(19)日本国特許庁(JP)　(12)特　許　公　報(B2)　(11)特許番号
特許第6539835号
(P6539835)
(45)発行日　令和1年7月10日(2019.7.10)　(24)登録日　令和1年6月21日(2019.6.21)

(51)Int.Cl.　F I
G16B 30/00 (2019.01)　GO6F 19/22

請求項の数 21　(全 43 頁)

(21)出願番号	特願2017-251051 (P2017-251051)
(22)出願日	平成29年12月27日 (2017.12.27)
(62)分割の表示	特願2016-116859 (P2016-116859)の分割
原出願日	平成23年12月20日 (2011.12.20)
(65)公開番号	特開2018-67350 (P2018-67350A)
(43)公開日	平成30年4月26日 (2018.4.26)
審査請求日	平成29年12月27日 (2017.12.27)
(31)優先権主張番号	13/373,550
(32)優先日	平成23年11月18日 (2011.11.18)
(33)優先権主張国	米国 (US)

(73)特許権者　592034548
ザ・リージェンツ・オブ・ザ・ユニバーシティー・オブ・カリフォルニア
THE REGENTS OF THE UNIVERSITY OF CALIFORNIA
アメリカ合衆国、カリフォルニア州　９４６０７－５２００、オークランド、フランクリン　ストリート　１１１１、フィフス・フロアー
(74)代理人　100102978
弁理士　清水　初志
(74)代理人　100102118
弁理士　春名　雅夫

最終頁に続く

(54)【発明の名称】ＢＡＭＢＡＭ：高スループット配列決定データの並列比較分析

(57)【特許請求の範囲】
【請求項１】
ゲノム差分配列分析システムであって、
患者の腫瘍の腫瘍ゲノム配列及び患者の健康な組織の生殖系列ゲノム配列を記憶する第１のデータベースと、
前記第１のデータベースと接続され、ソフトウェア指示を記憶する有形の一時的でないコンピュータ読取可能なメモリ、及び、少なくとも１つのプロセッサを含む配列解析エンジンと、
を含み、
前記プロセッサは、前記ソフトウェア指示の実行時に、
前記第１のデータベースから、前記腫瘍ゲノム配列の少なくとも一部を表す腫瘍ストリング、及び、前記生殖系列ゲノム配列の少なくとも一部を表す生殖系列ストリングを取得し、ここで、前記腫瘍ストリング及び前記生殖系列ストリングは少なくとも１つのゲノム位置で互いに重複し、
前記少なくとも１つのゲノム位置で、前記腫瘍ストリングと前記生殖系列ストリングとの間の差分を含む局所差分ストリングを生成し、ここで、前記局所差分ストリングはタンパク質コード配列を含み、
第２のデータベース内で、前記局所差分ストリングに基づいて、前記患者に関連づけられている差分配列オブジェクトを更新する、
よう構成可能なゲノム差分配列分析システム。

【請求項２】
前記腫瘍ゲノム配列は、前記患者の腫瘍のポリペプチドコード配列を含む、請求項１に記載のシステム。
【請求項３】
前記腫瘍ストリングは、前記患者の腫瘍のポリペプチドコード配列を含む、請求項２に記載のシステム。
【請求項４】
前記生殖系列ゲノム配列は、前記患者の腫瘍のポリペプチドコード配列を含む、請求項１に記載のシステム。
【請求項５】
前記生殖系列ストリングは、前記患者の腫瘍のポリペプチドコード配列を含む、請求項４に記載のシステム。
【請求項６】
前記差分配列オブジェクトは、ポリペプチドコード配列を含む、請求項１に記載のシステム。
【請求項７】
前記差分配列オブジェクトは、前記腫瘍ストリングと前記生殖系列ストリングとの間のタンパク質コード配列における差分を含む、請求項１に記載のシステム。
【請求項８】
前記差分配列オブジェクトは、前記腫瘍ストリングと前記生殖系列ストリングとの間の核酸配列における差分を含む、請求項１に記載のシステム。
【請求項９】
前記核酸配列における差分は、ＤＮＡ配列又はＲＮＡ配列を含む、請求項８に記載のシステム。
【請求項１０】
前記配列解析エンジンは、さらに、前記差分配列オブジェクトに基づいて、参照ゲノムに対する患者に特異的な偏差を計算するよう構成可能である、請求項１に記載のシステム。
【請求項１１】
前記患者に特異的な偏差が、患者に特異的な偏差プロフィールの一部である、請求項１０に記載のシステム。
【請求項１２】
前記第１のデータベースは、次のファイルフォーマット：ＢＡＭフォーマット及びＳＡＭフォーマットのうち、少なくとも１つに従って、前記腫瘍ゲノム配列及び前記生殖系列ゲノム配列を記憶するように構成されている、請求項１に記載のシステム。
【請求項１３】
前記差分配列オブジェクトは、前記腫瘍ゲノム配列と前記生殖系列ゲノム配列との間の差分ストリングの一群を含む、請求項１に記載のシステム。
【請求項１４】
前記少なくとも１つのゲノム位置は、既知の参照ゲノム配列に対するものである、請求項１に記載のシステム。
【請求項１５】
前記配列解析エンジンは、前記差分配列オブジェクト及び類似の既知の差分配列オブジェクトに基づいて患者に特異的な指示を生成するようにさらに構成可能である、請求項１に記載のシステム。
【請求項１６】
前記患者に特異的な指示は、治療、診断、予後診断、処置結果の予測、危険性評価、及び処方の少なくとも１つを含む、請求項１５に記載のシステム。
【請求項１７】
前記配列解析エンジンは、以下の管理機能：

前記差分配列オブジェクトを作成し、前記局所差分ストリングを含める機能、
前記局所差分ストリングに従って前記差分配列オブジェクトを修正する機能、
前記局所差分ストリングと共に前記差分配列オブジェクトを追加する機能、
前記差分配列オブジェクトの少なくとも一部を削除する機能、
前記差分配列オブジェクトの少なくとも一部をコピーする機能、及び
前記差分配列オブジェクトの少なくとも一部を解析する機能
のうちの少なくとも１つに従って、前記局所差分ストリングに基づいて前記差分配列オブジェクトを更新するようにさらに構成可能である、請求項１に記載のシステム。

【請求項１８】

前記差分配列オブジェクトは、メタデータ属性を含む、請求項１に記載のシステム。

【請求項１９】

前記メタデータ属性は、タイムスタンプ、試料型、患者名、組織型、組織の状態、新生物成長、倍数性、遺伝子のコピー数、反復のコピー数、反転、欠失、挿入、ウイルス挿入、体細胞突然変異、生殖系列突然変異、再編成、転位、及びヘテロ接合性の喪失
のうち、少なくとも１つを属性とする、請求項１８に記載のシステム。

【請求項２０】

前記差分配列オブジェクトは、前記メタデータ属性を介して照会可能である、請求項１８に記載のシステム。

【請求項２１】

前記患者の健康な組織の前記生殖系列ゲノム配列は、前記患者からの血液試料を表す、請求項１に記載のシステム。

【発明の詳細な説明】

【技術分野】

【０００１】

他出願との関連性

本出願は、その全体が参照により本明細書に組み込まれている、２０１１年１１月１８日出願の表題「Ｂａｍｂａｍ：高スループット配列決定データの並列比較分析（Ｂａｍｂａｍ：Ｐａｒａｌｌｅｌ　Ｃｏｍｐａｒａｔｉｖｅ　Ａｎａｌｙｓｉｓ　Ｏｆ　Ｈｉｇｈ－Ｔｈｒｏｕｇｈｐｕｔ　Ｓｅｑｕｅｎｃｉｎｇ　Ｄａｔａ）」の米国非仮特許出願第１３／３７３，５５０号に関連し、その優先権を主張するものである。

本発明は、以下の米国連邦政府関連機関からの助成金を使用して一部行われた：国立癌研究所番号１Ｕ２４ＣＡ１４３８５８－０１。米国連邦政府は本発明の一定の権利を有する。

本発明は、データを処理し、個体又は対象における生物学的経路の構成要素を同定し、それによって個体又は対象が障害又は疾患の危険性にあるかどうかを決定する方法に関する。この方法は、ＳＡＭ／ＢＡＭにフォーマット済みのファイルに記憶されているショートリードアラインメントを使用した、個体又は対象の腫瘍及び生殖系列配列決定データの比較分析を行うためのツールとして使用し得る。データを処理する方法は、全体的及び対立遺伝子特異的なコピー数を計算し、対立遺伝子不均衡の領域にわたって生殖系列配列をフェージングし、体細胞及び生殖系列配列の変異を発見し、体細胞及び生殖系列の構造的変動の領域を推量する。また、本発明は、その方法を使用して、癌、自己免疫疾患、細胞周期障害、又は他の障害に感受性があるかどうかを診断することにも関する。

【背景技術】

【０００２】

現代の癌処置の中心となる前提は、患者の診断、予後診断、危険性評価、及び処置応答の予測が、腫瘍のゲノム、転写及びエピゲノムの特徴と共に診断時に集められる関連性のある臨床的情報（たとえば、患者の病歴、腫瘍組織学及び段階）ならびに続く臨床的経過観察データ（たとえば、治療レジメン及び疾患再発事象）に基づいた癌の層別化によって改善させることができることである。

配列決定における近年の進歩は、個々の生物及び一生物の組織の両方、ならびに明確な

集団及び種でさえもの、大量のゲノム及びサブゲノムデータをもたらしている。このことは、様々な疾患のゲノムに基づく個別化処置又は診断、予後診断／危険性評価、ならびにゲノム、転写、及び／又は後成的情報を使用した処置応答の予測さえもの開発に拍車をかけた。

【０００３】

ゲノムデータの量が相当のレベルに達しているため、計算上の要件及び有意義な出力の生成の様式は難題となっている。たとえば、複数の腫瘍及び一致した正常な全ゲノムの配列が現在「癌ゲノムアトラス（Ｔｈｅ　Ｃａｎｃｅｒ　Ｇｅｎｏｍｅ　Ａｔｌａｓ）」（ＴＣＧＡ）などのプロジェクトから入手可能であり、関連性のある情報の抽出は困難である。これは、統計的に関連性のあるデータを得るための高いゲノム配列決定カバレッジ（たとえば３０倍を超えるもの）の必要性によって度合がさらに強まっている。圧縮形態でさえも、ゲノム情報はしばしば数百ギガバイトに達する場合があり、複数のそのような大きなデータセットを比較する解析は、ほとんどの場合で遅いかつ管理が困難であるが、しかし、第２の試料と比較して任意の所定の試料中で起こった多くのゲノム変化を発見するために絶対的に必要である。

【０００４】

乳癌は臨床的及びゲノム的に異種であり、いくつかの病理学的及び分子的に明確なサブタイプから構成される。慣用及び標的化治療学に対する患者の応答はサブタイプ間で異なり、マーカーによって導かれる治療戦略の開発の動機となっている。乳癌細胞系のコレクションは腫瘍中に見つかる分子サブタイプ及び経路の多くを反映しており、これは、候補治療化合物を用いた細胞系の処置が分子サブタイプ、経路、及び薬物応答の間の関連性の同定を導くことができることを示唆している。７７個の治療化合物の試験では、ほぼすべての薬物がこれらの細胞系にわたって示差的な応答を示し、約半数がサブタイプ、経路及び／又はゲノム異常に特異的な応答を示す。これらの観察は、臨床的薬物展開及び薬物を有効に組み合わせる取り組みに情報を与え得る、応答及び耐性の機構を示唆している。

【発明の概要】

【発明が解決しようとする課題】

【０００５】

現在、疾患及び障害の特徴づけ、診断、処置、及び結果の決定に使用することができる方法を提供する必要性が存在する。

【課題を解決するための手段】

【０００６】

本発明者らは、複数の膨大なファイルを処理する必要のない様式、かつゲノム異常に関して情報密度が比較的低い、同様に膨大な出力ファイルの生成を回避する様式で、有意義な出力の迅速な生成を可能にする、比較ゲノム解析の様々なシステム及び方法を発見した。

【０００７】

本発明の主題の一態様では、差分遺伝子配列オブジェクトを誘導する方法は、（ａ）第１の組織を表す第１の遺伝子配列ストリング及び（ｂ）第２の組織を表す第２の遺伝子配列ストリングを記憶する遺伝子データベースへのアクセスを提供するステップであって、第１及び第２の配列ストリングが複数の対応するサブストリングを有するステップを含む。別のステップでは、遺伝子データベースと連結させた配列解析エンジンへのアクセスを提供し、さらに別のステップでは、配列解析エンジンは、複数の対応するサブストリングのうちの少なくとも１つの既知の位置を使用して第１及び第２の配列ストリングを増分同期させることによって、局所アラインメントを生成する。さらなるステップでは、配列解析エンジンは、局所アラインメントを使用して、局所アラインメント内の第１と第２の配列ストリングとの間で局所差分ストリングを生成し、配列解析エンジンは、局所差分ストリングを使用して、差分配列データベース中の差分遺伝子配列オブジェクトを更新する。

最も好ましくは、第１及び第２の遺伝子配列ストリングは、第１及び第２の組織のゲノム、トランスクリプトーム、もしくはプロテオームの少なくとも１０％、より典型的には

少なくとも５０％、又はそれぞれ第１及び第２の組織のゲノム、トランスクリプトーム、もしくはプロテオームの実質的に全体でさえを表す。さらに、第１及び第２の組織は同じ生物学的実体を起源とする（たとえば、患者、健康な個体、細胞系、幹細胞、実験動物モデル、組換え細菌細胞、又はウイルス）ことも理解されたい。他方では、第１の組織は健康な組織であり得る一方で、第２の組織は患部組織（たとえば腫瘍組織）であり得る。さらなる企図される態様では、対応するサブストリングはホモ接合性又はヘテロ接合性の対立遺伝子を含む。

【０００８】

また、同期ステップは、複数のサブストリングのうちの少なくとも１つをアラインさせることを含み、アラインメントは、第１のストリング内の先験的に既知の位置に基づくことが一般的に好ましい。その代わりに又はそれに加えて、同期ステップは、複数のサブストリングのうちの少なくとも１つの既知の位置を含めた既知の参照ストリング（たとえばコンセンサス配列）に基づいて、複数のサブストリングのうちの少なくとも１つをアラインさせることを含み、かつ／又は、同期ステップは、複数のサブストリングのうちの少なくとも１つの長さよりも短い長さのウィンドウ内で複数のサブストリングのうちの少なくとも１つをアラインさせることを含む。所望される場合は、企図される方法は、第１の配列ストリングの全長にわたって第１及び第２の配列ストリングを繰り返し増分同期させるステップをさらに含み得る。

【０００９】

特に好ましい方法では、差分遺伝子配列オブジェクトは、少なくとも１つの染色体の複数の局所差分ストリングを表す、第１の組織の実質的に全ゲノムの複数の局所差分ストリングを表す、及び／又は差分遺伝子配列オブジェクトを記述するメタデータを含む属性を含む。特に好ましい属性は第１及び第２の組織のうちの少なくとも１つの状態である。たとえば、状態には、第１及び第２の組織のうちの少なくとも１つの生理的状態（たとえば、新生物成長、アポトーシス、分化状態、組織年齢、及び処置への応答性）、又は遺伝子状態（たとえば、倍数性、遺伝子のコピー数、反復のコピー数、反転、欠失、ウイルス遺伝子の挿入、体細胞突然変異、生殖系列突然変異、構造的再編成、転位、及びヘテロ接合性の喪失）が含まれ得る。また、適切な状態には、組織内のシグナル伝達経路（たとえば、成長因子シグナル伝達経路、転写因子シグナル伝達経路、アポトーシス経路、細胞周期経路、及びホルモン応答経路）に関する経路モデル情報も含まれる。遺伝子配列オブジェクトはファイルを含み、最も好ましくは、ファイルは標準化フォーマット（たとえばＳＡＭ／ＢＡＭフォーマット）に従うことがさらに企図される。

【００１０】

本発明の主題の別の態様では、本発明者らは、医療サービスを提供する方法も企図する。そのような方法では、医療記録記憶装置と情報的に連結させた解析エンジンへのアクセスを提供し、記憶装置は患者の差分遺伝子配列オブジェクトを記憶する。別のステップでは、解析エンジンは、患者の差分遺伝子配列オブジェクト中の局所差分ストリング又は複数の局所差分ストリングの一群の存在を使用して患者に特異的なデータセットを生成し、また、解析エンジンは、患者に特異的なデータセットに基づいて患者に特異的な指示も生成する。

特に好ましい方法では、医療記録記憶装置はスマートカードとして構成されており、患者によって保有され、かつ／又は医療提供者によって遠隔アクセス可能である。

最も典型的には、患者の差分遺伝子配列オブジェクトは、少なくとも２つの染色体、又は患者の実質的に全ゲノムでさえもの複数の局所差分ストリングを含む。その代わりに、又はそれに加えて、患者の差分遺伝子配列オブジェクトは、少なくとも２つの組織型を表す複数の局所差分ストリング、又は同じ組織において時間間隔を空けた少なくとも２つの結果（たとえば、同じ組織の時間間隔を空けた結果は、処置開始の前及び後から得られる）も含み得る。患者に特異的な指示は、診断、予後診断、処置結果の予測、処置戦略の推奨、及び／又は処方であることがさらに一般的に好ましい。

【００１１】

本発明の主題のさらに別の態様では、本発明者らは、集団を解析する方法であって、集団の医療記録データベース中の複数の差分遺伝子配列オブジェクトを獲得及び記憶するステップであって、記録データベースが解析エンジンと情報的に連結されているステップを含む方法を企図する。別のステップでは、解析エンジンは、複数の差分遺伝子配列オブジェクト内の複数の局所差分ストリングの一群を同定して一群記録を生成し、解析エンジンは一群記録を使用して集団解析記録を生成する。

そのような方法では、集団が、複数の血縁者ならびに／又は少なくとも１つの共通の特長（たとえば、病原体への曝露、有害物質への曝露、既往歴、処置歴、処置の成功、性別、種、及び／もしくは年齢）を共有することによって特徴づけられた複数のメンバーを含むことが一般的に企図される。また、適切な集団は、地理的位置、民族性、及び／又は職業を共有することによって特徴づけられた複数のメンバーも含み得る。したがって、集団解析記録は父系性又は母系性の確認を含むことを認識されたい。

【００１２】

本明細書中に提示する方法には、個々の患者の一群記録を集団解析記録と比較するステップがさらに含まれていてもよく、これがひいては患者に特異的な記録（たとえば、危険性評価又は患者が特定集団に属するという確認を示す）を作成し得ることがさらに企図される。また、患者に特異的な記録は診断、予後診断、処置結果の予測、処方、及び／又は処置戦略の推奨も含み得る。

【００１３】

したがって、本発明者らは、一ステップにおいて参照差分遺伝子配列オブジェクトを解析エンジンと情報的に連結させた医療記録データベースに記憶させる、人の差分遺伝子配列オブジェクトを解析する方法も企図する。その後、解析エンジンは、人の差分遺伝子配列オブジェクト中の複数の局所差分ストリングと参照差分遺伝子配列オブジェクト中の複数の局所差分ストリングとの間の偏差を計算して偏差記録を生成し、その後、解析エンジンは、偏差記録を使用して人に特異的な偏差プロフィールを生成する。

【００１４】

そのような方法では、参照差分遺伝子配列オブジェクトは、人の複数の局所差分ストリングから、又は人の複数の局所差分ストリングから計算することが好ましい。

【００１５】

本明細書中に提示する方法において、患者又は人は、状態、特に疾患又は障害を診断された患者又は人であり得ることを認識されたい。たとえば、企図される状態には、後天性免疫不全症候群（ＡＩＤＳ）、アジソン病、成人呼吸窮迫症候群、アレルギー、強直性脊椎炎、アミロイド症、貧血、喘息、アテローム性動脈硬化症、自己免疫性溶血性貧血、自己免疫性甲状腺炎、良性前立腺肥大、気管支炎、チェディアック－東症候群、胆嚢炎、クローン病、アトピー性皮膚炎、皮膚筋炎（ｄｅｒｍｎａｔｏｍｙｏｓｉｔｉｓ）、真性糖尿病、気腫、胎児赤芽球症、結節性紅斑、萎縮性胃炎、糸球体腎炎、グッドパスチャー症候群、痛風、慢性肉芽腫性疾患、グレーブス病、橋本甲状腺炎、過好酸球増加症、過敏性腸症候群、多発性硬化症、重症筋無力症、心筋又は心膜の炎症、骨関節炎、骨粗鬆症、膵炎、多嚢胞性卵巣症候群、多発性筋炎、乾癬、ライター症候群、関節リウマチ、強皮症、重症複合型免疫不全症（ＳＣＩＤ）、シェーグレン症候群、全身性アナフィラキシー、全身性エリテマトーデス、全身性硬化症、血小板減少性紫斑病、潰瘍性大腸炎、ブドウ膜炎、ウェルナー症候群、癌、血液透析、及び体外循環の合併症、ウイルス、細菌、真菌、寄生生物、原虫、及び蠕虫の感染症；腺癌、白血病、リンパ腫、黒色腫、骨髄腫、肉腫、奇形癌、特に、副腎、膀胱、骨、骨髄、脳、乳房、子宮頸部、胆嚢、神経節、胃腸管、心臓、腎臓、肝臓、肺、筋肉、卵巣、膵臓、副甲状腺、陰茎、前立腺、唾液腺、皮膚、脾臓、精巣、胸腺、甲状腺、及び子宮の癌、アカシジア、アルツハイマー病、健忘症、筋萎縮性側索硬化症（ＡＬＳ）、運動失調、双極性障害、緊張病、脳性麻痺、脳血管疾患、クロイツフェルト－ヤコブ病、認知症、鬱病、ダウン症候群、遅発性ジスキネジア、ジストニア、癲癇、ハンチントン病、多発性硬化症、筋ジストロフィー、神経痛、神経線維腫症、神経障害、パーキンソン病、ピック病、網膜色素変性症、統合失調症、季節性情動障害、老

