(19)日本国特許庁(JP)　(12)公表特許公報(A)　(11)特許出願公表番号
特表2021-501609
(P2021-501609A)
(43)公表日　令和3年1月21日(2021.1.21)

(51)Int.Cl.			F I		テーマコード（参考）
C12Q	1/04	(2006.01)	C12Q	1/04	4B063
C12Q	1/68	(2018.01)	C12Q	1/68	

審査請求 未請求　予備審査請求 未請求　（全 55 頁）

(21)出願番号　特願2020-544091 (P2020-544091)
(86)(22)出願日　平成30年11月2日 (2018.11.2)
(85)翻訳文提出日　令和2年6月3日 (2020.6.3)
(86)国際出願番号　PCT/CN2018/113640
(87)国際公開番号　WO2019/085988
(87)国際公開日　令和1年5月9日 (2019.5.9)
(31)優先権主張番号　62/580,906
(32)優先日　平成29年11月2日 (2017.11.2)
(33)優先権主張国・地域又は機関
米国 (US)

(71)出願人　512037244
ザ　チャイニーズ　ユニバーシティ　オブ　ホンコン
中華人民共和国　ホンコン　ニュー　テリトリーズ　シャーティン　ピチウ　ビルディング　ルーム　３０１　オフィス　オブ　リサーチ　アンド　ナレッジ　トランスファー　サービスィズ
(71)出願人　520154782
グレイル，インコーポレイティド
アメリカ合衆国，カリフォルニア　９４０２５，メンロー　パーク，オブライエン　ドライブ　１５２５
(74)代理人　100099759
弁理士　青木　篤

最終頁に続く

(54)【発明の名称】非侵襲的出生前検査および癌検出のために核酸サイズ範囲を使用すること

(57)【要約】
サイズバンド分析は、染色体領域がコピー数異常またはエピジェネティック変化を示すかを決定するために使用される。特定のサイズに焦点を合わせる代わりに、複数のサイズ範囲が分析され得る。特定のサイズの代わりに複数のサイズ範囲を使用することによって、方法は、より多くの配列リードを分析し得、臨床関連ＤＮＡが低画分の生体試料であり得る場合でも、染色体領域がコピー数異常を示すかを決定することができ得る。複数の範囲を使用することは、ゲノム領域における選択されたリードのサブセットではなく、ゲノム領域からの全ての配列リードの使用を可能にし得る。分析の精度は、類似またはより高い特異度でのより高い感度で増加され得る。分析は、同じ精度を達成するためにより少ない配列決定リードを含み、より効率的なプロセスをもたらし得る。

【特許請求の範囲】
【請求項１】
染色体領域が対象からの生体試料においてコピー数異常を示すかを決定する方法であって、前記生体試料が、臨床関連ＤＮＡ分子および他のＤＮＡ分子を含む無細胞ＤＮＡ分子の混合物を含み、前記方法が、
複数のサイズ範囲の各サイズ範囲について、
前記サイズ範囲に対応する前記生体試料からの無細胞ＤＮＡ分子の第１の量を測定することと、
コンピュータシステムによって、前記サイズ範囲に対応する無細胞ＤＮＡ分子の前記第１の量および前記サイズ範囲にないサイズを含む第２のサイズ範囲のＤＮＡ分子の第２の量を用いてサイズ比を計算することと、
複数のサイズ範囲の複数の参照サイズ比を含む参照サイズパターンを取得することであって、前記参照サイズパターンが、染色体領域にコピー数異常を有する対象またはコピー数異常を有さない対象からの複数の参照試料から決定される、取得することと、
複数の前記サイズ比と前記参照サイズパターンとを比較することと、
前記比較に基づいて前記染色体領域がコピー数異常を示すかを決定することと、を含む、方法。
【請求項２】
前記臨床関連ＤＮＡ分子が、胎児ＤＮＡまたは母体ＤＮＡを含む、請求項１に記載の方法。
【請求項３】
前記臨床関連ＤＮＡ分子が、腫瘍ＤＮＡを含み、前記他のＤＮＡ分子が、非腫瘍ＤＮＡを含む、請求項１に記載の方法。
【請求項４】
前記コピー数異常が、異数性である、請求項２に記載の方法。
【請求項５】
前記コピー数異常が、癌の徴候である、請求項３に記載の方法。
【請求項６】
前記複数のサイズ範囲の各サイズ範囲が、帯域幅によって特徴付けられる、請求項１に記載の方法。
【請求項７】
前記帯域幅が、５０ｂｐ～２００ｂｐの範囲にある、請求項６に記載の方法。
【請求項８】
各サイズ範囲が、前記複数のサイズ範囲のいずれかの他のサイズ範囲と重複していない、請求項１に記載の方法。
【請求項９】
各サイズ範囲が、前記複数のサイズ範囲のうちの少なくとも１つの他のサイズ範囲と重複している、請求項１に記載の方法。
【請求項１０】
前記サイズ比が、ｚスコアを含む、請求項１に記載の方法。
【請求項１１】
前記第２のサイズ範囲は、前記複数のサイズ範囲の各サイズ範囲よりも大きい範囲である、請求項１に記載の方法。
【請求項１２】
前記第２のサイズ範囲が、前記生体試料中の無細胞ＤＮＡ分子の全てのサイズ、または前記染色体領域中の無細胞ＤＮＡ分子の全てのサイズを含む、請求項１に記載の方法。
【請求項１３】
前記無細胞ＤＮＡ分子が、ゲノム領域に由来する、請求項１に記載の方法。
【請求項１４】
前記ゲノム領域が、染色体である、請求項１３に記載の方法。

【請求項１５】
前記ゲノム領域が、染色体腕である、請求項１３に記載の方法。
【請求項１６】
前記複数のサイズ比と前記参照サイズパターンとを比較することが、
前記複数のサイズ比の各サイズ比と前記対応するサイズ範囲での前記参照サイズ比とを比較することと、
各サイズ比が前記対応するサイズ範囲での前記参照サイズ比と統計的に類似することを決定することと、を含む、請求項１に記載の方法。
【請求項１７】
前記複数のサイズ比と前記参照サイズパターンとを比較することが、
前記複数のサイズ範囲について前記複数のサイズ比を含むサイズパターンを決定することと、
前記サイズパターンと前記参照サイズパターンとを比較することと、
前記サイズパターンが前記参照サイズパターンと類似の形状を有することを決定することと、を含む、請求項１に記載の方法。
【請求項１８】
前記参照サイズパターンが、コピー数異常を有する対象からの前記複数の参照試料から決定され、
前記方法が、
前記比較に基づいて前記染色体領域がコピー数異常を示すことを決定することをさらに含む、請求項１６に記載の方法。
【請求項１９】
前記参照サイズパターンを取得し、前記複数のサイズ比と前記参照サイズパターンとを比較することが、前記複数のサイズ比を機械学習モデルに入力することを含み、
前記機械学習モデルが、複数の参照試料からの複数の訓練サイズパターンを用いて訓練された、請求項１に記載の方法。
【請求項２０】
前記複数のサイズ比と前記参照サイズパターンとを比較することが、前記複数のサイズ比と前記複数の参照試料から決定される複数の閾値とを比較することを含む、請求項１に記載の方法。
【請求項２１】
対象からの生体試料において癌分類を決定する方法であって、前記生体試料が、腫瘍ＤＮＡ分子および非腫瘍ＤＮＡ分子を含む無細胞ＤＮＡ分子の混合物を含み、前記方法が、
複数のサイズ範囲の各サイズ範囲について、
前記サイズ範囲に対応する前記生体試料からのメチル化無細胞ＤＮＡ分子の第１の量を測定することと、
コンピュータシステムによって、前記サイズ範囲に対応するメチル化無細胞ＤＮＡ分子の前記第１の量および前記サイズ範囲にないサイズを含む第２のサイズ範囲のＤＮＡ分子の第２の量を用いてメチル化レベルを計算することと、
前記複数のサイズ範囲の複数のメチル化レベルを含む参照サイズパターンを取得することであって、前記参照サイズパターンが、癌を有する対象または癌を有さない対象からの複数の参照試料から決定される、取得することと、
複数の前記メチル化レベルと前記参照サイズパターンとを比較することと、
前記比較に基づいて癌のレベルを決定することと、を含む、方法。
【請求項２２】
前記第２の量が、メチル化無細胞ＤＮＡ分子のものである、請求項２１に記載の方法。
【請求項２３】
前記メチル化無細胞ＤＮＡ分子が、染色体腕由来である、請求項２１に記載の方法。
【請求項２４】
前記複数のメチル化レベルと前記参照サイズパターンとを比較することが、

前記複数のサイズ範囲の各メチル化レベルと前記対応するサイズ範囲での前記参照メチル化レベルとを比較することと、
各メチル化レベルが前記対応するサイズ範囲での前記参照メチル化レベルと統計的に類似することを決定することと、を含む、請求項２１に記載の方法。
【請求項２５】
前記複数のメチル化レベルと前記参照サイズパターンとを比較することとが、
前記複数のサイズ範囲について前記複数のメチル化レベルを含むサイズパターンを決定することと、
前記サイズパターンと前記参照サイズパターンとを比較することと、
前記サイズパターンが前記参照サイズパターンと類似の形状を有することを決定することと、を含む、請求項２１に記載の方法。
【請求項２６】
前記参照サイズパターンが、癌を有する対象からの前記複数の参照試料から決定され、
前記方法が、
前記対象が癌を有することを決定することをさらに含む、請求項２４に記載の方法。
【請求項２７】
メチル化無細胞ＤＮＡ分子の前記第１の量が、ゲノム領域由来である、請求項２１に記載の方法。
【請求項２８】
前記ゲノム領域が、染色体腕であり、前記染色体腕が、１ｐ、１ｑ、８ｐ、８ｑ、１３ｑ、および１４ｑからなる群から選択される、請求項２７記載の方法。
【請求項２９】
前記複数のメチル化レベルと前記参照サイズパターンとを比較することが、前記複数のメチル化レベルと前記複数の参照試料から決定される複数の閾値とを比較することを含む、請求項２１に記載の方法。
【請求項３０】
前記複数のサイズ範囲が、Ｍサイズ範囲を含み、
メチル化無細胞ＤＮＡ分子の前記第１の量を測定することが、前記サイズ範囲に対応し、Ｎゲノム領域の各ゲノム領域に対応するメチル化無細胞ＤＮＡ分子の前記第１の量を測定することを含み、
前記サイズ範囲に対応し、前記ゲノム領域に対応するメチル化無細胞ＤＮＡの前記第１の量および前記第２の量を用いて前記メチル化レベルを計算することが、Ｎ×Ｍメチル化レベルの測定ベクターを生成し、Ｎは１以上の整数であり、Ｍは１超の整数であり、
前記参照サイズパターンが、前記Ｎゲノム領域および前記Ｍサイズ範囲の参照メチル化レベルの参照ベクターを含み、前記参照サイズパターンが、癌を有する対象または癌を有さない対象からの複数の参照試料から決定され、
前記複数のメチル化レベルと前記参照サイズパターンとを比較することが、前記測定ベクターと前記参照ベクターとを比較することを含む、請求項２１に記載の方法。
【請求項３１】
対象からの生体試料において癌分類を決定する方法であって、前記生体試料が、腫瘍ＤＮＡ分子および非腫瘍ＤＮＡ分子を含む無細胞ＤＮＡ分子の混合物を含み、前記方法が、
Ｎゲノム領域の各ゲノム領域について
Ｍサイズ範囲の各サイズ範囲について
前記サイズ範囲に対応し、前記ゲノム領域に対応する前記生体試料からの無細胞ＤＮＡ分子の第１の量を測定することと、
コンピュータシステムによって、前記サイズ範囲に対応し、前記ゲノム領域に対応する無細胞ＤＮＡ分子の前記第１の量および前記サイズ範囲にないサイズを含む第２のサイズ範囲のＤＮＡ分子の第２の量を用いてサイズ比を計算し、これによってＮ×Ｍサイズ比の測定ベクターを生成することであって、Ｎが１以上の整数であり、Ｍが１超の整数である

、生成することと、
前記Ｎゲノム領域および前記Ｍサイズ範囲の参照サイズ比の参照ベクターを含む参照サイズパターンを取得することであって、前記参照サイズパターンが、癌を有する対象または癌を有さない対象からの複数の参照試料から決定される、取得することと、
前記測定ベクターと前記参照ベクターとを比較することと、
前記比較に基づいて癌のレベルを決定することと、を含む、方法。
【請求項３２】
各ゲノム領域が、染色体腕である、請求項３１に記載の方法。
【請求項３３】
前記参照サイズパターンが、機械学習モデルを用いて決定され、前記機械学習モデルが、サポートベクターマシン、決定木、単純ベイズ分類、ロジスティック回帰、クラスタリングアルゴリズム、主成分分析、特異値分解、ｔ分布型確率的近傍埋め込み、および人工ニューラルネットワークからなる群から選択される少なくとも１つを含む、請求項３１に記載の方法。
【請求項３４】
前記測定ベクターと前記参照ベクターとを比較することが、癌を有すると決定された個体および癌を有していないと決定された個体の異なるゲノム領域のサイズ比を含む訓練ベクターの訓練セットで訓練された機械学習モデルを使用することを含む、請求項３１に記載の方法。
【請求項３５】
前記癌が、肝細胞癌を含む、請求項３１に記載の方法。
【請求項３６】
前記癌のレベルが、癌の確率を含む、請求項３１に記載の方法。
【請求項３７】
前記参照サイズパターンを取得することおよび前記測定ベクターと前記参照ベクターとを比較することが、機械学習モデルを使用することを含み、
前記機械学習モデルが、複数の参照サイズパターンを用いて訓練され、
前記測定ベクターと前記参照ベクターとを比較することが、前記測定ベクターと前記参照ベクターとの類似性を特徴付けるカットオフ値を決定することを含み、
癌のレベルを決定することが、前記カットオフ値を使用する、請求項３１に記載の方法。
【請求項３８】
前記測定ベクターと前記参照ベクターとを比較することが、前記Ｎ×Ｍサイズ比と前記複数の参照試料から決定される複数の閾値とを比較することを含む、請求項３１に記載の方法。
【請求項３９】
請求項１に記載の操作を行うようにコンピュータシステムを制御するための複数の指示を保存する非一時的コンピュータ可読媒体を備える、コンピュータ製品。
【請求項４０】
システムであって、
請求項３９に記載のコンピュータ製品と、
前記非一時的コンピュータ可読媒体上に保存される指示を実行するための１つ以上のプロセッサと、を備える、システム。
【請求項４１】
前記方法のうちのいずれかを行うための手段を備える、システム。
【請求項４２】
前記方法のうちのいずれかを行うように構成された、システム。
【請求項４３】
前記方法のうちのいずれかのステップをそれぞれ行うモジュールを備える、システム。
【発明の詳細な説明】

【技術分野】
【０００１】
関連出願の相互参照
本出願は、２０１７年１１月２日に出願された「ＵＳＩＮＧ　ＮＵＣＬＥＩＣ　ＡＣＩＤ　ＳＩＺＥ　ＲＡＮＧＥ　ＦＯＲ　ＮＯＮＩＮＶＡＳＩＶＥ　ＰＲＥＮＡＴＡＬ　ＴＥＳＴＩＮＧ　ＡＮＤ　ＣＡＮＣＥＲ　ＤＥＴＥＣＴＩＯＮ」と題された米国仮特許出願第６２／５８０，９０６号からの優先権を主張し、その全内容は、全ての目的のために参照により本明細書に組み込まれる。
【背景技術】
【０００２】
妊婦の血漿および血清中の胎児に由来する循環無細胞ＤＮＡ（ｃｆＤＮＡ）の存在の実証（Ｌｏ　ｅｔ　ａｌ．，Ｌａｎｃｅｔ　１９９７；３５０：４８５－４８７）は、非侵襲的出生前検査（ＮＩＰＴ）の開発を通して出生前検査の実施を完全に転換した。ＮＩＰＴは、羊水穿刺および絨毛膜絨毛サンプリング（ＣＶＳ）などの侵襲的組織サンプリングと関連したリスクの回避において利点を有する。これまでのところ、ＮＩＰＴは、胎児ＲｈＤ血液群ジェノタイピング（Ｆｉｎｎｉｎｇ　ｅｔ　ａｌ．ＢＭＪ　２００８；３３６：８１６－８１８；Ｌｏ　ｅｔ　ａｌ．Ｎ　Ｅｎｇｌ　Ｊ　Ｍｅｄ　１９９８；３３９：１７３４－１７３８）、伴性障害についての胎児性決定（Ｃｏｓｔａ　ｅｔ　ａｌ．Ｎ．Ｅｎｇｌ．Ｊ．Ｍｅｄ．２００２；３４６：１５０２）、染色体異数性検出（Ｃｈｉｕ　ｅｔ　ａｌ．Ｐｒｏｃ　Ｎａｔｌ　Ａｃａｄ　Ｓｃｉ　Ｕ　Ｓ　Ａ　２００８；１０５：２０４５８－２０４６３；Ｆａｎ　ｅｔ　ａｌ．Ｎａｔｕｒｅ　２０１２；４８７：３２０－３２４；Ｃｈｉｕ　ｅｔ　ａｌ．ＢＭＪ　２０１１；３４２：ｃ７４０１；Ｂｉａｎｃｈｉ　ｅｔ　ａｌ．Ｎ．Ｅｎｇｌ．Ｊ．Ｍｅｄ．２０１４；３７０：７９９－８０８；Ｙｕ　ｅｔ　ａｌ．Ｐｒｏｃ．Ｎａｔｌ．Ａｃａｄ．Ｓｃｉ．Ｕ．Ｓ．Ａ．２０１４；１１１：８５８３－８；Ｎｏｒｔｏｎ　ｅｔ　ａｌ．Ｎ．Ｅｎｇｌ．Ｊ．Ｍｅｄ．２０１５；３７２：１５８９－１５９７）、および単一遺伝子障害の診断（Ｌａｍ　ｅｔ　ａｌ．Ｃｌｉｎ．Ｃｈｅｍ．２０１２；５８：１４６７－７５；Ｌｏ　ｅｔ　ａｌ．Ｓｃｉ．Ｔｒａｎｓｌ．Ｍｅｄ．２０１０；２：６１ｒａ９１－６１ｒａ９１；Ｍａ　ｅｔ　ａｌ．Ｇｅｎｅ　２０１４；５４４：２５２－２５８；Ｎｅｗ　ｅｔ　ａｌ．Ｊ．Ｃｌｉｎ．Ｅｎｄｏｃｒｉｎｏｌ．Ｍｅｔａｂ．２０１４；９９：Ｅ１０２２－Ｅ１０３０）に使用されている。特に、母体血漿ＤＮＡの超並列配列決定を用いる、一般的な染色体異数性についてのＮＩＰＴは、数十か国で臨床サービスに急速に採用され、毎年数百万人の妊婦によって使用されている（Ａｌｌｙｓｅ　ｅｔ　ａｌ．Ｉｎｔ．Ｊ．Ｗｏｍｅｎｓ．Ｈｅａｌｔｈ　２０１５；７：１１３－２６；Ｃｈａｎｄｒａｓｅｋｈａｒａｎ　ｅｔ　ａｌ．Ｓｃｉ　Ｔｒａｎｓｌ　Ｍｅｄ　２０１４；６：２３１ｆｓ１５）。
【０００３】
初期検証研究（Ｃｈｉｕ　ｅｔ　ａｌ．ＢＭＪ　２０１１；３４２：ｃ７４０１；Ｓｐａｒｋｓ　ｅｔ　ａｌ．Ａｍ．Ｊ．Ｏｂｓｔｅｔ．Ｇｙｎｅｃｏｌ．２０１２；２０６：３１９．ｅ１－９）では、ＮＩＰＴは、異数性の高いリスクで患者に実施され、９２％～１００％の高い陽性的中率（ＰＰＶ）が達成された。一般的には胎児ＤＮＡ画分と呼ばれる、特定の母体試料における胎児ＤＮＡの相対濃度は、ＮＩＰＴの精度の重要な決定因子である（Ｃｈｉｕ　ｅｔ　ａｌ．ＢＭＪ　２０１１；３４２：ｃ７４０１；Ｊｉａｎｇ　ｅｔ　ａｌ．Ｂｉｏｉｎｆｏｒｍａｔｉｃｓ　２０１２；２８：２８８３－２８９０、ｎｐｊ　Ｇｅｎｏｍｉｃ　Ｍｅｄ．２０１６；１：１６０１３）。２１トリソミー検出の感度は、胎児ＤＮＡ画分の低減とともに有意に減少するであろう（Ｃｈｉｕ　ｅｔ　ａｌ．ＢＭＪ　２０１１；３４２：ｃ７４０１；Ｃａｎｉｃｋ　ｅｔ　ａｌ．Ｐｒｅｎａｔ．Ｄｉａｇｎ．２０１３；３３：６６７－６７４）。したがって、トリソミー検出の偽陰性結果は、低い胎児ＤＮＡ画分を有する妊娠において発生し得る。例えば、Ｃａｎｉｃｋ　ｅｔ　ａｌは、ダウン症候群を有する２１２症例の中に、４つの偽陰性があり、その全てが４％～７％の胎児ＤＮＡ画分を有したことを報告した（Ｃａｎｉｃｋ　ｅｔ　ａｌ．Ｐｒ

ｅｎａｔ．Ｄｉａｇｎ．２０１３；３３：６６７－６７４）。
【０００４】
ＮＩＰＴを行う多くの実験室では、ある割合の分析において検査失敗またはノーコール結果が観察されることに留意することが重要である。いくつかの研究では、総実験室失敗率は、８．８％に達し得る（Ｐｏｒｒｅｃｏ ｅｔ ａｌ．Ａｍ．Ｊ．Ｏｂｓｔｅｔ．Ｇｙｎｅｃｏｌ．２０１４；２１１：３６５．ｅ１－３６５．ｅ１２）。ＮＩＰＴで結果を得ることができない主な理由の１つは、いくつかの試料における母体血漿ＤＮＡ中の低い胎児ＤＮＡ画分、通常は＜４％である（Ｇｉｌ ｅｔ ａｌ．Ｆｅｔａｌ Ｄｉａｇｎ．Ｔｈｅｒ．２０１４；３５：１５６－７３）。４％未満の胎児ＤＮＡ画分を有する患者では、異数性の有病率は４．７％であると報告され、これは全体コホートにおける０．４％の有病率と比較して有意により高かったことが実証された（Ｎｏｒｔｏｎ ｅｔ ａｌ．Ｎ．Ｅｎｇｌ．Ｊ．Ｍｅｄ．２０１５；３７２：１５８９－１５９７）。したがって、そのような検査失敗は、最終的にはＮＩＰＴの全体的な性能に悪影響を及ぼし得る。例えば、より高い検査失敗率がより低い実際のＰＰＶをもたらすことが例示された（Ｙａｒｏｎ Ｐｒｅｎａｔ．Ｄｉａｇｎ．２０１６；３６：３９１－６）。理論的推定（Ｙａｒｏｎ Ｐｒｅｎａｔ．Ｄｉａｇｎ．２０１６；３６：３９１－６）では、実験室での０．１％の失敗率は６７％の実際のＰＰＶを与えるが、１％の失敗率は、１６．７％の実際のＰＰＶを生じさせ、異数性の増加したリスクと関連していると報告された検査失敗を有する全てのこれらの患者が、胎児が実際に米国産科婦人科学会（ＡＣＯＧ）推奨からの推奨による異数体であるかを確認するために侵襲的検査を行うであろうことを想定する（Ｙａｒｏｎ Ｐｒｅｎａｔ．Ｄｉａｇｎ．２０１６；３６：３９１－６）。
【０００５】
妊娠の約２％が４％未満の胎児ＤＮＡ画分を有することが示されている（Ｗａｎｇ ｅｔ ａｌ．Ｐｒｅｎａｔ．Ｄｉａｇｎ．２０１３；３３：６６２－６６６）。１０～２１週間の胎児ＤＮＡの増加は、非常に微妙（胎児ＤＮＡ画分において週に約０．１％平均増加）であるため、低い胎児ＤＮＡ画分を示す第１の血液試料を有する患者の再採血が、十分な胎児ＤＮＡ画分の正当な理由となる可能性は低い（Ｗａｎｇ ｅｔ ａｌ．Ｐｒｅｎａｔ．Ｄｉａｇｎ．２０１３；３３：６６２－６６６）。加えて、そのような低胎児ＤＮＡ画分は、高母体重を有する女性において優先的に発生する。いくつかの研究では、胎児ＤＮＡ画分４％未満のために結果を報告することができないことは、５．９％に達し得る（Ｈａｌｌ ｅｔ ａｌ．ＰＬｏＳ Ｏｎｅ ２０１４；９：ｅ９６６７７）。
【０００６】
したがって、母体血漿中の低い胎児ＤＮＡ画分（例えば、４％未満）を有する妊婦についてのＮＩＰＴの性能を改善するためのアプローチを開発することは有用であろう。そのような改善は、一般的な染色体異数性（例えば、２１トリソミー、１８トリソミー、１３トリソミー、および性染色体異数性）およびサブ染色体異常（例えば、微小欠失および微小重複）についてのＮＩＰＴの性能に有益であろう。加えて、コピー数異常および癌についての検査の精度および効率を改善することは、同様のアプローチで対処することができる。これらおよびその他のニーズは、以下で対処される。
【発明の概要】
【０００７】
サイズバンド分析は、染色体領域がコピー数異常を示すか、または癌を検出するために使用されるかを決定するために使用される。特定のサイズに焦点を合わせる代わりに、複数のサイズ範囲が分析され得る。特定のサイズの代わりに複数のサイズ範囲を使用することによって、方法は、臨床関連ＤＮＡが低画分の生体試料であり得る場合でも、染色体領域がコピー数異常を示すかを決定することができ得る。複数の範囲を使用することは、ゲノム領域における選択されたリードのサブセットではなく、ゲノム領域からの全ての配列リードの使用を可能にし得る。分析の精度は、類似またはより高い特異度でのより高い感度で増加され得る。分析は、同じ精度を達成するためにより少ない配列決定リードを含み、より効率的なプロセスをもたらし得る。分析は、より低い画分の臨床関連ＤＮＡで行わ

れ得るため、分析は、妊娠または癌の初期段階で行われ得る。
【０００８】
本発明の実施形態の性質および利点に関するより良好な理解は、以下の詳細な説明および添付の図面を参照して得ることができる。
【図面の簡単な説明】
【０００９】
【図１】本発明の実施形態による血漿ＤＮＡサイズバンド分析の原理の概略図を示す。
【図２Ａ】本発明の実施形態による血漿ＤＮＡ断片のサイズについての異数体染色体について測定された胎児ＤＮＡ画分を示す。
【図２Ｂ】本発明の実施形態による正倍数性および２１トリソミー胎児からのＤＮＡを含む試料についてのサイズバンドのｚスコアを示す。
【図３－１】本発明の実施形態による４％の胎児ＤＮＡ画分を有する異なる個々の妊娠にわたる異数体染色体について測定されたゲノム表現（ＧＲ）のサイズバンドベースの変化するパターンを示す。
【図３－２】図３－１の説明に同じ。
【図３－３】図３－１の説明に同じ。
【図３－４】図３－１の説明に同じ。
【図３－５】図３－１の説明に同じ。
【図３－６】図３－１の説明に同じ。
【図３－７】図３－１の説明に同じ。
【図３－８】図３－１の説明に同じ。
【図３－９】図３－１の説明に同じ。
【図３－１０】図３－１の説明に同じ。
【図３－１１】図３－１の説明に同じ。
【図３－１２】図３－１の説明に同じ。
【図４Ａ】本発明の実施形態による正倍数体および２１トリソミー胎児での妊娠間のサイズバンドベースの変化するパターンのヒートマッププロットを示す。
【図４Ｂ】本発明の実施形態による正倍数体および２１トリソミー胎児での妊娠間のサイズバンドベースの変化するパターンのｔ－ＳＮＥ（ｔ分布型確率的近傍埋め込み）プロットを示す。
【図４Ｃ】本発明の実施形態による正倍数体および２１トリソミー胎児での妊娠間の従来のｚスコアアプローチを用いたｚスコア分布を示す。
【図５Ａ】本発明の実施形態による異なるサイズバンド間のｚスコアパターンを学習することによるニューラルネットワークベースのモデルの性能評価を示す。
【図５Ｂ】図５Ａの説明に同じ。
【図６】染色体領域が本発明の実施形態による対象からの生体試料においてコピー数異常を示すかを決定する方法を示す。
【図７】本発明の実施形態による肝細胞癌（ＨＣＣ）患者の血漿ＤＮＡにおける測定されたメチル化のサイズバンドベースの変化するパターンを示す。
【図８】本発明の実施形態による対象からの生体試料における癌分類を決定する方法を示す。
【図９】本発明の実施形態による肝細胞癌（ＨＣＣ）患者の血漿ＤＮＡにおける測定されたコピー数異常のサイズバンドベースの変化するパターンを示す。
【図１０】本発明の実施形態による癌検出のためのサイズバンドゲノム表現（ＧＲ）アプローチのワークフローを例示する。
【図１１Ａ】本発明の実施形態によるサイズバンドＧＲと従来のｚスコアアプローチとの比較を示す。
【図１１Ｂ】図１１Ａの説明に同じ。
【図１１Ｃ】図１１Ａの説明に同じ。
【図１２】本発明の実施形態による癌分類を決定する方法を示す。

