

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第4437050号
(P4437050)

(45) 発行日 平成22年3月24日(2010.3.24)

(24) 登録日 平成22年1月8日(2010.1.8)

(51) Int.Cl.

F I

G 0 6 F 19/00 (2006.01)

A 6 1 B 5/00 (2006.01)

G 0 6 Q 50/00 (2006.01)

G 0 6 F 19/00 6 0 0

A 6 1 B 5/00 G

G 0 6 F 17/60 1 2 6 G

請求項の数 5 (全 18 頁)

(21) 出願番号 特願2004-91104 (P2004-91104)
 (22) 出願日 平成16年3月26日(2004.3.26)
 (65) 公開番号 特開2005-276022 (P2005-276022A)
 (43) 公開日 平成17年10月6日(2005.10.6)
 審査請求日 平成18年4月19日(2006.4.19)

(73) 特許権者 000005108
 株式会社日立製作所
 東京都千代田区丸の内一丁目6番6号
 (74) 代理人 110000350
 ポレール特許業務法人
 (74) 代理人 100068504
 弁理士 小川 勝男
 (74) 代理人 100086656
 弁理士 田中 恭助
 (74) 代理人 100094352
 弁理士 佐々木 孝
 (72) 発明者 斎藤 聡
 埼玉県比企郡鳩山町赤沼2520番地 株
 式会社日立製作所 基礎研究所内

最終頁に続く

(54) 【発明の名称】 診断支援システム、診断支援方法および診断支援サービスの提供方法

(57) 【特許請求の範囲】

【請求項 1】

複数の個体の診療情報を格納する診療情報データベースと、集団の遺伝子多型に関する情報を格納する遺伝子多型情報データベースと、前記遺伝子多型情報データベースの情報を基礎として前記集団のハプロタイプブロックおよび前記各ハプロタイプブロック内におけるハプロタイプ頻度を推定するハプロタイプブロック推定処理プログラムと、前記集団の推定された前記各ハプロタイプブロック内におけるハプロタイプパターンおよび前記ハプロタイプ頻度を格納するハプロタイプ情報データベースと、前記ハプロタイプ情報データベースに格納された、異なる前記ハプロタイプパターン間の距離を計算し、該距離の近いパターンを同一の亜集団として前記集団を複数の亜集団に分割し、前記各個体の前記各亜集団への帰属度を計算する遺伝的構造推定処理プログラムと、分割された前記亜集団ごとの前記ハプロタイプ情報および前記各個体の前記各亜集団への帰属度情報を格納する遺伝的構造情報データベースと、前記遺伝的構造情報データベースに格納された前記各亜集団ごとに、前記各ハプロタイプブロックにおける前記ハプロタイプパターンと前記診療情報データベースに格納された前記各個体の診療情報との関連性を解析する関連性解析処理プログラムと、前記関連性解析処理プログラムによって得られた情報を格納する診断支援知識データベースと、前記診断支援知識データベースの情報を基礎として所定の個体が疾患に罹患する危険度を算出する罹患危険度算出処理プログラムとを有することを特徴とする診断支援システム。

【請求項 2】

前記距離が、各ハプロタイプ間における組み換えおよび突然変異の起こりやすさにより定義される請求項 1 に記載の診断支援システム。

【請求項 3】

遺伝子多型に関する情報を格納する遺伝子多型情報データベースの情報を基礎としてハプロタイプブロックおよび各ハプロタイプブロック内におけるハプロタイプ頻度を推定するステップ、推定された前記各ハプロタイプブロック内におけるハプロタイプパターンおよびハプロタイプ頻度をハプロタイプ情報データベースに格納するステップと、前記ハプロタイプ情報データベースに格納された、異なる前記ハプロタイプパターン間の距離を計算し、該距離の近いパターンを同一の亜集団として前記集団を複数の亜集団に分割し、前記各個体の前記各亜集団への帰属度を計算する遺伝的構造を推定するステップと、分割された前記亜集団ごとの前記ハプロタイプ情報および前記各個体の前記各亜集団への帰属度情報を遺伝的構造情報データベースに格納するステップと、前記遺伝的構造情報データベースに格納された前記各亜集団ごとに、前記各ハプロタイプブロックにおける前記ハプロタイプパターンと複数の個体の診療情報を格納する診療情報データベース内の前記各個体の診療情報との関連性を解析する関連性解析ステップと、前記関連性解析ステップで得られた情報を診断支援知識データベースに格納するステップと、前記診断支援知識データベースの情報を基礎として所定の個体が疾患に罹患する危険度を算出する罹患危険度算出ステップとよりなることを特徴とする診断支援方法。

10

【請求項 4】

請求項 3 に記載の診断支援方法であって、

20

前記距離が、各ハプロタイプ間における組み換えおよび突然変異の起こりやすさにより定義されることを特徴とする診断支援方法。

【請求項 5】

複数の個体の診療情報を格納する診療情報データベースと、遺伝子多型に関する情報を格納する遺伝子多型情報データベースと、前記遺伝子多型情報データベースの情報を基礎としてハプロタイプブロックおよび前記各ハプロタイプブロック内におけるハプロタイプ頻度を推定するハプロタイプブロック推定処理プログラムと、推定された前記各ハプロタイプブロック内におけるハプロタイプパターンおよび前記ハプロタイプ頻度を格納するハプロタイプ情報データベースと、前記ハプロタイプ情報データベースに格納された、異なる前記ハプロタイプパターン間の距離を計算し、該距離の近いパターンを同一の亜集団として前記集団を複数の亜集団に分割し、前記各個体の前記各亜集団への帰属度を計算する遺伝的構造推定処理プログラムと、分割された前記亜集団ごとの前記ハプロタイプ情報および前記各個体の前記各亜集団への帰属度情報を格納する遺伝的構造情報データベースと、前記遺伝的構造情報データベースに格納された前記各亜集団ごとに、前記各ハプロタイプブロックにおける前記ハプロタイプパターンと前記診療情報データベースに格納された前記各個体の診療情報との関連性を解析する関連性解析処理プログラムと、前記関連性解析処理プログラムによって得られた情報を格納する診断支援知識データベースと、前記診断支援知識データベースの情報を基礎として所定の個体が疾患に罹患する危険度を算出する罹患危険度算出処理プログラムとを有する診断支援システムに接続して受けることができる診断支援サービスの提供方法であって、

30

40

診断支援サービスを受けるものは、被診断者の個体から受け取った前記所定の個体遺伝子型データおよび診療情報を診断支援システムに送信して、

診断支援システムは、前記集団に存在する遺伝的構造に関する情報と、前記所定の個体の前記各亜集団への帰属度と、前記所定の個体が疾患に罹患する危険度とを算出して前記診断支援サービスを受けるものに提供することを特徴とする診断支援サービスの提供方法。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、臨床情報と遺伝子情報との関連性を解析し、臨床上有用な情報を抽出して提

50

示する診断支援システムおよび診断支援方法に関する。

【背景技術】

【0002】

ヒトゲノムプロジェクトは配列決定をほぼ終了し、ポストシーケンスの時代に突入した。今後は、蓄積された膨大な遺伝子情報の医学への有効利用が期待されている。遺伝子と疾患との関連性の解明が進むと、個人の遺伝子型を基礎として疾患の発症リスクを予測することが可能になり、個人の遺伝的体質に応じた疾患の予防、早期発見、治療を行なうことが可能になる。これらを実現するためには、臨床情報と遺伝子情報との関連性の解析が必要である。

