

(19) 日本国特許庁 (JP)

(12) 特 許 公 報 (B2)

(11) 特許番号

特許第6954899号

(P6954899)

(45) 発行日 令和3年10月27日 (2021. 10. 27)

(24) 登録日 令和3年10月4日 (2021. 10. 4)

(51) Int. Cl.

F I

C 1 2 Q 1/68 (2018. 01)

G O 1 N 33/50 (2006. 01)

C 1 2 Q 1/6869 (2018. 01)

C 1 2 Q 1/6813 (2018. 01)

C 1 2 N 15/10 (2006. 01)

C 1 2 Q 1/68

G O 1 N 33/50 P

C 1 2 Q 1/6869 Z

C 1 2 Q 1/6813 Z

C 1 2 N 15/10 Z

請求項の数 26 (全 45 頁)

(21) 出願番号 特願2018-524457 (P2018-524457)
 (86) (22) 出願日 平成28年12月2日 (2016. 12. 2)
 (65) 公表番号 特表2018-537086 (P2018-537086A)
 (43) 公表日 平成30年12月20日 (2018. 12. 20)
 (86) 国際出願番号 PCT/US2016/064611
 (87) 国際公開番号 WO2017/096158
 (87) 国際公開日 平成29年6月8日 (2017. 6. 8)
 審査請求日 令和1年12月2日 (2019. 12. 2)
 (31) 優先権主張番号 62/263, 532
 (32) 優先日 平成27年12月4日 (2015. 12. 4)
 (33) 優先権主張国・地域又は機関
 米国 (US)

(73) 特許権者 516380360
 1 O エックス ゲノミクス, インコーポレ
 イテッド
 アメリカ合衆国, カリフォルニア州 94
 566, プレザントン, コール センター
 パークウェイ 7068, スイート 4
 O 1
 (74) 代理人 100079108
 弁理士 稲葉 良幸
 (74) 代理人 100109346
 弁理士 大貫 敏史
 (74) 代理人 100117189
 弁理士 江口 昭彦
 (74) 代理人 100134120
 弁理士 内藤 和彦

最終頁に続く

(54) 【発明の名称】 核酸の解析のための方法及び組成物

(57) 【特許請求の範囲】

【請求項 1】

試料中の複数の核酸を解析する方法であって、前記方法は、

(a) 複数の核酸を含む試料を提供することと、
 (b) 前記試料の異なる領域が異なる地理的タグ又は地理的タグの異なる集団を受け取るように前記試料に複数の地理的タグを適用することと、
 (c) 前記試料を個別区画に区画することであって、各々の区画が前記複数の核酸からの核酸の部分及び前記複数の地理的タグからの地理的タグの部分を含む、ことと、
 (d) 各個別区画における前記核酸から配列情報を取得することと、
 (e) 各個別区画における前記地理的タグの特徴を同定することであって、各々の区画に
 対し、前記地理的タグの特徴が前記試料中の前記核酸のゲノム遺伝子座間の染色体間の相
 互作用に関する情報を提供する、ことと

を含み、

それによって前記複数の核酸を解析する、方法。

【請求項 2】

前記取得すること (d) に先立って、各個別区画において複数のタグ付けされた核酸断片を生成する、請求項 1 に記載の方法。

【請求項 3】

各個別区画における前記タグ付けされた核酸断片の各々は、前記区画内の前記核酸の部分の断片のコピー、及び区画特異的タグを含む、請求項 2 に記載の方法。

【請求項 4】

前記区画特異的タグは、オリゴヌクレオチドバーコードである、請求項 3 に記載の方法。

【請求項 5】

一つの個別区画内の前記タグ付けされた核酸断片の各々は、同じ区画特異的タグを含む、請求項 2 に記載の方法。

【請求項 6】

前記 (d) において取得する配列情報は、各区画における前記複数のタグ付けされた核酸断片をシーケンシングすることにより取得する、請求項 2 に記載の方法。

【請求項 7】

各区画はタグ付けされた粒子をさらに含み、該タグ付けされた粒子は、遊離可能に粒子に結合された複数の区画特異的タグを含む、請求項 1 に記載の方法。

【請求項 8】

前記区画特異的タグの各々は、オリゴヌクレオチドバーコードタグ及び無作為 N 量体オリゴヌクレオチドを含む、請求項 7 に記載の方法。

【請求項 9】

一つの個別区画内の区画特異的タグは同一であり、個別区画の間の区画特異的タグは異なる、請求項 7 に記載の方法。

【請求項 10】

各個別区画における前記区画特異的タグは、同じオリゴヌクレオチドバーコード及び異なる無作為 N 量体オリゴヌクレオチドを含む、請求項 8 に記載の方法。

【請求項 11】

前記粒子はビーズである、請求項 7 に記載の方法。

【請求項 12】

前記 (b) において、前記試料の異なる領域が地理的タグの異なる集団を受け取るように、前記地理的タグを前記試料に適用する、請求項 1 に記載の方法。

【請求項 13】

前記 (e) において同定する前記特徴は、各個別区画におけるタグの集団である、請求項 12 に記載の方法。

【請求項 14】

前記 (b) において、前記試料の異なる領域が異なる地理的タグを受け取るように、前記地理的タグを前記試料に適用する、請求項 1 に記載の方法。

【請求項 15】

前記 (e) において同定する前記特徴は、各個別区画における地理的タグの配列である、請求項 14 に記載の方法。

【請求項 16】

前記地理的タグは、オリゴヌクレオチドバーコードを含む、請求項 1 に記載の方法。

【請求項 17】

前記試料は、ホルマリン固定パラフィン試料である、請求項 1 に記載の方法。

【請求項 18】

前記取得すること (d) に先立って、前記試料を脱パラフィンする、請求項 17 に記載の方法。

【請求項 19】

各個別区画における核酸の量は、 10 ng/ml 未満である、請求項 1 に記載の方法。

【請求項 20】

各個別区画における核酸の量は、 1 ng/ml 未満である、請求項 1 に記載の方法。

【請求項 21】

前記試料は腫瘍試料である、請求項 1 に記載の方法。

【請求項 22】

前記地理的タグは粒子に結合されている、請求項 1 に記載の方法。

10

20

30

40

50

【請求項 2 3】

前記取得すること (d) は、ショートリード長シーケンシング反応及びロングリード長シーケンシング反応からなる群から選択されたシーケンシング反応を含む、請求項 1 に記載の方法。

【請求項 2 4】

前記シーケンシング反応はショートリードで高精度なシーケンシング反応である、請求項 2 3 に記載の方法。

【請求項 2 5】

前記核酸は RNA を含む、請求項 1 に記載の方法。

【請求項 2 6】

前記ビーズはゲルビーズである、請求項 1 1 に記載の方法。

【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、2015 年 12 月 4 日に提出された米国仮特許出願第 62 / 263, 532 号の利益を主張し、あらゆる目的のためにその全体が参照によって本明細書に組み入れられる。

【背景技術】**【0002】**

ポリヌクレオチドの配列決定は、腫瘍の遺伝的スクリーニングや遺伝子型決定のような医学的応用で用途が増え続けている。多数のポリヌクレオチド配列決定法は、ポリヌクレオチドの無作為断片化を含む、原試料の試料処理法に頼っている。これらの処理法は、処理能力及び効率という点で利点を提供するが、これらの処理された試料から得られる結果的な配列情報は、それらの配列を含有していた元の核酸分子のさらに広い線形 (二次元) 配列内での特定の配列の位置という点で重要なコンテキスト情報を欠き得る。原試料の三次元空間内での構造コンテキストも多数の試料の処理法及び配列決定法によって失われる。従って、特定される核酸配列の構造コンテキスト及び分子コンテキストを保持している配列決定法に対するニーズがある。

