選択により切り捨てられる。すなわち、患者の関連性が特定の閾値未満の変異は全くリストされない。距離又は統計的有意性の比較メトリックについて、より高い値の距離又は統計的有意性のメトリックが、診断対象に対する解析変異のより低い関連性を示すように、距離又は統計的有意性のメトリックの逆として関連性が定量化されてよい。様々な正規化又は他の調整を実行して、医師又は他の検討を行う医療関係者によって容易に理解される、定量的な関連性指標を提供することができる。

40

【0050】

別のアプローチでは、関連性が特定の閾値を超える変異のみがリスト内を含まれ、定量的な患者関連性の値は示されない。この場合、リスト上に変異が含まれていることが、その患者の関連性が閾値を超えていることを示す限りにおいて、患者の関連性はリスト内に暗黙的に含まれる。

50

【 0 0 5 1 】

図 1 の診断用遺伝子分析システムを、様々に実装することができる。あるアプローチでは、診断用遺伝子分析システムが、ウェブベースのサービスとして提供される。医師は患者の遺伝子データセット 1 4 をアップロードし、ウェブサイトは、処理コンポーネント 3 0、4 0 を具現化し、かつ関連する臨床解析データ 2 2、2 4 を保持するか臨床解析データ 2 2、2 4 へのアクセスを有するサーバを含む。サーバは、変異の識別処理を実行し、本明細書で開示される関連性評価動作を実行し、リスト 1 0 8 に基づいて患者レポート又は何らかの同様の出力を定式化し、患者のレポートを医師に返す。

【 0 0 5 2 】

別のサービスパラダイムでは、このサービスがメールを介して提供され、遺伝子データの取得を含む。この場合、医師は組織サンプル 1 0 を研究室にメール、配達業者又は別の便利な方法で送信し、研究室は、処理コンポーネント 3 0、4 0 を含み、関連する臨床解析データ 2 2、2 4 を保持するか臨床解析データ 2 2、2 4 へのアクセスを有し、そして遺伝子データ取得装置 1 2 も含む。全ての動作が、サービス研究室で行われ、リスト 1 0 8 に基づく患者レポート又は何らかの同様の出力が、メール、配達業者等により医師に返される。

10

【 0 0 5 3 】

例示の実施形態では、単一の臨床解析 2 0 が図示されている。これを、2 又はそれ以上の異なる臨床解析によって識別される変異を用いる診断処理に拡張することができる。20
が容易に認識されよう。そのような場合において、各変異についてのバックグラウンド分布は、変異と疾患の関連付けを公表した、臨床解析についての調査対象の母集団に対して計算される。

20

【 0 0 5 4 】

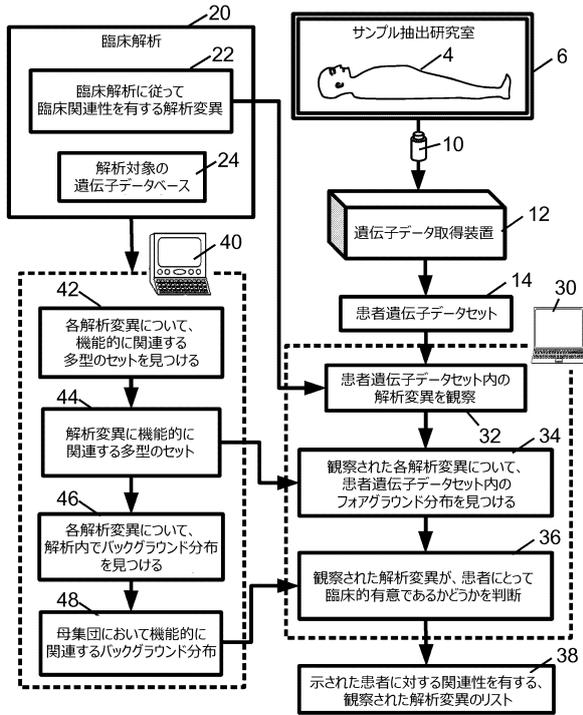
既に示されたように、処理コンポーネント 3 0、4 0 は、任意選択により、単一の処理コンポーネント（例えばネットワークサーバ、デスクトップコンピュータ、ノートブックコンピュータ、コンピュータのネットワークを備える「クラウド」コンピュータ等といった単一のコンピュータ）として統合されてもよい。加えて、開示される変異の識別及び関連する評価技術を、1 つ又は複数の適切な電子データ処理デバイス 3 0、4 0 によって実行可能な命令を格納する非一時的記憶媒体として具現化することができる。非一時的記憶媒体には、例えばハードディスクドライブ又は他の磁気記憶媒体；ランダムアクセスメモリ（RAM）、読取専用メモリ（ROM）、フラッシュメモリ又は他の電子記憶媒体；光ディスク又は他の光記憶媒体；その様々な組合せ等が含まれ得る。