【0003】

臨床情報と遺伝子情報との関連性の解析において強力な手法の一つに、遺伝統計学的解析法がある。遺伝統計学的解析法は、個人の遺伝子情報と疾患の有無とをデータとして、統計学を用いて疾患に関連する遺伝子を探索する方法であり、機序が未知である疾患の関連遺伝子も発見できる可能性があるため、次第に重要性を増している。遺伝統計学的解析法は、複数の遺伝子座（染色体上の遺伝子の位置）間の連鎖を利用して特定の形質に関係する遺伝子領域を探索する技術である。形質とは、個体レベルで観察される各種の形態的特徴のことで、疾患罹患の有無、身長、目や髪の色等が形質である。連鎖とは、「2つの異なる形質はそれぞれ分離独立して遺伝する」というメンデルの独立法則の例外である。

【0004】

ある2つの形質を規定する遺伝子座が染色体上で近い位置に存在しているとき、それらの遺伝子は分離独立せずに、つながったまま親から子へ遺伝する。この状態を、2つの遺伝子座が連鎖しているという。減数分裂の際に、両親から伝わった1対の染色体間に部分的な交換が起こり、子に伝える遺伝子の組み合わせが両親由来のものと異なる場合がある。この現象を組み換えという。

【0005】

1回の減数分裂においてある2つの遺伝子座間に組み換えが起こる確率を組み換え割合という。2つの遺伝子座間の距離が近いほど、組み換え割合は小さい。すなわち、連鎖する可能性が高い。遺伝統計学的解析法では、組み換えの情報を基礎として染色体上に網羅された遺伝子多型（一塩基多型（SNPs）やマイクロサテライト等）と疾患関連遺伝子との連鎖の有無を検定することによって、疾患関連遺伝子座を絞り込む。

【0006】

遺伝統計学的解析法は現在までにいくつかの手法が報告されている。単一遺伝子疾患については、大家系のデータを用いたパラメトリック連鎖解析によって、これまでに数多くの原因遺伝子が同定されてきた。今後の疾患原因遺伝子探索の研究においては、複数の遺伝的要因と環境要因とによって発症する多因子疾患（complex disease）の原因遺伝子の探索が主流になるものと考えられる。当初は、多因子疾患の原因遺伝子についても、多数の小家系のデータを用いたノンパラメトリック連鎖解析（罹患同胞対解析）によってその同定が可能と考えられていた。しかし、一般に浸透率（発症する確率）の低い多因子疾患の原因遺伝子を直接同定することは困難な場合が多い。最近では、その検出力の高さと解析のしやすさから、疾患集団と正常集団において着目する遺伝子多型の対立遺伝子（アレル）頻度を比較する相関解析（関連解析ともいう）が注目されている。

【0007】

従来の相関解析では、真に形質と関係している遺伝子を見落とししたり、目的とする形質とまったく関係がない遺伝子を誤って選択したりする可能性が比較的高かった。一般に、前者は偽陰性の問題、後者は偽陽性の問題として取り扱われる問題である。解析結果に偽陰性や偽陽性が生じる理由としては、単一の遺伝子多型もしくは狭い範囲の遺伝子多型によるハプロタイプのみを用いて遺伝子と形質との関係を解析していること、ハプロタイプを用いた解析を行なう際にハプロタイプブロックを考慮していないこと、対象とする集団に存在する多様性（これを遺伝的構造と呼ぶことにする）を考慮していないこと、等が挙げられる。

10

20

30

40

50

【0008】

ハプロタイプとは、連鎖している複数の座位における同じ親由来のアレルの組み合わせのことをいう。染色体上の近い距離に存在する複数の座位におけるアレルは、世代交代における組み換えの影響を受けずに連鎖した状態で次の世代へと伝達される。その結果、何世代にもおよぶ世代交代を経た後も、近い距離に存在する複数の座位にはお互いに相関関係が見られる。この状態を連鎖不平衡と呼ぶ。近年、例えば、非特許文献1（Gabriel SB et al.: The Structure of Haplotype Blocks in the Human Genome, Science, Vol.296, pp.2225 - 2229, 2002）等によって、ゲノム上には連鎖不平衡が比較的強い状態で保たれたハプロタイプブロックとよばれる部分と、組み換えが高頻度で起こるために座位間の連鎖不平衡を弱めるホットスポットとよばれる部分とが交互に存在することが報告されている。

10

【0009】

この事実、ハプロタイプブロックの位置を正確に推定することができれば、ハプロタイプブロック内の数個の座位の遺伝子型を測定するだけで正確なハプロタイプパターンを決定することが可能であることを意味している。また、この事実は同時に、ホットスポットを跨ぐような複数の座位を用いて解析を行なった場合には、遺伝学的には意味がない偽陽性の結果が多く出てしまうことを意味している。

【0010】

一般に、相関解析を行なう際には、注目する形質に応じて対象とする集団を群分けすることが多い。ある集団の中から多数の患者（case）と対照者（control）とをサンプルし、着目するアレルの頻度を患者群と対照者群とで比較し、アレル頻度に有意な差がみられる多型の座位を検出する症例対照研究（case-control study）が最も有名である。症例対照研究においては、患者の集団と対照者の集団とが注目する形質以外は完全にマッチした集団であることが前提となっている。

20

【0011】

しかし、この前提は常に成立しているわけではない。対象とする集団に遺伝的構造が存在する場合は特に問題となる。患者群と対照者群とを遺伝的に異なる、全く別の集団からサンプリングしてしまった場合などには、遺伝的構造が解析結果に大きく影響を与える。簡単な例を挙げて集団の遺伝的構造の影響を説明する。例えば、アメリカで鎌形赤血球症の患者群と対照者群を集めようとする、患者群にはアフリカ由来の人々が多く含まれ、対照者群にはヨーロッパ由来の人々が多く含まれるはずである。遺伝的構造の影響を考えずにこの2つの集団を比較すると、本来アフリカ人とヨーロッパ人とでアレル頻度に差がある多くの座位が、鎌形赤血球症の原因座位として検出されてしまう。このように、集団の遺伝的構造は解析結果に多くの偽陽性を生じさせる。また、集団の遺伝的構造は解析結果に偽陽性だけではなく偽陰性を生じさせることもある。

30

【0012】

【非特許文献1】 Gabriel SB et al.: The Structure of Haplotype Blocks in the Human Genome, Science, Vol.296, pp.2225 - 2229, 2002

【発明の開示】

【発明が解決しようとする課題】

40

【0013】

上述したように、相関解析を行なう際に、対象とする集団に存在するハプロタイプブロックの影響や遺伝的構造の影響を考慮しなかった場合、解析の際に多くの偽陰性や偽陽性が生じ、解析結果に多大な影響を与えるという問題があった。そこで、本発明では、ハプロタイプブロックおよび遺伝的構造の影響を考慮することによって、高精度な診断支援を行なうシステムを提供することを目的とする。

【課題を解決するための手段】

【0014】

本発明の診断支援システムおよび診断支援方法は、ハプロタイプブロック推定手段によって遺伝子多型情報を基礎として組み換えの位置を推定してハプロタイプブロックの位置

50

を推定し、ハプロタイプブロックごとに解析を行なうことによって、個体のハプロタイプパターンを高精度に推定する。推定されたハプロタイプ頻度情報や個体のハプロタイプパターン情報はハプロタイプ情報データベースに格納される。また、遺伝的構造推定手段によって個体のハプロタイプパターンによるクラスタリングを行ない、集団をいくつかの亜集団に分割することによって、集団に存在する遺伝的構造の影響を除去し、臨床情報と遺伝子情報との関連性を高精度に解析することを可能にする。遺伝的構造推定手段によって得られた結果は遺伝的構造情報データベースに格納され、遺伝的構造情報データベースと診療情報データベースとを用いて臨床情報と遺伝子情報との関連性を解析することによって、高精度な診断支援知識の提供が可能となる。臨床情報と遺伝子情報との関連性の解析によって得られた診断支援知識は診断支援知識データベースに格納され、罹患危険度算出手段によって診断支援知識データベースの情報を基礎として所定の個体が疾患に罹患する危険度を算出する。

