

(19) 日本国特許庁 (JP)

(12) 公表特許公報 (A)

(11) 特許出願公表番号

特表2021-536635

(P2021-536635A)

(43) 公表日 令和3年12月27日 (2021. 12. 27)

(51) Int.Cl.
G16H 10/60 (2018.01)F1
G16H 10/60テーマコード (参考)
5L099

審査請求 未請求 予備審査請求 未請求 (全 64 頁)

(21) 出願番号 特願2021-512545 (P2021-512545)
 (86) (22) 出願日 令和1年9月6日 (2019. 9. 6)
 (85) 翻訳文提出日 令和3年3月4日 (2021. 3. 4)
 (86) 国際出願番号 PCT/US2019/049942
 (87) 国際公開番号 WO2020/051445
 (87) 国際公開日 令和2年3月12日 (2020. 3. 12)
 (31) 優先権主張番号 62/728, 536
 (32) 優先日 平成30年9月7日 (2018. 9. 7)
 (33) 優先権主張国・地域又は機関
 米国 (US)

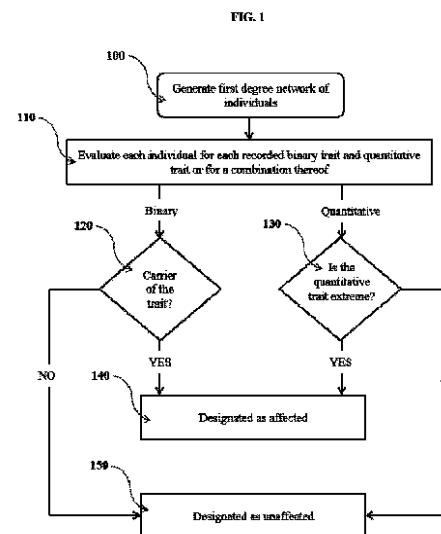
(71) 出願人 597160510
 リジェネロン・ファーマシューティカルズ
 ・インコーポレイテッド
 REGENERON PHARMACEU
 TICALS, INC.
 アメリカ合衆国10591-6707ニュ
 ヨーク州タリータウン、オールド・ソー
 ・ミル・リバー・ロード777番
 (74) 代理人 100105957
 弁理士 恩田 誠
 (74) 代理人 100068755
 弁理士 恩田 博宣
 (74) 代理人 100142907
 弁理士 本田 淳

最終頁に続く

(54) 【発明の名称】 家系エンリッチメント及び家系内の家族ベース解析のための方法及びシステム

(57) 【要約】

エンリッチされた家系を作成するための方法、非一時的コンピューターに実装された方法、及びシステムが提供される。また、疾患原因バリエーションを決定するための方法、非一時的コンピューターに実装された方法、及びシステムも提供される。



【特許請求の範囲】**【請求項 1】**

疾患原因バリエーションを同定するためのシステムであって、前記システムは、
データプロセッサと、前記データプロセッサに連結したメモリとを備え、前記プロセッサが、

コホートのシーケンシングデータに基づいて、個体の 1 親等ネットワークを作成することと、

前記 1 親等ネットワーク内の個体が罹患者であるか非罹患者であるかを同定することであって、少なくとも 1 つのバイナリー形質、極端な定量的形質、またはこれらの組合せを有する前記個体が罹患者として同定され、前記少なくとも 1 つのバイナリー形質または前記極端な定量的形質を有しない前記個体が非罹患者として同定される、前記同定することと、

罹患者または非罹患者の指定を含む前記個体を含む少なくとも 1 つのエンリッチされた家系を作成することと

を行うように構成される、前記システム。

【請求項 2】

前記プロセッサが、前記少なくとも 1 つのエンリッチされた家系の内部でまたはそれにわたって同時分離するバリエーション形質の対を同定するために分離解析を実施するようにさらに構成される、請求項 1 に記載のシステム。

【請求項 3】

前記プロセッサが、前記バリエーション形質の対を解析して前記疾患原因バリエーションを決定するために分離解析を実施するようにさらに構成される、請求項 1 に記載のシステム。

【請求項 4】

前記シーケンシングデータが、エキソームシーケンシングデータまたは全ゲノムシーケンシングデータまたは遺伝子型アレイデータからなる群より選択され得る、請求項 1 に記載のシステム。

【請求項 5】

前記プロセッサが、前記コホート内の前記少なくとも 1 つのバイナリー形質の有病率が 5 % を超える場合、前記同定された罹患者を非罹患者として同定するようにさらに構成される、請求項 1 に記載のシステム。

【請求項 6】

前記プロセッサが、前記個体の前記少なくとも 1 つの極端な定量的形質が前記コホートの極端な定量的形質の平均から 2 標準偏差未満である場合、前記同定された罹患者を非罹患者として同定するようにさらに構成される、請求項 1 に記載のシステム。

【請求項 7】

前記システムによって作成された前記エンリッチされた家系が、1 つの可能な構造を有する、請求項 1 ～ 6 のいずれか 1 項に記載のシステム。

【請求項 8】

前記システムによって作成された前記エンリッチされた家系が、共通の祖先を有する 3 例以上の罹患者を含む、請求項 1 ～ 7 のいずれか 1 項に記載のシステム。

【請求項 9】

前記システムによって作成された前記エンリッチされた家系が、非罹患と判定された親を有する 1 例以上の罹患者を含む、請求項 1 ～ 8 のいずれか 1 項に記載のシステム。

【請求項 10】

前記プロセッサが、前記分離解析を実施するために前記少なくとも 1 つの家系を優先順位付けするようにさらに構成される、請求項 1 ～ 9 のいずれか 1 項に記載のシステム。

【請求項 11】

前記少なくとも 1 つの家系を前記優先順位付けすることが、少なくとも 1 例の近親の非罹患者を含む前記少なくとも 1 つの家系を選択することを含む、請求項 10 に記載のシステム。

10

20

30

40

50

【請求項 1 2】

前記少なくとも 1 つの家系を前記優先順位付けすることが、少なくとも 2 例の罹患した同胞を含む前記少なくとも 1 つの家系を選択することを含む、請求項 1 0 に記載のシステム。

【請求項 1 3】

前記分離解析が、分離の優性遺伝モデルを形成する、請求項 1 0 に記載のシステム。

【請求項 1 4】

前記分離解析が、分離の劣性遺伝モデルを形成する、請求項 1 0 に記載の方法。

【請求項 1 5】

前記バリエーション形質の対を解析して前記疾患原因バリエーションを決定することが、家族ベースの関連性解析を用いて行われる、請求項 3 に記載の方法。

10

【請求項 1 6】

疾患原因バリエーションを同定する方法であって、

コホートのシーケンシングデータに基づいて、個体の 1 親等ネットワークを作成することと、

前記 1 親等ネットワーク内の個体を罹患者または非罹患者として同定することであって、少なくとも 1 つのバイナリー形質、極端な定量的形質、またはこれらの組合せによって罹患した前記個体が罹患者として同定され、前記少なくとも 1 つのバイナリー形質または前記極端な定量的形質によって罹患していない前記個体が非罹患者として同定される、前記同定することと、

20

罹患者または非罹患者の指定を含む前記個体を含む少なくとも 1 つのエンリッチされた家系を作出することと、

前記少なくとも 1 つのエンリッチされた家系の内部でまたはそれにわたって同時分離するバリエーション形質の対を同定するために分離解析を実施することと、

前記バリエーション形質の対を解析して前記疾患原因バリエーションを同定することとを含む、前記方法。

【請求項 1 7】

前記シーケンシングデータが、エキソームシーケンシングデータ、全ゲノムシーケンシングデータ、及び遺伝子型アレイデータから選択され得る、請求項 1 6 に記載の方法。

30

【請求項 1 8】

前記コホート内の前記少なくとも 1 つのバイナリー形質の有病率が 5 % を超える場合、前記罹患者として同定された前記個体が非罹患者として脱同定される、請求項 1 6 または請求項 1 7 に記載の方法。

【請求項 1 9】

前記個体の前記少なくとも 1 つの極端な定量的形質が前記コホートの極端な定量的形質の平均から 2 標準偏差未満である場合、前記罹患者として同定された前記個体が非罹患者として脱同定される、請求項 1 6 または請求項 1 7 に記載の方法。

【請求項 2 0】

前記エンリッチされた家系が、1 つの可能な構造を有する、請求項 1 6 ～ 1 9 のいずれか 1 項に記載の方法。

40

【請求項 2 1】

前記エンリッチされた家系が、共通の祖先を有する 3 例以上の罹患者を含む、請求項 1 6 ～ 2 0 のいずれか 1 項に記載の方法。

【請求項 2 2】

前記エンリッチされた家系が、非罹患と判定された親を有する 1 例以上の罹患者を含む、請求項 1 6 ～ 2 1 のいずれか 1 項に記載の方法。

【請求項 2 3】

前記分離解析を実施するために前記少なくとも 1 つの家系を優先順位付けすることをさらに含む、請求項 1 6 ～ 2 2 のいずれか 1 項に記載の方法。

50

【請求項 24】

前記少なくとも 1 つの家系を優先順位付けすることが、少なくとも 1 例の近親の非罹患者を含む前記少なくとも 1 つの家系を選択することを含む、請求項 23 に記載の方法。

【請求項 25】

前記少なくとも 1 つの家系を優先順位付けすることが、少なくとも 2 例の罹患した同胞を含む前記少なくとも 1 つの家系を選択することを含む、請求項 23 に記載の方法。

【請求項 26】

前記分離解析が、分離の優性遺伝モデルを形成する、請求項 23～25 のいずれか 1 項に記載の方法。

【請求項 27】

前記分離解析が、分離の劣性遺伝モデルを形成する、請求項 23～25 のいずれか 1 項に記載の方法。

【請求項 28】

前記バリエーション形質の対を解析して前記疾患原因バリエーションを同定することが、家族ベースの関連性解析を用いて行われる、請求項 16～27 のいずれか 1 項に記載の方法。

【請求項 29】

疾患原因バリエーションを同定するための方法をプロセッサに実施させるための命令を格納する (strong) 非一時的コンピューター可読媒体であって、

コホートのエクソームシーケンシングデータに基づいて、個体の 1 親等ネットワークを作成することと、

前記 1 親等ネットワーク内の個体が罹患者であるか非罹患者であるかを同定することであって、少なくとも 1 つのバイナリー形質、極端な定量的形質、またはこれらの組合せを有する前記個体が罹患者として同定され、前記少なくとも 1 つのバイナリー形質または前記極端な定量的形質を有しない前記個体が非罹患者として同定される、前記同定することと、

罹患者または非罹患者の指定を含む前記個体を含む少なくとも 1 つのエンリッチされた家系を作出することと、

前記少なくとも 1 つのエンリッチされた家系の内部でまたはそれにわたって同時分離するバリエーション形質の対を同定するために分離解析を実施することと、

前記バリエーション形質の対を解析して前記疾患原因バリエーションを決定することとを含む、前記非一時的コンピューター可読媒体。

【請求項 30】

前記シーケンシングデータが、エキソームシーケンシングデータ、全ゲノムシーケンシングデータ、及び遺伝子型アレイデータから選択され得る、請求項 29 に記載の非一時的コンピューター可読媒体。

【請求項 31】

前記コホート内の前記少なくとも 1 つのバイナリー形質の有病率が 5 % を超える場合、前記罹患者として同定された前記個体が非罹患者として脱同定される、請求項 29 または 30 に記載の非一時的コンピューター可読媒体。

【請求項 32】

前記個体の前記少なくとも 1 つの極端な定量的形質が前記コホートの極端な定量的形質の平均から 2 標準偏差未満である場合、前記罹患者として同定された前記個体が非罹患者として脱同定される、請求項 29 または 30 に記載の非一時的コンピューター可読媒体。

【請求項 33】

前記エンリッチされた家系が、1 つの可能な構造を有する、請求項 29～32 のいずれか 1 項に記載の非一時的コンピューター可読媒体。

【請求項 34】

前記エンリッチされた家系が、共通の祖先を有する 3 例以上の罹患者を含む、請求項 29～33 のいずれか 1 項に記載の非一時的コンピューター可読媒体。

【請求項 35】

前記エンリッチされた家系が、非罹患と判定された親を有する 1 例以上の罹患者を含む、請求項 29～34 のいずれか 1 項に記載の非一時的コンピューター可読媒体。

【請求項 36】

前記方法が、前記分離解析を実施するために前記少なくとも 1 つの家系を優先順位付けすることをさらに含む、請求項 29～35 のいずれか 1 項に記載の非一時的コンピューター可読媒体。

【請求項 37】

前記少なくとも 1 つの家系を前記優先順位付けすることが、少なくとも 1 例の近親の非罹患者を含む前記少なくとも 1 つの家系を選択することを含む、請求項 36 に記載の非一時的コンピューター可読媒体。

10

【請求項 38】

前記少なくとも 1 つの家系を前記優先順位付けすることが、少なくとも 2 例の罹患した同胞を含む前記少なくとも 1 つの家系を選択することを含む、請求項 36 に記載の非一時的コンピューター可読媒体。

【請求項 39】

前記分離解析が、分離の優性遺伝モデルを形成する、請求項 36～38 のいずれか 1 項に記載の非一時的コンピューター可読媒体。

【請求項 40】

前記分離解析が、分離の劣性遺伝モデルを形成する、請求項 36～38 のいずれか 1 項に記載の非一時的コンピューター可読媒体。

20

【請求項 41】

前記バリエーション形質の対を解析して前記疾患原因バリエーションを同定することが、家族ベースの関連性解析を用いて行われる、請求項 29～40 のいずれか 1 項に記載の非一時的コンピューター可読媒体。

【請求項 42】

エンリッチされた家系を作成するためのシステムであって、前記システムは、データプロセッサと、前記データプロセッサに連結したメモリーとを備え、前記プロセッサが、

コホートのシーケンシングデータに基づいて、個体の 1 親等ネットワークを作成することと、

30

前記 1 親等ネットワーク内の個体が罹患者であるか非罹患者であるかを同定することであって、少なくとも 1 つのバイナリー形質、極端な定量的形質、またはこれらの組合せを有する前記個体が罹患者として同定され、前記少なくとも 1 つのバイナリー形質または前記極端な定量的形質を有しない前記個体が非罹患者として同定される、前記同定することと、

罹患者または非罹患者の指定を含む前記個体を含む少なくとも 1 つのエンリッチされた家系を作成することと

を行うように構成される、前記システム。

【請求項 43】

前記シーケンシングデータが、エキソームシーケンシングデータまたは全ゲノムシーケンシングデータまたは遺伝子型アレイデータからなる群より選択され得る、請求項 42 に記載のシステム。

40

【請求項 44】

前記プロセッサが、前記コホート内の前記少なくとも 1 つのバイナリー形質の有病率が 5 % を超える場合、前記同定された罹患者を非罹患者として同定するようにさらに構成される、請求項 42 に記載のシステム。

【請求項 45】

前記プロセッサが、前記個体の前記少なくとも 1 つの極端な定量的形質が前記コホートの極端な定量的形質の平均から 2 標準偏差未満である場合、前記同定された罹患者を非罹患者として同定するようにさらに構成される、請求項 42 に記載のシステム。

50

【請求項 4 6】

前記システムによって作成された前記エンリッチされた家系が、1つの可能な構造を有する、請求項 4 2～4 5 のいずれか 1 項に記載のシステム。

【請求項 4 7】

前記システムによって作成された前記エンリッチされた家系が、共通の祖先を有する 3 例以上の罹患者を含む、請求項 4 2～4 6 のいずれか 1 項に記載のシステム。

【請求項 4 8】

前記システムによって作成された前記エンリッチされた家系が、非罹患者と判定された親を有する 1 例以上の罹患者を含む、請求項 4 2～4 7 のいずれか 1 項に記載のシステム。

【請求項 4 9】

エンリッチされた家系を作成するための方法であって、
コホートのシークエンシングデータに基づいて、個体の 1 親等ネットワークを作成することと、

前記 1 親等ネットワーク内の個体を罹患者または非罹患者として同定することであって、少なくとも 1 つのバイナリー形質、極端な定量的形質、またはこれらの組合せによって罹患した前記個体が罹患者として同定され、前記少なくとも 1 つのバイナリー形質または前記極端な定量的形質によって罹患していない前記個体が非罹患者として同定される、前記同定することと、

罹患者または非罹患者の指定を含む前記個体を含む少なくとも 1 つのエンリッチされた家系を作出することと

を含む、前記方法。

【請求項 5 0】

前記シークエンシングデータが、エキソームシークエンシングデータ、全ゲノムシークエンシングデータ、及び遺伝子型アレイデータから選択され得る、請求項 4 9 に記載の方法。

【請求項 5 1】

前記コホート内の前記少なくとも 1 つのバイナリー形質の有病率が 5 % を超える場合、前記罹患者として同定された前記個体が非罹患者として脱同定される、請求項 4 9 及び 5 0 のいずれか 1 項に記載の方法。

【請求項 5 2】

前記個体の前記少なくとも 1 つの極端な定量的形質が前記コホートの極端な定量的形質の平均から 2 標準偏差未満である場合、前記罹患者として同定された前記個体が非罹患者として脱同定される、請求項 4 9 及び 5 0 のいずれか 1 項に記載の方法。

【請求項 5 3】

前記エンリッチされた家系が、1つの可能な構造を有する、請求項 4 9～5 2 のいずれか 1 項に記載の方法。

【請求項 5 4】

前記エンリッチされた家系が、共通の祖先を有する 3 例以上の罹患者を含む、請求項 4 9～5 3 のいずれか 1 項に記載の方法。

【請求項 5 5】

前記エンリッチされた家系が、罹患していないと判定された親を有する 1 例以上の罹患者を含む、請求項 4 9～5 4 のいずれか 1 項に記載の方法。

【請求項 5 6】

エンリッチされた家系を作成するための方法をプロセッサに実施させるための命令を格納する (strong) 非一時的コンピューター可読媒体であって、

コホートのエクソームシークエンシングデータに基づいて、個体の 1 親等ネットワークを作成することと、

前記 1 親等ネットワーク内の個体が罹患者であるか非罹患者であるかを同定することであって、少なくとも 1 つのバイナリー形質、極端な定量的形質、またはこれらの組合せを有する前記個体が罹患者として同定され、前記少なくとも 1 つのバイナリー形質または前

10

20

30

40

50

記極端な定量的形質を有しない前記個体が非罹患者として同定される、前記同定することと、

罹患者または非罹患者の指定を含む前記個体を含む少なくとも1つのエンリッチされた家系を作成することと

を含む、前記非一時的コンピューター可読媒体。

【請求項57】

前記シーケンシングデータが、エキソームシーケンシングデータ、全ゲノムシーケンシングデータ、及び遺伝子型アレイデータから選択され得る、請求項56に記載の非一時的コンピューター可読媒体。

【請求項58】

前記コホート内の前記少なくとも1つのバイナリー形質の有病率が5%を超える場合、前記罹患者として同定された前記個体が非罹患者として脱同定される、請求項56または請求項57に記載の非一時的コンピューター可読媒体。

【請求項59】

前記個体の前記少なくとも1つの極端な定量的形質が前記コホートの極端な定量的形質の平均から2標準偏差未満である場合、前記罹患者として同定された前記個体が非罹患者として脱同定される、請求項56～58のいずれか1項に記載の非一時的コンピューター可読媒体。

【請求項60】

前記エンリッチされた家系が、1つの可能な構造を有する、請求項56～59のいずれか1項に記載の非一時的コンピューター可読媒体。

【請求項61】

前記エンリッチされた家系が、共通の祖先を有する3例以上の罹患者を含む、請求項56～60のいずれか1項に記載の非一時的コンピューター可読媒体。

【請求項62】

前記エンリッチされた家系が、非罹患と判定された親を有する1例以上の罹患者を含む、請求項56～61のいずれか1項に記載の非一時的コンピューター可読媒体。

【請求項63】

前記方法が、前記分離解析を実施するために前記少なくとも1つの家系を優先順位付けすることをさらに含む、請求項56～62のいずれか1項に記載の非一時的コンピューター可読媒体。

【発明の詳細な説明】

【技術分野】

【0001】

本開示は、概して、大規模集団コホートにおける家系エンリッチメントのための方法及びシステムに関する。より詳細には、本開示は、シーケンシングデータを用いて家系をエンリッチするために1親等家族ネットワークにおける罹患者を同定し、さらに、希少な遺伝子バリエーションを疾患及び疾患感受性に関連付けるために家系内及び家系間で同時分離するバリエーション形質の対を同定するための、システム及び方法に関する。

【0002】

関連出願の相互参照

本出願は、2018年9月7日に出願された米国仮特許出願第62/728,536号の利益を主張し、その内容の全体が参照により本明細書に援用される。

【背景技術】

【0003】

臨床研究者は、疾患の原因となる病原性バリエーションを同定するための努力を続けている。細胞ゲノムアレイ及び連鎖パネルのジェノタイピングは、それぞれ、コピー数バリエーションの同定と、大きなメンデル遺伝（特に優性）疾患ファミリー内の同時分離ハプロタイプの識別とにおいて、依然有用なアプローチとなっている。しかし、複雑な疾患で病原性バリエーションを発見する最適なアプローチは依然として明らかになっていない。

10

20

30

40

50

【0004】

家系を介しバリエーションの伝達を追跡することが現代遺伝学の基礎となっている。ほとんどの遺伝子障害は不均一性であり、少数から多数にわたる一連の遺伝子が疾患を引き起こす役割を担う。多くの希少障害の遺伝子異常については、依然説明されていない。古典的なポジショナルクローニング技法の場合、原因遺伝子が存在すると考えられる領域を同定するには相当な数の罹患家族を要し、希少障害においては、このような数の家族を常に利用できるわけではない。さらに、目的領域の同定だけでは不十分である。この領域内の遺伝子は全てシーケンシングしなければならない、これはかなりの労力を要する。次世代シーケンシングが登場したことで、候補遺伝子領域を選択する必要なしに患者の全ゲノムまたはエクソームを研究することが可能になっている。現在では、大規模な研究コホート内で希少な遺伝子バリエーションを発見しジェノタイピングすることができるが、これらのバリエーションの大部分は少数の個体にしか存在しない。集団ベースの遺伝子研究では、バリエーションの50%超が単一の個体にしか認められず、関連性のエビデンスを確立することが困難となっている。

10

【0005】

さらに、大規模な遺伝子コホートのゲノムワイドスクランニングにおいて、このような不均一性障害に希少バリエーションが及ぼす影響を調べることは特に困難である。配列バリエーションの疾患因果関係を明確に割り当てるのは不可能であることが多く、希少かつ重症疾患の多くの症例が根底に有するごく低頻度のバリエーションの場合には、特に不可能である。ただし、所与の遺伝子障害を共有する近親の個体セットが同定されるのであれば、この不均一性は大幅に低減され、家系内の罹患個体内で分離する特異的表現型を駆動する単一の遺伝子及びバリエーションに焦点を当てることができる。

20

【0006】

ヒトゲノム全体にわたって疾患座位に対し偏りのない探索を可能にするゲノムワイド関連性解析 (GWAS) の潜在可能性は、遺伝学において前例のない研究機会をもたらす。多数の対象にわたって同時に数十万の一塩基多型 (SNP) を調べることは、このような研究のデザイン及び解析における多くの統計的な課題を提起する。このような規模のジェノタイピングには、データ品質の問題を取り扱うための新たな方法論が必要である。同様に、関連性検定は何十万ものマーカーに対し計算されるが、これらの結果は多重比較に応じて調整されなければならない。これらの問題の大きさは、このような高密度のSNPセットをジェノタイピングする新たな技術能力が新規の遺伝子疾患座位の同定に至るのか、または技術的進歩が依然として活用されていないのかという問題を提起する。このようなゲノムワイド関連性研究に対するアプローチには少なくとも2つの方法があり、それは集団ベースの設計及び家族ベースの設計である。

