(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2017-519488
(P2017-519488A)
(43)公表日　平成29年7月20日(2017.7.20)

(51)Int.Cl.		F I			テーマコード（参考）
C12Q 1/68	***(2006.01)***	C12Q	1/68	A	4B063
G01N 33/53	***(2006.01)***	C12Q	1/68	Z	
G01N 33/574	***(2006.01)***	G01N	33/53	M	
C12N 15/09	*(2006.01)*	G01N	33/574	Z	
		C12N	15/00	A	

審査請求 未請求　予備審査請求 未請求　（全 199 頁）

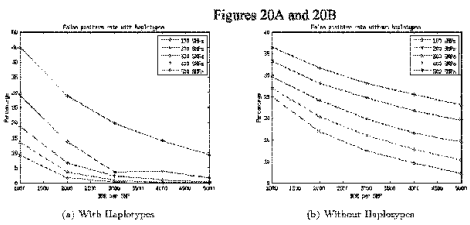
(21)出願番号　特願2016-563812 (P2016-563812)
(86)(22)出願日　平成27年4月21日 (2015.4.21)
(85)翻訳文提出日　平成28年12月20日 (2016.12.20)
(86)国際出願番号　PCT/US2015/026957
(87)国際公開番号　WO2015/164432
(87)国際公開日　平成27年10月29日 (2015.10.29)
(31)優先権主張番号　61/982,245
(32)優先日　平成26年4月21日 (2014.4.21)
(33)優先権主張国　米国(US)
(31)優先権主張番号　61/987,407
(32)優先日　平成26年5月1日 (2014.5.1)
(33)優先権主張国　米国(US)
(31)優先権主張番号　62/066,514
(32)優先日　平成26年10月21日 (2014.10.21)
(33)優先権主張国　米国(US)

(71)出願人　513156537
ナテラ，　インコーポレイテッド
アメリカ合衆国　カリフォルニア　９４０７０，　サン　カルロス，　インダストリアル　ストリート　２０１，　スイート　４１０
(74)代理人　100099759
弁理士　青木　篤
(74)代理人　100077517
弁理士　石田　敬
(74)代理人　100087871
弁理士　福本　積
(74)代理人　100087413
弁理士　古賀　哲次

最終頁に続く

(54)【発明の名称】変異の検出および染色体分節の倍数性

(57)【要約】
本発明は、アレルの相決定と、個別及び同時確率並びに最適合モデルの決定とに基づいて、染色体文節又は全長染色体の倍数性を検出するための、方法、システム、及びコンピュータ可読媒体を提供する。ある側面によれば、本発明は、妊娠中の胎児における癌又は染色体異常を検出するための、方法、システム、及びコンピュータ可読媒体を提供する。また、本発明は、倍数性決定により見いだされる多型座位のアレル不均衡率に基づいて、循環腫瘍核酸を検出するための方法を提供する。また、本発明は、増幅効率及びエラー率の推定値に基づいて一塩基多様体を検出するための方法と、複数の個体由来の対照試料の多様性アレル頻度の中央値に基づいて一塩基多様体を検出するための方法を提供する。
【選択図】図２０Ａ－２０Ｂ



【特許請求の範囲】
【請求項１】
個体の試料に含まれる染色体分節の倍数性を決定する方法であって、該方法は、
ａ．該試料中の該染色体分節にある多型座位の集合のうちの各座位に存在する各アレルの量を含むアレル頻度データを受け、
ｂ．該アレル頻度データの相を推定することによって該多型座位の集合に関する相決定済み（phased）アレル情報を作成し、
ｃ．該アレル頻度データを用いて、異なる倍数性状態に関する該多型座位のアレル頻度の個別確率を作成し、
ｄ．該個別確率と該相決定済みアレル情報を用いて該多型座位の集合に関する同時確率を作成し、
ｅ．該同時確率に基づき、染色体倍数性を示す最適合モデルを選択することにより、該染色体分節の倍数性を決定すること
とを含む、方法。
【請求項２】
前記アレル頻度データが核酸配列データから作成される、請求項１に記載の方法。
【請求項３】
前記アレル頻度データの誤差を修正し、修正済みアレル頻度データを前記個別確率の作成工程に使用することとをさらに含む、請求項１に記載の方法。
【請求項４】
修正される前記誤差がアレル増幅効率の偏りである、請求項３に記載の方法。
【請求項５】
前記個別確率が、前記多型座位の集合に関する異なる倍数性状態とアレル不均衡率の両方のモデルのセットを用いて作成される、請求項１に記載の方法。
【請求項６】
前記同時確率が、前記染色体分節にある多型座位間の連鎖を考慮することによって作成される、請求項１～５のいずれか一項に記載の方法。
【請求項７】
個体の試料の染色体倍数性を検出するための方法であって、該方法は、
ａ．該個体の染色体分節にある多型座位の集合に存在するアレルに関する核酸配列データを受け、
ｂ．該核酸配列データを用いて該多型座位の集合におけるアレル頻度を検出し、
ｃ．検出された該アレル頻度のアレル増幅効率の偏りを修正することによって、該多型座位の集合に関する修正済みアレル頻度を作成し、
ｄ．該核酸配列データの相を推定することによって該多型座位の集合に関する相決定済みアレル情報を作成し、
ｅ．該修正済みアレル頻度を該多型座位の集合に関する異なる倍数性状態とアレル不均衡率のモデルのセットと比較することにより、異なる倍数性状態に関する該多型座位のアレル頻度の個別確率を作成し、
ｆ．該個別確率を該染色体分節にある多型座位間の連鎖を考慮しながら組み合わせることにより、該多型座位の集合に関する同時確率を作成し、
ｇ．該同時確率に基づき、染色体異数性を示す最適合モデルを選択すること
とを含む、方法。
【請求項８】
前記染色体分節の第一の相同体と該染色体分節の第二の相同体の倍数性状態をモデリングすることによって個別確率を作成する、請求項１～５および７のいずれか一項に記載の方法であって、該倍数性状態が、
（１）いずれの細胞も該染色体分節の該第一の相同体または該第二の相同体の欠失または増幅を有していない状態、
（２）少なくとも一部の細胞が該染色体分節の該第一の相同体の欠失、または該第二の相

同体の増幅を有する状態、および
（３）少なくとも一部の細胞が該染色体分節の該第二の相同体の欠失、または該第一の相同体の増幅を有する状態
である、方法。
【請求項９】
修正されるエラーが環境不純物および遺伝型解析による不純物を含む、請求項７に記載の方法。
【請求項１０】
前記環境不純物および遺伝型解析による不純物が、前記試料中のホモ接合アレルに関して測定される、請求項９に記載の方法。
【請求項１１】
前記選択が、前記モデルに関して作成された前記相決定済みアレル情報と推定アレル頻度との差異の大きさを分析することによって実行される、請求項７に記載の方法。
【請求項１２】
アレル頻度の前記個別確率が、前記多型座位の集合におけるアレル頻度の期待値と観測値のベータ二項モデルに基づいて作成される、請求項１～５および７～１１のいずれか一項に記載の方法。
【請求項１３】
前記個別確率がベイズ分類器を用いて作成される、請求項７に記載の方法。
【請求項１４】
前記核酸配列データが、多重増幅反応によって生成された一連の増幅産物の複数のコピーの高処理ＤＮＡ配列決定を実行することによって作成され、ここで、該一連の増幅産物のうちの各増幅産物は、前記多型座位の集合のうちの少なくとも１つの多型座位にわたり、該多型座位の集合のうちの各多型座位が増幅される、請求項７に記載の方法。
【請求項１５】
前記多重増幅反応が制限的なプライマー条件で実行される、請求項１４に記載の方法。
【請求項１６】
前記試料に含まれる染色体分節の集合について染色体倍数性が決定される、請求項１４に記載の方法。
【請求項１７】
前記多重増幅反応が１００，０００プレックス以下の多重反応を含む、請求項１６に記載の方法。
【請求項１８】
前記試料の平均アレル不均衡率が０．４％～５％である、請求項７に記載の方法。
【請求項１９】
前記試料が、癌を有すると考えられる個体から得た血漿試料であり、前記方法が、前記選択に基づき該個体の腫瘍細胞にコピー数多様性が存在するかを決定することをさらに含む、請求項７に記載の方法。
【請求項２０】
前記試料が個体から得た血漿試料であり、前記方法が前記選択に基づき該個体に癌が存在するかを決定することをさらに含む、請求項１９に記載の方法。
【請求項２１】
一塩基多様性位置の集合に含まれる一塩基多様性位置の一塩基多様体を検出することをさらに含み、ここで、染色体異数性および／または該一塩基多様体の検出は前記試料中に循環腫瘍核酸が存在することを示す、請求項２０に記載の方法。
【請求項２２】
前記個体の腫瘍に関する前記染色体分節のハプロタイプ情報を受け、該ハプロタイプ情報を用いて、前記多型座位の集合に関する異なる倍数性状態とアレル不均衡率の前記モデルのセットを作成することをさらに含む、請求項１９に記載の方法。
【請求項２３】

前記試料が妊婦から得た血漿試料である、請求項７に記載の方法。
【請求項２４】
前記方法が、既知の平均アレル不均衡比を有する対照試料に対して該方法を実行することをさらに含む、請求項７に記載の方法。
【請求項２５】
前記修正済みアレル頻度を前記モデルのセットと比較する前に該修正済みアレル頻度のデータから異常値を除外することをさらに含む、請求項７に記載の方法。
【請求項２６】
個体の試料の染色体倍数性を検出するためのシステムであって、該システムは、
ａ．該試料中の該染色体分節にある多型座位の集合のうちの各座位に存在する各アレルの量を含むアレル頻度データを受けるよう構成された入力プロセッサと、
ｂ．モデラーであって、
ｉ．該アレル頻度データの相を推定することによって該多型座位の集合に関する相決定済みアレル情報を作成し、
ｉｉ．該アレル頻度データを用いて、異なる倍数性状態に関する該多型座位のアレル頻度の個別確率を作成し、
ｉｉｉ．該個別確率と該相決定済みアレル情報を用いて該多型座位の集合に関する同時確率を作成する
よう構成されたモデラーと、
ｃ．該同時確率に基づき、染色体倍数性を示す最適合モデルを選択することにより、該染色体分節の倍数性を決定するよう構成された、仮説管理部（hypothesis manager）
とを含む、システム。
【請求項２７】
前記アレル頻度データが核酸配列決定システムによって作成される、請求項２６に記載のシステム。
【請求項２８】
前記アレル頻度データの誤差を修正するよう構成された誤差修正部をさらに含み、ここで、修正済みアレル頻度データが前記個別確率の作成工程に用いられる、請求項２７に記載のシステム。
【請求項２９】
前記誤差修正部がアレル増幅効率の偏りを修正する、請求項２８に記載のシステム。
【請求項３０】
前記モデラーが、前記多型座位の集合に関する異なる倍数性状態とアレル不均衡率の両方のモデルのセットを用いて前記個別確率を作成する、請求項２６～２９のいずれか一項に記載のシステム。
【請求項３１】
前記モデラーが、前記染色体分節にある多型座位間の連鎖を考慮することによって前記同時確率を作成する、請求項３０に記載のシステム。
【請求項３２】
個体の試料の染色体倍数性を検出するためのシステムであって、該システムは、
ａ．該個体の染色体分節にある多型座位の集合に存在するアレルに関する核酸配列データを受け、該核酸配列データを用いて該多型座位の集合におけるアレル頻度を検出するよう構成された入力プロセッサと、
ｂ．検出された該アレル頻度の誤差を修正し、該多型座位の集合に関する修正済みアレル頻度を作成するよう構成された誤差修正部と、
ｃ．モデラーであって、
ｉ．該核酸配列データの相を推定することによって該多型座位の集合に関する相決定済みアレル情報を作成し、
ｉｉ．該相決定済みアレル情報を該多型座位の集合に関する異なる倍数性状態とアレル不均衡率のモデルのセットと比較することにより、異なる倍数性状態に関する該多型座位の

アレル頻度の個別確率を作成し、
ｉｉｉ．該個別確率を該染色体分節にある多型座位間の相対的距離を考慮しながら組み合わせることにより、該多型座位の集合に関する同時確率を作成する
よう構成されたモデラーと、
ｄ．該同時確率に基づき、染色体異数性を示す最適合モデルを選択するよう構成された仮説管理部
とを含む、システム。
【請求項３３】
前記仮説管理部が、前記モデルに関して作成された前記相決定済みアレル情報と推定アレル頻度との差異の大きさを分析することによって選択を行う、請求項３２に記載のシステム。
【請求項３４】
前記モデラーが、前記多型座位の集合におけるアレル頻度の期待値と観測値のベータ二項モデルに基づいてアレル頻度の前記個別確率を作成する、請求項３３に記載のシステム。
【請求項３５】
個体の試料の染色体倍数性を検出するための非一過性のコンピュータ可読媒体であって、該媒体はコンピュータ可読コードを含み、該コードを処理装置で実行することにより、該処理装置が、
ａ．該試料中の該染色体分節にある多型座位の集合のうちの各座位に存在する各アレルの量を含むアレル頻度データを受け、
ｂ．該アレル頻度データの相を推定することによって該多型座位の集合に関する相決定済みアレル情報を作成し、
ｃ．該アレル頻度データを用いて、異なる倍数性状態に関する該多型座位のアレル頻度の個別確率を作成し、
ｄ．該個別確率と該相決定済みアレル情報を用いて該多型座位の集合に関する同時確率を作成し、
ｅ．該同時確率に基づき、染色体倍数性を示す最適合モデルを選択することにより、該染色体分節の倍数性を決定する、
ように構成されるコンピュータ可読媒体。
【請求項３６】
前記アレル頻度データが核酸配列データから作成される、請求項３５に記載のコンピュータ可読媒体。
【請求項３７】
前記アレル頻度データの誤差を修正し、修正済みアレル頻度データを前記個別確率の作成工程に使用することをさらに含む、請求項３５に記載のコンピュータ可読媒体。
【請求項３８】
修正される前記誤差がアレル増幅効率の偏りである、請求項３７に記載のコンピュータ可読媒体。
【請求項３９】
前記個別確率が、前記多型座位の集合に関する異なる倍数性状態とアレル不均衡率の両方のモデルのセットを用いて作成される、請求項３５に記載のコンピュータ可読媒体。
【請求項４０】
前記同時確率が、前記染色体分節にある多型座位間の連鎖を考慮することによって作成される、請求項３５～３９のいずれか一項に記載のコンピュータ可読媒体。
【請求項４１】
個体の試料の染色体倍数性を検出するための非一過性のコンピュータ可読媒体であって、該媒体はコンピュータ可読コードを含み、該コードを処理装置で実行することにより、該処理装置が、
ａ．該個体の染色体分節にある多型座位の集合に存在するアレルに関する核酸配列データ

を受け、
ｂ．該核酸配列データを用いて該多型座位の集合におけるアレル頻度を検出し、
ｃ．検出された該アレル頻度のアレル増幅効率の偏りを修正することによって、該多型座位の集合に関する修正済みアレル頻度を作成し、
ｄ．該核酸配列データの相を推定することによって該多型座位の集合に関する相決定済みアレル情報を作成し、
ｅ．該修正済みアレル頻度を該多型座位の集合に関する異なる倍数性状態とアレル不均衡率のモデルのセットと比較することにより、異なる倍数性状態に関する該多型座位のアレル頻度の個別確率を作成し、
ｆ．該個別確率を該染色体分節にある多型座位間の連鎖を考慮しながら組み合わせることにより、該多型座位の集合に関する同時確率を作成し、
ｇ．該同時確率に基づき、染色体異数性を示す最適合モデルを選択する、
ように構成される、コンピュータ可読媒体。
【請求項４２】
前記選択が、前記モデルに関して作成された前記相決定済みアレル情報と推定アレル頻度との差異の大きさを分析することによって実行される、請求項４１に記載のコンピュータ可読媒体。
【請求項４３】
アレル頻度の前記個別確率が、前記多型座位の集合におけるアレル頻度の期待値と観測値のベータ二項モデルに基づいて作成される、請求項３５～３９および４１のいずれか一項に記載のコンピュータ可読媒体。
【請求項４４】
循環腫瘍核酸が個体の試料中に存在するかを決定するための方法であって、該方法は、
Ａ．該試料を分析して該個体の染色体分節にある多型座位の集合の倍数性を決定し、
Ｂ．該多型座位に存在するアレル不均衡のレベルを該倍数性決定に基づいて決定することを含み、ここで、０．４５％以上のアレル不均衡率は該試料に循環腫瘍核酸が存在することを示す、
方法。
【請求項４５】
一塩基多様性位置の集合に含まれる一塩基多様性位置の一塩基多様体を検出することをさらに含み、ここで、０．４５％以上のアレル不均衡率の検出および／または一塩基多様体の検出は、前記試料に循環腫瘍核酸が存在することを示す、請求項４４に記載の方法。
【請求項４６】
前記方法が、既知の平均アレル不均衡比を有する対照試料に対して該方法を実行することをさらに含む、請求項４４に記載の方法。
【請求項４７】
前記対照が前記個体の腫瘍から得た試料である、請求項４６に記載の方法。
【請求項４８】
前記分析が、癌において異数性を示すことが知られている染色体分節の集合を分析することを含む、請求項４６に記載の方法。
【請求項４９】
前記試料の分析が、多重ＰＣＲを実行することにより１０００～５０，０００個の同義座位および１００～１０００個の一塩基多様体位置にわたる各増幅産物を増幅することを含む、請求項４４～４８に記載の方法。
【請求項５０】
個体の試料に含まれるゲノム位置の集合に一塩基多様体が存在するかを決定するための方法であって、該方法は、
ａ．各ゲノム位置に関して、訓練データセットを用いてそのゲノム位置にわたる増幅産物に関するサイクル当たりの効率およびエラー率の推定値を作成し、
ｂ．該試料に含まれる各ゲノム位置に関する観測ヌクレオチド同一性情報を受け、

ｃ．各ゲノム位置に関するサイクル当たりの増幅効率およびエラー率の該推定値を個別に用いて、各ゲノム位置に関する該観測ヌクレオチド同一性情報を異なる多様体率のモデルと比較することにより、各ゲノム位置における１つ以上の実際の変異によって生じる一塩基多様体率の確率のセットを決定し、
ｄ．各ゲノム位置に関する該確率のセットから、最も可能性の高い実多様体率と信頼度を決定すること
とを含む、方法。
【請求項５１】
サイクル当たりの効率およびエラー率の前記推定値が、前記ゲノム位置にわたる一連の増幅産物に関して作成される、請求項５０に記載の方法。
【請求項５２】
前記観測ヌクレオチド同一性情報が、各ゲノム位置に関する全リード数の観測値と各ゲノム位置に関する多様体アレルのリード数の観測値を含む、請求項５０に記載の方法。
【請求項５３】
前記試料が血漿試料であり前記一塩基多様体が該試料の循環腫瘍ＤＮＡに存在する、請求項５０～５２のいずれか一項に記載の方法。
【請求項５４】
個体の試料に存在する一塩基多様体率を推定するための方法であって、該方法は、
ａ．ゲノム位置の集合において、訓練データセットを用いてそれらのゲノム位置にわたる１個以上の増幅産物に関するサイクル当たりの効率およびエラー率の推定値を作成し、
ｂ．該試料に含まれる各ゲノム位置に関する観測ヌクレオチド同一性情報を受け、
ｃ．該増幅産物のサイクル当たりの増幅効率とエラー率の該推定値を用いて、ある初期の実変異分子率を有する調査領域の全分子数と、背景エラー分子と、実変異分子とに関する平均と分散の推定値を作成し、
ｄ．平均および分散の該推定値を用いた分布を該試料の観測ヌクレオチド同一性情報に適合させることによって最も可能性の高い実際の一塩基多様体率を決定することにより、実際の変異によって生じる該試料の一塩基多様体率を決定すること
とを含む、方法。
【請求項５５】
前記試料が血漿試料であり前記一塩基多様体が該試料の循環腫瘍ＤＮＡに存在する、請求項５４に記載の方法。
【請求項５６】
個体の試験試料中に存在する１つ以上の一塩基多様体を検出するための方法であって、該方法は、
ａ．配列決定の実行時に作成された結果に基づき、複数の正常な個体の各々から得た複数の対照試料の多様性アレル頻度の中央値を、一塩基多様性位置の集合に含まれる各一塩基多様体位置について決定することにより、閾値未満である正常試料の多様体アレル頻度の中央値を有する、選定された一塩基多様体位置を確認し、該一塩基多様体位置の各々について異常値試料を除外した後、該一塩基多様体位置の各々に関する背景エラーを決定し、
ｂ．該試験試料に関する配列決定の実行時に作成されたデータに基づき、該試験試料について、選定された該一塩基多様体位置に関して観測リード深度による加重平均と分散を決定し、
ｃ．コンピュータを用いて、リード深度による加重平均が背景エラーに比べて統計的に有意である１つ以上の一塩基多様体位置を確認することによって、該１つ以上の一塩基多様体を検出すること
とを含む、方法。
【請求項５７】
前記試料が血漿試料であり、前記対照試料が血漿試料であり、検出された前記１つ以上の一塩基多様体が該試料の循環腫瘍ＤＮＡに存在する、請求項５６に記載の方法。
【請求項５８】

前記複数の対照試料が少なくとも２５個の試料を含む、請求項５６に記載の方法。
【請求項５９】
高処理配列決定の実行時に作成されたデータから異常値を除外することによって前記観測リード深度による加重平均と観測分散が算出される、請求項５６～５８のいずれか一項に記載の方法。
【請求項６０】
前記試験試料の各一塩基多様性位置に関する前記リード深度が少なくとも１００リードである、請求項５９に記載の方法。
【請求項６１】
前記配列決定の実行が、制限的なプライマー反応条件で実行される多重増幅反応を含む、請求項５９に記載の方法。
【発明の詳細な説明】
【技術分野】
【０００１】
関連する出願の参照
本願は、２０１４年４月２１日に提出された米国仮特許出願第６１／９８２，２４５号、２０１４年５月１日に提出された米国仮特許出願第６１／９８７，４０７号、２０１４年１０月２１日に提出された米国仮特許出願第６２／０６６，５１４号、２０１５年４月１０日に提出された米国仮特許出願第６２／１４６，１８８号、２０１５年４月１４日に提出された米国仮特許出願第６２／１４７，３７７号、および２０１５年４月１５日に提出された米国仮特許出願第６２／１４８，１７３号の優先権を主張するものであり、これら出願の教示についてその全内容を参照により本明細書に組み込む。
【０００２】
本発明は、主に染色体分節の倍数性を検出する方法およびシステム、および一塩基多様体を検出する方法およびシステムに関する。
【背景技術】
【０００３】
コピー数多様性（ＣＮＶ）は、典型的には１，０００塩基対（１ｋｂ）から２０メガ塩基対（ｍｂ）の長さの配列の重複および欠失を含むゲノムの構造多様性の主因として認識されている。染色体分節または染色体全体の欠失および重複は、病気に対する感受性または耐性等、様々な条件と関連づけられる。
【０００４】
ＣＮＶは、影響を受ける配列の長さに基づいて主に２つの型の内の１つに分類されることが多い。１つ目の型はコピー数多型（ＣＮＰ）を含む。ＣＮＰは、一般的な集団に共通して見られ、１％超の全頻度で発生する。ＣＮＰは、通常は小さく（ほとんどは１０キロ塩基対未満の長さ）、薬物解毒および免疫に重要なタンパク質をコードする遺伝子に多い。これらＣＮＰの下位集合ではコピー数が大きく異なる。その結果、遺伝子の特定の集合について染色体のコピー数が、個人間で大きく異なる（例えば、２、３、４、５等）可能性がある。免疫応答遺伝子に関連したＣＮＰは、近年、遺伝性の難病（乾癬、クローン病、および糸球体腎炎を含む）に対する感受性と関連づけられている。
【０００５】
もう一方のＣＮＶは、長さが約１０万塩基対から１００万超の塩基対まで、ＣＮＰよりもかなり長い比較的希な多様体を含む。いくつかの場合では、これらＣＮＶは、特定の個体を発生させた精子または卵子ができる過程で生じた可能性、あるいはある家族の数世代間でのみ受け継がれた可能性がある。このような希な大型構造多様体は、精神遅滞、発達遅滞、統合失調症、および自閉症を有する対象者において偏って観察されている。このような対象者の様子から、神経認知疾患においては遺伝した変異の他の形態（一塩基置換を含む）に比べてこのような希な大型ＣＮＶがより重要であるかもしれないと推測されている。
【０００６】

遺伝子のコピー数は癌細胞では変化している可能性がある。例えば、Ｃｈｒ１ｐの重複は乳癌で共通して見られる。また、ＥＧＦＲのコピー数は、正常な細胞に比べて非小細胞肺癌では多くなり得る。癌は、主要死亡原因の１つであり、癌の早期診断および治療が重要である。というのは、早期診断および治療は（例えば、沈静の確率やその持続を向上させることによって）患者の転帰を向上させることができるからである。早期診断はまた、患者が受ける治療法の量または数を低減させることができる。癌性細胞を破壊する現在の治療法の多くは正常細胞にも影響を及ぼし、その結果、様々な副作用（吐き気、嘔吐、血球数の低下、感染リスクの増加、抜け毛、および粘膜の潰瘍等）を起こす可能性がある。よって、癌の早期検出は、治療法（化学療法薬または放射線等）の数および／または量を低減させることができるため望ましい。

【０００７】

コピー数変異はまた、数種の精神障害、身体障害、および特発性学習障害と関連づけられている。無細胞ＤＮＡを用いた非侵襲性出生前検査（ＮＩＰＴ）により異常（例えば、胎児の１３番染色体、１８番染色体、および２１番染色体の三染色体性、三倍体性、および性染色体の異数性等）を検出することができる。亜染色体の微小欠失もまた、数種の精神障害および身体障害を引き起こし得るが、より小さいものであるためにその検出はより困難である。微小欠失症候群のうち、８つは凝集発生率が１０００分の１を超えており、胎児の常染色体の三染色体性と同程度見られる。

【０００８】

また、ＣＣＬ３Ｌ１の高コピー数がＨＩＶ感染への低感受性と関連づけられている。ＦＣＧＲ３Ｂ（ＣＤ１６細胞表面免疫グロブリン受容体）のコピー数が少ないと、全身性エリテマトーデスおよび類似の炎症性自己免疫疾患に対する感受性を増加させ得る。

【０００９】

このように、染色体分節または染色体全体の欠失および重複を検出する改良された方法が求められている。これら方法を用いて、癌等の病気の診断や病気の発症リスクの増加、あるいは胎児のＣＮＶを正確に診断することができるが好ましい。

【発明の概要】

【００１０】

本明細書の例示的態様では、個体の試料に含まれる染色体分節の倍数性を決定するための方法が提供される。該方法は以下の工程を含む。

ａ．該試料中の該染色体分節にある多型座位の集合のうちの各座位に存在する各アレルの量を含むアレル頻度データを受け、

ｂ．該アレル頻度データの相を推定することによって該多型座位の集合に関する相決定済み（ｐｈａｓｅｄ）アレル情報を作成し、

ｃ．該アレル頻度データを用いて、異なる倍数性状態に関する該多型座位のアレル頻度の個別確率を作成し、

ｄ．該個別確率と該相決定済みアレル情報を用いて該多型座位の集合に関する同時確率を作成し、

ｅ．該同時確率に基づき、染色体倍数性を示す最適合モデルを選択することにより、該染色体分節の倍数性を決定すること。

【００１１】

前記倍数性を決定するための方法の例示的な一態様では、前記データが核酸配列データ、特に高処理核酸配列データを用いて作成される。前記倍数性を決定するための方法のある例では、前記アレル頻度データが個別確率の作成に用いられる前に、エラーが修正される。特定の例示的な態様では、修正されるエラーがアレル増幅効率の偏りを含む。他の態様では、修正されるエラーが、環境不純物および遺伝型解析による不純物を含む。いくつかの態様では、修正されるエラーが、アレル増幅の偏り、環境不純物、および遺伝型解析による不純物を含む。

【００１２】

前記倍数性を決定するための方法の特定の態様では、前記個別確率が、前記多型座位の

集合に関する異なる倍数性状態とアレル不均衡率の両方のモデルのセットを用いて作成される。これらの態様および他の態様では、前記同時確率が、前記染色体分節にある多型座位間の連鎖を考慮することによって作成される。

【００１３】

したがって、本明細書では、これら態様のいくつかを組み合わせた例示的一態様において、個体の試料の染色体倍数性を検出するための方法が提供される。該方法は以下の工程を含む。

ａ．該個体の染色体分節にある多型座位の集合に存在するアレルに関する核酸配列データを受け、

ｂ．該核酸配列データを用いて該多型座位の集合におけるアレル頻度を検出し、

ｃ．検出された該アレル頻度のアレル増幅効率の偏りを修正することによって、該多型座位の集合に関する修正済みアレル頻度を作成し、

ｄ．該核酸配列データの相を推定することによって該多型座位の集合に関する相決定済みアレル情報を作成し、

ｅ．該修正済みアレル頻度を該多型座位の集合に関する異なる倍数性状態とアレル不均衡率のモデルのセットと比較することにより、異なる倍数性状態に関する該多型座位のアレル頻度の個別確率を作成し、

ｆ．該個別確率を該染色体分節にある多型座位間の連鎖を考慮しながら組み合わせることにより、該多型座位の集合に関する同時確率を作成し、

ｇ．該同時確率に基づき、染色体異数性を示す最適合モデルを選択すること。

【００１４】

本明細書において他の側面では、個体の試料の染色体倍数性を検出するためのシステムであって、該システムは、

ａ．該試料中の該染色体分節にある多型座位の集合のうちの各座位に存在する各アレルの量を含むアレル頻度データを受けるよう構成された入力プロセッサと、

ｂ．モデラーであって、

ｉ．該アレル頻度データの相を推定することによって該多型座位の集合に関する相決定済みアレル情報を作成し、

ｉｉ．該アレル頻度データを用いて、異なる倍数性状態に関する該多型座位のアレル頻度の個別確率を作成し、

ｉｉｉ．該個別確率と該相決定済みアレル情報を用いて該多型座位の集合に関する同時確率を作成するよう構成されたモデラーと、

ｃ．該同時確率に基づき、染色体倍数性を示す最適合モデルを選択することにより、該染色体分節の倍数性を決定するよう構成された、仮説管理部（hypothesis manager）とを含む。

【００１５】

このシステムの態様の特定の態様では、前記アレル頻度データが核酸配列決定システムによって作成されるデータである。特定の態様では、前記システムはさらに、前記アレル頻度データのエラーを修正するよう構成されたエラー修正部を含み、修正済みアレル頻度データが前記モデラーによって用いられて個別確率を作成する。特定の態様では、前記エラー修正部がアレル増幅効率の偏りを修正する。特定の態様では、前記モデラーが、前記多型座位の集合に関する異なる倍数性状態とアレル不均衡率の両方のモデルのセットを用いて前記個別確率を作成する。特定の例示的な態様では、前記モデラーが前記染色体分節にある多型座位間の連鎖を考慮することによって前記同時確率を作成する。

【００１６】

本明細書において、例示的一態様では、個体の試料の染色体倍数性を検出するためのシステムが提供される。該システムは、

ａ．該個体の染色体分節にある多型座位の集合に存在するアレルに関する核酸配列データを受け、該核酸配列データを用いて該多型座位の集合におけるアレル頻度を検出するよう構成された入力プロセッサと、

ｂ．検出された該アレル頻度のエラーを修正し、該多型座位の集合に関する修正済みアレル頻度を作成するよう構成されたエラー修正部と、
ｃ．モデラーであって、
ｉ．該核酸配列データの相を推定することによって該多型座位の集合に関する相決定済みアレル情報を作成し、
ｉｉ．該相決定済みアレル情報を該多型座位の集合に関する異なる倍数性状態とアレル不均衡率のモデルのセットと比較することにより、異なる倍数性状態に関する該多型座位のアレル頻度の個別確率を作成し、
ｉｉｉ．該個別確率を該染色体分節にある多型座位間の相対的距離を考慮しながら組み合わせることにより、該多型座位の集合に関する同時確率を作成するよう構成されたモデラーと、
ｄ．該同時確率に基づき、染色体異数性を示す最適合モデルを選択するよう構成された仮説管理部とを含む。

【００１７】

ある側面では、本発明は循環腫瘍核酸が個体の試料中に存在するかを決定するための方法を提供する。該方法は、
ａ．該試料を分析して該個体の染色体分節にある多型座位の集合の倍数性を決定し、
ｂ．該多型座位に存在するアレル不均衡のレベルを該倍数性決定に基づいて決定することを含み、ここで、０．４％、０．４５％、または０．５％以上のアレル不均衡率は該試料に循環腫瘍核酸が存在することを示す。

【００１８】

特定の態様では、前記循環腫瘍核酸が存在するかを決定するための方法が、一塩基多様性位置の集合に含まれる一塩基多様性位置の一塩基多様体を検出することをさらに含み、ここで、４５％以上のアレル不均衡率の検出および／または一塩基多様体の検出は前記試料に循環腫瘍核酸が存在することを示す。

【００１９】

特定の態様では、前記循環腫瘍核酸が存在するかを決定するための方法の分析工程は、癌で異数性を示すことが知られている染色体分節の集合を分析することを含む。特定の態様では、前記循環腫瘍核酸が存在するかを決定するための方法の分析工程は、倍数性について１，０００個から５０，０００個、または１００個から１０００個の多型座位を分析することを含む。

【００２０】

本明細書において、ある側面では、試料の一塩基多様体を検出する方法が提供される。したがって、本明細書において個体の試料に含まれるゲノム位置の集合に一塩基多様体が存在するかを決定するための方法が提供される。該方法は、
ａ．各ゲノム位置に関して、訓練データセットを用いてそのゲノム位置にわたる増幅産物に関する効率およびサイクル当たりのエラー率の推定値を作成し、
ｂ．該試料に含まれる各ゲノム位置に関する観測ヌクレオチド同一性情報を受け、
ｃ．各ゲノム位置に関する増幅効率およびサイクル当たりのエラー率の該推定値を個別に用いて、各ゲノム位置に関する該観測ヌクレオチド同一性情報を異なる多様体率のモデルと比較することにより、各ゲノム位置における１つ以上の実際の変異によって生じる一塩基多様体率の確率のセットを決定し、
ｄ．各ゲノム位置に関する該確率のセットから、最も可能性の高い実多様体率と信頼度を決定することとを含む。

【００２１】

前記一塩基多様体が存在するかを決定するための方法の例示的態様では、効率およびサイクル当たりのエラー率の前記推定値が、前記ゲノム位置にわたる一連の増幅産物に関して作成される。例えば、前記ゲノム位置にわたる２、３、４、５、１０、１５、２０、２５、５０、１００個以上の増幅産物が含まれる可能性がある。特定の態様では、前記１つ以上のＳＮＶを検出する方法における検出限界は０．０１５％、０．０１７％、または０

．０２％である。
【００２２】
前記一塩基多様体が存在するかを決定するための方法の例示的態様では、前記観測ヌクレオチド同一性情報が、各ゲノム位置に関する全リード数の観測値と各ゲノム位置に関する多様性アレルのリード数の観測値を含む。
【００２３】
前記一塩基多様体が存在するかを決定するための方法の例示的態様では、前記試料が血漿試料であり前記一塩基多様体が該試料の循環腫瘍ＤＮＡに存在する。
【００２４】
本明細書において、他の態様では、個体の試験試料中に存在する１つ以上の一塩基多様体を検出するための方法が提供される。この態様による方法は、以下の工程を含む。
ａ．配列決定の実行時に作成された結果に基づき、複数の正常な個体の各々から得た複数の対照試料の多様性アレル頻度の中央値を、一塩基多様性位置の集合に含まれる各一塩基多様性位置について決定することにより、閾値未満である正常試料の多様性アレル頻度の中央値を有する、選定された一塩基多様性位置を確認し、該一塩基多様性位置の各々について異常値試料を除外した後、該一塩基多様性位置の各々に関する背景エラーを決定し、
ｂ．該試験試料に関する配列決定の実行時に作成されたデータに基づき、該試験試料について、選定された該一塩基多様性位置に関して観測リード深度による加重平均と分散を決定し、
ｃ．コンピュータを用いて、リード深度による加重平均が背景エラーに比べて統計的に有意である１つ以上の一塩基多様性位置を確認することによって、該１つ以上の一塩基多様体を検出すること。
【００２５】
特定の態様では、前記１つ以上のＳＮＶを検出する方法において、前記試料が血漿試料であり、前記対照試料が血漿試料であり、検出された前記１つ以上の一塩基多様体が該試料の循環腫瘍ＤＮＡに存在する。特定の態様では、前記１つ以上のＳＮＶを検出する方法において、前記複数の対照試料が少なくとも２５個の試料を含む。特定の態様では、前記１つ以上のＳＮＶを検出する方法において、高処理配列決定の実行時に作成されたデータから異常値を除外することによって前記観測リード深度による加重平均と観測分散が算出される。特定の態様では、前記１つ以上のＳＮＶを検出する方法において、前記試験試料の各一塩基多様性位置に関する前記リード深度が少なくとも１００リードである。
【００２６】
特定の態様では、前記１つ以上のＳＮＶを検出する方法において、前記配列決定の実行が、制限的なプライマー反応条件で実行される多重増幅反応を含む。特定の態様では、前記１つ以上のＳＮＶを検出する方法において、前記検出限界が０．０１５％、０．０１７％、または０．０２％である。
【００２７】
一側面において、本発明は、ある個体に由来する１個以上の細胞のゲノム中の第二の相同染色体分節と比較して第一の相同染色体分節のコピー数の過剰出現があるかどうかを決定する方法を特徴とする。いくつかの態様では、前記方法は、前記第一の相同染色体分節にある多型座位の集合のうちの座位ごとに前記第一の相同染色体分節にあるその座位に存在するアレルの同一性を含む第一の相同染色体分節の相決定済み遺伝子データを得ることと、前記第二の相同染色体分節にある多型座位の集合のうちの座位ごとに前記第二の相同染色体分節にあるその座位に存在するアレルの同一性を含む第二の相同染色体分節の相決定済み遺伝子データを得ることと、前記多型座位の集合のうちの各座位に存在する各アレルについて、前記個体に由来する１個以上の細胞に由来するＤＮＡまたはＲＮＡの試料に存在する各アレルの量を含む、測定されたアレルの遺伝子データを得ることを含む。いくつかの態様では、前記方法は、前記個体に由来する１個以上の細胞のゲノムにおける前記第一の相同染色体分節の過剰出現度を規定する１つ以上の仮説の集合を列挙し、得られた前記試料の遺伝子データおよび相決定済み遺伝子データに基づいて前記１つ以上の仮説の

尤度を（コンピュータ等で）算出し、最も可能性が高い仮説を選択して、前記個体に由来する１個以上の細胞のゲノムにおける第一の相同染色体分節のコピー数の過剰出現度を決定することを含む。いくつかの態様では、前記相決定済みデータは、集団に基づくハプロタイプ頻度および／または測定された相決定済みデータ（例えば、前記個体または前記個体の血縁者に由来するＤＮＡまたはＲＮＡを含む試料を測定することによって得られた相決定済みデータ）を用いて推論された相決定済みデータを含む。

【００２８】

一側面において、本発明は、ある個体に由来する１個以上の細胞のゲノム中の第二の相同染色体分節と比較して第一の相同染色体分節のコピー数の過剰出現があるかどうかを決定する方法を提供する。いくつかの態様では、前記方法は、前記第一の相同染色体分節にある多型座位の集合のうちの座位ごとに第一の相同染色体分節についてその座位に存在するアレルの同一性を含む第一の相同染色体分節の相決定済み遺伝子データを得ることと、前記第二の相同染色体分節にある多型座位の集合のうちの座位ごとに前記第二の相同染色体分節にあるにその座位に存在するアレルの同一性を含む第二の相同染色体分節の相決定済み遺伝子データを得ることと、前記多型座位の集合のうちの各座位に存在する各アレルについて、前記個体に由来する１個以上の細胞に由来するＤＮＡまたはＲＮＡの試料に存在する各アレルの量を含む、測定されたアレルの遺伝子データを得ることを含む。いくつかの態様では、前記方法は、前記第一の相同染色体分節の過剰出現度を規定する１つ以上の仮説の集合を列挙し、前記仮説のそれぞれについて得られた相決定済み遺伝子データから前記試料の複数の座位について期待遺伝子データを算出し、前記試料について得られた該試料の遺伝子データと期待遺伝子データとのデータ適合を（コンピュータ等で）算出し、前記データ適合に応じて前記仮説の１つ以上をランク付けし、最高ランクの仮説を選択して、前記個体に由来する１個以上の細胞のゲノム中の前記第一の相同染色体分節のコピー数の過剰出現度を決定することを含む。

【００２９】

一側面において、本発明は、ある個体の１個以上の細胞のゲノム中の第二の相同染色体分節と比較して第一の相同染色体分節のコピー数の過剰出現があるかどうかを決定する方法を特徴とする。いくつかの態様では、前記方法は、前記第一の相同染色体分節にある多型座位の集合のうちの座位ごとに第一の相同染色体分節についてその座位に存在するアレルの同一性を含む第一の相同染色体分節の相決定済み遺伝子データを得ることと、前記第二の相同染色体分節にある多型座位の集合のうちの座位ごとに前記第二の相同染色体分節にあるその座位に存在するアレルの同一性を含む第二の相同染色体分節の相決定済み遺伝子データを得ることと、前記多型座位の集合のうちの各座位に存在する各アレルについて前記個体に由来する１個以上の目的細胞および１個以上の目的でない細胞に由来するＤＮＡまたはＲＮＡの試料に存在する各アレルの量を含む、測定されたアレルの遺伝子データを得ることを含む。いくつかの態様では、前記方法は、前記第一の相同染色体分節の過剰出現度を規定する１つ以上の仮説の集合を列挙し、前記仮説のそれぞれについて、得られた相決定済み遺伝子データから、前記１個以上の目的細胞に由来するＤＮＡまたはＲＮＡの前記試料中の総ＤＮＡまたはＲＮＡに対する１つ以上の可能な比率について前記試料中の前記複数の座位に関する期待遺伝子データを（コンピュータ等で）算出し、ＤＮＡまたはＲＮＡの可能な比率および仮説ごとに、前記試料の得られた遺伝子データと前記試料の期待遺伝子データとのデータ適合を（コンピュータ等で）算出し、前記データ適合に応じて前記仮説の１つ以上をランク付けし、最高ランクの仮説を選択して、前記個体に由来する１個以上の細胞のゲノムの第一の相同染色体分節のコピー数の過剰出現度を決定することを含む。

【００３０】

一側面において、本発明は、ある個体に由来する１個以上の細胞のゲノム中の第二の相同染色体分節と比較して第一の相同染色体分節のコピー数の過剰出現があるかどうかを決定する方法を特徴とする。いくつかの態様では、前記方法は、前記第一の相同染色体分節にある多型座位の集合のうちの座位ごとに第一の相同染色体分節についてその座位に存在

するアレルの同一性を含む第一の相同染色体分節の相決定済み遺伝子データを得ることと、前記第二の相同染色体分節にある多型座位の集合のうちの座位ごとに前記第二の相同染色体分節にあるその座位に存在するアレルの同一性を含む第二の相同染色体分節の相決定済み遺伝子データを得ることと、前記多型座位の集合のうちの各座位に存在する各アレルについて、前記個体に由来する１個以上の目的細胞および１個以上の目的でない細胞に由来するＤＮＡまたはＲＮＡの試料に存在する各アレルの量を含む、測定されたアレルの遺伝子データを得ることを含む。いくつかの態様では、前記方法は、前記第一の相同染色体分節の過剰出現度を規定する１つ以上の仮説の集合を列挙し、前記仮説のそれぞれについて、得られた相決定済み遺伝子データから、前記１個以上の目的細胞に由来するＤＮＡまたはＲＮＡの前記試料中の総ＤＮＡまたはＲＮＡに対する１つ以上の可能な比率について前記試料中の前記複数の座位に関する期待遺伝子データを（コンピュータ等で）算出し、前記複数の座位の各座位、ＤＮＡまたはＲＮＡの可能な比率および仮説ごとに、その仮説が正しい尤度を、その座位について得られた前記試料の遺伝子データと、そのＤＮＡまたはＲＮＡの可能な比率およびその仮説でのその座位の期待遺伝子データとの比較を（コンピュータ等で）算出し、各座位および可能な比率ごとに仮説の各座位の確率を組み合わせることによって各仮説の複合確率を決定し、最も高い複合確率を有する仮説を選択して、前記第一の相同染色体分節のコピー数の過剰出現度を決定することを含む。いくつかの態様では、すべての多型座位が同時に考慮されて特定の仮説の確率が算出され、最も高い確率を有する仮説が選択される。

【００３１】

一側面において、本発明は、胎児のゲノム中の目的の染色体分節のコピー数を決定するための方法を特徴とする。いくつかの態様では、前記方法は、前記胎児の少なくとも一方の生物学上の親の相決定済み遺伝子データを得ることを含み、ここで、前記相決定済み遺伝子データは、前記目的の染色体分節を含む相同染色体分節対中に、第一の相同染色体分節および第二の相同染色体分節にある多型座位の集合のうちの各座位に存在するアレルの同一性を含む。いくつかの態様では、前記方法は、胎児のＤＮＡまたはＲＮＡおよびその胎児の母親に由来する母親のＤＮＡまたはＲＮＡを含むＤＮＡまたはＲＮＡ混合試料中の、各座位に存在する各アレルの量を測定することにより目的の染色体分節にある多型座位の集合の遺伝子データを得ることを含む。いくつかの態様では、前記方法は、胎児のゲノムに存在する目的の染色体分節のコピー数を規定する１つ以上の仮説の集合を列挙することを含む。いくつかの態様では、前記方法は、一方または両方の親について、胎児のゲノム中のその親に由来する第一の相同染色体分節またはその部分のコピー数と、胎児のゲノム中のその親に由来する第二の相同染色体分節またはその部分のコピー数と、その胎児のゲノムに存在する目的の染色体分節の総コピー数を規定する１つ以上の仮説の集合を列挙することを含む。いくつかの態様では、前記方法は、前記仮説のそれぞれについて、得られた前記（両）親由来の相決定済み遺伝子データから前記混合試料中の複数の座位の期待遺伝子データを（コンピュータ等で）算出し、得られた混合試料の遺伝子データとその混合試料の期待遺伝子データとのデータ適合を（コンピュータで等で）算出し、前記データ適合により前記仮説の１つ以上をランク付けし、最高ランクの仮説を選択して、胎児のゲノムにおける目的の染色体分節のコピー数を決定することとを含む。

【００３２】

一側面において、本発明は、胎児のゲノム中の目的の染色体または染色体分節のコピー数を決定するための方法を特徴とする。いくつかの態様では、前記方法は、前記胎児の生物学上の少なくとも一方の親について相決定済み遺伝子データを得ることを含み、ここで前記相決定済み遺伝子データは、前記親の第一の相同染色体分節および第二の相同染色体分節にある多型座位の集合のうちの各座位に存在するアレルの同一性を含む。いくつかの態様では、前記方法は、胎児のＤＮＡまたはＲＮＡおよび前記胎児の母親由来の母親ＤＮＡまたはＲＮＡを含むＤＮＡまたはＲＮＡ混合試料において、各座位に存在する各アレルの量を測定することにより染色体または染色体分節にある多型座位の集合の遺伝子データを得ることを含む。いくつかの態様では、前記方法は、前記胎児のゲノムに存在する目的

の染色体または染色体分節のコピー数を規定する１つ以上の仮説の集合を列挙することを含む。いくつかの態様では、前記方法は、前記仮説のそれぞれについて、（ｉ）得られた前記（両）親由来の相決定済み遺伝子データと任意で（ｉｉ）前記胎児に目的の染色体または染色体分節のコピーをもたらした配偶子形成時に発生した可能性がある１つ以上の交差の確率から混合試料中の前記複数の座位の各座位に存在する各アレルの期待量の確率分布を（例えば、コンピュータ上で作成する等）作成し、前記仮説のそれぞれについて、（１）前記得られた混合試料の遺伝子データと（２）その仮説に対する混合試料中の複数の座位の各座位における各アレルの期待量の確率分布との適合を（コンピュータ等で）算出し、前記データ適合によって前記仮説の１つ以上をランク付けし、最高ランクの仮説を選択して、前記胎児のゲノムにおける目的の染色体分節のコピー数を決定することとを含む。

【００３３】

いくつかの態様では、前記方法は、前記胎児の母親の相決定済み遺伝子データを得ることを含む。いくつかの態様では、前記方法は、前記胎児のゲノム中の母親由来の第一の相同染色体分節またはその部分のコピー数と、前記胎児のゲノム中の母親由来の第二の相同染色体分節またはその部分のコピー数と、前記胎児のゲノムに存在する目的の染色体分節の総コピー数を規定する１つ以上の仮説の集合を列挙することを含む。いくつかの態様では、前記方法は、前記仮説のそれぞれについて、得られた母親由来の相決定済み遺伝子データから前記混合試料中の複数の座位について期待遺伝子データを算出することを含む。

【００３４】

いくつかの態様では、前記仮説のそれぞれについての期待遺伝子データは、前記混合試料中の母親のＤＮＡまたはＲＮＡおよび胎児のＤＮＡまたはＲＮＡの複数の座位の各座位に存在する１つ以上のアレルの同一性および量を含む。いくつかの態様では、前記方法は、前記混合試料中の胎児のＤＮＡまたはＲＮＡ分画および母親のＤＮＡまたはＲＮＡ分画を決定することによって期待遺伝子データを（コンピュータ等で）算出することを含む。いくつかの態様では、前記方法は、複数の座位の各座位について、得られた母親の相決定済み遺伝子データ中のその座位に存在するアレルの同一性および前記混合試料中の母親のＤＮＡまたはＲＮＡ分画を用いて、前記混合試料中の母親のＤＮＡまたはＲＮＡのその座位に対する１つ以上のアレルの期待量を算出することを含む。いくつかの態様では、前記方法は、仮説ごとに前記複数の座位の各座位について、前記胎児によって受け継がれたと前記仮説によって規定された母親由来の第一または第二の相同染色体分節のその座位に存在するアレルの同一性と、前記胎児によって受け継がれたと前記仮説により規定された母親由来の第一または第二の相同染色体分節のコピー数と、前記混合試料中の胎児のＤＮＡまたはＲＮＡの分画を用いて、前記混合試料中の母親から譲り受けた胎児のＤＮＡまたはＲＮＡにおけるその座位に対する１つ以上のアレルの期待量を（コンピュータ等で）算出することを含む。

【００３５】

いくつかの態様では、前記仮説のそれぞれについての期待遺伝子データは、前記混合試料中の母親のＤＮＡまたはＲＮＡおよび胎児のＤＮＡまたはＲＮＡの複数の座位の各座位に存在する１つ以上のアレルの同一性および量を含む。いくつかの態様では、前記方法は、前記混合試料の胎児のＤＮＡまたはＲＮＡの分画および母親のＤＮＡまたはＲＮＡの分画を決定することによって期待遺伝子データを算出することを含む。いくつかの態様では、前記方法は、前記複数の座位の各座位について、得られた母親の相決定済み遺伝子データ中のその座位に存在するアレルの同一性および前記混合試料中の母親のＤＮＡまたはＲＮＡ分画を用いて、前記混合試料中の母親のＤＮＡまたはＲＮＡのその座位に対する１つ以上のアレルの期待量を（コンピュータ等で）算出することを含む。いくつかの態様では、前記方法は、仮説ごとに前記複数の座位のうちの各座位について、前記胎児によって受け継がれたと仮説によって規定された母親由来の第一または第二の相同染色体分節に存在するアレルの同一性と、前記胎児によって受け継がれたと仮説によって規定された母親由来の第一または第二の相同染色体分節のコピー数と、前記胎児によって受け継がれたと仮

説によって規定された父親由来の第一または第二の相同染色体分節におけるその座位に存在し得る１つ以上のアレルの同一性と、前記胎児によって受け継がれたと仮説によって規定された父親由来の第一または第二の相同染色体分節のコピー数と、前記混合試料中の胎児のＤＮＡまたはＲＮＡの分画を用いて、前記混合試料中の母親から譲り受けた胎児のＤＮＡまたはＲＮＡにおけるその座位に対する１つ以上のアレルの期待量（コンピュータ等で）算出することを含む。いくつかの態様では、集団頻度を用いて父親由来の第一または第二の相同染色体分節のアレルの同一性を予測する。いくつかの態様では、前記父親由来の第一または第二の相同染色体分節の各座位に存在し得るアレルのそれぞれについてその確率が同じであると見なす。

【００３６】

いくつかの態様では、前記方法は、前記胎児の母親および父親の両方の相決定済み遺伝子データを得ることを含む。いくつかの態様では、前記方法は、前記胎児のゲノム中の母親由来の第一の相同染色体分節またはその部分のコピー数と、前記胎児のゲノム中の母親由来の第二の相同染色体分節のコピー数と、前記胎児のゲノム中の父親由来の第一の相同染色体分節またはその部分のコピー数と、前記胎児のゲノム中の父親由来の第二の相同染色体分節またはその部分のコピー数と、前記胎児のゲノム中に存在する目的の染色体分節の総コピー数を規定する１つ以上の仮説の集合を列挙することを含む。いくつかの態様では、前記方法は、前記仮説のそれぞれについて、得られた母親由来の相決定済み遺伝子データおよび父親由来の相決定済み遺伝子データから前記混合試料中の複数の座位の期待遺伝子データを（コンピュータ等で）算出することを含む。

【００３７】

いくつかの態様では、前記仮説のそれぞれについて、前記期待遺伝子データは、前記混合試料中の前記母親のＤＮＡまたはＲＮＡおよび胎児のＤＮＡまたはＲＮＡの複数の座位において各座位に存在する１つ以上のアレルの同一性と量を含む。いくつかの態様では、前記方法は、前記混合試料のＤＮＡまたはＲＮＡの分画と胎児の母親のＤＮＡまたはＲＮＡの分画を決定することによって期待遺伝子データを算出することを含む。いくつかの態様では、前記方法は、前記複数の座位の各座位について、得られた母親の相決定済み遺伝子データ中のその座位に存在するアレルの同一性および前記混合試料中の母親のＤＮＡまたはＲＮＡ分画を用いて、前記混合試料中の母親のＤＮＡまたはＲＮＡのその座位に対する１つ以上のアレルの期待量を（コンピュータ等で）算出することを含む。いくつかの態様では、前記方法は、仮説ごとに前記複数の座位の各座位について、前記胎児によって受け継がれたと仮説によって規定された母親由来の第一または第二の相同染色体分節のアレルの同一性と、前記胎児によって受け継がれたと仮説によって規定された母親由来の第一または第二の相同染色体分節のコピー数と、前記胎児によって受け継がれたと仮説によって規定された父親由来の第一または第二の相同染色体分節のその座位に存在するアレルの同一性と、前記胎児によって受け継がれたと仮説によって規定された父親由来の第一または第二の相同染色体分節のコピー数と、前記混合試料中の胎児のＤＮＡまたはＲＮＡの分画を用いて前記混合試料中の胎児のＤＮＡまたはＲＮＡの座位について１つ以上のアレルの期待量を（コンピュータ等で）算出することを含む。

【００３８】

いくつかの態様では、前記方法は、前記仮説のそれぞれについて、得られた前記（両）親由来の相決定済み遺伝子データから前記混合試料中の前記複数の座位の期待遺伝子データの確率分布を（コンピュータ等で）算出することを含む。いくつかの態様では、前記方法は、前記混合試料の第一の座位に存在する特定のアレルが前記親の第一の相同分節に存在し、かつ前記親の第一の相同分節の近隣した座位に存在するアレルが前記混合試料の得られた遺伝子データに観察される場合にその特定のアレルの確率分布における確率を増加させること、あるいは前記混合試料の第一の座位に存在する特定のアレルが前記親の第一の相同分節に存在し、かつ前記親の第一の相同分節の近隣した座位に存在するアレルが前記得られた混合試料の遺伝子データにおいて観察されない場合にその特定のアレルの確率分布の確率を低減させることを含む。いくつかの態様では、前記方法は、前記混合試料の

第二の座位に存在する特定のアレルが前記親の第二の相同分節に存在し、かつ前記親の第二の相同分節の近隣した座位に存在するアレルが前記得られた混合試料の遺伝子データに観察される場合にその特定のアレルの確率分布の確率を増加させること、あるいは前記混合試料の第二の座位に存在する特定のアレルが前記親の第二の相同分節に存在し、かつ前記親の第二の相同分節の近隣した座位に存在するアレルが前記得られた混合試料の遺伝子データに観察されない場合にその特定のアレルの確率分布の確率を低減させることを含む。

【００３９】

いくつかの態様では、前記方法は、前記胎児の母親および父親両方の相決定済み遺伝子データを得ることを含む。いくつかの態様では、前記方法は、前記胎児のゲノム中の前記母親由来の第一の相同染色体分節またはその部分のコピー数と、前記胎児のゲノム中の前記母親由来の第二の相同染色体分節またはその部分のコピー数と、前記胎児のゲノム中の前記父親由来の第一の相同染色体分節またはその部分のコピー数と、前記胎児のゲノム中の前記父親由来の第二の相同染色体分節またはその部分のコピー数と、前記胎児のゲノムに存在する目的の染色体分節の総コピー数を規定する１つ以上の仮説の集合を列挙することを含む。いくつかの態様では、前記方法は、前記仮説のそれぞれについて、前記母親および父親から得られた相決定済み遺伝子データから前記混合試料の複数の座位の期待遺伝子データの確率分布を（コンピュータ等で）算出することを含む。いくつかの態様では、前記方法は、前記混合試料の第一の座位に存在する特定のアレルが前記母親または父親の第一の相同分節に存在し、かつその親の第一の相同分節の近隣した座位に存在するアレルが前記得られた混合試料の遺伝子データに観察される場合にその特定のアレルの確率分布の確率を増加させること、あるいは前記混合試料の第一の座位に存在する特定のアレルが前記母親または父親の第一の相同分節に存在し、かつその親の第一の相同分節の近隣した座位に存在するアレルが前記得られた混合試料の遺伝子データに観察されない場合にはその特定のアレルの確率分布の確率を低減させることを含む。いくつかの態様では、前記方法は、前記混合試料の第二の座位に存在する特定のアレルが前記母親または父親の第二の相同分節に存在し、かつその親の第二の相同分節の近隣した座位に存在するアレルが前記得られた混合試料の遺伝子データに観察される場合にその特定のアレルの確率分布の確率を増加させること、あるいは前記混合試料の第二の座位に存在する特定のアレルが前記母親または父親の第二の相同分節に存在し、かつその親の第二の相同分節の近隣した座位に存在するアレルが前記得られた混合試料の遺伝子データに観察されない場合にその特定のアレルの確率分布の確率を低減させることを含む。

【００４０】

いくつかの態様では、前記第一の座位と第一の座位に近接した座位が共分離している。いくつかの態様では、前記第二の座位と第二の座位に近接した座位が共分離している。いくつかの態様では、前記第一の座位と第一の座位に近接した座位との間では、交差は起こらないと予測される。いくつかの態様では、前記第二の座位と第二の座位に近接した座位との間では、交差は起こらないと予測される。いくつかの態様では、前記第一の座位と第一の座位に近接した座位との距離は、５ｍｂ未満、１ｍｂ未満、１００ｋｂ未満、１０ｋｂ未満、１ｋｂ未満、０．１ｋｂ未満、または０．０１ｋｂ未満である。いくつかの態様では、前記第二の座位と第二の座位に近接した座位との距離は、５ｍｂ未満、１ｍｂ未満、１００ｋｂ未満、１０ｋｂ未満、１ｋｂ未満、０．１ｋｂ未満、または０．０１ｋｂ未満である。

【００４１】

いくつかの態様では、目的の染色体分節のコピーを前記胎児にもたらす配偶子の形成時に１つ以上の交差が起こり、その交差によって前記親由来の第一の相同分節の部分および第二の相同分節の部分を含む前記胎児のゲノム中の目的の染色体分節が生じる。いくつかの態様では、前記仮説の集合は、前記親に由来する第一の相同分節の部分および第二の相同分節の部分を含む前記胎児のゲノム中の目的の染色体分節のコピー数を規定する１つ以上の仮説を含む。

【００４２】
いくつかの態様では、前記混合試料の期待遺伝子データは、前記仮説のそれぞれについて、前記混合試料の複数の座位の各座位に存在する１つ以上のアレルの期待量を含む。
【００４３】
一側面において、本発明は、ある個体のゲノム（例えば、１個以上の細胞、無細胞ＤＮＡ、無細胞ＲＮＡ、癌が疑われる個体、胎児、または胚のゲノム）において第二の相同染色体分節と比較して第一の相同染色体分節のコピー数の過剰出現があるかどうかを、相決定済み遺伝子データを用いて決定する方法を特徴とする。いくつかの態様では、前記方法は、同時にまたは任意の順序で順に（ｉ）前記第一の相同染色体分節にある多型座位の集合のうちの座位ごとに第一の相同染色体分節についてその座位に存在するアレルの同一性を含む第一の相同染色体分節の相決定済み遺伝子データを得ることと、（ｉｉ）前記第二の相同染色体分節にある多型座位の集合のうちの座位ごとに前記第二の相同染色体分節にあるその座位に存在するアレルの同一性を含む第二の相同染色体分節の相決定済み遺伝子データを得ることと、（ｉｉｉ）前記個体に由来する１個以上の細胞に由来するＤＮＡまたはＲＮＡの試料または前記個体の遺伝的に異なる２個以上の細胞に由来する無細胞ＤＮＡまたはＲＮＡの混合試料中の、前記多型座位の集合のうちの各座位に存在する各アレルの量を含む測定されたアレルの遺伝子データを得ることを含む。いくつかの態様では、前記方法は、前記試料が由来する少なくとも１個の細胞においてヘテロ接合である前記多型座位の集合の１箇所以上の座位についてアレル比率を算出することを含む。いくつかの態様では、特定の座位の算出済みアレル比率は、前記アレルの１つの測定量をその座位のすべてのアレルの総測定量で除したものである。いくつかの態様では、前記方法は、ある座位の１つ以上の算出済みアレル比率を期待アレル比率（例えば、第一および第二の相同染色体分節が等しい割合で存在する場合にその座位に期待される比率）と比較することで第一の相同染色体分節のコピー数の過剰出現があるかどうかを決定することを含む。いくつかの態様では、期待比率は２アレル座位について０．５である。
【００４４】
出生前検査に関するいくつかの態様では、前記方法は、同時にまたは任意の順序で順に（ｉ）胎児（例えば、妊娠している母親の胎内の胎児）のゲノムの第一の相同染色体分節について、前記第一の相同染色体分節にある多型座位の集合のうちの座位ごとに前記第一の相同染色体分節にある各座位に存在するアレルの同一性を含む相決定済み遺伝子データを得ることと、（ｉｉ）前記胎児のゲノムの第二の相同染色体分節について、前記第二の相同染色体分節にある多型座位の集合のうちの座位ごとに前記第二の相同染色体分節にある各座位に存在するアレルの同一性を含む相決定済み遺伝子データを得ることと、（ｉｉｉ）胎児のＤＮＡまたはＲＮＡおよび母親のＤＮＡまたはＲＮＡを含む前記胎児の母親由来のＤＮＡまたはＲＮＡの混合試料（例えば、母親由来の血液試料に由来する、胎児の無細胞ＤＮＡまたはＲＮＡと母親の無細胞ＤＮＡまたはＲＮＡを含む無細胞ＤＮＡまたはＲＮＡの混合試料）において、前記多型座位の集合の各座位に存在する各アレルの量を含む測定されたアレルの遺伝子データを得ることを含む。いくつかの態様では、前記方法は、前記胎児においてヘテロ接合であり、かつ／または前記母親においてヘテロ接合である多型座位の集合の１箇所以上の座位についてアレル比率を算出することを含む。いくつかの態様では、特定の座位の算出済みアレル比率は、前記アレルの１つの測定量を前記座位のすべてのアレルの総測定量で除したものである。いくつかの態様では、前記方法は、ある座位の１つ以上の算出済みアレル比率を期待アレル比率（例えば、第一および第二の相同染色体分節が等しい割合で存在する場合にその座位に期待される比率）と比較することで前記第一の相同染色体分節のコピー数の過剰出現があるかどうかを決定することを含む。
【００４５】
いくつかの態様では、（ｉ）ある座位のすべてのアレルの総測定量で除した、前記第一の相同染色体にあるその座位に存在するアレルの測定量のアレル比率が、その座位の期待アレル比率より大きい場合、または（ｉｉ）ある座位のすべてのアレルの総測定量で除した、前記第二の相同染色体にあるその座位に存在するアレルの測定量のアレル比率がその

座位の期待アレル比率より小さい場合のいずれかであれば、算出済みアレル比率は前記第一の相同染色体分節のコピー数過剰出現を示す。いくつかの態様では、（ｉ）ある座位のすべてのアレルの総測定量で除した、前記第一の相同染色体にあるその座位に存在するアレルの測定量のアレル比率がその座位の期待アレル比率以下である場合、または（ｉｉ）ある座位のすべてのアレルの総測定量で除した、前記第二の相同染色体にあるその座位に存在するアレルの測定量のアレル比率がその座位の期待アレル比率以上である場合のいずれかであれば、算出済みアレル比率は前記第一の相同染色体分節のコピー数過剰出現でないことを示す。

【００４６】

いくつかの態様では、前記第一の相同染色体分節のコピー数の過剰出現があるかどうかの決定が、前記第一の相同染色体分節の過剰出現度を規定する１つ以上の仮説の集合を列挙することを含む。いくつかの態様では、少なくとも１個の細胞中でヘテロ接合である座位（例えば、前記胎児においてヘテロ接合である座位および／または前記母親においてヘテロ接合である座位）の予測アレル比率は、仮説ごとにその仮説によって過剰出現度が規定されたと仮定して推定される。いくつかの態様では、前記算出済みアレル比率と予測アレル比率を比較することで、仮説が正しい尤度が算出され、最も可能性が高い仮説が選択される。いくつかの態様では、仮説ごとに予測アレル比率を用いて検定統計量の期待分布が算出される。いくつかの態様では、前記算出済みアレル比率を用いて算出された検定統計量を、前記予測アレル比率を用いて算出された検定統計量の期待分布と比較することで、仮説が正しい尤度が算出され、最も可能性が高い仮説が選択される。いくつかの態様では、少なくとも１個の細胞中でヘテロ接合である座位（例えば、前記胎児においてヘテロ接合である座位および／または前記母親においてヘテロ接合である座位）の予測アレル比率に、前記第一の相同染色体分節の相決定済み遺伝子データ、前記第二の相同染色体分節の相決定済み遺伝子データ、および過剰出現度がその仮説によって規定されたと仮定して推定される。いくつかの態様では、前記算出済みアレル比率を前記予測アレル比率と比較することで、仮説が正しい尤度が算出され、最も可能性が高い仮説が選択される。

【００４７】

いくつかの態様では、１個以上の目的細胞に由来するＤＮＡ（またはＲＮＡ）の前記試料中の総ＤＮＡ（またはＲＮＡ）に対する比率が算出される。典型的な比率は、胎児のＤＮＡ（またはＲＮＡ）の前記試料中の総ＤＮＡ（またはＲＮＡ）に対する比率である。いくつかの態様では、胎児のＤＮＡの前記試料中の総ＤＮＡに対する比率は、前記胎児がアレルを有するがその母親はアレルを有していない１箇所以上の座位でアレルの量を測定することにより求められる。いくつかの態様では、胎児のＤＮＡの前記試料中の総ＤＮＡに対する比率は、１つ以上の母親のアレルと胎児のアレルのメチル化の違いを測定することで求められる。いくつかの態様では、前記第一の相同染色体分節の過剰出現度を規定する１つ以上の仮説の集合が列挙される。いくつかの態様では、少なくとも１個の細胞中でヘテロ接合である座位（例えば、前記胎児においてヘテロ接合である座位および／または前記母親においてヘテロ接合である座位）の予測アレル比率は、ＤＮＡまたはＲＮＡの算出済み比率を仮定して推定され、その仮説によって規定された過剰出現度が仮説ごとに推定される。いくつかの態様では、前記算出済みアレル比率を前記予測アレル比率と比較することで、仮説が正しい尤度が算出され、最も可能性が高い仮説が選択される。いくつかの態様では、前記予測アレル比率およびＤＮＡまたはＲＮＡの算出済み比率を用いて算出された検定統計量の期待分布は、仮説ごとに推定される。いくつかの態様では、前記算出済みアレル比率とＤＮＡまたはＲＮＡの算出済み比率を用いて算出された検定統計量を、前記予測アレル比率とＤＮＡまたはＲＮＡの算出済み比率を用いて算出された検定統計量の期待分布と比較することで、仮説が正しい尤度が求められ、最も可能性が高い仮説が選択される。

【００４８】

いくつかの態様では、前記方法は、前記第一の相同染色体分節の過剰出現度を規定する１つ以上の仮説の集合を列挙することを含む。いくつかの態様では、前記方法は、仮説ご

とに、（ｉ）過剰出現度がその仮説によって規定されたと仮定して少なくとも１個の細胞中でヘテロ接合である座位（例えば、前記胎児においてヘテロ接合である座位および／または前記母親においてヘテロ接合である座位）の予測アレル比率、または（ｉｉ）ＤＮＡまたはＲＮＡの１つ以上の起こりうる比率（例えば、胎児のＤＮＡまたはＲＮＡの前記試料中の総ＤＮＡまたはＲＮＡに対する比率）について、前記予測アレル比率および前記試料中の総ＤＮＡまたはＲＮＡに対する１個以上の目的細胞（例えば、胎児の細胞）に由来するＤＮＡまたはＲＮＡの起こりうる比率を用いて算出された検定統計量の期待分布のいずれかを推定することを含む。いくつかの態様では、（ｉ）前記予測アレル比率に対する算出済みアレル比率または（ｉｉ）前記算出済みアレル比率および前記ＤＮＡまたはＲＮＡの起こりうる比率を用いて算出された検定統計量のいずれかと、前記予測アレル比率および前記ＤＮＡまたはＲＮＡの起こりうる比率を用いて算出された検定統計量の期待分布とを比較することでデータ適合が算出される。いくつかの態様では、前記データ適合によって１つ以上の仮説がランク付けされ、最高ランクの仮説が選択される。いくつかの態様では、前記データ適合を算出する工程、前記仮説をランク付けする工程、最高ランクの仮説を選択する工程の１つ以上において技術またはアルゴリズム（例えば、検索アルゴリズム等）が用いられる。いくつかの態様では、前記データ適合は、ベータ二項分布に対する適合または二項分布への適合である。いくつかの態様では、前記技術またはアルゴリズムは、最尤推定、最大事後確率推定、ベイズ推定、動的推定（例えば、動的ベイズ推定等）、および期待値最大化推定からなる群から選択される。いくつかの態様では、前記方法は、前記技術またはアルゴリズムを前記得られた遺伝子データおよび前記期待遺伝子データに適用することを含む。

【００４９】

いくつかの態様では、前記方法は、１個以上の目的細胞に由来するＤＮＡまたはＲＮＡの前記試料中の総ＤＮＡまたはＲＮＡに対する比率について下限から上限までの範囲の起こりうる比率（例えば、胎児のＤＮＡまたはＲＮＡの前記試料中の総ＤＮＡまたはＲＮＡに対する比率）の区分を作成することを含む。いくつかの態様では、前記第一の相同染色体分節の過剰出現度を規定する１つ以上の仮説の集合が列挙される。いくつかの態様では、前記方法は、前記区分におけるＤＮＡまたはＲＮＡの起こりうる比率の各比率と各仮説について、（ｉ）前記ＤＮＡまたはＲＮＡの起こりうる比率およびその仮説によって規定された過剰出現度が与えられた場合は少なくとも１個の細胞中でヘテロ接合である座位（例えば、前記胎児においてヘテロ接合である座位および／または前記母親においてヘテロ接合である座位）の予測アレル比率、または（ｉｉ）前記予測アレル比率と前記ＤＮＡまたはＲＮＡの起こりうる比率を用いて算出された検定統計量の期待分布のいずれかを推定することを含む。いくつかの態様では、前記方法は、（ｉ）前記予測アレル比率に対する算出済みアレル比率または（ｉｉ）前記算出済みアレル比率および前記ＤＮＡまたはＲＮＡの起こりうる比率を用いて算出された検定統計量のいずれかと、前記予測アレル比率および前記ＤＮＡまたはＲＮＡの起こりうる比率を用いて算出された検定統計量の期待分布とを比較することで前記区分における前記ＤＮＡまたはＲＮＡの起こりうる比率の各比率と各仮説について、前記仮説が正しい尤度を算出することを含む。いくつかの態様では、前記区分において前記起こりうる比率の比率ごとにその仮説の確率を組み合わせることによって各仮説の複合確率が決定され、最も高い複合確率を有する仮説が選択される。いくつかの態様では、各仮説の複合確率は、前記起こりうる比率が正しい比率であるという尤度に基づいて特定の起こりうる比率について仮説の確率を加重平均することで各仮説が決定される。

【００５０】

一側面において、本発明は、相決定済み遺伝子データまたは相未決定（ｕｎｐｈａｓｅｄ）遺伝子データを用いてある個体由来の１個以上の細胞のゲノムの染色体または染色体分節のコピー数を決定するための方法を特徴とする。いくつかの態様では、前記方法は、各座位に存在する各アレルの量を測定することで試料中の前記染色体または染色体分節にある多型座位の集合の遺伝子データを得ることを含む。いくつかの態様では、前記試料は

、前記個体に由来する１個以上の細胞に由来するＤＮＡまたはＲＮＡの試料、または２個以上の遺伝的に異なる細胞に由来する無細胞ＤＮＡを含む、前記個体由来の無細胞ＤＮＡの混合試料である。いくつかの態様では、前記試料が由来している少なくとも１個の細胞中のヘテロ接合である座位についてアレル比率が算出される。いくつかの態様では、特定の座位の算出済みアレル比率は、そのアレルの１つの測定量を、前記座位のすべてのアレルの総測定量で除したものである。いくつかの態様では、特定の座位の算出済みアレル比率は、アレル（例えば、第一の相同染色体分節のアレル）の１つの測定量を、その座位の１つ以上の他のアレル（例えば、第二の相同染色体分節のアレル）の測定量で除したものである。いくつかの態様では、前記細胞の１つ以上のゲノムの染色体または染色体分節のコピー数を規定する１つ以上の仮説の集合が列挙される。いくつかの態様では、前記検定統計量に基づいて最も可能性が高い仮説が選択され、これにより前記細胞の１つ以上のゲノムの染色体または染色体分節のコピー数が決定される。

【００５１】

一側面において、本発明は、相決定済みまたは相未決定遺伝子データを用いて胎児（例えば、妊娠している母親の胎内の胎児）のゲノムの染色体または染色体分節のコピー数を決定する方法を特徴とする。いくつかの態様では、前記方法は、各座位に存在する各アレルの量を測定することで試料中の前記染色体または染色体分節にある多型座位の集合の遺伝子データを得ることに関する。いくつかの態様では、前記試料は、胎児のＤＮＡまたはＲＮＡおよびその胎児の母親に由来する母親のＤＮＡまたはＲＮＡ（例えば、母親由来の血液試料に由来する、胎児の無細胞ＤＮＡまたはＲＮＡおよび母親の無細胞ＤＮＡまたはＲＮＡを含む無細胞ＤＮＡまたはＲＮＡの混合試料）を含むＤＮＡの混合試料である。いくつかの態様では、前記胎児のヘテロ接合である座位および／または前記母親のヘテロ接合である座位についてアレル比率が算出される。いくつかの態様では、特定の座位の算出済みアレル比率は、前記アレルの１つの測定量を、その座位のすべてのアレルの総測定量で除したものである。いくつかの態様では、特定の座位の算出済みアレル比率が前記アレル（例えば、前記第一の相同染色体分節にあるアレル等）の１つの測定量を、その座位の１つ以上の他のアレル（例えば、第二の相同染色体分節にあるアレル等）の測定量で除したものである。いくつかの態様では、胎児のゲノムの染色体または染色体分節のコピー数を規定する１つ以上の仮説の集合が列挙される。いくつかの態様では、前記検定統計量に基づいて最も可能性が高い仮説が選択され、これによって前記胎児のゲノムの染色体または染色体分節のコピー数が決定される。

【００５２】

いくつかの態様では、前記検定統計量がある仮説の検定統計量分布に属する確率が上限閾値を超える場合にその仮説が選択され、前記検定統計量が１つ以上の仮説の検定統計量分布に属する確率が下限閾値未満である場合にそれら１つ以上の仮説が棄却される。あるいは、前記検定統計量ある仮説の検定統計量分布に属する確率が下限閾値と上限閾値の間である場合、またはその確率が十分に高い信頼度を持って決定されない場合には、その仮説は選択も棄却もされない。いくつかの態様では、前記第一の相同染色体分節のコピー数過剰出現は、前記第一の相同染色体分節の重複または前記第二の相同染色体分節の欠失によるものである。いくつかの態様では、１箇所以上の座位についてすべてのアレルの総測定量が、基準量と比較されて、前記第一の相同染色体分節のコピー数過剰出現が前記第一の相同染色体分節の重複または前記第二の相同染色体分節の欠失によるものであるかどうかが決定される。いくつかの態様では、１箇所以上の座位について前記算出済みアレル比率と前記期待アレル比率との差異の大きさを用いて前記第一の相同染色体分節のコピー数過剰出現が前記第一の相同染色体分節の重複または前記第二の相同染色体分節の欠失によるものかどうかが決定される。いくつかの態様では、前記第一の相同染色体分節のコピー数過剰出現がなく、かつ（例えば、細胞、無細胞ＤＮＡ、無細胞ＲＮＡ、個体、胎児、または胚のゲノム中の）第二の相同染色体分節の過剰出現がない場合に前記第一および第二の相同染色体分節が等しい割合で存在すると決定される。

【００５３】

いくつかの態様では、対象細胞の遺伝型が非対象細胞の遺伝型と異なっており、かつ対象細胞および非対象細胞が二染色体であると予測される１箇所以上の座位に存在する１つ以上のアレルの総量または相対量に基づき、１個以上の目的細胞に由来するＤＮＡの前記試料中の総ＤＮＡに対する比率が決定される。いくつかの態様では、この比率を用いて、前記第一の相同染色体分節のコピー数過剰出現が前記第一の相同染色体分節の重複または前記第二の相同染色体分節の欠失によるものかどうかが決定される。いくつかの態様では、この比率を用いて、重複している染色体分節または染色体の余剰コピー数が決定される。いくつかの態様では、前記相決定済み遺伝子データは確率データを含む。いくつかの態様では、前記胎児のゲノム中の第一の相同染色体分節および／または第二の相同染色体分節の相決定済み遺伝子データを得ることは、前記胎児の生物学上の親の一方または両方のゲノム中の第一の相同染色体分節および／または第二の相同染色体分節の相決定済み遺伝子データを得ることと、前記胎児がどの相同染色体分節をその生物学上の親のどちらから受け継いでいるかを推論することを含む。いくつかの態様では、胎児個体に前記第一の相同染色体分節または第二の相同染色体分節のコピーをもたらした配偶子の形成時に発生した可能性がある１つ以上の交差（例えば１つ、２つ、３つ、または４つの交差）の確率を用いて、前記胎児がどの相同染色体分節を生物学上の親のどちらから受け継いでいるかが推論される。いくつかの態様では、デジタルＰＣＲ、集団に基づくハプロタイプ頻度を用いたハプロタイプの推論、半数体細胞（例えば、精子または卵子）を用いたハプロタイプ決定、１つ以上の第一度親近に由来する遺伝子データを用いたハプロタイプ決定、およびこれらの組み合わせからなる群から選択される技術を用いて胎児の母親および／または父親の相決定済み遺伝子データが得られる。いくつかの態様では、前記個体に由来する試料中の欠失または重複に対応する領域の一部またはすべてについて相決定することで前記個体の相決定済み遺伝子データが得られる。いくつかの態様では、胎児または胎児の母親に由来する試料中の欠失または重複に対応する領域の一部またはすべてについて相決定することで胎児の相決定済み遺伝子データが得られる。いくつかの態様では、前記第一および第二の相同染色体分節の相決定済み遺伝子データを得ることは、これら染色体分節の一方に存在するアレルの同一性と、他方の染色体分節に存在するアレルの同一性を推論によって求めることを含む。いくつかの態様では、相未決定遺伝子データに由来する、第一の相同染色体分節に存在しないアレルは、第二の相同染色体分節に割り当てられる。例えば、ある個体の遺伝型が（ＡＢ，ＡＢ）であり、その個体の相決定済みデータが第一のハプロタイプが（Ａ，Ａ）であることを示している場合、もう一方のハプロタイプは（Ｂ，Ｂ）であると推論することができる。いくつかの態様では、ある座位でアレルが１つのみ測定されると、そのアレルは第一および第二の相同染色体分節の両方の一部であると決定される（例えば、ある座位で遺伝型がＡＡである場合、両方のハプロタイプはアレルＡを有する）。いくつかの態様では、前記個体の相決定済み遺伝子データは、例えば、組換えの起こりやすい箇所の配列と、任意で組換えの起こりやすい箇所に隣接する領域の配列を決定することで１つ以上の起こり得る染色体交差が起こったどうかの決定を含む。いくつかの態様では、本発明のプライマーライブラリーのいずれかを用いて組換え事象を検出して、ある個体のゲノムにどのハプロタイプブロックが存在するかが決定される。

【００５４】

いくつかの態様では、前記方法は、同時分布モデル（例えば、座位間の連鎖を考慮した同時分布モデル等）を用いること、連鎖分析を行うこと、二項分布モデルを用いること、ベータ二項分布モデルを用いること、および／または胎児になった胚を形成した配偶子を発生させた減数分裂時に起こった交差の尤度を用いること（例えば、染色体の異なる位置で染色体が交差する確率を用いて目的の染色体または染色体分節にある多型アレル間の依存度をモデル化すること等）を含む。

【００５５】

いくつかの態様では、無細胞ＤＮＡまたは無細胞ＲＮＡに関する１つ以上の算出済みアレル比率は、その無細胞ＤＮＡまたは無細胞ＲＮＡが由来している細胞における対応するＤＮＡまたはＲＮＡのアレル比率を示している。いくつかの態様では、無細胞ＤＮＡまた

は無細胞ＲＮＡに関する１つ以上の算出済みアレル比率は、前記個体のゲノムにおける対応するアレル比率を示している。いくつかの態様では、測定された遺伝子データが２つ以上の異なるアレルが前記試料中（例えば、無細胞ＤＮＡまたは無細胞ＲＮＡ試料中）のその座位に存在することを示している場合には、アレル比率が算出されるのみ、あるいは期待アレル比率と比較されるのみである。いくつかの態様では、前記座位が、前記試料が由来している細胞の少なくとも１個においてヘテロ接合である（例えば、胎児においてヘテロ接合である座位および／またはその母親においてヘテロ接合である座位等）場合には、アレル比率が算出されるのみ、あるいは期待アレル比率と比較されるのみである。いくつかの態様では、前記座位が胎児においてヘテロ接合である場合には、アレル比率が算出されるのみ、あるいは期待アレル比率と比較されるのみである。いくつかの態様では、アレル比率が算出されて、ホモ接合座位の期待アレル比率と比較される。例えば、検査される特定の個体（または胎児および妊娠している母親の両方）についてホモ接合が予測される座位のアレル比率を分析して、システムのノイズまたはエラーの量を決定してもよい。

【００５６】

いくつかの態様では、目的の染色体または染色体分節について少なくとも１０個、５０個、１００個、２００個、３００個、５００個、７５０個、１，０００個、２，０００個、３，０００個、４，０００個以上の（例えばＳＮＰ）座位が分析される。いくつかの態様では、目的の染色体または染色体分節のメガ塩基対（ｍｂ）当たりの（例えばＳＮＰ）座位の平均数は、メガ塩基対あたり少なくとも１個、１０個、２５個、５０個、１００個、１５０個、２００個、３００個、５００個、７５０個、１，０００個以上である。いくつかの態様では、目的の染色体または染色体分節のメガ塩基対当たりの座位（例えばＳＮＰ）の平均数は、メガ塩基対当たり１個と５００個の間（例えば、メガ塩基対当たり１個と５０個の間、５０個と１００個の間、１００個と２００個の間、２００個と４００個の間、２００個と３００個の間、または３００個と４００個、各両端の数値を含む）である。いくつかの態様では、起こり得る欠失または重複の複数の部分の座位が分析されて、１座位のみの分析や隣接するいくつかの座位のみの分析に比べてＣＮＶ決定の感度および／または特異性が高められる。いくつかの態様では、各座位に存在する最も共通する２つのアレルのみが測定される、あるいはこれらアレルのみが用いられて算出済みアレル比率が決定される。いくつかの態様では、５’→３’エキソヌクレアーゼ活性および／または低鎖置換活性を有するポリメラーゼ（例えば、ＤＮＡポリメラーゼ、ＲＮＡポリメラーゼ、または逆転写酵素等）を用いて座位の増幅が行われる。いくつかの態様では、（ｉ）前記試料中のＤＮＡまたはＲＮＡの配列決定を行うこと、（ｉｉ）前記試料中のＤＮＡまたはＲＮＡを増幅して、増幅されたＤＮＡを配列決定すること、または（ｉｉｉ）前記試料中のＤＮＡまたはＲＮＡを増幅して、ＰＣＲ生成物をライゲートして、ライゲートした生成物を配列決定することで前記測定されたアレルの遺伝子データが得られる。いくつかの態様では、前記試料に由来するＤＮＡまたはＲＮＡを複数の分画に分割して、（例えば、特定の分画のすべてのＤＮＡまたはＲＮＡは同じバーコードを有する等）各分画のＤＮＡまたはＲＮＡに異なるバーコードを付加して、任意でバーコード付きＤＮＡまたはＲＮＡを増幅して、分画を結合させ、結合させた分画のバーコード付きＤＮＡまたはＲＮＡを配列決定することで測定されたアレルの遺伝子データが得られる。いくつかの態様では、配列決定（例えば、ナノポア配列決定、Halcyon Molecular配列決定等）、ＳＮＰアレイ、リアルタイムＰＣＲ、TaqMan、Nanostring nCounter（登録商標）分析装置、差別的ＤＮＡポリメラーゼおよびリガーゼを用いたIllumina GoldenGate Genotyping Assay、ライゲーション媒介ＰＣＲ、連結逆方向プローブ（ＬＩＰ、環状化済みプローブ、予備環状化型プローブ、環状化型プローブ、パドロックプローブ、または分子反転プローブ（ＭＩＰ）と呼ばれることもある）のうちの１つ以上の方法を用いて、多型座位（例えばＳＮＰ）のアレルが同定される。いくつかの態様では、２つ以上（例えば、３または４）の対象増幅産物がライゲーションで結合されて、ライゲート生成物の配列決定が行われる。いくつかの態様では、同じ座位について異なるアレルの測定は、これらアレル間の代謝、アポトーシス、ヒストン、不活性、および／または増幅の違い（例えば、同じ座位の異なるアレル間

の増幅効率の違い等）についての調整が行われる。いくつかの態様では、この調整は、得られた遺伝子データのアレル比率を算出する前、あるいは測定遺伝子データを期待遺伝子データと比較する前に行われる。

【００５７】

いくつかの態様では、前記方法はまた、病気または障害の１種以上の危険因子の有無を決定することを含む。いくつかの態様では、前記方法はまた、病気または障害あるいはその病気または障害の危険性の増加に関連づけられた１種以上の多型または変異の有無を決定することを含む。いくつかの態様では、前記方法はまた、無細胞ＤＮＡ、無細胞ミトコンドリアＤＮＡ、無細胞核内ＤＮＡ、無細胞ＲＮＡ、ｍｉＲＮＡ、またはこれらの任意の組み合わせの総量を決定することを含む。いくつかの態様では、前記方法は、１つ以上の目的の無細胞ＤＮＡ、無細胞ミトコンドリアＤＮＡ、無細胞核内ＤＮＡ、無細胞ＲＮＡ、および／またはｍｉＲＮＡ分子（例えば、病気または障害あるいはその病気または障害の危険性の増加に関連づけられた多型または変異を有する分子等）の量を決定することを含む。いくつかの態様では、総ＤＮＡ中の腫瘍ＤＮＡ分画（例えば、総無細胞ＤＮＡ中の腫瘍無細胞ＤＮＡ分画、総無細胞ＤＮＡ中の特定の変異を有する腫瘍無細胞ＤＮＡ分画等）が決定される。いくつかの態様では、この腫瘍分画を用いて癌の段階が決定される（腫瘍分画が高い程、癌のより進行した段階と関連づけることができるため）。いくつかの態様では、前記方法はまた、ＤＮＡまたはＲＮＡの総量を決定することを含む。いくつかの態様では、前記方法は、メチル化された１種以上の目的のＤＮＡまたはＲＮＡ分子（例えば、病気または障害あるいはその病気または障害の危険性の増加に関連づけられた多型または変異を有する分子等）の量を決定することを含む。いくつかの態様では、前記方法は、ＤＮＡの完全性における変化の有無を決定することを含む。いくつかの態様では、前記方法はまた、ｍＲＮＡスプライシングの総量を決定することを含む。いくつかの態様では、前記方法は、ｍＲＮＡスプライシングの量を決定すること、または目的の１つ以上のＲＮＡ分子（例えば、病気または障害あるいはその病気または障害の危険性の増加に関連づけられた多型または変異を有する分子）に対応するもう１つのｍＲＮＡスプライシングを検出することを含む。

【００５８】

いくつかの態様では、本発明は、個体において癌の表現型を検出するための方法であって、前記癌の表現型が変異の集合のうちの少なくとも１つの変異の存在によって定義されることを特徴とする。いくつかの態様では、前記方法は、前記細胞の１つ以上が癌の表現型を有すると疑われる場合に前記個体に由来する１個以上の細胞に由来するＤＮＡまたはＲＮＡの試料についてＤＮＡまたはＲＮＡ測定を得ることと、前記ＤＮＡまたはＲＮＡ測定を分析して前記変異の集合のうちの変異ごとに前記細胞の少なくとも１つがその変異を有する尤度を決定することを含む。いくつかの態様では、前記方法は、（ｉ）前記変異の少なくとも１つについて、前記細胞の少なくとも１つがその変異を含む尤度が閾値よりも大きい場合、または（ｉｉ）前記変異の少なくとも１つの変異について、前記細胞の少なくとも１つがその変異を有する尤度が前記閾値未満である場合のいずれかであり、かつ複数の変異について、前記細胞の少なくとも１つが前記変異の少なくとも１つを有する尤度の組み合わせが前記閾値より大きい場合に前記個体が前記癌の表現型を有していると決定することを含む。いくつかの態様では、１個以上の細胞は、前記変異の集合内に下位集合または変異のすべてを有する。いくつかの態様では、前記変異の下位集合は、癌または癌の危険性の増加と関連づけられる。いくつかの態様では、前記試料は、無細胞ＤＮＡまたはＲＮＡを含む。いくつかの態様では、前記ＤＮＡまたはＲＮＡ測定は、１つ以上の染色体または染色体分節にある多型座位の集合における測定（例えば、各座位に存在する各アレルの量等）を含む。

【００５９】

一側面において、本発明は、哺乳類の病気または障害の治療、安定化、または予防のための治療法を選択する方法を特徴とする。いくつかの態様では、前記方法は、本明細書で開示した方法のいずれかを用いて、第二の相同染色体分節に比較して第一の相同染色体分

節のコピー数の過剰出現があるかどうかを決定することを含む。いくつかの態様では、前記哺乳類に対して治療法（例えば、第一の相同染色体分節の過剰出現と関連づけられた病気または障害の治療法等）が選択される。

【００６０】

一側面において、本発明は、哺乳類の病気または障害を防止、遅延、安定化、または治療するための方法を特徴とする。いくつかの態様では、前記方法は、本明細書で開示した方法のいずれかを用いて、第二の相同染色体分節と比較して第一の相同染色体分節のコピー数の過剰出現があるかどうかを決定することを含む。いくつかの態様では、治療法（例えば、第一の相同染色体分節の過剰出現と関連づけられた病気または障害の治療法）が選択されて、前記哺乳類に施される。

【００６１】

いくつかの態様では、病気または障害を治療、安定化、または防止することは、病気または障害の初期の発生またはそれに続いて起こる事象を防止または遅延すること、状態の消失からその再発までの無病生存時間を増加させること、状態に関連づけられた有害な症状を安定化または低下させること、あるいは状態の進行を阻害または安定化させることを含む。いくつかの態様では、治療された対象者の少なくとも２０、４０、６０、８０、９０、または９５％が、前記状態のすべての兆候が消え完全に沈静している。いくつかの態様では、ある状態と診断されて治療された後の対象者の生存時間は、（ｉ）未治療の対象者の平均生存時間または（ｉｉ）他の治療法で治療された対象者の平均生存時間の少なくとも２０、４０、６０、８０、１００、２００、または５００％延長されている。

【００６２】

いくつかの態様では、癌を治療、安定化、または防止することは、腫瘍（例えば、良性腫瘍または悪性腫瘍）の大きさを縮小する、あるいは安定化させること、腫瘍の増大を遅延または防止させること、腫瘍細胞数を低減または安定化させること、腫瘍の消失からその再発までの無病生存時間を増加させること、腫瘍の初期の発生またはそれに続く発生を防止すること、あるいは腫瘍と関連づけられた有害状態を低減または安定化させることを含む。一態様では、任意の標準的なアッセイを用いた測定において、治療後に残った癌性細胞数は、癌性細胞の初期数に比べて少なくとも１０、２０、４０、６０、８０、または１００％低減している。いくつかの態様では、本発明の治療法を施すことによって誘発された癌性細胞数は、非癌細胞数に比べて少なくとも２、５、１０、２０、または５０倍超、低減している。いくつかの態様では、治療法を施した後に存在する癌性細胞数は、対照（例えば、生理食塩水または緩衝剤の投与等）の投与後に存在する癌性細胞数に比べて少なくとも２、５、１０、２０、または５０倍低減している。いくつかの態様では、本発明の方法によって、標準的な方法を用いて決定した腫瘍の大きさが１０、２０、４０、６０、８０、または１００％縮小した。いくつかの態様では、治療された対象者の少なくとも１０、２０、４０、６０、８０、９０、または９５％が癌性細胞の検出されない完全な沈静を示した。いくつかの態様では、癌が全く再発しない、あるいは少なくとも２、５、１０、１５、または２０年間は再発しない。いくつかの態様では、癌と診断されて、本発明の治療法で治療された対象者の生存時間は、（ｉ）未治療の対象者の平均生存時間または（ｉｉ）他の治療法で治療された対象者の平均生存時間の少なくとも１０、２０、４０、６０、８０、１００、２００、または５００％延長されている。

【００６３】

一側面において、本発明は、哺乳類の病気または障害を治療、安定化、または予防するための臨床試験に含まれる対象者を層別化するための方法を特徴とする。いくつかの態様では、前記方法は、本明細書で開示した方法のいずれかを用いて、前記臨床試験の前、最中、または後に第二の相同染色体分節と比較して第一の相同染色体分節のコピー数の過剰出現があるかどうかを決定することを含む。いくつかの態様では、前記対象者のゲノムにおける第一の相同染色体分節の過剰出現の有無によって前記対象者が臨床試験のための下位群に振り分けられる。

【００６４】

いくつかの態様では、前記病気または障害は、癌、精神障害、学習障害（例えば、特発性学習障害）、精神遅滞、発達遅滞、自閉症、神経変性病または障害、統合失調症、身体障害、自己免疫疾患または障害、全身性エリテマトーデス、乾癬、クローン病、糸球体腎炎、ＨＩＶ感染、ＡＩＤＳ、およびこれらの組み合わせからなる群から選択される。いくつかの態様では、前記病気または障害は、ディ・ジョージ(DiGeorge)症候群、ディ・ジョージ２症候群、ディ・ジョージ/口蓋帆・心臓・顔(VCFS)症候群、プラダー・ウィリー(Prader-Willi)症候群、アンジェルマン(Angelman)症候群、ベックウィズ-ヴィーデマン(Beckwith-Wiedemann)症候群、１ｐ３６欠失症候群、２ｑ３７欠失症候群、３ｑ２９欠失症候群、９ｑ３４欠失症候群、１７ｑ２１．３１欠失症候群、猫鳴き（Cri-du-chat）症候群、ヤコブセン(Jacobsen)症候群、ミラー・ディッカー(Miller Dieker)症候群、フェラン-マクダーミド(Phelan-McDermid)症候群、スミス・マゲニス(Smith-Magenis)症候群、腫瘍-無虹彩症-泌尿生殖器奇形-精神遅滞(WAGR)症候群、ウォルフ・ヒルシュホーン(Wolf-Hirschhorn)症候群、ウィリアムズ(Williams)症候群、ウィリアムズ・ボイレン(Williams-Beuren)症候群、ミラー・ディッカー(Miller-Dieker)症候群、フェラン-マクダーミド(Phelan-McDermid)症候群、スミス・マゲニス(Smith-Magenis)症候群、ダウン症候群、エドワーズ(Edward)症候群、パトウ(Patau)症候群、クラインフェルター(Klinefelter)症候群、ターナー(Turner)症候群、４７，ＸＸＸ症候群、４７，ＸＹＹ症候群、ソトス(Sotos)症候群、およびこれらの組み合わせからなる群から選択される。いくつかの態様では、前記方法は、以下の染色体異常の１種以上の有無を決定する：零染色体性、一染色体性、片親性二染色体性、三染色体性、一致型三染色体性、部分一致型三染色体性、母親性三染色体性、親性三染色体性、三杯体性、モザイク型四染色体性、一致型四染色体性、部分一致型四染色体性、他の異数性、非均衡型転座、均衡型転座、挿入、欠失、組換え、およびこれらの組み合わせ。いくつかの態様では、前記染色体異常は、特定の染色体または染色体分節の最も一般的なコピー数との任意のズレであり、例えばヒト体細胞では、２コピーからずれているものはいずれも染色体異常と見なすことができる。いくつかの態様では、前記方法は、正倍数性の有無を決定する。いくつかの態様では、前記コピー数仮説は、単胎妊娠に関する１つ以上のコピー数仮説を含む。いくつかの態様では、前記コピー数仮説は、多胎妊娠、例えば双胎妊娠（例えば、同一性双生児、兄弟性双生児、消失性双生児）に関する１つ以上のコピー数仮説を含む。いくつかの態様では、前記コピー数仮説は、多胎妊娠の正倍数体であるすべての胎児、多胎妊娠の異数体（例えば、本明細書において開示される任意の異数性）であるすべての胎児、および／または多胎妊娠の正倍数体である１体以上の胎児、および多胎妊娠の異数性である１体以上の胎児を含む。いくつかの態様では、前記コピー数仮説は、同一性双生児（一卵性双生児ともいう）または兄弟性双生児（二卵性双生児ともいう）を含む。いくつかの態様では、前記コピー数仮説は、奇胎妊娠（例えば、完全奇胎妊娠、部分奇胎妊娠等）を含む。いくつかの態様では、前記目的の染色体分節は１本の染色体全体である。いくつかの態様では、前記染色体または染色体分節は、１３番染色体、１８番染色体、２１番染色体、Ｘ染色体、Ｙ染色体、これらの分節、およびこれらの組み合わせからなる群から選択される。いくつかの態様では、前記第一の相同染色体分節および前記第二の相同染色体分節は、前記目的の染色体分節を含む相同染色体分節の対である。いくつかの態様では、前記第一の相同染色体分節および前記第二の相同染色体分節は目的の相同染色体の対である。いくつかの態様では、信頼度は、ＣＮＶの決定または病気または障害の診断について算出される。

【００６５】

いくつかの態様では、前記欠失は、少なくとも０．０１ｋｂ、０．１ｋｂ、１ｋｂ、１０ｋｂ、１００ｋｂ、１ｍｂ、２ｍｂ、３ｍｂ、５ｍｂ、１０ｍｂ、１５ｍｂ、２０ｍｂ、３０ｍｂ、または４０ｍｂの欠失である。いくつかの態様では、前記欠失は、１ｋｂ～４０ｍｂ、例えば１ｋｂ～１００ｋｂ、１００ｋｂ～１ｍｂ、１～５ｍｂ、５～１０ｍｂ、１０～１５ｍｂ、１５～２０ｍｂ、２０～２５ｍｂ、２５～３０ｍｂ、または３０～４０ｍｂ（数値を含む）の欠失である。いくつかの態様では、前記染色体分節の１コピーが欠失しており、１コピーは存在する。いくつかの態様では、前記染色体分節の２コピーが