【図１３】本発明の実施形態による癌検出のためのサイズバンドメチル化密度（ＭＤ）アプローチのワークフローを例示する。
【図１４Ａ】本発明の実施形態によるサイズバンドＭＤと従来のｚスコアアプローチとの比較を示す。
【図１４Ｂ】図１４Ａの説明に同じ。
【図１４Ｃ】図１４Ａの説明に同じ。
【図１５】本発明の実施形態によるシステムを例示する。
【図１６】本発明の実施形態によるコンピュータシステムを示す。
【００１０】
用語
「試料」、「生体試料」、または「患者試料」という用語は、生きている対象または死んだ対象に由来する任意の組織または物質を含むことを意味する。生体試料は、無細胞試料であり得、これは、対象からの核酸分子と、場合によっては病原体、例えば、ウイルスからの核酸分子との混合物を含み得る。生体試料は一般に、核酸（例えば、ＤＮＡまたはＲＮＡ）またはその断片を含む。「核酸」という用語は一般に、デオキシリボ核酸（ＤＮＡ）、リボ核酸（ＲＮＡ）、またはそれらの任意のハイブリッドもしくは断片を指し得る。試料中の核酸は、無細胞核酸であってもよい。試料は、液体試料または固体試料（例えば、細胞または組織試料）であり得る。生体試料は、血液、血漿、血清、尿、膣液、水腫（例えば、精巣の）からの流体、膣洗浄流体、胸水、腹水、脳脊髄液、唾液、汗、涙、痰、気管支肺胞洗浄液、乳首からの排出液、体の異なる部分（例えば、甲状腺、乳腺）からの吸引液などの体液であり得る。便試料もまた、使用することができる。種々の実施形態では、無細胞ＤＮＡのために濃縮された生体試料（例えば、遠心分離プロトコルを介して得られた血漿試料）におけるＤＮＡの大部分は、無細胞であり得る（例えば、ＤＮＡの５０％、６０％、７０％、８０％、９０％、９５％、または９９％超は、無細胞であり得る）。遠心分離プロトコルは、例えば、３，０００ｇ×１０分で流体部分を得ることと、残留細胞を除去するために３０，０００ｇでさらに１０分間再遠心分離することと、を含み得る。
【００１１】
本明細書で使用されるとき、「遺伝子座（ｌｏｃｕｓ）」またはその複数形「遺伝子座（ｌｏｃｉ）」という用語は、ゲノムにわたって変動を有するヌクレオチド（または塩基対）の任意の長さの場所またはアドレスである。「配列リード」という用語は、核酸分子、例えば、ＤＮＡ断片の全部または一部から得られる配列を指す。一実施形態では、断片の一端のみが配列決定される。あるいは、断片の両端（例えば、各端から約３０ｂｐ）を配列決定して、２つの配列リードを生成することができる。対合された配列リードは、次いで参照ゲノムにアラインすることができ、これは断片の長さを提供することができる。さらに別の実施形態では、線状ＤＮＡ断片は、例えば、ライゲーションによって、環状化することができ、ライゲーション部位に及ぶ部分は、配列決定することができる。
【００１２】
本明細書で使用される「断片」（例えば、ＤＮＡ断片）という用語は、少なくとも３つの連続したヌクレオチドを含むポリヌクレオチドまたはポリペプチド配列の一部を指し得る。核酸断片は、親ポリペプチドの生体活性および／またはいくつかの特徴を保持することができる。核酸断片は、二本鎖または一本鎖、メチル化または非メチル化、インタクトまたはニック、他の高分子、例えば、脂質粒子、タンパク質と複合または非複合であり得る。腫瘍由来の核酸は、腫瘍細胞における病原体からの病原体核酸を含む、腫瘍細胞から放出される任意の核酸を指し得る。
【００１３】
「アッセイ」という用語は一般に、核酸の特性を決定するための技法を指す。アッセイ（例えば、第１のアッセイまたは第２のアッセイ）は一般に、試料中の核酸の量、試料中の核酸のゲノム同一性、試料中の核酸のコピー数変動、試料中の核酸のメチル化状態、試料中の核酸の断片サイズ分布、試料中の核酸の突然変異状態、または試料中の核酸の断片

化パターンを決定する技法を指す。当業者に既知の任意のアッセイは、本明細書で言及される核酸の特性のいずれかを検出するために使用され得る。核酸の特性には、配列、量、ゲノム同一性、コピー数、１つ以上のヌクレオチド位置でのメチル化状態、核酸のサイズ、１つ以上のヌクレオチド位置での核酸の突然変異、および核酸の断片化のパターン（例えば、核酸が断片化するヌクレオチド位置（複数可））が含まれる。「アッセイ」という用語は、「方法」という用語と交換可能に使用されてもよい。アッセイまたは方法は、特定の感度および／または特異度を有することができ、それらの診断ツールとしての相対的な有用性は、ＲＯＣ－ＡＵＣ統計を使用して測定することができる。

【００１４】

本明細書で使用される「ランダム配列決定」という用語は一般に、配列決定される核酸断片が、配列決定手順の前に、具体的に特定または既定されていない配列決定を指す。特定の遺伝子座を標的化するための配列特異的プライマーは、必要とされない。いくつかの実施形態では、アダプタが断片の端部に付加され、配列決定のためのプライマーがアダプタに結合される。よって、任意の断片は、同じ普遍的アダプタに結合する同じプライマーで配列決定することができ、よって、配列決定は、ランダムであり得る。超並列配列決定は、ランダム配列決定を使用して行われ得る。

【００１５】

「核酸」は、一本鎖または二本鎖のいずれかの形態のデオキシリボヌクレオチドまたはリボヌクレオチドおよびそれらのポリマーを指し得る。用語は、合成、自然発生、および非自然発生であり、参照核酸と同様の結合特性を有し、参照ヌクレオチドと同様の方法で代謝される、既知のヌクレオチド類似体または修飾されたバックボーン残基もしくは結合を含有する核酸を包含し得る。そのような類似体の例は、限定なく、ホスホロチオエート、ホスホラミダイト、メチルホスホネート、キラルメチルホスホネート、２－Ｏ－メチルリボヌクレオチド、ペプチド核酸（ＰＮＡ）を含み得る。

【００１６】

特に示されない限り、特定の核酸配列はまた、その保存的に修飾されたバリアント（例えば、縮退コドン置換）および相補的配列、ならびに明示的に示される配列を黙示的に包含する。具体的には、縮退コドン置換は、１つ以上の選択される（または全ての）コドン第３位が混合塩基および／またはデオキシイノシン残基で置換されている配列を生成することによって達成され得る（Ｂａｔｚｅｒ ｅｔ ａｌ．，Ｎｕｃｌｅｉｃ Ａｃｉｄ Ｒｅｓ．１９：５０８１（１９９１）；Ｏｈｔｓｕｋａ ｅｔ ａｌ．，Ｊ．Ｂｉｏｌ．Ｃｈｅｍ．２６０：２６０５－２６０８（１９８５）；Ｒｏｓｓｏｌｉｎｉ ｅｔ ａｌ．，Ｍｏｌ．Ｃｅｌｌ．Ｐｒｏｂｅｓ ８：９１－９８（１９９４））。核酸という用語は、遺伝子、ｃＤＮＡ、ｍＲＮＡ、オリゴヌクレオチド、およびポリヌクレオチドと交換可能に使用される。

【００１７】

自然発生リボヌクレオチドまたはデオキシリボヌクレオチドモノマーに言及することに加えて、「ヌクレオチド」という用語は、特に文脈が明示しない限り、ヌクレオチドが使用されている（例えば、相補的塩基へのハイブリダイゼーション）特定の文脈に関して機能的に同等である、誘導体および類似体を含む、それらの関連構造バリアントを指すと理解され得る。

【００１８】

「配列リード」は、核酸分子の任意の一部または全部から配列決定されるヌクレオチドの鎖を指す。例えば、配列リードは、生体試料に存在する核酸断片全体であり得る。同様に例として、配列リードは、核酸断片から配列決定されるヌクレオチドの短鎖（例えば、約２０～１５０個の塩基）、核酸断片の一端もしくは両端でのヌクレオチドの短鎖、または生体試料に存在する核酸断片全体の配列決定であり得る。配列リードは、例えば、配列決定技法を使用するか、もしくはプローブを使用する様々な方法で、例えば、ハイブリダイゼーションアレイもしくは捕捉プローブで、または単一プライマーもしくは等温増幅を使用するポリメラーゼ連鎖反応（ＰＣＲ）もしくは線形増幅などの増幅技法で、あるいは

質量分析などの生物物理学的測定に基づいて、入手され得る。配列リードは、単一分子配列決定から入手され得る。「単一分子配列決定」とは、テンプレートＤＮＡ分子のクローンコピーからの塩基配列情報を解釈する必要なしに配列リードを得るための、単一テンプレートＤＮＡ分子の配列決定を指す。単一分子配列決定は、ＤＮＡ分子の分子全体または一部のみを配列決定し得る。ＤＮＡ分子の大部分は、例えば、５０％、５５％、６０％、６５％、７０％、７５％、８０％、８５％、９０％、９５％、または９９％を超えて配列決定され得る。

【００１９】

「普遍的配列決定」という用語は、アダプタが断片の端部に付加され、配列決定のためのプライマーがアダプタに結合される配列決定を指す。よって、任意の断片を同じプライマーで配列決定することができ、よって、配列決定はランダムであり得る。

【００２０】

「臨床関連」ＤＮＡの例は、母体血漿中の胎児ＤＮＡおよび患者の血漿中の腫瘍ＤＮＡを含む。別の例は、移植患者の血漿中の移植片関連ＤＮＡの量の測定を含む。さらなる例は、対象の血漿中の造血および非造血ＤＮＡの相対量の測定を含む。この後者の実施形態は、造血組織および／または非造血組織が関与する病理学的プロセスまたは損傷を検出または監視または予知するために使用することができる。

【００２１】

「癌のレベル」（またはより一般的には「疾患のレベル」もしくは「病態のレベル」）という用語は、癌が存在するか（すなわち、存在または不在）、癌のステージ、腫瘍のサイズ、転移があるか、身体の総腫瘍負荷、治療に対する癌の応答、および／または癌の重症度の他の尺度（例えば、癌の再発）を指し得る。癌のレベルは、数字（例えば、確率）または記号、アルファベット文字、および色などの他のしるしであり得る。レベルは、ゼロであり得る。癌のレベルには、前悪性病態または前癌性病態（状態）も含まれ得る。癌のレベルは、様々な方法で使用することができる。例えば、スクリーニングにより、癌を有することを今まで知らなかった人物において癌が存在するかどうかをチェックすることができる。評価は、癌と診断されている人物を調べて、癌の進行を経時的に監視し、療法の有効性を研究し、または予後を決定することができる。一実施形態において、予後は、患者が癌で死亡する可能性、または特定の持続時間または特定の時間の後に癌が進行する可能性、または癌が転移する可能性として表すことができる。検出は、「スクリーニング」を意味することができ、または癌の示唆的な特徴（例えば、症状または他の陽性検査）を有する人物が癌を有するかどうかをチェックすることを意味し得る。「病理学のレベル」とは、病原体に関連する病理学のレベルを指すことができ、そのレベルは癌について上で説明したとおりであり得る。疾患／病態のレベルはまた、癌について上で説明したとおりであり得る。癌が病原体に関連している場合、癌のレベルは病理学のレベルの一種になり得る。

【００２２】

本明細書で使用される「染色体異数性」という用語は、二倍体ゲノムの量からの染色体の定量的な量の変動を意味する。変動は、増加または減少であり得る。これは、１つの染色体の全体または染色体の領域を含み得る。

【００２３】

本明細書で使用される「配列不均衡」または「異常」という用語は、参照量からの臨床関連染色体領域の量における少なくとも１つのカットオフ値によって定義される任意の有意な偏差を意味する。配列不均衡は、染色体投与量不均衡、対立遺伝子不均衡、突然変異投与量不均衡、コピー数不均衡、ハプロタイプ投与量不均衡、および他の同様の不均衡を含むことができる。一例として、対立遺伝子不均衡は、腫瘍が、欠失した遺伝子の１つの対立遺伝子もしくは増幅した遺伝子の１つの対立遺伝子、またはそのゲノムにおける２つの対立遺伝子の差動増幅を有し、これによって試料において特定の遺伝子座に不均衡を形成する場合、発生し得る。別の例として、患者は、腫瘍抑制遺伝子に遺伝性突然変異を有し得る。患者は次いで、腫瘍抑制遺伝子の非突然変異対立遺伝子が欠失している腫瘍を発

生させ得る。よって、腫瘍内では、突然変異投与量不均衡がある。腫瘍がそのＤＮＡを患者の血漿中に放出する場合、腫瘍ＤＮＡは、血漿中の患者の（正常細胞からの）構成的ＤＮＡと混合されるであろう。本明細書に記載される方法の使用を通して、血漿中のこのＤＮＡ混合物の突然変異投与量不均衡を検出することができる。異常は、染色体領域の欠失または増幅を含むことができる。

【００２４】

哺乳動物ゲノムにおける「ＤＮＡメチル化」とは、典型的には、ＣｐＧジヌクレオチド間でシトシン残基の５’炭素へのメチル基の付加（すなわち、５－メチルシトシン）を指す。ＤＮＡメチル化は、他の文脈、例えば、ＣＨＧおよびＣＨＨにおいてシトシンで生じ得、ここで、Ｈは、アデニン、シトシン、またはチミンである。シトシンメチル化はまた、５－ヒドロキシメチルシトシンの形態であり得る。Ｎ６－メチルアデニンなどの非シトシンメチル化も報告されている。

【００２５】

「分類」とは、試料の特定の特性と関連している任意の数（複数可）または他の特徴（複数可）を指す。例えば、「＋」という記号（または「陽性」という語）は、試料が欠失または増幅を有するものとして分類されることを意味し得る。分類は、二項（例えば、陽性または陰性）であるか、またはより多くのレベルの分類（例えば、１～１０または０～１のスケール）を有することができる。

【００２６】

「カットオフ」および「閾値」という用語は、操作に使用される所定の数を指すことができる。閾値または参照値は、特定の分類が適用される値より上または下の値、例えば、対象が病態を有するかどうか、または病態の重症度などの病態の分類であってもよい。カットオフは、試料または対象の特徴を参照して、または参照せずに、予め決定されてもよい。例えば、カットオフは、検査される対象の年齢または性別に基づいて選択され得る。カットオフは、検査データの出力後に、かつそれに基づいて選択され得る。例えば、試料の配列決定が特定の深度に達するときに、特定のカットオフが使用され得る。別の例として、１つ以上の病態の既知の分類および測定された特性値（例えば、メチル化レベル、統計的サイズ値、または数）を有する参照対象は、異なる病態および／または病態の分類（例えば、対象が病態を有するか）を区別するために参照レベルを決定するために、使用することができる。これらの用語のうちのいずれも、これらの文脈のうちのいずれにおいても使用することができる。当業者によって理解されるように、カットオフは、所望の感度および特異度を達成するように選択することができる。

【００２７】

「部位」（「ゲノム部位」とも呼ばれる）は、単一の塩基位置、または相関する塩基位置の群、例えば、ＣｐＧ部位、または相関する塩基位置のより大きい群であり得る、単一の部位に対応する。「遺伝子座」は、複数の部位を含む領域に対応し得る。遺伝子座は、遺伝子座をその脈絡における部位と等価にするであろうただ１つの部位を含むことができる。

【００２８】

各ゲノム部位（例えば、ＣｐＧ部位）に対する「メチル化指数」は、その部位におけるメチル化を、その部位をカバーするリード数の合計にわたって示す、（例えば、配列リードまたはプローブから決定されるような）ＤＮＡ断片の割合を指し得る。「リード」は、ＤＮＡ断片から得られた情報（例えば、部位のメチル化状態）に対応することができる。リードは、特定のメチル化状態のＤＮＡ断片と優先的にハイブリダイズする試薬（例えば、プライマーまたはプローブ）を使用して、得ることができる。典型的には、このような試薬は、ＤＮＡ分子のメチル化状態、例えば、バイサルファイト変換、またはメチル化感受性制限酵素、またはメチル化結合タンパク質、または抗メチルシトシン抗体に応じて、ＤＮＡ分子を特異的に修飾するか、または特異的に認識するプロセスによる処理後に適用される。別の実施形態において、メチルシトシンおよびヒドロキシメチルシトシンを認識する単一分子配列決定技術を使用して、メチル化状態を解明し、メチル化指数を決定する

ことができる。
【００２９】
領域の「メチル化密度」は、この領域における部位をカバーするリード数の合計で割ったメチル化を示す、領域内の部位におけるリード数を指し得る。この部位は、具体的な特徴を有し得、例えば、ＣｐＧ部位であり得る。したがって、領域の「ＣｐＧメチル化密度」は、この領域におけるＣｐＧ部位（例えば、特定のＣｐＧ部位、ＣｐＧアイランド内またはそれより大きな領域のＣｐＧ部位）をカバーするリード数の合計で割ったＣｐＧメチル化を示すリード数を指す。例えば、ヒトゲノム中の各１００ｋｂビンのメチル化密度は、１００ｋｂ領域へマップされた配列リードによって覆われた全てのＣｐＧ部位の割合として、ＣｐＧ部位の（メチル化されたシトシンに対応する）バイサルファイト処理後に変換されていないシトシンの総数から判定することができる。この分析はまた、５００ｂｐ、５ｋｂ、１０ｋｂ、５０ｋｂ、もしくは１Ｍｂなどの他のビンサイズに対して実施することができる。領域は、全ゲノム、または染色体、または染色体の一部（例えば、染色体腕）であり得る。ＣｐＧ部位のメチル化指数は、領域がそのＣｐＧ部位のみを含む場合、その領域のメチル化密度と同じである。「メチル化シトシンの割合」は、この領域における解析されたシトシン残基の総数、すなわち、ＣｐＧの脈絡外のシトシンを含む、メチル化されている（例えば、バイサルファイト変換後に未変換）ことが示されているシトシン部位「Ｃ」の数を指し得る。メチル化指数、メチル化密度、およびメチル化シトシンの割合は、「メチル化レベル」の例であり、これは、部位におけるメチル化リードの数を含む、他の比を含み得る。バイサルファイト変換とは別に、当業者に既知の他のプロセスは、これらに限定されないが、メチル化状態に敏感な酵素（例えば、メチル化感受性制限酵素）、メチル化結合タンパク質、メチル化状態に敏感なプラットフォームを用いる単一分子配列決定（例えば、ナノポア配列決定（Ｓｃｈｒｅｉｂｅｒ ｅｔ ａｌ．Ｐｒｏｃ Ｎａｔｌ Ａｃａｄ Ｓｃｉ ２０１３；１１０：１８９１０－１８９１５）およびＰａｃｉｆｉｃ Ｂｉｏｓｃｉｅｎｃｅｓ単一分子リアルタイム分析（Ｆｌｕｓｂｅｒｇ ｅｔ ａｌ．Ｎａｔ Ｍｅｔｈｏｄｓ ２０１０；７：４６１－４６５）によるもの）を含み、ＤＮＡ分子のメチル化状態を調べるために使用することができる。
【００３０】
「メチル化認識配列」とは、配列決定プロセス中にＤＮＡ分子のメチル化状態を確認することができる配列決定方法を指し、これにはバイサルファイト配列決定、またはメチル化感受性制限酵素消化、抗メチルシトシン抗体もしくはメチル化結合タンパク質を使用する免疫沈降、またはメチル化状態の解明を可能にする単一分子配列決定が含まれるが、これらに限定されない。「メチル化認識アッセイ」または「メチル化感受性アッセイ」には、ＭＳＰ、プローブに基づく調査、ハイブリダイゼーション、制限酵素消化とそれに続く密度測定、抗メチルシトシン免疫アッセイ、メチル化シトシンまたはヒドロキシメチルシトシンの割合の質量分析調査、配列決定を伴わない免疫沈降などの配列決定および非配列決定に基づく方法の両方が含まれ得る。
【００３１】
「分離値」（または相対存在量）は、２つのＤＮＡ分子量、２つの寄与率、または２つのメチル化レベル（試料（混合物）メチル化レベルおよび参照メチル化レベルなど）などの２つの値を含む差または比に対応する。分離値は、単純な差または比であり得る。例として、ｘ／ｙの直接比はｘ／（ｘ＋ｙ）と同様に分離値である。分離値は、他の因子、例えば、倍数因子を含むことができる。他の例として、値の関数の差または比、例えば、２つの値の自然対数（ｌｎ）の差または比を使用することができる。分離値は、差および／または比を含むことができる。メチル化レベルは、相対存在量の例であり、例えば、メチル化ＤＮＡ分子（例えば、特定の部位）と他のＤＮＡ分子（例えば、特定の部位にある他の全てのＤＮＡ分子または非メチル化ＤＮＡ分子）との相対存在量である。他のＤＮＡ分子量は、正規化因子として機能することができる。別の例として、全てまたは非メチル化ＤＮＡ分子の強度に対するメチル化ＤＮＡ分子の強度（例えば、蛍光または電界強度）を決定することができる。相対存在量は、体積あたりの強度も含むことができる。

【００３２】
「対照」、「対照試料」、「参照」、「参照試料」、「正常」、および「正常試料」という用語は、特定の病態を持たない試料、またはそうでなければ健康な試料を一般的に説明するために交換可能に使用され得る。一例では、本明細書に開示される方法は、腫瘍を有する対象に対して実施することができ、参照試料は、対象の健康な組織から採取された試料である。別の例では、参照試料は、疾患、例えば、癌または癌の特定ステージを有する対象から採取した試料である。参照試料は、対象またはデータベースから取得され得る。参照は一般に、対象からの試料を配列決定することから得られた配列リードをマップするために使用される参照ゲノムを指す。参照ゲノムは一般に、生体試料および生得的試料からの配列リードが整列および比較され得る、１倍体または２倍体ゲノムを指す。１倍体ゲノムについては、各遺伝子座において１つのヌクレオチドのみが存在する。２倍体ゲノムについては、ヘテロ接合遺伝子座を特定することができ、このような遺伝子座は２つの対立遺伝子を有し、いずれかのアレルが、遺伝子座へのアラインメントの一致を可能にし得る。参照ゲノムは、例えば、１つ以上のウイルスゲノムを含めることにより、ウイルスに対応し得る。

【００３３】
本明細書で使用される「健康である」という語句は一般に、良好な健康状態を有する対象を指す。このような対象は、悪性疾患または非悪性疾患が存在しないことを示す。「健常者」は、通常「健康である」とはみなされない、アッセイされる病態とは無関係の他の疾患または病態を有し得る。

【００３４】
「癌」または「腫瘍」という用語は交換可能に使用され、一般に、組織の異常な塊を指し、その塊の増殖は正常組織の増殖を上回り、協調されない。癌または腫瘍は、以下の特性に応じて「良性」または「悪性」と定義され得る。形態および機能性を含む細胞分化の程度、増殖速度、局所浸潤、および転移。「良性」腫瘍は一般に十分に分化しており、悪性腫瘍よりも特徴的に増殖が遅く、発生部位に局在したままである。さらに、良性腫瘍には、遠隔部位に浸潤、侵襲、または転移する能力を有さない。「悪性」腫瘍は一般に低分化（退形成）であり、周囲組織の進行性浸潤、侵襲、および破壊を伴う特徴的な急速な増殖を示す。さらに、悪性腫瘍には遠隔部位に転移する能力を有する。「ステージ」は、悪性腫瘍の進行状況を説明するために使用されることができる。初期ステージの癌または悪性腫瘍は、後期ステージの悪性腫瘍よりも体内の腫瘍量が少なく、一般的に症状が少なく、予後が良好で、治療成績が良好であることに関連している。後期または進行ステージの癌または悪性腫瘍は、多くの場合、遠隔転移および／またはリンパ拡散に関連している。

【００３５】
「偽陽性」（ＦＰ）という用語は、病態を有さない対象を指し得る。偽陽性とは一般に、腫瘍、癌、前癌性病態（例えば、前癌性病変）、限局性癌または転移性癌、非悪性疾患を有さない、またはそうでなければ健康である対象を指す。偽陽性という用語は一般に、病態を有さないが、本開示のアッセイまたは方法によって病態を有すると特定される対象を指す。

【００３６】
「感度」または「真陽性率」（ＴＰＲ）という用語は、真陽性の数を真陽性と偽陰性の数の合計で割ったものを指し得る。感度は、真に病態を有する集団の割合を正確に特定するアッセイまたは方法の能力を特徴付けることができる。例えば、感度は、癌を有する集団内の対象の数を正しく特定する方法の能力を特徴付けることができる。別の例では、感度は、癌を示す１つ以上のマーカーを正確に特定する方法の能力を特徴付けることができる。

【００３７】
「特異度」または「真陰性率」（ＴＮＲ）という用語は、真陰性の数を真陰性と偽陽性の数の合計で割ったものを指し得る。特異度は、真に病態を有さない集団の割合を正確に特定するアッセイまたは方法の能力を特徴付けることができる。例えば、特異度は、癌を

有さない集団内の対象の数を正しく特定する方法の能力を特徴付けることができる。別の例では、特異度は、癌を示す１つ以上のマーカーを正しく特定する方法の能力を特徴付けることができる。

【００３８】

「ＲＯＣ」または「ＲＯＣ曲線」という用語は、受信者動作特性曲線を指し得る。ＲＯＣ曲線は、二項分類システムの性能をグラフィカルに表現されることができる。任意の所与の方法について、様々な閾値設定で感度を特異度に対してプロットすることにより、ＲＯＣ曲線を生成することができる。対象における腫瘍の存在を検出する方法の感度および特異度は、対象の血漿試料中の腫瘍由来核酸の様々な濃度で決定され得る。さらに、得られた３つのパラメータ（感度、特異度、閾値設定など）のうちの少なくとも１つ、およびＲＯＣ曲線によって、任意の不明なパラメータの値または期待値を決定し得る。不明なパラメータは、ＲＯＣ曲線に適合した曲線を使用して決定され得る。「ＡＵＣ」または「ＲＯＣ－ＡＵＣ」という用語は一般に、受信者動作特性曲線下の領域を指す。このメトリックは、方法の感度と特異度の両方を考慮して、方法の診断的有用性の尺度を提供し得る。一般的に、ＲＯＣ－ＡＵＣの範囲は０．５～１．０であり、０．５に近い値は方法の診断的有用性が限られていること（例えば、低感度および／または低特異度）を示し、１．０に近い値は方法の診断的有用性が高いこと（例えば、高感度および／または高特異度）を示す。例えば、参照により本明細書に組み込まれる、Ｐｅｐｅ ｅｔ ａｌ，“Ｌｉｍｉｔａｔｉｏｎｓ ｏｆ ｔｈｅ Ｏｄｄｓ Ｒａｔｉｏ ｉｎ Ｇａｕｇｉｎｇ ｔｈｅ Ｐｅｒｆｏｒｍａｎｃｅ ｏｆ ａ Ｄｉａｇｎｏｓｔｉｃ，Ｐｒｏｇｎｏｓｔｉｃ，ｏｒ Ｓｃｒｅｅｎｉｎｇ Ｍａｒｋｅｒ，”Ａｍ．Ｊ．Ｅｐｉｄｅｍｉｏｌ ２００４，１５９（９）：８８２－８９０を参照されたい。尤度関数、オッズ比、情報理論、予測値、キャリブレーション（適合度を含む）、および再分類測定を使用して診断的有用性を特徴付ける追加のアプローチは、参照によりその全体が本明細書に組み込まれる、Ｃｏｏｋ，“Ｕｓｅ ａｎｄ Ｍｉｓｕｓｅ ｏｆ ｔｈｅ Ｒｅｃｅｉｖｅｒ Ｏｐｅｒａｔｉｎｇ Ｃｈａｒａｃｔｅｒｉｓｔｉｃ Ｃｕｒｖｅ ｉｎ Ｒｉｓｋ Ｐｒｅｄｉｃｔｉｏｎ，”Ｃｉｒｃｕｌａｔｉｏｎ ２００７，１１５：９２８－９３５に要約されている。