【発明の概要】**【0003】**

従って、本発明は、元の核酸分子の分子コンテキスト及び構造コンテキストを保持する配列情報を提供するための方法、システム及び組成物を提供する。

【0004】

一部の態様では、本開示は構造コンテキストを維持しながら核酸を解析する方法を提供する。そのような方法は、(a) 核酸を含有する試料を提供し、核酸が三次元構造を含むステップと、(b) 核酸三次元構造の一部も個別区画に分離されるように試料の一部を個別区画に分離するステップと、(c) 核酸から配列情報を入手し、それによって構造コンテキストを維持しながら核酸を解析するステップとを含む。

【0005】

一部の実施形態では、得られるステップ (c) からの配列情報には互いに空間的に近接し合う核酸の特定が含まれる。

【0006】

さらなる実施形態では、得られるステップ (c) からの配列情報には互いに空間的に近接し合う核酸の特定が含まれる。

【0007】

別のさらなる実施形態では、入手するステップ (c) はゲノムの遺伝子座間の染色体内の及び / または染色体間の相互作用に関する情報を提供する。

【0008】

別のさらなる実施形態では、入手するステップ (c) は染色体の高次構造に関する情報

10

20

30

40

50

を提供する。

【0009】

さらなる実施形態では、分離するステップ(b)に先立って、三次元構造の少なくとも一部が、三次元構造の範囲内で互いに近接し合う核酸の異なる部分を連結するように処理される。

【0010】

任意の実施形態では、核酸は分離するステップ(b)に先立って、試料から単離されない。

【0011】

任意の実施形態では、入手するステップ(c)に先立って、個別区画内での核酸にバーコードを付けて、複数のバーコード付き断片を形成し、その際、所与の個別区画内の断片はそれぞれ共通のバーコードを含むので、バーコードが所与の区画に由来する核酸を特定する。

10

【0012】

さらなる実施形態では、入手するステップ(c)は短い読み取り長さの配列決定反応及び長い読み取り長さの配列決定反応から成る群から選択される配列決定反応を含む。

【0013】

一部の態様では、本開示は、(a)空間的に隣接する核酸セグメントが連結されるように試料内で連結された核酸を形成するステップと、(b)連結された核酸を処理して複数のライゲーション産物を作り出し、その際、ライゲーション産物は空間的に隣接する核酸セグメントの一部を含有するステップと、(c)複数のライゲーション産物を個別区画に入れるステップと、(d)個別区画内のライゲーション産物にバーコードを付け、複数のバーコード付き断片を形成し、その際、所与の個別区画内の断片はそれぞれ共通のバーコードを含み、それによって各断片をそれが由来する連結された核酸に関連付けるステップと、(e)複数のバーコード付き断片から配列情報を入手し、それによって構造コンテキストを維持しながら試料に由来する核酸を解析するステップとを含む、構造コンテキストを維持しながら核酸を解析する方法を提供する。

20

【0014】

一部の態様では、本開示は、(a)空間的に隣接する核酸セグメントが連結されるように試料内で連結された核酸を形成するステップと、(b)連結された核酸を個別区画に入れるステップと、(c)連結された核酸を処理して複数のライゲーション産物を作り出し、その際、ライゲーション産物は空間的に隣接する核酸セグメントの一部を含有するステップと、(d)個別区画内のライゲーション産物にバーコードを付け、複数のバーコード付き断片を形成し、その際、所与の個別区画内の断片はそれぞれ共通のバーコードを含み、それによって各断片をそれが由来する連結された核酸に関連付けるステップと、(e)複数のバーコード付き断片から配列情報を入手し、それによって構造コンテキストを維持しながら試料に由来する核酸を解析するステップとを含む、構造コンテキストを維持しながら核酸を解析する方法を提供する。

30

【0015】

一部の態様では、本開示は、(a)試料内で核酸を架橋し、架橋された核酸を形成し、その際、架橋は空間的に隣接する核酸セグメント間で共有結合を形成するステップと、(b)架橋された核酸を個別区画に入れるステップと、(c)架橋された核酸を処理して複数のライゲーション産物を作り出し、その際、ライゲーション産物は空間的に隣接する核酸セグメントの一部を含有するステップと、(d)複数のライゲーション産物から配列情報を入手し、それによって構造コンテキストを維持しながら試料に由来する核酸を解析するステップとを含む、構造コンテキストを維持しながら核酸を解析する方法を提供する。

40

【0016】

任意の実施形態では、試料はホルマリン固定したパラフィン試料である。

【0017】

任意の実施形態では、個別区画はビーズを含む。さらなる実施形態では、ビーズはゲル

50

ビーズである。

【0018】

任意の実施形態では、試料は腫瘍試料を含む。

【0019】

任意の実施形態では、試料は腫瘍細胞と正常細胞との混合物を含む。

【0020】

任意の実施形態では、試料は核基質を含む。

【0021】

任意の実施形態では、核酸はRNAを含む。

【0022】

任意の実施形態では、試料における核酸の量は、5、10、15、20、25、30、35、40、45または50 ng/ml未満である。

【0023】

一部の態様では、本発明は構造コンテキストを維持しながら核酸を解析する方法を提供し、その際、該方法は、(a)核酸を含有する試料を提供するステップと、(b)試料の異なる地理的領域が異なるタグまたはタグの異なる集団を受け取るように試料にタグのライブラリを適用するステップと、(c)タグのライブラリの一部と核酸の一部も個別区画に分離されるように試料の一部を個別区画に分離するステップと、(d)核酸から配列情報を入手するステップと、(e)個別区画におけるタグまたはタグの集団を特定し、それによって構造コンテキストを維持しながら核酸を解析するステップとを含む。

【図面の簡単な説明】

【0024】

【図1】本明細書に記載されている方法による分子コンテキスト及び構造コンテキストの説明図である。

【図2】本明細書に記載されている過程の説明図である。

【図3】本明細書で開示されている方法及び組成物を用いて配列情報を検出するアッセイを実施するための典型的な作業の流れを説明する。

【図4】核酸試料をビーズと混ぜ合わせ、核酸とビーズを個別の液滴に区分する方法の模式図である。

【図5】染色体の核酸断片にバーコードを付け、増幅する方法の模式図である。

【図6】配列データを元の供給源核酸分子に帰属させることにおける核酸断片のバーコード付けの使用の模式図である。

【図7】例示的な試料調製の方法の模式図である。

【発明を実施するための形態】

【0025】

本発明の実践は、特に指示されない限り、有機化学、ポリマー技術、分子生物学（組換え技術を含む）、細胞生物学、生化学、及び免疫学の従来技法及び記載を採用してもよく、それらは当該技術分野の技量の範囲内にある。そのような従来技法には、ポリマーアレイの合成、ハイブリッド形成、ライゲーション、ファージディスプレイ、及び標識を用いたハイブリッド形成の検出が挙げられる。好適な技法の具体的な説明は本明細書で以下の実施例を参照して手に入れることができる。しかしながら、他の同等の従来手順も当然使用することができる。そのような従来技法及び記載は、たとえば、Genome

Analysis: A Laboratory Manual Series (Vols I-IV), Using Antibodies: A Laboratory Manual, Cells: A Laboratory Manual, PCR Primer: A Laboratory Manual, and Molecular Cloning: A Laboratory Manual (all from Cold Spring Harbor Laboratory Press), Stryer, L. (1995), Biochemistry, (4th, Ed.), Freeman, New York, Gait, "Oligonucleotide Synthesis: A

10

20

30

40

50

Practical Approach", 1984, IRL Press, London, Nelson and Cox, (2000), Lehninger, Principles of Biochemistry, 3rd Ed., W.H. Freeman Pub., New York, N.Y. 及び Berg, et al. (2002), Biochemistry, 5th Ed., W.H. Freeman Pub., New York, N.Y., のような標準の実験室マニュアルに見いだすことができ、それらはすべてあらゆる目的でその全体が参照によって本明細書に組み入れられる。

【0026】

本明細書及び添付のクレームで使用されるとき、文脈が明瞭に指示しない限り、単数形態「a」、「an」及び「the」は複数の指示対象を含むことに留意されたい。従って、たとえば、「ポリメラーゼ」への言及は1つの作用物質またはそのような作用物質の混合物を指し、「その方法」への言及は、当業者に既知の同等のステップ及び方法等への言及を含む。