人性認知症、脳卒中、トゥレット症候群、ならびに特に脳の腺癌、黒色腫、及び奇形癌を含めた癌が含まれる。

【００１６】

さらなる企図される状態には、腺癌、白血病、リンパ腫、黒色腫、骨髄腫、肉腫、奇形癌などの癌、特に、副腎、膀胱、骨、骨髄、脳、乳房、子宮頸部、胆嚢、神経節、胃腸管、心臓、腎臓、肝臓、肺、筋肉、卵巣、膵臓、副甲状腺、陰茎、前立腺、唾液腺、皮膚、脾臓、精巣、胸腺、甲状腺、及び子宮の癌；後天性免疫不全症候群（ＡＩＤＳ）、アジソン病、成人呼吸窮迫症候群、アレルギー、強直性脊椎炎、アミロイド症、貧血、喘息、アテローム性動脈硬化症、自己免疫性溶血性貧血、自己免疫性甲状腺炎、気管支炎、胆嚢炎、接触皮膚炎、クローン病、アトピー性皮膚炎、皮膚筋炎、真性糖尿病、気腫、リンパ球毒素を伴う偶発性リンパ球減少症、胎児赤芽球症、結節性紅斑、萎縮性胃炎、糸球体腎炎、グッドパスチャー症候群、痛風、グレーブス病、橋本甲状腺炎、過好酸球増加症、過敏性腸症候群、多発性硬化症、重症筋無力症、心筋又は心膜の炎症、骨関節炎、骨粗鬆症、膵炎、多発性筋炎、乾癬、ライター症候群、関節リウマチ、強皮症、シェーグレン症候群、全身性アナフィラキシー、全身性エリテマトーデス、全身性硬化症、血小板減少性紫斑病、潰瘍性大腸炎、ブドウ膜炎、ウェルナー症候群、癌、血液透析、及び体外循環の合併症、ウイルス、細菌、真菌、寄生生物、原虫、及び蠕虫の感染症、外傷、ブルトン型Ｘ連鎖無ガンマグロブリン血症（agammaglobinemia）、後天性免疫グロブリン血症(common variable immunodeficiency)（ＣＶＩ）、ディジョージ症候群（胸腺形成不全）、胸腺異形成、ＩｇＡ単独欠損症、重症複合型免疫不全症（ＳＣＩＤ）、血小板減少症及び湿疹を伴う免疫不全（ウィスコットーアルドリッチ症候群）、チェディアックー東症候群、慢性肉芽腫性疾患、遺伝性血管神経性浮腫、ならびにクッシング病関連の免疫不全などの免疫障害；腎尿細管性アシドーシス、貧血、クッシング症候群、軟骨形成不全性小人症、デュシェンヌ及びベッカー型筋ジストロフィー、癲癇、性腺形成不全、ＷＡＧＲ症候群（ウィルムス腫瘍、無虹彩症、泌尿生殖器奇形、及び精神遅滞）、スミスーマゲニス症候群、骨髄異形成症候群、遺伝性粘膜上皮異形成、遺伝性角皮症、シャルコーーマリーートゥース病及び神経線維腫症などの遺伝性神経障害、甲状腺機能低下症、水頭症、シデナム（Sydenham）舞踏病及び脳性麻痺などの発作性疾患、二分脊椎、無脳症、頭蓋脊椎披裂、先天性緑内障、白内障、感音性難聴、ならびに対象の任意の組織、器官、又は系、たとえば、脳、副腎、腎臓、骨格又は生殖器系に関与する細胞の成長と分化、胚形成、及び形態形成に関連する任意の障害などの発達障害も含まれる。

【００１７】

さらなる企図される状態には、性腺機能低下症、シーハン症候群、尿崩症、カルマン病、ハンドーシュラーークリスチャン病、レッテラーーシーベ病、サルコイドーシス、トルコ鞍空洞症候群、及び小人症を含めた、下垂体機能低下症に関連する障害などの内分泌障害；末端肥大症、巨人症、及び抗利尿ホルモン（ＡＤＨ）不適合分泌症候群（ＳＩＡＤＨ）を含めた下垂体機能亢進症；甲状腺種、粘液水腫、細菌感染症に関連する急性甲状腺炎、ウイルス感染症に関連する亜急性甲状腺炎、自己免疫性甲状腺炎（橋本病）、及びクレチン症を含めた、甲状腺機能低下症に関連する障害；甲状腺中毒症及びその様々な形態、グレーブス病、前脛骨粘液水腫、中毒性多結節性甲状腺種、甲状腺癌、及びプランマー病を含めた、甲状腺機能亢進症に関連する障害；コーン(Conn)病（慢性高カルシウム血症）を含めた副甲状腺機能亢進に関連する障害；アレルギー、喘息、急性及び慢性の炎症性肺疾患、ＡＲＤＳ、気腫、肺鬱血及び肺浮腫、ＣＯＰＤ、間質性肺疾患、肺癌などの呼吸器疾患；腺癌、白血病、リンパ腫、黒色腫、骨髄腫、肉腫、奇形癌などの癌、特に、副腎、膀胱、骨、骨髄、脳、乳房、子宮頸部、胆嚢、神経節、胃腸管、心臓、腎臓、肝臓、肺、筋肉、卵巣、膵臓、副甲状腺、陰茎、前立腺、唾液腺、皮膚、脾臓、精巣、胸腺、甲状腺、及び子宮の癌；ならびに後天性免疫不全症候群（ＡＩＤＳ）、アジソン病、成人呼吸窮迫症候群、アレルギー、強直性脊椎炎、アミロイド症、貧血、喘息、アテローム性動脈硬化症、自己免疫性溶血性貧血、自己免疫性甲状腺炎、気管支炎、胆嚢炎、接触皮膚炎、クローン病、アトピー性皮膚炎、皮膚筋炎、真性糖尿病、気腫、リンパ球毒素を伴う偶発性

リンパ球減少症、胎児赤芽球症、結節性紅斑、萎縮性胃炎、糸球体腎炎、グッドパスチャー症候群、痛風、グレーブス病、橋本甲状腺炎、過好酸球増加症、過敏性腸症候群、多発性硬化症、重症筋無力症、心筋又は心膜の炎症、骨関節炎、骨粗鬆症、膵炎、多発性筋炎、乾癬、ライター症候群、関節リウマチ、強皮症、シェーグレン症候群、全身性アナフィラキシー、全身性エリテマトーデス、全身性硬化症、血小板減少性紫斑病、潰瘍性大腸炎、ブドウ膜炎、ウェルナー症候群、癌、血液透析、及び体外循環の合併症、ウイルス、細菌、真菌、寄生生物、原虫、及び蠕虫の感染症、外傷などの免疫障害が含まれる。

【００１８】

また、本発明は、個体の危険性、特に、たとえば、それだけには限定されないが、疾患、障害、又は状態に対する個体の素因の危険性、個体の職場、住居、学校などでの危険性、毒素、発癌物質、突然変異原などへの個体の曝露の危険性、及び個体の食事習慣の危険性を決定するために使用し得るデータベースを生成する方法も提供する。さらに、本発明は、特定の個体、動物、植物、又は微生物を同定するために使用し得る方法を提供する。

【００１９】

一実施形態では、本発明は、差分遺伝子配列オブジェクトを誘導する方法であって、（ａ）第１の組織を表す第１の遺伝子配列ストリング及び（ｂ）第２の組織を表す第２の遺伝子配列ストリングを記憶する遺伝子データベースへのアクセスを提供するステップであって、第１及び第２の配列ストリングが複数の対応するサブストリングを有するステップと、遺伝子データベースと連結させた配列解析エンジンへのアクセスを提供するステップと、複数の対応するサブストリングのうちの少なくとも１つの既知の位置を使用して第１及び第２の配列ストリングを増分同期させることによって、配列解析エンジンを使用して局所アラインメントを生成するステップと、配列解析エンジンによって、局所アラインメントを使用して、局所アラインメント内の第１と第２の配列ストリングとの間で局所差分ストリングを生成するステップと、配列解析エンジンによって、局所差分ストリングを使用して、差分配列データベース中の差分遺伝子配列オブジェクトを更新するステップとを含む方法を提供する。好ましい実施形態では、第１及び第２の遺伝子配列ストリングは、それぞれ第１及び第２の組織のゲノム、トランスクリプトーム、又はプロテオームの少なくとも１０％を表す。好ましい代替の実施形態では、第１及び第２の遺伝子配列ストリングは、それぞれ第１及び第２の組織のゲノム、トランスクリプトーム、又はプロテオームの少なくとも５０％を表す。別の好ましい代替の実施形態では、第１及び第２の遺伝子配列ストリングは、それぞれ第１及び第２の組織のゲノム、トランスクリプトーム、又はプロテオームの実質的に全体を表す。別の好ましい実施形態では、対応するサブストリングはホモ接合性対立遺伝子を含む。好ましい代替の実施形態では、対応するサブストリングはヘテロ接合性対立遺伝子を含む。別のより好ましい実施形態では、遺伝子配列オブジェクトはファイルを含む。さらにより好ましい実施形態では、ファイルは標準化フォーマットに従う。最も好ましい実施形態では、ファイルはＳＡＭ／ＢＡＭフォーマットに従う。

【００２０】

好ましい実施形態では、同期ステップは、第１のストリング内の先験的に既知の位置に基づいて複数のサブストリングのうちの少なくとも１つをアラインさせることを含む。好ましい代替の実施形態では、同期ステップは、複数のサブストリングのうちの少なくとも１つの既知の位置を含む既知の参照ストリングに基づいて、複数のサブストリングのうちの少なくとも１つをアラインさせることを含む。より好ましい実施形態では、既知の参照ストリングはコンセンサス配列である。

【００２１】

別の好ましい実施形態では、同期ステップは、複数のサブストリングのうちの少なくとも１つの長さよりも短い長さのウィンドウ内で複数のサブストリングのうちの少なくとも１つをアラインさせることを含む。

別の好ましい実施形態では、差分遺伝子配列オブジェクトは、少なくとも１つの染色体の複数の局所差分ストリングを表す。

別の好ましい実施形態では、差分遺伝子配列オブジェクトは、第１の組織の実質的に全

ゲノムの複数の局所差分ストリングを表す。

【００２２】

さらに他の好ましい実施形態では、差分遺伝子配列オブジェクトは、差分遺伝子配列オブジェクトを記述するメタデータを含む属性を含む。より好ましい実施形態では、属性は、第１及び第２の組織のうちの少なくとも１つの状態を含む。さらにより好ましい実施形態では、状態は、第１及び第２の組織のうちの少なくとも１つの生理的状態を含む。最も好ましい実施形態では、生理的状態は、新生物成長、アポトーシス、分化状態、組織年齢、及び処置への応答性からなる群から選択される状態を含む。

【００２３】

より好ましい代替の実施形態では、状態は遺伝子状態を含む。最も好ましい実施形態では、遺伝子状態は、少なくとも１つの倍数性、遺伝子のコピー数、反復のコピー数、反転、欠失、ウイルス遺伝子の挿入、体細胞突然変異、生殖系列突然変異、構造的再編成、転位、及びヘテロ接合性の喪失からなる群から選択される状態を含む。

より好ましい代替の実施形態では、状態は、組織内のシグナル伝達経路に関連する経路モデル情報を含む。最も好ましい実施形態では、シグナル伝達経路は、成長因子シグナル伝達経路、転写因子シグナル伝達経路、アポトーシス経路、細胞周期経路、及びホルモン応答経路からなる群から選択される。

【００２４】

代替の実施形態では、第１及び第２の組織は同じ生物学的実体を起源とし、生物学的実体は、患者、健康な個体、細胞系、幹細胞、実験動物モデル、組換え細菌細胞、及びウイルスからなる群から選択される。代替の実施形態では、第１の組織は健康な組織であり、第２の組織は患部組織である。より好ましい実施形態では、患部組織は腫瘍組織を含む。

また、本発明は、第１の配列ストリングの全長にわたって第１及び第２の配列ストリングを繰り返し増分同期させるステップをさらに含む、本明細書中に開示した方法も提供する。

【００２５】

また、本発明は、医療サービスを提供する方法であって、医療記録記憶装置と情報的に連結させた解析エンジンへのアクセスを提供するステップであって、記憶装置が患者の差分遺伝子配列オブジェクトを記憶するステップと、解析エンジンによって、患者の差分遺伝子配列オブジェクト中の局所差分ストリング又は複数の局所差分ストリングの一群の存在を使用して患者に特異的なデータセットを生成するステップと、解析エンジンによって、患者に特異的なデータセットに基づいて患者に特異的な指示を生成するステップとを含む方法も提供する。好ましい実施形態では、医療記録記憶装置はスマートカードとして構成されており、患者によって保有される。別の好ましい実施形態では、医療記録記憶装置は医療提供者によって遠隔アクセス可能である。さらに他の好ましい実施形態では、患者の差分遺伝子配列オブジェクトは、少なくとも２つの染色体の複数の局所差分ストリングを含む。さらなる好ましい実施形態では、患者の差分遺伝子配列オブジェクトは、患者の実質的に全ゲノムの複数の局所差分ストリングを含む。別の好ましい実施形態では、患者の差分遺伝子配列オブジェクトは、少なくとも２つの組織型を表す複数の局所差分ストリング、又は同じ組織において時間間隔を空けた少なくとも２つの結果を含む。より好ましい実施形態では、同じ組織において時間間隔を空けた少なくとも２つの結果は、処置開始の前及び後から得られる。最も好ましい実施形態では、同じ組織において時間間隔を空けた少なくとも２つの結果は、処置開始の前及び後から得られる。

【００２６】

別の好ましい代替の実施形態では、本明細書中に開示した患者に特異的な指示は、診断、予後診断、処置結果の予測、処置戦略の推奨、及び処方からなる群から選択される。

また、本発明は、集団を解析する方法であって、集団の医療記録データベース中の複数の差分遺伝子配列オブジェクトを獲得及び記憶するステップであって、記録データベースが解析エンジンと情報的に連結されているステップと、解析エンジンによって、複数の差分遺伝子配列オブジェクト内の複数の局所差分ストリングの一群を同定して一群記録を生

成するステップと、解析エンジンによって、一群記録を使用して集団解析記録を生成するステップとを含む方法も提供する。好ましい実施形態では、集団は複数の血縁者を含む。好ましい代替の実施形態では、集団は、病原体への曝露、有害物質への曝露、既往歴、処置歴、処置の成功、性別、種、及び年齢からなる群から選択される少なくとも１つの共通の特長を共有することによって特徴づけられた複数のメンバーを含む。別の好ましい代替の実施形態では、集団は、地理的位置、民族性、及び職業からなる群から選択される少なくとも１つの共通の特長を共有することによって特徴づけられた複数のメンバーを含む。さらに好ましい代替の実施形態では、集団解析記録は父系性又は母系性の確認を含む。

【００２７】

代替の実施形態では、本明細書中に開示した方法は、個々の患者の一群記録を集団解析記録と比較するステップをさらに含む。好ましい実施形態では、個々の患者の一群記録を集団解析記録と比較するステップは、患者に特異的な記録を作成する。より好ましい実施形態では、患者に特異的な記録は、危険性評価又は患者が特定集団に属するという確認を含む。より好ましい代替の実施形態では、患者に特異的な記録は、診断、予後診断、処置結果の予測、処置戦略の推奨、及び処方を含む。

【００２８】

本発明はさらに、人の差分遺伝子配列オブジェクトを解析する方法であって、参照差分遺伝子配列オブジェクトを解析エンジンと情報的に連結させた医療記録データベースに記憶させるステップと、解析エンジンによって、人の差分遺伝子配列オブジェクト中の複数の局所差分ストリングと参照差分遺伝子配列オブジェクト中の複数の局所差分ストリングとの間の偏差を計算して偏差記録を生成するステップと、解析エンジンによって、偏差記録を使用して人に特異的な偏差プロフィールを生成するステップとを含む方法を提供する。好ましい実施形態では、参照差分遺伝子配列オブジェクトは、人の複数の局所差分ストリングから計算する。別の好ましい実施形態では、参照差分遺伝子配列オブジェクトは、人の複数の局所差分ストリングから計算する。

【００２９】

本明細書中に開示した様々な方法に関して、好ましい実施形態では、患者又は人は、状態を診断された患者又は人からなる群から選択され、状態は、疾患及び障害からなる群から選択される。より好ましい実施形態では、状態は、後天性免疫不全症候群（ＡＩＤＳ）、アジソン病、成人呼吸窮迫症候群、アレルギー、強直性脊椎炎、アミロイド症、貧血、喘息、アテローム性動脈硬化症、自己免疫性溶血性貧血、自己免疫性甲状腺炎、良性前立腺肥大、気管支炎、チェディアック－東症候群、胆嚢炎、クローン病、アトピー性皮膚炎、皮膚筋炎、真性糖尿病、気腫、胎児赤芽球症、結節性紅斑、萎縮性胃炎、糸球体腎炎、グッドパスチャー症候群、痛風、慢性肉芽腫性疾患、グレーブス病、橋本甲状腺炎、過好酸球増加症、過敏性腸症候群、多発性硬化症、重症筋無力症、心筋又は心膜の炎症、骨関節炎、骨粗鬆症、膵炎、多嚢胞性卵巣症候群、多発性筋炎、乾癬、ライター症候群、関節リウマチ、強皮症、重症複合型免疫不全症（ＳＣＩＤ）、シェーグレン症候群、全身性アナフィラキシー、全身性エリテマトーデス、全身性硬化症、血小板減少性紫斑病、潰瘍性大腸炎、ブドウ膜炎、ウェルナー症候群、癌、血液透析、及び体外循環の合併症、ウイルス、細菌、真菌、寄生生物、原虫、及び蠕虫の感染症；腺癌、白血病、リンパ腫、黒色腫、骨髄腫、肉腫、奇形癌、特に、副腎、膀胱、骨、骨髄、脳、乳房、子宮頸部、胆嚢、神経節、胃腸管、心臓、腎臓、肝臓、肺、筋肉、卵巣、膵臓、副甲状腺、陰茎、前立腺、唾液腺、皮膚、脾臓、精巣、胸腺、甲状腺、及び子宮の癌、アカシジア、アルツハイマー病、健忘症、筋萎縮性側索硬化症（ＡＬＳ）、運動失調、双極性障害、緊張病、脳性麻痺、脳血管疾患、クロイツフェルト－ヤコブ病、認知症、鬱病、ダウン症候群、遅発性ジスキネジア、ジストニア、癲癇、ハンチントン病、多発性硬化症、筋ジストロフィー、神経痛、神経線維腫症、神経障害、パーキンソン病、ピック病、網膜色素変性症、統合失調症、季節性情動障害、老人性認知症、脳卒中、トゥレット症候群、ならびに特に脳の腺癌、黒色腫、及び奇形癌を含めた癌からなる群から選択される。

【００３０】

別の好ましい実施形態では、状態は、腺癌、白血病、リンパ腫、黒色腫、骨髄腫、肉腫、奇形癌などの癌、特に、副腎、膀胱、骨、骨髄、脳、乳房、子宮頸部、胆嚢、神経節、胃腸管、心臓、腎臓、肝臓、肺、筋肉、卵巣、膵臓、副甲状腺、陰茎、前立腺、唾液腺、皮膚、脾臓、精巣、胸腺、甲状腺、及び子宮の癌；後天性免疫不全症候群（ＡＩＤＳ）、アジソン病、成人呼吸窮迫症候群、アレルギー、強直性脊椎炎、アミロイド症、貧血、喘息、アテローム性動脈硬化症、自己免疫性溶血性貧血、自己免疫性甲状腺炎、気管支炎、胆嚢炎、接触皮膚炎、クローン病、アトピー性皮膚炎、皮膚筋炎、真性糖尿病、気腫、リンパ球毒素を伴う偶発性リンパ球減少症、胎児赤芽球症、結節性紅斑、萎縮性胃炎、糸球体腎炎、グッドパスチャー症候群、痛風、グレーブス病、橋本甲状腺炎、過好酸球増加症、過敏性腸症候群、多発性硬化症、重症筋無力症、心筋又は心膜の炎症、骨関節炎、骨粗鬆症、膵炎、多発性筋炎、乾癬、ライター症候群、関節リウマチ、強皮症、シェーグレン症候群、全身性アナフィラキシー、全身性エリテマトーデス、全身性硬化症、血小板減少性紫斑病、潰瘍性大腸炎、ブドウ膜炎、ウェルナー症候群、癌、血液透析、及び体外循環の合併症、ウイルス、細菌、真菌、寄生生物、原虫、及び蠕虫の感染症、外傷、ブルトン型Ｘ連鎖無ガンマグロブリン血症、後天性免疫グロブリン血症（ＣＶＩ）、ディジョージ症候群（胸腺形成不全）、胸腺異形成、ＩｇＡ単独欠損症、重症複合型免疫不全症（ＳＣＩＤ）、血小板減少症及び湿疹を伴う免疫不全（ウィスコットーアルドリッチ症候群）、チェディアックー東症候群、慢性肉芽腫性疾患、遺伝性血管神経性浮腫、ならびにクッシング病関連の免疫不全などの免疫障害；腎尿細管性アシドーシス、貧血、クッシング症候群、軟骨形成不全性小人症、デュシェンヌ及びベッカー型筋ジストロフィー、癲癇、性腺形成不全、ＷＡＧＲ症候群（ウィルムス腫瘍、無虹彩症、泌尿生殖器奇形、及び精神遅滞）、スミスーマゲニス症候群、骨髄異形成症候群、遺伝性粘膜上皮異形成、遺伝性角皮症、シャルコーーマリーートゥース病及び神経線維腫症などの遺伝性神経障害、甲状腺機能低下症、水頭症、シデナム舞踏病及び脳性麻痺などの発作性疾患、二分脊椎、無脳症、頭蓋脊椎披裂、先天性緑内障、白内障、感音性難聴、ならびに対象の任意の組織、器官、又は系、たとえば、脳、副腎、腎臓、骨格又は生殖器系に関与する細胞の成長と分化、胚形成、及び形態形成に関連する任意の障害などの発達障害からなる群から選択される。

【００３１】

さらなる好ましい代替の実施形態では、状態は、性腺機能低下症、シーハン症候群、尿崩症、カルマン病、ハンドーシュラーークリスチャン病、レッテラーーシーベ病、サルコイドーシス、トルコ鞍空洞症候群、及び小人症を含めた、下垂体機能低下症に関連する障害などの内分泌障害；末端肥大症、巨人症、及び抗利尿ホルモン（ＡＤＨ）不適合分泌症候群（ＳＩＡＤＨ）を含めた下垂体機能亢進症；甲状腺種、粘液水腫、細菌感染症に関連する急性甲状腺炎、ウイルス感染症に関連する亜急性甲状腺炎、自己免疫性甲状腺炎（橋本病）、及びクレチン症を含めた、甲状腺機能低下症に関連する障害；甲状腺中毒症及びその様々な形態、グレーブス病、前脛骨粘液水腫、中毒性多結節性甲状腺種、甲状腺癌、及びプランマー病を含めた、甲状腺機能亢進症に関連する障害；コーン病（慢性高カルシウム血症）を含めた副甲状腺機能亢進に関連する障害；アレルギー、喘息、急性及び慢性の炎症性肺疾患、ＡＲＤＳ、気腫、肺鬱血及び肺浮腫、ＣＯＰＤ、間質性肺疾患、肺癌などの呼吸器疾患；腺癌、白血病、リンパ腫、黒色腫、骨髄腫、肉腫、奇形癌などの癌、特に、副腎、膀胱、骨、骨髄、脳、乳房、子宮頸部、胆嚢、神経節、胃腸管、心臓、腎臓、肝臓、肺、筋肉、卵巣、膵臓、副甲状腺、陰茎、前立腺、唾液腺、皮膚、脾臓、精巣、胸腺、甲状腺、及び子宮の癌；ならびに後天性免疫不全症候群（ＡＩＤＳ）、アジソン病、成人呼吸窮迫症候群、アレルギー、強直性脊椎炎、アミロイド症、貧血、喘息、アテローム性動脈硬化症、自己免疫性溶血性貧血、自己免疫性甲状腺炎、気管支炎、胆嚢炎、接触皮膚炎、クローン病、アトピー性皮膚炎、皮膚筋炎、真性糖尿病、気腫、リンパ球毒素を伴う偶発性リンパ球減少症、胎児赤芽球症、結節性紅斑、萎縮性胃炎、糸球体腎炎、グッドパスチャー症候群、痛風、グレーブス病、橋本甲状腺炎、過好酸球増加症、過敏性腸症候群、多発性硬化症、重症筋無力症、心筋又は心膜の炎症、骨関節炎、骨粗鬆症、膵炎、多発性筋炎、乾癬、ライター症候群、関節リウマチ、強皮症、シェーグレン症候群、全身