【００３９】

「約」または「およそ」という用語は、当業者によって決定される特定の値の許容誤差範囲内を意味し得、これは値の測定または決定方法、すなわち測定システムの制限について部分的に依存する。例えば、「約」は、当技術分野の慣例により、１以内または１を超える標準偏差を意味し得る。あるいは、「約」は、所与の値の最大２０％、最大１０％、最大５％、または最大１％の範囲を意味し得る。あるいは、特に生物学的システムまたはプロセスに関して、「約」または「およそ」という用語は、値の１桁以内、５倍以内、より好ましくは２倍以内を意味し得る。本出願および特許請求の範囲に特定の値が記載されている場合、特に明記しない限り、特定の値の許容誤差範囲内の「約」という用語を想定すべきである。「約」という用語は、当業者によって一般に理解されている意味を有し得る。「約」という用語は±１０％を指し得る。「約」という用語は、±５％を指し得る。

【００４０】

本明細書で使用される用語は、特定のケースのみを説明する目的のものであり、限定することを意図したものではない。本明細書で使用される場合、単数形「ａ」、「ａｎ」、および「ｔｈｅ」は、文脈上明らかに別途指示されない限り、複数形も含むことを意図している。「または」の使用は、それとは反対に具体的に示されない限り、「排他的なまたは」ではなく「包含的なまたは」を意味することが意図される。「～に基づいて」という用語は、「～に少なくとも部分的に基づいて」を意味することを意図している。さらに、「～を含む（ｉｎｃｌｕｄｉｎｇ）」、「～を含む（ｉｎｃｌｕｄｅｓ）」、「～を有する（ｈａｖｉｎｇ）」、「～を有する（ｈａｓ）」、「～とともに（ｗｉｔｈ）」という用語、またはその変形は、詳細な説明および／または特許請求の範囲のいずれかで使用される限りでは、そのような用語は、「～を含む（ｃｏｍｐｒｉｓｉｎｇ）」という用語と

同様の様態で包括的であることを意図している。
【発明を実施するための形態】
【００４１】
無細胞ＤＮＡのサイズベース分析は、染色体異数性および癌についての生体試料の分析に使用されている。しかしながら、以前のサイズベース技法では、生体試料が低いパーセンテージの臨床関連ＤＮＡを有する場合、統計的に有意な結果を得ることは、困難であり得る。臨床関連ＤＮＡの画分が低い場合、以前のサイズベース分析は、単一の分析技法として依存されるのではなく、別のタイプの分析の結果を確認するために使用され得る。本発明の実施形態は、サイズバンドの使用を含み、これは、より多くの無細胞ＤＮＡを分析に使用することを可能にし得、サイズのパターンを分析することを可能にし得る。結果として、サイズベース分析は、低画分の臨床関連ＤＮＡでも正確に行われ得る。
【００４２】
この研究では、我々は、ＮＩＰＴに必要な胎児ＤＮＡ画分の限度を下げるために、無細胞ＤＮＡのサイズ分析を適用することを目的とした。我々は、特異度に悪影響を与えることなく、ＮＩＰＴの感度を改善することを目的とする。同様の技法は、癌分析に適用され得る。特定のサイズの代わりに複数のサイズ範囲を使用することは、臨床関連ＤＮＡの画分が低い場合でも、生体試料の分析を可能にすることが見出された。実施形態は、染色体領域がコピー数異常（ＣＮＡ）を示すかを決定するためにサイズバンドを使用することを含み得る。ＣＮＡは、異数性または癌に関し得る。実施形態はまた、癌のレベルを決定するためにサイズバンドを使用することを含み得る。
【００４３】
Ｉ．サイズベース分析
母体血漿中の胎児由来分子が母体ＤＮＡ分子よりも短いことが実証されている（Ｃｈａｎ ｅｔ ａｌ．Ｃｌｉｎ Ｃｈｅｍ ２００４；５０：８８－９２；Ｌｏ ｅｔ ａｌ．Ｓｃｉ．Ｔｒａｎｓｌ．Ｍｅｄ．２０１０；２：６１ｒａ９１－６１ｒａ９１）。研究者らは、そのようなサイズの違いを利用して、ＮＩＰＴの母体血漿試料中の胎児ＤＮＡを濃縮した（Ｌｉ ｅｔ ａｌ．Ｃｌｉｎ Ｃｈｅｍ ２００４；５０：１００２－１０１１，ＪＡＭＡ ２００５；２９３：８４３－９；Ｌｕｎ ｅｔ ａｌ．Ｐｒｏｃ．Ｎａｔｌ．Ａｃａｄ．Ｓｃｉ．Ｕ．Ｓ．Ａ．２００８；１０５：１９９２０－５）。Ｙｕ ｅｔ ａｌ．は、胎児染色体異数性が、ペアエンド配列決定データにおいて異数体染色体から短い断片の異常な割合を決定することによって検出され得ることを例示した（Ｙｕ ｅｔ ａｌ．Ｐｒｏｃ．Ｎａｔｌ．Ａｃａｄ．Ｓｃｉ．Ｕ．Ｓ．Ａ．２０１４；１１１：８５８３－８）。そのようなアプローチは、母体血漿中のＤＮＡ分子の数と比較される場合、良好なＮＩＰＴ性能を達成することができる（Ｙｕ ｅｔ ａｌ．Ｐｒｏｃ．Ｎａｔｌ．Ａｃａｄ．Ｓｃｉ．Ｕ．Ｓ．Ａ．２０１４；１１１：８５８３－８）。
【００４４】
低胎児ＤＮＡ画分（例えば、＜４％）を有する妊婦における胎児染色体異常の非侵襲的検出の精度を改善するために、以前に検討されてきた１つの可能な方法は、インシリコサイズ選択または物理サイズ選択を通した短いＤＮＡ分子の選択的分析である（例えば、全ての目的で参照により本明細書に組み込まれる、２００８年７月２３日に出願された、ＷＯ２００９／０１３４９６）。これらの方法では、短い血漿ＤＮＡ分子からのデータまたは分子は、統計分析、疾患分類、および症例解釈の基礎を形成する。胎児由来ＤＮＡ分子は、母体由来のものと比較してより短いサイズ分布が短いため、短ＤＮＡ断片の選択的分析は、胎児由来ＤＮＡ分子を優先的に濃縮し、より高い胎児ＤＮＡ画分をもたらし得る。
【００４５】
胎児ＤＮＡ画分は、ＮＩＰＴ性能を左右する重要な因子であるため、これは、ＮＩＰＴの精度を潜在的に改善し得る。しかしながら、１５０ｂｐ未満の長さを有する配列決定されたリードのインシリコ選択は、有効な胎児ＤＮＡ画分を増加させ得るが、胎児ＤＮＡ画分とカウントされている分子の数との間のトレードオフのため、単一分子カウントによる異数性検出の感度は必ずしも増加させないであろうことが報告された（Ｆａｎ ｅｔ ａ

ｌ．Ｃｌｉｎ．Ｃｈｅｍ．２０１０；５６：１２７９－１２８６）。言い換えると、表１において示されるように、カウントされる血漿ＤＮＡ断片の数の著しい低減のため、短いＤＮＡを選択することでの以前のアプローチは、配列決定深度を増加させることなく感度を改善することができなかった。分析される血漿ＤＮＡ断片のサイズを低減することは、分析されるＤＮＡ断片の数を低減する。例えば、１００ｂｐ未満の長さのみが分析される場合、ＤＮＡ断片は、４８．５倍の低減を受ける。同時に、より小さな血漿ＤＮＡ断片に焦点を当てることによって、胎児ＤＮＡ画分が濃縮される。例えば、１００ｂｐ未満の長さについて、胎児ＤＮＡ画分は、１．７８倍の濃縮を有する。しかしながら、１．７８倍の濃縮は、分析されている血漿ＤＮＡ分子における４８．５倍の低減と比較して小さい。

【表１】

血漿 DNA サイズ（bp）	胎児 DNA 画分における濃縮倍率（x）	血漿 DNA 分子における低減倍率分析中（x）
<150	1.93	4.67
<120	2.04	21.2
<110	1.91	32.3
<100	1.78	48.5

表 1

【００４６】

一方で、我々は、特定の閾値未満、例えば、１５０ｂｐのサイズのＤＮＡ分子を活用することによって診断特異度を改善するために、別の血漿ＤＮＡサイズベースのアプローチ（米国特許８，６２０，５９３）を以前に開発した。この方法では、潜在的な異数性染色体に由来する血漿ＤＮＡ分子の平均サイズが、他の染色体に由来する血漿ＤＮＡ分子の平均サイズと比較される。胎児染色体異数性は、過大表現された染色体（例えば、トリソミック染色体）からの血漿ＤＮＡ分子の平均サイズの短縮または過小表現された染色体（例えば、モノソミック染色体）についての血漿ＤＮＡ分子の平均サイズの延長をもたらすので、このアプローチは、染色体異数性の非侵襲的な検出の特異度を改善することが示されている。しかしながら、そのようなアプローチは、カウントされる血漿ＤＮＡ分子の数の低減のため、感度を向上させることを期待することができなかった。

【００４７】

異数性染色体のコピー数変化を定量化するために、特定の短いＤＮＡ分子のインシリコ選択を使用することを試みる以前の取り組みがいくつかあった（Ｆａｎ　ｅｔ　ａｌ．Ｃｌｉｎ．Ｃｈｅｍ．２０１０；５６：１２７９－１２８６）。しかしながら、そのような特定のサイズを選択は、最終的な臨床分類に寄与するであろうＤＮＡ分子の数を低減し、したがって確率的変動を増加させるであろう。分析的に、そのような確率的変動の増加は、変動係数（ＣＶ）または標準偏差（ＳＤ）の増加として明示され得る。ポアソン分布によると、分析される分子の数の４倍低減毎に、ＣＶは２倍増加するであろう。一方で、循環胎児ＤＮＡの画分濃度の２倍増加毎に、胎児染色体異数性の正しい診断に到達するためにカウントする必要がある分子の数は、４倍減少するであろう。１５０ｂｐ未満の分子にサイズ選択を使用する場合、胎児ＤＮＡ画分は、約２倍増加するが、血漿ＤＮＡ分子の数

は、４．７倍減少するであろう。したがって、単純なサイズ選択による胎児ＤＮＡ画分の濃縮は、血漿ＤＮＡ分子の低減の有害な影響を効果的に相殺することができず、これは、単純なインシリコサイズ選択によるＮＩＰＴにおける一貫した改善がなかった重要な理由であり得る（Ｆａｎ　ｅｔ　ａｌ．Ｃｌｉｎ．Ｃｈｅｍ．２０１０；５６：１２７９－１２８６）。

【００４８】

ＩＩ．サイズパターン

この研究では、我々は、一連の異なるサイズ範囲にわたる分子数の詳細な変化するパターンを利用することによって、血漿ＤＮＡサイズ情報を組み込む新たな方法を開発し、これは、経験的データによれば、驚くべきことに、検査感度の改善をもたらした。血漿ＤＮＡ分子をより多くのサイズバンドに分画する場合、サイズバンドあたりの配列決定されたＤＮＡ分子がはるかに少なくなり、単独の各バンド内の血漿ＤＮＡ分子が感度を改善することができなかったため、これは直観に反する。１つの特定のバンドを単独で使用する代わりに、我々の新たなアプローチは、性能を改善するために異なるバンドにわたる関係を使用することである。

【００４９】

我々は、異数体染色体のゲノム表現（ＧＲ）の変化が、異なるサイズの血漿ＤＮＡ分子中に存在する測定された胎児ＤＮＡ画分に従って変動するであろうと推論した。無細胞胎児および母体ＤＮＡサイズは、２つの別個の断片化パターンを反映するため、我々は、影響を受けた染色体のＧＲ変化間の関係が、非ランダムな方法で異なるサイズ範囲（サイズバンド）に関連しているであろうと仮定した（Ｌｏ　ｅｔ　ａｌ．Ｓｃｉ．Ｔｒａｎｓｌ．Ｍｅｄ．２０１０；２：６１ｒａ９１－６１ｒａ９１）。したがって、我々は、異なるサイズバンドの中で異常染色体に由来するＧＲ値の詳細な変化する形状を分析するための新たなアプローチを開発した。このアプローチの概略原理は、図１において例示される。

【００５０】

図１は、血漿ＤＮＡサイズバンド分析の原理の概略図１００を示す。母体血漿は、それぞれ、胎児および母体細胞に由来する胎児ＤＮＡ分子（セクション１０４および分子１０６における赤波線）および母体ＤＮＡ分子（セクション１０８および分子１１０における黒波線）の混合物を含む。胎児ＤＮＡ分子は一般に、胎児ＤＮＡサイズプロファイルが母体ＤＮＡ分子のそれに対して左にシフトしていることによって証明されるように、母体のものよりも短い。したがって、測定された胎児ＤＮＡ画分は、異なるサイズバンドによって変化し、一般により短いサイズ範囲において濃縮するであろう。よって、トリソミック胎児を妊娠している女性について、参照群からのその導出をｚスコアによって測定することができる、測定されたゲノム表現（ＧＲ）は、異なるサイズバンドによって変動することが予想されるが、対照的に、正倍数体胎児の妊娠では特定の変化は起こらないであろう。

【００５１】

図１は、サイズバンドを離散バンドおよびスライディングウィンドウの両方として示す。頻度対サイズのグラフにおいて、異なる色付き列（例えば、列１１２）は、離散サイズ範囲に対応するサイズバンドを示す。ｚスコア（ｃｈｒ２１）対サイズのグラフ１１６および１１８において、色付き列（例えば、列１２２および列１２４）は、異なるサイズバンドのｚスコアを示す。ｚスコア対サイズのグラフにおける線１２６および１２８は、サイズバンドの結果をスライディングウィンドウとして示す。異数体胎児の妊娠では、線１２８は、特定のサイズを中心とするサイズバンドのｚスコアを示す。例えば、線１２８上に所与のｘ座標およびｙ座標を有するデータ点は、ｘ座標によって示されるサイズを中心とするサイズの範囲についてｙ座標によって示されるｚスコアを有する。各ｚスコアは、サイズバンド全体について計算されたプールされたｚスコアである。よって、正倍数体胎児の妊娠のグラフ１１６において、線１２６は、サイズバンドの結果をスライディングウィンドウとして示す。異数体胎児の妊娠のグラフ１１８において、線１２８は、サイズバンドの結果をスライディングウィンドウとして示す。

【００５２】
サイズバンドが離散またはスライディングウィンドウに基づくかにかかわらず、サイズバンドのｚスコアの形状またはパターンは、正倍数体胎児の妊娠と異数体胎児の妊娠との間で明確に異なる。例えば、グラフ１１６およびグラフ１１８において示されるように、異数体胎児の妊娠は、正倍数体胎児の妊娠におけるより周期的なパターンと比較した二峰性のパターンを示す。
【００５３】
異なるサイズバンドにわたる数のパターンは、胎児ＤＮＡ画分、腫瘍ＤＮＡ画分、または他の臨床関連ＤＮＡ画分に関連し得る。よって、異なるサイズバンドにわたる一連の分子数および異なるサイズバンドベースの読み出し間の関係を同時に定量化するこの新たなアプローチは、特定のサイズのＤＮＡ分子のみを使用するアプローチと比較して、血漿ＤＮＡサイズ特性を統合する場合、血漿ＤＮＡ分子を喪失しないであろう。そのような同時定量化は、特定のサイズカットオフ未満の単一読み出しのみの使用と比較して、精度を改善するであろう。血漿におけるコピー数変化のサイズバンドパターンは、これらに限定されないが、人工ニューラルネットワーク、ｋ最近傍アルゴリズム、サポートベクターマシン、および混合ガウスモデルなどのような機械学習アプローチの使用で認識することができる。
【００５４】
Ａ．サイズパターンデータ分析の検証
サイズパターン（すなわち、特定のサイズバンドでの無細胞ＤＮＡの量に関連する画分またはパラメータの形状）は、無細胞ＤＮＡの特性に依存し得る。例えば、サイズパターンは、図１におけるグラフ１１６および１１８のように、生体試料が異数体胎児からの無細胞ＤＮＡを含むかに依存し得る。第１に、異なるサイズのＤＮＡの胎児ＤＮＡ画分は、特定のサイズの無細胞ＤＮＡが、母体ＤＮＡと比較して胎児ＤＮＡについて濃縮されることを示すために分析される。第２に、異数体胎児を有する妊婦からのデータは、正倍数体胎児を有する妊婦からのデータに対するサイズバンドを用いて分析される。これらの分析は、ＣＮＡが異数体胎児の結果である場合を含め、ＣＮＡの差を区別するためにサイズパターンを分析することができることを確認する。
【００５５】
１．測定された胎児ＤＮＡ画分は、異なるサイズバンドによって変動する
胎児ＤＮＡ画分変化が断片サイズによって非ランダムな方法で変動しないであろうという仮説を検証するために、我々は、我々の以前の研究において記載されるデータを再分析した（Ｃｈａｎ　ｅｔ　ａｌ．Ｐｒｏｃ．Ｎａｔｌ．Ａｃａｄ．Ｓｃｉ．２０１６；１１３：Ｅ８１５９　－Ｅ８１６８）。
【００５６】
図２Ａは、５０～４００ｂｐの範囲である、血漿ＤＮＡ断片のサイズについての異数体染色体の測定された胎児ＤＮＡ画分を示す。Ｘ軸は、ＤＮＡ分子のサイズであり、Ｙ軸は、胎児ＤＮＡであるそのサイズでのＤＮＡ分子の画分である。例えば、１２０ｂｐのサイズで、胎児ＤＮＡ画分は、７０．５％であり、これは、１２０ｂｐのサイズを有するＤＮＡ分子のうち、７０．５％が胎児由来であり、それらの２９．５％が妊婦由来であることを意味する。胎児ＤＮＡ画分は、男性胎児を有する妊婦からの試料の染色体Ｙパーセンテージから決定された。胎児ＤＮＡ画分は、それぞれ、１２０ｂｐおよび２８０ｂｐのサイズで濃縮されていることがわかった。７０．５％の最大胎児ＤＮＡ画分は、１２０ｂｐのサイズで見られ、これは、１７．４％の胎児ＤＮＡ画分で２００ｂｐサイズでの最低のものよりも４倍高い。
【００５７】
２．血漿ＤＮＡにおけるＣＮＡは、異なるサイズバンドについて変動する
不均一なパターンを示す胎児ＤＮＡ画分の変化は、異数体染色体に由来する分子数の表現に影響を与えるであろう。異数体染色体は、異常な数の染色体を有する。胎児における異常な数の染色体は、母体ＤＮＡと比較して胎児ＤＮＡの量に影響するであろう。例えば

、２１トリソミーは、２つのみの代わりに、３つの２１染色体を有する。胎児が２１トリソミーを有する場合、胎児ＤＮＡは、正常正倍数体胎児よりも高い画分を有する。胎児ＤＮＡは多くの場合、母体ＤＮＡよりも短いため、２１トリソミーを有する胎児を妊娠している女性の母体試料は、正倍数体胎児を妊娠している女性の母体試料と比較して、より高い濃度の２１染色体からの短いＤＮＡを有する可能性が高いであろう。

【００５８】

図２Ｂは、２１トリソミー胎児の妊娠および正倍数体胎児の妊娠のサイズバンドスライディングウィンドウを用いてｚスコア結果を示す。サイズバンドスライディングウィンドウの帯域幅は、５０ｂｐであった。２１トリソミー胎児の妊娠は、４％の胎児ＤＮＡ画分を有した。図２Ｂにおいて見られるように、２１トリソミー胎児の１２０ｂｐ位置は、分析された全ての試料の中で最も高いｚスコアを有し、したがって測定されたコピー数異常の最も高い程度に対応した。異なるサイズバンドは、１２０ｂｐおよび他のサイズでのｚスコアの規模に影響するであろう。影響を受けた染色体のｚスコアの計算は、以下に記載される。

【００５９】

５０ｂｐ帯域幅を有するサイズバンドの中点が、長さｉに位置する（例えば、サイズバンドの中点が７５ｂｐのｉに位置し、バンドが５０～１００ｂｐの範囲にある）と仮定して、標的化染色体（例えば、２１染色体）にマップする配列決定リードのパーセンテージは、ゲノム表現ｉ（すなわち、GR_i）として示される、目的の特定のサイズ範囲（例えば、５０～１００ｂｐ）内のそのような断片を用いて計算することができる。長さｉのｚスコアは、下記式で計算され、

【数１】

$$Zスコア_i = \frac{GR_i - M_i}{SD_i}$$

式中、M_iおよびSD_iは、長さｉを中心とするサイズバンドの標的化染色体のゲノム表現の平均値および標準偏差を表し、これは、この研究では、正倍数体胎児を有する５０の妊娠から推測された。サイズの完全なスペクトルは、５０～４００ｂｐの範囲である、サイズプロファイルにおけるサイズバンドの中点の位置を動的に変化することによって調査されるであろう。

【００６０】

図２Ｂでは、我々は、２１トリソミー胎児での妊娠についてのサイズバンドベースのｚスコア曲線２０２において規則的な波状パターンを観察することができる。この観察は、異なるサイズバンドにおける胎児のＤＮＡ画分の変化を連想させた。しかしながら、正倍数体胎児での対照群においてそのようなパターンは示されなかった。特定のサイズバンドにおけるそのような変化の規模は、胎児ＤＮＡ画分の変化とは異なるように思われた。例えば、１２０　ｂｐでのｚスコアは、２８０ｂｐでのｚスコアよりもはるかに高かったが（図２Ｂ）、胎児ＤＮＡ画分は、これら２つのサイズ間で同等であった（図２Ａ）。ばらつきは、１６６ｂｐよりも短い長さと比較して、１６６ｂｐよりも長い長さでより急速に減少する分子数の結果であり得、高いサンプリング変動は、長い分子に存在するであろう。

【００６１】

図２Ｂはまた、ｘ軸上の「全て」とラベル付けされた値に対応する円として例示された、全てのサイズについてのｚスコアを示す。最も高い円である、赤色円２０４は、２１トリソミーに対応する。赤色円２０４は、３未満のＺスコアを有する。よって、全ての断片を使用し、カットオフとして３のｚスコアを使用する場合、この症例は、誤って正倍数体胎児として分類され、偽陰性の結果をもたらすであろう。対照的に、異なるサイズバンドに対して変動するｚスコアの変化の別個の形状を使用する場合、症例は、対照群との比較において２１トリソミー症例として正しく特定することができる。

【００６２】

Ｂ．サイズパターン分析の適用

サイズパターンデータは、正倍数体胎児または異数体胎児のいずれかを妊娠している女性について生成された。データは次いで、正倍数体胎児での妊娠と異数体胎児での妊娠との間を区別するためにサイズパターンを使用することができたかを決定するために、機械学習モデルを使用することを含む、異なる技法によって分析された。

【００６３】

１．血漿中のＣＮＡのサイズバンド形状は、低胎児画分を有する染色体異数性を知らせる

そのようなサイズバンドベースのｚスコアパターンを低胎児ＤＮＡ画分を有する他の試料に一般化することができるかを評価するために、我々は、各々２１トリソミーの胎児での４８症例および各々正倍数体胎児での６３症例を含む、各々男性胎児での追加の１１１個の母体血漿ＤＮＡ試料を分析した。胎児ＤＮＡ画分は、男性胎児に由来するＹ染色体配列を用いて推定された（Ｈｕｄｅｃｏｖａ ｅｔ ａｌ．ＰＬｏＳ Ｏｎｅ ２０１４；９：ｅ８８４８４；Ｃｈｉｕ ｅｔ ａｌ．ＢＭＪ ２０１１；３４２：ｃ７４０１）。４％以下の低胎児ＤＮＡ画分を有する十分な症例を有するために、トリソミック胎児での４８の妊娠についての各ペアエンド配列決定データセットは、４％胎児ＤＮＡ画分以下のレベルを達成するために正倍数体胎児での症例からの配列決定データセットとインシリコで混合された。

【００６４】

図３は、４％の胎児ＤＮＡ画分を有する異なる個々の妊娠にわたる異数体染色体について測定されたゲノム表現（ＧＲ）のサイズバンドベースの変化するパターンを示す。Ｙ軸は、ｚスコア値を示し、正倍数体胎児を妊娠している女性と比較した異数体胎児を妊娠している女性における測定されたＧＲについての導出の程度を示した。Ｘ軸は、異なるサイズバンドを示した。赤線（またより濃い線）は、トリソミック胎児での妊娠を表し、灰色線は、正倍数体胎児での妊娠を表した。

【００６５】

図３は、トリソミック胎児でのほとんど全ての症例が、正倍数体胎児での症例からのものと比較して、測定されたコピー数異常の一貫して異なるサイズバンドベースのパターンを示したことを示す。各症例では、２１トリソミー症例のサイズパターンについての線は、正倍数性症例のパターンとは明らかに異なり、これは、図２Ｂに示されるように、２１トリソミーが全てのサイズ断片についてｚスコアを用いることよりも容易に決定されることを可能にすることができる。

【００６６】

我々は、トリソミックおよび正倍数体症例を有する妊娠間のデータ構造を視覚化するために、ヒートマップおよびｔ－ＳＮＥ（ｔ分布確率的近傍埋め込み）アプローチをさらに使用した。図４Ａは、正倍数体および２１トリソミー胎児での妊娠間のサイズバンドベースの変化するパターンのヒートマッププロットを示す。青色（例えば、エリア４０２）は、正倍数体を示すサイズバンドの特徴についてであり、緑色（例えば、エリア４０４）は、２１トリソミーを示すサイズバンドの特徴についてである。図４Ａにおけるほぼ全ての症例（４６／４８、９６％）は、２１トリソミー胎児症例を一緒にクラスター化することを含む。同様に、正倍数体胎児を含む図４Ａにおけるほぼ全ての症例（６２／６３、９８％）は、一緒にクラスター化された。

【００６７】

図４Ｂは、正倍数体および２１トリソミー胎児での妊娠間のサイズバンドベースの変化するパターンのｔ－ＳＮＥプロットを示す。ｔ－ＳＮＥプロットは、機械学習から決定される２つの特徴に基づく。ｔ－ＳＮＥプロットは、２１トリソミー症例での妊娠を正倍数体症例での妊娠と容易に区別することができるという一貫した結果を与え（図４Ｂ）、血漿ＤＮＡ中の測定されたコピー数異常のサイズバンドベースの形状が４％などの低胎児ＤＮＡ画分を有する症例について染色体異数性を知らせることができたことを示した。

【００６８】

図４Ｃは、正倍数体および２１トリソミー胎児での妊娠間の従来のｚスコアアプローチ

を用いたｚスコア分布を示す。破線は、３のｚスコア閾値を示す。３のｚスコアカットオフを使用して、２１トリソミーの検出率は、４８％のみであろう。言い換えると、２１トリソミーの５２％は、偽陰性をもたらすであろう。加えて、図４Ｃは、１つの正倍数性妊娠が２１トリソミーについて偽陽性をもたらすであろうことを示す。従来のｚスコアアプローチは、図４Ｂにおけるｔ－ＳＮＥアプローチと比較してより低い感度および特異度をもたらし、任意の偽陽性または偽陰性を生成しなかった。

【００６９】

２．低胎児ＤＮＡ画分を有する症例を検出するための機械学習パターン認識

我々は、胎児コピー数異常を検出するためのサイズバンドベースのアプローチの使用をさらに実証するために、ニューラルネットワークモデルを利用した。我々は、試料を訓練および検査データセットに分けた。訓練データセットは、２１トリソミー胎児での３３の妊娠および正倍数体胎児での６３の症例を含み、検査データセットは、１５の２１トリソミー胎児および５０の生倍数体胎児を含有した。各々２０ニューロンを有する１つの層で構成されたニューラルネットワークは、サイズバンドにおける隠されたパターンを捕捉するモデルを学習するために使用された。その後、我々は、このモデルを検査データセットに適用した。