30

【0007】

集団ベース研究のサンプルサイズは、数千例の対象となっている (非特許文献1)。しかし、集団ベース研究は費用や時間がかかり、サンプルサイズが大きいことから表現型及び遺伝子型の不均一性に遭遇する可能性がある (非特許文献2および非特許文献3)。

【0008】

家族ベース解析は、目的表現型と共に同時分離する潜在的な中～大の影響を有する希少バリエーションを調べる際には特に有益であると考えられ、このようなバリエーションは、集団ベース解析では容易に検出されない場合がある。家族ベースの関連性研究は、検出力喪失の可能性のあるものの、その主な利点は、集団層別化による交絡バイアスを制御することにある (非特許文献4および非特許文献5)。

40

【0009】

数十万例の脱同定された個体を確認しシーケンシングするための大規模シーケンシングイニシアチブが多数存在し、例えば、DiscovEHR、UK Biobank、米国政府のAll of US (Precision Medicine Initiativeの一環)、TOPMed、ExAC/gnomAD、及びその他多数 (非特許文献6～9) が該当する。家系は、このようなタンパク質シーケンス情報の大規模なデー

50

タセットから作成することができ、研究者は、これを用いて形質及び障害の遺伝率及び遺伝モデルを決定することができる。正確な家系構造を知ること、疾患遺伝の遺伝子様式を正確に同定し、真の家系構造を必要とする、またはそこから恩恵を受ける強力な遺伝子解析ツールを利用することができる。しかし、脱同定された健康記録から正確な家系記録を直接入手するという課題が存在し、これが多くの強力な家族ベース解析の妨げとなっている。

【0010】

密接なペアワイズ関係は、PRIMUS及びCLAPPER（非特許文献10および非特許文献11）などのツールを用いて遺伝子データから直接家系構造を再構築するために使用することができる。推定された関係及び家系は極めて有用であるが、推定された関係及び家系の使用に関しては、推定された関係及び家系の構造の不正確さに敏感な解析では統計的不確実性が顕著になる懸念が存在する。

10

【0011】

プレジジョンメディシンのコホートは、家系情報を容易に有しない場合があるが、情報価値のある家系を遺伝子データから直接得て、伝統的なメンデル遺伝解析用の大規模コホートを作出することができる。家族単位内では遺伝子的原因が共有される可能性がより高いため、目的表現型を有する罹患者のためのエンリッチされた家系の同定は、このような表現型を駆動する原因（希少）バリエーションを同定する試みの中で使用することができる。家系エンリッチメント解析で使用される罹患個体のセットを定義することは不可欠であるため、家系エンリッチメントを可能にする方法またはシステムが必要である。これらのエンリッチされた家系は、目的表現型を有する近親の参加者のサブセットを定義し、次いでこれらのサブセットを調べて形質及び疾患の遺伝子的駆動要因を同定することに活用することができる。大規模な伝統的メンデル遺伝解析を可能にする高い情報価値を有し得る家系—表現型の対を同定するために、家系エンリッチメント用の改良されたバイオインフォマティクスツールが依然として必要とされている。

20

【0012】

エンリッチされた家系作成に対する方法及びシステムの発見は、創薬科学者が、ある特定のタンパク質及びそのバリエーションが正常な生理学または疾患の原因で果たす重要な役割を理解し、生化学及び生物学の両面においてこれらの機能を解明するための指針となり得る（非特許文献12）。

30

【0013】

本明細書で説明される方法及びシステムは、このような疾患原因バリエーション（複数可）の同定につながり得るエンリッチされた家系を提供し、よって創薬努力及び臨床研究努力を刺激するものとなる。

【先行技術文献】

【非特許文献】

【0014】

【非特許文献1】Szkló M.、Epidemiologic Reviews (1998) 20 (1) : 81-90

【非特許文献2】SorlieおよびWei.、Journal of American College of Cardiology (2011) 58 (19) : 2010-3

40

【非特許文献3】LairdおよびLange、Statistical Science (2009) 24 (4) : 388-397

【非特許文献4】Witte他、American Journal of epidemiology (1999) 149 (8) : 693-705

【非特許文献5】Thomas他、Cancer (2003) 97 (8) : 1894-1903

【非特許文献6】Dewey他、Science (2016) ; 254, aaf6814

【非特許文献7】Sudlow他、PLoS Med. (2015) 12, e100177

50

9

【非特許文献8】Collins他、(2016) New England Journal of Medicine (2015) 372, 793-795

【非特許文献9】Lek他、Nature (2016) 536, 285-291

【非特許文献10】Staples他、American Journal of Human Genetics (2014) 95, 553-564

【非特許文献11】KoおよびNielson、PLOS Genet. (2017) 13, e1006963

【非特許文献12】Lelle R., J. Assoc. Physicians India (2003) 51: 373-380

10

【発明の概要】

【課題を解決するための手段】

【0015】

1つの例示的な態様において、本開示は、エンリッチされた家系を作成するための方法であって、コホートのシークエンシングデータに基づいて個体の1親等ネットワークを作成することと、コホート内の個体を罹患者または非罹患者として同定することと、罹患者及び非罹患者を含むエンリッチされた家系を作成することによって作成する、方法を提供する。

【0016】

いくつかの例示的な実施形態において、エンリッチされた家系を作成するための方法は、家系内の個体を罹患者または非罹患者として同定することであって、少なくとも1つのバイナリー形質を有する個体が罹患者として同定され、当該少なくとも1つのバイナリー形質を有しない個体が非罹患者として同定される、同定することと、次いで、罹患者及び非罹患者のパターンがメンデル遺伝様式（例えば、常染色体優性、常染色体劣性、x連鎖優性、x連鎖劣性、またはy連鎖）と一致するかどうかを評価することとを含み得る。いくつかの特定の例示的な実施形態において、バイナリー形質は、世界保健機関（WHO）による医学分類リストである疾病及び関連保健問題の国際統計分類（ICD）を用いて定義することができる。この分類には、疾患、徴候及び症状、異常所見、病訴、社会的状況、ならびに傷害または疾患の外部原因に関するコードが含まれる。バイナリー形質の定義には、ICDの第9版または第10版を使用することができる。1つの例示的な実施形態において、特定のバイナリー形質に関して電子健康記録データが利用できない可能性のある個体、または特定のバイナリー形質に関して矛盾するもしくは信頼できないデータを有する個体は、医療記録に特定のバイナリー形質が存在するかしないかにかかわらず、未知の罹患者として判定され得る。

20

30

【0017】

いくつかの例示的な実施形態において、エンリッチされた家系を作成するための方法は、家系内の個体を罹患者または非罹患者として同定することであって、少なくとも1つの極端な定量的形質を有する個体が罹患者として同定され、当該少なくとも1つの極端な定量的形質を有しない個体が非罹患者として同定される、同定することと、次いで、罹患者及び非罹患者のパターンがいずれかのメンデル遺伝様式（例えば、常染色体優性、常染色体劣性、x連鎖優性、x連鎖劣性、またはy連鎖）と一致するかどうかを評価することとを含み得る。いくつかのパラメーターを使用して、極端な定量的形質の影響を受けているかどうかを定義することができる。例えば、最大年齢カットオフを使用してより早期の障害発症を定義すること、または、形質の正常な集団の測定値からの偏差の定義された統計的カットオフを超えた（例えば、集団平均を2標準偏差上回る）定量的形質の最小値もしくは最大値もしくは中央値を測定することができる。1つの例示的な実施形態において、特定の定量的形質に関して電子健康記録データが利用できない可能性のある個体、または特定の定量的形質に関して矛盾するもしくは信頼できないデータを有する個体は、医療記録に特定の定量的形質が存在するかしないかにかかわらず、未知の罹患者として判定され得る。

40

50

【0018】

いくつかの例示的な実施形態において、エンリッチされた家系を作成するための方法は、家系内の個体を罹患者または非罹患者として同定することであって、少なくとも1つのバイナリー形質、極端な定量的形質、またはこれらの組合せを有する個体が罹患者として同定され、少なくとも1つのバイナリー形質、極端な定量的形質、またはこれらの組合せを有しない個体が非罹患者として同定される、同定することを含み得る。バイナリー形質は、上記のような定義されたICDコードとすることができる。いくつかのパラメーターを使用して、上記のような極端な定量的形質を定義することができる。1つの例示的な実施形態において、特定のバイナリー形質、定量的形質、もしくはこれらの組合せに関して電子健康記録データが利用できない個体、または特定のバイナリー形質、定量的形質、もしくはこれらの組合せに関して矛盾するもしくは信頼できないデータを有する個体は、医療記録に特定の定量的形質が存在するかしないかにかかわらず、未知の罹患者として判定され得る。

10

【0019】

いくつかの例示的な実施形態において、エンリッチされた家系を作成するための方法は、家系内の個体を罹患者または非罹患者として同定することであって、少なくとも1つのバイナリー形質、極端な定量的形質、またはこれらの組合せを有する個体が罹患者として同定され、少なくとも1つのバイナリー形質、極端な定量的形質、またはこれらの組合せを有しない個体が非罹患者として同定される、同定することを含み得、このとき、少なくとも1つのバイナリー形質、極端な定量的形質、またはこれらの組合せは、2つ以上の類似のまたは補完的な形質を含み得る。

20

【0020】

いくつかの例示的な実施形態において、エンリッチされた家系を作成するための方法は、家系内の個体を罹患者または非罹患者として同定することであって、少なくとも1つのバイナリー形質、極端な定量的形質、またはこれらの組合せを有する個体が罹患者として同定され、少なくとも1つのバイナリー形質、極端な定量的形質、またはこれらの組合せを有しない個体が非罹患者として同定される、同定することを含み得、このとき、少なくとも1つのバイナリー形質、極端な定量的形質、またはこれらの組合せは、2つ以上の極端なまたは興味深い形質の共通部分を採用することを含み得る。

【0021】

いくつかの例示的な実施形態において、エンリッチされた家系を作成するための方法は、家系内の個体を罹患者として同定することであって、少なくとも1つのバイナリー形質、極端な定量的形質、またはこれらの組合せを有する個体が罹患者として同定される、同定することと、罹患者として判定された個体を、外部解析からの関連性結果の罹患キャリアとして定義することを含み得る。

30

【0022】

いくつかの例示的な実施形態において、エンリッチされた家系を作成するための方法は、コホートのシークエンシングデータに基づいて個体の1親等ネットワークを作成することを含む。シークエンシングデータは、全ゲノムシークエンシングデータ、エクソームシークエンシングデータ、または遺伝子型データを含み得る。

40

【0023】

いくつかの例示的な実施形態において、エンリッチされた家系を作成するための方法は、エクソームシークエンシングデータに基づいて個体の1親等ネットワークを作成することを含む。エクソームシークエンシングデータに基づく個体の1親等ネットワークは、以下を含めた集団の関連性を活用することによって作成することができる：複数のヒト対象から得られた核酸配列サンプルのデータセットから低品質の配列バリエーションを削除すること、サンプルのうちの1例以上の各々に対する祖先スーパークラス指定を確立すること、データセットから低品質のサンプルを削除すること、祖先スーパークラス内の対象の第1の家系同一性推定値を作成すること、対象の祖先スーパークラスとは無関係に第2の家系同一性推定値を作成すること、及び第2の家系同一性推定値のうちの1つ以上に基づき、

50

対象を主要な1親等家族ネットワークにクラスター化すること。

【0024】

いくつかの例示的な実施形態において、エンリッチされた家系を作成するための方法は、コホートのシークエンシングデータに基づいて個体の1親等ネットワークを作成することを含み、このときコホートは、複数の対象を含む任意のデータセットを含み得る。

【0025】

いくつかの例示的な実施形態において、エンリッチされた家系を作出するための方法はさらに、p値に基づいて家系をエンリッチすることを含む。エンリッチメントは、家系の「創始者アンカー分枝」または「分枝」を家系内の創始者の全ての子孫として定義し、二項検定を使用して分枝のバイナリー形質がエンリッチされているかを評価することを含み得る。バイナリー形質は、上記のようなICDを用いて定義することができる。また、エンリッチメントは、家系の「創始者アンカー分枝」または「分枝」を家系内の創始者の全ての子孫として定義し、t検定を使用して分枝の極端な定量的形質がエンリッチされているかを評価することも含み得る。いくつかのパラメーターを使用して、上記のような極端な定量的形質を定義することができる。さらに、エンリッチメントは、多重検定p値カットオフを適用することも含み得る。

【0026】

1つの例示的な態様において、本開示は、エンリッチされた家系を作成することによって疾患原因バリエーションを同定するための方法であって、コホートのシークエンシングデータに基づいて個体の1親等ネットワークを作成し、コホート内の個体を罹患者または非罹患者として同定し、罹患者及び非罹患者を含む少なくとも1つのエンリッチされた家系を作出し、当該少なくとも1つのエンリッチされた家系の内部でまたはそれにわたって同時分離するバリエーション形質の対を同定するために分離解析を実施し、バリエーション形質の対を解析して疾患原因バリエーションを同定することにより、エンリッチされた家系を作成する、方法を提供する。

【0027】

いくつかの例示的な実施形態において、疾患原因バリエーションを同定するための方法は、家系内の個体を罹患者または非罹患者として同定することであって、少なくとも1つのバイナリー形質を有する個体が罹患者として同定され、当該少なくとも1つのバイナリー形質を有しない個体が非罹患者として同定される、同定することと、次いで、罹患者及び非罹患者のパターンがメンデル遺伝様式（例えば、常染色体優性、常染色体劣性、x連鎖優性、x連鎖劣性、またはy連鎖）と一致するかどうかを評価することとを含み得る。いくつかの特定の例示的な実施形態において、バイナリー形質は、世界保健機関（WHO）による医学分類リストである疾病及び関連保健問題の国際統計分類（ICD）を用いて定義することができる。この分類には、疾患、徴候及び症状、異常所見、病訴、社会的状況、ならびに傷害または疾患の外部原因に関するコードが含まれる。バイナリー形質の定義には、ICDの第9版または第10版を使用することができる。1つの例示的な実施形態において、特定のバイナリー形質に関して電子健康記録データが利用できない可能性のある個体、または特定のバイナリー形質に関して矛盾するもしくは信頼できないデータを有する個体は、医療記録に特定のバイナリー形質が存在するかしないかにかかわらず、未知の罹患者として判定され得る。

【0028】

いくつかの例示的な実施形態において、疾患原因バリエーションを同定するための方法は、家系内の個体を罹患者または非罹患者として同定することであって、少なくとも1つの極端な定量的形質を有する個体が罹患者として同定され、当該少なくとも1つの極端な定量的形質を有しない個体が非罹患者として同定される、同定することと、次いで、罹患者及び非罹患者のパターンがいずれかのメンデル遺伝様式（例えば、常染色体優性、常染色体劣性、x連鎖優性、x連鎖劣性、またはy連鎖）と一致するかどうかを評価することとを含み得る。いくつかのパラメーターを使用して、極端な定量的形質の影響を受けているかどうかを定義することができる。例えば、最大年齢カットオフを使用してより早期の障害

10

20

30

40

50

発症を定義すること、または、形質の正常な集団の測定値からの偏差の定義された統計的カットオフを超えた（例えば、集団平均を2標準偏差上回る）定量的形質の最小値もしくは最大値もしくは中央値を測定することができる。1つの例示的な実施形態において、特定の定量的形質に関して電子健康記録データが利用できない可能性のある個体、または特定の定量的形質に関して矛盾するもしくは信頼できないデータを有する個体は、医療記録に特定の定量的形質が存在するかしないかにかかわらず、未知の罹患者として判定され得る。

【0029】

いくつかの例示的な実施形態において、疾患原因バリエーションを同定するための方法は、家系内の個体を罹患者または非罹患者として同定することであって、少なくとも1つのバイナリー形質、極端な定量的形質、またはこれらの組合せを有する個体が罹患者として同定され、少なくとも1つのバイナリー形質、極端な定量的形質、またはこれらの組合せを有しない個体が非罹患者として同定される、同定することを含み得る。バイナリー形質は、上記のような定義されたICDコードとすることができる。いくつかのパラメータを使用して、上記のような極端な定量的形質を定義することができる。1つの例示的な実施形態において、特定のバイナリー形質、定量的形質、もしくはこれらの組合せに関して電子健康記録データが利用できない個体、または特定のバイナリー形質、定量的形質、もしくはこれらの組合せに関して矛盾するもしくは信頼できないデータを有する個体は、医療記録に特定の定量的形質が存在するかしないかにかかわらず、未知の罹患者として判定され得る。

【0030】

いくつかの例示的な実施形態において、疾患原因バリエーションを同定するための方法は、家系内の個体を罹患者または非罹患者として同定することであって、少なくとも1つのバイナリー形質、極端な定量的形質、またはこれらの組合せを有する個体が罹患者として同定され、少なくとも1つのバイナリー形質、極端な定量的形質、またはこれらの組合せを有しない個体が非罹患者として同定される、同定することを含み得、このとき、少なくとも1つのバイナリー形質、極端な定量的形質、またはこれらの組合せは、2つ以上の類似のまたは補完的な形質を含み得る。

【0031】

いくつかの例示的な実施形態において、疾患原因バリエーションを同定するための方法は、家系内の個体を罹患者または非罹患者として同定することであって、少なくとも1つのバイナリー形質、極端な定量的形質、またはこれらの組合せを有する個体が罹患者として同定され、少なくとも1つのバイナリー形質、極端な定量的形質、またはこれらの組合せを有しない個体が非罹患者として同定される、同定することを含み得、このとき、少なくとも1つのバイナリー形質、極端な定量的形質、またはこれらの組合せは、2つ以上の極端なまたは興味深い形質の共通部分を採用することを含み得る。

【0032】

いくつかの例示的な実施形態において、疾患原因バリエーションを同定するための方法は、家系内の個体を罹患者として同定することであって、少なくとも1つのバイナリー形質、極端な定量的形質、またはこれらの組合せを有する個体が罹患者として同定される、同定することと、罹患者として判定された個体を定義することと、罹患者として判定された個体を、外部解析からの関連性結果の罹患キャリアとして定義することとを含み得る。

【0033】

いくつかの例示的な実施形態において、疾患原因バリエーションを同定するための方法は、コホートのシーケンシングデータに基づいて個体の1親等ネットワークを作成することを含む。シーケンシングデータは、全ゲノムシーケンシングデータ、エクソームシーケンシングデータ、または遺伝子型データを含み得る。

【0034】

いくつかの例示的な実施形態において、疾患原因バリエーションを同定するための方法は、エクソームシーケンシングデータに基づいて個体の1親等ネットワークを作成すること

を含む。エクソームシーケンシングデータに基づく個体の1親等ネットワークは、以下を含めた集団の関連性を活用することによって作成することができる：複数のヒト対象から得られた核酸配列サンプルのデータセットから低品質の配列バリエーションを削除すること、サンプルのうちの1例以上の各々に対する祖先スーパークラス指定を確立すること、データセットから低品質のサンプルを削除すること、祖先スーパークラス内の対象の第1の家系同一性推定値を作成すること、対象の祖先スーパークラスとは無関係に第2の家系同一性推定値を作成すること、及び第2の家系同一性推定値のうちの1つ以上に基づき、対象を主要な1親等家族ネットワークにクラスター化すること。

【0035】

いくつかの例示的な実施形態において、疾患原因バリエーションを同定するための方法は、コホートのシーケンシングデータに基づいて個体の1親等ネットワークを作成することを含み、このときコホートは、複数の対象を含む任意のデータセットを含み得る。

10

【0036】

いくつかの例示的な実施形態において、エンリッチされた家系を作出するための方法はさらに、p値に基づいて家系をエンリッチすることを含む。エンリッチメントは、家系の「創始者アンカー分枝」または「分枝」を家系内の創始者の全ての子孫として定義し、二項検定を使用して分枝のバイナリー形質がエンリッチされているかを評価することを含み得る。バイナリー形質は、上記のようなICDを用いて定義することができる。また、エンリッチメントは、家系の「創始者アンカー分枝」または「分枝」を家系内の創始者の全ての子孫として定義し、t検定を使用して分枝の極端な定量的形質がエンリッチされているかを評価することも含み得る。いくつかのパラメータを使用して、上記のような極端な定量的形質を定義することができる。さらに、エンリッチメントは、多重検定p値カットオフを適用することも含み得る。

20

【0037】

いくつかの例示的な実施形態において、疾患原因バリエーションを同定するための方法は、家系内で罹患者と共に同時分離するバリエーション形質の対を同定することと、表現型分離に基づいて少なくとも1つのエンリッチされた家系を見出すことを含む分離解析を実施することとを含み得る。この分離には、優性及び相加的分離モデルならびに劣性分離モデルが含まれ得る。1つの例示的な実施形態において、優性及び相加的分離モデルに基づいて少なくとも1つのエンリッチされた家系を見出すことは、1つの可能な構造と、共通の祖先を有する少なくとも3例の罹患者とを有する家系を選択することを含む。これはさらに、偽陽性を減らすために、1例以上の近親の非罹患者を有する少なくとも1つのエンリッチされた家系を選択することを含んでもよい。別の例示的な実施形態において、劣性分離モデルに基づいて少なくとも1つのエンリッチされた家系を見出すことは、1つの可能な構造と、非罹患の親を有する1例を超える罹患者とを有する家系を選択することを含む。これはさらに、偽陽性を減らすために、少なくとも2例の罹患した同胞を有する少なくとも1つのエンリッチされた家系を選択することを含んでもよい。

30

【0038】

いくつかの例示的な実施形態において、疾患原因バリエーションを同定するための方法は、分離の特定の遺伝モデルを形成するために分離解析を実施することを含む。分離の特定の遺伝モデルには、分離の優性遺伝モデルまたは分離の劣性遺伝モデルが含まれ得る。加えて、分離の特定の遺伝モデルには、他の遺伝様式に基づく分離の遺伝モデル、例えば、Y連鎖、多因子、またはミトコンドリア連鎖遺伝様式も含まれ得る。1つの例示的な実施形態において、疾患原因バリエーションを同定するための方法は、分離の優性遺伝モデルを形成するために分離解析を実施することであって、当該疾患原因バリエーションが、少なくとも1つのバイナリー形質、極端な定量的形質、またはこれらの組合せにおいて罹患者と共に分離する、実施することを含む。1つの例示的な実施形態において、疾患原因バリエーションを同定するための方法は、分離の劣性遺伝モデルを形成するために分離解析を実施することであって、当該疾患原因バリエーションが、所与の遺伝子において両アレルバリエーションキャリアである罹患者と共に分離する、実施することを含み、親の遺伝子データが入手可能な場

40

50

合、これらは同定された疾患原因バリエントに対しヘテロ接合性でなければならない。

【0039】

いくつかの例示的な実施形態において、疾患原因バリエントを同定するための方法は、少なくとも1つのエンリッチされた家系の内部でまたはそれにわたって同時分離するバリエント形質の対を同定するために分離解析を実施することを含み得る。1つの例示的な実施形態において、疾患原因バリエントを同定するための方法は、複数のエンリッチされた家系の内部でまたはそれにわたって同時分離するバリエント形質の対を同定するための分離解析を含む。