欠失している。いくつかの態様では、１本の染色体全体が欠失している。

【００６６】

いくつかの態様では、前記重複は、少なくとも０．０１ｋｂ、０．１ｋｂ、１ｋｂ、１０ｋｂ、１００ｋｂ、１ｍｂ、２ｍｂ、３ｍｂ、５ｍｂ、１０ｍｂ、１５ｍｂ、２０ｍｂ、３０ｍｂ、または４０ｍｂの重複である。いくつかの態様では、前記重複は、１ｋｂ～４０ｍｂ、例えば１ｋｂ～１００ｋｂ、１００ｋｂ～１ｍｂ、１～５ｍｂ、５～１０ｍｂ、１０～１５ｍｂ、１５～２０ｍｂ、２０～２５ｍｂ、２５～３０ｍｂ、３０～４０ｍｂ（数値を含む）の重複である。いくつかの態様では、前記染色体分節の重複は１回である。いくつかの態様では、前記染色体分節の重複は２回以上（例えば、２、３、４、５回）である。いくつかの態様では、１本の染色体全体が重複している。いくつかの態様では、第一の相同分節のある領域が欠失しており、第二の相同分節の同じ領域または他の領域が重複している。いくつかの態様では、検査したＳＮＶの少なくとも５０、６０、７０、８０、９０、９５、９６、９８、９９、１００％は転換型変異であり、転移型変異ではない。

【００６７】

いくつかの態様では、前記試料は、（ｉ）１個以上の目的細胞または（ｉｉ）１個以上の目的でない細胞に由来するＤＮＡおよび／またはＲＮＡを含む。いくつかの態様では、前記試料は、１個以上の目的細胞および１個以上の目的でない細胞に由来するＤＮＡおよび／またはＲＮＡを含む混合試料である。いくつかの態様では、前記対象細胞はＣＮＶ（例えば、目的の欠失または重複）を有する細胞であり、前記非対象細胞は目的のコピー数多様性を有しない細胞である。前記１個以上の目的細胞が癌細胞であり、前記１個以上の目的でない細胞が非癌細胞であるいくつかの態様では、前記方法は、前記癌細胞のうちの１個以上の癌細胞のゲノム中の前記第一の相同染色体分節のコピー数過剰出現があるかどうかを決定することを含む。前記１個以上の目的細胞が遺伝的に同一の癌細胞であり、前記１個以上の目的でない細胞が非癌細胞であるいくつかの態様では、前記方法は、前記癌細胞のゲノム中の前記第一の相同染色体分節のコピー数過剰出現があるかどうかを決定することを含む。前記１個以上の目的細胞が遺伝的に非同一の癌細胞であり、前記１個以上の目的でない細胞が非癌細胞であるいくつかの態様では、前記方法は、前記遺伝的に非同一の癌細胞のうちの１個以上の癌細胞のゲノム中の前記第一の相同染色体分節のコピー数過剰出現があるかどうかを決定することを含む。前記試料が１個以上の癌細胞および１つ以上の非癌細胞の混合物に由来する無細胞ＤＮＡを含むいくつかの態様では、前記方法は、前記癌細胞のうちの１個以上の癌細胞のゲノム中の前記第一の相同染色体分節のコピー数過剰出現があるかどうかを決定することを含む。前記１個以上の目的細胞が遺伝的に同一の胎児の細胞であり、前記１個以上の目的でない細胞が母親の細胞であるいくつかの態様では、前記方法は、前記胎児の細胞のゲノム中の前記第一の相同染色体分節のコピー数過剰出現があるかどうかを決定することを含む。前記１個以上の目的細胞が遺伝的に非同一の胎児の細胞であり、前記１個以上の目的でない細胞が母親の細胞であるいくつかの態様において、前記方法は、前記遺伝的に非同一の胎児の細胞のうちの１個以上の細胞のゲノム中の前記第一の相同染色体分節のコピー数過剰出現があるかどうかを決定することを含む。大部分の個体の細胞は細胞核ＤＮＡのほぼ同一の集合を含むので、いくつかの態様では用語「対象細胞」は、用語「個体」とは区別なく用いてもよい。癌性細胞は、宿主である個体とは異なる遺伝型を有する。この場合、その癌自体を個体と見なしてもよい。また、多くの癌は異種性であるが、これはある腫瘍の各細胞が同じ腫瘍の他の細胞と遺伝的に異なっていることを意味する。この場合、遺伝的に同一の領域をそれぞれ異なる個体と見なすことができる。あるいは、その癌を、異なるゲノムを有する細胞の混合物を含む単一の個体を見なしてもよい。非対象細胞は、通常、正倍数体であるが、必ずしもそうでない場合がある。

【００６８】

いくつかの態様では、前記試料は、母親の全血液試料またはその分画、母親の血液試料から単離した細胞、羊水穿刺試料、受胎試料の生成物、胎盤組織試料、絨毛膜絨毛試料、

胎盤膜試料、頸管粘液試料、または胎児由来の試料から得られる。いくつかの態様では、前記試料は、前記母親由来の血液試料またはその分画から得られた無細胞ＤＮＡを含む。いくつかの態様では、前記試料は、胎児の細胞および母親の細胞の混合物から得られた細胞核ＤＮＡを含む。いくつかの態様では、前記試料は、胎児の細胞が富化された、有核細胞を含む母親の血液の分画から得られる。いくつかの態様では、前記試料は、複数個の分画（例えば、２、３、４、５個以上の分画）に分割されて、本発明の方法を用いてそれぞれ分析される。各分画が同じ結果（例えば、目的の１つ以上のＣＮＶの有無）を示す場合は、その結果の信頼度が増す。各分画が異なる結果を示す場合は、その試料が再分析される、あるいは同じ対象者から他の試料が採取されて分析されることがある。

【００６９】

対象者の例としては、哺乳類（例えば、ヒトおよび獣医学の目的の哺乳類）等が挙げられる。いくつかの態様では、前記哺乳類は霊長類（例えば、ヒト、サル、ゴリラ、類人猿、キツネザル等）、ウシ、ウマ、ブタ、イヌ、またはネコである。

【００７０】

いくつかの態様では、前記方法のいずれかは、本発明の前記方法の結果（例えば、欠失または重複の有無）を開示する報告書（例えば、文書による報告書、または電子報告書）を作成することを含む。

【００７１】

いくつかの態様では、前記方法のいずれかは、本発明の前記方法の結果（例えば、欠失または重複の有無）に基づいて臨床行為を行うことを含む。本発明の前記方法の結果に基づき、胚または胎児が１つ以上の目的の多型または変異（例えば、ＣＮＶ）を有するいくつかの態様では、前記臨床行為は、追加試験（例えば、多型または変異の存在を確認するための試験等）を行うことを含むが、体外受精のための前記胚を移植すること、体外受精のための別の胚を移植すること、妊娠状態を終了させること、特別支援が必要な子供に備えること、あるいは遺伝性疾患の表現型提示の重症度を低減するように設計された介入を行うことは含まない。いくつかの態様では、前記臨床行為は、超音波診断、胎児の羊水穿刺、母親および／または父親に由来する遺伝子材料を受け継いだ胎児の羊水穿刺、胎児の絨毛膜絨毛生検、母親および／または父親に由来する遺伝子材料を受け継いた胎児の絨毛膜絨毛生検、体外受精、母親および／または父親に由来する遺伝子材料を受け継いだ１個以上の胚の着床前遺伝子診断、母親の染色体分析、父親の染色体分析、胎児の心エコー図検査（例えば、三染色体性２１、１８、または１３、一染色体性Ｘ、または微少欠失を有する胎児の心エコー図検査等）、およびこれらの組み合わせからなる群から選択される。いくつかの態様では、前記臨床行為は、一染色体性Ｘを有する生まれた子供への成長ホルモンの投与（例えば、約９ヶ月での投与開始）、２２ｑ欠失（例えば、ディ・ジョージ症候群）を持って生まれた子供へのカルシウム投与、４７，ＸＸＹを有する生まれた子供へのテストステロン等のアンドロゲンの投与（例えば、３ヶ月間、乳児または幼児への１ヶ月あたり２５ｍｇのエナント酸テストステロンの注射）、（例えば、三倍体胎児等の）完全奇胎妊娠または部分奇胎妊娠の女性に対する癌検査の実施、（例えば、三倍体胎児等の）完全奇胎妊娠または部分奇胎妊娠の女性に対する（化学療法薬等による）癌治療の実施、男子と判定された胎児（例えば、本発明の方法を用いて男子と判定された胎児）に対する１種以上のＸ関連遺伝性疾患（例えば、デュシェンヌ型筋ジストロフィー（ＤＭＤ）、副腎白質ジストロフィー、血友病等）についてのスクリーニング検査、Ｘ関連疾患の危険性がある男子の胎児の羊水穿刺の実施、先天性副腎過形成の危険性がある女子の胎児（例えば、本発明の方法を用いて女子と判定された胎児）を妊娠している女性へのデキサメタゾンの投与、先天性副腎過形成の危険性がある男子の胎児について羊水穿刺の実施、（生ワクチンの代わりに）死菌ワクチンの投与または２２ｑ１１．２欠失による免疫不全（または、免疫不全の疑いがある）生まれた子供へのある種のワクチン投与の回避、作業療法および／または物理治療の実施、教育への早期介入、新生児集中治療室および／または出産時に対応可能な小児専門医を有する三次医療センターへの赤ん坊の搬送、生まれた子供（例えば、ＸＸＸ、ＸＸＹ、またはＸＹＹを有する子供）への行動介入、およびこれらの

組み合わせからなる群から選択される。

【００７２】

いくつかの態様では、多胎妊娠（例えば、双生児）であると判定された女性に対して超音波または他のスクリーニング検査が行われて、胎児の２体以上が単一絨毛膜性であるかどうかが決定される。一卵性双生児は、単一の卵母細胞の排卵および受精と、それに続く受精卵の細胞分裂から発生する。胎盤形成は二絨毛膜性または単一絨毛膜性である可能性がある。二卵性双生児は、２つの卵母細胞の排卵および受精から発生するが、通常、二絨毛膜性胎盤が形成される。単一絨毛膜性双生児は双胎児間輸血症候群の危険性を有する。これら症候群によって胎児間の血液分布が不均衡になる可能性があり、その結果、双生児の成長や発達に差が生じ、場合によっては死産となる。よって、本発明の方法を用いて一卵性双生児と判定された双生児は、（例えば、超音波等によって）検査されて、単一絨毛膜性双生児であるかどうかが判定され、単一絨毛膜性双生児である場合は双胎児間輸血症候群の兆候についてこれらの双生児を（例えば、１６週目から隔週で超音波等により）監視することができることが望ましい。

【００７３】

本発明の前記方法の結果に基づいて胚または胎児が１つ以上の目的の多型または変異（例えば、ＣＮＶ）を有していないいくつかの態様では、前記臨床行為は、体外受精のための前記胚を移植すること、または妊娠状態を継続することを含む。いくつかの態様では、前記臨床行為は、超音波、羊水穿刺、絨毛膜絨毛生検、およびこれらの組み合わせからなる群から選択される行為を行って前記多型または変異がないことを確認するための追加検査である。

【００７４】

本発明の前記方法の結果に基づいてある個体が１つ以上の多型または変異（例えば、（癌等の）病気または障害または（癌等の）病気または障害の危険性の増加に関連づけられた多型または変異）を有するいくつかの態様では、前記臨床行為は、追加検査を行うこと、または病気または障害に対する１つ以上の治療法（例えば、癌の治療法、前記個体が診断された特定の種類の癌または特定の種類の変異に対する治療法、または本明細書で開示された治療法のいずれか）を施すことを含む。いくつかの態様では、前記臨床行為は、生検、手術、医療画像診断（例えば、マンモグラム、超音波等）、およびこれらの組み合わせからなる群から選択される、多型または変異の有無を確認するための追加検査である。

【００７５】

いくつかの態様では、前記追加検査は、多型または変異（例えば、ＣＮＶ）の有無を確認するための同じまたは異なる方法（例えば、本明細書で開示した方法のいずれか）を行うこと（例えば、検査された同じ試料の第二の分画、または同じ個体（例えば、癌の危険性が増加している同じ妊娠している母親、胎児、胚、または個体）に由来する異なる試料を検査する等）を含む。いくつかの態様では、多型または変異（例えば、ＣＮＶ）の確率が閾値を超えている個体に対して前記追加検査（例えば、可能性の高い多型または変異の存在を確認するための追加検査等）が行われる。いくつかの態様では、多型または変異（例えば、ＣＮＶ）の決定の信頼度またはＺ値が閾値を超えている個体に対して前記追加検査（例えば、可能性の高い多型または変異の存在を確認するための追加検査等）が行われる。いくつかの態様では、多型または変異（例えば、ＣＮＶ）の決定のための信頼度またはＺ値が最少閾値と最大閾値の間である個体について、前記追加検査（例えば、初期の結果が正しいことの信頼度を高めるための追加検査等）が行われる。いくつかの態様では、多型または変異（例えば、ＣＮＶ）の有無の決定のための信頼度が閾値未満（例えば、ＣＮＶの有無を十分な信頼度でもって決定できなかった結果である「呼び出しなし」等）である個体について、前記追加検査が行われる。Ｚ値は例えばChiu et al. BMJ 2011;342:c7401（参照によりその全体が本明細書に組み込まれる）で算出される。ここでは、一例として２１番染色体が用いられているが、検査試料中のいずれか他の染色体または染色体分節と置換されてもよい。

検査ケースの２１番染色体の百分率に対するＺ値＝（（検査ケースの２１番染色体の百分

率）－（基準対照の２１番染色体の百分率の平均値））／（基準対照の２１番染色体の百分率の標準偏差）
いくつかの態様では、最初の試料が品質管理ガイドラインを満たしてなかった、あるいは閾値未満の胎児の分画または腫瘍分画を含んでいた個体について、前記追加検査が行われる。いくつかの態様では、前記方法は、本発明の前記方法の結果、その結果の確率、その結果の信頼度、またはＺ値に基づいて追加検査を行う個体を選択し、前記個体（例えば、同じまたは別の試料）に前記追加検査を行うことを含む。いくつかの態様では、（例えば、癌等の）病気または障害と診断された対象者は、本発明の方法または前記病気または障害を検査するための周知の検査方法を用いて複数の時点で反復検査を受けて、前記病気または障害の進行、沈静、または再発が監視される。

【００７６】
一側面において、本発明は、本発明の方法の結果（例えば、欠失または重複の有無）を開示する報告書（例えば、文書による報告書、または電子報告書）を特徴とする。

【００７７】
様々な態様では、前記プライマー伸長反応すなわち前記ポリメラーゼ連鎖反応は、ポリメラーゼで１個以上のヌクレオチドを付加させることを含む。いくつかの態様では、前記プライマーは溶液中にある。いくつかの態様では、前記プライマーは溶液中にあり、固体の支持体に固定化されていない。いくつかの態様では、前記プライマーはマイクロアレイの一部ではない。様々な態様では、前記プライマー伸長反応すなわち前記ポリメラーゼ連鎖反応は、ライゲーション媒介ＰＣＲを含まない。様々な態様では、前記プライマー伸長反応すなわち前記ポリメラーゼ連鎖反応は、リガーゼで２本のプライマーを結合させることを含まない。様々な態様では、前記プライマーは、連結逆方向プローブ（ＬＩＰ、環状化済みプローブ、予備環状化型プローブ、環状化型プローブ、パドロックプローブ、または分子反転プローブ（ＭＩＰ）と呼ばれることもある）を含まない。

【００７８】
本明細書で記載された本発明の側面および態様は、本発明の側面または態様のうち、いずれか２つ以上の組み合わせを含む。

定義

【００７９】
「一塩基多型（ＳＮＰ）」は、同じ種の２個体間のゲノムで異なる場合がある１つのヌクレオチドを指す。この用語の使用は、各変異体が発生する頻度に関するいかなる限定も意味する意図はない。

【００８０】
「配列」は、ＤＮＡ配列または遺伝子配列を指す。この用語は、個体のＤＮＡ分子またはＤＮＡ鎖の一次的物理構造を指す場合がある。この用語は、ＤＮＡ分子に見られるヌクレオチド配列、またはこのＤＮＡ分子の相補鎖を指す場合がある。この用語は、コンピュータでの表示として、このＤＮＡ分子に含まれる情報を指す場合がある。

【００８１】
「座位」は、個体のＤＮＡに関する目的の特定の領域を指し、ＳＮＰ、挿入または欠失の可能性がある部位、あるいはその他のなんらかの関連する遺伝的多様性の部位を指す場合がある。疾病関連ＳＮＰはまた、疾病関連座位という場合がある。

【００８２】
「多型アレル」および「多型座位」はそれぞれ、ある種の個体間で遺伝型が異なるアレル、座位を指す。多型アレルのいくつかの例として、一塩基多型、短縦列型反復、欠失、重複、および逆位等が挙げられる。

【００８３】
「多型部位」は、多型領域に見られる、個体間で異なる特定のヌクレオチドを指す。

【００８４】
「変異」は、天然の核酸配列または基準核酸配列の変化を指し、挿入、欠失、重複、転座、置換、読み枠シフト変異、サイレント変異、ナンセンス変異、ミスセンス変異、点変

異、塩基転位変異、塩基転換変異、復帰変異、マイクロサテライト変更等が挙げられる。いくつかの態様では、核酸配列によってコードされるアミノ酸配列は、天然の配列から少なくとも１つのアミノ酸変更を有する。

【００８５】

「アレル」は、特定の座位を占める遺伝子群を指す。

【００８６】

「遺伝子データ」、「遺伝型データ」はいずれも、１つ以上の個体のゲノムの側面を記述したものを指す。この用語は、座位または座位の集合、配列の部分または配列全体、染色体の部分または染色体全体、あるいはゲノム全体を指す場合がある。この用語は、１個または複数のヌクレオチドの同一性を指す場合がある。また、ヌクレオチド配列、ゲノム中の異なる位置にあるヌクレオチド、またはこれらの組み合わせを指す場合がある。しかしながら、遺伝型データは典型的にはコンピュータ内のものであり、ある配列の物理的ヌクレオチドを、化学的にコードされた遺伝子データとして考慮することも可能である。遺伝型データは、個体「上」、「の」、「で」、「から」、または「に関する」遺伝型データという場合がある。遺伝型データは、遺伝子材料を測定する遺伝型決定プラットフォームから得られる測定出力を指す場合がある。

【００８７】

「遺伝子材料」および「遺伝子試料」は、ＤＮＡまたはＲＮＡを含む１つ以上の個体由来の物質（組織または血液等）を指す。

【００８８】

「信頼度」は、呼び出されたＳＮＰ、アレル、アレルの集合、染色体または染色体分節の決定されたコピー数、あるいは病気の有無の診断が、その個体の実際の遺伝子状態を正しく表していることについての統計尤度を指す。

【００８９】

「倍数性呼び出し」、および「染色体コピー数呼び出し」あるいは「コピー数呼び出し」（ＣＮＣ）は、ある細胞に存在する１つ以上の染色体または染色体分節の量および／または染色体同一性を決定する動作を指す場合がある。

【００９０】

「異数性」は、ある細胞に染色体数の誤り（例えば、完全な染色体の数の誤り、染色体分節の数の誤り（染色体分節の欠失または重複の存在等）が存在する状態を指す。ヒト体細胞の場合、この用語は、ある細胞が２２対の常染色体と１対の性染色体を含まない場合を指すことがある。ヒト配偶子の場合、この用語は、ある細胞が２３本の染色体のいずれか１つを含まない場合を指すことがある。単一染色体型の場合、この用語は、３つ以上または１つ以下の相同ではあるが同一ではない染色体コピーが存在する場合、あるいは同じ親由来の染色体コピーが２つ存在する場合を指すことがある。いくつかの態様では、染色体分節の欠失は微少欠失である。

【００９１】

「倍数性状態」は、ある細胞の１つ以上の染色体または染色体分節の量および／または染色体同一性を指す。

【００９２】

「染色体」は、単一染色体コピーを指す場合がある。この用語は、正常な体細胞には４６個存在するＤＮＡの単一分子（例えば「母親由来の１８番染色体」等）を意味する。染色体はまた、正常なヒト体細胞には２３個存在する染色体型（例えば「１８番染色体」等）を指すことがある。

【００９３】

「染色体同一性」は、染色体参照番号、すなわち染色体型を指す場合がある。正常なヒトは、番号付けされた常染色体の２２種の染色体型と、２種類の性染色体を有する。この用語は、染色体の親の起源を指すこともある。この用語は、親から受け継いだ特定の染色体を指すこともある。この用語は、染色体のその他の同定に用いられる特徴を指すこともある。

【００９４】
「アレルデータ」は、１つ以上のアレルの集合に関する遺伝型データの集合を指す。この用語は、相決定済みハプロタイプデータを指す場合がある。この用語は、ＳＮＰ同一性を指す場合がある。また、挿入、欠失、反復、および変異を含むＤＮＡの配列データを指す場合がある。この用語は、各アレルの親起源を含むことがある。
【００９５】
「アレル状態」は、１つ以上のアレルの集合における遺伝子の実際の状態を指す。この用語は、アレルデータによって記載された遺伝子の実際の状態を指す場合がある。
【００９６】
「アレル数」は、特定の座位にマッピングされる配列の数を指し、その座位が多型である場合は各アレルにマッピングされる配列の数を指す。各アレルが二進数で計測される場合、アレルの数は整数になる。アレルが確率で計測される場合、アレルの数は分数であってもよい。
【００９７】
「アレル数確率」は、特定の座位または多型座位にあるアレルの集合にマッピングされると思われる配列の数を、マッピングの確率と合わせたものを指す。各計測された配列のマッピング確率が二値（０または１）である場合は、アレル数は、アレル数確率に相当する。いくつかの態様では、アレル数確率は二値である場合がある。いくつかの態様では、アレル数確率はＤＮＡ測定と等しくなるように設定されていることがある。
【００９８】
「アレル分布」または「アレル数分布」は、座位の集合における各座位に存在する各アレルの相対量を指す。アレル分布は、個体、試料、または試料に対して行った測定の集合を指し得る。配列決定等のデジタルアレル測定の文脈では、このアレル分布は、多型座位の集合における各アレルについて特定のアレルにマッピングされるリードの数または確数を指す。ＳＮＰアレイ等、アナログアレル測定の文脈では、このアレル分布は、アレル強度および／またはアレル比率を指す。アレル測定は、確率として、すなわち、あるアレルがある配列リードに存在する尤度が０と１の間の分数として処理されてもよく、または二進法、すなわち、ある任意のリードがある特定のアレルのコピーとして０または１のいずれかと見なされるような方法で処理されてもよい。
【００９９】
「アレル分布パターン」は、文脈（親の文脈等）ごとに異なるアレル分布の集合を指す。あるアレル分布パターンは、ある倍数性状態を指す場合がある。
【０１００】
「アレルの偏り」は、ヘテロ接合座位に存在するアレルの測定比率がＤＮＡまたはＲＮＡの元の試料の比率と異なっている程度を指す。特定の座位に存在するアレルの偏りの程度は、測定されたその座位に存在する観察されたアレルの比率を元のＤＮＡまたはＲＮＡ試料のその座位に存在するアレルの比率で除したものに等しい。アレルの偏りは、増幅による偏り、精製による偏り、または異なるアレルに様々な影響を及ぼすいくつかのその他の現象による場合がある。
【０１０１】
「アレル不均衡率」は、ＳＮＶの場合、異常なＤＮＡの割合を指す。この割合は、典型的には多様性アレル頻度（ある座位に存在する多様性アレル数／その座位に存在する総アレル数）を用いて計測される。腫瘍における２つの相同体の量の差は類似しているので、ＣＮＶについて異常なＤＮＡの割合を平均アレル不均衡率（ＡＡＩ）を用いて測定する。平均アレル不均衡率は｜（Ｈ１－Ｈ２）｜／（Ｈ１＋Ｈ２）として定義され、ここでＨｉは試料中の相同体ｉの平均コピー数であり、Ｈｉ／（Ｈ１＋Ｈ２）は相同体ｉの分数の量、すなわち相同体比率である。最大相同体比率は、もっとも量が多い相同体の相同体比率である。
【０１０２】
「試験脱落率」は、すべてのＳＮＰを用いて推定した、リードのないＳＮＰの百分率で

ある。
【０１０３】
「単一アレル脱落（ＡＤＯ）率」は、ヘテロ接合ＳＮＰのみを用いて推定した、アレルが１つのみ存在するＳＮＰの百分率である。
【０１０４】
「プライマー」および「ＰＣＲプローブ」は、単一核酸分子（ＤＮＡ分子、ＤＮＡオリゴマー等）、あるいは同一またはほぼ同一の核酸分子（ＤＮＡ分子、ＤＮＡオリゴマー等）の集合体を指す。プライマーは、対象とする座位（例えば、対象とする多型座位または非多型座位）または汎用プライミング配列にハイブリダイズされるように設計された領域を含む。プライマーは、ＰＣＲ増幅を可能にするように設計されたプライミング配列を含んでいてもよい。プライマーはまた、分子バーコードを含んでいてもよい。プライマーは、個体分子ごとに異なるランダム領域を含んでいてもよい。
【０１０５】
「プライマーのライブラリー」は、２個以上のプライマーの集団を指す。様々な態様において、このライブラリーは、少なくとも１００個、２００個、５００個、７５０個、１，０００個、２，０００個、５，０００個、７，５００個、１０，０００個、２０，０００個、２５，０００個、３０，０００個、４０，０００個、５０，０００個、７５，０００個、または１００，０００個の異なるプライマーを含む。様々な態様において、このライブラリーは、少なくとも１００個、２００個、５００個、７５０個、１，０００個、２，０００個、５，０００個、７，５００個、１０，０００個、２０，０００個、２５，０００個、３０，０００個、４０，０００個、５０，０００個、７５，０００個、または１００，０００個の異なるプライマー対を含み、各プライマー対は、フォワード試験プライマーおよびリバース試験プライマーを含み、各試験プライマー対は対象座位にハイブリダイズされる。いくつかの態様では、プライマーのライブラリーは、少なくとも１００個、２００個、５００個、７５０個、１，０００個、２，０００個、５，０００個、７，５００個、１０，０００個、２０，０００個、２５，０００個、３０，０００個、４０，０００個、５０，０００個、７５，０００個、または１００，０００個の異なる個別のプライマーを含む。個別のプライマーはそれぞれ異なる対象座位にハイブリダイズされ、これら個別のプライマーはプライマー対の一部ではない。いくつかの態様では、ライブラリーは、（ｉ）プライマー対および（ｉｉ）プライマー対の一部ではない個別のプライマー（汎用プライマー等）の両方を含む。
【０１０６】
「異なるプライマー」とは、同一ではないプライマーを指す。
【０１０７】
「異なるプール」とは、同一ではないプールを指す。
【０１０８】
「異なる対象座位」とは、同一ではない対象座位を指す。
【０１０９】
「異なる増幅産物」とは、同一ではない増幅産物を指す。
【０１１０】
「ハイブリッドキャプチャー用プローブ」は、試料中の特定の目的ＤＮＡ配列の一本鎖と相補であることを意図してＰＣＲ、直接合成等の様々な方法で作成された、修飾可能な任意の核酸配列を指す。用意した試料に外来性のハイブリッドキャプチャー用プローブを添加して、変性再アニール法でハイブリダイズさせて、外来性－内在性断片の二本鎖を形成してもよい。これら二本鎖を様々な手段で試料から物理的に分離してもよい。
【０１１１】
「配列リード」は、例えば、クローン配列決定法を用いて測定したヌクレオチドの塩基配列を表すデータを指す。クローン配列決定法により、元の１つのＤＮＡ分子、そのクローン、またはそのクラスターを表す配列データが作成されてもよい。配列リードはまた、ヌクレオチドが正しく呼び出された確率を示す、関連づけられた質スコアを配列の各塩基

位置に有していてもよい。

【０１１２】

「配列リードのマッピング」とは、特定の器官のゲノム配列における配列リードの元の位置を決定する方法である。配列リードの元の位置は、そのリードのヌクレオチド配列およびゲノム配列の類似性に基づいている。

【０１１３】

「一致型コピー誤り」および「一致型染色体異数性（ＭＣＡ）」は、１個の細胞が２つの同一、またはほぼ同一の染色体を含む異数性状態を指す。この種の異数性は、減数分裂時に配偶子が形成されるときに生じる場合があり、減数分裂不分離誤りとも言う場合がある。この種の誤りは有糸分裂時に生じる場合がある。一致型三染色体性は、ある染色体のコピーが３つ、１個体に存在し、そのうちの２つが同一である場合を指すことがある。

【０１１４】

「部分一致型コピー誤り」および「非同一染色体異数性（ＵＣＡ）」は、１個の細胞が、相同ではあるが同一ではない可能性がある、同じ親由来の２つの染色体を含む異数性状態を指す。この種の異数性は、減数分裂時に生じることがあり、減数分裂誤りと言う場合がある。部分一致型三染色体性は、ある染色体のコピーが３つ、１個体に存在し、そのうちの２つが同じ親由来であり、相同であるが同一ではない場合を指すことがある。部分一致型三染色体性は、一方の親由来の２つの相同染色体が存在し、その染色体分節のある部分は同一であるが他の分節は単に相同である場合を指すことがある。

【０１１５】

「相同染色体」は、減数分裂に正常に対をなす遺伝子の同じ集合を含む染色体コピーを指す。

【０１１６】

「同一染色体」は、同じ遺伝子の集合を含み、各遺伝子が同一、またはほぼ同一のアレルの同じ集合を有する染色体コピーを指す。

【０１１７】

「アレル脱落（ＡＤＯ）」は、あるアレルの相同染色体の塩基対の集合のうち、少なくとも塩基対の一方が検出されない状況を指す。

【０１１８】

「座位脱落（ＬＤＯ）」は、あるアレルの相同染色体の塩基対の集合のうち、塩基対の両方が検出されない状況を指す。

【０１１９】

「ホモ接合」は、対応する染色体座位として類似のアレルを有することを指す。

【０１２０】

「ヘテロ接合」は、対応する染色体座位として類似しないアレルを有することを指す。

【０１２１】

「ヘテロ接合性率」は、集団におけるある座位にヘテロ接合アレルを有する個体の率を指す。ヘテロ接合性率はまた、個体あるいはＤＮＡまたはＲＮＡ試料におけるある座位に存在するアレルの期待比率または測定比率を指すことがある。

【０１２２】

「染色体領域」は染色体の分節、または完全な染色体を指す。

【０１２３】

「染色体の分節」は、染色体の部分を指し、１塩基対から完全な染色体までの大きさである可能性がある。

【０１２４】

「染色体」は、完全な染色体、または染色体の分節すなわち部分を指す。

【０１２５】

「コピー」は、染色体分節のコピー数を指す。この用語は、染色体分節の同一コピー、同一ではないが相同であるコピーを指す場合がある。同一ではないが相同であるコピーでは、染色体分節の各コピーが実質的に類似した座位集合を含み、１つ以上のアレルが異な

っている。Ｍ２コピー誤り等、異数性のいくつかの場合では、ある染色体分節のコピーのうち、一部が同一コピーであり、一部が同一ではないコピーである可能性がある。

【０１２６】

「ハプロタイプ」は、典型的には同じ染色体に受け継がれる複数の座位に存在するアレルの組み合わせを指す。ハプロタイプは、座位の所定の集合間で起こった組換え事象の数に応じて少なくとも２つの座位から完全な染色体までを指すことがある。ハプロタイプはまた、統計的に関連づけられる単一の染色分体上にあるＳＮＰの集合を指し得る。

【０１２７】

「ハプロタイプデータ」、および「相決定済みデータ」、「順序付け遺伝子データ」は、二倍体または倍数体における単一の染色体または染色体分節に由来するデータを指し、例えば、二倍体ゲノムにおける分離した母親または父親の染色体コピー等である。

【０１２８】

「相決定」は、ある順序づけされていない二倍体（または倍数性）遺伝子データにおいて個体のハプロタイプ遺伝子データを決定する動作を指す。この用語は、１つの染色体に見られるアレルの集合に関して、あるアレルの２つの遺伝子のうちのどちらと、個体の２つの相同染色体のいずれと関連づけられるかを決定する動作を指す場合がある。

【０１２９】

「相決定済みデータ」は、１つ以上のハプロタイプが決定された遺伝子データを指す。

【０１３０】

「仮説」は、起こりうる状態を指す（例えば、第一の相同染色体または染色体分節が第二の相同染色体または染色体分節と比較してコピー数過剰出現が起こりうる度合い、起こりうる欠失、起こりうる重複、１つ以上の染色体または染色体分節の集合で起こりうる倍数性状態、１箇所以上の座位の集合で起こりうるアレルの状態、起こりうる父系性関係、１つ以上の染色体または染色体分節の集合で起こりうるＤＮＡ、ＲＮＡ、胎児の分画、座位の集合に由来する遺伝子材料の量の集合等）。遺伝子状態が、仮説の各要素が仮説の他の要素との関係で真であるという相対的尤度、あるいは仮説が全体として真であるという相対的尤度を示す確率と結びつけられてもよい。この可能性の集合は１つ以上の要素を含んでいてもよい。

【０１３１】

「コピー数仮説」および「倍数性状態仮説」は、個体の染色体または染色体分節のコピー数に関する仮説を指す。この用語はまた、各染色体の同一性（各染色体の元の親、およびその親の２つの染色体のうち、どちらがその個体に存在しているかを含む）に関する仮説を指す場合がある。この用語はまた、関係がある個体が存在する場合、その個体由来の染色体または染色体分節が、ある個体のある染色体と遺伝的に対応するということについての仮説を指す場合がある。

【０１３２】

「関係がある個体」は、対象の個体と遺伝的に関係があり、かつ対象の個体とハプロタイプブロックを共有する任意の個体を指す。ある文脈では、関係がある個体は、対象の個体の遺伝学上の親、または親に由来した任意の遺伝子材料（精子、極体、胚、胎児、または子供）であってもよい。この用語はまた、同胞、親、または祖父母を指す場合がある。

【０１３３】

「同胞」は、遺伝学上の親が対象の個体と同じである任意の個体を指す。いくつかの態様では、この用語は、生まれた子供、胚、胎児、あるいは生まれた子供、胚、または胎児に由来する１個以上の細胞を指す場合がある。同胞はまた、どちらかの親に由来する半数体個体（例えば、精子、極体、またはハプロタイプ遺伝物質のその他の集合）を指す場合がある。ある個体が、それ自身の同胞と見なされる場合がある。

【０１３４】

「子供」は、胚、卵割球、または胎児を指す場合がある。本明細書で開示された態様では、記述した概念は、生まれた子供、胎児、胚、またはそれらに由来する細胞の集合である個体にも等しく適用される。用語「子供」を用いた場合、単に、子供として言及された

個体はその親からの遺伝的子孫であることを含意することを意味することがある。

【０１３５】

「胎児の」は、「その胎児の」または「その胎児に遺伝的に類似した胎盤領域」を指す。妊婦の胎盤のある部分は、胎児に遺伝的に類似しており、母親の血液中に見られる自由浮遊胎児ＤＮＡは、その胎児と一致する遺伝型を有する胎盤の一部に由来する場合がある。胎児の染色体の半分の遺伝子情報は、胎児の母親から受け継がれる。いくつかの態様では、胎児の細胞から得た、母親から受け継がれた染色体に由来するＤＮＡは、「母親由来」ではなく「胎児由来」と見なされる。

【０１３６】

「胎児由来のＤＮＡ」は、本来、その遺伝型が胎児のものと本質的に同じであった細胞の一部であったＤＮＡを指す。

【０１３７】

「母親由来のＤＮＡ」は、本来、その遺伝型が母親のものと本質的に同じであった細胞の一部であったＤＮＡを指す。

【０１３８】

「親」は、個体の遺伝学上の母親または父親を指す。個体は、典型的には二親、母親と父親を有するが、遺伝子キメラ現象または染色体キメラ現象の場合は必ずしも当てはまらないことがある。親は個体と見なされる場合がある。

【０１３９】

「親の文脈」は、対象の二親のうち、一方または両方の関連する２つの染色体のそれぞれについて、あるＳＮＰの遺伝子状態を指す。

【０１４０】

「母親の血漿」は、妊婦由来の血液の血漿部分を指す。

【０１４１】

「臨床判断」は、個体の健康または生存に影響を及ぼす結果を有する行動を取るか否かという任意の判断を指す。臨床判断はまた、さらなる試験を行う、中絶すること、妊娠状態を維持する、望ましくない表現型を緩和するための行動を取る、またはある表現型に備えるための行動を取ることを指す場合がある。

【０１４２】

「診断ボックス」は、本明細書で開示される方法の一側面あるいは複数の側面を行うように設計された１つの機械、またはそのような機械の組み合わせを指す。一態様では、診断ボックスは患者治療の点で設けられている場合がある。一態様では、診断ボックスは、目的の増幅と、その後の配列決定を行う場合がある。一態様では、診断ボックスは独立して、あるいは技術者の助けにより機能する場合がある。

【０１４３】

「インフォマティックスに基づく方法」は、大量のデータを解明するために統計学に大きく依存する方法を指す。出生前診断の文脈では、この用語は、（例えば、分子アレイや配列決定からの）大量の遺伝子データがある場合に、直接物理的にその状態を測定するのはなく、統計的に最も可能性のある状態を推測することにより、１つ以上の染色体または染色体分節の倍数性状態、１つ以上のアレルのアレル状態、または父系性を決定するように設計された方法を指す。本開示の一態様では、インフォマティックスに基づく技術は、本特許出願で開示されるものであることがある。本開示の一態様では、インフォマティックスに基づく技術は、PARENTAL SUPPORT（商標：Gen Security Network, Inc.で開発された染色体異数性のスクリーニング技術）である場合がある。

【０１４４】

「一次遺伝子データ」は、遺伝型決定プラットフォームによって出力されたアナログ強度信号を指す。ＳＮＰアレイの文脈では、一次遺伝子データは、任意の遺伝型呼び出しが行われる前の強度信号を指す。配列決定の文脈では、一次遺伝子データは、任意の塩基対の同一性が決定される前、および配列がゲノムにマッピングされる前に配列決定装置から出力される、クロマトグラムに類似したアナログ測定を指す。

【０１４５】
「二次遺伝子データ」は、遺伝型決定プラットフォームによって出力された処理済み遺伝子データを指す。ＳＮＰアレイの文脈では、二次遺伝子データは、ＳＮＰアレイ読取り器に関連づけられたソフトウェアによってなされるアレル呼び出しを指し、このソフトウェアは、あるアレルが試料中に存在するか否かを呼び出す。配列決定の文脈では、二次遺伝子データは、決定された配列の塩基対の同一性を指す。また、配列がゲノムにマッピングされた場合の塩基対の同一性を指し得る。
【０１４６】
「ある座位に対応するＤＮＡの優先富化またはある座位に存在するＤＮＡの優先富化」は、富化後のＤＮＡ混合物中のその座位に対応するＤＮＡ分子の百分率を、富化前のＤＮＡ混合物中のその座位に対応するＤＮＡ分子の百分率よりも高くする任意の方法を指す。この方法は、ある座位に対応するＤＮＡ分子の選択的増幅を含む場合がある。この方法は、その座位に対応しないＤＮＡ分子を除去することを含む場合がある。この方法は、複数の方法の組み合わせを含む場合がある。富化度は、富化後の混合物中のその座位に対応するＤＮＡ分子の百分率を、富化前の混合物中のその座位に対応するＤＮＡ分子の百分率で除したものとして定義される。優先富化が複数の座位について行われる場合がある。本開示のいくつかの態様では、富化度は、２０、２００、または２，０００より大きい。優先富化が複数の座位について行われる場合、富化度は、座位の集合におけるすべての座位の平均富化度を指すことがある。
【０１４７】
「増幅」は、ＤＮＡ分子またはＲＮＡ分子のコピー数を増加させる方法を指す。
【０１４８】
「選択的増幅」は、特定のＤＮＡ（またはＲＮＡ）分子のコピー数、またはＤＮＡ（またはＲＮＡ）の特定の領域に対応するＤＮＡ（またはＲＮＡ）分子のコピー数を増加させる方法を指す場合がある。この用語はまた、非対象ＤＮＡ（またはＲＮＡ）分子または領域の増加以上に特定の対象ＤＮＡ（またはＲＮＡ）分子または対象ＤＮＡ（またはＲＮＡ）領域のコピー数を増加させる方法を指す場合がある。選択的増幅は優先富化の方法であってもよい。
【０１４９】
「汎用プライミング配列」は、例えば、ライゲーション、ＰＣＲ、またはライゲーション媒介ＰＣＲによって対象ＤＮＡ（またはＲＮＡ）分子の集団に付加されてもよいＤＮＡ（またはＲＮＡ）配列を指す。一旦、汎用プライミング配列が対象分子の集団に付加されると、この汎用プライミング配列に特異的なプライマーを用いて、この対象集団を１対の増幅プライマーにより増幅させることができる。汎用プライミング配列は、通常、対象配列と関係がない。
【０１５０】
「汎用接着体」、「ライゲーション接着体」、または「ライブラリー標識」は、対象の二本鎖核酸分子の集団の５’末端および３’末端に共有結合することができる汎用プライミング配列を含む核酸分子である。接着体の付加により、ＰＣＲ増幅を開始させることができる対象集団の５’末端および３’末端に汎用プライミング配列が供給され、１対の増幅プライマーを用いてこの対象集団からすべての分子を増幅させる。
【０１５１】
「対象化」は、ＤＮＡ（またはＲＮＡ）混合物中の座位の集合に対応するＤＮＡ（またはＲＮＡ）分子を選択的に増幅または優先的に富化するために用いられる方法を指す。
【０１５２】
「同時分布モデル」は、変数の確率が結合される同じ確率空間で定義された複数のランダム変数を仮定していくつかのランダム変数で定義される事象の確率を定義するモデルを指す。いくつかの態様では、変数の確率が結合されない縮退した場合が用いられることがある。
【０１５３】

「癌関連遺伝子」は、癌の危険性の変化または予後の変化に関連した遺伝子を指す。癌を促進する癌関連遺伝子の例としては、癌遺伝子、細胞増殖、浸潤、または転移を高める遺伝子、アポトーシスを阻害する遺伝子、および血管形成促進遺伝子等が挙げられる。癌を阻害する癌関連遺伝子の例としては、腫瘍抑制遺伝子、細胞増殖、浸潤、または転移を阻害する遺伝子、アポトーシスを促進する遺伝子、および血管形成阻害遺伝子等が挙げられるが、これらに限定されない。

【０１５４】

「エストロゲン関連癌」は、エストロゲンによって調節される癌を指す。エストロゲン関連癌の例としては、乳癌および卵巣癌が挙げられるが、これらに限定されない。Ｈｅｒ２は多くのエストロゲン関連癌で過剰発現する（米国特許第６，１６５，４６４号、参照によりその全体が本明細書に組み込まれる）。

【０１５５】

「アンドロゲン関連癌」は、アンドロゲンによって調節される癌を指す。アンドロゲン関連癌の例としては、前立腺癌がある。

【０１５６】

「正常な発現量より高い」とは、対照としての対象者（例えば、癌等の病気や疾患がない対象者）のｍＲＮＡまたはタンパク質の平均発現量よりも高いことを指す。様々な態様では、この正常な発現量より高い発現量は、対照としての対象者における発現量に対して少なくとも２０、４０、５０、７５、９０、１００、２００、５００，あるいは１０００％、多くなっている。

【０１５７】

「正常な発現量より低い」とは、対照としての対象者（例えば、癌等の病気や疾患がない対象者）のｍＲＮＡまたはタンパク質の平均発現量よりも低いことを指す。様々な態様では、この正常な発現量より高い発現量は、対照としての対象者における発現量に対して少なくとも２０、４０、５０、７５、９０、９５、または１００％、低くなっている。いくつかの態様では、ｍＲＮＡまたはタンパク質の発現が検出不能である。

【０１５８】

「発現または活性を調節する」とは、例えば、対照条件に対してタンパク質または核酸配列の発現または活性を増加あるいは低減させることを指す。いくつかの態様では、発現または活性における調節は、少なくとも１０、２０、４０、５０、７５、９０、１００、２００、５００、または１０００％の増加あるいは低減である。様々な態様では、転写、翻訳、ｍＲＮＡまたはタンパク質の安定性、あるいは生体内の他の分子へのｍＲＮＡまたはタンパク質の結合が治療によって調節される。いくつかの態様では、ｍＲＮＡの量が標準的なノーザンブロット解析によって求められ、タンパク質の量が標準的なウエスタンブロット解析（例えば、本明細書において記載する解析や、例えばAusubel et al.（Current Protocols in Molecular Biology（分子生物学における最新のプロトコル），John Wiley & Sons, New York, July 11, 2013, 参照によりその全体が本明細書に組み込まれる）で記述された解析）によって求められる。一態様では、タンパク質の量は、標準的な方法を用いて酵素活性の量を測定することにより求められる。他の好ましい態様では、ｍＲＮＡ、タンパク質、または酵素活性の量は、タンパク質の機能型を発現していない対照細胞（例えば、ノンセンス変異の場合ホモ接合である細胞等）のそれらの量より多いが、２０倍、１０倍、５倍、または２倍以下である。さらに他の態様において、ｍＲＮＡ、タンパク質、または酵素活性の量は、対照細胞（例えば、非癌細胞、異常細胞増殖を誘発する条件、またはアポトーシスを阻害する条件にさらされなかった細胞、目的の病気または疾患を有しない対象者由来の細胞等）の基礎量より多いが、２０倍、１０倍、５倍、または２倍以下である。

【０１５９】

「ｍＲＮＡまたはタンパク質の発現または活性を調節するのに十分な投与量」は、対象者に投与するとｍＲＮＡまたはタンパク質の発現または活性が増加あるいは低減する治療法の量を指す。いくつかの態様では、発現または活性を低減させる化合物の場合、調節は

、治療を受けた対象者において、阻害薬を投与される前の同じ対象者と比べて、あるいは治療を受けていない対照としての対象者と比べて少なくとも１０％、３０％、４０％、５０％、７５％、または９０％、発現または活性を低減させることである。また、いくつかの態様では、発現または活性を増加させる化合物の場合、治療を受けた対象者のｍＲＮＡまたはタンパク質の発現量または活性量は、調節薬を投与する前の同じ対象者あるいは治療を受けていない対照としての対象者の発現量または活性量に比べて１．５倍、２倍、３倍、５倍、１０倍、または２０倍多くなっている。

【０１６０】

いくつかの態様では、化合物はＲＮＡまたはタンパク質の発現または活性を直接または間接に調節することがある。例えば、ある化合物は、目的ｍＲＮＡまたはタンパク質の発現または活性に直接または間接に影響を及ぼす分子（例えば、核酸、タンパク質、シグナル伝達分子、成長因子、サイトカイン、またはケモカイン）の発現または活性を調節することで目的のｍＲＮＡまたはタンパク質の発現または活性を間接的に調節することがある。いくつかの態様では、前記化合物は、細胞分裂を阻害する、あるいはアポトーシスを誘発する。前記治療法におけるこれらの化合物の例としては、非精製タンパク質または精製タンパク質、抗体、合成有機分子、天然有機分子、核酸分子、およびこれらの成分等が挙げられる。併用療法においてこれらの化合物は同時または順次に投与されてもよい。これらの化合物としては、例えば、シグナル伝達抑制分子が挙げられる。

【０１６１】

「精製」は、天然に伴っている他の成分から分離することを指す。通常、ある因子の少なくとも５０重量％が天然の状態で関連づけられたタンパク質、抗体、および天然有機分子を含まなければ、その因子は実質的に純粋である。いくつかの態様では、前記因子は少なくとも７５重量％、９０重量％、または９９重量％純粋である。実質的に純粋な因子は、化学合成、天然源からのその因子の分離、または天然にはその因子を生成しない組み換え宿主細胞におけるその因子の生成によって得られることがある。タンパク質および小分子は、例えば、Ausubel et al.（Current Protocols in Molecular Biology, John Wiley & Sons, New York, July 11, 2013, 参照によりその全体が本明細書に組み込まれる）に記載された標準的な技術を用いて当業者によって精製されることがある。いくつかの態様では、ポリアクリルアミドゲル電気泳動法、カラムクロマトグラフィー、光学濃度、ＨＰＬＣ分析、またはウエスタン分析（Ausubel et al., supra）を用いた測定において前記因子は開始材料の少なくとも２倍、５倍、または１０倍純粋である。精製の方法としては、例えば、免疫沈殿、カラムクロマトグラフィー（例えば、免疫親和性クロマトグラフィー）、磁性ビーズ免疫親和性精製、およびプレート結合抗体によるパニング等が挙げられる。

【０１６２】

本発明の他の特徴および利点は、以下の詳細な説明および請求項から明らかである。

【０１６３】

本特許すなわち出願書類は、少なくとも１つのカラー図面を含む。カラー図面付きの本特許あるいは特許出願公報の複製は、請求と必要手数料の支払いががあれば特許庁より提供される。

【０１６４】

本明細書において開示される態様は、添付の図面を参照してさらに説明される。いくつかの図面で、同様の構造には同様の参照符号が付けられている。ここに示めす図面は必ずしも縮尺通りではなく、むしろ本明細書において開示される態様の原則の説明に強調が置かれている。

【図面の簡単な説明】

【０１６５】

【図１Ａ－１Ｄ】ＳＮＰ数の増加に対し、リード深度（ＤＯＲ）が５００、腫瘍分画が１％の検定統計量ＳをＴ（ＳＮＰ数）で除した（「Ｓ／Ｔ」）異なるコピー数仮説に関する分布を示すグラフである。

【０１６６】
【図２Ａ－２Ｄ】ＳＮＰ数の増加に対し、ＤＯＲが５００、腫瘍分画が２％の場合の、異なるコピー数仮説に関するＳ／Ｔ分布を示すグラフである。
【０１６７】
【図３Ａ－３Ｄ】ＳＮＰ数の増加に対し、ＤＯＲが５００、腫瘍分画が３％の場合の、異なるコピー数仮説に関するＳ／Ｔ分布を示すグラフである。
【０１６８】
【図４Ａ－４Ｄ】ＳＮＰ数の増加に対し、ＤＯＲが５００、腫瘍分画が４％の場合の、異なるコピー数仮説に関するＳ／Ｔ分布を示すグラフである。
【０１６９】
【図５Ａ－５Ｄ】ＳＮＰ数の増加に対し、ＤＯＲが５００、腫瘍分画が５％の場合の、異なるコピー数仮説に関するＳ／Ｔ分布を示すグラフである。
【０１７０】
【図６Ａ－６Ｄ】ＳＮＰ数の増加に対し、ＤＯＲが５００、腫瘍分画が６％の場合の、異なるコピー数仮説に関するＳ／Ｔ分布を示すグラフである。
【０１７１】
【図７Ａ－７Ｄ】ＳＮＰ数の増加に対し、ＤＯＲが１０００、腫瘍分画が０．５％の場合の、異なるコピー数仮説に関するＳ／Ｔ分布を示すグラフである。
【０１７２】
【図８Ａ－８Ｄ】ＳＮＰ数の増加に対し、ＤＯＲが１０００、腫瘍分画が１％の場合の、異なるコピー数仮説に関するＳ／Ｔ分布を示すグラフである。
【０１７３】
【図９Ａ－９Ｄ】ＳＮＰ数の増加に対し、ＤＯＲが１０００、腫瘍分画が２％の場合の、異なるコピー数仮説に関するＳ／Ｔ分布を示すグラフである。
【０１７４】
【図１０Ａ－１０Ｄ】ＳＮＰ数の増加に対し、ＤＯＲが１０００、腫瘍分画が３％の場合の、異なるコピー数仮説に関するＳ／Ｔ分布を示すグラフである。
【０１７５】
【図１１Ａ－１１Ｄ】ＳＮＰ数の増加に対し、ＤＯＲが１０００、腫瘍分画が４％の場合の、異なるコピー数仮説に関するＳ／Ｔ分布を示すグラフである。
【０１７６】
【図１２Ａ－１２Ｄ】ＳＮＰ数の増加に対し、ＤＯＲが３０００、腫瘍分画が０．５％の場合の、異なるコピー数仮説に関するＳ／Ｔ分布を示すグラフである。
【０１７７】
【図１３Ａ－１３Ｄ】ＳＮＰ数の増加に対し、ＤＯＲが３０００、腫瘍分画が１％の場合の、異なるコピー数仮説に関するＳ／Ｔ分布を示すグラフである。
【０１７８】
【図１４】６種の微少欠失症候群を検出するための感度および特異性を示す表である。
【０１７９】
【図１５Ａ－１５Ｃ】正倍数性をグラフ表示したものである。ｘ軸は染色体に沿った個体の多型座位の直線位置を表し、ｙ軸はアレルの総リード数（Ａ＋Ｂ）の分画としてアレルＡのリード数を表す。母親の遺伝型および胎児の遺伝型は、プロットの右側に示されている。各プロットは母親の遺伝型に応じて色づけされている。例えば、赤色は母親の遺伝型ＡＡ、青色は母親の遺伝型ＢＢ、緑色は母親の遺伝型ＡＢを表す。図１５Ａは、２本の染色体が存在し、胎児の無細胞ＤＮＡ分画が０％であるときのプロットである。このプロットは、妊娠していない女性由来のものであり、よってその遺伝型はすべて母親のものであるパターンを表す。したがって、アレル群は１（ＡＡのアレル）、０．５（ＡＢのアレル）、および０（ＢＢのアレル）に集中している。図１５Ｂは、２本の染色体が存在し胎児の分画が１２％であるときのプロットである。アレルＡのリードの分画に含まれる胎児のアレルによって、アレルの点の位置がｙ軸に沿って上または下にシフトしている。図１５

Ｃは２本の染色体が存在し胎児の分画が２６％である時のプロットである。このパターンには、２本の赤い周辺バンドと２本の青い周辺バンドを含み、３本の緑色の中のバンドが容易に見える。

【０１８０】

【図１６Ａ－１６Ｂ】２２ｑ１１．２欠失症候群をグラフ表示したものである。図１６Ａは、母親の２２ｑ１１．２欠失の保有者についての図である（緑色のＡＢのＳＮＰがないことによって示されている）。図１６Ｂは、親から受け継がれた胎児の２２ｑ１１欠失についての図である（１本の赤い周辺バンドと１本の青い周辺バンドで示されている）。ｘ軸はＳＮＰの直線位置を表し、ｙ軸は総リード数におけるアレルＡのリードの分画を表す。各点は、単一のＳＮＰ座位を表す。

【０１８１】

【図１７】母親から受け継がれた猫鳴き欠失症候群のグラフ表示である（３本の緑色のバンドの代わりに２本の中央の緑色のバンドの存在によって示されている）。ｘ軸はＳＮＰの直線位置を表し、ｙ軸は総リード数におけるアレルＡのリードの分画を表す。各点は、単一のＳＮＰ座位を表す。

【０１８２】

【図１８】ウォルフ・ヒルシュホーン欠失症候群のグラフ表示である（１本の赤い周辺バンドと１本の青い周辺バンドによって示されている）。ｘ軸はＳＮＰの直線位置を表し、ｙ軸は総リード数におけるアレルＡのリードの分画を表す。各点は、単一のＳＮＰ座位を表す。

【０１８３】

【図１９Ａ】染色体または染色体分節の余剰コピーを表す、Ｘ染色体スパイク・イン（spike-in）実験のグラフ表示である。これらのプロットは、異なる量の父親由来のＤＮＡを娘由来のＤＮＡと混合したものを示しており、それぞれ、父親由来のＤＮＡが１６％（図１９Ａ）、１０％（図１９Ｂ）、１％（図１９Ｃ）、および０．１％（図１９Ｄ）である。ｘ軸はＸ染色体にあるＳＮＰの直線位置を表し、ｙ軸は総リード数（Ｍ＋Ｒ）におけるアレルＭのリードの分画を表す。各点は、アレルＭまたはＲを有する単一のＳＮＰ座位を示す。

【図１９Ｂ】同上。

【図１９Ｃ】同上。

【図１９Ｄ】同上。

【０１８４】

【図２０Ａ－２０Ｂ】ハプロタイプデータを用いた偽陰性率のグラフ（図２０Ａ）、およびハプロタイプデータを用いない偽陰性率のグラフ（図２０Ｂ）である。

【０１８５】

【図２１Ａ－２１Ｂ】ハプロタイプデータを用いたｐ＝１％での偽陽性率のグラフ（図２１Ａ）、およびハプロタイプデータを用いないｐ＝１％での偽陽性率のグラフ（図２１Ｂ）である。

【０１８６】

【図２２Ａ－２２Ｂ】ハプロタイプデータを用いたｐ＝１．５％での偽陽性率のグラフ（図２２Ａ）、およびハプロタイプデータを用いないｐ＝１．５％での偽陽性率のグラフ（図２２Ｂ）である。

【０１８７】

【図２３Ａ－２３Ｂ】ハプロタイプデータを用いたｐ＝２％での偽陽性率のグラフ（図２３Ａ）、およびハプロタイプデータを用いないｐ＝２％での偽陽性率のグラフ（図２３Ｂ）である。

【０１８８】

【図２４Ａ－２４Ｂ】ハプロタイプデータを用いたｐ＝２．５％での偽陽性率のグラフ（図２４Ａ）、およびハプロタイプデータを用いないｐ＝２．５％での偽陽性率のグラフ（図２４Ｂ）である。

【０１８９】
【図２５Ａ－２５Ｂ】ハプロタイプデータを用いたｐ＝３％での偽陽性率のグラフ（図２５Ａ）、およびハプロタイプデータを用いないｐ＝３％での偽陽性率のグラフ（図２５Ｂ）である。
【０１９０】
【図２６】第一のシミュレーション用の偽陽性率表である。
【０１９１】
【図２７】第一のシミュレーション用の偽陰性率表である。
【０１９２】
【図２８Ａ－２８Ｃ】図２８Ａは、正常な（非癌性）細胞株について、その座位の総数で除した基準数（１つのアレル（例えば、アレル「Ａ」）の数）のグラフである。図２８Ｂは、欠失を有する癌細胞株について、総数で除した基準数のグラフである。図２８Ｃは、上記正常細胞株と癌細胞株に由来するＤＮＡの混合物について、基準計数を総計数で除したグラフである。
【０１９３】
【図２９】腫瘍分画が４．３３％と推定された（ＤＮＡの４．３３％が腫瘍細胞である）段階ＩＩａの乳癌の患者に由来する血漿試料について、基準計数を総計数で除したグラフである。このグラフの緑色の部分は、ＣＮＶが存在しない領域を表す。このグラフの青色部分および赤色部分は、ＣＮＶが存在し、かつ期待アレル比率０．５から測定アレル比率が明らかに分離している領域を表す。青色はあるハプロタイプを示し、赤色は他のハプロタイプを示す。ＣＮＶの領域では、約６３６個のヘテロ接合ＳＮＰが分析された。
【０１９４】
【図３０】腫瘍分画が０．５８％と推定された段階ＩＩｂの乳癌の患者に由来する血漿試料について、基準計数を総計数で除したグラフである。このグラフの緑色の部分は、ＣＮＶが存在しない領域を表す。このグラフの青色部分および赤色部分は、ＣＮＶが存在するが、測定アレル比率が期待アレル比率０．５からはっきりとは分離していない領域を示す。この分析では、ＣＮＶの領域で８６個のヘテロ接合ＳＮＰが分析された。
【０１９５】
【図３１Ａ－３１Ｂ】腫瘍分画の最尤推定を示すグラフである。最尤推定値はグラフのピークで示され、図３１Ａでは４．３３％、図３１Ｂでは０．５８％である。
【０１９６】
【図３２Ａ－３２Ｂ】図３２Ａは、高腫瘍分画試料（４．３３％）および低腫瘍分画試料（０．５８％）について、様々な起こりうる腫瘍分画の対数オッズ比を比較したグラフである。対数オッズ比が０未満の場合、正倍数体仮説が最も可能性が高い。対数オッズ比が０超の場合、ＣＮＶの存在が最も可能性が高い。図３２Ｂは、低腫瘍分画試料（０．５８％）について、様々な起こりうる腫瘍分画で欠失の確率を欠失がない確率で除したもののグラフである。
【０１９７】
【図３３】低腫瘍分画試料（０．５８％）について、様々な起こりうる腫瘍分画の対数オッズ比を示すグラフである。図３３は、図３２Ａの低腫瘍分画試料のグラフを拡大したものである。
【０１９８】
【図３４】実施例６に記載する３つの異なる方法を用いた腫瘍生検材料中の一塩基多様体の検出限界を示すグラフである。
【０１９９】
【図３５】実施例６に記載する３つの異なる方法を用いた血漿試料中の一塩基多様体の検出限界を示すグラフである。
【０２００】
【図３６Ａ－３６Ｂ】ＣＮＶを検出するように設計された約２８，０００本のプライマーのライブラリーを用いてゲノムＤＮＡ（図３６Ａ）または単一の細胞に由来するＤＮＡ（

図３６Ｂ）を分析したグラフである。中央に存在する１本ではなく２本のバンドがＣＮＶの存在を示している。ｘ軸はＳＮＰの直線位置を表し、ｙ軸は総リード数におけるアレルＡのリードの分画を表す。

【０２０１】

【図３７Ａ－３７Ｂ】ＣＮＶを検出するように設計された約３，０００本のプライマーのライブラリーを用いてゲノムＤＮＡ（図３７Ａ）または単一の細胞に由来するＤＮＡ（図３７Ｂ）を分析したグラフである。中央に存在する１本ではなく２本のバンドがＣＮＶの存在を示している。ｘ軸はＳＮＰの直線位置を表し、ｙ軸は総リード数におけるアレルＡのリードの分画を表す。

【０２０２】

【図３８】約３，０００座位のＤＯＲにおける均一性を示すグラフである。

【０２０３】

【図３９】ゲノムＤＮＡおよび単一の細胞に由来するＤＮＡについて、誤り呼び出しの評価基準を比較した表である。

【０２０４】

【図４０】転移型変異および転換型変異の誤り率を示すグラフである。

【０２０５】

【図４１ａ－４１ｄ】ＰｌａｓｍＡｒｔで決定したＣｏＮＶＥＲＧｅの感度を示すグラフである。（ａ）は、２２ｑ１１．２欠失および一致型正常細胞株に由来するＤＮＡを含むＰｌａｓｍＡｒｔ試料のＣｏＮＶＥＲＧｅで算出された平均アレル不均衡率（ＡＡＩ）と実際の入力分画との相関関係である。（ｂ）は、０～９．０９％の腫瘍ＤＮＡ分画を含む、染色体２ｐおよび２ｑのＣＮＶおよび一致型正常ＨＣＣ２２１８ＢＬ細胞を含むＨＣＣ２２１８乳癌細胞に由来するＤＮＡを含むＰｌａｓｍＡｒｔ試料の算出されたＡＡＩと実際の腫瘍ＤＮＡ入力との相関関係である。（ｃ）は、０～５．６６％の腫瘍ＤＮＡ分画を含む、染色体１ｐおよび１ｑのＣＮＶおよび一致型正常ＨＣＣ１９５４ＢＬ細胞を含むＨＣＣ１９５４乳癌細胞に由来するＤＮＡを含むＰｌａｓｍＡｒｔ試料の算出されたＡＡＩと実際の腫瘍ＤＮＡ入力の相関関係である。（ｄ）は、（ｃ）で用いたＨＣＣ１９５４細胞のアレル頻度のプロットである。（ａ）、（ｂ）、および（ｃ）では、データ点および誤差範囲（エラーバー）は、それぞれ３～８個の反復の平均値および標準偏差（ＳＤ）を示す。

【０２０６】

【図４２】ＰｌａｓｍＡｒｔ基準に関する詳細の一例を示しており、下部には断片の寸法分布を示すグラフを含む。

【０２０７】

【図４３Ａ】左は、微少欠失および癌パネルの妥当性のためのＰｌａｓｍＡｒｔ合成ｃｔＤＮＡ標準試料から得た結果の希釈曲線である。右のパネルは、オッズ比のプロットとして、ＤＮＡ分画の腫瘍の最尤推定を示す。

【図４３Ｂ－Ｃ】図４３Ｂは、転換事象が検出されたプロットである。図４３Ｃは、転移事象が検出されたプロットである。

【０２０８】

【図４４】ｃｔＤＮＡの％が異なる様々な試料について、異なる染色体領域のＣＮＶを示すプロットである。

【０２０９】

【図４５】ｃｔＤＮＡ量（％）が異なる様々な卵巣癌試料について、異なる染色体領域のＣＮＶを示すプロットである。

【０２１０】

【図４６】ｃｔＤＮＡ中のＳＮＶ、またはＳＮＶおよび／またはＣＮＶの組み合わせを有する乳癌または肺癌の患者のパーセントを示す表である。

【０２１１】

【図４７】異なる段階の乳癌における、血漿試料中に含まれる腫瘍に特異的なＳＮＶおよ

び／またはＣＮＶ（％）を示すグラフと、その右側のデータ関連表である。
【０２１２】
【図４８】異なる下位段階の乳癌における、血漿試料中に含まれる腫瘍に特異的なＳＮＶおよび／またはＣＮＶ（％）を示すグラフと、その右側のデータ関連表である。
【０２１３】
【図４９】異なる段階の肺癌における、血漿試料中に含まれる腫瘍に特異的なＳＮＶおよび／またはＣＮＶ（％）を示すグラフと、その右側のデータ関連表である。
【０２１４】
【図５０】異なる下位段階の肺癌における、血漿試料中に含まれる腫瘍に特異的なＳＮＶおよび／またはＣＮＶ（％）を示すグラフと、その右側のデータ関連表である。
【０２１５】
【図５１Ａ】腫瘍のクローンおよび準クローンの不均一性について分析された一次肺腫瘍の組織学的所見／履歴を示す。
【図５１Ｂ】全ゲノムの配列決定およびＡｍｐｌｉＳＥＱによるアッセイによる生検された肺腫瘍の多様性アレル頻度（ＶＡＦ）の同一性を示す表である。
【０２１６】
【図５２】血漿に由来するｃｔＤＮＡを用いて、腫瘍の不均一性を克服するための一塩基付加（ＳＮＡ）変異のクローンおよび準クローンの両方の同定を示す。
【０２１７】
【図５３】ＡｍｐｌｉＳｅｑで見逃されたが、血漿由来のｃｔＤＮＡ中で同定されたＳＮＶ変異である一次腫瘍のＳＮＶを検出するための、ＡｍｐｌｉＳｅｑおよびｍｍＰＣＲ－ＮＧＳによるＶＡＦ呼び出しを比較した表である。
【０２１８】
【図５４】図５４Ａは、一次肺腫瘍におけるＶＡＦ（％）を示すプロットである。図５４Ｂは、ＡｍｐｌｉＳｅｑによるＶＡＦとＮａｔｅｒａによるＶＡＦを比較した直線回帰プロットである。
【０２１９】
【図５５】プライマー濃度が限定された場合のＳＮＶの８４プレックスＰＣＲプライマー反応のプール１／４のグラフである。
【０２２０】
【図５６】プライマー濃度が限定された場合のＳＮＶの８４プレックスＰＣＲプライマー反応のプール２／４のグラフである。
【０２２１】
【図５７】プライマー濃度が限定された場合のＳＮＶの８４プレックスＰＣＲプライマー反応のプール３／４グラフである。
【０２２２】
【図５８】プライマー濃度が限定された場合のＳＮＶの８４プレックスＰＣＲプライマー反応のプール４／４のグラフである。
【０２２３】
【図５９】１５ＰＣＲサイクルの８４プレックスＰＣＲ反応においてＳＮＶ転移および転換変異を検出した場合の検出限界（ＬＯＤ）とリード深度（ＤＯＲ）のプロットである。
【０２２４】
【図６０】２０ＰＣＲサイクルの８４プレックスＰＣＲ反応においてＳＮＶ転移および転換変異を検出した場合の検出限界（ＬＯＤ）とリード深度（ＤＯＲ）のプロットである。
【０２２５】
【図６１】２５ＰＣＲサイクルの８４プレックスＰＣＲ反応においてＳＮＶ転移および転換変異を検出した場合の検出限界（ＬＯＤ）とリード深度（ＤＯＲ）のプロットである。
【０２２６】
【図６２】腫瘍細胞のゲノムＤＮＡと単一細胞のゲノムＤＮＡの感度を比較したプロットである。上部は腫瘍細胞のゲノムＤＮＡを示す。下部は単一細胞のゲノムＤＮＡを示す。

【０２２７】
【図６３】図６３ａのＳＮＰを対象とした大多重ＰＣＲ（ｍｍＰＣＲ）アッセイにおける各種癌試料中のＣＮＶの分析の作業フローを示す。図６３ｂ～図６３ｆは、乳癌細胞株と一致型正常細胞株についてＣｏＮＶＥＲＧｅアッセイとマイクロアレイアッセイでの比較を示す。
【０２２８】
【図６４】生の冷凍（ＦＦ）乳癌試料およびホルマリン固定パラフィン埋め込み（ＦＦＰＥ）乳癌試料と一致型対照との比較を示す。図ａ～ｈは、乳癌細胞株と一致型軟膜ｇＤＮＡ対照試料に対するＣｏＮＶＥＲＧｅアッセイとマイクロアレイアッセイを比較したものである。
【０２２９】
【図６５】ＣｏＮＶＥＲＧｅアッセイを用いて染色体コピー数を反映させて、単一細胞中のＣＮＶを検出したアレル頻度のプロットである。図６５ａ～６５ｃは乳癌単一細胞の３つの反復の分析である。図６５ｄは、注目領域でのＣＮＶがないＢリンパ球細胞株の分析である。
【０２３０】
【図６６】ＣｏＮＶＥＲＧｅアッセイを用いて染色体コピー数を反映させて、実際の血漿試料中のＣＮＶを検出したアレル頻度のプロットである。図６６ａは、段階ＩＩの乳癌の血漿無細胞ＤＮＡ試料およびその一致型腫瘍生検材料ｇＤＮＡである。図６６ｂは、末期の卵巣癌の血漿無細胞ＤＮＡ試料およびその一致型腫瘍生検材料ｇＤＮＡである。図６６ｃは、５つの末期の卵巣癌血漿および一致型組織試料におけるＣＮＶ検出により求められた腫瘍の不均一性を示す表である。
【０２３１】
【図６７－１】乳癌の染色体位置および変異による変化を示す。
【図６７－２】同上。
【図６７－３】同上。
【図６７－４】同上。
【０２３２】
【図６８】３１６８ｍｍＰＣＲ反応に用いられたＳＮＰの多数アレル頻度（図６８Ａ）および少数アレル頻度（図６８Ｂ）を示す。
【０２３３】
【図６９】本発明の態様を行うのに有用なシステム設計Ｘ００の一例を示す。
【０２３４】
【図７０】本発明の態様を行うためのコンピュータシステムの一例を示す。　記載したこれらの図面に態様が開示されているが、考察で述べたように、他の態様もまた企図されている。本開示は、典型として例示的な態様を示したものであり、限定ではない。当業者であれば、本明細書において開示される態様の原則の範囲と主旨を逸脱しない範囲で、多くの他の修正および態様を工夫することが可能である。
【発明を実施するための形態】
【０２３５】
　一側面において、本発明は一般に、少なくとも特に染色体分節または完全な染色体の欠失または重複等のコピー数多様性の有無を決定する改良された方法に関する。前記方法は、小さい欠失または重複の検出に特に有用である。このような小さい欠失または重複は、関連する染色体分節から入手可能なデータ量が少ないために従来の方法を用いて高い特異性と感度で検出することが困難な可能性がある。前記方法は、改良された分析方法、改良されたバイオアッセイ方法、および改良された分析方法と改良されたバイオアッセイ方法の組み合わせを含む。本発明の方法はまた、検査される細胞または核酸分子に存在する百分率が小さい欠失または重複を検出するのに用いることができる。これにより、病気の発症前（例えば、前癌性段階等）あるいは病気の初期段階（例えば、多数の病気の細胞（例えば、癌細胞）に欠失または重複が蓄積する前等）に欠失または重複を検出することがで

きる。病気または障害と関連づけられた欠失または重複をより正確検出することにより、その病気または障害の診断、予知、防止、遅延、安定化、または治療を行うための改良された方法を提供することができる。数種の欠失または重複は、癌あるいはいくつかの精神障害または身体障害と関連づけられることが知られている。

【０２３６】

他の側面では、本発明は一般に、少なくとも特に一塩基多様性（ＳＮＶ）を検出する改良された方法に関する。これらの改良された方法は、改良された分析方法、改良されたバイオアッセイ方法、および改良された分析方法と改良されたバイオアッセイ方法の組み合わせを用いた改良された方法を含む。特定の例示的な態様において、前記方法を用いて、例えば、ＳＮＶが非常に低濃度（例えば、循環遊離ＤＮＡ試料等のＳＮＶ座位の正常なコピーの総数の１０％、５％、４％、３％、２．５％、２％、１％、０．５％、０．２５％、または０．１％未満等）で存在する試料において癌の検出、診断、監視、または病期分類が行われる。すなわち、特定の例示的な態様において、これらの方法は、遺伝子座位に存在する正常な多型アレルに対して比較的低い百分率の変異または多様体が存在する試料に特に好適である。最後に、コピー数多様性を検出するための改良された方法と一塩基多様性を検出するための改良された方法を組み合わせた方法が本明細書において開示される。

【０２３７】

癌等の病気の治療の成功は、初期診断、疾患の正しい病期分類、有効治療投薬計画の選択、および再発を防止または検出するための緊密な監視に依存することが多い。癌診断の場合、組織生検から得られた腫瘍材料の組織学的評価が、最も信頼性の高い方法であると考えられがちである。しかしながら、生検に基づく試料化が有する侵襲性のため、集団検診および定期検診の場合にはこのような試料化の実行が困難である。したがって、本発明の方法は、希望すれば、比較的低コストで短い所要時間で非侵襲的な方法で行えるという利点を有する。本発明の方法により用いられてもよい標的配列決定は、必要なリードがショットガン配列決定よりも少なく（例えば、４０００万リードではなく、数百万リード）、これによりコストが低減する。用いられてもよい多重ＰＣＲおよび次世代配列決定により、処理能力が増加し、コストが低下する。

【０２３８】

いくつかの態様では、前記方法を用いて個体中の欠失、重複、または一塩基多様体が検出される。欠失、重複、または一塩基多様体を有すると疑われる細胞または核酸を含む個体に由来する試料が分析されてもよい。いくつかの態様では、この試料は、欠失、重複、または一塩基多様体を有することが疑われる組織または器官（癌性が疑われる細胞または腫瘤等）に由来する。本発明の方法を用いて、欠失、重複、または一塩基多様体を有する細胞と欠失、重複、または一塩基多様体を有しない細胞の混合物中の１個の細胞または少数の細胞にのみ存在する欠失、重複、または一塩基多様体を検出することができる。いくつかの態様では、前記個体に由来する血液試料に由来する無細胞ＤＮＡまたは無細胞ＲＮＡが分析される。いくつかの態様では、無細胞ＤＮＡまたは無細胞ＲＮＡは、癌細胞等の細胞によって分泌される。いくつかの態様では、無細胞ＤＮＡまたは無細胞ＲＮＡは、癌細胞等、壊死またはアポトーシスを起こしている細胞によって放出される。本発明の方法を用いて、無細胞ＤＮＡまたは無細胞ＲＮＡに低い百分率でのみ存在する欠失、重複、または一塩基多様体を検出することができる。いくつかの態様では、胚に由来する１個以上の細胞が検査される。

【０２３９】

いくつかの態様では、前記方法は、胎児の非侵襲性または侵襲性出生前検査に用いられる。これらの方法は、染色体分節または染色体全体の欠失または重複（例えば、いくつかの精神障害または身体障害、学習障害、または癌に関連づけられることが知られている欠失または重複等）の有無を決定するのに用いることができる。いくつかの態様では、非侵襲出生前検査（ＮＩＰＴ）で妊娠している母親に由来する血液試料に由来する細胞、無細胞ＤＮＡ、または無細胞ＲＮＡが検査される。前記方法により、妊娠している母親由来の

細胞、無細胞ＤＮＡ、または無細胞ＲＮＡの量が多くても、胎児に由来する細胞、無細胞ＤＮＡ、または無細胞ＲＮＡの欠失または重複を検出することができる。いくつかの態様では、侵襲性出生前検査で胎児に由来する試料（例えば、ＣＶＳまたは羊水穿刺試料）に由来するＤＮＡまたはＲＮＡが検査される。前記試料が妊娠している母親由来のＤＮＡまたはＲＮＡで汚染されている場合は、前記方法を用いて胎児のＤＮＡまたはＲＮＡの欠失または重複を検出することができる。

【０２４０】

コピー数多様性の有無の決定に加えて、必要に応じて１種以上の他の因子を分析することができる。これらの因子を用いて、診断（例えば癌の有無または癌の危険性の増加の決定、癌の分類、または癌の病期分類）あるいは予後の正確度を高めることができる。これらの因子を用いて、対象者に有効と思われる特定の治療法または治療計画を選択することもできる。このような因子の例としては、多型または変異の有無、無細胞ＤＮＡ、無細胞ＲＮＡ、ミクロＲＮＡ（ｍｉＲＮＡ）の総量あるいは特定のどれかの変化（増加または低下）量、変化した（増加または低下した）腫瘍分画、メチル化の変化（増加または低下）量、変化した（増加または低下した）ＤＮＡの完全性、変化した（増加または低下した）あるいは別のｍＲＮＡスプライシング等が挙げられる。

【０２４１】

以下に、相決定済みデータ（例えば、推定または測定された相決定済みデータ）または相未決定データを用いて欠失または重複を検出するための方法、検査可能な試料、試料を調製、増幅、および定量するための方法、遺伝子データの相決定法、検出可能な多型，変異、核酸の変化、ｍＲＮＡスプライシングの変化、および核酸の量の変化、前記方法に由来する結果のデータベース、他の危険因子およびスクリーニング方法、診断または治療可能な癌、癌治療、治療法を検査するための癌モデル、および治療法を処方および実施するための方法について記載する。

相決定済みデータを用いて倍数性を決定するための方法の例

【０２４２】

本発明の前記方法のいくつかは、ＣＮＶの検出に相未決定データを用いた場合に比べて、相決定済みデータを用いた場合に擬似陰性率および偽陽性率が低下するという発見に一部基づいている（図２０Ａ～図２７）。この改良は、少ない量で存在するＣＮＶを含む試料にとって恩恵が多い。よって、（例えば、異なる座位のアレル比率が、同じまたは異なるハプロタイプが異常な量で存在している可能性を示すことを考慮せずに、ある染色体または染色体分節の１箇所以上の座位に存在するアレルの比率または集計値（例えば、平均値）を与えるための集計アレル比率を算出する方法等）相未決定データを用いた場合に比べて、相決定済みデータはＣＮＶ検出の正確度を高める。相決定済みデータを用いることにより、測定アレル比率と期待アレル比率との差がノイズによるものなのか、あるいはＣＮＶの存在によるものなのかをより正確に決定することができる。例えば、ある領域の大部分またはすべての座位の測定アレル比率と期待アレル比率との差が同じハプロタイプの過剰出現を示している場合、ＣＮＶが存在する可能性がより高い。あるハプロタイプのアレルの連鎖を用いると、測定済み遺伝子データが、（ランダムノイズではなく）同じハプロタイプの過剰出現と一致するかどうかを決定することができる。これに対して、測定アレル比率と期待アレル比率との差がノイズ（例えば、実験誤差）のみによるものである場合、いくつかの態様では、約半分の時間は第一のハプロタイプが過剰出現しており、残りの約半分の時間は第二のハプロタイプが過剰出現しているようである。

【０２４３】

ＳＮＰ間の連鎖、および胎児に成長した胚を形成した配偶子を発生させた減数分裂時に発生した交差の尤度を考慮することで正確度を高めることができる。１つ以上の仮説についてアレル測定の期待分布を作成する際に連鎖を用いると、連鎖を用いない場合に比べてより実際のものに一致したアレル測定の期待分布を作成することができる。例えば、互いに近接する２つのＳＮＰ１および２があり、母親の一方の相同体１にあるＳＮＰ１がＡ、ＳＮＰ２がＡ、他方の相同体２にあるＳＮＰ１がＢ、ＳＮＰ２がＢであるとする。父親の

両方の相同体にあるＳＮＰがどちらもＡであり、胎児のＳＮＰ１でＢが測定された場合、相同体２が胎児によって受け継がれたこと、したがって、胎児のＳＮＰ２にＢが存在する尤度がより高いことを示している。連鎖を考慮するモデルではこれを予測することができるが、連鎖を考慮しないモデルでは予測できない。あるいは、母親のＳＮＰ１がＡＢ、近隣のＳＮＰ２がＡＢである場合、この位置での母親の三染色体性に一致する２つの仮説、すなわち、一致型コピー誤り（胎児の初期の発達時の減数分裂ＩＩまたは有糸分裂での不分離）を含むという仮説および部分一致型コピー誤り（減数分裂Ｉでの不分離）を含むという仮説を用いることができる。一致型コピー誤りの三染色体性であれば、胎児が母親からＳＮＰ１にＡＡを受け継いだ場合、この胎児は、ＳＮＰ２にＡＢではなくＡＡまたはＢＢのいずれかを受け継いでいる可能性がより高い。部分一致型コピー誤りであれば、胎児は母親からＳＮＰ１および２の両方にＡＢを受け継いでいる。連鎖を考慮したＣＮＶ呼び出し法で作成したアレル分布仮説はこれらの予測を行うことができる。したがって、連鎖を考慮しないＣＮＶ呼び出し法よりも、より実際のアレル測定に一致させることができる。

【０２４４】

いくつかの態様では、相決定済み遺伝子データを用いて、（例えば、１個以上の細胞のゲノム中、あるいは無細胞ＤＮＡまたは無細胞ＲＮＡ中の）個体のゲノム中の第二の相同染色体分節に比較して第一の相同染色体分節のコピー数の過剰出現があるかどうかが決定される。過剰出現の例としては、第一の相同染色体分節の重複、または第二の相同染色体分節の欠失が挙げられる。いくつかの態様では過剰出現はない、というのは第一および第二の相同染色体分節が等しい割合（例えば、二倍体試料中に各分節のコピーが１つ）で存在するからである。いくつかの態様では、核酸試料の算出済みアレル比率を期待アレル比率と比較し、以下でさらに説明するように過剰出現があるかどうかが決定される。本明細書において、用語「第二の相同染色体分節と比較しての第一の相同染色体分節」は、ある染色体分節の第一の相同体とその染色体分節の第二の相同体を意味する。

【０２４５】

いくつかの態様では、前記方法は、前記第一の相同染色体分節にある多型座位の集合のうちの座位ごとに第一の相同染色体分節についてその座位に存在するアレルの同一性を含む第一の相同染色体分節の相決定済み遺伝子データを得ることと、前記第二の相同染色体分節にある多型座位の集合のうちの座位ごとに前記第二の相同染色体分節にあるその座位に存在するアレルの同一性を含む第二の相同染色体分節の相決定済み遺伝子データを得ることと、前記多型座位の集合のうちの各座位に存在する各アレルについて、前記個体由来の１個以上の目的細胞および１個以上の目的でない細胞に由来するＤＮＡまたはＲＮＡ試料中に存在する各アレルの量を含む測定されたアレルの遺伝子データを得ることとを含む。いくつかの態様では、前記方法は、第一の相同染色体分節の過剰出現度を規定する１つ以上の仮説の集合を列挙し、前記仮説のそれぞれについて、前記１個以上の目的細胞に由来するＤＮＡまたはＲＮＡの前記試料中の総ＤＮＡまたはＲＮＡに対する１つ以上の起こりうる比率について、得られた相決定済み遺伝子データから前記試料中の複数の座位の期待遺伝子データを算出し、ＤＮＡまたはＲＮＡの可能な比率ごと、および仮説ごとに、そのＤＮＡまたはＲＮＡの起こりうる比率とその仮説について得られた試料の遺伝子データと前記試料の期待遺伝子データのデータ適合を（コンピュータ等で）算出し、前記データ適合に応じて１つ以上の仮説をランク付けし、最高ランクの仮説を選択して、前記個体に由来する１個以上の細胞のゲノム中の第一の相同染色体分節のコピー数の過剰出現度を決定することとを含む。

【０２４６】

一側面において、本発明は、胎児のゲノム中の目的の染色体または染色体分節のコピー数を決定する方法を特徴とする。いくつかの態様では、前記方法は、前記胎児の少なくとも一方の生物学上の親の相決定済み遺伝子データを得ることを含み、ここで、前記相決定済み遺伝子データが前記親の第一の相同染色体分節および第二の相同染色体分節にある多型座位の集合のうちの各座位に存在するアレルの同一性を含む。いくつかの態様では、前

記方法は、各座位に存在する各アレルの量を測定することで、胎児のＤＮＡまたはＲＮＡおよびその胎児の母親に由来する母親のＤＮＡまたはＲＮＡを含むＤＮＡまたはＲＮＡの混合試料中の染色体または染色体分節にある多型座位の集合について遺伝子データを得ることを含む。いくつかの態様では、前記方法は、前記胎児のゲノムに存在する前記目的の染色体または染色体分節のコピー数を規定する１つ以上の仮説の集合を列挙することを含む。いくつかの態様では、前記方法は、前記仮説のそれぞれについて、（ｉ）得られた前記（両）親由来の相決定済み遺伝子データと任意で（ｉｉ）前記胎児に目的の染色体または染色体分節のコピーをもたらした配偶子の形成時に発生した可能性がある１つ以上の交差の確率から混合試料中の複数の座位の各座位に存在する各アレルの期待量の確率分布を（例えば、コンピュータ上等で）作成し、前記仮説のそれぞれについて、（１）前記得られた混合試料の遺伝子データと（２）その仮説ごとに混合試料中の複数の座位の各座位における各アレルの期待量の確率分布の適合を（コンピュータ等で）算出し、前記データ適合に応じて１つ以上の仮説をランク付けし、最高ランクの仮説を選択して、前記胎児のゲノム中の目的の染色体分節のコピー数を決定することとを含む。

【０２４７】

いくつかの態様では、前記方法は、本明細書で開示した方法のいずれか、または任意の周知の方法を用いて相決定済み遺伝子データを得ることを含む。いくつかの態様では、前記方法は、同時にまたは任意の順序で順に（ｉ）前記第一の相同染色体分節にある多型座位の集合のうちの座位ごとに第一の相同染色体分節についてその座位に存在するアレルの同一性を含む第一の相同染色体分節の相決定済み遺伝子データを得ることと、（ｉｉ）前記第二の相同染色体分節にある多型座位の集合のうちの座位ごとに前記第二の相同染色体分節にあるその座位に存在するアレルの同一性を含む第二の相同染色体分節の相決定済み遺伝子データを得ることと、（ｉｉｉ）前記個体に由来する１個以上の細胞に由来するＤＮＡの試料中の前記多型座位の集合のうちの各座位に存在する各アレルの量を含む測定されたアレルの遺伝子データを得ることとを含む。

【０２４８】

いくつかの態様では、前記方法は、前記試料が由来している少なくとも１個の細胞でヘテロ接合である（例えば、前記胎児においてヘテロ接合である座位および／または前記母親においてヘテロ接合である座位）多型座位の集合のうちの１箇所以上の座位についてアレル比率を算出することを含む。いくつかの態様では、特定の座位の算出済みアレル比率は、前記アレルの１つの測定量をその座位のすべてのアレルの総測定量で除したものである。いくつかの態様では、特定の座位の算出済みアレル比率は、前記アレル（例えば、前記第一の相同染色体分節にあるアレル等）の１つの測定量を、その座位の１つ以上の他のアレル（例えば、第二の相同染色体分節にあるアレル等）の測定量で除したものである。前記算出済みアレル比率は、本明細書で開示した方法のいずれかを用いて、あるいは任意の標準的な方法（例えば、本明細書に記載の算出済みアレル比率の任意の数学的変換等）で算出されてもよい。

【０２４９】

いくつかの態様では、前記方法は、ある座位の１つ以上の算出済みアレル比率を第一および第二の相同染色体分節が等しい割合で存在する場合にその座位に期待されるアレル比率と比較して前記第一の相同染色体分節のコピー数過剰があるかどうかを決定することを含む。いくつかの態様では、前記期待アレル比率は、ある座位に起こりうるアレルが存在について等しい尤度を有していると仮定する。特定の座位の算出済みアレル比率が前記アレルの１つの測定量をその座位のすべてのアレルの総測定量で除したものであるいくつかの態様では、対応する期待アレル比率は二対立遺伝子座位で０．５、三対立遺伝子座位で１／３である。いくつかの態様では、前記期待アレル比率は、すべての座位で０．５等、同じである。いくつかの態様では、前記期待アレル比率は、ある座位に起こりうるアレルが存在について異なる尤度（例えば、例えば、対象者が属する特定の集団における前記アレルのそれぞれの頻度に基づいた尤度（例えば、対象者の祖先に基づいた集団等））を有する可能性があると仮定する。このようなアレル頻度は公的に利用可能である（例えば、

以下を参照のこと。ハップマップ計画、Perlegen Human Haplotype Project（パーレジェン・ヒトハプロタイプ・プロジェクト）、ウェブページ（ncbi.nlm.nih.gov/projects/SNP/）、Sherry ST, Ward MH, Kholodov M, et al. dbSNP、ＮＣＢＩ遺伝的変異データベース、Nucleic Acids Res. 2001 Jan 1;29(1):308-11、それぞれは参照により本明細書に組み込まれる）。いくつかの態様では、前記期待アレル比率は、前記第一の相同染色体分節の過剰出現度を規定する特定の仮説について検査される特定の個体に期待されるアレル比率である。例えば、特定の個体の期待アレル比率は、（例えば、非癌性試料等の欠失または重複を有していないと思われる個体に由来する試料等）前記個体に由来する相決定済みまたは相未決定遺伝子データ、あるいは前記個体のひとり以上の親近者に由来するデータに基づいて決定されてもよい。いくつかの態様では、出生前検査の場合、前記期待アレル比率は、第一の相同染色体分節の過剰出現度を規定する特定の仮説について、妊娠している母親および胎児に由来するＤＮＡまたはＲＮＡを含む混合試料（例えば、母親の血漿、または母親由来の無細胞ＤＮＡおよび胎児由来の無細胞ＤＮＡを含む血清試料等）に期待されるアレル比率である。例えば、前記混合試料の期待アレル比率は、母親由来の遺伝子データおよび胎児について予測される遺伝子データ（例えば、母親および／または父親から受け継がれると思われるアレルについての予測等）に基づいて決定されてもよい。いくつかの態様では、母親のみに由来するＤＮＡまたはＲＮＡ試料に由来する相決定済みまたは相未決定遺伝子データ（例えば、母親の血液試料由来の軟膜等）を用いて、混合試料中の母親のＤＮＡまたはＲＮＡ由来のアレル、および胎児が母親から受け継いだ可能性がある（したがって、混合試料中の胎児のＤＮＡまたはＲＮＡに存在する可能性がある）アレルが決定される。いくつかの態様では、父親にのみ由来するＤＮＡまたはＲＮＡ試料に由来する相決定済みまたは相未決定遺伝子データを用いて、胎児が父親から受け継いだ可能性がある（したがって、混合試料中の胎児のＤＮＡまたはＲＮＡに存在する可能性がある）アレルが決定される。前記期待アレル比率は、本明細書で開示した方法のいずれか、または任意の標準的な方法（例えば、本明細書に記載の期待アレル比率の任意の数学的変換等）（２０１１年１１月１８日に提出された米国特許公開第２０１２／０２７０２１２号、参照によりその全体が本明細書に組み込まれる）を用いて算出されてもよい。

【０２５０】

いくつかの態様では、（ｉ）前記第一の相同染色体にあるその座位に存在するアレルの測定量をその座位のすべてのアレルの総測定量で除したアレル比率がその座位の期待アレル比率より大きい場合、または（ｉｉ）前記第二の相同染色体にあるその座位に存在するアレルの測定量をその座位のすべてのアレルの総測定量で除したアレル比率がその座位の期待アレル比率未満である場合のいずれかの場合、算出済みアレル比率は前記第一の相同染色体分節のコピー数過剰出現を示す。いくつかの態様では、算出済みアレル比率がその座位の期待比率よりも有意に大きい、または小さい場合は、その算出済みアレル比率は過剰出現を示すと見なされるのみである。いくつかの態様では、（ｉ）前記第一の相同染色体にあるその座位に存在するアレルの測定量をその座位のすべてのアレルの総測定量で除したアレル比率がその座位の期待アレル比率以下である場合、または（ｉｉ）前記第二の相同染色体にあるその座位に存在するアレルの測定量をその座位のすべてのアレルの総測定量で除したアレル比率がその座位の期待アレル比率以上である場合のいずれかの場合には、算出済みアレル比率は前記第一の相同染色体分節のコピー数過剰出現がないことを示す。いくつかの態様では、対応する期待比率と等しい算出済み比率は（過剰出現がないことを示すため）無視される。

【０２５１】

様々な態様では、以下の方法の１つ以上を用いて算出済みアレル比率の１つ以上が対応する期待アレル比率と比較される。いくつかの態様では、特定の座位について、差異の大きさにかかわらず前記算出済みアレル比率が期待アレル比率より大きいか、小さいかが決定される。いくつかの態様では、特定の座位について、算出済みアレル比率が期待アレル比率より大きいか小さいかにかかわらず算出済みアレル比率と期待アレル比率との差異の大きさが決定される。いくつかの態様では、特定の座位について、算出済みアレル比率が

期待アレル比率より大きいか、小さいかとその差異の大きさが決定される。いくつかの態様では、差異の大きさにかかわらず、算出済みアレル比率の平均値または加重平均値が期待アレル比率の平均値または加重平均値より大きいか、小さいかが決定される。いくつかの態様では、算出済みアレル比率の平均値または加重平均値が期待アレル比率の平均値または加重平均値より大きいか、小さいかにかかわらず、算出済みアレル比率の平均値または加重平均値と期待アレル比率の平均値または加重平均値との差異の大きさが決定される。いくつかの態様では、算出済みアレル比率の平均値または加重平均値が期待アレル比率の平均値または加重平均値より大きいか、小さいかとその差異の大きさが決定される。いくつかの態様では、算出済みアレル比率と期待アレル比率の差異の大きさの平均値または加重平均値が決定される。

【０２５２】

いくつかの態様では、１箇所以上の座位の算出済みアレル比率と期待アレル比率の差異の大きさを用いて、前記第一の相同染色体分節のコピー数過剰出現が前記細胞の１つ以上のゲノムにおける前記第一の相同染色体分節の重複または前記第二の相同染色体分節の欠失によるものであるかどうかが決定される。

【０２５３】

いくつかの態様では、以下の条件の１つ以上が満たされる場合は、前記第一の相同染色体分節のコピー数過剰出現が決定される。いくつかの態様では、前記第一の相同染色体分節のコピー数過剰出現を示す算出済みアレル比率の数が閾値より大きい。いくつかの態様では、前記第一の相同染色体分節のコピー数過剰出現でないことを示す算出済みアレル比率の数が閾値より小さい。いくつかの態様では、前記第一の相同染色体分節のコピー数過剰出現を示す算出済みアレル比率と対応する期待アレル比率との差異の大きさが閾値より大きい。いくつかの態様では、過剰出現を示すすべての算出済みアレル比率と対応する期待アレル比率との差異の大きさの和が閾値より大きい。いくつかの態様では、前記第一の相同染色体分節のコピー数過剰出現ではないことを示す算出済みアレル比率と対応する期待アレル比率との差異の大きさが閾値未満である。いくつかの態様では、前記第一の相同染色体に存在するアレルの測定量をその座位のすべてのアレルの総測定量で除した算出済みアレル比率の平均値または加重平均値が、少なくとも閾値で除した前記期待アレル比率の平均値または加重平均値より大きい。いくつかの態様では、前記第二の相同染色体に存在するアレルの測定量をその座位のすべてのアレルの総測定量で除した算出済みアレル比率の平均値または加重平均値が、少なくとも閾値で除した前記期待アレル比率の平均値または加重平均値未満である。いくつかの態様では、算出済みアレル比率と前記第一の相同染色体分節のコピー数過剰出現を予測するアレル比率とのデータ適合が閾値より小さい（データ適合が良好なことを示している）。いくつかの態様では、算出済みアレル比率と前記第一の相同染色体分節のコピー数過剰出現でないことを示すアレル比率とのデータ適合が閾値より大きい（データ適合がよくないことを示している）。

【０２５４】

いくつかの態様では、以下の条件の１つ以上が満たされる場合は、前記第一の相同染色体分節のコピー数過剰出現が決定される。いくつかの態様では、前記第一の相同染色体分節のコピー数過剰出現を示す算出済みアレル比率の数が閾値より小さい。いくつかの態様では、前記第一の相同染色体分節のコピー数過剰出現でないことを示す算出済みアレル比率の数が閾値より大きい。いくつかの態様では、前記第一の相同染色体分節のコピー数過剰出現を示す算出済みアレル比率と対応する期待アレル比率との差異の大きさが閾値より小さい。いくつかの態様では、前記第一の相同染色体分節のコピー数過剰出現でないことを示す算出済みアレル比率と対応する期待アレル比率との差異の大きさが閾値より大きい。いくつかの態様では、第一の相同染色体に存在するアレルの測定量をその座位のすべてのアレルの総測定量で除した算出済みアレル比率の平均値または加重平均値から期待アレル比率の平均値または加重平均値を引いたものが閾値より小さい。いくつかの態様では、期待アレル比率の平均値または加重平均値から前記第二の相同染色体に存在するアレルの測定量をその座位のすべてのアレルの総測定量で除した算出済みアレル比率の平均値また

は加重平均値を引いたものが、閾値より小さい。いくつかの態様では、算出済みアレル比率と前記第一の相同染色体分節のコピー数過剰出現を予測するアレル比率とのデータ適合が、閾値より大きい。いくつかの態様では、算出済みアレル比率と前記第一の相同染色体分節のコピー数過剰出現でないことを示すアレル比率とのデータ適合が、閾値より小さい。いくつかの態様では、前記閾値は、目的のＣＮＶを有することがわかっている試料および／またはＣＮＶがないことがわかっている試料を試験して決定される。

【０２５５】

いくつかの態様では、前記第一の相同染色体分節のコピー数過剰出現があるかどうかを決定することは、前記第一の相同染色体分節の過剰出現度を規定する１つ以上の仮説の集合を列挙することを含む。仮説の一例は、第一および第二の相同染色体分節が等しい割合（例えば、二倍体試料中、各分節のコピーが１つずつ）で存在する場合は過剰出現がないというものである。仮説の別の例は、１回以上重複された（例えば、第二の相同染色体分節のコピー数に比べて第一の相同染色体の余剰コピーが１つ、２つ、３つ、４つ、５つ、またはそれ以上ある）第一の相同染色体分節を含む。仮説の他の例は、第二の相同染色体分節の欠失を含む。他の例示的仮説は、第一および第二の相同染色体分節の両方の欠失である。いくつかの態様では、少なくとも１個の細胞中でヘテロ接合である座位（例えば、前記胎児においてヘテロ接合である座位および／または前記母親においてヘテロ接合である座位））の予測アレル比率は、仮説ごとに、過剰出現度が仮説によって規定されたと仮定して推定される。いくつかの態様では、前記算出済みアレル比率を前記予測アレル比率と比較して仮説が正しい尤度が算出され、最も可能性が高い仮説が選択される。

【０２５６】

いくつかの態様では、検定統計量の期待分布は、仮説ごとに前記予測アレル比率を用いて算出される。いくつかの態様では、前記算出済みアレル比率を用いて算出された検定統計量と前記予測アレル比率を用いて算出された検定統計量の期待分布を比較して仮説が正しい尤度が算出されて、最も可能性が高い仮説が選択される。

【０２５７】

いくつかの態様では、少なくとも１個の細胞中でヘテロ接合である座位（例えば、前記胎児においてヘテロ接合である座位および／または前記母親においてヘテロ接合である座位））の予測アレル比率は、第一の相同染色体分節の相決定済み遺伝子データ、第二の相同染色体分節の相決定済み遺伝子データ、および過剰出現度が仮説によって規定されたと仮定して推定される。いくつかの態様では、前記算出済みアレル比率と前記予測アレル比率を比較して仮説が正しい尤度が算出されて、最も可能性が高い仮説が選択される。