10

【0027】

定義されない限り、本明細書で使用される専門用語及び科学用語はすべて、本発明が属する技術分野の当業者によって一般に理解されるのと同じ意味を有する。本明細書で言及される出版物はすべて、その出版物に記載され、現在記載されている本発明と併せて使用されてもよい装置、組成物、製剤及び方法を記載し、開示する目的で参照によって本明細書に組み入れられる。

【0028】

20

値の範囲が提供される場合、文脈が明瞭に指示しない限り下限の単位の10分の1までの、その範囲及び他の言及される範囲の上限と下限の間のそれぞれ介在する値、またはその言及される範囲における介在する値が本発明の範囲内に包含されることが理解される。これらのさらに小さい範囲における上限及び下限は独立してそのさらに小さい範囲に含まれてもよく、本発明の範囲内にも包含され、言及される範囲における具体的に除外される端の対象となる。言及される範囲が一方又は双方の端を含む場合、それらの包含される端の双方を除外する範囲も本発明に含まれる。

【0029】

以下の記載では、多数の具体的な詳細が述べられて本発明のさらに十分な理解が提供される。しかしながら、本発明はこれらの具体的な詳細の1以上がない状態で実践されてもよいことが当業者に明らかであろう。他の例では、本発明を曖昧にするのを回避するために、当業者に周知である周知の特徴及び手順は記載されていない。

30

【0030】

本明細書で使用されるとき、用語「comprising（含む）」は組成物及び方法が引用された要素を含むが、他を除外しないことを意味するように意図される。組成物及び方法を定義するのに使用される場合、「consisting essentially of（から本質的に成る）」は組成物及び方法にとっての必須の重要性のうち、他の要素を除外することを意味するものとする。「consisting of（から成る）」は請求される組成物及び実質的な方法ステップについての他の成分のうち、微量を超える要素を除外することを意味するものとする。これらの移行用語のそれぞれによって定義される実施形態は本発明の範囲の範囲内である。従って、方法及び組成物は、追加のステップ及び成分を含む（comprising）ことができるか、あるいは重要な意味を持たないステップ及び組成物から本質的に成る（consisting essentially of）か、あるいは言及されている方法ステップまたは組成物のみを意図する（consisting of）。

40

【0031】

範囲を含む数字表示、たとえば、pH、温度、時間、濃度及び分子量はすべて、0.1の増分による（+）または（-）で変化する近似値である。常に明白に述べられるとは限らないが、数字表示にはすべて用語「約」が先行することが理解されるべきである。用語「約」には、たとえば、「X + 0.1」または「X - 0.1」のような「X」の軽微な増

50

分に加えて正確な値「X」も含まれる。常に明白に述べられるとは限らないが、本明細書に記載されている試薬は単に例示であって、そのようなものの同等物は当該技術分野で既知であることも理解されるべきである。

【0032】

I. 概説

本開示は遺伝物質の特徴付けのための方法、組成物及びシステムを提供する。一般に、本明細書に記載されている方法、組成物及びシステムは、それらが元々試料に存在したときのそれらの成分の構造コンテキストと同様に分子コンテキストに関する情報を保持しながら試料の成分を解析する方法を提供する。本明細書の議論の多くは核酸の解析という点であるが、本明細書で議論されている方法及びシステムは、タンパク質及び他の分子を含む試料の他の成分に適用するように適応させることができることが理解されるであろう。

10

【0033】

デオキシリボ核酸(DNA)は線形分子であり、そのようなものとしてゲノムは線形次元という点で記載され、評価されることが多い。しかしながら、染色体は剛性ではなく、2つのゲノム遺伝子座間の空間距離はゲノムの線形配列に沿ったその距離に常に対応する必要があるわけではない。多数のメガ塩基対によって分離された領域は3次元空間では直接隣接することができる。調節の観点から、ゲノム遺伝子座間の長距離相互作用を理解することが有用であってもよい。たとえば、遺伝子のエンハンサー、サイレンサー及びインシュレーターの要素は膨大なゲノム距離にわたって機能する可能性がある。配列読み取りの構造コンテキスト及び分子コンテキストの双方を保持する能力はそのような長距離相互作用を理解する能力を提供する。

20

【0034】

本明細書で使用されるとき、「構造コンテキストを保持すること」は、複数の配列読み取りまたは配列読み取りの複数の部分が試料内のそれら配列読み取りの元の三次元の相対位置に帰属可能であることを意味する。言い換えれば、配列読み取りは、試料における近隣の核酸(及び一部の状況では関連するタンパク質)に関してその試料の中での相対的な位置に関連付けることができる。それらの近隣の核酸が単一の元の核酸分子の線形配列の範囲内に物理的に位置していないとしても、この空間情報は本明細書で議論されている方法を介して利用することができる。図1における模式図を参照して、試料(101)では、配列(104)及び(105)は2つの異なる元の核酸分子(それぞれ(102)及び(103))の線形配列の範囲内に位置しているが、試料の範囲内で互いに空間的に近接して位置している。本明細書に記載されている方法及び組成物は、配列読み取りの構造コンテキストに関するその情報を保持する能力を提供するので、配列(104)及び(105)からの読み取りを、それらの配列読み取りが由来する元の核酸分子(102)及び(103)における原試料内でのその相対的な空間的近接性に帰属させるのを可能にする。

30

【0035】

本明細書で開示されている方法及び組成物はまた分子コンテキストを保持する配列情報も提供する。「分子コンテキストを保持すること」は本明細書で使用されるとき、複数の配列読み取りまたは配列読み取りの複数の部分が核酸の単一の元の分子に帰属可能であってもよいことを意味する。核酸のこの単一分子が種々の長さのいずれかであってもよい一方で、好ましい態様では、それは相対的に長い分子であり、長距離分子コンテキストの維持を可能にする。特に、単一の元の分子は好ましくは、典型的な短い読み取り配列の長さより実質的に長い、たとえば、200塩基より長い、少なくとも1000塩基以上長い、5000塩基以上長い、10,000塩基以上長い、20,000塩基以上長い、30,000塩基以上長い、40,000塩基以上長い、50,000塩基以上長い、60,000塩基以上長い、70,000塩基以上長い、80,000塩基以上長い、90,000塩基以上長い、または100,000塩基以上長いことが多く、場合によっては、1メガ塩基以上まで長い。

40

【0036】

一般に、本明細書に記載されている方法は構造コンテキスト及び分子コンテキストを維

50

持しながら核酸を解析することを含む。そのような解析には核酸を含有する試料を提供する方法が含まれ、核酸は三次元構造を含有する。核酸の三次元構造の一部も個別区画に分離されるように試料の一部が個別区画に分離され、互いに空間的に近接し合う核酸配列は同じ区画に分離される傾向があるので、後で得られた配列読み取りが元々同じ個体の元の核酸分子上になかった配列に由来するとしてもその空間的近接性の三次元情報を保持する。再び図1を参照して、核酸分子102及び103及び106を含有する試料101が、試料のサブセットが異なる個別区画に割り当てられるように個別区画に分離される場合、核酸分子102と103と106の間の物理的距離のため、核酸分子106よりも核酸分子102及び103が同じ区画に入る可能性が高い。従って、同じ個別区画内の核酸分子は原試料にて互いに空間的に近接し合っていたものである。従って、個別区画内の核酸から得られる配列情報は、たとえば、核酸の配列決定を介して核酸を解析する方法を提供し、それらの配列読み取りを元の核酸分子の構造コンテキストに戻って帰属させる。