【0040】

いくつかの例示的な実施形態において、疾患原因バリエントを同定するための方法は、家族構造に含まれない目的表現型について他の罹患者内の分離するバリエントまたは遺伝子を同定するために分離解析を実施することを含み得る。

10

【0041】

いくつかの例示的な実施形態において、疾患原因バリエントを同定するための方法は、分離解析を実施することを含み得、分離解析は、集団規模解析からの関連性結果と共にバリエント及び形質を相互参照することを含む。

【0042】

いくつかの例示的な実施形態において、疾患原因バリエントを同定するための方法は、以前から知られている原因バリエント及び原因遺伝子を同定するために分離解析を実施することを含み得る。

20

【0043】

いくつかの例示的な実施形態において、疾患原因バリエントを同定するための方法は、さらに、支持する家系／罹患者の数により、ならびに候補となる原因バリエント及び原因遺伝子の数により、エンリッチされた家系を優先順位付けすることを含み得る。

【0044】

いくつかの例示的な実施形態において、疾患原因バリエントを同定するための方法は、バリエント形質の対を解析することを含み得、これはさらに、家族ベースの関連性解析を保証するのに十分な家族データで罹患者のセットを同定することを含む。

【0045】

いくつかの例示的な実施形態において、疾患原因バリエントを同定するための方法は、バリエント形質の対を解析することを含み得、これは家系及び表現型の情報に基づいて、伝達不平衡試験（TDT）または必要に応じてその他の解析を実施することを含む。

30

【0046】

いくつかの例示的な実施形態において、疾患原因バリエントを同定するための方法は、いくつかの生理学的障害の疾患原因バリエントを同定するための方法を含み得る。

1つの例示的な態様において、本開示は、エンリッチされた家系を作成するための方法をプロセッサに実施させるための命令を格納する非一時的コンピューター可読媒体を提供し、これは、コホートのエクソームシーケンシングデータに基づいて個体の1親等ネットワークを作成することと、当該1親等ネットワーク内の個体を罹患者または非罹患者として同定することと、罹患者または非罹患者としての指定を含む当該個体を包含する少なくとも1つのエンリッチされた家系を作成することを含む。

40

【0047】

いくつかの例示的な実施形態において、エンリッチされた家系を作成するための方法をプロセッサに実施させるための命令を格納する非一時的コンピューター可読媒体は、当該家系内の個体が罹患者または非罹患者であるかどうかを同定することであって、少なくとも1つのバイナリー形質を有する当該個体が罹患者として同定され、当該少なくとも1つのバイナリー形質を有しない当該個体が非罹患者として同定される、同定することと、次いで、罹患者及び非罹患者のパターンがメンデル遺伝様式（例えば、常染色体優性、常染色体劣性、x連鎖優性、x連鎖劣性、またはy連鎖）と一致するかどうかを評価することを含む。いくつかの特定の例示的な実施形態において、バイナリー形質は、世界保健

50

機関（WHO）による医学分類リストである疾病及び関連保健問題の国際統計分類（ICD）を用いて定義することができる。この分類には、疾患、徴候及び症状、異常所見、病訴、社会的状況、ならびに傷害または疾患の外部原因に関するコードが含まれる。バイナリー形質の定義には、ICDの第9版または第10版を使用することができる。1つの例示的な実施形態において、特定のバイナリー形質に関して電子健康記録データが利用できない可能性のある個体、または特定のバイナリー形質に関して矛盾するもしくは信頼できないデータを有する個体は、医療記録に特定のバイナリー形質が存在するかしないかにかかわらず、未知の罹患者として判定され得る。

【0048】

いくつかの例示的な実施形態において、エンリッチされた家系を作成するための方法をプロセッサに実施させるための命令を格納する非一時的コンピューター可読媒体は、当該家系内の個体が罹患者または非罹患者であるかどうかを同定することであって、少なくとも1つの極端な定量的形質を有する当該個体が罹患者として同定され、当該少なくとも1つの極端な定量的形質を有しない当該個体が非罹患者として同定される、当該同定することと、次いで、罹患者及び非罹患者のパターンがメンデル遺伝様式（例えば、常染色体優性、常染色体劣性、x連鎖優性、x連鎖劣性、またはy連鎖）と一致するかどうかを評価することを含む。いくつかのパラメーターを使用して、極端な定量的形質の影響を受けているかどうかを定義することができる。例えば、最大年齢カットオフを使用してより早期の障害発症を定義すること、または、形質の正常な集団の測定値からの偏差の定義された統計的カットオフを超えた（例えば、集団平均を2標準偏差上回る）定量的形質の最小値もしくは最大値もしくは中央値を測定することができる。1つの例示的な実施形態において、特定の定量的形質に関して電子健康記録データが利用できない可能性のある個体、または特定の定量的形質に関して矛盾するもしくは信頼できないデータを有する個体は、医療記録に特定の定量的形質が存在するかしないかにかかわらず、未知の罹患者として判定され得る。

【0049】

いくつかの例示的な実施形態において、エンリッチされた家系を作成するための方法をプロセッサに実施させるための命令を格納する非一時的コンピューター可読媒体は、家系内の個体が罹患者または非罹患者であるかどうかを同定することであって、少なくとも1つのバイナリー形質、極端な定量的形質、またはこれらの組合せを有する個体が罹患者として同定され、少なくとも1つのバイナリー形質、極端な定量的形質、またはこれらの組合せを有しない個体が非罹患者として同定される、同定することを含む。バイナリー形質は、上記のような定義されたICDコードとすることができる。いくつかのパラメーターを使用して、上記のような極端な定量的形質を定義することができる。1つの例示的な実施形態において、特定のバイナリー形質、定量的形質、もしくはこれらの組合せに関して電子健康記録データが利用できない個体、または特定のバイナリー形質、定量的形質、もしくはこれらの組合せに関して矛盾するもしくは信頼できないデータを有する個体は、医療記録に特定の定量的形質が存在するかしないかにかかわらず、未知の罹患者として判定され得る。

【0050】

いくつかの例示的な実施形態において、エンリッチされた家系を作成するための方法をプロセッサに実施させるための命令を格納する非一時的コンピューター可読媒体は、家系内の個体が罹患者または非罹患者であるかどうかを同定することであって、少なくとも1つのバイナリー形質、極端な定量的形質、またはこれらの組合せを有する個体が罹患者として同定され、少なくとも1つのバイナリー形質、極端な定量的形質、またはこれらの組合せを有しない個体が非罹患者として同定される、同定することを含み、このとき、当該少なくとも1つのバイナリー形質、極端な定量的形質、またはこれらの組合せは、2つ以上の類似のまたは補完的な形質を含み得る。

【0051】

いくつかの例示的な実施形態において、エンリッチされた家系を作成するための方法を

プロセッサに実施させるための命令を格納する非一時的コンピューター可読媒体は、家系内の個体が罹患者または非罹患者であるかどうかを同定することであって、少なくとも1つのバイナリー形質、極端な定量的形質、またはこれらの組合せを有する個体が罹患者として同定され、少なくとも1つのバイナリー形質、極端な定量的形質、またはこれらの組合せを有しない個体が非罹患者として同定される、同定することを含み、このとき、当該少なくとも1つのバイナリー形質、極端な定量的形質、またはこれらの組合せは、2つ以上の極端なまたは興味深い形質の共通部分を採用することを含み得る。

【0052】

いくつかの例示的な実施形態において、エンリッチされた家系を作成するための方法をプロセッサに実施させるための命令を格納する非一時的コンピューター可読媒体はさらに、コホート内の個体が、少なくとも1つのバイナリー形質、極端な定量的形質、またはこれらの組合せを有する場合、当該個体を罹患者として同定することと、罹患者として判定された個体を、外部解析からの関連性結果の罹患キャリアとして定義することを含み得る。

10

【0053】

いくつかの例示的な実施形態において、エンリッチされた家系を作成するための方法をプロセッサに実施させるための命令を格納する非一時的コンピューター可読媒体は、コホートのシーケンシングデータに基づいて個体の1親等ネットワークを作成することを含む。シーケンシングデータは、全ゲノムシーケンシングデータ、エクソームシーケンシングデータ、または遺伝子型データを含み得る。

20

【0054】

いくつかの例示的な実施形態における、エクソームシーケンシングデータに基づいてエンリッチされた家系を作成するための方法をプロセッサに実施させるための命令を格納する、非一時的コンピューター可読媒体。エクソームシーケンシングデータに基づく個体の1親等ネットワークは、以下を含めた集団の関連性を活用することによって作成することができる：複数のヒト対象から得られた核酸配列サンプルのデータセットから低品質の配列バリエーションを削除すること、サンプルのうちの1例以上の各々に対する祖先スーパークラス指定を確立すること、データセットから低品質のサンプルを削除すること、祖先スーパークラス内の対象の第1の家系同一性推定値を作成すること、対象の祖先スーパークラスとは無関係に第2の家系同一性推定値を作成すること、及び第2の家系同一性推定値のうちの1つ以上に基づき、対象を主要な1親等家族ネットワークにクラスター化すること。

30

【0055】

いくつかの例示的な実施形態において、エンリッチされた家系を作成するための方法をプロセッサに実施させるための命令を格納する非一時的コンピューター可読媒体は、コホートのシーケンシングデータに基づいて個体の1親等ネットワークを作成することを含み得、このとき、当該コホートは、複数の対象を含む任意のデータセットを含み得る。

【0056】

いくつかの例示的な実施形態において、エンリッチされた家系を作成するための方法をプロセッサに実施させるための命令を格納する非一時的コンピューター可読媒体は、さらに、p値に基づいて家系をエンリッチすることを含み得る。エンリッチメントは、家系の「創始者アンカー分枝」または「分枝」を家系内の創始者の全ての子孫として定義し、二項検定を使用して分枝のバイナリー形質がエンリッチされているかを評価することを含み得る。バイナリー形質は、上記のようなICDを用いて定義することができる。また、エンリッチメントは、家系の「創始者アンカー分枝」または「分枝」を家系内の創始者の全ての子孫として定義し、t検定を使用して分枝の極端な定量的形質がエンリッチされているかを評価することを含み得る。いくつかのパラメーターを使用して、上記のような極端な定量的形質を定義することができる。さらに、エンリッチメントは、多重検定p値カットオフを適用することを含み得る。

40

【0057】

50

1つの例示的な態様において、本開示は、疾患原因バリエーションを同定するための方法をプロセッサに実施させるための命令を格納する非一時的コンピューター可読媒体を提供し、これは、コホートのエクソームシーケンシングデータに基づいて個体の1親等ネットワークを作成することと、当該1親等ネットワーク内の個体を罹患者または非罹患者として同定することと、罹患者または非罹患者としての指定を含む当該個体を含む少なくとも1つのエンリッチされた家系を作出することと、少なくとも1つのエンリッチされた家系の内部でまたはそれにわたって同時分離するバリエーション形質の対を同定するために分離解析を実施することと、当該バリエーション形質の対を解析して疾患原因バリエーションを決定することを含む。

【0058】

いくつかの例示的な実施形態において、疾患原因バリエーションを同定するための方法をプロセッサに実施させるための命令を格納する非一時的コンピューター可読媒体は、当該家系内の個体が罹患者または非罹患者であるかどうかを同定することであって、少なくとも1つのバイナリー形質を有する当該個体が罹患者として同定され、当該少なくとも1つのバイナリー形質を有しない当該個体が非罹患者として同定される、同定することと、次いで、罹患者及び非罹患者のパターンがメンデル遺伝様式（例えば、常染色体優性、常染色体劣性、x連鎖優性、x連鎖劣性、またはy連鎖）と一致するかどうかを評価することを含む。いくつかの特定の例示的な実施形態において、バイナリー形質は、世界保健機関（WHO）による医学分類リストである疾病及び関連保健問題の国際統計分類（ICD）を用いて定義することができる。この分類には、疾患、徴候及び症状、異常所見、病歴、社会的状況、ならびに傷害または疾患の外部原因に関するコードが含まれる。バイナリー形質の定義には、ICDの第9版または第10版を使用することができる。1つの例示的な実施形態において、特定のバイナリー形質に関して電子健康記録データが利用できない可能性のある個体、または特定のバイナリー形質に関して矛盾するもしくは信頼できないデータを有する個体は、医療記録に特定のバイナリー形質が存在するかしないかにかかわらず、未知の罹患者として判定され得る。

【0059】

いくつかの例示的な実施形態において、疾患原因バリエーションを同定するための方法をプロセッサに実施させるための命令を格納する非一時的コンピューター可読媒体は、当該家系内の個体が罹患者または非罹患者であるかどうかを同定することであって、少なくとも1つの極端な定量的形質を有する当該個体が罹患者として同定され、当該少なくとも1つの極端な定量的形質を有しない当該個体が非罹患者として同定される、当該同定することと、次いで、罹患者及び非罹患者のパターンがメンデル遺伝様式（例えば、常染色体優性、常染色体劣性、x連鎖優性、x連鎖劣性、またはy連鎖）と一致するかどうかを評価することを含む。いくつかのパラメーターを使用して、極端な定量的形質の影響を受けているかどうかを定義することができる。例えば、最大年齢カットオフを使用してより早期の障害発症を定義すること、または、形質の正常な集団の測定値からの偏差の定義された統計的カットオフを超えた（例えば、集団平均を2標準偏差上回る）定量的形質の最小値もしくは最大値もしくは中央値を測定することができる。1つの例示的な実施形態において、特定の定量的形質に関して電子健康記録データが利用できない可能性のある個体、または特定の定量的形質に関して矛盾するもしくは信頼できないデータを有する個体は、医療記録に特定の定量的形質が存在するかしないかにかかわらず、未知の罹患者として判定され得る。

【0060】

いくつかの例示的な実施形態において、疾患原因バリエーションを同定するための方法をプロセッサに実施させるための命令を格納する非一時的コンピューター可読媒体は、家系内の個体が罹患者または非罹患者であるかどうかを同定することであって、少なくとも1つのバイナリー形質、極端な定量的形質、またはこれらの組合せを有する個体が罹患者として同定され、少なくとも1つのバイナリー形質、極端な定量的形質、またはこれらの組合せを有しない個体が非罹患者として同定される、同定することを含む。バイナリー形質

は、上記のような定義された ICDコードとすることができる。いくつかのパラメーターを使用して、上記のような極端な定量的形質を定義することができる。1つの例示的な実施形態において、特定のバイナリー形質、定量的形質、もしくはこれらの組合せに関して電子健康記録データが利用できない個体、または特定のバイナリー形質、定量的形質、もしくはこれらの組合せに関して矛盾するもしくは信頼できないデータを有する個体は、医療記録に特定の定量的形質が存在するかしないかにかかわらず、未知の罹患者として判定され得る。

【0061】

いくつかの例示的な実施形態において、疾患原因バリエーションを同定するための方法をプロセッサに実施させるための命令を格納する非一時的コンピューター可読媒体は、家系内の個体が罹患者または非罹患者であるかどうかを同定することであって、少なくとも1つのバイナリー形質、極端な定量的形質、またはこれらの組合せを有する個体が罹患者として同定され、少なくとも1つのバイナリー形質、極端な定量的形質、またはこれらの組合せを有しない個体が非罹患者として同定される、同定することを含み、このとき、当該少なくとも1つのバイナリー形質、極端な定量的形質、またはこれらの組合せは、2つ以上の類似のまたは補完的な形質を含み得る。

【0062】

いくつかの例示的な実施形態において、疾患原因バリエーションを同定するための方法をプロセッサに実施させるための命令を格納する非一時的コンピューター可読媒体は、家系内の個体が罹患者または非罹患者であるかどうかを同定することであって、少なくとも1つのバイナリー形質、極端な定量的形質、またはこれらの組合せを有する個体が罹患者として同定され、少なくとも1つのバイナリー形質、極端な定量的形質、またはこれらの組合せを有しない個体が非罹患者として同定される、同定することを含み、このとき、当該少なくとも1つのバイナリー形質、極端な定量的形質、またはこれらの組合せは、2つ以上の極端なまたは興味深い形質の共通部分を採用することを含み得る。

【0063】

いくつかの例示的な実施形態において、疾患原因バリエーションを同定するための方法をプロセッサに実施させるための命令を格納する非一時的コンピューター可読媒体はさらに、コホート内の個体が、少なくとも1つのバイナリー形質、極端な定量的形質、またはこれらの組合せを有する場合、当該個体を罹患者として同定することと、罹患者として判定された個体を、外部解析からの関連性結果の罹患者キャリアとして定義することとを含み得る。

【0064】

いくつかの例示的な実施形態において、疾患原因バリエーションを同定するための方法をプロセッサに実施させるための命令を格納する非一時的コンピューター可読媒体は、コホートのシーケンシングデータに基づいて個体の1親等ネットワークを作成することを含む。シーケンシングデータは、全ゲノムシーケンシングデータ、エクソームシーケンシングデータ、または遺伝子型データを含み得る。

【0065】

いくつかの例示的な実施形態における、エクソームシーケンシングデータに基づいて疾患原因バリエーションを同定するための方法をプロセッサに実施させるための命令を格納する、非一時的コンピューター可読媒体。エクソームシーケンシングデータに基づく個体の1親等ネットワークは、以下を含めた集団の関連性を活用することによって作成することができる：複数のヒト対象から得られた核酸配列サンプルのデータセットから低品質の配列バリエーションを削除すること、サンプルのうちの1例以上の各々に対する祖先スーパークラス指定を確立すること、データセットから低品質のサンプルを削除すること、祖先スーパークラス内の対象の第1の家系同一性推定値を作成すること、対象の祖先スーパークラスとは無関係に第2の家系同一性推定値を作成すること、及び第2の家系同一性推定値のうちの1つ以上に基づき、対象を主要な1親等家族ネットワークにクラスター化すること。

10

20

30

40

50

【0066】

いくつかの例示的な実施形態において、疾患原因バリエーションを同定するための方法をプロセッサに実施させるための命令を格納する非一時的コンピューター可読媒体は、コホートのシーケンシングデータに基づいて個体の1親等ネットワークを作成することを含み得、このとき、当該コホートは、複数の対象を含む任意のデータセットを含み得る。

【0067】

いくつかの例示的な実施形態において、疾患原因バリエーションを同定するための方法をプロセッサに実施させるための命令を格納する非一時的コンピューター可読媒体は、さらに、p値に基づいて家系をエンリッチすることを含み得る。エンリッチメントは、家系の「創始者アンカー分枝」または「分枝」を家系内の創始者の全ての子孫として定義し、二項検定を使用して分枝のバイナリー形質がエンリッチされているかを評価することを含み得る。バイナリー形質は、上記のようなICDを用いて定義することができる。また、エンリッチメントは、家系の「創始者アンカー分枝」または「分枝」を家系内の創始者の全ての子孫として定義し、t検定を使用して分枝の極端な定量的形質がエンリッチされているかを評価することも含み得る。いくつかのパラメーターを使用して、上記のような極端な定量的形質を定義することができる。さらに、エンリッチメントは、多重検定p値カットオフを適用することも含み得る。

【0068】

いくつかの例示的な実施形態において、疾患原因バリエーションを同定するための方法をプロセッサに実施させるための命令を格納する非一時的コンピューター可読媒体は、家系内で罹患者と共に同時分離するバリエーション形質の対を同定することと、表現型分離に基づいて少なくとも1つのエンリッチされた家系を見出すことを含む分離解析を実施することとを含み得る。この分離には、優性及び相加的分離モデルならびに劣性分離モデルが含まれ得る。1つの例示的な実施形態において、優性及び相加的分離モデルに基づいて少なくとも1つのエンリッチされた家系を見出すことは、1つの可能な構造と、共通の祖先を有する少なくとも3例の罹患者とを有する家系を選択することを含む。これはさらに、偽陽性を減らすために、1例以上の近親の非罹患者を有する少なくとも1つのエンリッチされた家系を選択することを含んでもよい。別の例示的な実施形態において、劣性分離モデルに基づいて少なくとも1つのエンリッチされた家系を見出すことは、1つの可能な構造と、非罹患の親を有する1例を超える罹患者とを有する家系を選択することを含む。これはさらに、偽陽性を減らすために、少なくとも2例の罹患した同胞を有する少なくとも1つのエンリッチされた家系を選択することを含んでもよい。

【0069】

いくつかの例示的な実施形態において、疾患原因バリエーションを同定するための方法をプロセッサに実施させるための命令を格納する非一時的コンピューター可読媒体は、分離の特定の遺伝モデルを形成するために分離解析を実施することを含み得る。分離の特定の遺伝モデルには、分離の優性遺伝モデルまたは分離の劣性遺伝モデルが含まれ得る。加えて、分離の特定の遺伝モデルには、他の遺伝様式に基づく分離の遺伝モデル、例えば、Y連鎖、多因子、またはミトコンドリア連鎖遺伝様式も含まれ得る。1つの例示的な実施形態において、疾患原因バリエーションを同定するための方法は、分離の優性遺伝モデルを形成するために分離解析を実施することであって、当該疾患原因バリエーションが、少なくとも1つのバイナリー形質、極端な定量的形質、またはこれらの組合せにおいて罹患者と共に分離する、実施することを含む。1つの例示的な実施形態において、疾患原因バリエーションを同定するための方法は、分離の劣性遺伝モデルを形成するために分離解析を実施することであって、当該疾患原因バリエーションが、所与の遺伝子において両アレルバリエーションキャリアである罹患者と共に分離する、実施することを含み、親の遺伝子データが入手可能な場合、これらは同定された疾患原因バリエーションに対しヘテロ接合性でなければならない。

【0070】

いくつかの例示的な実施形態において、疾患原因バリエーションを同定するための方法をプロセッサに実施させるための命令を格納する非一時的コンピューター可読媒体は、少な

10

20

30

40

50

くとも1つのエンリッチされた家系の内部でまたはそれにわたって同時分離するバリエーション形質の対を同定するために分離解析を実施することを含み得る。1つの例示的な実施形態において、疾患原因バリエーションを同定するための方法は、複数のエンリッチされた家系の内部でまたはそれにわたって同時分離するバリエーション形質の対を同定するための分離解析を含む。

【0071】

いくつかの例示的な実施形態において、疾患原因バリエーションを同定するための方法をプロセッサに実施させるための命令を格納する非一時的コンピューター可読媒体は、家族構造に含まれない目的表現型について他の罹患者内の分離バリエーションまたは遺伝子を同定するために分離解析を実施することを含み得る。

10

【0072】

いくつかの例示的な実施形態において、疾患原因バリエーションを同定するための方法をプロセッサに実施させるための命令を格納する非一時的コンピューター可読媒体は、分離解析を実施することを含み得、これは、集団規模解析からの関連性結果と共にバリエーション及び形質を相互参照することを含む。

【0073】

いくつかの例示的な実施形態において、疾患原因バリエーションを同定するための方法をプロセッサに実施させるための命令を格納する非一時的コンピューター可読媒体は、以前から知られている原因バリエーション及び原因遺伝子を同定するために分離解析を実施することを含み得る。

20

【0074】

いくつかの例示的な実施形態において、疾患原因バリエーションを同定するための方法をプロセッサに実施させるための命令を格納する非一時的コンピューター可読媒体は、支持する家系／罹患者の数により、ならびに候補となる原因バリエーション及び原因遺伝子の数により、エンリッチされた家系を優先順位付けすることを含み得る。