【００７０】

図５は、ニューラルネットワークモデルの訓練データセットおよび検査データセットを示す。２１トリソミーの確率について０．７のカットオフで、我々は、それぞれ、１％、２％、３％、および４％の胎児ＤＮＡ画分について９８％の特異度で４０％、８０％、１００％、および１００％感度を達成することができたことがわかった。１％の低胎児ＤＮＡ画分でも、ニューラルネットワークモデルは、２１トリソミーの真陽性を特定する能力を示す。

【００７１】

ニューラルネットワークモデル以外の機械学習モデルは、対象における胎児異数性または癌の確率を決定することができるパターンおよび特徴を決定するために使用され得る。これらの機械学習モデルの訓練は、障害または臨床関連特徴によって影響されるものおよびそうでないものからの試料を含むデータセットを使用することができる。訓練に考慮され得るパラメータは、サイズバンドの帯域幅、サイズバンドの中心点、ＤＮＡ分子の量、ＤＮＡ分子の場所、エピゲノム信号（例えば、メチル化）、および他の変数を含む。

【００７２】

３．コピー数異常を検出するための例示的な方法

図６は、染色体領域が対象からの生体試料においてコピー数異常を示すかを決定する方法６００を示す。生体試料は、臨床関連ＤＮＡ分子および他のＤＮＡ分子を含む無細胞ＤＮＡ分子の混合物を含み得る。臨床関連ＤＮＡ分子は、胎児ＤＮＡまたは母体ＤＮＡを含み得る。臨床関連ＤＮＡ分子が胎児ＤＮＡを含む場合、他のＤＮＡは、母体ＤＮＡを含み得る。臨床関連ＤＮＡ分子が母体ＤＮＡを含む場合、他のＤＮＡは、胎児ＤＮＡを含み得る。臨床関連ＤＮＡは、腫瘍ＤＮＡを含み得、他のＤＮＡ分子は、非腫瘍ＤＮＡを含む。

【００７３】

ブロック６０２で、方法６００は、複数のサイズ範囲の各サイズ範囲についてのサイズ範囲に対応する生体試料からの無細胞ＤＮＡ分子の第１の量を測定することを含み得る。無細胞ＤＮＡ分子は、染色体または染色体の一部であり得る、特定のゲノム領域に由来し得る。例えば、ゲノム領域は、染色体腕であり得る。ゲノム領域は、ゲノムからの任意の領域であり得る。いくつかの実施形態では、無細胞ＤＮＡ分子は、複数のばらばらなまたは連続したゲノム領域に由来し得る。サイズ範囲は、本明細書に記載されるサイズバンドであり得る。

【００７４】

使用する特定のサイズ範囲は、機械学習モデルによって決定され得る。機械学習モデルは、データセットで訓練することができ、モデルは、変動することができ、コピー数異常または臨床病態を検出するために感度および特異度を最適化するために、その範囲（例え

ば、サイズ範囲の中心点位置および／または帯域幅）が使用される。データセットは、複数の参照サイズパターンを含み得る。機械学習モデルは、サイズ範囲の特定の帯域幅が有利であることを決定し得る。加えて、機械学習モデルは、特定のサイズ範囲が予測結果にとって他よりも重要であり得ることを決定し得る。例えば、サイズ範囲は、１００ｂｐ～１５０ｂｐの任意のサイズを中心とするスライドするサイズ範囲であると決定され得る。他の実施形態では、機械学習モデルは、離散した重複しないサイズ範囲が、スライドするサイズ範囲に対して改善した結果を提供し得ることを決定し得る。訓練セットの感度および／または特異度または他の精度に関するコスト関数は、機械学習モデルのパラメータおよび特徴選択（例えば、使用するサイズ範囲および特定のサイズ比）を更新するために使用することができる。検証データセットはまた、モデルの精度を確認するために使用することができる。

【００７５】

ブロック６０４で、方法６００は、複数のサイズ範囲の各サイズ範囲について、コンピュータシステムによって、サイズ範囲に対応する無細胞ＤＮＡ分子の第１の量およびサイズ範囲にないサイズを含む第２のサイズ範囲のＤＮＡ分子の第２の量を用いてサイズ比を計算することを含み得る。サイズ比は、無細胞ＤＮＡ分子のｚスコアまたは正規化された量（例えば、画分、パーセンテージ、または相対存在量）であり得る。例えば、サイズ比は、ゲノム表現（ＧＲ）であり得る。他の実施形態では、サイズ比は、ＧＲで計算されたｚスコア（例えば、図２Ｂにおける曲線２０２上の点でのｚスコア値）であり得る。

【００７６】

各サイズ範囲は、帯域幅を有し得、これは、サイズ範囲におけるサイズの範囲の数値を記載する。例えば、帯域幅は、５０ｂｐ～１００ｂｐ、１００ｂｐ～２００ｂｐ、２００ｂｐ～３００ｂｐ、または３００ｂｐ～４００ｂｐの範囲にあり得る。１００ｂｐを中心とする５０ｂｐの帯域幅を有するサイズ範囲は、７５ｂｐ～１２５ｂｐに及ぶであろう。各サイズ範囲は、複数のサイズ範囲の任意の他のサイズ範囲（例えば、図１における列１２２および列１２４などの別個のサイズバンド）と重複していない場合があり得る。他の実施形態では、各サイズ範囲は、複数のサイズ範囲のうちの少なくとも１つの他のサイズ範囲と重複し得る。この方法では、サイズ範囲は、スライディングウィンドウとみなされ得る。スライディングウィンドウは次いで、多くのサイズ（例えば、図１における線１２６または線１２８）にわたって連続するサイズ比値をもたらす。

【００７７】

第２のサイズ範囲は、複数のサイズ範囲の各サイズ範囲よりも大きい場合があり得る。第２のサイズ範囲は、無細胞ＤＮＡ分子の全てのサイズを含み得るか、または測定された無細胞ＤＮＡ分子についてのゲノム領域における無細胞ＤＮＡ分子の全てのサイズを含み得る。第２のサイズ範囲は、ブロック６０２における測定された無細胞ＤＮＡ分子と同じゲノム領域（例えば、同じ染色体（複数可）または染色体腕（複数可））からの無細胞ＤＮＡ分子を含み得る。第２のサイズ範囲はまた、ブロック６０２における測定された無細胞ＤＮＡ分子についてのゲノム領域以外のゲノム領域からの無細胞ＤＮＡ分子を含み得る。例えば、２１トリソミーでは、ブロック６０２で測定された無細胞分子は、染色体２１に由来し得る。この場合では、第２のサイズ範囲は、他の染色体（例えば、参照として、またはゲノム全体にわたって機能する異なる染色体）からの無細胞ＤＮＡ分子を含み得る。方法６００は次いでまた、第２のサイズ範囲にある無細胞ＤＮＡ分子の量を測定することを含み得る。

【００７８】

ブロック６０６で、方法６００は、複数のサイズ範囲の複数の参照サイズ比を含む参照サイズパターンを得ることを含み得る。参照サイズパターンは、コピー数異常を有する対象から、または染色体領域にコピー数異常を有さない対象からの複数の参照試料から決定され得る。例えば、検査されているコピー数異常が胎児異数性に関する場合、参照試料は、正倍数体胎児を有することが知られている対象からのものであり得る。他の実施形態では、参照試料は、胎児異数性を有することが知られている対象からのものであり得る。複

数のサイズ範囲の各参照サイズ比は、生体試料の代わりに参照試料を除いて、ブロック６０４において計算されたサイズ比と同じ方法で決定され得る。例えば、図２Ｂにおいて、参照試料のサイズパターンは、曲線２０２を除いて、図２Ｂにおける曲線のいずれか１つであり得る。参照サイズパターンは、参照試料の全てのサイズパターンの統計的表現であり得る。例えば、参照サイズパターンは、全てのサイズパターンの平均（平均値、中央値、または最頻値）であり得る。例えば、この平均化された参照サイズパターンは、図１における線１２６であり得る。

【００７９】

ブロック６０８で、方法６００は、複数のサイズ比と参照サイズ比とを比較することを含み得る。複数のサイズ比と参照サイズパターンとを比較することは、複数のサイズ比の各サイズ比と対応するサイズ範囲での参照サイズ比とを比較することを含み得る。例えば、複数のサイズ比は、図１における線１２８を構成する点であり得る。いくつかの場合では、複数のサイズ比は、線１２８の一部のみを構成し得る。参照サイズパターンが図１における線１２６であると仮定し、複数のサイズ比と参照サイズパターンとを比較することは、線１２８の点と線１２６の参照点との間の統計的比較を含み得る。

【００８０】

各サイズ範囲の各サイズ比は、対応するサイズ範囲での参照サイズ比と統計的に類似することが決定され得る。統計的類似性は、閾値を用いて決定され得る。閾値は、サイズ比が参照サイズ比にどれだけ近い必要があるかを示し得る。閾値は、参照サイズ比からの特定の数の標準偏差（例えば、１、２、または３）であり得る。いくつかの実施形態では、全てのサイズ比が参照サイズ比と統計的に類似する必要はない。代わりに、サイズ比の最小数は、統計的に類似し得る。例えば、サイズ比の８０％、８５％、９０％、または９５％は、対応する参照サイズ比と統計的に類似し得る。

【００８１】

複数のサイズ比と参照サイズパターンとを比較することは、複数のサイズ比と複数の参照試料から決定される複数の閾値とを比較することを含み得る。例えば、各サイズ範囲は、参照試料の標準偏差に基づき得る、異なる閾値を有し得る。単一のサイズ範囲はまた、異なる閾値を有し得、各閾値は、サイズ比が参照試料とは異なるという異なる確実性レベルに関連している。比較することは、超えられた閾値の数をカウントすることと、数が量または画分（例えば、０．５、０．６、０．７、０．８、または０．９）を超えるかを決定することを含み得る。数が量を超える場合、コピー数異常は、染色体領域によって示されることが決定され得る。

【００８２】

いくつかの実施形態では、複数のサイズ比と参照サイズパターンとを比較することは、複数のサイズ範囲の複数のサイズ比を含むサイズパターンを決定することを含み得る。サイズパターンは、サイズ比とサイズ範囲とを関連付けるグラフであり得る。例えば、サイズパターンは、図１における線１２８、図２Ｂにおける曲線２０２、または図３における２１トリソミー線のいずれかであり得る。サイズパターンは、参照サイズパターンと類似の形状を有すると決定され得る。類似の形状を決定することは、サイズパターンの勾配（例えば、一次導関数）および／または変曲点（例えば、二次導関数）が参照サイズパターンのものと類似することを決定することを含み得る。勾配または変曲点の類似性は、統計的有意性（例えば、特定の数の標準偏差）を示し得る、閾値を用いて決定され得る。

【００８３】

いくつかの実施形態では、複数のサイズ比と参照サイズパターンとを比較することは、ニューラルネットワークを含む、機械学習を用いる比較を含み得る。機械学習モデルは、サイズ比を計算する方法、サイズ比と参照サイズパターンとを比較する方法、および／またはサイズパターンが参照サイズパターンに類似するかを決定する方法を決定するために使用することができる。サイズ比を計算する方法は、サイズ範囲の帯域幅ならびに第２のサイズ範囲のサイズおよび帯域幅を決定することを含み得る。サイズ比と参照サイズパターンとを比較する方法は、異なるサイズ範囲の重み付け、およびサイズパターンの０次、

１次、または２次導関数を使用するかを決定することが含み得る。サイズパターンが参照パターンに類似するかを決定する方法は、類似性の閾値を決定することを含み得る。

【００８４】

参照サイズパターンを取得し、複数のサイズ比と参照サイズパターンとを比較することは、複数のサイズ比を機械学習モデルに入力することを含み得る。機械学習モデルは、複数の参照試料からの複数の訓練サイズパターンを用いて訓練され得る。訓練された機械学習モデル（例えば、ニューラルネットワーク）は、染色体領域に異常を有する試料の確率を出力し得る。

【００８５】

ブロック６１０で、方法６００は、比較に基づいて染色体領域がコピー数異常を示すかを決定することを含み得る。コピー数異常は、２１トリソミー、１８トリソミー、１３トリソミー、および性染色体異数性を含む、異数性であり得る。コピー数異常は、癌の徴候であり得る。方法６００はまた、癌について対象を治療すること、または異数性について計画を立てることを含み得る。

【００８６】

参照サイズパターンが、コピー数異常を有する対象からの複数の参照試料から決定され、比較が、サイズ比またはサイズパターンが参照サイズパターンに類似することを示す場合、染色体領域は、コピー数の異常を示すことが決定され得る。比較がサイズ比またはサイズパターンと参照サイズパターンとの間の差異を示す場合、染色体領域は、コピー数異常を示さないことが決定され得る。いくつかの実施形態では、コピー数異常を示す確率が決定され得る。確率は、サイズ比またはサイズパターンと参照サイズパターンとの類似度または非類似度と相関され得る。確率は、ニューラルネットワークまたは本明細書に記載される任意のモデルを含む、機械学習モデルを用いて決定され得る。

【００８７】

あるいは、参照サイズパターンが、コピー数異常を有さない対象からの複数の参照試料から決定され、比較が、サイズ比またはサイズパターンが参照サイズパターンに類似することを示す場合、染色体領域は、コピー数の異常を示さないことが決定され得る。比較がサイズ比またはサイズパターンと参照サイズパターンとの間の差異を示す場合、染色体領域は、コピー数異常を示すことが決定され得る。

【００８８】

Ｃ．低胎児画分での改善した精度

血漿ＤＮＡにおける測定されたコピー数異常のサイズバンドベースのパターンを活用することによってアプローチの性能をベンチマークするために、我々はまた、従来のｚスコア（Ｃｈｉｕ ｅｔ ａｌ．Ｐｒｏｃ Ｎａｔｌ Ａｃａｄ Ｓｃｉ ＵＳＡ ２００８；１０５：２０４５８－２０４６３）およびサイズ選択方法を用いて、４％、３％、２％、および１％などの異なる胎児ＤＮＡ画分にわたる特異度および感度を計算した。胎児ＤＮＡは、１２０ｂｐで母体血漿ＤＮＡ中に存在する測定された胎児ＤＮＡ画分の最大値を与えたため（図２Ａ）、我々は、約１２０ｂｐでのサイズバンドが全てのＤＮＡ断片を使用するよりも良好な性能を与えるであろうと仮定した。この目的のために、我々は、１０５～１５５ｂｐのサイズバンドを選択し、対応するＺスコアを計算した。

【００８９】

表２は、サイズ選択ありおよびなしの従来のカウントベースの方法と比較した、サイズバンドベースのパターン認識の性能を示す。血漿ＤＮＡにおける測定されたコピー数異常のサイズバンドベースのパターンの使用は、従来のＺスコアおよびサイズ選択アプローチとの比較において優れた性能を与えた。例えば、我々の研究では、３％の胎児ＤＮＡ画分で、測定されたコピー数異常のサイズバンドベースのパターンの認識は、９８％の特異度で１００％感度を与えた。比較として、従来のカウントベースのアプローチは、１０％の感度および９８％の特異度のみを与えた。１５０ｂｐ未満の断片のサイズ選択を使用して、感度は、４３％に改善した。しかしながら、１２０ｂｐまでのさらに短いサイズの断片を選択すると、感度は２０％に低減した。これは、本発明において提案される方法が、サ

イズ選択を用いる既存のアプローチよりもはるかに良好な分析性能を提供することを示す。

【表２】

胎児DNA画分	従来のカウントベースのアプローチ						測定されたコピー数異常のサイズバンドベースのパターン（新たに発明されたアプローチ）	
	サイズ選択あり (<120bp)		サイズ選択あり (<150bp)		サイズ選択なし			
	特異度	感度	特異度	感度	特異度	感度	特異度	感度
4%	96%	47%	100%	75%	98%	48%	98%	100%
3%	96%	20%	100%	43%	98%	10%	98%	100%
2%	96%	16%	100%	14%	98%	2%	98%	80%
1%	96%	6%	100%	6%	98%	2%	98%	40%

表２

【００９０】

増加した精度に加えて、本発明の実施形態は、低減した量の配列決定を可能にし得る。サイズパターンアプローチは、特定のサイズの配列リードを破棄することを含まない場合があり得、結果として、所定の配列決定深度での配列リードが分析に使用される。サイズパターンアプローチはそこで、特定のサイズ範囲でより多くのリードを提供するために追加の配列決定を必要としない場合があり得る。さらに、特定の低レベルの胎児画分でより高い配列決定深度でも、サイズバンドまたはサイズパターンを使用しないアプローチは、２１トリソミーを依然として正確に決定しない場合があり得る。サイズバンドまたはサイズパターンが分析されない場合、低胎児画分は、２１トリソミーと正倍数性症例との間の統計的に有意なサイズ差をもたらさない場合があり得る。さらに、サイズバンドまたはサイズパターンなしでのサイズ選択を用いる既存のアプローチは、他の技法を補完するために使用され得るが、サイズバンドまたはサイズパターンを使用する実施形態は、独立して、２１トリソミーまたはコピー数異常を決定するために使用され得る。

【００９１】

この研究では、我々は、例えば、２％に及ぶ、低胎児ＤＮＡ画分を有する妊婦についてＮＩＰＴを行うことを可能にする新規方法を開発した。ニューラルネットワークモデルまたは他の機械学習モデルを訓練するためにより多くの試料が使用され、我々は、検出限界をさらに低下させると予想するであろう。我々は、母体血漿ＤＮＡにおけるコピー数変化の程度が、トリソミックおよび正倍数体胎児での妊娠間の異なるサイズバンドに関連して別個のパターンを示すであろうという事実を活用した。これは、２％未満の胎児ＤＮＡ画分に及ぶ胎児染色体異数性の非侵襲的検出の限度を下げることによって、幅広い人口カバレッジを達成するための重要なステップである。従来のアプローチを使用して、４％未満の胎児ＤＮＡ画分を含む妊娠は、ＮＩＰＴには適さず、一般的には、報告することができ

ない結果または検査失敗が出されるであろう。

【００９２】

我々の新たなアプローチは、検出下限のために偽陰性率を低減するだけでなく、異数性を有するリスクが４％未満の胎児ＤＮＡ画分を有する妊娠において増加するであろうことを示す多くの報告があったため、実際のＰＰＶも改善する可能性を有する（Ｎｏｒｔｏｎ　ｅｔ　ａｌ．Ｎ．Ｅｎｇｌ．Ｊ．Ｍｅｄ．２０１５；３７２：１５８９－１５９７）。以前に、一部の研究者は、低胎児ＤＮＡ画分を有する妊娠が、遺伝的カウンセリングを受けるべきであり、異数性の増加したリスクのために包括的な超音波評価および診断検査を提供されるべきであると主張する（Ｙａｒｏｎ　Ｐｒｅｎａｔ．Ｄｉａｇｎ．２０１６；３６：３９１－３９６）。胎児ＤＮＡ画分は一般に、母体重と逆相関しているため（Ｗａｎｇ　ｅｔ　ａｌ．Ｐｒｅｎａｔ．Ｄｉａｇｎ．２０１３；３３：６６２－６６６；Ｈｕｄｅｃｏｖａ　ｅｔ　ａｌ．ＰＬｏＳ　Ｏｎｅ　２０１４；９：ｅ８８４８４）、高ボディマス指数を有する妊娠は、低胎児ＤＮＡ画分でのシナリオに敏感に取り組むためのそのようなサイズバンドベースのアプローチの能力から特に恩恵を受けるであろう。我々の新たなアプローチの別の使用は、胎児ＤＮＡ画分が一般的により低い場合、ＮＩＰＴが妊娠の初期（例えば、妊娠１０週前）に行われることを可能にすることであろう。

【００９３】

Ｄ．腫瘍学におけるメチル化レベル分析

コピー数異常（ＣＮＡ）はまた、多くの癌で存在する。結果として、ＣＮＡは、対象における癌のレベルを決定するために使用され得る。加えて、癌患者はしばしば、特定のゲノム領域においてより高いレベルのメチル化を示す。したがって、メチル化マーカーはまた、癌のレベルを決定するためにサイズバンド分析と組み合わせて使用され得る。

【００９４】

１．メチル化でのサイズパターン分析

我々は、メチル化などの他のタイプの癌関連異常もまた、非癌対象から区別することができた特定のサイズバンドベースのパターンを構成するために使用され得るであろうと推論した。したがって、我々はまた、上記のようなＨＣＣ患者からの４つの血漿ＤＮＡ試料をさらに分析した。我々は、健康な対象の臓器ではメチル化されていないと考えられるが、癌患者でははるかに高いメチル化される可能性を有する領域についてのメチル化レベルを定量化するために、これに限定されないが、標的化バイサルファイト配列決定を使用した。我々は、健康な対象との比較においてメチロミック異常に関してサイズバンド関連パターンを調査するために、本明細書に記載されるサイズバンドベースのアプローチを適用した。メチル化は、２０１３年３月１５日に出願された米国特許出願第１３／８４２，２０９号（２０１７年８月１５日の米国特許第９，７３２，３９０号として発行）および２０１５年７月２０日に出願された米国特許出願第１４／８０３，６９２号にさらに記載され、両方の内容は、全ての目的で参照により本明細書に組み込まれる。

【００９５】

図７は、肝細胞癌（ＨＣＣ）患者の血漿ＤＮＡにおける測定されたメチル化のサイズバンドベースの変化するパターンを示す。ｚスコアは、ＨＣＣを有さないことが知られている健康な対象からの参照試料の平均メチル化レベル平均値を計算し、平均メチル化レベルに関連する標準偏差を計算することによって計算される。各サイズバンドのＺスコアは、そのサイズバンドでのメチル化レベルと平均メチル化レベル平均値との間の差、および標準偏差によって除算された差として計算される。図７における破線は、＋３または－３のｚスコアを示し、これは、平均メチル化レベル平均値からの統計的有意性を示すために使用され得る。

【００９６】

赤線またはより濃い線７０２、７０４、７０６、および７０８は、早期ＨＣＣ（ｅＨＣＣ）を表し、灰色線は、ＨＣＣを有さない慢性Ｂ型肝炎ウイルス（ＨＢＶ）キャリアを表した。図７において、我々は、ＨＣＣ患者（線７０２、７０４、７０６、および７０８）に関連したメチロミック異常の別個のサイズバンドパターンを確認することができ、これ

は、ＨＣＣ０１、ＨＣＣ０２、およびＨＣＣ０３においてＨＢＶキャリア（灰色線）から癌患者を特定することを可能にした。線７０２、７０４、および７０６は、ＨＢＶ試料の灰色線からかなり高く見える少なくとも２つのピークを有するパターンを示す。線７０８は、灰色線により近いが、依然としてＨＢＶ試料の灰色線よりも高い２つのピークを有する。「全て」とラベル付けされた各グラフにおける右端のデータは、サイズバンドに関係なく、全てのデータについてのプールされたＺスコアである。ＨＣＣ０４について、ランダムでないサイズバンドベースの湾曲パターンは、全ての断片の使用で全体的な異常メチル化度（円７１０によって表される）よりも情報が豊富であることがわかった。異なるグラフにおいて異なるゲノム領域が使用された。染色体腕１ｑは、ＨＣＣ０１およびＨＣＣ０４に使用され、１０ｐは、ＨＣＣ０２に使用され、１９ｑは、ＨＣＣ０３に使用された。他の実施形態では、癌を検出するために、例えば、これらに限定されないが、低メチル化、点突然変異、ヒドロキシメチル化、断片化末端などのサイズバンドベースの変化するパターンも使用することができた。

【００９７】

２．癌のレベルを決定するための例示的な方法

図８は、対象からの生体試料における癌のレベルを決定する方法８００を示す。生体試料は、無細胞ＤＮＡ分子の混合物を含み得る。無細胞ＤＮＡ分子は、腫瘍ＤＮＡ分子および非腫瘍ＤＮＡ分子を含み得る。

【００９８】

ブロック８０２で、方法８００は、複数のサイズ範囲の各サイズ範囲についてのサイズ範囲に対応する生体試料からのメチル化無細胞ＤＮＡ分子の第１の量を測定することを含み得る。メチル化無細胞ＤＮＡ分子は、染色体腕に由来し得る。サイズ範囲に対応するメチル化無細胞ＤＮＡ分子の量を測定することは、無細胞ＤＮＡ分子がメチル化されていることを除いて、方法６００または本明細書に記載される任意の他の方法に記載されるように行われ得る。メチル化無細胞ＤＮＡ分子の第１の量は、１つ以上のゲノム領域に由来し得る。ゲノム領域は、染色体腕、例えば、１ｐ、１ｑ、８ｐ、８ｑ、１３ｑ、または１４ｐであり得る。ゲノム領域の様々な組み合わせが使用され得る。使用する特定の領域は、既知の癌のレベルを有する試料の訓練セットに対して癌のレベルを決定するための領域の様々な組み合わせの精度を分析することによって決定することができる。

【００９９】

ブロック８０４で、方法８００は、各サイズ範囲について、コンピュータシステムによって、サイズ範囲に対応するメチル化無細胞ＤＮＡ分子の第１の量およびサイズ範囲にないサイズを含む第２のサイズ範囲のＤＮＡ分子の第２の量を用いてメチル化レベルを計算することを含み得る。第２の量は、メチル化無細胞ＤＮＡ分子のものであり得る。これらまたは他の実施形態では、第２の量は、非メチル化無細胞ＤＮＡ分子を含み得る。

【０１００】

メチル化レベルは、１つ以上の部位でメチル化または非メチル化されるＤＮＡ分子のＤＮＡ分子のｚスコアまたは正規化された量（例えば、画分、パーセンテージ、または相対存在量）であり得る。例えば、メチル化レベルは、第２の量に対する第１の量の比であり得る。他の実施形態では、メチル化レベルは、ｚスコアであり得る。Ｚスコアは、第２の量に対するサイズ範囲に対応する無細胞ＤＮＡ分子の量の比を用いて計算され得る。次いで、計算された比と平均比平均値との間の差を、標準偏差で割って、Ｚスコアを決定する。平均比平均値は、対照群（例えば、非癌患者、参照試料、または癌に関連しないゲノム領域）の平均メチル化レベルであってもよい。メチル化レベルがｚスコアである場合、サイズ範囲のメチル化レベルは、図７における線７０２、７０４、７０６、および７０８上の任意の点であり得る。

【０１０１】

ブロック８０６で、方法８００は、複数のサイズ範囲の複数の参照メチル化レベルを含む参照サイズパターンを得ることを含み得る。複数のサイズ範囲は、機械学習アルゴリズムによって決定され得、方法６００について記載されるものと同じ方法で決定され得る。

参照サイズパターンは、癌を有する対象または癌を有さない対象からの複数の参照試料から決定され得る。例えば、参照試料は、ＨＣＣまたは任意のタイプの癌を有さないことが知られている患者に由来し得る。参照サイズパターンは、ＨＣＣを有さない慢性ＨＢＶキャリアからのデータに基づき得る。例えば、参照サイズパターンは、図７におけるＨＢＶについての灰色線のいずれかであり得る。いくつかの実施形態では、参照サイズパターンは、方法６００で説明されるように、参照試料についての全てのサイズパターンの統計的表現であり得る。

【０１０２】

ブロック８０８で、方法８００は、複数のメチル化レベルと参照サイズパターンとを比較することを含み得る。方法８００は、複数のサイズ比の各メチル化レベルと対応するサイズ範囲での参照メチル化レベルとを比較することを含み得る。メチル化レベルと参照サイズパターンとを比較することは、サイズの代わりにメチル化レベルであることを除いて、サイズ比が方法６００における参照サイズパターンと比較される方法で行われ得る。方法８００は、各メチル化レベルが対応するサイズ範囲での参照メチル化レベルと統計的に類似することを決定することを含み得る。いくつかの実施形態では、方法８００は、各メチル化レベルまたはいくつかのメチル化レベルが、対応するサイズ範囲での参照メチル化レベルと統計的に異なることを決定することを含み得る。

【０１０３】

いくつかの実施形態では、複数のメチル化レベルと参照サイズパターンとを比較することは、複数のサイズ範囲の複数のメチル化レベルを含むサイズパターンを決定することを含み得る。サイズパターンは、参照サイズパターンと比較され得る。サイズパターンは、参照サイズパターンと類似の形状を有すると決定され得る。方法８００における参照サイズパターンとの比較は、方法６００における参照サイズパターンとの比較に類似し得る。