30

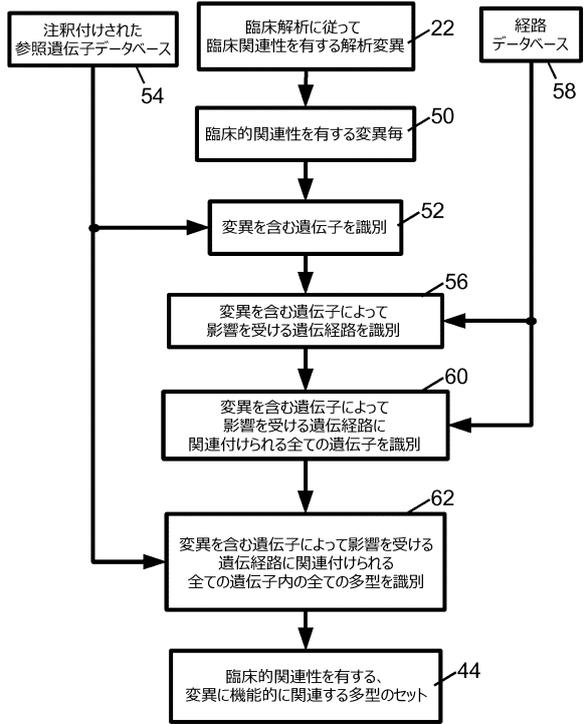
【 0 0 5 5 】

本発明は、好ましい実施形態との関連で説明されてきた。前述の詳細な説明を読んで理解することにより、明らかに、修正及び変更が他者には思い浮かぶであろう。本発明は、そのような修正及び変更が添付の特許請求の範囲又はその均等な範囲内にあるときに、これらの修正及び変更を全て含むものとして解釈されるように意図される。