【0075】

いくつかの例示的な実施形態において、疾患原因バリエーションを同定するための方法をプロセッサに実施させるための命令を格納する非一時的コンピューター可読媒体は、バリエーション形質の対を解析することを含み得、これはさらに、家族ベースの関連性解析を保証するのに十分な家族データで罹患者のセットを同定することを含む。

30

【0076】

いくつかの例示的な実施形態において、疾患原因バリエーションを同定するための方法をプロセッサに実施させるための命令を格納する非一時的コンピューター可読媒体は、バリエーション形質の対を解析することを含み得、これは家系及び表現型の情報に基づいて、伝達不平衡試験(TDT)または必要に応じてその他の解析を実施することを含む。

【0077】

いくつかの例示的な実施形態における、いくつかの生理学的障害についての疾患原因バリエーションを同定するための方法をプロセッサに実施させるための命令を格納する、非一時的コンピューター可読媒体。

【0078】

1つの例示的な態様において、本開示は、エンリッチされた家系を作成するためのシステムであって、データプロセッサと、当該データプロセッサに連結したメモリーとを備え、当該プロセッサが、コホートのシークエンシングデータに基づいて個体の1親等ネットワークを作成することと、当該1親等ネットワーク内の個体が罹患者または非罹患者であるかを同定することと、罹患者または非罹患者としての指定を含む当該個体を含む少なくとも1つのエンリッチされた家系を作成することとを行うように構成された、システムを提供する。

40

【0079】

いくつかの例示的な実施形態において、エンリッチされた家系を作成するためのシステムは、データプロセッサと、当該データプロセッサに連結したメモリーとを備え、当

50

該プロセッサは、当該家系内の個体が罹患者または非罹患者であるかどうかを同定することであって、少なくとも1つのバイナリー形質を有する当該個体が罹患者として同定され、当該少なくとも1つのバイナリー形質を有しない当該個体が非罹患者として同定される、同定することと、次いで、罹患者及び非罹患者のパターンがメンデル遺伝様式（例えば、常染色体優性、常染色体劣性、x連鎖優性、x連鎖劣性、またはy連鎖）と一致するかどうかを評価することとを行うように構成されている。いくつかの特定の例示的な実施形態において、バイナリー形質は、世界保健機関（WHO）による医学分類リストである疾病及び関連保健問題の国際統計分類（ICD）を用いて定義することができる。この分類には、疾患、徴候及び症状、異常所見、病訴、社会的状況、ならびに傷害または疾患の外部原因に関するコードが含まれる。バイナリー形質の定義には、ICDの第9版または第10版を使用することができる。1つの例示的な実施形態において、特定のバイナリー形質に関して電子健康記録データが利用できない可能性のある個体、または特定のバイナリー形質に関して矛盾するもしくは信頼できないデータを有する個体は、医療記録に特定のバイナリー形質が存在するかしなにかにかかわらず、未知の罹患者として判定され得る。

10

【0080】

いくつかの例示的な実施形態において、エンリッチされた家系を作成するためのシステムは、データプロセッサと、当該データプロセッサに連結したメモリとを備え、当該プロセッサは、当該家系内の個体が罹患者または非罹患者であるかどうかを同定することであって、少なくとも1つの極端な定量的形質を有する当該個体が罹患者として同定され、当該少なくとも1つの極端な定量的形質を有しない当該個体が非罹患者として同定される、同定することと、次いで、罹患者及び非罹患者のパターンがメンデル遺伝様式（例えば、常染色体優性、常染色体劣性、x連鎖優性、x連鎖劣性、またはy連鎖）と一致するかどうかを評価することとを行うように構成されている。いくつかのパラメータを使用して、極端な定量的形質の影響を受けているかどうかを定義することができる。例えば、最大年齢カットオフを使用してより早期の障害発症を定義すること、または、形質の正常な集団の測定値からの偏差の定義された統計的カットオフを超えた（例えば、集団平均を2標準偏差上回る）定量的形質の最小値もしくは最大値もしくは中央値を測定することができる。1つの例示的な実施形態において、特定の定量的形質に関して電子健康記録データが利用できない可能性のある個体、または特定の定量的形質に関して矛盾するもしくは信頼できないデータを有する個体は、医療記録に特定の定量的形質が存在するかしなにかにかかわらず、未知の罹患者として判定され得る。

20

30

【0081】

いくつかの例示的な実施形態において、エンリッチされた家系を作成するためのシステムは、データプロセッサと、当該データプロセッサに連結したメモリとを備え、当該プロセッサは、家系内の個体が罹患者または非罹患者であるかどうかを同定することであって、少なくとも1つのバイナリー形質、極端な定量的形質、またはこれらの組合せを有する個体が罹患者として同定され、少なくとも1つのバイナリー形質、極端な定量的形質、またはこれらの組合せを有しない個体が非罹患者として同定される、同定することとを行うように構成されている。バイナリー形質は、上記のような定義されたICDコードとすることができる。いくつかのパラメータを使用して、上記のような極端な定量的形質を定義することができる。1つの例示的な実施形態において、特定のバイナリー形質、定量的形質、もしくはこれらの組合せに関して電子健康記録データが利用できない個体、または特定のバイナリー形質、定量的形質、もしくはこれらの組合せに関して矛盾するもしくは信頼できないデータを有する個体は、医療記録に特定の定量的形質が存在するかしなにかにかかわらず、未知の罹患者として判定され得る。

40

【0082】

いくつかの例示的な実施形態において、エンリッチされた家系を作成するためのシステムは、データプロセッサと、当該データプロセッサに連結したメモリとを備え、当該プロセッサは、家系内の個体が罹患者または非罹患者であるかどうかを同定すること

50

であって、少なくとも1つのバイナリー形質、極端な定量的形質、またはこれらの組合せを有する個体が罹患者として同定され、少なくとも1つのバイナリー形質、極端な定量的形質、またはこれらの組合せを有しない個体が非罹患者として同定される、同定することを行うように構成されており、このとき、当該少なくとも1つのバイナリー形質、極端な定量的形質、またはこれらの組合せは、2つ以上の類似のまたは補完的な形質を含み得る。

【0083】

いくつかの例示的な実施形態において、エンリッチされた家系を作成するためのシステムは、データプロセッサと、当該データプロセッサに連結したメモリとを備え、当該プロセッサは、家系内の個体が罹患者または非罹患者であるかどうかを同定すること
10
であって、少なくとも1つのバイナリー形質、極端な定量的形質、またはこれらの組合せを有する個体が罹患者として同定され、少なくとも1つのバイナリー形質、極端な定量的形質、またはこれらの組合せを有しない個体が非罹患者として同定される、同定することを行うように構成されており、このとき、当該少なくとも1つのバイナリー形質、極端な定量的形質、またはこれらの組合せは、2つ以上の極端なまたは興味深い形質の共通部分を採用することを含み得る。

【0084】

いくつかの例示的な実施形態において、エンリッチされた家系を作成するためのシステムは、データプロセッサと、当該データプロセッサに連結したメモリとを備え、当該プロセッサは、コホート内の個体が、少なくとも1つのバイナリー形質、極端な定量的形質、またはこれらの組合せを有する場合、当該個体を罹患者として同定することと、
20
罹患者として判定された個体を、外部解析からの関連性結果の罹患キャリアとして定義することを行うように構成されている。

【0085】

いくつかの例示的な実施形態において、エンリッチされた家系を作成するためのシステムは、データプロセッサと、当該データプロセッサに連結したメモリとを備え、当該プロセッサは、コホートのシークエンシングデータに基づいて個体の1親等ネットワークを作成するように構成されている。シークエンシングデータは、全ゲノムシークエンシングデータ、エクソームシークエンシングデータ、または遺伝子型データを含み得る。

【0086】

いくつかの例示的な実施形態において、エンリッチされた家系を作成するためのシステムは、データプロセッサと、当該データプロセッサに連結したメモリとを備え、当該プロセッサは、エクソームシークエンシングデータに基づいて個体の1親等ネットワークを作成するように構成されている。エクソームシークエンシングデータに基づく個体の1親等ネットワークは、以下を含めた集団の関連性を活用することによって作成することができる：複数のヒト対象から得られた核酸配列サンプルのデータセットから低品質の配列バリエーションを削除すること、サンプルのうちの1例以上の各々に対する祖先スーパー
30
クラス指定を確立すること、データセットから低品質のサンプルを削除すること、祖先スーパークラス内の対象の第1の家系同一性推定値を作成すること、対象の祖先スーパークラスとは無関係に第2の家系同一性推定値を作成すること、及び第2の家系同一性推定値
40
のうちの1つ以上に基づき、対象を主要な1親等家族ネットワークにクラスター化すること。

【0087】

いくつかの例示的な実施形態において、エンリッチされた家系を作成するためのシステムは、データプロセッサと、当該データプロセッサに連結したメモリとを備え、当該プロセッサは、コホートのシークエンシングデータに基づいて個体の1親等ネットワークを作成するように構成されており、このときコホートは、複数の対象を含む任意のデータセットを含み得る。

【0088】

いくつかの例示的な実施形態において、エンリッチされた家系を作成するためのシステ

10

20

30

40

50

ムは、データプロセッサと、当該データプロセッサに連結したメモリとを備え、当該プロセッサは、さらに、p 値に基づいて家系をエンリッチすることを含むように構成されている。エンリッチメントは、家系の「創始者アンカー分枝」または「分枝」を家系内の創始者の全ての子孫として定義し、二項検定を使用して分枝のバイナリー形質がエンリッチされているかを評価することを含み得る。バイナリー形質は、上記のような ICD を用いて定義することができる。また、エンリッチメントは、家系の「創始者アンカー分枝」または「分枝」を家系内の創始者の全ての子孫として定義し、t 検定を使用して分枝の極端な定量的形質がエンリッチされているかを評価することも含み得る。いくつかのパラメータを使用して、上記のような極端な定量的形質を定義することができる。さらに、エンリッチメントは、多重検定 p 値カットオフを適用することも含み得る。

10

【0089】

1つの例示的な態様において、本開示は、疾患原因バリエーションを同定するためのシステムであって、データプロセッサと、当該データプロセッサに連結したメモリとを備え、当該プロセッサが、コホートのシーケンシングデータに基づいて個体の1親等ネットワークを作成することと、当該1親等ネットワーク内の個体が罹患者または非罹患者であるかを同定することと、罹患者または非罹患者としての指定を含む当該個体を含む少なくとも1つのエンリッチされた家系を作成することとを行うように構成された、システムを提供する。

【0090】

いくつかの例示的な実施形態において、疾患原因バリエーションを同定するためのシステムは、データプロセッサと、当該データプロセッサに連結したメモリとを備え、当該プロセッサは、当該家系内の個体が罹患者または非罹患者であるかどうかを同定することであって、少なくとも1つのバイナリー形質を有する当該個体が罹患者として同定される、当該少なくとも1つのバイナリー形質を有しない当該個体が非罹患者として同定される、同定することと、次いで、罹患者及び非罹患者のパターンがメンデル遺伝様式（例えば、常染色体優性、常染色体劣性、x連鎖優性、x連鎖劣性、またはy連鎖）と一致するかどうかを評価することとを行うように構成されている。いくつかの特定の例示的な実施形態において、バイナリー形質は、世界保健機関（WHO）による医学分類リストである疾病及び関連保健問題の国際統計分類（ICD）を用いて定義することができる。この分類には、疾患、徴候及び症状、異常所見、病訴、社会的状況、ならびに傷害または疾患の外部原因に関するコードが含まれる。バイナリー形質の定義には、ICDの第9版または第10版を使用することができる。1つの例示的な実施形態において、特定のバイナリー形質に関して電子健康記録データが利用できない可能性のある個体、または特定のバイナリー形質に関して矛盾するもしくは信頼できないデータを有する個体は、医療記録に特定のバイナリー形質が存在するかしなにかにかかわらず、未知の罹患者として判定され得る。

20

30

【0091】

いくつかの例示的な実施形態において、疾患原因バリエーションを同定するためのシステムは、データプロセッサと、当該データプロセッサに連結したメモリとを備え、当該プロセッサは、当該家系内の個体が罹患者または非罹患者であるかどうかを同定することであって、少なくとも1つの極端な定量的形質を有する当該個体が罹患者として同定される、当該少なくとも1つの極端な定量的形質を有しない当該個体が非罹患者として同定される、同定することと、次いで、罹患者及び非罹患者のパターンがメンデル遺伝様式（例えば、常染色体優性、常染色体劣性、x連鎖優性、x連鎖劣性、またはy連鎖）と一致するかどうかを評価することとを行うように構成されている。いくつかのパラメータを使用して、極端な定量的形質の影響を受けているかどうかを定義することができる。例えば、最大年齢カットオフを使用してより早期の障害発症を定義すること、または、形質の正常な集団の測定値からの偏差の定義された統計的カットオフを超えた（例えば、集団平均を2標準偏差上回る）定量的形質の最小値もしくは最大値もしくは中央値を測定することができる。1つの例示的な実施形態において、特定の定量的形質に関して電子健康記録データが利用できない可能性のある個体、または特定の定量的形質に関して矛盾するもしくは

40

50

は信頼できないデータを有する個体は、医療記録に特定の定量的形質が存在するかしないかにかかわらず、未知の罹患者として判定され得る。

【0092】

いくつかの例示的な実施形態において、疾患原因バリエーションを同定するためのシステムは、データプロセッサと、当該データプロセッサに連結したメモリとを備え、当該プロセッサは、家系内の個体が罹患者または非罹患者であるかどうかを同定することであって、少なくとも1つのバイナリー形質、極端な定量的形質、またはこれらの組合せを有する個体が罹患者として同定され、少なくとも1つのバイナリー形質、極端な定量的形質、またはこれらの組合せを有しない個体が非罹患者として同定される、同定を行うように構成されている。バイナリー形質は、上記のような定義されたICDコードとすることができる。いくつかのパラメータを使用して、上記のような極端な定量的形質を定義することができる。1つの例示的な実施形態において、特定のバイナリー形質、定量的形質、もしくはこれらの組合せに関して電子健康記録データが利用できない個体、または特定のバイナリー形質、定量的形質、もしくはこれらの組合せに関して矛盾するもしくは信頼できないデータを有する個体は、医療記録に特定の定量的形質が存在するかしないかにかかわらず、未知の罹患者として判定され得る。

10

【0093】

いくつかの例示的な実施形態において、疾患原因バリエーションを同定するためのシステムは、データプロセッサと、当該データプロセッサに連結したメモリとを備え、当該プロセッサは、家系内の個体が罹患者または非罹患者であるかどうかを同定することであって、少なくとも1つのバイナリー形質、極端な定量的形質、またはこれらの組合せを有する個体が罹患者として同定され、少なくとも1つのバイナリー形質、極端な定量的形質、またはこれらの組合せを有しない個体が非罹患者として同定される、同定を行うように構成されており、このとき、当該少なくとも1つのバイナリー形質、極端な定量的形質、またはこれらの組合せは、2つ以上の類似のまたは補完的な形質を含み得る。

20

【0094】

いくつかの例示的な実施形態において、疾患原因バリエーションを同定するためのシステムは、データプロセッサと、当該データプロセッサに連結したメモリとを備え、当該プロセッサは、家系内の個体が罹患者または非罹患者であるかどうかを同定することであって、少なくとも1つのバイナリー形質、極端な定量的形質、またはこれらの組合せを有する個体が罹患者として同定され、少なくとも1つのバイナリー形質、極端な定量的形質、またはこれらの組合せを有しない個体が非罹患者として同定される、同定を行うように構成されており、このとき、当該少なくとも1つのバイナリー形質、極端な定量的形質、またはこれらの組合せは、2つ以上の極端なまたは興味深い形質の共通部分を採用することを含み得る。

30

【0095】

いくつかの例示的な実施形態において、疾患原因バリエーションを同定するためのシステムは、データプロセッサと、当該データプロセッサに連結したメモリとを備え、当該プロセッサは、コホート内の個体が、少なくとも1つのバイナリー形質、極端な定量的形質、またはこれらの組合せを有する場合、当該個体を罹患者として同定することと、罹患者として判定された個体を、外部解析からの関連性結果の罹患キャリアとして定義することとを行うように構成されている。

40

【0096】

いくつかの例示的な実施形態において、疾患原因バリエーションを同定するためのシステムは、データプロセッサと、当該データプロセッサに連結したメモリとを備え、当該プロセッサは、コホートのシーケンシングデータに基づいて個体の1親等ネットワークを作成するように構成されている。シーケンシングデータは、全ゲノムシーケンシングデータ、エクソームシーケンシングデータ、または遺伝子型データを含み得る。

【0097】

いくつかの例示的な実施形態において、疾患原因バリエーションを同定するためのシステム

50

は、データプロセッサと、当該データプロセッサに連結したメモリとを備え、当該プロセッサは、エクソームシーケンシングデータに基づいて個体の1親等ネットワークを作成するように構成されている。エクソームシーケンシングデータに基づく個体の1親等ネットワークは、以下を含めた集団の関連性を活用することによって作成することができる：複数のヒト対象から得られた核酸配列サンプルのデータセットから低品質の配列バリエーションを削除すること、サンプルのうちの1例以上の各々に対する祖先スーパークラス指定を確立すること、データセットから低品質のサンプルを削除すること、祖先スーパークラス内の対象の第1の家系同一性推定値を作成すること、対象の祖先スーパークラスとは無関係に第2の家系同一性推定値を作成すること、及び第2の家系同一性推定値のうちの1つ以上に基づき、対象を主要な1親等家族ネットワークにクラスター化すること。

10

【0098】

いくつかの例示的な実施形態において、疾患原因バリエーションを同定するためのシステムは、データプロセッサと、当該データプロセッサに連結したメモリとを備え、当該プロセッサは、コホートのシーケンシングデータに基づいて個体の1親等ネットワークを作成するように構成されており、このときコホートは、複数の対象を含む任意のデータセットを含み得る。

【0099】

いくつかの例示的な実施形態において、疾患原因バリエーションを同定するためのシステムは、データプロセッサと、当該データプロセッサに連結したメモリとを備え、当該プロセッサは、さらに、p値に基づいて家系をエンリッチすることを含むように構成されている。エンリッチメントは、家系の「創始者アンカー分枝」または「分枝」を家系内の創始者の全ての子孫として定義し、二項検定を使用して分枝のバイナリー形質がエンリッチされているかを評価することを含み得る。バイナリー形質は、上記のようなICDを用いて定義することができる。また、エンリッチメントは、家系の「創始者アンカー分枝」または「分枝」を家系内の創始者の全ての子孫として定義し、t検定を使用して分枝の極端な定量的形質がエンリッチされているかを評価することも含み得る。いくつかのパラメーターを使用して、上記のような極端な定量的形質を定義することができる。さらに、エンリッチメントは、多重検定p値カットオフを適用することも含み得る。

20

【0100】

いくつかの例示的な実施形態において、疾患原因バリエーションを同定するためのシステムは、データプロセッサと、当該データプロセッサに連結したメモリとを備え、当該プロセッサは、家系内で罹患者と共に同時分離するバリエーション形質の対を同定することと、表現型分離に基づいて少なくとも1つのエンリッチされた家系を見出すことを含む分離解析を実施することとを行うように構成されている。この分離には、優性及び相加的分離モデルならびに劣性分離モデルが含まれ得る。1つの例示的な実施形態において、優性及び相加的分離モデルに基づいて少なくとも1つのエンリッチされた家系を見出すことは、1つの可能な構造と、共通の祖先を有する少なくとも3例の罹患者とを有する家系を選択することを含む。これはさらに、偽陽性を減らすために、1例以上の近親の非罹患者を有する少なくとも1つのエンリッチされた家系を選択することを含んでもよい。別の例示的な実施形態において、劣性分離モデルに基づいて少なくとも1つのエンリッチされた家系を見出すことは、1つの可能な構造と、非罹患の親を有する1例を超える罹患者とを有する家系を選択することを含む。これはさらに、偽陽性を減らすために、少なくとも2例の罹患した同胞を有する少なくとも1つのエンリッチされた家系を選択することを含んでもよい。

30

40

【0101】

いくつかの例示的な実施形態において、疾患原因バリエーションを同定するためのシステムは、データプロセッサと、当該データプロセッサに連結したメモリとを備え、当該プロセッサは、分離の特定の遺伝モデルを形成するために分離解析を実施するように構成されている。分離の特定の遺伝モデルには、分離の優性遺伝モデルまたは分離の劣性遺

50

伝モデルが含まれ得る。加えて、分離の特定の遺伝モデルには、他の遺伝様式に基づく分離の遺伝モデル、例えば、Y連鎖、多因子、またはミトコンドリア連鎖遺伝様式も含まれ得る。1つの例示的な実施形態において、疾患原因バリエントを同定するための方法は、分離の優性遺伝モデルを形成するために分離解析を実施することであって、当該疾患原因バリエントが、少なくとも1つのバイナリー形質、極端な定量的形質、またはこれらの組合せにおいて罹患者と共に分離する、実施することを含む。1つの例示的な実施形態において、疾患原因バリエントを同定するための方法は、分離の劣性遺伝モデルを形成するために分離解析を実施することであって、当該疾患原因バリエントが、所与の遺伝子において両アレルバリエントキャリアである罹患者と共に分離する、実施することを含み、親の遺伝子データが入手可能な場合、これらは同定された疾患原因バリエントに対しヘテロ接合性でなければならない。

10

【0102】

いくつかの例示的な実施形態において、疾患原因バリエントを同定するためのシステムは、データプロセッサと、当該データプロセッサに連結したメモリとを備え、当該プロセッサは、少なくとも1つのエンリッチされた家系の内部でまたはそれにわたって同時分離するバリエント形質の対を同定するために分離解析を実施するように構成されている。1つの例示的な実施形態において、疾患原因バリエントを同定するための方法は、複数のエンリッチされた家系の内部でまたはそれにわたって同時分離するバリエント形質の対を同定するための分離解析を含む。

【0103】

いくつかの例示的な実施形態において、疾患原因バリエントを同定するためのシステムは、データプロセッサと、当該データプロセッサに連結したメモリとを備え、当該プロセッサは、家族構造に含まれない目的表現型について他の罹患者内の分離するバリエントまたは遺伝子を同定するために分離解析を実施するように構成されている。

20

【0104】

いくつかの例示的な実施形態において、疾患原因バリエントを同定するためのシステムは、データプロセッサと、当該データプロセッサに連結したメモリとを備え、当該プロセッサは、集団規模解析からの関連性結果と共にバリエント及び形質を相互参照することを含む分離解析を実施するように構成されている。

【0105】

いくつかの例示的な実施形態において、疾患原因バリエントを同定するためのシステムは、データプロセッサと、当該データプロセッサに連結したメモリとを備え、当該プロセッサは、以前から知られている原因バリエント及び原因遺伝子を同定するために分離解析を実施するように構成されている。

30

【0106】

いくつかの例示的な実施形態において、疾患原因バリエントを同定するためのシステムは、データプロセッサと、当該データプロセッサに連結したメモリとを備え、当該プロセッサは、支持する家系／罹患者の数により、ならびに候補となる原因バリエント及び原因遺伝子の数により、エンリッチされた家系を優先順位付けするように構成されている。

40

【0107】

いくつかの例示的な実施形態において、疾患原因バリエントを同定するためのシステムは、データプロセッサと、当該データプロセッサに連結したメモリとを備え、当該プロセッサは、バリエント形質の対を解析するように構成され、さらに、家族ベースの関連性解析を保証するのに十分な家族データで罹患者のセットを同定することを含む。