混合試料の使用

【０２５８】

なお、多くの態様で、試料は１個以上の目的細胞および１個以上の目的でない細胞に由来するＤＮＡまたはＲＮＡを含む混合試料である。いくつかの態様では、対象細胞は、目的の欠失または重複等、ＣＮＶを有する細胞であり、非対象細胞は、目的のコピー数多様性を有しない細胞（例えば、目的の欠失または重複を有する細胞および検査対象の欠失または重複のいずれも有しない細胞の混合物）である。いくつかの態様では、前記対象細胞は、病気または障害あるいは病気または障害の危険性の増加に関連づけられた細胞（例えば、癌細胞等）であり、前記非対象細胞は、病気または障害あるいは病気または障害の危険性の増加に関連づけられていない細胞（例えば、非癌性細胞等）である。いくつかの態様では、前記対象細胞はすべて、同じＣＮＶを有する。いくつかの態様では、２個以上の対象細胞は異なるＣＮＶを有する。いくつかの態様では、前記対象細胞の１個以上は、少なくとも他の１個の対象細胞には見られない病気または障害あるいは病気または障害の危険性の増加に関連づけられたＣＮＶ、多型、または変異を有する。そのようないくつかの態様では、ある試料に由来する総細胞のうちの、病気または障害あるいは病気または障害の危険性の増加に関連づけられた細胞の分画は、その試料でこれらＣＮＶ、多型、または変異のうち最も高い頻度の分画と同じかそれ以上であると仮定される。例えば、細胞の６％がＫ－ｒａｓ変異を有し、８％がＢＲＡＦ変異を有する場合は、細胞の少なくとも８％

が癌性であると仮定される。

【０２５９】

いくつかの態様では、前記試料中の前記１個以上の目的細胞に由来するＤＮＡ（またはＲＮＡ）の総ＤＮＡ（またはＲＮＡ）に対する比率が算出される。いくつかの態様では、前記第一の相同染色体分節の過剰出現度を規定する１つ以上の仮説の集合が列挙される。いくつかの態様では、少なくとも１個の細胞中でヘテロ接合である座位（例えば、前記胎児においてヘテロ接合である座位および／または前記母親においてヘテロ接合である座位））の予測アレル比率は、ＤＮＡまたはＲＮＡの算出済み比率および過剰出現度が仮説によって規定されたと仮定して仮説ごとに推定される。いくつかの態様では、前記算出済みアレル比率と前記予測アレル比率を比較して仮説が正しい尤度が算出されて、最も可能性が高い仮説が選択される。

【０２６０】

いくつかの態様では、前記予測アレル比率およびＤＮＡまたはＲＮＡの算出済み比率を用いて算出された検定統計量の期待分布が、仮説ごとに推定される。いくつかの態様では、前記算出済みアレル比率とＤＮＡまたはＲＮＡの算出済み比率を用いて算出された検定統計量を前記予測アレル比率とＤＮＡまたはＲＮＡの算出済み比率を用いて算出された検定統計量の期待分布と比較して仮説が正しい尤度が決定され、最も可能性が高い仮説が選択される。

【０２６１】

いくつかの態様では、前記方法は、前記第一の相同染色体分節の過剰出現度を規定する１つ以上の仮説の集合を列挙することを含む。いくつかの態様では、前記方法は、仮説ごとに、（ｉ）その仮説によって規定された過剰出現度が与えられた場合は少なくとも１個の細胞中でヘテロ接合である座位（例えば、前記胎児においてヘテロ接合である座位および／または前記母親においてヘテロ接合である座位））の予測アレル比率、または（ｉｉ）ＤＮＡまたはＲＮＡの１つ以上の起こりうる比率について、前記予測アレル比率、および前記１個以上の目的細胞に由来する前記ＤＮＡまたはＲＮＡの、前記試料中の総ＤＮＡまたはＲＮＡに対する起こりうる比率を用いて算出された検定統計量の期待分布のいずれかを推定することを含む。いくつかの態様では、（ｉ）前記予測アレル比率に対する算出済みアレル比率、または（ｉｉ）前記算出済みアレル比率および前記ＤＮＡまたはＲＮＡの起こりうる比率を用いて算出された検定統計量のいずれかと、前記予測アレル比率および前記ＤＮＡまたはＲＮＡの起こりうる比率を用いて算出された検定統計量の期待分布とを比較して、データ適合が算出される。いくつかの態様では、前記データ適合に応じて１つ以上の仮説がランク付けされて、最高ランクの仮説が選択される。いくつかの態様では、前記データ適合を算出する工程、前記仮説をランク付けする工程、最高ランクの仮説を選択する工程の１つ以上において技術またはアルゴリズム（例えば、検索アルゴリズム等）が用いられる。いくつかの態様では、前記データ適合はベータ二項分布に対する適合または二項分布への適合である。いくつかの態様では、前記技術またはアルゴリズムが、最尤推定、最大事後確率推定、ベイズ推定、動的推定（例えば、動的ベイズ推定等）、および期待値最大化推定からなる群から選択される。いくつかの態様では、前記方法は、前記技術またはアルゴリズムを前記得られた遺伝子データおよび前記期待遺伝子データに適用することを含む。

【０２６２】

いくつかの態様では、前記方法は、前記１個以上の目的細胞に由来するＤＮＡまたはＲＮＡの前記試料中の総ＤＮＡまたはＲＮＡに対する比率について、下限から上限までの範囲の起こりうる比率の区分を作成することを含む。いくつかの態様では、前記第一の相同染色体分節の過剰出現度を規定する１つ以上の仮説の集合が列挙される。いくつかの態様では、前記方法は、前記区分におけるＤＮＡまたはＲＮＡの起こりうる比率の各比率と各仮説について、（ｉ）前記ＤＮＡまたはＲＮＡの起こりうる比率およびその仮説によって規定された過剰出現度が与えられた場合は少なくとも１個の細胞中でヘテロ接合である座位（例えば、前記胎児においてヘテロ接合である座位および／または前記母親においてヘ

テロ接合である座位）の予測アレル比率、または（ｉｉ）前記予測アレル比率と前記ＤＮＡまたはＲＮＡの起こりうる比率を用いて算出された検定統計量の期待分布のいずれかを推定することを含む。いくつかの態様では、前記方法は、（ｉ）前記予測アレル比率に対する算出済みアレル比率または（ｉｉ）前記算出済みアレル比率および前記ＤＮＡまたはＲＮＡの起こりうる比率を用いて算出された検定統計量のいずれかと、前記予測アレル比率および前記ＤＮＡまたはＲＮＡの起こりうる比率を用いて算出された検定統計量の期待分布とを比較することで前記区分における前記ＤＮＡまたはＲＮＡの起こりうる比率の各比率と各仮説について、前記仮説が正しい尤度を算出することを含む。いくつかの態様では、前記区分の中で起こりうる比率の比率ごとにその仮説の確率を組み合わせて各仮説の複合確率が決定され、最も高い複合確率を有する仮説が選択される。いくつかの態様では、各仮説の複合確率は、前記起こりうる比率が正しい比率であるという尤度に基づいて特定の起こりうる比率について仮説の確率を加重平均することで各仮説が決定される。

【０２６３】

いくつかの態様では、最尤推定、最大事後確率推定、ベイズ推定、動的推定（例えば、動的ベイズ推定等）、および期待値最大化推定からなる群から選択される技術を用いて、前記１個以上の目的細胞に由来するＤＮＡまたはＲＮＡの前記試料中の総ＤＮＡまたはＲＮＡに対する比率が推定される。いくつかの態様では、前記１個以上の目的細胞に由来するＤＮＡまたはＲＮＡの前記試料中の総ＤＮＡまたはＲＮＡに対する比率は、目的のＣＮＶの２つ以上（または、すべて）について同じであると仮定される。いくつかの態様では、前記１個以上の目的細胞に由来するＤＮＡまたはＲＮＡの前記試料中の総ＤＮＡまたはＲＮＡに対する比率は、目的のＣＮＶごとに算出される。

不十分な相決定済みデータを用いる方法の例

【０２６４】

当然のことだが、多くの態様で、不十分な相決定済みデータが用いられている。例えば、第一の相同染色体分節および／または第二の相同染色体分節にある１箇所以上の座位にどのアレルが存在するかについて、１００％確実に分かっていない場合がある。いくつかの態様では、個体の起こりうるハプロタイプ（例えば、集団に基づいたハプロタイプ頻度に基づくハプロタイプ等）の優先順位が各仮説の確率の算出に用いられる。いくつかの態様では、起こりうるハプロタイプの優先順位は、遺伝子データの相決定を行う他の方法または他の対象者（例えば、優先順位の高い対象者等）に由来する相決定済みデータを用いて調整されて、前記個体の相決定に基づいたインフォマティックスに用いられた集団データが改良される。

【０２６５】

いくつかの態様では、前記相決定済み遺伝子データは、相決定済み遺伝子データの２つ以上の起こりうる集合に関する確率データを含み、相決定済みデータの起こりうる集合のそれぞれは、前記第一の相同染色体分節にある多型座位の集合のうちの各座位に存在するアレルの起こりうる同一性と、前記第二の相同染色体分節にある多型座位の集合のうちの各座位に存在するアレルの起こりうる同一性を含む。いくつかの態様では、前記仮説の少なくとも１つの確率は、相決定済み遺伝子データの起こりうる集合のそれぞれについて決定される。いくつかの態様では、前記仮説の複合確率は、相決定済み遺伝子データの起こりうる集合のそれぞれについて仮説の確率を組み合わせて決定され、最も高い複合確率を有する仮説が選択される。

【０２６６】

本明細書で開示された前記方法のいずれかまたは任意の周知の方法を用いて、（例えば、集団に基づくハプロタイプ頻度を用いて最も可能性の高い相を推論する等）本発明の方法に用いるための不十分な相決定済みデータが作成されてもよい。いくつかの態様では、より小さい分節のハプロタイプを確率的に組み合わせて相決定済みデータが得られる。例えば、第一の領域のあるハプロタイプと同じ染色体の他の領域の他のハプロタイプの起こりうる組み合わせに基づいて起こりうるハプロタイプを決定することができる。異なる領域にある特定のハプロタイプが同じ染色体にある同じ、より大きいハプロタイプブロック

の一部である確率は、例えば、集団に基づくハプロタイプ頻度および／または異なる領域間の周知の組換え率を用いて決定することができる。

【０２６７】

いくつかの態様では、二染色体性の帰無仮説に単一仮説棄却試験が用いられる。いくつかの態様では、二染色体性仮説の確率が算出されて、その確率が所定の閾値未満（例えば、１，０００分の１未満）である場合には二染色体性の仮説が棄却される。帰無仮説が棄却される場合は、不十分な相決定済みデータのエラーによるもの、あるいはＣＮＶの存在によるものである可能性がある。いくつかの態様では、（例えば、バイオインフォマティックスに基づいて推論された相決定済みデータではなく、実際の相決定済みデータを得るための本明細書に開示された分子相決定法のいずれかから得られた相決定済みデータ等）より正確な相決定済みデータが得られる。いくつかの態様では、前記のより正確な相決定済みデータを用いて二染色体性仮説の確率が再算出されて、それでも二染色体性仮説が棄却されるかどうかが決定される。この仮説の棄却は、前記染色体分節の重複または欠失が存在することを示す。必要であれば、閾値を調整することで偽陽性率を変更することができる。

相決定済みデータを用いて倍数性を決定する実施形態の別の例

【０２６８】

本明細書の例示的態様では、個体の試料中の染色体分節の倍数性を決定する方法が提供される。前記方法は、以下の工程を含む。

ａ．前記試料中の前記染色体分節にある多型座位の集合のうちの各座位に存在する各アレルの量を含むアレル頻度データを受け、

ｂ．前記アレル頻度データの相を推定することによって前記多型座位の集合に関する相決定済みアレル情報を作成し、

ｃ．前記アレル頻度データを用いて、異なる倍数性状態に関する前記多型座位のアレル頻度の個別確率を作成し、

ｄ．前記個別確率と前記相決定済みアレル情報を用いて前記多型座位の集合に関する同時確率を作成し、

ｅ．前記同時確率に基づき、染色体倍数性を示す最適合モデルを選択することにより、前記染色体分節の倍数性を決定すること。

【０２６９】

本明細書で開示されるように、当該分野で周知の方法によりアレル頻度データ（本明細書では測定されたアレルの遺伝子データともいう）を作成することができる。例えば、定量ＰＣＲまたはマイクロアレイを用いて前記データを作成することができる。例示的一態様では、前記データは、核酸配列データ、特に高処理核酸配列データを用いて作成される。

【０２７０】

ある具体例では、前記アレル頻度データを用いて個別確率が作成される前に前記アレル頻度データのエラーが修正される。例示的の特定の態様では、前記修正されたエラーはアレル増幅効率の偏りを含む。他の態様では、前記修正されたエラーは、環境不純物および遺伝型解析による不純物を含む。いくつかの態様では、修正されたエラーは、アレル増幅による偏り、環境不純物、および遺伝型解析による不純物を含む。

【０２７１】

特定の態様では、前記個別確率は、前記多型座位の集合に関する異なる倍数性状態とアレル不均衡率の両方のモデルのセットを用いて作成される。これらの態様および他の態様では、前記同時確率は、前記染色体分節にある多型座位間の連鎖を考慮することによって作成される。

【０２７２】

したがって、これらの態様のいくつかを組み合わせた本明細書の例示的一態様では、個体の試料中の染色体倍数性を検出する方法が提供される。前記方法は以下の工程を含む。

ａ．前記個体の染色体分節にある多型座位の集合に存在するアレルに関する核酸配列デー

タを受け、
ｂ．前記核酸配列データを用いて前記多型座位の集合におけるアレル頻度を検出し、
ｃ．検出された前記アレル頻度のアレル増幅効率の偏りを修正することによって、前記多型座位の集合に関する修正済みアレル頻度を作成し、
ｄ．前記核酸配列データの相を推定することによって前記多型座位の集合に関する相決定済みアレル情報を作成し、
ｅ．前記修正済みアレル頻度を前記多型座位の集合に関する異なる倍数性状態とアレル不均衡率のモデルのセットと比較することにより、異なる倍数性状態に関する前記多型座位のアレル頻度の個別確率を作成し、
ｆ．前記個別確率を前記染色体分節にある多型座位間の連鎖を考慮しながら組み合わせることにより、前記多型座位の集合に関する同時確率を作成し、
ｇ．前記同時確率に基づき、染色体異数性を示す最適合モデルを選択すること。
【０２７３】
本明細書に開示されるように、前記多型座位の集合に関する異なる倍数性状態と平均アレル不均衡率の両方のモデルまたは仮説の集合を用いて前記個別確率を生成することができる。例えば、特定の例では、前記染色体分節の第一の相同体および前記染色体分節の第二の相同体の倍数性状態をモデル化することで個別確率が生成される。モデル化される倍数性状態は以下のものを含む。
【０２７４】
（１）すべての細胞は、前記染色体分節の第一の相同体または第二の相同体に欠失あるいは増幅がない。
【０２７５】
（２）少なくともいくつかの細胞は、前記染色体分節の第一の相同体の欠失、または第二の相同体の増幅を有する。
【０２７６】
（３）少なくともいくつかの細胞は、前記染色体分節の第二の相同体の欠失、または第一の相同体の増幅を有する。
【０２７７】
なお、上記モデルはまた、あるモデルを抑制するのに用いられる仮説と言うこともできる。したがって、上に示されたのは、用いることができる３つの仮説である。
【０２７８】
モデル化された平均アレル不均衡率は、前記染色体分節の実際の平均アレル不均衡率を含む平均アレル不均衡率の任意の範囲を含む。例えば、特定の例示的態様では、モデル化された平均アレル不均衡率の範囲は、下限が０、０．１、０．２、０．２５、０．３、０．４、０．５、０．６、０．７５、１、２、２．５、３、４、および５％、上限が１、２、２．５、３、４、５、１０、１５、２０、２５、３０、４０、５０、６０、７０、８０、９０、９５、および９９％の範囲であってもよい。この範囲のモデル化の間隔は、用いる計算能力と分析に使える時間に応じて任意の間隔であってもよい。例えば、間隔を０．０１、０．０５、０．０２、または０．１としてモデル化することができる。
【０２７９】
特定の例示的態様では、前記試料は、染色体分節の平均アレル不均衡率が０．４％～５％である。特定の態様では、前記平均アレル不均衡率は小さい。これら態様では、平均アレル不均衡率は通常１０％未満である。特定の例示的態様では、前記アレル不均衡率は、下限が０．２５、０．３、０．４、０．５、０．６、０．７５、１、２、２．５、３、４、および５％、上限が１、２、２．５、３、４、および５％である。他の例示的態様では、前記平均アレル不均衡率は、下限が０．４、０．４５、０．５、０．６、０．７、０．８、０．９、または１．０％、上限が０．５、０．６、０．７、０．８、０．９、１．０、１．５、２．０、３．０、４．０、または５．０％である。例えば、ある例では、前記試料の平均アレル不均衡率は、０．４５～２．５％である。他の例では、前記平均アレル不均衡率は０．４５、０．５、０．６、０．８、０．８、０．９、または１．０の感度で

検出される。本発明の方法に用いる低アレル不均衡率の試料の例としては、循環腫瘍ＤＮＡを有する癌を有する個体に由来する血漿試料、または循環胎児ＤＮＡを有する妊婦に由来する血漿試料が挙げられる。

【０２８０】

なお、ＳＮＶの場合、異常なＤＮＡの割合は、通常、変異体アレル頻度（ある座位に存在する変異体アレル数／その座位に存在する総アレル数）を用いて測定される。腫瘍における２つの相同体の量の差は類似しているので、ＣＮＶについて異常なＤＮＡの割合を平均アレル不均衡率（ＡＡＩ）を用いて測定する。平均アレル不均衡率は｜（Ｈ１－Ｈ２）｜／（Ｈ１＋Ｈ２）で定義され、ここでＨｉは試料中の相同体ｉの平均コピー数であり、Ｈｉ／（Ｈ１＋Ｈ２）は相同体ｉの分数の量、すなわち相同体比率である。最大相同体比率は、最も量が多い相同体の相同体比率である。

【０２８１】

試験脱落率は、すべてのＳＮＰを用いて推定された、リードのないＳＮＰの百分率である。単一アレル脱落（ＡＤＯ）率は、ヘテロ接合ＳＮＰのみを用いて推定された、アレルが１つだけ存在するＳＮＰの百分率である。遺伝型信頼度は、Ｂアレルリードであった各ＳＮＰに存在するリード数に対する二項分布を適合させて、そのＳＮＰの注目領域の倍数性状態を用いて各遺伝型の確率を推定して決定することができる。

【０２８２】

腫瘍組織試料の場合、染色体異数性（この段落ではＣＮＶとして例示されている）は、アレル頻度分布間の遷移によって示すことができる。血漿試料では、前記腫瘍試料から推定されたハプロタイプ情報を用いて、同じ個体に由来する腫瘍試料もＣＮＶを有する領域で血漿ＣＮＶを探す最尤アルゴリズムによりＣＮＶを同定することができる。このアルゴリズムは、（１）すべての細胞は正常である（アレル不均衡率がない）、（２）いくつか／すべての細胞は相同体１の欠失または相同体２の増幅を有する、または（３）いくつか／すべての細胞は相同体２の欠失または相同体１の増幅を有するという３つの仮説集合について、すべてのアレル不均衡率にわたって期待アレル頻度を０．０２５％の間隔でモデル化することができる。本明細書において例示されるように、ＳＮＰ座位の連鎖を考慮した特定の例示的態様では、すべてのヘテロ接合ＳＮＰのアレル頻度の期待値と観測値のベータ二項モデルに基づき、ベイズ分類器を用いてＳＮＰごとに各仮説の尤度を決定することができ、次いで、複数のＳＮＰの同時尤度を算出することができる。次に、最も可能性の高い仮説を選択することができる。

【０２８３】

腫瘍の平均Ｎ個のコピーを有する染色体領域について考える。ここで、ｃは、ある二染色体の領域における正常な細胞および腫瘍細胞の混合物に由来する血漿中のＤＮＡ分画を示す。以下のようにＡＡＩが算出される。

【数１】

$$AAI = \frac{c|N-2|}{2+c(N-2)}$$

【０２８４】

特定の例では、前記アレル頻度データを用いて個別確率を作成する前に、前記アレル頻度データのエラーが修正される。本明細書ではエラーの異なる種類および／または偏りの補正が開示される。特定の例示的態様では、修正されたエラーはアレル増幅効率の偏りである。他の態様では、修正されたエラーは、環境不純物および遺伝型解析による不純物を含む。いくつかの態様では、修正されたエラーは、アレル増幅による偏り、環境不純物、および遺伝型解析による不純物を含む。

【０２８５】

当然ではあるが、アレル増幅効率の偏りは、検査試料を含む実験による決定の一部としてアレルに対して決定することができる、あるいは効率が算出されているアレルを含む試料群を用いて異なる時点で決定することができる。環境不純物および遺伝型解析による不純物は、通常、検査試料の分析と同じ一操作で決定される。

【０２８６】

特定の態様では、環境不純物および遺伝型解析による不純物は、試料中のホモ接合のアレルについて決定される。当然ではあるが、ある座位が集団の中で比較的高いヘテロ接合性を有しているために分析に選択されたとしても、ある個体に由来する任意の試料中のいくつかの座位はヘテロ接合であり、他の座位はホモ接合である。いくつかの態様では、ある個体のヘテロ接合座位を用いて染色体分節の倍数性が決定されてもよいが、ホモ接合座位を用いて環境不純物および遺伝型解析による不純物を算出することができるので有益である。

【０２８７】

特定の例では、前記選択は、前記モデルに関して相決定済みアレル情報と推定アレル頻度との差異の大きさを分析することで行われる。

【０２８８】

ある例では、アレル頻度の個別確率は、前記多型座位の集合のアレル頻度の期待値と観測値のベータ二項モデルに基づいて作成される。ある例では、前記個別確率はベイズ分類器を用いて作成される。

【０２８９】

特定の例示的態様では、前記核酸配列データは、多重増幅反応によって生成された一連の増幅産物の複数のコピーの高処理ＤＮＡ配列決定を実行することによって作成され、ここで、前記一連の増幅産物のうちの各増幅産物は、前記多型座位の集合のうちの少なくとも１つの多型座位にわたり、前記多型座位の集合のうちの各多型座位が増幅される特定の態様では、前記多重増幅反応は、反応の少なくとも１／２が制限的なプライマー条件で実行される。いくつかの態様では、多重反応の１／１０、１／５、１／４、１／３、１／２、またはすべての反応で制限的なプライマー濃度が用いられる。本明細書では、ＰＣＲ等、増幅反応で制限的なプライマー条件を得られると思われる因子が提供される。

【０２９０】

特定の態様では、本明細書で提供される方法は、複数の染色体にわたって複数の染色体分節の倍数性を検出する。したがって、これらの態様における染色体倍数性は、前記試料中の染色体分節の集合について決定される。これらの態様では、より多くの多重増幅反応が必要である。したがって、これらの態様での多重増幅反応は、例えば、２，５００～５０，０００回の多重反応を含む可能性がある。特定の態様では、多重反応は、下限が１００、２００、２５０、５００、１，０００、２，５００、５，０００、１０，０００、２０，０００、２５，０００、５０，０００個、上限が２００、２５０、５００、１，０００、２，５００、５，０００、１０，０００、２０，０００、２５，０００、５０，０００、および１００，０００個の範囲で行われる。

【０２９１】

例示的態様では、前記多型座位の集合は、高いヘテロ接合性を示すことが知られている座位の集合である。しかしながら、任意の個体のこれら座位のいくつかはホモ接合であることが期待される。特定の例示的態様では、本発明の方法は、ある個体のホモ接合座位とヘテロ接合座位の両方の核酸配列情報を利用する。例えば、ある個体のホモ接合座位が誤り補正に用いられて、ヘテロ接合座位が試料のアレル不均衡率の決定に用いられる。特定の態様では、前記多型座位の少なくとも１０％は、前記個体のヘテロ接合座位である。

【０２９２】

本明細書で開示されるように、集団においてヘテロ接合であることがわかっている対象ＳＮＰ座位の分析が優先される。したがって、特定の態様では、前記多型座位の少なくとも１０、２０、２５、５０、７５、８０、９０、９５、９９、または１００％が集団でヘテロ接合であることがわかっている多型座位が選択される。

【０２９３】
本明細書で開示されるように、特定の態様では、前記試料は妊婦に由来する血漿試料である。
【０２９４】
いくつかの例では、前記方法は、既知の平均アレル不均衡比を有する対照試料に前記方法を行うことをさらに含む。前記対照は、前記染色体分節の異数性を示す特定のアレル状態の平均アレル不均衡比が０．４～１０％として、胎児または腫瘍に由来する循環遊離ＤＮＡに予測されるような低濃度で存在する試料中のアレルの平均アレル不均衡率を再現することができる。
【０２９５】
本明細書で開示されるように、いくつかの態様では、前記対照としてＰｌａｓｍＡｒｔ対照が用いられる。したがって、ある側面では、前記対照は、染色体異数性を示すことがわかっている核酸試料を断片化して前記個体の血漿中を循環するＤＮＡ断片の大きさにすることを含む方法で作成された試料である。ある側面では、染色体分節に異数性がない対照が用いられる。
【０２９６】
例示的態様では、１つ以上の対照に由来するデータは、検査試料と共に前記方法で分析することができる。前記対照の例としては、染色体異数性を含むと疑われていない個体に由来する別の試料、あるいはＣＮＶまたは染色体異数性を含むことが疑われる試料が挙げられる。例えば、検査試料が循環遊離腫瘍ＤＮＡを含むと疑われる血漿試料である場合、この血漿試料と共に前記対象者の腫瘍に由来する対照試料にも前記方法を実行することができる。本明細書で開示されるように、前記対照試料は、染色体異数性を示すことがわかっているＤＮＡ試料を断片化して調製することができる。特に前記試料が癌に苦しむ個体に由来する場合、このような断片化により、アポトーシス細胞のＤＮＡ組成物を再現したＤＮＡ試料を得ることができる。前記対照試料に由来するデータは、染色体異数性の検出の信頼度を高める。
【０２９７】
倍数性を決定する方法の特定の態様では、前記試料は、癌を有すると疑われる個体に由来する血漿試料である。これらの態様では、前記方法は、前記個体の腫瘍細胞にコピー数多様性が存在するかどうかを前記選択に基づいて決定することをさらに含む。これらの態様では、前記試料はある個体に由来する血漿試料であってもよい。これらの態様では、前記方法は、前記個体に癌が存在するかどうかを前記選択に基づいて決定することをさらに含んでいてもよい。
【０２９８】
染色体分節の倍数性を決定するこれらの態様は、一連の一塩基多様性位置に含まれる一塩基多様性位置の一塩基多様体を検出することをさらに含んでいてもよく、ここで、染色体異数性および／または前記一塩基変異の検出は前記試料中に循環腫瘍核酸が存在することを示す。
【０２９９】
これらの態様は、前記個体の腫瘍に関する前記染色体分節のハプロタイプ情報を受け、前記ハプロタイプ情報を用いて、前記多型座位の集合に関する異なる倍数性状態とアレル不均衡率の前記モデルのセットを作成することをさらに含んでいてもよい。
【０３００】
本明細書で開示されるように、倍数性を決定する方法の特定の態様は、初期または修正済みアレル頻度を前記モデルのセットと比較する前に前記初期または修正済みアレル頻度のデータから異常値を除外することをさらに含んでいてもよい。例えば、特定の態様では、前記染色体分節の他の座位の平均値より少なくとも２または３標準偏差大きい、あるいは小さい座位アレル頻度は、モデル化に用いられる前にデータから除外される。
【０３０１】
本明細書の記載から明らかなように、本明細書で提供される態様の多くは、染色体分節

の倍数性を決定する方法を含むが、不十分相決定済みデータまたは完全な相決定済みデータを用いることが好ましい。また、当然だが、本明細書で提供される方法は、倍数性を検出する従来の方法に比べて改良された多くの特徴を有し、これら特徴の多くの様々な組み合わせを用いることができる。

【０３０２】

本明細書の特定の態様では、図６９～図７０に示すように、本発明の任意の方法を実行するコンピュータシステムおよびコンピュータ可読媒体が提供される。これらは、倍数性を決定する方法を実行するためのシステムおよびコンピュータ可読媒体を含む。したがって、システムの非限定的な例である態様では、本明細書で提供される方法のいずれかが本開示内容を用いたシステムとコンピュータ可読媒体を用いて実行できることとが示される。本明細書の他の側面では、個体の試料の染色体倍数性を検出するためのシステムが提供される。前記システムは、以下のものを含む。

ａ．前記試料中の前記染色体分節にある多型座位の集合のうちの各座位に存在する各アレルの量を含むアレル頻度データを受けるよう構成された入力プロセッサと、

ｂ．モデラーであって、

ｉ．前記アレル頻度データの相を推定することによって前記多型座位の集合に関する相決定済みアレル情報を作成し、

ｉｉ．前記アレル頻度データを用いて、異なる倍数性状態に関する前記多型座位のアレル頻度の個別確率を作成し、

ｉｉｉ．前記個別確率と前記相決定済みアレル情報を用いて前記多型座位の集合に関する同時確率を作成するよう構成されたモデラーと、

ｃ．前記同時確率に基づき、染色体倍数性を示す最適合モデルを選択することにより、前記染色体分節の倍数性を決定するよう構成された、仮説管理部とを含む。

【０３０３】

このシステムの態様の特定の態様では、前記アレル頻度データは核酸配列決定システムにより作成されたデータである。特定の態様では、前記システムはさらに、前記アレル頻度データのエラーを修正するように構成されたエラー修正部を含み、修正済みアレル頻度データが前記モデラーによって用いられて個別確率を作成する。特定の態様では、前記エラー修正部はアレル増幅効率の偏りを修正する。特定の態様では、前記モデラーが、前記多型座位の集合に関する異なる倍数性状態とアレル不均衡率の両方のモデルのセットを用いて前記個別確率を作成する。特定の例示的態様では、前記モデラーは、前記染色体分節にある多型座位間の連鎖を考慮することによって前記同時確率を作成する。

【０３０４】

本明細書の例示的一態様では、以下のものを含む、個体の試料の染色体倍数性を検出するシステムが提供される。

ａ．前記個体の染色体分節にある多型座位の集合に存在するアレルに関する核酸配列データを受け、前記核酸配列データを用いて前記多型座位の集合におけるアレル頻度を検出するよう構成された入力プロセッサと、

ｂ．検出された前記アレル頻度のエラーを修正し、前記多型座位の集合に関する修正済みアレル頻度を作成するよう構成されたエラー修正部と、

ｃ．モデラーであって、

ｉ．前記核酸配列データの相を推定することによって前記多型座位の集合に関する相決定済みアレル情報を作成し、

ｉｉ．前記相決定済みアレル情報を前記多型座位の集合に関する異なる倍数性状態とアレル不均衡率のモデルのセットと比較することにより、異なる倍数性状態に関する前記多型座位のアレル頻度の個別確率を作成し、

ｉｉｉ．前記個別確率を前記染色体分節にある多型座位間の相対的距離を考慮しながら組み合わせることにより、前記多型座位の集合に関する同時確率を作成するよう構成されたモデラーと、

ｄ．前記同時確率に基づき、染色体異数性を示す最適合モデルを選択するよう構成された

仮説管理部。

【０３０５】

本明細書で提供される特定の例示的なシステムの態様では、前記多型座位の集合は、１０００個から５０，０００個の多型座位を含む。本明細書で提供される特定の例示的なシステムの態様では、前記多型座位の集合は、ヘテロ接合が起こりやすい箇所座位として知られている１００座位を含む。本明細書で提供される特定の例示的なシステムの態様では、前記多型座位の集合は、組換え起こりやすい箇所の０．５ｋｂ以内の１００座位を含む。

【０３０６】

本明細書で提供される特定の例示的なシステムの態様では、前記最適合モデルは、前記染色体分節の第一の相同体と前記染色体分節の第二の相同体の以下の倍数性状態を分析する。

【０３０７】

（１）いずれの細胞も前記染色体分節の前記第一の相同体または前記第二の相同体の欠失または増幅を有していない状態、

【０３０８】

（２）一部またはすべての細胞が前記染色体分節の前記第一の相同体の欠失、または前記第二の相同体の増幅を有する状態、および

【０３０９】

（３）一部またはすべての細胞が前記染色体分節の前記第二の相同体の欠失、または前記第一の相同体の増幅を有する状態。

【０３１０】

本明細書で提供される特定の例示的なシステムの態様では、前記修正されたエラーは、アレル増幅効率の偏り、不純物、および／または配列決定エラーを含む。本明細書で提供される特定の例示的なシステムの態様では、前記不純物は、環境不純物、および遺伝型解析による不純物を含む。本明細書で提供される特定の例示的なシステムの態様では、前記環境不純物および遺伝型解析による不純物はホモ接合アレルについて決定される。

【０３１１】

本明細書で提供される特定の例示的なシステムの態様では、前記仮説管理部は、前記モデルに関して作成された前記相決定済みアレル情報と推定アレル頻度との差異の大きさを分析するように構成される。本明細書で提供される特定の例示的なシステムの態様では、前記モデラーは、前記多型座位の集合におけるアレル頻度の期待値と観測値のベータ二項モデルに基づいてアレル頻度の個別確率を作成する。本明細書で提供される特定の例示的なシステムの態様では、前記モデラーは、ベイズ分類器を用いて個別確率を作成する。

【０３１２】

本明細書で提供される特定の例示的なシステムの態様では、前記核酸配列データは、多重増幅反応によって生成された一連の増幅産物の複数のコピーの高処理ＤＮＡ配列決定を実行することで作成され、ここで、前記一連の増幅産物のうちの各増幅産物は、前記多型座位の集合のうちの少なくとも１つの多型座位にわたり、前記多型座位の集合のうちの各多型座位が増幅される。本明細書で提供される特定の例示的なシステムの態様では、前記多重増幅反応は、これら反応の少なくとも１／２で制限的なプライマー条件で実行される。本明細書で提供される特定の例示的なシステムの態様では、前記試料は平均アレル不均衡率が０．４％～５％である。

【０３１３】

本明細書で提供される特定の例示的なシステムの態様では、前記試料は癌を有すると疑われる個体に由来する血漿試料であり、前記仮説管理部はさらに、前記個体の腫瘍細胞にコピー数多様性が存在するかどうかを前記最適合モデルに基づいて決定するように構成される。

【０３１４】

本明細書で提供される特定の例示的なシステムの態様では、前記試料はある個体に由来

する血漿試料であり、前記仮説管理部はさらに、前記個体に癌があるかどうかを前記最適合モデルに基づいて決定するように構成される。これらの態様では、前記仮説管理部はさらに、一塩基多様性位置の集合のうちのある一塩基多様性位置に存在する一塩基多様体を検出するように構成されてもよく、染色体異数性、一塩基多様体の一方または両方の検出は前記試料中に循環腫瘍核酸が存在することを示す。

【０３１５】

本明細書で提供される特定の例示的なシステムの態様では、前記入力プロセッサはさらに、前記個体の腫瘍に関する前記染色体分節のハプロタイプ情報を受けるように構成され、前記モデラーは前記ハプロタイプ情報を用いて前記多型座位の集合の異なる倍数性状態とアレル不均衡率のモデルのセットを作成するように構成される。

【０３１６】

本明細書で提供される特定の例示的なシステムの態様では、前記モデラーは０％～２５％のアレル不均衡率にわたってモデルを作成する。

【０３１７】

本明細書で提供される方法のいずれかが、非一過性のコンピュータ可読媒体に記憶されたコンピュータ可読コードにより実行することができることは明らかである。したがって、本明細書で提供される一態様は、個体の試料中の染色体倍数性を検出する非一過性のコンピュータ可読媒体であって、処理装置によって実行されると、前記処理装置は以下のことを実行する。

ａ．前記染色体分節にある多型座位の集合のうちの各座位で前記試料中に存在する各アレルの量を含むアレル頻度データを受け、

ｂ．前記アレル頻度データの相を推定することによって前記多型座位の集合に関する相決定済み立遺伝子情報を作成し、

ｃ．前記アレル頻度データを用いて、異なる倍数性状態に関する前記多型座位のアレル頻度の個別確率を作成し、

ｄ．前記個別確率と前記相決定済みアレル情報を用いて前記多型座位の集合に関する同時確率を作成し、

ｅ．前記同時確率に基づき、染色体倍数性を示す最適合モデルを選択することにより、前記染色体分節の倍数性を決定する。

【０３１８】

コンピュータ可読媒体の特定の態様では、前記アレル頻度データは核酸配列データから作成される。コンピュータ可読媒体の特定の態様はさらに、アレル頻度データのエラーを修正し、個別確率の作成工程ごとに修正済みアレル頻度データを用いることを含む。コンピュータ可読媒体の特定の態様では、前記修正されたエラーはアレル増幅効率の偏りである。コンピュータ可読媒体の特定の態様では、前記個別確率は、前記多型座位の集合に関する異なる倍数性状態とアレル不均衡率の両方のモデルのセットを用いて作成される。コンピュータ可読媒体の特定の態様では、前記前記同時確率は、前記染色体分節にある多型座位間の連鎖を考慮することによって作成される。

【０３１９】

本明細書で提供される特定の一態様は、個体の試料中の染色体倍数性を検出する非一過性のコンピュータ可読媒体であって、処理装置により実行されると、前記処理装置は以下のことを実行する。

ａ．前記個体の染色体分節にある多型座位の集合に存在するアレルに関する核酸配列データを受け、

ｂ．前記核酸配列データを用いて前記多型座位の集合におけるアレル頻度を検出し、

ｃ．検出された前記アレル頻度のアレル増幅効率の偏りを修正することによって、前記多型座位の集合に関する修正済みアレル頻度を作成し、

ｄ．前記核酸配列データの相を推定することによって前記多型座位の集合に関する相決定済みアレル情報を作成し、

ｅ．前記修正済みアレル頻度を前記多型座位の集合に関する異なる倍数性状態とアレル不

均衡率のモデルのセットと比較することにより、異なる倍数性状態に関する前記多型座位のアレル頻度の個別確率を作成し、
ｆ．前記個別確率を前記染色体分節にある多型座位間の連鎖を考慮しながら組み合わせることにより、前記多型座位の集合に関する同時確率を作成し、
ｇ．前記同時確率に基づき、染色体異数性を示す最適合モデルを選択する。
【０３２０】
コンピュータ可読媒体の特定の例示的態様では、前記選択は、前記モデルに関して作成された前記相決定済みアレル情報と推定アレル頻度との差異の大きさを分析することで実行される。
【０３２１】
コンピュータ可読媒体の特定の例示的態様では、前記アレル頻度の個別確率は、前記多型座位の集合におけるアレル頻度の期待値と観測値のベータ二項モデルに基づいて決定される。
【０３２２】
なお、本明細書で提供される態様の方法のいずれかは、非一過性のコンピュータ可読媒体に記憶されたコードを実行することで実行することができる。
癌を検出する実施形態の例
【０３２３】
特定の側面では、本発明は癌を検出する方法を提供する。なお、前記試料は、癌を有すると疑われる個体に由来する腫瘍試料または液体試料例えば、血漿等）であってもよい。前記方法は、試料中の総ＤＮＡの分画として少ない量のこれら遺伝子変化を有する試料中の、一塩基変化（例えばＳＮＶ等）やコピー数変化（例えば、ＣＮＶ等）等の遺伝子変異を検出するのに特に有効である。したがって、試料中の癌に由来するＤＮＡまたはＲＮＡを検出する感度が格別に高い。前記方法は、本明細書で提供されるＣＮＶおよびＳＮＶを検出するための改良のいずれかまたはすべてを組み合わせて、この格別に高い感度を達成することができる。
【０３２４】
したがって、本明細書の特定の態様では、循環腫瘍核酸が個体の試料中に存在するかを決定するための方法、および処理装置によって実行されると前記処理装置が前記方法を実行するコンピュータ可読コードを含む非一過性のコンピュータ可読媒体が提供される。前記方法は、以下の工程を含む。
ｃ．前記試料を分析して前記個体の染色体分節にある多型座位の集合の倍数性を決定し、
ｄ．前記倍数性の決定に基づいて前記多型座位に存在する平均アレル不均衡率を決定し、０．４％、０．４５％、０．５％、０．６％、０．７％、０．７５％、０．８％、０．９％、または１％以上の平均アレル不均衡率は、前記試料中の循環腫瘍核酸（例えば、ｃｔＤＮＡ等）の存在を示す。
【０３２５】
ある例では、０．４、０．４５、または０．５％超の平均アレル不均衡率は、ｃｔＤＮＡの存在を示す。特定の態様では、前記循環腫瘍核酸が存在するかを決定するための方法は、一塩基多様性位置の集合に含まれるある一塩基多様性位置の一塩基多様体を検出することをさらに含み、０．５％以上のアレル不均衡率を検出すること、および前記一塩基多様体を検出することの一方または両方は前記試料に循環腫瘍核酸が存在することを示す。当然ではあるが、染色体倍数性またはＣＮＶを検出するために提供される前記方法のいずれかを用いて、アレル不均衡率、通常は平均アレル不均衡率を決定することができる。なお、ＳＮＶを検出するために本明細書で提供される前記方法のいずれかを用いて、本発明のこの側面に係る単一のヌクレオチドを検出することができる。
【０３２６】
特定の態様では、前記循環腫瘍核酸が存在するかを決定するための方法はさらに、既知の平均アレル不均衡比を有する対照試料に前記方法を実行することを含む。例えば、前記対照は、前記個体の腫瘍に由来する試料であってもよい。いくつかの態様では、前記対照

は分析中の試料に期待される平均アレル不均衡率を有する。例えば、ＡＡＩが０．５％～５％、または平均アレル不均衡比が０．５％である。

【０３２７】

特定の態様では、前記循環腫瘍核酸が存在するかを決定する方法の分析工程は、癌で異数性を示すことが知られている染色体分節の集合を分析することを含む。特定の態様では、前記循環腫瘍核酸が存在するかを決定する方法の分析工程は、倍数性について１，０００～５０，０００個、または１００～１，０００個の多型座位を分析することを含む。特定の態様では、前記循環腫瘍核酸が存在するかを決定するための方法の分析工程は、１００～１，０００個の一塩基多様性位置を分析することを含む。例えば、これらの態様では、前記分析工程は、多重ＰＣＲを行って、１，０００～５０，０００個の同義座位および１００～１，０００個の一塩基多様性位置にわたる増幅産物を増幅することを含んでいてもよい。この多重反応は、単一の反応、あるいは様々な多重下位反応のプールとして設定することができる。本明細書で提供される多重反応法（例えば、本明細書に開示される大多重ＰＣＲ等）は、増幅反応を実行して多重化、したがって感度の向上を補助する例示的プロセスを提供する。

【０３２８】

特定の態様では、前記多重ＰＣＲ反応の少なくとも１０％、２０％、２５％、５０％、７５％、９０％、９５％、９８％、９９％、または１００％で、前記反応は制限的なプライマー条件で実行される。本明細書で提供される大多重反応を実行するための改良された条件を用いることができる。

【０３２９】

特定の側面では、前記循環腫瘍核酸が個体の試料中に存在するかを決定する方法、およびそのすべての態様は、システムを用いて実行することができる。本開示は、前記方法を実行するための特定の機能および構造的特徴に関する教示を提供する。非限定的な例として、前記システムは以下を含む。

ａ．前記試料に由来するデータを分析して、前記個体中の染色体分節にある多型座位の集合の倍数性を決定する入力プロセッサと、

ｂ．前記倍数性の決定に基づいて前記多型座位に存在するアレル不均衡率を決定するモデラーであって、０．５％以上のアレル不均衡率は循環の存在を示す。

一塩基多様体を検出する実施形態の例

【０３３０】

本明細書の特定の側面では、試料中の一塩基多様体を検出する方法が提供される。本明細書で提供される改良された方法は、試料中のＳＮＶを検出限界０．０１５、０．０１７、０．０２、０．０５、０．１、０．２、０．３、０．４、または０．５パーセントで検出することができる。ＳＮＶを検出するすべての態様は、システムを用いて実行することができる。本開示は、前記方法を実行するための特定の機能および構造的特徴に関する教示を提供する。さらに、処理装置により実行されると前記処理装置が本明細書で提供されるＳＮＶを検出する方法を実行するコンピュータ可読コードを含む非一過性のコンピュータ可読媒体を含む態様が本明細書で提供される。

【０３３１】

したがって、本明細書で提供される一態様は、個体の試料に含まれるゲノム位置の集合に一塩基多様体が存在するかを決定する方法であって、前記方法は、

ａ．各ゲノム位置に関して、訓練データセットを用いてそのゲノム位置にわたる増幅産物に関する効率およびサイクル当たりのエラー率の推定値を作成し、

ｂ．前記試料に含まれる各ゲノム位置に関する観測ヌクレオチド同一性情報を受け、

ｃ．各ゲノム位置に関する増幅効率およびサイクル当たりのエラー率の前記推定値を個別に用いて、各ゲノム位置に関する前記観測ヌクレオチド同一性情報を異なる変異率のモデルと比較することにより、各ゲノム位置における１つ以上の実際の変異によって生じる一塩基多様体率の確率のセットを決定し、

ｄ．各ゲノム位置に関する前記確率のセットから、最も可能性の高い実変異率と信頼度を

決定することとを含む。

【０３３２】

前記一塩基多様体が存在するかを決定する方法の例示的態様では、効率およびサイクル当たりのエラー率の前記推定値は、前記ゲノム位置にわたる一連の増幅産物に関して作成される。例えば、前記ゲノム位置にわたる２、３、４、５、１０、１５、２０、２５、５０、１００個、またはそれ以上の増幅産物が含まれていてもよい。

【０３３３】

前記一塩基多様体が存在するかを決定する方法の例示的態様では、前記観測ヌクレオチド同一性情報は、各ゲノム位置に関する全リード数の観測値と各ゲノム位置に関する多様性アレルのリード数の観測値を含む。

【０３３４】

前記一塩基多様体が存在するかを決定する方法の例示的態様では、前記試料は血漿試料であり前記一塩基多様体が前記試料の循環腫瘍ＤＮＡに存在する。

【０３３５】

本明細書の他の態様では、ある個体に由来する試料中に存在する一塩基多様体のパーセントを推定する方法が提供される。前記方法は、以下の工程を含む。

ａ．ゲノム位置の集合において、訓練データセットを用いてそれらのゲノム位置にわたる１個以上の増幅産物に関する効率およびサイクル当たりのエラー率の推定値を作成し、

ｂ．前記試料に含まれる各ゲノム位置に関する観測ヌクレオチド同一性情報を受け、

ｃ．前記増幅産物の増幅効率とサイクル当たりのエラー率の前記推定値を用いて、ある初期の実変異分子率を有する調査領域の全分子数と、背景エラー分子と、実変異分子とに関する平均と分散の推定値を作成し、

ｄ．平均および分散の前記推定値を用いた分布を前記試料の観測ヌクレオチド同一性情報に適合させることによって最も可能性の高い実際の一塩基多様体率を決定することにより、実際の変異によって生じる前記試料の一塩基多様体率を決定すること。

【０３３６】

ある個体に由来する試料中に存在する一塩基多様体のパーセントを推定する方法の一例では、前記試料は血漿試料であり、前記一塩基多様体が前記試料の循環腫瘍ＤＮＡに存在する。

【０３３７】

本発明のこの態様の訓練データセットは、典型的には健康な個体、好ましくは健康な個体群に由来する試料を含む。ある例示的態様では、前記訓練データセットは、１つ以上の検査試料の分析と同じ日に、または同じ一操作で分析される。例えば、２、３、４、５、１０、１５、２０、２５、３０、３６、４８、９６、１００、１９２、２００、２５０、５００、１０００人、またはそれ以上の健康な個体に由来する試料を用いて訓練データセットを作成することができる。多数（例えば９６人以上）の健康な個体がデータに利用可能である場合、検査試料に前記方法を実行する前に検査を行った場合でも増幅効率推定値の信頼度が高まる。ＰＣＲエラー率は増幅産物ごとにあるので、ＳＮＶの塩基位置について作成された核酸配列情報だけでなく、ＳＮＶの周辺の増幅領域全体について作成された核酸配列データを用いることができる。例えば、５０個体に由来する試料と前記ＳＮＶ周辺の２０塩基対の増幅産物の配列決定を用いて、１０００塩基のリードに由来するエラー頻度データを用いてエラー頻度率を決定することができる。

【０３３８】

通常、増幅効率は、増幅された分節の増幅効率について平均値と標準偏差を推定して、それを分布モデル（例えば、二項分布またはベータ二項分布等）に適合させて推定される。既知のサイクル数のＰＣＲ反応についてエラー率が決定されて、サイクル当たりのエラー率が推定される。

【０３３９】

特定の例示的態様では、検査データ集合の開始分子を推定することはさらに、リード観察数がリード推定数よりも大きく異なる場合は工程（ｂ）で推定された分子の開始数を用

いて前記検査データ集合の効率推定値を更新することを含む。前記推定値は新しい効率および／または開始分子ごとに更新されてもよい。

【０３４０】

分子、背景エラー分子、および実際の変異分子の総数を推定するのに用いられる検索空間は、あるＳＮＶ位置の塩基のコピーの０．１％、０．２％、０．２５％、０．５％、１％、２．５％、５％、１０％、１５％、２０％、または２５％（下限）から１％、２％、２．５％、５％、１０％、１２．５％、１５％、２０％、２５％、５０％、７５％、９０％、または９５％（上限）がＳＮＶ塩基である検索空間を含んでいてもよい。前記方法が循環腫瘍ＤＮＡを検出するある例では、より低い範囲として、下限が０．１％、０．２％、０．２５％、０．５％、また１％、上限が１％、２％、２．５％、５％、１０％、１２．５％、または１５％の範囲を血漿試料に用いることができる。腫瘍試料にはより高い範囲が用いられる。

【０３４１】

分布では、総分子中の誤り分子（背景エラーおよび実際の変異）の総数を適合させて、検索空間における起こりうる実際の変異のそれぞれの尤度すなわち確率が算出される。この分布は二項分布またはベータ二項分布であってもよい。

【０３４２】

最も可能性の高い実際の変異は、最も可能性の高い実際の変異の百分率を決定し、前記分布の適合に由来するデータを用いて信頼度を算出して決定される。本明細書で提供される前記方法の臨床解釈の限定を意図しない具体例として、中央変異率が高いと、ＳＮＶを陽性と決定するのに必要な信頼度（パーセント）は低くなる。例えば、最も可能性の高い仮説を用いたある試料中のＳＮＶの中央変異率が５％であり、信頼度（パーセント）が９９％である場合、陽性のＳＮＶ呼び出しが実行されることがある。一方、この例で、最も可能性の高い仮説を用いたある試料中のＳＮＶの中央変異率が１％であり、信頼度（パーセント）が５０％である場合、特定の状況では、陽性のＳＮＶ呼び出しは実行されないことがある。当然ではあるが、データの臨床解釈は、感度、特異性、有病率、および代替え品の利用可能性に応じたものである。

【０３４３】

例示的一態様では、前記試料は、循環ＤＮＡ試料（例えば、循環腫瘍ＤＮＡ試料等）である。

【０３４４】

本明細書の他の態様では、個体の試験試料中に存在する１つ以上の一塩基多様体を検出する方法が提供される。この態様による方法は以下の工程を含む。

ｄ．配列決定の実行時に作成された結果に基づき、複数の正常な個体の各々から得た複数の対照試料の多様性アレル頻度の中央値を、一塩基多様性位置の集合に含まれる各一塩基多様性位置について決定することにより、閾値未満である正常試料の多様性アレル頻度の中央値を有する、選定された一塩基多様性位置を確認し、前記一塩基多様性位置の各々について異常値試料を除外した後、前記一塩基多様性位置の各々に関する背景エラーを決定し、

ｅ．該試験試料に関する配列決定の実行時に作成されたデータに基づき、該試験試料について、選定された該一塩基多様性位置に関して観測リード深度による加重平均と分散を決定し、

ｆ．コンピュータを用いて、リード深度による加重平均が背景エラーに比べて統計的に有意である１つ以上の一塩基多様性位置を確認することによって、前記１つ以上の一塩基多様体を検出すること。

【０３４５】

特定の態様では、前記１つ以上のＳＮＶを検出する方法において、前記試料は血漿試料であり、前記対照試料は血漿試料であり、検出された前記１つ以上の一塩基多様体は前記試料の循環腫瘍ＤＮＡに存在する。特定の態様では、前記１つ以上のＳＮＶを検出する方法において、前記複数の対照試料は少なくとも２５個の試料を含む。ある例示的態様では

、前記複数の対照試料は、下限が少なくとも５、１０、１５、２０、２５、５０、７５、１００、２００、または２５０個の試料であり、上限が１０、１５、２０、２５、５０、７５、１００、２００、２５０、５００、および１０００個の試料である。

【０３４６】

特定の態様では、前記１つ以上のＳＮＶを検出する方法において、高処理配列決定の実行時に作成されたデータから異常値を除外することによって前記観測リード深度による加重平均と観測分散が算出される。特定の態様では、前記１つ以上のＳＮＶを検出する方法において、前記試験試料の各一塩基多様性位置に関する前記リード深度は少なくとも１００リードである。

【０３４７】

特定の態様では、前記１つ以上のＳＮＶを検出する方法において、前記配列決定の実行が、制限的なプライマー反応条件で実行される多重増幅反応を含む。本明細書で提供される多重増幅反応を実行する改良された方法を用いて、例示のこれら態様が実行される。

【０３４８】

理論により限定するものではないが、本実施形態の方法は、実行に特定の人工産物を説明するため、検査試料と同じ配列決定の一操作で配列決定された正常な血漿試料を用いた背景エラーモデルを利用している。閾値より大きい（例えば、０．１％、０．２％、０．２５％、０．５％、０．７５％、および１．０％より大きい）正常な中心多様性アレル頻度を有するノイズと思われる位置は除外される。

【０３４９】

ノイズおよび不純物を説明するため、異常値を有する試料は前記モデルから繰り返し除外される。各遺伝子座位の各塩基置換について、前記エラーのリード深度による加重平均と標準偏差が算出される。特定の例示的態様では、少なくとも閾値数のリード（例えば、少なくとも２、３、４、５、６、７、８、９、１０、１５、２０、２５、５０、１００、２５０、５００、または１０００個の変異体リード）を有し、特定の態様の背景エラーモデルに対するＺ値が２．５、５、７．５、または１０より大きい一塩基多様性位置を有する試料（例えば、腫瘍または無細胞血漿試料等）は、変異の候補として計測される。

【０３５０】

特定の態様では、前記一塩基多様性位置の各一塩基多様性位置の配列決定の実行において、下限が１００、２５０、５００、１，０００、２０００、２５００、５０００、１０，０００、２０，０００、２５，０００、５０，０００、または１００，０００リード超、上限が２０００、２５００、５，０００、７，５００、１０，０００、２５，０００、５０，０００、１００，０００、２５０，０００、または５００，０００リード超のリード深度が達成される。前記配列決定の一操作は、通常、高処理配列決定の一操作である。例示的態様で、前記検査試料について生成された平均値または中央値は、リード深度により加重平均される。したがって、ある多様性アレルの決定が１０００リード中１個の多様性アレルが検出される試料中に実際に起きる尤度は、１０，０００リード中１個の多様性アレルが検出される試料よりも高く加重平均される。ある多様性アレル（すなわち変異）の決定は１００％の信頼度で行えないので、同定された一塩基多様体は候補の多様体または候補の変異と見なすことができる。

【０３５１】

相決定済みデータの分析の検定統計量の例

以下に、遺伝的に同一ではない２個以上の細胞に由来するＤＮＡまたはＲＮＡを含む混合試料であると分かっているか、疑わしい試料に由来する相決定済みデータの分析について、検定統計量の例が記載される。目的ＤＮＡまたはＲＮＡの分画（例えば、目的のＣＮＶを有するＤＮＡまたはＲＮＡ分画、あるいは癌細胞等の目的の細胞に由来するＤＮＡまたはＲＮＡの分画）をｆとする。いくつかの態様では、出生前検査の場合、ｆは、胎児と母親のＤＮＡ、ＲＮＡ、または細胞の混合物中の胎児のＤＮＡ、ＲＮＡ、または細胞の分画を指す。ＤＮＡの２つのコピーが目的細胞のそれぞれによって生じたと仮定すると、ｆは目的細胞に由来するＤＮＡの分画を指す。ｆは、欠失または重複した分節に存在する目

的細胞に由来するＤＮＡ分画とは異なる。

【０３５２】

各ＳＮＰの起こりうるアレル値をＡおよびＢとする。ＡＡ、ＡＢ、ＢＡ、およびＢＢを用いて、起こりうるすべての順序付けられたアレル対を示す。いくつかの態様では、順序付けられたアレルＡＢまたはＢＡを有するＳＮＰが分析される。N_iをｉ番目のＳＮＰの配列リード数、A_iおよびB_iをそれぞれアレルＡおよびＢを示すｉ番目のＳＮＰのリード数とする。以下のように仮定する。

$N_i = A_i + B_i$

アレル比率R_iは以下のように定義される。

【数２】

$$R_i \triangleq \frac{A_i}{N_i}$$

Ｔを対象のＳＮＰ数とする。

【０３５３】

一般性を失うことなく、いくつかの態様は、単一の染色体分節に注目する。さらなる透明性のために、本明細書では、用語「第二の相同染色体分節と比較される第一の相同染色体分節」は、ある染色体分節の第一の相同体およびその染色体分節の第二の相同体を意味する。このようないくつかの態様では、対象ＳＮＰのすべてが目的の分節染色体に含まれる。他の態様では、起こりうるコピー数多様性について複数の染色体分節が分析される。

【０３５４】

マップ推定

この方法は、順序づけられたアレルを介して相決定の知識を利用して、対象分節の欠失または重複を検出する。各ＳＮＰｉについて、以下の通り定義する。

【数３】

$$X_i \triangleq \begin{cases} 1 & R_i < 0.5 \text{ かつ SNPiAB} \\ 0 & R_i \geqq 0.5 \text{ かつ SNPiAB} \\ 0 & R_i < 0.5 \text{ かつ SNPiBA} \\ 1 & R_i \geqq 0.5 \text{ かつ SNPiBA} \end{cases}$$

【０３５５】

よって、以下のように定義される。

【数４】

$$S \triangleq \sum_{\text{すべての SNP}} X_i.$$

様々なコピー数仮説（例えば、二染色体性の仮説、第一または第二の相同体の欠失の仮説、あるいは第一または第二の相同体の重複の仮説等）におけるX_iおよびＳの分布を以下で記載する。

【０３５６】

二染色体性仮説
対象分節は欠失も重複もしていないというこの仮説では、
【数５】

$$X_i = \begin{cases} 0 & wp\ 1-p\left(\frac{1}{2}, N_i\right) \\ 1 & wp\ p\left(\frac{1}{2}, N_i\right) \end{cases}$$

ここで
【数６】

$$p(b,n) \triangleq Pr\left\{X \sim Bino(b,n) \geq \frac{n}{2}\right\}.$$

リード深度定数Ｎを仮定すると、パラメータを有する二項分布Ｓが得られる。
【数７】

$$\boldsymbol{p\left(\frac{1}{2}, N\right)} \quad \text{とＴ}$$

【０３５７】
欠失仮説
第一の相同体が欠失している（すなわち、ＡＢのＳＮＰがＢになり、ＢＡのＳＮＰがＡになっている）という仮説では、R_iは、ＡＢのＳＮＰについてパラメータ｛１－（１／（２－ｆ））｝およびＴ、ＢＡのＳＮＰについて（１／（２－ｆ））およびＴを有する二項分布を有する。したがって、
【数８】

$$X_i = \begin{cases} 0 & wp\ 1-p\left(\frac{1}{2-f}, N_i\right) \\ 1 & wp\ p\left(\frac{1}{2-f}, N_i\right) \end{cases}$$

リード深度定数Ｎを仮定すると、以下のパラメータを有する二項分布Ｓが得られる。
【数９】

$$\boldsymbol{p\left(\frac{1}{2-f}, N\right)} \quad \text{とＴ}$$

【０３５８】
第二の相同体が欠失している（すなわち、ＡＢのＳＮＰがＡになり、ＢＡのＳＮＰがＢになっている）という仮説では、R_iは、ＡＢのＳＮＰについてパラメータ（１／（２－ｆ））およびＴ、ＢＡのＳＮＰについて｛１－（１／（２－ｆ））｝およびＴを有する二

項分布を有する。したがって、
【数１０】

$$X_i = \begin{cases} 0 & wp\ p\left(\frac{1}{2-f}, N_i\right) \\ 1 & wp\ 1-p\left(\frac{1}{2-f}, N_i\right) \end{cases}$$

リード深度定数Ｎを仮定すると、以下のパラメータを有する二項分布Ｓが得られる。
【数１１】

$$\mathbf{1-p\left(\frac{1}{2-f}, N\right)} \quad \text{とＴ}$$

【０３５９】
重複仮説
第一の相同体が重複している（すなわち、ＡＢのＳＮＰがＡＡＢとなり、ＢＡのＳＮＰがＢＢＡとなっている）という仮説では、Ｒ$_i$は、ＡＢのＳＮＰについてパラメータ（１＋ｆ）／（２＋ｆ）およびＴ、ＢＡのＳＮＰについて｛１－（（１＋ｆ）／（２＋ｆ））｝およびＴを有する二項分布を有する。したがって、
【数１２】

$$X_i = \begin{cases} 0 & wp\ p\left(\frac{1+f}{2+f}, N_i\right) \\ 1 & wp\ 1-p\left(\frac{1+f}{2+f}, N_i\right) \end{cases}$$

リード深度定数Ｎを仮定すると、以下のパラメータを有する二項分布Ｓが得られる。
【数１３】

$$\mathbf{1-p\left(\frac{1+f}{2+f}, N\right)} \quad \text{とＴ}$$

【０３６０】
第二の相同体が重複している（すなわち、ＡＢのＳＮＰがＡＢＢとなり、ＢＡのＳＮＰがＢＡＡとなっている）という仮説では、Ｒ$_i$は、ＡＢのＳＮＰについてパラメータ｛１－（（１＋ｆ）／（２＋ｆ））｝およびＴ、ＢＡのＳＮＰについて（１＋ｆ）／（２＋ｆ）およびＴを有する二項分布を有する。したがって、

【数１４】

$$X_i = \begin{cases} 0 & wp\ 1-p\left(\frac{1+f}{2+f}, N_i\right) \\ 1 & wp\ p\left(\frac{1+f}{2+f}, N_i\right) \end{cases}$$

リード深度定数Ｎを仮定すると、以下のパラメータを有する二項分布Ｓが得られる。

【数１５】

$$p\left(\frac{1+f}{2+f}, N\right) \quad \text{とＴ}$$

【０３６１】

分類

上の項で示したように、Ｘ$_{i}$は、以下の二項ランダム変数である。

【数１６】

$$Pr\{X_1 = 1\} = \begin{cases} p\left(\frac{1}{2}, N_i\right) & \text{二染色体と仮定} \\ p\left(\frac{1}{2-f}, N_i\right) & \text{相同体１の欠失} \\ 1-p\left(\frac{1}{2-f}, N_i\right) & \text{相同体２の欠失} \\ 1-p\left(\frac{1+f}{2+f}, N_i\right) & \text{相同体１の重複} \\ p\left(\frac{1+f}{2+f}, N_i\right) & \text{相同体２の重複} \end{cases}$$

【０３６２】

これにより、各仮説での検定統計量Ｓの確率を算出することができる。測定されたデータを仮定して各仮説の確率を算出することができる。いくつかの態様では、最も高い確率を有する仮説が選択される。必要に応じて、リード深度定数Ｎで各Ｎ$_{i}$を近似する、またはリード深度を定数Ｎまで切り捨てることで、Ｓの分布を簡略化することができる。この簡略化によって以下が得られる。

【数１７】

$$S\sim\begin{cases} Bino\left(p\left(\frac{1}{2},N\right),T\right) & \text{二染色体と仮定} \\ Bino\left(p\left(\frac{1}{2-f},N\right),T\right) & \text{相同体１の欠失} \\ Bino\left(1-p\left(\frac{1}{2-f},N\right),T\right) & \text{相同体２の欠失} \\ Bino\left(1-p\left(\frac{1+f}{2+f},N\right),T\right) & \text{相同体１の重複} \\ Bino\left(p\left(\frac{1+f}{2+f},N\right),T\right) & \text{相同体２の重複} \end{cases}$$

【０３６３】

測定されたデータを仮定して、例えば、最尤推定、最大事後確率推定、またはベイズ推定等のアルゴリズム（例えば、検索アルゴリズム）を用いてデータの最適合を生成する値のｆ等、最も可能性の高い値のｆを選択することで、ｆの値を推定することができる。いくつかの態様では、複数の染色体分節が分析されて、各分節のデータに基づいてｆの値が推定される。対象細胞のすべてがこれら重複または欠失を有する場合、これらの異なる分節のデータに基づくｆの推定値は類似している。いくつかの態様では、癌と非癌性ＤＮＡまたはＲＮＡのメチル化の違い（低メチル化または高メチル化）に基づいて癌細胞由来のＤＮＡまたはＲＮＡ分画を決定することで、ｆは実験的に測定される。

【０３６４】

いくつかの態様では、胎児の核酸と母親の核酸の混合試料の場合、ｆの値は胎児の分画、すなわち、前記試料中のＤＮＡ（またはＲＮＡ）の総量に対する胎児のＤＮＡ（またはＲＮＡ）の分画である。いくつかの態様では、母親および胎児の両方で二染色体性が予測される少なくとも１本の染色体にある多型座位の集合について母親の血液試料（またはその分画）由来の遺伝型データを得ることと、前記染色体について異なる起こりうる胎児の分画にそれぞれ対応する複数の仮説を作成し、起こりうる胎児の分画について前記染色体にある多型座位の集合で前記血液試料中の期待アレル測定に関するモデルを構築し、前記モデルと前記血液試料またはその分画に由来するアレル測定を用いて胎児の分画に関する各仮説の相対確率を算出し、最も高い確率を有する仮説に対応する胎児の分画を選択することで前記血液試料中の胎児の分画を決定することにより、前記胎児の分画は決定される。いくつかの態様では、母親のある多型座位の第一のアレルがホモ接合であり、父親のその多型座位の（ｉ）第一のアレルと第二のアレルがヘテロ接合である、または（ｉｉ）第二のアレルがホモ接合である多型座位を同定し、同定された多型座位のそれぞれについて前記血液試料中に検出された第二のアレルの量を用いて前記血液試料中の胎児の分画を決定することにより、前記胎児の分画は決定される（例えば、２０１２年３月２９日に提供された米国特許公開第２０１２／０１８５１７６号および２０１３年３月１３日に提出された米国特許公開第２０１４／００６５６２１号を参照のこと。これらはそれぞれ、参照によりその全内容が本明細書に組み込まれる）。

【０３６５】

胎児の分画を決定する他の方法は、高処理ＤＮＡ配列決定装置を用いて多数の多型（例えば、ＳＮＰ）遺伝子座位に存在するアレルを計測して、可能性のある胎児の分画をモデル化することを含む（例えば、米国特許公開第２０１２／０２６４１２１号を参照のこと。その全内容が参照により本明細書に組み込まれる）。胎児の分画を算出する他の方法は、Sparks et al.," Noninvasive prenatal detection and selective analysis of cell-free DNA obtained from maternal blood: evaluation for trisomy 21 and trisomy 18

（母親の血液から得た無細胞ＤＮＡの非侵襲性出生前検査および選択的分析：三染色体２１および１８の評価）" Am J Obstet Gynecol 2012;206:319.e1-9にも見られ、その全内容が参照により本明細書に組み込まれる。いくつかの態様では、胎児の分画は、胎児において特定の座位がメチル化される、あるいは優先的にメチル化され、これらの同じ座位は母親では非メチル化される、優先的に非メチル化されると仮定するメチル化アッセイを用いて決定される（例えば、米国特許第７，７５４，４２８号、第７，９０１，８８４号、および第８，１６６，３８２号を参照のこと。その全内容が参照により本明細書に組み込まれる）。

【０３６６】

図１Ａ～１３Ｄは、ＳＮＰ数の増加に対し、各種リード深度および腫瘍分画の場合の異なるコピー数仮説に関する検定統計量ＳをＴ（ＳＮＰ数）で除した（「Ｓ／Ｔ」）分布）（ここでｆは総ＤＮＡ中の腫瘍ＤＮＡ分画である）を示すグラフである。

【０３６７】

単一仮説の棄却

前記二染色体性仮説のＳの分布はｆに依存しない。よって、測定されたデータの確率は、ｆを算出せずに二染色体性仮説について算出することができる。二染色体性の帰無仮説に単一仮説棄却試験を用いることができる。いくつかの態様では、二染色体性仮説でのＳの確率が算出されて、その確率が所定の閾値未満（例えば、１，０００分の１未満）であると、この二染色体性仮説は棄却される。これは、前記染色体分節の重複または欠失が存在することを示す。必要に応じて、閾値を調整して偽陽性率を変更することができる。

【０３６８】

相決定済みデータ分析法の例

以下に、遺伝的に同一ではない２個以上の細胞に由来するＤＮＡまたはＲＮＡを含む混合試料であるとわかっているか疑わしい試料に由来するデータを分析する例示的方法が記載される。いくつかの態様では、相決定済みデータが用いられる。いくつかの態様では、前記方法は、算出済みアレル比率ごとに、前記算出済みアレル比率が前記期待アレル比率より大きいか小さいかを決定し、かつ特定の座位でのその差異の大きさを決定することを含む。いくつかの態様では、尤度分布は、特定の仮説に関してある座位に存在するアレル比率について決定され、算出済みアレル比率が尤度分布の中央に近いほど、その仮説が正しい可能性はより高くなる。いくつかの態様では、前記方法は、各座位で仮説が正しい尤度を決定することを含む。いくつかの態様では、前記方法は、各座位で仮説が正しい尤度を決定し、各座位の仮説の確率とその仮説を組み合わせることを含み、最も高い複合確率が選択される。いくつかの態様では、前記方法は、座位ごとに、および１個以上の目的細胞に由来するＤＮＡまたはＲＮＡの前記試料中の総ＤＮＡまたはＲＮＡに対する可能な比率ごとに、仮説が正しい尤度を決定することを含む。いくつかの態様では、各仮説の複合確率は、各座位の仮説の確率と各可能な比率を組み合わせて決定され、最も高い複合確率を有する仮説が選択される。

【０３６９】

一態様では、以下の仮説が考えられる。H_{11}（すべての細胞は正常）、H_{10}（相同体１のみを有する、相同体２の欠失の細胞が存在する）、H_{01}（相同体２のみを有する相同体１の細胞が存在する）、H_{21}（相同体１の重複を有する細胞が存在する）、H_{12}（相同体２の重複を有する細胞が存在する）。癌細胞またはモザイク細胞（または、対象細胞に由来するＤＮＡまたはＲＮＡ分画）等の対象細胞の分画ｆについて、ヘテロ接合（ＡＢまたはＢＡ）ＳＮＰの期待アレル比率は以下のように得られる。

式（１）

【数１８】

$$
\begin{aligned}
r(AB,H_{11}) = r(BA,H_{11}) &= 0.5,\\
r(AB,H_{10}) = r(BA,H_{01}) &= \frac{1}{2-f},\\
r(AB,H_{01}) = r(BA,H_{10}) &= \frac{1-f}{2-f},\\
r(AB,H_{21}) = r(BA,H_{12}) &= \frac{1+f}{2+f},\\
r(AB,H_{12}) = r(BA,H_{21}) &= \frac{1}{2+f}.
\end{aligned}
$$

偏り、不純物、および配列決定誤りの修正

【０３７０】

前記ＳＮＰでの観察D_sは、各アレルが存在する元のマッピングリードn_A^0およびn_B^0からなる。よって、アレルＡおよびＢの増幅での予測偏りを用いた修正リードn_Aおよびn_Bを求めることができる。

【０３７１】

c_aを環境不純物（例えば、空気または環境中のＤＮＡに由来する不純物等）とし、$r(c_a)$を環境不純物に対するアレル比率（ここでは始めは０．５とする）とする。また、c_gを遺伝型解析による不純物率（例えば、他の試料に由来する不純物）とし、$r(c_g)$を不純物のアレル比率とする。s_e（Ａ，Ｂ）およびs_e（Ｂ，Ａ）は、（例えば、アレルＢが存在する場合に誤ってアレルＡを検出することにより）あるアレルを別のアレルとして呼び出した配列決定エラーとする。

【０３７２】

環境不純物、遺伝型解析による不純物、および配列決定誤りを修正して、ある期待アレル比率ｒに対する観察アレル比率$q(r, c_a, r(c_a), c_g, r(c_g), s_e(A,B), s_e(B,A))$を求めることができる。

【０３７３】

不純物の遺伝型はわからないので、集団頻度を用いて$P(r(c_g))$を求めることができる。より具体的には、ｐをアレルの１つ（基準アレルと言う場合もある）に対する集団頻度とする。すると、$P(r(c_g)=0)=(1-p)^2$、$P(r(c_g)=0)=2p(1-p)$、および$P(r(c_g)=0)=p^2$である。$r(c_g)$に対する条件付き期待値を用いて$E[q(r, c_a, r(c_a), c_g, r(c_g), s_e(A,B), s_e(B,A))]$を決定することができる。なお、環境不純物および遺伝型解析による不純物はホモ接合ＳＮＰを用いて決定されるので、欠失または重複の有無によって影響されない。また、必要に応じて、基準染色体を用いて環境不純物および遺伝型解析による不純物を測定することができる。

【０３７４】

各ＳＮＰの可能性

以下の式により、アレル比率ｒを仮定して、n_Aおよびn_Bを観察する確率が求められる。

式（２）：

【数１９】

$$P(n_A,n_B|r)=p_{bino}(n_A;n_A+n_B,r)=\binom{n_A+n_B}{n_A}r^{n_A}(1-r)^{n_B}.$$

【０３７５】
D_sをＳＮＰのデータとする。各仮説ｈ∈｛H_{11}，H_{01}，H_{10}，H_{21}，H_{12}｝について、式（１）でｒ＝ｒ（ＡＢ，ｈ）またはｒ＝ｒ（ＢＡ，ｈ）と仮定して、ｒ（c_g）に対する条件付き期待値を求めて、観察アレル比率Ｅ［ｑ（ｒ，c_a，ｒ（c_a），c_g，ｒ（c_g））］を決定することができる。よって、式（２）でｒ＝Ｅ［ｑ（ｒ，c_a，ｒ（c_a），c_g，ｒ（c_g），s_e（Ａ，Ｂ），s_e（Ｂ，Ａ））］と仮定して、Ｐ（D_s｜ｈ，ｆ）を求めることができる。
【０３７６】
検索アルゴリズム
いくつかの態様では、（例えば、平均値より少なくとも２または３標準偏差大きい、または小さいアレル比率を有するＳＮＰを無視あるいは除外することで）異常値と思われるアレル比率を有するＳＮＰは無視される。この方法の利点は、モザイクが存在する百分率が高い場合に前記アレル比率の可変性が高くてもよいので、モザイクによるこのＳＮＰの削除が確実に起こらないようにする。
【０３７７】
Ｆ＝｛f_1，….，f_N｝を、モザイクの百分率（例えば、腫瘍分画等）の検索空間とする。各ＳＮＰのＰ（D_s｜ｈ，ｆ）およびｆ∈Ｆを求めて、すべてのＳＮＰの尤度を組み合わせることができる。
【０３７８】
前記アルゴリズムは、各仮説の各ｆについて調べる。ある検索方法を用いて、ｆの範囲Ｆ＊（ここで欠失または重複の仮説の信頼度が欠失および重複がない仮説の信頼度よりも高い）がある場合、モザイクが存在すると結論づけられる。いくつかの態様では、Ｆ＊でのＰ（D_s｜ｈ，ｆ）の最尤推定値が決定される。必要に応じて、ｆ∈Ｆ＊の条件付き期待値が決定されてもよい。必要に応じて、各仮説の信頼度が決定されてもよい。
【０３７９】
さらなる態様
いくつかの態様では、二項分布の代わりにベータ二項分布が用いられる。いくつかの態様では、基準染色体または染色体分節を用いて試料のベータ二項分布の特定のパラメータが決定される。
【０３８０】
シミュレーションを用いた理論的性能
必要に応じて、所定のリード深度（ＤＯＲ）で基準リード数をＳＮＰに無作為に割り当てて、アルゴリズムの理論的性能を評価してもよい。正常な場合、二項確率パラメータにｐ＝０．５が用いられ、欠失または重複についてｐはそれに応じて改訂される。各シミュレーションの入力パラメータの例としては、（１）ＳＮＰ（Ｓ）の数、（２）ＳＮＰごとの定数ＤＯＲ（Ｄ）、（３）ｐ、および（４）実験数である。
【０３８１】
第１のシミュレーション実験
この実験ではＳ∈｛５００，１０００｝、Ｄ∈｛５００，１０００｝、およびｐ∈｛０％，１％，２％，３％，４％，５％｝に注目した。各設定で１，０００回のシミュレーション実験を行った（相を用いた２４，０００回の実験と、相を用いない２４，０００回の実験）。二項分布（必要に応じて他の分布を用いることができる）からリード数のシミュレーションを行った。偽陽性率（この場合はｐ＝０％）と偽陰性率（この場合はｐ＞０％

）は、相情報を用いて、および相情報を用いずに決定された。偽陽性率が図２６に示されている。特にＳ＝１０００、Ｄ＝１０００の場合に相情報は非常に有益である。Ｓ＝５００、Ｄ＝５００でも、前記アルゴリズムは、試験条件から相のある最も高い偽陽性率、または相のない最も高い偽陽性率を有する。偽陰性率は図２７に示されている。

【０３８２】

相情報は、モザイク百分率が低い（≦３％）場合に特に有用である。相情報がない場合、ｐ＝１％では偽陰性が多く観察された。というのは、欠失に関する信頼度は、H_{10}とH_{01}に等しく機会を割り当てて決定され、ある仮説を支持する小さな偏差は、他の仮説からの低尤度を補償するには不十分だからである。これは重複にも適用される。前記アルゴリズムは、ＳＮＰ数に比べてリード深度の感度がより高いと思われる。相情報を用いた結果では、連続した多くのヘテロ接合ＳＮＰで完全な相情報が利用可能であると仮定する。必要に応じて、より小さい分節にあるハプロタイプを確率的に組み合わせてハプロタイプ情報を得ることができる。

【０３８３】

第２のシミュレーション実験

この実験では、Ｓ∈｛１００，２００，３００，４００，５００｝、Ｄ∈｛１０００，２０００，３０００，４０００，５０００｝、およびｐ∈｛０％，１％，１．５％，２％，２．５％，３％｝注目して、各設定で１０，０００回のランダム実験を行った。偽陽性率（この場合はｐ＝０％）と偽陰性率（この場合はｐ＞０％）は、相情報を用いて、および相情報を用いずに決定された。偽陰性率は、ハプロタイプ情報を用いてＤ≧３０００およびＮ≧２００の場合は１０％未満であるが、Ｄ＝５０００およびＮ≧４００の場合に同じ性能が得られる（図２０Ａおよび２０Ｂ）。偽陰性率間の差は、モザイクの百分率が小さい場合に特に顕著であった（図２１Ａ～２５Ｂ）。例えば、ｐ＝１％では、ハプロタイプデータを用いなければ２０％未満の偽陰性率は得られないが、Ｎ≧３００およびＤ≧３０００では偽陰性率は０％に近づく。ｐ＝３％では、ハプロタイプデータを用いて０％の偽陰性率が観察されるが、ハプロタイプデータを用いなければ同じ性能を得るのにＮ≧３００およびＤ≧３０００が必要である。

【０３８４】

相決定済みデータを用いずに欠失および重複を検出する方法の例

いくつかの態様では、相未決定の遺伝子データを用いて、個体のゲノム（例えば、１個以上の細胞、あるいは無細胞ＤＮＡまたは無細胞ＲＮＡのゲノム）中の第二の相同染色体分節と比較して第一の相同染色体分節のコピー数の過剰出現があるかどうかが決定される。いくつかの態様では、相決定済み遺伝子データが用いられるが、相決定は無視される。いくつかの態様では、前記ＤＮＡまたはＲＮＡ試料は、２個以上の遺伝的に異なる細胞を含む個体に由来する無細胞ＤＮＡまたは無細胞ＲＮＡを含む無細胞ＤＮＡまたは無細胞ＲＮＡの混合試料である。いくつかの態様では、前記方法は、前記座位のそれぞれについて算出済みアレル比率と期待アレル比率の差異の大きさを用いる。

【０３８５】

いくつかの態様では、前記方法は、各座位に存在する各アレルの量を測定して、前記個体に由来する１個以上の細胞に由来するＤＮＡまたはＲＮＡの試料中の染色体または染色体分節にある多型座位の集合の遺伝子データを得ることを含む。いくつかの態様では、前記試料が由来する少なくとも１個の細胞においてヘテロ接合である座位（例えば、前記胎児においてヘテロ接合である座位および／または前記母親においてヘテロ接合である座位）についてアレル比率が算出される。いくつかの態様では、特定の座位の算出済みアレル比率は、前記アレルの１つの測定量をその座位のすべてのアレルの総測定量で除したものである。いくつかの態様では、特定の座位の算出済みアレル比率は、前記アレル（例えば、前記第一の相同染色体分節にあるアレル等）の１つの測定量を、その座位の１つ以上の他のアレル（例えば、第二の相同染色体分節にあるアレル等）の測定量で除したものである。前記算出済みアレル比率と前記期待アレル比率は、本明細書で開示した方法のいずれかを用いて、または任意の標準的な方法（例えば、本明細書に記載した算出済みアレル比

率または期待アレル比率の任意の数学的変換等）を用いて算出されてもよい。
【０３８６】
いくつかの態様では、検定統計量は、前記座位のそれぞれについて算出済みアレル比率と期待アレル比率の差異の大きさに基づいて算出される。いくつかの態様では、前記検定統計量Δは下記式を用いて算出される。
【数２０】

$$\Delta = \frac{\sum_{\text{すべての座位}}(\delta_i - \mu_i)}{\sqrt{\sum_{\text{すべての座位}} \sigma_i^2}} \quad \Delta = \frac{\sum_{All\ Loci}(\delta_i - \mu_i)}{\sqrt{\sum_{All\ Loci} \sigma_i^2}}$$

ここで、δ_iは、ｉ番目の座位の算出済みアレル比率と期待アレル比率の差異の大きさであり、
μ_iはδ_iの平均値であり、
σ_i^2はδ_iの標準偏差である。
【０３８７】
例えば、期待アレル比率が０．５であるとき、δ_iを次のように定義することができる。
【数２１】

$$\delta_i \triangleq \left|\frac{1}{2} - R_i\right|$$

R_iが二項ランダム変数であるという事実を用いて、μ_iおよびσ_iの値を計算することができる。いくつかの態様では、前記標準偏差はすべての座位で同じと仮定される。いくつかの態様では、前記標準偏差の平均値または加重平均値、あるいは前記標準偏差の推定値はσ_i^2の値に用いられる。いくつかの態様では、前記検定統計量は正常な分布を有すると仮定される。例えば、中心極限定理は、座位数（例えば、ＳＮＰ数Ｔ）が増加するにつれてΔの分布は標準正規偏差に近づくことを示す。
【０３８８】
いくつかの態様では、前記細胞の１つ以上のゲノム中の染色体または染色体分節のコピー数を規定する１つ以上の仮説の集合が列挙される。いくつかの態様では、前記検定統計量に基づいて最も可能性の高い仮説が選択されて、前記細胞の１つ以上のゲノム中の染色体または染色体分節のコピー数が決定される。いくつかの態様では、その仮説について前記検定統計量がその検定統計量の分布に属している確率が上限閾値より大きい場合は仮説が選択され、その仮説について前記検定統計量がその検定統計量の分布に属している確率が下限閾値より小さい場合は前記仮説の１つ以上の仮説が棄却される。あるいは、その仮説について前記検定統計量がその検定統計量の分布に属している確率が下限閾値と上限閾値の間である、または前記確率が十分に高い信頼度で決定されていない場合は仮説は選択も棄却もされない。いくつかの態様では、実験的分布（例えば、訓練データ（例えば、二倍体試料等のコピー数が既知である試料、または特定の欠失または重複を有することが既知の試料等）に由来する分布等）から上限閾値および／または下限閾値が決定される。このような実験的分布を用いて単一仮説棄却試験の閾値を選択することができる。
【０３８９】
なお、検定統計量ΔはＳから独立しており、よって、いずれも必要に応じて独立して用いることができる。
【０３９０】

アレル分布またはパターンを用いた欠失および重複を検出する方法の例

この項は、第二の相同染色体分節と比較して第一の相同染色体分節のコピー数過剰出現があるかどうかを決定する方法を含む。いくつかの態様では、前記方法は、（ｉ）前記個体の１個以上の細胞（例えば、癌細胞）のゲノムに存在する染色体または染色体分節のコピー数を指定する複数の仮説、または（ｉｉ）前記個体の１個以上の細胞のゲノム中の第二の相同染色体分節と比較して第一の相同染色体分節のコピー数の過剰出現度を規定する複数の仮説を列挙することを含む。いくつかの態様では、前記方法は、前記染色体または染色体分節にある複数の多型座位（例えば、ＳＮＰ座位）について前記個体に由来する遺伝子データを得ることを含む。いくつかの態様では、前記仮説のそれぞれについて、前記個体の期待される遺伝型の確率分布が作成される。いくつかの態様では、前記個体の得られた遺伝子データと前記個体の期待される遺伝型の確率分布のデータ適合が算出される。いくつかの態様では、前記データ適合に応じて１つ以上の仮説がランク付けされて、最高ランクの仮説が選択される。いくつかの態様では、前記データ適合を算出する工程、前記仮説をランク付けする工程、最高ランクの仮説を選択する工程の１つ以上において技術またはアルゴリズム（例えば、検索アルゴリズム等）が用いられる。いくつかの態様では、前記データ適合が、ベータ二項分布に対する適合または二項分布への適合である。いくつかの態様では、前記技術またはアルゴリズムが、最尤推定、最大事後確率推定、ベイズ推定、動的推定（例えば、動的ベイズ推定等）、および期待値最大化推定からなる群から選択される。いくつかの態様では、前記方法は、前記技術またはアルゴリズムを前記得られた遺伝子データおよび前記期待遺伝子データに適用することを含む。

【０３９１】

いくつかの態様では、前記方法は、（ｉ）前記個体の１個以上の細胞（例えば、癌細胞）のゲノムに存在する染色体または染色体分節のコピー数を指定する複数の仮説、またはまたは（ｉｉ）前記個体の１個以上の細胞のゲノム中の第二の相同染色体分節と比較して第一の相同染色体分節のコピー数の過剰出現度を規定する複数の仮説を列挙することを含む。いくつかの態様では、前記方法は、前記染色体または染色体分節にある複数の多型座位（例えば、ＳＮＰ座位）について前記個体に由来する遺伝子データを得ることを含む。いくつかの態様では、前記遺伝子データは、前記複数の多型座位のアレル数を含む。いくつかの態様では、仮説ごとに、前記染色体または染色体分節にある前記複数の多型座位にある期待アレル数について同時分布モデルが作成される。いくつかの態様では、前記同時分布モデルと前記試料で測定されたアレル数を用いて前記仮説の１つ以上の仮説の相対確率が決定されて、最も高い確率を有する仮説が選択される。

【０３９２】

いくつかの態様では、アレルの分布またはパターン（例えば、算出済みアレル比率のパターン等）を用いてＣＮＶ（例えば、欠失または重複）の有無が決定される。必要に応じて、このパターンに基づいて前記ＣＮＶの親起源が決定されてもよい。母親から受け継がれた重複は母親由来の染色体分節の余剰コピーであり、母親から受け継がれた欠失は母親由来の染色体分節コピーが欠けており、存在する染色体分節の唯一のコピーが父親に由来するものである。パターンの例が図１５Ａ～１９Ｄに示されており、以下でさらに説明される。

【０３９３】

目的の染色体分節の欠失の有無を決定するため、前記アルゴリズムは、１本の染色体につき多数のＳＮＰについて２つの起こりうるアレルのそれぞれに由来する配列計測の分布を考慮する。なお、重要なことだが、前記アルゴリズムのいくつかの態様は視覚化に適さない手法を用いている。よって、例示のために、関連する傾向をよりわかりやすく視覚化することができるように、そのデータを２つの最も可能性の高いアレルＡおよびＢの比率として簡単に図１５Ａ～図１８に示す。この簡単な例示は、前記アルゴリズムの起こりうる特徴のいくつかを考慮していない。例えば、アレル比率を表示するなんらかの視覚化方法でデータを表示できない前記アルゴリズムの２つの態様は、１）連鎖不平衡、すなわち１個のＳＮＰの測定が近隣のＳＮＰの同一性の可能性に及ぼす影響を利用する能力、およ

び２）プラットフォーム特性と増幅偏りを仮定したＳＮＰのアレル測定の期待分布を説明する非ガウスデータモデルの使用である。また、前記アルゴリズムの簡易版は、各ＳＮＰで最も共通するアレルを考慮するのみで、他の起こりうるアレルは無視される。

【０３９４】

ゲノム試料および母親の血液試料で目的の欠失が検出された。いくつかの態様では、実施例１の多重ＰＣＲと配列決定方法を用いてゲノム試料および母親の血漿試料が分析される。検査されたゲノムＤＮＡ症候群試料は、対象領域ではヘテロ接合ＳＮＰがなく、このアッセイが一染色体性（発症）と二染色体性（非発症）を区別することができることが確認された。母親の血液試料の無細胞ＤＮＡの分析は、２２ｑ１１．２欠失症候群、猫鳴き欠失症候群、ウォルフ・ヒルシュホーン欠失症候群、および図１４の胎児の他の欠失症候群を検出することができた。

【０３９５】

図１５Ａ～１５Ｃは、前記試料が完全に母親由来である（胎児の無細胞ＤＮＡは存在しない）場合（図１５Ａ）、中濃度（１２％）の胎児の無細胞ＤＮＡ分画を含む場合（図１５Ｂ）、高濃度（２６％）の胎児の無細胞ＤＮＡ分画を含む場合（図１５Ｃ）の２本の染色体の有無を示すデータを示す。ｘ軸は、染色体に沿った個体の多型座位の直線位置を表し、ｙ軸はアレルの総リード数（Ａ＋Ｂ）の分画としてアレルＡのリード数を表す。母親の遺伝型および胎児の遺伝型は、プロットの右側に示されている。各プロットは母親の遺伝型に応じて色づけされている。例えば、赤は母親の遺伝型ＡＡ、青は母親の遺伝型ＢＢ、緑は母親の遺伝型ＡＢを表す。測定は母親の血液から単離された総無細胞ＤＮＡに対して行われ、前記無細胞ＤＮＡは母親と胎児両方の無細胞ＤＮＡを含む。よって、各点は、そのＳＮＰをもたらした胎児と母親のＤＮＡの組み合わせを示している。したがって、母親の無細胞ＤＮＡの割合が０％から１００％に増加すると、母親の遺伝型および胎児の遺伝型に応じてプロット内でいくつかの点は多少上下する。

【０３９６】

すべての場合で、母親と胎児の両方でアレルＡがホモ接合である（ＡＡ）のＳＮＰが、プロットの上限と密接に関連づけられることがわかる。というのは、アレルＢが存在しないためアレルＡのリードの分画が高いからである。逆に、母親と胎児の両方でアレルＢがホモ接合であるＳＮＰは、プロットの下限と密接に関連づけられることがわかる。というのは、アレルＢのみが存在するため、アレルＡのリードの分画が低いからである。プロットの上限および下限に密接に関連づけられない点は、母親、胎児、または両方がヘテロ接合であるＳＮＰを表す。これらの点は、胎児の欠失または重複を同定するのに有益であるが、両親の遺伝と母親の遺伝を決定するのにも有益な可能性がある。これらの点は、母親の遺伝型と胎児の遺伝型の両方および胎児の分画に基づいて分かれており、ｙ軸に沿った各点の正確な位置は化学量論と胎児の分画の両方に依存する。例えば、母親がＡＡで胎児がＡＢである座位は、アレルＡのリードの分画が異なっていると予測され、したがって、胎児の分画によってｙ軸の位置が異なる。

【０３９７】

図１５Ａは、妊娠していない女性のデータであり、遺伝型が完全に母親のものであるパターンを示す。このパターンは、点の「集合」を含む。赤色の集合はプロットの上部（ここでは母親の遺伝型がＡＡであるＳＮＰ）と密接に関連づけられて、青い色の集合はプロットの底部（ここでは母親の遺伝型がＢＢのＳＮＰ）と密接に関連づけられて、単一の中央の緑色の集合（ここでは母親の遺伝型がＡＢのＳＮＰ）と密接に関連づけられている。図１５Ｂでは、胎児のアレルがアレルＡのリードの分画に含まれているのでいくつかのアレルの点の位置がｙ軸に沿って上下に移動している。図１５Ｃでは、２本の赤色の周辺バンドと２本の青色の周辺バンド、３本の緑色の中のバンドを含むパターンが容易に見える。これら３本の緑色の中のバンドは、母親においてヘテロ接合であるＳＮＰに対応しており、プロットの上部（赤色）と底部（青色）の両方にある２本の「周辺」バンドは、それぞれ母親においてホモ接合であるＳＮＰに対応している。

【０３９８】

図１６Ａに２２ｑ１１．２欠失の保有者（この欠失を有する母親）の分析を示す。この欠失の保有者は、この領域にヘテロ接合ＳＮＰを有していない。というのは、この保有者のみがこの領域の１コピーを有するからである。よって、この欠失は緑色のＡＢのＳＮＰがないことにより示される。図１６Ｂに、両親から受け継がれた胎児の２２ｑ１１欠失の分析を示す。この胎児のみが染色体分節の単一コピーを受け継ぎ（両親から受け継がれた欠失の場合、胎児に存在するコピーは母親に由来する）、よってこの分節の各座位に単一アレルを受け継いでいる場合、この胎児のヘテロ接合性は起こりえない。したがって、唯一起こりうる胎児のＳＮＰの同一性はＡまたはＢである。なお、周辺バンドの内側がなくなっている。両親から受け継がれた欠失の場合、特徴的なパターンは、母親ではヘテロ接合であるＳＮＰを示す２本の中の緑色のバンドを含み、母親ではホモ接合であり、それぞれプロットの上限および下限（１および０）と密接に関連づけられているＳＮＰを示す単一の赤色のバンドと青色のバンドを有する。

【０３９９】

図１７に母親から受け継がれた猫鳴き欠失症候群の分析を示す。３本の緑色のバンドの代わりに２本の中の緑色のバンドと、２本の赤色の周辺バンドと２本の青色の周辺バンドがある。母親と胎児の両方が欠失を有しているため、母親から受け継がれた欠失（例えば、デュシェンヌ型筋ジストロフィーの母親の保有者）もまた、母親と胎児のＤＮＡ混合試料（例えば、血漿試料）中の欠失領域において少量の信号に基づいて検出することができる。

【０４００】

図１８は、両親から受け継がれたウォルフ・ヒルシュホーン欠失症候群のプロットであり、１本の赤い周辺バンドと１本の青い周辺バンドによって示されている。

【０４０１】

必要に応じて、欠失または重複（例えば、癌と関連づけられたＣＮＶ等）を有すると疑われる個体に由来する試料について同様のプロットを作成することができる。このようなプロットでは、ＣＮＶを有しない細胞の遺伝型に基づいて以下の色コードを用いることができる。赤色はＡＡの遺伝型を示し、青色はＢＢの遺伝型を示し、緑色はＡＢの遺伝型を示す。いくつかの態様では、欠失の場合、そのパターンは、前記個体ではヘテロ接合である（緑色の上のバンドは欠失のない細胞に由来するＡＢおよび欠失がある細胞に由来するＡを示し、緑色の下のバンドは欠失がない細胞に由来するＡＢおよび欠失がある細胞に由来するＢを示す）ＳＮＰを表す２本の中の緑色のバンドを含み、前記個体ではホモ接合であり、それぞれプロットの上限および下限（１および０）と密接に関連づけられているＳＮＰを示す単一の赤色のバンドと青色のバンドを有する。いくつかの態様では、欠失を有する細胞、ＤＮＡ、またはＲＮＡの分画が増加するにつれて、２本の緑色のバンドの分離も増加する。

【０４０２】

多胎妊娠を同定および分析する方法の例

いくつかの態様では、本発明の方法のいずれかを用いて双胎妊娠等の多胎妊娠の存在が検出されるが、胎児のうちの少なくとも一方は、少なくともう一方の胎児と遺伝的に異なっている。いくつかの態様では、検査した座位のいくつか（またはすべて）で異なるアレル、異なるアレル比率、または異なるアレル分布を有する２体の胎児の存在に基づいて二卵性双生児が同定される。いくつかの態様では、試料（例えば、血漿試料）中に同じまたは異なる胎児の分画を有する可能性がある２体の胎児について各座位（例えば、ＳＮＰ座位）の期待アレル比率を決定することで二卵性双生児が同定される。いくつかの態様では、母親の遺伝型と遺伝型集団頻度を条件として、特定の対の胎児の起こりうる遺伝型のいくつか、あるいはすべてを考慮して特定の対の胎児の分画（ここでｆ１は胎児１の胎児の分画であり、ｆ２は胎児２の胎児の分画）の尤度が算出される。２体の胎児の遺伝型と一人の母親の遺伝型の混合物を胎児の分画と組み合わせて、あるＳＮＰの期待アレル比率が決定される。例えば、母親がＡＡで、胎児１がＡＡで、胎児２がＡＢである場合、このＳＮＰのアレルＢの総分画は、ｆ２の半分である。尤度の算出では、胎児の遺伝型の起こ

りうる組み合わせのすべてに基づいてすべてのＳＮＰが期待アレル比率とどの程度適合するかが算出される。データに最も適合する１対の胎児の分画（ｆ１、ｆ２）が選択される。胎児の特定の遺伝型を算出することは必要ではない。その代わり、例えば、統計的組み合わせで起こりうる遺伝型のすべてを考慮することができる。いくつかの態様では、前記方法が単生児と同一性双生児を区別しない場合、超音波を用いて単生児か同一性双胎妊娠であるかを決定することができる。超音波により双胎妊娠が検出されると、同一性双胎妊娠であると仮定することができる。というのは、兄弟性双胎妊娠であれば上述のＳＮＰ分析に基づいて検出されるはずであるからである。

【０４０３】

いくつかの態様では、事前検査（例えば、超音波等）に基づいて妊娠している母親が多胎妊娠（例えば、双胎妊娠）であると分かっている。本発明の方法のいずれかを用いて、多胎妊娠が同一性双生児または兄弟性双生児を含むかどうかを決定することができる。例えば、同一性双生児（単胎妊娠としての同じアレル比率）または二卵性双生児（例えば、上述のアレル比率の算出）に期待されることと測定されたアレル比率とを比較することができる。いくつかの同一性双生児は単一絨毛膜性双生児であり、双胎児間輸血症候群を有する危険性がある。よって、本発明の方法を用いて、同一性双生児であると判定された双生児は、（例えば、超音波等によって）検査されて、単一絨毛膜性双生児であるかどうかが判定され、単一絨毛膜性双生児である場合はこれらの双生児を（例えば、１６週目から隔週で超音波等により）監視できることが望ましい。

【０４０４】

いくつかの態様では、本発明の方法のいずれかを用いて、多胎妊娠（例えば、双胎妊娠等）の胎児のいずれかが異数体であるかどうかが決定される。双生児の異数性検査は、胎児の分画の推定値から始まる。いくつかの態様では、上述のようにデータに最も適合する１対の胎児の分画（ｆ１、ｆ２）が選択される。いくつかの態様では、起こりうる胎児の分画の範囲にわたり、パラメータ対（ｆ１、ｆ２）について最尤推定が行われる。いくつかの態様では、ｆ２の範囲は、０からｆ１である。というのは、ｆ２は小さい方の胎児分画として定義されるからである。（ｆ１、ｆ２）の対を仮定して、ＳＮＰ座位等の座位の集合で観察されたアレル比率からデータの尤度が算出される。いくつかの態様では、前記データ尤度は母親の遺伝型、入手可能であれば父親の遺伝型、集団頻度、得られた胎児の遺伝型の確率を反映する。いくつかの態様では、ＳＮＰは独立して仮定される。胎児の推定分画対は、最も高いデータ尤度を生成するものである。ｆ２が０である場合、このデータは、同一性双生児を示す胎児の遺伝型の１集合のみによって最もよく説明される。この場合、ｆ１は組み合わされた胎児の分画である。そうでなければ、ｆ１とｆ２は、個々の双生児の胎児の分画の推定値である。（ｆ１、ｆ２）の最適推定値が確立されると、必要に応じて母親の遺伝型と胎児の遺伝型の任意の組み合わせについて血漿中のアレルＢの総分画を予測することができる。個々の配列リードを個々の胎児に割り当てる必要はない。２つの仮説のデータ尤度を比較する他の最尤推定値を用いて倍数性検査が行われる。いくつかの態様では、同一性双生児の場合、（ｉ）双生児の両方が正倍数体であるという仮説、および（ｉｉ）双生児の両方が三染色体であるという仮説が考慮される。いくつかの態様では、二卵性双生児の場合、（ｉ）双生児の両方が正倍数体であるとい仮説、および（ｉｉ）双生児の少なくとも一方が三染色体であるという仮説が考慮される。二卵性双生児に対する三染色体性仮説は、低い胎児の分画に基づいている。というのは、高い胎児の分画を有する双生児の三染色体性も検出されるからである。二染色体性仮説または三染色体性仮説のいずれかで条件付けられた各対象ゲノム座位でのリード期待数を予測する方法を用いて倍数性尤度が算出される。二染色体性基準染色体の要件は存在しない。リード期待数の分散モデルは、個々の対象座位の性能と、座位間の相関関係を考慮する（例えば、２０１４年６月５日に提出された米国シリアル第６２／００８，２３５号、および２０１４年８月４日日に提出された米国シリアル第６２／０３２，７８５号、それぞれ参照にその全内容が本明細書に組み込まれる）。小さい双生児が胎児の分画ｆ１を有する場合、その双生児の三染色体性を検出する能力は、同じ胎児の分画で単胎妊娠における三染色体性を