【０１０４】

メチル化無細胞ＤＮＡ分子の第１の量が、複数のゲノム領域からのものである場合、メチル化レベルは、ゲノムにおけるそれらの場所に基づいて分析され得る。複数のメチル化レベルは、多次元ベクターを含み得る。多次元ベクターは、Ｎ×Ｍであり得、Ｎは、サイズ範囲の数であり、Ｍは、ゲノム領域の数である。ゲノム領域は、染色体、染色体腕、または染色体腕の一部であり得る。参照サイズパターンは、同様に多次元ベクター（例えば、サイズＮ×Ｍ）であり得る。複数のメチル化レベルは、機械学習モデルまたは他の技法を用いて参照サイズパターンと比較され得る。多次元ベクターおよびメチル化レベルの使用は、以下に記載される（例えば、図１３、１４Ａ、１４Ｂ、および１４Ｃ）。

【０１０５】

ブロック８１０で、方法８００は、比較に基づいて癌のレベルを決定することを含み得る。癌のレベルは、対象が癌を有するか否か、癌の可能性、または腫瘍サイズを含み得る。

【０１０６】

参照サイズパターンが、癌を有する対象からの複数の参照試料から決定され、比較が、類似のメチル化レベルまたは類似の形状の決定を含む場合、対象は、癌を有すると決定され得る。この参照サイズパターンでは、比較が、異なるメチル化レベルまたは異なる形状の決定を含む場合、対象は、癌を有さないと決定され得る。参照サイズパターンが、癌を有さない複数の参照試料から決定され、比較が、異なるメチル化レベルまたは形状の決定を含む場合、対象は、癌を有すると決定され得る。参照サイズパターンが、癌を有さない複数の参照試料から決定され、比較が、類似のメチル化レベルまたは形状の決定を含む場合、対象は、癌を有さないと決定され得る。

【０１０７】

Ｅ．サイズバンドマトリックスでのパターン分析

様々な癌で、染色体腕を含む、特定のゲノム領域は、コピー数異常を有する可能性がより高くなり得る。考えられるコピー数異常について染色体腕によってサイズ範囲を分析することは、癌の確率の決定または癌の検出に役立てるために使用され得る。機械学習モデ

ルは、異なる染色体領域（例えば、腕）でのサイズ特徴のパターンに基づいて癌分類子を決定するために使用することができる。

【０１０８】

１．サイズパターン分析

癌患者の血漿中の腫瘍由来ＤＮＡのサイズプロファイルは、非腫瘍由来ＤＮＡ分子とは異なることが示され、前者は一般に、より多くの短いＤＮＡ分子からなるため（Ｊｉａｎｇ　ｅｔ　ａｌ．Ｐｒｏｃ．Ｎａｔｌ．Ａｃａｄ．Ｓｃｉ．２０１５；１１２：Ｅ１３１７－Ｅ１３２５）、我々は、本発明に記載されるサイズバンドベースのアプローチが、コピー数異常（ＣＮＡ）およびメチロミック異常などの癌関連異常を検出するために有用であろうと推論した。例として、我々は、サイズバンドベースのパターン認識を、早期肝細胞癌（ＨＣＣ）患者の４つの血漿ＤＮＡ試料およびＨＣＣ癌を有さない６７個の慢性Ｂ型肝炎（ＨＢＶ）キャリア（ＨＢＶキャリア）に適用した。健康な対照の３０個の血漿ＤＮＡ試料は、ＨＣＣ患者およびＨＢＶキャリアのＣＮＡおよびメチロミック異常を呼び出すために使用されたコピー数変化の正常参照範囲を構築するために使用された。

【０１０９】

図９は、肝細胞癌（ＨＣＣ）患者の血漿ＤＮＡにおける測定されたコピー数異常のサイズバンドベースの変化するパターンを示す。赤線は、早期ＨＣＣ（ｅＨＣＣ）を表し、灰色線はＨＣＣを有さない慢性Ｂ型肝炎ウイルス（ＨＢＶ）キャリアを表した。我々は、ＨＣＣ癌を有する患者における測定されたＣＮＡのサイズバンドパターンの曲線（赤またはより濃い線９０２、９０４、９０６、および９０８）が、ＨＢＶキャリアを有する患者についての曲線（灰色線）とは異なったことを観察する。例えば、ＨＣＣ０１およびＨＣＣ０３症例は、それぞれ、１３ｑおよび１ｐ染色体腕上でコピー増加を有した。

【０１１０】

ＨＣＣ０１およびＨＣＣ０３では、我々は、２１０ｂｐに中点を有するサイズバンドが、コピー数変化を示すその左側および右側に対する転換点の傾向があった、非ランダムな波状サイズバンドベースのパターンを一貫して検出することができ、１２０ｂｐ付近のサイズバンドパターンは、「ベルカーブ」の傾向を示した。１４ｑ欠失を受けたＨＣＣ０２症例について、逆「ベルカーブ」が存在した。ＨＣＣ０４症例について、我々が全ての断片にＺスコアを使用した場合、我々は、３未満であり、非癌患者のＺスコアの範囲内に十分に含まれるＺスコアを有する円９１０によって示されるように、癌を検出することができなかった。しかしながら、我々がサイズバンドベースのアプローチを利用した場合、我々は、ランダムなサイズバンドベースのパターン（灰色線）を示す非癌患者からＨＣＣ０４を区別することができた。対照的に、そのような非ランダムな異なるサイズバンドベースパターンは、対照群には存在しなかった。異なる染色体腕は、異なるサイズのパターンを示す。サイズパターンは、染色体腕に特異的なサイズパターンを参照する必要があり得る。

【０１１１】

２．サイズバンドＧＲマトリックスでの癌分類子

癌細胞は一般に、任意の染色体腕で発生するであろうコピー数異常を有し、これは、腫瘍細胞がＤＮＡを癌患者の血液循環に流す場合に血漿において反映されるであろう。腫瘍由来無細胞ＤＮＡ分子は、バックグラウンド正常無細胞ＤＮＡとの比較において異なるサイズ特性を有することが示されている（例えば、腫瘍無細胞ＤＮＡ分子は、正常細胞に由来するバックグラウンド無細胞ＤＮＡよりも短い）ので、異なるサイズ範囲にわたる相対腫瘍ＤＮＡ画分は、変動するであろう。よって、癌患者の血漿中に存在する異なるサイズ範囲にわたるコピー数異常の測定された程度は、異なるサイズ範囲にわたる相対腫瘍ＤＮＡ画分の関数であろう。

【０１１２】

我々は、異なるサイズ範囲にわたる測定されたコピー数異常の詳細なパターンを捕捉することが、癌患者と非癌患者との区別における性能を改善するであろうと提案した。パターンは、複数の領域も含むことができる。

【０１１３】
図１０は、本発明の実施形態による癌検出のためのサイズバンドゲノム表現（ＧＲ）アプローチのワークフローを例示する。ステージ１０１０で、我々は、配列決定された無細胞ＤＮＡ断片を参照ゲノムにマップした。ステージ１０２０で、配列決定された断片は、異なる染色体腕にマップされる。
【０１１４】
ステージ１０３０で、配列決定された断片は、異なるサイズ範囲（サイズバンド）にさらに分類される。例えば、サイズ範囲は、これらに限定されないが、３５－７５ｂｐ、４０－８０ｂｐ、４５－８５ｂｐ、５０－９０ｂｐ、５５－９５ｂｐ、６０－１００ｂｐ、６５－１０５ｂｐ、７０－１１０ｂｐ、７５－１１５ｂｐ、８０－１２０ｂｐ、８５－１２５ｂｐ、９０－１３０ｂｐ、９５－１３５ｂｐ、１００－１４０ｂｐ、１０５－１４５ｂｐ、１１０－１５０ｂｐ、１１５－１５５ｂｐ、１２０－１６０ｂｐ、１２５－１６５ｂｐ、１３０－１７０ｂｐ、１３５－１７５ｂｐ、１４０－１８０ｂｐ、１４５－１８５ｂｐ、１５０－１９０ｂｐ、１５５－１９５ｂｐ、１６０－２００ｂｐ、１６５－２０５ｂｐ、１７０－２１０ｂｐ、１７５－２１５ｂｐ、１８０－２２０ｂｐ、１８５－２２５ｂｐ、１９０－２３０ｂｐ、１９５－２３５ｂｐ、２００－２４０ｂｐ、２０５－２４５ｂｐ、２１０－２５０ｂｐ、２１５－２５５ｂｐ、２２０－２６０ｂｐ、２２５－２６５ｂｐ、２３０－２７０ｂｐ、２３５－２７５ｂｐ、２４０－２８０ｂｐ、２４５－２８５ｂｐ、２５０－２９０ｂｐ、２５５－２９５ｂｐ、２６０－３００ｂｐ、２６５－３０５ｂｐ、２７０－３１０ｂｐ、２７５－３１５ｂｐ、２８０－３２０ｂｐ、２８５－３２５ｂｐ、２９０－３３０ｂｐ、２９５－３３５ｂｐ、３００－３４０ｂｐ、３０５－３４５ｂｐ、３１０－３５０ｂｐ、３１５－３５５ｂｐ、３２０－３６０ｂｐ、３２５－３６５ｂｐ、３３０－３７０ｂｐ、３３５－３７５ｂｐ、３４０－３８０ｂｐ、３４５－３８５ｂｐ、３５０－３９０ｂｐ、３５５－３９５ｂｐ、３６０－４００ｂｐ、３６５－４０５ｂｐ、３７０－４１０ｂｐ、３７５－４１５ｂｐ、３８０－４２０ｂｐ、および３８５－４２５ｂｐを含み得る。そのようなサイズ範囲は、全ての他の実施形態にも使用され得る。
【０１１５】
特定のサイズ範囲内の分子の群について、各染色体腕にマップされた配列決定された断片の割合が計算され、本明細書ではゲノム表現（ＧＲ）と呼ばれる。ＧＲは、サイズ範囲内の特定の領域（またはゲノム全体）に対応する全てのＤＮＡ断片の割合である。ステージ１０３０は、様々なサイズ範囲、異なる染色体腕、癌を有することが知られている試料、および癌を有さないことが知られている試料のＧＲを示す。
【０１１６】
例として、各染色体腕が７１のサイズ範囲を含み、常染色体が合計３９の染色体腕を有する場合、サイズ範囲および染色体腕は、２，７６９次元ベクターをもたらす。ステージ１０４０は、可能な多次元ベクターを示す表（「サイズバンドＧＲマトリックス」）を示す。第１の行１０４２は、癌試料１に対応し、７１×Ｎ次元ベクターを示し、ここで、Ｎは、染色体腕の数である。表は、癌のＭ試料および非癌のＰ試料を示す。
【０１１７】
ステージ１０５０で、多次元ベクターおよび多次元ベクターから形成されるサイズバンドＧＲマトリックスは、癌分類モデルを訓練するために使用することができる。機械学習アルゴリズムまたは深層学習アルゴリズムは、これらに限定されないが、サポートベクターマシン（ＳＶＭ）、決定木、単純ベイズ分類、ロジスティック回帰、クラスタリングアルゴリズム、主成分分析（ＰＣＡ）、特異値分解（ＳＶＤ）、ｔ分布型確率的近傍埋め込み（ｔＳＮＥ）、人工ニューラルネットワーク、および分類子のセットを構成し、それらの予測の加重投票によって新たなデータ点を分類するアンサンブル方法を含む、癌分類子を訓練するために使用することができた。癌分類子が訓練されると、新たな患者について癌の確率を予測することができる。
【０１１８】

訓練データは、癌および非癌対象を含むことができる。無細胞ＤＮＡ測定（サイズバンドＧＲ、メチル化など）をモデル化する機械学習アルゴリズムは、癌と非癌対象との間の最良の分離を与える分類境界を構成するために（例えば、ロジスティック回帰式などの、線形または非線形式で編成された訓練された重みおよび係数のセットを用いて）使用することができる。癌関連データ点への最適分類境界からの無細胞ＤＮＡ測定を含む新たな試料の入力ベクターの偏差は、癌である可能性を示すであろう。そのような偏差は、正規化または０～１のスケール内の癌の確率に変換することができる。確率が高いほど、癌になる可能性が高くなる。特定の閾値を超える（例えば、＞０．６）癌の確率は、癌での陽性検査とみなすことができる。

【０１１９】

肝細胞癌について、１ｐ、１ｑ、８ｐ、および８ｑは、一般的にはコピー数に関して異常であったことが報告された（Ｐｒｏｃ　Ｎａｔｌ　Ａｃａｄ　Ｓｃｉ　ＵＳＡ．２０１５　Ｍａｒ　１７；１１２（１１）：Ｅ１３１７－２５）。よって、サイズバンド癌検出の性能を例示するために、我々は、多数の健康な対照（ＣＴＲ）、ＨＢＶキャリア（ＨＢＶ）、肝硬変対象（肝硬変）、初期ＨＣＣ（ｅＨＣＣ）、中期ＨＣＣ（ｉＨＣＣ）、および進行期ＨＣＣ（ａＨＣＣ）を配列決定するために、超並列配列決定プラットフォームを使用した。訓練データセットについて、我々は、限られた数の進行期ＨＣＣ患者を配列決定し、次いで進行期ＨＣＣ患者の配列決定結果と非ＨＣＣ対象のものとを人為的に混合して、０．０１％～５０％の範囲である腫瘍ＤＮＡ画分の広いカバレッジを有する十分なＨＣＣ陽性患者および非ＨＣＣ対象を含有する訓練データセットを形成した。この目的のために、４０１人のＨＣＣ患者は、使用される配列決定リードの割合を変動することによって３４人のＨＢＶ、１０人のＣＴＲ、および９人のａＨＣＣ対象をランダムに繰り返し混合することによって作成され、１７５人の非ＨＣＣ患者は、３４人のＨＢＶ、１５人の肝硬変、および１０人のＣＴＲ対象をランダムに繰り返し混合することによって作成された。ＳＶＭアルゴリズムは、そのような４０１人のＨＣＣ患者および１７５人のｉｎ－ＨＣＣ患者を用いて癌分類子を訓練するために使用された。

【０１２０】

ステージ１０６０で、訓練された癌分類モデルは、新たな試料が癌を有するか、または癌を有さないかを予測するために使用することができる。癌の確率は、モデルによって決定され得、閾値を超える確率は、癌の陽性検査とみなされる。

【０１２１】

癌を検出するためのサイズバンドアプローチおよび従来のｚスコアアプローチは、３０のＣＴＲ、１９のＨＢＶ、１４の肝硬変、３６のｅＨＣＣ、および１１のｉＨＣＣ対象を含む検査データセットに適用された。

【０１２２】

図１１Ａは、癌を検出するためのサイズバンドアプローチの結果を示す。ＳＶＭは、癌分類子を訓練するために使用された。ｅＨＣＣおよびｉＨＣＣ対象の両方は、０．６０を超える中央値の癌の確率を有し、ｉＨＣＣは、ｅＨＣＣよりも高い確率を有した。ＣＴＲ、ＨＢＶ、および肝硬変対象は、０．２０未満の確率中央値を示した。癌を検出するためのサイズバンドアプローチは、９５％の特異度で６４％感度を有した。赤点線は、９５％特異度に対応する。

【０１２３】

図１１Ｂは、癌を検出するための従来のｚスコアアプローチの結果を示す。赤点線は、９５％特異度に対応し、これは、約４．２のｚスコアであった。染色体腕１ｐ、１ｑ、８ｐ、および８ｑが例として使用された。検査試料の各腕のＧＲが計算された。対応する平均値および標準偏差も計算された。各腕ｚスコアは、（ＧＲ－平均値）／標準偏差として計算されるであろう。絶対ｚスコアは、４つの染色体腕に対応する４つの絶対ｚスコアの合計と等しかった。ｉＨＣＣ対象は、ＣＴＲ、ＨＢＶ、肝硬変、およびｅＨＣＣ対象よりも著しく高い癌の絶対ｚスコア中央値を有した。ｉＨＣＣの絶対ｚスコア中央値は、他の対象の絶対ｚスコアよりも高かったが、いくらかのｉＨＣＣ対象のｚスコアは、他の対象

とかなり類似していた。しかしながら、ｅＨＣＣの絶対ｚスコアの中央値は、ＣＴＲ、ＨＢＶ、および肝硬変対象のものよりもわずかに高いのみであり、３のｚスコア閾値レベルとほぼ同じであった。従来のｚスコアアプローチは、９５％の特異度で５１％感度を有した。よって、サイズバンドアプローチは、従来のｚスコアアプローチに対して優れた感度を示す。

【０１２４】

図１１Ｃは、受信者動作特性曲線（ＲＯＣ）分析での従来のｚスコアアプローチに対するサイズバンドアプローチの優位性を示す（０．８４対０．８２）。

【０１２５】

３．サイズバンドゲノム表現（ＧＲ）マトリックスでの例示的な方法

図１２は、対象からの生体試料における癌分類を決定する例示的な方法１２００を示す。生体試料は、腫瘍ＤＮＡ分子および非腫瘍ＤＮＡ分子を含む無細胞ＤＮＡ分子の混合物を含み得る。

【０１２６】

ブロック１２０２で、生体試料からの無細胞ＤＮＡ分子の第１の量が測定され得る。無細胞ＤＮＡ分子の第１の量は、Ｍ範囲の各サイズ範囲およびＮゲノム領域の各ゲノム領域に対応し得る。複数のサイズ範囲は、方法６００または方法８００で記載されるように決定され得る。各ゲノム領域は、染色体腕であり得る。

【０１２７】

ブロック１２０４で、サイズ比は、サイズ範囲にないサイズを含む第２のサイズ範囲の無細胞ＤＮＡ分子の第１の量および無細胞ＤＮＡ分子の第２の量を用いて計算され得る。サイズ比は、方法６００のように計算され得るが、サイズ比は、特定のゲノム領域（例えば、染色体腕）についてのものであり得る。例として、サイズ比は、図１０における行１００４のゲノム表現ＧＲ１、ＧＲ２、ＧＲ３、．．．ＧＲ７１のいずれかであり得る。サイズ比を計算することは、Ｎ×Ｍサイズ比の測定ベクターを生成し得る。Ｎは、１以上の整数であり得る。ＮおよびＭは、２、３、４、５、または６超を含む、１を超える整数であり得る。

【０１２８】

ブロック１２０６で、参照サイズパターンが取得され得る。参照サイズパターンは、Ｎゲノム領域およびＭサイズ範囲についての参照サイズ比の参照ベクターを含み得る。参照サイズパターンは、癌を有する対象または癌を有さない対象からの複数の参照試料から決定され得る。参照サイズパターンは、機械学習モデルを用いて決定され得る。

【０１２９】

機械学習モデルは、癌を有する個体からの複数のゲノム領域の各々でのサイズ比を含むサイズ比の訓練セットを用いて決定され得る。癌分類子は、機械学習アルゴリズムまたは深層学習アルゴリズムを用いて決定され得る。機械学習モデルまたは深層学習アルゴリズムは、サポートベクターマシン（ＳＶＭ）、決定木、単純ベイズ分類、ロジスティック回帰、クラスタリングアルゴリズム、主成分分析（ＰＣＡ）、特異値分解（ＳＶＤ）、ｔ分布型確率的近傍埋め込み（ｔＳＮＥ）、人工ニューラルネットワーク、または本明細書に記載される任意のアルゴリズムを含み得る。訓練セットは、癌を有すると決定された個体および癌を有さないと決定された個体の異なるゲノム領域でのサイズ比を含み得る。機械学習モデルは、図１０における癌分類子であり得る。

【０１３０】

ブロック１２０８で、測定ベクターは、参照ベクターと比較され得る。比較は、機械学習モデルを用いて比較され得る。比較は、測定ベクターと参照ベクターとの類似性に基づく値をもたらし得る。

【０１３１】

測定ベクターと参照ベクターとを比較することは、Ｎ×Ｍサイズ比と複数の参照試料から決定される複数の閾値とを比較することを含み得る。例えば、各サイズ範囲は、参照試料の標準偏差に基づき得る、異なる閾値を有し得る。したがって、Ｎ×Ｍ閾値があり得る

。単一のサイズ範囲はまた、異なる閾値を有し得、各閾値は、サイズ比が参照試料とは異なるという異なる確実性レベルに関連している。比較することは、超えられた閾値の数をカウントすることと、比較に基づいて癌のレベルを決定することと、を含み得る。超えられたより高いレベルの閾値は、測定ベクターと参照ベクターとの間のより大きな差を示し得る。

【０１３２】

ブロック１２１０で、癌のレベルは、比較に基づいて決定され得る。癌は、肝細胞癌を含み得る。癌は、結腸直腸癌、肺癌、鼻咽頭癌、卵巣癌、胃癌、および血液癌を含み得る。方法１２００は、癌と非癌対象との間の区別を可能にし得る。対象は、測定ベクターと参照ベクターとの類似性に基づく値に基づいて、癌を有するか、または高い癌の可能性を有すると分類され得る。類似性に基づく値は、カットオフ値と比較され得る。カットオフ値をより大きく超える類似性に基づく値は、より高い癌の可能性または重症度を示し得る。方法は、対象が癌を有するか、または高い癌の可能性を有すると分類される場合、癌を治療することをさらに含み得る。

【０１３３】

方法１２００は、癌の代わりに自己免疫障害のレベルを決定するように適合され得る。自己免疫障害は、全身性エリテマトーデス（ＳＬＥ）を含み得る。ＤＮＡ断片のサイズは、その内容が全ての目的のために参照により本明細書に組み入れられる、２０１４年９月１９日に出願された、米国特許出願公開第２０１５／００８７５２９Ａ１号に記載されるように、ＳＬＥに関連することが見出されている。自己免疫障害のレベルは、測定ベクターと参照ベクターとを比較することによって決定され得る。参照ベクターは、参照サイズパターンからのものであり得る。参照サイズパターンは、健康な対象または既知の自己免疫障害のレベルを有する対象からの試料から決定され得る。方法１２００は、自己免疫障害ありおよびなしの対象間の区別を可能にし得る。

【０１３４】

４．サイズバンドメチル化密度（ＭＤ）マトリックスでの癌分類子

癌細胞は一般に、任意のゲノム領域において発生するであろう特定のメチル化パターンを有する。例えば、癌細胞では、Ａｌｕリピート領域は、非悪性細胞と比較して優先的により少なくメチル化され得、ＣｐＧアイランド領域は、非悪性細胞と比較して優先的によりメチル化され得る。そのような癌関連異常メチル化信号は、腫瘍細胞がＤＮＡを血液循環に流す場合に癌患者の血漿において反映され得る。上記で説明されるように、様々なサイズ範囲にわたる相対腫瘍ＤＮＡ画分は変動する。よって、癌患者の血漿中に存在する異なるサイズ範囲にわたる癌関連メチル化レベルの測定された程度は、異なるサイズ範囲にわたる相対腫瘍ＤＮＡ画分の関数であろう。

【０１３５】

我々は、異なるサイズ範囲にわたる測定されたメチル化異常の詳細なパターンを捕捉することが、癌患者と非癌患者との区別における性能を改善するであろうと提案した。

【０１３６】

図１３は、本発明の実施形態による癌検出のためのサイズバンドメチル化密度（ＭＤ）アプローチのワークフローを例示する。ステージ１３１０で、我々は、配列決定されたバイサルファイト変換無細胞ＤＮＡ断片を参照ゲノムに、Ｍｅｔｈｙ－Ｐｉｐｅ（Ｊｉａｎｇ　ｅｔ　ａｌ．，ＰＬｏＳ　Ｏｎｅ．２０１４；９（６）：ｅ１００３６０）または他のメチル化対応アライナーを用いてマップした。ステージ１３２０で、異なる差動メチル化領域にマップされた配列決定された断片が配置される。

【０１３７】

ステージ１３３０で、配列決定された断片は、異なるサイズ範囲（サイズバンド）にさらに分類される。例えば、サイズ範囲は、図１０についてステージ１０３０に記載されるサイズ範囲を含む、本明細書に記載される任意のサイズ範囲を含み得る。

【０１３８】

特定のサイズ範囲内の分子の群について、目的の領域上の配列決定されたＣｐＧ（例え

ば、ＡｌｕリピートまたはＣｐＧアイランド）の割合が計算され、メチル化レベルを反映するメチル化密度（ＭＤ）をもたらすであろう。領域は、造血細胞（例えば、Ｔ細胞、Ｂ細胞、好中球、マクロファージ、赤芽細胞など）、肝臓細胞、および結腸細胞を含む、肝臓癌細胞と他の正常細胞との間で異なるメチル化レベルを示し得る。ステージ１３３０は、異なるサイズ範囲、異なるゲノム領域、癌を有することが知られている試料、および癌を有さないことが知られている試料のＭＤを示す。

【０１３９】

例として、各領域が、７１のサイズ範囲を含み、合計３２，４５０の領域があり、肝臓癌細胞と他の正常細胞との間の作動的メチル化を示す場合、サイズ範囲およびゲノム領域は、２，３０３，９５０次元ベクターをもたらす。ステージ１３４０は、可能な多次元ベクターを示す表（「サイズバンドＭＤマトリックス」）を示す。表の第１の行１３４２は、癌試料１に対応し、７１×Ｎ次元ベクターを示し、ここで、Ｎは、ゲノム領域の数である。表は、癌のＭ試料および非癌のＰ試料を示す。

【０１４０】

ステージ１３５０で、多次元ベクターおよび多次元ベクターから形成されるサイズバンドＭＤマトリックスは、癌分類モデルを訓練するために使用することができる。訓練は、図１０のステージ１０５０を含む、例えば、本明細書に記載されるような、分類を行う任意の好適な機械学習モデルによるものであり得る。癌分類子が訓練されると、新たな患者の癌を示す試料の確率を予測することができる。特定の閾値を超える（例えば、＞０．６）癌の確率は、癌での陽性検査とみなすことができる。

【０１４１】

サイズバンドメチル化レベルの使用での癌検出の性能を例示するために、我々は、多数の健康な対照（ＣＴＲ）、ＨＢＶキャリア（ＨＢＶ）、肝硬変対象（肝硬変）、初期ＨＣＣ（ｅＨＣＣ）、中期ＨＣＣ（ｉＨＣＣ）、および進行期ＨＣＣ（ａＨＣＣ）を配列決定するために、超並列配列決定プラットフォームを使用した。訓練データセットについて、我々は、限られた数の進行期ＨＣＣ患者を配列決定し、次いで進行期ＨＣＣ患者の配列決定結果と非ＨＣＣ対象のものとを人為的に混合して、０．０１％～５０％の範囲である腫瘍ＤＮＡ画分の広いカバレッジを有する十分なＨＣＣ陽性患者および非ＨＣＣ対象を含有する訓練データセットを形成した。この目的のために、１４０人のＨＣＣ患者は、使用される配列決定リードの割合を変動することによって２７人のＨＢＶおよび７人のａＨＣＣ対象をランダムに繰り返し混合することによって作成され、１４０人の非ＨＣＣ患者は、７人のＨＢＶおよび２０人のＣＴＲ対象をランダムに繰り返し混合することによって作成された。ＳＶＭアルゴリズムは、そのような１４０人のＨＣＣ患者および１４０人の非ＨＣＣ患者を用いて癌分類子を訓練するために使用された。

【０１４２】

ステージ１３６０で、訓練された癌分類モデルは、新たな試料が癌を有するか、または癌を有さないかを予測するために使用することができる。癌の確率は、モデルによって決定され得、閾値を超える確率は、癌の陽性検査とみなされる。

【０１４３】

図１４Ａ、１４Ｂ、および１４Ｃは、本発明の実施形態によるサイズバンドＭＤと従来のｚスコアアプローチとの比較を示す。図１４Ａは、サイズバンドＭＤアプローチの結果を示す。図１４Ｂは、従来のｚスコアアプローチの結果を示す。

【０１４４】

図１４Ａおよび１４Ｂは、２７人のＨＢＶ、３６人のｅＨＣＣ、および１１人のｉＨＣＣ対象を含む検査データセットにおいて、癌を検出するためのサイズバンドメチル化アプローチが従来のｚスコアアプローチよりも優れていたことを示す。従来のｚスコアアプローチは、次の方法で行われた。（１）目的の全ての領域に由来する総断片についての（「Ｘ」によって示される）プールされたメチル化レベルが計算され、（２）プールされたメチル化レベルの平均値（Ｍ）、および非癌群におけるプールされたメチル化レベルの標準偏差（ＳＤ）が計算され、（３）次いで従来のメチル化ｚスコアが、ｚスコア＝（Ｘ－Ｍ