性アナフィラキシー、全身性エリテマトーデス、全身性硬化症、血小板減少性紫斑病、潰瘍性大腸炎、ブドウ膜炎、ウェルナー症候群、癌、血液透析、及び体外循環の合併症、ウイルス、細菌、真菌、寄生生物、原虫、及び蠕虫の感染症、外傷などの免疫障害からなる群から選択される。

【００３２】

本発明はさらに、差分遺伝子配列オブジェクトを誘導する方法であって、（ａ）第１の組織を表す第１の遺伝子配列ストリング及び（ｂ）第２の組織を表す第２の遺伝子配列ストリングを記憶する遺伝子データベースへのアクセスを提供するステップであって、第１及び第２の配列ストリングが複数の対応するサブストリングを有するステップと、遺伝子データベースと連結させた配列解析エンジンへのアクセスを提供するステップと、配列解析エンジンを使用して、複数の対応するサブストリングのうちの少なくとも１つの既知の位置を使用して第１及び第２の配列ストリングを増分同期させることによって、局所アラインメントを生成するステップと、配列解析エンジンによって、局所アラインメントを使用して、局所アラインメント内の第１と第２の配列ストリングとの間で局所差分ストリングを生成するステップと、配列解析エンジンによって、局所差分ストリングを使用して、差分配列データベース中の差分遺伝子配列オブジェクトを作成し、それによって差分配列オブジェクトを誘導するステップとを含む方法を提供する。

【００３３】

本発明はさらに、第１の遺伝子配列と第２の配列との間の臨床的に関連性のある相違を表す差分遺伝子配列オブジェクトを作成するための変換方法であって、（ｉ）（ａ）第１の組織を表す第１の遺伝子配列ストリング及び（ｂ）第２の組織を表す第２の遺伝子配列ストリングを記憶する遺伝子データベースへのアクセスを提供するステップであって、第１及び第２の配列ストリングが複数の対応するサブストリングを有するステップと、（ｉｉ）遺伝子データベースと連結させた配列解析エンジンへのアクセスを提供するステップと、（ｉｉｉ）配列解析エンジンを使用して、複数の対応するサブストリングのうちの少なくとも１つの既知の位置を使用して第１及び第２の配列ストリングを増分同期させることによって、局所アラインメントを生成するステップと、（ｉｖ）配列解析エンジンによって、局所アラインメントを使用して、局所アラインメント内の第１と第２の配列ストリングとの間で局所差分ストリングを生成するステップと、（ｖ）配列解析エンジンによって、局所差分ストリングを使用して、差分配列データベース中の差分遺伝子配列オブジェクトを作成し、それによって差分配列オブジェクトを誘導するステップであって、差分配列オブジェクトが客観的な情報をユーザに提供するステップとを含む方法を提供する。

好ましい実施形態では、客観的な情報は、遺伝的に関連性のある情報、代謝的に関連性のある情報、毒物学的に関連性のある情報、臨床的に意味のある情報、時間的に関連性のある情報、地理的に関連性のある情報、職業上の危険性に関連性のある情報、生活史に関連性のある情報などからなる群から選択される。

【００３４】

本発明の主題の様々な目的、特長、態様及び利点は、同様の数字は同様の構成要素を表す添付の図面と共に、以下の好ましい実施形態の詳細な説明から、より明らかとなるであろう。

【図面の簡単な説明】

【００３５】

【図１】「ＢａｍＢａｍ」データフローの模式図を例示する図である。

【図２】対立遺伝子特異的なコピー数の計算の全体像を例示する図である。

【図３】構造的変異コールの全体像を例示する図である。

【図４】構造的再編成が起こったゲノムの位置を同定するための例示的な方法を例示する図である。

【図５】例示的な腫瘍特異的ゲノムブラウザを例示する図である。

【図６】本発明の主題に従って差分遺伝子配列オブジェクトを生成するための例示的なコンピュータシステムの模式図である。

【図７】差分遺伝子配列オブジェクトを誘導する方法の模式図である。
【図８】患者に特異的な指示の形態で医療サービスを提供する方法の模式図である。
【図９】遺伝学における相違に関して集団を解析する方法の模式図である。
【図１０】人の差分遺伝子配列オブジェクトを解析する方法の模式図である。
【発明を実施するための形態】
【００３６】
本文書中に開示されている実施形態は例示的及び典型的なものであり、本発明を限定することを意図しない。本発明の特許請求の範囲の範囲から逸脱せずに、他の実施形態を利用することができ、また構造的変化を行うことができる。
本明細書中及び添付の特許請求の範囲中で使用する単数形「１つの（ａ）」、「１つの（ａｎ）」、及び「その（ｔｈｅ）」には、内容により明らかにそうでないと指示される場合以外は、複数形の言及が含まれる。したがって、たとえば、「１つの対立遺伝子（ａｎ　ａｌｌｅｌｅ）」への言及には複数のそのような対立遺伝子対が含まれ、「１つのクラスター（ａ　ｃｌｕｓｔｅｒ）」への言及は１つ又は複数のクラスター及びその均等物への言及であり、その他も同様である。
【００３７】
本明細書中で使用する用語「キュレーションした」とは、分子生物学的、生化学的、生理的、解剖学的、ゲノム、トランスクリプトミクス、プロテオミクス、メタボロミクス、ＡＤＭＥ、及び生物情報学的な技法などの当分野で周知の方法を使用した科学的及び／又は臨床的原理に従って試験、解析、及び同定した生体分子及び／又は非生体分子の組の間の関係性を意味する。関係性は、生化学的経路、遺伝経路、代謝経路、遺伝子調節経路、遺伝子転写経路、遺伝子翻訳経路、ｍｉＲＮＡによって調節される経路、擬似遺伝子によって調節される経路などの生化学的なものであり得る。本発明者らは、第１及び第２の組織試料（たとえば健康及び患部の組織）からのそれぞれのより大きな遺伝子配列ストリングの、複数の比較的小さなゲノム配列サブストリング（たとえば配列決定の実行からのショートリード）を得るシステム及び方法を開発した。その後、対応するサブストリングのうちの少なくとも１つの１つ又は複数の既知の位置を使用して遺伝子配列ストリングを増分同期して、局所アラインメントを生成する。その後、そのように生成された局所アラインメントを解析して（典型的には参照ゲノム配列を使用）、局所アラインメント内の第１と第２の配列ストリングとの間で局所差分ストリングを生成し、これはしたがって有意な差分情報（典型的には参照ゲノム配列に対する）を含有する。その後、局所差分ストリング、及び最も典型的には複数の局所差分ストリングを使用して、一部分又は全ゲノムでさえもの差分遺伝子配列オブジェクトを作成する。
【００３８】
したがって、２つの非常に大きなファイルを処理して別の非常に大きな中間ファイル（又は出力ファイルでさえ）を生成する代わりに、ゲノム全体にわたる解析を複数の顕著により小さな部分中で達成することができ、前記より小さな部分は、１つ又は複数のサブストリングのゲノム内の既知の位置を使用して参照ゲノムとアラインすることを認識されたい。別の角度から見ると、サブストリングの既知の位置及び参照ゲノム配列を使用した配列ストリングの増分同期によってアラインメントを行い、参照ゲノムに対して関連性のある変化のみを含む出力ファイルを生成することができる。したがって、処理速度は顕著に改善され、有意義な出力を生成するために必要なデータの量は劇的に減る。さらに、企図されるシステム及び方法は、とりわけ、ハプロタイプ決定／体細胞及び生殖系列の変異コール、ならびに対立遺伝子特異的なコピー数の決定をさらに可能にする。さらに、本明細書中に提示するシステム及び方法は、ＳＡＭ／ＢＡＭフォーマットの配列情報と共に使用するために適している。
【００３９】
たとえば、複数の配列決定断片（たとえば、ドナーの腫瘍試料及び同じドナーの対応する非腫瘍試料からのショートリード）を同じ参照ゲノムとアラインさせ、これを用いて試料から配列決定断片を組織化する。その後、ＢＡＭＢＡＭは同じ患者からの２つの配列決

定断片データセット（一方は腫瘍から、他方は対応する正常な「生殖系列」組織から）及び参照ゲノムを使用し、同じゲノム位置に重複するどちらのデータセット中のすべての配列（参照ゲノム及びサブストリング中のアノテーションに基づく）も同時に処理されるように、データセットを読み込む。これが、そのようなデータを処理するために最も効率的である一方で、それぞれのデータセットを単独で処理し、その後にのみ結果を統合する、連続様式では達成が困難又は不可能であろう複雑な解析を可能にする方法である。

【００４０】

したがって、ＢＡＭＢＡＭは２つのファイルから同時に増分的に読み込み、常にそれぞれのＢＡＭファイルと他のファイルとの同調を保ち、２つのファイル間の共通のゲノム位置のすべてに重複するゲノムリードをパイルアップさせることを認識されたい。それぞれのパイルアップの対について、ＢＡＭＢＡＭは一連の解析を実行した後、パイルアップを廃棄して次の共通のゲノム位置へと移動する。この様式で処理することによって、コンピュータのＲＡＭ使用量は劇的に減り、処理速度はファイルシステムが２つのファイルを読み込むことができる測度によって主に制限される。このことは、一台のコンピュータ又はコンピュータクラスタにわたって実行するために十分に柔軟である一方で、ＢＡＭＢＡＭが膨大な量のデータを素早く処理することを可能にする。これらのファイルをＢＡＭＢＡＭで処理することの別の重要な利点は、その出力がかなり最小限であることであり、典型的にはそれぞれのファイル中に見つかる重要な相違のみが含まれることである。これにより、本質的には患者の腫瘍及び生殖系列のゲノムの間の全ゲノム差分解析であるものが生成され、それぞれのファイルにすべてのゲノム情報を別々に記憶した場合にかかるよりもはるかに少ないディスク記憶量を要する。

【００４１】

以下の記述はコンピュータ／サーバベースの経路解析システムに引き寄せられているが、様々な代替の構成も適切とみなされており、個々に又は集合的に動作するサーバ、インタフェース、システム、データベース、エージェント、ピア、エンジン、コントローラ、又は他の種の計算装置を含めた様々な計算装置を用い得ることに注意されたい。計算装置は、有形の、一時的でないコンピュータ読取可能な記憶媒体（たとえば、ハードドライブ、ソリッドステートドライブ、ＲＡＭ、フラッシュ、ＲＯＭなど）上に記憶されたソフトウェア指示を実行するように構成されているプロセッサを備えることを理解されたい。好ましくは、ソフトウェア指示は、開示した機器に関して以下に記述した役割、責任、又は他の機能性を提供するように計算装置を構成する。特に好ましい実施形態では、様々なサーバ、システム、データベース、又はインタフェースは、標準化プロトコル又はアルゴリズム、場合によってはＨＴＴＰ、ＨＴＴＰＳ、ＡＥＳ、公開－私用鍵交換、ウェブサービスＡＰＩ、既知の金融取引プロトコル、又は他の電子情報交換方法に基づくものを使用してデータを交換する。データ交換は、好ましくは、パケット交換ネットワーク、インターネット、ＬＡＮ、ＷＡＮ、ＶＰＮ、又は他の種のパケット交換ネットワーク上で実施する。

【００４２】

さらに、以下の記述は、本発明の主題の多くの実施形態の例を提供する。それぞれの実施形態は本発明の要素の単一の組合せを表すが、本発明の主題には、開示した要素のすべての可能な組合せが含まれるとみなされる。したがって、第１の実施形態が要素Ａ、Ｂ、及びＣを含み、第２の実施形態が要素Ｂ及びＤを含む場合、本発明の主題には、明白に開示されていない場合でも、Ａ、Ｂ、Ｃ、又はＤの他の残りの組合せも含まれるとみなされる。

内容により明らかにそうでないと指示される場合以外は、本明細書中で使用する用語「連結されている」には、直接連結（互いに接触している２つの要素が互いに連結されている）及び間接連結（少なくとも１つの追加の要素が２つの要素の間に位置している）がどちらも含まれることを意図する。したがって、用語「連結されている」及び「連結させた」は同義に使用される。本文書内では、「連結させた」とは、「通信可能に連結させた」ことも意味すると解釈されるべきである。

【００４３】
高スループットデータは癌組織中の分子変化の包括的な見解を提供している。新しい技術は、ゲノムコピー数変動、遺伝子発現、ＤＮＡメチル化、ならびに腫瘍試料及び癌細胞系の後成学の状態の同時ゲノム全体アッセイを可能にする。
癌ゲノムアトラス（ＴＣＧＡ）、Ｓｔａｎｄ Ｕｐ Ｔｏ Ｃａｎｃｅｒ（ＳＵ２Ｃ）及び多数の他の研究などの研究が、様々な腫瘍のために近い将来に計画されている。現在のデータセットの解析により、患者間の遺伝子変化は異なる場合があるが、多くの場合に共通の経路を含むことが判明している。したがって、癌進行に関与している関連性のある経路を同定し、これらが様々な患者においてどのように変更されているかを検出することが重要である。

【００４４】
複数の完全に配列決定された腫瘍及び癌ゲノムアトラス（ＴＣＧＡ）などのプロジェクトからの一致した正常ゲノムの公開に伴って、これらの膨大なデータセットを効率的に解析することができるツールの大きな必要性が存在する。

【００４５】
そのために、本発明者らは、ＳＡＭ／ＢＡＭにフォーマット済みのファイル中に含有されるアラインしたショートリードデータを使用して患者の腫瘍及び生殖系列のゲノムからのそれぞれのゲノム位置を同時に解析するツールである、ＢａｍＢａｍを開発した（SAMtools library; Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R; 1000 Genome Project Data Processing Subgroup. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009 Aug 15;25(16):2078-9. Epub 2009 Jun 8）。ＢａｍＢａｍは、ＳＡＭｔｏｏｌｓライブラリとインタフェースして、ＳＡＭ／ＢＡＭにフォーマット済みのファイルからのショートリードアラインメントを使用して患者の腫瘍及び生殖系列のゲノムを同時に解析する。本開示中では、ＢａｍＢａｍツールは、情報ストリングを含む配列を比較するために使用する配列解析エンジンであり得る。一実施形態では、情報ストリングは、生物学的情報、たとえばポリヌクレオチド配列又はポリペプチド（ｐｏｌｙｐｅｔｉｄｅ）配列を含む。別の実施形態では、生物学的情報は、発現データ、たとえばｍＲＮＡ転写物又はｒＲＮＡ又はｔＲＮＡ又はペプチド又はポリペプチド又はタンパク質の相対濃度レベルを含むことができる。別の実施形態では、生物学的情報は、たとえばそれだけには限定されないが、リン酸化、硫酸化、アセチル化（ａｃｔｙｌａｔｉｏｎ）、メチル化、グリコシル化（ｇｌｙｃｏｓｉｌａｔｉｏｎ）、シアリル化（ｓｉａｌａｔｉｏｎ）、グリコシルホスファチジルイノシトールを用いた修飾、又はプロテオグリカンを用いた修飾などのタンパク質修飾の相対量であり得る。

【００４６】
この処理方法は、ＢａｍＢａｍが、腫瘍及び生殖系列のゲノムの両方中の全体的なコピー数を効率的に計算し、構造的変動の領域（たとえば染色体転座）を推量すること、全体的及び対立遺伝子特異的なコピー数を効率的に計算すること、ヘテロ接合性の喪失（ＬＯＨ）を示す領域を推量すること、体細胞及び生殖系列配列の変異（たとえば点突然変異）ならびに構造的再編成（たとえば染色体融合をどちらも発見することを可能にする。さらに、２つのゲノム配列を同時に比較することによって、ＢａｍＢａｍは、生殖系列配列の変異から体細胞のものを即時に区別する、腫瘍ゲノム中の対立遺伝子特異的なコピー数の変更を計算する、及び腫瘍ゲノム中で対立遺伝子の割合がシフトした染色体領域にわたって生殖系列ハプロタイプをフェージングすることもできる。これらの解析をすべて単一のツールへと一緒にすることによって、研究者は、ＢａｍＢａｍを使用して、患者の腫瘍ゲノム内の、多くの場合は特定の対立遺伝子に対して起こった多くの種類のゲノム変更を発見することができる、これは、腫瘍化の潜在的な駆動因子の同定を助ける。

【００４７】
発見された変異が体細胞変異であるか（すなわち、腫瘍中でのみ見つかる変異配列）生殖系列変異（すなわち、遺伝性又は遺伝的である変異配列）であるかを決定するためには

、腫瘍及び一致した正常ゲノムを何らかの方法で比較することを必要とする。これは、腫瘍及び生殖系列の両方のすべてのゲノム位置でのデータを要約し、その後、解析のために結果を合わせることによって、順次行うことができる。残念ながら、全ゲノムのＢＡＭファイルはその圧縮形態で数百ギガバイトであるため（非圧縮では１～２テラバイト）、後の解析のために記憶しておく必要のある中間結果は非常に大きなものとなり、統合及び解析が遅い。

【００４８】

この問題を回避するために、ＢａｍＢａｍは２つのファイルから同時に読み込み、常にそれぞれのＢＡＭファイルと他のファイルとの同調を保ち、２つのファイル間の共通のゲノム位置のすべてに重複するゲノムリードをパイルアップさせる。それぞれのパイルアップの対について、ＢａｍＢａｍは上記列挙した一連の解析を実行した後、パイルアップを廃棄して次の共通のゲノム位置へと移動する。この方法でこれらの膨大なＢＡＭファイルを処理することによって、コンピュータのＲＡＭ使用量が最小限となり、処理速度はファイルシステムが２つのファイルを読み込むことができる測度によって主に制限される。このことは、一台のコンピュータ又はコンピュータクラスタにわたって実行するために十分に柔軟である一方で、ＢａｍＢａｍが膨大な量のデータを素早く処理することを可能にする。これらのファイルをＢａｍＢａｍで処理することの別の重要な利点は、その出力がかなり最小限であることであり、それぞれのファイル中に見つかる重要な相違のみからなることである。これにより、本質的には患者の腫瘍及び生殖系列のゲノムの間の全ゲノム差分（ｄｉｆｆ）が生成され、それぞれのファイルにすべてのゲノム情報を別々に記憶した場合にかかるよりもはるかに少ないディスク記憶量を要する。

【００４９】

ＢａｍＢａｍは、大きな配列決定データセットを調査して、その生殖系列と比較してそれぞれの腫瘍内で起こる高品質なゲノム事象の組を生成するための、計算効率の良い方法である。これらの結果は腫瘍の染色体ダイナミクスへの瞥見をもたらし、腫瘍の最終状態及びそれらをもたらした事象の理解を改善させる。ＢａｍＢａｍデータフローの例示的なスキームを図１に示す。

【００５０】

本発明の特定の例示的な一実施形態は、差分遺伝子配列オブジェクトの作成及び使用である。本明細書中で使用するオブジェクトとは、ＢａｍＢａｍ技法からインスタンス化したデジタルオブジェクトを表し、参照配列（たとえば第１の配列（ｓｅｒｑｕｅｎｃｅ））と解析配列（たとえば第２の配列）との間の相違を反映する。オブジェクトは、多くの様々な市場において難所とみなされ得る。市場観点からそのようなオブジェクトの使用及び管理に関連する以下の要因を考慮し得る：

・オブジェクトは動的であり得、パラメータ（たとえば、時間、地理的地域、遺伝系図、種など）のベクターに関して変化することができる

・オブジェクトは、互いのオブジェクト又は参照配列に対して「距離」があるとみなされる場合がある。距離は、関連性の次元に応じて測定することができる。たとえば、距離は、時間に関する仮定上の標準からの偏差又はドリフトであり得る。

・オブジェクトは危険性の指標であり得る：疾患を発生する危険性、曝露に対する脆弱性、ある場所で働く危険性など。

・オブジェクトは利害関係者に提示するために管理することができる：医療提供者、保険会社、患者など。

・グラフオブジェクトとして提示することができる

・統計様式で提示することができる：１人の人、集団、正準の人など

・参照配列をオブジェクトから生成して、正規化した配列を形成することができる。正規化した配列は、測定したオブジェクトから誘導したコンセンサスに基づいて構築することができる。