10

【0037】

さらなる例では、構造コンテキスト(「地理的コンテキスト」とも呼ばれる)はタグ(たとえば、バーコードのオリゴヌクレオチド)を使用することによって維持されて試料の地理をコードしてもよい。一部の状況では、これには、バーコード付きの配列(たとえば、mRNA配列)の収集をコードするウイルスライブラリを試料に注入することが含まれ得る。バーコードは能動的過程によってまたは拡散によって試料を移動する。次いで本明細書に記載されている及び当該技術分野で既知である方法に従って試料がさらに処理されると、バーコードは構造上の位置に相関して試料内での同じ地理的位置から核酸配列を特定することができる。バーコードが能動的過程を介して試料に分配される例では、同じバーコードを持つ配列は地理的に関連付けられてもよく、及び/または同じ過程を介して関連付けられてもよい。理解されるように、構造コンテキストをコードするタグを使用するこのシステムは単独で使用することができ、または構造コンテキスト及び分子コンテキストをさらに保持するために個別区画を利用する本明細書に記載されている方法と組み合わせて使用することができる。空間的位置をコードするタグと同じ個別区画に分離される分子を特定するためのバーコードとを使用する例では、試料は本質的にタグ付けされる、または一方のセットのバーコードが空間的位置を特定するのに使用され、もう一方のセットのバーコードが区画専用である「二重バーコード付け」が行われる。そのような例では、バーコードの双方のセットを用いて、試料から生成される配列読み取りの構造コンテキスト及び分子コンテキストを保持する情報を提供することができる。

20

30

【0038】

一部の例では、核酸から得られる配列情報はゲノム遺伝子座間の染色体内及び/または染色体間の相互作用に関する情報を提供する。さらなる例では、配列情報は染色体の高次構造に関する情報を含む。

【0039】

さらなる例では、個別区画への分離に先立って、その三次元構造内で互いに近接し合う配列の領域が互いに結合されるように試料における核酸を処理してそれらの三次元構造の異なる領域を連結してもよい。従って、試料の個別区画への分離はそれらの連結した領域を同じ区画に分離し、それによってそれら核酸に由来する配列読み取りの構造コンテキストが保持されることをさらに保証する。

40

【0040】

一部の状況では、核酸の連結は、空間的に近接する分子を架橋するのに使用される当該技術分野で既知の方法を用いて達成されてもよい。そのような架橋剤には、アルキル化剤、シスプラチン、亜酸化窒素、ブソラレン、アルデヒド、アクロレイン、グリオキサール、四酸化オスミウム、カルボジイミド、塩化第二水銀、亜鉛塩、ピクリン酸、重クロム酸カリウム、エタノール、メタノール、アセトン、酢酸等が挙げられるが、それらに限定されない。具体的な例では、あらゆる目的で及び特に核酸分子を連結することに関する教示すべてについてその全体が参照によって本明細書に組み入れられる、たとえば、Decker, et al., "Capturing chromosome conforma

50

tion", Science, 295:1306-1311, (2002), 及び Berkum, et al., J. Vis. Exp. (39), e1869, doi:10.3791/1869 (2010) に記載される「Hi-C」プロトコルのようなゲノムの三次元構造の解析のために考案されたプロトコルを用いて、核酸が連結される。そのようなプロトコルには一般に、空間的に密に近接するゲノム遺伝子座が連結されるように試料を架橋することによって分子のライブラリを作り出すことが関与する。さらなる実施形態では、架橋間の介在DNAループが消化で取り去られ、次いで配列内領域がライブラリへの付加のために逆架橋される。消化ステップと逆架橋ステップは、試料を個別区画に区画化するステップに先立って存在してもよく、または分離するステップの後の区画の中に存在してもよい。

10

【0041】

別のさらなる例では、核酸は区画内での核酸すべてに共通のバーコードを提供するタグ付けステップまたはバーコード付けステップを経てもよい。理解されるように、このバーコード付けは上述された核酸の連結/架橋のステップの有無にかかわらず存在してもよい。本明細書で開示されているバーコード付け技法の使用は、ゲノム領域についての個々の構造コンテキスト及び分子コンテキストを提供する独特の能力を付与し、すなわち、特定の配列読み取りを個々の試料の核酸分子に帰属させることによって、及び変異による協調組み立てを介して、複数の試料核酸分子の間で、及び/または特定の染色体に対してさら広範囲のまたはさらに長距離の推論されるコンテキストを提供する。用語「ゲノム領域」または「領域」は本明細書で使用されるとき、ゲノム及び/または染色体の任意の定義された長さを指す。たとえば、ゲノム領域は1を超える染色体間の関連性（すなわち、たとえば、相互作用）を指してもよい。ゲノム領域はまた完全な染色体または部分的な染色体も包含することもできる。加えて、ゲノム領域は、染色体上（すなわち、たとえば、オープンリーディングフレーム及び/または調節遺伝子）または遺伝子間の非コーディング領域における具体的な核酸配列を含むことができる。

20

【0042】

バーコード付けの使用は、たとえば、血流にて循環している腫瘍DNAの検出及び特徴付けのための、試料から抽出された全核酸集団の少数構成成分と主要構成成分との間を区別する能力を高める追加の利点を付与し、且つ、任意の増幅ステップの間での増幅の偏りを減らすまたは排除する。加えて、マイクロ流体工学の形式の実施は、DNAの極端に少ない試料容量及び少ない入力量で作業する能力と同様に多数の試料区画（液滴）を迅速に処理してゲノム規模でのタグ付けを促す能力を付与する。

30

【0043】

ゲノムの全体の領域または選択した領域から配列情報を得る能力を提供することに加えて、本明細書に記載されている方法及びシステムはまた、US 2014/0316,383; 14/316,398; 14/316,416; 14/316,431; 14/316,447; 及び14/316,463に記載されるような、ハプロタイプの位相化、構造上の変動及びコピー数の変動の特定を含むがこれらに限定しない、ゲノム物質の他の特徴付けも提供することができ、あらゆる目的のために、特にゲノム物質の特徴付けに関して記述されたすべての説明、図面及び作業例について参照によって本明細書に組み入れられる。

40

【0044】

一般に、本発明の方法は、図2で説明されているようなステップを含み、本明細書でさらに詳細に議論されている本発明の方法の図式的概観を提供する。理解されるように、図2で要点が述べられている方法は、必要に応じて及び本明細書に記載されているように変更されてもよく、または改変されてもよい例示的な実施形態である。図2に示すように、本明細書に記載されている方法は、試料の核酸が互いに空間的に近接して核酸を連結するように処理される任意のステップ201を含んでもよい。その予備的な処理ステップ(201)の有無にかかわらず、本明細書に記載されている方法はほとんどの実施例で核酸を含有する試料が区画化されるステップ(202)を含むであろう。一般に、対象とするゲ

50

ノム領域に由来する核酸を含有する各区画はバーコードを含有する断片を生じる処理(203)を経るであろう。次いでそれらの断片は配列決定(205)に先立ってプールされる(204)。(205)からの配列読み取りは、一般に区画専用のバーコード(203)のために、元の構造コンテキスト及び分子コンテキスト(206)に帰属させることができる。各区画は一部の例では1を超える核酸を含んでもよく、一部の例では、数百の核酸分子を含有するであろう。ステップ203のバーコード付き断片は当該技術分野で既知の方法を用いて生成することができ、一部の例では、オリゴヌクレオチドが個別区画内に試料と共に含まれる。そのようなオリゴヌクレオチドは、試料の多数の異なる領域を無作為に刺激を与えるように意図される無作為配列を含んでもよく、またはそれらは試料の標的とされる領域の上流に刺激を与えるように標的化された具体的なプライマー配列を含んでもよい。さらなる例では、これらのオリゴヌクレオチドはバーコード配列も含有するため、複製過程は原試料核酸の得られる複製された断片にもバーコードを付ける。試料を増幅し、バーコード付けすることにおけるこれらのバーコードオリゴヌクレオチドの使用についての特に洗練された方法は、US 2014/0316,383; 14/316,398; 14/316,416; 14/316,431; 14/316,447; 及び14/316,463にて詳細に記載され、あらゆる目的で及び特にオリゴヌクレオチドにバーコードを付け、増幅することに関する教示すべてについてその全体が参照によって本明細書に組み入れられる。区画にも含有される伸長反応試薬、たとえば、DNAポリメラーゼ、ヌクレオシド三リン酸、補因子(たとえば、 Mg^{2+} または Mn^{2+} 等)が次いで、試料を鋳型として用いてプライマー配列を伸長し、プライマーがアニーリングされる鋳型の鎖に対して相補性の断片を作り出し、相補性の断片はオリゴヌクレオチドとその関連するバーコード配列を含む。試料の異なる部分に対する複数のプライマーのアニーリング及び伸長は試料の重複する相補性断片の大きなプールを生じることができ、それぞれがそれが作り出された区画を示すそれ自体のバーコード配列を持っている。場合によっては、これらの相補性断片は区画に存在するオリゴヌクレオチドによって刺激を与えられる鋳型としてそれ自体使用されて再びバーコード配列を含む相補体の相補体を作り出してよい。さらなる例では、この複製過程は、第1の相補体が複製される場合、それはその末端にてまたはその近傍で2つの相補性配列を作り、ヘアピン構造または部分的なヘアピン構造の形成を可能にするように構成され、それはさらなる反復コピーを生じる基礎である分子の能力を低減する。本明細書に記載されている方法及びシステムの利点は、区画専用のまたは試料専用のバーコードをコピーされた断片に結合することが配列決定された断片の元の分子コンテキストを維持し、それらをその元の区画、従ってその元の試料の核酸分子に帰属させることである。