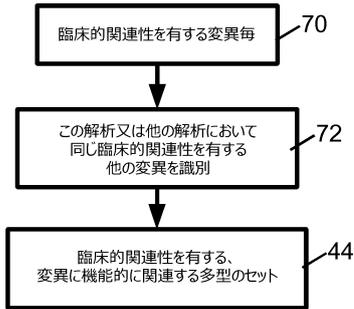
【図1】



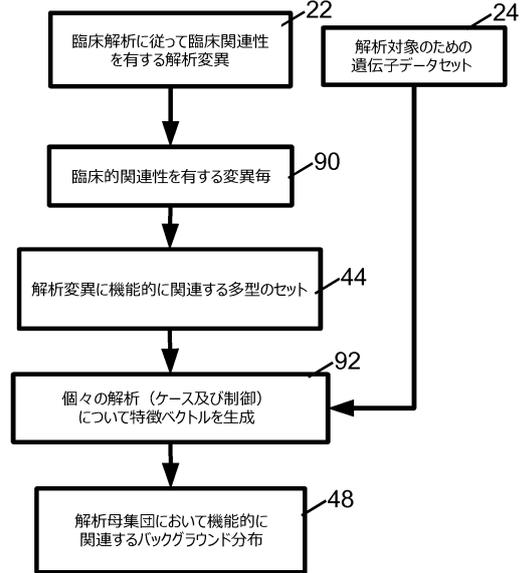
【図2】



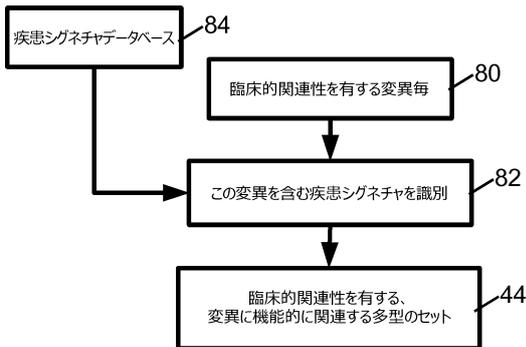
【図3】



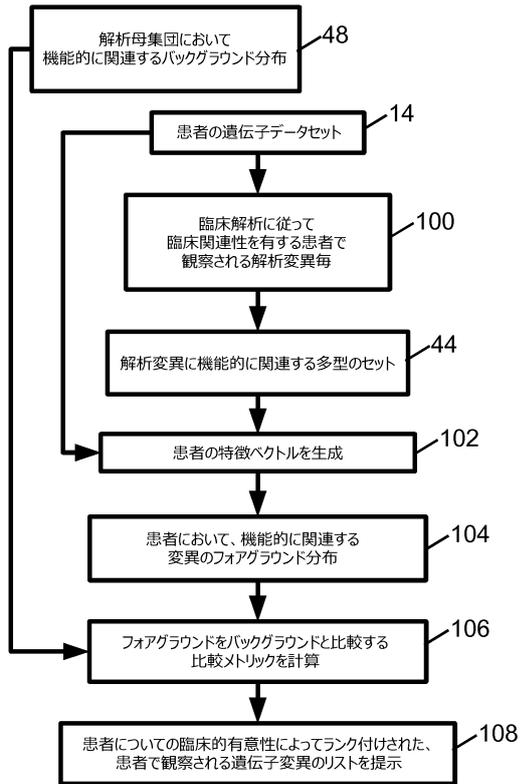
【図5】



【図4】



【 図 6 】



フロントページの続き

- (74)代理人 100091214
弁理士 大貫 進介
- (72)発明者 カマラカラン, シタールトン
オランダ国, 5 6 5 6 アーエー アインドーフエン, ハイ・テク・キャンパス・ビルディング
5
- (72)発明者 ヴアラダン, ヴィナイ
オランダ国, 5 6 5 6 アーエー アインドーフエン, ハイ・テク・キャンパス・ビルディング
5
- (72)発明者 パネルジェー, ニランジャナ
オランダ国, 5 6 5 6 アーエー アインドーフエン, ハイ・テク・キャンパス・ビルディング
5
- (72)発明者 ヤネフスキー, エンジェル
オランダ国, 5 6 5 6 アーエー アインドーフエン, ハイ・テク・キャンパス・ビルディング
5
- (72)発明者 ディミトロワ, ネヴェンカ
オランダ国, 5 6 5 6 アーエー アインドーフエン, ハイ・テク・キャンパス・ビルディング
5

審査官 宮久保 博幸

- (56)参考文献 特表2004-500836(JP,A)
特表2003-521024(JP,A)
特開2007-004211(JP,A)
特開2007-102709(JP,A)
特表2012-502398(JP,A)
Carbonell, J., A map of human microRNA variation uncovers unexpectedly high levels of
variability, Genome medicine, 2012年 8月24日, Vol.4, No.8, 62

- (58)調査した分野(Int.Cl., DB名)
- | | |
|---------|-----------|
| G 0 6 F | 1 9 / 1 8 |
| C 1 2 Q | 1 / 6 8 |
| G 0 6 F | 1 9 / 2 4 |