検出する能力と同等である。これは、いくつかの態様において三染色体性を検出する前記方法の一部が遺伝型に依存しておらず、多胎妊娠か単胎妊娠かを区別しないためである。単に決定された胎児の分画に応じてリード数の増加を検索するのみである。

【０４０５】

いくつかの態様では、前記方法は、（例えば、上述のように）ＳＮＰ座位に基づいて双生児の存在を検出することを含む。双生児が検出された場合、ＳＮＰを用いて、上述のように各胎児（ｆ１、ｆ２）の胎児の分画が決定される。いくつかの態様では、高い信頼度二染色体性呼び出しを有する試料を用いてＳＮＰあたりの増幅偏りが決定される。いくつかの態様では、高い信頼度二染色体性呼び出しを有するこれら試料は、１つ以上の目的の試料と同一操作で分析される。いくつかの態様では、前記ＳＮＰあたりの増幅偏りを用いて、２体の双生児の胎児の分画の低い方を仮定して１つ以上の目的の染色体または染色体分節（例えば、２１番染色体等）、あるいは二染色体性仮説と三染色体性仮説に関するリード分布がモデル化される。２つのモデルと目的の染色体または染色体分節の測定量を仮定して二染色体性または三染色体性の尤度すなわち確率が算出される。

【０４０６】

いくつかの態様では、陽性の異数性呼び出し（例えば、三染色体性呼び出し）の閾値は、低いほうの胎児の分画を有する双生児に基づいて設定される。この場合、他の双生児が陽性である、または両方が陽性である場合は、総染色体出現数は明らかに前記閾値より大きい。

【０４０７】

計測方法／定量方法の例

いくつかの態様では、１つ以上の計測方法（定量方法ともいう）を用いて１つ以上のＣＮＳ（例えば、染色体分節または染色体全体の欠失または重複等）が検出される。いくつかの態様では、１つ以上の計測方法を用いて、前記第一の相同染色体分節のコピー数過剰出現が前記第一の相同染色体分節の重複または前記第二の相同染色体分節の欠失によるものであるかどうかが決定される。いくつかの態様では、１つ以上の計測方法を用いて、（例えば、１個、２個、３個、４個、あるいはそれ以上の余剰コピーがあるかどうか等）重複している染色体分節または染色体の余剰コピー数が決定される。いくつかの態様では、１つ以上の計測方法を用いて、多くの重複と小さい方の腫瘍分画を有する試料と、少ない方の重複と大きい方の腫瘍分画を有する試料が区別される。例えば、１つ以上の計測方法を用いて、染色体の４つの余剰コピーと１０％の腫瘍分画を有する試料と、染色体の２つの余剰コピーと２０％の腫瘍分画を有する試料が区別されてもよい。例示的な方法は、例えば米国公開第２００７／０１８４４６７号、第２０１３／０１７２２１１号、および第２０１２／０００３６３７号、米国特許第８，４６７，９７６号、第７，８８８，０１７号、第８，００８，０１８号、第８，２９６，０７６号、および第８，１９５，４１５号、２０１４年６月５日に提出された米国シリアル第６２／００８，２３５号、および２０１４年８月４日に提出された米国シリアル第６２／０３２，７８５号に開示されており、それぞれ参照によりその全内容が本明細書に組み込まれる。

【０４０８】

いくつかの態様では、前記計測方法は、１つ以上の所定の染色体または染色体分節をマッピングするＤＮＡ配列に基づくリード数を計測することを含む。このようないくつかの方法は、特定の染色体または染色体分節にマッピングされるＤＮＡ配列リード数の基準値（カットオフ値）の作成を含み、前記基準値を超えるリード数は特定の遺伝的異常を示す。

【０４０９】

いくつかの態様では、１箇所以上の座位のすべてのアレルの総測定量（例えば、多型座位または非多型座位の総量等）が基準量と比較される。いくつかの態様では、前記基準量は、特定のコピー数仮説の（ｉ）閾値または（ｉｉ）期待量である。いくつかの態様では、前記基準量（ＣＮＶが存在しない場合）は、欠失または重複がないと分かっている、あるいは期待される１つ以上の染色体または染色体分節について１箇所以上の座位のすべて

のアレルの総測定量である。いくつかの態様では、前記基準量（ＣＮＶが存在する場合）は、欠失または重複を有すると分かっている、あるいは期待される１つ以上の染色体または染色体分節について１箇所以上の座位のすべてのアレルの総測定量である。いくつかの態様では、前記基準量は、１つ以上の基準染色体または染色体分節について１箇所以上の座位のすべてのアレルの総測定量である。いくつかの態様では、前記基準量は、２つ以上の異なる染色体、染色体分節、または異なる試料について決定された値の平均値または中央値である。いくつかの態様では、ランダム（例えば、大量並行ショットガン配列決定）または標識化配列決定を用いて、１つ以上の多型座位または非多型座位の量が決定される。

【０４１０】

いくつかの態様では、前記方法は、基準量を用いて、（ａ）目的の染色体または染色体分節にある遺伝子材料の量を測定し、（ｂ）工程（ａ）で得た量と基準量を比較し、（ｃ）前記比較に基づいて欠失または重複の有無を同定することを含む。

【０４１１】

いくつかの態様では、前記方法は、基準染色体または染色体分節を用いて、試料に由来するＤＮＡまたはＲＮＡの配列決定を行って、対象座位に対して並べる複数の配列タグを得ることを含む。いくつかの態様では、前記配列タグは、特定の対象座位（例えば、長さが１５～１００個のヌクレオチド）に割り当てるのに十分な長さである。前記対象座位は、前記試料中で異常な分布を有すると疑われる少なくとも１つの第一の染色体または染色体分節および前記試料中に正常に分散していると仮定される少なくとも１つの第二の染色体または染色体分節を含む複数の異なる染色体または染色体分節に由来する。いくつかの態様では、前記複数の配列タグは、対応する対象座位に割り当てられる。いくつかの態様では、前記第一の染色体または染色体分節の対象座位に並べる配列タグ数と、前記第二の染色体または染色体分節の対象座位に並べる配列タグ数が決定される。いくつかの態様では、これらの配列タグ数を比較して、前記第一の染色体または染色体分節の異常な分布（例えば、欠失または重複）の有無が決定される。

【０４１２】

いくつかの態様では、ＣＮＶの決定にｆの値（例えば、胎児の分画または腫瘍分画等）を用いて、２本の染色体または染色体分節の量の観察された差異と、特定のタイプのＣＮＶについてｆの値を仮定して期待される差異を比較する（例えば、米国公開第２０１２／０１９００２０号、米国公開第２０１２／０１９００２１号、米国公開第２０１２／０１９０５５７号、米国公開第２０１２／０１９１３５８号、それぞれ参照によりその全内容が本明細書に組み込まれる）。例えば、胎児の分画が増加するにつれて、前記胎児を宿している母親由来の血液試料中の二染色体の基準染色体分節の量と前記胎児で重複している染色体分節の量の差も増加する。また、腫瘍分画が増加するにつれて、二染色体の基準染色体分節の量と腫瘍で重複している染色体分節の量の差が増加する。いくつかの態様では、前記方法は、基準染色体または染色体分節（例えば、二染色体であると期待される、あるいは分かっている染色体または染色体分節）に対する目的の染色体または染色体分節の相対頻度と、ｆの値を比較して、前記ＣＮＶの尤度を決定することを含む。例えば、前記第一の染色体または染色体分節と前記基準染色体または染色体分節の量の差が、起こりうるそれぞれのＣＮＶ（例えば、目的の染色体分節の１つまたは２つの余剰コピー等）についてｆの値を仮定して期待されるものと比較することができる。

【０４１３】

以下の予測例では、計測方法／定量方法を用いて前記第一の相同染色体分節の重複と前記第二の相同染色体分節の欠失を区別することが例示される。宿主の正常な二染色体のゲノムを基準値と見なすと、正常な細胞と癌細胞の混合物の分析により混合物中の基準値と癌ＤＮＡの平均差が得られる。例えば、アッセイの対象である染色体領域で欠失を有する細胞に由来する試料中のＤＮＡが１０％であるとする。いくつかの態様では、定量法により、その領域に対応するリードの量が、正常試料に期待される量の９５％であると期待されることが分かる。これは、対象領域の欠失を有する腫瘍細胞のそれぞれで２つの対象の

染色体領域の１つが失われており、その領域にマッピングされるＤＮＡの総量が｛９０％（正常細胞の場合）＋（１／２×１０％）（腫瘍細胞の場合）＝９５％｝であるからである。あるいは、いくつかの態様では、アレル法により、ヘテロ接合座位に存在するアレルの比率が平均で１９：２０であることが分かる。ここで、アッセイの対象である染色体領域を５倍局所増幅した細胞に由来する試料中のＤＮＡが１０％とする。いくつかの態様では、定量法により、その領域に対応するリードの量が、正常試料に期待される量の１２５％であると期待されることが分かる。これは、５倍の局所増幅を有する腫瘍細胞のそれぞれで２つの対象の染色体領域の１つがその対象領域にわたって５回余分にコピーされており、その領域にマッピングされるＤＮＡの総量が｛９０％（正常細胞の場合）＋（（２＋５）×１０％／２）（腫瘍細胞の場合）＝１２５％｝であるからである。また、いくつかの態様では、アレル法により、ヘテロ接合座位に存在するアレルの比率が平均で２５：２０であることが分かる。なお、アレル法のみを用いて、１０％の無細胞ＤＮＡを有する試料中の染色体領域を５倍局所増幅したものは、４０％の無細胞ＤＮＡを有する試料中の同じ領域の欠失と同じように見える。これら２つの場合では、欠失の場合に過小出現したハプロタイプは、局所重複を有する場合ではＣＮＶがないハプロタイプであると思われる。また、欠失の場合でＣＮＶがないハプロタイプは、局所重複がある場合では過剰出現したハプロタイプと思われる。このアレル法によって作成された尤度と定量法により作成された尤度を組み合わせて、２つの可能性が区別される。

【０４１４】

基準試料を用いた計測方法／定量方法の例

１つ以上の基準試料を用いる定量方法の例が、２０１４年６月５日に提出された米国シリアル第６２／００８，２３５号、および２０１４年８月４日に提出された米国シリアル第６２／０３２，７８５号に記載されており、参照によりその全体が本明細書に組み込まれる。いくつかの態様では、腫瘍ＤＮＡの最も高い分画を有する試料を選択する、ゼロ最も近いＺ値を有する試料を選択する、最も高い信頼度または尤度を有するＣＮＶがないに対応する仮説とデータが適合する試料を選択する、正常であることが既知の試料を選択する、癌を有する尤度が最も低い個体（例えば、低年齢、乳癌のスクリーニングにおける男性、家族歴がない等の個体）に由来する試料を選択する、ＤＮＡの入力量が最も高い試料を選択する、ノイズ比率に対して最も高い信号を有する試料を選択する、癌を有する尤度と相関関係があると思われる他の基準に基づいて試料を選択する、あるいはいくつかの基準の組み合わせを用いて試料を選択することのいずれかにより、１つ以上の染色体または目的の染色体（例えば、正常試料等）にＣＮＶを有していない可能性が最も高い１つ以上の基準試料が同定される。基準集合が一旦選択されると、これらの場合が二染色体であると仮定して、各座位について、ＳＮＰあたりの偏り、すなわち、実験に特異的な増幅および他の処理による偏りを推定することができる。次いで、この実験に特異的な偏り推定値を用いて、目的の染色体（例えば、２１番染色体の座位等）や、必要に応じて他の染色体座位、二染色体性が２１番染色体と仮定される下位集合の一部ではない試料の測定での偏り修正することができる。一旦、倍数性が未知であるこれら試料で偏りが修正されると、同じまたは異なる方法を用いてこれら試料のデータに対して２回目の分析が行われて、前記個体（例えば、胎児）が三染色体性２１を患っているかどうかを決定することができる。例えば、倍数性が未知の残りの試料に定量方法を用いることができ、２１番染色体に関する修正された測定遺伝子データを用いてＺ値を算出することができる。また、２１番染色体の倍数性状態の予備推定値の一部として、胎児の分画（または癌を有すると疑われる個体に由来する試料の腫瘍分画）を算出することができる。この胎児の分画の場合に対して、二染色体性（二染色体性仮説）の場合に期待される修正済みリードの割合と、三染色体性（三染色体性仮説）の場合に期待される修正済みリードの割合を算出することができる。また、胎児の分画が前もって測定されなかった場合、それぞれの胎児の分画について二染色体性仮説と三染色体性仮説の集合を作成することができる。各場合に、種々のＤＮＡ座位の選択と測定における期待統計分散を仮定して、修正済みリードの割合の期待分布を算出することができる。修正済みリードの観察割合と、修正済みリードの期待割合の分

布と比較して、倍数性が未知の試料のそれぞれについて二染色体性仮説と三染色体性仮説の尤度比率を算出することができる。修正済み倍数性状態として最も高い算出済み尤度を有する仮説に関連づけられた倍数性状態を選択することができる。

【０４１５】

いくつかの態様では、癌を有する尤度が十分に低い試料の下位集合を選択して、試料の対照集合として作用させてもよい。この下位集合は固定数であってもよく、閾値未満の試料のみを選択することに基づく変数であってもよい。試料の下位集合に由来する定量的データを組み合わせ、平均、あるいは加重平均を用いて組み合わせてもよい。加重平均は、正常試料の尤度に基づいている。前記定量的データを用いて、対照試料の即席バッチで試料の配列決定で増幅対象の座位あたりの偏りを決定してもよい。前記座位あたりの偏りはまた、試料の他のバッチに由来するデータを含んでいてもよい。前記試料の下位集合がＣＮＶを含んでおらず、観察されたいずれかの過剰増幅または過小増幅が増幅および／または配列決定、あるいは他の偏りによるものと仮定すると、前記座位あたりの偏りは、他の座位と比べてその座位で観察される相対的な過剰増幅または過小増幅を示してもよい。前記座位あたりの偏りは、増幅産物のＧＣ含量を考慮してもよい。座位あたりの偏りを算出するために、座位を座位群に分けてもよい。一旦、前記複数の座位の各座位について座位あたりの偏りが算出されると、各座位の定量測定を調整してその座位の偏り効果を除去して前記試料の下位集合に含まれない１つ以上の試料についての配列決定データと、任意で前記試料の下位集合に含まれる１つ以上の試料を修正してもよい。例えば、患者の下位集合で平均の２倍のリード深度を有するＳＮＰ１が観察された場合、前記調整はＳＮＰ１に対応するリード数を半分の数に置き換えることを含んでいてもよい。目的の座位がＳＮＰである場合は、前記調整は、その座位に存在するアレルのそれぞれに対応するリード数を半分にすることを含んでいてもよい。一旦、１つ以上の試料中のそれぞれで配列決定データが調節されると、１つ以上の染色体領域でＣＮＶを検出するための方法を用いてそのデータを分析してもよい。

【０４１６】

ある例では、試料Ａは、定量方法を用いて分析される正常な細胞と癌性細胞の混合物に由来する増幅されたＤＮＡの混合物である。以下に、起こりうるデータの例を示す。２２番染色体にある長腕(q-arm)領域は、その領域にマッピングされると期待されるＤＮＡの９０％のみを有することが分かっている。ＨＥＲ２遺伝子に対応する局所領域は、その領域にマッピングされると期待されるＤＮＡの１５０％を有することが分かっている。５番染色体の短腕(p-arm)は、そこにマッピングされると期待されるＤＮＡの１０５％を有することが分かっている。臨床医は、前記試料が２２番染色体にある長腕領域の欠失とＨＥＲ２遺伝子の重複を有すると推論することが可能である。前記臨床医は、２２ｑ欠失は乳癌に共通すること、および両方の染色体の２２ｑ領域の欠失を有する細胞は通常、生き残らないことから、前記試料中のＤＮＡの約２０％は２本の染色体の１本で２２ｑ欠失を有する細胞に由来すると推論することが可能である。前記臨床医はまた、腫瘍細胞に由来する混合試料に由来するＤＮＡが、ＨＥＲ２領域および２２ｑ領域が同質である遺伝的腫瘍細胞の集合に由来する場合、これらの細胞はＨＥＲ２領域の５倍の重複を含むと推論することが可能である。

【０４１７】

ある例では、試料Ａはまた、アレル法を用いて分析される。以下に、起こりうるデータの例を示す。２２番染色体の長腕の同じ領域にある２つのハプロタイプは、４：５の比率で存在する。ＨＥＲ２遺伝子に対応する局所領域の２つのハプロタイプは１：２の比率で存在する。５番染色体の短腕にある２つのハプロタイプは、２０：２１の比率で存在する。ゲノムのすべての他のアッセイ領域は、いずれのハプロタイプでも統計的に有意な過剰を有していない。臨床医は、前記試料が、２２ｑ領域、ＨＥＲ２領域、および５短腕にＣＮＶを有する腫瘍に由来するＤＮＡを含むと推論することが可能である。２２ｑ欠失が乳癌で非常に一般的であるという知識および／またはゲノムの２２ｑ領域にマッピングされるＤＮＡ量の過小出現を示す定量分析に基づいて、前記臨床医は２２ｑ欠失を有する腫瘍

の存在を推論することが可能である。ＨＥＲ２増幅が乳癌で非常に一般的であるという知識および／またはゲノムのＨＥＲ２領域にマッピングされるＤＮＡ量の過剰出現を示す定量分析に基づいて、前記臨床医はＨＥＲ２増幅を有する腫瘍の存在を推論することが可能である。

【０４１８】

基準の染色体または染色体分節の例

一部の態様では、本明細書に記載の方法のいずれかを、基準となる１つ以上の染色体または染色体分節に対しても実行し、その結果を対象である１つ以上の染色体または染色体分節の結果と比較する。

【０４１９】

一部の態様では、基準の染色体または染色体分節をＣＮＶが存在しないと期待される内容に関する対照として用いる。一部の態様では、基準は、対象と同一の染色体または染色体分節に欠失または重複がないことがわかっているないしは期待される１つ以上の別の試料から得られた、対象と同一の染色体または染色体分節である。一部の態様では、基準は、試験対象試料から得られ二染色体性であると期待される、別の染色体または染色体分節である。一部の態様では、基準は、試験対象である同一の試料に含まれる目的の染色体のうちの１つから得られた別の分節である。たとえば、基準は欠失または重複の可能性のある領域の外側にある１つ以上の分節であってもよい。試験対象である同一の染色体に基準があることによって、異なる染色体間で見られる可変性（染色体間で見られる代謝、アポトーシス、ヒストン、不活化、および／または増幅の違いなど）を避けることができる。試験対象の染色体と同一の染色体に存在しＣＮＶを含まない分節を分析することによって相同体間で見られる代謝、アポトーシス、ヒストン、不活化、および／または増幅の違いを確認することもでき、ＣＮＶを含まない相同体間での可変性のレベルを、可能性のあるＣＮＶによる結果と比較するために確認することが可能である。一部の態様では、可能性のあるＣＮＶについてのアレル比の算出値と期待値との差異の大きさは、基準についての対応する大きさよりも大きく、それによってＣＮＶが存在することが確認される。

【０４２０】

一部の態様では、基準となる染色体または染色体分節を、対象である特定の欠失または重複などのＣＮＶが存在すると期待される内容に関する対照として用いる。一部の態様では、基準は、対象と同一の染色体または染色体分節に欠失または重複があることがわかっているか期待される１つ以上の別の試料から得られた、対象と同一の染色体または染色体分節である。一部の態様では、基準は、試験対象試料から得られＣＮＶが存在することがわかっているか期待される、別の染色体または染色体分節である。一部の態様では、可能性のあるＣＮＶについてのアレル比の算出値と期待値との差異の大きさは、そのＣＮＶに関する基準についての対応する大きさと類似し（たとえば、有意差はない）、それによってＣＮＶが存在することが確認される。一部の態様では、可能性のあるＣＮＶについてのアレル比の算出値と期待値との差異の大きさは、そのＣＮＶに関する基準についての対応する大きさよりも小さく（たとえば有意に小さい）、それによってＣＮＶが存在しないことが確認される。一部の態様では、癌細胞（あるいは癌細胞に由来する無細胞ＤＮＡまたは無細胞ＲＮＡなどのＤＮＡまたはＲＮＡ）の遺伝型が非癌性細胞（あるいは非癌性細胞に由来する無細胞ＤＮＡまたは無細胞ＲＮＡなどのＤＮＡまたはＲＮＡ）の遺伝型とは異なる１つ以上の座位を用いて腫瘍率を決定する。腫瘍率を用いて、第一の相同染色体分節のコピー数過剰出現の原因が第一の相同染色体分節の重複であるか第二の相同染色体分節の欠失であるかを決定することができる。腫瘍率を用いて、重複した染色体分節または染色体の余分なコピー数（たとえば１個、２個、３個、４個、またはそれ以上の余分なコピーがあるか）を決定し、たとえば染色体の余剰コピー数が４個であり腫瘍率が１０％の試料を、染色体の余剰コピー数が２個であり腫瘍率が２０％である試料と区別することもできる。腫瘍率を用いて、可能性のあるＣＮＶについて観測されたデータが期待されたデータにどの程度適合しているかを決定することもできる。一部の態様では、ＣＮＶによる過剰出現の程度を利用して、個体に対し特定の治療または治療計画を選択する。たとえば、

一部の治療薬は、染色体分節のコピーが少なくとも４個、６個、またはそれ以上である場合にのみ有効である。

【０４２１】

一部の態様では、腫瘍率の決定に用いられる１つ以上の座位は基準の染色体または染色体分節に存在する。基準の染色体または染色体分節は、たとえば、二染色体性であることがわかっているか期待される染色体または染色体分節、一般的な癌細胞またはある個体が有していることがわかっているか有している可能性の高い特定の種類の癌では重複または欠失が起こることが稀な染色体または染色体分節、あるいは異数性であるとは考えられない染色体または染色体分節（欠失または重複した場合、細胞死につながると期待される分節）などである。一部の態様では、本発明の方法のいずれかを用いて、基準の染色体または染色体分節が癌細胞と非癌性細胞の両方で二染色体性であることを確認する。一部の態様では、二染色体性検出の信頼度が高い１つ以上の染色体または染色体分節を用いる。

【０４２２】

腫瘍率の決定に用いることのできる例示の座位は、個体の癌細胞（あるいは癌細胞に由来する無細胞ＤＮＡまたは無細胞ＲＮＡなどのＤＮＡまたはＲＮＡ）には存在するが非癌性細胞（あるいは非癌性細胞に由来するＤＮＡまたはＲＮＡ）には存在しない多型または変異（ＳＮＰなど）を含む。一部の態様では、個体から採取した試料（血漿試料または腫瘍生検など）に含まれる、癌細胞（または癌細胞に由来するＤＮＡまたはＲＮＡ）が非癌性細胞（あるいは非癌性細胞に由来するＤＮＡまたはＲＮＡ）には存在しないアレルを有する多型座位を特定し；特定した多型座位のうちの少なくとも１つ以上の座位における癌細胞特有のアレルの量を用いて試料の腫瘍率を決定することにより、腫瘍率を決定する。一部の態様では、非癌性細胞ではその多型座位が第一のアレルのホモ接合型であり、癌細胞ではその多型座位が（ｉ）第一のアレルと第二のアレルのヘテロ接合型または（ｉｉ）第二のアレルのホモ接合型である。一部の態様では、非癌性細胞ではその多型座位が第一のアレルと第二のアレルのヘテロ接合型であり、癌細胞は（ｉ）その多型座位に１個または２個の第三のアレルのコピーを有する。一部の態様では、癌細胞は非癌性細胞には存在しないアレルのコピーを１個のみ有することが推測されるか知られている。たとえば、非癌性細胞の遺伝型がＡＡで癌細胞がＡＢであり、ある試料においてその座位のシグナルのうち５％がＢアレルに由来し９５％がＡアレルに由来する場合、その試料の腫瘍率は１０％である。一部の態様では、癌細胞は非癌性細胞には存在しないアレルのコピーを２個有することが推測されるか知られている。たとえば、非癌性細胞の遺伝型がＡＡで癌細胞がＢＢであり、ある試料においてその座位のシグナルのうち５％がＢアレルに由来し９５％がＡアレルに由来する場合、その試料の腫瘍率は５％である。一部の態様では、癌細胞が非癌性細胞に存在しないアレルを有する複数の座位を分析し、その癌細胞ではそれらの座位のうちのどの座位がヘテロ接合型でありどの座位がホモ接合型であるかを決定する。たとえば、非癌性細胞がＡＡである座位の場合、Ｂアレルに由来するシグナルが一部の座位では約５％であり一部の座位では約１０％である場合、その癌細胞は、Ｂアレルが約５％である座位ではヘテロ接合型であり、Ｂアレルが約１０％である座位ではホモ接合型であると推測される（腫瘍率が約１０％であることが示される）。

【０４２３】

腫瘍率の決定に用いることのできる座位の例としては、癌細胞と非癌性細胞が１種類のアレルを共有する座位（癌細胞がＡＢであり非癌性細胞がＢＢであるか、癌細胞がＢＢであり非癌性細胞がＡＢである座位など）が挙げられる。混合試料（癌細胞と非癌性細胞に由来するＤＮＡまたはＲＮＡを含有する）におけるＡシグナルの量、Ｂシグナルの量、またはＡシグナルのＢシグナルに対する比を（ｉ）癌細胞のみに由来するＤＮＡまたはＲＮＡを含有する試料、あるいは（ｉｉ）非癌性細胞のみに由来するＤＮＡまたはＲＮＡを含有する試料の対応する値と比較し、値の差を用いてその混合試料の腫瘍率を決定する。

【０４２４】

一部の態様では、腫瘍率の決定に用いることのできる座位は、（ｉ）癌細胞のみに由来するＤＮＡまたはＲＮＡを含有する試料、および／あるいは（ｉｉ）非癌性細胞のみに由

来するＤＮＡまたはＲＮＡを含有する試料の遺伝型に基づいて選択される。一部の態様では、混合試料の分析に基づいて座位を選択する（各アレルの絶対量または相対量が、ある特定の座位において癌細胞も非癌性細胞も同一の遺伝型を有する場合に期待される量と異なる座位など）。たとえば、癌細胞と非癌性細胞が同一の遺伝型を有する場合、その座位は、全細胞がＡＡであれば０％のＢシグナル、全細胞がＡＢであれば５０％のＢシグナル、あるいは全細胞がＢＢであれば１００％のＢシグナルを発生すると期待される。Ｂシグナルの値が他の値であった場合、癌細胞と非癌性細胞の遺伝型がその座位では異なり、したがってその座位を腫瘍率の決定に用いることができることがわかる。

【０４２５】

一部の態様では、１つ以上の座位のアレルに基づいて算出した腫瘍率を、本明細書に開示する計測方法のうちの１つ以上を用いて算出した腫瘍率と比較する。

表現型の検出または複数の変異の分析のための方法の例

【０４２６】

一部の態様では、前記方法は、疾患または障害（癌など）あるいは疾患または障害のリスク増加に関連する変異の集合について試料を分析することを含む。ある方法のシグナル対ノイズ比を改善するのに使用でき、腫瘍を異なる臨床的部分集合に分類することができるクラス（ＭまたはＣなどの癌のクラス）間の事象には強い相関がある。たとえば、１つ以上の染色体または染色体分節の数個の変異（数個のＣＮＶなど）に関する境界の結果を複合的に考慮すると、非常に強いシグナルとなるかも知れない。一部の態様では、対象とする複数（たとえば２個、３個、４個、５個、８個、１０個、１２個、１５個、またはそれ以上）の多型または変異の有無を決定することによって、癌などの疾患または障害、あるいは癌などの疾患または障害のリスク増加の有無の決定の感度および／または特異度が高くなる。一部の態様では、複数の染色体にわたる事象間の相関を利用し、各染色体を個別に調べるよりも有効にシグナルを調べる。腫瘍の分類を最善に行うため、方法自体の設計を最適化することができる。このことは、１つの特定の変異／ＣＮＶに対する感度が最も高いと考えられる、再発の早期発見・検診に非常に有用となり得る。一部の態様では、事象は常に相関するのではなく、相関確率を有する。一部の態様では、非対角項を有するノイズ共分散行列を用いる行列による推定の構築が行われる。

【０４２７】

一部の態様では、本発明は個体の表現型（癌の表現型など）を検出するための方法を特徴とし、ここで、表現型は変異の集合のうちの少なくとも１つの変異の存在によって定義される。一部の態様では、前記方法は、個体から採取した１つ以上の細胞から得たＤＮＡまたはＲＮＡの試料に関するＤＮＡ測定値またはＲＮＡ測定値を得ること（ここで、前記細胞のうちの１つ以上は前記表現型を有すると考えられる）と；該ＤＮＡ測定値またはＲＮＡ測定値を分析することによって、前記変異の集合のうちの各変異について、該細胞のうちの少なくとも１つがその変異を有する尤度を決定することとを含む。一部の態様では、前記方法は、（ｉ）前記変異のうちの少なくとも１つについて、前記細胞のうちの少なくとも１つがその変異を有する尤度が閾値よりも高い場合、または（ｉｉ）該変異のうちの少なくとも１つについて、該細胞のうちの少なくとも１つがその変異を有する尤度が閾値よりも低く、複数の変異について、該細胞のうちの少なくとも１つが該変異のうちの少なくとも１つを有する結合尤度が閾値よりも高い場合に、前記個体が前記表現型を有すると決定することを含む。一部の態様では、１つ以上の細胞は、変異の集合に含まれる変異の部分集合またはすべての変異を有する。一部の態様では、変異の部分集合は癌または癌のリスク増加に関連する。一部の態様では、変異の集合は癌変異のＭクラスに含まれる変異の部分集合またはすべての変異を含む（Ciriello, Nat Genet. 45(10):1127-1133, 2013, doi: 10.1038/ng.2762（この文献の全内容は参照によって本明細書に組み込まれる））。一部の態様では、変異の集合は、癌変異のＣクラスに含まれる変異の部分集合またはすべての変異を含む（Ｃｉｒｉｅｌｌｏによる上記文献）。一部の態様では、試料は無細胞ＤＮＡまたは無細胞ＲＮＡを含む。一部の態様では、ＤＮＡ測定値またはＲＮＡ測定値は、対象とする１つ以上の染色体または染色体分節に存在する多型座位の集合での測定値

（各座位の各アレルの量など）を含む。

【０４２８】

実父確定検査または遺伝的関連性検査のための方法の例

本発明の方法は、実父確定検査またはその他の遺伝的関連性検査の精度を改善するために利用可能である（たとえば、２０１１年１２月２２日出願の米国特許出願公開第２０１２／０１２２７０１号明細書を参照；この文献の全内容は参照によって本明細書に組み込まれる）。たとえば多重ＰＣＲ法によって、本明細書に記載の、擬父が胎児の実父であるかを判定するPARENTAL SUPPORTアルゴリズムに使用するために数千の多型座位（ＳＮＰなど）を分析することが可能である。一部の態様では、本発明は、擬父が、妊娠中の母親が懐胎している胎児の実父であるかを確証する方法を特徴とする。一部の態様では、前記方法は、擬父に関する相決定済み遺伝データを得ることを含む（たとえば、本明細書に記載の別の方法を用いて遺伝データを相決定することによって）。ここで、相決定済み遺伝データは、擬父の第一の相同染色体分節と第二の相同染色体分節にある多型座位の集合の各座位に存在するアレルの同一性を含む。一部の態様では、前記方法は、胎児の母親から採取した、胎児のＤＮＡと母親のＤＮＡを含有するＤＮＡの混合試料に含まれる染色体または染色体分節にある多型座位の集合の遺伝データを、各座位での各アレルの量を測定することによって得ることを含む。一部の態様では、前記方法は、擬父に関する相決定済み遺伝データから、ＤＮＡの混合試料の期待遺伝データをコンピュータで算出することと；ＤＮＡの混合試料に関して得られた遺伝データをＤＮＡの混合試料の期待遺伝データと比較することによって、擬父が胎児の実父である確率をコンピュータで決定することと；決定された、擬父が胎児の実父である確率を用いて、擬父が胎児の実父であるか否かを確証することとを含む。一部の態様では、前記方法は、胎児の実母の相決定済み遺伝データを得ることを含む（たとえば本明細書に記載の別の方法を用いて遺伝データを相決定することによって）。ここで、相決定済み遺伝データは、母親の第一の相同染色体分節と第二の相同染色体分節にある多型座位の集合の各座位に存在するアレルの同一性を含む。一部の態様では、前記方法は、胎児に関する相決定済み遺伝データを得ることを含む（たとえば本明細書に記載の別の方法を用いて遺伝データを相決定することによって）。ここで、相決定済み遺伝データは、胎児の第一の相同染色体分節と第二の相同染色体分節にある多型座位の集合の各座位に存在するアレルの同一性を含む。一部の態様では、前記方法は、擬父に関する相決定済み遺伝データを用い、かつ母親に関する相決定済み遺伝データおよび／または胎児に関する相決定済み遺伝データを用いて、ＤＮＡの混合試料の期待される遺伝データをコンピュータで算出することを含む。

【０４２９】

一部の態様では、本発明は、擬父が、妊娠中の母親が懐胎している胎児の実父であるかを確証する方法を特徴とする。一部の態様では、前記方法は、擬父に関する相決定済み遺伝データを得ることを含む（たとえば本明細書に記載の別の方法を用いて遺伝データを相決定することによって）。ここで、相決定済み遺伝データは、擬父の第一の相同染色体分節と第二の相同染色体分節にある多型座位の集合の各座位に存在するアレルの同一性を含む。一部の態様では、前記方法は、胎児の母親から採取した、胎児のＤＮＡと母親のＤＮＡを含有するＤＮＡの混合試料に含まれる染色体または染色体分節にある多型座位の集合の遺伝データを、各座位での各アレルの量を測定することによって得ることを含む。一部の態様では、前記方法は、（ｉ）多型座位において胎児のＤＮＡには存在するが母親のＤＮＡには存在しないアレルを確認すること、および／または（ｉ）多型座位において胎児のＤＮＡにも母親のＤＮＡにも存在しないアレルを確認することを含む。一部の態様では、前記方法は擬父が胎児の実父である確率をコンピュータで決定することであって、該決定は（１）（ｉ）多型座位において胎児のＤＮＡには存在するが母親のＤＮＡには存在しないアレルを（ｉｉ）擬父から得た遺伝物質中の対応する多型座位に存在するアレルと比較すること、および／または（２）（ｉ）多型座位において胎児のＤＮＡにも母親のＤＮＡにも存在しないアレルを（ｉｉ）擬父から得た遺伝物質中の対応する多型座位に存在するアレルと比較し、擬父が胎児の実父である決定された確率を用いて擬父が胎児の実父で

あるかどうかを確認することを含む。

【０４３０】

一部の態様では、擬父が胎児の実父であるかどうかを決定するための上記方法を用いて、胎児の血縁者（祖父母、兄弟姉妹、おじまたはおばなど）とされる者が実際に胎児の実の血縁者であるかを決定する（たとえば擬父の遺伝データの代わりに血縁者とされる者の遺伝データを用いることによって）。

【０４３１】

方法の組み合わせの例

結果の精度を上げるため、ＣＮＶの有無を検出するための２つ以上の方法（本発明の方法のうちの任意の方法または任意の既知の方法など）を実行する。一部の態様では、疾患または障害あるいは疾患または障害のリスク増加の有無を示す因子を分析するための１つ以上の方法（本明細書に記載の方法のうちの任意の方法または任意の既知の方法など）を実行する。

【０４３２】

一部の態様では、標準の数学的手法を用いて２つ以上の方法の間の共分散および／または相関を算出する。標準の数学的手法を用いて２つ以上の試験に基づく特定の仮説の複合確率を決定してもよい。手法の例としては、メタ解析、独立な検定に関するＦｉｓｈｅｒの複合確率検定、既知の共分散を用いて非独立なｐ値を統合するＢｒｏｗｎの方法、未知の共分散を用いて非独立なｐ値を統合するＫｏｓｔの方法が挙げられる。第一の方法によって決定した尤度が第二の方法に関して決定した尤度と直交する、すなわち非関連である場合、尤度の結合は単に乗算と正規化によって、すなわち以下のような式を用いることによって可能である：

$R_{comb} = R_1 R_2 / [R_1 R_2 + (1 - R_1)(1 - R_2)]$

R_{comb}は結合尤度であり、R_1およびR_2は個別尤度である。たとえば、方法１による三染色体性の尤度が９０％であり、方法２による三染色体性の尤度が９５％である場合、この２つの方法の結果を組み合わせることにより、臨床医は、胎児が（０．９０）（０．９５）／［（０．９０）（０．９５）＋（１－０．９０）（１－０．９５）］＝９９．４２％の尤度で三染色体性であると結論付けることができる。第一の方法と第二の方法が直交でない場合、すなわち２つの方法間に相関がある場合も尤度を結合することができる。

【０４３３】

複数の因子または変数を分析する方法の例は、２０１１年９月２０日発行の米国特許第８，０２４，１２８号明細書；２００６年７月３１日出願の米国特許出願公開第２００７／００２７６３６号明細書；２００６年１２月６日出願の米国特許出願公開第２００７／０１７８５０１号明細書（これらの文献それぞれの全内容は参照によって本明細書に組み込まれる）に開示されている。

【０４３４】

様々な態様では、特定の仮説または診断の複合確率は８０％超、８５％超、９０％超、９２％超、９４％超、９６％超、９８％超、９９％超、または９９．９％超であるか、その他の何らかの閾値よりも高い。

検出限界

【０４３５】

一部の態様では、本発明の方法の変異（ＳＮＶまたはＣＮＶなど）の検出限界は、１０％以下、５％以下、２％以下、１％以下、０．５％以下、０．１％以下、０．０５％以下、０．０１％以下、または０．００５％以下である。一部の態様では、本発明の方法の変異（ＳＮＶまたはＣＮＶなど）の検出限界は１５％～０．００５％であり、たとえば１０％～０．００５％、１０％～０．０１％、１０％～０．１％、５％～０．００５％、５％～０．０１％、５％～０．１％、１％～０．００５％、１％～０．０１％、１％～０．１％、０．５％～０．００５％、０．５％～０．０１％、０．５％～０．１％、または０．１～０．０１（それぞれ、範囲の上限値と下限値を含む）。一部の態様では、検出限界は

、試料（無細胞ＤＮＡまたは無細胞ＲＮＡの試料など）に含まれる、その座位を有するＤＮＡ分子またはＲＮＡ分子のうち１０％以下、５％以下、２％以下、１％以下、０．５％以下、０．１％以下、０．０５％以下、０．０１％以下、または０．００５％以下に存在する変異（ＳＮＶまたはＣＮＶなど）が検出される（または検出可能である）ような検出限界である。たとえば、その座位を有するＤＮＡ分子またはＲＮＡ分子のうち１０％以下、５％以下、２％以下、１％以下、０．５％以下、０．１％以下、０．０５％以下、０．０１％以下、または０．００５％以下が座位に変異を有する場合でも、（たとえばその座位の野生型すなわち非変異型またはその座位にある別の変異ではなく）その変異を検出することができる。一部の態様では、検出限界は、試料（無細胞ＤＮＡまたは無細胞ＲＮＡの試料など）に含まれるＤＮＡ分子またはＲＮＡ分子のうち１０％以下、５％以下、２％以下、１％以下、０．５％以下、０．１％以下、０．０５％以下、０．０１％以下、または０．００５％以下に存在する変異（ＳＮＶまたはＣＮＶなど）が検出される（または検出可能である）ような検出限界である。ＣＮＶが欠失である一部の態様では、試料に含まれる、欠失を含むか含まないかわからない対象領域を有するＤＮＡ分子またはＲＮＡ分子のうち１０％以下、５％以下、２％以下、１％以下、０．５％以下、０．１％以下、０．０５％以下、０．０１％以下、または０．００５％以下にしか欠失が存在しない場合でも、その欠失を検出することができる。ＣＮＶが欠失である一部の態様では、試料に含まれるＤＮＡ分子またはＲＮＡ分子のうち１０％以下、５％以下、２％以下、１％以下、０．５％以下、０．１％以下、０．０５％以下、０．０１％以下、または０．００５％以下にしか欠失が存在しない場合でも、その欠失を検出することができる。ＣＮＶが重複である一部の態様では、存在する余分な重複ＤＮＡまたはＲＮＡが、試料に含まれる、重複があるかないかわからない対象領域を有するＤＮＡ分子またはＲＮＡ分子のうち１０％以下、５％以下、２％以下、１％以下、０．５％以下、０．１％以下、０．０５％以下、０．０１％以下、または０．００５％以下である場合でも、その重複を検出することができる。ＣＮＶが重複である一部の態様では、存在する余分な重複ＤＮＡまたはＲＮＡが、試料に含まれるＤＮＡ分子またはＲＮＡ分子のうち１０％以下、５％以下、２％以下、１％以下、０．５％以下、０．１％以下、０．０５％以下、０．０１％以下、または０．００５％以下である場合でも、その重複を検出することができる。実施例６に検出限界の算出方法の例を示す。一部の態様では、実施例６の「ＬＯＤ－ｚｓ５．０－ｍｒ５」の方法を用いる。

【０４３６】

試料の例

本発明のいずれかの側面の態様では、試料は欠失または重複を有すると考えられる細胞（癌細胞だと考えられる細胞など）に由来する細胞内および／または細胞外の遺伝物質を含む。一部の態様では、試料は、欠失または重複を有する細胞、ＤＮＡ、またはＲＮＡ（癌の細胞、ＤＮＡ、またはＲＮＡなど）を含むと考えられる任意の組織または体液を含む。ＤＮＡまたはＲＮＡを含む任意の試料について、これらの方法の一部として用いる遺伝学的測定を行うことができる。試料の例としては、たとえば、組織、血液、血清、血漿、尿、毛髪、涙、唾液、皮膚、指の爪、糞便、胆汁、リンパ液、頸管粘液、精液、あるいは、核酸を含むその他の細胞または物質が挙げられるが、これらに限定されるものではない。試料は任意の細胞型を含んでもよく、任意の細胞型に由来するＤＮＡまたはＲＮＡを用いてもよい（癌を有すると考えられる任意の器官または組織の細胞、あるいはニューロンなど）。一部の態様では、試料は核内ＤＮＡおよび／またはミトコンドリアＤＮＡを含む。一部の態様では、試料は本明細書に開示された任意の対象個体から得られる。一部の態様では、対象個体は生後の個体、懐胎中の胎児、受胎産物試料などの懐胎中でない胎児、胎芽、またはその他の任意の個体である。

【０４３７】

試料の例としては、無細胞ＤＮＡまたは無細胞ＲＮＡを含有するものが挙げられる。一部の態様では、細胞を破砕する工程を必要とせずに分析するために無細胞ＤＮＡを利用することができる。無細胞ＤＮＡは多様な組織（たとえば血液、血漿、リンパ液、腹水、脳

脊髄液などの液状組織）から入手可能である。無細胞ＤＮＡは胎児の細胞に由来するＤＮＡで構成される場合があり、胎児の細胞に由来するＤＮＡと母親の細胞に由来するＤＮＡで構成される場合もある。全血から遠心分離によって細胞物質を除去して単離した血漿から無細胞ＤＮＡを単離する場合もある。無細胞ＤＮＡは、標的細胞（癌細胞など）に由来するＤＮＡと非標的細胞（非癌性細胞など）に由来するＤＮＡの混合物であってもよい。

【０４３８】

一部の態様では、試料は、たとえば癌のＤＮＡ（またはＲＮＡ）と非癌性のＤＮＡ（またはＲＮＡ）の混合物などのＤＮＡ（またはＲＮＡ）の混合物を含有するないしは含有すると考えられる。一部の態様では、試料に含まれる細胞の少なくとも０．５％、１％、３％、５％、７％、１０％、１５％、２０％、３０％、４０％、５０％、６０％、７０％、８０％、９０％、９２％、９４％、９５％、９６％、９８％、９９％、または１００％が癌細胞である。一部の態様では、試料に含まれるＤＮＡ（無細胞ＤＮＡなど）またはＲＮＡ（無細胞ＲＮＡなど）の少なくとも０．５％、１％、３％、５％、７％、１０％、１５％、２０％、３０％、４０％、５０％、６０％、７０％、８０％、９０％、９２％、９４％、９５％、９６％、９８％、９９％、または１００％が癌細胞に由来する。様々な態様では、試料に含まれる癌細胞の割合は０．５％～９９％であり、たとえば、１％～９５％、５％～９５％、１０％～９０％、５％～７０％、１０％～７０％、２０％～９０％、または２０％～７０％である（それぞれ、範囲の上限値と下限値を含む）。一部の態様では、試料中の癌細胞あるいは癌細胞由来のＤＮＡまたはＲＮＡを濃縮する。試料中の癌細胞を濃縮する一部の態様では、濃縮した試料に含まれる細胞の少なくとも０．５％、１％、２％、３％、４％、５％、６％、７％、１０％、１５％、２０％、３０％、４０％、５０％、６０％、７０％、８０％、９０％、９２％、９４％、９５％、９６％、９８％、９９％、または１００％が癌細胞である。試料中の癌細胞由来のＤＮＡまたはＲＮＡを濃縮する一部の態様では、濃縮した試料に含まれるＤＮＡまたはＲＮＡの少なくとも０．５％、１％、２％、３％、４％、５％、６％、７％、１０％、１５％、２０％、３０％、４０％、５０％、６０％、７０％、８０％、９０％、９２％、９４％、９５％、９６％、９８％、９９％、または１００％が癌細胞に由来する。一部の態様では、細胞選別（蛍光活性化細胞選別（Fluorescent Activated Cell Sorting; FACS）など）によって癌細胞を濃縮する（Barteneva et. al., Biochim Biophys Acta., 1836(1):105-22, Aug 2013. doi: 10.1016/j.bbcan.2013.02.004. Epub 2013 Feb 24；およびIbrahim et al., Adv Biochem Eng Biotechnol. 106:19-39, 2007（これらの文献それぞれの全内容は参照によって本明細書に組み込まれる））。

【０４３９】

本発明のいずれかの側面の一部の態様では、試料は少なくとも部分的に胎児に由来すると考えられる任意の組織を含む。一部の態様では、試料は胎児に由来する細胞内および／または細胞外の遺伝物質、不純物である細胞内および／または細胞外の遺伝物質（胎児の母親に由来する遺伝物質など）、またはその組み合わせを含む。一部の態様では、試料は胎児に由来する細胞内遺伝物質、不純物である細胞内遺伝物質、またはその組み合わせを含む。

【０４４０】

一部の態様では、試料は懐胎中の胎児に由来する。一部の態様では、試料は、受胎産物試料、または胎児の死亡後に任意の胎児の組織から得た試料など、懐胎中でない胎児に由来する試料である。一部の態様では、試料は母親の全血試料、母親の血液試料から単離した細胞、母親の血漿試料、母親の血清試料、羊水穿刺による試料、胎盤組織試料（たとえば絨毛、脱落膜、または胎盤膜）、頸管粘液試料、あるいは胎児に由来するその他の試料である。一部の態様では、試料に含まれる細胞の少なくとも３％、５％、７％、１０％、１５％、２０％、３０％、４０％、５０％、６０％、７０％、８０％、９０％、９２％、９４％、９５％、９６％、９８％、９９％、または１００％が母親の細胞である。様々な態様では、試料に含まれる母親の細胞の割合は５％～９９％であり、たとえば１０％～９５％、２０％～９５％、３０％～９０％、３０％～７０％、４０％～９０％、４０％～７

０％、５０％～９０％、または５０％～８０％である（それぞれ、範囲の上限値と下限値を含む）。

【０４４１】

一部の態様では、試料中の胎児の細胞を濃縮する。試料中の胎児の細胞を濃縮する一部の態様では、濃縮した試料に含まれる細胞の少なくとも０．５％、１％、２％、３％、４％、５％、６％、７％、またはそれ以上が胎児の細胞である。一部の態様では、試料に含まれる胎児の細胞の割合は０．５％～１００％であり、たとえば１％～９９％、５％～９５％、１０％～９５％、１０％～９５％、２０％～９０％、または３０％～７０％である（それぞれ、範囲の上限値と下限値を含む）。一部の態様では、試料中の胎児のＤＮＡを濃縮する。試料中の胎児のＤＮＡを濃縮する一部の態様では、濃縮した試料に含まれるＤＮＡの少なくとも０．５％、１％、２％、３％、４％、５％、６％、７％、またはそれ以上が胎児のＤＮＡである。一部の態様では、試料に含まれる胎児のＤＮＡの割合は０．５％～１００％であり、たとえば１％～９９％、５％～９５％、１０％～９５％、１０％～９５％、２０％～９０％、または３０％～７０％である（それぞれ、範囲の上限値と下限値を含む）。

【０４４２】

一部の態様では、試料は単一細胞を含むか、単一細胞に由来するＤＮＡおよび／またはＲＮＡを含む。一部の態様では、複数の個体細胞（たとえば同一の対象または異なる対象から得た少なくとも５個、１０個、２０個、３０個、４０個、または５０個の細胞）を並行して分析する。一部の態様では、同一の個体から採取した複数の試料から得た細胞を組み合わせる。それによって、試料を個々に分析するのに比べて作業量が減る。複数の試料を組み合わせることにより、複数の組織の癌検査を同時に行うこともできる（これを利用して、より徹底的な癌検診を提供したり、癌が他の組織に転移した可能性があるかどうかを決定したりすることができる）。

【０４４３】

一部の態様では、試料は単一細胞または少数の細胞を含む（たとえば２個、３個、５個、６個、７個、８個、９個、または１０個）。一部の態様では、試料は１個～１００個、１００個～５００個、または５００個～１，０００個の細胞を含む（それぞれ、範囲の上限値と下限値を含む）。一部の態様では、試料は１ピコグラム～１０ピコグラム、１０ピコグラム～１００ピコグラム、１００ピコグラム～１ナノグラム、１ナノグラム～１０ナノグラム、１０ナノグラム～１００ナノグラム、または１００ナノグラム～１マイクログラムのＲＮＡおよび／またはＤＮＡを含む（それぞれ、範囲の上限値と下限値を含む）。

【０４４４】

一部の態様では、試料をパラフィンフィルムに包埋する。一部の態様では、試料をホルムアルデヒドなどの防腐剤と共に保存し、必要に応じてパラフィンに包んでもよいが、ＤＮＡの架橋を起こし、ＰＣＲに利用できるＤＮＡが減る場合がある。一部の態様では、試料はホルムアルデヒド固定パラフィン包埋（formaldehyde fixed-paraffin embedded; FFPE）試料である。一部の態様では、試料は新たに採取した試料（たとえば分析の１日～２日前に採取した試料）である。一部の態様では、試料を分析まで凍結しておいてもよい。一部の態様では、試料は過去に採取した試料である。

【０４４５】

上記の試料を本発明の任意の方法に用いることができる。

【０４４６】

試料調製方法の例

一部の態様では、前記方法は、ＤＮＡおよび／またはＲＮＡの単離または精製を含む。この目的は当技術分野で公知の複数の標準的な方法によって達成される。一部の態様では、試料を遠心分離して様々な層を分離してもよい。一部の態様では、ＤＮＡまたはＲＮＡを濾過によって単離してもよい。一部の態様では、ＤＮＡまたはＲＮＡの調製は、増幅、分離、クロマトグラフィーによる精製、液液分離、単離、選択的濃縮、選択的増幅、標的化増幅、あるいは当技術分野で公知ないしは本明細書に記載の複数のその他の技術のいず

れかを含んでもよい。ＤＮＡの単離に関する一部の態様では、ＲＮＡ分解酵素を用いてＲＮＡを分解する。ＲＮＡの単離に関する一部の態様では、ＤＮＡ分解酵素（Invitrogen（Carlsbad, CA, USA）のDNAse Iなど）を用いてＤＮＡを分解する。一部の態様では、RNeasy Mini Kit（Qiagen）を用いて、製造者のプロトコルに従いＲＮＡを単離する。一部の態様では、低分子ＲＮＡをmirVana（商標）PARIS（商標） Kit（Ambion（Austin, TX, USA））を用いて、製造者のプロトコルに従い単離する（Gu et al., J. Neurochem. 122:641-649, 2012；この文献の全内容は参照によって本明細書に組み込まれる））。必要に応じてＲＮＡの濃度および純度をNanoVue（商標）（GE Healthcare（Piscataway, NJ, USA）で決定してもよく、必要に応じてＲＮＡの完全性を2100 Bioanalyzer（Agilent Technologies（Santa Clara, CA, USA））で測定してもよい（Gu et al., J. Neurochem. 122:641-649, 2012；この文献の全内容は参照によって本明細書に組み込まれる）。一部の態様では、TRIzol（商標）またはRNAlater（登録商標）（Ambion）を用いて保存中のＲＮＡを安定化する。

【０４４７】

一部の態様では、汎用タグを付けたアダプターを付加してライブラリを作製する。ライゲーションの前に試料ＤＮＡを平滑末端化してもよく、次いで１個のアデノシン塩基を３’末端に付加する。ライゲーションの前にＤＮＡを制限酵素またはその他の何らかの切断方法によって切断してもよい。ライゲーションの際、試料断片の３’末端のアデノシンと、アダプターの相補的な３’末端のチロシンの突出により、ライゲーション効率を高めることができる。一部の態様では、Agilent SureSelectのライゲーションキットを用いてアダプターのライゲーションを行う。一部の態様では、汎用プライマーを用いてライブラリを増幅する。ある態様では、サイズ分離によって、あるいはAGENCOURT（登録商標）AMPURE（登録商標）ビーズなどの製品またはその他の類似の方法によって、増幅したライブラリを分画する。一部の態様では、ＰＣＲ増幅によって標的座位を増幅する。一部の態様では、増幅したＤＮＡを配列決定する（たとえばILLUMINA（登録商標）のGAIIXシーケンサまたはHiSeq（商標）シーケンサを用いる配列決定など）。一部の態様では、配列決定エラーを低減するため、増幅後のＤＮＡの両末端から配列決定を行う。増幅後のＤＮＡの一方の末端から配列決定を行ったときに特定の塩基に配列エラーがある場合、（増幅後のＤＮＡの配列決定を同じ末端から複数回行う場合に比べて）増幅後のＤＮＡのもう一方の末端から配列決定を行った場合に相補性の塩基に配列エラーが起こる可能性がより低くなる。

【０４４８】

一部の態様では、全ゲノム増幅（ＷＧＡ）によって核酸試料を増幅する。ＷＧＡに利用可能な方法は複数ある：ライゲーションを介したＰＣＲ（ligation-mediated PCR（LM-PCR））、縮重オリゴヌクレオチドプライマーＰＣＲ（degenerate oligonucleotide primer PCR（DOP-PCR））、多重置換増幅（multiple displacement amplification（MDA））である。ＬＭ－ＰＣＲでは、アダプターと呼ばれる短いＤＮＡ配列をライゲートすることにより、ＤＮＡの末端を平滑化する。これらのアダプターは、汎用的な増幅配列を含み、これを用いてＰＣＲによりＤＮＡを増幅する。ＤＯＰ－ＰＣＲでは、同じく汎用的な増幅配列を含むランダムプライマーをアニーリングと初回のＰＣＲで用いる。次いで第２回のＰＣＲにより、汎用プライマー配列をさらに用いて配列を増幅する。ＭＤＡではｐｈｉ－２９ポリメラーゼを用いる。ｐｈｉ－２９ポリメラーゼはＤＮＡを複製し、単一細胞の分析に用いられてきた、加工性が高く非特異的な酵素である。一部の態様では、ＷＧＡを行わない。

【０４４９】

一部の態様では、選択的な増幅または濃縮により、標的座位を増幅または濃縮する。一部の態様では、増幅および／または選択的濃縮技術は、ライゲーションを介したＰＣＲなどのＰＣＲ、ハイブリダイゼーションや分子反転プローブまたはその他の環状化プローブによる断片の捕捉を含んでもよい。一部の態様では、リアルタイム定量ＰＣＲ（real-time quantitative PCR = RT-qPCR）、デジタルＰＣＲ、またはエマルジョンＰＣＲ、単一ア

レルの塩基伸長反応（single allele base extension reaction）の後に質量分析を行う（Hung et al., J Clin Pathol 62:308-313, 2009；この文献の全内容は参照によって本明細書に組み込まれる）。一部の態様では、ＤＮＡの選択的濃縮にハイブリッド捕捉プローブとのハイブリダイゼーションによる捕捉を用いる。一部の態様では、増幅または選択的濃縮のための方法は、標的配列と正しくハイブリダイズした際に少数のヌクレオチドによってヌクレオチドプローブの３’末端または５’末端が多型アレルの多型部位から隔離されるようなプローブを使用することを含んでもよい。この隔離により、一方のアレルの優先的増幅（アレルの偏りという）が減少する。これは、正しくハイブリダイズしたプローブの３’末端または５’末端がアレルの多型部位と直接隣接するか非常に近くなるようなプローブを使用することを含む方法に対する改善点である。一態様では、ハイブリダイズする領域が多型部位を含む可能性があるか、含むことが確実であるプローブを排除する。ハイブリダイゼーションの部位に多型部位があることにより、一部のアレルで不均等なハイブリダイゼーションが起こるか、全体のハイブリダイゼーションが阻害され、結果としてあるアレルが優先的に増幅される可能性がある。これらの態様は、試料が単一の個体から得た純粋なゲノム試料であっても個体の混合物であっても、試料の各多型座位での元のアレル頻度をより良好に保つ点で、標的化増幅および／または選択的濃縮を含むその他の方法に対する改善である。

【０４５０】

一部の態様では、ＰＣＲ（ミニＰＣＲという）を用いて非常に短い増幅産物を生成する（２０１２年１１月２１日出願の米国特許出願第１３／６８３，６０４号（米国特許出願公開第２０１３／０１２３１２０号）明細書、２０１１年１１月１８日出願の米国特許出願第１３／３００，２３５号（米国特許出願公開第２０１２／０２７０２１２号）明細書、２０１４年５月１６日出願の米国特許出願第６１／９９４，７９１号明細書（これらの文献それぞれの全内容は参照によって本明細書に組み込まれる））。無細胞ＤＮＡ（母親の血清中の胎児の無細胞ＤＮＡ、壊死またはアポトーシスによって放出された癌の無細胞ＤＮＡなど）は高度に断片化される。胎児の無細胞ＤＮＡの場合、断片のサイズは概ねガウス関数的に分布する（平均１６０ｂｐ、標準偏差１５ｂｐ、最小サイズ約１００ｂｐ、最大サイズ約２２０ｂｐ）。１つの特定の標的座位の多型部位は、その座位に由来する多様な断片の開始位置から終位置までの任意の位置を占め得る。無細胞ＤＮＡ断片は短いため、両方のプライマー部位が存在する尤度、すなわち順方向プライマー部位と逆方向プライマー部位両方を含む長さＬの断片の尤度は、増幅産物の長さと断片の長さの比である。理想的な条件では、増幅産物が４５ｂｐ、５０ｂｐ、５５ｂｐ、６０ｂｐ、６５ｂｐ、または７０ｂｐである試験法は、それぞれ、利用可能な鋳型断片分子の７２％、６９％、６６％、６３％、５９％、５６％からの増幅に成功する。癌を有すると考えらえる個体から採取した試料から得た無細胞ＤＮＡの場合に最も好ましいある態様では、無細胞ＤＮＡは、最大の増幅産物長が８５ｂｐ、８０ｂｐ、７５ｂｐ、または７０ｂｐ（ある好ましい態様では７５ｂｐ）であり、かつ、融解温度が５０℃～６５℃（ある好ましい態様では５４℃～６０．５℃）であるプライマーを用いて増幅される。増幅産物長は順方向プライミング部位の５’末端と逆方向プライミング部位の５’末端の間の距離である。当業者が一般的に用いるものよりも増幅産物長が短いことにより、短い配列リードで済むため、所望の多型座位の測定をより効率的にすることができる。一態様では、増幅産物の実質的な画分は１００ｂｐ未満、９０ｂｐ未満、８０ｂｐ未満、７０ｂｐ未満、６５ｂｐ未満、６０ｂｐ未満、５５ｂｐ未満、５０ｂｐ未満、または４５ｂｐ未満である。

【０４５１】

一部の態様では、直接多重ＰＣＲ、逐次ＰＣＲ、入れ子ＰＣＲ、二重入れ子ＰＣＲ（doubly nested PCR）、片側入れ子・片側半入れ子ＰＣＲ（one-and-a-half sided nested PCR）、完全入れ子ＰＣＲ（fully nested PCR）、片側完全入れ子ＰＣＲ（one sided fully nested PCR）、片側入れ子ＰＣＲ（one-sided nested PCR）、半入れ子ＰＣＲ（hemi-nested PCR）、半入れ子ＰＣＲ、三重半入れ子ＰＣＲ（triply hemi-nested PCR）、準入れ子ＰＣＲ（semi-nested PCR）、片側準入れ子ＰＣＲ（one sided semi-nested PCR）、

逆方向準入れ子ＰＣＲ（reverse semi-nested PCR）法、または片側ＰＣＲを用いて増幅を行う。これらのＰＣＲ法は以下の文献に記載されている：２０１２年１１月２１日出願の米国特許出願第１３／６８３，６０４号（米国特許出願公開第２０１３／０１２３１２０号）明細書；２０１１年１１月１８日出願の米国特許出願第１３／３００，２３５号（米国特許出願公開第２０１２／０２７０２１２号）明細書；２０１４年5月１６日出願の米国特許出願第６１／９９４，７９１号明細書（これらの文献の全内容は参照によって組み込まれる）。必要に応じて、これらの方法のいずれかをミニＰＣＲに用いることができる。

【０４５２】

必要に応じ、時間の観点から、長さが２００ヌクレオチド超、３００ヌクレオチド超、４００ヌクレオチド超、５００ヌクレオチド超、または１，０００ヌクレオチド超の断片からの増幅を減らすようにＰＣＲ増幅の伸長工程を制限してもよい。このことによって、断片化されたＤＮＡまたはより短いＤＮＡ（胎児のＤＮＡ、あるいはアポトーシスまたは壊死を経た癌細胞から得たＤＮＡなど）を濃縮でき、試験性能を改善することができる。

【０４５３】

一部の態様では多重ＰＣＲを用いる。一部の態様では、核酸試料中の標的座位を増幅する方法は、（ｉ）核酸試料を、少なくとも１００個、２００個、５００個、７５０個、１，０００個、２，０００個、５，０００個、７，５００個、１０，０００個、２０，０００個、２５，０００個、３０，０００個、４０，０００個、５０，０００個、７５，０００個、または１００，０００個の異なる標的座位と同時にハイブリダイズするプライマーのライブラリと接触させて反応混合物を生成することと；（ｉｉ）反応混合物をプライマー伸長反応条件（たとえばＰＣＲ条件）に供して標的増幅産物を含む増幅産物を生成することを含む。一部の態様では、標的座位の少なくとも５０％、６０％、７０％、８０％、９０％、９５％、９６％、９７％、９８％、９９％、または９９．５％を増幅する。様々な態様では、増幅産物の６０％未満、５０％未満、４０％未満、３０％未満、２０％未満、１０％未満、５％未満、４％未満、３％未満、２％未満、１％未満、０．５％未満、０．２５％未満、０．１％未満、または０．０５％未満がプライマー二量体である。一部の態様では、プライマーは溶液状である（たとえば固相でなく液相に溶解している）。一部の態様では、プライマーは溶液状であり、固体担体に固定されない。一部の態様では、プライマーはマイクロアレイの一部ではない。一部の態様では、プライマーは分子反転プローブ（ＭＩＰ）を含まない。

【０４５４】

一部の態様では、２個以上（たとえば３個～４個）の標的増幅産物（たとえば本明細書に開示したミニＰＣＲ法で得た増幅産物）を共にライゲートし、次いでライゲーション後の産物を配列決定する。複数の増幅産物を組み合わせて１つのライゲーション産物を得ることにより、後の配列決定工程の効率が向上する。一部の態様では、標的増幅産物の長さはライゲーション前で１５０塩基対未満、１００塩基対未満、９０塩基対未満、７５塩基対未満、または５０塩基対未満である。選択的濃縮および／または増幅は、各個体分子に異なるタグ、分子バーコード、増幅用タグ、および／または配列決定用タグをつけることを含んでもよい。一部の態様では、増幅産物を配列決定（たとえば高処理配列決定による）、またはＳＮＰアレイ、ILLUMINA（登録商標）INFINIUM（登録商標）アレイ、AFFIMETRIX GeneChip（登録商標）などのアレイにハイブリダイズすることによって分析する。一部の態様では、Genia社が開発したナノポア配列決定技術（たとえばワールドワイドウェブのgeniachip.com/technologyを参照；この全内容は参照によって本明細書に組み込まれる）などのナノポア配列決定を用いる。一部の態様では二重鎖配列決定を用いる（Schmitt et al., "Detection of ultra-rare mutations by next-generation sequencing," Proc Natl Acad Sci U S A. 109(36): 14508-14513, 2012；この文献の全内容は参照によって本明細書に組み込まれる）。この方法により、ＤＮＡ二重鎖の二本鎖の各々を個々にタグを付けて配列決定することによるエラーが大きく低減される。二本鎖は相補的であるため、実変異は両方の鎖の同じ位置に見られる。これに対し、ＰＣＲまたは配列決定のエラ

ーの場合、一方の鎖のみに変異があることになるため、技術的なエラーとして無視することができる。一部の態様では、方法は、ランダムであるが相補的な二本鎖ヌクレオチド配列（Duplex Tagという）で二本鎖ＤＮＡの両方の鎖を標識付けすることを伴う。まずランダム化された一本鎖ヌクレオチド配列を１つのアダプター鎖に導入し、次いで反対の鎖をＤＮＡポリメラーゼで伸長して相補的な二本鎖タグを生成することによって、二本鎖タグ配列を標準的な配列決定アダプターに組み込む。タグを付けたアダプターを、切断されたＤＮＡにライゲートした後、個々に標識を付けた鎖をアダプターの両端の非対称プライマー部位からＰＣＲ増幅し、ペアエンド配列決定を行う。一部の態様では、試料（ＤＮＡ試料またはＲＮＡ試料など）を異なるウェル（たとえばWaferGen SmartChip（商標））などの複数の画分に分ける。試料を異なる画分（たとえば少なくとも５個、１０個、２０個、５０個、７５個、１００個、１５０個、２００個、または３００個の画分）に分けることにより、変異を有する分子の割合が、一部のウェルでは試料全体に比べて高くなるため、分析の感度を向上することができる。一部の態様では、各画分は５００個未満、４００個未満、２００個未満、１００個未満、５０個未満、２０個未満、１０個未満、５個未満、２個未満、または１個未満のＤＮＡ分子またはＲＮＡ分子を有する。一部の態様では、各画分に含まれる分子を個別に配列決定する。一部の態様では、同じバーコード（ランダムな配列またはヒト以外の配列など）を同じ画分のすべての分子に付加し（たとえばそのバーコードを含むプライマーを用いた増幅、あるいはバーコードをライゲートすることによって）、異なる画分の分子には異なるバーコードを付加する。バーコードを付けた分子をプールしておき、まとめて配列決定することができる。一部の態様では、分子を入れ子ＰＣＲなどによって増幅してからプールし、配列決定する。一部の態様では、１個の順方向プライマーと２個の逆方向プライマー、または２個の順方向プライマーと１個の逆方向プライマーを用いる。

【０４５５】

一部の態様では、試料（無細胞ＤＮＡまたは無細胞ＲＮＡの試料など）に含まれるＤＮＡ分子またはＲＮＡ分子の１０％未満、５％未満、２％未満、１％未満、０．５％未満、０．１％未満、０．０５％未満、０．０１％未満、または０．００５％未満に存在する変異（ＳＮＶまたはＣＮＶなど）が検出される（ないしは検出可能である）。一部の態様では、試料（たとえば血液試料から得た無細胞ＤＮＡまたは無細胞ＲＮＡの試料など）に含まれる１，０００個未満、５００個未満、１００個未満、５０個未満、２０個未満、１０個未満、５個未満、４個未満、３個未満、または２個未満の元のＤＮＡ分子またはＲＮＡ分子（増幅前）に存在する変異（ＳＮＶまたはＣＮＶ）が検出される（ないしは検出可能である）。一部の態様では、試料（たとえば血液試料から得た無細胞ＤＮＡまたは無細胞ＲＮＡなど）に含まれる元のＤＮＡ分子またはＲＮＡ分子（増幅前）１個のみに存在する変異（ＳＮＶまたはＣＮＶなど）が検出される（ないしは検出可能である）。

【０４５６】

たとえば、変異（一塩基多様体（ＳＮＶ）など）の検出限界が０．１％である場合、０．０１％に存在する変異は、画分を１００個のウェルなどの複数の画分に分けることによって検出可能である。ほとんどのウェルは変異のコピーを有していない。変異を有するわずかなウェルでは、リードのうちのはるかに高い割合で変異が存在する。一例では、標的座位のＤＮＡの初期のコピーは２０，０００個あり、これらのコピーのうち２個が対象のＳＮＶを含む。試料を１００個のウェルに分けた場合、９８個のウェルがＳＮＶを有し、２個のウェルが０．５％でＳＮＶを有する。各ウェル内のＤＮＡにバーコードを付け、増幅し、他のウェルのＤＮＡと共にプールし、配列決定することができる。ＳＮＶを有していないウェルを用いて、背景の増幅／配列決定エラーの割合を測定して異常値ウェルからのシグナルがノイズの背景レベルを超えているかどうか決定することができる。

【０４５７】

一部の態様では、１つ以上の対象染色体（たとえば１３番染色体、１８番染色体、２１番染色体、Ｘ染色体、Ｙ染色体、またはそれらの任意の組み合わせ）に対するプローブを有するアレイ（特にマイクロアレイ）などのアレイを用いて増幅産物を検出する。当然の

ことながら、たとえば、市販のＳＮＰ検出マイクロアレイ（たとえば、Illumina社（San Diego, CA）のGoldenGate（登録商標）、DASL（登録商標）、Infinium（登録商標）、またはCytoSNP-12の遺伝型決定試験法、あるいはAffymetrix社のＳＮＰ検出マイクロアレイ製品（OncoScan（商標）マイクロアレイなど）が利用可能である。一部の態様では、単一細胞からのアレイデータ分析の精度を向上するため、胎芽または胎児の実親のうち一方または両方に関する相決定済み遺伝データを用いる。

【０４５８】

配列決定を含む一部の態様では、リード深度は所与の座位にマッピングされる配列決定リードの数である。リード深度を総リード数に対して正規化してもよい。試料のリード深度に関して一部の態様では、リード深度は標的座位に対する平均リード深度である。座位のリード深度に関して一部の態様では、リード深度はシーケンサが測定した、その座位にマッピングされるリード数である。一般的に、座位のリード深度が深いほど、その座位でのアレル比は元のＤＮＡ試料のアレル比に近くなる傾向にある。リード深度は多様な異なる形で表現可能であり、百分率または比が挙げられるがこれらに限定するものではない。このように、たとえば、Illumina（登録商標）HiSeqなどの超並列型ＤＮＡシーケンサ（highly parallel DNA sequencer）（たとえば１００万クローンの配列を作製する）では、１つの座位の配列決定を３，０００回行うと、その座位のリード深度は３，０００リードになる。その座位でのリードの割合は、３，０００を総リードの１００万で割り、すなわち総リードの０．３％である。

【０４５９】

一部の態様ではアレルデータを得る。ここで、アレルデータは、多型座位の特定のアレルのコピー数を示す定量的測定値を含む。一部の態様では、アレルデータは、多型座位で観測されるアレルの各々のコピー数を示す定量的測定値を含む。一般的に、対象多型座位の、可能性のあるすべてのアレルについて定量的測定値を得る。たとえば、先の段落に記載したＳＮＰ座位またはＳＮＶ座位のアレルの決定方法（たとえば、マイクロアレイ、定量ＰＣＲ、ＤＮＡ配列決定（高処理ＤＮＡ配列決定など）のいずれかを用いて、多型座位の特定のアレルのコピー数の定量的測定値を作成することができる。この定量的測定値を、本明細書ではアレル頻度データまたは測定された遺伝アレルデータという。アレルデータを用いる方法を定量的アレル方法ということもある。このような方法は、アレル同一性を考慮せずに非多型座位の定量的データまたは多型座位の定量的データのみを用いる定量的方法と対照的である。高処理配列決定によってアレルデータを測定する場合、アレルデータは一般的に、対象座位にマッピングされる各アレルのリード数を含む。

【０４６０】

一部の態様では、非アレルデータを得る。ここで、非アレルデータは特定の座位のコピー数を示す定量的測定値を含む。座位は多型であっても非多型であってもよい。座位が非多型である一部の態様では、非アレルデータはその座位に存在し得る個々のアレルの相対量または絶対量に関する情報を含まない。非アレルデータのみを用いる（すなわち、各断片のアレル同一性を考慮せずに非多型のアレルの定量データまたは多型座位の定量データを用いる）方法を定量的方法と呼ぶ。一般的に、対象多型座位の、可能性のあるすべてのアレルについて定量的測定値を得、全体として１つの値をその座位のすべてのアレルに関する測定量と関連づける。ある多型座位に関する非アレルデータを、その座位の各アレルに関する定量的アレルデータを合計することによって得てもよい。高処理配列決定によってアレルデータを測定する場合、非アレルデータは一般的に、対象座位にマッピングされるリード数を含む。配列決定測定値は、アレルの同一性にかかわらず、その座位に存在するアレルの各々の相対数および／または絶対数を示す可能性があり、非アレルデータは、その座位にマッピングされるリードの合計を含む。一部の態様では、配列決定測定値の同一の集合を用いてアレルデータと非アレルデータの両方を作成できる。一部の態様では、対象染色体でのコピー数を決定する方法の一部としてアレルデータを用い、作成された非アレルデータを、対象染色体でのコピー数を決定する別の方法の一部として用いることができる。一部の態様では、この２つの方法は統計的に直交であり、これらを組み合わせる

ことにより、対象染色体でのコピー数をより高精度に決定できる。

【０４６１】

一部の態様では、遺伝データを得ることは、（ｉ）実験技術によってＤＮＡ配列情報を取得すること（たとえば自動高処理ＤＮＡシーケンサの使用によって）または（ｉｉ）以前に実験技術によって得た情報を取得することを含み、ここで、情報は、たとえばコンピュータを用いてインターネットを通じて、あるいは配列決定装置からの電子送信によって、電子送信される。

【０４６２】

試料調製、増幅、および定量の方法のさらなる例が２０１２年１１月２１日出願の米国特許出願第１３／６８３，６０４号（米国特許出願公開第２０１３／０１２３１２０号）明細書、および２０１４年5月１６日出願の米国特許出願第６１／９９４，７９１号明細書（これらの文献の全内容は参照によって本明細書に組み込まれる）に記載されている。本明細書に開示した任意の試料の分析にこれらの方法を用いることができる。

【０４６３】

無細胞ＤＮＡの定量方法の例

必要に応じ、無細胞ＤＮＡまたは無細胞ＲＮＡの量または濃度を標準の方法で測定することができる。一部の態様では、無細胞ミトコンドリアＤＮＡの量または濃度を決定する。一部の態様では、核内ＤＮＡに由来する無細胞ＤＮＡ（無細胞核内ＤＮＡ）の量または濃度を決定する。一部の態様では、無細胞ミトコンドリアＤＮＡと無細胞ｎＤＮＡの量または濃度を同時に決定する。

【０４６４】

一部の態様では、定量ＰＣＲを用いて無細胞核内ＤＮＡおよび／または無細胞ミトコンドリアＤＮＡを測定する（Kohler et al. “Levels of plasma circulating cell free nuclear and mitochondrial DNA as potential biomarkers for breast tumors.” Mol Cancer 8:105, 2009, 8:doi:10.1186/1476-4598-8-105；この文献の全内容は参照によって本明細書に組み込まれる）。たとえば、無細胞核内ＤＮＡの１つ以上の座位（たとえばグリセルアルデヒド－３－リン酸デヒドロゲナーゼ（ＧＡＰＤＨ）など）と無細胞ミトコンドリアＤＮＡの１つ以上の座位（ＡＴＰアーゼ８（MTATP 8））を多重定量ＰＣＲで測定することができる。一部の態様では、蛍光標識ＰＣＲを用いて無細胞核内ＤＮＡおよび／または無細胞ミトコンドリアＤＮＡを測定する（Schwarzenbach et al., “Evaluation of cell-free tumour DNA and RNA in patients with breast cancer and benign breast disease.” Mol Biosys 7:2848-2854, 2011；この文献の全内容は参照によって組み込まれる）。必要に応じ、データの正規性分布をシャピロ－ウィルク検定などの標準の方法で決定することができる。必要に応じ、マン－ホイットニーのＵ検定などの標準の方法で無細胞核内ＤＮＡおよび無細胞ミトコンドリアＤＮＡの濃度を比較することができる。一部の態様では、マン－ホイットニーのＵ検定またはクラスカル－ワリス検定などの標準の方法で無細胞核内ＤＮＡおよび／または無細胞ミトコンドリアＤＮＡの濃度を他の確立された予後因子と比較する。

【０４６５】

ＲＮＡの増幅、定量、分析方法の例

以下の方法例のいずれかを用いて、無細胞ＲＮＡ、細胞内ＲＮＡ、細胞質ＲＮＡ、コーディング細胞質ＲＮＡ、非コーディング細胞質ＲＮＡ、ｍＲＮＡ、ｍｉＲＮＡ、ミトコンドリアＲＮＡ、ｒＲＮＡ、またはｔＲＮＡなどのＲＮＡを増幅し、必要に応じて定量してもよい。一部の態様では、ｍｉＲＮＡはワールドワイドウェブのmirbase.orgで利用できるmiRBaseデータベース（その全内容は参照によって本明細書に組み込まれる）に掲載されている任意のｍｉＲＮＡ分子である。ｍｉＲＮＡ分子の例としては、ｍｉＲ－５０９、ｍｉＲ－２１、ｍｉＲ－１４６ａが挙げられる。

【０４６６】

一部の態様では、逆転写酵素を用いる多重ライゲーション依存性プローブ増幅（reverse-transcriptase multiplex ligation-dependent probe amplification = RT-MLPA）を用

いてＲＮＡを増幅する。一部の態様では、ハイブリダイズするプローブの各セットはＳＮＰにわたる２つの短い合成オリゴヌクレオチドと１つの長いオリゴヌクレオチドで構成される（Li et al., Arch Gynecol Obstet. “Development of noninvasive prenatal diagnosis of trisomy 21 by RT-MLPA with a new set of SNP markers,” July 5, 2013, DOI 10.1007/s00404-013-2926-5; Schouten et al. “Relative quantification of 40 nucleic acid sequences by multiplex ligation-dependent probe amplification.” Nucleic Acids Res 30:e57, 2002; Deng et al. (2011) “Non-invasive prenatal diagnosis of trisomy 21 by reverse transcriptase multiplex ligation-dependent probe amplification,” Clin, Chem. Lab Med. 49:641-646, 2011（これらの文献それぞれの全内容は参照によって本明細書に組み込まれる））。

【０４６７】

一部の態様では、逆転写酵素ＰＣＲによってＲＮＡを増幅する。一部の態様では、先に述べたように、SYBR GREEN Iを用いた一段階リアルタイム逆転写酵素ＰＣＲのようなリアルタイム逆転写酵素ＰＣＲによってＲＮＡを増幅する（Li et al., Arch Gynecol Obstet. “Development of noninvasive prenatal diagnosis of trisomy 21 by RT-MLPA with a new set of SNP markers,” July 5, 2013, DOI 10.1007/s00404-013-2926-5; Lo et al., “Plasma placental RNA allelic ratio permits noninvasive prenatal chromosomal aneuploidy detection,” Nat Med 13:218-223, 2007; Tsui et al., Systematic micro-array based identification of placental mRNA in maternal plasma: towards non-invasive prenatal gene expression profiling. J Med Genet 41:461-467, 2004; Gu et al., J. Neurochem. 122:641-649, 2012（これらの文献のそれぞれの全内容は参照によって本明細書に組み込まれる））。

【０４６８】

一部の態様では、マイクロアレイを用いてＲＮＡを検出する。たとえば、Agilent TechnologiesのヒトｍｉＲＮＡマイクロアレイを製造者のプロトコルに従って使用することができる。簡潔に述べると、単離したＲＮＡを脱リン酸化し、ｐＣｐ－Ｃｙ３とライゲートする。Sanger miRBase release 14.0に基づき、標識を付けたＲＮＡを精製し、ヒトの成熟ｍｉＲＮＡに対するプローブを含むｍｉＲＮＡアレイにハイブリダイズさせる。このアレイを洗浄し、マイクロアレイスキャナ（Agilent TechnologiesのG2565BA）を用いてスキャンする。各ハイブリダイゼーションシグナルの強度をAgilent extraction software v9.5.3で評価する。標識付け、ハイブリダイゼーション、およびスキャニングは、Agilent miRNA Microarray Systemのプロトコルに従って実行可能である（Gu et al., J. Neurochem. 122:641-649, 2012；この文献の全内容は参照によって本明細書に組み込まれる）。

【０４６９】

一部の態様では、TaqMan（登録商標）試験法を用いてＲＮＡを検出する。試験法の一例は１５７種類のTaqMan（登録商標）MicroRNA Assayを含むTaqMan（登録商標） Array Human MicroRNA Panel v1.0 (Early Access)（Applied Biosystems（商標））であり、各逆転写プライマーとＰＣＲ用プライマーとTaqMan（登録商標）プローブとを含む。（Chim et al., “Detection and characterization of placental microRNAs in maternal plasma,” Clin Chem. 54(3):482-90, 2008（この文献の全内容は参照によって組み込まれる））。

【０４７０】

必要に応じ、１つ以上のｍＲＮＡのｍＲＮＡスプライシングパターンを標準の方法で決定できる（Fackenthall and Godley, Disease Models & Mechanisms 1: 37-42, 2008, doi:10.1242/dmm.000331（この文献の全内容は参照によって組み込まれる）。たとえば、高密度マイクロアレイおよび／または高処理ＤＮＡ配列決定を用いてｍＲＮＡスプライス多様体を検出できる。

【０４７１】

一部の態様では、全トランスクリプトームショットガン配列決定またはアレイを用いて

トランスクリプトームを測定する。

【０４７２】

増幅方法の例

同一反応容量（たとえば、すべての標的座位を同時に増幅する試料多重ＰＣＲ反応の一部）内での近接または隣接する標的座位の増幅による干渉を最小限にするか防止する改良型ＰＣＲ増幅法も開発されている。これらの方法を用いて近接または隣接する標的座位を同時に増幅することができ、干渉を避けるために近接する標的座位を個別に増幅できるように別の反応容量に分離する必要がある場合に比べて時間とコストを低減できる。

【０４７３】

一部の態様では、５’→３’エキソヌクレアーゼ活性および／または鎖置換活性の低いポリメラーゼ（たとえばＤＮＡポリメラーゼ、ＲＮＡポリメラーゼ、または逆転写酵素）を用いて標的座位の増幅を行う。一部の態様では、５’→３’エキソヌクレアーゼ活性のレベルを低くすることにより、近接するプライマーの分解（たとえばプライマー伸長時に伸張されなかったプライマーや１個以上のヌクレオチドが付加されたプライマーが生じること）を低減または防止する。一部の態様では、鎖置換活性を低くすることにより、近接するプライマーの置換（たとえばプライマー伸長時に伸張されなかったプライマーや１個以上のヌクレオチドが付加されたプライマーが生じること）を低減または防止する。一部の態様では、互いに隣接する標的座位（たとえば標的座位間に塩基が存在しない場合）または近接する標的座位（たとえば座位が互いに５０塩基以内、４０塩基以内、３０塩基以内、２０塩基以内、１５塩基以内、１０塩基以内、９塩基以内、８塩基以内、７塩基以内、６塩基以内、５塩基以内、４塩基以内、３塩基以内、２塩基以内、または１塩基以内の位置に存在する場合）を増幅する。一部の態様では、１つの座位の３’末端は次の下流側座位の５’末端から５０塩基以内、４０塩基以内、３０塩基以内、２０塩基以内、１５塩基以内、１０塩基以内、９塩基以内、８塩基以内、７塩基以内、６塩基以内、５塩基以内、４塩基以内、３塩基以内、２塩基以内、または１塩基以内の位置に存在する。

【０４７４】

一部の態様では、少なくとも１００個、２００個、５００個、７５０個、１，０００個、２，０００個、５，０００個、７，５００個、１０，０００個、２０，０００個、２５，０００個、３０，０００個、４０，０００個、５０，０００個、７５，０００個、または１００，０００個の異なる標的座位を同時増幅などによって１つの反応容量で増幅する。一部の態様では、増幅産物の少なくとも５０％、６０％、７０％、８０％、９０％、９５％、９６％、９７％、９８％、９９％、または９９．５％が標的増幅産物である。様々な態様では、標的増幅産物である増幅産物の量は５０％～９９．５％であり、たとえば６０％～９９％、７０％～９８％、８０％～９８％、９０％～９９．５％、または９５％～９９．５％（それぞれ、範囲の上限値と下限値を含む）。一部の態様では、標的座位の少なくとも５０％、６０％、７０％、８０％、９０％、９５％、９６％、９７％、９８％、９９％、または９９．５％を同時増幅などによって１つの反応容量で増幅する（たとえば、増幅前の量に比べて少なくとも５倍、１０倍、２０倍、３０倍、５０倍、または１００倍に増幅する）。様々な態様では、増幅される標的座位の量は５０％～９９．５％であり、たとえば、６０％～９９％、７０％～９８％、８０％～９９％、９０％～９９．５％、９５％～９９．９％、または９８％～９９．９９％（それぞれ、範囲の上限値と下限値を含む）である（たとえば増幅前に比べて少なくとも５倍、１０倍、２０倍、３０倍、５０倍、または１００倍に増幅される）。一部の態様では、非標的増幅産物の生成をさらに少なくすることができる（たとえば、第一のプライマー対のうち順方向プライマーと第二のプライマー対のうち逆方向プライマーから形成される増幅産物がより少ない）。このような望ましくない非標的増幅産物は、従来の増幅方法によって、たとえば第一のプライマー対のうち逆方向プライマーおよび／または第二のプライマー対のうち順方向プライマーが分解かつ／または置換された場合に生成される可能性がある。

【０４７５】

一部の態様では、これらの方法により、ポリメラーゼの５’→３’エキソヌクレアーゼ

活性および／または鎖置換活性が低いため、伸長されるプライマーに結合したポリメラーゼが近接するプライマー（次の下流プライマーなど）を分解かつ／または置換する可能性がより低くなり、使われる伸長時間がより長くなる。様々な態様では、伸長されるプライマーに付加されるヌクレオチドの数が、同一鎖上にあるプライマー結合部位の３’末端と次の下流プライマー結合部位の５’末端との間に存在するヌクレオチドの数の８０％以上、９０％以上、９５％以上、１００％以上、１１０％以上、１２０％以上、１３０％以上、１４０％以上、１５０％以上、１７５％以上、または２００％以上になるようなポリメラーゼ伸長率が得られる反応条件（伸長時間・温度など）を用いる。

【０４７６】

一部の態様では、ＤＮＡポリメラーゼを用い、ＤＮＡを鋳型としてＤＮＡ増幅産物を生成する。一部の態様では、ＲＮＡポリメラーゼを用いてＤＮＡを鋳型としてＲＮＡ増幅産物を生成する。一部の態様では、逆転写酵素を用いてＲＮＡを鋳型としてｃＤＮＡ増幅産物を生成する。

【０４７７】

一部の態様では、ポリメラーゼの５’→３’エキソヌクレアーゼ活性のレベルが低いことは、同じ条件下、同じ量のテルムス・アクウァーティクス（テルムス属）ポリメラーゼ（一般に使用される好熱性細菌由来のＤＮＡポリメラーゼである「Ｔａｑ」ポリメラーゼ（PDB 1BGX, EC 2.7.7.7））（Murali et al., “Crystal structure of Taq DNA polymerase in complex with an inhibitory Fab: the Fab is directed against an intermediate in the helix-coil dynamics of the enzyme,” Proc. Natl. Acad. Sci. USA 95:12562-12567, 1998（この文献の全内容は参照によって本明細書に組み込まれる））の活性の８０％未満、７０％未満、６０％未満、５０％未満、４０％未満、３０％未満、２０％未満、１０％未満、５％未満、１％未満、または０．１％未満であるということである。一部の態様では、ポリメラーゼの鎖置換活性が低いということは、同じ条件下、同じ量のＴａｑポリメラーゼの活性の８０％未満、７０％未満、６０％未満、５０％未満、４０％未満、３０％未満、２０％未満、１０％未満、５％未満、１％未満、または０．１％未満であるということである。

【０４７８】

一部の態様では、ポリメラーゼはPHUSION（登録商標）DNA Polymeraseであり、たとえばPHUSION（登録商標）High Fidelity DNA Polymerase（M0530S, New England BioLabs, Inc.）またはPHUSION（登録商標）Hot Start Flex DNA Polymerase（M0535S, New England BioLabs, Inc.；Frey and Suppman BioChemica. 2:34-35, 1995; Chester and Marshak Analytical Biochemistry. 209:284-290, 1993（これらの文献それぞれの全内容は参照によって本明細書に組み込まれる））である。PHUSION（登録商標）DNA Polymeraseは、処理能力を高めるドメインを融合した、ピロコックス属に類似の酵素である。PHUSION（登録商標）DNA Polymeraseは５’→３’ポリメラーゼ活性と３’→５’エキソヌクレアーゼ活性を有し、末端が平滑な産物を生成する。PHUSION（登録商標）DNA Polymeraseには、５’→３’エキソヌクレアーゼ活性および鎖置換活性はない。

【０４７９】

一部の態様では、ポリメラーゼはQ5（登録商標）DNA Polymeraseであり、たとえばQ5（登録商標）High-Fidelity DNA Polymerase（M0491S, New England BioLabs, Inc.）またはQ5（登録商標）Hot Start High-Fidelity DNA Polymerase（M0493S, New England BioLabs, Inc.）である。Q5（登録商標）High-Fidelity DNA Polymeraseは処理能力を高めるＳｓｏ７ｄドメインを融合した、３’→５’エキソヌクレアーゼ活性を有し正確性が高い耐熱性ＤＮＡポリメラーゼである。Q5（登録商標）High-Fidelity DNA Polymeraseには５’→３’エキソヌクレアーゼ活性と鎖置換活性はない。

【０４８０】

一部の態様では、ポリメラーゼはT4 DNA Polymerase（M0203S, New England BioLabs, Inc.；Tabor and Struh. (1989). “DNA-Dependent DNA Polymerases,” (Ausebel et al. (Ed.), Current Protocols in Molecular Biology. 3.5.10-3.5.12. New York: John W

iley & Sons, Inc., 1989に記載）；およびSambrook et al. Molecular Cloning: A Laboratory Manual. (2nd ed.), 5.44-5.47. Cold Spring Harbor: Cold Spring Harbor Laboratory Press, 1989（これらの文献それぞれの全内容は参照によって本明細書に組み込まれる））。T4 DNA Polymeraseは５’→３’方向のＤＮＡ合成を触媒し、鋳型とプライマーを必要とする。この酵素は、DNA Polymerase Iに見られる活性に比べてはるかに高い３’→５’エキソヌクレアーゼ活性を有する。T4 DNA Polymeraseには５’→３’エキソヌクレアーゼ活性と鎖置換活性はない。

【０４８１】

一部の態様では、ポリメラーゼはSulfolobus DNA Polymerase IV （M0327S, New England BioLabs, Inc.；（Boudsocq,. et al. (2001). Nucleic Acids Res., 29:4607-4616, 2001；およびMcDonald, et al. (2006). Nucleic Acids Res., 34:1102-1111, 2006（これらの文献それぞれの全内容は参照によって本明細書に組み込まれる））である。Sulfolobus DNA Polymerase IVはＹファミリーに属する耐熱性の損傷バイパス型ＤＮＡポリメラーゼであり、多様なＤＮＡ鋳型の損傷を越えて効率的にＤＮＡを合成する（McDonald, J. P. et al. (2006). Nucleic Acids Res.,. 34, 1102-1111（この文献の全内容は参照によって本明細書に組み込まれる））。Sulfolobus DNA Polymerase IVには５’→３’エキソヌクレアーゼ活性と鎖置換活性はない。

【０４８２】

一部の態様では、ＳＮＰのある領域にプライマーが結合する場合、異なるアレルに対し異なる効率で結合かつ増幅してもよく、１個のアレルのみに対して結合かつ増幅してもよい。ヘテロ接合型の対象者の場合、アレルのうち１つをプライマーによって増幅しなくてもよい。一部の態様では、各アレルに対してプライマーを設計する。たとえば、２つのアレルが存在する場合（たとえば２アレルのＳＮＰ）、標的座位の同じ位置に結合する２個のプライマー（たとえば「Ａ」アレルに結合する順方向プライマーと「Ｂ」アレルに結合する順方向プライマー）を用いることができる。ｄｂＳＮＰデータベースなどの標準の方法を用いて、ヘテロ接合体率の高いＳＮＰホットスポットなどの既知のＳＮＰの位置を決定することができる。

【０４８３】

一部の態様では、増幅産物はサイズが類似している。一部の態様では、標的増幅産物の長さの範囲は１００ヌクレオチド未満、７５ヌクレオチド未満、５０ヌクレオチド未満、２５ヌクレオチド未満、１５ヌクレオチド未満、１０ヌクレオチド未満、または５ヌクレオチド未満である。一部の態様（断片化したＤＮＡまたはＲＮＡの標的座位の増幅など）では、標的増幅産物の長さは５０ヌクレオチド～１００ヌクレオチドであり、たとえば６０ヌクレオチド～８０ヌクレオチド、または６０ヌクレオチド～７５ヌクレオチド（それぞれ、範囲の上限と下限を含む）である。一部の態様（エキソン全体または遺伝子全体の複数の標的座位の増幅など）では、標的増幅産物の長さは１００ヌクレオチド～５００ヌクレオチドであり、たとえば１５０ヌクレオチド～４５０ヌクレオチド、２００ヌクレオチド～４００ヌクレオチド、２００ヌクレオチド～３００ヌクレオチド、または３００ヌクレオチド～４００ヌクレオチド（それぞれ、範囲の上限値と下限値を含む）である。

【０４８４】

一部の態様では、複数の標的座位を、各標的座位に対する順方向プライマーと逆方向プライマーを含むプライマー対を用いてその反応容量で増幅されるよう同時に増幅する。一部の態様では、標的座位ごとに１個のプライマーを用いて一回のＰＣＲを実行し、次いで標的座位ごとにプライマー対を用いて二回目のＰＣＲを実行する。たとえば、すべてのプライマーが同じ鎖に結合するように標的座位ごとに１個のプライマーを用いて一回目のＰＣＲを実行してもよい（たとえば各標的座位に対して順方向プライマーを用いる）。このことにより、ＰＣＲによって直線的な増幅が行われ、配列や長さの違いによる増幅産物間の増幅の偏りを低減または排除することができる。一部の態様では、次いで増幅産物を各標的座位に対する順方向プライマーと逆方向プライマーを用いて増幅する。

プライマー設計方法の例

【０４８５】
必要に応じ、プライマー二量体を形成する尤度の低いプライマーを用いて多重ＰＣＲを実行してもよい。特に、高度多重化ＰＣＲを行うと、プライマー二量体形成などの非生産的な副反応によるＤＮＡ産物の割合が非常に高くなる場合が多い。一態様では、非生産的な副反応を起こす尤度が最も高い特定のプライマーをプライマーライブラリから除外し、ゲノムにマッピングされる増幅ＤＮＡの割合をより高くするプライマーライブラリを得てもよい。問題のあるプライマー、すなわち二量体を形成する尤度が特に高いプライマーを除外する工程により、予想外にも、その後の配列決定による分析のために非常に高度に多重化したＰＣＲが可能になった。

【０４８６】
マッピングされないプライマー二量体または他の悪影響を及ぼすプライマー産物の量を最小限にしたライブラリを得るためにプライマーを選択する方法は複数存在する。経験的なデータにより、少数の「悪い」プライマーはマッピングされないプライマー二量体副反応の大量発生の原因となることが示されている。これらの「悪い」プライマーを除外することにより、標的座位にマッピングされる配列リードの割合を上げることができる。「悪い」プライマーを特定する１つの方法は、標的化増幅によって増幅したＤＮＡの配列決定データを調べることであり、最も高頻度で認められるプライマー二量体を除外して、ゲノムにマッピングされない副産物ＤＮＡをもたらす可能性が極めて低いプライマーライブラリを得ることができる。プライマーの様々な組み合わせの結合エネルギーを算出できる公的に入手可能なプログラムも存在し、結合エネルギーが最も高いプライマーの組み合わせを除外することによっても、ゲノムにマッピングされない副産物ＤＮＡをもたらす可能性が極めて低いプライマーライブラリが得られる。

【０４８７】
プライマーの選択に関する一部の態様では、候補標的座位に対し１個以上のプライマーまたはプライマー対を設計することにより、候補プライマーの初期ライブラリを作製する。候補標的座位（ＳＮＰなど）の集合は、その標的座位に関する所望のパラメータ（標的母集団内のＳＮＰの頻度、またはＳＮＰのヘテロ接合体率）についての公的に入手可能な情報に基づいて選択可能である。一態様では、Primer3プログラム（ワールドワイドウェブのprimer3.sourceforge.net; libprimer3 release 2.2.3（その全内容は参照によって本明細書に組み込まれる））を用いてＰＣＲプライマーを設計してもよい。必要であれば、特定のアニーリング温度範囲内でアニーリングし、特定の範囲のＧＣ含量を有し、特定のサイズ範囲を有し、特定のサイズ範囲の標的増幅産物を生成し、かつ／またはその他のパラメータ特性を有するようにプライマーを設計することができる。候補標的座位ごとに複数のプライマーまたはプライマー対から出発することによって、大多数またはすべての標的座位についてプライマーまたはプライマー対がライブラリに残る可能性が高くなる。一態様では、選択基準は標的座位ごとに少なくとも１個のプライマー対がライブラリに残ることを前提としてもよい。このようにして、最終プライマーライブラリを用いて大多数またはすべての標的座位を増幅する。このことは、ゲノム上の多数の位置の欠失または重複のスクリーニング、あるいは疾患または疾患のリスク増加に関与する多数の配列（多型またはその他の変異など）のスクリーニングなどの用途にとって望ましい。ライブラリのプライマー対が、別のプライマー対によって生成された標的増幅産物と重複する標的増幅産物を生成する場合、それらのプライマー対のうちの１つを、干渉を防止するためにライブラリから除外してもよい。

【０４８８】
一部の態様では、候補プライマーのライブラリから、考えられる２個のプライマーの組み合わせの大部分またはすべてについて、「不適切度スコア」（スコアが高いほど適切度が低いことを示す）を計算する（コンピュータによる計算など）。様々な態様では、ライブラリに含まれる候補プライマーの考えられる組み合わせのうち少なくとも８０％、９０％、９５％、９８％、９９％、または９９．５％について不適切度スコアを計算する。各不適切度スコアは、少なくとも部分的に２個の候補プライマーが二量体を形成する尤度に