【0045】

試料は、区画化ステップに先立ってビーズに遊離可能に結合されるオリゴヌクレオチドタグのセットと組み合わせられることが多い。核酸にバーコードを付ける方法は当該技術分野で知られており、本明細書に記載されている。一部の例では、方法は、Amini, et al, 2014, Nature Genetics, Advance Online Publication)に記載されているように利用され、あらゆる目的で及び特に核酸にバーコードまたは他のオリゴヌクレオチドタグを結合することに関連する教示すべてについてその全体が参照によって本明細書に組み入れられる。本出願に記載されている方法及びシステムに従って核酸を処理し、配列決定する方法はUS 2014/0316,383; 14/316,398; 14/316,416; 14/316,431; 14/316,447; 及び14/316,463にもさらに詳細に記載され、あらゆる目的で及び特に核酸を処理すること及び配列決定すること及びゲノム物質の他の特徴付けに関して記述された説明、図面及び作業例すべてについてその全体が参照によって本明細書に組み入れられる。

【0046】

上記の作業の流れに加えて、チップに基づく及び溶液に基づく捕捉方法の双方を含む方法を用いて、さらなる解析、特に配列決定のために、標的とされるゲノム領域は濃縮され

10

20

30

40

50

、単離されまたは分離されてもよく、すなわち、「壊されて」もよい。そのような方法は、対象とするゲノム領域または対象とするゲノム領域の近傍もしくは隣接する領域に対して相補性であるプローブを利用する。たとえば、ハイブリッド（またはチップに基づく）捕捉では、対象とする領域を総合して網羅する配列を持つ捕捉プローブ（普通一本鎖オリゴヌクレオチド）を含有するマイクロアレイが表面に固定される。ゲノムDNAは断片化され、さらに、末端修復のような処理を受けて平滑末端及び／または、たとえば、汎用プライミング配列のような追加の特徴の付加を作り出してもよい。これらの断片はマイクロアレイ上のプローブとハイブリッド形成する。ハイブリッド形成しなかった断片は洗い流され、所望の断片が溶出され、またはさもなければ、配列決定もしくは他の解析のために表面で処理されるので、表面上に残っている断片の集団は対象とする標的とされる領域を含有する断片について濃縮される（たとえば、捕捉プローブに含有されるものに相補性である配列を含む領域）。濃縮された断片の集団は、当該技術分野で既知の増幅技法を用いてさらに増幅されてもよい。そのような標的とされるプルダウン濃縮法のための例示的な方法は、2015年10月29日に出願されたUS N 14 / 927, 297に記載され、あらゆる目的で及び特に記述された説明、図面及び実施例を含む標的とされるプルダウン濃縮法及び配列決定法に関する教示すべてについてその全体が参照によって本明細書に組み入れられる。標的とされるゲノム領域の集団はさらに、これらの標的とされる領域の適用範囲を増やす方法を用いて、上述のプルダウン法に先立って濃縮されてもよい。そのような適用範囲の増加は、たとえば、たとえば2015年2月24日に出願されたUS N 62 / 119, 996に記載されているものを含む、標的とされる増幅法を用いて達成されてもよく、あらゆる目的で及び特に核酸分子の標的とされる適用範囲に関する教示すべてについて参照によって本明細書に組み入れられる。

【0047】

具体的な例では、本明細書に記載されている方法は、ゲノムの選択された領域が配列決定に先立って増幅されるステップを含む。一般に当該技術分野で既知の方法（限定しないでPCR増幅を含む）を用いて実施されるこの増幅は、ゲノムの選択された領域の少なくとも1X、10X、20X、50X、100X、200X、500X、1000X、1500X、2000X、5000X、または10000Xの適用範囲を提供し、それによって、それらの選択された領域のデノボ配列決定を可能にする核酸の量を提供する。さらなる実施形態では、増幅はゲノムの選択された領域の少なくとも1X～20X、50X～100X、200X～1000X、1500X～5000X、5000X～10,000X、1000X～10000X、1500X～9000X、2000X～8000X、2500X～7000X、3000X～6500X、3500X～6000X、4000X～5500Xの適用範囲を提供する。

【0048】

増幅は一般に、ゲノムの選択された領域の範囲内またはその近傍の配列に相補性のプライマーの伸長を介して実施される。場合によっては、対象とする領域をうめるように設計されるプライマーのライブラリが使用され 言い換えれば、プライマーのライブラリはゲノムの選択された領域に沿った具体的な距離で領域を増幅するように設計される。一部の例では、選択的な増幅は、ゲノムの選択された領域に沿って10、15、20、25、50、100、200、250、500、750、1000または10000塩基ごとに相補性であるプライマーを利用する。別のさらなる例では、プライマーのタイル状ライブラリは距離の混合を捕捉するように設計され その混合は距離の無作為の混合であることができ、または選択された領域の特定の部分もしくは比率が様々なプライマー対によって増幅されるように賢く設計することができる。本明細書に記載されている方法に従って使用するためのゲノムの標的とされる適用範囲のさらなる情報は、たとえば、2015年4月13日に出願されたUS N 62 / 146, 834にて提供され、あらゆる目的で及び特にゲノムの標的とされる適用範囲に関する教示すべてについてその全体が参照によって本明細書に組み入れられる。

【0049】

一般に、本明細書に記載されている方法及びシステムは、たとえば、配列決定のような解析のために核酸を提供する。配列決定の情報は、極端に少ない配列決定の誤差率と短い読み取り配列決定法の高い処理能力の利点を有する方法を用いて得られる。上述されているように、核酸の配列決定は通常、配列読み取りまたは配列読み取りの一部の構造コンテキスト及び分子コンテキストを守る方法で実施される。それによって、複数の配列読み取りまたは配列読み取りの複数の部分が、原試料における他の核酸と比較した空間的位置（構造コンテキスト）に、及び核酸の単一の元の分子の線形配列に沿った配列読み取りの位置（分子コンテキスト）に帰属可能であることを意味する。核酸のこの単一の分子が種々の長さのものであってもよい一方で、好ましい態様では、それは相対的に長い分子であり、長距離分子コンテキストの維持を可能にする。特に、単一の元の分子は好ましくは、典型的な短い読み取り配列の長さより実質的に長く、たとえば、200塩基よりも長く、少なくとも1000塩基以上長い、5000塩基以上長い、10,000塩基以上長い、20,000塩基以上長い、30,000塩基以上長い、40,000塩基以上長い、50,000塩基以上長い、60,000塩基以上長い、70,000塩基以上長い、80,000塩基以上長い、90,000塩基以上長い、または100,000塩基以上長いことが多く、場合によっては、1メガ塩基以上までである。