10

【発明の効果】

【0015】

本発明の診断支援システムおよび診断支援方法は、ハプロタイプブロック推定アルゴリズムによって組み換えの位置を推定してハプロタイプブロックの位置を推定し、ハプロタイプブロックごとに解析を行なうことによって、個体のハプロタイプパターンを高精度に推定することを可能にする。また、遺伝的構造推定アルゴリズムによって個体のハプロタイプパターンによるクラスタリングを行ない、集団をいくつかの亜集団に分割することによって、集団に存在する遺伝的構造の影響を除去し、臨床情報と遺伝子情報との関連性を高精度に解析することを可能にする。

20

【発明を実施するための最良の形態】

【0016】

図1は、本発明の診断支援システムの構成例を示す図である。本発明の診断支援システム111は、いわゆるパソコン等の電子計算機を主体として構成される。システムバス5に処理装置1、メモリ2、入力装置3、表示装置4および外部記憶装置10が接続される。外部記憶装置10内には、複数の個体（被診断者）の診療情報を格納する診療情報データベース11、複数の個体（被診断者）の遺伝子多型に関する情報を格納する遺伝子多型情報データベース12、該遺伝子多型情報データベース12の情報を基礎としてハプロタイプブロックの位置を推定し、ハプロタイプブロックごとに集団のハプロタイプ頻度および個体のハプロタイプパターンを推定して得られたハプロタイプブロックごとに集団のハプロタイプ頻度情報と個体のハプロタイプパターンを格納するハプロタイプ情報データベース14、該ハプロタイプ情報データベース14の情報を基礎として集団の遺伝的構造を推定し、ハプロタイプブロックごとに個体のハプロタイプパターンによるクラスタリングを行ない、集団をいくつかの亜集団に分割するとともに、各個体の各亜集団への帰属度を推定して得られた、分割された亜集団ごとのハプロタイプ情報および各個体の各亜集団への帰属度情報を格納する遺伝的構造情報データベース16、前記診療情報データベース11および遺伝的構造情報データベース16の情報を基礎として、亜集団のハプロタイプブロックごとに個体のハプロタイプパターンと形質との関連性を解析し、疾患に罹患する危険度を算出する関連性解析によって得られた知識を格納する診断支援知識データベース18と、前記遺伝子多型情報データベース12の情報から前記ハプロタイプ情報データベース14の情報を導出するためのハプロタイプブロック推定処理プログラム13、前記ハプロタイプ情報データベース14の情報から前記遺伝的構造情報データベース16の情報を導出する遺伝的構造推定処理プログラム15、前記診療情報データベース11および前記遺伝的構造情報データベース16の情報から前記診断支援知識データベース18の情報を導出する関連性解析処理プログラム17、前記診断支援知識データベース18の情報を基礎として所定の個体が疾患に罹患する危険度を算出する罹患危険度算出処理プログラム19が内蔵される。もちろん、これらの他に、電子計算機としての機能を果たすために必要とされるデータベースおよび処理プログラムが備えられる。

30

40

【0017】

50

ここで、上述のデータベースは集団のデータが扱われるものであり、診断支援知識データベース18の情報は、その集団に対して有効なものである。また、これらのデータベースの内容は診断を受けた人のデータの蓄積により、より充実したものとなる。

【0018】

本発明の診断支援システムは、ハプロタイプブロック推定処理プログラム13によって遺伝子多型情報を基礎として組み換えの位置を推定してハプロタイプブロックの位置を推定し、ハプロタイプブロックごとに解析を行なうことによって、個体のハプロタイプパターンを高精度に推定する。推定されたハプロタイプ頻度情報や個体のハプロタイプパターン情報はハプロタイプ情報データベース14に格納される。また、遺伝的構造推定手段15によって個体のハプロタイプパターンによるクラスタリングを行ない、集団をいくつかの亜集団に分割することによって、集団に存在する遺伝的構造の影響を除去し、臨床情報と遺伝子情報との関連性を高精度に解析することを可能にする。遺伝的構造推定処理プログラム15によって得られた結果は遺伝的構造情報データベース16に格納され、遺伝的構造情報データベース16と診療情報データベース11とを用いて臨床情報と遺伝子情報との関連性を解析することによって、高精度な診断支援知識の提供が可能となる。臨床情報と遺伝子情報との関連性の解析によって得られた診断支援知識は診断支援知識データベース18に格納され、罹患危険度算出処理プログラム19によって診断支援知識データベース18の情報を基礎として所定の個体が疾患に罹患する危険度を算出する。

【0019】

診療情報データベース11には、個体の氏名、住所、生年月日、家族構成等の基本データや、個体の既往歴、家族歴、主訴、所見、検査結果、生活習慣、症状経過、治療経過、薬剤の処方に関する情報等の臨床データや、インフォームドコンセントに関するデータ等を格納する。遺伝子多型情報データベース12には、多型に関する基本情報（位置、測定方法、多型タイプ（SNP、STRP等）、アレル頻度等）や、個体の遺伝子多型測定結果（塩基配列パターン、ホモ、ヘテロ等）や、検査に用いた検体の識別情報や、保存状態等の検体管理データ等を格納する。

【0020】

次に、ハプロタイプブロック推定処理プログラム13について説明する。先に述べたように、ハプロタイプブロック内では比較的強い状態で連鎖不平衡が保たれている。また、例えば、先に述べた非特許文献1に示されるように、ハプロタイプブロック内ではハプロタイプの多様性が比較的小さいことも知られている。ハプロタイプブロックの位置を推定するためには、ゲノム上のある領域における連鎖不平衡の強さを定義する必要がある。

【0021】

一般に、連鎖不平衡の強さは2つの座位間の連鎖不平衡係数 D' を用いて表されることが多い。本発明では、例えば、ある領域における複数の座位の連鎖不平衡係数が次式の条件を満たすような場合に、その領域をハプロタイプブロックとして定義する。

$$\min(|D'|) > 0.8$$

推定したハプロタイプブロックについて、各ハプロタイプブロック内における集団のハプロタイプ頻度および個体のハプロタイプパターンを推定する。個体が持つ2つのハプロタイプの組み合わせのことをディプロタイプ形と呼ぶことにする。遺伝子型データから個体のディプロタイプ形を推定するための手法は現在までにいくつか提案されている。代表的なものとしては、例えば、文献：Excoffier L & Slatkin M: Maximum-likelihood estimation of molecular haplotype frequencies in a diploid population, Mol Biol Evol, Vol. 12, pp. 921-927, 1995に示すようなEMアルゴリズムを用いた手法や、文献：Stephens M et al.: A new statistical method for haplotype reconstruction from population data, Am J Hum Genet, Vol. 68, pp. 978-989, 2001に示すようなPHASE法等がある。

【0022】

EMアルゴリズムを用いて母集団のハプロタイプ頻度と個体のディプロタイプ形とを推定する方法について以下で説明する。いま、 n 個の個体からなるサンプル集団を考える。