基づいている。必要に応じ、不適切度スコアはさらに、標的座位のヘテロ接合体率、標的座位での配列（たとえば多型）に関連する疾患罹患率、標的座位での配列（たとえば多型）に関連する疾患浸透度、標的座位に対する候補プライマーの特異性、候補プライマーのサイズ、標的増幅産物の融解温度、標的増幅産物のＧＣ含量、標的増幅産物の増幅効率、標的増幅産物のサイズ、および組換えホットスポットの中心からの距離からなる群より選択される、他の１つ以上のパラメータに基づいてもよい。一部の態様では、標的座位に対する候補プライマーの特異性は、候補プライマーの設計上の増幅対象である標的座位以外の座位に対して結合・増幅することによってミスプライムする尤度を含む。一部の態様では、ミスプライムする１個以上またはすべての候補プライマーをライブラリから除外する。一部の態様では、選択する候補プライマーの数を増加させるため、ミスプライムする可能性のある候補プライマーをライブラリから除外しない。複数の因子について考慮する場合、不適切度スコアを多様なパラメータの加重平均に基づいて計算してもよい。パラメータには、プライマーを使用する特定の用途にとっての重要性に基づいて異なる重みを割り当てることができる。一部の態様では、不適切度スコアの最も高いプライマーをライブラリから除外する。除外したプライマーが、１つの標的座位にハイブリダイズするプライマー対を構成するプライマーである場合、そのプライマー対を構成するもう一方のプライマーをライブラリから除外してもよい。必要に応じて、プライマーを除外する工程を繰り返してもよい。一部の態様では、ライブラリに残っているプライマーの組み合わせ候補の不適切度スコアがすべて最小閾値以下になるまで選択方法を実行する。一部の態様では、ライブラリに残っている候補プライマーの数が所望の数に減少するまで選択方法を実行する。

【０４８９】

様々な態様では、不適切度スコアを計算した後、２個の候補プライマーの最大数の組み合わせの一部である、不適切度スコアが第一の最小閾値を超える候補プライマーをライブラリから除外する。この工程では、第一の最小閾値以下の相互作用は重要性が低いため、このような相互作用を無視する。除外したプライマーが、１つの標的座位にハイブリダイズするプライマー対を構成するプライマーである場合、そのプライマー対を構成するもう一方のプライマーをライブラリから除外してもよい。必要に応じて、プライマーを除外する工程を繰り返してもよい。一部の態様では、ライブラリに残っているプライマーの組み合わせ候補の不適切度スコアがすべて第一の最小閾値以下になるまで選択方法を実行する。ライブラリに残っている候補プライマーの数が所望の数よりも多い場合、第一の最小閾値をより低い第二の最小閾値まで下げ、プライマーを除外する工程を繰り返すことによってプライマーの数を減少させてもよい。ライブラリに残っている候補プライマーの数が所望の数よりも少ない場合には、第一の最小閾値をより高い第二の最小閾値まで上げ、元の候補プライマーライブラリでプライマーを除外する工程を繰り返すことによって方法を続けることが可能であり、このようにしてより多くの候補プライマーをライブラリに残すことができる。一部の態様では、ライブラリに残っているプライマーの組み合わせ候補の不適切度スコアがすべて第二の最小閾値以下になるか、ライブラリに残っている候補プライマーの数が所望の数に減少するまで選択方法を実行する。

【０４９０】

必要に応じて、別のプライマー対が生成する標的増幅配列と重複する標的増幅産物を生成するプライマー対を別の増幅反応に分けることができる。（標的増幅産物が重複しているため候補標的座位を分析から外すのではなく）候補標的座位をすべて分析することが望ましい用途では、複数のＰＣＲ増幅反応を行うことが望ましくなる場合がある。

【０４９１】

これらの選択方法により、所望されるプライマー二量体低減を達成するためにライブラリから除外する必要のある候補プライマーの数が最小限になる。ライブラリから除外する候補プライマーの数がより少ないことによって、得られたプライマーライブラリを用いてより多くの（またはすべての）標的座位を増幅することができる。

【０４９２】

多数のプライマーを複合することにより、含まれ得る試験法には相当な制約が課される。意図せず相互作用する試験法では、偽の増幅産物が発生する。ミニＰＣＲのサイズ制約によってさらに制約が生じる可能性がある。一態様では、可能性のある非常に多数（約５００～１００万超）のＳＮＰ標的から出発し、各ＳＮＰを増幅するためのプライマーの設計を試みることが可能である。プライマーを設計することが可能な場合、ＤＮＡ二重鎖の形成について公開されている熱力学的パラメータを使用し、考えられるすべてのプライマー対の偽プライマー二本鎖形成の尤度を評価することによって、偽の生成物を形成する可能性のあるプライマー対の特定を試みることが可能である。相互作用に関連するスコアリング機能によってプライマー相互作用をランク付けしてもよく、相互作用スコアが最も悪いプライマーを、所望のプライマー数が得られるまで排除する。ヘテロ接合性である可能性の高いＳＮＰが最も有用である場合、試験法のリストをランク付けし、最もヘテロ接合性に適合する試験法を選択することも可能である。相互作用スコアの高いプライマーはプライマー二量体を形成する可能性が最も高いことが実験によって実証されている。高度多重化では、偽相互作用をすべて排除することは不可能であるが、相互作用スコアが最も高いプライマーまたはプライマー対は反応全体を支配して目的の標的からの増幅を大きく制限する可能性があるため、これらをコンピュータで除外することが必須である。発明者らはこの手順を実行し、最大１０，０００個、場合によってはそれを超えるプライマーの多重プライマーセットを作製した。この手順による改善は大きく、ＰＣＲ産物をすべて配列決定で確認すると、最も悪いプライマーを除外しない反応では１０％の増幅率であったのに対し、標的産物は８０％超、９０％超、９５％超、９８％超、さらには９９％の増幅率である。前述の部分的に準入れ子の方法と組み合わせると、増幅産物の９０％以上、さらには９５％以上が標的配列にマッピングされ得る。

【０４９３】

当然ながら、どのＰＣＲプローブが二量体を形成する可能性が高いかを決定する方法は他にもある。一態様では、最適化されていないプライマーセットを用いて増幅したＤＮＡのプールを分析すれば、問題のあるプライマーを決定するのに十分である場合がある。たとえば、配列決定によって分析を行ってもよく、最も多く存在する二量体を、二量体を結成する可能性が最も高いプライマーと判断し、そのプライマーを除外してもよい。一態様では、プライマー設計の方法を本明細書に記載のミニＰＣＲ法と組み合わせて使用してもよい。

【０４９４】

プライマー上にタグを使用することにより、プライマー二量体生成物の増幅および配列決定を低減できる。一部の態様では、プライマーは、タグと共にループ構造を形成する内部領域を含む。特定の態様では、プライマーは、標的座位に特異的な５’領域と、標的座位に非特異的かつループ構造を形成する内部領域と、標的座位に特異的な３’領域とを含む。一部の態様では、ループ領域は、鋳型ＤＮＡの連続領域または隣接領域に結合するように設計された２つの結合領域の間に位置してもよい。様々な態様では、３’領域は少なくとも７ヌクレオチド長である。一部の態様では、３’領域は７～２０ヌクレオチド長であり、たとえば７～１５ヌクレオチド長、または７～１０ヌクレオチド長（それぞれ、範囲の上限と下限を含む）。様々な態様では、プライマーは、標的座位に非特異的な５’領域（タグまたは汎用プライマー結合部位など）に続いて、標的座位に特異的な領域と、標的座位に非特異的かつループ構造を形成する内部領域と、標的座位に特異的な３’領域を含む。必要な標的特異的配列を２０塩基対未満、１５塩基対未満、１２塩基対未満、さらには１０塩基対未満に短縮するためにタグプライマーを利用することができる。これは、標準的なプライマー設計を用いても標的配列をプライマー結合部位で断片化するときに偶然起こる可能性もあるし、プライマーの構造として設計することもできる。この方法の利点としては、ある最大の増幅産物長に関して設計可能な試験法の数が増えることと、プライマー配列の「情報として無益な」配列決定を短縮することが挙げられる。この方法を内部標識法と組み合わせて用いてもよい。

【０４９５】

一態様では、多重標的化ＰＣＲ増幅における非生産的な産物の相対量を、アニーリング温度を上げることによって低減できる。標的特異性プライマーと同じタグを用いてライブラリを増幅する場合、タグがプライマー結合に役立つため、アニーリング温度をゲノムＤＮＡの場合に比べて高くすることができる。一部の態様では、必要に応じて、アニーリング時間を長くすると共にプライマー濃度を低くする。一部の態様では、アニーリング時間は３分超、５分超、８分超、１０分超、１５分超、２０分超、３０分超、６０分超、１２０分超、２４０分超、４８０分超、さらには９６０分超であってもよい。ある例示の態様では、プライマー濃度を低くすると共にアニーリング時間を長くする。様々な態様では、伸長時間を通常よりも長くする（たとえば３分超、５分超、８分超、１０分超、または１５分超）。一部の態様では、プライマー濃度はわずか５０ｎＭ、２０ｎＭ、１０ｎＭ、５ｎＭ、１ｎＭ、および１ｎＭ未満である。こうすることにより、意外なことに、高度多重化（マルチプレックス）反応（たとえば、１，０００プレックス反応、２，０００プレックス反応、５，０００プレックス反応、１０，０００プレックス反応、２０，０００プレックス反応、５０，０００プレックス反応、さらには１００，０００プレックス反応）のための堅牢な性能が得られる。一態様では、アニーリング時間を長くして増幅を１サイクル、２サイクル、３サイクル、４サイクル、または５サイクル実行し、続いて、より一般的なアニーリング時間でタグ付きプライマーを用いたＰＣＲサイクルを行う。

【０４９６】

標的位置を選択するには、候補プライマー対設計のプールから出発し、プライマー対の間で起こる可能性のある有害な相互作用の熱力学的モデルを作成し、次いでこのモデルを用いて、プールに含まれる他の設計と適合しない設計を排除することができる。

【０４９７】

一態様では、本発明は、標的座位（たとえば、疾患または障害、あるいは癌などの疾患または障害のリスク増加に関連する多型または変異を含む可能性のある座位）の数を減少させ、かつ／または検出される疾患負荷を増加させる（たとえば、検出される多型または変異の数を増加させる）方法を特徴とする。一部の態様では、前記方法は、癌などの疾患または障害を有する対象者の各座位を多型または変異（一塩基の多様性、挿入、または欠失、あるいは本明細書に記載のそれ以外の任意の多様性）の頻度すなわち反復性によってランク付けする（たとえば降順にランク付けする）ことを含む。一部の態様では、ＰＣＲプライマーを一部の座位またはすべての座位に対して設計する。プライマーライブラリ用のＰＣＲプライマーの選択では、頻度すなわち反復性の高い座位（より高いランクの座位）に対するプライマーは頻度すなわち反復性の低い座位（より低いランクの座位）に対するプライマーよりも有利である。一部の態様では、このパラメータを、本明細書に記載の不適切度スコアの計算におけるパラメータの１つとして含む。必要に応じて、ライブラリに含まれる、他の設計とは適合しないプライマー（ランクの高い座位に対するプライマーなど）を別のＰＣＲライブラリ／プールに含めることができる。一部の態様では、複数（たとえば２個、３個、４個、５個、またはそれ以上）のライブラリ／プールを別のＰＣＲ反応に用いて、すべてのライブラリ／プールで表されるすべて（または大部分）の座位を増幅することができる。一部の態様では、１つ以上のライブラリ／プールに十分なプライマーが含まれるまでこの方法を継続することにより、プライマー全体によって疾患または障害について所望の疾患負荷を（たとえば、疾患負荷の少なくとも８０％、８５％、９０％、９５％、または９９％の検出によって）とらえることができる。

【０４９８】

プライマーライブラリの例

一側面では、本発明はプライマー（本発明の任意の方法によって候補プライマーのライブラリから選択されるプライマーなど）のライブラリを特徴とする。一部の態様では、ライブラリは、少なくとも１００個、２００個、５００個、７５０個、１，０００個、２，０００個、５，０００個、７，５００個、１０，０００個、２０，０００個、２５，０００個、３０，０００個、４０，０００個、５０，０００個、７５，０００個、または１００，０００個の異なる標的座位と１つの反応容量で同時にハイブリダイズする（または同

時にハイブリダイズ可能である）、あるいはこれらを同時に増幅する（または同時に増幅できる）プライマーを含む。様々な態様では、ライブラリは１００個～５００個、５００個～１，０００個、１，０００個～２，０００個、２，０００個～５，０００個、５，０００個～７，５００個、７，５００個～１０，０００個、１０，０００個～２０，０００個、２０，０００個～２５，０００個、２５，０００個～３０，０００個、３０，０００個～４０，０００個、４０，０００個～５０，０００個、５０，０００個～７５，０００個、または７５，０００個～１００，０００個（それぞれ、範囲の上限と下限を含む）の異なる標的座位を１つの反応容量で同時に増幅する（または同時に増幅できる）プライマーを含む。様々な態様では、ライブラリは、１，０００個～１００，０００個（たとえば、１，０００個～５０，０００個、１，０００個～３０，０００個、１，０００個～２０，０００個、１，０００個～１０，０００個、２，０００個～３０，０００個、２，０００個～２０，０００個、２，０００個～１０，０００個、５，０００個～３０，０００個、５，０００個～２０，０００個、または５，０００個～１０，０００個（それぞれ、範囲の上限と下限を含む））の異なる標的座位を１つの反応容量で同時に増幅する（または同時に増幅できる）プライマーを含む。一部の態様では、ライブラリは、増幅産物のうちプライマー二量体が６０％未満、４０％未満、３０％未満、２０％未満、１０％未満、５％未満、４％未満、３％未満、２％未満、１％未満、０．５％未満、０．２５％未満、０．１％未満、または０．５％になるように標的座位を１つの反応容量で同時に増幅する（または同時に増幅できる）プライマーを含む。様々な態様では、プライマー二量体である増幅産物の量は０．５％～６０％であり、たとえば、０．１％～４０％、０．１％～２０％、０．２５％～２０％、０．２５％～１０％、０．５％～２０％、０．５％～１０％、１％～２０％、または１％～１０％（それぞれ、範囲の上限と下限を含む）である。一部の態様では、プライマーは、増幅産物のうち少なくとも５０％、６０％、７０％、８０％、９０％、９５％、９６％、９７％、９８％、９９％、または９９．５％が標的増幅産物となるように標的座位を１つの反応容量で同時に増幅する（または同時に増幅できる）。様々な態様では、標的増幅産物である増幅産物の量は５０％～９９．５％であり、たとえば６０％～９９％、７０％～９８％、８０％～９８％、９０％～９９．５％、または９５％～９９．５％（それぞれ、範囲の上限と下限を含む）である。一部の態様では、プライマーは、標的座位のうち少なくとも５０％、６０％、７０％、８０％、９０％、９５％、９６％、９７％、９８％、９９％、または９９．５％が増幅される（たとえば、増幅前の量に比べて少なくとも５倍、１０倍、２０倍、３０倍、５０倍、または１００倍に増幅される）ように標的座位を１つの反応容量で同時に増幅する（または同時に増幅できる）。様々な態様では、増幅される（たとえば、増幅前の量に比べて少なくとも５倍、１０倍、２０倍、３０倍、５０倍、または１００倍に増幅される）標的座位の量は５０％～９９．５％であり、たとえば６０％～９９％、７０％～９８％、８０％～９９％、９０％～９９．５％、９５％～９９．９％、または９８％～９９．９９％（それぞれ、範囲の上限と下限を含む）である。一部の態様では、プライマーライブラリは、少なくとも１００個、２００個、５００個、７５０個、１，０００個、２，０００個、５，０００個、７，５００個、１０，０００個、２０，０００個、２５，０００個、３０，０００個、４０，０００個、５０，０００個、７５，０００個、または１００，０００個のプライマー対を含み、ここで、各プライマー対は、順方向試験プライマーと逆方向試験プライマーを含み、各試験プライマー対が標的座位にハイブリダイズする。一部の態様では、プライマーライブラリは、少なくとも１００個、２００個、５００個、７５０個、１，０００個、２，０００個、５，０００個、７，５００個、１０，０００個、２０，０００個、２５，０００個、３０，０００個、４０，０００個、５０，０００個、７５，０００個、または１００，０００の個別のプライマーを含み、プライマーはそれぞれ異なる標的座位にハイブリダイズする。ここで、個別のプライマーはプライマー対の一部ではない。

【０４９９】

様々な態様では、各プライマーの濃度は、１００ｎＭ未満、７５ｎＭ未満、５０ｎＭ未満、２５ｎＭ未満、２０ｎＭ未満、１０ｎＭ未満、５ｎＭ未満、２ｎＭ未満、または１ｎ

Ｍ未満、あるいは５００μＭ未満、１００μＭ未満、１０μＭ未満、または１μＭ未満である。様々な態様では、各プライマーの濃度は１μＭ～１００ｎＭであり、たとえば１μＭ～１ｎＭ、１ｎＭ～７５ｎＭ、２ｎＭ～５０ｎＭ、または５ｎＭ～５０ｎＭ（それぞれ、範囲の上限と下限を含む）である。様々な態様では、プライマーのＧＣ含量は３０％～８０％であり、たとえば、４０％～７０％、または５０％～６０％（それぞれ、範囲の上限と下限を含む）である。一部の態様では、プライマーのＧＣ含量の範囲は、３０％未満、２０％未満、１０％未満、または５％未満である。一部の態様では、プライマーのＧＣ含量の範囲は５％～３０％であり、たとえば５％～２０％、または５％～１０％（それぞれ、範囲の上限と下限を含む）である。一部の態様では、試験プライマーの融解温度（T_m）は４０℃～８０℃であり、たとえば５０℃～７０℃、５５℃～６５℃、または５７℃～６０．５℃（それぞれ、範囲の上限と下限を含む）である。一部の態様では、Primer3プログラム（libprimer3 release 2.2.3）を用いて、組み込まれているSantaLuciaパラメータを用いてT_m値を計算する（ワールドワイドウェブのprimer3.sourceforge.net）。一部の態様では、プライマーの融解温度の範囲は１５℃未満、１０℃未満、５℃未満、３℃未満、または１℃未満である。一部の態様では、プライマーの融解温度の範囲は１℃～１５℃であり、たとえば、１℃～１０℃、１℃～５℃、または１℃～３℃（それぞれ、範囲の上限と下限を含む）である。一部の態様では、プライマーの長さは１５～１００ヌクレオチドであり、たとえば、１５～７５ヌクレオチド、１５～４０ヌクレオチド、１７～３５ヌクレオチド、１８～３０ヌクレオチド、または２０～６５ヌクレオチド（それぞれ、範囲の上限と下限を含む）である。一部の態様では、プライマーの長さの範囲は、５０ヌクレオチド未満、４０ヌクレオチド未満、３０ヌクレオチド未満、２０ヌクレオチド未満、１０ヌクレオチド未満、または５ヌクレオチド未満である。一部の態様では、プライマーの長さの範囲は５～５０ヌクレオチドであり、たとえば、５～４０ヌクレオチド、５～２０ヌクレオチド、または５～１０ヌクレオチド（それぞれ、範囲の上限と下限を含む）である。一部の態様では、標的増幅産物の長さは５０～１００ヌクレオチドであり、たとえば、６０～８０ヌクレオチド、または６０～７５ヌクレオチド長（それぞれ、範囲の上限と下限を含む）である。一部の態様では、標的増幅産物の長さの範囲は、５０ヌクレオチド未満、２５ヌクレオチド未満、１５ヌクレオチド未満、１０ヌクレオチド未満、または５ヌクレオチド未満である。一部の態様では、標的増幅産物の長さの範囲は、５～５０ヌクレオチドであり、たとえば、５～２５ヌクレオチド、５～１５ヌクレオチド、または５～１０ヌクレオチド（それぞれ、範囲の上限と下限を含む）である。一部の態様では、ライブラリはマイクロアレイを含まない。一部の態様では、ライブラリはマイクロアレイを含む。

【０５００】

一部の態様では、一部（たとえば、少なくとも８０％、９０％、または９５％）またはすべてのアダプターまたはプライマーは、隣接するヌクレオチド間に天然のホスホジエステル結合以外の１つ以上の結合を有する。このような結合の例としては、ホスホルアミド結合、ホスホロチオアート結合、ホスホロジチオアート結合が挙げられる。一部の態様では、一部（少なくとも８０％、９０％、または９５％）またはすべてのアダプターまたはプライマーは、チオホスファート（モノチオホスファートなど）を３’末端のヌクレオチドと３’末端から二番目のヌクレオチドとの間に有する。一部の態様では、一部（たとえば、少なくとも８０％、９０％、または９５％）またはすべてのアダプターまたはプライマーは、チオホスファート（モノチオホスファートなど）を３’末端の最後の２個、３個、４個、または５個のヌクレオチドの間に有する。一部の態様では、一部（たとえば、少なくとも８０％、９０％、または９５％）またはすべてのアダプターまたはプライマーは、チオホスファート（モノチオホスファートなど）を、３’末端の最後の１０個のヌクレオチドのうち少なくとも１個、２個、３個、４個、または５個のヌクレオチドの間に有する。一部の態様では、このようなプライマーは切断または分解される可能性がより低い。一部の態様では、プライマーは酵素切断部位（プロテアーゼ切断部位など）を含まない。

【０５０１】

多重ＰＣＲ法とライブラリのさらなる例は、２０１２年１１月２１日出願の米国特許出願第１３／６８３，６０４号（米国特許出願公開第２０１３／０１２３１２０号）明細書、および２０１４年5月１６日出願の米国特許出願第６１／９９４，７９１号明細書（これらの文献それぞれの全内容は参照によって本明細書に組み込まれる）。これらの方法およびライブラリは、本明細書に開示の任意の試料の分析に利用可能であり、本発明の任意の方法に利用可能である。

【０５０２】

組換え検出用ライブラリのプライマーの例

一部の態様では、プライマーライブラリに含まれるプライマーを、１つ以上の既知の組換えホットスポットで組換え（ヒトの相同染色体間の交叉など）が起こっているかどうかを決定するよう設計する。染色体間でどのような交叉が起こったかを知ることで、個体に関するより正確な相決定済み遺伝データを決定できる。組換えホットスポットは、組換え事象が集中しやすい染色体局所領域である。組換えホットスポットは、平均よりも組換え頻度が低い領域である「コールドスポット」に隣接している場合が多い。組換えホットスポットは類似の形態を共有する傾向があり、約１ｋｂ～２ｋｂの長さである。ホットスポットの分布は、ＧＣ含量および反復要素の分布と正の相関を有する。一部のホットスポット活性には、部分的に縮重した１３マーのモチーフであるＣＣＮＣＣＮＴＮＮＣＣＮＣが関与する。ＰＲＤＭ９と呼ばれる亜鉛フィンガータンパク質がこのモチーフに結合し、その位置で組換えを開始することがわかっている。複数の組換えホットスポットの中心間の距離の平均は約８０ｋｂであると報告されている。一部の態様では、複数の組換えホットスポットの中心間の距離は約３ｋｂ～約１００ｋｂの範囲である。公知のデータベースにはヒトの既知の組換えホットスポットが多数含まれている（HUMHOTデータベース、International HapMap Projectデータベースなど（たとえば、以下を参照：Nishant et al., “HUMHOT: a database of human meiotic recombination hot spots,” Nucleic Acids Research, 34: D25-D28, 2006, Database issue；Mackiewicz et al., “Distribution of Recombination Hotspots in the Human Genome - A Comparison of Computer Simulations with Real Data” PLoS ONE 8(6): e65272, doi:10.1371/journal.pone.0065272；およびワールドワイドウェブのhapmap.ncbi.nlm.nih.gov/downloads/index.html.en（それぞれの全内容は参照によって本明細書に組み込まれる））。

【０５０３】

一部の態様では、プライマーライブラリに含まれるプライマーを組換えホットスポット（たとえばヒトの既知の組換えホットスポット）またはその付近に集中させる。一部の態様では、対応する増幅産物を用いて組換えホットスポット内またはその付近の配列を決定し、組換えがその特定のホットスポットで起こったかどうか（たとえば、その増幅産物の配列が、組換えが起こった場合に期待される配列であるか、組換えが起こらなかった場合に期待される配列であるか）を決定できる。一部の態様では、プライマーを、組換えホットスポット（と、必要に応じ、組換えホットスポットに隣接する配列）の一部または全体を増幅するように設計する。一部の態様では、ロングリード配列決定（Illumina社が開発したMoleculo Technologyを用いて約１０ｋｂ以下の配列を決定する配列決定など）またはペアエンド配列決定を用いて組換えホットスポットの一部または全体を配列決定する。組換え事象が起こったかどうかの認識を用いて、どのハプロタイプブロックがホットスポットに隣接しているかを決定する。必要であれば、特定のハプロタイプブロックの存在を、そのハプロタイプブロック内の領域に特異的なプライマーを用いて確認できる。一部の態様では、既知の組換えホットスポットの間に交叉がないことが推測される。一部の態様では、プライマーライブラリに含まれるプライマーを染色体の末端またはその付近に集中させる。たとえば、このようなプライマーを用いて、染色体の末端の特定のアームまたはセクションが存在するかどうかを決定することができる。一部の態様では、プライマーライブラリに含まれるプライマーを組換えホットスポットまたはその付近と染色体の末端またはその付近に集中させる。

【０５０４】

一部の態様では、プライマーライブラリは、組換えホットスポット（たとえばヒトの既知の組換えホットスポット）に特異的かつ／または組換えホットスポットの付近（たとえば組換えホットスポットの５’末端または３’末端から１０ｋｂ以内、８ｋｂ以内、５ｋｂ以内、３ｋｂ以内、２ｋｂ以内、１ｋｂ以内、または０．５ｋｂ以内）の領域に特異的な１個以上のプライマー（たとえば、少なくとも５個、１０個、５０個、１００個、２００個、５００個、７５０個、１，０００個、２，０００個、５，０００個、７，５００個、１０，０００個、２０，０００個、２５，０００個、３０，０００個、４０，０００個、または５０，０００個の異なるプライマーまたは異なるプライマー対）を含む。一部の態様では、少なくとも１個、５個、１０個、２０個、４０個、６０個、８０個、１００個、または１５０個の異なるプライマー（またはプライマー対）が、同じ組換えホットスポットに特異的であるか、同じ組換えホットスポットまたはその付近の領域に特異的である。一部の態様では、少なくとも１個、５個、１０個、２０個、４０個、６０個、８０個、１００個、または１５０個の異なるプライマー（またはプライマー対）が組換えホットスポットの間の領域（たとえば組換えが起こった可能性の低い領域）に特異的であり、これらのプライマーを用いてハプロタイプブロック（たとえば、組換えが起こったかどうかに応じて期待されるハプロタイプブロック）の存在を確認できる。一部の態様では、プライマーライブラリに含まれるプライマーのうち少なくとも１０％、２０％、３０％、４０％、５０％、６０％、７０％、８０％、または９０％が組換えホットスポットに特異的かつ／または組換えホットスポットの付近（たとえば組換えホットスポットの５’末端または３’末端から１０ｋｂ以内、８ｋｂ以内、５ｋｂ以内、３ｋｂ以内、２ｋｂ以内、１ｋｂ以内、または０．５ｋｂ以内）の領域に特異的である。一部の態様では、プライマーライブラリを用いて、５個以上、１０個以上、５０個以上、１００個以上、２００個以上、５００個以上、７５０個以上、１，０００個以上、２，０００個以上、５，０００個以上、７，５００個以上、１０，０００個以上、２０，０００個以上、２５，０００個以上、３０，０００個以上、４０，０００個以上、または５０，０００個以上の異なる組換えホットスポット（たとえばヒトの既知の組換えホットスポット）で組換えが起こったかどうかを決定する。一部の態様では、組換えホットスポットまたはその付近の領域に対するプライマーが標的とする領域は、ゲノム上のその部分に概ね均等に広がる。一部の態様では、少なくとも１個、５個、１０個、２０個、４０個、６０個、８０個、１００個、または１５０個の異なるプライマー（またはプライマー対）が、染色体の末端またはその付近の領域（たとえば、染色体の末端から２０ｍｂ以内、１０ｍｂ以内、５ｍｂ以内、１ｍｂ以内、０．５ｍｂ以内、０．１ｍｂ以内、０．０１ｍｂ以内、または０．００１ｍｂ以内の領域）に特異的である。一部の態様では、プライマーライブラリに含まれるプライマーのうち少なくとも１０％、２０％、３０％、４０％、５０％、６０％、７０％、８０％、または９０％が染色体の末端またはその付近の領域（たとえば、染色体の末端から２０ｍｂ以内、１０ｍｂ以内、５ｍｂ以内、１ｍｂ以内、０．５ｍｂ以内、０．１ｍｂ以内、０．０１ｍｂ以内、または０．００１ｍｂ以内の領域）に特異的である。一部の態様では、少なくとも１個、５個、１０個、２０個、４０個、６０個、８０個、１００個、または１５０個の異なるプライマー（またはプライマー対）が、染色体内の微小欠失である可能性のある部分に含まれる領域に特異的である。一部の態様では、プライマーライブラリに含まれるプライマーのうち少なくとも１０％、２０％、３０％、４０％、５０％、６０％、７０％、８０％、または９０％が、染色体内の微小欠失である可能性のある部分に含まれる領域に特異的である。一部の態様では、プライマーライブラリに含まれるプライマーのうちの少なくとも１０％、２０％、３０％、４０％、５０％、６０％、７０％、８０％、または９０％が、染色体内の組換えホットスポット、組換えホットスポットの付近の領域、染色体の末端またはその付近の領域、あるいは微小欠失である可能性のある部分に含まれる領域に特異的である。

【０５０５】

キットの例

一側面では、本発明はキット（たとえば、本明細書に記載の任意の方法によって染色体

分節または染色体全体の欠失および／または重複を検出するために核酸試料中の標的座位を増幅するためのキット）を特徴とする。一部の態様では、キットは本発明の任意のプライマーライブラリを含むことができる。一態様では、キットは複数の内側順方向プライマー、必要に応じて複数の内側逆方向プライマー、必要に応じて外側順方向プライマーおよび外側逆方向プライマーを含む。ここで、プライマーの各々は、標的である染色体または染色体分節（必要に応じて、追加の染色体または染色体分節）の標的部位（たとえば多型部位）の１つに接する上流および／または下流のＤＮＡ領域にハイブリダイズするよう設計される。一部の態様では、キットは、たとえば本明細書に記載の任意の方法によって１つ以上の染色体分節または染色体全体の１つ以上の欠失および／または重複を検出するためにプライマーライブラリを用いて標的座位を増幅するための説明書を含む。

【０５０６】

ある態様では、本発明のキットは、染色体異数性およびＣＮＶ決定を検出するためのプライマー対（たとえば、ＣＮＶの検出（CoNVERGe = Copy Number Variant Events Revealed Genotypically）および／またはＳＮＶなどの染色体異数性の検出のための超多重化反応用のプライマー対）を提供する。これらの態様では、キットは、同送される少なくとも１００個、２００個、２５０個、３００個、５００個、１０００個、２０００個、２５００個、３０００個、５０００個、１０，０００個、２０，０００個、２５，０００個、２８，０００個、５０，０００個、または７５，０００個、かつ最大で２００個、２５０個、３００個、５００個、１０００個、２０００個、２５００個、３０００個、５０００個、１０，０００個、２０，０００個、２５，０００個、２８，０００個、５０，０００個、７５，０００個、または１００，０００個の範囲のプライマー対を含むことができる。１つの容器（１つの管または箱など）あるいは複数の管または箱などにプライマー対を入れることができる。ある態様では、販売者がプライマー対をあらかじめ限定して共に販売し、他の態様では、顧客が注文する遺伝子標的および／またはプライマーを選択して販売者がプライマープールを作製し、１つの管や複数の管に入れずに顧客に発送する。ある例示の態様では、キットは、ＣＮＶおよびＳＮＶ、特に少なくとも１種の癌と相関することが知られているＣＮＶおよびＳＮＶを検出するためのプライマーを含む。

【０５０７】

本発明の一部の態様による循環ＤＮＡ検出用キットは、循環ＤＮＡ検出のための標準および／または対照を含む。たとえば、ある態様では、標準および／または対照を、増幅反応を実行するのに用いる本明細書に記載のプライマー（CoNVERGeを実行するためのプライマーなど）と併せて販売、必要に応じて発送・梱包する。ある態様では、対照はＤＮＡなどのポリヌクレオチドを含む。ポリヌクレオチドとしては、ＣＮＶなどの１つ以上の染色体異数性を示し、かつ／または１つ以上のＳＮＶを含む単離ゲノムＤＮＡが挙げられる。ある態様では、標準および／または対照はPlasmArt標準と呼ばれ、ＣＮＶを示す（特に、特定の遺伝性疾患、および癌などの特定の病状において）ことがわかっているゲノムの領域と同一の配列と、血漿に本来見られる無細胞ＤＮＡ断片のサイズ分布を反映するサイズ分布とを有するポリヌクレオチドを含む。PlasmArt標準を作製するための方法の例を、本明細書に記載の実施例に示す。一般的には、染色体異数性を有することがわかっている起源から得たゲノムＤＮＡを単離し、断片化し、精製してサイズ選択する。

【０５０８】

したがって、上に要約した方法で調製した単離ポリヌクレオチド試料を、染色体異数性および／またはＳＮＶを示さないことがわかっているＤＮＡ試料に、生体内の無細胞ＤＮＡにみられる濃度と類似の濃度（たとえば、ＤＮＡが液体中、０．０１％～２０％、０．１％～１５％、または４％～１０％）で添加することにより、人工無細胞ＤＮＡポリヌクレオチド標準および／または対照を作製できる。これらの標準／対照は、試験法の設計、特性評価、開発、および／または検証のための対照として、また、試験（ＣＬＩＡ（Clinical Laboratory Improvement Amendment；臨床検査改善修正法案）認定された研究所で実施される癌試験など）の際の品質管理標準として、および／あるいは研究用試験キットまたは診断試験キットに含まれる標準として使用可能である。

【０５０９】
正規化／修正方法の例
一部の態様では、異なる座位、染色体分節、または染色体に関する測定値を、偏り（ＧＣ含量の違いによる偏り、または増幅効率におけるその他の違いによる偏りなど）や配列決定エラーに応じて調整する。一部の態様では、同じ座位の異なるアレルに関する測定値を、アレル間で見られる代謝、アポトーシス、ヒストン、不活化、および／または増幅の違いに応じて調整する。一部の態様では、ＲＮＡ内の同じ座位の異なるアレルに関する測定値を、異なるＲＮＡアレル間で見られる転写速度または安定性の違いに応じて調整する。

【０５１０】
遺伝データの相決定方法の例
一部の態様では、遺伝データを相決定するための本明細書に記載の方法または任意の既知の方法で遺伝データを相決定する（たとえば、２００９年２月９日出願の国際公開第２００９／１０５５３１号パンフレット、および２００９年８月４日出願の国際公開第２０１０／０１７２１４号パンフレット；２０１２年１１月２１日出願の米国特許出願公開第２０１３／０１２３１２０号明細書；２０１０年１０月７日出願の米国特許出願公開第２０１１／００３３８６２号明細書；２０１０年８月１９日出願の米国特許出願公開第２０１１／００３３８６２号明細書；２０１１年２月３日出願の米国特許出願公開第２０１１／０１７８７１９号明細書；２００８年３月１７日出願の米国特許第８，５１５，６７９号明細書；２００６年１１月２２日出願の米国特許出願公開第２００７／０１８４４６７号明細書；２００８年３月１７日出願の米国特許出願公開第２００８／０２４３３９８号明細書；および２０１４年５月１６日出願の米国特許出願第６１／９９４，７９１号明細書を参照；これらの文献それぞれの全内容は参照によって本明細書に組み込まれる）。一部の態様では、対象ＣＮＶを含むとわかっているか、考えられる１つ以上の領域について相を決定する。一部の態様では、ＣＮＶ領域に隣接する１つ以上の領域および／または１つ以上の基準領域についても相を決定する。一態様では、個体（たとえば、本発明の方法による試験の対象である個体、あるいは懐胎中の胎児または胎芽の親などの血縁者）の遺伝データを、その個体から採取した半数体組織を測定することにより（たとえば１個以上の精子または卵子を測定することにより）、推測によって相を決定する。一態様では、１人以上の第一度近親者（個体の両親（たとえば個体の父親の精子を用いる）または兄弟姉妹など）の測定遺伝型データを用いた推測によって個体の遺伝データを相決定する。

【０５１１】
一態様では、希釈（ＤＮＡまたはＲＮＡを１つ以上のウェルに希釈して入れる）によって、たとえばデジタルＰＣＲを用いて個体の遺伝データを相決定する。一部の態様では、各ウェルに存在する各ハプロタイプのコピーが約１個のみだと期待されるようになるまでＤＮＡまたはＲＮＡを希釈し、次いで１つ以上のウェルに含まれるＤＮＡまたはＲＮＡを測定する。一部の態様では、染色体が緊密な束になる有糸分裂期に細胞を停止させ、微小流体工学を用いて個々の染色体を別個のウェルに入れる。ＤＮＡまたはＲＮＡは希釈されているため、２個以上のハプロタイプが同じ画分（または管）に入っている可能性はない。したがって、事実上、管には単一のＤＮＡ分子が入っているため、単一のＤＮＡ分子またはＲＮＡ分子のハプロタイプを決定できる。一部の態様では、前記方法は、ＤＮＡまたはＲＮＡの試料を複数の画分に分割する（それらの画分のうち少なくとも１つが染色体対のうちの１つの染色体または１つの染色体分節を含むように）ことと、少なくとも１つの画分に含まれるＤＮＡ試料またはＲＮＡ試料の遺伝型を解析する（たとえば２個以上の多型座位の存在を決定する）ことによってハプロタイプを決定することとを含む。一部の態様では、遺伝型解析は、配列決定（ショットガン配列決定または単分子配列決定など）、多型座位を検出するＳＮＰアレイ、または多重ＰＣＲを含む。一部の態様では、遺伝型解析は、ＳＮＰアレイを用いて多型座位（たとえば、少なくとも１００個、２００個、５００個、７５０個、１，０００個、２，０００個、５，０００個、７，５００個、１０，０００個、２０，０００個、２５，０００個、３０，０００個、４０，０００個、５０，０

００個、７５，０００個、または１００，０００個の異なる多型座位）を検出することを含む。一部の態様では、遺伝型解析は多重ＰＣＲの利用を含む。一部の態様では、前記方法は、画分に含まれる試料を、少なくとも１００個、２００個、５００個、７５０個、１，０００個、２，０００個、５，０００個、７，５００個、１０，０００個、２０，０００個、２５，０００個、３０，０００個、４０，０００個、５０，０００個、７５，０００個、または１００，０００個の異なる多型座位（ＳＮＰなど）と同時にハイブリダイズするプライマーのライブラリと接触させて反応混合物を生成することと；反応混合物をプライマー伸長反応条件に供して増幅産物を生成し、この増幅産物を高処理シーケンサで測定して配列決定データを作成する。一部の態様では、ＲＮＡ（たとえばｍＲＮＡ）を配列決定する。ｍＲＮＡはエキソンのみを含むため、ｍＲＮＡを配列決定することにより、ゲノム内の長い距離（たとえば数メガベース）にわたって多型座位（ＳＮＰなど）のアレルを決定できる。一部の態様では、個体のハプロタイプを染色体選別によって決定する。例示の染色体選別方法は、染色体が緊密な束になる有糸分裂期に細胞を停止させ、微小流体工学を利用して個々の染色体を別個のウェルに入れることを含む。別の方法は、蛍光活性化細胞選別（ＦＡＣＳ）を介した単一の染色体の選別によって単一の染色体を収集することを含む。標準の方法（配列決定またはアレイなど）を用いて単一の染色体上のアレルを特定し、個体のハプロタイプを決定することができる。

【０５１２】

一部の態様では、たとえばIllumina社の開発したMoleculo Technologyを利用して、ロングリード配列決定によって個体のハプロタイプを決定する。一部の態様では、ライブラリ調製工程は、ＤＮＡを切断して約１０ｋｂのサイズなどの断片に断片化し、断片を希釈してウェルに入れる（１つのウェルに約３，０００個の断片が入るように）ことと、各ウェル内の断片を長距離ＰＣＲ（long-range PCR）によって増幅し、短い断片に切断してその断片にバーコードを付けることと、各ウェルのバーコードを付けた断片を共にプールし、すべてを配列決定することとを含む。配列決定後、計算工程は、断片に付けたバーコードに基づいて各ウェルのリードを分けて断片に分類し、重複するヘテロ接合性のＳＮＶの断片を集めてハプロタイプブロックを得ることと、相決定済みの基準パネルに基づいてそのブロックを統計的に相決定して長いハプロタイプコンティグを作製することとを含む。

【０５１３】

一部の態様では、個体の血縁者から得たデータを用いて個体のハプロタイプを決定する。一部の態様では、ＳＮＰアレイを用いて、個体と個体の血縁者のＤＮＡまたはＲＮＡの試料に少なくとも１００個、２００個、５００個、７５０個、１，０００個、２，０００個、５，０００個、７，５００個、１０，０００個、２０，０００個、２５，０００個、３０，０００個、４０，０００個、５０，０００個、７５，０００個、または１００，０００個の異なる多型座位が存在することを決定する。一部の態様では、前記方法は、個体および／または個体の血縁者から採取したＤＮＡ試料を、少なくとも１００個、２００個、５００個、７５０個、１，０００個、２，０００個、５，０００個、７，５００個、１０，０００個、２０，０００個、２５，０００個、３０，０００個、４０，０００個、５０，０００個、７５，０００個、または１００，０００個の異なる多型座位（ＳＮＰなど）と同時にハイブリダイズするプライマーのライブラリと接触させて反応混合物を得ることと；反応混合物をプライマー伸長反応条件に供して増幅産物を生成し、この増幅産物を高処理シーケンサで測定して配列決定データを作成することを含む。

【０５１４】

一態様では、集団に基づくハプロタイプ頻度を用いて最も尤度の高い相を推測するコンピュータプログラムを利用して個体の遺伝データを相決定する（HapMapに基づく相決定など）。たとえば、母集団における既知のハプロタイプブロック（たとえば公知のHapMap ProjectとPerlegen Human Haplotype Projectについて作成したもの）を利用する統計学的手法により、二倍体データから半数体データの集合を直接的に演繹することができる。ハプロタイプブロックは、本質的には、多様な集団において反復的に見られる相関するアレルである。これらのハプロタイプブロックの多くは古く一般的であるため、これらを用い

て二倍体の遺伝型からハプロタイプを予測してもよい。この課題を達成する公的に利用可能なアルゴリズムには、不完全な系統樹の方法、共役事前確率に基づくベイズの方法、集団遺伝学による事前確率が含まれる。このようなアルゴリズムの一部では隠れマルコフモデルを用いる。

【０５１５】

一態様では、遺伝型データからハプロタイプを推定するアルゴリズム（たとえば局在性ハプロタイプのクラスタリングを用いたアルゴリズム）を用いて個体の遺伝データを相決定する（たとえば、Browning and Browning, “Rapid and Accurate Haplotype Phasing and Missing-Data Inference for Whole-Genome Association Studies By Use of Localized Haplotype Clustering” Am J Hum Genet. Nov 2007; 81(5): 1084-1097を参照；この文献の全内容は参照によって本明細書に組み込まれる）である。例示のプログラムは、Beagle version: 3.3.2またはversion 4（ワールドワイドウェブのhfaculty.washington.edu/browning/beagle/beagle.htmlから利用可能であり、その全内容は参照によって本明細書に組み込まれる）である。

【０５１６】

一態様では、遺伝型データからハプロタイプを推定するアルゴリズム（距離に伴う連鎖不平衡の減衰、遺伝型解析済みのマーカーの順序と間隔、欠落したデータの補完、組換え率の推定、またはそれらの組み合わせを用いるアルゴリズムなど）を用いて個体の遺伝データを相決定する（たとえば、Stephens and Scheet, “Accounting for Decay of Linkage Disequilibrium in Haplotype Inference and Missing-Data Imputation” Am. J. Hum. Genet. 76:449-462, 2005を参照；この文献の全内容は参照によって本明細書に組み込まれる）。例示のプログラムは、PHASE v.2.1またはv2.1.1.（ワールドワイドウェブのstephenslab.uchicago.edu/software.htmlから入手可能；この全内容は参照によって本明細書に組み込まれる）。

【０５１７】

一態様では、集団の遺伝型データからハプロタイプを推定するアルゴリズム（たとえば、隠れマルコフモデルに従いクラスターの構成要素を染色体に沿って継続的に変更できるアルゴリズム）を用いて個体の遺伝データを相決定する。この手法は柔軟であり、連鎖不平衡の「ブロックのような」パターンと、距離に伴う連鎖不均衡の減少を考慮する（たとえば、Scheet and Stephens, “A fast and flexible statistical model for large-scale population genotype data: applications to inferring missing genotypes and haplotypic phase.” Am J Hum Genet, 78:629-644, 2006を参照；この文献の全内容は参照によって本明細書に組み込まれる）。例示のプログラムはfastPHASE（ワールドワイドウェブのstephenslab.uchicago.edu/software.htmlより入手可能であり、全内容が参照によって本明細書に組み込まれる）である。

【０５１８】

一態様では、たとえば以下の参考データセットのうち１つ以上を用いる方法などの遺伝型補完方法によって個体の遺伝データを相決定する：HapMapデータセット、複数のＳＮＰチップで遺伝型解析した対照のデータセット、１０００人ゲノムプロジェクト（1,000 Genomes Project）による密に分類された試料。例示の手法は、複数の基準パネルに及ぶ情報を組み合わせて精度を高める柔軟なモデリングフレームワークである（たとえば、Howie, Donnelly, and Marchini (2009) “A flexible and accurate genotype imputation method for the next generation of genome-wide association studies.” PLoS Genetics 5(6): e1000529, 2009を参照；この文献の全内容は参照によって本明細書に組み込まれる）。例示のプログラムはIMPUTEまたはIMPUTE version 2（IMPUTE2とも呼ばれる）（ワールドワイドウェブのmathgen.stats.ox.ac.uk/impute/impute_v2.htmlで入手可能であり、全内容が参照によって本明細書に組み込まれる）である。

【０５１９】

一態様では、ハプロタイプを推測するアルゴリズム（たとえば、StephensがPHASE v2.1で開発したアルゴリズムなど、組換えを含む合祖の遺伝モデルに基づきハプロタイプを推

測するアルゴリズム）を用いて個体の遺伝データを相決定する。アルゴリズムの主な改善点は、二分木を用いて各個体に関する候補ハプロタイプの集合を表現することによるものである。これらの二分木による表現は、（１）PHASE v2.1で行われた冗長演算を避けることによってハプロタイプの事後確率の計算が高速化し、（２）最も尤度の高い経路（すなわちハプロタイプ）を迅速に探索することによってハプロタイプ推測問題における指数関数の面を克服する（たとえば、Delaneau, Coulonges and Zagury, “Shape-IT: new rapid and accurate algorithm for haplotype inference,” BMC Bioinformatics 9:540, 2008 doi:10.1186/1471-2105-9-540を参照；この文献の全内容は参照により本明細書に組み込まれる）。例示のプログラムはSHAPEIT（ワールドワイドウェブのmathgen.stats.ox.ac.uk/genetics_software/shapeit/shapeit.htmlで入手可能であり、全内容が参照により本明細書に組み込まれる）。

【０５２０】

一態様では、集団の遺伝型データからハプロタイプを推定するアルゴリズム（たとえばハプロタイプ断片の頻度を使用してより長いハプロタイプについて経験に基づく確率を得るアルゴリズム）を用いて個体の遺伝データを相決定する。一部の態様では、アルゴリズムは、ハプロタイプの局所的一貫性が最大になるようにハプロタイプを再構築する（例えば、Eronen, Geerts, and Toivonen, “HaploRec: Efficient and accurate large-scale reconstruction of haplotypes,” BMC Bioinformatics 7:542, 2006を参照；この文献の全内容は参照によって本明細書に組み込まれる）。例示のプログラムはHaploRec（たとえばHaploRec version 2.3.（ワールドワイドウェブのcs.helsinki.fi/group/genetics/haplotyping.htmlより入手可能；その全内容は参照によって本明細書に組み込まれる）。

【０５２１】

一態様では、集団の遺伝型データからハプロタイプを推定するアルゴリズム（たとえば、分割・結合（partition-ligation）法と期待値最大化に基づくアルゴリズムを使用するアルゴリズムなど）を用いて個体の遺伝データを相決定する（たとえば、Qin, Niu, and Liu, “Partition-Ligation-Expectation-Maximization Algorithm for Haplotype Inference with Single-Nucleotide Polymorphisms,” Am J Hum Genet. 71(5): 1242-1247, 2002を参照；この文献の全内容は参照により本明細書に組み込まれる）。例示的なプログラムは、PL-EM（ワールドワイドウェブのpeople.fas.harvard.edu/~junliu/plem/click.htmlで入手可能であり、全内容が参照により本明細書に組み込まれる）。

【０５２２】

一態様では、集団の遺伝型データからハプロタイプを推定するアルゴリズム（たとえば遺伝型を相決定してハプロタイプを得ることとブロック分割とを同時に行うためのアルゴリズムなど）を用いて個体の遺伝データを相決定する。一部の態様では、期待値最大化アルゴリズムを用いる（たとえば、Kimmel and Shamir, “GERBIL: Genotype Resolution and Block Identification Using Likelihood,” Proceedings of the National Academy of Sciences of the United States of America (PNAS) 102: 158-162, 2005を参照；この文献の全内容は参照により本明細書に組み込まれる）。例示的なプログラムはGERBILであり、GEVALT version 2 program（ワールドワイドウェブのacgt.cs.tau.ac.il/gevalt/で入手可能であり、全内容が参照によって本明細書に組み込まれる）の一部として入手可能である。

【０５２３】

一態様では、集団の遺伝型データからハプロタイプを推定するアルゴリズム（たとえば、ＥＭアルゴリズムを用い、相を特定しない遺伝型測定値に基づいてハプロタイプ頻度のＭＬ推定値を算出するアルゴリズム）を用いて個体の遺伝子データを相決定する。このアルゴリズムでは、一部の遺伝型の測定値が（たとえばＰＣＲの不具合によって）欠落することを考慮している。このアルゴリズムによって、個々のハプロタイプの多重補完が可能である（たとえば、Clayton, D. (2002), "SNPHAP: A Program for Estimating Frequencies of Large Haplotypes of SNPs"を参照；この文献の全内容は参照によって本明細書に組み込まれる）。例示的なプログラムはSNPHAP（ワールドワイドウェブのgene.cimr.cam.

ac.uk/clayton/software/snphap.txtで入手可能であり、全内容が参照により本明細書に組み込まれる）である。

【０５２４】

一態様では、集団の遺伝型データからハプロタイプを推定するアルゴリズム（たとえば、ＳＮＰ対について収集した遺伝型統計値に基づいてハプロタイプを推測するためのアルゴリズム）を用いて個体の遺伝データを相決定する。このソフトウェアを用いて、多数の長いゲノム配列（たとえばＤＮＡアレイから得たもの）を比較的正確に相決定できる。例示のプログラムは、入力された遺伝型の行列を受け、対応するハプロタイプの行列を出力する（たとえば、Brinza and Zelikovsky, “2SNP: scalable phasing based on 2-SNP haplotypes,” Bioinformatics.22(3):371-3, 2006を参照；この文献の全内容は参照によって本明細書に組み込まれる）。例示のプログラムは、2SNPである（ワールドワイドウェブのalla.cs.gsu.edu/~software/2SNPより入手可能であり、全内容が参照によって本明細書に組み込まれる）。

【０５２５】

様々な態様では、染色体または染色体分節の様々な位置で染色体が交叉する確率に関するデータを用いて（たとえば、HapMapデータベースで見ることができるような組換えデータを用いて、任意の間隔について組換えリスクスコアを作成することによって）個体の遺伝データを相決定し、その染色体または染色体分節に存在する多型アレル間の従属性をモデル化する。一部の態様では、それぞれ染色体または染色体分節の有し得る異なる状態に関する複数の仮説（たとえば、個体に由来する１個以上の細胞のゲノムにおいて、第一の相同染色体分節のコピー数が第二の相同染色体分節のコピー数に比べて過剰出現している；第一の相同染色体分節が重複している；第二の相同染色体分節が欠失している；あるいは第一と第二の相同染色体分節が均等に出現している、など）を（コンピュータなどで）作成し；染色体上の多型座位におけるアレル数の期待値のモデル（たとえば同時分布モデル）を各仮説について（コンピュータなどで）構築し；前記の同時分布モデルおよびアレル数を用いて各仮説の条件付き確率を（コンピュータなどで）決定し；最も確率の高い仮説を選択する。一部の態様では、アレル数に関する同時分布モデルの構築と、各仮説が正しい相対的確率を決定する工程を、基準染色体の使用を必要としない方法で実行する。

【０５２６】

一態様では、個体の１人以上の血縁者（１人以上の親、兄弟姉妹、子、胎児、胎芽、祖父母、おじ、おば、いとこなど）の遺伝データを用いて個体の遺伝データを相決定する。一態様では、個体の１人以上（たとえば１人、２人、３人、またはそれ以上）の遺伝的子孫（胎芽、胎児、生後の子、または流産物試料など）の遺伝データを用いて個体の遺伝データを相決定する。一態様では、親（懐胎中の胎児または胎芽の親など）の遺伝データを、もう一方の親に関する相決定済みのハプロタイプデータと、両方の親の１人以上の遺伝的子孫の相決定前の遺伝データを用いて相決定する。

【０５２７】

一部の態様では、個体（癌を有すると考えられる個体、胎児、または胎芽）から採取した試料（たとえば腫瘍生検などの生検、血液試料、血漿試料、血清試料、あるいは、大部分または全体が、対象ＣＮＶを有する細胞、ＤＮＡ、またはＲＮＡである可能性の高い別の試料）を分析し、対象ＣＮＶ（欠失または重複など）を含むとわかっているないしは考えられる１つ以上の領域について、相を決定する。一部の態様では、試料は高い腫瘍率（たとえば３０％、４０％、５０％、６０％、７０％、８０％、９０％、９５％、９８％、９９％、または１００％）を有する。一部の態様では、胎児またはその妊娠中の母親から採取した試料（たとえば母親の全血試料、母親の血液試料から単離した細胞、母親の血漿試料、母親の血清試料、羊水穿刺による試料、胎盤組織試料、胎児死亡後の胎児の組織、胎児に由来するその他の試料、あるいは、大部分または全体が、対象ＣＮＶを有する細胞、ＤＮＡ、またはＲＮＡである可能性の高い別の試料）を分析し、対象ＣＮＶ（欠失または重複など）を含むとわかっているないしは考えられる１つ以上の領域について、相を決定する。一部の態様では、試料は胎児画分が多い（たとえば２５％、３０％、４０％、５

０％、６０％、７０％、８０％、９０％、９５％、９８％、９９％、または１００％）。

【０５２８】

一部の態様では、試料はハプロタイプ不均衡またはいずれかの異数性を有する。一部の態様では、試料は２種のＤＮＡの任意の混合物を含む。ここで、この２種は、２つのハプロタイプの比が異なり、かつ少なくとも１つのハプロタイプを共有する。たとえば、胎児と母親の場合、母親が１：１、胎児が１：０（これに加えて父親のハプロタイプを有する）である。たとえば、腫瘍の場合、正常組織が１：１、腫瘍組織が１：０または１：２、１：３、１：４などである。一部の態様では、少なくとも１０個、１００個、５００個、１，０００個、２，０００個、３，０００個、５，０００個、８，０００個、または１０，０００個の多型座位を分析し、一部またはすべての座位のアレルの相を決定する。一部の態様では、試料は異数性になるような処理（たとえば長期間の細胞培養による異数性誘発など）を施した細胞または組織に由来する。

【０５２９】

一部の態様では、試料中のＤＮＡまたはＲＮＡの大部分または全体が対象ＣＮＶを有する。一部の態様では、対象ＣＮＶを有する１つ以上の標的細胞に由来するＤＮＡまたはＲＮＡの、試料中の全ＤＮＡまたはＲＮＡに対する比が、少なくとも８０％、８５％、９０％、９５％、または１００％である。欠失を有する試料に関して、欠失を有する細胞（あるいはＤＮＡまたはＲＮＡ）に存在するハプロタイプは１つのみである。この第一のハプロタイプは、標準の方法で欠失の領域に存在するアレルの同一性を決定することによって決定可能である。その欠失を有する細胞（あるいはＤＮＡまたはＲＮＡ）のみを含む試料では、これらの細胞に存在する第一のハプロタイプに由来するシグナルのみが存在する。欠失を有さない少量の細胞（少量の非癌性細胞など）（あるいはＤＮＡまたはＲＮＡ）では、これらの細胞（あるいはＤＮＡまたはＲＮＡ）に見られる第二のハプロタイプに由来するシグナルを無視することができる。欠失を有さない個体に由来するその他の細胞、ＤＮＡ、またはＲＮＡに存在する第二のハプロタイプは、推測によって決定可能である。たとえば、欠失を有さない個体に由来する細胞の遺伝型が（ＡＢ、ＡＢ）であり、その個体に関する相決定済みデータによって第一のハプロタイプは（Ａ、Ａ）であると示されている場合、もう一方のハプロタイプは（Ｂ、Ｂ）であると推測できる。

【０５３０】

欠失を有する細胞（あるいはＤＮＡまたはＲＮＡ）と欠失を有さない細胞（あるいはＤＮＡまたはＲＮＡ）の両方が存在する試料の場合も相を決定できる。たとえば、図１８または図２９に似た図を作成できる。図１８または図２９では、横軸は染色体に沿った個々の座位の線形位置を表し、縦軸はＡアレルのリード数をアレル全体（Ａ＋Ｂ）のリードにおける比率として表す。欠失に関する一部の態様では、パターンは、個体がヘテロ接合性であるＳＮＰを表す２つの中央バンドを含む（上のバンドは欠失のない細胞に由来するＡＢと欠失を有する細胞に由来するＡを表し、下のバンドは欠失のない細胞に由来するＡＢと欠失を有する細胞に由来するＢを表す）。一部の態様では、これらの２つのバンドの分離は、欠失を有する細胞、ＤＮＡまたはＲＮＡの比率が増加するほど大きくなる。したがって、Ａアレルの同一性を用いて第一のハプロタイプを決定し、Ｂアレルの同一性を用いて第二のハプロタイプを決定することができる。

【０５３１】

重複を有する試料の場合、重複を有する細胞（あるいはＤＮＡまたはＲＮＡ）に関するハプロタイプの余分なコピーが存在する。重複領域のこのハプロタイプを標準の方法で決定し、重複領域での存在量がより多いアレルの同一性を決定することができる。あるいは、重複していない領域のハプロタイプを標準の方法で決定し、存在量がより少ないアレルの同一性を決定することができる。１つのハプロタイプを決定すると、それ以外のハプロタイプを推測によって決定することができる。

【０５３２】

重複を有する細胞（あるいはＤＮＡまたはＲＮＡ）と重複のない細胞（あるいはＤＮＡまたはＲＮＡ）の両方が存在する試料の場合にも、欠失について上記した方法と類似の方

法を用いて相を決定できる。たとえば、図１８および図２９と似た図を作成することができる。図１８または図２９では、横軸は染色体に沿った個々の座位の線形位置を表し、縦軸はＡアレルのリード数をアレル全体（Ａ＋Ｂ）のリードにおける比率として表す。欠失に関する一部の態様では、パターンは、個体がヘテロ接合性であるＳＮＰを表す２つの中央バンドを含む（上のバンドは重複のない細胞に由来するＡＢと重複を有する細胞に由来するＡＡＢを表し、下のバンドは重複のない細胞に由来するＡＢと欠失を有する細胞に由来するＡＢＢを表す）。一部の態様では、これらの２つのバンドの分離は、重複を有する細胞、ＤＮＡまたはＲＮＡの比率が増加するにつれて大きくなる。したがって、Ａアレルの同一性を用いて第一のハプロタイプを決定し、Ｂアレルの同一性を用いて第二のハプロタイプを決定することができる。一部の態様では、１つ以上のＣＮＶ領域の相（たとえば、測定した領域に含まれる多型座位の少なくとも５０％、６０％、７０％、８０％、９０％、９５％、または１００％の相）を、癌を有することがわかっている個体から採取した試料（たとえば腫瘍生検または血漿試料）について決定し、癌の進行を監視する（たとえば寛解または再発を監視する）ために同じ個体のその後の試料の分析に使用する。一部の態様では、腫瘍率の高い試料（たとえば腫瘍負荷の高い個体から採取した腫瘍生検または血漿試料）を用いて、腫瘍率がより低いその後の試料（癌の治療を受けている個体または寛解中の個体から採取した血漿試料など）の分析に使用する相決定済みデータを得る。

【０５３３】

出生前診断に関する別の態様では、親の相決定済みのハプロタイプデータによって、父親に由来する複数の相同体の存在が検出される。これは、複数の胎児に由来する遺伝物質が母親の血液試料中に存在することを意味する。胎児において正倍数性であると期待される染色体に注目することによって、胎児が三染色体性を有する可能性を排除できる。また、胎児のＤＮＡが現在の父親に由来しないかどうかを決定することもできる。

【０５３４】

一部の態様では、本明細書に記載の方法のうち２つ以上を用いて個体の遺伝データを相決定する。一部の態様では、生物情報学的手法（たとえば、母集団に基づくハプロタイプ頻度を用いて最も尤度の高い相を推測すること）と分子生物学的手法（たとえば、分子の相決定方法であって、生物情報学に基づく推測の相決定済みデータではなく実際の相決定済みデータを得る、本明細書に記載のいずれかの方法）の両方を用いる。一部の態様では、対象者（以前の対象者など）から得た相決定済みデータを用いて、母集団データを精緻化する。たとえば、他の対象者から得た相決定済みデータを母集団データに加え、別の対象者について考えられるハプロタイプに関する事前確率を計算することができる。一部の態様では、他の対象者（以前の対象者など）から得た相決定済みデータを用いて、別の対象者について考えられるハプロタイプに関する事前確率を計算する。

【０５３５】

一部の態様では、確率的データを使用してもよい。たとえば、試料中のＤＮＡ分子の表現には確率的性質があることや、増幅や測定の様々な差があることから、２個の異なる座位、または所与の座位に存在する異なるアレルから測定したＤＮＡ分子の相対数は、混合物または個体に含まれる分子の相対数を表すとは限らない。常染色体上の所与の座位において正常な二倍体性の個体の遺伝型を、その個体の血漿に由来するＤＮＡの配列決定によって決定しようとした場合、１種類のアレルのみが確認される（ホモ接合性）か、ほぼ同数の２種類のアレルが確認される（ヘテロ接合性）ことが期待される。そのアレルにおいて１０個のＡアレル分子が確認され２個のＢアレル分子が確認された場合、その個体がその座位においてホモ接合性で２個のＢアレル分子はノイズまたは不純物によるものであるのかどうか、あるいは、その個体がヘテロ接合性でＢアレルの分子数がより少ないのは血漿中のＤＮＡ分子数のランダムな統計的変動、増幅の差、不純物、またはあらゆる他の原因によるものであるのかどうかは明らかではない。この場合、個体がホモ接合性である確率と、それに対応する、その個体がヘテロ接合性である確率を計算することができ、これらの確率的遺伝型をさらに計算に使用できる。

【０５３６】

当然ながら、所与のアレル比について、その比が個体におけるＤＮＡ分子の比を厳密に表す尤度は、観測した分子の数が多いほど大きくなる。たとえば、１００個のＡ分子と１００個のＢ分子を測定した場合、実際の比が５０％である尤度は、１０個のＡ分子と１０個のＢ分子を測定した場合よりも相当高い。一態様では、ベイズの理論をデータの詳細なモデルと併用し、ある観測値に基づいて特定の仮説が正しいという尤度を決定する。たとえば、２つの仮説、すなわち三染色体性である個体に対応する仮説と二染色体性である個体に対応する仮説を検討する場合、二染色体性の仮説が正しいという確率は、２種類のアレル分子をそれぞれ１００個観測した場合には、その２種類のアレル分子をそれぞれ１０個観測した場合に比べて相当高くなる。差、不純物または他の何らかのノイズ源によってデータにノイズが多くなるか、所与の座位での観測数が少なければ、観測データに基づいて最尤仮説が真である確率は低くなる。実際には、多くの座位にわたる確率を集約し、最尤仮説が正しい仮説であると決定する信頼度を高めることができる。一部の態様では、確率を組換えとは無関係に単純に集約する。一部の態様では、交叉を考慮して計算する。

【０５３７】

一態様では、確率的に相決定したデータを用いてコピー数多様性を決定する。一部の態様では、確率的に相決定したデータは、HapMapデータベースなどのデータ源から得た、集団に基づくハプロタイプブロック頻度データである。一部の態様では、確率的に相決定したデータは、分子的方法（たとえば、染色体の個々の分節を一反応あたり一分子になるまで希釈する、希釈による相決定）で相決定したハプロタイプデータであるが、確率論的ノイズのため、ハプロタイプの同一性を絶対的に知ることはできない。一部の態様では、確率的に相決定したデータは、ハプロタイプの同一性を高い確度で知ることができる、分子的方法で得たハプロタイプデータである。

【０５３８】

特定の染色体分節に欠失を有する細胞を個体が有するかどうかを個体から採取した血漿ＤＮＡを測定することによって判断することを医師が望むような、仮定の事例について考える。医師は、血漿ＤＮＡの由来する細胞がすべて二倍体であり同一の遺伝型である場合にヘテロ接合性の座位では２種類のアレルの各々について観測されるＤＮＡ分子の相対数はＡアレル５０％、Ｂアレル５０％に集中する１つの分布になるという知識を利用できる。しかし、血漿ＤＮＡの由来する細胞の画分が特定の染色体分節に欠失を有する場合、ヘテロ接合性の座位では２種類のアレルの各々について観測されるＤＮＡ分子の相対数は２つの分布に分かれると期待される。すなわち、１つは染色体分節のうちＢアレルを含む座位が欠失している場合で、Ａアレルは５０％超に集中し、もう１つは染色体分節のうちＡアレルを含む座位が欠失している場合で、Ａアレルは５０％未満に集中する。血漿ＤＮＡが由来する細胞が欠失を含む割合が大きいほど、これらの２つの分布は５０％から遠くなる。

【０５３９】

次の仮定の事例では、個体の体内細胞の一部に染色体領域の欠失があるかどうかを決定したい臨床医を想定する。臨床医はその個体からバキュテナー（真空採血管）または他の種類の採決管に血液を採取し、その血液を遠心分離して血漿層を単離することができる。臨床医は、ＤＮＡを血漿から単離し、おそらく標的化増幅またはその他の増幅、座位の捕捉（locus capture）による方法、サイズに基づく選択的濃縮（size enrichment）、または他の濃縮法によってＤＮＡの標的座位を濃縮することができる。臨床医は、ＳＮＰの集合におけるアレル数の測定などにより、すなわちアレル頻度データの作成により、定量ＰＣＲ、配列決定、マイクロアレイ、または試料中のＤＮＡ量を測定するその他の方法などの試験法を用いて、濃縮かつ／または増幅されたＤＮＡを分析することができる。臨床医が無細胞血漿ＤＮＡを標的化増幅法で増幅して増幅後のＤＮＡを配列決定し、染色体分節で確認され癌を示す６種類のＳＮＰにおいて以下の考えられるデータ例を得、個体がこれらのＳＮＰにおいてヘテロ接合性であった場合のデータ分析について、考察する。

【０５４０】

ＳＮＰ１：Ａアレルが４６０リード；Ｂアレルが５４０リード（Ａが４６％）

ＳＮＰ２：Ａアレルが５３０リード；Ｂアレルが４７０リード（Ａが５３％）
ＳＮＰ３：Ａアレルが４０リード；Ｂアレルが６０リード（Ａが４０％）
ＳＮＰ４：Ａアレルが４６リード；Ｂアレルが５４リード（Ａが４６％）
ＳＮＰ５：Ａアレルが５２０リード；Ｂアレルが４８０リード（Ａが５２％）
ＳＮＰ６：Ａアレルが２００リード；Ｂアレルが２００リード（Ａが５０％）

【０５４１】

このデータのセットからは、全ての細胞が二染色体性であり個体が正常である場合と、細胞（血漿中の無細胞ＤＮＡはこの細胞のＤＮＡから得られる）の一部が染色体に欠失または重複を有し、ＤＮＡ個体が癌を有し得る場合とを区別するのは難しいかもしれない。たとえば、最も尤度の高い２つの仮説は、個体がこの染色体分節に欠失を有し腫瘍率が６％であるということと、欠失した染色体分節がＳＮＰの集合について遺伝型（Ａ、Ｂ、Ａ、Ａ、Ｂ、Ｂ）または（Ａ、Ｂ、Ａ、Ａ、Ｂ、Ａ）を有するということである。ＳＮＰの集合に関する個体の遺伝型をこのように表現する場合、括弧内の最初の文字はＳＮＰ１のハプロタイプの遺伝型に対応し、二番目以降の文字はＳＮＰ２以降のハプロタイプの遺伝型にそれぞれ対応する。

【０５４２】

個体のその染色体分節におけるハプロタイプを決定する方法を用いて、２つの染色体分節のうちの１つのハプロタイプが（Ａ、Ｂ、Ａ、Ａ、Ｂ、Ｂ）であることがわかった場合、このことは最尤仮説に合致し、その個体がその分節に欠失を有しており癌または前癌性の細胞を有するという尤度の算出値は、相当増加する。一方、その個体がハプロタイプ（Ａ、Ａ、Ａ、Ａ、Ａ、Ａ）を有することがわかった場合、その個体がその染色体分節に欠失を有する尤度は相当減少し、欠失を有さないという仮説の尤度がより高くなるかもしれない（実際の尤度の値は、特にシステムのノイズの測定値など、他のパラメータに依存する）。

【０５４３】

個体のハプロタイプを決定する方法は多数存在し、その多くを本明細書の他の箇所に記載する。一部をここに列挙するが、これが網羅的であるということではない。１つの方法は、任意の所定の反応容量中に存在する各染色体領域由来の分子が約１個になるまで個体のＤＮＡ分子を希釈し、次いで配列決定などの方法で遺伝型を測定する、生物学的方法である。別の方法は情報科学に基づく方法で、様々なハプロタイプの頻度に関連する母集団データを確率的な方法で使用することができる。別の方法は、個体と、その個体とハプロタイプブロックを共有すると期待される、関連する１人以上の個体の二倍体データを測定し、ハプロタイプブロックを推測する方法である。別の方法は、欠失した分節または重複した分節を高濃度に有する組織試料を採取し、アレル不均衡に基づいてハプロタイプを決定することであり、たとえば、欠失を有する腫瘍組織の試料から得た遺伝型測定値を使用してその欠失領域の相決定済みデータを決定することができ、このデータを使用して、癌が切除後に再増殖したかどうかを判定することができる。

【０５４４】

実際には、所定の染色体分節上で一般的に２０個超、５０個超、１００個超、５００個超、１，０００個超、または５，０００個超のＳＮＰを測定する。

【０５４５】

相の決定、アレル比の予測、胎児の遺伝データ再構築の方法の例

一側面では、本発明は胎児の１つ以上のハプロタイプを決定するための方法を特徴とする。様々な態様では、この方法により、どの多型座位（ＳＮＰなど）が胎児に遺伝したかを決定し、どの相同体（組換え事象を含む）が胎児に存在するかを再現する（ことによって多型座位間の配列を推定する）。必要に応じ、本質的には胎児の全ゲノムを再現できる。胎児のゲノムに曖昧さ（交叉の間隔など）が残る場合、必要に応じて別の多型座位を分析することにより、この曖昧さを最小限にすることができる。様々な態様では、あらゆる曖昧さを所望のレベルまで低減するような密度で染色体の１つ以上を含むよう、多型座位を選択する。この方法は、胎児の対象となる多型またはその他の変異（欠失または重複な

ど）の検出に有用性が高い、というのも、この方法によって、対象となる多型またはその他の変異を胎児のゲノムで検出するのではなく、連鎖（たとえば、胎児のゲノムにおける連鎖した多型座位の存在）に基づいた検出が可能になるためである。たとえば、親が嚢胞性線維症（ＣＦ）に関連する変異を保有している場合、胎児の母親に由来する母親のＤＮＡと胎児に由来する胎児のＤＮＡとを含む核酸試料を分析し、そのＣＦ変異を有するハプロタイプを胎児のＤＮＡが有するかどうかを決定できる。特に、多型座位を分析して、ＣＦ変異を含むハプロタイプを胎児のＤＮＡが有するかどうかを、胎児のＤＮＡ中のＣＦ変異自体を検出する必要なく決定することができる。従って、変異を直接検出する必要なく１つ以上の変異（疾患関連変異など）について検査するのに役立つ。

【０５４６】

一部の態様では、前記方法は、たとえば本明細書に記載の任意の方法で親のハプロタイプ（たとえば胎児の母親または父親のハプロタイプ）を決定することを含む。一部の態様では、母親または父親の血縁者からのデータを使用せずにこの決定を行う。一部の態様では、希釈による方法の後に本明細書に記載のＳＮＰ遺伝型解析または配列決定を行うことによって親のハプロタイプを決定する。一部の態様では、母親（または父親）の血縁者から得たデータを用いて、本明細書に記載の任意の方法で母親（または父親）のハプロタイプを決定する。一部の態様では、父親と母親両方についてハプロタイプを決定する。

【０５４７】

この親のハプロタイプデータを用いて、親のハプロタイプが胎児に遺伝したかどうかを決定できる。一部の態様では、胎児の母親に由来する母親のＤＮＡと胎児に由来する胎児のＤＮＡとを含む核酸試料を、ＳＮＰアレイを用いて分析し、少なくとも１００個、２００個、５００個、７５０個、１，０００個、２，０００個、５，０００個、７，５００個、１０，０００個、２０，０００個、２５，０００個、３０，０００個、４０，０００個、５０，０００個、７５，０００個、または１００，０００個の異なる多型座位）を検出する。一部の態様では、胎児の母親に由来する母親のＤＮＡと胎児に由来する胎児のＤＮＡとを含む核酸試料を、少なくとも１００個、２００個、５００個、７５０個、１，０００個、２，０００個、５，０００個、７，５００個、１０，０００個、２０，０００個、２５，０００個、３０，０００個、４０，０００個、５０，０００個、７５，０００個、または１００，０００個の異なる多型座位（ＳＮＰなど）と同時にハイブリダイズするプライマーのライブラリと接触させて反応混合物を生成することによって分析する。一部の態様では、反応混合物をプライマー伸長反応条件に供して増幅産物を生成する。一部の態様では、増幅産物を高処理シーケンサで測定し、配列決定データを作成する。

【０５４８】

様々な態様では、先に述べたように、染色体または染色体分節の異なる位置における、染色体の交叉確率に関するデータを用いて（たとえば、HapMapデータベースで見ることができるような組換えデータを用いて任意の間隔について組換えリスクスコアを作成することで）、胎児のハプロタイプを決定し、その染色体または染色体分節に存在する多型アレル間の従属性をモデル化する。一部の態様では、前記方法は、ＳＮＰ（対象とする遺伝子または変異に隣接するＳＮＰなど）の物理的距離と、各位置固有の組換えの尤度および母親の血漿の遺伝学的測定の観測データを基に得た組換えデータとを考慮し、胎児の最も尤度の高い遺伝型を得る。次いで、これらのＳＮＰから得た標的化配列決定データまたはＳＮＰアレイデータに対してPARENTAL SUPPORT（商標）を実行し、どの相同体が両親から胎児に遺伝したのかを決定してもよい（たとえば、米国特許出願第１１／６０３，４０６号（米国特許出願公開第２００７／０１８４４６７号）明細書、米国特許出願第１２／０７６，３４８号（米国特許出願公開第２００８／０２４３３９８号）明細書、米国特許出願第１３／１１０，６８５号（米国特許出願公開第２０１１／０２８８７８０号）明細書、国際出願ＰＣＴ／ＵＳ０９／５２７３０号明細書（国際公開第２０１０／０１７２１４号パンフレット）、国際出願ＰＣＴ／ＵＳ１０／０５０８２４号明細書（国際公開第２０１１／０４１４８５号パンフレット）、米国特許出願第１３／３００，２３５号（米国特許出願公開第２０１２／０２７０２１２号）明細書、米国特許出願第１３／３３５，０４３

号（米国特許出願公開第２０１２／０１２２７０１号）明細書、米国特許出願第１３／６８３，６０４号明細書、米国特許出願第１３／７８０，０２２号明細書を参照；これらの文献それぞれの全内容は参照によって本明細書に組み込まれる）。

【０５４９】

１つの座位において可能性のあるアレルがＡとＢである一般的な例を想定する。特定のアレルを任意にＡまたはＢと識別する。特定のＳＮＰに関する親の遺伝型（遺伝的背景という）は、母親の遺伝型｜父親の遺伝型として表される。つまり、母親がホモ接合性であり父親がヘテロ接合性である場合、遺伝的背景はＡＡ｜ＡＢである。同様に、同一のアレルについて両親ともホモ接合性である場合、親の遺伝型はＡＡ｜ＡＡと表される。また、胎児がＡＢまたはＢＢの状態を有することは決してなく、Ｂアレルの配列リード数は少ないため、これを用いて、微量のＤＮＡ不純物や配列決定エラーなど、試験法と遺伝型解析プラットフォームのノイズ応答を決定できる。これらのノイズ応答は、期待される遺伝データプロファイルをモデル化するのに有用である。可能性のある遺伝的背景（母親｜父親）はＡＡ｜ＡＡ、ＡＡ｜ＡＢ、ＡＢ｜ＡＡ、ＡＢ｜ＡＢ、ＡＡ｜ＢＢの５種類のみである。他の遺伝的背景も、対称であるため同等である。同一のアレルについて両親がホモ接合性であるＳＮＰは、ノイズや汚染のレベルを決定するためにのみ情報的価値を有する。同一のアレルについて両親がホモ接合性でないＳＮＰは、胎児画分およびコピー数の数値を決定するうえで情報的価値を有する。

【０５５０】

ＳＮＰｉにおける各アレルのリード数を$N_{A,i}$および$N_{B,i}$とし、その座位における親の遺伝的背景をC_iとすると、特定の染色体に関するデータセットは$N_{AB}=\{N_{A,i}, N_{B,i}\}$（ｉ＝１．．．Ｎ）および$C=\{C_i\}$（ｉ＝１．．．Ｎ）で表される。胎児のゲノムの一部または全体を再現するため、必要に応じて、胎児が異数性（たとえば、染色体または染色体分節の欠失または余分なコピーなど）を有するかどうかを決定することができる。試験対象である個々の染色体または染色体分節それぞれについて、染色体の総数、各染色体の親起源、受精して子を形成した配偶子の形成中に組換えが起こった親の染色体上の位置に関する１つ以上の仮説の集合をＨとする。仮説Ｐ（Ｈ）の確率は、HapMapデータベースのデータや、倍数性状態のそれぞれに関連する事前の情報を用いて計算することができる。

【０５５１】

さらに、試料に含まれる胎児の無細胞ＤＮＡ画分をＦとする。可能性のあるＨ、Ｃ、Ｆの集合が与えられれば、分子試験法および配列決定プラットフォームのノイズ源のモデル化に基づき、N_{AB}とＰ（N_{AB}｜Ｈ，Ｆ，Ｃ）を計算できる。目的はＰ（Ｈ，Ｆ｜N_{AB}）が最大となる仮説Ｈと胎児画分Ｆを得ることである。標準的なベイズの統計的手法を用い、Ｆが０～１の場合の一様な確率分布を想定し、ＨおよびＦに関してＰ（N_{AB}｜Ｈ，Ｆ，Ｃ）Ｐ（Ｈ）の確率を最大化するという観点で、これを再計算することができる。現在ではこれらはすべてコンピュータで計算可能である。特定のコピー数と胎児の比率（たとえば、三染色体性かつＦ＝１０％）に関連し、可能なすべての親起源と交叉位置を含む、すべての仮説の確率を集約する。最も確率の高いコピー数仮説を試験結果として選択し、その仮説に関連する胎児画分を胎児画分とする。その仮説に関連する確率は結果の精度の計算値である。

【０５５２】

一部の態様では、アルゴリズムはコンピュータシミュレーションで、可能性のある胎児の遺伝パターン、試料のパラメータ、方法における増幅や測定の人為的結果によって生じ得る、非常に多数の仮説的配列決定データセットを作成する。より具体的には、このアルゴリズムは、まず、多数のＳＮＰにおける親の遺伝型と、HapMapデータベースから得た交叉頻度データを利用し、可能性のある胎児の遺伝型を予測する。次いで、可能性のある胎児の遺伝型の各々を用い、多様なパラメータ（胎児画分、リード深度プロファイルの期待値、試料中に存在する胎児のゲノム等価物、各ＳＮＰにおける増幅の偏りの期待値、複数のノイズパラメータなど）を考慮して、胎児を懐胎している母親から採取した混合試料か

ら測定される配列決定データについて、期待されるデータプロファイルを予測する。データモデルは、これらの仮説のそれぞれについて特定のパラメータのセットに基づいて期待される配列決定データまたはＳＮＰアレイデータを表す。モデル化したデータと測定データとの間のデータ適合性が最も良い仮説を選択する。

【０５５３】

必要に応じて、どのハプロタイプが胎児に遺伝したかについての結果を用いて、胎児に由来するＤＮＡまたはＲＮＡについてアレル比の期待値を計算できる。アレル比の期待値は、母親に由来する核酸と胎児に由来する核酸の両方を含む混合試料についても計算可能である（これらのアレル比は、各アレルの総量（試料中の母親の核酸と胎児の核酸に由来するアレルの量を含む）の測定に関して期待される値を示す。第一の相同染色体分節の過剰出現の程度を特定する異なる仮説についてアレル比の期待値を計算することができる。

【０５５４】

一部の態様では、本方法は、胎児が以下の状態の１つ以上を有するかどうかを決定することを含む：嚢胞性線維症、ハンチントン病、脆弱Ｘ症候群、サラセミア、筋ジストロフィー（例えば、デュシェンヌ型筋ジストロフィー）、アルツハイマー、ファンコニー貧血、ゴーシェ病、ムコ脂質症ＩＶ型、ニーマン・ピック病、テイ・サックス病、鎌状赤血球貧血、パーキンソン病、捻転ジストニア、および癌。一部の態様では、胎児のハプロタイプを、１３番染色体、１８番染色体、２１番染色体、Ｘ染色体、およびＹ染色体からなる群より選択される１つ以上の染色体について決定する。一部の態様では、胎児のハプロタイプを胎児の染色体すべてについて決定する。様々な態様では、前記方法によって本質的には胎児の全ゲノムを決定する。一部の態様では、胎児のゲノムの少なくとも３０％、４０％、５０％、６０％、７０％、８０％、９０％、または９５％についてハプロタイプを決定する。一部の態様では、胎児のハプロタイプの決定は、少なくとも１００個；２００個；５００個；７５０個；１，０００個；２，０００個；５，０００個；７，５００個；１０，０００個；２０，０００個；２５，０００個；３０，０００個；４０，０００個；５０，０００個；７５，０００個；または１００，０００個の異なる多型座位にどのアレルが存在するかに関する情報を含む。一部の態様では、この方法を用いて胎芽のハプロタイプまたはアレル比を決定する。

【０５５５】

アレル比の予測方法の例

試料についてアレル比の期待値を計算する方法の例を以下に説明する。表１は、母親と胎児の核酸を含む混合試料（母親の血液試料など）のアレル比の期待値である。これらのアレル比の期待値は、混合試料中の母親の核酸と胎児の核酸のアレルの量を含む、各アレルの総量の測定値について期待される値を示す。一態様では、共分離すると期待される２個の近接座位（たとえば、染色体の交叉がないと期待される２個の座位）において母親はヘテロ接合性である。したがって、母親は（ＡＢ、ＡＢ）である。ここで、１つのハプロタイプについて母親が（Ａ、Ａ）であると母親の相決定済みデータが示していると仮定する。したがって、他方のハプロタイプについて、母親は（Ｂ、Ｂ）であることが推測できる。表１は、胎児画分が２０％である場合の異なる仮説に関するアレル比の期待値を示す。この例では、父親のデータに関する知識を全く想定せず、ヘテロ接合体率を５０％と仮定する。２個のＳＮＰの各々について、アレル比の期待値を（Ａのリードの割合／総リード数）として表す。母親の相決定済みデータ（１つのハプロタイプが（Ａ、Ａ）であり１つが（Ｂ、Ｂ）であるという知識）を用いた場合と、母親の相決定済みデータを用いない場合の両方について、これらの比を計算する。表１は、それぞれの親からの胎児の染色体分節のコピー数に関する様々な仮説を含む。

【表１－１】

表１：母親と胎児の核酸の混合試料について期待される遺伝データ

コピー数仮説	母親の相決定済みデータを用いた場合のアレル比の期待値	母親の相決定済みデータを用いない場合のアレル比の期待値
一染色体性 （母親のコピーの欠落）	(0.444; 0.444) (0.444; 0.555) (0.555; 0.444) (0.555; 0.555)	(0.444; 0.444) (0.444; 0.555) (0.555; 0.444) (0.555; 0.555)
一染色体性 （父親のコピーの欠落）	(0.444; 0.444) (0.555; 0.555)	(0.444; 0.444) (0.444; 0.555) (0.555; 0.444) (0.555; 0.555)
二染色体性	(0.40; 0.40) (0.40; 0.50) (0.50; 0.40) (0.50; 0.50) (0.50; 0.60) (0.60; 0.50) (0.60; 0.60)	(0.40; 0.40) (0.40; 0.50) (0.40; 0.60) (0.50; 0.40) (0.50; 0.50) (0.50; 0.60) (0.60; 0.40) (0.60; 0.50) (0.60; 0.60)

【表１－２】

コピー数仮説	母親の相決定済みデータを用いた場合のアレル比の期待値	母親の相決定済みデータを用いない場合のアレル比の期待値
三染色体性（母親のコピーの過剰；一致型）	(0.36; 0.36) (0.36; 0.45) (0.45; 0.36) (0.45; 0.45) (0.54; 0.54) (0.54; 0.63) (0.63; 0.54) (0.63; 0.63)	(0.36; 0.36) (0.36; 0.45) (0.36; 0.54) (0.36; 0.63) (0.45; 0.36) (0.45; 0.45) (0.45; 0.54) (0.45; 0.63) (0.54; 0.36) (0.54; 0.45) (0.54; 0.54) (0.54; 0.63) (0.63; 0.36) (0.63; 0.45) (0.63; 0.54) (0.63; 0.63)
三染色体性（母親のコピーの過剰；不一致型）	(0.45, 0.45) (0.45; 0.54) (0.54; 0.45) (0.54; 0.54)	(0.36; 0.36) (0.36; 0.45) (0.36; 0.54) (0.36; 0.63) (0.45; 0.36) (0.45; 0.45) (0.45; 0.54) (0.45; 0.63) (0.54; 0.36) (0.54; 0.45) (0.54; 0.54) (0.54; 0.63) (0.63; 0.36) (0.63; 0.45) (0.63; 0.54) (0.63; 0.63)

【表１－３】

コピー数仮説	母親の相決定済みデータを用いた場合のアレル比の期待値	母親の相決定済みデータを用いない場合のアレル比の期待値
三染色体性（父親のコピーの過剰；一致型）	(0.36; 0.36) (0.36; 0.54) (0.54; 0.36) (0.54; 0.54) (0.45; 0.45) (0.45; 0.63) (0.63; 0.45) (0.63; 0.63)	(0.36; 0.36) (0.36; 0.45) (0.36; 0.54) (0.36; 0.63) (0.45; 0.36) (0.45; 0.45) (0.45; 0.54) (0.45; 0.63) (0.54; 0.36) (0.54; 0.45) (0.54; 0.54) (0.54; 0.63) (0.63; 0.36) (0.63; 0.45) (0.63; 0.54) (0.63; 0.63)
三染色体性（父親のコピーの過剰；不一致型）	(0.36; 0.36) (0.36; 0.45) (0.36; 0.54) (0.36; 0.63) (0.45; 0.36) (0.45; 0.45) (0.45; 0.54) (0.45; 0.63) (0.54; 0.36) (0.54; 0.45) (0.54; 0.54) (0.54; 0.63) (0.63; 0.36) (0.63; 0.45) (0.63; 0.54) (0.63; 0.63)	(0.36; 0.36) (0.36; 0.45) (0.36; 0.54) (0.36; 0.63) (0.45; 0.36) (0.45; 0.45) (0.45; 0.54) (0.45; 0.63) (0.54; 0.36) (0.54; 0.45) (0.54; 0.54) (0.54; 0.63) (0.63; 0.36) (0.63; 0.45) (0.63; 0.54) (0.63; 0.63)

【０５５６】

相決定済みデータを用いることで、可能性のあるアレル比の期待値の数を減らすうえ、アレル比の期待値の各々の事前確率を、最尤の結果が正しい可能性を高めるように変更する。可能性のないアレル比の期待値または仮説を排除することにより、正しい仮説が選択される可能性が高くなる。一例として、アレル比の測定値が（０．４１、０．５９）だとする。相決定済みデータを用いない場合、最も尤度の高い仮説は二染色体性仮説（アレル比の測定値が、二染色体性の場合のアレル比の期待値（０．４０、０．６０）に近いため）だと推測されると考えられる。しかし、相決定済みデータを用いた場合、二染色体性仮説に関するアレル比の期待値（０．４０、０．６０）を排除し、三染色体性仮説をより尤度の高いものとして選択できる。

【０５５７】

アレル比の測定値が（０．４、０．４）だと仮定する。ハプロタイプ情報がなければ、各ＳＮＰにおける母親の欠失の確率は０．５×Ｐ（Ａの欠失）＋０．５×Ｐ（Ｂの欠失）である。したがって、Ａが欠失しているように見えるが（胎児における欠失）、欠失の尤度は２つの平均値である。胎児画分が十分に高い場合、最尤仮説を決定することができる

。胎児画分が非常に低い場合、平均化は欠失仮説を排除するかもしれないが、ハプロタイプ情報があれば、相同体１が欠失している確率Ｐ（Ａの欠失）がより高くなり、測定データとの適合がより良好になる。必要に応じ、２個の座位の間の交叉の確率も考慮できる。

【０５５８】

相決定済みデータを用いて尤度を組み合わせる別の説明例では、２個の連続するＳＮＰ１とＳＮＰ２について検討し、D_1とD_2をこれらのＳＮＰにおけるアレルデータとする。これらの２個のＳＮＰに関する尤度を組み合わせる方法の例を以下に示す。２個の連続するヘテロ接合性のＳＮＰが、同一の相同体において同一のアレルを有する（すなわち、ＳＮＰが両方ともＡＢであるか、ＳＮＰが両方ともＢＡである）確率をｃで表す。したがって、１－ｃはＳＮＰの一方がＡＢで他方がＢＡである確率である。たとえば、仮説H_{10}とアレル不均衡値ｆを考慮する。まず、ＳＮＰがすべてＡＢまたはＢＡであると想定してすべての尤度を計算すると仮定する。次に、２個の連続するＳＮＰの尤度を以下のように組み合わせることができる：

【数２２】

$$Lik(D_1, D_2 \mid H_{10}, f) =$$

【数２３】

$$Lik(D_1 \mid H_{10}, f) \times c \times Lik(D_2 \mid H_{10}, f) + Lik(D_1 \mid H_{10}, f) \times (1-c) \times Lik(D_2 \mid H_{01}, f).$$
$$Lik(D_1 \mid H_{10}, f) \times c \times Lik(D_2 \mid H_{10}, f) + Lik(D_1 \mid H_{10}, f) \times (1-c) \times Lik(D_2 \mid H_{01}, f).$$

これを再帰的に行い、すべてのＳＮＰに関する結合尤度

【数２４】

$$Lik(D_1, \ldots, D_N \mid H_{10}, f)$$

を決定できる。

【０５５９】

変異の例

癌などの疾患または障害、あるいは癌などの疾患または障害のリスク増加（たとえば正常のリスクレベルを超える）に関連する変異の例としては、一塩基多様体（ＳＮＶ）、複数塩基の変異、欠失（２～３０００万塩基対の領域の欠失など）、重複、または縦列反復が挙げられる。一部の態様では、変異は、無細胞ＤＮＡ、無細胞ミトコンドリアＤＮＡ、核内ＤＮＡに由来する無細胞ＤＮＡ（無細胞核内ＤＮＡ）、細胞内ＤＮＡ、ミトコンドリアＤＮＡなどのＤＮＡに存在する。一部の態様では、無細胞ＲＮＡ、細胞内ＲＮＡ、細胞質ＲＮＡ、コーディング細胞質ＲＮＡ、非コーディング細胞質ＲＮＡ、ｍＲＮＡ、ｍｉＲＮＡ、ミトコンドリアＲＮＡ、ｒＲＮＡ、またはｔＲＮＡなどのＲＮＡに存在する。一部の態様では、疾患または障害（癌など）を有する対象者では、疾患または障害（癌など）を有さない対象者に比べて高頻度で変異が存在する。一部の態様では、変異（原因変異など）は癌を示す。一部の態様では、変異は疾患または障害に原因として関与するドライバー変異である。一部の態様では、変異は原因変異ではない。たとえば、一部の癌では複数の変異が蓄積するが、その一部は原因変異ではない。原因変異ではない変異（たとえば、疾患または障害を有する対象者において、疾患または障害を有さない対象者に比べて高頻度で存在する変異など）も、疾患または障害の診断に有用となり得る。一部の態様では、変異は１つ以上のマイクロサテライトにおけるヘテロ接合性の消失（ＬＯＨ）である。

【０５６０】
一部の態様では、対象者が有することがわかっている１つ以上の多型または変異について対象者をスクリーニングする（たとえば、その多型または変異の存在、それらの多型または変異を有する細胞、ＤＮＡ、またはＲＮＡの量の変化、あるいは癌の寛解または再発について試験するため）。一部の態様では、対象者がリスクを有することがわかっている１つ以上の多型または変異について対象者（たとえばその多型または変異を有する血縁者がいる対象者）をスクリーニングする。一部の態様では、癌などの疾患または障害に関連する多型または変異のパネル（たとえば、少なくとも５個、１０個、５０個、１００個、２００個、３００個、５００個、７５０個、１，０００個、１，５００個、２，０００個、または５，０００個の多型または変異）について対象者をスクリーニングする。

【０５６１】
Abaan et al., "The Exomes of the NCI-60 Panel: A Genomic Resource for Cancer Biology and Systems Pharmacology", Cancer Research, July 15, 2013およびワールドワイドウェブのdtp.nci.nih.gov/branches/btb/characterizationNCI60.html（それぞれの全内容は参照によって本明細書に組み込まれる）には、癌に関連する多くのコード化多様体が記載されている。ヒト癌細胞株パネルＮＣＩ－６０は肺癌、肺癌、大腸癌、脳癌、卵巣癌、乳癌、前立腺癌、腎癌、白血病、および黒色腫を代表する６０種の異なる細胞株からなる。これらの細胞株に認められた遺伝的多様性は２種類である：正常な集団に見られるＩ型多様体；および癌に特異的なＩＩ型多様体。

【０５６２】
多型または変異（欠失または重複など）の例は、以下の遺伝子のうち１つ以上のものである：TP53, PTEN, PIK3CA, APC, EGFR, NRAS, NF2, FBXW7, ERBBs, ATAD5, KRAS, BRAF, VEGF, EGFR, HER2, ALK, p53, BRCA, BRCA1, BRCA2, SETD2, LRP1B, PBRM, SPTA1, DNMT3A, ARID1A, GRIN2A, TRRAP, STAG2, EPHA3/5/7, POLE, SYNE1, C20orf80, CSMD1, CTNNB1, ERBB2. FBXW7, KIT, MUC4, ATM, CDH1, DDX11, DDX12, DSPP, EPPK1, FAM186A, GNAS, HRNR, KRTAP4-11, MAP2K4, MLL3, NRAS, RB1, SMAD4, TTN, ABCC9, ACVR1B, ADAM29, ADAMTS19, AGAP10, AKT1, AMBN, AMPD2, ANKRD30A, ANKRD40, APOBR, AR, BIRC6, BMP2, BRAT1, BTNL8, C12orf4, C1QTNF7, C20orf186, CAPRIN2, CBWD1, CCDC30, CCDC93, CD5L, CDC27, CDC42BPA, CDH9, CDKN2A, CHD8, CHEK2, CHRNA9, CIZ1, CLSPN, CNTN6, COL14A1, CREBBP, CROCC, CTSF, CYP1A2, DCLK1, DHDDS, DHX32, DKK2, DLEC1, DNAH14, DNAH5, DNAH9, DNASE1L3, DUSP16, DYNC2H1, ECT2, EFHB, RRN3P2, TRIM49B, TUBB8P5, EPHA7, ERBB3, ERCC6, FAM21A, FAM21C, FCGBP, FGFR2, FLG2, FLT1, FOLR2, FRYL, FSCB, GAB1, GABRA4, GABRP, GH2, GOLGA6L1, GPHB5, GPR32, GPX5, GTF3C3, HECW1, HIST1H3B, HLA-A, HRAS, HS3ST1, HS6ST1, HSPD1, IDH1, JAK2, KDM5B, KIAA0528, KRT15, KRT38, KRTAP21-1, KRTAP4-5, KRTAP4-7, KRTAP5-4, KRTAP5-5, LAMA4, LATS1, LMF1, LPAR4, LPPR4, LRRFIP1, LUM, LYST, MAP2K1, MARCH1, MARCO, MB21D2, MEGF10, MMP16, MORC1, MRE11A, MTMR3, MUC12, MUC17, MUC2, MUC20, NBPF10, NBPF20, NEK1, NFE2L2, NLRP4, NOTCH2, NRK, NUP93, OBSCN, OR11H1, OR2B11, OR2M4, OR4Q3, OR5D13, OR8I2, OXSM, PIK3R1, PPP2R5C, PRAME, PRF1, PRG4, PRPF19, PTH2, PTPRC, PTPRJ, RAC1, RAD50, RBM12, RGPD3, RGS22, ROR1, RP11-671M22.1, RP13-996F3.4, RP1L1, RSBN1L, RYR3, SAMD3, SCN3A, SEC31A, SF1, SF3B1, SLC25A2, SLC44A1, SLC4A11, SMAD2, SPTA1, ST6GAL2, STK11, SZT2, TAF1L, TAX1BP1, TBP, TGFBI, TIF1, TMEM14B, TMEM74, TPTE, TRAPPC8, TRPS1, TXNDC6, USP32, UTP20, VASN, VPS72, WASH3P, WWTR1, XPO1, ZFHX4, ZMIZ1, ZNF167, ZNF436, ZNF492, ZNF598, ZRSR2, ABL1, AKT2, AKT3, ARAF, ARFRP1, ARID2, ASXL1, ATR, ATRX, AURKA, AURKB, AXL, BAP1, BARD1, BCL2, BCL2L2, BCL6, BCOR, BCORL1, BLM, BRIP1, BTK, CARD11, CBFB, CBL, CCND1, CCND2, CCND3, CCNE1, CD79A, CD79B, CDC73, CDK12, CDK4, CDK6, CDK8, CDKN1B, CDKN2B, CDKN2C, CEBPA, CHEK1, CIC, CRKL, CRLF2, CSF1R, CTCF, CTNNA1, DAXX, DDR2, DOT1L, EMSY (C11orf30), EP300, EPHA3, EPHA5, EPHB1, ERBB4, ERG, ESR1, EZH2, FAM123B (WTX), FAM46C, FANCA, FANCC, FANCD2, FANCE, FANCF, FANCG, FANCL, FGF10, FGF14, FGF19, FGF23, FGF3, FGF4, FGF6, FGFR1, FGFR2, FGFR3, FG