【0050】

上記で言及したように、本明細書に記載されている方法及びシステムはさらに長い核酸の短い配列読み取りについての個々の分子コンテキストを提供する。本明細書で使用されるとき、個々の分子コンテキストは、配列読み取りそれ自体の範囲内には含まれない、たとえば、隣接するまたは近傍の配列への関係のような、特定の配列読み取りの域を超える配列コンテキストを指し、従って通常、それらは短い配列読み取りには、たとえば、対合した読み取りについて約150塩基または約300塩基の読み取りには全体的にまたは部分的に含まれないようなものである。特に好まれる態様では、方法及びシステムは短い配列読み取りについて長距離の配列コンテキストを提供する。そのような長距離のコンテキストは、1kbより長い、5kbより長い、10kbより長い、15kbより長い、20kbより長い、30kbより長い、40kbより長い、50kbより長い、60kbより長い、70kbより長い、80kbより長い、90kbより長いまたは100kbよりさらに長い、それ以上の互いの距離の範囲内である配列読み取りに対する所与の配列読み取りの関係または連関を含む。理解されるように、長距離の個々の分子コンテキストを提供することによって、その個々の分子コンテキストの範囲内で変異の位相化情報を導き出すこともでき、たとえば、特定の長い分子上の変異は一般に定義によって位相化されるであろう。

【0051】

長距離の個々の分子コンテキストを提供することによって、本発明の方法及びシステムはさらにはるかに長い推論された分子コンテキスト（本明細書では「長い仮想の単一分子読み取り」とも呼ばれる）も提供する。配列コンテキストは、本明細書に記載されているように、完全なゲノム配列の異なる（一般にキロベース規模で）範囲にわたる断片の連鎖をマッピングすることまたは提供することを含むことができる。これらの方法は、短い配列読み取りを個々のさらに長い分子または連結された分子のコンティグにマッピングすることと同様に、たとえば、連続する決定された配列が1kbより長い、5kbより長い、10kbより長い、15kbより長い、20kbより長い、30kbより長い、40kbより長い、50kbより長い、60kbより長い、70kbより長い、80kbより長い、90kbより長いまたはさらに100kbより長い、個々の分子のそのような決定された配列を有するさらに長い個々の分子の大きな部分の長距離の配列決定を含む。配列コンテキスト、短い配列のさらに長い核酸への帰属と同様に、たとえば、双方の個々の長い核酸分子または連結された核酸分子もしくはコンティグの収集には、さらに長い核酸区間に対して短い配列をマッピングして高いレベルの配列コンテキストを提供することと同様にこれらのさらに長い核酸を介して短い配列から組み立てられる配列を提供することの双方が含まれてもよい。

【 0 0 5 2 】

さらに、長い個々の分子に関連する長距離の配列コンテキストを利用してもよい一方で、そのような長距離の配列コンテキストを有することはさらに長距離の配列コンテキストを推論させる。一例として、上述されている長距離の分子コンテキストを提供することによって、異なる元の分子に由来する長い配列の間で重複する変異部分、たとえば、位相化された変異、転座配列等を特定することができ、これらの分子の間の推論される連結を可能にする。そのような推論される連結または分子コンテキストは本明細書では「推論されるコンティグ」と呼ばれる。場合によっては、位相配列の文脈で議論されるとき、たとえば、重複している位相化変異により個々の元の分子よりも実質的に長い長さの位相化コンティグを推論することができる場合、推論されるコンティグは一般に位相化された配列を表してもよい。これらの位相化コンティグは本明細書では「位相ブロック」と呼ばれる。

10

【 0 0 5 3 】

さらに長い単一の分子読み取り（たとえば、上述したような「長い仮想単一分子読み取り」）で開始することによって、短い読み取り配列決定技術または位相化された配列決定への他のアプローチを用いて帰属できることよりも長い推論コンティグまたは位相ブロックを導き出すことができる。たとえば、公開された米国特許出願第 2 0 1 3 - 0 1 5 7 8 7 0 号を参照のこと。特に、本明細書に記載されている方法及びシステムを使用して、少なくとも約 1 0 k b、少なくとも約 2 0 k b、少なくとも約 5 0 k b の N 5 0（言及される N 5 0 の数より大きいブロックの長さの合計がブロックの長さすべての合計の 5 0 % である）を有する推論されるコンティグまたは位相ブロックの長さを得ることができる。さらに好ましい態様では、少なくとも約 1 0 0 k b、少なくとも約 1 5 0 k b、少なくとも約 2 0 0 k b、及び多くの場合、少なくとも約 2 5 0 k b、少なくとも約 3 0 0 k b、少なくとも約 3 5 0 k b、少なくとも約 4 0 0 k b、及び場合によっては、少なくとも約 5 0 0 k b 以上の N 5 0 を有する推論されるコンティグまたは位相ブロックの長さが達成される。さらに他の場合、2 0 0 k b を超える、3 0 0 k b を超える、4 0 0 k b を超える、5 0 0 k b を超える、1 M b を超えるまたはさらに 2 M b を超える位相ブロックの最大長が得られてもよい。

20

【 0 0 5 4 】

1 つの態様では、且つ本明細書で上述及び後述されている方法のいずれかと併せて、本明細書に記載されている方法及びシステムは、試料核酸またはその断片を個別の区分または区画（本明細書では相互交換可能に区画と呼ぶ）に入れるまたは区画化する区分化を提供し、その際、各区画は他の区画の内容物からのそれ自体の内容物の分離を維持する。特徴の、たとえば、核酸の配列情報の特定の区分内に含まれる試料核酸への、特に元々区画に入れられてもよい連続する試料核酸の相対的に長い区間へのその後の帰属を可能にするために、独特の識別子、たとえば、バーコードが、区分化されたまたは区画化された試料核酸を保持する区画に予め、続いて、または付随して送達されてもよい。このその後の帰属は、原試料の三次元内で互いに密接であった核酸が同じ区画に入れられる可能性が高いので、原試料におけるそれら試料核酸の元の構造コンテキストへの帰属をさらに可能にする。従って、配列読み取りの区画（及びそれら区画内に含有される核酸）への帰属は、その配列読み取りが由来した元の核酸分子に沿った線形位置に関して分子コンテキストを提供するだけでなく、原試料の三次元コンテキストで互いに空間的に密に近接し合った核酸からの配列読み取りを特定する構造コンテキストも提供する。

30

40

【 0 0 5 5 】

本明細書に記載されている方法で利用される試料核酸は通常、解析される試料全体、たとえば、染色体全体、エキソームまたは他の大きなゲノム部分の多数の重複している部分を表す。これらの試料核酸はゲノム全体、個々の染色体、エキソーム、アンプリコン、または対象とする種々の異なる核酸のいずれかを含んでもよい。試料核酸は通常、核酸が連続する核酸分子の相対的に長い断片または区間にて区画に存在するように区画化される。通常、試料核酸のこれらの断片は、1 k b より長く、5 k b より長く、1 0 k b より長く、1 5 k b より長く、2 0 k b より長く、3 0 k b より長く、4 0 k b より長く、5 0 k

50

bより長く、60 kbより長く、70 kbより長く、80 kbより長く、90 kbより長くまたはさらに100 kbより長くてもよく、それはさらに長距離の、上述の構造コンテキスト及び分子コンテキストを可能にする。