）／ＳＤによって定義される。ＳＶＭは、癌分類子を訓練するために使用された。図１４Ａにおけるサイズバンドメチル化アプローチは、９２．５％の特異度で７４．５％の感度を有した。対照的に、図１４Ｂにおける従来のｚスコアアプローチは、９２．５％の特異度で、より低い感度、６５．９％の感度を有した。増加した感度は、重要な利益をもたらし得る。早期癌の早期検出は一般に、より良好な治療転帰と関連している。ｅＨＣＣ群およびｉＨＣＣ群の両方は、治療可能なステージであるとみなされる。したがって、治療可能な症例における感度の任意の増加は、臨床的影響を有し、患者の非常に異なる生存プロファイルにつながり得る。

【０１４５】

図１４Ｃは、受信者動作特性曲線（ＲＯＣ）分析におけるサイズバンドメチル化アプローチの優位性を示す（ＳＶＭ：０．８９ＡＵＣ対ｚスコア：０．８７ＡＵＣ）。

【０１４６】

したがって、ゲノム表現（ＧＲ）を有する多次元ベクターの使用（例えば、図１０～１２）は、ＧＲの代わりにメチル化密度を用いる分析に適合され得る。

【０１４７】

Ｆ．追加サイズパターン適用

サイズバンドベースのパターンは、血漿ＤＮＡに見られるそれらの異常の起源を知らせるであろう。例として、妊娠の文脈において、コピー数異常が母親に由来する場合、母体ＤＮＡ断片が胎児ＤＮＡよりも長いため、サイズバンドパターンは、胎児由来のものと比較して逆方向に発生するであろう（Ｙｕ ｅｔ ａｌ．Ｃｌｉｎ．Ｃｈｅｍ．２０１７；６３：４９５－５０２）。サイズバンドベースの分子診断はまた、癌などの他の臨床病態における無細胞ＤＮＡの分析に適用することができ（Ｊｉａｎｇ ｅｔ ａｌ．Ｐｒｏｃ．Ｎａｔｌ．Ａｃａｄ．Ｓｃｉ．２０１５；１１２：Ｅ１３１７－Ｅ１３２５）、点突然変異、サブ染色体異常、およびエピジェネティック異常の検出を向上させることを含む。臨床病態は、移植された組織または臓器に対する免疫応答の存在を決定することを含み得る。

【０１４８】

さらに、全身性エリテマトーデス（ＳＬＥ）患者の血漿ＤＮＡ中に存在する見かけのコピー数変化（Ｃｈａｎ ｅｔ ａｌ．Ｐｒｏｃ．Ｎａｔｌ．Ａｃａｄ．Ｓｃｉ．２０１４；１１１：Ｅ５３０２－Ｅ５３１１）は、特定の細胞における真のコピー数の変化ではなく、特定のＤＮＡ配列への抗ＤＮＡ抗体の優先的な結合によるものである可能性が高いので、ＳＬＥなどの血漿ＤＮＡ中に存在する血漿ＤＮＡ交絡異常を区別することも可能になるであろう。よって、サイズバンドベース分析は、ＳＬＥ患者の血漿中に存在する測定されたコピー数異常について異なるサイズバンドに関してランダムな形状変化を認めることが予想されるであろう。

【０１４９】

実施形態は、患者における疾患または病態のレベルまたは確率を決定した後に、患者における疾患または病態を治療することを含み得る。治療には、本明細書で言及される参考文献に記載される任意の治療を含む、任意の好適な療法、薬物、または外科手術が含まれ得る。参考文献における治療に関する情報は、参照により本明細書に組み込まれる。

【０１５０】

ＩＩＩ．材料および方法

試料収集および処理

この後ろ向き研究のために分析された匿名化されたデータは、Ｃｈｉｎｅｓｅ Ｕｎｉｖｅｒｓｉｔｙ ｏｆ Ｈｏｎｇ ＫｏｎｇのＵｎｉｖｅｒｓｉｔｙ Ｐａｔｈｏｌｏｇｙ Ｓｅｒｖｉｃｅ（ＵＰＳ）における既存の患者データから得られた。１６１個の試料からなる患者データは、ＵＰＳの実験室で開発された検査の結果として生成された。腫瘍切除のためにＤｅｐａｒｔｍｅｎｔ ｏｆ Ｓｕｒｇｅｒｙ ｏｆ ｔｈｅ Ｐｒｉｎｃｅ Ｗａｌｅｓ Ｈｏｓｐｉｔａｌ，Ｈｏｎｇ Ｋｏｎｇに入院したＨＣＣを有する匿名化された患者が動員された。全ての血液は、手術前に収集された。匿名化されたＨＢＶキ

ャリアおよび肝硬変対象は、Ｄｅｐａｒｔｍｅｎｔ ｏｆ Ｍｅｄｉｃｉｎｅ ａｎｄ Ｔｈｅｒａｐｅｕｔｉｃｓ ｏｆ ｔｈｅ Ｐｒｉｎｃｅ ｏｆ Ｗａｌｅｓ Ｈｏｓｐｉｔａｌ，Ｈｏｎｇ Ｋｏｎｇから動員された。試料は、血液を遠心分離して血漿を得ることによって得られた。簡潔には、末梢血試料は、その後４℃で１０分間１，６００ｇで遠心分離された、ＥＤＴＡ含有チューブに収集された。血漿部分は、さらなる分析まで－８０℃で保存された無細胞血漿を得るために、４℃で１０分間１６，０００ｇで再遠心分離された。ＤＮＡは、ＱＩＡａｍｐ ＤＳＰ ＤＮＡ Ｂｌｏｏｄ Ｍｉｎｉ Ｋｉｔ（Ｑｉａｇｅｎ）を用いて、４～１０ｍＬの血漿から抽出された。血漿ＤＮＡは、ＳｐｅｅｄＶａｃ Ｃｏｎｃｅｎｔｒａｔｏｒ（Ｓａｖａｎｔ ＤＮＡ１２０、Ｔｈｅｒｍｏ Ｓｃｉｅｎｔｉｆｉｃ）で、試料あたり７５μＬの最終体積に濃縮された。

【０１５１】

配列決定ライブラリー調製およびＤＮＡ配列決定

抽出された血漿ＤＮＡを使用して、インデックス付きＤＮＡライブラリーは、製造業者の指示に従ってペアエンド配列決定試料調製キットで構成された。このステップでは、血漿二本鎖ＤＮＡ分子は、末端修復して鈍端を形成し、同時に余分なＡ塩基が追加された。ＰＣＲ増幅を支援し、フローセルにアニーリングし、配列決定を促進することができるアダプタは、配列決定ライブラリーを形成するためにＡタグ付き二本鎖血漿ＤＮＡ分子にライゲートされた。ライブラリーは、以前に記載されるように、各端について３６または５０または７５サイクルの使用で、ペアエンドモードで配列決定することができる（Ｙｕ ｅｔ ａｌ．Ｐｒｏｃ．Ｎａｔｌ．Ａｃａｄ．Ｓｃｉ．Ｕ．Ｓ．Ａ．２０１４；１１１：８５８３－８）。

【０１５２】

配列アラインメント

各試料からの配列は、以前に記載されるように、短オリゴヌクレオチドアラインメントプログラム２（ＳＯＡＰ２）（Ｌｉ ｅｔ ａｌ．Ｂｉｏｉｎｆｏｒｍａｔｉｃｓ ２００９；２５：１９６６－１９６７）を用いてヒト参照ゲノム（ｈｇ１９）にアラインされた（Ｙｕ ｅｔ ａｌ．Ｐｒｏｃ．Ｎａｔｌ．Ａｃａｄ．Ｓｃｉ．Ｕ．Ｓ．Ａ．２０１４；１１１：８５８３－８）。平均で、各試料は、１２００万個の一意にマップされたペアエンドリードを得た（範囲：１０００万～１５００万）。

【０１５３】

メチル化レベル

配列リードの部位のメチル化状態は、本明細書に記載されるように得ることができる。例えば、ＤＮＡ分子は、ＤＮＡ分子の配列リードを使用して分析することができ、この場合、配列決定はメチル化認識である。例えば、メチル化認識配列決定は、バイサルファイト配列決定、または配列決定に先立つメチル化感受性制限酵素消化、抗メチルシトシン抗体もしくはメチル化結合タンパク質を使用する免疫沈降、またはメチル化状態の解明を可能にする単一分子配列決定を含むことができるが、これらに限定されない。他のメチル化認識アッセイも使用することができる。

【０１５４】

配列リードはそれぞれ、生体試料の無細胞ＤＮＡ分子のメチル化状態を含むことができる。メチル化状態は、特定のシトシン残基が、５－メチルシトシンまたは５－ヒドロキシメチルシトシンであるかどうかを含み得る。配列リードは、様々な方法、様々な配列決定技術、ＰＣＲ技術（例えば、リアルタイムまたはデジタル）、アレイ、および断片の配列を特定するためのその他の好適な技術で得ることができる。リアルタイムＰＣＲは、例えば、部位でメチル化されたＤＮＡの数に比例する強度信号として、ＤＮＡ群を集合的に分析する例である。配列リードは、２つの部位の相互の近接度と配列リードの長さに応じて、２つ以上の部位をカバーすることができる。

【０１５５】

分析は、メチル化認識配列決定から配列リードを受信することによって行うことができ、したがって、分析は、ＤＮＡから以前に得られたデータにのみ分析を行うことができる

。他の実施形態では、分析は、実際の配列決定、またはＤＮＡ分子の特性の測定を実行する他の能動的ステップを含み得る。配列決定は、例えば、超並列配列決定または次世代配列決定を使用して、単一分子配列決定を使用して、および／または二本鎖もしくは一本鎖ＤＮＡ配列決定ライブラリー調製プロトコル、ならびに本明細書に記載の他の技術を使用して、様々な方法で実行することができる。配列決定の一部として、配列リードの一部が細胞核酸に対応し得ることが可能である。

【０１５６】

配列決定は、例えば本明細書に記載されるような標的化配列決定であり得る。例えば、ウイルスからの核酸分子に対して生体試料を濃縮することができる。ウイルス由来の核酸分子の生体試料の濃縮には、ウイルスの一部またはウイルスのゲノム全体に結合する捕捉プローブの使用を含むことができる。他の実施形態は、ウイルスの特定の遺伝子座に特異的なプライマーを使用することができる。生体試料は、ヒトゲノムの一部、例えば常染色体の領域由来の核酸分子について濃縮され得る。図１は、そのような捕獲プローブの例を示す。他の実施形態では、配列決定はランダム配列決定を含むことができる。

【０１５７】

配列決定装置による配列決定後、配列リードは、例えば、有線または無線通信または取り外し可能な記憶装置を介して配列決定を実行する配列決定装置に通信可能に結合され得るコンピュータシステムによって受信され得る。いくつかの実施形態では、核酸断片の両端を含む１つ以上の配列リードが受信され得る。ＤＮＡ分子の位置は、ＤＮＡ分子の１つ以上の配列リードをヒトゲノムのそれぞれの部分、例えば、特異的メチル化領域（ＤＭＲ）などの特定の領域にマップ（アライン）することにより決定することができる。一実施態様では、リードが目的の領域にマップされない場合、リードを無視することができる。他の実施形態では、特定のプローブ（例えば、ＰＣＲまたは他の増幅後）は、特定の蛍光色などを介して位置を示すことができる。特定は、無細胞ＤＮＡ分子が１つ以上の部位のセットのうちの１つに対応するものであり得、すなわち、１つ以上の部位でメチル化されたＤＮＡの量が必要とされる全てであるため、特定の部位が不明となり得る。

【０１５８】

配列決定およびアラインメント後、個々のＣｐＧ部位のメチル化状態はよって、ＣｐＧ文脈におけるシトシン残基でのメチル化配列リード「Ｍ」（メチル化）の数および非メチル化配列リード「Ｕ」（非メチル化）の数から推測することができた。バイサルファイト配列決定データを使用して、母体血液、胎盤、および母体血漿のメチローム全体が構成された。母体血漿中の特定の遺伝子座のメチル化ＣｐＧ密度平均値（メチル化密度、ＭＤとも呼ばれる）は、下記方程式を用いて計算することができ、

【数２】

$$MD = \frac{M}{M+U}$$

式中、Ｍは、メチル化リードの数であり、Ｕは、遺伝子座内のＣｐＧ部位での非メチル化リードの数である。遺伝子座内に２つ以上のＣｐＧ部位がある場合、ＭおよびＵは、部位にわたる数に対応する。

【０１５９】

代替として、メチル化アッセイは、Ｉｎｆｉｎｉｕｍ ＨＤメチル化アッセイプロトコルに従って、バイサルファイト変換されたゲノムＤＮＡで行うことができる。ハイブリダイズされたビーズチップは、Ｉｌｌｕｍｉｎａ ｉＳｃａｎ装置でスキャンすることができる。ＤＮＡメチル化データは、内部制御に対する正規化およびバックグラウンド減算とともに、ＧｅｎｏｍｅＳｔｕｄｉｏ（ｖ２０１１．１）Ｍｅｔｈｙｌａｔｉｏｎ Ｍｏｄｕｌｅ（ｖ１．９．０）ソフトウェアによって分析された。個々のＣｐＧ部位のメチル化指数は、ベータ値（β）によって表すことができ、これは、メチル化および非メチル化対立遺伝子間の蛍光強度の比を用いて計算され得る。

【数３】

$$\beta = \frac{\text{メチル化対立遺伝子の強度}}{\text{非メチル化対立遺伝子の強度+メチル化強度+100}}$$

【０１６０】

胎児ＤＮＡ画分の計算

男性胎児を有する妊娠において、母体血漿試料中の胎児ＤＮＡ画分（ｆ）は、Ｙ染色体（ｃｈｒＹ％）にアラインされたリードの割合から決定することができる。以前の研究では、女性胎児を有する妊婦の血漿中の少数の配列は、Ｙ染色体に誤ってアラインされていたことが示された（Ｃｈｉｕ　ｅｔ　ａｌ．Ｐｒｏｃ　Ｎａｔｌ　Ａｃａｄ　Ｓｃｉ　ＵＳＡ　２００８；１０５：２０４５８－２０４６３）。したがって、男性胎児を有する妊婦の血漿中のｃｈｒＹ％は、男性胎児に由来するＹ染色体リードおよびＹ染色体にミスアラインされた母体リードの混合物であった（Ｃｈｉｕ　ｅｔ　ａｌ．ＢＭＪ　２０１１；３４２：ｃ７４０１）。男性胎児を有する妊娠におけるｃｈｒＹ％とｆとの間の関係は、下記方程式を用いて表すことができ、

$chrY\% = chrY_{男性}\% \times f - chrY_{女性}\% \times (1 - f)$

式中、$chrY_{男性}\%$は、１００％男性ＤＮＡを含有する血漿試料におけるＹ染色体にアラインされたリードの割合であり、ｃｈｒＹ女性％は、１００％女性ＤＮＡを含有する血漿試料におけるＹ染色体にアラインされたリードの割合である。

【０１６１】

特定の実施形態の具体的な詳細は、本発明の実施形態の趣旨および範囲から逸脱することなく、任意の好適な様態で組み合わせることができる。しかしながら、本発明の他の実施形態は、各個々の態様、またはこれらの個々の態様の具体的な組み合わせに関する具体的な実施形態を対象とし得る。

【０１６２】

ＩＶ．例示的なシステム

図１５は、本発明の一実施形態によるシステム１５００を例示する。示されるシステムは、試料ホルダ１５１０内の無細胞ＤＮＡ分子などの試料１５０５を含み、試料１５０５は、アッセイ１５０８と接触して物理的特性１５１５の信号を提供することができる。試料ホルダの例は、アッセイのプローブおよび／もしくはプライマー、または液滴が（アッセイを含む液滴とともに）移動するチューブを含む、フローセルであり得る。試料からの蛍光強度値などの物理的特性１５１５は、検出器１５２０によって検出される。検出器は、データ信号を構成するデータ点を得るために、間隔（例えば、周期的な間隔）を空けて測定を行うことができる。一実施形態において、アナログデジタル変換器は、検出器からのアナログ信号をデジタル形態へと複数回変換する。試料ホルダ１５１０および検出器１５２０は、アッセイデバイス、例えば、本明細書に記載される実施形態に従って配列決定を行う配列決定デバイスを形成することができる。データ信号１５２５は、検出器１５２０から論理システム１５３０へ送信される。データ信号１５２５は、ローカルメモリ１５３５、外部メモリ１５４０、またはストレージデバイス１５４５に保存され得る。

【０１６３】

論理システム１５３０は、コンピュータシステム、ＡＳＩＣ、マイクロプロセッサなどであってもよいか、またはそれらを含んでもよい。それはまた、ディスプレイ（例えば、モニタ、ＬＥＤディスプレイなど）、およびユーザ入力デバイス（例えば、マウス、キーボード、ボタンなど）を含み得るか、またはそれらに連結され得る。論理システム１５３０および他の構成要素は、スタンドアローンもしくはネットワーク接続されたコンピュータシステムの一部であってもよく、または検出器１５２０および／もしくは試料ホルダ１５１０を含むデバイス（例えば、配列決定デバイス）に直接取り付けられても組み込まれてもよい。論理システム１５３０はまた、プロセッサ１５５０において実行するソフトウェアを含み得る。論理システム１５３０は、本明細書に記載される方法のいずれかを行うようにシステム１５００を制御するための指示を保存するコンピュータ可読媒体を含み得

る。例えば、論理システム１５３０は、配列決定または他の物理的操作が行われるように、試料ホルダ１５１０を含むシステムにコマンドを提供することができる。そのような物理的操作は、特定の順序で、例えば、試薬が特定の順序で追加および除去されるように、行うことができる。そのような物理的操作は、試料を入手してアッセイを行うために使用され得るように、例えば、ロボットアームを含む、ロボットシステムによって行われ得る。

【０１６４】

本明細書で言及されるコンピュータシステムのうちのいずれも、任意の好適な数のサブシステムを利用してもよい。そのようなサブシステムの例は、図１６においてコンピュータ装置１６００において示されている。いくつかの実施形態において、コンピュータシステムは、単一のコンピュータ装置を含み、サブシステムは、コンピュータ装置の構成要素であり得る。他の実施形態において、コンピュータシステムは、各々がサブシステムであり、内部構成要素を備える、複数のコンピュータ装置を含むことができる。

【０１６５】

図１６に示されるサブシステムは、システムバス１６７５を介して相互接続される。プリンタ１６７４、キーボード１６７８、固定ディスク１６７９、ディスプレイアダプタ１６８２に結合しているモニタ１６７６などの追加のサブシステムなどが示されている。Ｉ／Ｏコントローラ１６７１に連結した、周辺および入力／出力（Ｉ／Ｏ）デバイスは、シリアルポート１６７７などの、当該技術分野で既知の任意の数の手段によって、コンピュータシステムへ接続することができる。例えば、シリアルポート１６７７または外部インターフェース１６８１（例えば、Ｅｔｈｅｒｎｅｔ、Ｗｉ－Ｆｉなど）を使用して、Ｉｎｔｅｒｎｅｔなどの広域ネットワーク、マウス入力デバイス、またはスキャナに、コンピュータ装置１６００を接続することができる。システムバス１６７５を介した相互接続は、中央プロセッサ１６７３が各サブシステムと通信し、システムメモリ１６７２または固定ディスク１６７９からの指示の実行、およびサブシステム間の情報の交換を制御することを可能にする。システムメモリ１６７２および／または固定ディスク１６７９は、コンピュータ可読媒体を具現化し得る。本明細書に言及される値のうちのいずれも、１つの構成要素から別の構成要素へ出力することができ、ユーザへ出力することができる。

【０１６６】

コンピュータシステムは、例えば、外部インターフェース１６８１によって、または内部インターフェースによって一緒に接続された、複数の同じ構成要素またはサブシステムを含むことができる。いくつかの実施形態では、コンピュータシステム、サブシステム、または装置は、ネットワーク上で通信することができる。かかる例において、１つのコンピュータをクライアント、別のコンピュータをサーバとみなすことができ、各々が、同じコンピュータシステムの一部であり得る。クライアントおよびサーバは各々、複数のシステム、サブシステム、または構成要素を含むことができる。

【０１６７】

本発明の実施形態のうちのいずれも、ハードウェア（例えば、特定用途向け集積回路またはフィールドプログラマブルゲートアレイ）を使用して、かつ／またはモジュラー様式もしくは統合様式で一般にプログラム可能なプロセッサとともにコンピュータソフトウェアを使用して、制御論理の形態で実装することができることを理解されたい。本開示および本明細書に提供される教示に基づいて、当業者は、ハードウェア、およびハードウェアとソフトウェアとの組み合わせを使用して、本発明の実施形態を実装するための他の方法および／または方法を認識および理解するであろう。

【０１６８】

本出願に記載されるソフトウェア構成要素または機能のうちのいずれも、例えば、従来のまたはオブジェクト指向技法を使用する、例えば、Ｊａｖａ、Ｃ＋＋、Ｐｙｔｈｏｎ、またはＰｅｒｌなどの任意の好適なコンピュータ言語を用いてプロセッサによって実行されるソフトウェアコードとして実装されてもよい。ソフトウェアコードは、記憶および／または伝送のために、コンピュータ可読媒体上に一連の命令またはコマンドとして記憶さ

れ得、好適な媒体としては、ランダムアクセスメモリ（ＲＡＭ）、リードオンリーメモリ（ＲＯＭ）、磁気媒体（ハードドライブもしくはフロッピーディスク等）、または光学媒体（コンパクトディスク（ＣＤ）もしくはＤＶＤ（デジタル多用途ディスク）等）、およびフラッシュメモリ等が挙げられる。コンピュータ可読媒体は、かかる記憶または送信デバイスの任意の組み合わせであってもよい。

【０１６９】

かかるプログラムはまた、コード化され、インターネットを含む様々なプロトコルに従う有線ネットワーク、光ネットワーク、および／または無線ネットワークを介した送信に適合した搬送波信号を使用して送信されてもよい。したがって、本発明の一実施形態に従うコンピュータ可読媒体は、そのようなプログラムでコードされたデータ信号を使用して作製されてもよい。プログラムコードでコード化されたコンピュータ可読媒体は、互換性のあるデバイスでパッケージ化されていてもよく、または（例えば、インターネットダウンロードを介して）他のデバイスとは別個に提供されてもよい。任意のそのようなコンピュータ可読媒体は、単一のコンピュータプログラム製品（例えば、ハードドライブ、ＣＤ、またはコンピュータシステム全体）上またはその内部に存在してもよく、システムまたはネットワーク内の異なるコンピュータプログラム製品上またはその内部に存在してもよい。コンピュータシステムは、モニタ、プリンタ、または本明細書に記載の結果のうちのいずれかをユーザへ提供するための他の好適なディスプレイを含み得る。

【０１７０】

本明細書に記載される方法のうちのいずれも、ステップを行うように構成することができる、１つ以上のプロセッサを含むコンピュータシステムで全体的または部分的に行われ得る。したがって、実施形態は、本明細書に説明される方法のうちのいずれかのステップを実行するように構成されたコンピュータシステムを対象とし得、潜在的には異なる構成要素がそれぞれのステップまたはそれぞれのステップ群を実行する。番号付けされたステップとして提示されるものの、本明細書における方法のステップは、同時にまたは異なる順序で行うことができる。加えて、これらのステップの部分は、他の方法からの他のステップの部分と併用することができる。また、あるステップの全てまたは部分は、任意選択的であってもよい。さらに、本方法のうちのいずれかのステップのうちのいずれも、モジュール、回路、またはこれらのステップを実施するための他の手段で実施することができる。

【０１７１】

本発明の例示的実施形態の上の説明は、例解および説明目的で提示されている。包括的であること、または本発明を説明された正確な形態に限定することは意図されず、多くの修正および変更が、先の教示に鑑みて可能である。

【０１７２】

前述の記載では、説明の目的で、本技術の様々な実施形態の理解を提供するために、多数の詳細が述べられてきた。しかしながら、特定の実施形態は、これらの詳細の一部なしで、または追加の詳細を用いて実施され得ることが当業者には明らかであろう。

【０１７３】

いくつかの実施形態を記載してきたが、本発明の精神から逸脱することなく、様々な修正、代替構成、および均等物が使用され得ることが当業者によって認識されるであろう。加えて、本発明を不必要に不明瞭にすることを避けるために、多数の周知のプロセスおよび要素は記載されていない。加えて、任意の特定の実施形態の詳細は、その実施形態の変形に常に存在するとは限らず、他の実施形態に追加されてもよい。

【０１７４】

値の範囲が提供される場合、文脈が明確に別段に示さない限り、その範囲の上限と下限との間の各介在する値も、下限の１０分の１まで具体的に開示されていると理解される。示された範囲における任意の示された値または介在する値と、その示された範囲における任意の他の示されたまたは介在する値との間の各より小さな範囲が包含される。これらのより小さな範囲の上限および下限は、範囲に独立して含まれるか除外されてもよく、どち

らか一方、両方の限度がより小さな範囲に含まれるか、またはどちらも含まれない各範囲も、述べられた範囲における特に除外された限度を条件として、本発明に包含される。述べられた範囲が一方または両方の限度を含む場合、それらのいずれかまたは両方を除外する範囲も含まれる。

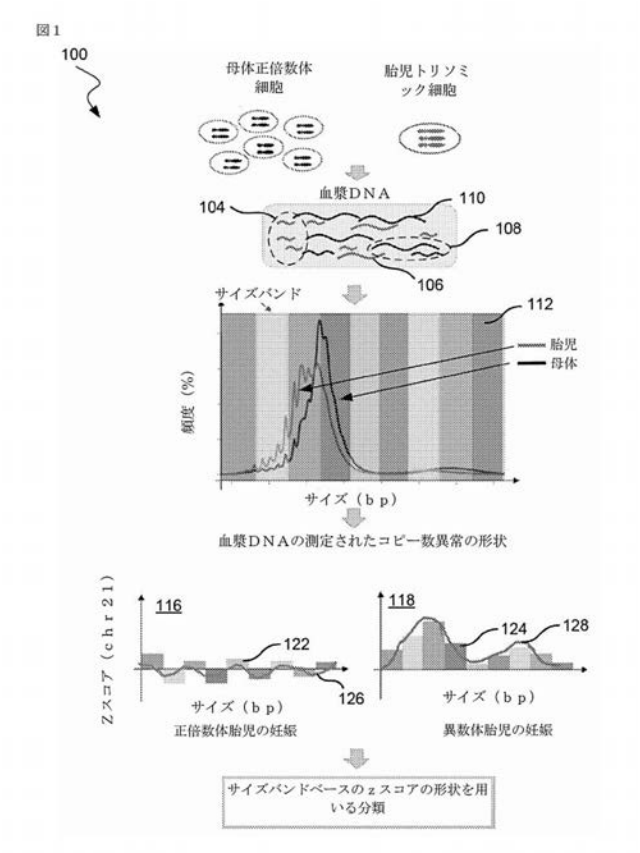
【０１７５】

本明細書および添付の特許請求の範囲で使用されるとき、単数形「ａ」、「ａｎ」、および「ｔｈｅ」は、文脈が明らかに別段に指示しない限り、複数の参照物を含む。よって、例えば、「方法」への言及は、複数のそのような方法を含み、「粒子」への言及は、当業者に知られている１つ以上の粒子およびそれらの同等物への言及を含む。本発明は、ここで明確化および理解の目的で詳細に記載されてきた。しかしながら、添付の特許請求の範囲の範囲内で特定の変更および修正が実施され得ることが理解されるであろう。

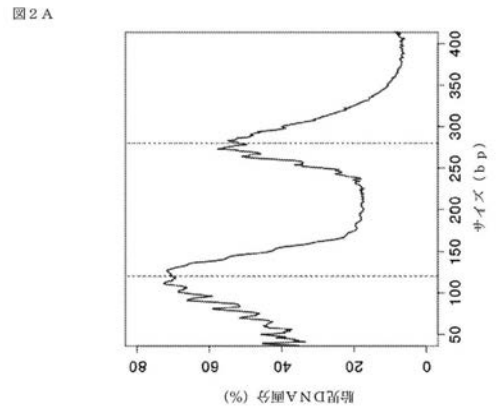
【０１７６】

本明細書で引用される全ての刊行物、特許、および特許出願は、全ての目的のためにそれらの全体が参照により本明細書に組み込まれる。いかなるものも、先行技術であるとは認められていない。