この集団において、連鎖する複数のマーカー座位におけるハプロタイプを考え、その母集団における頻度を $F = (F_1, F_2, \dots, F_M)$ とする。 M は可能なハプロタイプの総数である。例えば、マーカー座位が全て SNP 座位である場合、座位数を L とすると $M = 2^L$ である。各個体の、連鎖する複数のマーカー座位における遺伝子型の観察データを $G = (G_1, G_2, \dots, G_n)$ とする。多くの場合、 G_i は不完全データである。したがって、 G_i に対応するディプロタイプ形は 1 つに定まらないことが多い。このような場合は、可能なディプロタイプ形の上の確率分布（これをディプロタイプ分布と呼ぶ）を定義する。個体 i ($i = 1, 2, \dots, n$) について、 G_i に対応するディプロタイプ形を D_{ij} ($j = 1, 2, \dots, m_i$) とする。ここで、 m_i は G_i に対して可能なディプロタイプの数であり、 m_i の最大値は M である。

10

【0023】

図 2 は、母集団のハプロタイプ頻度と個体のディプロタイプ形とを推定するハプロタイプブロック推定処理プログラム 13 の例を示す図である。

【0024】

ステップ 21：まず、可能な M 個のハプロタイプ（それぞれ $H_1, H_2, \dots, H_M$ とする）に対して、ハプロタイプ頻度の初期値 $F^{(0)}$ を与える。ハプロタイプ頻度の合計は 1 である。

【0025】

次に、 $t = 0, 1, 2, \dots$ について、以下のステップ 22～ステップ 25 によって $F^{(t)}$ から $F^{(t+1)}$ を計算する。

20

【0026】

ステップ 22：各ディプロタイプ形 D_{ij} は 2 つのハプロタイプ H_l, H_m によって構成されている。ただし、 $1 \leq l \leq M, 1 \leq m \leq M$ である。母集団のハプロタイプ頻度 $F^{(t)}$ が与えられているとき、 D_{ij} が得られる確率は式 (1) の通りである。

【0027】

【数 1】

$$\Pr(D_{ij}) = \begin{cases} F_l^{(t)^2} & l = m \text{ のとき} \\ 2F_l^{(t)}F_m^{(t)} & l \neq m \text{ のとき} \end{cases} \quad \text{--- (1)}$$

30

【0028】

したがって、遺伝子型の観察データ G_i のもとでの、個体 i のディプロタイプ形が D_{ij} である事後確率 $\Pr(D_{ij} | G_i)$ は、ベイズの定理より、式 (2) となる。

【0029】

【数 2】

$$\Pr(D_{ij} | G_i) = \frac{\Pr(D_{ij})\Pr(G_i | D_{ij})}{\sum_{k=1}^{m_i} \Pr(D_{ik})\Pr(G_i | D_{ik})} = \frac{\Pr(D_{ij})}{\sum_{k=1}^{m_i} \Pr(D_{ik})} \quad \text{--- (2)}$$

40

【0030】

これを全ての j ($j = 1, 2, \dots, m_i$) について計算すれば個体 i のディプロタイプ分布が定まる。これをサンプル集団における全ての個体について適用する。

【0031】

ステップ 23：個体のディプロタイプ分布が定まると、サンプル集団における全ての個体のディプロタイプ分布から母集団のハプロタイプ頻度の期待値を計算することができる。母集団のハプロタイプ頻度の期待値は、式 (3) となる。

【0032】

50

【数 3】

$$E[F_i^{(t)}] = \frac{1}{2n} \sum_{j=1}^n \sum_{k=1}^{m_i} \Pr(D_{jk}|G_j) N_{D_{jk}i} \quad \text{--- (3)}$$

【0033】

ここで、 $N_{D_{jk}i}$ はディプロタイプ形 D_{jk} の中に含まれる H_i の数（すなわち 0, 1, 2 のいずれか）である。

【0034】

ステップ 24：このとき、全体の尤度は、個体ごとの全てのディプロタイプ形の尤度を結合し、さらに全ての個体の尤度を結合することによって、式（4）で表すことができる。

【0035】

【数 4】

$$L(F^{(t)}) = \Pr(G|F^{(t)}) = \prod_{i=1}^n \sum_{j=1}^{m_i} \Pr(D_{ij}) \quad \text{--- (4)}$$

【0036】

ステップ 25： $F^{(t+1)} = E[F^{(t)}]$ として F を更新する。 $L(F)$ の値が収束したか否か判定する。 $L(F^{(t+1)}) - L(F^{(t)}) < \beta$ を満足すれば収束としてステップ 26 に進み、満足しなければ、ステップ 22 に戻り、ステップ 25 まで繰り返す。ここで β は閾値である。

【0037】

ステップ 26：収束した時点での $E[F] = F^{(EM)}$ を母集団におけるハプロタイプ頻度の最尤推定値とし、このときの $\Pr(D|G)$ を母集団におけるハプロタイプ頻度の最尤推定値のもとでの個体のディプロタイプ分布とする。

【0038】

ハプロタイプ情報データベース 14 には、前述したように、遺伝子多型情報データベース 12 の情報を基礎として、ハプロタイプブロックの位置を推定し、ハプロタイプブロックごとに集団のハプロタイプ頻度および個体のハプロタイプパターンを推定して得られたハプロタイプブロックごとに集団のハプロタイプ頻度情報と個体のハプロタイプパターンを格納するとともに、ハプロタイプブロックの設定に必要な基本情報と、各ハプロタイプブロック内におけるハプロタイプパターンおよびハプロタイプ頻度情報とを格納する。

【0039】

図 3 は、ハプロタイプブロックの設定に必要な基本情報の格納データ例を示す図である。例えば、遺伝子 $GENE_1$ については、SNP 多型である多型 POL_1 および多型 POL_2 と STRP 多型である POL_3 がテーブルに登録されており、 POL_1 , POL_2 および POL_3 がハプロタイプブロック HB_1 を構成していることを示している。図 3 に示したデータ以外にも、例えば、ハプロタイプブロックの長さ、ハプロタイプブロックを構成する多型の選択基準（アレル頻度やアミノ酸変異の有無等）、連鎖不平衡係数、ハプロタイプブロックを構成する多型が存在する遺伝子の位置等を格納してもよい。

【0040】

図 4 は、各ハプロタイプブロック内におけるハプロタイプパターンおよびハプロタイプ頻度情報の格納例を示す図である。例えば、ハプロタイプブロック HB_1 内には、ハプロタイプ HT_1 , ハプロタイプ HT_2 , ハプロタイプ HT_3 およびハプロタイプ HT_4 の 4 つのハプロタイプが存在し、各ハプロタイプの母集団における頻度はそれぞれ

10

20

30

40

50

0.50, 0.28, 0.15および0.07であることを示している。

【0041】

図5は、個体ごとのハプロタイプパターンの格納例を示す図である。例えば、個体PERSON_1はハプロタイプブロックHB_1についてはハプロタイプHT_1を2つ有しており（2つのハプロタイプHT_1から構成されるディプロタイプ形を有しており）、そのディプロタイプ形を有している確率が1.00であることを示している。同様に、個体PERSON_1はハプロタイプブロックHB_2については2つのハプロタイプHT_5から構成されるディプロタイプ形（確率0.95）またはハプロタイプHT_5およびハプロタイプHT_6から構成されるディプロタイプ形（確率0.05）を有しており、ハプロタイプブロックHB_mについては2つのハプロタイプHT_Yから構成されるディプロタイプ形（確率1.00）を有していることを示している。