FR4, FLT3, FLT4, FOXL2, GATA1, GATA2, GATA3, GID4 (C17orf39), GNA11, GNA13, GNAQ, GNAS, GPR124, GSK3B, HGF, IDH1, IDH2, IGF1R, IKBKE, IKZF1, IL7R, INHBA, IRF4, IRS2, JAK1, JAK3, JUN, KAT6A (MYST3), KDM5A, KDM5C, KDM6A, KDR, KEAP1, KLHL6, MAP2K2, MAP2K4, MAP3K1, MCL1, MDM2, MDM4, MED12, MEF2B, MEN1, MET, MITF, MLH1, MLL, MLL2, MPL, MSH2, MSH6, MTOR, MUTYH, MYC, MYCL1, MYCN, MYD88, NF1, NFKBIA, NKX2-1, NOTCH1, NPM1, NRAS, NTRK1, NTRK2, NTRK3, PAK3, PALB2, PAX5, PBRM1, PDGFRA, PDGFRB, PDK1, PIK3CG, PIK3R2, PPP2R1A, PRDM1, PRKAR1A, PRKDC, PTCH1, PTPN11, RAD51, RAF1, RARA, RET, RICTOR, RNF43, RPTOR, RUNX1, SMARCA4, SMARCB1, SMO, SOCS1, SOX10, SOX2, SPEN, SPOP, SRC, STAT4, SUFU, TET2, TGFBR2, TNFAIP3, TNFRSF14, TOP1, TP53, TSC1, TSC2, TSHR, VHL, WISP3, WT1, ZNF217, ZNF703、およびその組み合わせ（Su et al., J Mol Diagn 2011, 13:74-84; DOI:10.1016/j.jmoldx.2010.11.010; and Abaan et al., "The Exomes of the NCI-60 Panel: A Genomic Resource for Cancer Biology and Systems Pharmacology", Cancer Research, July 15, 2013；この文献の全内容は参照によって本明細書に組み込まれる）。一部の態様では、重複は乳癌に関連する染色体１ｐ（「Ｃｈｒ１ｐ」）の重複である。一部の態様では、１つ以上の多型または変異はＢＲＡＦに存在する（Ｖ６００Ｅ変異など）。一部の態様では、１つ以上の多型または変異はＫ－ｒａｓに存在する。一部の態様では、Ｋ－ｒａｓとＡＰＣに１つ以上の多型または変異の組み合わせが存在する。一部の態様では、Ｋ－ｒａｓとｐ５３に１つ以上の多型または変異の組み合わせが存在する。一部の態様では、ＡＰＣとｐ５３に１つ以上の多型または変異の組み合わせが存在する。一部の態様では、Ｋ－ｒａｓ、ＡＰＣ、ｐ５３に１つ以上の多型または変異の組み合わせが存在する。一部の態様では、Ｋ－ｒａｓとＥＧＦＲに１つ以上の多型または変異の組み合わせが存在する。例示の多型または変異は、以下のマイクロＲＮＡのうち１つ以上に存在する：miR-15a, miR-16-1, miR-23a, miR-23b, miR-24-1, miR-24-2, miR-27a, miR-27b, miR-29b-2, miR-29c, miR-146, miR-155, miR-221, miR-222, and miR-223（Calin et al. “A microRNA signature associated with prognosis and progression in chronic lymphocytic leukemia.” N Engl J Med 353:1793-801, 2005；この文献の全内容は参照によって本明細書に組み込まれる）。

【０５６３】

一部の態様では、欠失は少なくとも０．０１ｋｂ、０．１ｋｂ、１ｋｂ、１０ｋｂ、１００ｋｂ、１ｍｂ、２ｍｂ、３ｍｂ、５ｍｂ、１０ｍｂ、１５ｍｂ、２０ｍｂ、３０ｍｂ、または４０ｍｂの欠失である。一部の態様では、欠失は、１ｋｂ～４０ｍｂ（たとえば１ｋｂ～１００ｋｂ、１００ｋｂ～１ｍｂ、１ｍｂ～５ｍｂ、５ｍｂ～１０ｍｂ、１０ｍｂ～１５ｍｂ、１５ｍｂ～２０ｍｂ、２０ｍｂ～２５ｍｂ、２５ｍｂ～３０ｍｂ、または３０ｍｂ～４０ｍｂ）（それぞれ、範囲の上限と下限を含む）の欠失である。

【０５６４】

一部の態様では、重複は少なくとも０．０１ｋｂ、０．１ｋｂ、１ｋｂ、１０ｋｂ、１００ｋｂ、１ｍｂ、２ｍｂ、３ｍｂ、５ｍｂ、１０ｍｂ、１５ｍｂ、２０ｍｂ、３０ｍｂ、または４０ｍｂの重複である。一部の態様では、重複は、１ｋｂ～４０ｍｂ（たとえば１ｋｂ～１００ｋｂ、１００ｋｂ～１ｍｂ、１ｍｂ～５ｍｂ、５ｍｂ～１０ｍｂ、１０ｍｂ～１５ｍｂ、１５ｍｂ～２０ｍｂ、２０ｍｂ～２５ｍｂ、２５ｍｂ～３０ｍｂ、または３０ｍｂ～４０ｍｂ）（それぞれ、範囲の上限と下限を含む）である。

【０５６５】

一部の態様では、縦列反復は２個～６０個のヌクレオチドの反復であり、たとえば２個～６個、７個～１０個、１０個～２０個、２０個～３０、３０個～４０個、４０個～５０個、または５０個～６０個（それぞれ、範囲の上限と下限を含む）のヌクレオチドの反復である。一部の態様では、縦列反復は２個のヌクレオチドの反復（ジヌクレオチド反復）である。一部の態様では、縦列反復は３個のヌクレオチドの反復（トリヌクレオチド反復）である。

【０５６６】

一部の態様では、多型または変異は予後のものである。予後変異の例としてはＫ－ｒａ

ｓ変異（たとえば大腸癌の手術後の疾患再発の標識であるＫ－ｒａｓ変異）（Ryan et al.” A prospective study of circulating mutant KRAS2 in the serum of patients with colorectal neoplasia: strong prognostic indicator in postoperative follow up, ” Gut 52:101-108, 2003；およびLecomte T et al. Detection of free-circulating tumor-associated DNA in plasma of colorectal cancer patients and its association with prognosis,” Int J Cancer 100:542-548, 2002；これらの文献それぞれの全内容は参照によって本明細書に組み込まれる）が挙げられる。

【０５６７】

一部の態様では、多型または変異は特定の治療に対する反応の変化（たとえば有効性または副作用の増加または減少）に関連する。例としては、ＥＧＦＲに基づく非小細胞肺癌治療に対する反応の低下に関するＫ－ｒａｓ変異が挙げられる（Wang et al.“Potential clinical significance of a plasma-based KRAS mutation analysis in patients with advanced non-small cell lung cancer,” Clin Canc Res16:1324-1330, 2010；この文献の全内容は参照によって本明細書に組み込まれる）。

【０５６８】

Ｋ－ｒａｓは多くの癌において活性化される癌遺伝子である。Ｋ－ｒａｓ変異の例としては、コドン１２、１３、および６１の変異である。無細胞ＤＮＡのＫ－ｒａｓ変異は、膵臓癌、肺癌、結腸直腸癌、膀胱癌および胃癌において確認されている（Fleischhacker & Schmidt “Circulating nucleic acids (CNAs) and caner - a survey,” Biochim Biophys Acta 1775:181-232, 2007；この文献の全内容は参照によって本明細書に組み込まれる）。

【０５６９】

ｐ５３は多くの癌において変異し腫瘍の進行に寄与する腫瘍抑制因子である（Levine & Oren “The first 30 years of p53: growing ever more complex. Nature Rev Cancer, ” 9:749-758, 2009；この文献の全内容は参照によって本明細書に組み込まれる）。Ｓｅｒ２４９など多くの異なるコドンが変異し得る。無細胞ＤＮＡのｐ５３変異は、乳癌、肺癌、卵巣癌、膀胱癌、胃癌、膵臓癌、結腸直腸癌、腸癌および肝細胞癌において確認されている（Fleischhacker & Schmidt “Circulating nucleic acids (CNAs) and caner - a survey,” Biochim Biophys Acta 1775:181-232, 2007；この文献の全内容は参照によって本明細書に組み込まれる）。

【０５７０】

ＢＲＡＦはＲａｓの下流の癌遺伝子である。ＢＲＡＦ変異は膠細胞腫瘍、黒色腫、甲状腺癌および肺癌で確認されている（Dias-Santagata et al. BRAF V600E mutations are common in pleomorphic xanthoastrocytoma: diagnostic and therapeutic implications. PLOS ONE 2011;6:e17948, 2011; Shinozaki et al. Utility of circulating B-RAF DNA mutation in serum for monitoring melanoma patients receiving biochemotherapy. Clin Canc Res 13:2068-2074, 2007；およびBoard et al. Detection of BRAF mutations in the tumor and serum of patients enrolled in the AZD6244 (ARRY-142886) advanced melanoma phase II study. Brit J Canc 2009;101:1724-1730；これらの文献それぞれの全内容は参照によって本明細書に組み込まれる）。ＢＲＡＦのＶ６００Ｅ変異は、たとえば黒色腫腫瘍で発生し、進行期により多くみられる。Ｖ６００Ｅ変異は無細胞ＤＮＡに検出されている。

【０５７１】

ＥＧＦＲは細胞増殖に寄与し、多くの癌で誤制御される（Downward J. Targeting RAS signalling pathways in cancer therapy. Nature Rev Cancer 3:11-22, 2003；およびLevine & Oren “The first 30 years of p53: growing ever more complex. Nature Rev Cancer,” 9:749-758, 2009（これらの文献の全内容は参照によって本明細書に組み込まれる））。ＥＧＦＲ変異の例としては、肺癌患者に確認されたエキソン１８～２１の変異が挙げられる。無細胞ＤＮＡのＥＧＦＲ変異が肺癌患者に確認されている（Jia et al.“Prediction of epidermal growth factor receptor mutations in the plasma/pleural ef

fusion to efficacy of gefitinib treatment in advanced non-small cell lung cancer," J Canc Res Clin Oncol 2010;136:1341-1347, 2010；この文献の全内容は参照によって本明細書に組み込まれる）。

【０５７２】

乳癌に関連する多型または変異の例としては以下が挙げられる：マイクロサテライトのヘテロ接合性の消失（ＬＯＨ）（Kohler et al. "Levels of plasma circulating cell free nuclear and mitochondrial DNA as potential biomarkers for breast tumors," Mol Cancer 8:doi:10.1186/1476-4598-8-105, 2009；この文献の全内容は参照によって本明細書に組み込まれる）；ｐ５３変異（例えば、エキソン５～８における変異）（Garcia et al." Extracellular tumor DNA in plasma and overall survival in breast cancer patients," Genes, Chromosomes & Cancer 45:692-701, 2006；この文献の全内容は参照によって本明細書に組み込まれる）；ＨＥＲ２変異（Sorensen et al. "Circulating HER2 DNA after trastuzumab treatment predicts survival and response in breast cancer," Anticancer Res30:2463-2468, 2010；この文献の全内容は参照によって本明細書に組み込まれる）；ＰＩＫ３ＣＡ、ＭＥＤ１、およびＧＡＳ６の多型または変異（Murtaza et al. "Non-invasive analysis of acquired resistance to cancer therapy by sequencing of plasma DNA," Nature 2013;doi:10.1038/nature12065, 2013；この文献の全内容は参照によって本明細書に組み込まれる）。

【０５７３】

無細胞ＤＮＡレベルとＬＯＨの増加は、全生存期間および無病生存期間の減少に関連する。ｐ５３変異（エキソン５～８）は全生存期間の減少に関連する。循環ＨＥＲ２無細胞ＤＮＡレベルの減少は、ＨＥＲ２陽性乳房腫瘍患者のＨＥＲ２標的治療に対する、より良好な応答と関連する。ＰＩＫ３ＣＡの活性化変異、ＭＥＤ１の切断、およびＧＡＳ６のスプライシング変異は、治療に対する耐性をもたらす。

【０５７４】

結腸直腸癌に関連する多型または変異の例としては、ｐ５３、ＡＰＣ、Ｋ－ｒａｓ、およびチミジル酸合成酵素の変異、ならびにｐ１６遺伝子のメチル化が挙げられる（Wang et al. "Molecular detection of APC, K-ras, and p53 mutations in the serum of colorectal cancer patients as circulating biomarkers," World J Surg 28:721-726, 2004；Ryan et al. "A prospective study of circulating mutant KRAS2 in the serum of patients with colorectal neoplasia: strong prognostic indicator in postoperative follow up," Gut 52:101-108, 2003；Lecomte et al. "Detection of free-circulating tumor-associated DNA in plasma of colorectal cancer patients and its association with prognosis," Int J Cancer 100:542-548, 2002；Schwarzenbach et al. "Molecular analysis of the polymorphisms of thymidylate synthase on cell-free circulating DNA in blood of patients with advanced colorectal carcinoma," Int J Cancer 127:881-888, 2009（これらの文献それぞれの全内容は参照によって本明細書に組み込まれる））。術後の血清中でのＫ－ｒａｓ変異の検出は疾患の再発の強力な予測因子である。Ｋ－ｒａｓ変異およびｐ１６遺伝子のメチル化の検出は、生存率の低下および疾患の再発率の増加に関連する。Ｋ－ｒａｓ、ＡＰＣ、および／またはｐ５３の変異の検出は、再発および／または転移と関連する。無細胞ＤＮＡを用いるチミジル酸合成酵素（フルオロピリミジンに基づく化学療法の標的）遺伝子における多型（ＬＯＨ、ＳＮＰ、縦列反復数の変動性、欠失など）は、治療反応と関連している可能性がある。

【０５７５】

肺癌（非小細胞肺癌など）に関連する多型または変異の例としてはＫ－ｒａｓ（コドン１２の変異など）およびＥＧＦＲの変異が挙げられる。予後変異の例としては、全生存期間および無増悪生存期間の延長に関連するＥＧＦＲ変異（エキソン１９の欠失またはエキソン２１の変異）が挙げられ、Ｋ－ｒａｓ変異（コドン１２および１３）は、無増悪生存期間の延長に関連する（Jian et al. "Prediction of epidermal growth factor receptor mutations in the plasma/pleural effusion to efficacy of gefitinib treatment i

n advanced non-small cell lung cancer,” J Canc Res Clin Oncol 136:1341-1347, 2010；Wang et al. “Potential clinical significance of a plasma-based KRAS mutation analysis in patients with advanced non-small cell lung cancer,” Clin Canc Res 16:1324-1330, 2010（これらの文献それぞれの全内容は参照によって本明細書に組み込まれる）。治療への応答を示す多型または変異の例としては、治療への応答を改善するＥＧＦＲ変異（エキソン１９の欠失またはエキソン２１の変異）と、治療への応答を低下させるＫ－ｒａｓ変異（コドン１２および１３）が挙げられる。抵抗性を付与するＥＦＧＲ変異が確認されている（Murtaza et al. “Non-invasive analysis of acquired resistance to cancer therapy by sequencing of plasma DNA,” Nature doi:10.1038/nature12065, 2013；この文献の全内容は参照によって本明細書に組み込まれる）。

【０５７６】

黒色腫（ブドウ膜黒色腫など）に関連する多型または変異の例としては、ＧＮＡＱ、ＧＮＡ１１、ＢＲＡＦ、およびｐ５３の変異が挙げられる。ＧＮＡＱおよびＧＮＡ１１の変異の例はＲ１８３変異およびＱ２０９変異である。ＧＮＡＱまたはＧＮＡ１１のＱ２０９変異は骨への転移に関連する。転移期／進行期の黒色腫患者ではＢＲＡＦのＶ６００Ｅ変異を検出できる。ＢＲＡＦのＶ６００Ｅ変異は黒色腫の標識である。化学療法後のＢＲＡＦのＶ６００Ｅ変異の存在は、治療反応がないことと関連する。

【０５７７】

膵癌に関連する多型または変異の例としては、Ｋ－ｒａｓおよびｐ５３の変異（たとえばｐ５３Ｓｅｒ２４９）が挙げられる。ｐ５３Ｓｅｒ２４９はＢ型肝炎感染症および肝細胞癌、卵巣癌、非ホジキンリンパ腫にも関連する。

【０５７８】

本発明の方法を用いれば、試料中に低頻度で存在する多型または変異も検出可能である。たとえば、１００万分の１の頻度で存在する多型または変異を、１０００万リードの配列決定を行うことで観測できる。必要に応じて、所望の感度のレベルに応じて配列決定リード数を変更できる。一部の態様では、配列決定リード数を増やして試料を再分析するか、対象者の別の試料を分析することによって感度を高めることができる。たとえば、癌または癌のリスク増加に関連する多型または変異が検出されないか少数（たとえば１個、２個、３個、４個、または５個）しか検出されない場合、その試料を再分析するか、別の試料を試験する。

【０５７９】

一部の態様では、複数の多型または変異は癌または転移性癌の必要条件である。このような場合、複数の多型または変異についてスクリーニングすることにより、癌または転移性癌をより正確に診断できる。癌または転移性癌の必要条件である複数の多型の部分集合を対象者が有するような一部の態様では、対象者を後に再検査し、新たな変異が起こっていないかを確認できる。

【０５８０】

複数の多型または変異が癌または転移性癌の必要条件であるような一部の態様では、それぞれの多型または変異の頻度を比較し、それらが同様の頻度で起こっているかを確認できる。たとえば、２つの変異（「Ａ」および「Ｂ」とする）が癌の必要条件である場合、一部の細胞はいずれも有さず、一部の細胞はＡを有し、一部はＢを有し、一部はＡとＢを有する。ＡとＢが同様の頻度で観測される場合、対象者はＡもＢも含む細胞を有する可能性がより高い。ＡとＢが異なる頻度で観測される場合、対象者は異なる細胞集団を有する可能性がより高い。

【０５８１】

複数の多型または変異が癌または転移性癌の必要条件であるような一部の態様では、対象者に存在するこのような多型または変異の数または同一性を用いて、その対象者が疾患または障害を有する尤度またはそれらを近いうちに有する尤度を予測できる。多型または変異が所定の順序で発生する傾向にあるような一部の態様では、対象者を定期的にスクリーニングし、他の多型または変異が起こったかどうかを確認してもよい。

【０５８２】
一部の態様では、複数（たとえば２個、３個、４個、５個、８個、１０個、１２個、１５個、またはそれ以上）の多型または変異の有無を決定することによって、癌などの疾患または障害、あるいは癌などの疾患または障害のリスク増加の有無の決定の感度および／または特異度が高くなる。
【０５８３】
一部の態様では、多型または変異を直接検出する。一部の態様では、多型または変異を、その多型または変異に結合している１つ以上の配列（たとえばＳＮＰなどの多型座位）を検出することにより、間接的に検出する。
核酸変化の例
【０５８４】
一部の態様では、癌などの疾患または障害あるいは癌などの疾患または障害のリスク増加に関連するような、ＲＮＡまたはＤＮＡの完全性の変化（断片化した無細胞ＲＮＡまたは無細胞ＤＮＡのサイズの変化、あるいはヌクレオソーム構成の変化）がある。一部の態様では、癌などの疾患または障害あるいは癌などの疾患または障害のリスク増加に関連するような、ＲＮＡまたはＤＮＡのメチル化パターンの変化（たとえば腫瘍抑制遺伝子の過剰メチル化）がある。たとえば、腫瘍抑制遺伝子のプロモーター領域のＣｐＧ島のメチル化が局所的な遺伝子発現抑制を誘発すると示唆されている。肝臓癌、肺癌、乳癌の患者ではｐ１６腫瘍抑制遺伝子の異常なメチル化が起こる。メチル化が起こりやすいその他の腫瘍抑制遺伝子（ＡＰＣ、Ｒａｓ結合ドメインファミリータンパク質（ＲＡＳＳＦ１Ａ）、グルタチオンＳ－転移酵素Ｐ１（ＧＳＴＰ１）、ＤＡＰＫなど）は多様な種類の癌（たとえば、上咽頭癌、結腸直腸癌、肺癌、食道癌、前立腺癌、膀胱癌、黒色腫、急性白血病）で検出されている。ｐ１６などの特定の腫瘍抑制遺伝子のメチル化は、癌形成における初期事象と言われており、したがって早期癌検診に有用である。
【０５８５】
一部の態様では、メチル化感受性制限酵素による消化を用いる亜硫酸水素塩変換または亜硫酸水素塩に基づかない手法でメチル化パターンを決定する（Hung et al., J Clin Pathol 62:308-313, 2009；この文献の全内容は参照によって本明細書に組み込まれる）。亜硫酸水素塩変換では、メチル化シトシンはシトシンのままであり、非メチル化シトシンはウラシルに変換される。メチル化配列は未変化のままであるが、メチル化感受性制限酵素（たとえばＢｓｔＵＩ）は非メチル化ＤＮＡ配列を特定の認識部位（たとえば、ＢｓｔＵＩの場合は５’－ＣＧＶＣＧ－３’）で切断する。一部の態様では、未変化のメチル化配列を検出する。一部の態様では、ステムループプライマーを用いて、制限酵素消化した非メチル化断片を選択的に増幅し、酵素消化していないメチル化ＤＮＡは共増幅しない。
【０５８６】
ｍＲＮＡスプライシングの変化の例
一部の態様では、ｍＲＮＡスプライシングの変化は癌などの疾患または障害、あるいは癌などの疾患または障害のリスク増加に関連する。一部の態様では、ｍＲＮＡスプライシングの変化は、癌または癌のリスク増加に関連する以下の核酸のうちの１つ以上に見られる：ＤＮＭＴ３Ｂ、ＢＲＣＡ１、ＫＬＦ６、Ｒｏｎ、およびＧｅｍｉｎ５。一部の態様では、検出されたｍＲＮＡスプライス多様体は癌などの疾患または障害に関連する。一部の態様では、複数のｍＲＮＡスプライス多様体が健常細胞（非癌性細胞など）によって生成されるが、ｍＲＮＡスプライス多様体の相対量の変化が癌などの疾患または障害に関連する。一部の態様では、ｍＲＮＡスプライシングの変化はｍＲＮＡ配列の変化（スプライス部位の変異など）、スプライシング因子のレベルの変化、利用できるスプライシング因子の量の変化（たとえば、スプライシング因子が反復配列に結合することによって利用可能なスプライシング因子の量が減少する）、スプライシング制御の変化、または腫瘍微小環境などに起因する。
【０５８７】
スプライシング反応は、スプライソソームと呼ばれる、複数のタンパク質とＲＮＡとの

複合体によって行われる。（Fackenthall and Godley, Disease Models & Mechanisms 1: 37-42, 2008, doi:10.1242/dmm.000331；この文献の全内容は参照によって本明細書に組み込まれる）。スプライソソームは、イントロンとエキソンとの境界を認識し、介在するイントロンを２回のエステル転移反応で除去する。その結果、２個の隣接するエキソンは結合する。結合が正しく行われなければ、正常なタンパク質コード能力が損なわれるため、この反応の忠実度は精巧でなければならない。たとえば、翻訳中のアミノ酸の同一性や順序を指定する三塩基コドンの読み枠がエキソンの読み飛ばしによって保存された場合、選択的にスプライスされたｍＲＮＡは重要なアミノ酸残基を欠失したタンパク質を特定する可能性がある。より一般的には、エキソンを読み飛ばすと翻訳読み枠が中断され、未成熟終止コドンが生じる。これらのｍＲＮＡは、ナンセンス変異依存ｍＲＮＡ分解として知られる処理によって、一般的には少なくとも９０％分解される。その結果、このような欠陥のあるメッセージが蓄積して切断型タンパク質産物を生成する可能性が低減される。ミススプライスされたｍＲＮＡがこの経路を回避すると、切断型、変異型、または不安定なタンパク質が生成される。

【０５８８】

選択的スプライシングは、同じゲノムＤＮＡから数種または多種の異なる転写物を発現する手段であり、特定のタンパク質について利用可能なエキソンの一部が含まれるため生じる。１つ以上のエキソンが排除されると、コードされたタンパク質からは特定のタンパク質ドメインが失われることになり、タンパク質機能の消失または獲得が起こることがある。数種類の選択的スプライシングが開示されている：エキソンの読み飛ばし；選択的５’スプライシング部位または選択的３’スプライス部位；相互排他的エキソン；イントロン保持（他に比べてはるかに珍しい）。他の研究者は、生物情報学的手法を用いて癌細胞と正常細胞における選択的スプライシングのレベルを比較し、癌が正常細胞よりも低レベルの選択的スプライシングを示すと決定した。さらに、選択的スプライシング事象の種類の分布は、癌細胞と正常細胞で異なった。癌細胞では正常細胞に比べてエキソンの読み飛ばしは少ないが選択的５’スプライス部位および選択的３’スプライス部位やイントロン保持は多いことが実証されている。エキソン化現象（他の組織によって主にイントロンとして使用される配列がエキソンとして使用されること）を試験すると、癌細胞におけるエキソン化に関連する遺伝子は、ｍＲＮＡプロセシングと優先的に関連しており、癌細胞と異常ｍＲＮＡスプライス形態の発生との間に直接的関係があることが示された。

【０５８９】

ＤＮＡまたはＲＮＡの濃度変化の例

一部の態様では、１種以上のＤＮＡ（無細胞ＤＮＡ、無細胞ミトコンドリアＤＮＡ、無細胞核内ＤＮＡ、細胞内ＤＮＡ、ミトコンドリアＤＮＡなど）またはＲＮＡ（無細胞ＲＮＡ、細胞内ＲＮＡ、細胞質ＲＮＡ、コーディング細胞質ＲＮＡ、非コーディング細胞質ＲＮＡ、ｍＲＮＡ、ｍｉＲＮＡ、ミトコンドリアＲＮＡ、ｒＲＮＡ、またはｔＲＮＡなど）の総量または総濃度に変化が見られる。一部の態様では１種以上の特定のＤＮＡ（無細胞ＤＮＡ、無細胞ミトコンドリアＤＮＡ、無細胞核内ＤＮＡ、細胞内ＤＮＡ、ミトコンドリアＤＮＡなど）またはＲＮＡ（無細胞ＲＮＡ、細胞内ＲＮＡ、細胞質ＲＮＡ、コーディング細胞質ＲＮＡ、非コーディング細胞質ＲＮＡ、ｍＲＮＡ、ｍｉＲＮＡ、ミトコンドリアＲＮＡ、ｒＲＮＡ、またはｔＲＮＡなど）の分子の量または濃度に変化が見られる。一部の態様では、１個のアレルは、対象座位にある別のアレルよりも多く発現する。例示のｍｉＲＮＡは、遺伝子発現を制御する２０～２２ヌクレオチドの短いＲＮＡ分子である。一部の態様では、トランスクリプトームに変化（１つ以上のＲＮＡ分子の同一性または量の変化など）が見られる。

【０５９０】

一部の態様では、無細胞ＤＮＡまたは無細胞ＲＮＡの総量または濃度の増加は癌などの疾患または障害、あるいは癌などの疾患または障害のリスク増加に関連する。一部の態様では、ある種類のＤＮＡ（無細胞ＤＮＡ、無細胞ミトコンドリアＤＮＡ、無細胞核内ＤＮＡ、細胞内ＤＮＡ、またはミトコンドリアＤＮＡなど）またはＲＮＡ（無細胞ＲＮＡ、細

胞内ＲＮＡ、細胞質ＲＮＡ、コーディング細胞質ＲＮＡ、非コーディング細胞質ＲＮＡ、ｍＲＮＡ、ｍｉＲＮＡ、ミトコンドリアＲＮＡ、ｒＲＮＡ、またはｔＲＮＡなど）の総濃度は、健常な（癌を有していない）対象者のその種類のＤＮＡまたはＲＮＡの総濃度に比べて少なくとも２倍、３倍、４倍、５倍、６倍、７倍、８倍、９倍、１０倍、またはそれ以上に増加する。一部の態様では、無細胞ＤＮＡの総濃度が７５ｎｇ／ｍＬ～１００ｎｇ／ｍＬ、１００ｎｇ／ｍＬ～１５０ｎｇ／ｍＬ、１５０ｎｇ／ｍＬ～２００ｎｇ／ｍＬ、２００ｎｇ／ｍＬ～３００ｎｇ／ｍＬ、３００ｎｇ／ｍＬ～４００ｎｇ／ｍｇＬ、４００ｎｇ／ｍＬ～６００ｎｇ／ｍＬ、６００ｎｇ／ｍＬ～８００ｎｇ／ｍＬ、８００ｎｇ／ｍＬ～１，０００ｎｇ／ｍＬ（それぞれ、範囲の上限と下限を含む）、あるいは１００ｎｇ／ｍＬ超であり、たとえば２００ｎｇ／ｍＬ超、３００ｎｇ／ｍＬ超、４００ｎｇ／ｍＬ超、５００ｎｇ／ｍＬ超、６００ｎｇ／ｍＬ超、７００ｎｇ／ｍＬ超、８００ｎｇ／ｍＬ超、９００ｎｇ／ｍＬ超、または１，０００ｎｇ／ｍＬ超であることは、癌、癌のリスク増加、良性でなく悪性である腫瘍のリスク増加、癌が寛解する可能性の低下、あるいは癌の予後の悪化を示す。一部の態様では、癌などの疾患または障害、あるいは癌などの疾患または障害のリスク増加に関連する１つ以上の多型／変異（欠失または重複）を有するある種類のＤＮＡ（無細胞ＤＮＡ、無細胞ミトコンドリアＤＮＡ、無細胞核内ＤＮＡ、細胞内ＤＮＡ、またはミトコンドリアＤＮＡなど）またはＲＮＡ（無細胞ＲＮＡ、細胞内ＲＮＡ、細胞質ＲＮＡ、コーディング細胞質ＲＮＡ、非コーディング細胞質ＲＮＡ、ｍＲＮＡ、ｍｉＲＮＡ、ミトコンドリアＲＮＡ、ｒＲＮＡ、またはｔＲＮＡなど）の量は、その種類のＤＮＡまたはＲＮＡの総量の少なくとも２％、３％、４％、５％、６％、７％、８％、９％、１０％、１１％、１２％、１４％、１６％、１８％、２０％、または２５％である。一部の態様では、ある種類のＤＮＡ（無細胞ＤＮＡ、無細胞ミトコンドリアＤＮＡ、無細胞核内ＤＮＡ、細胞内ＤＮＡ、またはミトコンドリアＤＮＡなど）またはＲＮＡ（無細胞ＲＮＡ、細胞内ＲＮＡ、細胞質ＲＮＡ、コーディング細胞質ＲＮＡ、非コーディング細胞質ＲＮＡ、ｍＲＮＡ、ｍｉＲＮＡ、ミトコンドリアＲＮＡ、ｒＲＮＡ、またはｔＲＮＡなど）の総量の少なくとも２％、３％、４％、５％、６％、７％、８％、９％、１０％、１１％、１２％、１４％、１６％、１８％、２０％、または２５％が、癌などの疾患または障害、あるいは癌などの疾患または障害のリスク増加に関連する特定の多型または変異（欠失または重複など）を有する。

【０５９１】

一部の態様では、無細胞ＤＮＡを内包する。一部の態様では、無細胞ＤＮＡを内包しない。

【０５９２】

一部の態様では、腫瘍ＤＮＡのＤＮＡ全体に対する比率（たとえば腫瘍の無細胞ＤＮＡの無細胞ＤＮＡ全体に対する比率、または特定の変異を有する腫瘍の無細胞ＤＮＡの無細胞ＤＮＡ全体に対する比率など）を決定する。一部の態様では、変異が一塩基多様体、コピー数多様体、示差的メチル化、またはそれらの組み合わせである可能性のある場合に、複数の変異について腫瘍ＤＮＡの比率を決定してもよい。一部の態様では、腫瘍率の計算値が最も高い１つの変異または変異の集合について計算した平均腫瘍率を、試料中の実際の腫瘍率とする。一部の態様では、すべての変異について計算した平均腫瘍率を試料中の実際の腫瘍率とする。一部の態様では、（腫瘍率が高いほど、進行した癌の段階と関連づけることができるため）この腫瘍率を用いて癌を病期分類する。一部の態様では、大きな腫瘍ほど血漿中の腫瘍ＤＮＡ率と相関し得るため、腫瘍率を用いて癌をサイズ分類する。一部の態様では、腫瘍率を用いて１つ以上の変異を有する腫瘍の割合をサイズ分類する、というのは血漿中の腫瘍率の測定値と所定の変異遺伝型を有する組織のサイズとの間には相関があるかもしれないからである。たとえば、所定の変異遺伝型を有する組織のサイズは、特定の変異に着目して計算できる腫瘍ＤＮＡの比率と相関しているかも知れない。

【０５９３】

データベースの例

本発明はさらに、本発明の方法から得た１つ以上の結果を含むデータベースを特徴とす

る。たとえば、データベースは、１人以上の対象者に関する以下のいずれかの情報を含む記録を有してもよい：確認された任意の多型／変異（ＣＮＶなど）；多型／変異と疾患または障害、あるいは疾患または障害のリスク増加との任意の既知の関連性；コードされたｍＲＮＡまたはタンパク質の発現または活性レベルに対する多型／変異の影響；疾患または障害に関連するＤＮＡ、ＲＮＡ、または細胞（たとえば、疾患または障害に関連する多型／変異を有するＤＮＡ、ＲＮＡ、または細胞）の、試料中のＤＮＡ、ＲＮＡ、または細胞全体に対する比率；多型／変異の確認に使用した試料（血液試料または特定の組織から得た試料など）の入手源；疾患細胞数；後の反復試験（たとえば疾患または障害の進行や寛解を監視するために試験を繰り返すこと）の結果；疾患または障害に関する他の試験の結果；対象者が診断された疾患または障害の種類；施される治療；その治療に対する反応；その治療の副作用；症状（疾患または障害に関連する症状など）；寛解の長さと数；生存期間の長さ（最初の試験から死亡までの時間の長さ、または診断から死亡までの時間の長さなど）；死因；およびそれらの組み合わせ。

【０５９４】

一部の態様では、データベースは、１人以上の対象者に関する以下のいずれかの情報を含む記録を有する：確認された任意の多型／変異；多型／変異と癌または癌のリスク増加との任意の既知の関連性；コードされたｍＲＮＡまたはタンパク質の発現または活性レベルに対する多型／変異の影響；癌のＤＮＡ、ＲＮＡ、または細胞の、試料中のＤＮＡ、ＲＮＡ、または細胞全体に対する比率；多型／変異の確認に使用した試料（血液試料または特定の組織から得た試料など）の入手源；癌細胞数；腫瘍のサイズ；後の反復試験（たとえば癌の進行や寛解を監視するために試験を繰り返すこと）の結果；癌に関する他の試験の結果；対象者が診断された癌の種類；施される治療；その治療に対する反応；その治療の副作用；症状（癌に関連する症状など）；寛解の長さと数；生存期間の長さ（最初の試験から死亡までの時間の長さ、または癌の診断から死亡までの時間の長さなど）；死因；およびそれらの組み合わせ。一部の態様では、治療反応は以下のいずれかを含む：腫瘍（たとえば、良性腫瘍または癌性腫瘍）のサイズの縮小または安定；腫瘍のサイズ増大の遅れまたは防止；腫瘍細胞数の低下または安定；腫瘍の消失から再発までの無病生存期間の延長；最初または後の腫瘍発生の予防；腫瘍に関連する有害な症状の低減または安定化；またはそれらの組み合わせ。一部の態様では、癌などの疾患または障害に関する１つ以上の他の試験の結果（たとえば、スクリーニング検査、医用画像、または組織試料の顕微鏡検査の結果）が含まれる。

【０５９５】

このような一側面では、本発明は、少なくとも５個、１０個、10^2個、10^3個、10^4個、10^5個、10^6個、10^7個、10^8個、またはそれ以上の記録を含む電子データベースを特徴とする。一部の態様では、データベースは少なくとも５人、１０人、10^2人、10^3人、10^4人、10^5人、10^6人、10^7人、10^8人、またはそれ以上の異なる対象者に関する記録を含む。

【０５９６】

別の態様では、本発明は、本発明のデータベースとユーザインタフェースとを含むコンピュータを特徴とする。一部の態様では、ユーザインタフェースは、１つ以上の記録に含まれている情報の一部またはすべてを表示可能である。一部の態様では、ユーザインタフェースは、以下を表示できる：（ｉ）コンピュータに記録が保存されている多型または変異を含むと確認された１つ以上の癌の種類；（ｉｉ）コンピュータに記録が保存されている特定の種類の癌において確認された１つ以上の多型または変異；（ｉｉｉ）コンピュータに記録が保存されている特定の種類の癌あるいは特定の多型または変異に関する予後情報；（ｉｖ）コンピュータに記録が保存されている多型または変異を有する癌に対して有用な１つ以上の化合物またはその他の治療法；（ｖ）コンピュータに記録が保存されているｍＲＮＡまたはタンパク質の発現または活性を調節する１つ以上の化合物；ならびに（ｖｉ）コンピュータに記憶された化合物によって発現または活性が調節される１つ以上のｍＲＮＡ分子またはタンパク質。コンピュータの内部構成要素としては、一般的に、メモ

リに接続されたプロセッサが挙げられる。外部構成要素としては、通常、大容量記憶装置（たとえばハードディスクドライブ）と；ユーザ入力装置（たとえば、キーボードおよびマウス）と；ディスプレイ（たとえばモニター）と；コンピュータシステムを他のコンピュータに接続してデータの共有およびタスク処理を可能にするネットワークリンクとを含む。操作中、プログラムをこのシステムのメモリにロードしてもよい。

【０５９７】

別の態様では、本発明は本発明のいずれかの方法の１つ以上の工程を含む、コンピュータで実行される処理を特徴とする。

【０５９８】

リスク因子の例

一部の態様では、癌などの疾患または障害に関する１つ以上のリスク因子についても対象者を評価する。リスク因子の例としては、その疾患または障害に関する家族歴、生活習慣（喫煙、発癌性物質への暴露など）、１種以上のホルモンまたは血清タンパク質（肝癌ではα胎児性タンパク質（ＡＦＰ）、結腸直腸癌では癌胎児性抗原（ＣＥＡ）、あるいは前立腺癌では前立腺特異抗原（ＰＳＡ））のレベルが挙げられる。一部の態様では、腫瘍のサイズおよび／または数を測定し、対象者の予後の決定または対象者への治療の選択に用いる。

【０５９９】

スクリーニング方法の例

必要に応じ、癌などの疾患または障害の有無を確認でき、あるいは、癌などの疾患または障害を任意の標準の方法で分類できる。たとえば、癌などの疾患または障害は、特定の徴候と症状の存在、腫瘍生検、スクリーニング検査、または医用画像（乳房Ｘ線像または超音波画像など）などの複数の方法で検出可能である。可能性のある癌が検出されると、それを組織試料の顕微鏡検査で診断してもよい。一部の態様では、診断を受けた対象者は、その疾患または障害の進行あるいは寛解または再発を監視するために、その疾患または障害について本発明の方法または既知の試験による反復試験を複数の時点で受ける。

【０６００】

癌の例

本発明のいずれかの方法で診断、予知、安定化、治療、または予防できる癌の例としては、固形腫瘍、癌腫、肉腫、リンパ腫、白血病、胚細胞腫瘍、または芽細胞腫が挙げられる。様々な態様では、癌は、急性リンパ性白血病、急性骨髄性白血病、副腎皮質癌、ＡＩＤＳ関連癌、ＡＩＤＳ関連リンパ腫、肛門癌、虫垂癌、星細胞腫（小児小脳星細胞腫または小児大脳星細胞腫）、基底細胞癌、胆管癌（肝外胆管癌など）、膀胱癌、骨腫瘍（骨肉腫または悪性線維性組織球腫など）、脳幹神経膠腫、脳癌（小脳星細胞腫、大脳星細胞腫／悪性神経膠腫、上衣腫、髄芽腫、テント上原始神経外胚葉性腫瘍、または視経路および視床下部神経膠腫（visual pathway and hypothalamic glioma）など）、膠芽腫、乳癌、気管支腺腫または気管支カルチノイド、バーキットリンパ腫、カルチノイド腫瘍（小児カルチノイド腫瘍または消化管カルチノイド腫瘍など）、癌腫、中枢神経系リンパ腫、小脳星細胞腫または小脳悪性神経膠腫（小児小脳星細胞腫または小児小脳悪性神経膠腫など）、子宮頚癌、小児癌、慢性リンパ性白血病、慢性骨髄性白血病、慢性骨髄増殖性疾患、直腸癌、皮膚Ｔ細胞性リンパ腫、線維形成性小円形細胞腫瘍、子宮内膜癌、上衣腫、食道癌、ユーイング肉腫、ユーイング腫瘍ファミリーの腫瘍、頭蓋外胚細胞腫瘍（小児頭蓋外胚細胞腫瘍など）、性腺外生殖細胞腫瘍、眼癌（眼球内黒色腫または網膜芽細胞腫、眼癌など）、胆嚢癌、胃癌、消化管カルチノイド腫瘍、消化管間質腫瘍、胚細胞腫瘍（頭蓋外胚細胞腫瘍、性腺外生殖細胞腫瘍、または卵巣胚細胞腫瘍など）、妊娠性絨毛性疾患腫瘍、神経膠腫（脳幹神経膠腫、小児大脳星細胞腫、または小児視経路および視床下部神経膠腫など）、胃カルチノイド、有毛細胞白血病、頭頚部癌、悪性心臓腫瘍（heart cancer）、肝細胞（肝臓）癌、ホジキンリンパ腫、下咽頭癌、視床下部および視経路神経膠腫（小児視経路神経膠腫）、島細胞癌（内分泌島細胞癌または膵島細胞癌など）、カポジ肉腫、腎癌、咽頭癌、白血病（たとえば急性リンパ性白血病、急性骨髄性白血病、慢性リンパ性白

血病、慢性骨髄性白血病、または有毛細胞白血病）、口唇癌または口腔癌、脂肪肉腫、肝癌（非小細胞癌または小細胞癌など）、肺癌、リンパ腫（ＡＩＤＳ関連リンパ腫、バーキットリンパ腫、皮膚Ｔ細胞性リンパ腫、ホジキンリンパ腫、非ホジキンリンパ腫、または中枢神経系リンパ腫など）、マクログロブリン血症（ワルデンシュトレーム・マクログロブリン血症、骨悪性線維性組織球腫または骨肉腫、髄芽腫（小児髄芽腫など）、黒色腫、メルケル細胞癌、中皮腫（成人または小児中皮腫）、潜伏性口腔癌を伴う転移頚部扁平上皮癌、多発性内分泌腫瘍症候群（小児多発性内分泌腫瘍症候群など）、多発性骨髄腫または形質細胞腫瘍、菌状息肉症、骨髄異形成症候群、骨髄異形成または骨髄増殖性疾患、骨髄性白血病（慢性骨髄性白血病など）、骨髄性白血病（成人急性骨髄性白血病または小児急性骨髄性白血病など）、骨髄増殖性疾患（慢性骨髄増殖性疾患など）、鼻腔癌または副鼻腔癌、上咽頭癌、神経芽腫、口腔癌、中咽頭癌、骨肉腫または骨悪性線維性組織球腫、卵巣癌、卵巣上皮癌、卵巣胚細胞腫瘍、卵巣低悪性度腫瘍、膵癌（膵島細胞癌など）、副鼻腔癌または鼻腔癌、副甲状腺癌、陰茎癌、咽頭癌、褐色細胞腫、松果体星細胞腫、松果体胚細胞腫、松果体芽腫またはテント上原始神経外胚葉性腫瘍（小児松果体芽腫または小児テント上原始神経外胚葉性腫瘍など）、下垂体腺腫、形質細胞腫、胸膜肺芽腫、原発性中枢神経系リンパ腫、癌、直腸癌、腎細胞癌、尿管癌（腎盂移行上皮癌または尿管移行上皮癌など）、網膜神経膠腫、横紋筋肉腫（小児横紋筋肉腫など）、唾液腺癌、肉腫（ユーイング腫瘍ファミリーの肉腫、カポジ肉腫、軟部肉腫、または子宮肉腫など）、セザリー症候群、皮膚癌（非黒色腫皮膚癌、黒色腫、またはメルケル細胞皮膚癌など）、小腸癌、扁平上皮癌、テント上原始神経外胚葉性腫瘍（小児テント上原始神経外胚葉性腫瘍など）、Ｔ細胞リンパ腫（皮膚Ｔ細胞リンパ腫など）、精巣癌、咽頭癌、胸腺腫（小児胸腺腫など）、胸腺腫または胸腺癌、甲状腺癌（小児甲状腺癌など）、絨毛性腫瘍（妊娠性絨毛性腫瘍など）、原発不明癌（成人または小児の原発不明癌など）、尿道癌（子宮内膜癌、子宮癌など）、子宮肉腫、腟癌、視経路および視床下部神経膠腫（小児視経路および視床下部神経膠腫など）、外陰癌、ワルデンシュトレーム・マクログロブリン血症、あるいはウィルムス腫瘍（小児ウィルムス腫瘍など）である。様々な態様では、癌は転移した癌または転移していない癌である。

【０６０１】

癌は、ホルモン関連癌またはホルモン依存性癌（たとえば、エストロゲン関連癌またはアンドロゲン関連癌）であってもなくてもよい。本発明の方法および／または組成物を使用して良性腫瘍または悪性腫瘍を診断、予知、安定化、治療、または予防してもよい。

【０６０２】

一部の態様では、対象者は癌症候群を有する。癌症候群は１つ以上の遺伝子における遺伝的変異により、影響を受けた個体が癌を発症しやすくなる遺伝的障害であり、これらの癌の発症を早めることもある。癌症候群は多くの場合、癌の発症だけでなく複数の独立した原発腫瘍の生涯の発症リスクを高める。これらの症候群の多くは、細胞の癌細胞化の予防に関与する腫瘍抑制遺伝子の変異によって引き起こされる。影響を受ける可能性のある他の遺伝子はＤＮＡ修復遺伝子、癌遺伝子、および血管の生成（血管新生）に関与する遺伝子である。遺伝性癌症候群の一般例は、遺伝性乳癌・卵巣癌症候群、および遺伝性非ポリポーシス大腸癌（リンチ症候群）である。

【０６０３】

一部の態様では、１つ以上の多型または変異をＫ－ｒａｓ、ｐ５３、ＢＲＡ、ＥＧＦＲ、またはＨＥＲ２に有する対象者に、それぞれＫ－ｒａｓ、ｐ５３、ＢＲＡ、ＥＧＦＲ、またはＨＥＲ２を標的とする治療を施す。

【０６０４】

本発明の方法は、任意の種類の細胞、組織、または器官の悪性または良性の腫瘍の治療に一般的に応用できる。

【０６０５】

治療法の例

必要に応じ、癌などの疾患または障害、あるいは癌などの疾患または障害のリスク増加

を安定化、治療、または予防するための任意の治療法を対象者（たとえば、本発明のいずれかの方法で癌または癌のリスクを有すると確認された対象者）に施すことができる。様々な態様では、治療法は癌などの疾患または障害に対する既知の治療法または併用療法（細胞毒性薬、標的治療、免疫療法、ホルモン治療、放射線治療、癌細胞または癌細胞になる可能性のある細胞の除去手術、幹細胞移植、骨髄移植、光線力学治療、対症療法、またはそれらの組み合わせなど）である。一部の態様では、治療（予防的薬物療法など）によって、癌などの疾患または障害のリスクの高い対象者の癌などの疾患または障害を予防し、その進行を遅らせ、あるいは重症度を低下させる。

【０６０６】

一部の態様では、標的治療は、癌の増殖と生存に寄与する、癌の特異的な遺伝子、タンパク質、または組織環境を標的とする治療法である。この種の治療は、癌細胞の増殖と拡散を阻害しつつ正常細胞の損傷を制限し、通常、他の癌薬物療法よりも副作用が少なくなる。

【０６０７】

より多く成功している方法の１つは、腫瘍周辺の血管新生（新しい血管の増殖）を標的とすることであった。ベバシズマブ（アバスチン（登録商標））、レナリドミド（レブラミド（登録商標））、ソラフェニブ（ネクサバール（登録商標））、スニチニブ（スーテント（登録商標））、サリドマイド（サロミド（登録商標））などの標的治療は血管新生を妨害する。別の例は、ＨＥＲ２を過剰発現する癌（一部の乳癌など）に対するＨＥＲ２標的治療（トラスツズマブまたはラパチニブなど）の利用である。一部の態様では、モノクローナル抗体を用いて癌細胞外部で特定の標的を阻害する。その例は、アレムツズマブ（キャンパス－１Ｈ（登録商標））、ベバシズマブ、セツキシマブ（アービタックス（登録商標））、パニツムマブ（ベクティビックス（登録商標））、ペルツズマブ（オムニターグ（登録商標））、リツキシマブ（リツキサン（登録商標））、およびトラスツズマブである。一部の態様では、モノクローナル抗体トシツモマブ（ベキサール（登録商標））を用いて腫瘍を放射線照射する。一部の態様では、経口小分子が癌細胞内部の癌の過程を阻害する。例としては、ダサチニブ（スプリセル（登録商標））、エルロチニブ（タルセバ（登録商標））、ゲフィチニブ（イレッサ（登録商標））、イマチニブ（グリベック（登録商標））、ラパチニブ（タイカーブ（登録商標））、ニロチニブ（タシグナ（登録商標））、ソラフェニブ、スニチニブ、テムシロリムス（トーリセル（登録商標））が挙げられる。一部の態様では、プロテアソーム阻害剤（多発性骨髄腫薬ボルテゾミブ（ベルケイド（登録商標））は、細胞内の他のタンパク質を分解する酵素と呼ばれる特殊なタンパク質を妨害する。

【０６０８】

一部の態様では、癌と闘うために本来の生体防御力を高めるように免疫療法を設計する。例示の免疫療法の種類では、体内で産生された物質または研究室で生成された物質を用いて免疫系機能を増強、標的化、または回復する。

【０６０９】

一部の態様では、ホルモン治療によって体内のホルモン量を低下させて癌を治療する。一部の乳癌や前立腺癌などの数種類の癌は、ホルモンと呼ばれる生体内の天然化学物質の存在下で増殖し拡散する。様々な態様では、ホルモン療法によって前立腺癌、乳癌、甲状腺癌、および生殖器系の癌を治療する。

【０６１０】

一部の態様では、治療は、罹患した骨髄を造血幹細胞と呼ばれる非常に特殊な細胞で置き換える幹細胞移植を含む。造血幹細胞は血流中と骨髄に見られる。

【０６１１】

一部の態様では、治療は、光増感剤と呼ばれる特殊な薬剤を光と共に用いて癌細胞を殺す光線力学療法を含む。これらの薬剤は、ある種の光で活性化されると作用する。

【０６１２】

一部の態様では、治療は、癌細胞または癌細胞になる可能性のある細胞の除去手術（腫

瘍摘出術または乳房切除術など）を含む。たとえば、乳癌感受性遺伝子の変異（ＢＲＣＡ１またはＢＲＣＡ２遺伝子変異）を有する女性は、リスク低減卵管卵巣摘出術（卵管および卵巣の除去）および／またはリスク低減両側乳房切除（両乳房の除去）によって乳癌および卵巣癌のリスクを低減することができる。一部の癌の治療を含む非常に慎重な外科手術には、非常に強力で精密な光線であるレーザーを刃（メス）の代わりに用いることができる。

【０６１３】

癌の進行を遅らせる、阻止する、または除去する治療（疾患標的治療ともいう）に加え、癌治療の重要な部分は、対象者の症状と疼痛や嘔気などの副作用を軽減することである。これには、対象者の身体的、感情的、社会的な欲求を支援すること、すなわち緩和ケアまたは支持療法と呼ばれる方法が含まれる。多くの場合、患者は疾患標的治療と症状を緩和するための治療を同時に受ける。

【０６１４】

例示の治療には以下が含まれる：アクチノマイシンＤ、アドセトリス（登録商標）、アドリアマイシン、アルデスロイキン、アレムツズマブ、アリムタ（登録商標）、Amsidine（登録商標）、アムサクリン、アナストロゾール、アレジア（登録商標）、アリミデックス（登録商標）、アロマシン（登録商標）、アスパラギナーゼ、アバスチン（登録商標）、ベバシズマブ、ビカルタミド、ブレオマイシン、Bondronat（登録商標）、Bonefos（登録商標）、ボルテゾミブ、Busilvex（登録商標）、ブスルファン、カンプト（登録商標）、カペシタビン、カルボプラチン、カルムスチン、カソデックス（登録商標）、セツキシマブ、Chimax（登録商標）、クロランブシル、シメチジン、シスプラチン、クラドリビン、クロドロン酸、クロファラビン、クリサンタスパーゼ、シクロホスファミド、酢酸シプロテロン、Cyprostat（登録商標）、シタラビン、シトキサン（登録商標）、Dacarbozine、ダクチノマイシン、ダサチニブ、ダウノルビシン、デキサメタゾン、ジエチルスチルベストロール、ドセタキセル、ドキソルビシン、Drogenil（登録商標）、Emcyt（登録商標）、エピルビシン、Eposin（登録商標）、アービタックス（登録商標）、エルロチニブ、Estracyte（登録商標）、エストラムスチン、Etopophos（登録商標）、エトポシド、Evoltra（登録商標）、エキセメスタン、フェアストン（登録商標）、フェマーラ（登録商標）、フィルグラスチム、フルダラ（登録商標）、フルダラビン、フルオロウラシル、フルタミド、ゲフィチニブ、ゲムシタビン、ジェムザール（登録商標）、グリベック（Gleevec, Glivec）（登録商標）、Gonapeptyl Depot（登録商標）、ゴセレリン、ハラベン（登録商標）、ハーセプチン（登録商標）、Hycamptin（登録商標）、ヒドロキシカルバミド、イバンドロン酸、イブリツモマブ、イダルビシン、イフォスファミド、インターフェロン、メシル酸イマチニブ、イレッサ（登録商標）、イリノテカン、ジェブタナ（登録商標）、Lanvis（登録商標）、ラパチニブ、レトロゾール、リューケラン（登録商標）、リュープリン（登録商標）、Leustat（登録商標）、ロムスチン、マブキャンパス（登録商標）、マブセラ（登録商標）、Megace（登録商標）、メゲストロール、メトトレキサート、ミトキサントロン、マイトマイシン、Mutulane、ミレラン（登録商標）、ナベルビン（登録商標）、Neulasta（登録商標）、ニューポジェン（登録商標）、ネクサバール（登録商標）、Nipent（登録商標）、ノルバデックスＤ（登録商標）、ノバントロン（登録商標）、オンコビン（登録商標）、パクリタキセル、パミドロン酸、ＰＣＶ、ペメトレキセド、ペントスタチン、パージェタ（登録商標）、プロカルバジン、プロベンジ（登録商標）、プレドニゾロン、Prostrap（登録商標）、ラルチトレキセド、リツキシマブ、スプリセル（登録商標）、ソラフェニブ、Soltamox（登録商標）、ストレプトゾシン、スチルベストロール、Stimuvax（登録商標）、スニチニブ、スーテント（登録商標）、Tabloid（登録商標）、タガメット（登録商標）、Tamofen（登録商標）、タモキシフェン、タルセバ（登録商標）、タキソール（登録商標）、タキソテール（登録商標）、テガフール・ウラシル、テモダール（登録商標）、テモゾロミド、サリドマイド、Thioplex（登録商標）、チオテパ、チオグアニン、Tomudex（登録商標）、トポテカン、トレミフェン、トラスツズマブ、トレチノイン、トレオスルファン、トリエチレンホスホラミド、トリプトレリン、

タイバーブ（登録商標）、Uftoral（登録商標）、ベルケイド（登録商標）、ベプシド（登録商標）、ベサノイド（登録商標）、ビンクリスチン、ビノレルビン、ザーコリ（登録商標）、ゼローダ（登録商標）、ヤーボイ（登録商標）、ザクティマ（登録商標）、ザノサー（登録商標）、Zavedos（登録商標）、Zevelin、ゾラデックス（登録商標）、ゾレドロン酸、ゾメタ（登録商標）（ゾレドロン酸）、およびザイティガ（登録商標）。

【０６１５】

対象者がｍＲＮＡまたはタンパク質の変異型（たとえば癌に関連する型）と野生型（たとえば癌に関連しない型）両方を発現する場合、治療は、野生型の発現または活性の阻害に比べて変異型の発現または活性を少なくとも２倍、５倍、１０倍、または２０倍阻害することが好ましい。複数の治療薬を同時または逐次的に使用することによって癌の発生を著しく低減し、治療を受け治療抵抗性になる癌の数を減少させることができる。さらに、併用療法の一部として用いられる治療薬は、その治療薬を単独使用する場合の相当用量に比べて癌の治療に必要な用量が少ない。併用療法では各化合物の用量が少ないため、化合物が起こし得る有害な副作用の重症度が低減される。

【０６１６】

一部の態様では、癌のリスク増加を有すると確認された対象者は、特定のリスク因子を回避するか生活習慣を変えることで、それ以上の何らかの癌リスクを低減することができる。

【０６１７】

一部の態様では、多型、変異、リスク因子、またはそれらの任意の組み合わせを用いて対象者の治療計画を選択する。一部の態様では、対象者の癌のリスクが高いほど、あるいは予後が悪いほど、多い用量または多数の治療法を選択する。

【０６１８】

単独療法または併用療法に用いるその他の化合物

必要に応じ、癌などの疾患または障害、あるいは癌などの疾患または障害のリスク増加を安定化、治療、または予防するための追加の化合物は、天然の産物と合成（または半合成）抽出物の大きなライブラリ、または化合物ライブラリから、当業者に公知の方法に従って特定することができる。創薬・医薬品開発の分野の当業者には明らかであるが、試験抽出物や化合物の正確な入手源は本発明の方法にとって決定的ではない。したがって、実質的にはあらゆる数の化学抽出物または化合物を、特定の種類の癌または特定の対象者の細胞に対する効果、あるいは癌関連分子（たとえば、特定の種類の癌で活性または発現が変化したことがわかっている癌関連分子）の活性または発現に対する効果について、スクリーニングすることができる。粗抽出物が癌関連分子の活性または発現を調節するとわかれば、当技術分野で公知の方法で陽性の先導抽出物をさらに分画し、観測された効果の原因となる化学成分を単離してもよい。

【０６１９】

治療法を試験するための試験法および動物モデルの例

本明細書に開示する治療法のうちの１つ以上を、癌などの疾患または障害に及ぼす影響について、細胞株（たとえば、癌または癌のリスク増加を有すると診断された対象で確認された１つ以上の変異を有する細胞株）あるいはその疾患または障害の動物モデル（ＳＣＩＤマウスモデルなど）を用いて試験することができる（Jain et al., Tumor Models In Cancer Research, ed. Teicher, Humana Press Inc., Totowa, N.J., pp. 647-671, 2001；この文献の全内容は参照によって本明細書に組み込まれる）。また、癌などの疾患または障害、あるいは癌などの疾患または障害のリスクの増加を安定化、治療、または予防するための特定の治療の有効性を決定するのに使用できる標準の試験法および動物モデルは多数存在する。ヒトを対象とする標準の臨床試験で治療を試験することもできる。

【０６２０】

特定の対象者に対し好ましい治療を選択するために、対象者で変異している１つ以上の遺伝子の発現または活性に対する化合物の効果を試験することができる。たとえば、特定のｍＲＮＡ分子またはタンパク質の発現を調節する化合物の能力を、標準的なノーザン分

析、ウエスタン分析、またはマイクロアレイ分析を用いて検出できる。一部の態様では、（ｉ）対象者で（たとえば対象者から採取した試料で）正常レベルよりも発現レベルまたは活性レベルの高い、癌を促進するｍＲＮＡ分子またはタンパク質の発現または活性を阻害するか、（ｉｉ）対象者で正常レベルよりも発現レベルまたは活性レベルの低い、癌を阻害するｍＲＮＡ分子またはタンパク質の発現または活性を促進する、１種以上の化合物を選択する。（ｉ）対象者の、癌に関連する変異を有するｍＲＮＡ分子またはタンパク質の最大数を調節し、かつ（ｉｉ）対象者の、癌に関連する変異を有さないｍＲＮＡ分子またはタンパク質の最小数を調節する、単独療法または併用療法を選択する。一部の態様では、選択された単独療法または併用療法は高い薬物有効性を有し、有害な副作用を起こしたとしてもごくわずかにしか起こさない。

【０６２１】

上記の対象者に特異的な分析の代わりに、ＤＮＡチップを用いて特定の種類の初期または後期の癌（たとえば乳癌細胞）でのｍＲＮＡ分子の発現を正常組織での発現と比較することができる（Marrack et al., Current Opinion in Immunology 12, 206-209, 2000；Harkin, Oncologist. 5:501-507, 2000；およびPelizzari et al., Nucleic Acids Res. 28(22):4577-4581, 2000（これらの文献それぞれの全内容は参照によって本明細書に組み込まれる））。この分析に基づいて、その種類の癌で発現量が変化するｍＲＮＡまたはタンパク質の発現を調節するように、この種類の癌を有する対象者に対する単独療法または併用療法を選択できる。

【０６２２】

発現特性は、特定の対象者または対象者群の治療法の選択に使用できるだけでなく、治療中に起こるｍＲＮＡおよび／またはタンパク質発現の変化の監視にも使用できる。例えば、発現特性を用いて、癌関連遺伝子の発現が正常レベルに戻ったかどうかを決定し、戻っていない場合には、治療に含まれる１種以上の化合物の用量を、対応する癌関連遺伝子の発現レベルに対する治療の効果を増減するように変更することができる。また、この分析は、治療が他の遺伝子（たとえば、有害な副作用に関連する遺伝子）の発現に影響を及ぼすかどうかを決定するために使用できる。必要に応じて、望ましくない副作用を予防または軽減するために、治療の用量または組成を変更することができる。

【０６２３】

製剤と投与法の例

癌などの疾患または障害、あるいは癌などの疾患または障害のリスク増加を安定化、治療、あるいは予防するために、組成物を当業者に公知の任意の方法をで製剤かつ投与できる（たとえば、米国特許第８，３８９，５７８号明細書および米国特許第８，３８９，５５７号明細書を参照；これらの文献それぞれの全内容は参照によって本明細書に組み込まれる）。製剤および投与のための一般的な技術は"Remington: The Science and Practice of Pharmacy,” 21st Edition, Ed. David Troy, 2006, Lippincott Williams & Wilkins, Philadelphia, Paに記載されている（この文献の全内容は参照によって本明細書に組み込まれる）。液剤、懸濁剤、錠剤、カプセル剤、丸剤、散剤、顆粒剤、ゲル剤、軟膏剤、坐剤、注射剤、吸入剤、および噴霧剤がこのような製剤の例である。例えば、当技術分野で公知の別の方法も用いて徐放性の経口製剤を調製することができる。たとえば、有効成分の適切な徐放性剤形は、基質錠剤（matrix tablet）またはカプセル剤組成物であってもよい。基質を形成する適切な材料としては、たとえば、ワックス類（たとえば、カルナウバロウ、ミツロウ、パラフィンワックス、セレシン、シェラックワックス、脂肪酸、脂肪アルコールなど）、油、硬化油脂（たとえば、硬化ナタネ油、硬化ヒマシ油、硬化牛脂、硬化ヤシ油、硬化ダイズ油）、および重合体（たとえばヒドロキシプロピルセルロース、ポリビニルピロリドン、ヒドロキシプロピルメチルセルロース、ポリエチレングリコールなど）などが挙げられる。他の適切な基質錠剤材料としては、微結晶性セルロース、粉末セルロース、ヒドロキシプロピルセルロース、エチルセルロースと、他の担体と賦形剤である。錠剤は顆粒、被覆粉末、またはペレットを含んでもよい。錠剤は多層型であってもよい。必要に応じ、完成した錠剤を被覆してもよいししなくてもよい。

【０６２４】
このような組成物の一般的な投与経路としては、経口投与、舌下投与、口腔投与、局所投与、経皮投与、吸入投与、非経口投与（たとえば皮下、静脈内、筋肉内、あるいは胸骨内の注射または点滴技術）、直腸投与、膣投与、鼻腔内投与が挙げられるがこれらに限定するものではない。好ましい態様では、治療は徐放装置を用いて実行される。本発明の組成物は、組成物に含有される活性成分が組成物の投与時に生物学的に利用可能であるように処方される。組成物は、１つ以上の投与単位の形態をとってもよい。組成物は、１種、２種、３種、４種、またはそれ以上の活性成分を含有してもよく、必要に応じて１種、２種、３種、４種、またはそれ以上の不活性成分を含有してもよい。

【０６２５】
代替的な態様
本明細書に記載の任意の方法は、物理的な形式（コンピュータ画面、または紙への印刷）のデータ出力を含んでもよい。本発明の任意の方法を、医師がそれに基づいて判断できるような形式で、すぐ使用可能なデータを出力することと組み合わせてもよい。本明細書に記載した対象個体に関する遺伝データ決定の一部の態様を、出生前診断の状況での胎児の潜在的な染色体異常（欠失または重複など）またはそれがないことを（必要に応じて、中絶するかしないかの判断と組み合わせて）医療専門家が通知することと組み合わせてもよい。本明細書に記載の態様の一部を、すぐ使用可能なデータを出力し、臨床治療を行うという臨床判断をすること、または行わないという臨床判断をすることと組み合わせてもよい。

【０６２６】
一部の態様で、本発明の任意の方法の結果（たとえば、欠失または重複の有無）を開示する報告を作成する方法を本明細書に記載する。本発明の方法で得た結果を含む報告を作成してもよく、それを医師に電子送信してもよいし、出力装置（デジタル式の報告）に表示してもよい。あるいは、文書での報告書（たとえば報告書の印刷物）を医師に提供してもよい。また、記載した方法を、臨床治療を行うという臨床判断を実際にすること、または何もしないという臨床判断をすることと組み合わせてもよい。

【０６２７】
特定の態様では、本発明は、本明細書に開示された多重ＰＣＲ法を用いて同一の試料からＣＮＶとＳＮＶの両方を検出するための試薬、キット、方法、ならびにこのような方法を実行するためのコード化された命令を有するコンピュータシステムおよびコンピュータ媒体を提供する。特定の好ましい形態では、試料は、循環腫瘍ＤＮＡを含むと考えられる単一細胞試料または血漿試料である。これらの態様は、本明細書に開示された高感度多重ＰＣＲ法を用いてＣＮＶおよびＳＮＶについて単一細胞または血漿から得たＤＮＡ試料を調べることにより、特にＣＮＶを示す癌（乳癌、卵巣癌、肺癌など）の場合に、ＣＮＶまたはＳＮＶのいずれかについて個々に調査する場合に比べて癌の検出を改善できるという発見を活用する。特定の例示的態様によるＣＮＶ分析法は、５０個～１００，０００個、５０個～１０，０００個、または５０個～１，０００個のＳＮＰについて調査し、ＳＮＶ分析法は、５０個～１０００個のＳＮＶまたは５０個～５００個のＳＮＶ、あるいは５０個～２５０個のＳＮＶについて調べる。本明細書に記載の、癌（たとえば、乳癌、肺癌、卵巣癌など、ＣＮＶをおよびＳＮＶを示すとわかっている癌）を有すると考えられる対象者の血漿でＣＮＶおよび／またはＳＮＶを検出する方法は、遺伝的構成の観点で不均質な癌細胞集団で構成されることの多い腫瘍のＣＮＶおよび／またはＳＮＶを検出するという利点を提供する。したがって、特定の腫瘍領域のみの分析に着目する従来の方法は多くの場合、他の腫瘍領域に存在するＣＮＶまたはＳＮＶを見落とす可能性がある。血漿試料は、任意のＣＮＶおよび／またはＳＮＶを検出するために調査できる液体細胞診試料として機能する。

【０６２８】
コンピュータアーキテクチャの例
図６９は、本発明の態様を実行するのに有用な例示のシステム構造Ｘ００を示す。シス

テム構造Ｘ００は、１つ以上の研究所情報システム（「ＬＩＳ」）Ｘ０４に接続された分析プラットフォームＸ０８を含む。図６９に示すように、分析プラットフォームＸ０８は、ネットワークＸ０２を通じてLIS X04に接続されてもよい。ネットワークＸ０２は、１つ以上のネットワークタイプ（ＬＡＮ、ＷＡＮ、インターネットなどの任意の組み合わせを含む）の１つ以上のネットワークを含んでもよい。ネットワークＸ０２は、システム構造Ｘ００のいずれかまたはすべての構成要素間の接続を含んでもよい。分析プラットフォームＸ０８は、代替的または追加的にLIS X06に直接接続されてもよい。一態様では、分析プラットフォームＸ０８は、LIS X04からサービス型ソフトウェア（software-as-as-service）の様式で提供された遺伝データを分析し（LIS X04は第三者ＬＩＳである）、また、分析プラットフォームＸ０８は、LIS X06から完全サービス（full-service）様式または自己完結（in-house）様式で提供された遺伝データを分析する（LIS X06と分析プラットフォームＸ０８は同じ当事者に管理される）。分析プラットフォームＸ０８がネットワークＸ０２を通じて情報を提供する一態様では、分析プラットフォームＸ０８はサーバであってもよい。

【０６２９】

一態様では、例示の研究所情報システムＸ０４は、遺伝データを収集、管理、かつ／または保存する１つ以上の公的機関または私的機関を含む。関連技術の当業者には明らかであるが、遺伝データを保護する方法または規格は公知であり、情報セキュリティの様々な技術とポリシー（たとえばユーザ名／パスワード、トランスポート・レイヤー・セキュリティ（Transport Layer Security; TLS）、セキュア・ソケット・レイヤー（Secure Sockets Layer; SSL）、および／または通信セキュリティを提供するその他の暗号化プロトコル）を利用して実行可能である。

【０６３０】

一態様では、例示のシステム構造Ｘ００は、サービス指向アーキテクチャとして動作し、クライアントサーバモデルを使用する。関連分野の当業者には明らかであるが、クライアントサーバモデルによってLIS X04と分析プラットフォームＸ０８は多様な形態で相互作用できる。システムアーキテクチャＸ００は、多様なタイプのネットワークＸ０２上に分散されてもよく、かつ／またはクラウド・コンピューティング・アーキテクチャとして動作してもよい。クラウド・コンピューティング・アーキテクチャは、任意のタイプの分散型ネットワークアーキテクチャを含んでもよい。一例として、クラウド・コンピューティング・アーキテクチャは、サービス型ソフトウェア（software as a service; SaaS）、サービス型インフラストラクチャ（infrastructure as a service; IaaS）、サービス型プラットフォーム（platform as a service; PaaS）、サービス型ネットワーク（network as a service; NaaS）、サービス型データ（data as a service; DaaS）、サービス型データベース（database as a service; DBaaS）、サービス型バックエンド（backend as a service; BaaS）、サービス型テスト環境（test environment as a service; TEaaS）、サービス型ＡＰＩ（API as a service; APIaaS）、サービス型統合プラットフォーム（integration platform as a service; IPaaS）などを提供するのに有用であるが、これらに限定されるものではない。

【０６３１】

例示の態様では、LIS X04およびＸ０６は、それぞれコンピュータ、装置、インタフェースなど、またはそれらの任意のサブシステムを含む。LIS X04およびＸ０６は、オペレーティングシステム（ＯＳ）、すなわち様々な機能（たとえば、ローカルアクセス、メモリへのアクセス、および／またはネットワークＸ０２などを通じたアクセスが可能なデータへのアクセスおよび／またはナビゲーション）を実行するためにインストールされるアプリケーションを含むことができる。一態様では、LIS X04は、アプリケーション・プログラミング・インターフェイス（「ＡＰＩ」）を介して分析プラットフォームＸ０８にアクセスする。LIS X04はさらに、ＡＰＩとは独立して動作可能な１つ以上のネイティブアプリケーション（直接実行可能なプログラム）を含んでもよい。

【０６３２】

例示の態様では、分析プラットフォームＸ０８は、入力プロセッサＸ１２、仮説管理部Ｘ１４、モデラーＸ１６、誤差修正部Ｘ１８、機械学習部Ｘ２０、および出力プロセッサＸ１８のうちの１つ以上を含む。入力プロセッサＸ１２は、LIS X04および／またはＸ０６からの入力を受け、処理する。処理は、LIS X04および／またはＸ０６から受けた任意の入力の構文解析、コード変換、翻訳、適合、またはその他の方法での処理の動作を含むが、これらに限定されるものではない。LIS X04およびＸ０６によってアクセス可能となる１つ以上のデータストリーム、データフィード、データベース、または他のデータソースを介して入力を受けてもよい。誤差修正部Ｘ１８によって上記の誤差修正機構を実行することでデータの誤差を修正してもよい。

【０６３３】

例示の態様では、仮説管理部Ｘ１４は、モデルおよび／またはアルゴリズムとして表される遺伝的分析に関する仮説に従って処理できる状態で入力プロセッサＸ１２から受けた入力を受けるように構成される。このようなモデルおよび／またはアルゴリズムを、たとえば、動態統計値、リアルタイム統計値、および／または履歴統計値、あるいは他の指標に基づいて確率を作成するためにモデラーＸ１６で使用してもよい。このような手法のモデルおよび／またはアルゴリズムの導出や、そのモデルやアルゴリズムへの入力に使用されるデータは、たとえば遺伝データソースＸ１０から仮説管理部Ｘ１４取り込むことができる。遺伝データソースＸ１０は、たとえば核酸シーケンサを含んでもよい。仮説管理部Ｘ１４は、たとえば、そのモデルおよび／またはアルゴリズムに入力する必要のある変数に基づいて仮説を構築するように構成されてもよい。入力されたモデルおよび／またはアルゴリズムを、上記の１つ以上の仮説を作成するためにモデラーＸ１６で使用してもよい。仮説管理部Ｘ１４は、最も尤度の高い仮説に基づいて、特定の値、値の範囲、または推定値を上記の出力として選択することができる。モデラーＸ１６は、機械学習部Ｘ２０によって訓練されたモデルおよび／またはアルゴリズムに従って動作してもよい。たとえば、機械学習部Ｘ２０は、上記のような分類アルゴリズムを訓練セットデータベース（図示せず）に適用することによってこのようなモデルおよび／またはアルゴリズムを開発してもよい。特定の態様では、機械学習部は、１つ以上の対照試料を分析し、本明細書に記載のＳＮＶ検出方法に有用な訓練データセットを作成する。

【０６３４】

仮説管理部Ｘ１４が特定の出力を確認すると、出力プロセッサＸ２２は情報を要求している特定のＬＩＳ１０４または１０６にその出力を戻してもよい。

【０６３５】

本開示の様々な態様は、ソフトウェア、ファームウェア、ハードウェア、またはそれらの組み合わせによってコンピュータ計算装置で実行できる。図７０は、意図した態様またはその一部をコンピュータ可読コードとして実施できる例示のコンピュータシステムＹ００を示す。この例示のコンピュータシステムＹ００に関して様々な態様を説明する。

【０６３６】

図７０の態様では、タスクの処理はプロセッサＹ０２（２台以上の場合もある、以下同じ）によって実行される。しかし、当然ながら、ここでは様々な種類の処理技術（programmable logic array; PLA）、特定用途向け集積回路（application-specific integrated circuits ; ASIC）、マルチコアプロセッサ、マルチプロセッサ、または分散型プロセッサなどの様々なタイプの処理技術など）を利用してもよい。特定の処理タスクを支援するために、画像、マルチメディア、または数学的処理能力などの追加の専用処理リソースを用いてもよい。このような処理リソースは、ハードウェア、ソフトウェア、またはそれらの適切な組み合わせであってもよい。たとえば、プロセッサＹ０２は、画像処理ユニット（graphics-processing unit; GPU）であってもよい。一態様では、ＧＰＵは、計算の多いアプリケーションを電子装置で迅速に処理するよう設計された特殊電子回路であるプロセッサである。ＧＰＵは、計算の多いデータなどの大きなデータブロックの並列処理に効率的な、高度に並列な構造を有してもよい。これに代えるかこれに加えて、プロセッサＹ０２は画像最適化のない特殊な並列処理を行ってもよく、このような並列プロセッサが本

明細書に記載の計算の多い機能を実行する。プロセッサＹ０２は処理アクセラレータ（たとえばＤＳＰまたは他の専用プロセッサ）を含むことができる。

【０６３７】

コンピュータシステムＹ００はさらにメインメモリＹ３０を含み、二次メモリＹ４０も含んでもよい。メインメモリＹ３０は、揮発性メモリでも不揮発性メモリでもよく、複数のチャネルに分かれていてもよい。二次メモリＹ４０は、たとえば、ハードディスクドライブＹ５０、着脱式記憶ドライブＹ６０、および／またはメモリスティックなどの不揮発性メモリを含んでもよい。着脱式記憶ドライブＹ６０としては、フロッピー（登録商標）ディスクドライブ、磁気テープドライブ、光ディスクドライブ、フラッシュメモリなどを挙げることができる。着脱式記憶ドライブＹ６０は周知の方法で着脱式記憶装置４７０からの読込みおよび／または着脱式記憶装置４７０への書込みを行う。着脱式記憶装置Ｙ７０としては、着脱式記憶ドライブＹ６０が読込みと書込みを行うフロッピーディスク（登録商標）、磁気テープ、光ディスクなどを挙げることができる。関連技術分野の当業者には当然であるが、着脱式記憶装置Ｙ７０は、コンピュータソフトウェアおよび／またはデータを記憶したコンピュータ使用可能な記憶媒体を含む。

【０６３８】

代替的な態様では、二次メモリＹ４０は、コンピュータプログラムまたはその他の命令をコンピュータシステムＹ００にロードできるようにするその他の類似の手段を含んでもよい。このような手段は、たとえば着脱式記憶装置Ｙ７０とインタフェース（図示せず）を含んでもよい。このような手段の例としては、プログラムカートリッジとカートリッジインタフェース（ビデオゲーム装置に見られるものなど）、着脱式メモリチップ（ＥＰＲＯＭまたはＰＲＯＭなど）とそれに関連するソケット、ならびに、ソフトウェアとデータを着脱式記憶装置Ｙ７０からコンピュータシステムＹ００へ転送できるようにするその他の着脱式記憶装置Ｙ７０とインタフェースが挙げられる。

【０６３９】

コンピュータシステムＹ００はさらに、メモリ制御装置Ｙ７５を含んでもよい。メモリ制御装置Ｙ７５は、メインメモリＹ３０および二次メモリＹ４０へのデータアクセスを制御する。一部の態様では、図７０に示すようにメモリ制御装置Ｙ７５がプロセッサＹ１０の外部にあってもよい。他の態様では、メモリ制御装置Ｙ７５は直接的にプロセッサＹ１０の一部であってもよい。たとえば、多くのＡＭＤ（商標）プロセッサおよびＩｎｔｅｌ（商標）プロセッサは、プロセッサＹ１０（図７０には図示せず）と同一のチップの一部である一体型のメモリ制御装置を使用する。

【０６４０】

コンピュータシステムＹ００はさらに、通信・ネットワークインタフェースＹ８０を含んでもよい。通信・ネットワークインタフェースＹ８０によって、ソフトウェアおよびデータをコンピュータシステムＹ００と外部装置との間で転送できる。通信・ネットワークインタフェースＹ８０は、モデム、通信ポート、ＰＣＭＣＩＡスロットおよびカードなどを含んでもよい。通信・ネットワークインタフェースＹ８０を介して転送されるソフトウェアおよびデータは信号の形態であり、通信・ネットワークインタフェースＹ８０が受信できる電子信号、電磁気信号、光学信号または他の信号であってよい。これらの信号は、通信経路Ｙ８５を介して通信・ネットワークインタフェースＹ８０へ供給される。通信経路Ｙ８５はワイヤまたはケーブル、光ファイバ、電話回線、携帯電話リンク、ＲＦリンク、またはその他の通信チャネルを用いて信号を伝達する。

【０６４１】

通信・ネットワークインタフェースＹ８０によって、コンピュータシステムＹ００はＬＡＮ、ＷＡＮ、インターネットなどの通信ネットワークまたは媒体を通じて通信できる。通信・ネットワークインタフェースＹ８０は、有線接続または無線接続によって遠隔のサイトまたはネットワークと相互作用できる。

【０６４２】

本明細書では、「コンピュータプログラム媒体」、「コンピュータ使用可能媒体」およ

び「非一過性媒体」という用語は、一般的に、着脱式記憶装置Ｙ７０、着脱式記憶ドライブＹ６０、ハードディスクドライブＹ５０にインストールされたハードディスクなどの実体のある媒体を指す。通信経路Ｙ８５を通じて伝達される信号も本明細書に記載の論理を具体化できる。コンピュータプログラム媒体およびコンピュータ使用可能媒体は、メインメモリＹ３０および二次メモリＹ４０などのメモリ（これらはメモリ半導体（たとえばＤＲＡＭなど）であってもよい）を指すこともできる。これらのコンピュータプログラム製品は、コンピュータシステムＹ００にソフトウェアを提供するための手段である。

【０６４３】

コンピュータプログラム（コンピュータ制御論理ともいう）は、メインメモリＹ３０および／または二次メモリＹ４０に記憶されている。コンピュータプログラムは、通信・ネットワークインタフェースＹ８０を介して受信されてもよい。このようなコンピュータプログラムを実行すると、コンピュータシステムＹ００は本明細書に記載の態様を実行できる。特に、コンピュータプログラムを実行すると、プロセッサＹ１０は本明細書に開示した処理を実行できる。したがって、このようなコンピュータプログラムはコンピュータシステムＹ００の制御装置になる。ソフトウェアを使用して態様を実施する場合、ソフトウェアはコンピュータプログラム製品に保存されていてもよく、たとえば、着脱式記憶ドライブＹ６０、インタフェース、ハードドライブＹ５０または通信・ネットワークインタフェースＹ８０を用いてコンピュータシステムＹ００にロードされる。

【０６４４】

コンピュータシステムＹ００はさらに、キーボード、モニタ、ポインティング装置、タッチスクリーンなどの入力／出力／表示装置Ｙ９０を含んでもよい。

【０６４５】

当然ながら、様々な態様のシミュレーション、合成および／または製造の一部を、一般的なプログラミング言語（ＣまたはＣ＋＋など）、ハードウェア記述言語（hardware description language; HDL）（たとえばVerilog HDL、VHDL、Altera HDL（AHDL））、または他の利用可能なプログラミングツールなどのコンピュータ可読コードを用いて実行してもよい。このコンピュータ可読コードは、半導体、磁気ディスク、光ディスク（ＣＤ－ＲＯＭ、ＤＶＤ－ＲＯＭなど）などの任意の既知のコンピュータ使用可能媒体に配置できる。このように、インターネットなどの通信ネットワークを介してコードを送信することができる。

【０６４６】

これらの態様は、任意のコンピュータ使用可能媒体に記憶されたソフトウェアを含むコンピュータプログラム製品にも関する。このようなソフトウェアを１つ以上のデータ処理装置で実行すると、データ処理装置は本明細書に記載したように動作する。態様によっては、現在公知であるか将来公知になる、任意のコンピュータ使用可能媒体またはコンピュータ可読媒体、ならびにコンピュータ使用可能記憶媒体またはコンピュータ可読記憶媒体を使用する。コンピュータ使用可能媒体またはコンピュータ可読媒体の例としては、一次記憶装置（たとえば、任意の種類のランダムアクセスメモリ）、二次記憶装置（たとえば、ハードドライブ、フロッピーディスク（登録商標）、ＣＤ－ＲＯＭ、ＺＩＰディスク、テープ、磁気記憶装置、光学記憶装置、ＭＥＭＳ、ナノ技術記憶装置など）、および通信媒体（たとえば、有線および無線通信ネットワーク、ローカルエリアネットワーク、ワイドエリアネットワーク、イントラネットなど）が挙げられるが、これらに限定されるものではない。コンピュータ使用可能媒体またはコンピュータ可読媒体としては、任意の形態の一過性の（信号を含む）媒体または非一過性の（信号を含まない）媒体を挙げることができる。非一過性媒体としては、上記の物理的な記憶装置（たとえば一次記憶装置および二次記憶装置）が挙げられるが、これらに限定されるものではない。

【０６４７】

当然ながら、本明細書に開示された態様はいずれも本明細書に開示された他の任意の態様と組み合わせて使用できる。

【実施例】

【０６４８】
本明細書に開示の態様を以下の実施例で説明する。これらの実施例は本開示の理解を助けるために記載するものであり、当然ながら、後に記載する特許請求の範囲で定義する開示の範囲をいかなる形でも限定するものではない。以下の実施例は、記載した態様の実施方法を当業者に完全に開示し具体化するために記載するものであり、本開示の範囲を限定することを意図するものではなく、以下の実験のみが実施したすべての実験であることを表すことを意図するものでもない。使用する数値（たとえば量や温度など）については正確にするよう努めたが、実験誤差および偏差が考慮されるべきである。別段の指示がない限り、「部」は容量部であり、温度は摂氏温度である。当然ながら、実験が示すよう意図された基本的側面を変えることなく、記載した方法を変更してもよい。

【０６４９】
実施例１
試料の調製および増幅の方法の例は、２０１２年１１月２１日出願の米国特許出願第１３／６８３，６０４号（米国特許公開第２０１３／０１２３１２０号）明細書、２０１４年５月１６日出願の米国特許出願第６１／９９４，７９１号明細書に記載されている（これらの文献の全内容は参照によって本明細書に組み込まれる）。これらの方法を用いて、本明細書に開示された任意の試料を分析できる。

【０６５０】
一実験では、血漿試料を調製し、半入れ子・１９，４８８プレックスのプロトコルで増幅した。試料を以下の方法で調製した：最大２０ｍＬの血液を遠心分離してバフィーコート（軟膜）と血漿を単離した。このバフィーコートから血液試料のゲノムＤＮＡを調製した。ゲノムＤＮＡは、唾液試料からも調製できる。血漿中の無細胞ＤＮＡをQIAGEN社のCIRCULATING Nucleic Acid kitを使用して単離し、製造者の指示に従いＴＥ緩衝液５０μＬで溶離した。精製した血漿ＤＮＡ４０μＬの各分子の末端に汎用的なライゲーションアダプターを付加し、アダプターに特異的なプライマーを用いてライブラリを９サイクルにわたって増幅した。AGENCOURT（登録商標）AMPUREビーズを用いてライブラリを精製し、５０μＬのＤＮＡ懸濁用緩衝液で溶離した。

【０６５１】
タグ付きの標的特異的な逆方向プライマー１９，４８８個（プライマー濃度７．５ｎＭ）とアダプター特異的な順方向プライマーの１つのライブラリ（濃度５００ｎＭ）とを用いて、６μＬのＤＮＡを１５サイクルのＳＴＡＲ１で増幅した（最初にポリメラーゼ活性化のために９５℃で１０分間；次いで９６℃で３０秒間、６５℃で１分間、５８℃で６分間、６０℃で８分間、６５℃で４分間および７２℃で３０秒間、を１５サイクル；最後の伸長を７２℃で２分間）。

【０６５２】
半入れ子ＰＣＲプロトコルでは、逆方向タグ（濃度１０００ｎＭ）と、１９，４８８個の標的特異的な順方向プライマーの各々（濃度２０ｎＭ）を用いて、ＳＴＡＲ１産物の希釈物を１５サイクルにわたって増幅する２回目の増幅（ＳＴＡＲ２）を行った（最初にポリメラーゼ活性化のために９５℃で１０分間；次いで９５℃で３０秒間、６５℃で１分間、６０℃で５分間、６５℃で５分間、７２℃で３０秒間、を１５サイクル；最後の伸長を７２℃で２分間）。

【０６５３】
次に、タグ特異的順方向プライマーとバーコード付き逆方向プライマー（１μＭ）を用いて、一定分量のＳＴＡＲ２産物を標準のＰＣＲで１２サイクルにわたって増幅し、バーコード付き配列のライブラリを作製した。各ライブラリの一定分量を、バーコードの異なる配列ライブラリと混合し、スピンカラムで精製した。

【０６５４】
このように、単一ウェルでの反応に１９，４８８個のプライマーを用いた。プライマーは、１番染色体、２番染色体、１３番染色体、１８番染色体、２１番染色体、Ｘ染色体、およびＹ染色体に見られるＳＮＰを標的とするように設計した。次いで増幅産物をILLUMI

NA（登録商標）GAIIXシーケンサで配列決定した。必要に応じ、配列決定リード数を増やし、増幅し配列決定する標的ＳＮＰの数を増やすことができる。

【０６５５】

関連するゲノムＤＮＡ試料を、準入れ子にした外側順方向プライマーとタグ付き逆方向プライマー（７．５ｎＭ）１９，４８８対を用いて、増幅した（ＳＴＡＲ１）。ＳＴＡＲ２の熱条件および構成とバーコーディングＰＣＲについては、半入れ子のプロトコルと同様とした。

【０６５６】

実施例２

プライマー選択方法の例は、２０１２年１１月２１日出願の米国特許出願第１３／６８３，６０４号（米国特許公開第２０１３／０１２３１２０号）明細書、および２０１４年５月１６日出願の米国特許出願第６１／９９４，７９１号明細書に記載されている（これらの全内容は参照によって本明細書に組み込まれる）。これらの方法を用いて、本明細書に開示された任意の試料を分析できる。

【０６５７】

以下の実験では、本発明の任意の多重ＰＣＲ法に使用できるプライマーライブラリの設計法および選択方法の例を説明する。目的は、候補プライマーの初期ライブラリから、多数の標的座位（または標的座位の部分集合）を１つの反応容量で同時に増幅するのに使用できるプライマーを選択することである。候補標的座位の初期の集合について、プライマーを各標的座位に対して設計または選択する必要はなく、最も望ましい標的座位の大部分についてプライマーを設計・選択することが好ましい。

【０６５８】

工程１

候補標的座位（ＳＮＰなど）の集合を、標的に関する所望のパラメータについての公的に利用可能な情報（たとえば、対象集団におけるＳＮＰの頻度、ＳＮＰのヘテロ接合体率など）（ワールドワイドウェブのncbi.nlm.nih.gov/projects/SNP/；およびSherry ST, Ward MH, Kholodov M, et al. dbSNP: the NCBI database of genetic variation. Nucleic Acids Res. 2001 Jan 1;29(1):308-11（それぞれの全内容は参照によって本明細書に組み込まれる）に基づいて選択した。各候補座位について、１つ以上のＰＣＲプライマー対をPrimer 3 programワールドワイドウェブのprimer3.sourceforge.net; libprimer3 release 2.2.3（全内容は参照によって本明細書に組み込まれる））で設計した。特定の標的座位に対するＰＣＲプライマーの実現可能な設計がない場合、その標的座位を除外して検討した。

【０６５９】

必要に応じ、標的座位の大部分またはすべてについて「標的座位スコア（スコアが高いほど適切度が高い）を計算することができる。（たとえば、その標的座位に求める多様なパラメータの加重平均に基づいて計算した標的座位スコア）。パラメータには、プライマーを使用する特定の用途にとっての重要性に基づいて異なる重みを割り当ててよい。パラメータの例としては、標的座位のヘテロ接合体率、標的座位での配列（たとえば多型）に関連する疾患罹患率、標的座位での配列（たとえば多型）に関連する疾患浸透度、標的座位の増幅に用いる候補プライマーの特異性、標的座位の増幅に用いる候補プライマーのサイズ、標的増幅産物のサイズが挙げられる。一部の態様では、標的座位に対する候補プライマーの特異性は、候補プライマーの設計上の増幅対象である標的座位以外の座位に対して結合・増幅することによってミスプライムする尤度を含む。一部の態様では、ミスプライムする１個以上またはすべての候補プライマーをライブラリから除外する。

【０６６０】

工程２

工程１で得た他のすべての標的座位について、標的座位に対する各プライマーとすべてのプライマーでの熱力学的相互作用スコアを計算した（たとえば、以下を参照：Allawi, H. T. & SantaLucia, J., Jr. (1998), "Thermodynamics of Internal C-T Mismatches i

n DNA", Nucleic Acids Res. 26, 2694-2701; Peyret, N., Seneviratne, P. A., Allawi, H. T. & SantaLucia, J., Jr. (1999), "Nearest-Neighbor Thermodynamics and NMR of DNA Sequences with Internal A-A, C-C, G-G, and T-T Mismatches", Biochemistry 38, 3468-3477; Allawi, H. T. & SantaLucia, J., Jr. (1998), "Nearest-Neighbor Thermodynamics of Internal A-C Mismatches in DNA: Sequence Dependence and pH Effects", Biochemistry 37, 9435-9444.; Allawi, H. T. & SantaLucia, J., Jr. (1998), "Nearest Neighbor Thermodynamic Parameters for Internal G-A Mismatches in DNA", Biochemistry 37, 2170-2179; and Allawi, H. T. & SantaLucia, J., Jr. (1997), "Thermodynamics and NMR of Internal G-T Mismatches in DNA", Biochemistry 36, 10581-10594; MultiPLX 2.1 (Kaplinski L, Andreson R, Puurand T, Remm M. MultiPLX: automatic grouping and evaluation of PCR primers. Bioinformatics. 2005 Apr 15;21(8):1701-2（これらの文献それぞれの全内容は参照によって本明細書に組み込まれる））。この工程の結果、相互作用スコアの２次元の行列を得た。相互作用スコアにより２個の相互作用するプライマーを含むプライマー二量体の尤度を予測した。相互作用スコアは以下のように計算した：

相互作用スコア＝最大（－ΔＧ２，０．８×（－ΔＧ１））

（式中、

「ΔＧ２」は、ＰＣＲによって両端が伸長可能な（すなわち、各プライマーの３’末端はもう一方のプライマーにアニーリングする）二量体に関するギブズのエネルギー（二量体を切断するのに必要なエネルギー）であり、

「ΔＧ１」は、少なくとも一方の末端がＰＣＲで伸長可能な二量体に関するギブズのエネルギーである。）

【０６６１】

工程３

各標的座位について、２つ以上のプライマー対設計がある場合、１つの設計を以下の方法で選択した：

１．座位に対する各プライマー対設計について、その設計に含まれる２個のプライマーと、他のすべての標的座位に対するすべての設計に含まれるすべてのプライマーとの最悪の場合の（最も高い）相互作用スコアを見つける。

２．最悪の場合の相互作用スコアが最も良い（最も低い）設計を選択する。

【０６６２】

工程４

各節点が１つの座位とそれに関連するプライマー対設計を表すようなグラフを描いた（たとえば、最大クリーク問題）。各２つの節点の間に枝を設けた。各枝に、その枝が結ぶ２つの節点に関連するプライマーの間の最悪の場合の（最も高い）相互作用スコアと同等の重みを割り当てた。

【０６６３】

工程５

必要に応じ、２個の異なる標的座位に関する各対の設計について、一方の設計のプライマーの１つともう一方の設計のプライマーの１つが重複する標的領域にアニールする場合、その２つの設計に関する節点の間に枝を追加する。これらの枝の重みは、工程４で割り当てた重みの最大値と等しく設定した。したがって、工程５は、多重ＰＣＲ反応時に重複する標的領域にアニールして互いに干渉するプライマーがライブラリに含まれるのを防止する。

【０６６４】

工程６

初期の相互作用スコアの閾値を以下のように計算した：

重み閾値＝最大（枝重み）－０．０５×（最大（枝重み）－最小（枝重み））

（式中、「最大（枝重み）」はグラフ中の枝の重みの最大値であり；

「最小（枝重み）」はグラフ中の枝の重みの最小値である。

初期の閾値範囲を以下のように設定した：
最大重み閾値＝最大（枝重み）
最小重み閾値＝最小（枝重み）
【０６６５】
工程７
工程５と同じ節点の集合からなり、重み閾値を越えた重み付きの枝のみを含む新しいグラフを作成した。このように、この工程では、スコアが重み閾値以下の相互作用は無視する。
【０６６６】
工程８
枝が残らなくなるまで節点（と、削除された節点を結ぶすべての枝）を工程７のグラフから削除した。以下の手順を繰り返すことで節点を削除した。
１．次数が最も高い（枝の数が最も多い）節点を探す。複数ある場合は任意に１つを選ぶ。
２．上で選んだ節点とその節点に接続しているすべての節点からなる節点の集合を定義する。ただし、上で選んだ節点よりも次数が低い節点は排除する。
３．その集合から、工程１で得た標的座位スコア（スコアが低いほど適切度が低い）が最も低い節点を選び、グラフから削除する。
【０６６７】
工程９
グラフに残る節点の数が、多重ＰＣＲプールに必要な標的座位の数を（許容される公差の範囲内で）満たしていれば、工程１０で方法を継続する。
【０６６８】
グラフに残る節点の数が多すぎるか少なすぎる場合、二分探索を実行し、どのような閾値であれば所望の数の節点が残るかを決める。グラフの節点が多すぎる場合、重みの閾値の値の範囲を以下のように調節する：
最大重み閾値＝重み閾値
他の場合（グラフの節点が少なすぎる場合）には、重みの閾値を以下のように調節する。
最小重み閾値＝重み閾値
次に、重みの閾値を以下のように調節する：
重み閾値＝（最大重み閾値＋最小重み閾値）／２
工程７～９を繰り返す。
【０６６９】
工程１０
グラフに残っている節点に関連するプライマー対設計をプライマーライブラリに選択した。このプライマーライブラリは、本発明の任意の方法に使用できる。
【０６７０】
必要に応じて、（プライマー対でなく）１個のプライマーのみを標的座位の増幅に用いるプライマーライブラリを得るためにこのプライマー設計・選択方法を実行できる。この場合、節点は標的座位ごとに（プライマー対ではなく）１個のプライマーを表す。
【０６７１】
実施例３
必要に応じ、本発明の方法を試験して染色体または染色体分節の欠失または重複の検出能力を評価することができる。以下の実験を実施し、父親から遺伝したＸ染色体またはＸ染色体の分節の過剰出現の検出を、母親から遺伝したＸ染色体またはＸ染色体分節の場合と比較して実証する。この試験法は、染色体または染色体分節の欠失または重複を模倣するよう設計される。異なる量の父親由来ＤＮＡ（ＸＹ性染色体を有する）をその父親の娘由来のＤＮＡ（ＸＸ性染色体を有する）と混合し、父親由来のＸ染色体の余剰量を分析した（図１９Ａ～図１９Ｄ）。
【０６７２】

父親と娘の細胞系からＤＮＡを抽出し、Qubit（商標）を用いて定量した。父親の細胞系ＡＧ１６７８２、ｃＡＧ１６７８２－２－Ｆおよび娘の細胞系ＡＧ１６７７７、ｃＡＧ１６７７７－２－Ｐを使用した。父親のＸ染色体に関するハプロタイプを決定するため、Ｘ染色体には存在するがＹ染色体には存在しないＳＮＰを検出した。つまり、そのＳＮＰには父親のＸ染色体からのシグナルは存在するがＹ染色体からのシグナルは存在しないと考えられる。娘にはこのハプロタイプが父親から遺伝した。娘のもう一方のＸ染色体のハプロタイプは母親から遺伝したものである。母親のハプロタイプは、娘の細胞系のＤＮＡで父親から遺伝したのではないＳＮＰを母親のハプロタイプに割り当てることによって決定できる。

【０６７３】

父親由来のＸ染色体の過剰発現が検出される可能性があるかを決定するために、異なる量の父親細胞系由来のＤＮＡを、娘の細胞系由来のＤＮＡと混合した。全体のＤＮＡ使用量は、約７５ｎｇ（約２５ｋコピー）のゲノムＤＮＡであった。約３，４５６個のＳＮＰを、Ｘ染色体およびＹ染色体の試験のために直接多重ＰＣＲで増幅した。増幅産物を、７ｂｐのバーコードを用いて、Ｒａｐｉｄ／ＨＴモードで５０ｂｐシングルリード配列決定によって配列決定した。リード数はＳＮＰあたり約１０Ｋであった。

【０６７４】

図１９Ａ～図１９Ｄに示すように、父親のＤＮＡからのモザイク現象が検出された。図から、過剰出現した染色体分節または染色体全体が検出される可能性があることがわかる。

【０６７５】

本明細書で引用した全ての特許、特許出願、および公開されたその引用文献の全内容は参照によって本明細書に組み込まれる。本開示の方法は、その特定の態様に関連して記載されているが、当然ながら、修正を加えることができる。さらに、本願は、本開示の方法の任意の変更、使用、または適応を包含し、本開示を発展させたこれらの内容についても、本開示の方法が属する技術分野で公知または慣習的な実践の範囲内であり、添付の特許請求項の範囲にも包含することを意図する。本発明のいずれの態様も、試料中のＤＮＡおよび／またはＲＮＡの分析によって実施できる。たとえば、ＤＮＡに関して本明細書に開示した方法のいずれも、たとえば逆転写でＲＮＡをＤＮＡに変換する工程を含むことによって、ＲＮＡに適応させることができる。

【０６７６】

実施例４

本実施例では、乳癌関連のコピー数多様性を非侵襲的に無細胞腫瘍ＤＮＡに基づいて検出するための方法の例について説明する。乳癌検診には乳房Ｘ線撮影が含まれるが、偽陽性率が高く、一部の癌は見落とされる。腫瘍由来の循環無細胞ＤＮＡ（無細胞循環腫瘍ＤＮＡ）の癌関連のＣＮＶを分析することによって、より迅速、安全、かつ高精度の検診が可能である。ＳＮＰに基づく超多重ＰＣＲ（massively multiplex PCR）法を用いて、乳癌患者の血漿から単離した無細胞循環腫瘍ＤＮＡをＣＮＶについてスクリーニングした。癌でＣＮＶが見られることの多い１番染色体、２番染色体、２２番染色体（たとえば、乳癌の試料の４９％は２２ｑの欠失を有する）で３，１６８個のＳＮＰ超多重ＰＣＲ試験法を標的化するように超多重ＰＣＲ試験法を設計した。乳癌患者（ＩＩａ期１名、ＩＩｂ期４名、ＩＩＩｂ期１名）から採取した６個の血漿試料を分析した。各試料は、標的染色体のうち１つ以上にＣＮＶを有していた。この試験法では、０．５８％の無細胞循環腫瘍ＤＮＡ率で正しく検出されたＩＩｂ期の試料１個を含む６個の血漿試料すべてにＣＮＶが確認された（図３０、図３１Ｂ、図３２Ａ、図３２Ｂ、図３３）。検出に必要なのは８６個のヘテロ接合性のＳＮＰのみであった。ＩＩａ期の試料も４．３３％の無細胞循環腫瘍ＤＮＡ率で、６３６個のヘテロ接合性のＳＮＰを用いて正しく検出された（図２９、図３１Ａ、および図３２Ａ）。このことから、染色体アームの局所的なＣＮＶまたは全体のＣＮＶ（どちらも癌では一般的である）を容易に検出できることがわかる。