【0108】

いくつかの例示的な実施形態において、疾患原因バリエントを同定するためのシステムは、データプロセッサと、当該データプロセッサに連結したメモリとを備え、当該プロセッサは、バリエント形質の対を解析するように構成され、家系及び表現型の情報に基づいて、伝達不平衡試験（TDT）または必要に応じてその他の解析を実施すること

50

を含む。

【0109】

いくつかの例示的な実施形態において、疾患原因バリエーションを同定するためのシステムは、データプロセッサと、当該データプロセッサに連結したメモリとを備え、当該プロセッサは、いくつかの生理学的障害についての疾患原因バリエーションを同定するように構成されている。

【0110】

本明細書で説明される方法及びシステムは、(i) 疾患を引き起こす分子機構をより良好に理解できるようにし、(ii) より良好に疾患の分類及び管理ができるように導き、(iii) 重要な遺伝子バリエーションに関連する異なる代謝を同定できるようにし（がん細胞の代謝変化に関連する不可欠な酵素またはタンパク質または受容体を新薬開発の標的として使用）、(iv) がんなどの疾患向けに、将来の臨床経過及び生存期間の予測する助けとなり得る精緻化されたクラス予測を提供し、そして(v)（望ましいが欠損した遺伝子の増大、または（アンチセンスオリゴヌクレオチドもしくは転写因子デコイ、または特異的アプタマーを通じての）有害な遺伝子の遮断により）疾患を引き起こす遺伝子欠損を同定することによって遺伝子療法を設計することができる。

【図面の簡単な説明】

【0111】

【図1】家系エンリッチメントを実施するための本発明の例示的な実施形態のフローチャートである。

【図2】家系エンリッチメントを実施するための本発明の例示的な実施形態のフローチャートである。

【図3】例示的な動作環境である。

【図4】本開示の方法を実施するために構成された複数のシステムコンポーネントを示している。

【図5】例示的な実施形態に従って確認されたDiscovEHRコホートからの最初の92Kのシーケンシングされた個体におけるIBD0対IBD1プロットを示している。

【図6】原発性血栓形成傾向表現型（Phe10__D685、ICD10CM D68.5）におけるDiscovEHRコホートからのいくつかのエンリッチされた家系を示しており、家系エンリッチメントは、例示的な実施形態に従って実施されている。

【図7】A及びBは、遺伝性出血性末梢血管拡張症表現型（Phe10__I780、ICD10CM I78.0）における2つのエンリッチされた家系を示しており、家系エンリッチメントは、例示的な実施形態に従って実施されている。

【図8】遺伝性出血性末梢血管拡張症表現型（Phe10__I780、ICD10CM I78.0）におけるバリエーションの分離を実証するエンリッチされた家系を含むDiscovEHRコホートからの家系を示しており、家系エンリッチメント及び分離解析は、例示的な実施形態に従って実施されている。

【図9】気腫表現型におけるDiscovEHRコホートからのいくつかのエンリッチされた家系を示しており、家系エンリッチメントは、例示的な実施形態に従って実施されている。

【図10】腎臓移植表現型（Phe9__V420、ICD9CM V42.0）におけるDiscovEHRコホートからのエンリッチされた家系を示しており、家系エンリッチメントは、例示的な実施形態に従って実施されている。

【図11】末期腎疾患表現型（Phe9__5856、ICD9CM 585.6）におけるDiscovEHRコホートからのいくつかのエンリッチされた家系を示しており、家系エンリッチメントは、例示的な実施形態に従って実施されている。

【図12】遺伝性運動感覚性ニューロパチー表現型（シャルコー・マリー・トゥース病）（Phe10__G600、ICD10CM G60.0）表現型におけるDiscovEHRコホートからのエンリッチされた家系を示している。

10

20

30

40

50

【図13】様々な組織内でコードされるトロポミオシン2（TMP2）遺伝子の100万個当たりの転写産物（TPM）の遺伝子発現データを示す図表である。

【図14】双極性障害におけるDiscovEHRコホートからのエンリッチされた家系を示しており、家系エンリッチメント及び分離解析は、例示的な実施形態に従って実施されている。

【図15】様々な組織内でコードされる20番染色体オープンリーディングフレーム203（C20orf203）の100万個当たりの転写産物（TPM）の遺伝子発現データを示す図表である。

【図16】双極性障害表現型におけるDiscovEHRコホートからのエンリッチされた家系を示しており、家系エンリッチメントは、例示的な実施形態に従って実施されている。

【図17】双極性障害表現型におけるDiscovEHRコホートからのエンリッチされた家系を示しており、家系エンリッチメントは、例示的な実施形態に従って実施されている。

【図18】双極性障害表現型におけるDiscovEHRコホートからのエンリッチされた家系を示しており、家系エンリッチメントは、例示的な実施形態に従って実施されている。

【図19】様々な組織内のマイクロセファリン1（MCPH1）の100万個当たりの転写産物（TPM）の遺伝子発現データを示す図表である。

【図20】家族性サラセミア表現型におけるDiscovEHRコホートからのエンリッチされた家系を示しており、家系エンリッチメントは、例示的な実施形態に従って実施されている。

【図21】アルカリホスファターゼ外来患者中心傾向値におけるDiscovEHRコホートからのエンリッチされた家系を示しており、家系エンリッチメントは、例示的な実施形態に従って実施されている。

【発明を実施するための形態】

【0112】

「a」という用語は、「少なくとも1つ」を意味するものと理解されるべきであり、「約（about）」及び「およそ（approximately）」という用語は、当業者が理解するであろう標準的なばらつきを許容するものと理解されるべきであり、範囲が示される場合、終点が含まれる。

【0113】

家族ベースの関連性研究では症例対照デザインが使用され、症例は病院または疾病登録から得られる。対照は、非血縁者（例えば、集団もしくは病院／登録ベース）または症例の家族（例えば、親もしくは同胞）のいずれかとする事ができる。症例及び対照における所与のアレルの出現率を比較して、遺伝子と疾患との間に「関連性」が存在するか調べる。大規模な一塩基多型（SNP）ジェノタイピングが利用できることに伴い、関連性研究はますます一般的になりつつあり、焦点を絞った候補遺伝子研究からゲノムワイドの関連性研究へと急速に拡大している。

【0114】

次世代シーケンシング戦略の出現により、このような疾患における遺伝子欠損を解明する見通しが明るくなった。現在、全ゲノム（およそ30億塩基対）のシーケンシングは数日という期間で行うことができ、コストは急速に低下していることから、通例的な研究ツールとしてアクセス可能になっている。ゲノムのタンパク質コード部分のシーケンシングはエクソームシーケンシングと称され、疾患原因遺伝子を見出すのにより一層効率的である。というのは、エクソームはゲノムのわずかな部分（およそ38Mb）に相当するに過ぎず、またエクソンはメンデル遺伝子における既知の変異の大部分を有するためである（Albert et al. Nature Methods (2007) 4:903-905; Gnirke et al. Nature Biotechnology (2009) 27:182-189; Hodges et al. Nature

10

20

30

40

50

Genetics (2007) 9:1522–1527; Majewski et al. Journal of Medical Genetics (2011) 48:580–589)。そのため、エクソームシーケンシングは、候補遺伝子または経路の先験的知識を伴わずに遺伝子的原因が疑われる障害の変異を探索するのに非常に適している。

【0115】

大規模なヒトシーケンシング研究の多くは、付随する表現型が豊富な電子健康記録（EHR）を有する一貫した医療集団からサンプルを収集し、EHR及びゲノム配列データを組み合わせてトランスレーショナルな発見及びプレジジョンメディシンを触媒することを目指している。このようなプロジェクトから得られたデータは、形質及び疾患のある特定の遺伝子的駆動要因を同定するために使用することができる。

10

【0116】

異なるソース集団から集団の層別化を引き起こす様々なアレル頻度を有する症例及び対照が得られた場合、偽の関連性が検出される可能性がある（Cardon and Palmer. Lancet (2003) 361 (9357): 598–604)。このような交絡からどの程度のバイアスが生じるかに関しては議論がある（Wacholder et al. Cancer Epidemiology, Biomarkers & Prevention (2002) 11 (6): 513–520; Thomas and Witte. Cancer Epidemiology, Biomarkers & Prevention (2002) 11 (6): 502–512; Gorroochurn et al. Human Heredity (2004) 58 (1): 40–48)。集団の層別化は、家族ベースの研究設計を使用することによって回避され得る。親及びその子孫または同胞を研究する場合、各家族内の症例及び対照は同じソース集団から生じる。一般的な家族ベースの症例対照デザインは、親トリオ（例えば、伝達不平衡試験（TDT）アプローチ）及び同胞対照である。他の近親者（例えば、いとこ）を研究しても、多数の異なる家族を同時に研究してもよい。

20

【0117】

大規模コホート内での家族の同定は、所与の形質が家族ベースの遺伝子研究を受け入れられるだけの十分な情報価値のある罹患者からなる家系を同定することを伴う。家系は、家族内の所与の目的表現型と共に同時分離する潜在的な中～大の影響を有する希少バリエーションを調べる際には特に有益である。これらの家系は、目的表現型を有する近親の参加者のサブセットを定義し、次いでこれらのサブセットを調べて形質及び疾患の遺伝子的駆動要因を同定することに活用することができる。

30

【0118】

本開示は、少なくとも部分的には、複数の対象のゲノムサンプルのデータセット内の個体の1親等ネットワークに関する情報が、とりわけ、希少な遺伝子バリエーションと疾患との間の関連性を調べるのを可能にするという認識に基づく。

【0119】

本明細書で説明される方法は、様々なタイプのゲノムサンプルのデータセットに適用され得る。データセットのタイプの非限定的な例としては、シングル医療ネットワーク集団；マルチ医療ネットワーク集団；人種的、文化的、もしくは社会的に同種もしくは異種の集団；年齢が混合した集団もしくは年齢が同種の集団；地理的に集中もしくは分散した集団；またはこれらの組み合わせが挙げられる。データセットは、様々なタイプの遺伝子変異を有し得る。評価され得る遺伝子バリエーションのタイプの非限定的な例としては、点変異、挿入、欠失、逆位、重複、及び多量体化が挙げられる。遺伝子バリエーションを取得することができる手段の非限定的な例としては、以下のステップが挙げられる。

40

【0120】

・サンプル調製及びシーケンシング（Dewey et al. (2016), Science 354, aaf6814–1からaaf6814–10）。

・シーケンシングが完了したら、各シーケンシング実行からの生データは、自動化解析のために、ローカルバッファーストレージに集められ、DNAexusプラットフォーム

50

ーム (Reid et al. (2014); BMC Bioinformatics 15, 30) にアップロードすることができる。

【0121】

・サンプルレベルのリードファイルは、CASA VA (Illumina Inc., San Diego, CA) で作成し、BWA-memによりGRCh38にアラインメントすることができる (Li and Durbin (2009); Bioinformatics 25, 1754-176; Li (2013); arXiv q-bio. GN)。

【0122】

・得られたBAMファイルは、GATK (McKenna et al. (2010); Genome Res. 20, 1297-1303) 及びPicardを用いて処理して、推定インデルの周囲のリード値をソートし、複製物をマークし、局所的な再アラインメントを実施することができる。

【0123】

・シーケンシングしたバリエーションは、Ensembl 85遺伝子定義を用いてsnpEFF (Cingolani et al. (2012); Fly (Austin) 6, 80-92) でアノテーションして転写産物及び遺伝子に対する機能的影響を決定することができる。

【0124】

本明細書で説明される方法は、生理学的障害の原因となる疾患原因バリエーションを同定するために適用することができる。非限定的な例としては、心理的障害、血液関連障害、疼痛関連障害、ホルモン関連障害、肺疾患、歯科障害、生殖関連障害、精神障害、運動障害、心臓血管障害、循環障害、自己免疫疾患、炎症性疾患、腎疾患、肝臓障害、遺伝性出血性末梢血管拡張症、運動感覚性ニューロパチー、家族性大動脈瘤、甲状腺癌、色素性緑内障、家族性高コレステロール血症、またはこれらの組合せが挙げられる。

【0125】

方法は、前述のステップのいずれかに限定されず、配列バリエーションの取得は、任意の好適な手段によって行われ得ることを理解されたい。

また、本開示は、少なくとも部分的には、複数の対象のゲノムサンプルのデータセット内で1親等近親者に関する情報から作成された家系が、家族内で分離する希少バリエーションを同定するための情報を提供することができるという認識にも基づいている。

【0126】

1親等近親者を同定するために使用され得るいくつかの統計的方法が開発されている。このような非限定的な例の1つは、個人がデータセット内の異なるタイプの家族関係を同定する場合の家系同一性 (IBD) 推定値の計算によるものであり、PRIMUS (Staples et al. (2014), Am. J. Hum. Genet. 95, 553-564) を使用してペアワイズ関係を異なる家族クラスに分類し、家系を再構築することができる。データセット間での推定1親等関係のみを含むものとする。例えば、エクソーム配列決定データを含むデータセットから1親等近親者を同定するには、2018年9月7日に出願された同時係属中の米国特許出願公開第20190205502号明細書 (表題 "SYSTEMS AND METHODS FOR LEVERAGING RELATEDNESS IN GENOMIC DATA ANALYSIS") で説明されている方法を利用することができる (当該文献はその全体が参照により本明細書に援用される)。

【0127】

複数の対象のゲノムサンプルのデータセットから家系を作成するには、いくつかのアプローチが利用可能であり、例えば、COP (非近交系家系の構築 (Constructing Outbred Pedigrees)) 及びCIP (近交系家系の構築 (Constructing Inbred Pedigrees))、IPED (遺伝パスベース家系の再構築 (Inheritance Path-based Pedigree

10

20

30

40

50

Reconstruction))及びIPED2、PREPARE(親族の分割(Partitioning of Relatives))、ならびに家系再構築及び最大限に非血縁のセットの同定(Pedigree Reconstruction and Identification of the Maximally Unrelated Set)(PRIMUS)が利用可能である(Riester et al. Bioinformatics(2009)25:2134-2139;Hadfield et al. Molecular Ecology(2006)15:3715-3730;Marshall et al. Molecular Ecology(1998)7:639-655;Cussens et al. Genetic Epidemiology(2013)37:69-83;He et al. Journal of Computational Biology(2013)20:780-792;Kirkpatrick et al. Journal of Computational Biology(2011)18:1481-1493;Staples et al. Genetic Epidemiology(2013)37:136-141;Shem-Tov and Halperin. PLoS Computational Biology(2014)10:e1003610)。他の方法、例えば、PLINK、KING、及びKINSHIPも使用され得る。

【0128】

本開示は、前述のデータセット、1親等の近親者を特定する方法、及び／または家系を作成する方法のいずれかに限定されず、複数の対象のゲノムサンプルのデータセットの取得及び処理は、当技術分野で知られている任意の好適な方法によって実施できることを理解されたい。

【0129】

また、本開示は、少なくとも部分的には、データセット内の罹患者及び非罹患者を決定することにより家系を形成し、家系を精緻化してエンリッチされた家系を形成することは、とりわけ、希少な遺伝子バリエーションと疾患との間の関連性を見出すための下流解析にとって不可欠であるという認識にも基づいている。

【0130】

データセット内の罹患者は、少なくとも1つのバイナリー形質もしくは極端な定量的形質またはこれらの組合せの存在に基づいてデータセット内の個体を同定することにより、定義することができる。

【0131】

いくつかの例示的な実施形態において、バイナリー形質は、疾病及び関連保健問題の国際統計分類リスト(ICD)からの3文字コードを用いて定義される。いくつかの特定の例示的な実施形態において、ICDの第9または第10改訂版からの3文字コードを使用して、バイナリー形質を定義した。バイナリー形質は、さらに、ICDの第9または第10改訂版からの4文字コードを用いてさらに定義され得る。個体の表現型が、説明されたバイナリー形質を有する場合、個体を「罹患者」と判定することができる。いくつかの例示的な実施形態において、コホート内で5%を超える有病率を有するバイナリー形質を有する個体は、以前に「罹患者」と判定された場合であっても「非罹患者」と判定されることがある。さらに、個体が医療記録内に形質の非存在または存在の徴候を有する場合、及び個体が矛盾する記録を有する場合、その個体は未知の罹患者と判定される。

【0132】

いくつかの例示的な実施形態において、極端な定量的形質は、集団内での当該形質の分布に基づいて、当該形質の値が極端に高いまたは低い個体を採用することによって定義される。例えば、各形質の値に対しZスコアを計算し、当該形質のZスコアが、極端に高いまたは低い形質値についてそれぞれ2より上または-2より下である場合、個体を「罹患者」としてラベル付けする。さらに、個体が医療記録内に形質の非存在または存在の徴候を有する場合、及び個体が矛盾する記録を有する場合、その個体は未知の罹患者と判定される。

10

20

30

40

50

【0133】

罹患者を含む家系をさらに精緻化してエンリッチされた家系を作成することができる。
家系は、表現型分離またはp値に基づいてエンリッチすることができる。

図1は、1親等ネットワークからの個体が罹患者及び非罹患者と判定される例示的な実施形態のフローチャートである。個体の1親等ネットワークは、ステップ100において、任意の好適な手段によって複数のヒト対象から作成される。ネットワーク内のあらゆる個体は、110において、記録された各バイナリー形質もしくは記録された各定量的形質またはこれらの組合せについて評価され得る。ネットワーク内のあらゆる個体は、ステップ120において、記録された各バイナリー形質について評価され得、バイナリー形質の影響を受けている場合、ステップ140において「罹患者」として分類される。反対に、個体が、検討されている特定のバイナリー形質の影響を受けていない場合、個体は、ステップ150において「非罹患者」として分類される。ネットワーク内のあらゆる個体は、ステップ130において、記録された各定量的形質について評価され得、定量的形質の影響を受けている場合、ステップ140において「罹患者」として分類される。反対に、個体が、検討されている特定の定量的形質の影響を受けていない場合、個体は、ステップ150において「非罹患者」として分類される。

10

【0134】

図2は、1親等ネットワークからの個体が罹患者及び非罹患者と判定される別の例示的な実施形態のフローチャートである。ステップ100において、任意の適切な手段によって複数の対象から個体の1親等ネットワークを作成した後、ネットワーク内のあらゆる個体は、110において、記録された各バイナリー形質もしくは記録された各定量的形質またはこれらの組合せについて評価され得る。さらに、記録された各バイナリー形質もしくは記録された各定量的形質またはこれらの組合せのいずれかを有するあらゆる個体は、ステップ155において、バイナリー形質または定量的形質の存在に基づいて評価される。ステップ155の後のステップ160では、個体を分類することができる。個体を罹患者として分類するために使用されるバイナリー形質がコホート内で5%を超える有病率を有する場合、罹患者は、ステップ190において「非罹患者」として分類され得、個体を罹患者として分類するために使用されるバイナリー形質が5%未満の罹患率を有する場合、罹患者は、ステップ180において「罹患者」として分類され得る。同様に、ステップ170は、個体を再分類することができる。個体を罹患者として分類するために使用される定量的形質が、コホートの平均の定量的形質より2標準偏差超大きい場合、個体は、ステップ180において「罹患者」として分類され、そうでない場合、個体は、ステップ190において「非罹患者」として分類される。

20

30

【0135】

家系内または家系間での表現型分離は、優性及び相加的分離モデルまたは劣性分離モデルのいずれかを作成することができる。優性及び相加的分離モデルへの表現型分離を伴う家系についてのいくつかの例示的な実施形態において、1つの可能な構造及び共通の祖先を有する3例を超える罹患者を有する家系を使用して、エンリッチされた家系を作成することができる。さらに、エンリッチされた家系は、偽陽性を減らすために1例または複数例の近親の非罹患者（複数可）を有する家系を選択することにより、分離解析について優先順位付けすることができる。

40

【0136】

劣性分離モデルへの表現型分離を伴う家系についてのいくつかの例示的な実施形態において、エンリッチされた家系を作成するために、1つの可能な構造と非罹患の親を有する1例を超える罹患者とを有する家系が使用される。さらに、エンリッチされた家系は、2例以上の罹患した同胞を有する家系を選択することにより、エンリッチされた家系を分離解析について優先順位付けすることができる。

【0137】

いくつかの例示的な実施形態において、2つ以上の表現型が類似のまたは相補的なバイナリー形質または極端な定量的形質からの罹患者を併合して、これらの形質全てを包含す

50

る障害についての罹患者を形成することができる。例えば、双極性障害についてエンリッチされた家系を探す場合、双極性障害の遺伝子的原因は一部の個体では単極性としてのみ顕在化する場合があるため、単極性障害も考慮することができる。

【0138】

いくつかの例示的な実施形態において、2つ以上の極端なまたは興味深いバイナリー形質または極端な定量的形質を有する罹患者を選択して、これらの2つ以上の形質全てを包含する障害についての罹患者を形成することができる。2つ以上の極端なまたは興味深い形質を有する罹患者の共通部分を採用することにより、より均質な個体のサブセットを同定することができる。例えば、喘息及びCOPDの両方を有する個体のエンリッチされた家系を得るために、喘息及びCOPDの両方を有する患者の共通部分を罹患者とみなす。

10

【0139】

本開示は、前述の障害または分離モデルのいずれにも限定されず、家系エンリッチメントは、少なくとも1つのバイナリー形質、極端な定量的形質、またはこれらの組合せに基づいて任意の障害または分離モデルに対して行われ得ることを理解されたい。

【0140】

代替的に、エンリッチされた家系は、p値に基づいて決定することができる。いくつかの例示的な実施形態において、家系の創始者アンカー分枝を同定する際に、二項検定を行って、家系がバイナリー形質についてエンリッチされているかを評価する。他の例示的な実施形態において、家系の創始者アンカー分枝を同定する際に、t検定を行って、家系が極端な定量的形質についてエンリッチされているかどうかを評価する。加えて、偽陽性を取り除くために、多重検定補正p値カットオフが設定される。

20

【0141】

本開示は、少なくとも部分的には、所与の表現型、付随する（例えば、希少な）バリエーションを有する罹患者個体についてエンリッチされた家系が、目的表現型と共に分離しそれを駆動し得るという認識に基づいている。このような遺伝子的原因は家族単位内で共有される可能性がより高いため、目的表現型を有する罹患者についてエンリッチされた家系を同定することは、これらの表現型を駆動する偶然の（例えば、希少な）変異を同定する一助となり得る。

【0142】

エンリッチされた家系が同定されたら、根底にある遺伝子的原因は、分離解析及び家族ベースの関連性解析を行うことによって決定することができる。家系によっては、罹患者と共に分離する既知の疾患原因変異が存在するであろう。残りの家系は、複数の家系にわたって罹患者を分離しているバリエーション及び遺伝子、または家系に含まれていないデータセット内の罹患者によって優先順位付けすることができる。いずれにしても、これらの分離解析からの結果は、候補バリエーションのリストを含むことができる。

30

【0143】

分離解析は、様々な程度の一般性のモデルを試験することによって実施され得る。様々な制約（例えば、優性遺伝または劣性遺伝）を有するモデルは、最も一般的なモデルと比較することができ、モデル内の全てのパラメーターを推定して、どのモデル（複数可）がデータに最も適合するかを調べる。大きな家系を有する家族及び多くの罹患者は、遺伝子が重要であること証明することと、特定の遺伝子を同定することとの両方において、特に情報価値がある。