10

【0042】

次に、遺伝的構造推定処理プログラム15について説明する。本発明では、集団の遺伝的構造を推定するために、個体のハプロタイプパターンによるクラスタリングを行ない、集団をいくつかの亜集団に分割する。本発明では、各ハプロタイプ間における突然変異および組み換えの起こりやすさによって決定される距離を新しく定義し、この距離を用いて個体のクラスタリングを行なう。以下では、本発明のクラスタリング手法について述べる。

【0043】

図6は、あるハプロタイプブロック内においてハプロタイプ1～ハプロタイプ5に示す5つのハプロタイプが観察された例を説明する図である。各ハプロタイプ間の距離を計算するために、まず図6に示すようなハプロタイプ進化系統樹を作成する。ハプロタイプ進化系統樹を作成する方法は、例えば、文献：McPeck MS & Strahs A: Assessment of linkage disequilibrium by the decay of haplotype sharing, with application to fine-scale genetic mapping, Am J Hum Genet, Vol. 65, pp. 858 - 875, 1999に示される方法等、現在までにいくつか報告されている。

20

【0044】

本発明では、進化系統樹の枝（edge）が1回の突然変異もしくは1回の組み換えによる進化を表すように進化系統樹を作成する。図6のハプロタイプ1からハプロタイプ5への進化のように、1回の突然変異もしくは1回の組み換えだけでは進化を表すことができない場合は、実際には観察されていない補助的なハプロタイプを挿入して進化系統樹を作成する。図6のハプロタイプ6は、この補助的なハプロタイプの例である。

30

【0045】

次に、作成した系統樹のそれぞれの枝について、その進化が組み換えによるものなのか突然変異によるものなのかを決定する。例えば、図6において、ハプロタイプ1からハプロタイプ4への進化は組み換えによるものと考えられるが、ハプロタイプ1からハプロタイプ2への進化やハプロタイプ1からハプロタイプ3への進化は、突然変異による進化と組み換えによる進化の両方が考えられる。

【0046】

あるハプロタイプH_Sが別のハプロタイプH_Tへと進化したときの尤度は、式（5）で表される。

40

【0047】

【数5】

$$\begin{aligned} \Pr(H_T|H_S) = & \Pr(H_T|H_S, \text{mut.})\Pr(\text{mut.}|\text{mut. or rec.}) \\ & + \Pr(H_T|H_S, \text{rec.})\Pr(\text{rec.}|\text{mut. or rec.}) \end{aligned} \quad \text{---- (5)}$$

【0048】

ここで、mut. は突然変異（mutation）を、rec. は組み換え（recombination）を

50

示す。(式) 5は、あるハプロタイプ H_S が別のハプロタイプ H_T へと進化したときの尤度はその進化が突然変異によるものであると仮定した場合の尤度と組み換えによるものであると仮定した場合の尤度との和で表されることを示している。ここで、ある座位 j における突然変異率を γ_j 、ハプロタイプにおける k 番目のギャップの組み換え割合を θ とすると、 $\Pr(\text{mut.} | \text{mut.} \text{ or } \text{rec.}) = A / (A + B)$ であり、また、 $\Pr(\text{rec.} | \text{mut.} \text{ or } \text{rec.}) = B / (A + B)$ である。ただし、 A は式(6)、 B は式(7)に示す通りである。

【0049】

【数6】

$$A = \sum_j \gamma_j \prod_{i \neq j} (1 - \gamma_j) \quad \text{--- (6)}$$

10

【0050】

【数7】

$$B = \sum_k \theta_k \prod_{i \neq k} (1 - \theta_k) \quad \text{--- (7)}$$

20

【0051】

図6におけるハプロタイプ1からハプロタイプ4への進化のように、ハプロタイプを構成する多型が2つ以上の座位で異なる場合は、その進化が組み換えによるものであることが明らかであり、 $\Pr(H_T | H_S, \text{mut.}) = 0$ である。組み換えによる進化の場合は、例えば、図6のハプロタイプ1からハプロタイプ4への進化の場合、ハプロタイプ1およびハプロタイプ4の右側に共通する部分ハプロタイプG C C C T C T A T上のどのギャップ(両端を含む)において組み換えが起こっても、見かけ上は同じハプロタイプが形成される。そこで、 H_S と H_T とが、 k_0 番目のギャップまでは見かけ上同じアレルから構成されており(これをIBS(identical by state)という)、それ以降の部分で異なっているとすると、組み換えによる進化の場合の尤度は式(8)と表される。

30

【0052】

【数8】

$$\Pr(H_T | H_S, \text{rec.}) = \sum_{k=0}^{k_0} \Pr(H_T | H_S, \text{rec.}, R = k) \Pr(R = k) \quad \text{--- (8)}$$

【0053】

いま、 H_S が L 個の座位によって構成されていることとし、 H_S のうちの座位 $m, m+1, \dots, n$ の部分で構成される部分ハプロタイプを $H_S^{\{m:n\}}$ と表すことにする。 H_T についても同様に表すことにすると、式(9)となる。

40

【0054】

【数9】

$$\begin{aligned} & \Pr(H_T | H_S, \text{rec.}, R = k) \Pr(R = k) \\ &= \Pr(H_T^{1:k} \text{ IBD to } H_S^{1:k}, H_T^{(k+1):L} | H_T^{1:k} \text{ IBS to } H_S^{1:k}) \\ &= \Pr(H_T^{1:k} \text{ IBD to } H_S^{1:k} | H_T^{1:k} \text{ IBS to } H_S^{1:k}) \Pr(H_T^{(k+1):L}) \quad \text{--- (9)} \end{aligned}$$

【0055】

ここで、ある2つのハプロタイプがIBD(identical by descent)であるとは、同祖

50

由来のアレルを共有していることを示す。また、ある2つのハプロタイプが見かけ上はIBSであっても実際にはIBDである場合もあるため、これをIBS*と表すことにする。

【0056】

ベイズの定理を適用すると、式(10)となる。

【0057】

【数10】

$$\left. \begin{aligned} & \Pr(H_T^{1:k} \text{ IBD to } H_S^{1:k} | H_T^{1:k} \text{ IBS to } H_S^{1:k}) \\ &= \Pr(H_T^{1:k} \text{ IBD to } H_S^{1:k}) \\ & \quad / [\Pr(H_T^{1:k} \text{ IBD to } H_S^{1:k}) + \Pr(H_T^{1:k} \text{ IBS* to } H_S^{1:k}) \Pr(H_T^{1:k} | H_T^{1:k} \text{ IBS* to } H_S^{1:k})] \end{aligned} \right\} (10) \quad 10$$

【0058】

ここで、式(11)と仮定することができ、

【0059】

【数11】

$$\Pr(H_T^{1:k} \text{ IBD to } H_S^{1:k}) = \Pr(H_T^{1:k} \text{ IBS* to } H_S^{1:k}) = \frac{1}{2} \quad \text{--- (11)} \quad 20$$

【0060】

式(12)は $H_T^{\{1:k\}}$ の頻度であるから、式(10)の値は容易に計算することができる。

【0061】

【数12】

$$\Pr(H_T^{1:k} | H_T^{1:k} \text{ IBS* to } H_S^{1:k}) \quad \text{--- (12)}$$

【0062】

本発明では、式(5)で表される尤度を各ハプロタイプ間の距離として新しく定義し、この距離を用いて個体のクラスタリングを行なうこととした。したがって、k番目のハプロタイプブロックについて H_{ka_k} 、 H_{kb_k} のハプロタイプをもつ個体と H_{kc_k} 、 H_{kd_k} のハプロタイプをもつ個体との距離 d_k を、式(13)のように定義する。