・オブジェクトは、単一遺伝子のアラインメントではなく大きなサブゲノム又はゲノム情報を代表しており、アノテーションされている／標準のソフトウェアによって読み込み

可能なメタデータを含有する。
・オブジェクトは、検出することができる内部パターン又は構造を有することができる：ある状態に相関する突然変異の組は別の箇所での第２の突然変異の組と相関している場合がある；一群の差分パターンはホットスポットである可能性がある；多変量の解析又は他のＡＩ技法を使用して相関を同定する；ホットスポットの有意性（たとえば、存在、非存在など）を検出する
・１人の人に関連するオブジェクトはセキュリティーキーとして使用することができる
【００５１】
差分配列オブジェクトの更新（更新には、作成、修正、変更、削除などが含まれる）は、
・鋳型に基づくことができる
・ｄｅ　ｎｏｖｏオブジェクトであり得る
・既存のオブジェクトであり得る
【００５２】
例示的な代替の実施形態では、本方法を使用して、処置に対する患者の応答性を確認及び予測することができる：予想、仮定、予測、実測など。
例示的な代替の実施形態では、本方法を使用して患者に特異的な指示を提供することができる：処方、推奨、予後診断など。
一実施形態では、本方法を使用して、たとえば、癌組織の検出、癌組織の病期決定、転移性組織の検出など；それだけには限定されないが、アルツハイマー病、筋萎縮性側索硬化症（ＡＬＳ）、パーキンソン病、統合失調症、癲癇、及びそれらの合併症などの神経障害の検出；ディジョージ症候群、自閉症、多発性硬化症などの自己免疫障害、糖尿病等の発達障害；それだけには限定されないが、ウイルス感染症、細菌感染症、真菌感染症、リーシュマニア、住血吸虫症、マラリア、サナダムシ、象皮症、線虫、紐虫（ｎｅｍａｔｉｎｅｓ）による感染症などの感染症処置の、様々な診断及び治療の応用に使用することができる臨床的情報を提供し得る。
【００５３】
一実施形態では、本方法を使用して、遺伝子又はタンパク質の発現の変更に関連する状態のための、メッセンジャーＲＮＡ（ｍＲＮＡ）、リボソームＲＮＡ（ｒＲＮＡ）、トランスファーＲＮＡ（ｔＲＮＡ）、マイクロＲＮＡ（ｍｉＲＮＡ）、アンチセンスＲＮＡ（ａｓＲＮＡ）などへの変更及び／又は修飾を含めた遺伝子構造の変更、遺伝子突然変異、遺伝子生化学的修飾を検出及び定量するための臨床的情報を提供し得る。発現の変更に関連する状態、疾患又は障害には、後天性免疫不全症候群（ＡＩＤＳ）、アジソン病、成人呼吸窮迫症候群、アレルギー、強直性脊椎炎、アミロイド症、貧血、喘息、アテローム性動脈硬化症、自己免疫性溶血性貧血、自己免疫性甲状腺炎、良性前立腺肥大、気管支炎、チェディアック－東症候群、胆嚢炎、クローン病、アトピー性皮膚炎、皮膚筋炎、真性糖尿病、気腫、胎児赤芽球症、結節性紅斑、萎縮性胃炎、糸球体腎炎、グッドパスチャー症候群、痛風、慢性肉芽腫性疾患、グレーブス病、橋本甲状腺炎、過好酸球増加症、過敏性腸症候群、多発性硬化症、重症筋無力症、心筋又は心膜の炎症、骨関節炎、骨粗鬆症、膵炎、多嚢胞性卵巣症候群、多発性筋炎、乾癬、ライター症候群、関節リウマチ、強皮症、重症複合型免疫不全症（ＳＣＩＤ）、シェーグレン症候群、全身性アナフィラキシー、全身性エリテマトーデス、全身性硬化症、血小板減少性紫斑病、潰瘍性大腸炎、ブドウ膜炎、ウェルナー症候群、癌、血液透析、及び体外循環の合併症、ウイルス、細菌、真菌、寄生生物、原虫、及び蠕虫の感染症；腺癌、白血病、リンパ腫、黒色腫、骨髄腫、肉腫、奇形癌、特に、副腎、膀胱、骨、骨髄、脳、乳房、子宮頸部、胆嚢、神経節、胃腸管、心臓、腎臓、肝臓、肺、筋肉、卵巣、膵臓、副甲状腺、陰茎、前立腺、唾液腺、皮膚、脾臓、精巣、胸腺、甲状腺、及び子宮の癌が含まれる。診断的アッセイでは、変更された遺伝子発現を検出するために、ハイブリダイゼーション又は増幅技術を使用して、患者からの生体試料中における遺伝子発現を標準試料と比較し得る。この比較のための定性的又は定量的方法は当分野で周知である。

【００５４】
別の実施形態では、本方法を使用して、遺伝子又はタンパク質の発現の変更に関連する障害のための、メッセンジャーＲＮＡ（ｍＲＮＡ）、リボソームＲＮＡ（ｒＲＮＡ）、トランスファーＲＮＡ（ｔＲＮＡ）、マイクロＲＮＡ（ｍｉＲＮＡ）、アンチセンスＲＮＡ（ａｓＲＮＡ）などへの変更及び／又は修飾を含めた遺伝子構造の変更、遺伝子突然変異、遺伝子生化学的修飾を検出及び定量するための臨床的情報を提供し得る。発現の変更に関連する障害には、アカシジア、アルツハイマー病、健忘症、筋萎縮性側索硬化症（ＡＬＳ）、運動失調、双極性障害、緊張病、脳性麻痺、脳血管疾患、クロイツフェルト－ヤコブ病、認知症、鬱病、ダウン症候群、遅発性ジスキネジア、ジストニア、癲癇、ハンチントン病、多発性硬化症、筋ジストロフィー、神経痛、神経線維腫症、神経障害、パーキンソン病、ピック病、網膜色素変性症、統合失調症、季節性情動障害、老人性認知症、脳卒中、トゥレット症候群、ならびに特に脳の腺癌、黒色腫、及び奇形癌を含めた癌が含まれる。

【００５５】
一実施形態では、本方法を使用して、哺乳動物タンパク質の発現又は活性の変更に関連する状態の臨床的情報を提供し得る。そのような状態の例には、それだけには限定されないが、後天性免疫不全症候群（ＡＩＤＳ）、アジソン病、成人呼吸窮迫症候群、アレルギー、強直性脊椎炎、アミロイド症、貧血、喘息、アテローム性動脈硬化症、自己免疫性溶血性貧血、自己免疫性甲状腺炎、良性前立腺肥大、気管支炎、チェディアック－東症候群、胆嚢炎、クローン病、アトピー性皮膚炎、皮膚筋炎、真性糖尿病、気腫、胎児赤芽球症、結節性紅斑、萎縮性胃炎、糸球体腎炎、グッドパスチャー症候群、痛風、慢性肉芽腫性疾患、グレーブス病、橋本甲状腺炎、過好酸球増加症、過敏性腸症候群、多発性硬化症、重症筋無力症、心筋又は心膜の炎症、骨関節炎、骨粗鬆症、膵炎、多嚢胞性卵巣症候群、多発性筋炎、乾癬、ライター症候群、関節リウマチ、強皮症、重症複合型免疫不全症（ＳＣＩＤ）、シェーグレン症候群、全身性アナフィラキシー、全身性エリテマトーデス、全身性硬化症、血小板減少性紫斑病、潰瘍性大腸炎、ブドウ膜炎、ウェルナー症候群、癌、血液透析、及び体外循環の合併症、ウイルス、細菌、真菌、寄生生物、原虫、及び蠕虫の感染症；腺癌、白血病、リンパ腫、黒色腫、骨髄腫、肉腫、奇形癌、特に、副腎、膀胱、骨、骨髄、脳、乳房、子宮頸部、胆嚢、神経節、胃腸管、心臓、腎臓、肝臓、肺、筋肉、卵巣、膵臓、副甲状腺、陰茎、前立腺、唾液腺、皮膚、脾臓、精巣、胸腺、甲状腺、及び子宮の癌、アカシジア、アルツハイマー病、健忘症、筋萎縮性側索硬化症、運動失調、双極性障害、緊張病、脳性麻痺、脳血管疾患、クロイツフェルト－ヤコブ病、認知症、鬱病、ダウン症候群、遅発性ジスキネジア、ジストニア、癲癇、ハンチントン病、多発性硬化症、筋ジストロフィー、神経痛、神経線維腫症、神経障害、パーキンソン病、ピック病、網膜色素変性症、統合失調症、季節性情動障害、老人性認知症、脳卒中、トゥレット症候群、ならびに特に脳の腺癌、黒色腫、及び奇形癌を含めた癌が含まれる。

【００５６】
さらに別の実施形態では、本方法を使用して、遺伝子又はタンパク質の発現の変更に関連する障害のための、メッセンジャーＲＮＡ（ｍＲＮＡ）、リボソームＲＮＡ（ｒＲＮＡ）、トランスファーＲＮＡ（ｔＲＮＡ）、マイクロＲＮＡ（ｍｉＲＮＡ）、アンチセンスＲＮＡ（ａｓＲＮＡ）などへの変更及び／又は修飾を含めた遺伝子構造の変更、遺伝子突然変異、遺伝子生化学的修飾を検出及び定量するための臨床的情報を提供し得る。そのような障害の例には、それだけには限定されないが；腺癌、白血病、リンパ腫、黒色腫、骨髄腫、肉腫、奇形癌などの癌、特に、副腎、膀胱、骨、骨髄、脳、乳房、子宮頸部、胆嚢、神経節、胃腸管、心臓、腎臓、肝臓、肺、筋肉、卵巣、膵臓、副甲状腺、陰茎、前立腺、唾液腺、皮膚、脾臓、精巣、胸腺、甲状腺、及び子宮の癌；後天性免疫不全症候群（ＡＩＤＳ）、アジソン病、成人呼吸窮迫症候群、アレルギー、強直性脊椎炎、アミロイド症、貧血、喘息、アテローム性動脈硬化症、自己免疫性溶血性貧血、自己免疫性甲状腺炎、気管支炎、胆嚢炎、接触皮膚炎、クローン病、アトピー性皮膚炎、皮膚筋炎、真性糖尿病、気腫、リンパ球毒素を伴う偶発性リンパ球減少症、胎児赤芽球症、結節性紅斑、萎縮性

胃炎、糸球体腎炎、グッドパスチャー症候群、痛風、グレーブス病、橋本甲状腺炎、過好酸球増加症、過敏性腸症候群、多発性硬化症、重症筋無力症、心筋又は心膜の炎症、骨関節炎、骨粗鬆症、膵炎、多発性筋炎、乾癬、ライター症候群、関節リウマチ、強皮症、シェーグレン症候群、全身性アナフィラキシー、全身性エリテマトーデス、全身性硬化症、血小板減少性紫斑病、潰瘍性大腸炎、ブドウ膜炎、ウェルナー症候群、癌、血液透析、及び体外循環の合併症、ウイルス、細菌、真菌、寄生生物、原虫、及び蠕虫の感染症、外傷、ブルトン型Ｘ連鎖無ガンマグロブリン血症、後天性免疫グロブリン血症（ＣＶＩ）、ディジョージ症候群（胸腺形成不全）、胸腺異形成、ＩｇＡ単独欠損症、重症複合型免疫不全症（ＳＣＩＤ）、血小板減少症及び湿疹を伴う免疫不全（ウィスコットーアルドリッチ症候群）、チェディアックー東症候群、慢性肉芽腫性疾患、遺伝性血管神経性浮腫、ならびにクッシング病関連の免疫不全などの免疫障害；腎尿細管性アシドーシス、貧血、クッシング症候群、軟骨形成不全性小人症、デュシェンヌ及びベッカー型筋ジストロフィー、癲癇、性腺形成不全、ＷＡＧＲ症候群（ウィルムス腫瘍、無虹彩症、泌尿生殖器奇形、及び精神遅滞）、スミスーマゲニス症候群、骨髄異形成症候群、遺伝性粘膜上皮異形成、遺伝性角皮症、シャルコーーマリーートゥース病及び神経線維腫症などの遺伝性神経障害、甲状腺機能低下症、水頭症、シデナム舞踏病及び脳性麻痺などの発作性疾患、二分脊椎、無脳症、頭蓋脊椎披裂、先天性緑内障、白内障、感音性難聴、ならびに対象の任意の組織、器官、又は系、たとえば、脳、副腎、腎臓、骨格又は生殖器系に関与する細胞の成長と分化、胚形成、及び形態形成に関連する任意の障害などの発達障害が含まれる。

【００５７】

別の実施形態では、本方法を使用して、遺伝子又はタンパク質の発現の変更に関連する障害のための、メッセンジャーＲＮＡ（ｍＲＮＡ）、リボソームＲＮＡ（ｒＲＮＡ）、トランスファーＲＮＡ（ｔＲＮＡ）、マイクロＲＮＡ（ｍｉＲＮＡ）、アンチセンスＲＮＡ（ａｓＲＮＡ）などへの変更及び／又は修飾を含めた遺伝子構造の変更、遺伝子突然変異、遺伝子生化学的修飾を検出及び定量するための臨床的情報を提供し得る。そのような障害の例には、それだけには限定されないが、性腺機能低下症、シーハン症候群、尿崩症、カルマン病、ハンドーシュラーークリスチャン病、レッテラーーシーベ病、サルコイドーシス、トルコ鞍空洞症候群、及び小人症を含めた、下垂体機能低下症に関連する障害などの内分泌障害；末端肥大症、巨人症、及び抗利尿ホルモン（ＡＤＨ）不適合分泌症候群（ＳＩＡＤＨ）を含めた下垂体機能亢進症；甲状腺種、粘液水腫、細菌感染症に関連する急性甲状腺炎、ウイルス感染症に関連する亜急性甲状腺炎、自己免疫性甲状腺炎（橋本病）、及びクレチン症を含めた、甲状腺機能低下症に関連する障害；甲状腺中毒症及びその様々な形態、グレーブス病、前脛骨粘液水腫、中毒性多結節性甲状腺種、甲状腺癌、及びプランマー病を含めた、甲状腺機能亢進症に関連する障害；コーン病（慢性高カルシウム血症）を含めた副甲状腺機能亢進に関連する障害；アレルギー、喘息、急性及び慢性の炎症性肺疾患、ＡＲＤＳ、気腫、肺鬱血及び肺浮腫、ＣＯＰＤ、間質性肺疾患、肺癌などの呼吸器疾患；腺癌、白血病、リンパ腫、黒色腫、骨髄腫、肉腫、奇形癌などの癌、特に、副腎、膀胱、骨、骨髄、脳、乳房、子宮頸部、胆嚢、神経節、胃腸管、心臓、腎臓、肝臓、肺、筋肉、卵巣、膵臓、副甲状腺、陰茎、前立腺、唾液腺、皮膚、脾臓、精巣、胸腺、甲状腺、及び子宮の癌；ならびに後天性免疫不全症候群（ＡＩＤＳ）、アジソン病、成人呼吸窮迫症候群、アレルギー、強直性脊椎炎、アミロイド症、貧血、喘息、アテローム性動脈硬化症、自己免疫性溶血性貧血、自己免疫性甲状腺炎、気管支炎、胆嚢炎、接触皮膚炎、クローン病、アトピー性皮膚炎、皮膚筋炎、真性糖尿病、気腫、リンパ球毒素を伴う偶発性リンパ球減少症、胎児赤芽球症、結節性紅斑、萎縮性胃炎、糸球体腎炎、グッドパスチャー症候群、痛風、グレーブス病、橋本甲状腺炎、過好酸球増加症、過敏性腸症候群、多発性硬化症、重症筋無力症、心筋又は心膜の炎症、骨関節炎、骨粗鬆症、膵炎、多発性筋炎、乾癬、ライター症候群、関節リウマチ、強皮症、シェーグレン症候群、全身性アナフィラキシー、全身性エリテマトーデス、全身性硬化症、血小板減少性紫斑病、潰瘍性大腸炎、ブドウ膜炎、ウェルナー症候群、癌、血液透析、及び体外循環の合併症、ウイルス、細菌、真菌、寄生生物、原虫、及び蠕虫の感染症、外傷などの免疫障害が含まれる。ポ

リヌクレオチド配列は、変更された核酸配列発現を検出するために、サザンもしくはノーザン分析、ドットブロット、又は他の膜ベースの技術；ＰＣＲ技術；ディップスティック、ピン、及びＥＬＩＳＡアッセイ；ならびに患者からの体液又は組織を利用したマイクロアレイにおいて使用し得る。そのような定性的又は定量的方法は当分野で周知である。

【００５８】

本発明の特徴づけ及び最良の形態

「ＢａｍＢａｍ」は、大きな配列決定データセットを調査して、その生殖系列と比較してそれぞれの腫瘍内で起こる高品質なゲノム事象の組を生成するための、計算効率の良い方法である。これらの結果は腫瘍の染色体ダイナミクスへの瞥見をもたらし、腫瘍の最終状態及びそれらをもたらした事象の理解を改善させる。

【００５９】

診断学

本明細書中に記載の方法は、遺伝子又はタンパク質の発現の変更に関連する状態、疾患、又は障害のために、メッセンジャーＲＮＡ（ｍＲＮＡ）、リボソームＲＮＡ（ｒＲＮＡ）、トランスファーＲＮＡ（ｔＲＮＡ）、マイクロＲＮＡ（ｍｉＲＮＡ）、アンチセンスＲＮＡ（ａｓＲＮＡ）などへの変更及び／又は修飾を含めた遺伝子構造の変更、遺伝子突然変異、遺伝子生化学的修飾を検出及び定量するために使用し得る。また、本明細書中に記載の方法は、変更された遺伝子発現、非存在／存在対過剰、ｍＲＮＡの発現を検出及び定量するため、又は治療行為中のｍＲＮＡレベルを監視するためにも使用し得る。発現の変更に関連する状態、疾患又は障害には、特発性肺動脈高血圧、二次性肺高血圧、細胞増殖性障害、特に未分化乏突起神経膠腫、星細胞腫、オリゴ星細胞腫、膠芽細胞腫、髄膜腫、神経節神経腫、神経新生物、多発性硬化症、ハンチントン病、乳腺癌、前立腺腺癌、胃腺癌、転移性神経内分泌癌、非増殖性及び増殖性の乳腺線維嚢胞症、胆嚢の胆嚢炎及び胆石症、骨関節炎、ならびに関節リウマチ；後天性免疫不全症候群（ＡＩＤＳ）、アジソン病、成人呼吸窮迫症候群、アレルギー、強直性脊椎炎、アミロイド症、貧血、喘息、アテローム性動脈硬化症、自己免疫性溶血性貧血、自己免疫性甲状腺炎、良性前立腺肥大、気管支炎、チェディアック－東症候群、胆嚢炎、クローン病、アトピー性皮膚炎、皮膚筋炎、真性糖尿病、気腫、胎児赤芽球症、結節性紅斑、萎縮性胃炎、糸球体腎炎、グッドパスチャー症候群、痛風、慢性肉芽腫性疾患、グレーブス病、橋本甲状腺炎、過好酸球増加症、過敏性腸症候群、多発性硬化症、重症筋無力症、心筋又は心膜の炎症、骨関節炎、骨粗鬆症、膵炎、多嚢胞性卵巣症候群、多発性筋炎、乾癬、ライター症候群、関節リウマチ、強皮症、重症複合型免疫不全症（ＳＣＩＤ）、シェーグレン症候群、全身性アナフィラキシー、全身性エリテマトーデス、全身性硬化症、血小板減少性紫斑病、潰瘍性大腸炎、ブドウ膜炎、ウェルナー症候群、血液透析、体外循環、ウイルス、細菌、真菌、寄生生物、原虫、及び蠕虫の感染症；プロラクチン産生障害、卵管疾患、排卵不良、及び子宮内膜症を含めた不妊症、発情周期破壊、月経周期破壊、多嚢胞性卵巣症候群、卵巣過剰刺激症候群、子宮内膜又は卵巣腫瘍、子宮筋腫、自己免疫障害、子宮外妊娠、ならびに奇形発生；乳癌、乳腺線維嚢胞症、及び乳汁漏出；精子形成破壊、異常精子生理学、良性前立腺肥大、前立腺炎、ペイロニー病、性交不能症、女性化乳房；光線性角化症、動脈硬化症、滑液包炎、硬変、肝炎、混合性結合組織病（ＭＣＴＤ）、骨髄線維症、発作性夜間ヘモグロビン尿症、真性赤血球増加症、原発性血小板血症、癌の合併症、腺癌、白血病、リンパ腫、黒色腫、骨髄腫、肉腫、奇形癌、特に、副腎、膀胱、骨、骨髄、脳、乳房、子宮頸部、胆嚢、神経節、胃腸管、心臓、腎臓、肝臓、肺、筋肉、卵巣、膵臓、副甲状腺、陰茎、前立腺、唾液腺、皮膚、脾臓、精巣、胸腺、甲状腺、及び子宮の癌を含めた癌が含まれる。別の態様では、本発明の核酸。

【００６０】

本明細書中に記載の方法は、遺伝子又はタンパク質の発現の変更に関連する障害のために、メッセンジャーＲＮＡ（ｍＲＮＡ）、リボソームＲＮＡ（ｒＲＮＡ）、トランスファーＲＮＡ（ｔＲＮＡ）、マイクロＲＮＡ（ｍｉＲＮＡ）、アンチセンスＲＮＡ（ａｓＲＮＡ）などへの変更及び／又は修飾を含めた遺伝子構造の変更、遺伝子突然変異、遺伝子生

化学的修飾を検出及び定量するために使用し得る。また、本明細書中に記載の方法は、変更された遺伝子発現、ｍＲＮＡの非存在、存在、もしくは過剰発現を検出及び定量するため、又は治療行為中のｍＲＮＡレベルを監視するためにも使用し得る。発現の変更に関連する障害には、アカシジア、アルツハイマー病、健忘症、筋萎縮性側索硬化症、運動失調、双極性障害、緊張病、脳性麻痺、脳血管疾患、クロイツフェルトーヤコブ病、認知症、鬱病、ダウン症候群、遅発性ジスキネジア、ジストニア、癲癇、ハンチントン病、多発性硬化症、筋ジストロフィー、神経痛、神経線維腫症、神経障害、パーキンソン病、ピック病、網膜色素変性症、統合失調症、季節性情動障害、老人性認知症、脳卒中、トゥレット症候群、ならびに特に脳の腺癌、黒色腫、及び奇形癌を含めた癌が含まれる。

【００６１】

遺伝子発現に関連する状態、疾患又は障害の診断の基礎を提供するために、正常又は標準の発現プロフィールを確立する。これは、動物又はヒトのいずれかの正常対象から採取した生体試料を、ハイブリダイゼーション又は増幅の条件下でプローブと合わせることによって達成し得る。標準ハイブリダイゼーションは、正常対象を使用して得られた値を、既知の量の実質的に精製された標的配列を使用した実験からの値と比較することによって定量し得る。この様式で得られた標準値を、特定の状態、疾患、又は障害について症候性である患者からの試料から得られた値と比較し得る。特定の状態に近づく標準値からの偏差を使用してその状態を診断する。

また、そのようなアッセイは、動物研究及び臨床治験における特定の治療処置レジメンの有効性を評価するため、又は個々の患者の処置を監視するために使用し得る。状態の存在が確立され、処置プロトコルが開始された後、診断的アッセイを定期的に繰り返して、患者中の発現のレベルが正常対象中で観察されるレベルに近づき始めるかどうかを決定し得る。また、アッセイは、腫瘍の存在、腫瘍の非存在、又は臨床的処置もしくは治療を受けている個体の寛解状態を示す及び／又は同定する、メッセンジャーＲＮＡ（ｍＲＮＡ）、リボソームＲＮＡ（ｒＲＮＡ）、トランスファーＲＮＡ（ｔＲＮＡ）、マイクロＲＮＡ（ｍｉＲＮＡ）、アンチセンスＲＮＡ（ａｓＲＮＡ）などへの変更及び／又は修飾を含めた遺伝子構造、遺伝子突然変異、遺伝子生化学的修飾を検出、定量（ｑｕａｍｔｉｆｙ）、又は測定するためにも使用し得る。連続的なアッセイから得られた結果を使用して、数日間から数カ月の範囲の期間にわたる処置の有効性を示し得る。

【００６２】

本明細書中に開示した方法は、以前に同定されていない、又は特定の臨床的疾患、障害、もしくは状態と関連づけられていない、メッセンジャーＲＮＡ（ｍＲＮＡ）、リボソームＲＮＡ（ｒＲＮＡ）、トランスファーＲＮＡ（ｔＲＮＡ）、マイクロＲＮＡ（ｍｉＲＮＡ）、アンチセンスＲＮＡ（ａｓＲＮＡ）などへの変更及び／又は修飾を含めた遺伝子構造の変化、遺伝子突然変異、遺伝子生化学的修飾を検出する、定量する、及び相関させるためにも使用し得る。代替の方法では、本明細書中に開示した方法は、新規の臨床的疾患、障害、又は状態を同定するために使用し得る。その後、遺伝子構造、遺伝子突然変異、及び遺伝子生化学的修飾における新規の変化を核酸配列又はタンパク質配列の既知の化学的及び生化学的特性と比較してよく、臨床的疾患、障害、又は状態と相関するものを使用して、臨床的使用のための細胞代謝に関する新しいデータベース及び知識を作成し得る。