40

【0144】

所与の表現型の遺伝子的原因の同定を助けるために家系構造を使用する方法は、典型的には、関連性マッピング、連鎖解析、またはこの両方に対する革新的なバリエーションを伴う。このような方法としては、MORGAN、pVAASST、FBAT (www.hsph.harvard.edu/fbat/fbat.htm)、QTDT (csg.sph.umich.edu/abecasis/qt dt/)、ROADTRIPS、rareIBD、及びRV-GDTが挙げられる。使用する適切な方法は、表現型、遺伝様式、祖先の背景、家系の構造／サイズ、家系の数、及び非血縁のデータセットのサイズに

50

依存する。遺伝子－表現型の関連性を直接調べるために血縁関係及び家系を使用することに加えて、これらは、追加のまたは改善されたデータを作成するために他の多くの方法でも使用され得、すなわち、家系を意識した補完、家系を意識した段階分け、メンデル則のエラーチェック、複雑なヘテロ接合性ノックアウト検出及び *de novo* 変異コーリング、ならびにバリエーションコーリングの検証でも使用され得る。

【0145】

本発明によって説明または例示される方法はいずれも、コンピューターに実装された方法として、及び／またはシステムとして、実施することができる。当業者に知られている任意の好適なコンピューターシステムが、この目的で使用され得る。

【0146】

図3は、本発明の方法及びシステムが動作し得る例示的な環境200の様々な態様を示している。本発明の方法は、デジタル及びアナログ両方の機器を用いる様々なタイプのネットワーク及びシステムで使用され得る。本明細書では機能的な説明が提供され、それぞれの機能は、ソフトウェア、ハードウェア、またはソフトウェアとハードウェアとの組合せによって実行され得る。

【0147】

環境200は、ローカルデータ／処理センター210を含むことができる。ローカルデータ／処理センター210は、1つ以上のコンピューティングデバイス間の通信を容易にするために、ローカルエリアネットワークなどの1つ以上のネットワークを含むことができる。1つ以上のコンピューティングデバイスを使用して、生物学的データを格納、処理、解析、出力、及び／または視覚化することができる。環境200は、任意選択で、医療データプロバイダー220を含むことができる。医療データプロバイダー220は、生物学的データの1つ以上のソースを含むことができる。例えば、医療データプロバイダー220は、1例以上の患者の医療情報にアクセスする1つ以上の医療システムを含むことができる。医療情報は、例えば、病歴、医療専門家の観察及び所見、臨床検査報告、診断、医師の指示、処方、バイタルサイン、体液バランス、呼吸機能、血液パラメーター、心電図、X線、CTスキャン、MRIデータ、臨床検査結果、診断、予後、評価、入院及び退院記録、ならびに患者登録情報を含むことができる。医療データプロバイダー220は、1つ以上のコンピューティングデバイス間の通信を容易にするために、ローカルエリアネットワークなどの1つ以上のネットワークを含むことができる。1つ以上のコンピューティングデバイスを使用して、医療情報を格納、処理、解析、出力、及び／または視覚化することができる。医療データプロバイダー220は、医療情報を脱同定し、脱同定された医療情報をローカルデータ／処理センター210に提供することができる。脱同定された医療情報は、各患者に対する固有の識別子を含むことができ、これにより、医療情報を脱同定された状態で維持しながら、ある患者の医療情報を別の患者から区別することができる。脱同定された医療情報は、患者の識別情報が患者の特定の医療情報と結びつけられることを防止する。ローカルデータ／処理センター210は、脱同定された医療情報を解析して、（例えば、国際疾病分類「ICD」及び／または最新専門用語集（*Current Procedural Terminology*）「CPT」コードを割り当てることにより）各患者に1つ以上の表現型を割り当てること

【0148】

環境200は、NGSシーケンシング設備230を含むことができる。NGSシーケンシング設備230は、1つ以上のシーケンサー（例えば、*Illumina HiSeq 2500*、*Pacific Biosciences PacBio RS I* など）を含むことができる。1つ以上のシーケンサーは、エクソームシーケンシング、全エクソームシーケンシング、RNA-seq、全ゲノムシーケンシング、標的化シーケンシングなどのために構成することができる。例示的な態様において、医療データプロバイダー220は、脱同定された医療情報に関連する患者から生物学的サンプルを提供することができる。ユニークな識別子は、生物学的サンプルと、生物学的サンプルに対応する脱同定された医療情報との間の関連性を維持するために使用され得る。NGS

10

20

30

40

50

シーケンシング設備 230 は、生物学的サンプルに基づいて各患者のエクソームをシーケンシングすることができる。シーケンシング前に生物学的サンプルを保存するために、NGS シーケンシング設備 230 は、バイオバンク（例えば、Liconic Instruments 製）を含むことができる。生物学的サンプルは、チューブ（各チューブは患者に関連付けられる）で受け取ることができ、各チューブはバーコード（または他の識別子）を含むことができ、このバーコードはスキャンされて、サンプルを自動的にローカルデータ／処理センター 210 に記録することができる。NGS シーケンシング設備 230 は、均一なデータ及び有効なノンストップ動作を保証するために、1 つ以上のシーケンシング段階で使用するための 1 つ以上のロボットを含むことができる。NGS シーケンシング設備 230 は、このようにして年当たり数万個のエクソームをシーケンシングすることができる。1 つの態様において、NGS シーケンシング設備 230 は、月当たり少なくとも 1000、2000、3000、4000、5000、6000、7000、8000、9000、10,000、11,000、または 12,000 個の全エクソームをシーケンシングする機能的能力を有する。

10

【0149】

NGS シーケンシング設備 230 によって作成された生物学的データ（例えば、生のシーケンシングデータ）は、ローカルデータ／処理センター 210 に転送することができる。次いで、リモートデータ／処理センター 240 に生物学的データを転送することができる。リモートデータ／処理センター 240 は、1 つ以上のコンピューティングデバイスを含むクラウドベースのデータストレージ及び処理センターを含むことができる。ローカルデータ／処理センター 210 及び NGS シーケンシング設備 230 は、1 つ以上の大容量ファイバー回線を介してリモートデータ／処理センター 240 との間でデータを直接通信することができるが、その他のデータ通信システム（例えば、インターネット）も企図されている。例示的な態様において、リモートデータ／処理センター 240 は、サードパーティーシステム、例えば、Amazon Web Services（DNAexus）を含むことができる。リモートデータ／処理センター 240 は、解析ステップの自動化を容易にし、1 つ以上の共同作業 250 と安全な方式でデータを共有することを可能にする。ローカルデータ／処理センター 210 から生物学的データを受信すると、リモートデータ／処理センター 240 は、バイオインフォマティクスツールを用いた一次及び二次データ解析のために自動化された一連のパイプラインステップを実行することができ、その結果、各サンプルに対するアノテーション付きバリエーションファイルが得られる。このようなデータ解析（例えば、遺伝子型）からの結果は、ローカルデータ／処理センター 210 に返され、例えば、実験室情報管理システム（LIMS）に統合することができ、各生物学的サンプルの状態を維持するように構成することができる。

20

30

【0150】

次いで、ローカルデータ／処理センター 210 は、NGS シーケンシング設備 230 及びリモートデータ／処理センター 240 を介して得られた生物学的データ（例えば、遺伝子型）を、脱同定された医学的情報（同定された表現型を含む）と組み合わせて利用して、遺伝子型と表現型との間の関連性を同定することができる。例えば、ローカルデータ／処理センター 210 は、ある表現型がある特定の疾患領域（例えば、心血管疾患における極端な血中脂質）で治療可能性を有し得ると定義される場合に、表現型ファーストアプローチを適用することができる。別の例は、典型的な併存症の範囲から保護されていると思われる個人を同定するための肥満患者の研究である。別のアプローチは、遺伝子型及び仮説、例えば、遺伝子 X が疾患 Y の原因、または疾患 Y からの保護に関与しているという仮説から始めることである。

40

【0151】

例示的な態様において、1 以上の共同作業 250 は、インターネット 260 などのネットワークを介して、生物学的データ及び／または脱同定された医療情報の一部または全てにアクセスすることができる。

【0152】

50

図4に示されている例示的な態様において、ローカルデータ／処理センター210及び／またはリモートデータ／処理センター240のうちの1つ以上は、遺伝子データ構成要素300、表現型データ構成要素310、遺伝子バリエーション表現型関連性データ構成要素320、及び／またはデータ解析構成要素330のうちの1つ以上を含む1つ以上のコンピューティングデバイスを含むことができる。遺伝子データ構成要素300、表現型データ構成要素310、及び／または遺伝子バリエーション表現型関連性データ構成要素320は、配列データの品質評価、参照ゲノムへのリードアラインメント、バリエーションの同定、バリエーションのアノテーション、表現型の同定、バリエーション表現型関連性の同定、データ可視化、これらの組合せなどのうちの1つ以上のために構成することができる。

【0153】

例示的な態様において、構成要素のうちの1つ以上は、完全なハードウェア実施形態、完全なソフトウェア実施形態、またはソフトウェア及びハードウェアの態様を組み合わせた実施形態の形態をとることができる。さらに、方法及びシステムは、ストレージ媒体に具現化されたコンピューター可読プログラム命令（例えば、非一時的コンピューターソフトウェア）を有するコンピューター可読ストレージ媒体上のコンピュータープログラム製品の形態をとることができる。より詳細には、本発明の方法及びシステムは、ウェブで実装されたコンピューターソフトウェアの形態をとることができる。ハードディスク、CD-ROM、光ストレージ装置、または磁気ストレージ装置を含めた任意の好適なコンピューター可読ストレージ媒体を利用することができる。

【0154】

例示的な態様において、遺伝子データコンポーネント300は、1つ以上の遺伝子バリエーションに機能的にアノテーションするように構成することができる。遺伝子データ構成要素300は、1つ以上の遺伝子バリエーションの保存、解析、受信などを行うように構成することもできる。1つ以上の遺伝子バリエーションは、1例以上の患者（対象）から得られた配列データ（例えば、生の配列データ）からアノテーションされ得る。例えば、1つ以上の遺伝子バリエーションは、少なくとも100,000例、200,000例、300,000例、400,000例、または500,000例の対象の各々からアノテーションされ得る。1つ以上の遺伝子バリエーションに機能的にアノテーションした結果、遺伝子バリエーションデータが作成される。例として、遺伝子バリエーションデータは、1つ以上のバリエーションコールフォーマット（VCF）ファイルを含むことができる。VCFファイルは、SNP、インデル、及び／または構造バリエーションコールを表すテキストファイル形式である。バリエーションは、転写産物／遺伝子に対する機能的影響について評価され、潜在的な機能喪失（pLoF）候補が同定される。バリエーションは、Ensembl75遺伝子定義を用いてsnpeffでアノテーションされ、次いで機能的アノテーションが各バリエーション（及び遺伝子）についてさらに処理される。

【0155】

本明細書で示されるような数字及び／または文字を用いた方法ステップの連続的なラベル付けは、方法またはその任意の実施形態を特定の示された順序に限定するように意図されていない。

【0156】

特許、特許出願、公開特許出願、アクセッション番号、技術論文、及び学術論文を含めた様々な刊行物が本明細書全体で引用されている。これらの引用された参考文献の各々は、その全体があらゆる目的において参照により本明細書に援用される。

【0157】

以下の実施例を参照することにより、本開示はより十分に理解されるであろう。実施例は、本開示をより詳細に説明するために示されるものである。実施例は、例示を意図するものであり、本開示の範囲を限定するものと解釈されるべきではない。

【実施例】

【0158】

実施例1

個体及びサンプル

MyCode Community Health Initiativeに参加することに同意した93,368例の脱同定されたGeisinger Health System (GHS) 参加者のシーケンシングを行った。このイニシアチブの一環で、個体は、Geisinger Institutional Review Boardによって承認されたプロトコルの下、Regeneron GHS DiscovEHR共同研究の一部としてのゲノム解析及びGHS EHRのデータへのリンクを含む幅広い将来の研究のために、血液及びDNAサンプルを提供することに同意した。実施した全ての解析は、参加者の同意及びIRBの承認に従って行われた。各参加者のエクソームは、対応する脱同定されたEHRに関連付けられている。DiscovEHRの研究は、研究参加者として特に家族を対象としたものではなかったが、慢性的な健康問題のために医療システムと頻繁に相互作用する（かつ互いに近親者の可能性がある）成人、そしてGHSからのCoronary Catheterization Laboratory及びBariatric Serviceの参加者について暗黙のうちにエンリッチされた。

10

【0159】

実施例2

サンプル調製、シーケンシング、バリエントコーリング、及びサンプルQC

最初の61Kサンプル（「VCRomeセット」）に対するサンプル調製及びシーケンシングは、以前に説明されている（Dewey et al. Science (2016) 354: aaf6814）。残りの31Kサンプルのセットを同じプロセスで調製し、ただし、NimbleGenプローブ捕捉の代わりに、IDTのxGenプローブのわずかに改変されたバージョンを使用し、NimbleGen VCRome捕捉試薬で十分に網羅されているが標準的なxGenプローブではほとんど網羅されていないゲノムの領域を捕捉するための補足的プローブを加えたことを例外として行った。捕捉したフラグメントをストレプトアビジン結合ビーズに結合させ、非特異的DNAフラグメントを、製造業者（IDT）の推奨プロトコルに従って一連のストリンジェント洗浄で除去した。第2のサンプルセットは「xGenセット」と称された。Genome Analysis Toolkit (GATK; Web Resources) を用いてバリエントコールを生成した。GATKは、推定インデルの周囲の各サンプルのアラインメントし重複マークしたリードを局所的に再アラインメントするために使用した。INDEL再アラインメントし重複マークしたリードは、GATKのHaplotype Callerを用いて処理して、サンプルがゲノムバリエントコールフォーマット（gVCF）におけるゲノム参照から変化した全てのエクソン位置を同定した。ジェノタイピングは、各サンプル上のGATKのGenotype GYCF、ならびに参照と比較して一塩基バリエント（SNV）及びインデルの両方を同定する単一のサンプルバリエントコールフォーマット（VCF）ファイルを出力する50種のランダムに選択されたサンプルのトレーニングセットを用いて遂行した。単一サンプルVCFファイルを使用して、両方のセットの単一サンプルVCFファイルからの全ての可変部位を含む擬似サンプルを作出した。独立したpVCFファイルは、200個の単一サンプルgVCFファイルを擬似サンプルと共にジョイントコールして、2つの捕捉セットにわたって全ての可変部位で各サンプルに対しコールまたはコールなしを強制することにより、VCRomeセットに対し作出した。200サンプル全てのpVCFファイルを組み合わせるVCRome pVCFファイルを作出し、次いでこのプロセスを繰り返してxGen pVCFファイルを作成した。VCRome及びxGen pVCFファイルを組み合わせる結合pVCFを作成した。GRCh38に対する配列リードは、Ensembl 85遺伝子定義を使用することによってアラインメント及びアノテーションされたバリエントであった。遺伝子の定義は54,214の転写産物に限定されており、これはアノテーション付きの開始及び停止でタンパク質をコードする遺伝子19,467個に対応する。先に説明したサンプルQCプロセスの後、92,455個のエクソームが解析用に残った。

20

30

40

【0160】

50

実施例 3

主要構成要素及び祖先の推定

PLINK v1.910を使用して結合データセットをHapMap 318と併合し、参照SNPダスターIDに基づいて両データセットに存在するSNPを保持した。解析は、以下のPLINKフィルター「`maf 0.1 geno 0.05 snps-only-hwe 0.00001`」を適用することにより、マイナーアレル頻度 $>10\%$ 、遺伝子型消失 $<5\%$ 、及びハーディ・ワインベルグ平衡 p 値 >0.00001 を有する高品質な共通のSNPに限定された。HapMap 3サンプルにおける主要構成要素(PC)を計算し、次いでPLINKを使用することによりこれらのPCにデータセット内の各サンプルを投影した。HapMap 3サンプルにおけるPCを使用して、5つの祖先スーパークラス：アフリカ人(AFR)、混合アメリカ人(AMR)、東アジア人(EAS)、ヨーロッパ人(EUR)、南アジア人(SAS)の各々に対しカーネル密度推定量(KDE)を訓練した。各サンプルがスーパークラスの各々に属する尤度を推定するためにKDEを計算した。各サンプルに対し、尤度に基づいた祖先スーパークラスを割り当てた。サンプルが尤度 >0.3 の2つの先祖グループを有する場合、サンプルを、EURよりもAFR、EURよりもAMR、EASよりもAMR、EURよりもSAS、及びAFRよりもAMRに優先して割り当て、それ以外を「不明」とした。ゼロまたは2つを超える祖先グループが十分に高い尤度を有する場合、サンプルは祖先について「不明」と割り当てた。未知の祖先を有するサンプルは、祖先に基づく家系同一性(IBD)計算から除外した。

10

20

【0161】

実施例 4

IBD推定

以下のフラグ：`--maf 0.1 --geno 0.05 --snps-only --hwe 0.00001`を用いて完全なデータセットにPLINKを実行することにより、高品質な共通のバリエーションをフィルタリングした。次いで、2方向アプローチを採用してエクソームデータから正確なIBD推定値を得た。第1に、個体間のIBD推定値は、祖先解析から決定されたのと同じ祖先スーパークラス(例えば、AMR、AFR、EAS、EUR及びSAS)内で計算した。

【0162】

第2に、異なる祖先を有する個体間の1親等関係を把握するため、`--min 0.3` PLINKオプションを用いて全ての個人間でIBD推定値を計算した。次いで個体を、ネットワークノードが個体でありエッジが1親等関係である1親等家族ネットワークにグループ分けした。各1親等家族ネットワークをprePRIMUSパイプライン(Staples et al. (2014); Am. J. Hum. Genet. 95, 553-564)によって実行し、これによりサンプルの祖先を適切な祖先マイナーアレル頻度とマッチさせてIBD推定を改善した。このプロセスは、各家族ネットワーク内の個体間の1親等関係を正確に推定した(最小PI_HATは0.15)。

30

【0163】

実施例 5

92Kヒトエクソームのコホート内の関係推定及び関係性説明

92,455例のDiscovEHRデータセットから、43例の一卵双胎児、16,476例の親子関係、10,479例の完全同胞関係、及び39,000例の2親等関係が同定された(図5)。個体をノードとして、関係をエッジとして処理して無向グラフを作成した。1親等関係のみを用いて、12,594の結びついた構成要素を同定した。これを1親等ネットワークと称する。DiscovEHRコホート内の個体の39%が、データセット内で少なくとも1例の1親等近親者を有していた。

40

【0164】

【表 1】

表 1 (D i s c o v E H R データセットの祖先の内訳)

祖先クラス	サンプル数	人の%
E U R	88634	95.9%
A F R	1984	2.1%
A M R	959	1.0%
S A S	196	0.2%
E A S	194	0.2%
不明	488	0.5%

10

【 0 1 6 5 】

【表 2】

表 2 (D i s c o v E H R データセットの 1 親等関係に關与する
個体の祖先的背景の完全な内訳)

関係	祖先	カウント
一卵双胎児	E U R - E U R	42
一卵双胎児	S A S - S A S	1
親 - 子	E U R - E U R	16028
親 - 子	A F R - A F R	115
親 - 子	A F R - E U R	86
親 - 子	A M R - E U R	83
親 - 子	A M R - A M R	43
親 - 子	E U R - 不明	43
親 - 子	不明 - 不明	20
親 - 子	A F R - 不明	13
親 - 子	A M R - 不明	13
親 - 子	E A S - 不明	13
親 - 子	S A S - S A S	11
親 - 子	A F R - A M R	5
親 - 子	E U R - S A S	2
親 - 子	E A S - S A S	1
完全同胞	E U R - E U R	10364
完全同胞	A F R - A F R	155
完全同胞	A M R - E U R	24
完全同胞	A M R - A M R	16
完全同胞	不明 - 不明	10
完全同胞	A M R - 不明	4
完全同胞	S A S - S A S	2
完全同胞	E A S - E A S	1
完全同胞	E A S - 不明	1
完全同胞	E U R - S A S	1
完全同胞	E U R - 不明	1

【0166】

実施例 6

家系の再構築

D i s c o v E H R コホート内で同定された全ての 1 親等家族ネットワークを P R I M U S v 1 . 9 . 0 で再構築した。合わせた I B D 推定値を、遺伝的に導出された性及び E H R で報告された年齢と共に P R I M U S に提供した。P I _ H A T > 0 . 3 7 5 の関係性カットオフを指定して、1 親等ネットワークに対する再構築を限定した。

【0167】

これらの家系間においてメンデル様式で分離する 3 0 0 超の電子健康記録 (E H R) 由来の表現型がデータセットから見出され、大規模な伝統的メンデル解析を可能にする 2 ,

000件超の情報価値を有し得る家系—表現型の対が得られた。

【0168】

実施例7

家系エンリッチメント

1親等家族ネットワークからの個体は、少なくとも1つのバイナリー形質、極端な定量的形質、またはこれらの組合せについて「罹患者」または「非罹患者」と判定された。これらの罹患者のセットを家系と交差させて、家族ベースの分離解析を受け入れられるように十分な罹患者でエンリッチされた家系を同定した。

【0169】

データセットから2,978件の形質—家系エンリッチメント対が認識された（優性2,596件、劣性382件）。これらの形質—家系エンリッチメント対の中で、981件の家系内に1,015件の異なる形質を有する3,975例の罹患者が存在した。形質のうちの50%超が2つ以上の家系でエンリッチされ、357件の形質が3つ以上の家系でエンリッチされた。

【0170】

加えて、2,978件の形質—家系エンリッチメント対のうち1,911件は、809件の異なる形質及び673件の家系を伴うバイナリー形質—家系エンリッチメント対であった。バイナリー形質—家系エンリッチメント対において、最もエンリッチされた家系は、う歯に対するものであった（N=46）。さらに、2,978件の形質—家系エンリッチメント対のうち、1,067件は206件の異なる形質及び581件の家系を伴う定量的形質—家系エンリッチメント対であった。定量的形質—家系エンリッチメント対において、最もエンリッチされた家系は、高トリグリセリド__Med__LabValueに対するものであった（N=19）。

【0171】

7.1 原発性血栓形成傾向

原発性血栓形成傾向は、血栓形成（過凝固状態）に至る止血機構の遺伝性障害である。これは、一般的には静脈系に影響を及ぼす（例えば、深部静脈血栓症、肺塞栓症）。

【0172】

集団内の個体を、原発性血栓形成傾向（Ph e 1 0 __D 6 8 5、I C D 1 0 4 D）のバイナリー形質に基づいて罹患者であると判定した。

実施例6に挙げた方法を用いて再構築した家系（表3及び4）から、1親等家系をフィルターにかけて、可能な構造が1つだけではなく、共通の祖先を有する3例未満の原発性血栓形成傾向罹患者を有する全ての家系を取り除いて、原発性血栓形成傾向についてエンリッチされた家系を生成した。コホート内では、原発性血栓形成傾向（Ph e 1 0 __D 6 8 5、I C D 1 0 C M D 6 8 . 5）の有病率は1.3%であった。

【0173】

このようにして、原発性血栓形成傾向についてエンリッチされたいくつかの家系を同定した（図6を参照）。

【0174】

10

20

30

【表 3】

表 3

シリアル N o .	罹患者の数	共通の祖先 を有する 罹患者数の 最大値	共通の祖先 を有する情 報価値のあ る罹患者数 の最大値	罹患した 同胞対の数	家系作成 数
1	3	3	3	1	2
2	3	3	4	2	3
3	3	3	3	0	3
4	2	2	2	1	3
5	4	4	4	0	3
6	3	3	3	3	4
7	3	3	3	0	4
8	3	3	3	1	3