【０６７７】

さらに感度を評価するため、３Ｍｂの癌細胞系由来の２２ｑＣＮＶを有する２２種類の人工混合物を、正常細胞系由来のＤＮＡと混合し（５：９５）、０．４３％～７．３５％の無細胞循環腫瘍ＤＮＡ率を再現した（図２８Ａ～２８Ｃ）。この方法では、これらの試料の１００％でＣＮＶが正確に検出された。このように、ＣＮＶを示すことがわかっている非無細胞ＤＮＡ源（腫瘍細胞系など）によって生成された断片化ポリヌクレオチド混合物を含む単離ポリヌクレオチド試料を、他のＤＮＡ試料に、生体内の無細胞ＤＮＡにみられる濃度と類似の濃度（たとえば、ＤＮＡが液体中０．０１％～２０％、０．１％～１５％、または４％～１０％）で添加することにより、人工無細胞ＤＮＡポリヌクレオチド標準／対照を作製できる。これらの標準／対照は、試験法の設計、特性評価、開発、および／または検証のための対照として、また、試験（ＣＬＩＡ（Clinical Laboratory Improvement Amendment；臨床検査改善修正法案）認定された研究所で実施される癌試験など）の際の品質管理標準として、および／あるいは研究用試験キットまたは診断試験キットに含まれる標準として使用可能である。重要なことであるが、乳癌と卵巣癌を含む多数の癌では点変異よりもＣＮＶがより高頻度に見られる。総合すると、この結果により、このＳＮＰに基づく超多重ＰＣＲ法が、これらの癌を非侵襲的に検出するための、よりコスト効率の高い方法を提供することが裏付けられる。

【０６７８】

実施例５

本実施例では、ＳＮＰを標的とする超多重ＰＣＲを使用して乳癌試料でコピー数変異を検出するための方法の例について説明する。腫瘍組織でのＣＮＶの評価は、一般的にＳＮＰマイクロアレイまたはａＣＧＨを含む。これらの方法は全ゲノム分解能が高いが、大量の使用材料が必要であり、固定費が高く、ホルムアルデヒド固定パラフィン包埋（ＦＦＰＥ）試料ではうまく機能しない。本実施例では次世代配列決定（next generation sequencing; NGS）を用いる２８，０００プレックスＳＮＰ標的化ＰＣＲを使用し、１ｐ、１ｑ、２ｐ、２ｑ、４ｐ１６、５ｐ１５、７ｑ１１、１５ｑ、１７ｐ、２２ｑ１１、２２ｑ１３および１３番染色体、１８番染色体、２１番染色体を乳癌試料でＣＮＶを検出するための標的とした。異数性または微小欠失を有する９６個の試料について精度を検証した。単一分子感度は単一細胞を分析することによって確立した。１７個の乳癌試料（新鮮凍結腫瘍組織が１５個、ＦＦＰＥ腫瘍組織が２個、腫瘍細胞系と正常細胞系の対応対が５対）を分析し、１６個（両方のＦＦＰＥ試料を含む）は、１個～１５個（平均：７．８個）の標的に完全または部分的なＣＮＶを有することが認められた。腫瘍がヘテロ接合性である兆候を認めた。１個のＣＮＶを有する３個の組織はすべて、乳癌で最も頻度の高い細胞遺伝学的異常である１ｑの重複を有していた。ＣＮＶの頻度が最も高い領域は１ｑ、７ｐ、および２２ｑ１であった。１つの腫瘍組織（９個のＣＮＶを有する）のみが、ＬＯＨを含む領域を有していた。このＬＯＨは、他の８個のＣＮＶを有さない隣接する推定正常組織でも検出された。一方、ＬＯＨを有し全体のＣＮＶ発生率（平均：１２．８）が高い領域を細胞系で５つ以上検出した。したがって、超多重化ＰＣＲは標的化による方式でＣＮＶを調査するための経済的かつ高処理な手法を提供し、分析が困難なＦＦＰＥ組織などの試料に利用できる。

【０６７９】

実施例６

本実施例では、本発明のいずれかの方法の検出限界を計算するための方法の例について説明する。これらの方法を用いて、腫瘍生検（図３４）および血漿試料（図３５）における一塩基多様体（ＳＮＶ）の検出限界を計算した。

【０６８０】

第１の方法（図３４および図３５の「ＬＯＤ－ｍｒ５」）では、配列決定データに実際にＳＮＶが存在するという十分な確信を得るための最小のＳＮＶ観測回数として選択した最低５個のリードに基づいて検出限界を計算する。検出限界は、リード深度（ＤＯＲ）の観測値がこの最小値５を超えるかどうかに基づく。図３４および図３５の灰色の線は、ＤＯＲによって検出限界が制限されるＳＮＶを示す。これらの場合、測定したリードの数は

、試験法の誤差限界に達するのに十分ではなかった。必要に応じて、ＤＯＲを上げることにより、これらのＳＮＶの検出限界を改善できる（数値が低くなる）。

【０６８１】

第二の方法（図３４および図３５の「ＬＯＤ－ｚｓ５．０」）では、ｚスコアに基づいて検出限界を計算する。ｚスコアは、観測エラー率が背景平均エラーから離れている標準偏差の数である。必要に応じて、異常値を除外してｚスコアを再計算し、この処理を繰り返すことができる。ｚスコアの計算にはエラー率の最終加重平均と標準偏差を用いる。ＤＯＲが高いほど精度は高くなるため、平均値をＤＯＲで加重する。

【０６８２】

この実施例で用いたｚスコア計算例では、それぞれのゲノム座位と置換の種類について、同じ配列決定実行の他のすべての試料から、リード深度で加重した背景平均エラーおよび標準偏差を計算した。試料が背景平均から５標準偏差である場合、その試料は背景分布では考慮しなかった。図３４および図３５のオレンジ色の線は、検出限界がエラー率によって制限されるＳＮＶを示す。これらのＳＮＶについては、最小リード値５に達するのに十分なリードが得られ、検出限界はエラー率で制限された。必要に応じて、検出の限界は、試験法を最適化してエラー率を低減することによって改善できる。

【０６８３】

第三の方法（図３４および図３５の「ＬＯＤ－ｚｓ５．０－ｍｒ５」）では、上記の２つの計量の最大値に基づき、検出限界を算出する。

【０６８４】

図３４に示す腫瘍試料の分析の場合、平均検出限界は０．３６％であり、検出限界の中央値は０．２８％であった。ＤＯＲで制限されたＳＮＶ（灰色の線）の数は９３４であった。エラー率で制限されたＳＮＶ（オレンジ色の線）の数は７３８であった。

【０６８５】

図３５の血漿試料中のｃＤＮＡの分析の場合、平均検出限界は０．２４％であり、検出限界の中央値は０．０９％であった。ＤＯＲで制限されたＳＮＶ（灰色の線）の数は７３２であった。エラー率で制限されたＳＮＶ（オレンジ色の線）の数は９２１であった。

【０６８６】

実施例７

本実施例では、同一の単一細胞からのＣＮＶおよびＳＮＶの検出について説明する。以下のプライマーライブラリを用いた：ＣＮＶ検出用の約２８，０００個のプライマーのライブラリ；ＣＮＶの検出用の約３，０００個のプライマーのライブラリ；およびＳＮＶの検出用のプライマーのライブラリ。単一細胞の分析のため、細胞を一滴あたり３個または４個の細胞が存在するまで段階希釈した。個々の細胞をピペットで取り、ＰＣＲ管に入れた。以下の条件で、プロテアーゼＫ、塩、およびＤＴＴを用いて細胞を溶解した：５６℃で２０分間、９５℃で１０分間、次いで４℃で保持。ゲノムＤＮＡを分析するため、分析した単一細胞と同じ細胞系由来のＤＮＡを、購入するか、細胞を増殖させてＤＮＡを抽出することによって得た。

【０６８７】

約２８，０００個のプライマーのライブラリを用いた増幅に以下のＰＣＲ条件を用いた：反応容量４０μＬ、各プライマー７．５ｎＭ、２×マスターミックス（ＭＭ）。一部の態様では、QIAGEN Multiplex PCR Kitをマスターミックスに用いた（QIAGEN社のカタログ番号２０６１４３；たとえばワールドワイドウェブのqiagen.com/products/catalog/assay-technologies/ end-point-pcr-and-rt-pcr-reagents/qiagen-multiplex-pcr-kitを参照（この全内容は参照によって本明細書に組み込まれる）。キットには、２×QIAGEN（登録商標）Multiplex PCR Master Mix（終濃度3mM $MgCl_2$、３×０．８５ｍＬ）と、５×Q-Solution（１×２．０ｍＬ）と、ＲＮＡｓｅを含まない遊離水（２×１．７ｍＬ）とを含む。QIAGEN（登録商標）Multiplex PCR Master Mix（ＭＭ）は、ＫＣｌおよび$(NH_4)_2SO_4$と、ＰＣＲ添加剤と、Factor MP（鋳型での局所的プライマー濃度を上げる）の組み合わせを含む。Factor MPは特異的に結合したプライマーを安定化し、たとえばHotStar

Tag（登録商標）DNA Polymeraseによるプライマーの伸長効率を良くする。HotStarTaq（登録商標）DNA PolymeraseはTaq（登録商標）DNA Polymeraseの改変型であり、環境温度ではポリメラーゼ活性を有さない。以下の熱サイクル条件を用いて初回のＰＣＲを行った：９５℃で１０分間；９６℃で３０秒間、６５℃で２９分間、７２℃で３０秒間、を２５サイクル；次いで７２℃で２分間、４℃で保持。二回目のＰＣＲでは、反応容量を１０μＬとし、１×ＭＭと５ｎＭの各プライマーを用いた。以下の熱サイクル条件を用いた：９５℃で１５分間；９４℃で３０秒間、６５℃で１分間、６０℃で５分間、６５℃で５分間、７２℃で３０秒間、を２５サイクル；次いで７２℃で２分間、４℃で保持。

【０６８８】

約３，０００個のプライマーのライブラリに関して、例示の反応条件は１０μＬの反応容量（２×ＭＭ、70mM TMAC、各プライマー２ｎＭ）を含む。ＳＮＶを検出するためのプライマーのライブラリの場合、例示の反応条件は、１０μＬの反応容量（２×ＭＭ、4mM EDTA、各プライマー７．５ｎＭ）を含む。例示の熱サイクル条件としては以下が挙げられる：９５℃で１５分間；９４℃で３０秒間、６５℃で１５分間、７２℃で３０秒間、を２０サイクル；次いで７２℃で２分間、４℃で保持。

【０６８９】

増幅産物にバーコードを付加した。１回の配列決定は、試料あたり等しいリード数で行った。

【０６９０】

図３６Ａおよび図３６Ｂは、ＣＮＶを検出するよう設計した約２８，０００個のプライマーのライブラリを用いたゲノムＤＮＡ（図３６Ａ）または単一細胞由来のＤＮＡ（図３６Ｂ）の分析結果を示す。試料あたり約４００万リードを測定した。１つの中央バンドではなく２つの中央バンドが存在することから、ＣＮＶが存在することが示される。単一細胞に由来する３種類のＤＮＡ試料について、マッピングされたリードの割合はそれぞれ８９．９％、９４．０％、９３．４％であった。２種類のゲノムＤＮＡ試料について、マッピングされたリードの割合はそれぞれ９９．１％であった。

【０６９１】

図３７Ａおよび図３７Ｂは、ＣＮＶを検出するよう設計した約３，０００個のプライマーのライブラリを用いたゲノムＤＮＡ（図３７Ａ）または単一細胞由来のＤＮＡ（図３７Ｂ）の分析結果を示す。試料ごとに約１２０万リードを測定した。１つの中央バンドでなく２つの中央バンドが存在することから、ＣＮＶが存在することが示される。単一細胞に由来する３種類のＤＮＡ試料について、マッピングされたリードの割合はそれぞれ９８．２％、９８．２％、９７．９％であった。２種類のゲノムＤＮＡ試料について、マッピングされたリードの割合はそれぞれ９８．８％であった。図３８は、これらの約３，０００個の剤に関するＤＯＲの均一性を示す。

【０６９２】

ＳＮＶの検出について、真陽性の変異の検出率は単一細胞由来のＤＮＡとゲノムＤＮＡの場合で似ていた。縦軸を単一細胞に対する真陽性の変異の検出率、横軸をゲノムＤＮＡの真陽性の変異とするグラフでは、R^2＝０．９８３４でｙ＝１．００７６ｘ－０．３０８８の曲線にフィットした。図３９は、ゲノムＤＮＡと単一細胞由来のＤＮＡの場合では検出エラーの計測値が似ていることを示す。図４０は、転移変異の検出に関するエラー率が、塩基転換変異を検出する場合よりも大きかったことを示しており、可能であれば転移変異ではなく塩基転換変異を検出対象に選択することが望ましいとわかる。

【０６９３】

実施例８

本実施例では、本明細書に開示の染色体異数性およびＣＮＶ決定のための超多重化ＰＣＲ法（CoNVERGeという（Copy Number Variant Events Revealed Genotypically)をさらに検証し、無細胞腫瘍循環ＤＮＡ試料のＰＣＲに関するPlasmArt基準の開発と使用についてさらに説明する。PlasmArt基準は、ＣＮＶを示すことがわかっているゲノムの領域と同一の配列を有し、血漿中に天然に見られる無細胞ＤＮＡ断片のサイズ分布を反映するサイズ

分布を有する、ポリヌクレオチドを含む。

【０６９４】

試料の採取

ヒトの乳癌細胞株（ＨＣＣ３８、ＨＣＣ１１４３、ＨＣＣ１３９５、ＨＣＣ１９３７、ＨＣＣ１９５４およびＨＣＣ２２１８）と、対応する正常細胞株（ＨＣＣ３８ＢＬ、ＨＣＣ１１４３ＢＬ、ＨＣＣ１３９５ＢＬ、ＨＣＣ１９３７ＢＬ、ＨＣＣ１９５４ＢＬ、およびＨＣＣ２２１８ＢＬ）をAmerican Type Culture Collection（ATCC）から入手した。２１番染色体トリソミーのＢリンパ球細胞（ＡＧ１６７７７）と、父親と子の対になったディジョージ症候群（ＤＧＳ）細胞株（それぞれ、ＧＭ１０３８３およびＧＭ１０３８２）をCoriell Cell Repository（Camden, NJ）から入手した。ＧＭ１０３８２細胞は、父親の２２ｑ１１．２領域のみを有する。

【０６９５】

１６名の乳癌患者の腫瘍組織をGeneticist(Glendale, CA)から入手した１１個の新鮮凍結（ＦＦ）試料と、North Shore-LIJ（Manhasset、NY）から入手した５個のホルマリン固定パラフィン包埋（ＦＦＰＥ）試料）。患者８名の対応するバフィーコート試料と、患者９名の一致した血漿試料を得た。５名の卵巣癌患者のＦＦ腫瘍組織と、対応したバフィーコート試料と血漿試料をNorth Shore-LIJから入手した。８個の乳房腫瘍のＦＦ試料に関して、組織小片を分析用に切り取った。試料採取のためにNorthshore/LIJ IRBとKharkiv National Medical University Ethics Committeeの機関審査委員会の承認書を取得し、患者全員からインフォームドコンセントを得た。

【０６９６】

血液試料をＥＤＴＡ管に採取した。QIAamp（登録商標）Circulating Nucleic Acid Kit (Qiagen, Valencia, CA)を用いて血漿１ｍＬから循環腫瘍ＤＮＡを単離した。

【０６９７】

一典型的方法に従いPlasmAt標準を作製するため、まず９×１０^{6}個の細胞を低張溶解緩衝液（20mM Tris-Cl (pH7.5), 10mM NaCl, 3mM $MgCl_2$）を用いて氷上で１５分間かけて溶解した。次いで１０％のIGEPAL（登録商標）CA-630（Sigma, St. Louis, MO）を終濃度が０．５％になるまで添加した。４℃で３分間、３，０００ｇで遠心分離した後、ペレット化した核を1x micrococcal nuclease (MNase) Buffer（New England BioLabs, Ipswich, MA）に再懸濁した後、１０００ＵのMNase（New England BioLabs）を添加し、次いで３７℃で５分間インキュベートした。ＥＤＴＡを最終濃度が１５ｍＭになるように添加することによって反応を停止させた。未消化の染色質を、２，０００ｇで１分間の遠心分離によって除去した。断片化ＤＮＡをDNA Clean & Concentrator（登録商標）-500 kit（Zymo Research, Irvine, CA）で精製した。ＭＮａｓｅ消化によって得られたモノヌクレオソームＤＮＡも精製し、AMPure（登録商標）XP磁気ビーズ（Beckman Coulter, Brea, CA）を用いてサイズ選別した。Bioanalyzer DNA 1000 chip（Agilent, Santa Clara, CA）でＤＮＡ断片のサイズ分類と定量を行った。

【０６９８】

様々な濃度の無細胞循環腫瘍ＤＮＡをモデル化するために、ＨＣＣ１９５４およびＨＣＣ２２１８癌細胞由来のPlasmArtの異なる画分を、対応する正常細胞株（それぞれ、ＨＣＣ１９５４ＢＬおよびＨＣＣ２２１８ＢＬ）由来のものと混合した。各濃度３個の試料を分析した。同様に、血漿ＤＮＡの３．５Ｍｂの局所領域のアレル不均衡をモデル化するため、母親由来の２２ｑ１１．２欠失を有する子のＤＮＡと父親のＤＮＡとの比の異なるＤＮＡ混合物からPlasmArtを作製した。父親のＤＮＡのみを含む試料を陰性対照として使用した。各濃度８個の試料を分析した。

【０６９９】

したがって、CoNVERGeの感度と再現性を評価するため、特にＣＮＶに関する異常なＤＮＡの割合または平均アレル不均衡（average allelic imbalance; AAI）が低い場合、これを、過去に特性評価した異常な試料で構成されるＤＮＡ混合物（異常な試料を対応する正常試料に用量設定したもの）に含まれるＣＮＶを検出するために用いた。

【０７００】
混合物は、断片サイズ分布が天然の無細胞ＤＮＡ（上記を参照）に近い人工無細胞ＤＮＡ（「PlasmArt」という）で構成した。図４２は、染色体アーム１ｐ、１ｑ、２ｐ、および２ｑのＣＮＶに着目し、染色体のアーム癌細胞株由来の例示のPlasmArtのサイズ分布と無細胞ＤＮＡのサイズ分布の比較を図示する。第一のペアでは、３Ｍｂの局所的なＣＮＶ（２２ｑ１１．２領域の欠失）を有する息子の腫瘍ＤＮＡ試料を、父親由来の対応する正常な試料に、全体の無細胞ＤＮＡの割合が０％～１．５％となるように用量設定して添加した（図４１ａ）。CoNVERGeは、推定ＡＡＩが＞０．３５％（ＡＡＩが≧０．５％＋／－０．２％の混合物中）の既知の異常に対応するＣＮＶを確認し再現性を示したが、異常ＤＮＡが０．２５％の場合、８個のうち６個の複製物ではＣＮＶを検出できず、８個の陰性対照試料すべてについて≦０．０５％という値を報告した。CoNVERGeによるＡＡＩ推定値は、高い直線性（Ｒ２＝０．９４０）と再現性（誤差分散＝０．０８７）を示した。試験法は、同一試料で異なる増幅レベルに対して高感度であった。これらのデータに基づき、ＡＡＩ０．４５％の保守的な検出閾値をその後の分析に用いることができた。このカットオフ値（閾値）を用いて、PlasmArt合成無細胞循環腫瘍ＤＮＡを既知の濃度で添加して約０．５％～約３．５％の合成癌血漿を作製する、別の実験を行った。陰性血漿も対照として用いた。すべての合成癌血漿で０．４５％超の推定値が得られ、陰性血漿に関する測定値は０．４５％を大きく下回った（図４３Ａ～Ｃ）。図４３Ａの右図は腫瘍の最尤度であり、ＤＮＡ率の推定値の結果をオッズ比のグラフとして示す。図４３Ｂは塩基転換事象の検出に関するグラフである。図４３Ｃは転移事象の検出に関するグラフである。

【０７０１】
さらに、対応する腫瘍細胞株と正常細胞株のペアから調製した。それぞれ１番染色体または２番染色体にＣＮＶを有する２種類の既知濃度のPlasmArtを評価した（図４１ｂ、図４１ｃ）。陰性対照では、すべての値が＜０．４５％であり、既知のＤＮＡ使用量とCoNVERGeによる算出値との間には高い直線性（Ｒ２＝０．９５２（ＨＣＣ１９５４株、１ｐ）、Ｒ２＝０．９９３（ＨＣＣ１９５４株、１ｑ）、Ｒ２＝０．９７７（ＨＣＣ２２１８株、２ｐ）、Ｒ２＝０．９６７（ＨＣＣ２２１８株、２ｑ）および再現性（誤差分散＝０．１９０（ＨＣＣ１９５４株、１ｐ）、０．０２９（ＨＣＣ１９５４株、１ｑ）、０．２５０（ＨＣＣ２２１８株、２ｐ）、０．３５０（ＨＣＣ２２１８株、２ｑ）が認められた。１対の試料の領域１ｐおよび１ｑに関する回帰の傾きの差は、同じ試料の領域１ｐおよび１ｑのＢアレル頻度（ＢＡＦ）で観察されたコピー数の相対差と相関する。このことは、CoNVERGeによって計算されたＡＡＩ推定値の相対的な正確性を示す（図４１ｃ、４１ｄ）。

【０７０２】
試料を処理するためのワークフローを図６３に示す。CoNvergeは、ＦＦＰＥ試料、新鮮凍結試料、単一細胞、生殖細胞系対照、無細胞循環腫瘍ＤＮＡなどの多様な試料源に応用できる。６個のヒト乳癌細胞株および対応する正常細胞株にCoNVERGeを応用し、体細胞ＣＮＶを検出できるかどうかを評価した。アームレベルでのＣＮＶと局所的なＣＮＶは６個の癌細胞株すべてに存在したが、それらに対応する正常細胞株には存在せず、例外としてＨＣＣ１１４３株の２番染色体では正常細胞株が１：１の相同体比から偏差が見られる（図６３ｂ）。これらの結果を別のプラットフォームで検証するため、CytoSNP-12マイクロアレイ分析を行い、すべての試料について一貫した結果を得た（図６３ｄ、図６３ｅ）。さらに、CoNVERGeおよびCytoSNP-12マイクロアレイで確認したＣＮＶに関する相同体比の最大値は強い直線的相関を示した（Ｒ２＝０．９８７、Ｐ＜０．００１）（図６３ｆ）。

【０７０３】
次に、CoNVERGeを新鮮凍結（ＦＦ）（図６４ａ）およびホルマリン固定パラフィン包埋（ＦＦＰＥ）乳房腫瘍組織サンプル（図６４ｂ、６４ｄ）に応用した。いずれの種類の試料でも、数個のアームレベルのＣＮＶおよび局所的なＣＮＶが存在したが、対応するバフィーコート試料のＤＮＡではＣＮＶは検出されなかった。CoNVERGeの結果は、同じ試料のマイクロアレイ分析結果と高度に相関していた（図６４ｅ～ｈ；Ｒ２＝０．９０９（ＦＦ

）；Ｐ＜０．００１（CytoSNP-12）；Ｒ２＝０．９９２（ＦＦＰＥ）；Ｐ＜０．００１（OncoScan））。CoNVERGeは、マイクロアレイ法が適切でないレーザーキャプチャ顕微解剖（laser capture microdissection; LCM）試料から抽出した少量のＤＮＡについても一貫した結果を示した。

【０７０４】

CoNVERGeを用いた単一細胞でのＣＮＶ検出

この超多重化ＰＣＲ法の応用性の限界を試験するために、上記６個の癌細胞株と、標的領域にＣＮＶを有さないＢリンパ球細胞株から、単一細胞を単離した。これらの単一細胞実験から得たＣＮＶプロファイルは３個の複製物で一貫しており、かつ細胞約２０，０００個の大容量の試料から抽出したゲノムＤＮＡ（ｇＤＮＡ）から得たＣＮＶプロファイルとも一致していた（図６５）。配列決定リードのないＳＮＰの数を基準に、大容量試料での平均試験ドロップアウト率は０．４８％（範囲：０．４１％～０．６０％）であった。これは、合成上または試験法設計上の不具合による。単一細胞の場合、追加の平均試験ドロップアウト率の観測値は０．３９％（範囲：０．１９％～０．６７％）であった。失敗のなかった（すなわち試験ドロップアウトが起こらなかった）単一細胞試験法の場合、ヘテロ接合性のＳＮＰのみを用いて計算した平均単一試験ドロップアウト率は、わずか０．０５％（範囲：０．００～０．４３％）であった。さらに、信頼度の高い遺伝型（すなわち、少なくとも９８％の信頼度で決定されたＳＮＰ遺伝型）を有するＳＮＰの割合は、単一細胞試料および大容量試料の両方で類似し、単一細胞試料の遺伝型は大容量試料と一致した（平均９９．５２％、範囲：９２．６３％～１００．００％）。

【０７０５】

腫瘍細胞の試料とは異なり、単一細胞では、アレル頻度は染色体コピー数を直接反映すると期待されるが、ＴＨおよび腫瘍細胞でない不純物によって誤解を招く可能性がある。Ｂアレル頻度１／ｎおよび（ｎ－１）／ｎは、ｎ個の染色体コピーが領域に存在することを示す。単一細胞試料とゲノムＤＮＡ試料について、染色体コピー数をアレル頻度のグラフ上に示す（図６５）。

【０７０６】

血液試料へのCoNVERGeの応用

実際の血漿試料でのCoNVERGeのＣＮＶ検出能力を調べるため、この手法を、２名のステージＩＩの乳癌患者と５名の後期卵巣癌患者のそれぞれから採取した腫瘍生検と対応対にした無細胞ＤＮＡに応用した。７名の患者全員で、ＦＦ腫瘍組織にも対応する血漿試料にもＣＮＶが検出された（図６６）。図６７はＳＮＶ乳癌変異のリストである。試験した５つの領域にわたり、合計で３２個のＣＮＶ（ＡＡＩレベル≧０．４５％）が７個の血漿試料で検出された（範囲：０．４８％～１２．９９％ＡＡＩ）。これはゲノムの約２０％にあたる。当然ながら、他の直交方法が存在しないため、血漿中のＣＮＶの存在は確認できない。

【０７０７】

ＡＡＩ推定値は腫瘍のＢアレル頻度と相関するように見えるが、腫瘍がヘテロ接合性であるため、直接の比例関係は期待されない。たとえば、試料ＢＣ５（図６６ａ）では、図６６ａの左上部の楕円は、Ｎ＝１１と両立するＢアレル頻度を有する領域を示す。このことを血漿試料に基づくＡＡＩの計算と組み合わせると、２つの領域に関してｃが２．３３％および２．６７％と推定される。血漿中の他の領域を用いてｃを推定することによって４．４６％～９．５３％の値が得られ、腫瘍のヘテロ接合性の存在がはっきりとわかる。

【０７０８】

これらのデータは試料のうちの大部分で血漿にＣＮＶが検出できることを示し、腫瘍にＣＮＶが多くみられるほどＣＮＶが無細胞ＤＮＡで観測される可能性が高いことを示唆する。さらに、CoNVERGeは従来の腫瘍生検では発見できなかったかもしれないＣＮＶを液体細胞診試料から検出した。

【０７０９】

実施例９

本実施例では、様々な種類の試料をCoNVERGeで分析するために用いる試料調製方法の一例の詳細を説明する。

【０７１０】

２８，０００プレックスＰＣＲに関する単一細胞ＣＮＶプロトコル

多重化ＰＣＲは、多数の標的を１つの反応で同時に増幅できる。標的ＳＮＰを、１０％の最小の集団である少数アレル頻度を持つ各ゲノム領域で確認した（1000 Genomes Projectデータ：２０１２年４月３０日発表）。各ＳＮＰについて、準入れ子の複数のプライマーを、最大長７５ｂｐの増幅産物を生成し、かつ融解温度が５４℃～６０．５℃となるように設計した。プライマーの考えられるすべての組み合わせについてプライマー相互作用スコアを計算した。スコアの高いプライマーを排除してプライマー二量体産物が形成される可能性を低減した。候補ＰＣＲ試験法を、標的ＳＮＰのマイナーアレル頻度と、観測されたヘテロ接合体率と、HapMapに存在するかどうかと、増幅産物の長さとを基準にランク付けして選択した。

【０７１１】

一実験では、単一細胞試料を調製し、超多重化ＰＣＲの２８，０００プレックスプロトコルを用いて増幅した。試料は以下の方法で調製した：単一細胞を分析するため、１滴あたり３個または４個の細胞が存在するように細胞を段階希釈した。個々の細胞をピペットで取り、ＰＣＲ管に入れた。この細胞をプロテアーゼＫ、塩、ＤＴＴを用いて、以下の条件で溶解した。すなわち、５６℃で２０分間、９５℃で１０分間、次いで４℃で保持。ゲノムＤＮＡを分析するため、分析した単一細胞と同じ細胞株由来のＤＮＡを、購入するか、細胞を増殖させてＤＮＡを抽出することによって得た。Qiagen（登録商標）mp-PCRマスターミックス（終濃度２×ＭＭ）とプライマー（濃度７．５ｎＭ）を含む４０μＬの反応容量で、半入れ子にした逆方向プライマーを含む２８Ｋプライマー対について、以下の条件でＤＮＡを増幅した。すなわち９５℃で１０分間の後、（９６℃で３０秒間、６５℃で２９分間、７２℃で３０秒間）を２５サイクル、７２℃で２分間、次いで４℃で保持。増幅産物を１：２００で水に希釈し、２μＬをＳＴＡＲ２（反応容量１０μＬ（１×ＭＭ、プライマー濃度５ｎＭ））に添加し、内側順方向プライマーとタグ特異的な逆方向プライマーを半入れ子にして用いてＰＣＲを実施した。すなわち、９５℃で１５分間の後、（９４℃で３０秒間、６５℃で１分間、６０℃で５分間、６５℃で５分間、７２℃で３０秒間）を２５サイクル、７２℃で２分間、最後に４℃で保持。

【０７１２】

全長配列のタグとバーコードを増幅産物に付加し、アダプターに特異的なプライマーを用いて９サイクルかけて増幅した。配列決定の前にバーコードライブラリ産物をプールし、QIAquick（登録商標）PCR Purification Kit (Qiagen)を用いて精製し、Qubit（登録商標）dsDNA BR Assay Kit（Life Technologies）を用いて定量した。Illumina HiSeq 2500 sequencerを用いて増幅産物を配列決定した。

【０７１３】

血液／血漿試料からのＤＮＡ抽出

血液試料をＥＤＴＡ管に採取した。全血試料を遠心分離して３つの層に分離した：血液試料の上層５５％は血漿であり無細胞ＤＮＡを含み、中間層であるバフィーコートはＤＮＡ（全体の＜１％）を有する白血球を含み、また採取した血液試料の４５％である下層には赤血球が含まれていたが、この画分には赤血球が除核されているためＤＮＡは存在しなかった。QIAamp（登録商標）Circulating Nucleic Acid Kit、Qia-Amp（Qiagen, Valencia, CA）を製造元のプロトコルに従って用い、少なくとも１ｍＬの血漿から循環腫瘍ＤＮＡを単離した。

【０７１４】

染色体１ｐ、１ｑ、２ｐ、２ｑ、２２ｑ１１の３，１６８プレックスに関する血漿ＣＮＶプロトコル

血漿ＤＮＡライブラリを調製し、超多重ＰＣＲ３，１６８プレックスプロトコルを用いて増幅した。試料は以下のように調製した：最大２０ｍＬの血液を遠心分離してバフィー

コートと血漿を単離した。無細胞循環腫瘍ＤＮＡを血漿から抽出し、ライブラリを調製した。ＤＮＡを５０μＬのＴＥ緩衝液で溶離した。超多重ＰＣＲへの使用量は、約１２００ｎｇの使用量で増幅・精製したNatera血漿ライブラリのうち６．７μＬであった。血漿ＤＮＡを、Qiagen（登録商標）mp-PCRマスターミックス（最終濃度２×ＭＭ）と３，１６８プレックスプライマープールのタグ付きプライマー（濃度２ｎＭ）を含む２０μＬの反応容量でＰＣＲ増幅した。すなわち、９５℃で１０分間の後、（９６℃で３０秒、６５℃で２０分間、７２℃で３０秒間）を２５サイクル、さらに７２℃で２分間、最後に４℃で保持した。増幅産物を１：２０００で水に希釈し、１μＬを１０μＬの反応容量のバーコーディングＰＣＲに添加した。バーコードは、タグ特異的なプライマーを用いた１２サイクルのＰＣＲ増幅によって付加する。複数の試料の産物をプールし、次いでQIAquick PCR Purification Kit (Qiagen)で精製し、５０μＬのＤＮＡ懸濁用緩衝液で溶離した。２８，０００プレックスＰＣＲに関する単一細胞のＣＮＶプロトコルについて記載したのと同様に試料を次世代配列決定で配列決定する。

【０７１５】

血漿由来のＳＮＶパネルの乳癌に関する実現可能性

乳癌患者の血液試料由来の無細胞ＤＮＡを調製し、４つの８４プレックスプールに分けた３３６プライマー対を用いて増幅した。染色体１ｐ、１ｑ、２ｐ、２ｑ、２２ｑ１１の３，１６８プレックスに関する血漿ＣＮＶプロトコルに記載したのと同様にNatera血漿ライブラリを調製した。ＤＮＡを５０μＬのＴＥ緩衝液で溶離した。超多重ＰＣＲへの使用量は、使用量約６００ｎｇでは増幅・精製したNatera血漿ライブラリのうち２．５μＬであった。図６８Ａ～Ｂは、３１６８プレックスの超多重ＰＣＲ反応に用いたＳＮＰの多数アレル頻度と少数アレル頻度を示す。横軸はＳＮＰの数（左から順に染色体１ｑ、１ｐ、２ｑ、２ｐ、２２ｑ）を示す。ＳＮＰを1000 Genomes map for Humans, Group 19とｄｂＳＮＰから選択して対象を選んだが、1000 Genomesから選択したＳＮＰのみを用いてスクリーニングし、少数アレル頻度を得た。８４プレックスプライマープール（反応容量１０μＬ、Qiagen mp-PCRマスターミックス（最終濃度２×ＭＭ）、4mM EDTA、プライマー濃度７．５ｎＭ（合計１．２６μＭ））の４つの反応を並行して行い、血漿ＤＮＡをＰＣＲ増幅した。すなわち９５℃で１５分間の後、（９４℃で３０秒間、６５℃で１５分間、７２℃で３０秒間）を２５サイクル、７２℃で２分間、最後に４℃で保持した。４つのサブプールの増幅産物をそれぞれ１：２００で水に希釈し、１μＬを１０μＬの反応容量（Q5 HS HF master mix（終濃度１×）、１μＭの各バーコーディングプライマーを含む）のバーコーディングＰＣＲに加え、プールの各々を以下の条件で増幅した。すなわち９８℃で１分間の後、（９８℃で１０秒間、７０℃で１０秒間、６０℃で３０秒間、６５℃で１５秒間、７２℃で１５秒間）を２５サイクル、７２℃で２分間、最後に４℃で保持した。ライブラリをQIAquick（登録商標）PCR Purification Kit（Qiagen）で精製し、５０μＬのＤＮＡ懸濁用緩衝液で溶離した。試料をペアエンド配列決定で配列決定した。

【０７１６】

実施例１０

本実施例では、配列決定データを分析してＳＮＶを確認するための方法の一例に関して詳細を説明する。

【０７１７】

ＳＮＶ確認方法１：この態様では、実行回ごとの特異的な人為的結果を説明するため、同じ実行回で配列決定された正常な血漿試料を用いて背景エラーモデルを構築した。一部の態様では、５個、１０個、１５個、２０個、２５個、３０個、４０個、５０個、１００個、１５０個、２００個、２５０個、または２５０個超の正常な血漿試料を同じ実行回の配列決定で分析した。一部の例示の態様では、２０個、２５個、４０個、または５０個の正常な血漿試料を同じ実行回の配列決定で分析する。多様体アレル頻度の中央値がカットオフ値を超える正常値である、ノイズの多い位置を除外する。たとえば、一部の態様ではこのカットオフ値は＞０．１％、０．２％、０．２５％、０．５％、１％、２％、５％または１０％である。一部の例示の態様では、多様体アレル頻度の中央値が０．５％超の正

常値である、ノイズの多い位置を除外する。異常値試料を反復的にモデルから除外し、ノイズと不純物を説明する。一部の態様では、ｚスコアが５超、６超、７超、８超、９超、または１０超の試料をデータ分析から除外した。すべてのゲノムの座位の各塩基置換について、リード深度で加重した、エラーの平均値および標準偏差を計算した。腫瘍または無細胞血漿試料の、少なくとも５個の多様体のリードがありｚスコアが背景エラーモデルに対して１０であった位置を候補変異と呼んだ。

【０７１８】

ＳＮＶ確認方法２：本実施例では、血漿中の無細胞循環腫瘍ＤＮＡデータを用いて一塩基多様体（ＳＮＶ）を決定することを目的とする。ＰＣＲ工程を確率過程としてモデル化し、訓練セットを使用してパラメータを推定し、別個の試験セットを使用して最終ＳＮＶ検出を行う。主要な考え方は、複数のＰＣＲサイクルにわたる誤差の伝搬を決定し、背景誤差の平均と分散を計算し、背景誤差を実変異と区別することである。

【０７１９】

各塩基について以下のパラメータを推定する：

ｐ＝効率（各リードが各サイクルで複製される確率）

p_e＝変異タイプｅに関するサイクル当たりのエラー率（タイプｅのエラーが起こる確率）

X_0＝最初の分子数

リードをＰＣＲ工程の過程で複製するにつれ、より多くのエラーが発生する。したがって、リードのエラーの特性は元のリードからの分離度によって決定される。ｋ回の複製を経て生成されたリードをk^{th}世代とする。

各塩基について以下の変数を定義する：

X_{ij}＝ＰＣＲのｊサイクル目で生成される世代のリード数

Y_{ij}＝ｊ回目のサイクル終了時の世代ｉのリード数の合計

X_{ij}^{e}＝ＰＣＲのｊサイクル目で生成される、変異ｅを有する世代ｉのリード数

また、正常な分子分子X_0の他に、変異ｅを有する$f_e X_0$個の分子がＰＣＲ工程の開始時に存在する（したがってｆｅ／（１＋ｆｅ）は最初の混合物に含まれる変異分子の比率である）。ｊ－１サイクル目の世代ｉ－１のリード総数が与えられると、ｊサイクル目で生成される世代ｉのリード数は、試料のサイズ$Y_{i-1,j-1}$および確率パラメータｐを含む二項分布を有する。したがって、$E(X_{ij,}|Y_{i-1,j-1},p)=pY_{i-1,j-1}$かつ$Var(X_{ij,}|Y_{i-1,j-1},p)=p(1-p)Y_{i-1,j-1}$である。

【０７２０】

また、

【数２５】

$$Y_{ij}=\sum_{k=i}^{j}X_{ik}$$

である。したがって、回帰、シミュレーション、または類似の方法によって$E(X_{ij,})$を決定することができる。同様に、$Var(X_{ij})=E(Var(X_{ij,}|p))+Var(E(X_{ij,}|p))$を、ｐの分布を用いて決定できる。最後に、$E(X_{ij}^{e}|Y_{i-1,j-1},p_e)=p_eY_{i-1,j-1}$であり、$Var(X_{ij}^{e}|Y_{i-1,j-1},p)=p_e(1-p_e)Y_{i-1,j-1}$であり、これらを用いて$E(X_{ij}^{e})$と$Var(X_{ij}^{e})$を計算できる。

【０７２１】

２０．

６＋．２アルゴリズム

【０７２２】

まず、アルゴリズムはトレーニングセットを用いてサイクル当たりの効率およびエラー

率を推定する。ｎをＰＣＲサイクル数の合計とする。

【０７２３】

各塩基の総リード数R_bは、$(1+p_b)^n X_0$（式中、p_bは塩基ｂでの効率である）によって近似できる。次に$(R_b/X_0)^{1/n}$を用いて$1+p_b$を近似できる。次に、すべての訓練用試料にわたってp_bの平均および標準偏差を決定し、各塩基の確率分布のパラメータ（正規分布、ベータ分布、または類似の分布など）を推定することができる。

【０７２４】

同様に、各塩基ｂでのエラーｅのリードR_b^eの数を用いてp_eを推定できる。すべての訓練用試料のエラー率の平均と標準偏差を決定した後、この平均値と標準偏差値を使用してパラメータを推定した確率分布（正規分布、ベータ分布、または類似の分布など）を近似する。

【０７２５】

次に、試験データについて、各塩基の最初の開始コピーを以下のように推定する：

【数２６】

$$\int_0^1 \frac{R_b}{(1+p_b)^n} f(p_b)\,dp_b$$

（式中、ｆ（．）は訓練セットから得た推定分布である）。

【０７２６】

したがって、確率過程において使用するパラメータを推定した。次に、これらの推定値を使用して、各サイクルで生成される分子の平均と分散を推定できる（なお、推定は正常な分子、エラー分子、および変異分子について個別に行う）。

【０７２７】

最後に、確率論的方法（最尤法または類似の方法など）を使用することによって、エラー分子、変異分子、および正常分子の分布に最も適合する最良のf_e値を決定できる。より具体的には、最終リードにおける様々なf_e値について誤差分子の全体分子に対する比の期待値を推定し、これらの値のそれぞれについてデータの尤度を決定し、次に最も尤度の高い値を選択する。

【０７２８】

ある態様では、上記方法２を以下のように実行する：

ａ）サイクル当たりのＰＣＲ効率とエラー率を、訓練データセットを用いて推定し；

ｂ）工程（ａ）で推定した効率の分布を用いて、各塩基の試験データセットに関する開始分子の数を推定し；

ｃ）必要であれば、工程（ｂ）で推定した開始分子数を用いて、試験データセットに関する効率の推定値を更新し；

ｄ）工程（ａ）、（ｂ）、（ｃ）で推定した試験セットデータとパラメータを用いて、（ある初期の実変異分子率を有する調査領域に関する）総分子数、背景エラー分子数および実変異分子数に関する平均および分散を推定し；

ｅ）分布を分子全体におけるエラー分子全体（背景エラー分子および実変異分子）に適合させることによって、調査領域における各実変異率に関する尤度を計算し；

ｆ）最も尤度の高い実変異率を決定し、工程（ｅ）で得たデータを用いて信頼度を計算する。

【０７２９】

実施例１１

本実施例では、循環ＤＮＡ中のＣＮＶを検出することによって癌を検出するための、本明細書に記載の多重化ＰＣＲ・CoNVERGe法を用いた結果を記載する。本明細書に記載の染色体１ｐ、１ｑ、２ｐ、２ｑ、および２２ｑ１１の３，１６８プレックスに関する血漿Ｃ

ＮＶプロトコルを使用した。２１名の乳癌患者（ステージＩ～ＩＩＩＢ）から採取した血漿を分析した。図４４に示す結果から、ＡＡＩ≧０．４５％としたすべてのサンプルでＣＮＶが検出され、わずか６２個のヘテロ接合性のＳＮＰのみが必要だったことがわかる。類似のプロトコルを用いて、卵巣癌患者から採取した血漿を分析した。図４５に示すように、０．４５％のカットオフ値を用いて１００％の卵巣癌検出率を達成した。５個の試料のそれぞれには対応する腫瘍試料も存在した。

【０７３０】

実施例１２

本実施例では、血漿中のＣＮＶとＳＮＶの存在を試験することで癌検出能力が劇的に改善したことを示す。上記の実施例に記載の方法でＣＮＶおよびＳＮＶを検出した。試料は実施例９の適切なプロトコルに従って調製した。ＳＮＶに関する上記方法１を用いてＳＮＶを確認した。図４６に示すように、ステージＩ～ＩＩＩの癌患者の血漿をＣＮＶおよびＳＮＶの両方について分析することによって、ＳＮＶ単独について試験する場合と比較して、乳癌および肺癌の検出感度が大きく改善される。ＳＮＶを単独で分析した場合、癌の７１％が血漿試料で検出されたが、ＳＮＶおよび／またはＣＮＶの有無について分析することによって、検出率は分析対象の母集団のうち乳癌では８３％、肺癌では９２％に上昇した。ＴＣＧＡおよびＣＯＳＭＩＣデータセットで確認されたすべてのＳＮＶおよびＣＮＶについて考慮する場合、診断負荷の期待値は、乳癌では９７％を超え、肺癌では９８％を超える。

【０７３１】

さらに、実施例９に記載の血漿試料調製方法と上記のＳＮＶ確認方法１を用いて、病期の異なる４１名癌患者の試料を分析した。図４７に示すように、乳癌患者の腫瘍ＤＮＡ中のＣＮＶとＳＮＶについて試験した場合、ＳＮＶについては０．２％無細胞循環腫瘍ＤＮＡ、ＣＮＶについては０．４５％無細胞循環腫瘍ＤＮＡの定量限界を用いて、ステージＩの６０％、ステージＩＩの８８％、ステージＩＩＩの１００％の乳癌が検出された。

【０７３２】

図４８に示すように、無細胞循環腫瘍ＤＮＡ中のＣＮＶとＳＮＶについて、病期の小区分が異なる４１名の乳癌患者の試料に注目して試験した場合、ＳＮＶについては０．２％無細胞循環腫瘍ＤＮＡ、ＣＮＶについては０．４５％無細胞循環腫瘍ＤＮＡの定量限界を用いて、ステージＩの６０％、ステージＩＩの１００％、ステージＩＩＡの９０％、ステージＩＩＢの８０％、ステージＩＩＩ、ＩＩＩＡ、ＩＩＩＢの１００％の乳癌が検出された。図４９に示すように、２４個の肺癌患者の試料から得た無細胞循環腫瘍ＤＮＡでＣＮＶとＳＮＶについて試験した場合、ＳＮＶについては０．２％無細胞循環腫瘍ＤＮＡ、ＣＮＶについては０．４５％無細胞循環腫瘍ＤＮＡの定量限界を用いて、ステージＩの８８％、ステージＩＩの１００％、ステージＩＩＩの１００％の肺癌が検出された。図５０に示すように、病期の小区分が異なる２４個の肺癌患者の試料に着目して無細胞循環腫瘍ＤＮＡでＣＮＶとＳＮＶについて試験した場合、ＳＮＶについては０．２％無細胞循環腫瘍ＤＮＡ、ＣＮＶについては０．４５％無細胞循環腫瘍ＤＮＡの定量限界を用いて、ステージＩＢの肺癌患者では８２％の検出率であったことを除けば、すべての病期小区分で１００％の検出率が得られた。

【０７３３】

実施例１３

本実施例では、無細胞循環腫瘍ＤＮＡ中のＳＮＶの検出が、生検試料で多様体アレルを識別する際の腫瘍の異質性に起因する検出限界を克服することを実証する。３名の小細胞肺癌患者の試料のTRACERx試料と１名の線癌・肺癌患者の試料（腫瘍生検と、対応する術前血漿試料を採取した）を腫瘍の異質性の分析に使用した。試料は、Cancer Research UK Lung Cancer Centre of Excellence, University College London Cancer Institute（London WC1E 6BT, UK）から入手した。試料は、ＳＮＶ変異分析用の原発性肺癌試料であった。各患者の癌を有する肺全体の様々な領域から２個～３個の生検を採取した（図５１Ａ）。根本的なクローンの異質性を識別するため、各生検試料を全エキソン配列決定（Illu

mina（登録商標）HiSeq200; Illumina、San Diego、CA）した後、PGM（登録商標）の装置でAmpliSeq（登録商標） sequencing（Ion Torrent、South San Francisco、CA）することによって試験した。ＳＮＶ分析の後、各生検試料について多様体アレル頻度（ＶＡＦ）を決定した（図５１Ｂ）。

【０７３４】

４名の患者それぞれから採取した血漿試料を用いて無細胞循環腫瘍ＤＮＡを単離し、クローンＳＮＶ変異とサブクローンＳＮＶ変異が血漿で確認され、腫瘍の異質性が克服された（図５２）。クローン集団では、試験したすべての生検試料と血漿のＶＡＦの検出があったが、サブクローン集団では、少なくとも１個の（すべてではない）生検試料でＶＡＦアレルの検出があった。これらの血漿を各患者の無細胞循環腫瘍ＤＮＡに見られるＳＮＶの累積代表値とみなした。配列決定により確認したすべてのＳＮＶについて対応するＰＣＲ試験法を設計できたわけではない。

【０７３５】

腫瘍の異質性を確認するためのAmpliSeq（商標）（Swanton）と超多重化ＰＣＲ／次世代配列決定の試験方法を比較するため、Nateraは、生検とそれに対応する血漿由来無細胞循環腫瘍ＤＮＡの両方で各ＳＮＶ変異についてＶＡＦを検出するためのＰＣＲ試験法を設計した（図５３）。空欄は入手可能な生検試料を示し、値がゼロの場合、ＶＡＦが検出されなかったことを示す。以下の１１個の遺伝子はAmpliSeq（商標）のＦＰ（偽陽性）／ＦＮ（偽陰性）試験法で最初は陰性として認識された（偽のＶＡＦ検出）がNateraのＴＰ（真陽性）／ＴＮ（真陰性）試験法と多重化ＰＣＲ／次世代配列決定による分析方法で正しく検出された：Ｌ１２：ＣＹＦＩＰ１、ＦＡＴ１、ＭＬＬＴ４、およびＲＡＳＡ１；Ｌ１３：ＨＥＲＣ４、ＪＡＫ２、ＭＳＨ２、ＭＴＯＲ、およびＰＬＣＧ２；Ｌ１５：ＧＡＢＲＧ１；Ｌ１７：ＴＲＩＭ６７。予想外にも、AmpliSeq（商標）の配列決定生データを再試験すると、これらの結果が検証された。AmpliSeq（商標）配列決定生データのファイルから、このデータがＰＧＭ（登録商標）やIllumina（登録商標）の検出可能閾値の設定に満たないことがわかった。データから、３８個のうち１６個の多様体が血漿で検出され、Ｌ１２患者試料ではクローンＳＮＶ変異を主に有していた数個の生検試料があったことが確認できた：Ｌ１２：ＢＲＩＰ１、ＣＡＲＳ、ＦＡＴ１、ＭＬＬＴ４、ＮＦＥ２Ｌ２、ＴＰ５３、ＴＰ５３（以下でも同様：Ｌ１３患者：ＥＧＦＲ、ＥＧＦＲ、ＴＰ５３；およびＬ１５患者：ＫＤＭ６Ａ、ＲＯＳ１）。さらに２名の患者が合計で４個のサブクローン多様性変異を血漿中に有していることがわかった（Ｌ１２:ＣＩＣ、ＫＤＭ６Ａ；およびＬ１７：ＮＦ１、ＴＲＩＭ６７）。これらの結果が図５４Ａに要約されている。図５４Ａは、図５３に記載した各試料に関する、各試験方法による平均ＶＡＦ値の箱ひげ図である。図５４Ｂは、各試験法のＶＡＦ値の試料平均の直線回帰図で表した直接比較である。

【０７３６】

実施例１４

本実施例では、次世代配列決定に続くワークフローでプライマー量が多重ＰＣＲの反応物を制限するような低いプライマー濃度を用いることによって、増幅反応のプール全体のリード深度（したがって検出限界）が改善されることを実証する。全体の反応容量を２０μＬでなく１０μＬとした以外は上記の実施例９に従い、３，１６８プレックスパネルを用いていくつかの実験を血漿ＣＮＶについて行った。さらに、ＰＣＲを１５サイクル、２０サイクル、または２５サイクルかけて実施した。プライマー濃度を２ｎＭにした以外は実施例９のプロトコルに従い、乳癌試料で４つの８４プレックスプールを用いて他の実験を実施し、ＰＣＲ増幅を１５サイクル、２０サイクル、または２５サイクル実施した。

【０７３７】

理論によって制限するものではないが、プライマーを制限した多重ＰＣＲでは、多重ＰＣＲ後にマルチリード配列決定（たとえばIllumina（登録商標）のHiSeqまたはMiSeqシステム、あるいはIon Torrent PGM（登録商標）またはProton（商標）システムでの配列決定）を行う場合に、以下の考察に基づき、リード深度の均一性が改善されると考えられている。多重ＰＣＲの増幅の一部の効率が他の増幅よりも低い場合、通常の多重ＰＣＲでは

リード深度（depth of read;「ＤＯＲ」という）の値が広範囲になる。しかし、プライマー量を制限し、かつプライマーを使い果たすのに必要なサイクル数よりも多いサイクル数で多重ＰＣＲを行う場合、増幅効率が高いほど倍増が停止し（使用するプライマーがなくなっているため）、効率が低いほど倍増が継続する。その結果、すべての増幅産物に関して増幅産物量が同様になる。このことによってＤＯＲの分布がはるかに均一になる。

【０７３８】

以下の計算を用いて、所定量のプライマーと開始核酸鋳型を必要とすると考えられるサイクル数を決定する：

【０７３９】

所定の開始ＤＮＡ使用量を各標的１００ｋコピー（10^5；これは増幅したライブラリを用いて容易に得られる）と仮定する；

【０７４０】

例示の濃度として２ｎＭの各プライマーを用いると仮定する（ただし、たとえば０．２ｎＭ、０．５ｎＭ、１ｎＭ、１．５ｎＭ、２ｎＭ、２．５ｎＭ、５ｎＭ、または１０ｎＭなどの他の濃度も使用できると考えらえる）；

【０７４１】

各プライマーについてプライマー分子数を計算する：2×10^{-9}（モル濃度２ｎＭ）$\times10\times10^{-6}$（反応容量１０μＬ）$\times6\times10^{23}$（モル当たりの分子数、アボガドロ数）$=12\times10^{9}$；

【０７４２】

プライマーをすべて使うのに必要な増幅倍数を計算する：12×10^{9}（プライマー分子の数）／10^{5}（各標的のコピー数）$=12\times10^{4}$；

【０７４３】

各サイクルの効率が１００％と仮定し、この増幅倍数を達成するのに必要なサイクル数を計算する：ｌｏｇ２（12×10^{4}）＝１７サイクル（各サイクルでコピー数が倍になるためｌｏｇ２とする）。

【０７４４】

したがって、この条件（使用量１００ｋコピー、プライマー２ｎＭ、反応容量１０μＬ、各サイクルのＰＣＲ効率は１００％と仮定する）の場合、プライマーは１７サイクルのＰＣＲで消費される。

【０７４５】

しかし、重要な前提として、産物の一部は１００％の効率ではないため、効率の測定を行わない場合（いずれにせよ少数を用いる場合にしか実用的ではないが）、プライマーを使いきるのに１７サイクルを超えるサイクル数が必要となる場合がある。

【０７４６】

図５５～図５８は、ＳＮＶの４つの８４プレックスＰＣＲプライマープールに関する結果を示す。各プールについて、サイクルを１５サイクルから２０サイクル、２５サイクルへと増やすにつれＤＯＲ効率の改善が認められた。３，１６８プレックスパネルを用いた実験でも似た結果が得られた（図５９～図６１）。リード深度を上げるにつれて検出限界は下がった（すなわち、ＳＮＶ感度が高くなった）。さらに、感度は一貫して、塩基転換変異の検出の場合の方が転移変異の検出の場合よりも高かった。プライマーを制限した多重ＰＣＲをマルチリード配列決定の前に行うと追加のサイクルでＤＯＲ効率をさらに上げることができる可能性が高い。

【０７４７】

したがって、本明細書に記載の一側面に、以下の方法が挙げられる：核酸試料中の複数の標的座位を増幅する方法であって、（ｉ）核酸試料をプライマーのライブラリおよび他のプライマー伸長成分と接触させて反応混合物を生成すること（ここで、反応混合物中の他のプライマー伸長反応成分と比較した各プライマーの相対量によって、プライマーが限界濃度で存在するような反応が生じ、プライマーは複数の異なる標的座位とハイブリダイズする）；（ｉｉ）反応混合物を、プライマーライブラリ中のプライマーが消費またはす

べて使用されるのに十分なサイクル数でプライマー伸長反応条件に供して標的増幅産物を含む増幅産物を生成することを含む、方法。たとえば、複数の異なる標的座位としては、少なくとも２個、３個、５個、１０個、２５個、５０個、１００個、２００個、２５０個、５００個、１，０００個、２，０００個、５，０００個、７，５００個、１０，０００個、２０，０００個、２５，０００個、３０，０００個、４０，０００個、５０，０００個、７５，０００個、または１００，０００個の異なる標的座位であり、最大で５０個、１００個、２００個、２５０個、５００個、１，０００個、２，０００個、５，０００個、７，５００個、１０，０００個、２０，０００個、２５，０００個、３０，０００個、４０，０００個、５０，０００個、７５，０００個、１００，０００個、２００，０００個、２５０，０００個、５００，０００個、および１，０００，０００個の異なる標的座位を用いて反応混合物を生成することができる。

【０７４８】

例示の態様は、プライマーの量を律速量に決定することを含む。この計算は、一般的には標的分子数の推定および／または決定を含み、実行した増幅サイクル数の分析および／または決定を含む。たとえば、例示の態様では、各プライマーの濃度は、１００ｎＭ未満、７５ｎＭ未満、５０ｎＭ未満、２５ｎＭ未満、１０ｎＭ未満、５ｎＭ未満、２ｎＭ未満、１ｎＭ未満、０．５ｎＭ未満、０．２５ｎＭ未満、０．２ｎＭ未満、または０．１ｎＭ未満である。様々な態様では、プライマーのＧＣ含量は３０％～８０％であり、たとえば４０％～７０％または５０％～６０％（それぞれ、範囲の上限と下限を含む）である。一部の態様では、プライマーのＧＣ含量の範囲（たとえば、最大ＧＣ含量－最小ＧＣ含量（たとえば８０％－６０％＝２０％の範囲））は、３０％未満、２０％未満、１０％未満、または５％未満である。一部の態様では、プライマーの融解温度（T_m）は、４０℃～８０℃であり、たとえば５０℃～７０℃、５５℃～６５℃、または５７℃～６０．５℃（それぞれ、範囲の上限と下限を含む）である。一部の態様では、プライマーの融解温度の範囲は、２０℃未満、１５℃未満、１０℃未満、５℃未満、３℃未満、または１℃未満である。一部の態様では、プライマーの長さは、１５～１００ヌクレオチドであり、たとえば、１５～７５ヌクレオチド、１５～４０ヌクレオチド、１７～３５ヌクレオチド、１８～３０ヌクレオチド、２０～６５ヌクレオチド（それぞれ、範囲の上限と下限を含む）である。一部の態様では、プライマーは、内部ループ構造を形成するタグなどの、標的特異的ではないタグを含む。一部の態様では、タグは２つのＤＮＡ結合部位に存在する。様々な態様では、プライマーは標的座位に特異的な５’領域と、標的座位に特異的ではなくループ構造を形成する内部領域と、標的座位に特異的な３’領域とを含む。様々な態様では、３’領域の長さは少なくとも７ヌクレオチドである。一部の態様では、３’領域の長さは７～２０ヌクレオチドであり、たとえば、７～１５ヌクレオチド、または７～１０ヌクレオチド（それぞれ、範囲の上限と下限を含む）である。様々な態様では、試験プライマーは、標的座位に特異的ではない５’領域（タグまたは汎用プライマー結合部位など）と、それに続いて標的座位特異的な領域と、標的座位に特異的ではなくループ構造を形成する内部領域と、標的座位に特異的な３’領域とを含む。一部の態様では、プライマーの長さの範囲は５０ヌクレオチド未満、４０ヌクレオチド未満、３０ヌクレオチド未満、２０ヌクレオチド未満、１０ヌクレオチド未満、または５ヌクレオチド未満である。一部の態様では、標的増幅産物の長さは５０～１００ヌクレオチドであり、たとえば６０～８０ヌクレオチド、または６０～７５ヌクレオチド（それぞれ、範囲の上限と下限を含む）である。一部の態様では、標的増幅産物の長さの範囲は、１００ヌクレオチド未満、７５ヌクレオチド未満、５０ヌクレオチド未満、２５ヌクレオチド未満、１５ヌクレオチド未満、１０ヌクレオチド未満、または５ヌクレオチド未満である。

【０７４９】

本発明のいずれかの側面の様々な態様では、プライマー伸長条件は、ポリメラーゼ連鎖反応（ＰＣＲ）条件である。様々な態様では、アニーリング工程の長さは、３分超、５分超、８分超、１０分超、または１５分超、かつ２４０分未満、１２０分未満、６０分未満、または３０分未満である。様々な態様では、伸長工程の長さは、３分超、５分超、８分

超、１０分超、または１５分超であり、２４０分未満、１２０分未満、６０分未満、または３０分未満である。

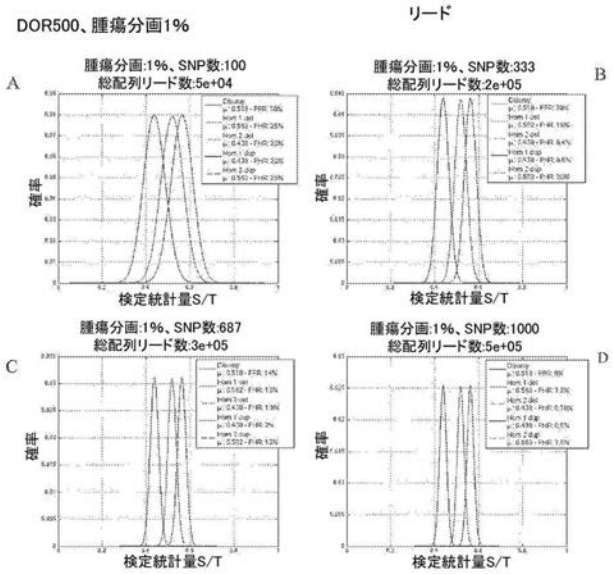
【０７５０】

実施例１５

本実施例では、本発明のＳＮＶ検出方法が単一細胞分析（単一分子分析ともいう）でモザイク現象を確認する能力を実証する。図６２は、実施例９に記載の２８Ｋプレックスプライマーセットを用いた２８Ｋ単一細胞法に従い腫瘍細胞のゲノムＤＮＡと単一細胞／分子を使用して得た多重ＰＣＲの結果を示す。この方法を用いて、４．７Ｍリード以上のリードのうち８５％超をマッピングした（標的あたり約１６７リード）。図の下部から、細胞にモザイク現象が見られたことがわかる。

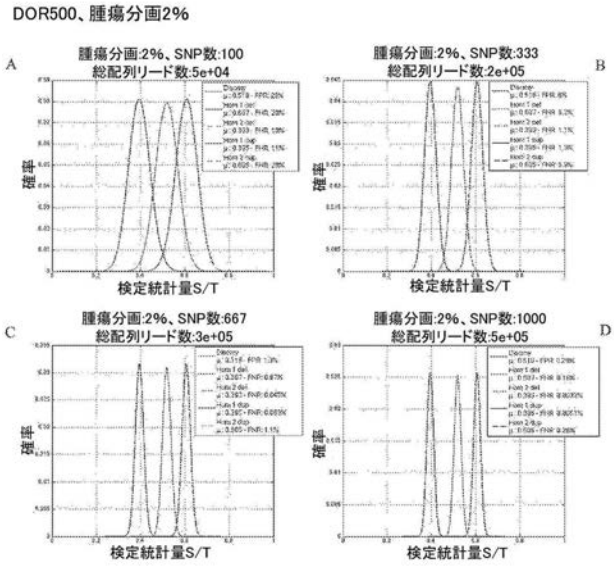
【図１Ａ－１Ｄ】

Figures 1A-1D



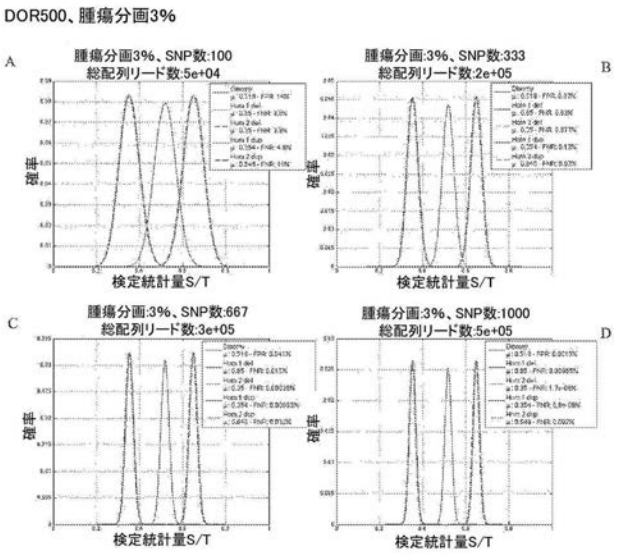
【図２Ａ－２Ｄ】

Figures 2A-2D



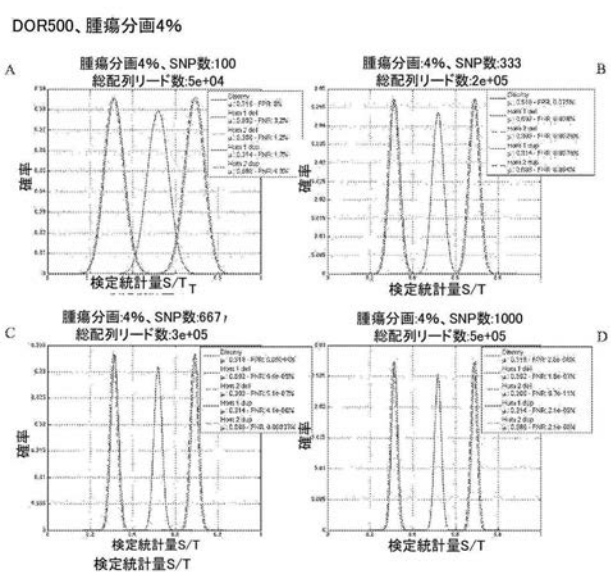
【図３Ａ－３Ｄ】

Figures 3A-3D



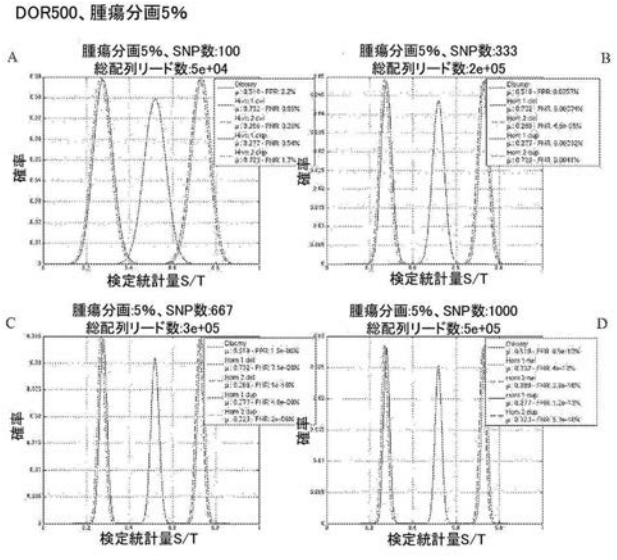
【図４Ａ－４Ｄ】

Figures 4A-4D



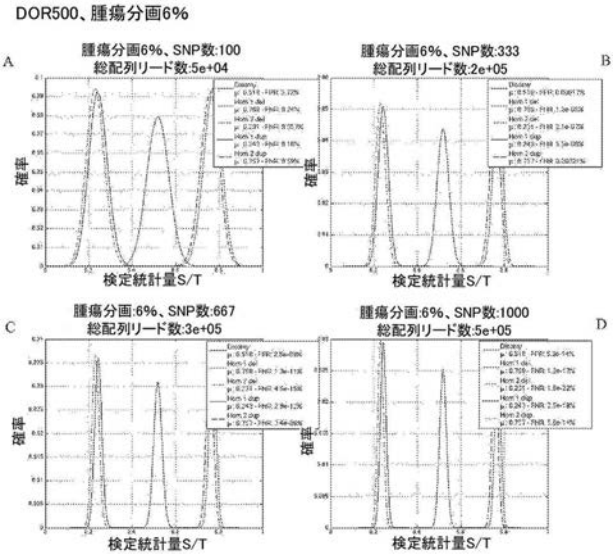
【図５Ａ－５Ｄ】

Figures 5A-5D



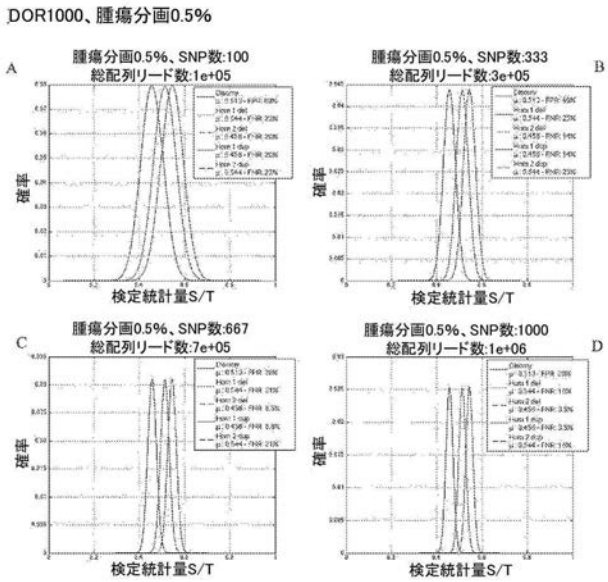
【図６Ａ－６Ｄ】

Figures 6A-6D



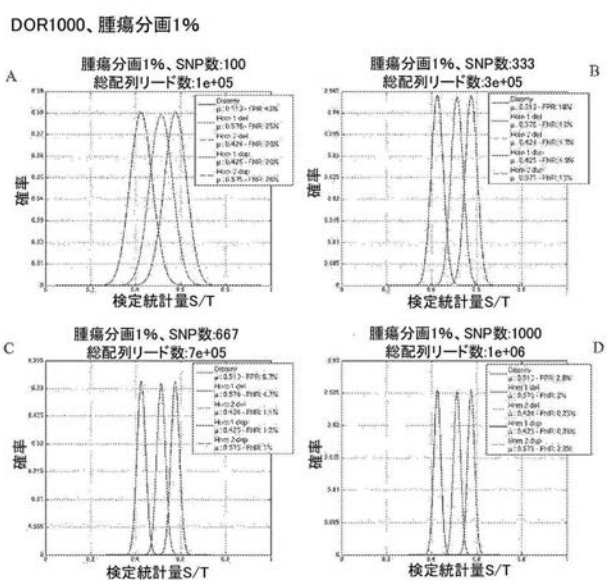
【図７Ａ－７Ｄ】

Figures 7A-7D



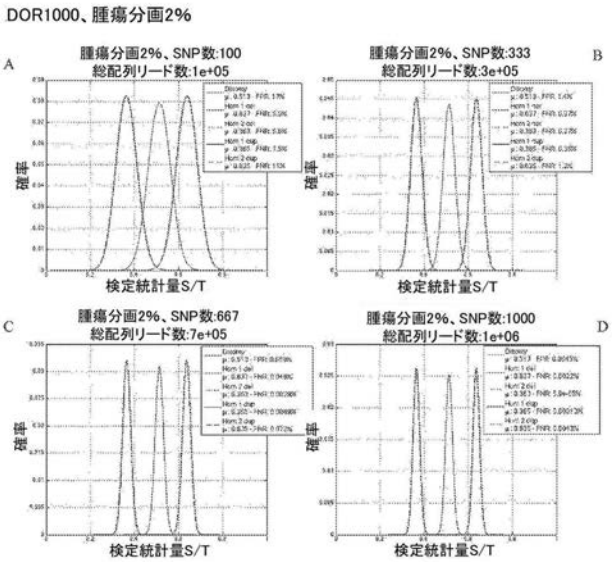
【図８Ａ－８Ｄ】

Figures 8A-8D



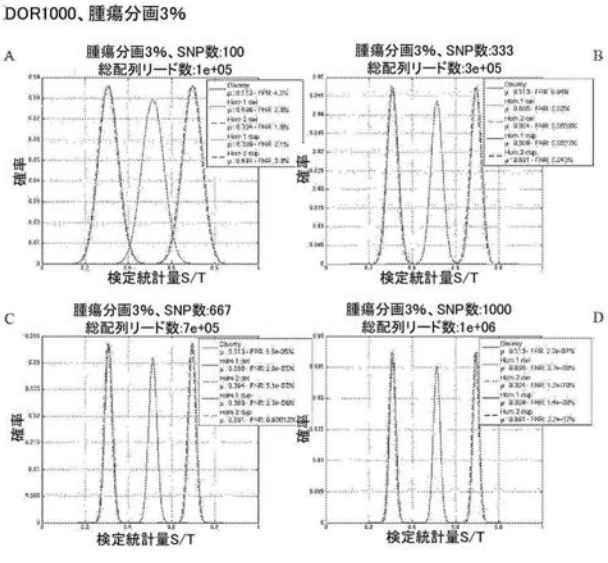
【図９Ａ－９Ｄ】

Figures 9A-9D



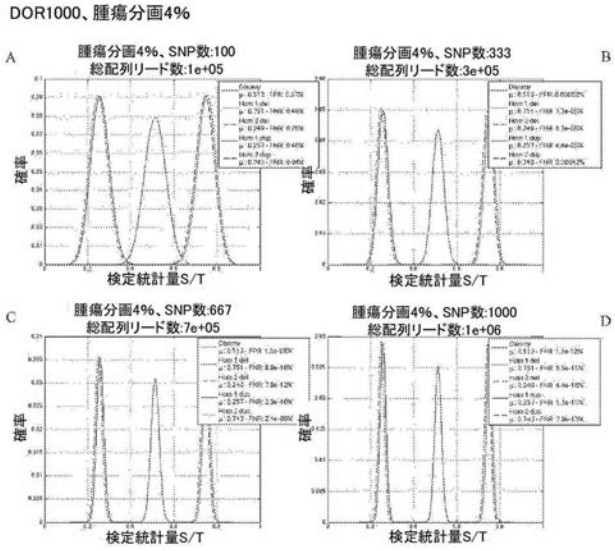
【図１０Ａ－１０Ｄ】

Figures 10A-10D



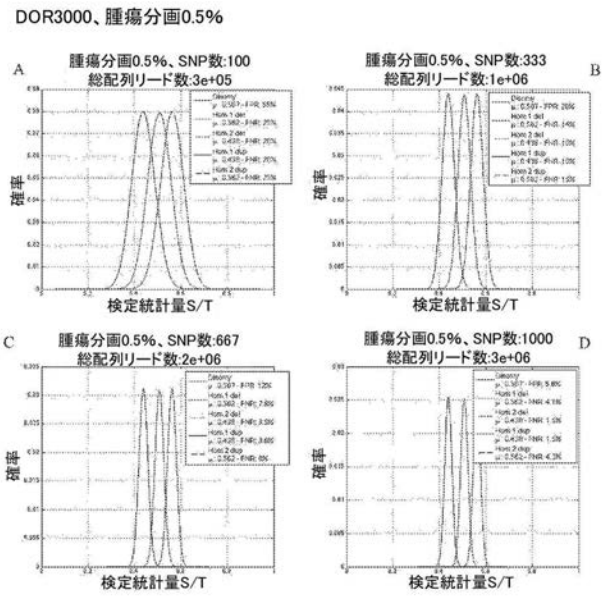
【図１１Ａ－１１Ｄ】

Figures 11A-11D



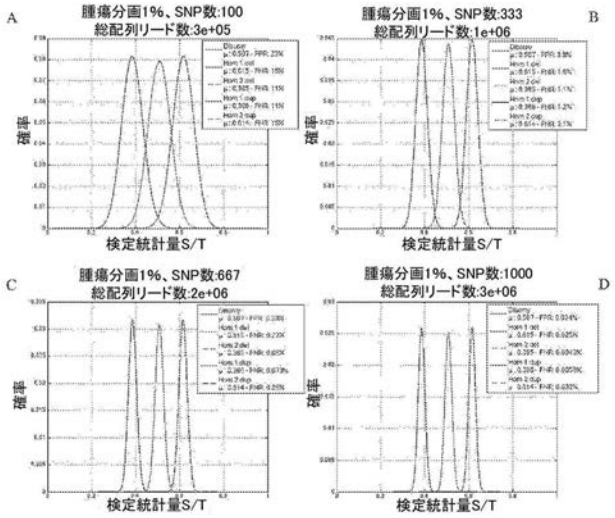
【図１２Ａ－１２Ｄ】

Figures 12A-12D



【図１３Ａ－１３Ｄ】

Figures 13A-13D

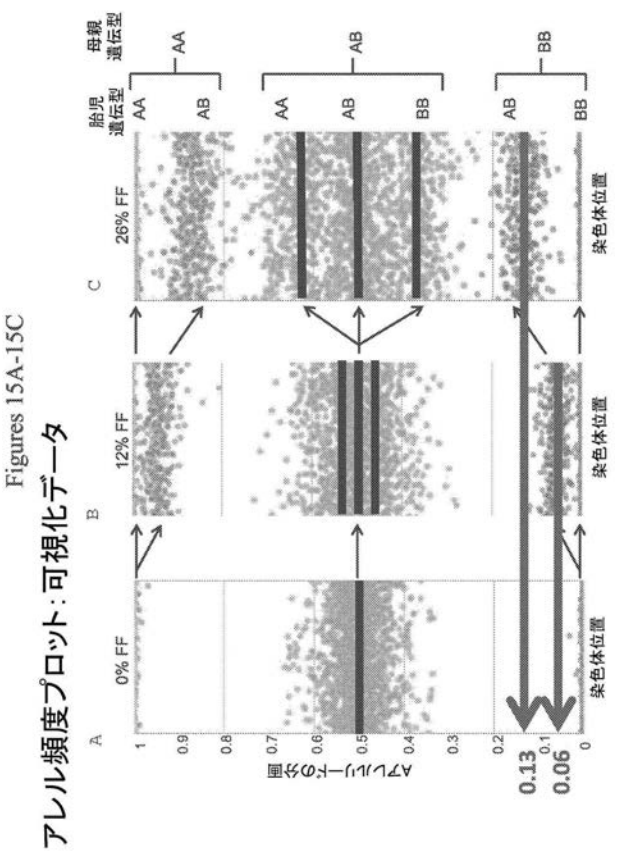


【図１４】

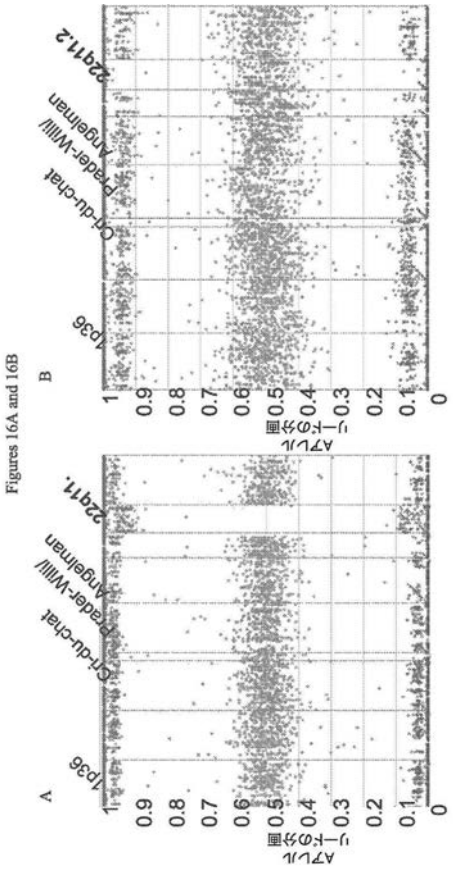
Figure 14

症候群	感度	特異性
22q11.2 Deletion/ DiGeorge	97.8% (45/46) [CI: 88.5 – 99.95%]	99.2% (392/395) [CI: 97.8 - 99.8%]
Angelman	100% (21/21) [CI: 83.9 – 100%]	100% (418/418) [CI: 99.1 – 100%]
Cri du chat	100% (24/24) [CI: 85.7 – 100%]	99.8% (415/416) [CI: 98.7 – 99.99%]
Monosomy 1p36	100% (1/1) [CI: 2.5 – 100%]	100% (438/438) [CI: 99.2 – 100%]
Prader-Willi	100% (15/15) [CI: 78.2 – 100%]	100% (424/424) [CI: 99.1 – 100%]
Wolf-Hirschhorn	100% (2/2) [CI: 15.8 – 100%]	99.8% (437/438) [CI: 98.7 – 100%]

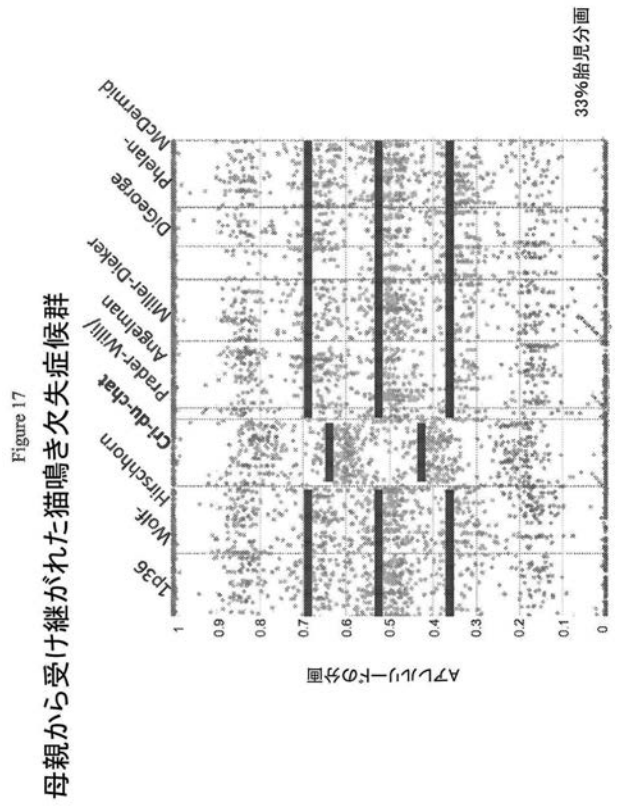
【図１５Ａ－１５Ｃ】



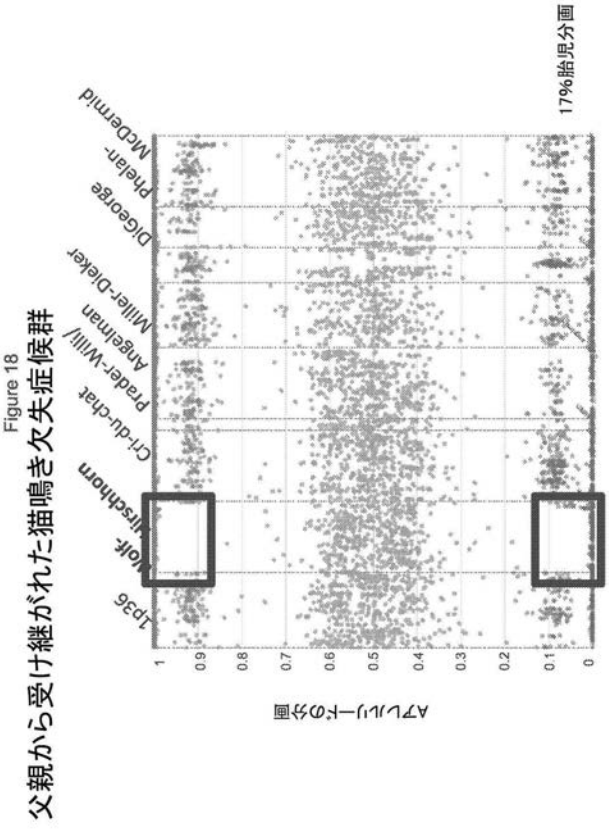
【図１６Ａ－１６Ｂ】



【図１７】



【図１８】



【図１９Ａ】

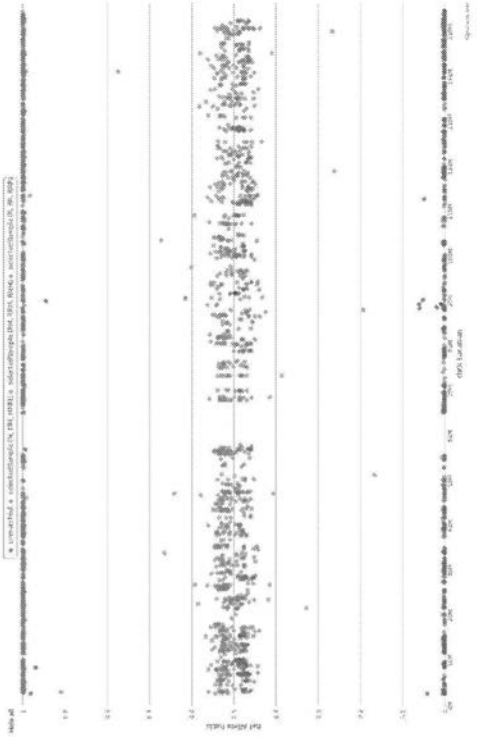


FIG. 19A 娘中16%父親

【図１９Ｂ】

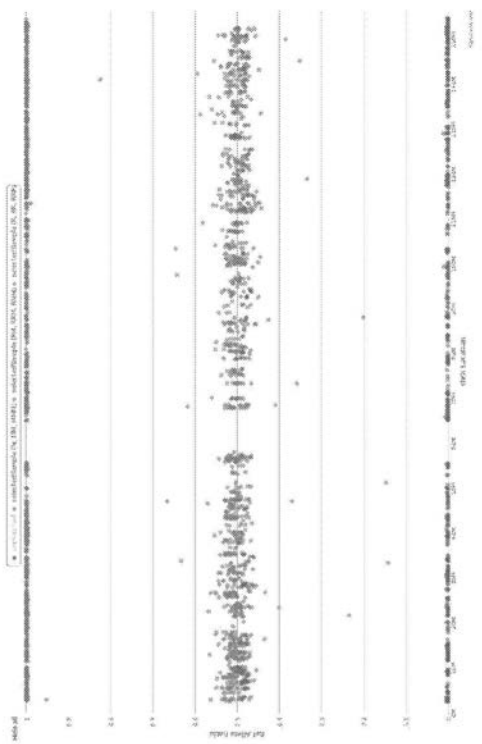


FIG. 19B 娘中10%父親

【図１９Ｃ】

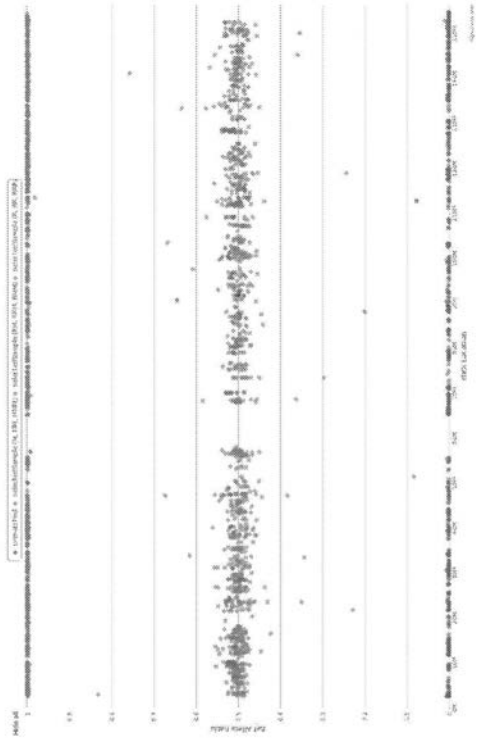


FIG. 19C 娘中1%父親

【図１９Ｄ】

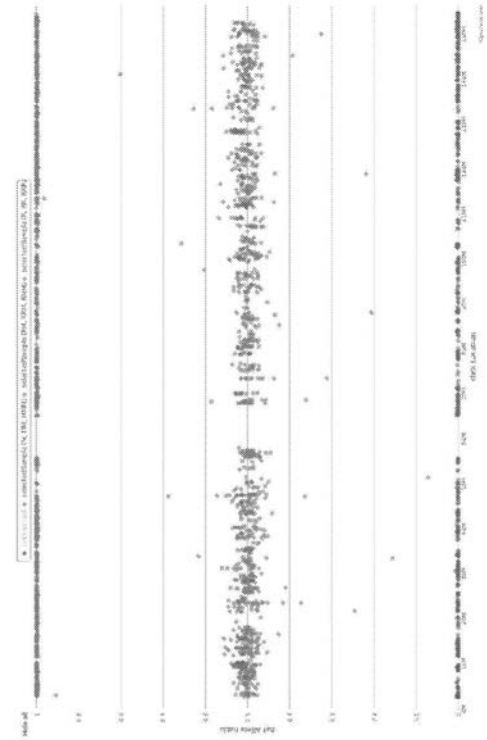
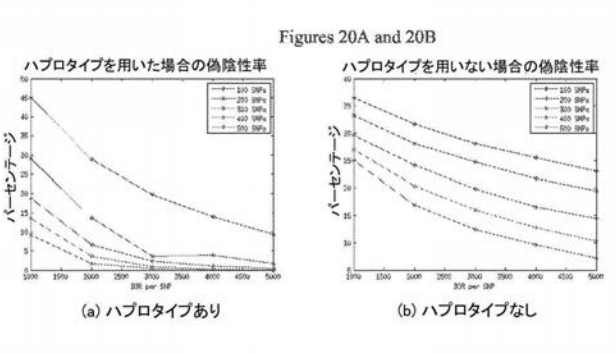
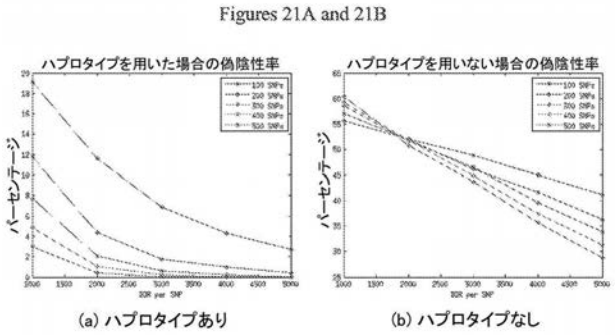


FIG. 19D 娘中0.1%父親

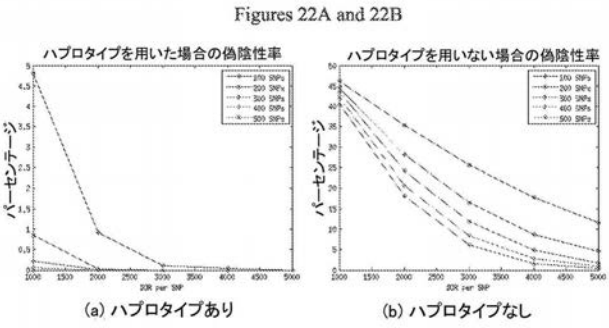
【図２０Ａ－２０Ｂ】



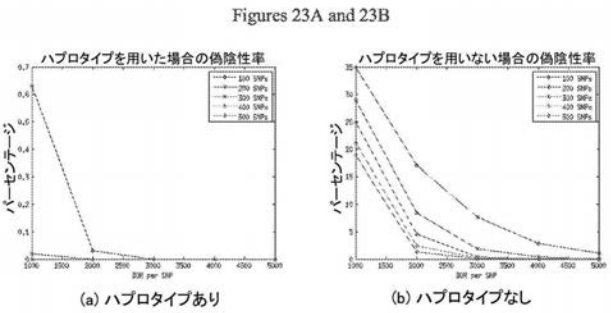
【図２１Ａ－２１Ｂ】



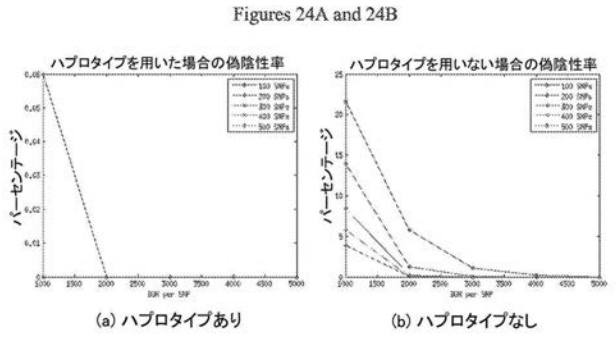
【図２２Ａ－２２Ｂ】



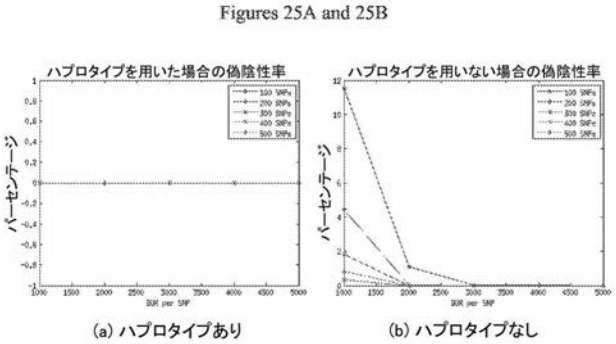
【図２３Ａ－２３Ｂ】



【図２４Ａ－２４Ｂ】



【図２５Ａ－２５Ｂ】



【図２６】

Figure 26

	S = 1000, D = 1000	S = 500, D = 1000	S = 1000, D = 500	S = 500, D = 500
相あり	1.50%	8.80%	10.20%	22.30%
相なし	10.60%	10.80%	17.90%	20.80%

【図２７】

Figure 27

	S = 1000, D = 1000	S = 500, D = 1000	S = 1000, D = 500	S = 500, D = 500
$p = 1\%$, 相あり	0.00%	0.10%	3.00%	9.70%
$p = 1\%$, 相なし	74.20%	73.70%	74.60%	76.00%
$p = 2\%$, 相あり	0.00%	0.00%	0.00%	0.00%
$p = 2\%$, 相なし	13.90%	14.80%	41.00%	42.70%
$p = 3\%$, 相あり	0.00%	0.00%	0.00%	0.00%
$p = 3\%$, 相なし	0.00%	0.00%	5.70%	5.00%
$p = 4\%$, 相あり	0.00%	0.00%	0.00%	0.00%
$p = 4\%$, 相なし	0.00%	0.00%	0.00%	0.00%
$p = 5\%$, 相あり	0.00%	0.00%	0.00%	0.00%
$p = 5\%$, 相なし	0.00%	0.00%	0.00%	0.00%

【図２８Ａ－２８Ｃ】

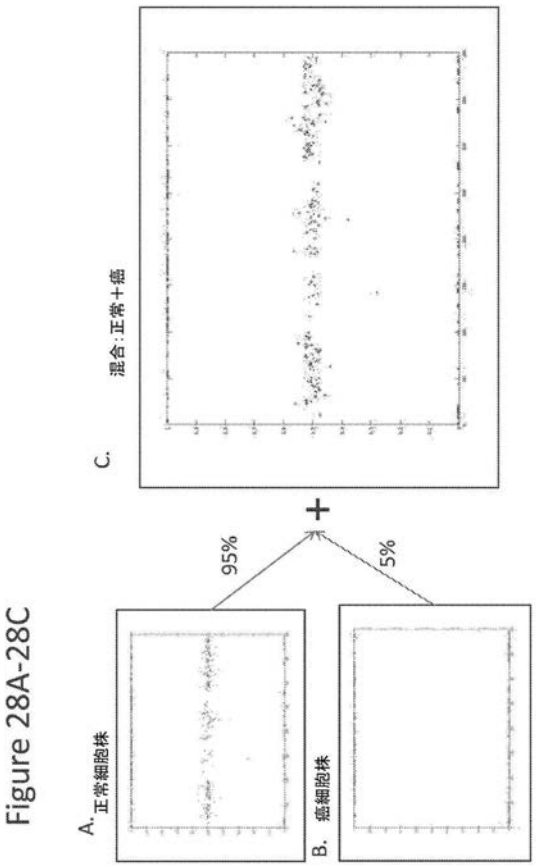
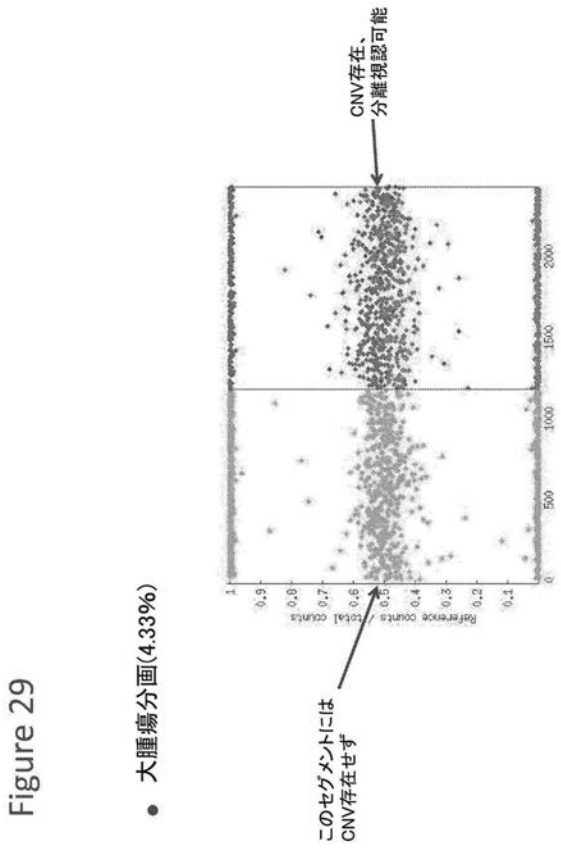