【００６３】

モデル系

ヒトと同様の毒性応答を示し、曝露条件がヒトでの曝露に関連性がある、動物モデルをバイオアッセイとして使用し得る。哺乳動物が最も一般的なモデルであり、ほとんどの毒性研究は、低コスト、入手可能性、及び豊富な参照毒性学が理由でラット又はマウスなどのげっ歯類に対して行う。同系交配させたげっ歯類株は、目的遺伝子の過少発現又は過剰発現の生理的結果を調査するため、ならびに疾患の診断及び処置方法を開発するための、好都合なモデルを提供する。また、特定の遺伝子（たとえば乳中に分泌されるもの）を過剰発現するように同系交配させた哺乳動物は、その遺伝子によって発現されたタンパク質の好都合な供給源としても役割を果たし得る。

【００６４】
毒性学
毒性学とは、生物系に対する薬物の効果の研究である。大多数の毒性研究が、ヒトの健康に対するこれらの薬剤の効果の予測を助けるために、ラット又はマウスに対して行われている。生理学、行動、恒常性プロセス、及び致死性における定性的及び定量的変化の観察を使用して、毒性プロフィールを作成し、薬剤に曝露させた後のヒトの健康に対する結果を評価する。
【００６５】
遺伝毒性学は、薬剤が遺伝子突然変異を生じる能力を同定及び解析する。遺伝毒性剤は、通常は、核酸との相互作用を促進する共通の化学又は物理特性を有しており、染色体異常が子孫へと伝えられる場合に最も有害である。毒性学研究は、受胎前の親に、妊娠中の母親、又は発生中の生物のいずれかに投与した場合に子孫における構造的又は機能的異常の頻度を増加させる薬剤を同定し得る。統計的要件を満たすために必要な生物数を生じる、その短い生殖周期が理由で、マウス及びラットがこれらの試験において最も頻繁に使用される。
急性毒性試験は対象に対する薬剤の単一の投与に基づいて、薬剤の症候学又は致死性を決定する。３つの実験を実施する：（ａ）初期の用量範囲発見実験、（ｂ）有効用量の範囲を狭めるための実験、及び（ｃ）用量応答曲線を確立するための最終実験。
長期毒性試験は薬剤の繰返し投与に基づく。ラット及びイヌをこれらの研究において一般的に使用して、様々なファミリー中の種からのデータを提供する。発癌を例外として、薬剤を高用量濃度で３～４カ月の期間の間１日１回投与することで、成体動物における毒性のほとんどの形態が明らかとなるという多数の証拠が存在する。
１年間以上の期間をもつ慢性毒性試験は、毒性の非存在又は薬剤の発癌潜在性のどちらかを実証するために使用する。研究をラットに対して実行する場合、最小で３つの試験群及び１つの対照群を使用し、動物を最初及び実験全体にわたって間隔的に検査及び監視する。
【００６６】
トランスジェニック動物モデル
目的遺伝子を過剰発現又は過少発現するトランスジェニックげっ歯類を同系交配させ、ヒト疾患をモデリングするため、又は治療剤もしくは毒性剤を試験するために使用し得る。（参照により本明細書に組み込まれている米国特許第４，７３６，８６６号、第５，１７５，３８３号、及び第５，７６７，３３７号を参照。）一部の事例では、導入された遺伝子は、胎児発達中又は出産後の特定の時点で特定の組織型中で活性化させ得る。導入遺伝子の発現は、実験薬物療法を用いたチャレンジの前、その間、及びその後でのトランスジェニック動物における表現型又は組織特異的なｍＲＮＡ発現の解析によって監視する。
【００６７】
胚性幹細胞
げっ歯類の胚から単離した胚性幹細胞（ＥＳ）は、胚を形成する潜在性を保持する。ＥＳ細胞をキャリアの胚内に入れた際、これらは正常な発達を再開し、生産動物のすべての組織に貢献する。ＥＳ細胞は、実験用ノックアウト及びノックインげっ歯類株の作製に使用される好ましい細胞である。マウス１２９／ＳｖＪ細胞系などのマウスＥＳ細胞は初期マウス胚に由来し、当分野で周知の培養条件下で成長させる。ノックアウト株用のベクターは、ｉｎ　ｖｉｖｏでの転写及び／又は翻訳を破壊するマーカー遺伝子を含むように改変された疾患遺伝子候補を含有する。ベクターは、電気穿孔、リポソーム送達、微量注入などの当分野で周知の形質転換方法によってＥＳ細胞内に導入する。内在げっ歯類遺伝子は、細胞分裂中の相同組換え及び組込みによって、破壊された疾患遺伝子によって置き換えられる。形質転換されたＥＳ細胞を同定し、好ましくはＣ５７ＢＬ／６マウス株からのものなどのマウス細胞胚盤胞内に微量注入する。胚盤胞を偽妊娠した雌親へと外科的に移植し、生じるキメラ子孫を遺伝子型決定し、繁殖させて、ヘテロ接合性又はホモ接合性の株を生成する。

また、ＥＳ細胞は、神経細胞、造血系、及び心筋細胞などの様々な細胞種及び組織のｉｎ　ｖｉｔｒｏでの分化を研究するためにも使用されている（Bain et al. (1995) Dev. Biol. 168: 342-357、Wiles and Keller (1991) Development 111: 259-267、及びKlug et al. (1996) J. Clin. Invest. 98: 216-224）。最近の進展により、ヒト胚盤胞に由来するＥＳ細胞は、内肺葉、中胚葉、及び外胚葉（ectodermnal）細胞種を含めた８つの別々の細胞系列へと分化するようにｉｎ　ｖｉｔｒｏで操作し得ることも実証されている（Thomson (1998) Science 282: 1145-1147）。

【００６８】

ノックアウト解析

遺伝子ノックアウト解析では、ヒト疾患遺伝子候補の領域を、ネオマイシンホスホトランスフェラーゼ遺伝子（ｎｅｏ、たとえばCapecchi (1989) Science 244: 1288-1292を参照）などの非哺乳動物遺伝子を含むように酵素的に改変させる。挿入されたコード配列は標的遺伝子の転写及び翻訳を破壊し、疾患候補タンパク質の生化学的合成を妨げる。改変された遺伝子を培養胚性幹細胞内に形質転換させ（上述）、形質転換細胞をげっ歯類胞胚内に注入し、胞胚を偽妊娠した雌親へと移植する。トランスジェニック子孫を異種交配させて、ホモ接合性の同系交配系を得る。

【００６９】

ノックイン解析

胚発生の初期に存在する全能性ＥＳ細胞を使用して、ヒト疾患のノックインヒト化動物（ブタ）又はトランスジェニック動物モデル（マウスもしくはラット）を作製することができる。ノックイン技術では、ヒト遺伝子の領域を動物ＥＳ細胞内に注入し、ヒト配列は組換えによって動物細胞ゲノム内に組み込まれる。組み込まれたヒト遺伝子を含有する全能性ＥＳ細胞は上述のように取り扱う。同系交配させた動物を研究及び処置して、類似のヒト状態に関する情報を得る。これらの方法は、いくつかのヒト疾患をモデリングするために使用されている。（たとえば、Lee et al. (1998) Proc. Natl. Acad. Sci. 95: 11371-11376、Baudoin et al. (1998) Genes Dev. 12: 1202-1216、及びZhuang et al. (1998) Mol. Cell Biol. 18: 3340-3349を参照）。

【００７０】

非ヒト霊長類モデル

動物試験の分野は、生理学、遺伝学、化学、薬理学及び統計学などの基礎科学からのデータ及び方法を扱う。これらのデータは、ヒトの健康に関連している可能性があるため、非ヒト霊長類に対する治療剤の効果を評価することにおいて卓越している。ワクチン及び薬物の評価におけるヒトの代用としてサルが使用されており、その応答は、同様の条件下でのヒトの曝露に関連性がある。カニクイザル（Macaca fascicularis、Macaca mulata）及びコモンマーモセット（Callithrix jacchus）がこれらの調査で使用されている最も一般的な非ヒト霊長類（ＮＨＰ）である。ＮＨＰのコロニーの育成及び維持には多大な費用が関連するため、初期研究及び毒性学研究は、通常はげっ歯類モデルで実施する。薬物嗜癖などの行動測定を使用する研究では、ＮＨＰが第一選択肢の試験動物である。さらに、ＮＨＰ及び個々のヒトは多くの薬物及び毒素に対して異なる感受性を示し、これらの薬剤の「高（extensive）代謝者」及び「低（poor）代謝者」として分類することができる。

【００７１】

本発明の例示的な使用

個別化医療は、特定の処置を、利益を受ける可能性が最も高い患者へ送達することを保証する。本発明者らは、治療化合物の約半数が、臨床的に関連性のある転写又はゲノムの乳癌サブタイプのうちの１つ又は複数において優先的に有効であることを示した。これらの発見は、乳癌処置において応答関連の分子サブタイプを定義する重要性を支援している。また、本発明者らは、細胞系に関する転写及びゲノムデータの経路の組込みは、観察されたサブタイプ特異的な応答の機構的な説明を提供するサブネットワークを明らかにすることも示す。細胞系と腫瘍との間のサブネット活性の比較分析は、サブタイプ特異的なサブネットワークの大多数が細胞系と腫瘍との間で保存的であることを示す。これらの解析

は、十分に特徴づけられた細胞系パネルにおける実験化合物の前臨床スクリーニングが、初期段階の臨床治験において感度上昇に使用することができる候補の応答関連分子シグネチャを同定することができるという考えを支援する。本発明者らは、このｉｎ　ｖｉｔｒｏ評価手法は、応答性の腫瘍サブタイプが、化合物の臨床開発が開始される前に同定され、それによって費用を減らし、最終的なＦＤＡ認可の確率を増加させ、場合によっては、応答する見込みのない患者を処置することに関連する毒性を回避する可能性を増加させることを提案する。本研究では、本発明者らは転写性サブタイプを定義する分子シグネチャのみを評価し、再発性ゲノムコピー数異常（ＣＮＡ）を選択した。本発明者らは、遺伝子突然変異、メチル化及び選択的スプライシングなどの追加の分子特長が解析に含められるにつれて、この手法の能力及び制度が増加すると予想する。同様に、細胞系パネルのサイズを大きくすることは、パネル内のより一般的でない分子パターンを評価する能力を増加させ、ヒト乳癌中に存在する多様性のより完全な範囲を表す確率を増加させる。

【００７２】

ここでは、本発明者らは、腫瘍（体細胞）及び生殖系列の一致した配列決定データセットの迅速な比較を可能にする、ＢａｍＢａｍと呼んでいる新しいソフトウェアツールを開示する。ＢａｍＢａｍによって出力される結果は可変性であり、それぞれの患者の試料によって含有される体細胞及び生殖系列の変異の徹底的なカタログが生成される。このカタログは、研究者に、腫瘍の発生中に起こった重要な変化を素早く見つける能力を提供するが、患者の生殖系列中に存在する、疾患に対する素因を示し得る高品質な変異も提供する。ＢａｍＢａｍのさらなる改善は、腫瘍化の駆動因子を指示し得る、同じゲノム領域中で起こっている複数の種類の変異（たとえば、遺伝子の一方の対立遺伝子が欠失しており、他方の対立遺伝子がブレイクポイントによる切断突然変異を含有する）を具体的に検索する方法からなる。また、本発明者らは、ゲノム対を超えるものを処理するＢａｍＢａｍの能力を拡大すること、及び、研究者に、自身の解析方法をＢａｍＢａｍのパイプラインに差し込む能力を提供することも計画している。

さらなる実施形態では、ポリヌクレオチド核酸は未だ開発されていない任意の分子生物学技法において使用してよいが、ただし、その新しい技法は、それだけには限定されないが、トリプレット遺伝暗号及び特異的塩基対相互作用などの特性を含めた現在知られている核酸分子の特性に依存するものである。

【００７３】

図６は、遺伝子配列解析の生態系１００を例示し、これには、１つ又は複数のデータベースと連結させた配列解析エンジン１４０が、場合によってはネットワーク１１５（たとえば、ＬＡＮ、ＷＡＮ、ＶＰＮ、インターネットなど）上で含まれる。好ましいデータベースには、１つ又は複数の組織の遺伝子配列ストリングを記憶する遺伝子データベース１１０、局所差分ストリングを表す差分遺伝子配列オブジェクトを記憶する差分配列データベース１２０、及び患者、人、集団、又は他の種の実体に関連する１つ又は複数の医療記録を記憶する医療記録データベース１３０が含まれる。また、医療記録データベース１３０は、１つ又は複数の差分遺伝子配列オブジェクト、場合によっては患者、人、集団又は他の群に関連するものも記憶することができる。

【００７４】

本発明の主題の一態様は、差分遺伝子配列オブジェクトの管理を含むとみなされる。遺伝子配列ストリングの解析によって、解析エンジン１４０は差分ストリング又は一群の差分ストリング１４５を作成することができる。差分ストリング１４５は差分遺伝子配列オブジェクトへと変換することができ、これは立ち代って差分配列データベース１２０又は医療記録データベース１３０に記憶させることができる。配列オブジェクトは、オブジェクトの性質を説明する１つ又は複数の属性を用いてタグづけすることができる。属性例には、オブジェクト作成のタイムスタンプ、試料を患者から採取した際のタイムスタンプ、患者名、人口統計学的情報、組織型（たとえば、健康、疾患、腫瘍、器官組織など）、又は他の特長が含まれ得る。属性は、解析エンジン１４０によって活用されて、医療記録データベース１３０中の医療記録に関連する特徴間の１つ又は複数の相関を確立することが

できる。

【００７５】

差分遺伝子配列オブジェクトの管理は広範囲の役割又は責任をカバーする。上述のように、一態様にはそのようなオブジェクトの作成が含まれる。また、解析エンジン１４０は、好ましくは、所望に応じて配列オブジェクトを更新、解析、修正、時間追跡、削除、コピー、分割、追加、又は他の様式で操作するようにも構成されている。さらに、解析エンジン１４０は、差分遺伝子配列オブジェクト管理インタフェース、場合によっては出力装置１９０上のものを提供することができる。たとえば、一部の実施形態では、生態系１００は、インターネット上で利用可能な１つ又は複数のウェブサーバを備える有料サービスとして動作する。そのような実施形態では、ブラウザを備えたコンピュータは、解析エンジン１４０とインタフェースして、差分遺伝子配列オブジェクトを管理又はそれと相互作用することができる。

【００７６】

一部の実施形態では、以下にさらに記述するように、解析エンジン１４０は、遺伝子データベース１１０から得られた遺伝子配列ストリングを解析するように構成されている。好ましくは、遺伝子配列ストリングは少なくとも２つの異なる組織試料内で関連している。解析エンジン１４０は、配列ストリング中の対応するサブストリングの少なくとも１つの既知の位置を使用して少なくとも２つの配列を増分同期することによって、１つ又は複数の局所アラインメント１４３を生成する。さらに、解析エンジン１４０は、局所アラインメントを使用して、遺伝子配列ストリング間の１つもしくは複数の局所差分ストリング１４５又は一群の差分ストリング１４５を生成する。その後、解析エンジン１４０は、差分ストリング１４５を使用して、差分配列データベース１２０又は医療記録データベース１３０中の差分遺伝子配列オブジェクトを更新することができる。その後、差分配列オブジェクトをさらなる解析のために使用することができる。

【００７７】

一部の実施形態では、解析エンジン１４０は、特定の患者、人、個体、家族、集団、又は他の群の差分遺伝子配列オブジェクトを記憶する医療記録データベース１３０と通信可能に連結する。解析エンジン１４０は患者の差分配列オブジェクトを入手し、患者の配列オブジェクトに関連する局所差分ストリング又は一群の差分ストリングの存在に基づいて患者に特異的なデータセットを生成する。その後、解析エンジン１４０は、患者に特異的なデータセットを活用して、１つ又は複数の患者に特異的な指示１５１を生成又は他の様式で生じることができる。たとえば、患者の特異的局所差分ストリングの解析によって、解析エンジン１４０は患者の特異的差分ストリングと既知の状態との間に相関が存在するかどうかを決定することができ、立ち代ってこれは指示にマッピングすることができる。企図される指示には、診断、予後診断、推奨処置、予測、処方、又は他の種の指示が含まれ得る。

【００７８】

さらに他の実施形態では、解析エンジン１４０は医療記録データベース１３０に記憶されている差分遺伝子配列オブジェクトを入手し、配列オブジェクトは個体の集団に関連している。解析エンジン１４０は複数の配列オブジェクトから一群の局所差分ストリングを同定し、前記一群から一群記録１５２を生成する。一群記録１５２は、集団に関連する局所差分ストリングに関連する情報（たとえば、属性、特性、メタデータ、特徴など）の提示を備える。解析エンジン１４０は一群記録１５２を使用して集団解析記録１５３を生成する。したがって、差分遺伝子配列オブジェクトを集団セグメントへとマッピングすることができる。

さらに別の実施形態は、差分遺伝子配列オブジェクトを使用して人の遺伝子配列が参照試料から偏差する程度を決定する解析エンジン１４０を含む。参照差分遺伝子配列オブジェクト、場合によっては実在の人又は正準の人を表すものは、医療記録データベース１３０中に医療記憶として記憶させることができる。解析エンジン１４０は、人に関連する様々な配列オブジェクトからの人の局所差分ストリングと参照差分遺伝子配列オブジェクト

からの局所差分ストリングとの間の偏差を計算する。偏差を計算した後、解析エンジン１４０は偏差又は逸脱を表す偏差記録１５４を生成する。システム中の他の記録と同様、偏差記録１５４には、記録中の情報の特徴を反映する属性（たとえば、人名、タイムスタンプ、試料型など）も含めることができる。その後、解析エンジン１４０は、偏差記録１５４を活用して、人の遺伝子配列が参照差分スティング（sting）からどのように偏差するか、又はその度合を示す、人に特異的な偏差プロフィール１５５を生成することができる。

【００７９】

解析の種類又は生成された結果（たとえば、患者指示１５１、集団解析１５３、人に特異的なプロフィール１５５など）にかかわらず、解析エンジン１４０は、出力装置１９０が結果を提示するようにさらに構成することができる。出力装置１９０は、好ましくは解析エンジン１４０と連結させた計算装置、場合によってはネットワーク１１５上のものを備える。出力装置１９０の例には、携帯電話、案内所、医療現場でのコンピュータ端末、保険会社のコンピュータ、プリンタ、画像診断装置、ゲノムブラウザ、又は他の種の装置が含まれる。

したがって、本発明の主題に従ったシステムを使用することは、典型的には遺伝子データベースを含む。既に上述したように、遺伝子データベースは物理的には一台のコンピュータ上に位置していてよいが、分散データベースも本明細書における使用に適切であるとみなされることを理解されたい。さらに、そのようなデータベースが、それぞれ第１及び第２の組織を表す、複数の対応するサブストリングを有する第１及び第２の遺伝子配列ストリングの記憶及び検索が可能であり限りは、データベースの特定のフォーマットは本発明の主題を限定しないことも理解されたい。

【００８０】

同様に、第１及び第２の遺伝子配列ストリングが、ゲノム中での位置が既知である１つ又は複数の対応するサブストリングを含む限りは、第１及び第２の遺伝子配列ストリングの特定のフォーマットは本発明の主題を限定しないことに注意されたい。したがって、適切なデータフォーマットには単純なＡＳＣＩＩ又はバイナリコードを含み、配列ストリングは、現在知られている配列解析ツールにおいて一般的に用いられている仕様に従ってフォーマットし得る。したがって、特に好ましいフォーマットには、ＥＭＢＬ、ＧＣＧ、ｆａｓｔａ、ＳｗｉｓｓＰｒｏｔ、ＧｅｎＢａｎｋ、ＰＩＲ、ＡＢＩ、及びＳＡＭ／ＢＡＭフォーマットが含まれる。

【００８１】

解析

解析及び試料の特定の性質に応じて、遺伝子配列ストリングの種類は相当に変動する場合があり、配列は核酸配列（ＤＮＡ又はＲＮＡ）及びタンパク質配列であり得ることを指摘すべきである。最も典型的には、しかし、遺伝子配列ストリングは、解析下の第１及び第２の組織のゲノム、トランスクリプトーム、及び／又はプロテオームの顕著な一部分を表す核酸ストリングとなる。たとえば、第１及び第２の遺伝子配列ストリングは、第１及び第２の組織のゲノム、トランスクリプトーム、又はプロテオームの少なくとも１０％、より典型的には少なくとも２５％、より典型的には少なくとも５０％、より典型的には少なくとも７０％でさえ、最も典型的には少なくとも９０％、又は実質的に全体（少なくとも９８％）でさえを表すことが企図される。したがって、本明細書中に提示するシステム及び方法は、第１及び第２の組織との間の有意差の迅速かつ高度に包括的な全体像を可能にする一方で、コンパクトかつ情報価値のある出力ファイルを生成することを理解されたい。

【００８２】

調査下にある組織の種類に応じて、複数の種類の解析を行うことができることに注意されたい。たとえば、第１及び第２の組織が同じ生物学的実体を起源とする場合は、健康な組織を異なる健康な組織に対して比較し得るか、又は健康な組織を対応する患部組織（たとえば腫瘍組織）に対して比較し得る。したがって、生物学的実体は健康な個体又は疾患

もしくは障害を診断された個体であり得る。他方では、細胞系（不死化又は初代）に由来する場合は、薬物の遺伝的影響又は後成的影響を迅速に同定し得る。同様に、第１及び第２の組織が幹細胞に由来する場合は、発達中の胚の遺伝的組成の変化又は遺伝的可変性を解析し得る。さらなる企図される例では、第１及び第２の組織は、疾患の進行又は処置の効果を調査するための、実験動物モデルからのものであり得る。あるいは、第１及び第２の組織は、酵母、組換え細菌細胞、及び／又はウイルスからのものでさえあり得る。

したがって、対応するサブストリングの性質は相当に変動し、採取した組織の種類及びゲノムのカバレッジ量に少なくとも部分的に依存することを認識されたい。しかし、ゲノムカバレッジが比較的高いことが典型的には好ましく、ほとんどの事例では全ゲノムを解析する。したがって、対応するサブストリングには、典型的にはホモ接合性及びヘテロ接合性の対立遺伝子が含まれる。