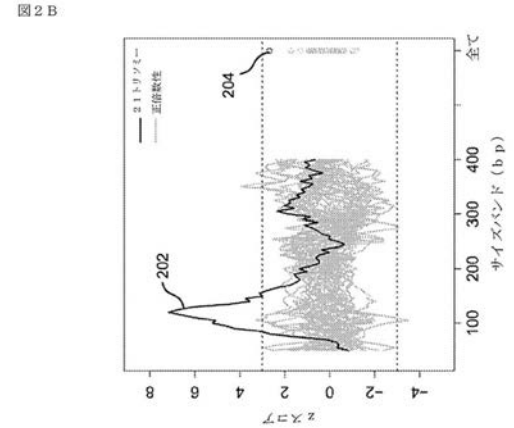
【図１】



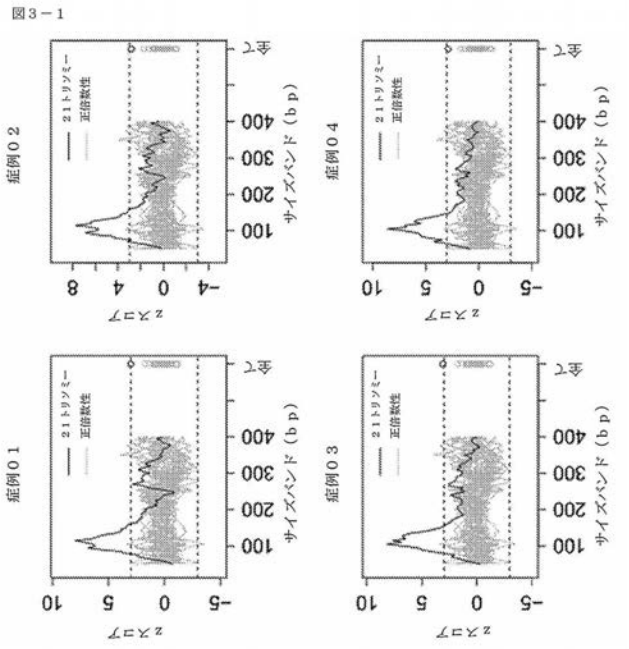
【図２Ａ】



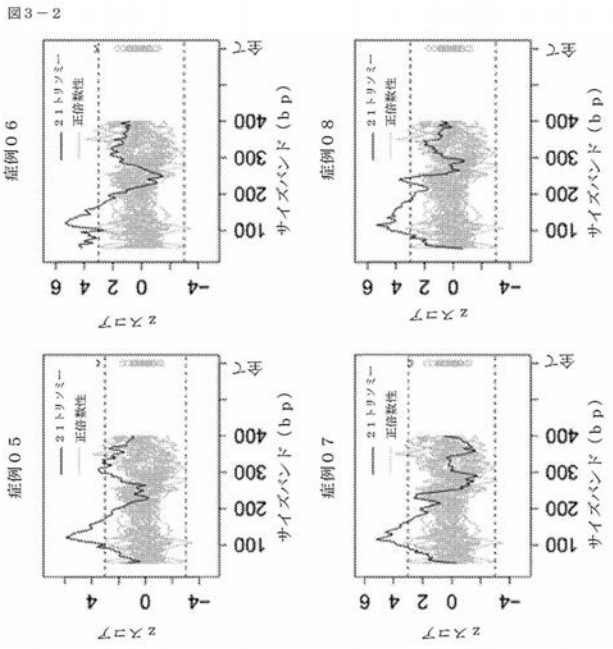
【図２Ｂ】



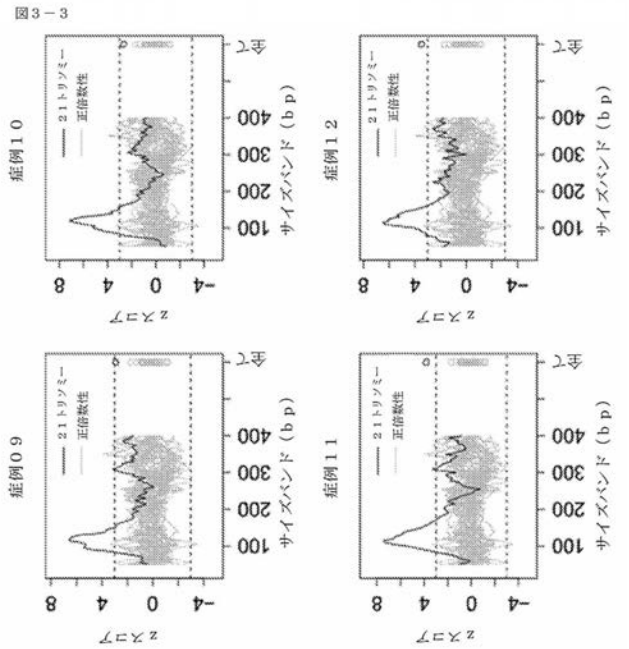
【図３－１】



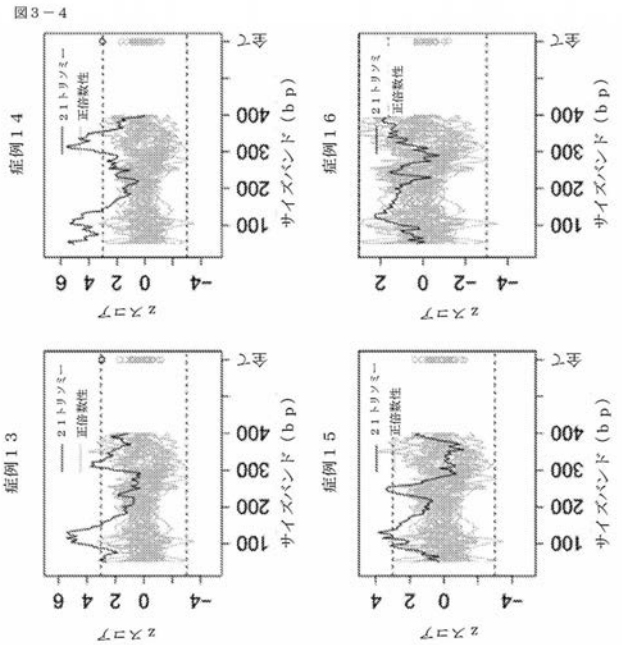
【図３－２】



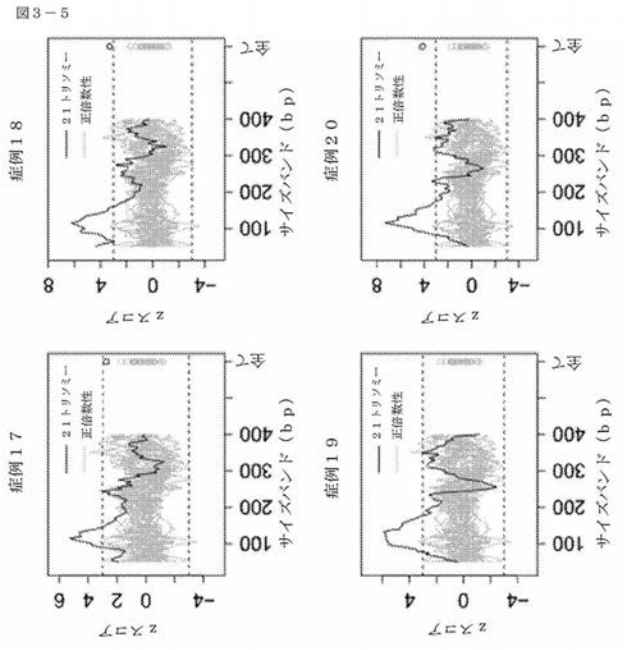
【図３－３】



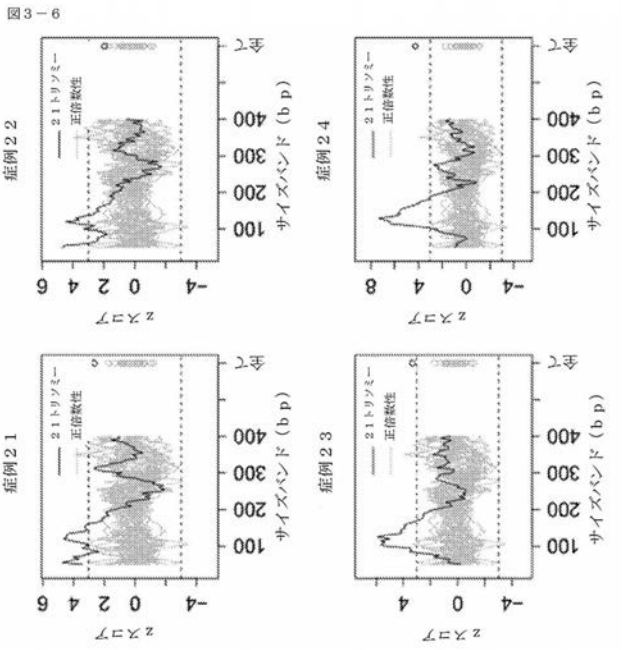
【図３－４】



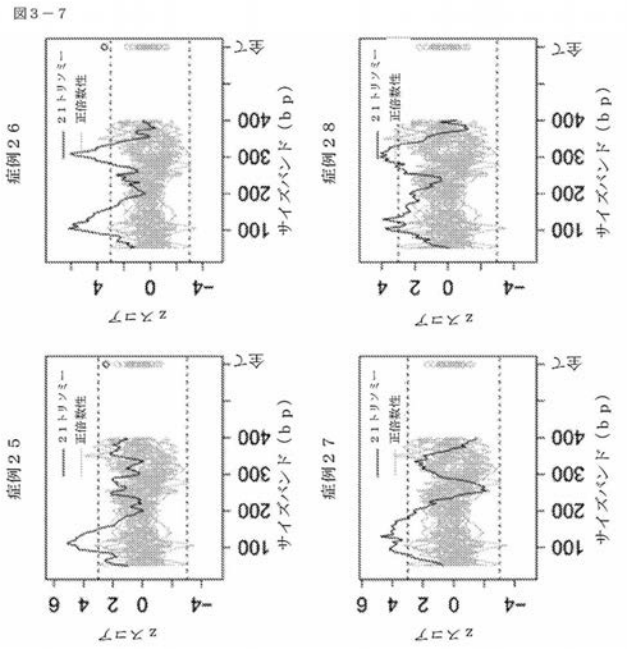
【図３－５】



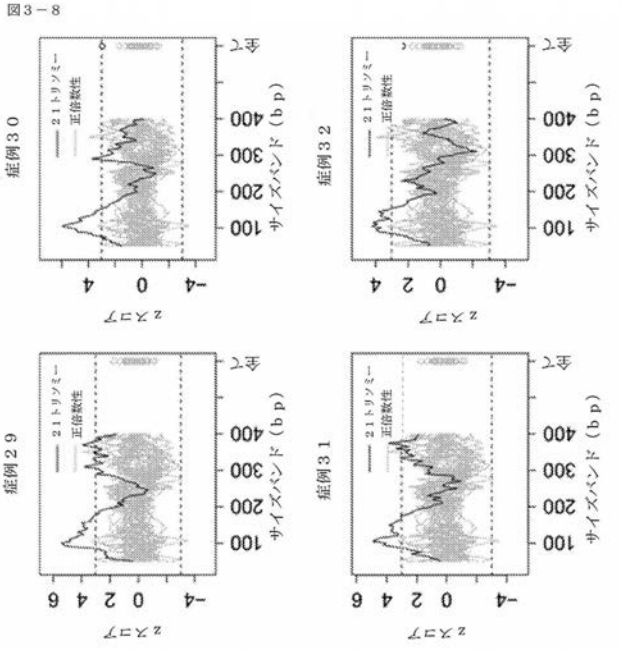
【図３－６】



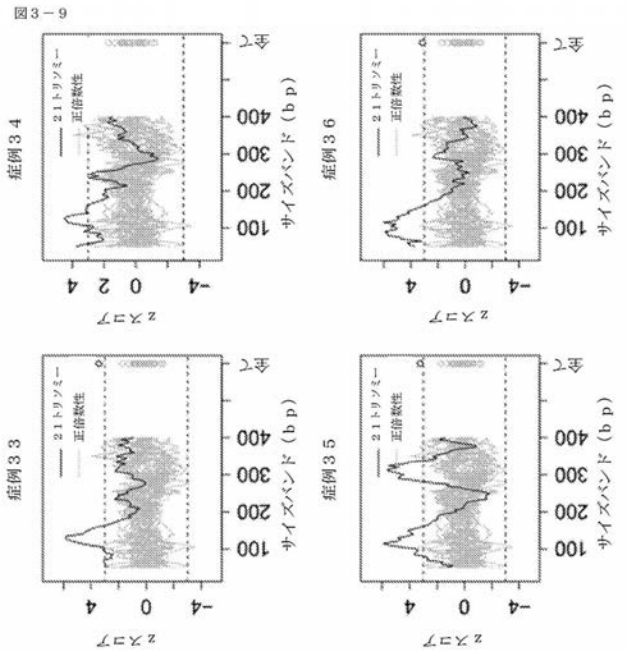
【図３－７】



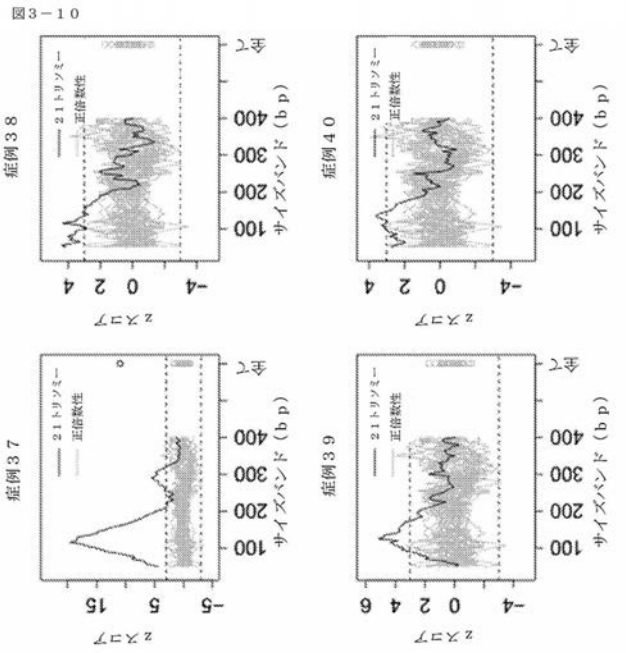
【図３－８】



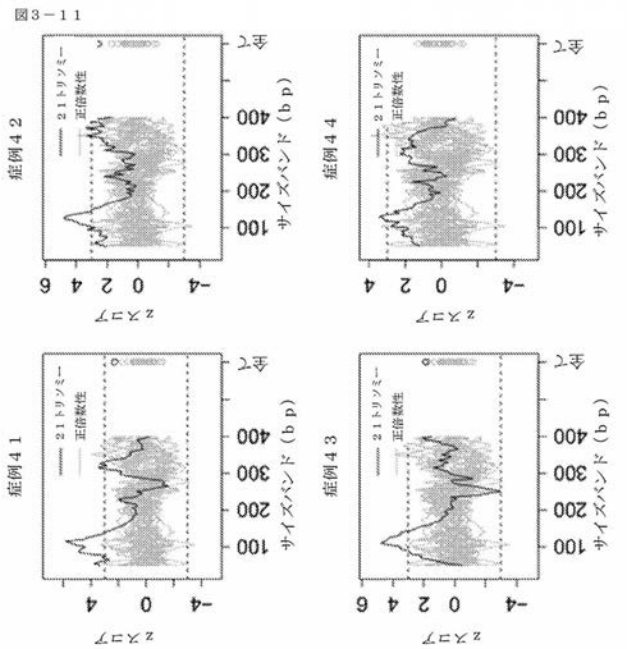
【図３－９】



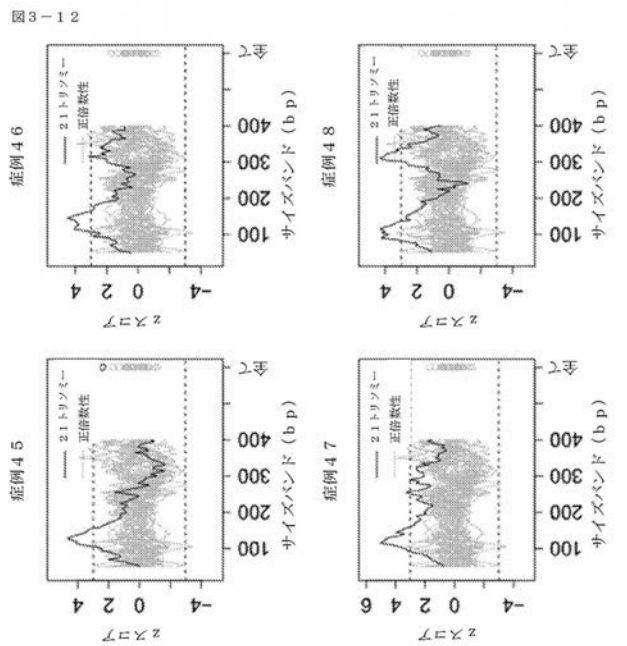
【図３－１０】



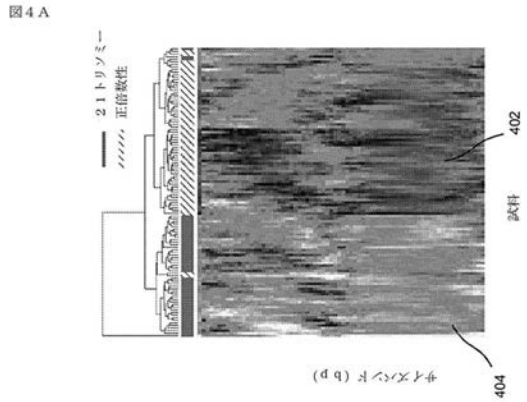
【図３－１１】



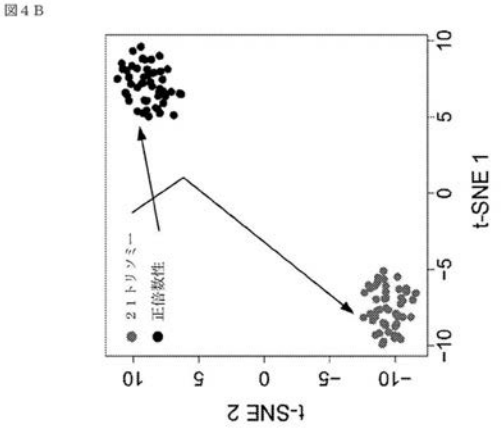
【図３－１２】



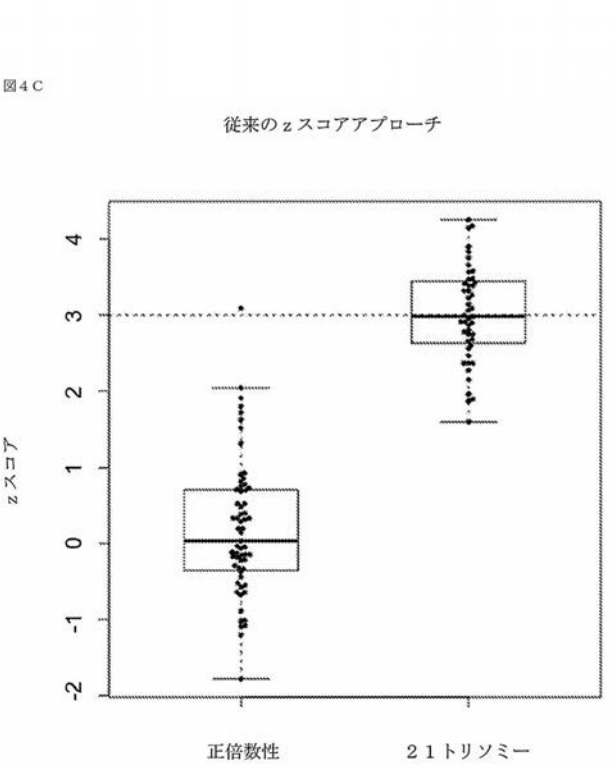
【図４Ａ】



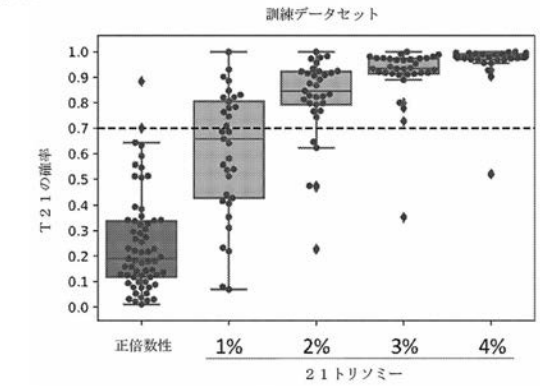
【図４Ｂ】



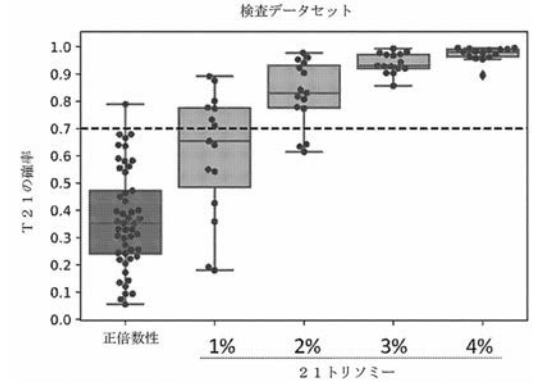
【図４Ｃ】



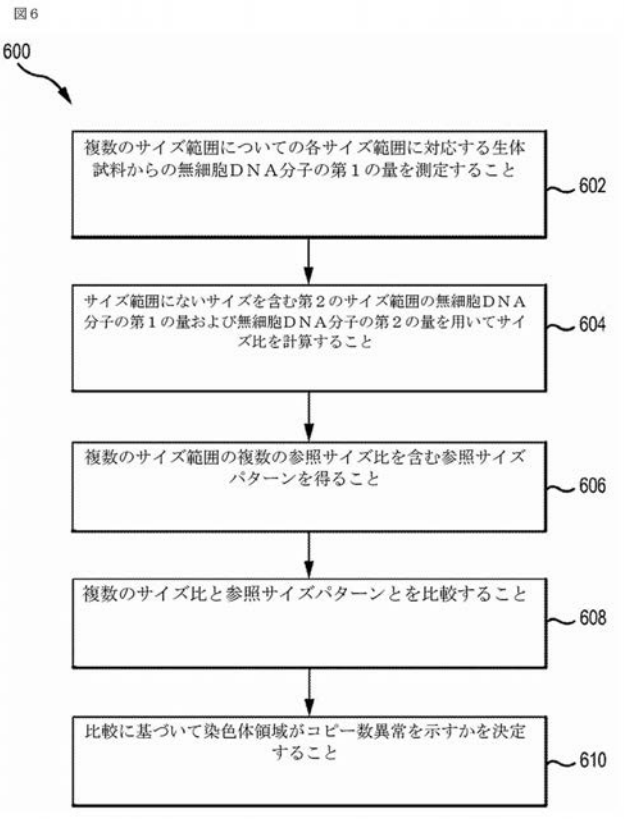
【図５Ａ】



【図５Ｂ】



【図６】



【図７】

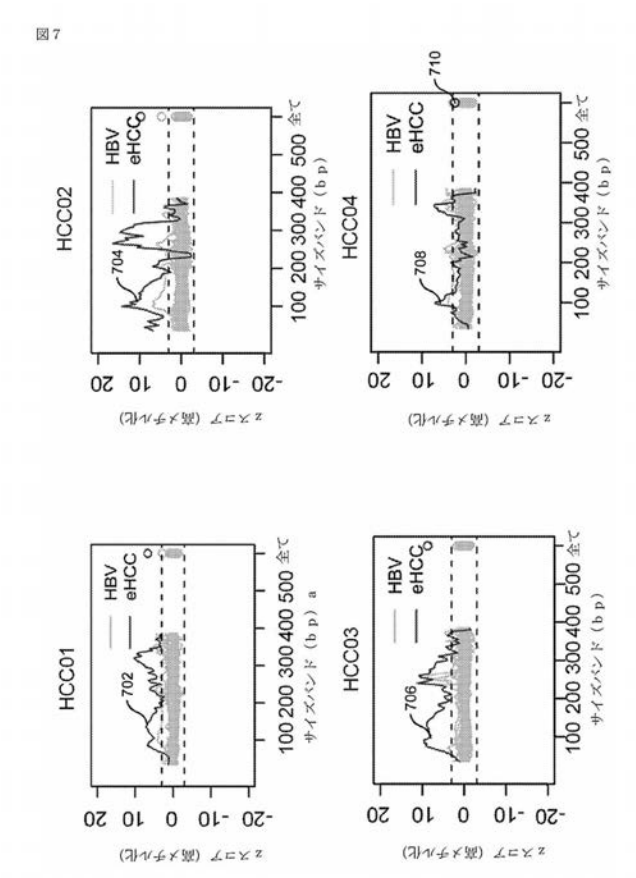


図7
HCC02
HCC04
HCC01
HCC03
HBV
eHCC
704
708
702
706
710
サイズバンド（ｂｐ）
ｚスコア（再メチル化）

【図８】

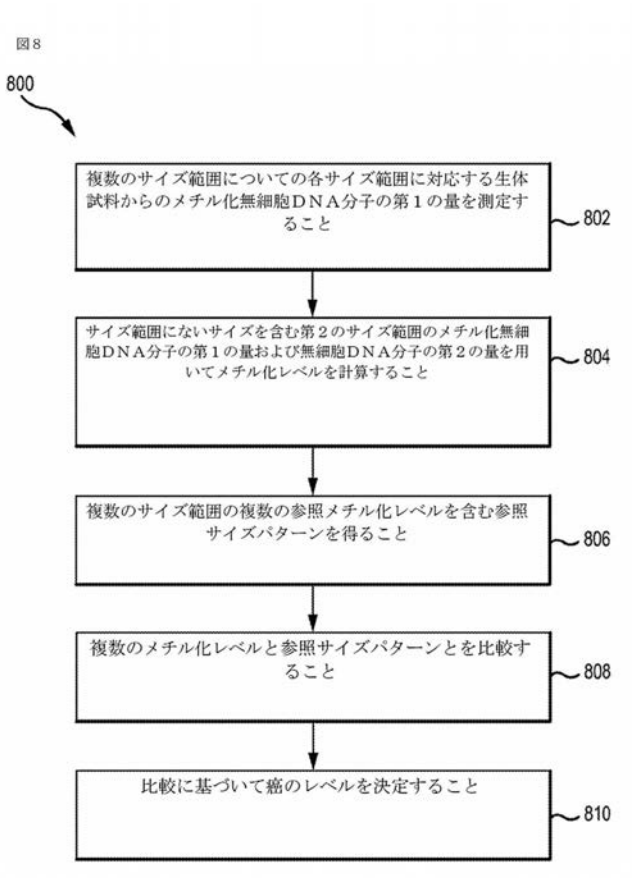


図8
800
複数のサイズ範囲についての各サイズ範囲に対応する生体試料からのメチル化無細胞ＤＮＡ分子の第１の量を測定すること
802
サイズ範囲にないサイズを含む第２のサイズ範囲のメチル化無細胞ＤＮＡ分子の第１の量および無細胞ＤＮＡ分子の第２の量を用いてメチル化レベルを計算すること
804
複数のサイズ範囲の複数の参照メチル化レベルを含む参照サイズパターンを得ること
806
複数のメチル化レベルと参照サイズパターンとを比較すること
808
比較に基づいて癌のレベルを決定すること
810