10

【 0 1 7 5】

20

【表 4】

表 4

シリアル N o .	罹患者 作成数	非罹患者 の数	未知の 罹患者の 数	家系内の サンプル 数	年齢が一 貫した可 能な家系 の数	遺伝 様式
1	2	3	0	6	1	優性
2	2	11	0	14	1	優性
3	2	6	0	9	1	優性
4	1	5	1	8	1	劣性
5	3	0	0	4	1	優性
6	1	29	2	34	1	優性
7	3	23	1	27	0	優性
8	2	6	0	9	1	優性

30

【 0 1 7 6】

40

7. 2 遺伝性出血性末梢血管拡張症

遺伝性出血性末梢血管拡張症（H T T）は、全身の血管に影響を及ぼし（血管異形成を引き起こす）、出血傾向をもたらす希少な常染色体優性遺伝障害である。（この状態は、オスラー・ウェーバー・ランデュ病（O W R D）としても知られており、この2つの用語は互換的に使用される）。H H Tは、皮膚粘膜末梢血管拡張及び動静脈奇形（A V M）によって顕在化し、これは重篤な病的状態及び死亡の潜在的原因である。病変は、鼻咽頭、中枢神経系（C N S）、肺、肝臓、及び脾臓、ならびに尿路、胃腸管（G I）、結膜、体幹、腕、及び手指に影響を及ぼし得る。

【 0 1 7 7】

集団内の個体をH T T（P h e 1 0 _ I 7 8 0、I C D 1 0 C M I 7 8 . 0）のバイ

50

ナリー形質に基づいて罹患者であると判定した。

H T T に対し実施例 6 で挙げた方法を用いて、2 つの家系を再構築した（表 5 及び 6 を参照）。両方の家系が共に、共通の祖先及び 1 つの可能な構造を有する 3 例の H H T 罹患者を有した。さらに、コホート内において、H T T の有病率は 0. 0 % であった。

【0178】

【表 5】

表 5

シリアル No.	罹患者の数	共通の祖先 を有する 罹患者数の 最大値	共通の祖先 を有する情 報価値のあ る罹患者数 の最大値	罹患した 同胞対の数	家系 作成数
1	3	3	3	0	3
2	3	3	3	0	3

10

【0179】

【表 6】

表 6

シリアル No.	罹患者 作成数	非罹患者 の数	未知の 罹患者の 数	家系内の サンプル 数	年齢が一 貫した可 能な家系 の数	遺伝 様式
1	3	1	1	5	1	優性
2	2	13	1	17	1	優性

20

【0180】

H T T におけるバイナリー形質についてエンリッチされた 2 件の家系を使用して、希少バリエーション分離解析を実施した（図 7 A 及び 7 B を参照）。

図 7 A で示される H T T についてエンリッチされた家系では、分離及び関連性の解析により、S M A D 4 遺伝子のバリエーションが家系における H T T 表現型と共に同時分離することが示された（表 7 を参照）。S M A D 4（S M A D ファミリーメンバー 4）は、シグナル伝達タンパク質の S M A D ファミリーのメンバーである。S m a d タンパク質は、形質転換成長因子（T G F）ーベータシグナル伝達に応答して、膜貫通型セリントレオニン受容体キナーゼによってリン酸化され活性化される。S M A D 4 は、他の活性化 S m a d タンパク質と共にホモマー複合体及びヘテロマー複合体を形成し、次いで核内に蓄積し標的遺伝子の転写を調節し、B M P シグナル伝達経路の重要な構成要素である。S M A D 4 の変異または欠失は、遺伝子障害である遺伝性出血性末梢血管拡張症候群（H H T）及びマイア症候群；ならびに若年性ポリポーシス症候群（染色体 18 q 21 上の S M A D 4 遺伝子のヘテロ接合性変異）を含む家族性がん感受性障害と関連している。S M A D 4 は腫瘍抑制因子として作用し、上皮細胞の増殖を阻害する。また、血管新生を低減し、血管透過性亢進を増大させることにより、腫瘍への阻害効果を有し得る。S M A D 4 の体細胞変異が膵臓癌で同定されている。

30

40

【0181】

【表 7】

表 7

遺伝子	遺伝子名	NT 変化	AA 変化	TGP 頻度	EXA C頻度	RGC 頻度	保存	予測	遺伝子 型
<i>SMAD4</i>	SMAD ファミリー メンバー 4	c.1242_1245 delAGA C	p. D415Efs *20	該当 なし	該当 なし	0.00001 1	該当 なし	有害	HET=4/ HOM=0

10

【0182】

図 7 B 及び図 8 で示される H T T についてエンリッチされた家系では、分離及び関連性の解析により、アクチビン A 受容体 I I 型様 1 (A C V R L 1) 遺伝子のバリエーションが家系における H T T 表現型と共に同時分離することが示された (表 8 を参照)。A C V R L 1 遺伝子は、リガンドの T G F ベータスーパーファミリーに対する I 型細胞表面受容体をコードし、受容体セリン／トレオニンキナーゼのサブファミリーを形成する他の密接に関連する A L K またはアクチビン受容体様キナーゼタンパク質と共に類似のドメイン構造を共有する。A C V R L 1 の変異は、ランデュ・オスラー・ウェーバー症候群 2 としても知られる出血性末梢血管拡張症 2 及び肺動脈高血圧と関連している。患者は、結膜の末梢血管拡張、鼻粘膜の末梢血管拡張 (しばしば疾患の最初の徴候として鼻出血に至る)、口腔の末梢血管拡張、様々な臓器の動静脈奇形、皮膚の末梢血管拡張、貧血を示し、肺動脈高血圧を発症する患者もいる。H H T 2 の内臓所見には、肺動静脈奇形 (P A V M)、脳 A V M、脊髄 A V M、肝臓 A V M、A V M による胃腸管出血、及び肝硬変が含まれた。H H T 2 の神経学的徴候としては、発作、虚血性卒中、片頭痛、脳動静脈奇形、及び脳内出血が挙げられる。

20

【0183】

【表 8】

表 8

遺伝子	遺伝子名	NT 変化	AA 変化	TGP 頻度	EXA C頻度	RGC 頻度	保存	予測	遺伝子 型
<i>ACVRL1</i>	アクチビン A 受容体様 1 型	c.C853T	p.L285F	該当 なし	該当 なし	0.00001 1	保存 的	有害	HET=4/ HOM= 0

30

【0184】

7. 3 肺活量測定による G O L D ステージ 2 ～ 4 の患者の肺気腫

40

肺気腫は、息切れを起こす肺の状態であり、慢性閉塞性肺疾患 (C O P D) を構成する疾患の 1 つである。肺気腫の患者では、肺の中にある気嚢 (肺胞) が損傷を受ける。経時的に気嚢の内壁が弱くなり破裂し、多数の小さな気腔の代わりに大きな気腔が作られる。これにより、肺の表面積が減少し、ひいては血流に到達する酸素の量が減少する。呼気の際に損傷した肺胞が適切に機能せず、古い空気が閉じ込められ、酸素を豊富に含む新鮮な空気が入る余地がなくなる。

【0185】

「肺活量測定による G O L D ステージ 2 ～ 4 患者の肺気腫」についてのバイナリー形質を肺機能検査の定量的形質から導出した。電子医療記録で報告された複数の発生率に基づいた非喫煙 C O P D 患者の高信頼性セットを使用した。肺機能検査における定量的形質の

50

1つは、「最も最近の肺活量測定による気管支拡張薬投与前の50パーセント努力肺活量における努力呼気速度から50パーセント努力肺活量における努力吸気速度」を用いて定義された。集団の形質の平均は0で、標準偏差は0.27であった。エンリッチメントは、定量的形質の下限を用いて実施した。肺機能検査の別の定量的形質は、「最も最近の肺活量測定による1秒間の予測される気管支拡張薬投与後の努力呼気肺活量のパーセント」を用いて定義された。集団の形質の平均は81.89で、標準偏差は20.84であった。エンリッチメントは、定量的形質の下限を用いて実施した。

【0186】

1親等家族ネットワークからの肺活量測定によるGOLDステージ2～4の患者における気腫のバイナリー形質についてエンリッチされた家系を単離した（図9を参照）。コホート内では、この特定の表現型の有病率は1.8%であった。家系は1つのみの可能な構造を有し、共通の祖先を有する3例の罹患者を含んでいた。

【0187】

7.4 腎臓移植

腎臓移植（Phe9__V420、ICD9DM V42.0）におけるバイナリー形質についてエンリッチされた家系を1親等家族ネットワークから単離した。この特定の表現型の有病率は0.8%であった。

【0188】

1親等家系は1つのみの可能な構造を有し、共通の祖先を有する4例の罹患者を有した。必要な基準を含む家系を同定した（図10及び表9を参照）。

【0189】

【表9】

表9

形質カテゴリー	健康状態及び保健サービスとの接触に影響する因子
罹患者の数	5
共通の祖先を有する罹患者数の最大値	5
共通の祖先を有する情報価値のある罹患者数の最大値	5
罹患した同胞対の数	2
家系作成数	2
罹患者作成数	2
非罹患者の数	3
未知の罹患者の数	0
家系内のサンプル数	8
年齢が一貫した可能な家系の数	1
遺伝様式	優性

【0190】

7.5 末期腎疾患

集団内の個体を、末期腎疾患（Phe10__5856、ICD9CM 585.6）のバイナリー形質に基づいて罹患者であると判定した。末期腎疾患についてエンリッチされたいくつかの家系が同定された（図11）。

【0191】

7.6 遺伝性運動感覚性ニューロパチー（シャルコー・マリー・トゥース病）

シャルコー・マリー・トゥース病（CMT）は最も一般的な遺伝性神経性障害の1つであり、米国ではおよそ2,500人に1人が罹患している。当該疾患は、遺伝性運動感覚性ニューロパチー（HMSN）または腓骨筋萎縮症としても知られており、末梢神経に影

響を及ぼす障害の群を含む。

【0192】

集団内の個体を遺伝性運動感覚性ニューロパチー（P h e 1 0 _ G 6 0 0、I C D 1 0 C M G 6 0 . 0）のバイナリー形質に基づいて罹患者であると判定した。コホート内では、この特定の表現型の有病率は0.1%であった。

【0193】

実施例6から再構築した家系から、遺伝性運動感覚性ニューロパチーの1親等家系は、1つの可能な構造及び共通の祖先を有する3例の罹患者を有した（図12及び表10を参照）。

【0194】

【表10】

表10

形質カテゴリー	神経系及び感覚器官の疾患
罹患者の数	3
共通の祖先を有する罹患者数の最大値	3
共通の祖先を有する情報価値のある罹患者数の最大値	3
罹患した同胞対の数	1
家系作成数	3
罹患者作成数	2
非罹患者の数	1
未知の罹患者の数	1
家系内のサンプル数	5
年齢が一貫した可能な家系の数	1
遺伝様式	優性

【0195】

遺伝性運動感覚性ニューロパチーについてエンリッチされた家系では、分離解析及び関連性解析により、トロポミオシン2（ベータ）（T P M 2）遺伝子のバリエントが、家系内の遺伝性運動感覚性ニューロパチー表現型と共に同時分離することが示された（表11）。T P M 2は、アクチンフィラメント結合タンパク質ファミリーのメンバーであるベータトロポミオシンをコードし、主に遅い1型筋線維で発現している。T P M 2の変異は他のサルコメアトロポミオシンタンパク質の発現を変化させ、キャップ病（c a p d i s e a s e）、ネマリンミオパチー、及び遠位関節拘縮症候群を引き起こし得る。

【0196】

【表11】

表11

遺伝子	遺伝子名	NT変化	AA変化	TGP頻度	EXAC頻度	RGC頻度	保存	予測	遺伝子型
TPM2	トロポミオシン2（ベータ）	c.G479A	p.R160H	該当なし	0.00002	0.00001	保存的	有害	HET=6/HOM=0

【0197】

様々な組織内でコードされた TPM2 の 100 万個当たりの転写産物 (TPM) の遺伝子発現データは、動脈、S 字結腸、食道－胃腸管接合部、食道－筋、及び骨格筋で高い発生率を示した (図 13 を参照)。

【0198】

家系内の罹患者の患者記録 (表 12 を参照) から、この家族は遺伝性運動感覚性ニューロパチーの証拠を示さず、むしろ TPM2 の変異によるネマリンミオパチー 4 型であることが示唆された (Donner et al. Neuromuscular Disorders (2009) 19:348-351)。

【0199】

【表 12】

10

表 12

シリアル No.	表現型の注記
1	足根管症候群、遺伝性末梢神経ニューロパチー、先天性遺伝性筋ジストロフィー、後天性足部変形、先天性遺伝性筋ジストロフィー、多発性関節拘縮、筋力低下 (全身性)、四肢の痛み、腓骨筋萎縮、息切れ
2	腓骨筋萎縮、後天性内反凹足変形、遺伝性進行性筋ジストロフィー、部位不明の単神経炎、関節、足首、及び足部の痛み、四肢の痛み、他に分類されるその他の疾患における多発神経障害、分類不能な遺伝性及び特発性末梢神経障害
3	腓骨筋萎縮、歩行異常、遺伝性末梢神経障害、その他の錐体外路疾患及び異常な運動障害、後天性足部変形

20

【0200】

7. 7 双極性障害

双極性障害または「躁うつ病」は、感情の高揚 (躁状態または軽躁状態) 及び消沈 (うつ状態) を含めた極端な気分の変化を引き起こす。人口の約 2.6% (米国人の成人 570 万例) が毎年この障害に罹患している。

30

【0201】

集団中の個体を、双極性障害及び単極性障害に基づいて罹患者であると判定した。双極性障害の ICD-10 コードは F31 であり、ICD-9 コードは 296.4~296.7 である。患者のサブセット (35~40%) はリチウムの処方を受ける。単極/大うつ病性障害の ICD-10 コードは F32、F33、F39 であり、ICD-9 コードは 296.2/、3/、9 (家族ネットワーク内の 2 親等) である。自閉症 (ICD-10 コード F84) 及び精神遅滞 (ICD-10 コード F70.9、F71.9、F72.9、F73.9、F79.9) を有する個体を罹患者セットから除外した。コホート内の双極性障害 (F319:3.2%) 及び単極性障害 (F31、F32、及び F33:それぞれ 0.0%、4.1%、及び 2.1%) の有病率は 5% 未満であった。

40

【0202】

双極性障害のバイナリー形質についてエンリッチされた家系を 1 親等家族ネットワークから単離した。

1 親等の家系を評価して、1 つのみの可能な構造を有し、共通の祖先を有する少なくとも 3 例の罹患者を有することを確認した (図 14 を参照)。エンリッチされた家系に実施した分離解析で、表現型と共に同時分離する可能なバリエーションのリストが作成された (表 13)。表現型と共に同時分離するバリエーション C20orf203 は有害であり、非保存的である。

【0203】

50

【表 1 3 - 1】

表 1 3

遺伝子	遺伝子名	N T 変化	A A 変化	T G P 頻度	E X A C 頻度	R G C 頻度	保存	予測	遺伝子型
<i>C20orf203</i>	20 番染色体オープンリーディングフレーム 203	c.113delC	p.T38fs	該当なし	該当なし	1.11E-05	該当なし	有害	HET=4/ HOM=0
<i>TUBGCP6</i>	チューブリンガンマ複合体関連タンパク質 6	c.C2119T	p.Q707X	該当なし	該当なし	1.11E-05	保存的	有害	HET=4/ HOM=0
<i>SEC14L3</i>	SEC14 様脂質結合 3	c.A346T	p.K116X	該当なし	8.38E-06	1.11E-05	保存的	有害	HET=4/ HOM=0
<i>TMPRSS4</i>	膜貫通プロテアーゼ、セリン 4	c.C358T	p.Q120X	該当なし	8.27E-06	2.21E-05	保存的	許容	HET=8/ HOM=0
<i>CARD10</i>	カスパーゼ動員ドメインファミリーメンバー 10	c.C550A	p.R184S	該当なし	該当なし	3.88E-05	保存的	有害	HET=1 4/HOM=0
<i>LMAN1L</i>	レクチン、マンノース結合 1 様	c.C568T	p.R190W	該当なし	4.79E-05	0.000144	保存的	許容	HET=5 2/HOM=0
<i>C10orf71</i>	10 番染色体オープンリーディングフレーム 71	c.C3715A	p.P2139T	該当なし	該当なし	1.38E-05	保存的	有害	HET=5/ HOM=0
<i>TMEM181</i>	膜貫通タンパク質 181	c.G1657A	p.A553T	該当なし	該当なし	1.38E-05	保存的	許容	HET=5/ HOM=0

10

20

30

40

【表 1 3 - 2】

<i>MAN2 B1</i>	マンノシダーゼアルファクラス 2B メンバー 1	c.G2094C	p.Q698H	該当なし	4.94E-05	7.75E-05	中立	有害	HET=28/HOM=0
<i>IFT17 2</i>	鞭毛内輸送 172	c.G2681A	p.R894H	該当なし	該当なし	3.05E-05	保存的	有害	HET=11/HOM=0
<i>PIWIL 1</i>	piwi 様 RNA 介在遺伝子サイレンシング 1	c.G152A	p.R51Q	該当なし	8.26E-06	3.05E-05	保存的	有害	HET=11/HOM=0
<i>ACVR L1</i>	アクチビン A 受容体様 1 型	c.G948C	p.E316D	該当なし	4.16E-05	0.000125	中立	有害	HET=45/HOM=0
<i>TTLL1 0</i>	チューブリンチロシンリガーゼ様 10	c.G211A	p.E71K	0.0002	3.37E-05	1.66E-05	中立	許容	HET=6/HOM=0
<i>CDBF A2T2</i>	CBFA2/RUNX1 転座パートナー 2	c.C135A	p.N45K	該当なし	8.24E-06	0.000105	保存的	有害	HET=38/HOM=0
<i>ZBTB 2</i>	ジンクフィンガー及び BTB ドメイン含有 2	c.A1493G	p.H498R	該当なし	1.65E-05	3.05E-05	保存的	許容	HET=11/HOM=0

【0205】

FLJ33706 (代替遺伝子記号 C2orf203) は、ニコチン中毒の原因となる可能なバリエーションとして同定されている。20 番染色体オープンリーディングフレーム 203 (C2orf203) の 100 万個あたりの転写産物 (TPM) の遺伝子発現データは、種々の組織でコードされるが、主に小脳半球及び脳の小脳で発現する (図 15)。連鎖研究により、FLJ33706 の 3' UTR に位置する SNP である rs17123507 が、ニコチン中毒に対する感受性と有意に関連していることが確認されている (Liet al. PLoS Computational Biology (2010) 6:e1000734)。

【0206】

さらに、2つのよりエンリッチされた家系が同定された (図 16 及び 17; 表 14 及び表 15 を参照)。いずれの家系も 1つのみの可能な構造を有し、共通の祖先を有する 3例超の罹患者を有した。

【0207】

【表14】

表14

シリアル No.	罹患者の数	共通の祖先 を有する 罹患者数の 最大値	共通の祖先 を有する 情報価値の ある罹患者 数の最大値	罹患した 同胞対の数	家系 作成数
1	5	3	3	0	2
2	4	3	2	0	2

10

【0208】

【表15】

表15

シリアル No.	罹患者 作成数	非罹患者 の数	未知の 罹患者の 数	家系内の サンプル 数	年齢が一 貫した可 能な家系 の数	遺伝 様式
1	2	2	0	7	1	優性
2	2	3	1	8	1	優性

20

【0209】

加えて、双極性障害についてバイナリー形質がエンリッチされた別の家系は、1つのみの可能な構造のみを有し、共通の祖先を有する3人以上の罹患者を有した（図18を参照）。

【0210】

エンリッチされた家系に実施したバリエーション解析で、表現型と共に同時分離する可能なバリエーションのリストが作成された（表16）。

30

【0211】

【表 16-1】

表 16

遺伝子	遺伝子名	NT 変化	AA 変化	TGP 頻度	EXA C頻度	RG C頻度	保存	予測	遺伝 子型
<i>MC PHI</i>	マイクロセ ファリン1	c.2453- 1G>C)		該当 なし	1.675E- 05	3.32E -05	保存的	有害	HET= 12/HO M=0
<i>NOL 8</i>	核小体タン パク質	c.38del G	p.G13 fs	該当 なし	166E- 05	3.60E -05	該当 なし	有害	HET= 13/HO M=0
<i>INS M2</i>	INSM 転写 リプレッサ ー2	c.1031_ 1035del	p.344 _345del	該当 なし	該当 なし	5.54E -05	該当 なし	有害	HET= 20/HO M=0
<i>CDR T15 L2</i>	CMT1A 複 製転写産物 15 様 2	c.345_3 46del	p.115 _1116 del	該当 なし	0.0002	0.000 235	該当 なし	有害	HET= 85/HO M=0
<i>SHP K</i>	セドヘプツ ロキナーゼ	c.C355 T	p.R11 9X	0.00059 9	0.0005	0.000 684	中立	許容	HET= 245/H OM=0
<i>INS M2</i>	INSM 転写 リプレッサ ー2	c.A686 G	p.D22 9G	該当 なし	5.56E- 05	0.000 13	保存的	有害	HET= 47/HO M=0
<i>DAA M1</i>	ディシュベ ルト関連形 態形成活性 化因子1	c.T247 4G	p.F82 5C	該当 なし	6.59E- 05	0.000 191	保存的	有害	HET= 69/HO M=0
<i>APB A2</i>	アミロイド ベータ前駆 体タンパク 質結合ファ ミリーA メ ンバー2	c.C141 7T	p.R47 3C	0.0002	8.46E- 06	9.97E -05	保存的	有害	HET= 36/HO M=0
<i>FAM 107 B</i>	配列類似性 107 メンバ ーB のファ ミリー	c.A155 G	p.E52 G	該当 なし	該当 なし	5.26E -05	保存的	有害	HET= 19/HO M=0
<i>AKR 1C3</i>	アルドケト レダクター ゼファミリー ー1 メンバ ーC3	c.G512 T	p.R17 1M	該当 なし	8.24E- 06	6.64E -05	中立	有害	HET= 24/HO M=0
<i>PLE KH G3</i>	プレクスト リン相同性 及び RhoG F ドメイン 含有 G3	c.G80T	p.G27 V	該当 なし	1.87E- 05	3.05E -05	保存的	有害	HET= 11/HO M=0

【表 16-2】

遺伝子	遺伝子名	NT 変化	AA 変化	TGP 頻度	EXA C頻度	RG C頻度	保存	予測	遺伝子型
<i>PDS S1</i>	プレニル (デカプレ ニル) ニリ ン酸シタ ーゼ、サブ ユニット1	c.G113 8A	p.A38 0T	該当 なし	4.95E- 05	8.30E -05	保存的	有害	HET= 30/HO M=0
<i>CC M2</i>	CCM2 足場 タンパク質	c.G391 A	p.D13 1N	0.0002	0.0003	0.000 221	保存的	有害	HET= 80/HO M=0
<i>LRR C37 B</i>	ロイシンリ ッチリピー ト含有 37B	c.A218 8C	p.K73 0Q	該当 なし	4.98E- 05	9.41E -05	中立	許容	HET= 34/HO M=0
<i>EML 4</i>	棘皮類微小 管関連タン パク質様4	c.A736 T	p.N24 6Y	該当 なし	1.65E- 05	3.88E -05	保存的	有害	HET= 14/HO M=0
<i>ARH GEF 10</i>	Rho グアニ ンヌクレオ チド交換因 子 10	c.C309 7T	p.H10 33Y	該当 なし	該当 なし	1.66E -05	保存的	有害	HET= 6/HO M=0
<i>STA T5A</i>	シグナル伝 達性転写因 子 5A	c.G779 A	p.R26 0Q	該当 なし	3.34E- 05	0.000 188	保存的	有害	HET= 66/HO M=1
<i>RP1 L1</i>	色素性網膜 炎 1 様 1	c.G149 3T	p.S49 81	該当 なし	8.76E- 06	4.43E -05	中立	許容	HET= 16/HO M=0
<i>ATA D2</i>	ATPアー ゼファミリ ー、AAA ドメイン含 有 2	c.C221 9A	p.S74 0Y	該当 なし	該当 なし	0.000 102	保存的	許容	HET= 37/HO M=0
<i>RBM 43</i>	RNA 結合 モチーフタ ンパク質 43	c.C35G	p.A12 G	該当 なし	3.33E- 05	0.000 147	中立	許容	HET= 53/HO M=0