【0063】

【数13】

$$\left. \begin{aligned} d_k = \frac{1}{8} [& \Pr(H_{kc_k} | H_{ka_k}) + \Pr(H_{ka_k} | H_{kc_k}) + \Pr(H_{kd_k} | H_{ka_k}) + \Pr(H_{ka_k} | H_{kd_k}) \\ & + \Pr(H_{kc_k} | H_{kb_k}) + \Pr(H_{kb_k} | H_{kc_k}) + \Pr(H_{kd_k} | H_{kb_k}) + \Pr(H_{kb_k} | H_{kd_k})] \end{aligned} \right\} (13) \quad 40$$

【0064】

ハプロタイプブロックの数をmとすると、2つの個体間の距離dは、全てのハプロタイプブロックにおける距離を結合して、式(14)となる。

【0065】

【数14】

$$d = \frac{1}{m} \sum_{k=1}^m d_k \quad \text{--- (14)}$$

【0066】

次に、個体の帰属度の推定方法、すなわち、遺伝的構造推定処理プログラム 15 について説明する。本発明では、各個体が、先に述べたクラスタリング手法によって生成された亜集団のうちのどの亜集団に属するかという情報を、個体の帰属度として定義する。

【0067】

図 7 は、個体の帰属度を推定する遺伝的構造推定処理プログラム 15 を示す図である。

【0068】

ステップ 71：図 6 を参照して説明した方法によって、各ハプロタイプごとにハプロタイプ間の距離を決定する。

【0069】

ステップ 72：ハプロタイプ間の距離にもとづくクラスタリングを行なう。

10

【0070】

ステップ 73：ステップ 72 の結果から、n 個の個体からなる集団が N 個の亜集団に分割されたとする。このとき、ある個体 i がある亜集団 j に分類されているとすると、個体 i の亜集団 j への帰属度は 100% であり、個体 i の亜集団 j 以外の亜集団への帰属度は 0% である。ハプロタイプブロックの数を m とすると、全体の尤度は式 (15) と表すことができる。

【0071】

【数 15】

$$L(N) = \prod_{i=1}^n \sum_{j=1}^N \prod_{k=1}^m \Pr(D|G)_{jk}^{(i)} Q_j^{(i)} \quad \text{--- (15)}$$

20

【0072】

ここで、 $\Pr(D|G)$ は個体の最尤ディプロタイプ分布であり、式 (16) はある亜集団 j の k 番目のハプロタイプブロックにおける個体 i の最尤ディプロタイプ分布を示す。

【0073】

【数 16】

$$\Pr(D|G)_{jk}^{(i)} \quad \text{--- (16)}$$

30

【0074】

ステップ 74：L(N) の値が収束したか否か判定する。L(N_{k-1}) - L(N_k) < β を満足すれば収束としてステップ 75 に進み、満足しなければ、ステップ 71 に戻り、ステップ 74 まで繰り返す。ここで β は閾値である。

また、式 (17) は個体 i の亜集団 j への帰属度である。

【0075】

【数 17】

$$Q_j^{(i)} \quad \text{--- (17)}$$

40

【0076】

ステップ 75：式 (15) で表される尤度が最大となるときの N が亜集団数の最尤推定値である。この最尤推定値をパラメータとして採用する。

【0077】

ステップ 76：式 (15) で表される尤度を基礎として個体の各亜集団への帰属度を計算する。例えば、N_{k} 個の亜集団があり、次の連結ステップで亜集団 N_{1} と亜集団 N_{1+1} とが連結されて N_{k-1} 個の亜集団が形成されたとすると、

50

このステップにおいて尤度に変化がなく、かつこのときに尤度が最大となる場合には、亜集団 $N_{\{1\}}$ および亜集団 $N_{\{1+1\}}$ に分類されている全ての個体について、亜集団 $N_{\{1\}}$ および亜集団 $N_{\{1+1\}}$ への帰属度をそれぞれ50%ずつとする。

【0078】

遺伝的構造情報データベース16には、先にも述べたように、各亜集団におけるハプロタイプパターンおよびハプロタイプ頻度情報と、個体ごとの各亜集団への帰属度情報とを格納する。

【0079】

図8は、各亜集団におけるハプロタイプパターンおよびハプロタイプ頻度情報の格納例を示す図である。例えば、亜集団SUBPOP_1および亜集団SUBPOP_2内にハプロタイプブロックHB_1、HB_2がある。ここで、亜集団SUBPOP_1内にはハプロタイプHT_1、ハプロタイプHT_2、ハプロタイプHT_3およびハプロタイプHT_4の4つのハプロタイプが存在し、亜集団SUBPOP_2内には別のハプロタイプHT_7、ハプロタイプHT_8、ハプロタイプHT_9の3つのハプロタイプが存在することを示している。

【0080】

一方、図4を参照して分かるように、例えば、ハプロタイプブロックHB_1内には、ハプロタイプHT_1、ハプロタイプHT_2、ハプロタイプHT_3およびハプロタイプHT_4の4つのハプロタイプが存在し、各ハプロタイプの母集団における頻度はそれぞれ0.50、0.28、0.15および0.07であることを示している。また、ハプロタイプブロックHB_1内に、別のハプロタイプHT_7、ハプロタイプHT_8およびハプロタイプHT_9の3つのハプロタイプが存在し、各ハプロタイプの母集団における頻度はそれぞれ0.34、0.33および0.33であることを示している。

【0081】

図9は、個体ごとの各亜集団への帰属度情報の格納例を示す図である。例えば、個体PERSON_1は亜集団SUBPOP_1への帰属度は1.00（100%と百分率で表記してもよい）であり、個体PERSON_2は亜集団SUBPOP_1への帰属度が0.50（50%）、亜集団SUBPOP_3への帰属度が0.50（50%）であることを示している。

【0082】

次に、関連性解析処理プログラム17によって、診療情報データベース11および遺伝的構造情報データベース16の情報を基礎として、各亜集団のハプロタイプブロックごとに個体のハプロタイプパターンと形質との関連性を解析する手順について説明する。関連性解析処理プログラム17は、特定のハプロタイプを所有する個体の群と所有しない個体の群との間の形質を比較して（例えば、疾患の発症の有無を比較して）両群間のオッズ比等を計算し、特定のハプロタイプを所有する個体の群が特定のハプロタイプを所有しない個体の群と比較してどの程度疾患を発症するリスクが高まるかどうかを推定する。

【0083】

本発明では、例えば、特定のハプロタイプを所有する個体の群の、特定のハプロタイプを所有しない個体の群に対する疾患発症のオッズ比をハプロタイプ相対リスクとして定義する。多くの場合、特定のハプロタイプの所有の有無、疾患の発症の有無（臨床イベントの有無や薬剤の副作用の有無等でもよい）によって2×2分割表を作成し、この2×2分割表の独立性の検定（ χ^2 乗検定やFisherの直接確率法を用いる）によって特定のハプロタイプの所有の有無が疾患の発症の有無に与える影響を計算する。形質がいくつかのカテゴリに分割できないような場合は、t検定やWilcoxonテスト等を実施し、特定のハプロタイプを所有する個体の群と所有しない個体の群との形質の差を比較してもよい。