【図９】

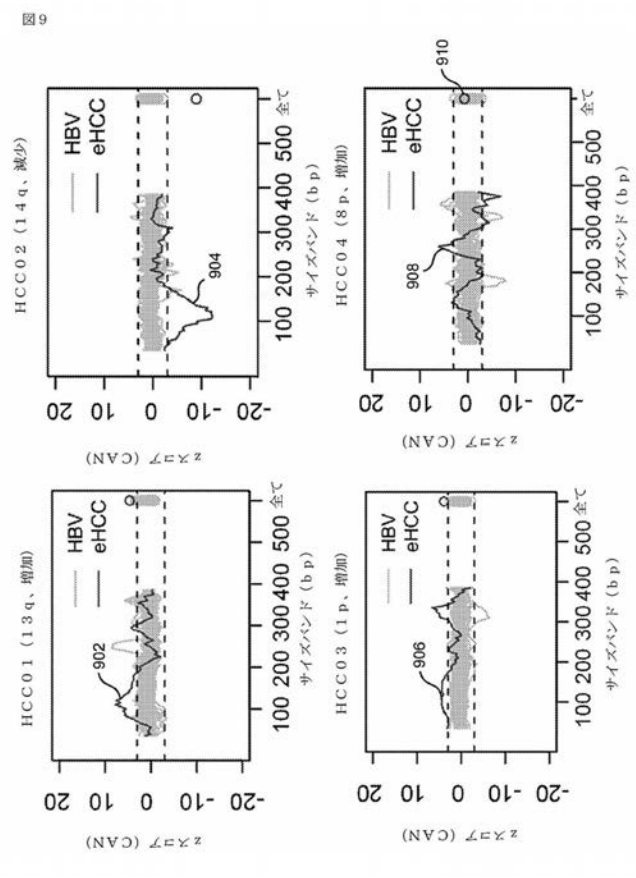


図9
HCC02（14q、減少）
HCC04（8p、増加）
HCC01（13q、増加）
HCC03（1p、増加）
HBV
eHCC
904
908
902
906
910
サイズバンド（ｂｐ）
ｚスコア（CAN）

【図１０】

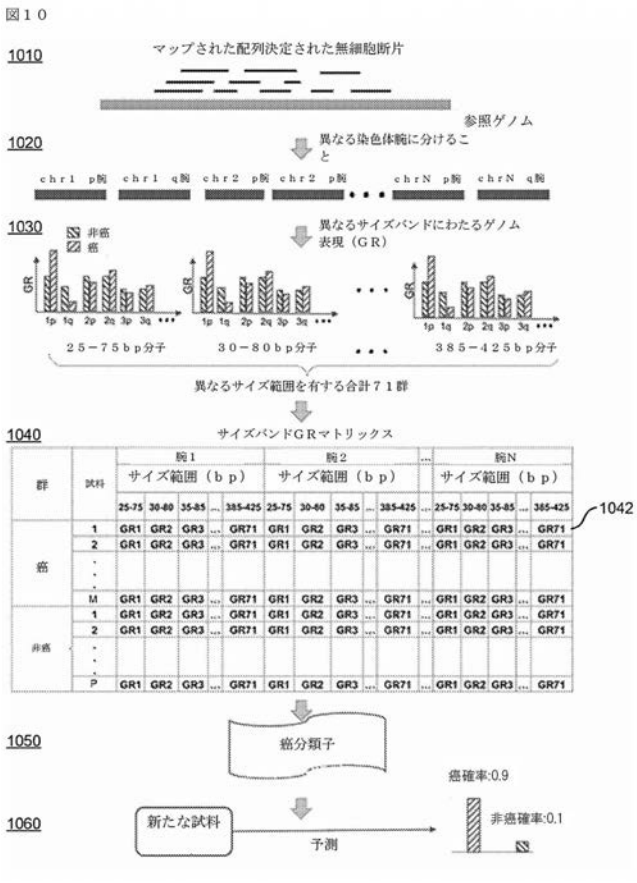
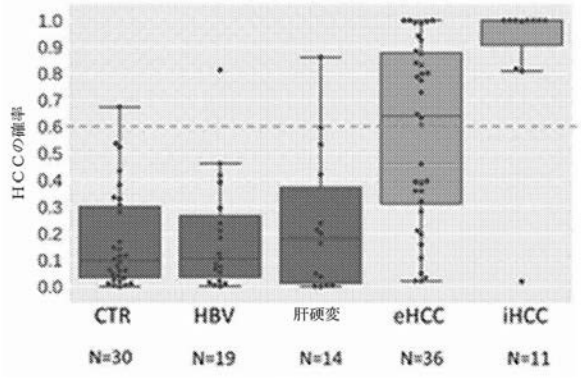


図10
1010
マップされた配列決定された無細胞断片
参照ゲノム
1020
異なる染色体腕に分けること
chr1 p腕
chr1 q腕
chr2 p腕
chr2 p腕
chrN p腕
chrN q腕
1030
異なるサイズバンドにわたるゲノム表現（GR）
非癌
癌
GR
25－75ｂｐ分子
30－80ｂｐ分子
385－425ｂｐ分子
異なるサイズ範囲を有する合計７１群
1040
サイズバンドGRマトリックス
群
試料
腕1
腕2
腕N
サイズ範囲（ｂｐ）
25-75
30-80
35-85
385-425
GR1
GR2
GR3
GR71
癌
非癌
1042
1050
癌分類子
癌確率:0.9
1060
新たな試料
予測
非癌確率:0.1

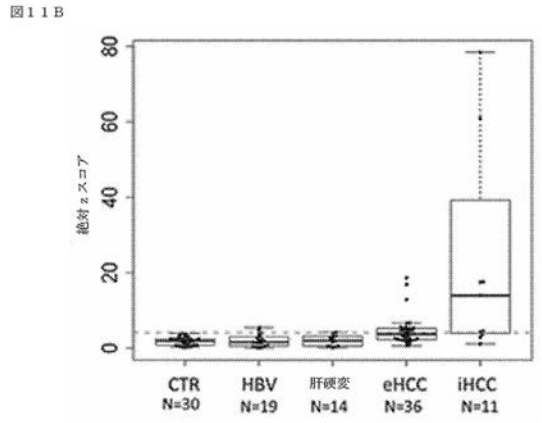
【図１１Ａ】

図１１Ａ



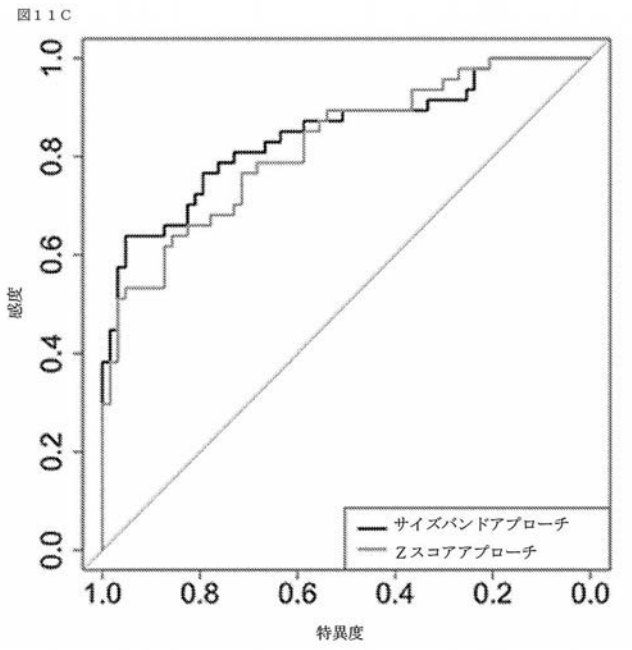
【図１１Ｂ】

図１１Ｂ



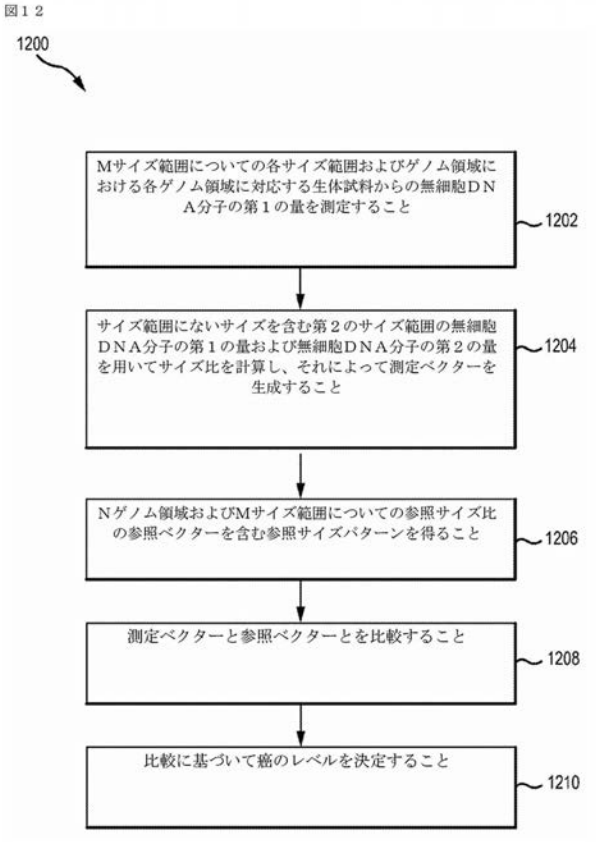
【図１１Ｃ】

図１１Ｃ



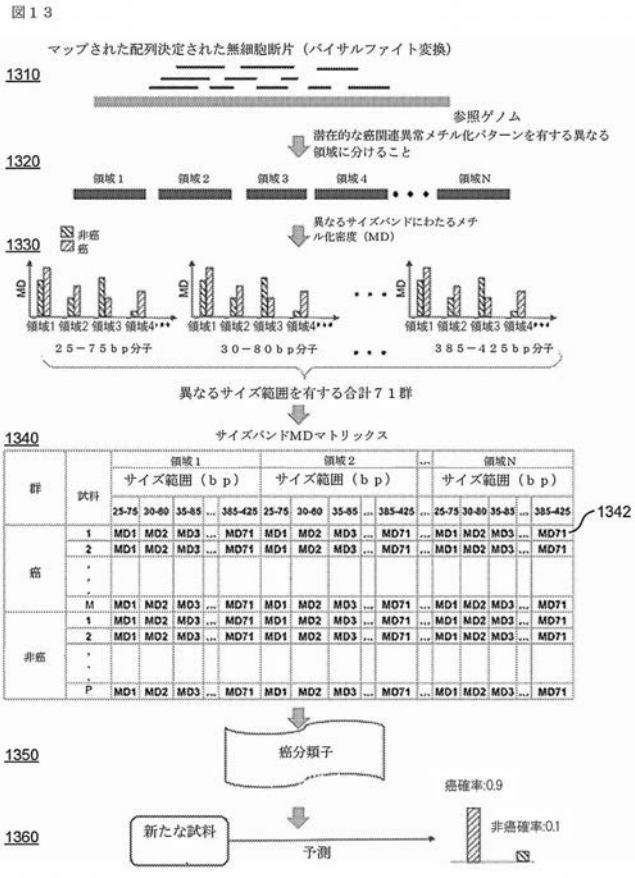
【図１２】

図１２

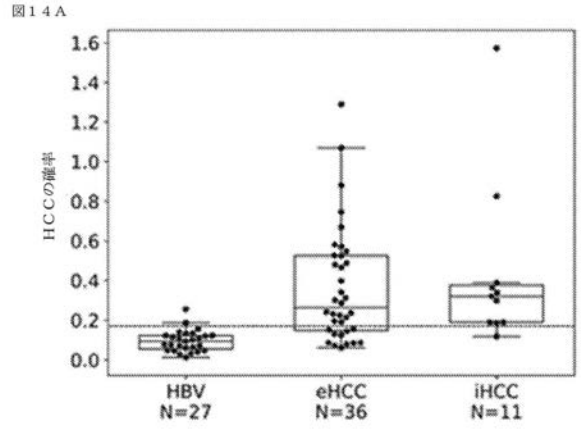


【図１３】

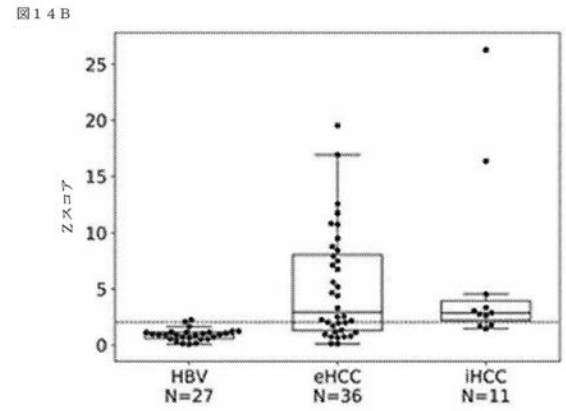
図１３



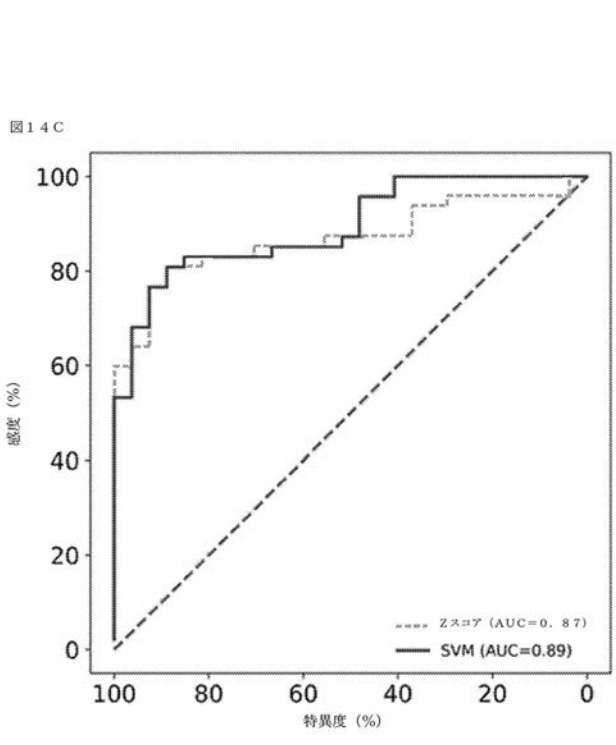
【図１４Ａ】



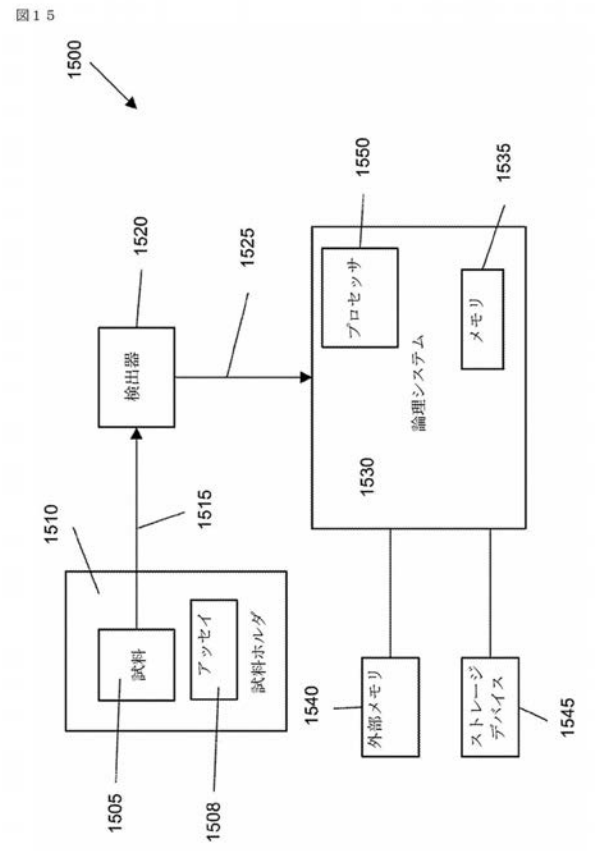
【図１４Ｂ】



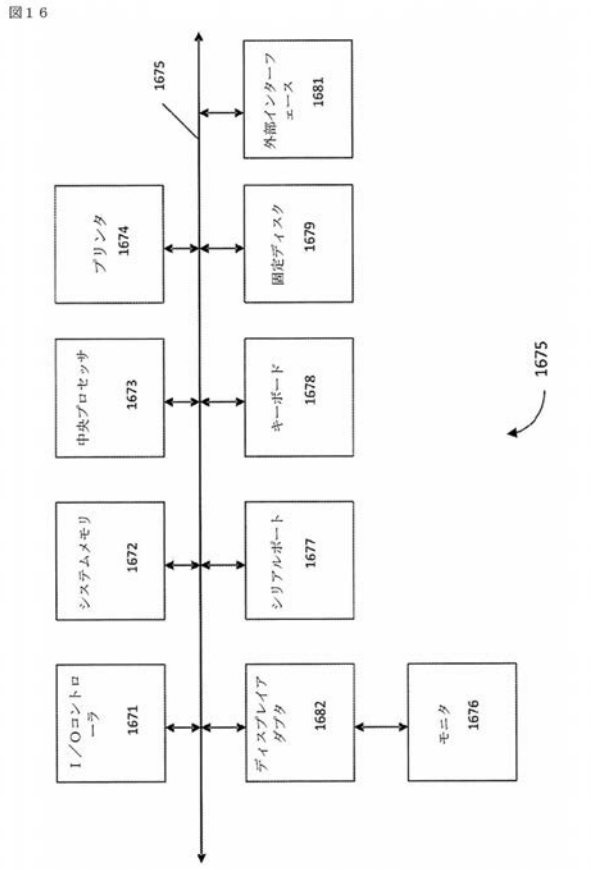
【図１４Ｃ】



【図１５】



【図１６】



【国際調査報告】

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2018/113640

A. CLASSIFICATION OF SUBJECT MATTER

C12Q 1/68(2018.01)i

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

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)

CNABS, VEN, WOTXT, USTXT, EPTXT, SIPOABS, DWPI, CNKI, GOOGLE, PubMed, ISI Web of Knowledge: THE CHINESE UNIVERSITY OF HONG KONG, size range, copy number aberration,cell-free DNA, size ratio, size pattern, fetal, maternal, z-score, methylation, cancer, tumor

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	YU, S.C.Y. et al. "Size-based molecular diagnostics using plasma DNA for noninvasive prenatal testing" *PNAS*, Vol. 111, No. 23, 10 June 2014 (2014-06-10), pages 8583–8588	1-20
Y	YU, S.C.Y. et al. "Size-based molecular diagnostics using plasma DNA for noninvasive prenatal testing" *PNAS*, Vol. 111, No. 23, 10 June 2014 (2014-06-10), pages 8583–8588	21-43
X	WO 2016112850 A1 (THE CHINESE UNIVERSITY OF HONG KONG) 21 July 2016 (2016-07-21) claims 1-39	1-20, 31-43
Y	WO 2016008451 A1 (THE CHINESE UNIVERSITY OF HONG KONG) 21 January 2016 (2016-01-21) claims 1-59	21-43
A	WO 2016116033 A1 (THE CHINESE UNIVERSITY OF HONG KONG) 28 July 2016 (2016-07-28) claims 1-30	1-20

☑ Further documents are listed in the continuation of Box C. ☑ See patent family annex.

* Special categories of cited documents:
"A" document defining the general state of the art which is not considered to be of particular relevance
"E" earlier application or patent but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed
"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
07 January 2019	**01 February 2019**
Name and mailing address of the ISA/CN **National Intellectual Property Administration, PRC** **6, Xitucheng Rd., Jimen Bridge, Haidian District, Beijing 100088** **China** Facsimile No. **(86-10)62019451**	Authorized officer **CHEN,Ying** Telephone No. **86-(010)-53961975**

Form PCT/ISA/210 (second sheet) (January 2015)

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2018/113640

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	WO 2009051842 A2 (THE JOHNS HOPKINS UNIVERSITY) 23 April 2009 (2009-04-23) the whole document	1-43
A	WO 2014043763 A1 (THE CHINESE UNIVERSITY OF HONG KONG) 27 March 2014 (2014-03-27) the whole document	1-43
A	CN 106886688 A (UNIV HONG KONG CHINESE) 23 June 2017 (2017-06-23) the whole document	1-43
A	CN 105243295 A (UNIV HONG KONG CHINESE) 13 January 2016 (2016-01-13) the whole document	1-43

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/CN2018/113640

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
WO	2016112850	A1	21 July 2016	CN	107406885	A	28 November 2017
				EP	3245299	A4	04 July 2018
				SG	11201705685V	A	30 August 2017
				US	2016201142	A1	14 July 2016
				TW	201639969	A	16 November 2016
				JP	2018512042	A	10 May 2018
				CA	2973025	A1	21 July 2016
				AU	2016207171	A1	20 July 2017
				KR	20170106372	A	20 September 2017
				EP	3245299	A1	22 November 2017
WO	2016008451	A1	21 January 2016	CN	106795562	A	31 May 2017
				TW	201617455	A	16 May 2016
				EP	3169813	A1	24 May 2017
				EP	3169813	A4	03 January 2018
				HK	1232256	A1	05 January 2018
				US	2016017419	A1	21 January 2016
WO	2016116033	A1	28 July 2016	EP	3247806	A1	29 November 2017
				CA	2974239	A1	28 July 2016
				CN	107208155	A	26 September 2017
				TW	201641694	A	01 December 2016
				EP	3247806	A4	20 June 2018
				US	2016217251	A1	28 July 2016
				AU	2016208914	A1	10 August 2017
WO	2009051842	A2	23 April 2009	WO	2009051842	A3	13 August 2009
WO	2014043763	A1	27 March 2014	EP	3354747	A1	01 August 2018
				ES	2665273	T3	25 April 2018
				DK	2898100	T3	26 February 2018
				EP	2898100	B1	22 November 2017
				AU	2017251832	A1	16 November 2017
				JP	2015536639	A	24 December 2015
				AU	2013317708	A1	09 April 2015
				JP	6317354	B2	25 April 2018
				MX	2015003571	A	22 June 2015
				HK	1207124	A1	22 January 2016
				KR	20150082228	A	15 July 2015
				CA	2884066	A1	27 March 2014
				PT	2898100	T	26 February 2018
				AU	2013317708	C1	23 November 2017
				NZ	706269	A	26 August 2016
				EP	2898100	A1	29 July 2015
				TW	201418474	A	16 May 2014
				NZ	717423	A	25 August 2017
				SG	11201501927V	A	29 April 2015
				ZA	201501772	B	26 October 2016
				EA	201500327	A1	30 July 2015
				HK	1206797	A1	15 January 2016
				JP	2018121644	A	09 August 2018
				IL	237495	A	30 April 2015
				EP	2898100	A4	11 May 2016

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/CN2018/113640

Patent document cited in search report	Publication date (day/month/year)	Patent family member(s)	Publication date (day/month/year)
		SG 10201705198U A	28 July 2017
		CN 104781422 A	15 July 2015
		PL 2898100 T3	29 June 2018
		BR 112015006183 A2	04 July 2017
		PH 12015500547 A1	04 May 2015
		AU 2013317708 B2	03 August 2017
		IL 237495 D0	30 April 2015
CN 106886688 A	23 June 2017	JP 6151739 B2	21 June 2017
		HR P20160493 T1	15 July 2016
		EP 2557517 A2	13 February 2013
		SG 183062 A1	30 August 2012
		JP 2017148073 A	31 August 2017
		EP 2557520 A2	13 February 2013
		HK 1182195 A1	05 January 2018
		KR 20160113145 A	28 September 2016
		EP 2557518 A2	13 February 2013
		AU 2008278839 A1	29 January 2009
		EP 2557520 A3	29 April 2015
		HK 1144024 A1	09 May 2014
		CN 103849684 A	11 June 2014
		EP 2557517 A3	05 November 2014
		CY 1114773 T1	14 December 2016
		EA 201201551 A1	30 August 2013
		KR 20180100713 A	11 September 2018
		WO 2009013496 A1	29 January 2009
		EP 3067807 A1	14 September 2016
		AU 2008278843 A1	29 January 2009
		HK 1177768 A1	26 August 2016
		EP 2557518 A3	08 October 2014
		HK 1224033 A1	11 August 2017
		CA 3009992 A1	29 January 2009
		EP 2183692 A1	12 May 2010
		DK 2183693 T3	20 January 2014
		JP 6383837 B2	29 August 2018
		KR 101829564 B1	14 February 2018
		EP 2557519 A2	13 February 2013
		US 2009029377 A1	29 January 2009
		EA 201000231 A1	30 June 2010
		EP 2557519 A3	23 July 2014
		US 2014256560 A1	11 September 2014
		HR P20140009 T1	14 March 2014
		CY 1117525 T1	26 April 2017
		US 8706422 B2	22 April 2014
		KR 101829565 B1	29 March 2018
		KR 20160113148 A	28 September 2016
		EP 2527471 A3	26 December 2012
		MX 2010000846 A	21 April 2010
		EP 2183693 A1	12 May 2010
		CN 107083424 A	22 August 2017
		SI 2183693 T1	28 February 2014

Form PCT/ISA/210 (patent family annex) (January 2015)

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/CN2018/113640

Patent document cited in search report	Publication date (day/month/year)	Patent family member(s)	Publication date (day/month/year)
		KR 20180114251 A	17 October 2018
		KR 20100058503 A	03 June 2010
		EP 2527471 A2	28 November 2012
		KR 20160030404 A	17 March 2016
		DK 2514842 T3	30 May 2016
		IL 203311 A	31 July 2014
CN 105243295 A	13 January 2016	CN 103403182 A	20 November 2013
		US 2013040824 A1	14 February 2013
		CN 108899091 A	27 November 2018
		KR 20180095941 A	28 August 2018
		US 9965585 B2	08 May 2018
		DK 2646579 T3	07 August 2017
		JP 2018019696 A	08 February 2018
		EP 2646579 B1	14 June 2017
		HK 1190437 A1	15 September 2016
		IL 226272 D0	31 July 2013
		WO 2012071621 A1	07 June 2012
		MX 2013006075 A	02 July 2013
		KR 101891847 B1	24 August 2018
		EA 201300649 A1	30 October 2013
		SG 10201509766Y A	30 December 2015
		AU 2011335891 A1	02 May 2013
		MX 349568 B	03 August 2017
		KR 20140091650 A	22 July 2014
		US 2017220736 A1	03 August 2017
		TW 201307570 A	16 February 2013
		TW I532843 B	11 May 2016
		IL 226272 A	29 February 2016
		EP 2646579 A1	09 October 2013
		NZ 611599 A	29 May 2015
		AU 2011335891 B2	26 March 2015
		JP 6196157 B2	20 September 2017
		SG 190344 A1	28 June 2013
		JP 2014502155 A	30 January 2014
		ZA 201304625 B	28 September 2016
		IL 250931 D0	30 April 2017
		EP 2646579 A4	04 June 2014
		CN 105243295 B	17 August 2018
		US 2014227699 A1	14 August 2014
		BR 112013013421 A2	06 September 2016
		TW 201615841 A	01 May 2016
		US 8741811 B2	03 June 2014
		CA 2817370 A1	07 June 2012
		HK 1217553 A1	13 January 2017
		EP 3243910 A1	15 November 2017
		IL 243823 D0	21 April 2016
		CN 103403182 B	25 November 2015

Form PCT/ISA/210 (patent family annex) (January 2015)

フロントページの続き

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

（特許庁注：以下のものは登録商標）
１．ＪＡＶＡ
２．Ｐｙｔｈｏｎ

(74)代理人　100123582
弁理士　三橋　真二
(74)代理人　100117019
弁理士　渡辺　陽一
(74)代理人　100141977
弁理士　中島　勝
(74)代理人　100150810
弁理士　武居　良太郎
(74)代理人　100138210
弁理士　池田　達則
(72)発明者　ロー　ユクーミン　デニス
中華人民共和国，ホンコン，ニュー　テリトリーズ，シャーティン　ピチウ　ビルディング，シー／オー　テクノロジー　ライセンシング　オフィス　ルーム　３２８
(72)発明者　チウ　ロッサ　ワイ　クン
中華人民共和国，ホンコン，ニュー　テリトリーズ，シャーティン　ピチウ　ビルディング，シー／オー　テクノロジー　ライセンシング　オフィス　ルーム　３２８
(72)発明者　チャン　クワン　チー
中華人民共和国，ホンコン，ニュー　テリトリーズ，シャーティン　ピチウ　ビルディング，シー／オー　ナレッジ　トランスファー　オフィス　ルーム　３０１
(72)発明者　チアン　ペイヨン
中華人民共和国，ホンコン，ニュー　テリトリーズ，シャーティン　ピチウ　ビルディング，シー／オー　ナレッジ　トランスファー　オフィス　ルーム　３０１

Ｆターム(参考)　4B063 QA19 QQ02 QQ28 QQ42 QQ62 QR08 QR32 QR42 QR56 QR62
QR66 QR72 QS03 QS24 QS25 QS34 QS36 QX02