【００８３】

サブストリングの種類にかかわらず、同期は、第１のストリング内の先験的に既知の位置に基づいて複数のサブストリングのうちの少なくとも１つをアラインさせるステップを含むことが一般的に好ましい。様々な生物（特にヒト）の数々のゲノムが既に実質的に完全にアノテーションされており、未知の配列でさえもしばしば少なくとも推定上の機能がアノテーションされており、かつ実質的に（直鎖状）配列のゲノム全体が知られているため、参照ゲノムに関する先験的に既知の位置の数は高い。したがって、参照ゲノム内のアノテーションの知識は有効かつ正確な同期のロードマップとして役割を果たすであろう。もちろん、参照ゲノムの性質は必ずしも単一の健康な組織のゲノムに限定されず、参照ゲノムは任意の定義された（実測又は計算した）ゲノム構造であり得ることを理解されたい。たとえば、参照ゲノムを複数人の健康な個体（の典型的には単一組織）から構築して、コンセンサス参照配列を生成し得る。あるいは、参照ストリングは、同じ（もしくは異なる）個体の複数の組織のコンセンサス、又は患部組織試料のコンセンサス（同じもしくは複数の患者からのもの）に基づき得る。

【００８４】

したがって、差分遺伝子配列オブジェクトは、参照組織に対する１つ又は複数の試料組織の情報を提供することを認識されたい。したがって、参照ストリングの選択肢に応じて、差分遺伝子配列オブジェクトの情報内容は相当に変動し得る。たとえば、差分遺伝子配列オブジェクトは、試料が（参照ストリングによって定義される）特定の部分集団に一致する、又は試料が疾患もしくは状態に関連しているもしくはしていない可能性のある複数のミスマッチを有するという情報を提供し得る。

【００８５】

本発明の主題のさらに好ましい態様では、同期は、複数のサブストリングのうちの少なくとも１つの長さよりも短い長さのウィンドウ内でサブストリングをアラインさせることによっても行い得る。最も好ましくは、同期は、第１の配列ストリングの全長にわたって第１及び第２の配列ストリングを繰り返して増分同期させることによって行う。したがって、異なる観点から見ると、同期は、２つの半分を増分的に一致させてアラインメントを生じるジッパーのものに類似の様式で行う。同じイメージを使用すると、その後、閉じたジッパーのミスマッチの部分のみを差分遺伝子配列オブジェクトに反映させる。

したがって、差分遺伝子配列オブジェクトは、１つ又は複数の局所差分ストリング、典型的には少なくともゲノムの定義された一部分（たとえば少なくとも１つの染色体）、より典型的には第１又は第２の組織の実質的に全ゲノムを表すことを認識されたい。もちろん、既に知られている位置及び／又は決定された参照ストリングからの偏差に基づいて、差分遺伝子配列オブジェクトには、差分遺伝子配列オブジェクトを記述するメタデータを有する１つ又は複数の属性が含まれることに注意されたい。たとえば、属性は、第１及び／又は第２の組織の状態を説明するものであり得る。状態が生理的状態である場合は、メタデータは新生物成長、アポトーシス、分化状態、組織年齢、及び／又は組織の処置に対する応答性を反映し得る。他方では、状態が遺伝子状態である場合は、メタデータは倍数性、遺伝子のコピー数、反復のコピー数、反転、欠失、ウイルス遺伝子の挿入、体細胞突

然変異、生殖系列突然変異、構造的再編成、転位、及び／又はヘテロ接合性の喪失を反映し得る。同様に、状態には組織内のシグナル伝達経路に関連する経路モデル情報（たとえば、薬物に対する予想される応答性、受容体の欠損など）が含まれ、特に企図される経路には、シグナル伝達経路（たとえば、成長因子シグナル伝達経路、転写因子シグナル伝達経路、アポトーシス経路、細胞周期経路、ホルモン応答経路など）が含まれる。

【００８６】

本明細書中に提示するシステム及び方法によって提供される出力情報は、参照ストリングからの複数の偏差を示す単一の差分遺伝子配列オブジェクト、もしくは参照ストリングからの個々の偏差を示す複数の差分遺伝子配列オブジェクト、又はその任意の合理的な組合せの形態であり得る。最も典型的には、差分遺伝子配列オブジェクトは電子フォーマットのものであり、したがってコンピュータ読取可能なファイルとして検索及び／又は転送される。容易に認識されるように、ファイルは最も好ましくは標準化されており、フォーマットはＳＡＭ／ＢＡＭフォーマットに従うことが特に好ましい。

したがって、上記に鑑みて、差分遺伝子配列オブジェクトを様々な様式で使用してよく、差分遺伝子配列オブジェクトは医療、集団解析、及び個別化医療における数々の応用に特に適切であることを理解されたい。

【００８７】

たとえば、１つ又は複数の差分遺伝子配列オブジェクトが個体について既知である場合、患者の差分遺伝子配列オブジェクト中の局所差分ストリング又は複数の局所差分ストリングの一群に基づく、患者に特異的なデータセットを生成してよく、その後、患者に特異的なデータセットを使用して、患者に特異的な指示を生成する。典型的な例では、本発明者らは、解析エンジンが患者の差分遺伝子配列オブジェクトを記憶する医療記録記憶装置と連結されている、医療サービスを提供する方法を企図する。その後、解析エンジンは、患者の差分遺伝子配列オブジェクト中の１つもしくは複数の局所差分ストリング又は複数の局所差分ストリングの一群を使用して患者に特異的なデータを生成し、患者に特異的なデータセットに基づいて患者に特異的な指示を生成する。

【００８８】

医療記録記憶装置は数々の様式で構成されていてよく、また、患者によって携帯可能であり得る（たとえば患者が保有するスマートカード）、患者によってアクセス可能であり得る（たとえばスマートフォンを介して）、又は患者もしくは患者の医療従事者によってアクセス可能であるサーバ上に遠隔で記憶され得ることを理解されたい。上記の記述から理解されるように、患者の差分遺伝子配列オブジェクトには任意の数の局所差分ストリング（すなわち、ゲノム中の特定の位置での、参照ゲノムに対する配列偏差）が含まれていてよく、局所差分ストリングは、ゲノムの定義された領域、１つもしくは複数の染色体、又はゲノム全体にわたってさえ位置し得る。同様に、差分遺伝子配列オブジェクトは、少なくとも２つの組織型（たとえば健康対疾患）、又は同じ組織において時間間隔を空けた少なくとも２つの結果（たとえば、特定のレジメンにおける特定の薬物を用いた処置の前及び処置が開始された後）を表す、複数の局所差分ストリングを含み得る。

【００８９】

したがって、かつ異なる観点から見ると、全ゲノム（又はその一部分［たとえば染色体もしくは近接する配列ストレッチ］）についての医学的に関連性のある情報は、１つ又は複数の局所差分ストリングを有する偏差記録として表すことができ、また、この情報を使用して、局所差分ストリングと関連する又はそれについての処置の選択肢、診断、及び／又は予後診断を含有するデータベースに対して比較することができることに注意されたい。複数の局所差分ストリングが存在する場合は、選択された局所差分ストリングの組合せは状態、素因、又は疾患の指標である場合があり、そのような複数の特異的局所差分ストリングの一群を使用して患者に特異的なデータを生成してよく、その後、これを使用して患者に特異的な指示を生成することに注意されたい。したがって、患者に特異的な指示の性質は相当に変動し、診断、予後診断、処置結果の予測、処置戦略の推奨、及び／又は処方であり得る。

【００９０】
企図される差分遺伝子配列オブジェクトのさらに別の好ましい使用では、本発明者らは、遺伝子分析は個体において可能であるだけでなく、本明細書中に提示するシステム及び方法を使用して、集団全体にわたる解析も迅速かつ有効な様式で実施することができることを発見した。たとえば、集団を解析する方法では、複数の差分遺伝子配列オブジェクト（たとえば複数人の個体用）を集団の医療記録データベースに記憶させ、解析エンジンが複数の差分遺伝子配列オブジェクト内の複数の局所差分ストリングの一群を同定して（たとえば、多型性、後成的変化などに基づく）一群記録を生成し、その後、これを使用して集団解析記録を生成する。
【００９１】
たとえば、一群記録を、血縁者、同じ民族又は人種のメンバー、同じ職業で働く集団、選択された地理的位置に住む集団について作成することができる。あるいは、集団は、病原体もしくは有害物質への曝露、既往歴、処置歴、処置の成功、性別、種、及び／又は年齢を共有するメンバーを有することによっても定義し得る。したがって、一群記録は、個体が一群記録によって定義される１つ又は複数の特定の群に属するという確認を可能にする、ゲノム全体の解析ツールであることを認識されたい。したがって、一群記録及び関連する方法は、一群記録に鑑みて、父系性又は母系性を決定するために有用であり得る、又は、患者に特異的な記録を生成するために有用であり得る。たとえば、患者に特異的な記録は、疾患もしくは状態に対する素因、又は特定の薬物もしくは他の薬剤に対する感度を明らかにし得る。したがって、患者に特異的な記録は、危険性評価及び／又は患者が特定集団に属するという確認を提示し得る。あるいは、患者に特異的な記録には、典型的には患者の一群記録と集団解析記録との比較に少なくとも部分的に基づく、診断、予後診断、処置結果の予測、処置戦略の推奨、及び／又は処方が含まれ得る。
【００９２】
企図される差分遺伝子配列オブジェクトのさらに好ましい使用では、参照差分遺伝子配列オブジェクトを（たとえば上記のようにコンセンサス記録として）生成し、データベースに記憶させる。その後、人の差分遺伝子配列オブジェクト中の複数の局所差分ストリングと参照差分遺伝子配列オブジェクト中の複数の局所差分ストリングとの間の偏差を決定してその人の個々の偏差記録を生成し、これを使用して人に特異的な偏差プロフィールを生成することができる。したがって、１つ又は複数の生理的パラメータ（たとえば医師によって指示される一般的な全血球数）を使用する代わりに、人の（好ましくは）全ゲノムの差分遺伝子配列オブジェクトを参照差分遺伝子配列オブジェクトと比較して、有意により包括的な情報採取に達する。最も典型的には、その後、人に特異的な偏差プロフィールを参照差分遺伝子配列オブジェクトの正常又は参照記録に対してマッチングさせて、特定の状態又は疾患に一致するとして人を正確かつ素早く同定する。
【００９３】
したがって、異なる観点から見ると、本明細書中に提示するシステム及び方法は、ゲノム、トランスクリプトーム、及び／又はプロテオームの修飾が少なくとも部分的な原因である疾患又は状態の診断又は解析において特に有用であることを理解されたい。他の疾患及び状態のうち、特に企図される疾患及び状態には、後天性免疫不全症候群（ＡＩＤＳ）、アジソン病、成人呼吸窮迫症候群、アレルギー、強直性脊椎炎、アミロイド症、貧血、喘息、アテローム性動脈硬化症、自己免疫性溶血性貧血、自己免疫性甲状腺炎、良性前立腺肥大、気管支炎、チェディアック－東症候群、胆嚢炎、クローン病、アトピー性皮膚炎、皮膚筋炎、真性糖尿病、気腫、胎児赤芽球症、結節性紅斑、萎縮性胃炎、糸球体腎炎、グッドパスチャー症候群、痛風、慢性肉芽腫性疾患、グレーブス病、橋本甲状腺炎、過好酸球増加症、過敏性腸症候群、多発性硬化症、重症筋無力症、心筋又は心膜の炎症、骨関節炎、骨粗鬆症、膵炎、多嚢胞性卵巣症候群、多発性筋炎、乾癬、ライター症候群、関節リウマチ、強皮症、重症複合型免疫不全症（ＳＣＩＤ）、シェーグレン症候群、全身性アナフィラキシー、全身性エリテマトーデス、全身性硬化症、血小板減少性紫斑病、潰瘍性大腸炎、ブドウ膜炎、ウェルナー症候群、癌、血液透析、及び体外循環の合併症、ウイル

ス、細菌、真菌、寄生生物、原虫、及び蠕虫の感染症；腺癌、白血病、リンパ腫、黒色腫、骨髄腫、肉腫、奇形癌、特に、副腎、膀胱、骨、骨髄、脳、乳房、子宮頸部、胆嚢、神経節、胃腸管、心臓、腎臓、肝臓、肺、筋肉、卵巣、膵臓、副甲状腺、陰茎、前立腺、唾液腺、皮膚、脾臓、精巣、胸腺、甲状腺、及び子宮の癌、アカシジア、アルツハイマー病、健忘症、筋萎縮性側索硬化症（ＡＬＳ）、運動失調、双極性障害、緊張病、脳性麻痺、脳血管疾患、クロイツフェルトーヤコブ病、認知症、鬱病、ダウン症候群、遅発性ジスキネジア、ジストニア、癲癇、ハンチントン病、多発性硬化症、筋ジストロフィー、神経痛、神経線維腫症、神経障害、パーキンソン病、ピック病、網膜色素変性症、統合失調症、季節性情動障害、老人性認知症、脳卒中、トゥレット症候群、ならびに特に脳の腺癌、黒色腫、及び奇形癌を含めた癌；腺癌、白血病、リンパ腫、黒色腫、骨髄腫、肉腫、奇形癌などの癌、特に、副腎、膀胱、骨、骨髄、脳、乳房、子宮頸部、胆嚢、神経節、胃腸管、心臓、腎臓、肝臓、肺、筋肉、卵巣、膵臓、副甲状腺、陰茎、前立腺、唾液腺、皮膚、脾臓、精巣、胸腺、甲状腺、及び子宮の癌；後天性免疫不全症候群（ＡＩＤＳ）、アジソン病、成人呼吸窮迫症候群、アレルギー、強直性脊椎炎、アミロイド症、貧血、喘息、アテローム性動脈硬化症、自己免疫性溶血性貧血、自己免疫性甲状腺炎、気管支炎、胆嚢炎、接触皮膚炎、クローン病、アトピー性皮膚炎、皮膚筋炎、真性糖尿病、気腫、リンパ球毒素を伴う偶発性リンパ球減少症、胎児赤芽球症、結節性紅斑、萎縮性胃炎、糸球体腎炎、グッドパスチャー症候群、痛風、グレーブス病、橋本甲状腺炎、過好酸球増加症、過敏性腸症候群、多発性硬化症、重症筋無力症、心筋又は心膜の炎症、骨関節炎、骨粗鬆症、膵炎、多発性筋炎、乾癬、ライター症候群、関節リウマチ、強皮症、シェーグレン症候群、全身性アナフィラキシー、全身性エリテマトーデス、全身性硬化症、血小板減少性紫斑病、潰瘍性大腸炎、ブドウ膜炎、ウェルナー症候群、癌、血液透析、及び体外循環の合併症、ウイルス、細菌、真菌、寄生生物、原虫、及び蠕虫の感染症、外傷、ブルトン型Ｘ連鎖無ガンマグロブリン血症、後天性免疫グロブリン血症（ＣＶＩ）、ディジョージ症候群（胸腺形成不全）、胸腺異形成、ＩｇＡ単独欠損症、重症複合型免疫不全症（ＳＣＩＤ）、血小板減少症及び湿疹を伴う免疫不全（ウィスコットーアルドリッチ症候群）、チェディアックー東症候群、慢性肉芽腫性疾患、遺伝性血管神経性浮腫、ならびにクッシング病関連の免疫不全などの免疫障害；腎尿細管性アシドーシス、貧血、クッシング症候群、軟骨形成不全性小人症、デュシェンヌ及びベッカー型筋ジストロフィー、癲癇、性腺形成不全、ＷＡＧＲ症候群（ウィルムス腫瘍、無虹彩症、泌尿生殖器奇形、及び精神遅滞）、スミスーマゲニス症候群、骨髄異形成症候群、遺伝性粘膜上皮異形成、遺伝性角皮症、シャルコーーマリーートゥース病及び神経線維腫症などの遺伝性神経障害、甲状腺機能低下症、水頭症、シデナム舞踏病及び脳性麻痺などの発作性疾患、二分脊椎、無脳症、頭蓋脊椎披裂、先天性緑内障、白内障、感音性難聴、ならびに対象の任意の組織、器官、又は系、たとえば、脳、副腎、腎臓、骨格又は生殖器系に関与する細胞の成長と分化、胚形成、及び形態形成に関連する任意の障害などの発達障害、ならびに性腺機能低下症、シーハン症候群、尿崩症、カルマン病、ハンドーシュラーークリスチャン病、レッテラーーシーベ病、サルコイドーシス、トルコ鞍空洞症候群、及び小人症を含めた、下垂体機能低下症に関連する障害などの内分泌障害；末端肥大症、巨人症、及び抗利尿ホルモン（ＡＤＨ）不適合分泌症候群（ＳＩＡＤＨ）を含めた下垂体機能亢進症；甲状腺種、粘液水腫、細菌感染症に関連する急性甲状腺炎、ウイルス感染症に関連する亜急性甲状腺炎、自己免疫性甲状腺炎（橋本病）、及びクレチン症を含めた、甲状腺機能低下症に関連する障害；甲状腺中毒症及びその様々な形態、グレーブス病、前脛骨粘液水腫、中毒性多結節性甲状腺種、甲状腺癌、及びプランマー病を含めた、甲状腺機能亢進症に関連する障害；コーン病（慢性高カルシウム血症）を含めた副甲状腺機能亢進に関連する障害；アレルギー、喘息、急性及び慢性の炎症性肺疾患、ＡＲＤＳ、気腫、肺鬱血及び肺浮腫、ＣＯＰＤ、間質性肺疾患、肺癌などの呼吸器疾患；腺癌、白血病、リンパ腫、黒色腫、骨髄腫、肉腫、奇形癌などの癌、特に、副腎、膀胱、骨、骨髄、脳、乳房、子宮頸部、胆嚢、神経節、胃腸管、心臓、腎臓、肝臓、肺、筋肉、卵巣、膵臓、副甲状腺、陰茎、前立腺、唾液腺、皮膚、脾臓、精巣、胸腺、甲状腺、及び子宮の癌；ならびに後天性免疫不全症候群（ＡＩＤＳ）、アジソ

ン病、成人呼吸窮迫症候群、アレルギー、強直性脊椎炎、アミロイド症、貧血、喘息、アテローム性動脈硬化症、自己免疫性溶血性貧血、自己免疫性甲状腺炎、気管支炎、胆嚢炎、接触皮膚炎、クローン病、アトピー性皮膚炎、皮膚筋炎、真性糖尿病、気腫、リンパ球毒素を伴う偶発性リンパ球減少症、胎児赤芽球症、結節性紅斑、萎縮性胃炎、糸球体腎炎、グッドパスチャー症候群、痛風、グレーブス病、橋本甲状腺炎、過好酸球増加症、過敏性腸症候群、多発性硬化症、重症筋無力症、心筋又は心膜の炎症、骨関節炎、骨粗鬆症、膵炎、多発性筋炎、乾癬、ライター症候群、関節リウマチ、強皮症、シェーグレン症候群、全身性アナフィラキシー、全身性エリテマトーデス、全身性硬化症、血小板減少性紫斑病、潰瘍性大腸炎、ブドウ膜炎、ウェルナー症候群、癌、血液透析、及び体外循環の合併症、ウイルス、細菌、真菌、寄生生物、原虫、及び蠕虫の感染症、外傷などの免疫障害が含まれる。

【００９４】

解析実施形態の例

図７～１０に関する以下の記述は、上述の解析の実施形態の例を提供する。

図７は、上記及び図８～１０に関して記述したようにさらなる解析に使用することができる、差分遺伝子配列オブジェクトを誘導する方法２００を例示する。方法２００は、遺伝子データベースへのアクセスを提供することを含むステップ２１０から始まる。好ましい遺伝子データベースは、少なくとも、１つの組織からの第１の遺伝子配列ストリング及び第２の場合によっては異なる組織からの第２の遺伝子配列ストリングを記憶する。それぞれの遺伝子配列ストリングは、好ましくは１つ又は複数の対応するサブストリングを含む。

【００９５】

ステップ２２０は、場合によってはネットワーク上又は１つもしくは複数の応用プログラムインタフェース（ＡＰＩ）を介して、遺伝子データベースと連結させた配列解析エンジンへのアクセスを提供するステップを含む。ステップ２３０は、好ましくは、解析エンジンが、対応するサブストリングのうちの１つの少なくとも１つの既知の位置を使用することによって第１及び第２の遺伝子配列ストリングを増分同期することによって局所アラインメントを生成することを含む。局所アラインメントの生成はいくつかの技法を使用して行うことができる。たとえば、ステップ２３１は、遺伝子配列ストリングのうちの１つ内の先験的に既知の位置に基づいてサブストリングのうちの少なくとも１つをアラインさせることを含むことができる。さらに、ステップ２３３は、サブストリングのうちの少なくとも１つについての既知の位置を含む既知の参照ストリングに基づいて、サブストリングをアラインさせることを含むことができる。さらに、ステップ２３５は、サブストリング自体の長さよりも短い長さのウィンドウ内でサブストリングをアラインさせることを含むことができる。さらに別の例は、ストリングのうちの少なくとも１つの全長の全体にわたって遺伝子配列ストリングを繰り返し増分同期させることを含むステップ２３７を含む。

【００９６】

局所アラインメントをどのように達成するかにかかわらず、方法２００は、ステップ２４０で解析エンジンが局所アラインメントを使用して局所アラインメント内の遺伝子配列ストリング間の局所差分ストリングを生成することによって続く。最後に、ステップ２５０で、解析エンジンは局所差分ストリングを使用して、差分配列データベース中の差分遺伝子配列オブジェクトを更新する。その後、差分遺伝子配列オブジェクトをさらなる再調査又は解析のために使用することができる。

【００９７】

たとえば、図８は、差分遺伝子配列オブジェクトに基づく医療サービスを提供する方法３００を例示する。ステップ３１０は、記憶装置（たとえば、ハードドライブ、ソリッドステートドライブ、ファイルシステム、携帯電話のメモリ、メモリカードなど）を備えた医療記録データベースと情報的に連結させた解析エンジンへのアクセスを提供することを含む。医療記録データベースは、好ましくは、１人又は複数人の患者の差分遺伝子配列オ

ブジェクトを記憶する。

【００９８】

ステップ３２０は、解析エンジンが患者の差分遺伝子配列オブジェクト中の局所差分ストリング又は一群の局所差分ストリングの存在を使用して患者に特異的なデータセットを生成することを含む。さらに、ステップ３３０での解析エンジンは、患者に特異的なデータセットに基づいて患者に特異的な指示を生成する。たとえば、解析エンジンは、患者に特異的なデータセット内の患者の局所差分ストリングの属性を、類似の差分ストリングを有する既知の状態と比較することができる。したがって、解析エンジンは、場合によっては診断、予後診断、処置結果の予測、処置戦略に関する推奨、危険性評価、処方、又は他の種の指示を含めた、１つ又は複数の患者に特異的な指示を生成することができる。

また、図９に例示するように、差分遺伝子配列オブジェクトは、集団を解析するための方法４００内でも使用することができる。ステップ４１０は、医療記録データベース中の差分遺伝子配列オブジェクトを得る又は記憶することを含み、医療記録データベースは人の集団全体にわたる情報を記憶する。医療記録データベース中の記録は、集団の属性（たとえば、人口統計学、民族性、病気、地理、労働条件、曝露など）に応じて構築されたクエリによって得ることができることを理解されたい。たとえば、ヨーロッパ系の郵便番号内に住むすべての男性を標的とするクエリを実行依頼することによって、差分遺伝子配列オブジェクトの結果組を作成することができる。好ましくは、医療記録データベースは解析エンジンと通信可能に連結されている。