【0213】

表 17 に挙げたバリエーションのうち、マイクロセファリン 1 (MCPH1) は原発性小頭症の病原性バリエーションとして報告されている。種々の組織内でコードされた MCPH1 の 100 万個当たりの転写産物 (TPM) の遺伝子発現データは、いくつかの組織内で高い発現率を示した (図 19 を参照)。

【0214】

原発性小頭症 1 型は、年齢関連平均より 3 標準偏差超低い頭囲を特徴とする。脳重量は著明に減少し、大脳皮質は不均衡に小さい。罹患個体は、重度の知的障害を有する。また、一部の MCPH1 患者は、成長遅延、低身長、ならびに細胞遺伝学的調製物及び低品質の中期 G バンドで検出される多数の前期様細胞によって示されるような誤調節された染色体凝縮も示す。

【0215】

【表17】

表17

遺伝子	遺伝子名	NT 変化	AA 変化	TG P頻度	EXAC 頻度	RGC 頻度	保存	予測	遺伝子 型
MC PHI	マイクロ セファリ ン1	c.2453 -1G>C	スプライ シング	該当 なし	0.000016 67	0.0000 3321	保存 的	有害	HET=1 2/HOM =0

10

【0216】

7. 8 サラセミア

サラセミアは、体内のヘモグロビン及び赤血球が正常よりも少ないことを特徴とする遺伝性の血液障害である。サラセミアにおけるヘモグロビンの低下及び赤血球の減少が貧血を引き起こし、患者を疲労状態にし得る。

【0217】

サラセミアのICD 10コードはD56である。

サラセミアのバイナリー形質についてエンリッチされた家系を1親等家族ネットワークから単離した。

20

【0218】

1親等の家系を評価して、1つのみの可能な構造を有し、共通の祖先を有する少なくとも3例の罹患者を有することを確認した（図20を参照）。2つのエンリッチされた家系が同定された（図20を参照）。いずれの家系も1つのみの可能な構造を有し、3例以上の罹患者を有した。

【0219】

エンリッチされた家系に実施したバリエーション解析で、表現型と共に同時分離するHBB遺伝子の可能なバリエーションのリストが作成された。HBB遺伝子は、ベータグロビンと呼ばれるタンパク質の合成を命令する。ベータグロビンは、赤血球の内部にあるヘモグロビンという大きなタンパク質の構成要素（サブユニット）である。成人の場合、ヘモグロビンは、通常4つのタンパク質サブユニット、すなわち、ベータグロビンの2つのサブユニットと、アルファグロビンと呼ばれる別のタンパク質の2つのサブユニットとからなり、アルファグロビンはHBAと呼ばれる別の遺伝子から生成される。これらのタンパク質サブユニットの各々は、ヘムと呼ばれる鉄含有分子に付着（結合）しており、各ヘムはその中心に鉄分子を含み、鉄分子は1つの酸素分子に結合することができる。赤血球内のヘモグロビンは、肺の酸素分子に結合する。次にこれらの細胞は血流を介して移動し、全身の組織に酸素を送達する。HBB遺伝子に関連する疾患としては、ベータサラセミア及び鎌状赤血球貧血が挙げられる。

30

【0220】

表現型と共に同時分離するHBB遺伝子内で同定された2つの変異は、Gln40におけるストップゲイン変異及びGly84におけるフレームシフト変異であった（関連性解析のp値は $< 3.1 \times 10^{-19}$ ）。これらの同定された変異を研究し、この知見を用いて家族性サラセミアを治療するための可能な治療アプローチをさらに発展させることができる。

40

【0221】

7. 10 アルカリホスファターゼ低下外来患者の中心傾向値

アルカリホスファターゼの通常の臨床検査は、症候性の患者の診断目的と無症候性の患者のスクリーニング目的との両方において、病院でかなり頻繁に実施される。アルカリホスファターゼ酵素は全身の組織に存在するが、肝臓及び骨の疾患を有する患者では非常に

50

頻繁に上昇する。

【0222】

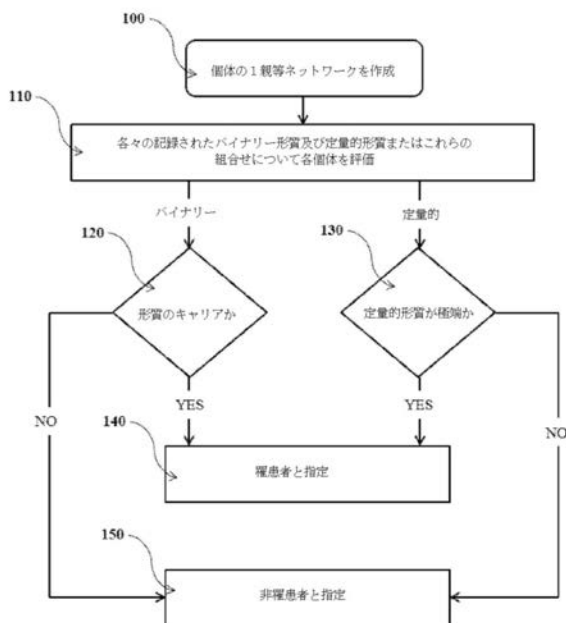
アルカリホスファターゼレベル低下についてエンリッチされた家系を作出し、この家系を評価して、1つのみの可能な構造を有し、共通の祖先を有する少なくとも3例の罹患者を有することを確認した（図21を参照）。

【0223】

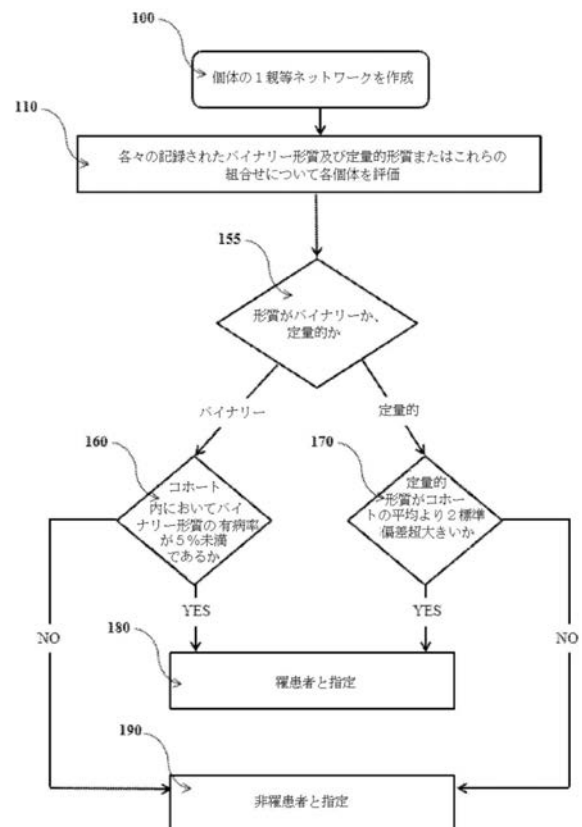
エンリッチされた家系に実施したバリエーション解析により、ALPL遺伝子のミスセンス変異が表現型と共に同時分離することが示された。ALPL遺伝子は、組織非特異的アルカリホスファターゼ（TNALP）と呼ばれる酵素の合成を命令する。この酵素は、骨及び歯の成長及び発達に重要な役割を果たしている。この酵素は、他の多くの組織でも活性であり、特に肝臓及び腎臓で活性である。この酵素はホスファターゼとして作用し、これは、他の分子から酸素及びリン原子のクラスター（リン酸基）を取り除くことを意味する。TNALPは、発育中の骨及び歯にカルシウム及びリンなどのミネラルが沈着する石灰化プロセスに必須である。石灰化は、強く硬い骨、ならびに咀嚼及び歯ぎしりに耐えられる歯を形成するために不可欠である。ALPL遺伝子内で同定されたヘテロ接合性ミスセンス変異は、Leu275における変異（Leu275Pro）であった（図21を参照）（関連性解析のp値は $< 7.2 \times 10^{-27}$ ）。関連性の結果及びメンデルの分離は、バリエーションと定量的形質の減少との間の関連性についてのある程度独立した証拠を示している。

10

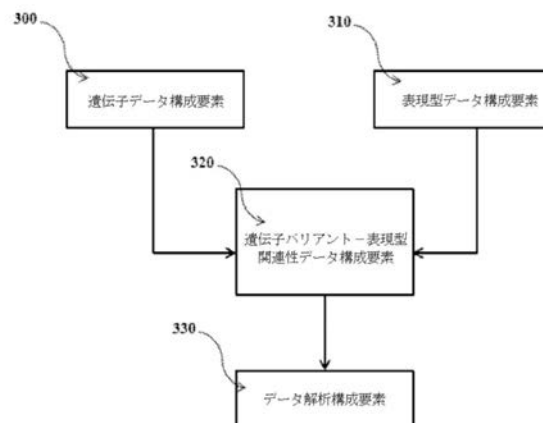
【図1】



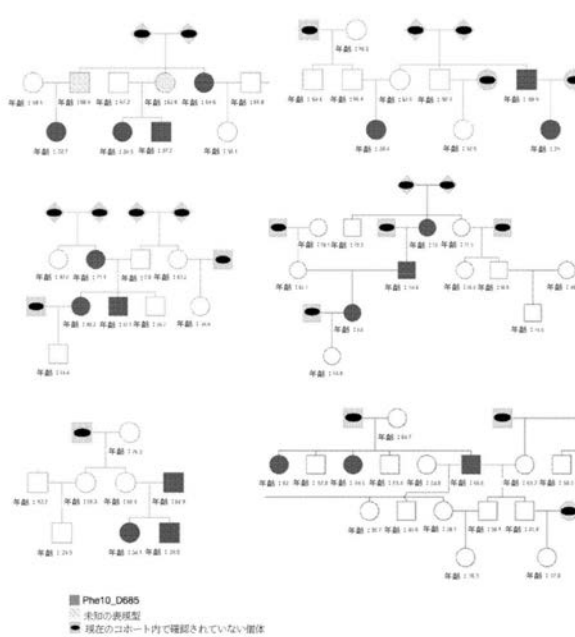
【図2】



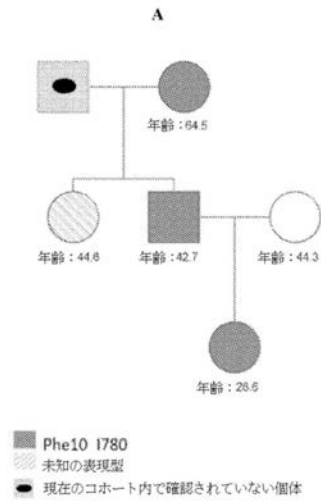
【図 4】



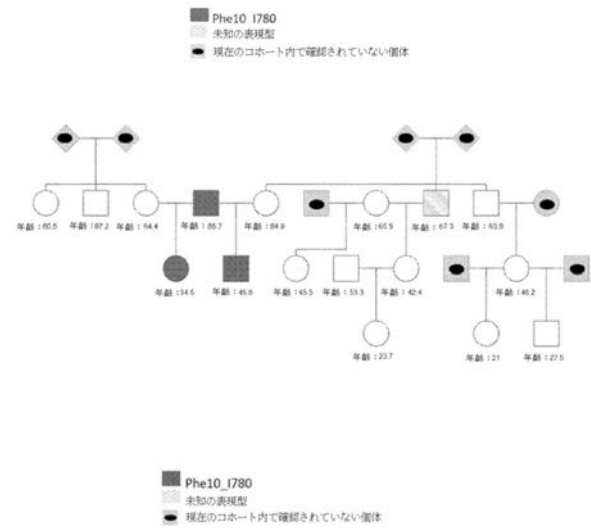
【图 6】



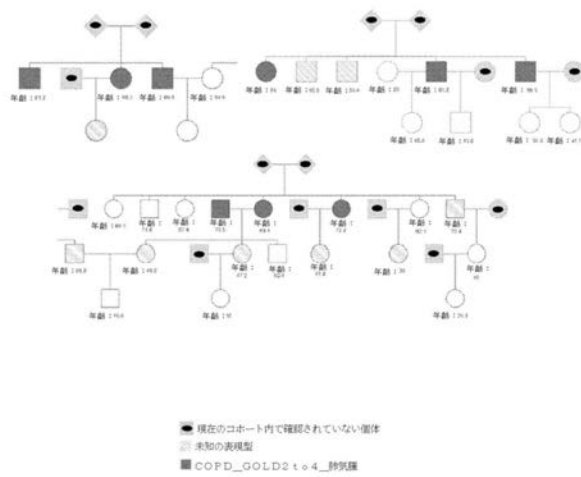
【図 7】



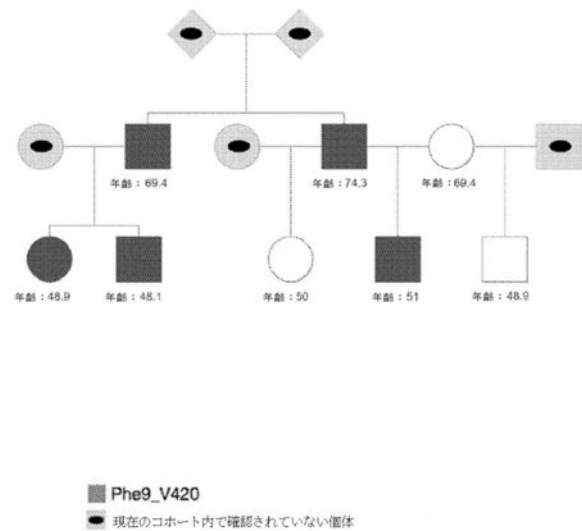
【図 8】



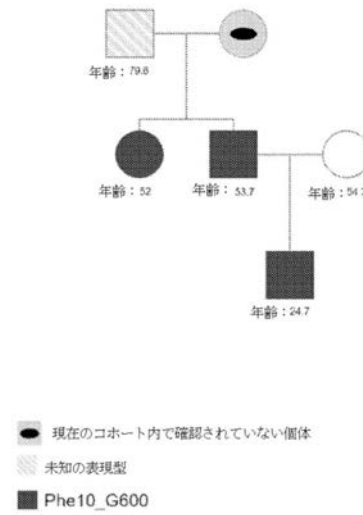
【図 9】



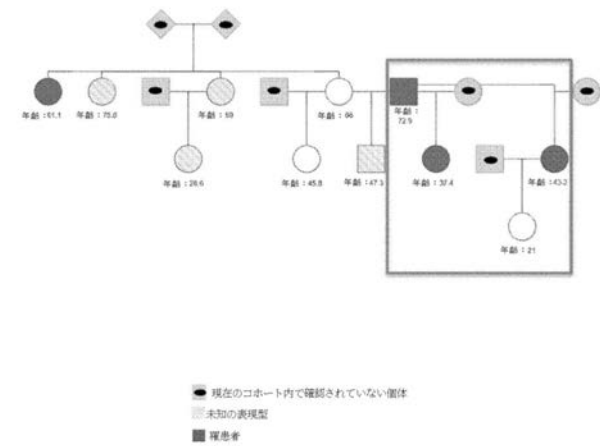
【図 10】



【図 1 2】



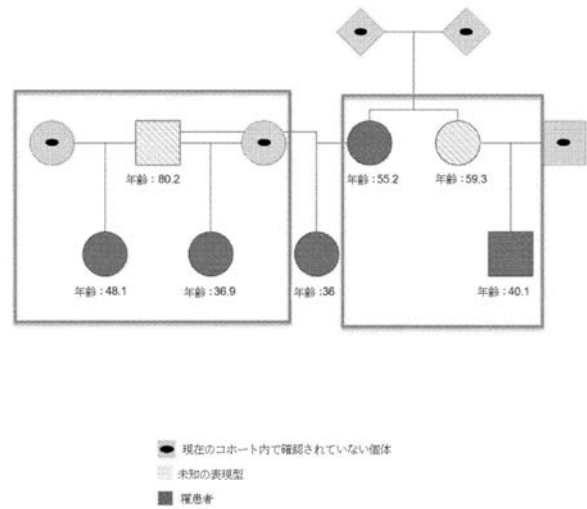
【図 14】



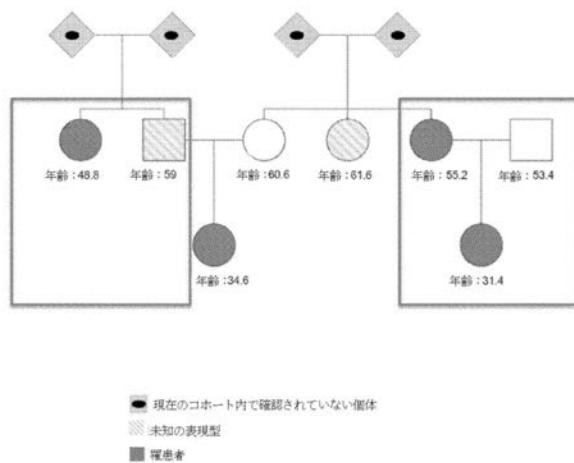
【図 15】



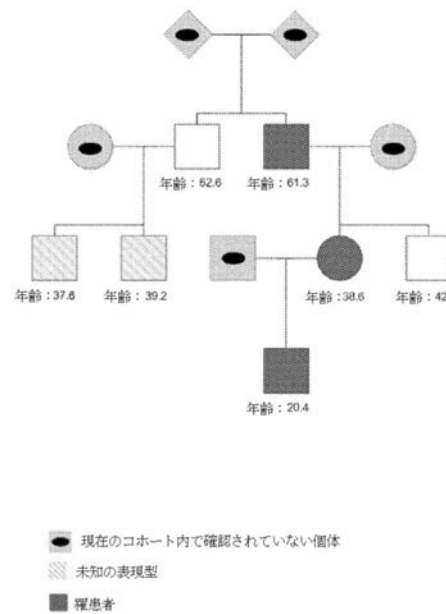
【図 16】



【図 17】



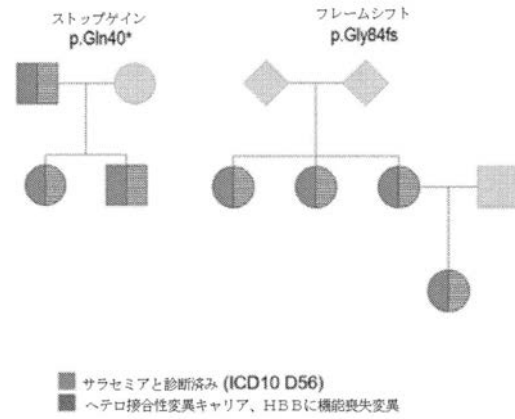
【図 18】



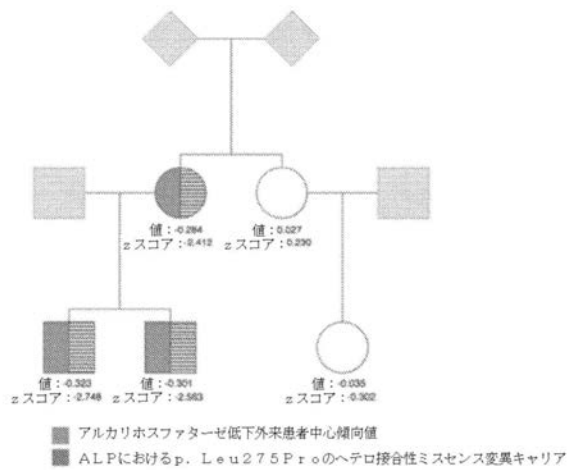
【図 19】



【図 20】



【図 21】



【国際調査報告】

INTERNATIONAL SEARCH REPORT

International application No
PCT/US2019/049942

A. CLASSIFICATION OF SUBJECT MATTER

INV. G16B20/20
ADD.

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
G16B

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

EPO-Internal, WPI Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 2017/286594 A1 (REID JEFFREY [US] ET AL) 5 October 2017 (2017-10-05) paragraphs [0013], [0018], [0174], [0297]	1-63
Y	----- US 2014/080718 A1 (HYLAND FIONA [US] ET AL) 20 March 2014 (2014-03-20) paragraphs [0031], [0032], [0058], [0079], [0089] ----- -/--	1-63

☒ Further documents are listed in the continuation of Box C.

☒ See patent family annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"Z" document member of the same patent family

Date of the actual completion of the international search

11 November 2019

Date of mailing of the international search report

18/11/2019

Name and mailing address of the ISA/

European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel: (+31-70) 340-2040,
Fax: (+31-70) 340-3016

Authorized officer

Schmidt, Karsten

INTERNATIONAL SEARCH REPORT

International application No

PCT/US2019/049942

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	JEFFREY STAPLES ET AL: "PRIMUS: Rapid Reconstruction of Pedigrees from Genome-wide Estimates of Identity by Descent", AMERICAN JOURNAL OF HUMAN GENETICS, vol. 95, no. 5, 1 November 2014 (2014-11-01), pages 553-564, XP055332199, US ISSN: 0002-9297, DOI: 10.1016/j.ajhg.2014.10.005 the whole document -----	1-63

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No

PCT/US2019/049942

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 2017286594	A1	05-10-2017	AU 2017242028 A1 06-09-2018 CA 3018186 A1 05-10-2017 CN 109155149 A 04-01-2019 EP 3437001 A1 06-02-2019 JP 2019515369 A 06-06-2019 KR 20180132727 A 12-12-2018 SG 11201808261R A 30-10-2018 US 2017286594 A1 05-10-2017 WO 2017172958 A1 05-10-2017
US 2014080718	A1	20-03-2014	US 2014080718 A1 20-03-2014 WO 2014043298 A1 20-03-2014

フロントページの続き

(81)指定国・地域 AP(BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), EA(AM, AZ, BY, KG, KZ, RU, TJ, TM), EP(AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OA(BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG), AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT

(74)代理人 100152489

弁理士 中村 美樹

(72)発明者 ステープルズ、ジェフリー

アメリカ合衆国 10591 ニューヨーク州 タリータウン オールド ソー ミル リバー
ロード 777

(72)発明者 ゴンザガーハウレギ、クラウドディア

アメリカ合衆国 10591 ニューヨーク州 タリータウン オールド ソー ミル リバー
ロード 777

(72)発明者 リード、ジェフリー

アメリカ合衆国 10591 ニューヨーク州 タリータウン オールド ソー ミル リバー
ロード 777

(72)発明者 ハベガー、ルーカス

アメリカ合衆国 10591 ニューヨーク州 タリータウン オールド ソー ミル リバー
ロード 777

Fターム(参考) 5L099 AA15