【0056】

試料核酸はまた通常、それによって所与の区画がゲノム遺伝子座の2つの重複する断片を含む非常に低い可能性を有するレベルで区画化される。これは通常、区画化過程の間で低い入力量及び/または濃度で試料核酸を提供することによって達成される。その結果、好ましい場合では、所与の区画は出発試料核酸の多数の長い重複しない断片を含んでもよい。次いで異なる区画における試料核酸を独特の識別子に関連付け、その際、任意の所与の区画については、それに含有される核酸は同じ独特の識別子を持つが、異なる区画は異なる独特の識別子を含んでもよい。さらに、区画化ステップは試料成分を非常に少ない容量の区画または液滴に割り当てるので、上記で言及されたような所望の割り当てを達成するために、たとえば、試験管またはマルチウェルプレートのウェルにて多い容量の処理で必要とされるような、試料の実質的な希釈を行う必要がない。さらに、本明細書に記載されているシステムはそのような高レベルのバーコード多様性を採用するので、上記で提供されたように、多数のゲノム同等物の間で多様なバーコードを割り当てることができる。特に、前述したように、マルチウェルプレート法(たとえば、米国特許公開出願第2013-0079231号及び第2013-0157870号を参照のこと)は通常、百から数百の異なるバーコード配列を操作するにすぎず、バーコードを異なる細胞/核酸に帰属させることができるためには試料の限界希釈法を採用する。従って、それらは一般に100よりはるかに少ない細胞を操作し、それは通常、1:10及び確実には1:100を軽く超える桁でゲノム:(バーコードの種類)の比を提供することになる。一方、本明細書に記載されているシステムは、ゲノム当たりさらに改善されたバーコードの多様性をさらに提供する一方で、高レベルのバーコード多様性、たとえば、10,000、100,000、500,000等を超える多様なバーコードの種類のため、多数のゲノム(たとえば、アッセイ当たり100ゲノムを超える、アッセイ当たり500ゲノムを超える、アッセイ当たり1000ゲノムを超える、またはそれ以上の桁で)を負荷することも可能にしながら、1:50以下、1:100以下、1:1000以下またはさらに小さい比の桁であるゲノム:(バーコードの種類)の比で操作することができる。

【0057】

さらなる例では、個別区画に分割された試料の一部と共に含まれるオリゴヌクレオチドは少なくとも第1と第2の領域を含んでもよい。第1の領域は、所与の区画内でのオリゴヌクレオチドの間では実質的に同一のバーコード配列であってもよいが、異なる区画の間では、異なるバーコード配列であってもよく、多くの場合異なるバーコード配列であるバーコード領域であってもよい。第2の領域は、区画内の試料の範囲内で核酸に刺激を与えるのに使用することができるN量体(無作為なN量体または特定の配列を標的とするように設計されたN量体)であってもよい。場合によっては、N量体が特定の配列を標的とするように設計される場合、それは特定の染色体(たとえば、第1、13、18または21染色体)または染色体の領域、たとえば、エキソームもしくは他の標的とされる領域を標的とするように設計されてもよい。場合によっては、N量体は、特定の遺伝子または遺伝領域、たとえば、疾患または障害(たとえば、癌)に関連する遺伝子または遺伝領域を標的とするように設計されてもよい。区画の範囲内では、核酸の長さに沿って異なる場所で核酸試料に刺激を与えるのに第2のN量体を用いて、増幅反応を実施してもよい。増幅の結果、各区画は、同一のまたはほぼ同一のバーコードに結合されている、且つ各区画で核酸の重複する、さらに小さい断片を表してもよい核酸の増幅産物を含有してもよい。バーコードは、核酸のセットが同じ区画を起源としたので、核酸の同じ鎖を起源とした可能性もあったことを示すマーカーの役割を果たすことができる。増幅に続いて、核酸はプールされ、配列決定され、配列決定アルゴリズムを用いて並べられてもよい。さらに短い配列読み取りは、その関連するバーコード配列のせいで、試料核酸の単一で長い断片に対して並べられ、それに帰属させられてもよいので、その配列で特定される変異のすべては単一

10

20

30

40

50

の元の断片及び単一の元の染色体に帰属させることができる。さらに、複数の長い断片にわたって複数の同一場所に配置された変異を並べることによって、その染色体の寄与をさらに特徴付けることができる。従って、ゲノム配列の長距離にわたって解析できるにつれて、次いで特定の遺伝的変異の位相化に関する結論、たとえば、ゲノムの十分に特徴付けられていない領域の区間にわたる配列情報の特定が引き出されてもよい。そのような情報は、一般に同一の核酸鎖または異なる核酸鎖に存在する特定されたセットの遺伝的変異であるハプロタイプを特定するのににも有用であってもよい。コピー数の変動もこの方法で特定されてもよい。

【 0 0 5 8 】

記載されている方法及びシステムは、現在の核酸の配列決定技術及びそれに関連する試料の調製法に対して十分な利点を提供する。集団試料の調製法及び配列決定法は、試料における主要構成成分を主として特定し、特徴付ける傾向があり、ゲノムの十分に特徴付けられていない領域もしくは高度に多形の領域に由来する少数の構成成分、たとえば、1本の染色体が寄与する遺伝物質、または抽出された試料にて全DNAの小さな比率を構成する、1またはわずかな細胞に由来する、もしくは血流で循環する断片化した腫瘍細胞のDNA分子に由来する物質を特定し、特徴付けるようには設計されない。本明細書に記載されている方法には、これらの少数構成成分に由来する遺伝物質を増やす選択的な増幅法が含まれ、この遺伝物質の分子コンテキストを保持する能力はさらに、これらの構成成分の遺伝的特徴付けを提供する。記載されている方法及びシステムはまた、さらに大きな試料内に存在する集団を検出する有意な利点も提供する。したがって、それらはハプロタイプ及びコピー数の変動を評価するのに特に有用であり 本明細書で開示されている方法は、試料調製の間に導入される偏りのため、核酸標的の集団にて十分に特徴付けられていないまたは十分に表されていないゲノムの領域にわたって配列情報を提供するのにも有用である。

【 0 0 5 9 】

本明細書で開示されているバーコード付け法の使用は、所与のセットの遺伝子マーカーについての個々の分子コンテキストを提供する独特の能力、すなわち、所与のセットの遺伝子マーカー（単一マーカーとは対照的に）を個々の試料の核酸分子に帰属させ、且つ変異による協調組み立てを介して、複数の試料核酸分子の間で、及び/または特定の染色体に対してさらに広範囲のまたはさらに長距離の推論された個々の構造コンテキスト及び分子コンテキストを提供する独特の能力を付与する。これらの遺伝子マーカーには特定の遺伝子座、たとえば、SNPのような変異が含まれてもよく、またはそれらは短い配列を含んでもよい。さらに、バーコード付けの使用は、たとえば、血流で循環する腫瘍DNAの検出及び特徴付けのために、試料から抽出される全核酸集団の少数構成成分と主要構成成分との間を区別する、及び任意の増幅ステップの間で増幅の偏りを減らすまたは排除する能力を促進する追加の利点を付与する。加えて、マイクロ流体工学形式の実施は、極端に少ない試料容量及び少ない入力量でのDNAで作業する能力と同様に多数の試料の区画（液滴）を迅速に処理してゲノム規模のタグ付けを円滑にする能力を付与する。

【 0 0 6 0 】

前述したように、本明細書に記載されている方法及びシステムの利点は、それらが普遍的に利用できる短い読み取り配列決定技術を介して所望の成績を達成できることである。そのような技術は、十分に特徴付けられ、高度に効果的であるプロトコール及び試薬システムと共に研究社会で容易に利用でき、広く普及されている利点を有する。これらの短い読み取り配列決定技術には、たとえば、Illumina, Inc. (GAIIx, NextSeq, MiSeq, HiSeq, X10), Ion Torrent division of Thermo-Fisher (Ion Proton and Ion PGM) から入手可能なもの、ピロ配列決定法等が挙げられる。

【 0 0 6 1 】

特に有利なのは、本明細書に記載されている方法及びシステムがこれらの短い読み取り配列決定技術を利用し、関連する低い誤差率と高い処理能力でそのようにすることである

。特に、本明細書に記載されている方法及びシステムは上述したように所望の個々の分子読み取り長さ及び分子コンテキストを達成するが、個々の配列決定読み取りは、1000bpより短い、500bpより短い、300bpより短い、200bpより短い、150bpより短い、またはさらに短いメイトペアの伸長を除外し、且つそのような個々の分子読み取り長さについての配列決定誤差率は、5%未満、1%未満、0.5%未満、0.1%未満、0.05%未満、0.01%未満、0.005%未満、またはさらに0.001%未満である。