【００９９】

ステップ４２０は、解析エンジンが複数の差分遺伝子配列オブジェクト内の一群の局所差分ストリングを同定することを含む。たとえば、一群には、特定の個体集団、場合によっては同じ地理的地域を訪れた個体の局所差分ストリングが含まれ得る。解析エンジンは、一群に関する情報を含む一群記録をさらに生成する。

ステップ４３０は、解析エンジンが一群記録を使用して集団解析記録を生成することを含み、これは１つ又は複数の出力装置上に提示させることができる。集団解析記録の例には、父系性又は母系性の確認、祖先情報、人口指標、又は他の集団情報が含まれ得る。

【０１００】

一部の実施形態では、方法４００は、解析エンジンが医療記録データベース内の患者に関連する差分遺伝子配列オブジェクトから誘導した個々の患者の一群記録を１つ又は複数の生成された集団解析記録と比較するステップ４４０を含む。したがって、患者の遺伝子状態を「正規化した」集団に対して比較することができる。さらに、ステップ４４５では、解析エンジンは情報から患者に特異的な記録を生成することができる。たとえば、患者に特異的な記録には、特定の集団の範囲内にある患者の危険性評価が含まれ得、又は既に記述した患者指示が含まれ得る。

差分遺伝子配列オブジェクトの別の使用が図１０の方法５００によって表されている。方法５００は、人の差分遺伝子配列オブジェクトを使用して、既知の参照に対する人に特異的な偏差プロフィールを誘導することを表している。ステップ５１０は、解析エンジンと通信可能に連結させた医療記録データベースに参照差分遺伝子配列オブジェクトを記憶させることを含む。参照差分遺伝子配列オブジェクトは、集団もしくは集団セグメントの統計的平均、正準の人、別の人、又は他の種の参照であり得る。

【０１０１】

ステップ５２０は、解析エンジンが人の差分遺伝子配列オブジェクトのうちの１つ又は複数と少なくとも１つの参照差分遺伝子配列オブジェクトとの間の偏差を計算することを含む。解析エンジンは、偏差を、偏差を説明する属性を含む偏差記録へとさらに変換することができる。偏差記録には、偏差の１つ又は複数の次元（たとえば、相違の数、相違の長さなど）に関連する情報が含まれ得ることを理解されたい。

ステップ５３０では、解析エンジンは偏差記録を使用して人に特異的な偏差プロフィールを生成する。解析エンジンは、望ましいフォーマットに従ってプロフィールを提示するように、１つ又は複数の計算装置をさらに構成することができる。一部の実施形態では、

偏差プロフィールは、素人が容易に読めるグラフ様式で人に提示することができる一方で、提示する情報は、遺伝学者、医師、保険会社、又は他の実体に提示した場合はより複雑であり得る。

本発明は、本発明の特定の態様及び実施形態を例示する目的のみで含まれ、限定するものとしては含まれない、以下の例を参照することにより容易に理解されるであろう。

【実施例】

【０１０２】

（例１）

参照ゲノムを介したデータセットの同期

すべてのショートリードを同じ参照ゲノムとアラインさせ、これにより、参照ゲノムが複数の関連する試料からの配列データを組織化するための自然な方法となる。ＢａｍＢａｍは、２つのショートリード配列決定データセット、すなわち、同じ患者からの、一方は腫瘍及び他方は一致した正常（「生殖系列」）ならびに参照ゲノムを取り込み、両方のデータセット中の同じゲノム位置に重複するすべての配列が同時に処理するために利用可能であるように、これらのデータセットを読み込む。これが、そのようなデータを処理するために最も効率的である一方で、それぞれのデータセットを単独で処理し、その後にのみ結果を統合する、連続様式では達成が困難又は不可能であろう複雑な解析を可能にする方法である。

そのような方法は、２つより多くの関連する配列決定データセットに容易に拡張可能である。たとえば、３つの試料、すなわち、一致した正常、腫瘍、及び再発を配列決定した場合、この方法を使用して腫瘍及び再発試料に特異的な変化ならびに再発のみに特異的な変化を検索することができ、再発腫瘍が、それが由来すると推定される元の腫瘍からいくらか変化していることが示唆される。また、この同じ方法を使用して、子供、父親、及び母親からの配列決定された試料を考慮して子供のゲノムの遺伝性の部分を決定することもできる。

【０１０３】

（例２）

体細胞及び生殖系列の変異コール

ＢａｍＢａｍは、ゲノム全体にわたる配列データを対のファイルで同期させて保持するため、腫瘍及び生殖系列の両方のＢＡＭファイルからの配列決定データならびにヒト参照を必要とする複雑な突然変異モデルを容易に実行することができる。このモデルは、生殖系列の遺伝子型（生殖系列の読取及び参照ヌクレオチドを考慮）ならびに腫瘍の遺伝子型（生殖系列の遺伝子型、単純突然変異モデル、腫瘍試料中の正常組織を汚染している割合の推定、及び腫瘍配列データを考慮）の両方の同時確率を最大にすることを目的とする。

【０１０４】

最適な腫瘍及び生殖系列の遺伝子型を見出すために、本発明者らは、

【数１】

$$P(D_g, D_t, G_g, G_t \mid \alpha, r) = P(D_g \mid G_g) P(G_g \mid r) P(D_t \mid G_g, G_t, \alpha) P(G_t \mid G_g) \quad (1)$$

［式中、ｒは観察された参照対立遺伝子であり、αは正常の汚染の割合であり、腫瘍及び生殖系列の遺伝子型はＧｔ＝（t_1，t_2）及びＧｇ＝（g_1，g_2）によって定義され、t_1，t_2，g_1，g_2∈｛Ａ，Ｔ，Ｃ，Ｇ｝である］によって定義される尤度を最大にすることを目的とする。腫瘍及び生殖系列配列データは、それぞれ一組のリード

【０１０５】

【数２】

$$D_t = \left\{ d_t^1, d_t^2, \cdots, d_t^m \right\}$$

及び

【数３】

$$D_g = \left\{ d_g^1, d_g^2, \cdots, d_g^n \right\}$$

として定義され、観察されたベースは

【数４】

$$d_t^i, d_g^i \in \{A, T, C, G\}$$

である。モデル中で使用するすべてのデータはユーザに定義されたベース及びマッピング品質閾値を上回っていなければならない。

【０１０６】

生殖系列の遺伝子型を考慮した生殖系列対立遺伝子の確率は、４つのヌクレオチドにわたる多項式としてモデリングされている：

【数５】

$$P(D_g|G_g) = \frac{n!}{{}^{n}A^{!n}T^{!n}G^{!n}C^{!}} \prod_i^n P(d_g^i|G_g),$$

［式中、ｎはこの位置での生殖系列のリードの合計数であり、n_A、n_G、n_C、n_Tは、それぞれの観察された対立遺伝子をサポートするリードである］。ベース確率

【数６】

$$P(d_g^i|G_g)$$

は独立していると仮定され、遺伝子型G_gによって表される２つの親対立遺伝子のどちらかに由来する一方で、シーケンサーのおおよそのベース誤差率も組み込まれている。生殖系列の遺伝子型に対する事前分布は、参照ベースを

【数７】

$$P(G_g|r=a) = \{\mu_{aa}, \mu_{ab}, \mu_{bb}\},$$

［式中、μ_{aa}はこの位置がホモ接合性の参照、μ_{ab}はヘテロ接合性の参照、μ_{bb}はホモ接合性の非参照である確率である］として条件づける。この時点で、生殖系列の事前分布は、既知の遺伝されたＳＮＰに関する情報をまったく取り込んでいない。

【０１０７】

腫瘍リードの組の確率は、再度、多項式

【数８】

$$P(D_t|G_t, G_g, \alpha) = \frac{n!}{{}^{n}A^{!n}T^{!n}G^{!n}C^{!}} \prod_i^n P(d_g^i|G_t, G_g, \alpha),$$

［式中、ｍはこの位置での生殖系列のリードの合計数であり、m_A、m_G、m_C、m_Tは、腫瘍データセット中のそれぞれの観察された対立遺伝子をサポートするリードである］として定義され、それぞれの腫瘍リードの確率は、腫瘍及び生殖系列の遺伝子型の両方から誘導されるベース確率の混合であり、これは、正常の汚染の割合αによって、

【数９】

$$P(d_t^i|G_t, G_g, \alpha) = \alpha P(d_t^i|G_t) + (1-\alpha) P(d_t^i|G_g)$$

として調節され、腫瘍の遺伝子型の確率は、生殖系列の遺伝子型からの単純な突然変異モデルによって定義される

【数１０】

$$P(G_t|G_g) = \max[P(t_1|g_1)P(t_2|g_2), P(t_1|g_2)P(t_2|g_1)]$$

［式中、突然変異なし（たとえば$t_1 = g_1$）の確率は最大限であり、転位（すなわちＡ→Ｇ、Ｔ→Ｃ）の確率は塩基転換（すなわちＡ→Ｔ、Ｔ→Ｇ）よりも４倍可能性が高い］。多項式分布のすべてのモデルパラメータ、α、μ_{aa}、μ_{ab}、μ_{bb}、及びベース確率

【数１１】

$$P(d^i|G)$$

はユーザが定義可能である。

【０１０８】

選択された腫瘍及び生殖系列の遺伝子型

【数１２】

$$G_t^{\max}, G_g^{\max},$$

は（１）を最大にするものであり、

【数１３】

$$\frac{P\left(D_g, D_t, G_g^{\max}, G_t^{\max} \middle| \alpha, \gamma\right)}{\sum_{i,j} P\left(D_g, D_t, G_g = i, G_t = j \middle| \alpha, \gamma\right)}$$

によって定義される事後確率を使用して、推量された遺伝子型の対における信頼性をスコアづけすることができる。腫瘍及び生殖系列の遺伝子型が異なる場合は、推定上の体細胞突然変異がそのそれぞれの信頼性と共に報告される。

【０１０９】

腫瘍及び生殖系列の遺伝子型の同時尤度を最大にすることは、特に一方又は両方の配列データセットが特定のゲノム位置の低いカバレッジを有する状況において、どちらの推量された遺伝子型の正確さも改善させるための助けとなる。ＭＡＱ及びＳＮＶＭｉｘなどの、単一の配列決定データセットを解析する他の突然変異コールアルゴリズムは、非参照又は突然変異対立遺伝子へのサポートが低い場合に誤りを起こす可能性がより高い（Li, H., et al. (2008) Mapping short DNA sequencing reads and calling variants using mapping quality scores, Genome Research, 11, 1851-1858、Goya, R. et al. (2010) SNVMix: predicting single nucleotide variants from next- generation sequencing of tumors, Bioinformatics, 26, 730-736）。

所定のゲノム位置でのすべてのリードから対立遺伝子のサポートを収集することに加えて、リードに関する情報を収集し（正方向又は逆方向のどちらの鎖にリードがマッピングされるか、リード内の対立遺伝子の位置、対立遺伝子の平均品質など）、偽陽性の呼び出しを選択的にフィルタ除去するために使用する。本発明者らは、変異をサポートする対立遺伝子のすべてについて鎖及び対立遺伝子の位置のランダム分布を予想し、分布がこのランダム分布から有意に非対称である（すなわち、すべての変異体対立遺伝子がリードの後部末端付近に見つかる）場合は、変異コールが疑わしいと示唆される。

【０１１０】

（例３）

全体的及び対立遺伝子特異的なコピー数

腫瘍又は生殖系列のどちらかのデータ中のカバレッジに応じてウィンドウのゲノム幅を拡大及び縮小させるダイナミックウィンドウ生成手法を使用して、全体的な体細胞コピー数を計算する。処理はゼロ幅のウィンドウで開始する。腫瘍又は生殖系列のどちらかの配列データからのそれぞれの固有のリードを腫瘍計数Ｎｔ又は生殖系列計数Ｎｇへと集計する。それぞれのリードの開始及び停止の位置がウィンドウの領域を定義し、新しいリードが現在のウィンドウの境界を越えるごとに拡大する。腫瘍又は生殖系列のどちらかの計数がユーザに定義された閾値を越えた際、ウィンドウサイズ及び位置、ならびにＮｔ、Ｎｇ、及び相対カバレッジＮｔを記録する。局所的なリードカバレッジに応じてＮｇウィンドウのサイズをあつらえることで、低いカバレッジの領域（たとえば反復領域）で大きなウィンドウ又は体細胞増幅を示す領域で小さなウィンドウが作成され、それにより、単位複製配列のゲノムの分解能が増加され、増幅の境界を定義する能力が増加される。

【０１１１】

示したように、生殖系列中でヘテロ接合性であるとみなされた位置のみを含める以外は同様に、対立遺伝子特異的なコピー数を計算する（図２を参照）。ヘテロ接合性とは、それぞれの親が１つの対立遺伝子に寄与する、２つの異なる対立遺伝子を有すると考えられる生殖系列中の位置として定義される。同じゲノム近隣中のデータを総計するために、全体的なコピー数について上述したものと同じダイナミックウィンドウ生成技法を使用して

多数(majority)及び少数(minority)のコピー数を計算する。本明細書中では、ヘテロ接合性部位での多数の対立遺伝子は、そのゲノム位置に重複する腫瘍データセット中に最大数のサポートリードを有する対立遺伝子であると定義される一方で、少数の対立遺伝子は最も少ないサポートを有する対立遺伝子であると定義される。腫瘍及び生殖系列の両方のデータ中で多数の対立遺伝子に帰されるすべての計数は多数コピー数の計算に数えられ、少数の対立遺伝子についても同様である。その後、多数及び少数の対立遺伝子の計数を生殖系列データ中の両対立遺伝子の計数Ｎｇによって正規化して、多数及び少数のコピー数を計算する。

対立遺伝子特異的なコピー数を使用して、ヘテロ接合性の喪失（コピーの中立及びコピーの欠失のどちらも）ならびに単一の対立遺伝子に特異的な増幅又は欠失を示すゲノム領域を同定する。この最後の点は、潜在的に疾患を引き起こす対立遺伝子を、腫瘍配列データ中で増幅されるもの又は欠失していないもののどちらかとして区別することを助けるために、特に重要である。さらに、ヘミ接合性の欠失を経験する領域（たとえば１つの親染色体アーム）を使用して、配列決定された腫瘍試料中の正常な汚染物質の量を直接推定することができ、これは、上述の生殖系列及び腫瘍の遺伝子型モデリングを改善させるために使用することができる。

【０１１２】

図２は対立遺伝子特異的なコピー数の計算の全体像を示す。ヘテロ接合性の遺伝子型を有する位置は、生殖系列の変異コールアルゴリズムによって決定される生殖系列及び腫瘍の配列決定データをどちらも使用して決定する。これらの位置に重複するすべての富取を収集し、ヘテロ接合性の遺伝子型中の２つの対立遺伝子のそれぞれのリードサポートが腫瘍及び生殖系列のどちらにも見つかる。多数の対立遺伝子は最も高いサポートを有する対立遺伝子であると決定され、多数のコピー数は、生殖系列中のその位置での全体的なリードの数によって正規化することによって計算する。

【０１１３】

（例４）

遺伝子型のフェージング

ＢａｍＢａｍは、腫瘍中の大スケールのゲノムの増幅又は欠失によって引き起こされる対立遺伝子不均衡を利用することによって、生殖系列中に見つかるすべてのヘテロ接合性の位置をフェージングすることを試みる。多数票ベースのコールを腫瘍配列データ中のすべての位置で選択して、腫瘍中に存在するフェージングしたハプロタイプを構築する。多数票はショートリードのプール中で観察される最も豊富な対立遺伝子を選択し、欠失事象後に腫瘍中に留まる対立遺伝子又は増幅事象の複製された対立遺伝子が選択されるはずである。それぞれの位置において、生殖系列の対立遺伝子の状態も同定し、位置は、必須のリードサポートを有する対立遺伝子が１つしか存在しない場合はホモ接合性、少なくとも２つの対立遺伝子が必要なリードサポートを有する場合はヘテロ接合性であるとみなされる。腫瘍のハプロタイプは２つの親ハプロタイプのうちの一方を表すと仮定され、２つ目の親ハプロタイプは、腫瘍ハプロタイプに属さない生殖系列対立遺伝子の配列として生じる。この手順は、腫瘍中の対立遺伝子の割合にかかわらずゲノム全体で使用するため、本発明者らは、遺伝子型のハプロタイプの割当ては、多数と少数の対立遺伝子との間で均等にバランスがとれている領域中で本質的にランダムであると予想する。生殖系列配列の正確なフェージングは、腫瘍中の単一のゲノム事象（たとえば局所の増幅又は欠失）から生じる一貫した対立遺伝子不均衡を示す領域中でのみ起こる。腫瘍由来のハプロタイプの妥当性確認は、腫瘍由来のハプロタイプをＨａｐＭａｐプロジェクトから入手可能なフェージングした遺伝子型と比較することによって達成することができる（International HapMap Consortium (2007), Nature, 7164: 851-861）。

【０１１４】

（例５）

ペアエンドクラスタリングを使用した構造的変動の推量

推定上の染色体内及び染色体間の再編成を同定するために、ＢａｍＢａｍは、ペア中の

それぞれのリードが参照配列の本質的に異なる領域にマッピングされる、不調和ペアリードを検索する。染色体内の不調和ペアは、異常に大きな挿入サイズを有する（すなわち、ペアリードを隔てる参照上のゲノムの距離がユーザに定義された閾値を越える）もの、又は不正確な配向（すなわち反転）でマッピングされるものである。染色体間の不調和ペアは、異なる染色体にマッピングされるペアリードによって定義される。ショートリードライブラリの調製におけるＰＣＲ増幅ステップの結果でしかない多数のリードによってサポートされる再編成のコールを回避するために、他のペアとして同一の位置とアラインするすべての不調和ペアエンドリードを除去する。このプロセスの全体像を図３に示す。

【０１１５】

すべての不調和ペアエンドリードを、そのゲノム位置に応じてクラスタリングして、ブレイクポイントが存在すると考えられる大体のゲノム領域を定義する。総計処理は、推定上のブレイクポイントの両側で他のリードと重複する固有のリードを一緒にグループ化することからなる。すべての重複するリードの鎖の配向は一致しなければならない、又はペアのクラスターが含まれていてはならない。クラスター中の重複する不調和ペアの数がユーザに定義された閾値を越えた際に、再編成を説明するブレイクポイントが定義される。生殖系列及び腫瘍のどちらのデータセット中にも同じ位置に再編成が存在する場合は、これらを以下のように比較する。生殖系列中で観察される構造的変動が何らかの方法で腫瘍中で反転されて参照と正確に一致していることは甚だしいため、生殖系列の再編成は、腫瘍及び生殖系列のデータセットが同じ再編成をサポートすることを必要とする。他方で、体細胞の再編成は、腫瘍配列決定データ中のみでしか観察される必要がなく、生殖系列データセット中に実質的に存在しない。これらの要件を満たす再編成は後処理解析及び可視化のために記憶される一方で、満たさないものは、配列決定機器、試料調製（全ゲノム増幅など）、又は用いたショートリードマッピングアルゴリズムの系統的バイアスのいずれかによって引き起こされた人為再編成として廃棄される。

【０１１６】

図３は構造的変異コールの全体像を示す。推定上の構造的変異の初期同定は、ＢａｍＢａｍによって、どちらのリードも参照ゲノムに完全にマッピングされるが、異常な非参照的な様式でマッピングされる、不調和にマッピングされたリードペアを使用して同定する。その後、ＢａｍＢａｍによって見つかった推定上のブレイクポイントを、ｂｒｉｄｇｅｔと呼ばれるプログラムによって、任意の利用可能なスプリットリードを使用して洗練される。

【０１１７】

（例６）

スプリットリードを使用した構造的変動の洗練

ＢａｍＢａｍによって最初に見つけられたブレイクポイントは、その性質によりブレイクポイントの実際の接合点と重複することができない完全にマッピングされたリードを使用するという点で、かつリードは参照（又は、体細胞の再編成の場合は生殖系列データセット）中に存在しない配列を表すため、概算である。ブレイクポイントの位置の知識を洗練させるためにＢｒｉｄｇｅｔと呼ばれるプログラムを開発し、これは図４に要約されている。

【０１１８】

Ｂｒｉｄｇｅｔは、ＢａｍＢａｍによって見つけられたブレイクポイントの概算が与えられ、完全にマッピングされたメイトによって推定上のブレイクポイント付近に固定されているすべてのアラインしていないリードを検索する。これらのマッピングされていないリードのそれぞれが、再編成のブレイクポイントの接合点に重複する「スプリットリード」となる潜在性を有する。ブレイクポイントの両側を取り囲む局在ゲノム配列を固有のタイル（現在のタイルサイズ＝１６ｂｐ）の組へと分割し、参照ゲノム中でのタイルの配列及びその位置タイルデータベースを構築する。それぞれのアラインしていないリードについて、リードを同じサイズのタイルへと分割し、リード内でのその位置を注記することによって、同様のタイルデータベースを構築する。参照タイルデータベースとアラインして

いないタイルデータベースとを比較して、参照中でのそれぞれのアラインしていないタイルのゲノム位置が決定される。参照及びアラインしていないリードのどちらにおいても近接しているタイルの最大組を、ブレイクポイントの各側それぞれに１つ決定することによって、これらの位置の「二重スパニング組」を計算する。

【０１１９】

参照座標中での「二重スパニング組」の最小及び最大のゲノム位置は、ブレイクポイントの位置及び配列の配向（又は鎖状）を正確に決定する。ブレイクポイントの左及び右の境界を説明する情報を用いて、再編成された配列が完全に定義される、すなわち、左側は（染色体＝ｃｈｒ１、位置＝１０００ｂｐ、鎖＝正方向）によって定義され、右側は（染色体＝ｃｈｒ５、位置＝５００，０００ｂｐ、鎖＝逆）によって定義される。また、ブレイクポイントの配列相同性（すなわち、ブレイクポイントのどちらの境界でも同一であると観察されるが、アラインしたリード中では２つの配列の接合点で１回のみ観察される、「ＣＡ」などの短い配列）もこれらの二重スパニング組から決定される。

それぞれのアラインしていないリードについて、二重スパニング組はブレイクポイントの潜在的な位置を決定する。それぞれのアラインしていないリードはブレイクポイントについてわずかに異なる位置を決定し得るため（ブレイクポイント付近の配列エラー、反復参照などに起因）、二重スパニング組から決定されたすべてのブレイクポイント位置を使用して可能な接合点配列を生成する。すべてのマッピングされていないリードをこれらの可能な接合点配列のそれぞれに対して新たにアラインさせ、そのアラインメントにおける全体的な改善を、リードがどのくらい良好に元の配列とアラインしたかに対して比較する。アラインメントスコアの最大の改善を与える接合点配列が、真の再編成の最良の候補として判断される。この最良の接合点配列がアラインメントスコア皆無又はそれに近い改善しかもたらさない場合は、この接合点配列は真の再編成を表している可能性が低いため廃棄する。この場合、スプリットリード確認の欠如は、ＢａｍＢａｍによって見つけられた元の構造的再編成が人為的なものである可能性の証拠であることも決定され得る。