Figure 28A-28C

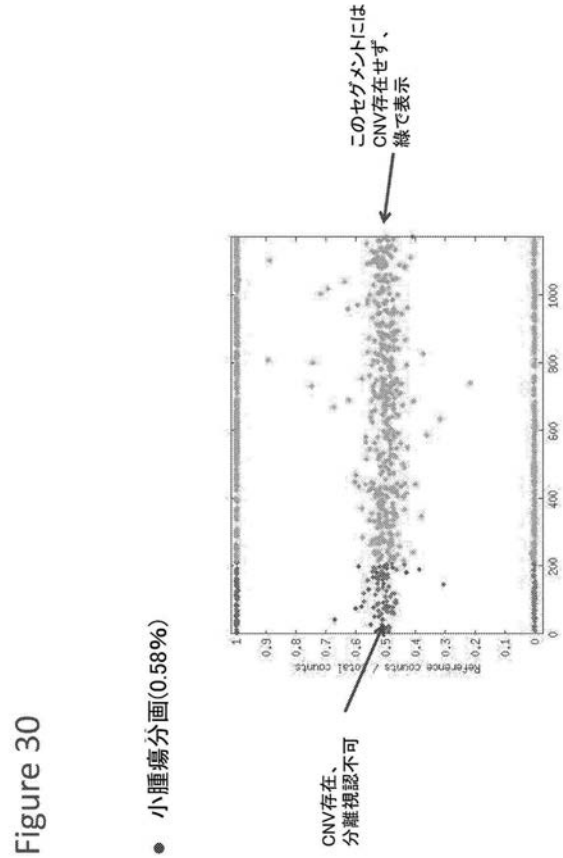
【図２９】

Figure 29

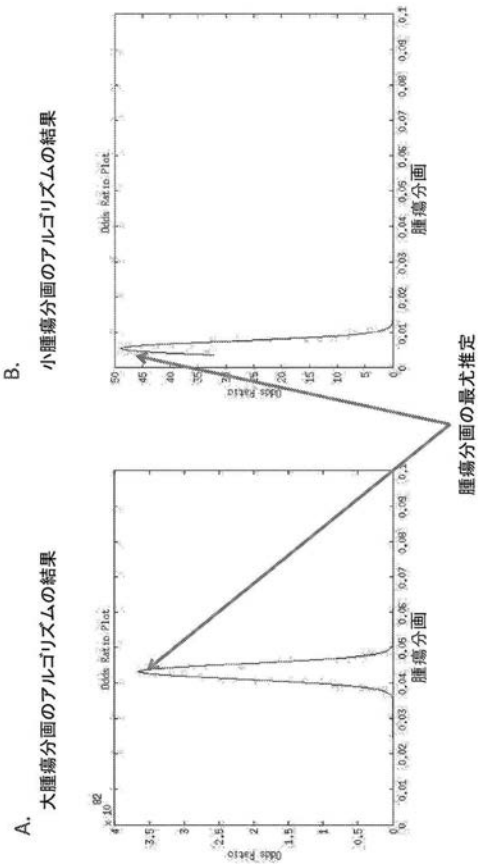


【図３０】

Figure 30

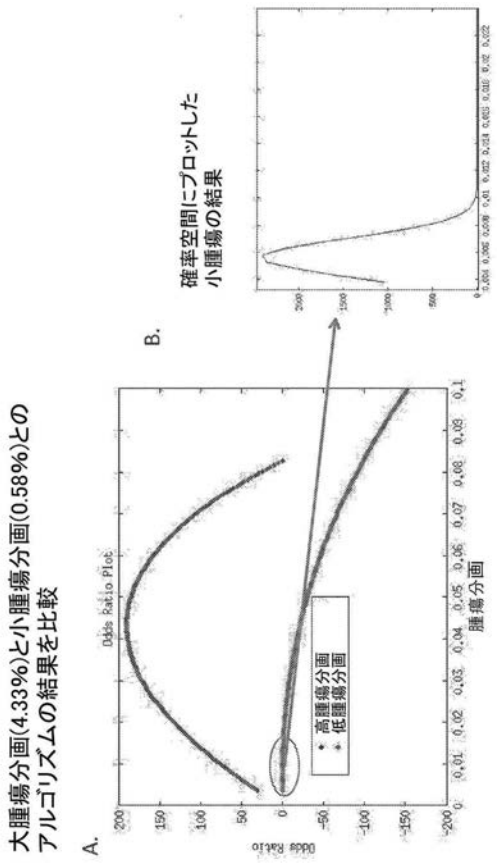


【図３１Ａ－３１Ｂ】



Figures 31A and 31B

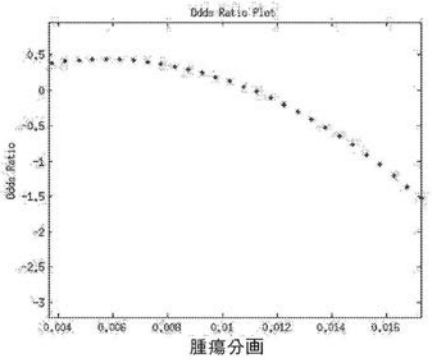
【図３２Ａ－３２Ｂ】



Figures 32A and 32B

【図３３】

Figure 33



【図３４】

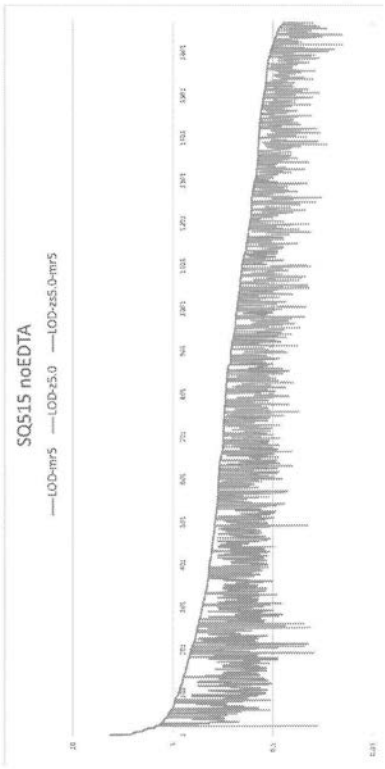


Figure 34

【図３５】

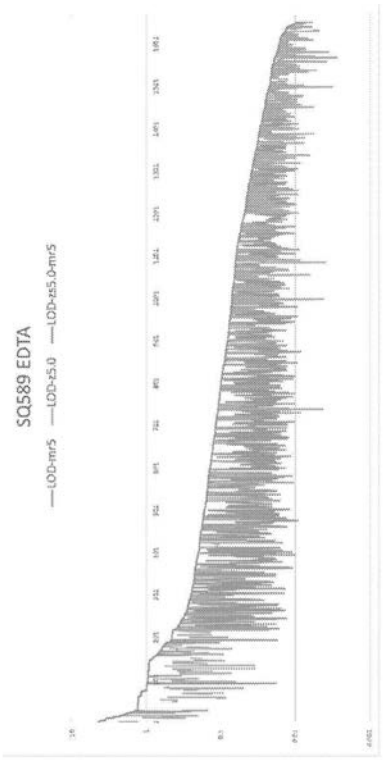
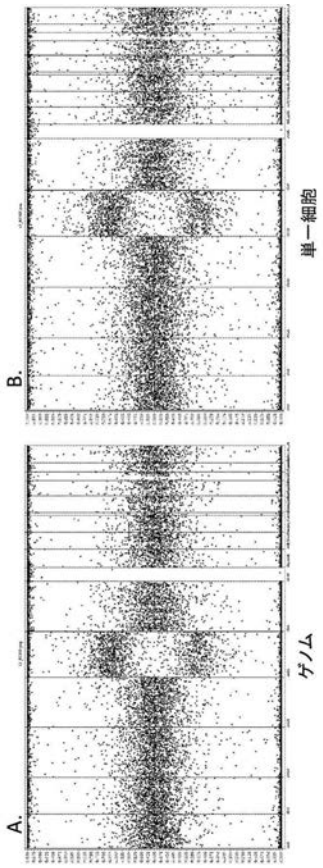


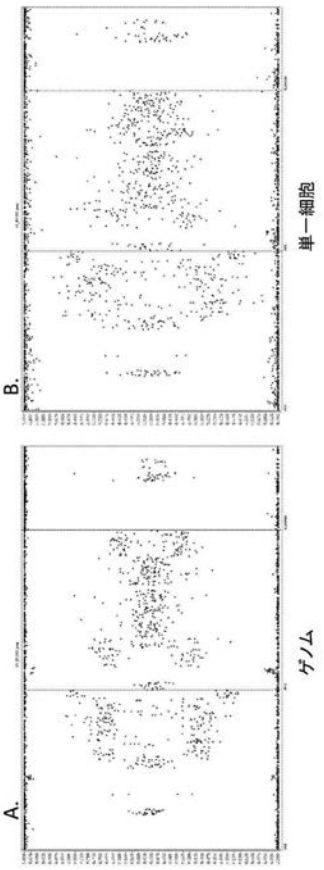
Figure 35

【図３６Ａ－３６Ｂ】



Figures 36A and 36B

【図３７Ａ－３７Ｂ】



Figures 37A and 37B

【図３８】

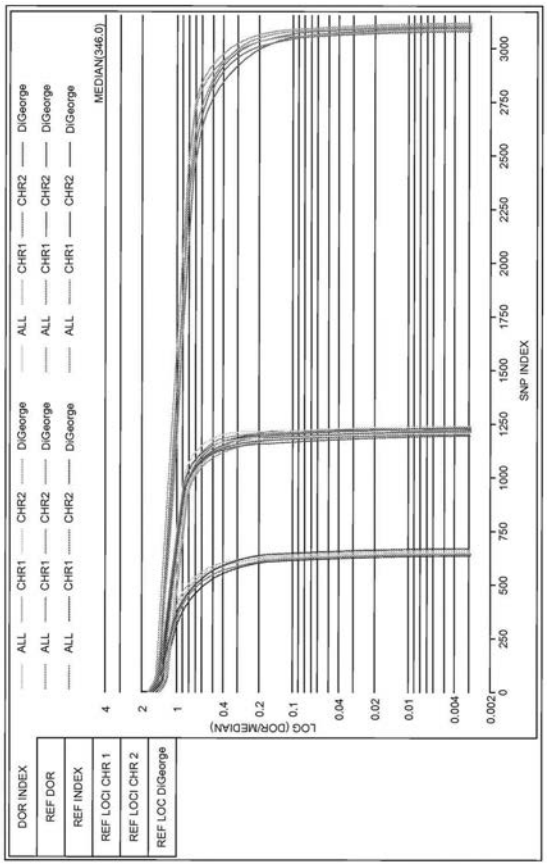


FIG. 38

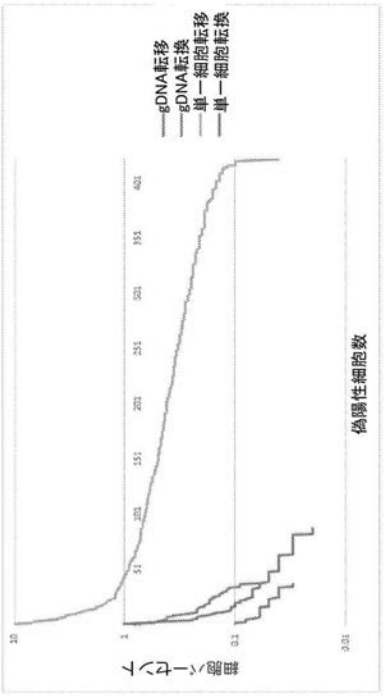
【図３９】

Figure 39

	gDNA	単一細胞
総数	75	510
平均値	0.15 %	0.51 %
中央値	0.09 %	0.33 %
最大値	1.03 %	10 %
標準偏差	0.16%	0.79 %
95パーセンタイル値	0.43 %	1.22 %
90パーセンタイル値	0.37 %	0.92 %

【図４０】

Figure 40



【図４１ａ－４１ｄ】

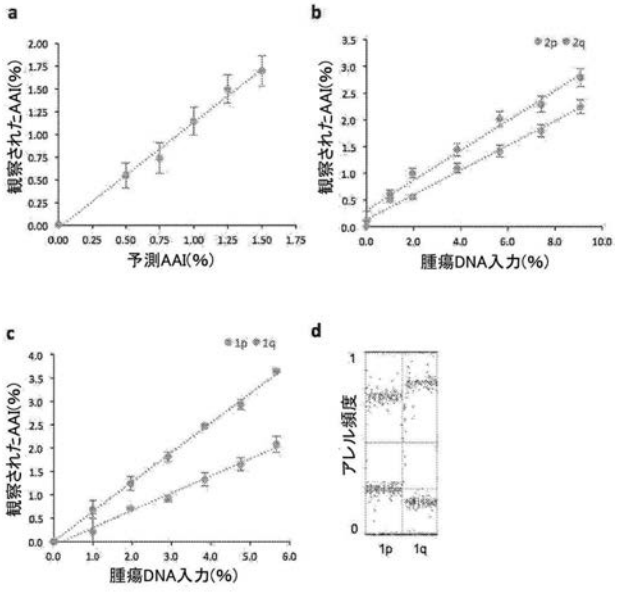


FIG. 41

【図４２】

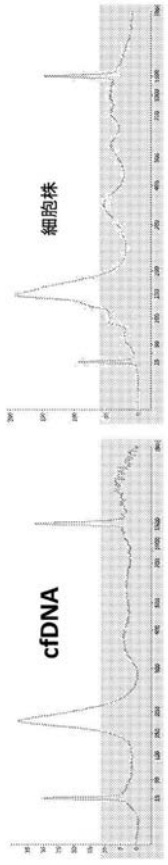


FIG. 42

【図４３Ａ】

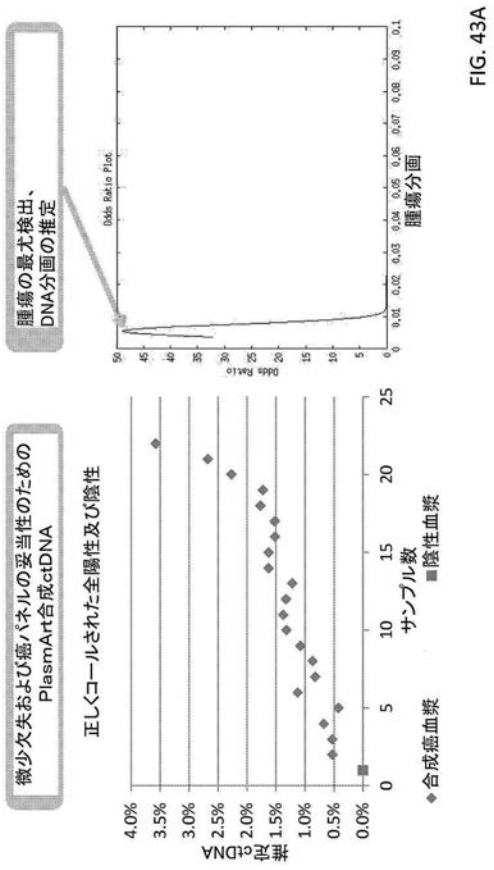
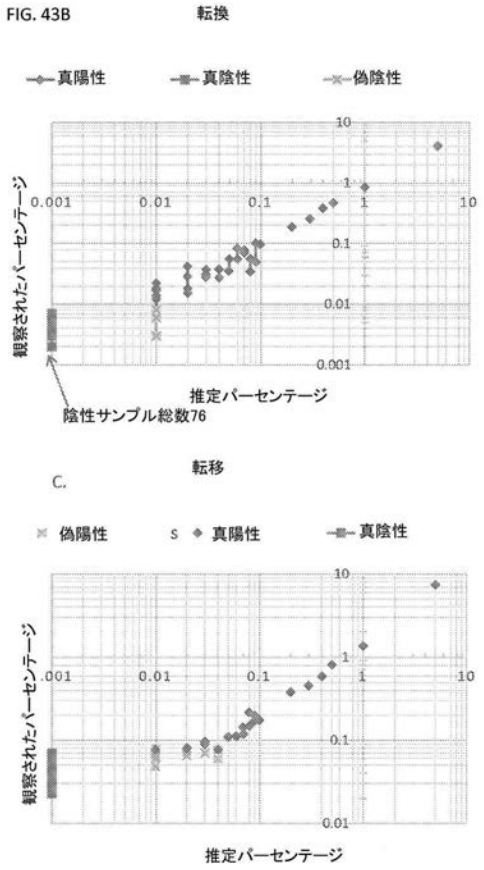


FIG. 43A

【図４３Ｂ－Ｃ】



【図４４】

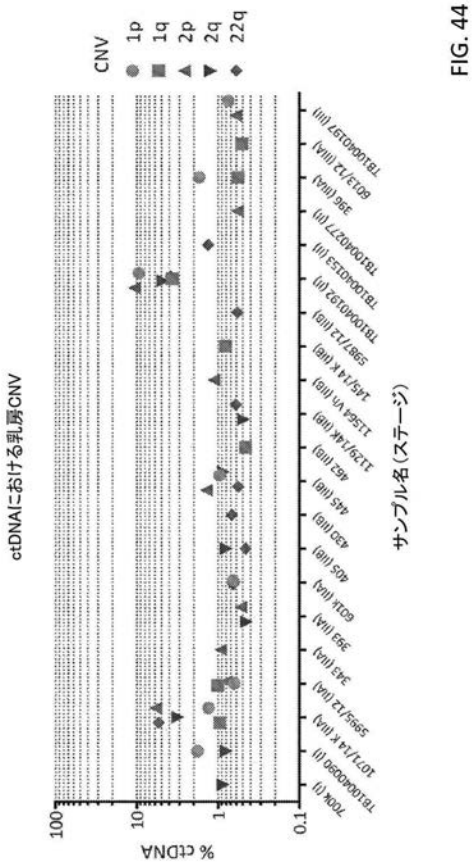


FIG. 44

【図４５】

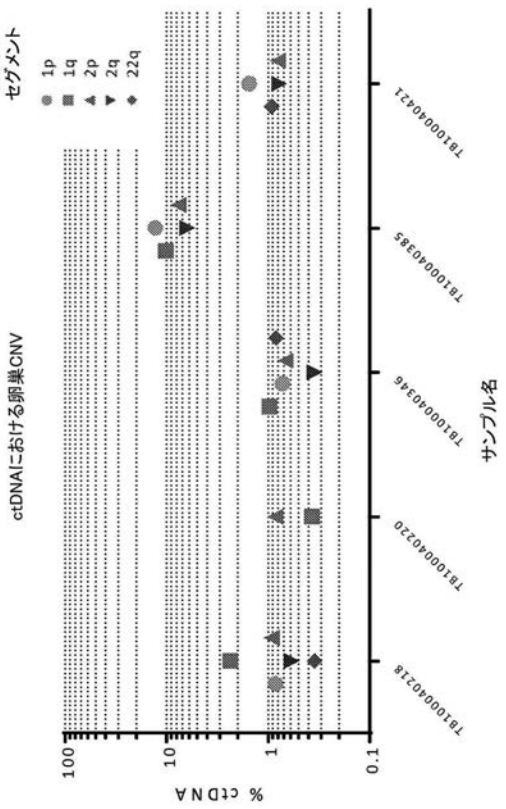


FIG. 45

【図４６】

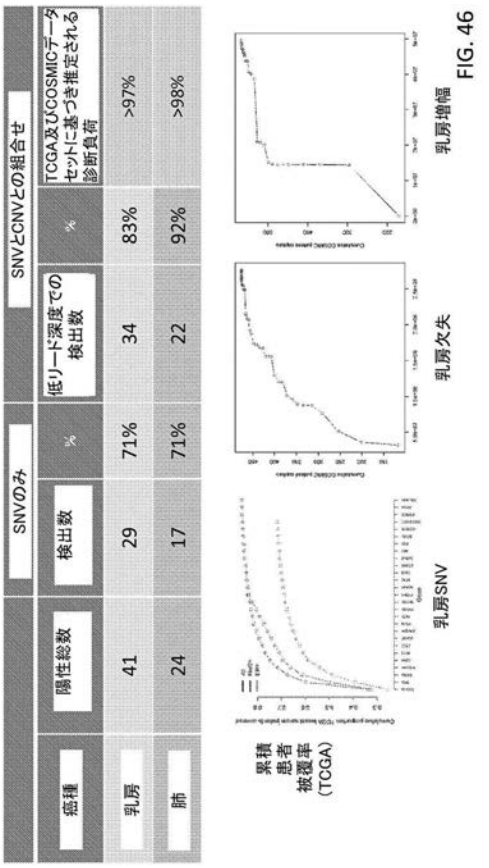


FIG. 46

【図４７】

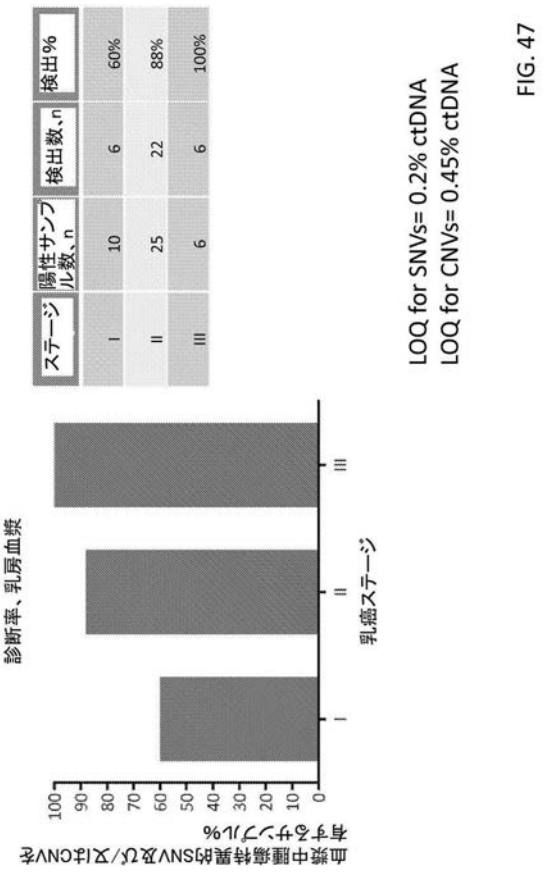


FIG. 47

【図４８】

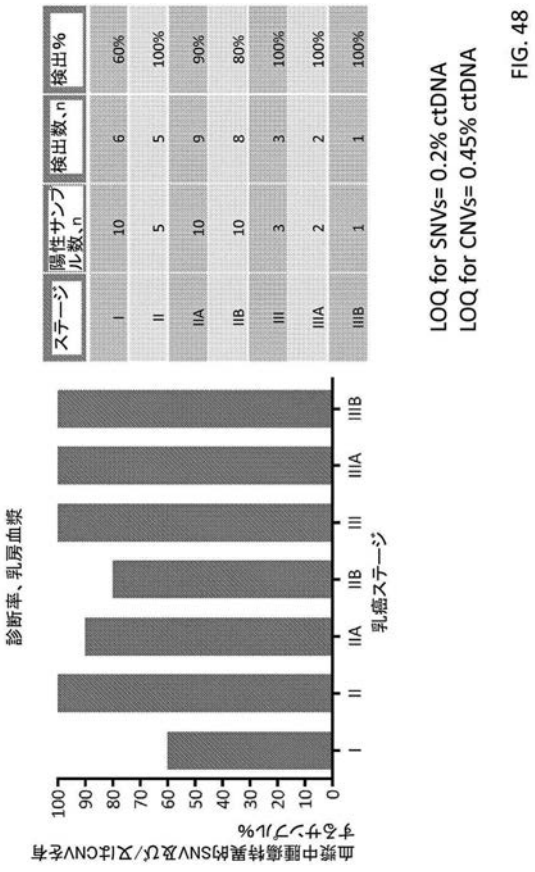


FIG. 48

【図４９】

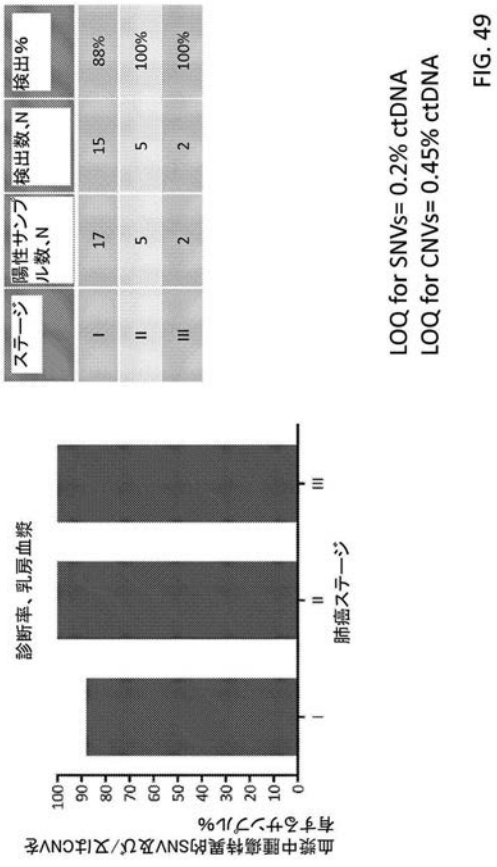


FIG. 49

【図５０】

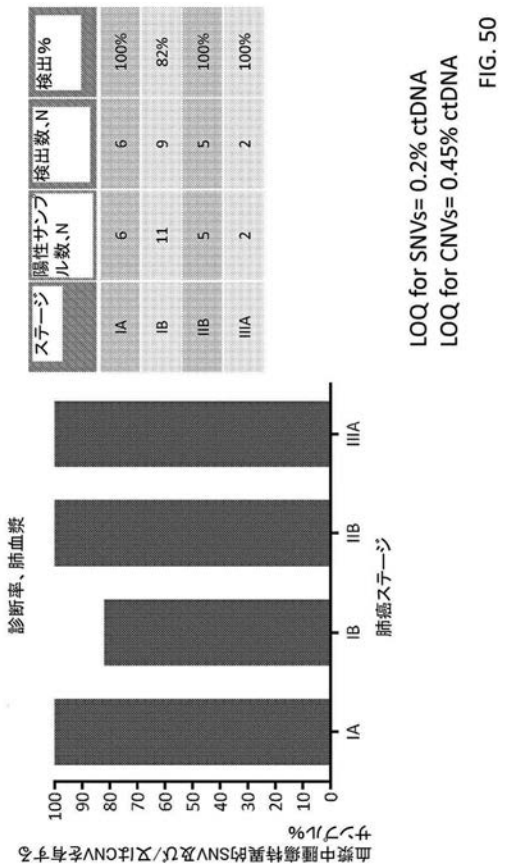


FIG. 50

【図５１Ａ】

Fig. 51A

	組織学	ステージ	年齢	直径(mm)	喫煙者（箱年数）	
L12	SCC	IB	69	40	Yes (?)	
L13	SCC	IA	68	30	Yes (100)	
L15	SCC	IB	68	50	Yes (50)	
L17	Adeno	IIB	61	20	Yes (48)	

【図５１Ｂ】

Fig. 51B

サンプル	遺伝子	Chr	ChrStart	VAF R1	VAF R2	VAF R3
L12	BRIP1	chr17	59924572	14	6	8
L12	CARS	chr11	3062181	22	11	16
L12	CIC	chr19	42797381	0	0	7
L12	CYFIP1	chr15	22940733	0	6	0
L12	FAT1	chr4	187519147	8	4	0
L12	KDM6A	chrX	44921898	10	5	0
L12	MLLT4	chr6	168347475	9	3	0
L12	NFE2L2	chr2	178098801	69	31	50
L12	RASA1	chr5	86642517	7	0	0
L12	TP53	chr17	7578406	23	9	17
L12	TP53	chr17	7578190	22	8	16
L13	EGFR	chr7	55241708	21	25	55
L13	EGFR	chr7	55242511	20	17	48
L13	HERC4	chr10	69793756	0	3	6
L13	JAK2	chr9	5022084	0	11	7
L13	KMT2C	chr7	151947008	0	0	4
L13	MSH2	chr2	47693816	5	0	11
L13	MTOR	chr1	11292495	2	0	3
L13	PLCG2	chr16	81942036	5	0	6
L13	TP53	chr17	7579509	7	4	16
L15	ALK	chr2	29940530	6	2	
L15	GABRG1	chr4	46060315	13	0	
L15	KDM6A	chrX	44922755	18	6	
L15	MLL2	chr12	49443815	5	0	
L15	ROS1	chr6	117687379	10	3	
L15	SLC39A4	chr8	145638322	14	0	
L15	TP53	chr17	7578254	19	3	
L15	ZFHX4	chr8	77776735	13	5	
L15	ZMYM4	chr1	35827319	14	2	
L17	BRCA2	chr13	32914959	16	28	0
L17	KRAS	chr12	25398284	13	22	0
L17	NF1	chr17	29653134	20	25	0
L17	NF1	chr17	29528088	0	0	20
L17	PAX8	chr2	113984793	9	0	0
L17	TP53	chr17	7577610	0	0	16
L17	TP53	chr17	7577079	20	35	0
L17	TRIM67	chr1	231299607	0	50	0
L17	TRIP11	chr14	92471631	7	0	0

【図５２】

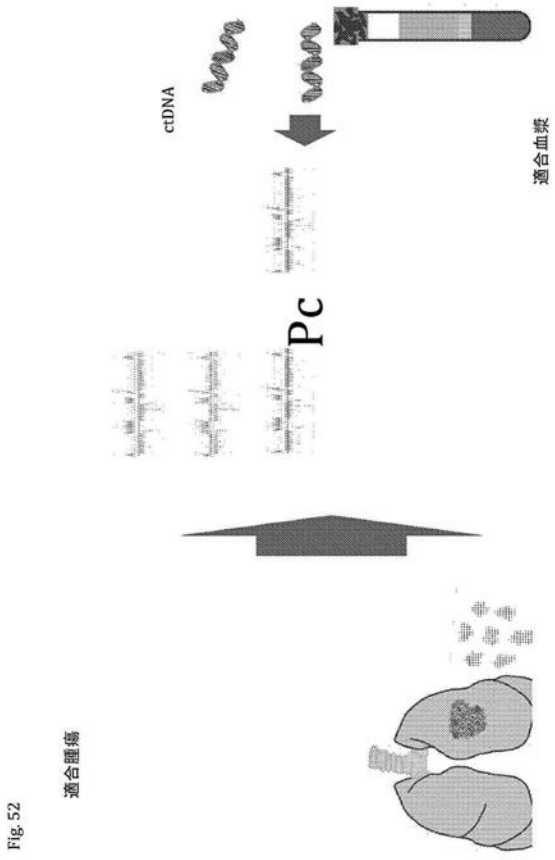


Fig. 52

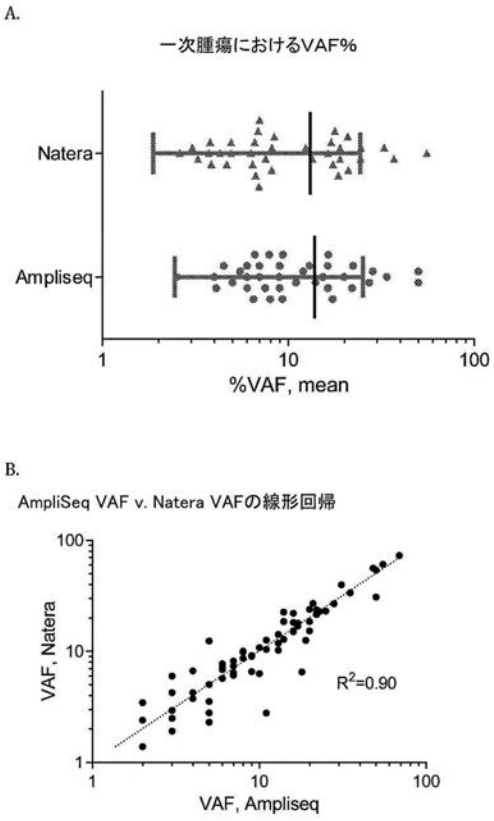
【図５３】

Fig. 53

サンプル	遺伝子	Chrom	chromStart	Swanton			Natera			
				VAF R1	VAF R2	VAF R3	VAF R1	VAF R2	VAF R3	VAF Plasma
L12	BRIP1	chr17	59924572	14	6	8	22.6	7.74	10.07	3.71
L12	CARS	chr11	3062181	22	11	16	22.39	10.47	18.21	5.02
L12	CIC	chr19	42797381	0	0	7	0	0	8.17	1.71
L12	CYFIP1	chr15	22940733	0	6	0	0	0	0	0
L12	FAT1	chr4	187519147	8	4	0	9.81	4.26	0.72	0.57
L12	KDM6A	chrX	44921898	10	5	0	6.33	3.54	0	0.28
L12	MLLT4	chr6	168347475	9	3	0	9.04	4.26	0.73	0.95
L12	NFE2L2	chr2	178098801	69	31	50	73.12	39.92	53.94	23.25
L12	RASA1	chr5	86642517	7	0	0	7.37	0	0.35	0
L12	TP53	chr17	7578406	23	9	17	22.93	9.17	16.97	4.89
L12	TP53	chr17	7578190	22	8	16	21.48	8.67	22.09	5.77
L13	EGFR	chr7	55241708	21	25	55	27.06	23.14	60.79	1.16
L13	EGFR	chr7	55242511	20	17	48	23.93	18.01	56.31	1.09
L13	HERC4	chr10	69793756	6	3	6	0.16	1.92	5.72	0
L13	JAK2	chr9	5022084	8	11	7	0.27	2.79	6.09	0
L13	KMT2C	chr7	151947008	0	0	4	0	0	6.69	0
L13	MSH2	chr2	47693816	5	0	11	5.04	3.73	12.67	0
L13	MTOR	chr1	11292495	2	0	3	2.4	1.38	6.01	0
L13	PLCG2	chr16	81942036	5	0	6	2.8	1.7	6.89	0
L13	TP53	chr17	7579509	7	4	16	6.11	3.76	15.4	0.4
L15	ALK	chr2	29940530	6	2		7.2	1.39		0
L15	GABRG1	chr4	46060315	13	6		11.84	2.08		0
L15	KDM6A	chrX	44922755	18	6		6.54	0.91		0.17
L15	MLL2	chr12	49443815	5	0		12.41	0		
L15	ROS1	chr6	117687379	10	3		10.82	2.94		0.15
L15	SLC39A4	chr8	145638322	14	0		18.59	0		0
L15	TP53	chr17	7578254	19	3		12.6	2.5		0
L15	ZFHX4	chr8	77776735	13	5		10.26	2.3		0
L15	ZMYM4	chr1	35827319	14	2		12.84	3.46		0
L17	BRCA2	chr13	32914959	16	28		15.03	27		0
L17	KRAS	chr12	25398284	13	22		14.22	23.79		0
L17	NF1	chr17	29653134	20	25		18.6	23.1		0
L17	NF1	chr17	29528088	0	0	20	0	0	17.79	0.37
L17	PAX8	chr2	113984793	9	0	0	6.6	0		0
L17	TP53	chr17	7577610	0	0	16	0	0	18.97	0
L17	TP53	chr17	7577079	20	35	0	15.36	33.7		0
L17	TRIM67	chr1	231299607	6	50	0	18.38	30.95		0.57
L17	TRIP11	chr14	92471631	7	0	0	6.44	0		0

【図５４】

Fig. 54



【図５５】

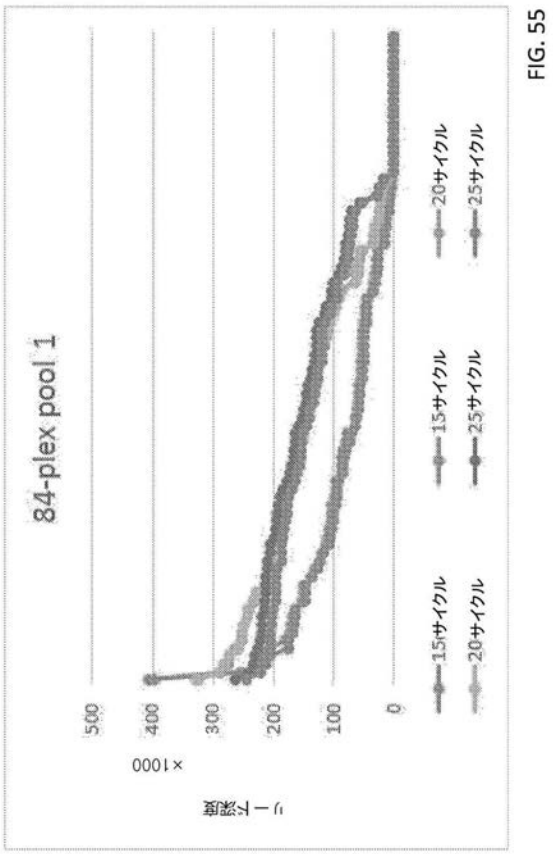


FIG. 55

【図５６】

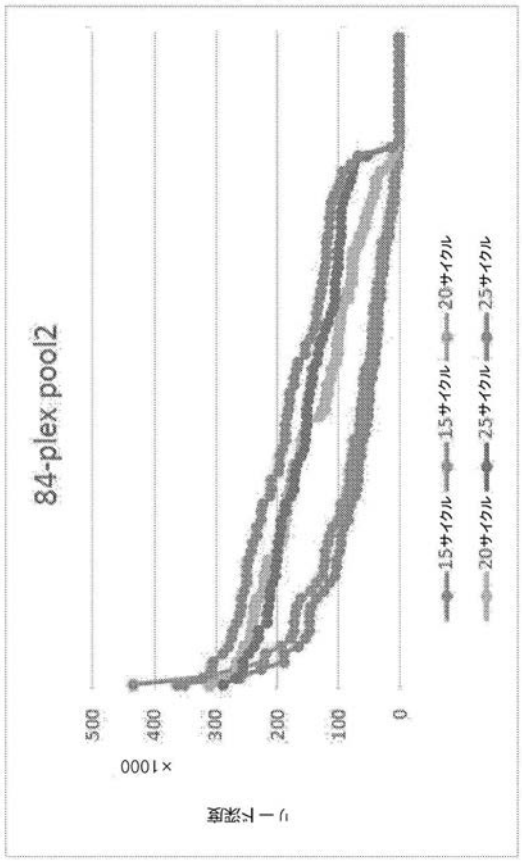


FIG. 56

【図５７】

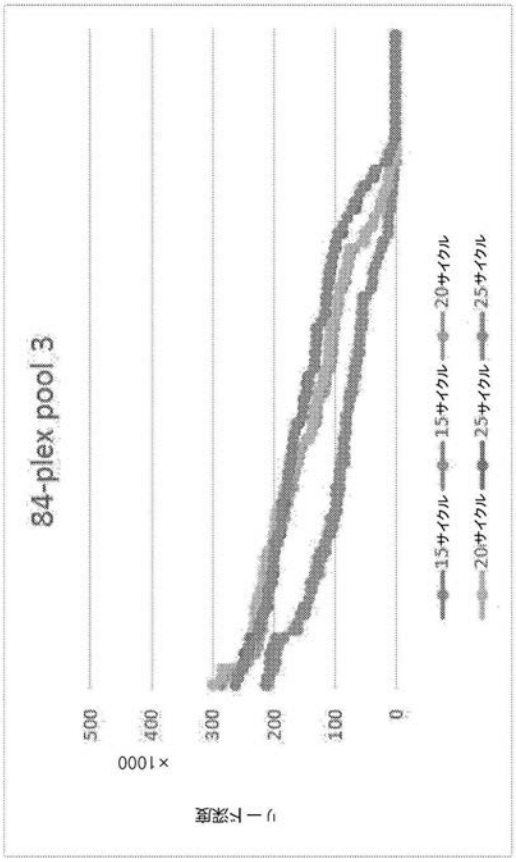


FIG. 57

【図５８】

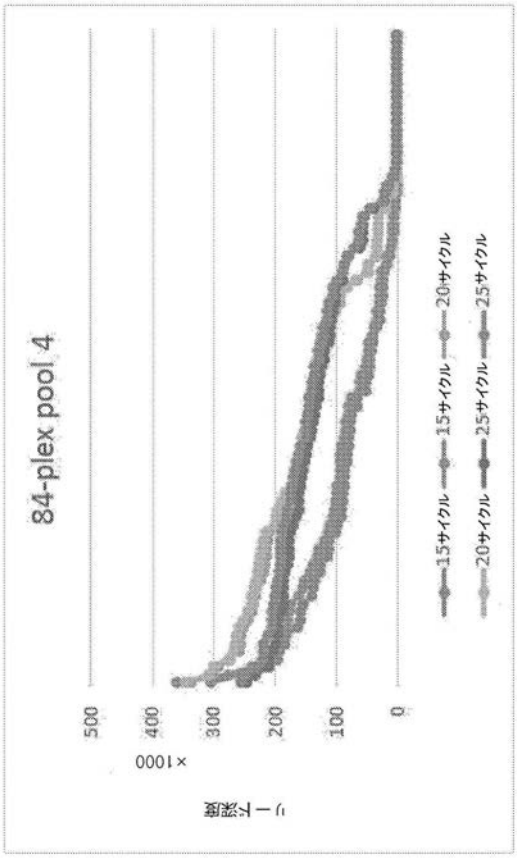


FIG. 58

【図５９】

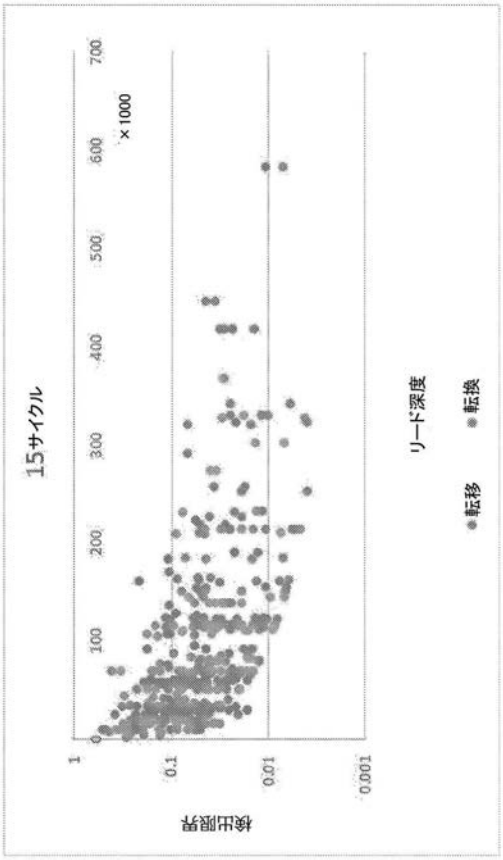


FIG. 59

【図６０】

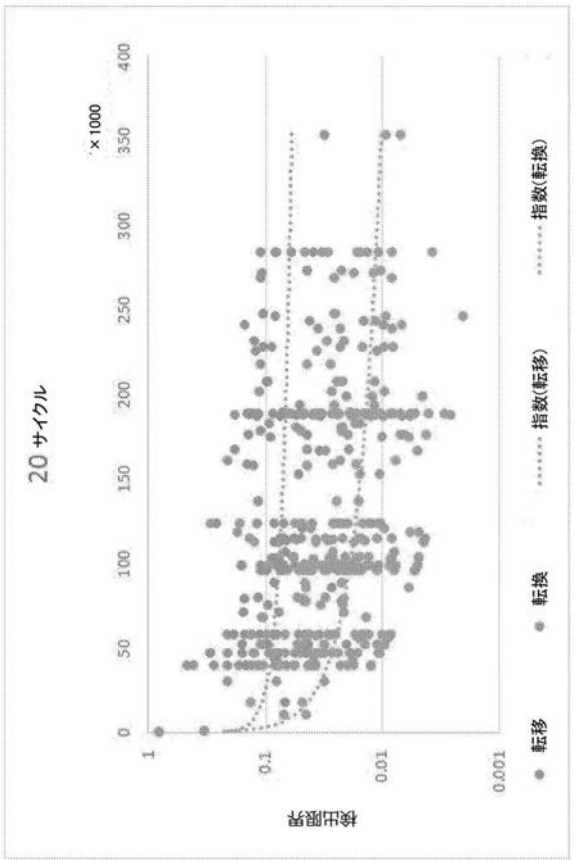


FIG. 60

【図６１】

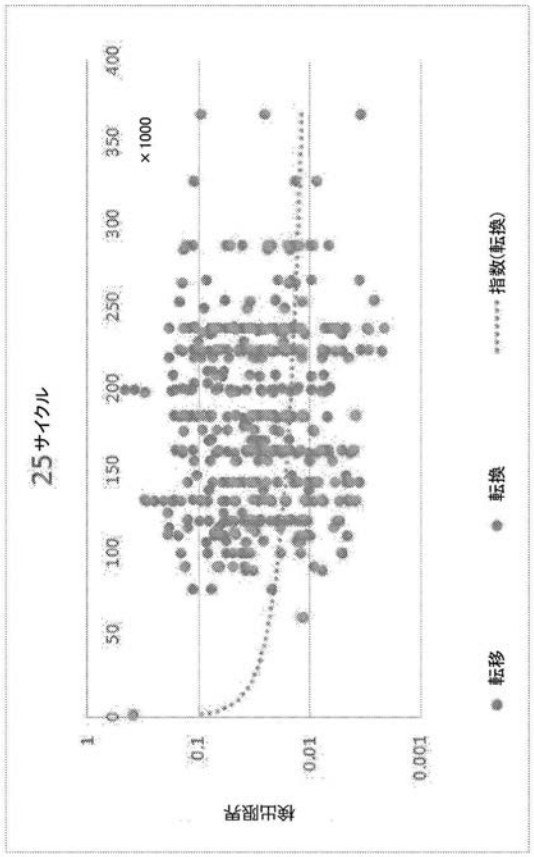


FIG. 61

【図６２】

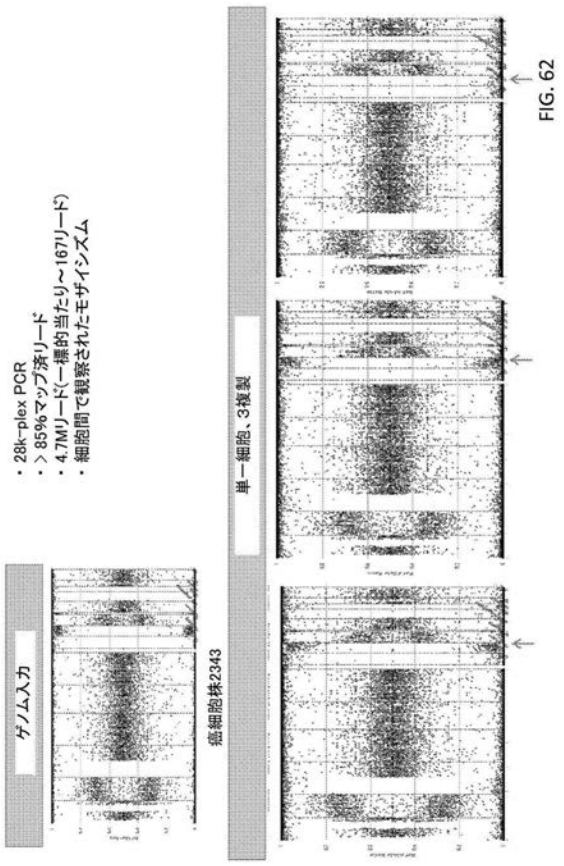


FIG. 62

【図６３】

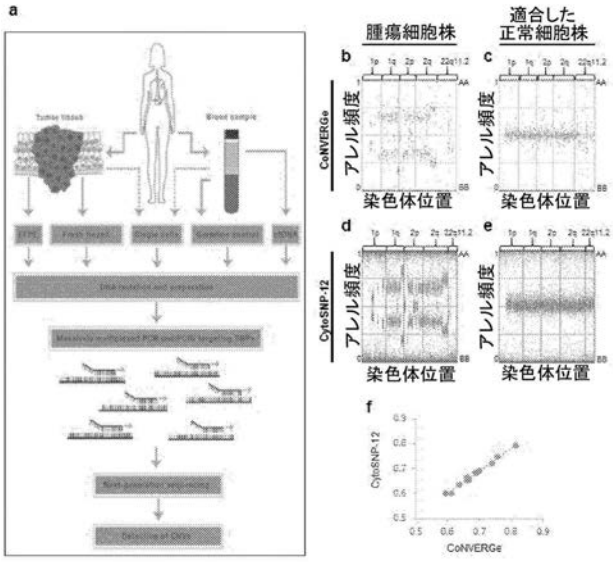


FIG. 63

【図６４】

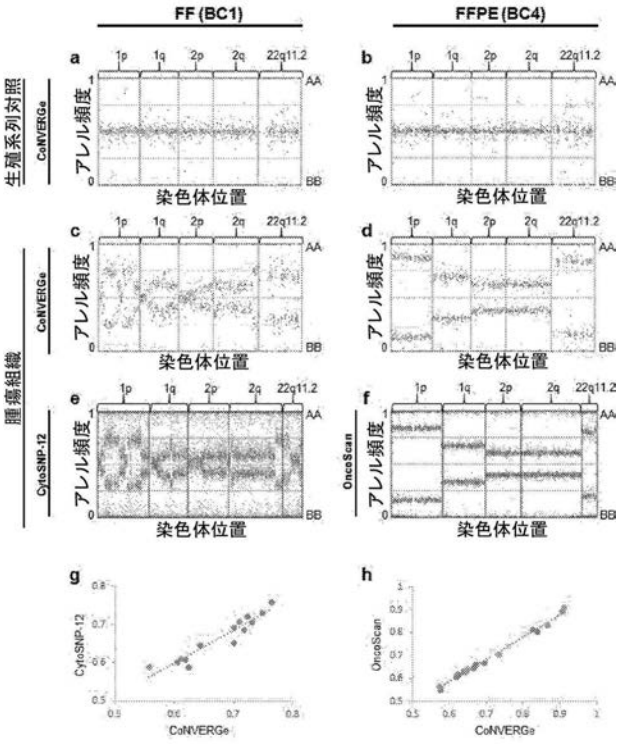


FIG. 64

【図６５】

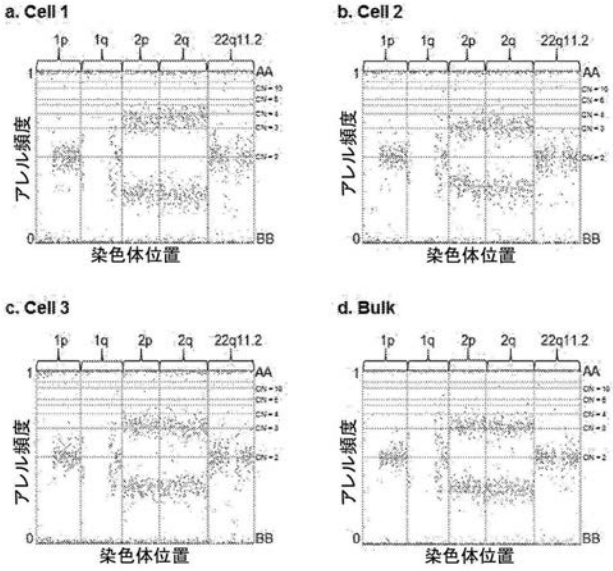


FIG. 65

【図６６】

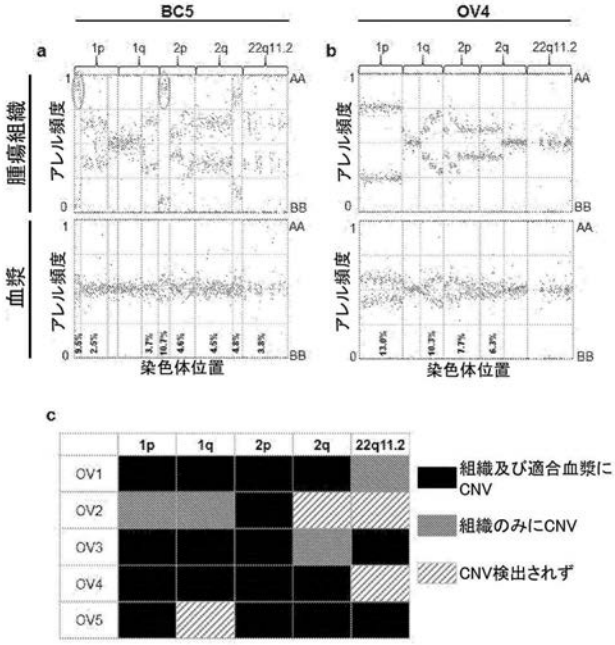


FIG. 66

【図６７－１】

FIG. 67

Chr No.:Position_Mutation	Chr No.:Position_Mutation	Chr No.:Position_Mutation
9:163985-163985_T>C	9:5050714-5050714_T>C	17:7577587-7577587_A>T
1:17275337-17275337_C>T	9:21971111-21971111_C>T	17:7577590-7577590_A>T
1:23689289-23689289_T>G	9:32632175-32632175_C>T	17:7577592-7577592_C>A
1:26774085-26774085_G>A	9:108097968-108097968_G>A	17:7577598-7577598_A>G
1:36226206-36226206_A>G	9:130931780-130931780_C>T	17:7577599-7577599_G>C
1:43032078-43032078_C>T	9:141071078-141071078_A>G	17:7577604-7577604_G>T
1:43912067-43912067_G>A	10:37486216-37486216_C>G	17:7578177-7578177_G>C
1:64515405-64515405_C>T	10:46254776-46254776_A>C	17:7578183-7578183_G>T
1:99771528-99771528_G>A	10:47207813-47207813_A>G	17:7578188-7578188_G>T
1:110173662-110173662_G>A	10:50678650-50678650_A>T	17:7578190-7578190_A>G
1:115256528-115256528_A>C	10:51853633-51853633_C>T	17:7578191-7578191_T>C
1:115256529-115256529_A>G	10:72358324-72358324_C>T	17:7578196-7578196_T>A
1:120458147-120458147_C>T	10:81058322-81058322_C>G	17:7578202-7578202_T>A
1:145323656-145323656_A>T	10:89624275-89624275_C>T	17:7578203-7578203_G>T
1:148343792-148343792_G>T	10:89624305-89624305_T>G	17:7578204-7578204_T>G
1:151149263-151149263_A>G	10:89692790-89692790_G>A	17:7578206-7578206_A>G
1:152186837-152186837_G>A	10:89692794-89692794_A>T	17:7578207-7578207_T>C
1:152189016-152189016_C>T	10:89692835-89692835_G>A	17:7578208-7578208_A>G
1:152328936-152328936_C>G	10:89692839-89692839_T>G	17:7578212-7578212_C>T
1:157805906-157805906_G>A	10:89692883-89692883_C>G	17:7578215-7578215_T>A
1:158609792-158609792_A>T	10:89692893-89692893_C>A	17:7578216-7578216_T>C
1:186276240-186276240_T>C	10:89692900-89692900_G>C	17:7578224-7578224_A>T
1:198675866-198675866_A>C	10:89692905-89692905_G>A	17:7578225-7578225_C>A
1:202699028-202699028_A>G	10:89692923-89692923_G>A	17:7578235-7578235_A>G
1:225306960-225306960_T>G	10:89692980-89692980_A>G	17:7578236-7578236_T>G
1:227288919-227288919_C>T	10:89692998-89692998_G>T	17:7578239-7578239_G>T
1:228560700-228560700_T>C	10:123258034-123258034_T>A	17:7578244-7578244_G>C
1:235884036-235884036_A>G	10:127548395-127548395_C>G	17:7578245-7578245_C>A
1:247614573-247614573_C>T	11:534288-534288_G>A	17:7578253-7578253_G>T
1:248402566-248402566_T>A	11:1093430-1093430_C>A	17:7578254-7578254_G>A
2:32820108-32820108_A>C	11:1643049-1643049_T>G	17:7578257-7578257_G>T
2:61719472-61719472_G>A	11:1651157-1651157_A>G	17:7578259-7578259_T>G
2:107049425-107049425_T>G	11:48171648-48171648_A>T	17:7578260-7578260_G>T
2:107460402-107460402_G>A	11:49854989-49854989_G>A	17:7578262-7578262_G>C
2:118715997-118715997_C>G	11:55541605-55541605_G>A	17:7578263-7578263_C>T
2:119752091-119752091_G>A	11:55861593-55861593_G>A	17:7578265-7578265_T>C
2:129075877-129075877_C>A	11:60666746-60666746_A>C	17:7578266-7578266_A>T
2:165986535-165986535_A>G	11:64544046-64544046_A>G	17:7578268-7578268_T>C
2:178098803-178098803_G>A	11:66335548-66335548_T>C	17:7578269-7578269_C>T
2:179434772-179434772_C>T	11:71932638-71932638_C>T	17:7578271-7578271_A>G
2:179635206-179635206_C>G	11:94204875-94204875_T>G	17:7578272-7578272_C>T
2:197641324-197641324_C>G	11:103182692-103182692_G>A	17:7578275-7578275_C>T
2:198266834-198266834_A>G	11:108117798-108117798_C>T	17:7578280-7578280_C>T
2:198363501-198363501_G>A	11:108175462-108175462_G>A	17:7578281-7578281_C>T
2:209113113-209113113_C>T	11:108200961-108200961_G>A	17:7578284-7578284_G>T
2:238672328-238672328_C>T	12:4627253-4627253_G>A	17:7578371-7578371_G>A
3:1269641-1269641_A>G	12:9453702-9453702_C>T	17:7578374-7578374_G>C
3:19924193-19924193_G>A	12:9574020-9574020_A>G	17:7578375-7578375_C>G

【図６７－２】

Chr No.:Position_Mutation	Chr No.:Position_Mutation	Chr No.:Position_Mutation
3:25832827-25832827_T>A	12:9583286-9583286_T>C	17:7578380-7578380_G>A
3:38104250-38104250_C>G	12:12630318-12630318_C>T	17:7578382-7578382_C>T
3:41266100-41266100_T>A	12:21953991-21953991_G>A	17:7578388-7578388_G>C
3:41266101-41266101_C>G	12:22677465-22677465_G>A	17:7578389-7578389_C>T
3:41266104-41266104_G>T	12:25380275-25380275_A>T	17:7578392-7578392_G>C
3:41266113-41266113_C>A	12:25380276-25380276_A>T	17:7578393-7578393_T>A
3:41266124-41266124_A>G	12:25380277-25380277_C>A	17:7578394-7578394_A>G
3:44612650-44612650_G>T	12:25398281-25398281_G>A	17:7578395-7578395_C>A
3:58191274-58191274_G>C	12:25398282-25398282_G>T	17:7578400-7578400_C>G
3:108788596-108788596_C>T	12:25398284-25398284_G>A	17:7578402-7578402_C>A
3:138022433-138022433_C>T	12:25398285-25398285_G>C	17:7578403-7578403_G>A
3:149245633-149245633_T>G	12:30888067-30888067_G>A	17:7578404-7578404_T>A
3:172533490-172533490_A>C	12:31231425-31231425_T>G	17:7578406-7578406_G>A
3:178916854-178916854_G>A	12:31242869-31242869_G>A	17:7578407-7578407_C>A
3:178916876-178916876_G>A	12:50745677-50745677_A>T	17:7578409-7578409_G>A
3:178916936-178916936_G>A	12:50746243-50746243_A>C	17:7578410-7578410_A>T
3:178916941-178916941_G>A	12:52385715-52385715_C>T	17:7578413-7578413_G>T
3:178916944-178916944_A>G	12:56482341-56482341_G>T	17:7578415-7578415_T>A
3:178916946-178916946_G>C	12:91502039-91502039_G>A	17:7578416-7578416_G>A
3:178917478-178917478_G>A	12:101745882-101745882_A>T	17:7578419-7578419_G>A
3:178921548-178921548_G>A	13:28919595-28919595_G>A	17:7578420-7578420_G>A
3:178921552-178921552_A>T	13:32968854-32968854_C>A	17:7578423-7578423_G>A
3:178921553-178921553_T>A	13:36700097-36700097_C>T	17:7578428-7578428_C>T
3:178927974-178927974_G>A	13:48953760-48953760_C>T	17:7578431-7578431_C>T
3:178927980-178927980_T>C	13:49033955-49033955_A>T	17:7578433-7578433_C>G
3:178928079-178928079_G>A	14:20215706-20215706_G>T	17:7578437-7578437_C>T
3:178936050-178936050_T>C	14:21873400-21873400_G>A	17:7578440-7578440_A>G
3:178936074-178936074_C>G	14:44973867-44973867_C>T	17:7578442-7578442_A>C
3:178936082-178936082_G>A	14:63784405-63784405_C>T	17:7578443-7578443_T>A
3:178936083-178936083_A>T	14:102323093-102323093_T>C	17:7578448-7578448_C>A
3:178936091-178936091_G>A	14:105246551-105246551_G>A	17:7578449-7578449_G>A
3:178936092-178936092_A>C	15:22743235-22743235_A>G	17:7578452-7578452_A>G
3:178936093-178936093_G>C	15:33954810-33954810_G>A	17:7578454-7578454_C>T
3:178936094-178936094_C>A	15:66774131-66774131_G>A	17:7578455-7578455_G>C
3:178936095-178936095_A>G	15:75047325-75047325_G>A	17:7578457-7578457_G>A
3:178936097-178936097_G>A	15:83014132-83014132_G>C	17:7578460-7578460_T>G
3:178937019-178937019_A>G	15:84909434-84909434_G>A	17:7578461-7578461_G>T
3:178938860-178938860_A>C	15:102516424-102516424_G>T	17:7578463-7578463_G>C
3:178938934-178938934_G>A	16:1004605-1004605_T>C	17:7578464-7578464_C>G
3:178947827-178947827_G>T	16:2049640-2049640_T>C	17:7578466-7578466_C>T
3:178951957-178951957_G>A	16:3820773-3820773_C>T	17:7578467-7578467_A>C
3:178951964-178951964_G>C	16:4432029-4432029_A>C	17:7578469-7578469_G>A
3:178952004-178952004_C>T	16:28507445-28507445_G>C	17:7578471-7578471_C>T
3:178952013-178952013_G>T	16:29110458-29110458_T>C	17:7578472-7578472_C>T
3:178952018-178952018_A>K	16:56782199-56782199_G>A	17:7578475-7578475_C>T
3:178952019-178952019_C>T	16:68772218-68772218_C>T	17:7578476-7578476_C>T
3:178952020-178952020_C>T	16:68844139-68844139_G>A	17:7578478-7578478_C>A
3:178952030-178952030_G>C	17:7573982-7573982_G>T	17:7578479-7578479_C>T
3:178952072-178952072_A>G	17:7574002-7574002_G>C	17:7578490-7578490_T>A

【図６７－３】

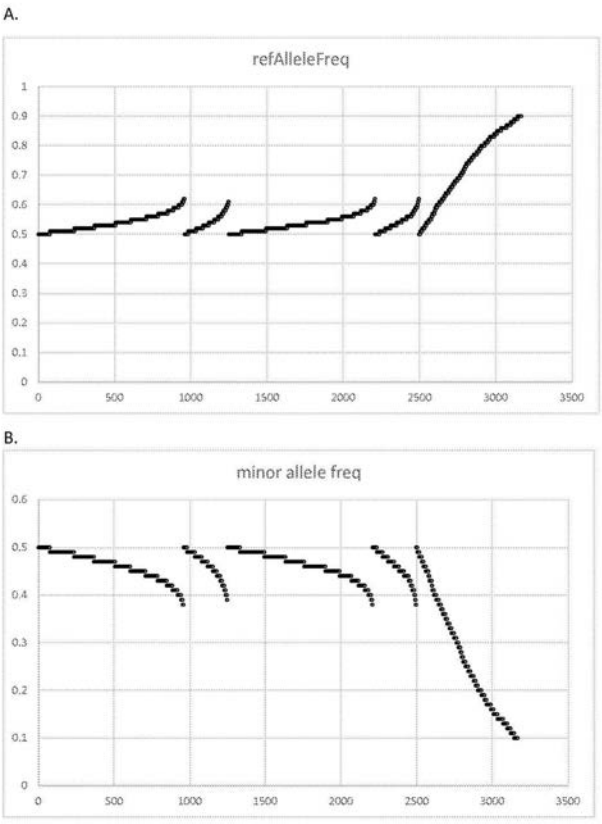
Chr No.:Position_Mutation	Chr No.:Position_Mutation	Chr No.:Position_Mutation
3:178952074-178952074_G>C	17:7574003-7574003_C>T	17:7578492-7578492_G>A
3:178952075-178952075_A>T	17:7574012-7574012_G>T	17:7578493-7578493_G>A
3:178952077-178952077_T>A	17:7574017-7574017_G>T	17:7578496-7578496_T>A
3:178952081-178952081_G>A	17:7574018-7574018_C>T	17:7578498-7578498_G>T
3:178952082-178952082_C>T	17:7576855-7576855_C>T	17:7578499-7578499_A>C
3:178952084-178952084_C>T	17:7576857-7576857_T>G	17:7578500-7578500_C>T
3:178952085-178952085_A>G	17:7576883-7576883_A>G	17:7578502-7578502_T>C
3:178952088-178952088_A>G	17:7576897-7576897_C>T	17:7578506-7578506_C>A
3:178952090-178952090_G>C	17:7577022-7577022_C>T	17:7578507-7578507_C>A
3:178952091-178952091_G>C	17:7577036-7577036_C>A	17:7578508-7578508_G>A
3:178952100-178952100_C>A	17:7577046-7577046_G>T	17:7578509-7578509_T>C
3:192516720-192516720_C>G	17:7577052-7577052_C>A	17:7578513-7578513_G>A
3:195452649-195452649_C>T	17:7577058-7577058_G>T	17:7578514-7578514_A>G
3:195505836-195505836_C>G	17:7577069-7577069_G>A	17:7578517-7578517_C>T
3:195506597-195506597_C>T	17:7577079-7577079_G>T	17:7578518-7578518_G>A
3:195506940-195506940_C>G	17:7577081-7577081_A>C	17:7578524-7578524_C>T
3:195511945-195511945_C>T	17:7577082-7577082_G>C	17:7578525-7578525_C>G
4:11400898-11400898_G>A	17:7577085-7577085_G>A	17:7578526-7578526_G>A
4:15443812-15443812_C>A	17:7577088-7577088_A>C	17:7578527-7578527_T>G
4:40356408-40356408_C>A	17:7577090-7577090_G>C	17:7578528-7578528_T>G
4:46930361-46930361_A>C	17:7577091-7577091_C>T	17:7578529-7578529_T>G
4:48597671-48597671_G>A	17:7577093-7577093_G>A	17:7578530-7578530_T>C
4:55593613-55593613_T>A	17:7577094-7577094_C>T	17:7578532-7578532_T>A
4:55593661-55593661_T>C	17:7577095-7577095_C>A	17:7578534-7578534_G>T
4:55594212-55594212_T>C	17:7577096-7577096_A>G	17:7578535-7578535_A>G
4:55594221-55594221_A>G	17:7577097-7577097_G>A	17:7578536-7578536_A>G
4:55599320-55599320_G>C	17:7577099-7577099_G>A	17:7578541-7578541_T>G
4:71469002-71469002_C>G	17:7577100-7577100_A>T	17:7578542-7578542_C>T
4:83778206-83778206_C>T	17:7577102-7577102_G>A	17:7578548-7578548_C>T
4:88537243-88537243_C>A	17:7577105-7577105_C>G	17:7578550-7578550_C>G
4:88537249-88537249_T>C	17:7577106-7577106_C>A	17:7578551-7578551_T>A
4:107845202-107845202_G>A	17:7577107-7577107_T>G	17:7578553-7578553_A>G
4:144336805-144336805_A>G	17:7577108-7577108_G>T	17:7578554-7578554_T>G
4:153245446-153245446_C>T	17:7577111-7577111_C>T	17:7579312-7579312_G>C
4:153247168-153247168_A>G	17:7577112-7577112_G>C	17:7579313-7579313_C>G
4:153247288-153247288_G>H	17:7577113-7577113_T>G	17:7579329-7579329_A>G
4:153247366-153247366_G>T	17:7577114-7577114_G>T	17:7579349-7579349_T>G
4:153249456-153249456_G>T	17:7577115-7577115_T>G	17:7579355-7579355_T>C
4:164507069-164507069_C>T	17:7577117-7577117_T>C	17:7579358-7579358_G>C
4:170321765-170321765_C>T	17:7577118-7577118_G>T	17:7579366-7579366_C>G
4:175896768-175896768_C>T	17:7577120-7577120_G>A	17:7579377-7579377_C>T
5:13762932-13762932_G>A	17:7577121-7577121_C>T	17:7579378-7579378_C>A
5:26906161-26906161_G>A	17:7577123-7577123_T>C	17:7579414-7579414_G>A
5:67591128-67591128_G>C	17:7577124-7577124_G>T	17:7579503-7579503_G>T
5:112173917-112173917_C>T	17:7577127-7577127_G>T	17:7579521-7579521_G>A
5:112175118-112175118_C>T	17:7577129-7577129_T>G	17:7579547-7579547_C>T
5:112175219-112175219_A>T	17:7577130-7577130_T>C	17:7579882-7579882_G>C
5:112175681-112175681_G>T	17:7577138-7577138_G>A	17:11554600-11554600_G>T
5:112175786-112175786_G>T	17:7577141-7577141_G>A	17:11998898-11998898_C>T

【図６７－４】

Chr No.:Position_Mutation	Chr No.:Position_Mutation	Chr No.:Position_Mutation
5:112176020-112176020_G>T	17:7577142-7577142_G>T	17:12011144-12011144_C>T
5:126791225-126791225_C>T	17:7577144-7577144_T>C	17:37868208-37868208_C>T
5:129030517-129030517_C>T	17:7577148-7577148_C>G	17:37880220-37880220_T>C
5:131931452-131931452_A>T	17:7577501-7577501_C>T	17:37880261-37880261_G>T
5:135394826-135394826_G>A	17:7577505-7577505_A>T	17:37881000-37881000_G>C
5:140683389-140683389_C>T	17:7577506-7577506_G>T	17:37881332-37881332_G>A
5:170239186-170239186_A>G	17:7577508-7577508_A>G	17:39240627-39240627_T>C
5:180335598-180335598_T>G	17:7577509-7577509_G>A	17:39274087-39274087_C>G
6:10756728-10756728_C>T	17:7577511-7577511_T>A	17:39274157-39274157_C>T
6:26032069-26032069_G>A	17:7577517-7577517_T>C	17:39305800-39305800_A>T
6:28497279-28497279_G>A	17:7577518-7577518_A>T	17:39595484-39595484_C>T
6:29910622-29910622_C>T	17:7577520-7577520_T>C	17:39673185-39673185_G>A
6:93956553-93956553_C>G	17:7577521-7577521_A>G	17:41245274-41245274_G>C
6:112452259-112452259_G>A	17:7577524-7577524_A>C	17:45219311-45219311_A>G
6:130497111-130497111_C>T	17:7577526-7577526_T>C	17:48777241-48777241_T>A
6:150001395-150001395_C>T	17:7577529-7577529_T>A	17:58288421-58288421_G>A
6:170871013-170871013_A>G	17:7577532-7577532_C>A	17:61958402-61958402_C>T
7:2577781-2577781_A>G	17:7577533-7577533_C>G	18:29470816-29470816_G>A
7:6426892-6426892_C>T	17:7577534-7577534_G>A	18:45368211-45368211_C>G
7:27832791-27832791_A>G	17:7577535-7577535_G>C	18:48575671-48575671_C>G
7:43591930-43591930_G>A	17:7577536-7577536_A>T	18:48604754-48604754_G>T
7:55259482-55259482_C>A	17:7577538-7577538_G>A	19:1223125-1223125_C>G
7:55259515-55259515_T>G	17:7577539-7577539_C>T	19:22836805-22836805_G>A
7:77379331-77379331_T>G	17:7577541-7577541_A>T	19:40383905-40383905_C>T
7:100643088-100643088_A>G	17:7577543-7577543_G>A	19:49926533-49926533_C>G
7:100677279-100677279_G>C	17:7577544-7577544_T>G	19:51274851-51274851_A>C
7:138268672-138268672_C>T	17:7577545-7577545_A>G	19:56369444-56369444_G>A
7:140453136-140453136_T>A	17:7577547-7577547_G>T	20:3209652-3209652_T>G
7:140453137-140453137_G>A	17:7577548-7577548_G>A	20:6751034-6751034_A>G
7:140481411-140481411_G>T	17:7577549-7577549_C>G	20:29623210-29623210_C>T
7:140481412-140481412_G>C	17:7577550-7577550_G>T	20:29623224-29623224_C>A
7:140481417-140481417_G>T	17:7577551-7577551_G>C	20:29628236-29628236_G>C
7:140500165-140500165_T>C	17:7577552-7577552_G>C	20:29628263-29628263_A>G
7:151970859-151970859_G>A	17:7577554-7577554_A>T	20:29632643-29632643_T>C
7:151970951-151970951_G>A	17:7577555-7577555_C>A	20:31671497-31671497_C>T
8:2796250-2796250_C>T	17:7577556-7577556_G>A	20:34242060-34242060_G>A
8:2855562-2855562_A>G	17:7577557-7577557_T>A	20:57484420-57484420_C>T
8:3009020-3009020_C>G	17:7577559-7577559_C>A	20:57484421-57484421_G>A
8:3205603-3205603_G>A	17:7577565-7577565_A>C	21:10951281-10951281_G>A
8:3245034-3245034_G>T	17:7577566-7577566_A>G	21:32127654-32127654_G>A
8:10467948-10467948_C>T	17:7577568-7577568_G>A	22:16449539-16449539_T>C
8:77768353-77768353_C>T	17:7577569-7577569_T>A	22:22899234-22899234_T>C
8:89086854-89086854_G>A	17:7577570-7577570_G>A	22:29092948-29092948_C>T
8:101083610-101083610_T>C	17:7577571-7577571_T>A	22:30398972-30398972_G>C
8:109796543-109796543_G>A	17:7577574-7577574_A>G	23:15821891-15821891_C>G
8:116426722-116426722_G>T	17:7577575-7577575_T>A	23:66765161-66765161_A>T
8:121344960-121344960_G>A	17:7577577-7577577_A>G	23:78011285-78011285_G>A
8:144940290-144940290_G>C	17:7577580-7577580_A>G	23:105153675-105153675_C>A
8:144940331-144940331_G>A	17:7577581-7577581_T>A	17:7577586-7577586_T>G

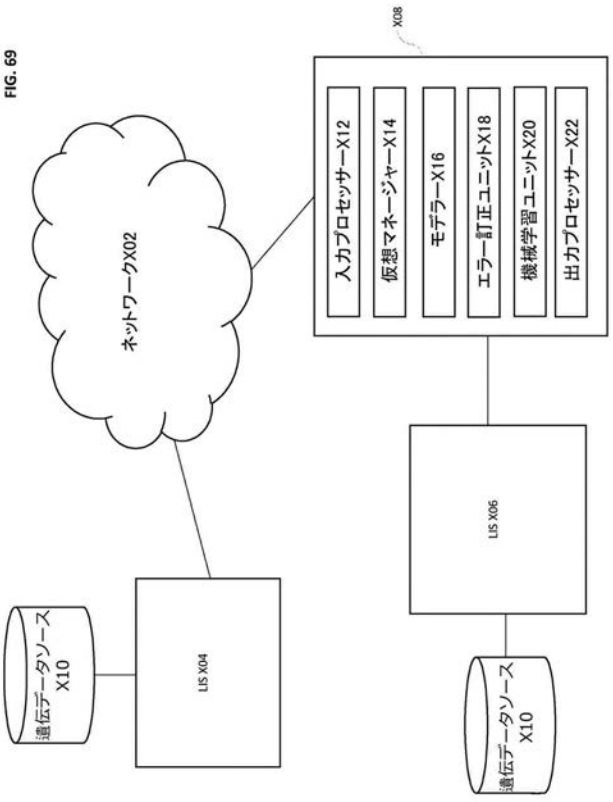
【図６８】

FIG. 68

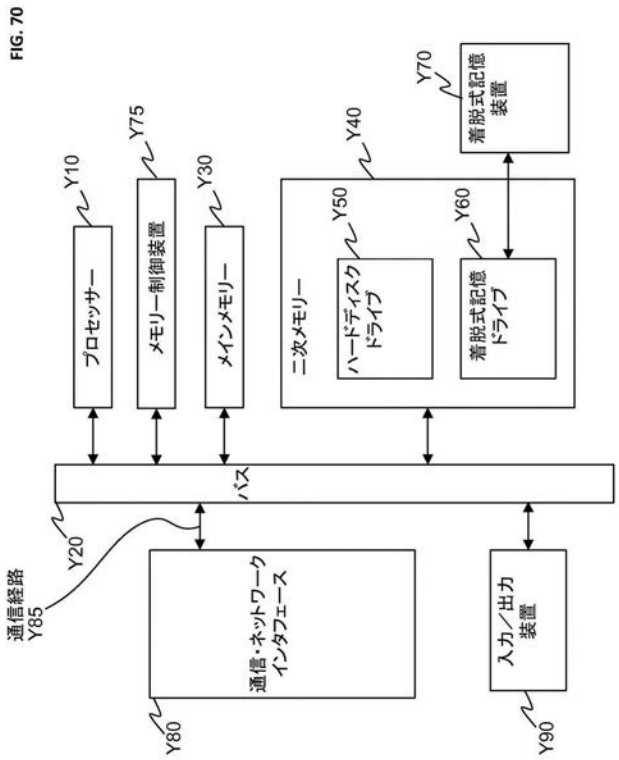


【図６９】

FIG. 69



【図７０】



【手続補正書】
【提出日】平成28年12月21日(2016.12.21)
【手続補正１】
【補正対象書類名】特許請求の範囲
【補正対象項目名】全文
【補正方法】変更
【補正の内容】
【特許請求の範囲】
【請求項１】
個体の試料に含まれる染色体分節の倍数性を決定する方法であって、該方法は、
ａ．該試料中の該染色体分節にある多型座位の集合のうちの各座位に存在する各アレルの量を含む、該試料から得られたアレル頻度データを受け、
ｂ．該アレル頻度データの相を推定することによって該多型座位の集合に関する相決定済み（phased）アレル情報を作成し、
ｃ．該アレル頻度データを用いて、異なる倍数性状態に関する該多型座位のアレル頻度の個別確率を作成し、ここで前記個別確率は、前記多型座位の集合に関する異なる倍数性状態とアレル不均衡率の両方のモデルのセットを用いて作成され、
ｄ．該個別確率と該相決定済みアレル情報を用いて該多型座位の集合に関する同時確率を作成し、ここで前記同時確率は、前記染色体分節にある多型座位間の連鎖を考慮することによって作成され、
ｅ．該同時確率に基づき、染色体倍数性を示す最適合モデルを選択することにより、該染色体分節の倍数性を決定すること
とを含み、
ここで該試料は、癌を有すると考えられる個体から得た血液試料、血清試料、又は血漿試料であり、

ここで該方法は、前記選択に基づき該個体の腫瘍細胞にコピー数多様性が存在するかを決定することをさらに含む、方法。
【請求項２】
前記アレル頻度データが核酸配列データから作成される、請求項１に記載の方法。
【請求項３】
前記アレル頻度データの誤差を修正することと、修正済みアレル頻度データを前記個別確率の作成工程に使用することとをさらに含み、修正される前記誤差がアレル増幅効率の偏りである、請求項１に記載の方法。
【請求項４】
個体の試料の染色体倍数性を検出するための方法であって、該方法は、
ａ．該個体の染色体分節にある多型座位の集合に存在するアレルに関する核酸配列データを受け、
ｂ．該核酸配列データを用いて該多型座位の集合におけるアレル頻度を検出し、
ｃ．検出された該アレル頻度のアレル増幅効率の偏りを修正することによって、該多型座位の集合に関する修正済みアレル頻度を作成し、
ｄ．該核酸配列データの相を推定することによって該多型座位の集合に関する相決定済みアレル情報を作成し、
ｅ．該修正済みアレル頻度を該多型座位の集合に関する異なる倍数性状態とアレル不均衡率のモデルのセットと比較することにより、異なる倍数性状態に関する該多型座位のアレル頻度の個別確率を作成し、
ｆ．該個別確率を該染色体分節にある多型座位間の連鎖を考慮しながら組み合わせることにより、該多型座位の集合に関する同時確率を作成し、
ｇ．該同時確率に基づき、染色体倍数性を示す最適合モデルを選択すること
とを含み、
ここで該試料は、癌を有すると考えられる個体から得た血液試料、血清試料、又は血漿試料であり、
ここで該方法は、前記選択に基づき該個体の腫瘍細胞にコピー数多様性が存在するかを決定することをさらに含む、方法。
【請求項５】
前記染色体分節の第一の相同体と該染色体分節の第二の相同体の倍数性状態をモデリングすることによって個別確率を作成する、請求項１～４のいずれか一項に記載の方法であって、該倍数性状態が、
（１）いずれの細胞も該染色体分節の該第一の相同体または該第二の相同体の欠失または増幅を有していない状態、
（２）少なくとも一部の細胞が該染色体分節の該第一の相同体の欠失、または該第二の相同体の増幅を有する状態、および
（３）少なくとも一部の細胞が該染色体分節の該第二の相同体の欠失、または該第一の相同体の増幅を有する状態
である、方法。
【請求項６】
修正されるエラーが環境不純物および遺伝型解析による不純物を含み、前記試料中のホモ接合アレルに関して測定される、請求項５に記載の方法。
【請求項７】
前記選択が、前記モデルに関して作成された前記相決定済みアレル情報と推定アレル頻度との差異の大きさを分析することによって実行され、差異の大きさが最も小さいモデルが選択される、請求項５に記載の方法。
【請求項８】
アレル頻度の前記個別確率が、前記多型座位の集合におけるアレル頻度の期待値と観測値のベータ二項モデルに基づいて作成される、請求項１～７のいずれか一項に記載の方法。

【請求項９】
前記個別確率がベイズ分類器を用いて作成される、請求項１～７のいずれか一項に記載の方法。
【請求項１０】
前記核酸配列データが、多重増幅反応によって生成された一連の増幅産物の複数のコピーの高処理ＤＮＡ配列決定を実行することによって作成され、ここで、該一連の増幅産物のうちの各増幅産物は、前記多型座位の集合のうちの少なくとも１つの多型座位にわたり、該多型座位の集合のうちの各多型座位が増幅される、請求項２又は４に記載の方法。
【請求項１１】
前記多重増幅反応が制限的なプライマー条件で実行される、請求項１０に記載の方法。
【請求項１２】
前記試料に含まれる染色体分節の集合について染色体倍数性が決定される、請求項１０に記載の方法。
【請求項１３】
前記多重増幅反応が１，０００～１００，０００プレックスの多重反応を含む、請求項１０に記載の方法。
【請求項１４】
０．４５％以上の平均アレル不均衡率が、該個体の腫瘍細胞にコピー数多様性が存在することを示す、請求項１～７のいずれか一項に記載の方法。
【請求項１５】
前記試料が、固形腫瘍を有していたことが知られている個体から得た血漿試料である、請求項１～７のいずれか一項に記載の方法。
【請求項１６】
標的となる複数の一塩基多様体のセットのうち何れかの一塩基多様体が前記試料に存在するか否かを決定することをさらに含み、ここで、該コピー数多様性および／または該一塩基多様体の検出は前記試料中に循環腫瘍核酸が存在することを示す、請求項１～７のいずれか一項に記載の方法。
【請求項１７】
前記個体の腫瘍に関する前記染色体分節のハプロタイプ情報を受けることと、該ハプロタイプ情報を用いて、前記多型座位の集合に関する異なる倍数性状態とアレル不均衡率の前記モデルのセットを作成することをさらに含む、請求項１５に記載の方法。
【請求項１８】
前記方法が、既知の平均アレル不均衡比を有する対照試料に対して該方法を実行することをさらに含む、請求項１～７のいずれか一項に記載の方法。
【請求項１９】
前記修正済みアレル頻度を前記モデルのセットと比較する前に該修正済みアレル頻度のデータから異常値を除外することをさらに含む、請求項１～７のいずれか一項に記載の方法。
【請求項２０】
個体の試料の染色体倍数性を検出するためのシステムであって、該システムは、
ａ．該試料中の該染色体分節にある多型座位の集合のうちの各座位に存在する各アレルの量を含む、該試料から得られたアレル頻度データを受けるよう構成された入力プロセッサと、
ｂ．モデラーであって、
ｉ．該アレル頻度データの相を推定することによって該多型座位の集合に関する相決定済みアレル情報を作成し、
ｉｉ．該アレル頻度データを用いて、異なる倍数性状態に関する該多型座位のアレル頻度の個別確率を作成し、ここで前記個別確率は、前記多型座位の集合に関する異なる倍数性状態とアレル不均衡率の両方のモデルのセットを用いて作成され、
ｉｉｉ．該個別確率と該相決定済みアレル情報を用いて該多型座位の集合に関する同時確

率を作成し、ここで前記同時確率は、前記染色体分節にある多型座位間の連鎖を考慮することによって作成される
よう構成されたモデラーと、
ｃ．該同時確率に基づき、染色体倍数性を示す最適合モデルを選択することにより、該染色体分節の倍数性を決定するよう構成された、仮説管理部（hypothesis manager）
とを含み、
ここで該試料は、癌を有すると考えられる個体から得た血液試料、血清試料、又は血漿試料であり、
ここで該仮説管理部はさらに、前記選択に基づき該個体の腫瘍細胞にコピー数多様性が存在するかを決定するよう構成される、システム。
【請求項２１】
前記アレル頻度データが核酸配列決定システムによって作成される、請求項２０に記載のシステム。
【請求項２２】
前記アレル頻度データの誤差を修正するよう構成された誤差修正部をさらに含み、ここで、修正済みアレル頻度データが前記個別確率の作成工程に用いられ、前記誤差修正部がアレル増幅効率の偏りを修正する、請求項２１に記載のシステム。
【請求項２３】
個体の試料の染色体倍数性を検出するためのシステムであって、該システムは、
ａ．該個体の染色体分節にある多型座位の集合に存在するアレルに関する、該試料から得られた核酸配列データを受け、該核酸配列データを用いて該多型座位の集合におけるアレル頻度を検出するよう構成された入力プロセッサと、
ｂ．検出された該アレル頻度の誤差を修正し、該多型座位の集合に関する修正済みアレル頻度を作成するよう構成された誤差修正部と、
ｃ．モデラーであって、
ｉ．該核酸配列データの相を推定することによって該多型座位の集合に関する相決定済みアレル情報を作成し、
ｉｉ．該相決定済みアレル情報を該多型座位の集合に関する異なる倍数性状態とアレル不均衡率のモデルのセットと比較することにより、異なる倍数性状態に関する該多型座位のアレル頻度の個別確率を作成し、
ｉｉｉ．該個別確率を該染色体分節にある多型座位間の相対的距離を考慮しながら組み合わせることにより、該多型座位の集合に関する同時確率を作成する
よう構成されたモデラーと、
ｄ．該同時確率に基づき、染色体倍数性を示す最適合モデルを選択するよう構成された仮説管理部
とを含み、
ここで該試料は、癌を有すると考えられる個体から得た血液試料、血清試料、又は血漿試料であり、
ここで該仮説管理部はさらに、前記選択に基づき該個体の腫瘍細胞にコピー数多様性が存在するかを決定するよう構成される、システム。
【請求項２４】
前記仮説管理部が、前記モデルに関して作成された前記相決定済みアレル情報と推定アレル頻度との差異の大きさを分析することによって選択を行い、差異の大きさが最も小さいモデルを選択する、請求項２３に記載のシステム。
【請求項２５】
前記モデラーが、前記多型座位の集合におけるアレル頻度の期待値と観測値のベータ二項モデルに基づいてアレル頻度の前記個別確率を作成する、請求項２４に記載のシステム。
【請求項２６】
個体の試料の染色体倍数性を検出するための非一過性のコンピュータ可読媒体であって

、該媒体はコンピュータ可読コードを含み、該コードを処理装置で実行することにより、該処理装置が、
ａ．該試料中の該染色体分節にある多型座位の集合のうちの各座位に存在する各アレルの量を含む、該試料から得られたアレル頻度データを受け、
ｂ．該アレル頻度データの相を推定することによって該多型座位の集合に関する相決定済みアレル情報を作成し、
ｃ．該アレル頻度データを用いて、異なる倍数性状態に関する該多型座位のアレル頻度の個別確率を作成し、ここで前記個別確率は、前記多型座位の集合に関する異なる倍数性状態とアレル不均衡率の両方のモデルのセットを用いて作成され、
ｄ．該個別確率と該相決定済みアレル情報を用いて該多型座位の集合に関する同時確率を作成し、ここで前記同時確率は、前記染色体分節にある多型座位間の連鎖を考慮することによって作成され、
ｅ．該同時確率に基づき、染色体倍数性を示す最適合モデルを選択することにより、該染色体分節の倍数性を決定する、
ように構成され、
ここで該試料は、癌を有すると考えられる個体から得た血液試料、血清試料、又は血漿試料であり、
ここで該コードを該処理装置で実行することにより、該処理装置がさらに、前記選択に基づき該個体の腫瘍細胞にコピー数多様性が存在するかを決定するよう構成される、コンピュータ可読媒体。
【請求項２７】
前記アレル頻度データが核酸配列データから作成される、請求項２６に記載のコンピュータ可読媒体。
【請求項２８】
前記アレル頻度データの誤差を修正することと、修正済みアレル頻度データを前記個別確率の作成工程に使用することをさらに含み、修正される前記誤差がアレル増幅効率の偏りである、請求項２６に記載のコンピュータ可読媒体。
【請求項２９】
個体の試料の染色体倍数性を検出するための非一過性のコンピュータ可読媒体であって、該媒体はコンピュータ可読コードを含み、該コードを処理装置で実行することにより、該処理装置が、
ａ．該個体の染色体分節にある多型座位の集合に存在するアレルに関する、該試料から得られた核酸配列データを受け、
ｂ．該核酸配列データを用いて該多型座位の集合におけるアレル頻度を検出し、
ｃ．検出された該アレル頻度のアレル増幅効率の偏りを修正することによって、該多型座位の集合に関する修正済みアレル頻度を作成し、
ｄ．該核酸配列データの相を推定することによって該多型座位の集合に関する相決定済みアレル情報を作成し、
ｅ．該修正済みアレル頻度を該多型座位の集合に関する異なる倍数性状態とアレル不均衡率のモデルのセットと比較することにより、異なる倍数性状態に関する該多型座位のアレル頻度の個別確率を作成し、
ｆ．該個別確率を該染色体分節にある多型座位間の連鎖を考慮しながら組み合わせることにより、該多型座位の集合に関する同時確率を作成し、
ｇ．該同時確率に基づき、染色体倍数性を示す最適合モデルを選択する、
ように構成され、
ここで該試料は、癌を有すると考えられる個体から得た血液試料、血清試料、又は血漿試料であり、
ここで該コードを該処理装置で実行することにより、該処理装置がさらに、前記選択に基づき該個体の腫瘍細胞にコピー数多様性が存在するかを決定するよう構成される、コンピュータ可読媒体。

【請求項３０】
前記選択が、前記モデルに関して作成された前記相決定済みアレル情報と推定アレル頻度との差異の大きさを分析することによって実行され、差異の大きさが最も小さいモデルが選択される、請求項２９に記載のコンピュータ可読媒体。
【請求項３１】
アレル頻度の前記個別確率が、前記多型座位の集合におけるアレル頻度の期待値と観測値のベータ二項モデルに基づいて作成される、請求項２６～２９のいずれか一項に記載のコンピュータ可読媒体。
【請求項３２】
前記方法が、標的となる複数の一塩基多様体のセットのうち何れかの一塩基多様体が前記試料に存在するか否かを決定することをさらに含み、
前記コピー数多様性および／または前記一塩基多様体の検出は、前記試料中に循環腫瘍核酸が存在することを示し、
前記核酸配列データが、多重増幅反応によって生成された一連の増幅産物の複数のコピーの高処理ＤＮＡ配列決定を実行することによって作成され、
該一連の増幅産物のうちの各増幅産物は、前記多型座位の集合のうちの少なくとも１つの多型座位にわたり、
該多型座位の集合のうちの各多型座位が増幅され、
前記血液試料、血清試料、又は血漿試料と同一の試料を用いて、該個体の腫瘍の細胞にコピー数多様性または一塩基多様体又はその両方が存在するか否かが決定される、請求項２～７の何れか一項に記載の方法。

【国際調査報告】

INTERNATIONAL SEARCH REPORT

International application No
PCT/US2015/026957

A. CLASSIFICATION OF SUBJECT MATTER
INV. C12Q1/68 G06F19/16
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)
C12Q G06F

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, BIOSIS, EMBASE, 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	S. Y. SU ET AL: "Inferring combined CNV/SNP haplotypes from genotype data", BIOINFORMATICS, vol. 26, no. 11, 1 June 2010 (2010-06-01), pages 1437-1445, XP055199108, ISSN: 1367-4803, DOI: 10.1093/bioinformatics/btq157 page 1441, left-hand column, paragraph 2 page 1439 - page 1440 ----- -/--	1,6

[X] Further documents are listed in the continuation of Box C. [X] 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

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
8 July 2015	28/09/2015

Name and mailing address of the ISA/	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Fax: (+31-70) 340-3016	Knudsen, Henrik

1

Form PCT/ISA/210 (second sheet) (April 2005)

INTERNATIONAL SEARCH REPORT

International application No
PCT/US2015/026957

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 2013/130848 A1 (NATERA INC [US]; RABINOWITZ MATTHEW [US]; BANER JOHAN [US]; GEMELOS GE) 6 September 2013 (2013-09-06) pages 161-162 pages 133-134 page 40 page 138, paragraph 1 page 143, lines 12-13 page 148, lines 17-20 page 159 page 180 pages 135-142 page 103, lines 17-19 page 91 page 68, last paragraph	1-43
X	WO 2014/018080 A1 (NATERA INC [US]; ZIMMERMANN BERNHARD [US]; HILL MATTHEW M [US]; LACROU) 30 January 2014 (2014-01-30) figure 29	1
A	WO 2012/071621 A1 (UNIV HONG KONG CHINESE [CN]; LO YUK MING DENNIS [CN]; CHAN KWAN CHEE [) 7 June 2012 (2012-06-07) abstract	18-22
A	H CHRISTINA FAN ET AL: "Whole-genome molecular haplotyping of single cells", NATURE BIOTECHNOLOGY, vol. 29, no. 1, 1 January 2011 (2011-01-01), pages 51-57, XP055026438, ISSN: 1087-0156, DOI: 10.1038/nbt.1739 abstract	1-43
A	RYAN TEWHEY ET AL: "The importance of phase information for human genomics", NATURE REVIEWS GENETICS, vol. 12, no. 3, 1 March 2011 (2011-03-01), pages 215-223, XP055199104, ISSN: 1471-0056, DOI: 10.1038/nrg2950 tables 1-2	1-43
A	BRYAN HOWIE ET AL: "Fast and accurate genotype imputation in genome-wide association studies through pre-phasing", NATURE GENETICS, vol. 44, no. 8, 22 July 2012 (2012-07-22), pages 955-959, XP055151106, ISSN: 1061-4036, DOI: 10.1038/ng.2354 abstract	1-43

1

Form PCT/ISA/210 (continuation of second sheet) (April 2005)

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US2015/026957

Box No. II Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

see additional sheet

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.

2. ☐ As all searchable claims could be searched without effort justifying an additional fees, this Authority did not invite payment of additional fees.

3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

1-43

Remark on Protest

☐ The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.

☐ The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.

☐ No protest accompanied the payment of additional search fees.

International Application No. PCT/US2015/026957

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

This International Searching Authority found multiple (groups of) inventions in this international application, as follows:

1. claims: 1-43

Method of determining ploidy of a chromosomal segment in a sample from an individual based on allele frequency data, phasing of the alleles, generating individual and joint probabilities and selecting a best fit for use in pregnant women or with patients suspected of suffering from cancer. Computer systems and media for use with the method.

---

2. claims: 44-49

Method for determining circulating tumour nucleic acids by determining a ploidy and a level of allelic imbalance higher than 0.45%

---

3. claims: 50-55

Method for determining whether a single nucleotide variant is present or the percent of single nucleotide variants at a genomic position by estimating the efficiency and error rate per cycle for an amplicon spanning the single nucleotide variant and receiving observed nucleotide identity information

---

4. claims: 56-61

Method for detecting one or more single nucleotide variants by determining a median variant allele frequency for a plurality of control samples from each of a plurality of control individuals; identifying selected single nucleotide variants having variant median allele frequencies below a threshold value and to determine background error; and determining an observed depth of read weighted mean and variance.

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No
PCT/US2015/026957

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 2013130848 A1	06-09-2013	NONE	
WO 2014018080 A1	30-01-2014	AU 2012385961 A1	12-02-2015
		CA 2877493 A1	30-01-2014
		CN 104685064 A	03-06-2015
		IL 236435 A	26-02-2015
		KR 20150038216 A	08-04-2015
		WO 2014018080 A1	30-01-2014
WO 2012071621 A1	07-06-2012	AU 2011335891 A1	02-05-2013
		CA 2817370 A1	07-06-2012
		CN 103403182 A	20-11-2013
		EA 201300649 A1	30-10-2013
		EP 2646579 A1	09-10-2013
		JP 2014502155 A	30-01-2014
		KR 20140091650 A	22-07-2014
		NZ 611599 A	29-05-2015
		SG 190344 A1	28-06-2013
		TW 201307570 A	16-02-2013
		US 2013040824 A1	14-02-2013
		US 2014227699 A1	14-08-2014
		WO 2012071621 A1	07-06-2012

フロントページの続き

(31)優先権主張番号 62/146,188
(32)優先日 平成27年4月10日(2015.4.10)
(33)優先権主張国 米国(US)
(31)優先権主張番号 62/147,377
(32)優先日 平成27年4月14日(2015.4.14)
(33)優先権主張国 米国(US)
(31)優先権主張番号 62/148,173
(32)優先日 平成27年4月15日(2015.4.15)
(33)優先権主張国 米国(US)
(31)優先権主張番号 61/994,791
(32)優先日 平成26年5月16日(2014.5.16)
(33)優先権主張国 米国(US)

(81)指定国 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,DK,DM,DO,DZ,EC,EE,EG,ES,FI,GB,GD,GE,GH,GM,GT,HN,HR,HU,ID,IL,IN,IR,IS,JP,KE,KG,KN,KP,KR,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,TZ,UA,UG,US

(74)代理人 100117019
弁理士 渡辺 陽一
(74)代理人 100150810
弁理士 武居 良太郎
(74)代理人 100141977
弁理士 中島 勝
(72)発明者 ジョシュア バビアーツ
アメリカ合衆国 カリフォルニア ９４０７０，サン カルロス，インダストリアル ロード ２０１，スイート ４１０
(72)発明者 チューダー ポンピリウ コンスタンティン
アメリカ合衆国 カリフォルニア ９４０７０，サン カルロス，インダストリアル ロード ２０１，スイート ４１０
(72)発明者 レーン エー．ユーバンク
アメリカ合衆国 カリフォルニア ９４０７０，サン カルロス，インダストリアル ロード ２０１，スイート ４１０
(72)発明者 ジョージ ゲメロス
アメリカ合衆国 カリフォルニア ９４０７０，サン カルロス，インダストリアル ロード ２０１，スイート ４１０
(72)発明者 マシュー ミカー ヒル
アメリカ合衆国 カリフォルニア ９４０７０，サン カルロス，インダストリアル ロード ２０１，スイート ４１０
(72)発明者 フセイン エサー キルキツラー
アメリカ合衆国 カリフォルニア ９４０７０，サン カルロス，インダストリアル ロード ２０１，スイート ４１０
(72)発明者 マシュー ラビノビッツ
アメリカ合衆国 カリフォルニア ９４０７０，サン カルロス，インダストリアル ロード ２０１，スイート ４１０

(72)発明者　オヌール　サカリヤ
アメリカ合衆国　カリフォルニア　９４０７０，サン　カルロス，インダストリアル　ロード　２０１，スイート　４１０
(72)発明者　スティルミル　シグルヨンソン
アメリカ合衆国　カリフォルニア　９４０７０，サン　カルロス，インダストリアル　ロード　２０１，スイート　４１０
(72)発明者　ベルンハルト　ジンマーマン
アメリカ合衆国　カリフォルニア　９４０７０，サン　カルロス，インダストリアル　ロード　２０１，スイート　４１０

Ｆターム(参考) 4B063 QA01 QA12 QA13 QA19 QQ03 QQ42 QR32 QR55 QR62 QR77
QS24 QS34

专利名称(译)	检测染色体片段的突变和倍性		
公开(公告)号	JP2017519488A	公开(公告)日	2017-07-20
申请号	JP2016563812	申请日	2015-04-21
申请(专利权)人(译)	花生酱公司		
[标]发明人	ジョシュアバビアーツ チューダーポンピリウコンスタンティン レーンエーユーバンク ジョージゲメロス マシューミカーヒル フセインエサーキルキツラー マシューラビノビッツ オヌールサカリヤ スティルミルシグルヨンソン ベルンハルトジンマーマン		
发明人	ジョシュア バビアーツ チューダー ポンピリウ コンスタンティン レーン エー.ユーバンク ジョージ ゲメロス マシュー ミカー ヒル フセイン エサー キルキツラー マシュー ラビノビッツ オヌール サカリヤ スティルミル シグルヨンソン ベルンハルト ジンマーマン		
IPC分类号	C12Q1/68 G01N33/53 G01N33/574 C12N15/09		
CPC分类号	C12Q1/6869 C12Q2539/10 G16B15/00 G16B20/00 G16B20/10 G16B20/20 G16B25/20 G16B40/20 G16H50/20 C12Q1/6806 C12Q1/6886 C12Q2600/156 C12Q2600/158 C12Q2600/16 C12Q2600/172 G06F19/00 G06N7/005 G06N20/00 G16B25/00 G16B40/00 G16H10/40		
FI分类号	C12Q1/68.A C12Q1/68.Z G01N33/53.M G01N33/574.Z C12N15/00.A		
F-TERM分类号	4B063/QA01 4B063/QA12 4B063/QA13 4B063/QA19 4B063/QQ03 4B063/QQ42 4B063/QR32 4B063/QR55 4B063/QR62 4B063/QR77 4B063/QS24 4B063/QS34		
代理人(译)	青木 笃 石田 敬 渡边洋一 武井良太郎 中岛胜		
优先权	61/982245 2014-04-21 US 61/987407 2014-05-01 US 62/066514 2014-10-21 US 62/146188 2015-04-10 US 62/147377 2015-04-14 US 62/148173 2015-04-15 US 61/994791 2014-05-16 US		
其他公开文献	JP6659575B2		

外部链接	Espacenet

摘要(译)

本发明提供了一种用于检测患有癌症或怀疑患有癌症的受试者的血液，血清或血浆样品中的一个或多个突变或遗传变异的方法，该方法包括：鉴定肿瘤样品中的多个突变或遗传变异。 通过整个外显子组测序对受试者的研究； 从受试者收集血液，血清或血浆样品，并从血液，血清，血浆或肿瘤样品中分离无细胞DNA； 从无细胞DNA扩增对应于突变或遗传变异的多个基因座以获得扩增子； 对扩增子进行测序以获得序列读数；从序列读数中检测无细胞DNA中存在的一种或多种突变或遗传变异； 任选地，其中无细胞DNA包括循环肿瘤DNA。