【0084】

関連性解析処理プログラム17によって得られた知識は、診断支援知識データベース18に格納される。

【0085】

図10は、診断支援知識データベース18の記述例を示す図である。各亜集団におけるハプロタイプ相対リスク情報の格納例を示している。ハプロタイプ相対リスクは、疾患の発症の有無や臨床イベントの有無、検査結果の正常・異常、薬剤の副作用の有無等、様々な臨床データについて定義することが可能であり、ここでは、心疾患、糖尿病、疾患Xの発症の有無に対する各亜集団ごとのハプロタイプ相対リスク情報の格納例を示している。例えば、ハプロタイプHT__1は亜集団SUBPOP__1内では心疾患に対する相対リスクが1.50であり、糖尿病、疾患Xに対する相対リスクがそれぞれ1.35, 1.00であることを示している。また、同時に、ハプロタイプHT__1は亜集団SUBPOP__2内では心疾患に対する相対リスクが2.00であり、糖尿病、疾患Xに対する相対リスクがそれぞれ1.89, 1.00と変化することを示している。

10

【0086】

罹患危険度算出処理プログラム19は、遺伝的構造情報データベース16および診断支援知識データベース18を参照して所定の個体が疾患に罹患する危険度を算出する。個体iがある疾患に罹患する危険度 R_i は、ハプロタイプブロックの数をm、集団内に存在する亜集団の数をN、亜集団jのハプロタイプブロックkにおける個体iのハプロタイプ相対リスクを r_{ijk} とすると、式(18)と表すことができる。

【0087】

【数18】

$$R_i = \prod_{k=1}^m \sum_{j=1}^N r_{ijk} Q_j \quad \text{---- (18)}$$

20

【0088】

図11は、外部診療機関112から接続パス31, 32およびインターネット30を介して本発明の診断支援システム111にアクセスし、本発明の診断支援システム111を利用した診断支援を受ける場合のシステム例を示す図である。外部診療機関112も、いわゆるパソコン等の電子計算機を備え、システムバス5に処理装置1、メモリー2、入力装置3、表示装置4および外部記憶装置10が接続される。ただし、外部診療機関112は、本発明のように、大規模な母集団のデータを扱うものではないので、複数の個体（被診断者）の診療情報を格納する診療情報データベース113および複数の個体（被診断者）の遺伝子多型に関する情報を格納する遺伝子多型情報データベース114は小規模のもので良い。単に、被診断者の診断に際して、個別に、本発明の診断支援システム111を利用した診断支援を受けることだけであれば、診療情報データベース113および遺伝子多型情報データベース114は無くても良い。尤も、本発明の診断支援システム111は、これを利用する外部診療機関112が、被診断者のデータを収集してこれを提供してくれることで、データを充実させてシステムをより完全なものにしていくことが望ましい。外部診療機関112が本発明の診断支援システム111を利用した診断支援を受ける場合は、外部診療機関112は診療情報データベース113および遺伝子多型情報データベース114から個体の遺伝子データおよび形質データを抽出して本発明の診断支援システム111へ送付する。外部診療機関112が診療情報データベース113および遺伝子多型情報データベース114を持たないときは、これらの情報を入力装置3から入力して、本発明の診断支援システム111に送付するものとすれば良い。本発明の診断支援システム111は、これらのデータを基礎に、算出された疾患に対する罹患危険度情報、遺伝的構造情報、個体の各亜集団への帰属度情報等を依頼元の外部診療機関112へ提供する。計算機の処理フローは特に説明するまでもない。

30

40

【図面の簡単な説明】

【0089】

【図1】本発明の診断支援システムの構成例を示す図である。

【図2】母集団のハプロタイプ頻度と個体のディプロタイプ形とを推定するハプロタイプ

50

ブロック推定処理プログラム 13 の例を示す図である。

【図 3】ハプロタイプブロックの設定に必要な基本情報の格納データ例を示す図である。

【図 4】各ハプロタイプブロック内におけるハプロタイプパターンおよびハプロタイプ頻度情報の格納例を示す図である。

【図 5】個体ごとのハプロタイプパターンの格納例を示す図である。

【図 6】あるハプロタイプブロック内においてハプロタイプ 1～ハプロタイプ 5 に示す 5 つのハプロタイプが観察された例を説明する図である。

【図 7】個体の帰属度を推定する遺伝的構造推定処理プログラム 15 を示す図である。

【図 8】各亜集団におけるハプロタイプパターンおよびハプロタイプ頻度情報の格納例を示す図である。

10

【図 9】個体ごとの各亜集団への帰属度情報の格納例を示す図である。

【図 10】診断支援知識データベース 18 の記述例を示す図である。

【図 11】外部診療機関 112 から接続パス 31, 32 およびインターネット 30 を介して本発明の診断支援システム 111 にアクセスし、本発明の診断支援システム 111 を利用した診断支援を受ける場合のシステム例を示す図である。

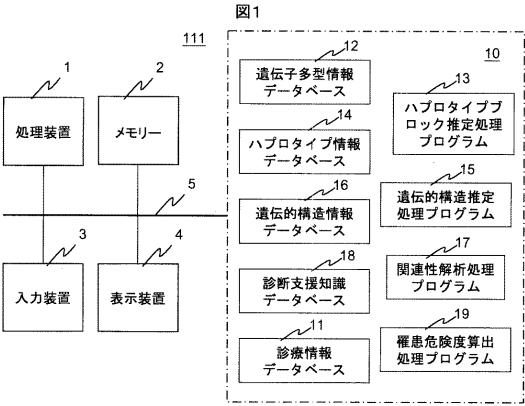
【符号の説明】

【0090】

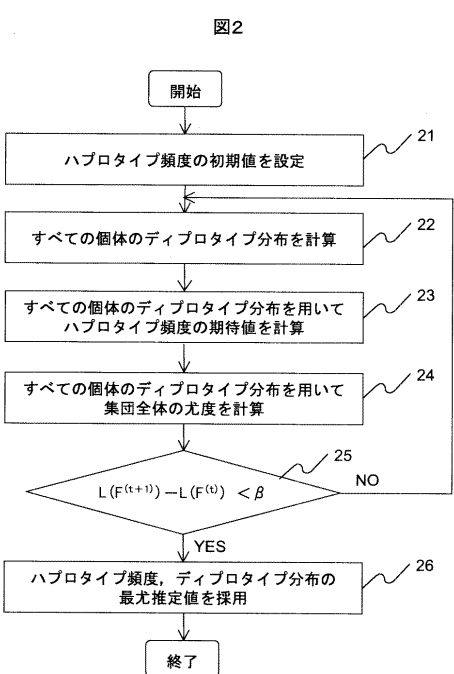
1…処理装置、2…メモリー、3…入力装置、4…表示装置、5…システムバス、10…外部記憶装置、11…診療情報データベース、12…遺伝子多型情報データベース、13…ハプロタイプブロック推定処理プログラム、14…ハプロタイプ情報データベース、15…遺伝的構造推定処理プログラム、16…遺伝的構造情報データベース、17…関連性解析処理プログラム、18…診断支援知識データベース、19…罹患危険度算出処理プログラム、21…ハプロタイプ頻度初期値設定ステップ、22…ディプロタイプ分布計算ステップ、23…尤度計算ステップ、24…ハプロタイプ頻度期待値計算ステップ、25…ハプロタイプ頻度・ディプロタイプ分布最尤推定値採用ステップ、71…ハプロタイプ間距離決定ステップ、72…クラスタリング実行ステップ、73…尤度計算ステップ、74…パラメータ採用ステップ、75…帰属度計算ステップ、111…診断支援システム、112…外部診療機関、113…外部診療の診療情報データベース、114…外部診療機関の遺伝子多型情報データベース。

20

【図 1】



【図 2】



【図 3】

図3

	遺伝子多型ID	遺伝子名	多型タイプ	ハプロタイプブロックID
1	POL_1	GENE_1	SNP	HB_1
2	POL_2	GENE_1	SNP	HB_1
3	POL_3	GENE_1	STRP	HB_1
4	POL_4	GENE_2	SNP	HB_2
5	POL_5	GENE_2	SNP	HB_3
...	...	...	...	...
X	POL_X	GENE_M	SNP	HB_m