【０１２０】

図４は、構造的再編成が起こったゲノムの位置を正確に同定するための例示的な方法を示す。タイル（又はｋｍｅｒ）を潜在的なスプリットリード及び参照ゲノムのどちらについても決定する。二重スパニング組を決定し（この図の下部の濃赤及び紫色のボックスとして表す）、これは再編成された配列をどのように構築するかを完全に定義している。二重スパニング組はスプリットリード中の配列エラー又はＳＮＰに対して強い。

【０１２１】

（例７）

腫瘍特異的ゲノムブラウザ

ＢａｍＢａｍによって出力されたすべての結果を可視化するために、図５に示すように、単一の腫瘍試料中に見つかったすべてのゲノム変異体を、その一致した正常に対して同時に表示する、腫瘍ゲノムブラウザを開発した。これは、全体的及び対立遺伝子特異的なコピー数、染色体内及び染色体間の再編成、ならびに突然変異及び小さなインデルを表示することができる。これは、データを直鎖状及び環状のどちらのプロットでも示し、染色体間の再編成を表示するためには後者がはるかにより良好に適している。

データを１つの画像で一緒に表示することによって、ユーザは単一の試料のデータを素早くナビゲートし、コピー数の変化と構造的変動との間の関係性を理解することができる。たとえば、大きな染色体内欠失型の再編成は、ブレイクポイントの間の領域中にコピー数にそれと調和した低下を有するはずである。また、突然変異データをコピー数データと共に表示することで、ユーザは、体細胞突然変異が続いて増幅されたかどうか、又は野生型対立遺伝子が腫瘍中で欠失していたかどうかを理解することができ、どちらの重要なデータ点も、この試料の腫瘍化におけるゲノム座位の重要性を示唆している。

【０１２２】

図５は例示的な腫瘍特異的ゲノムブラウザを示す。ブラウザは、ＢａｍＢａｍによって発見されたすべての高レベルの体細胞相違を１つの画像中で示し、複数の明確なデータセ

ットを合成して、腫瘍のゲノムの全体像を与えることを可能にする。ブラウザは、ゲノム領域を迅速にズームイン及びズームアウトすることができ、わずか数クリックで上記に示した全ゲノムビューから単一ベース分解能へと移行する。

【０１２３】

（例８）

コンピュータ要件

ＢａｍＢａｍ及びＢｒｉｄｇｅｔはどちらもＣで書かれており、標準のＣライブラリ及び最新のＳＡＭｔｏｏｌｓソースコード（ｈｔｔｐ：／／ｓａｍｔｏｏｌｓ．ｓｏｕｒｃｅｆｏｒｇｅ．ｎｅｔから入手可能）のみを必要とする。単一の処理として、又はクラスター全体にわたる一連のジョブに分割して（たとえば１つの染色体あたり１つのジョブで）実行し得る。それぞれが数十億個の１００ｂｐのリードを含有する一対の２５０ＧＢのＢＡＭファイルを処理するＢａｍＢａｍは、その全ゲノム解析を単一の処理として約５時間、又は控えめなクラスター（２４ノード）では約３０分間で終える。ＢａｍＢａｍのコンピュータ要件はごくわずかであり、単一のゲノム位置に重複するリードのデータを記憶するために十分なＲＡＭ、及び腫瘍又は生殖系列のゲノムのどちらかに見つかる良好にサポートされた変異を記憶するために十分なディスク空き容量のみを必要とする。

Ｂｒｉｄｇｅｔも非常に控えめなコンピュータ要件を有する。一台のマシン上での実行時間は典型的には１秒未満であり、これには、参照配列及びブレイクポイント近隣のすべての潜在的なスプリットリードを集め、参照及びスプリットリードの両方のタイルデータベースを構築し、すべての二重スパニング組を決定し、潜在的な接合点配列を構築し、すべてのスプリットリードを参照及びそれぞれの接合点配列の両方に対して再度アラインさせ、最良の接合点配列を決定するために必要な時間が含まれる。高度に増幅された又は多数のマッピングされていないリードを有する領域はＢｒｉｄｇｅｔの実行時間を増やすが、これはＢｒｉｄｇｅｔの容易な並行化能力によって軽減し得る。

【０１２４】

（例９）

ゲノムＤＮＡの単離

血液又は他の組織試料（２～３ｍｌ）を患者から採取し、ＥＤＴＡ含有チューブ中、－８０℃で使用時まで保管する。ゲノムＤＮＡを血液試料から、製造者の指示に従って、ＤＮＡ単離キットを使用して抽出する（ＰＵＲＥＧＥＮＥ、Ｇｅｎｔｒａ　Ｓｙｓｔｅｍｓ、ミネソタ州Ｍｉｎｎｅａｐｏｌｉｓ）。ＤＮＡの純度を、ベックマン分光光度計で測定した２６０及び２８０ｎｍでの吸光度の比（１ｃｍ光路、A_{260}／A_{280}）として測定する。

【０１２５】

（例１０）

ＳＮＰの同定

患者のＤＮＡ試料からの一遺伝子の領域を、ＰＣＲによって、その領域用に特異的に設計されたプライマーを使用して増幅する。上述のように当業者に周知の方法を使用して、ＰＣＲ産物を配列する。配列追跡において同定されたＳＮＰはＰｈｒｅｄ／Ｐｈｒａｐ／Ｃｏｎｓｅｄソフトウェアによって確認し、ＮＣＢＩ　ＳＮＰデータバンクに受託された既知のＳＮＰと比較する。

【０１２６】

（例１１）

統計分析

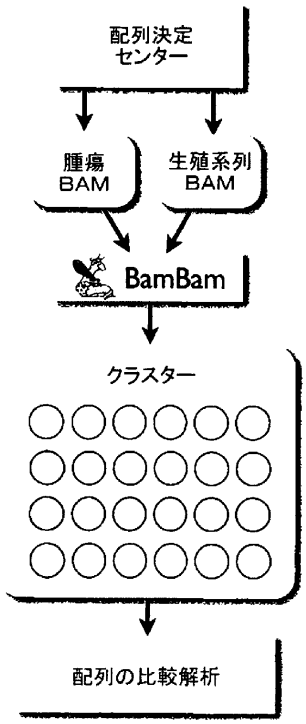
値は平均±ＳＤとして表す。χ^2解析（ウェブキー二乗計算機（Web Chi Square Calculator）、Georgetown Linguistics、ジョージタウン大学、ワシントンＤＣ）を使用して、正常対象と障害を有する患者との間の遺伝子型頻度の相違を評価する。事後解析を伴った一方向ＡＮＯＶＡを示したように行って、様々な患者群間の血行動態を比較する。

【０１２７】

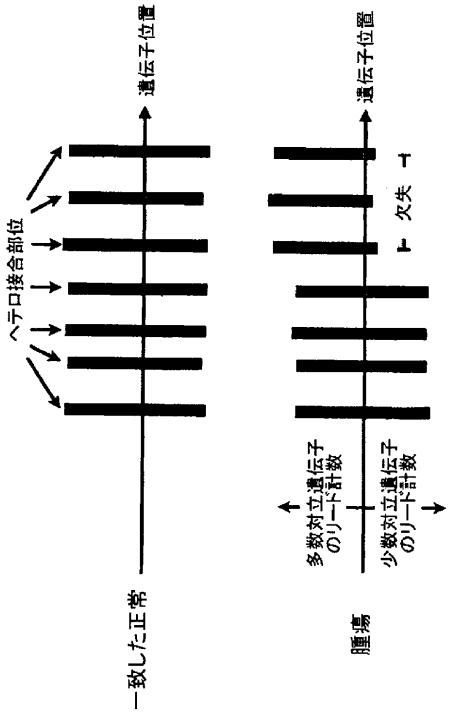
当業者には、本明細書中に記載の本発明の概念から逸脱せずに、既に記載したもの以外

にさらに多くの改変が可能であることが明らかであろう。したがって、本発明の主題は、添付の特許請求項の範囲の範囲以外に限定されるものではない。さらに、明細書及び特許請求項の範囲を解釈するにあたって、すべての用語はコンテキストと一貫した最大限可能な様式で解釈されるべきである。具体的には、用語「含む（comprises）」及び「含むこと（comprising）」とは、要素、構成要素、又はステップに非排他的な様式で言及すると解釈されるべきであり、これは、言及した要素、構成要素、もしくはステップが存在し得るもしくはそれを利用し得る、又は明確に言及していない他の要素、構成要素、もしくはステップと組み合わせ得ることを示す。明細書の特許請求項がＡ、Ｂ、Ｃ．．．及びＮからなる群から選択されるもののうちの少なくとも１つに言及する場合は、文章は、群からの１つの要素のみを必要とし、Ａ＋Ｎ又はＢ＋Ｎなどではないと解釈されるべきである。

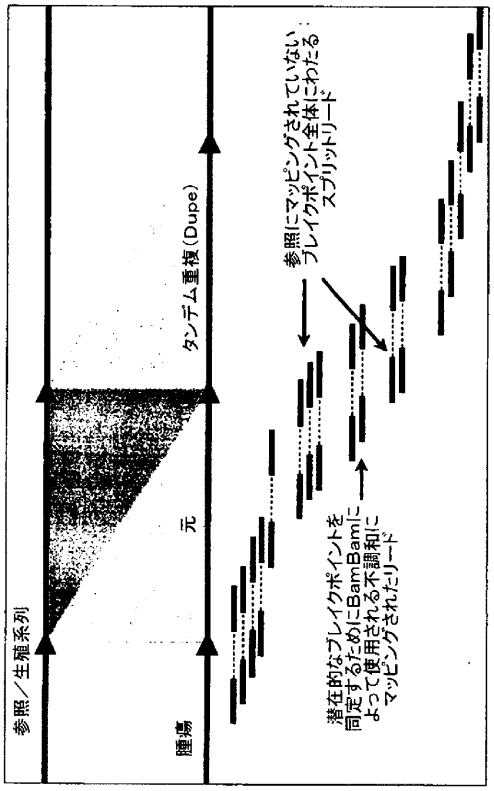
【図１】



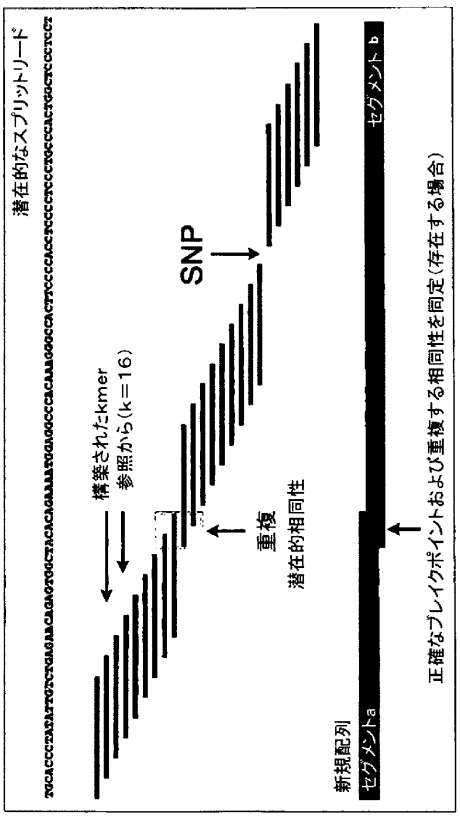
【図２】



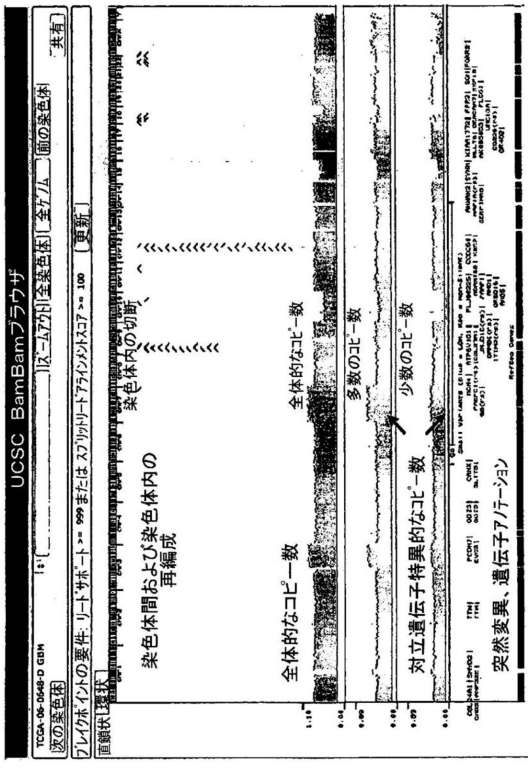
【図３】


参照／生殖系列
腫瘍
元
タンデム重複(Dupe)
参照にマッピングされていない
ブレイクポイント全体にわたる
スプリットリード
潜在的なブレイクポイントを
同定するためにBamBamに
よって使用されるが不調和に
マッピングされたリード

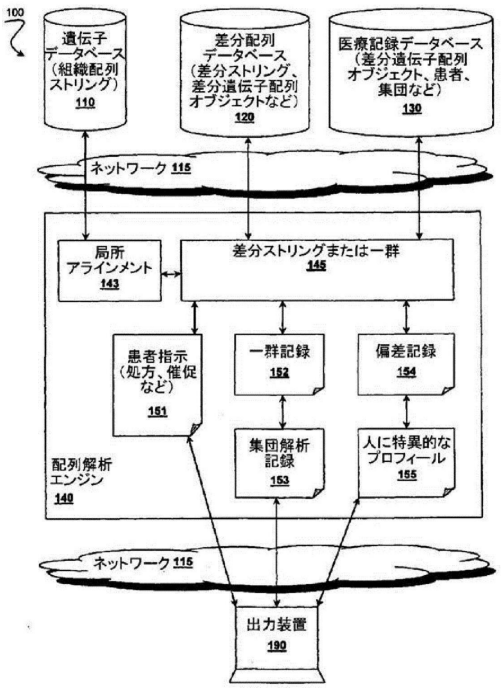
【図４】


潜在的なスプリットリード
構築されたkmer
参照から(k=16)
SNP
重複
潜在的相同性
新規配列
セグメントa
セグメントb
正確なブレイクポイントおよび重複する相同性を同定(存在する場合)

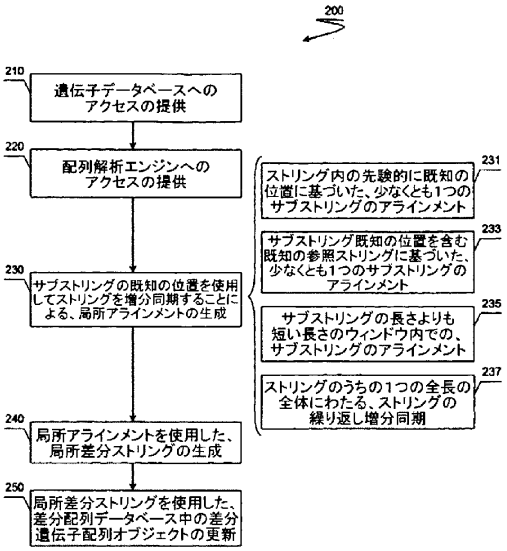
【図５】


UCSC BamBamブラウザ
染色体間および染色体内の
再編成
染色体内の切断
全体的なコピー数
多数のコピー数
少数のコピー数
対立遺伝子特異的なコピー数
突然変異、遺伝子アノテーション

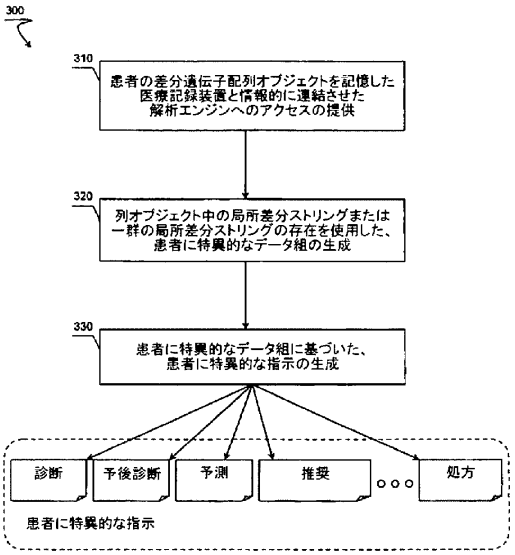
【図６】


100
遺伝子
データベース
(組織配列
ストリング)
110
差分配列
データベース
(差分ストリング、
差分遺伝子配列
オブジェクトなど)
120
医療記録データベース
(差分遺伝子配列
オブジェクト、患者、
集団など)
130
ネットワーク 115
局所
アラインメント
143
差分ストリングまたは一群
145
患者指示
(処方、催促
など)
161
一群記録
152
偏差記録
154
集団解析
記録
153
人に特異的な
プロフィール
155
配列解析
エンジン
140
ネットワーク 115
出力装置
190

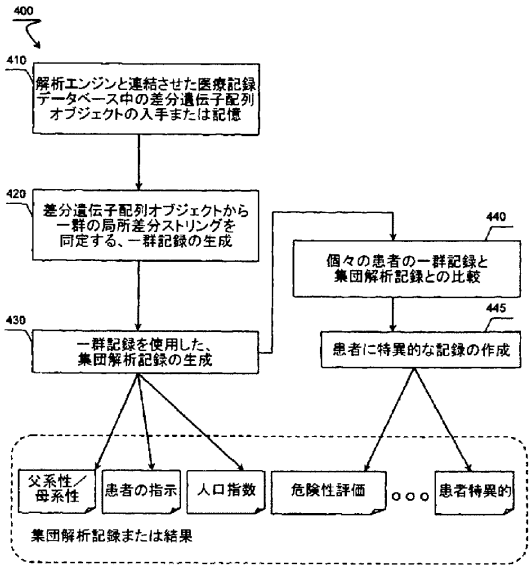
【図７】


200
210
遺伝子データベースへのアクセスの提供
220
配列解析エンジンへのアクセスの提供
230
サブストリングの既知の位置を使用してストリングを増分同期することによる、局所アラインメントの生成
240
局所アラインメントを使用した、局所差分ストリングの生成
250
局所差分ストリングを使用した、差分配列データベース中の差分遺伝子配列オブジェクトの更新
231
ストリング内の先験的に既知の位置に基づいた、少なくとも1つのサブストリングのアラインメント
233
サブストリング既知の位置を含む既知の参照ストリングに基づいた、少なくとも1つのサブストリングのアラインメント
235
サブストリングの長さよりも短い長さのウィンドウ内での、サブストリングのアラインメント
237
ストリングのうちの1つの全長の全体にわたる、ストリングの繰り返し増分同期

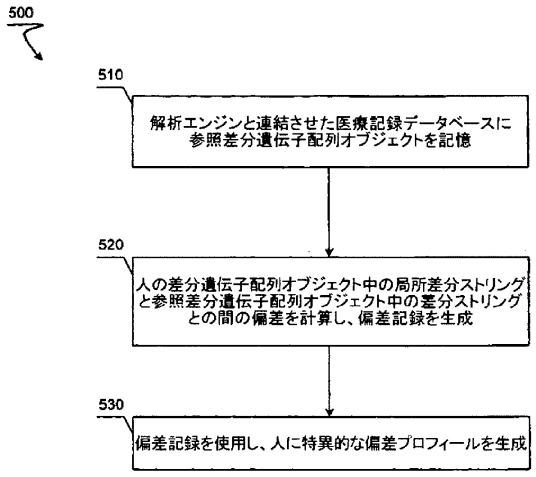
【図８】


300
310
患者の差分遺伝子配列オブジェクトを記憶した医療記録装置と情報的に連結させた解析エンジンへのアクセスの提供
320
列オブジェクト中の局所差分ストリングまたは一群の局所差分ストリングの存在を使用した、患者に特異的なデータ組の生成
330
患者に特異的なデータ組に基づいた、患者に特異的な指示の生成
診断
予後診断
予測
推奨
処方
患者に特異的な指示

【図９】


400
410
解析エンジンと連結させた医療記録データベース中の差分遺伝子配列オブジェクトの入手または記憶
420
差分遺伝子配列オブジェクトから一群の局所差分ストリングを同定する、一群記録の生成
430
一群記録を使用した、集団解析記録の生成
440
個々の患者の一群記録と集団解析記録との比較
445
患者に特異的な記録の作成
父系性／母系性
患者の指示
人口指数
危険性評価
患者特異的
集団解析記録または結果

【図１０】


500
510
解析エンジンと連結させた医療記録データベースに参照差分遺伝子配列オブジェクトを記憶
520
人の差分遺伝子配列オブジェクト中の局所差分ストリングと参照差分遺伝子配列オブジェクト中の差分ストリングとの間の偏差を計算し、偏差記録を生成
530
偏差記録を使用し、人に特異的な偏差プロフィールを生成

フロントページの続き

(74)代理人 100160923
弁理士 山口 裕孝
(74)代理人 100119507
弁理士 刑部 俊
(74)代理人 100142929
弁理士 井上 隆一
(74)代理人 100148699
弁理士 佐藤 利光
(74)代理人 100128048
弁理士 新見 浩一
(74)代理人 100129506
弁理士 小林 智彦
(74)代理人 100205707
弁理士 小寺 秀紀
(74)代理人 100114340
弁理士 大関 雅人
(74)代理人 100114889
弁理士 五十嵐 義弘
(74)代理人 100121072
弁理士 川本 和弥
(72)発明者 サンボーン ジョン ザカリー
アメリカ合衆国 カリフォルニア州 ９５０６４ サンタ クルーズ ユニバーシティ オブ カリフォルニア サンタ クルーズ内
(72)発明者 ハウスラー ディヴィッド
アメリカ合衆国 カリフォルニア州 ９５０６４ サンタ クルーズ ユニバーシティ オブ カリフォルニア サンタ クルーズ内

審査官 塩田 徳彦

(56)参考文献 特開２０１０－２０４８３８（ＪＰ，Ａ）
特表２００３－５２７８５５（ＪＰ，Ａ）
特表２００４－５０１６６９（ＪＰ，Ａ）
特表２００８－５２６７７５（ＪＰ，Ａ）
米国特許出願公開第２００６／０２７１３０９（ＵＳ，Ａ１）
米国特許出願公開第２００４／０１５３２５５（ＵＳ，Ａ１）
Aaron R. Quilan et al., BEDTools: a flexible suite of utilities for comparing genomic features, Bioinformatics, ２０１０年 3月, Vol.26 No.6, p.841-842

(58)調査した分野(Int.Cl., ＤＢ名)
Ｇ１６Ｂ ５／００ － ９９／００