【図 5】

図5

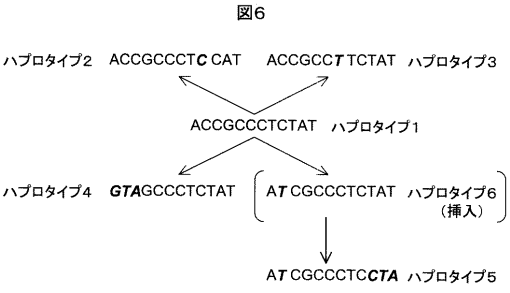
No.	個体ID	ハプロタイプブロックID	ハプロタイプ 1	ハプロタイプ 2	確率
1	PERSON_1	HB_1	HT_1	HT_1	1.00
2	PERSON_1	HB_2	HT_5	HT_5	0.95
3	PERSON_1	HB_2	HT_5	HT_6	0.05
...	...	...	...	...	...
p	PERSON_1	HB_m	HT_Y	HT_Y	1.00
p+1	PERSON_2	HB_1	HT_2	HT_3	0.90
p+2	PERSON_2	HB_1	HT_1	HT_4	0.10
p+3	PERSON_2	HB_2	HT_5	HT_5	0.20
...	...	...	...	...	...
q	PERSON_2	HB_m	HT_Y	HT_Y	1.00
...	...	...	...	...	...
r	PERSON_n	HB_m	HT_Y	HT_Y	1.00

【図 4】

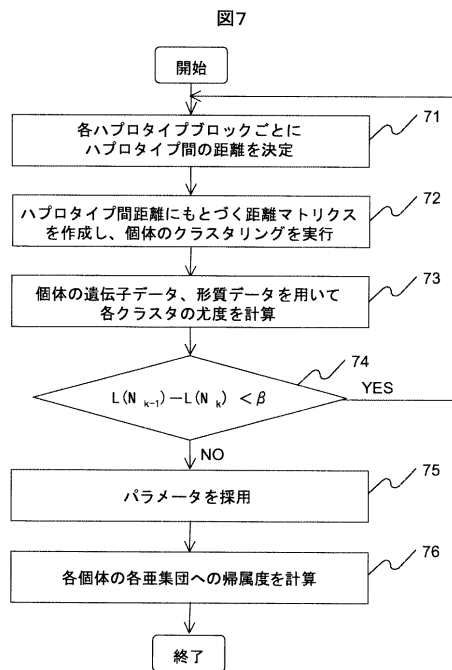
図4

No.	ハプロタイプID	ハプロタイプブロックID	ハプロタイプ頻度
1	HT_1	HB_1	0.50
2	HT_2	HB_1	0.28
3	HT_3	HB_1	0.15
4	HT_4	HB_1	0.07
5	HT_5	HB_2	0.88
6	HT_6	HB_2	0.12
7	HT_7	HB_1	0.34
8	HT_8	HB_1	0.33
9	HT_9	HB_1	0.33
10	HT_10	HB_2	1.00
...	...	...	...
Y	HT_Y	HB_m	0.40

【図 6】



【図 7】



【図 8】

図8

No.	ハプロタイプ ID	ハプロタイプ ブロックID	ハプロタイプ 頻度	亜集団ID
1	HT_1	HB_1	0.50	SUBPOP_1
2	HT_2	HB_1	0.28	SUBPOP_1
3	HT_3	HB_1	0.15	SUBPOP_1
4	HT_4	HB_1	0.07	SUBPOP_1
5	HT_5	HB_2	0.88	SUBPOP_1
6	HT_6	HB_2	0.12	SUBPOP_1
7	HT_7	HB_1	0.34	SUBPOP_2
8	HT_8	HB_1	0.33	SUBPOP_2
9	HT_9	HB_1	0.33	SUBPOP_2
10	HT_10	HB_2	1.00	SUBPOP_2
...	...	...	...	...
Y	HT_Y	HB_m	0.40	SUBPOP_N

【図 9】

図9

No.	個体ID	亜集団ID	帰属度
1	PERSON_1	SUBPOP_1	1.00
2	PERSON_1	SUBPOP_2	0.00
3	PERSON_1	SUBPOP_3	0.00
...	...	...	...
p	PERSON_1	SUBPOP_N	0.00
p+1	PERSON_2	SUBPOP_1	0.50
p+2	PERSON_2	SUBPOP_2	0.00
p+3	PERSON_2	SUBPOP_3	0.50
...	...	...	...
q	PERSON_2	SUBPOP_N	0.00
...	...	...	...
r	PERSON_n	SUBPOP_N	1.00

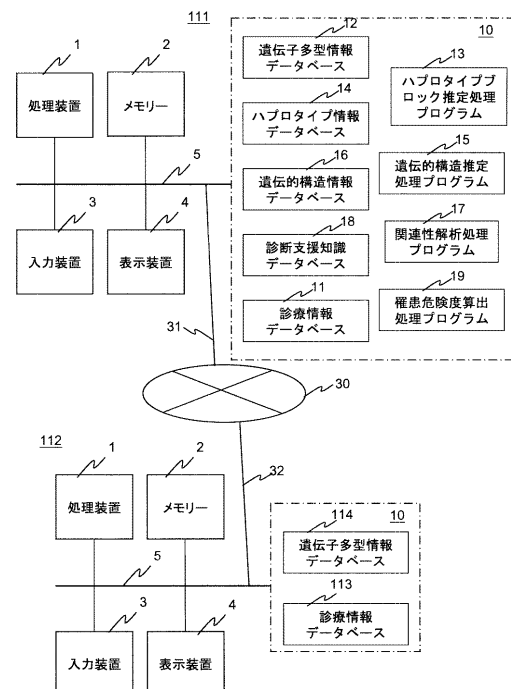
【図 10】

図10

No.	ハプロタイプ ID	亜集団ID	ハプロタイプ 相対リスク (心疾患)	ハプロタイプ 相対リスク (糖尿病)	ハプロタイプ 相対リスク (疾患X)
1	HT_1	SUBPOP_1	1.50	1.35	1.00
2	HT_1	SUBPOP_2	2.00	1.89	1.00
3	HT_2	SUBPOP_1	1.00	0.15	1.57
4	HT_3	SUBPOP_2	1.00	0.07	0.50
5	HT_3	SUBPOP_3	1.00	0.88	0.88
6	HT_3	SUBPOP_4	1.00	0.88	0.88
7	HT_4	SUBPOP_1	1.00	2.00	1.00
8	HT_4	SUBPOP_3	1.00	1.67	1.00
9	HT_5	SUBPOP_2	1.00	1.00	1.00
10	HT_6	SUBPOP_1	3.00	2.55	1.00
...	...	...	...	...	...
s	HT_Y	SUBPOP_N	0.80	0.75	0.90

【図 11】

図11



フロントページの続き

- (72)発明者 光山 訓
東京都国分寺市東恋ヶ窪一丁目280番地 株式会社日立製作所 中央研究所内
- (72)発明者 伴 秀行
東京都国分寺市東恋ヶ窪一丁目280番地 株式会社日立製作所 中央研究所内

審査官 宮久保 博幸

- (56)参考文献 特開2003-052383 (JP, A)
特表2003-521024 (JP, A)

- (58)調査した分野(Int.Cl., DB名)
- | | |
|--------|---------------|
| G06F | 19/00 |
| G06F | 17/30 |
| G06Q | 10/00 - 50/00 |
| PubMed | |