【0062】

II. 作業の流れの概説

1つの例示的な態様では、本開示に記載されている方法及びシステムは試料を個別区画に入れることまたは区画化することを提供し、その際、各区画は他の区画における内容物からのそれ自体の内容物の分離を維持する。本明細書でさらに詳細に議論されるように、試料は、たとえば、細胞または組織の試料のような患者に由来する試料を含んでもよく、それは核酸を含有することができ、特定の状況では、関連するタンパク質を同様に含有することができる。具体的な態様では、本明細書に記載されている方法で使用する試料には、ホルマリンで固定し、パラフィンで包埋した(FFP E)細胞及び組織の試料等、と同様に試料の分解のリスクが高い他の試料の種類が含まれる。

【0063】

本明細書で使用されるとき、区画は、穴等を介する種々の異なる形態、たとえば、ウェル、試験管、マイクロウェルまたはナノウェルを含んでもよい入れ物または容器を指す。好まれる態様では、しかしながら、区画は流体流れの中で流動可能である。これらの容器は、内部の流体の中心もしくは芯を取り囲む外側のバリアを有する、たとえば、マイクロカプセルまたはマイクロ容器で構成されてもよく、またはそれらは、そのマトリクス内に物質を同伴する及び/または保持することができる多孔性マトリクスであってもよい。しかしながら、好ましい態様では、これらの区画は非水性の連続位相、たとえば、油相内に水性流体の液滴を含んでもよい。種々の異なる容器は、たとえば、2013年8月13日に出版された米国特許出願第13/966,150号に記載されている。同様に、非水性のまたは油の連続相にて安定な液滴を作り出すエマルジョン系は、たとえば、公開米国特許出願第2010-0105112号にて詳細に記載されている。特定の例、マイクロ流体の導管ネットワークが本明細書に記載されているような区画を生成するのに特に適する。そのようなマイクロ流体装置の例には、その完全な開示があらゆる目的で全体として参照によって本明細書に組み入れられる2014年4月4日に出版された米国仮特許出願第61/977,804号にて詳細に記載されているものが挙げられる。それを介して細胞の水性混合物が非水性の流体に押し出される多孔性の膜を含む、代替のメカニズムも個々の細胞の区画化で採用されてもよい。そのようなシステムは一般に、たとえば、Nanomi, Incから入手可能である。

【0064】

エマルジョンにおける液滴の場合、試料物質の個別区画への区画化は一般に、区画化流体、たとえば、フッ素化された油の非水性の流れも流す接合部に水性の試料を含有する流れを流すことによって達成されてもよいので、水性の液滴は流れている流れの区画化流体の中で作り出され、そのような液滴は試料物質を含む。以下に記載されているように、区画、たとえば、液滴も通常、同時に区画化されるバーコードオリゴヌクレオチドを含む。特定の区画内での試料物質に相対量は、たとえば、水性の流れにおける試料の濃度、水性の流れ及び/または非水性の流れの流速等を含む、システムの種々のパラメータを制御することによって調整されてもよい。本明細書に記載されている区画は極端に少ない容量を有することを特徴とすることが多い。たとえば、液滴に基づく区画の場合、液滴は、1000pL未満、900pL未満、800pL未満、700pL未満、600pL未満、500pL未満、400pL未満、300pL未満、200pL未満、100pL未満、50pL未満、20pL未満、10pL未満、またはさらに1pL未満である全体容量を有してもよい。ビーズと共に同時に区画化される場合、区画内の試料流体の容量は、上述の

容量の90%未満、上述の容量の80%未満、70%未満、60%未満、50%未満、40%未満、30%未満、20%未満、またはさらに10%未満であってもよいことが理解されるであろう。場合によっては、少ない反应用量の区画の使用は、非常に少ない量の出発試薬、たとえば、入力核酸で反応を実施することにおいて特に有利である。少ない入力核酸で試料を解析する方法及びシステムは、その完全な開示が全体として参照によって本明細書に組み入れられる2014年6月26日に出願された米国仮特許出願第62/017,580号(代理人整理番号43487-727.101)にて提示されている。

【0065】

分解に供され、及び/または低濃度の対象とする成分を含有する試料を伴う状況では、試料はさらに区画化に先立って、または区画化の範囲内で処理されて核酸及び/または関連するタンパク質をさらなる解析のためにさらに遊離してもよい。たとえば、FFPE試料に含有される核酸は一般に、当該技術分野で既知の方法を用いて抽出される。さらに長い核酸分子を単離するには、そのような試料は、ホルムアルデヒド付加体を取り除くために有機触媒の添加によって処理されてもよい(たとえば、全体として及び特にFFPE試料の処理と加工に関する教示すべてについて参照によって本明細書に組み入れられる、Karmakar, et al., (2015), Nature Chemistry, DOI: 10.1038/NCHEM.2307を参照のこと)。

【0066】

いったん試料がそれぞれの区画に導入されると、区画内の試料核酸は増幅に供されて、その後の適用(たとえば、本明細書に記載されている及び当該技術分野で既知の配列決定法)のために核酸の量を増やしてもよい。特定の実施形態では、この増幅はゲノム配列の異なる部分に向けられるプライマーのライブラリで実施されるので、得られる増幅産物は元の核酸分子の小区分に由来する配列を表す。選択されたゲノム領域が対象になる実施形態では、標的とされる適用範囲について対象とされるゲノムの領域がゲノムの他の領域に比べて高い比率で存在するように、この増幅は1回以上の選択的な増幅を含んでもよい(とはいえ、理解されるように、ゲノムのそれら他の領域はデノボ適用範囲の対象ではないので、それらはほんの少しの程度に増幅されてもよい)。特定の実施形態では、増幅は、ゲノム全体のまたはゲノムの選択領域の少なくとも1x、2x、5x、10x、20x、30x、40x、または50xの適用範囲を提供する。さらなる実施形態では、区画内の核酸のすべてが増幅されるが、選択されたゲノム領域は、ゲノムの他の部分からよりもそれら選択されたゲノム領域から少なくとも1~5、2~10、3~15、4~20、5~25、6~30、7~35、8~40、9~45、または10~50倍多いアンプリコンが生じるように標的とされる方法で増幅される。

【0067】

上述した増幅と同時にまたはそれに続いて、区画内の核酸(またはその断片)に、それら核酸の特徴付けの際にそれらがそのそれぞれの起源に由来したように帰属させられてもよいように独特の識別子が提供される。従って、試料核酸は通常独特の識別子(たとえば、バーコード配列)と共に同時に区画化される。特に好ましい態様では、独特の識別子は、それらの試料に結合されてもよい核酸バーコード配列を含むオリゴヌクレオチドの形態で提供される。所与の区画におけるオリゴヌクレオチドの関係では、それに含有される核酸バーコード配列は同一であるが、異なる区画での関係では、オリゴヌクレオチドは異なるバーコード配列を有することができ、好ましくは有するように、オリゴヌクレオチドは区画化される。例示的な態様では、場合によっては、2以上の異なるバーコード配列が存在してもよいが、1つの核酸バーコード配列のみが所与の区画に関連付けられるであろう。

【0068】

核酸バーコード配列は通常、オリゴヌクレオチドの配列の中に6~約20以上のヌクレオチドを含むであろう。これらのヌクレオチドは完全に連続する、すなわち、隣接するヌクレオチドの単一区間であってもよいし、またはそれらは1以上のヌクレオチドによって分離される2以上の別々の部分配列に分離されてもよい。通常、分離された部分配列は通

10

20

30

40

50

常、長さ約 4 ～ 約 16ヌクレオチドであってもよい。

【0069】

同時に区画化されたオリゴヌクレオチドも通常、区画化された核酸の処理に有用な他の機能的な配列を含む。これらの配列には、たとえば、関連するバーコード配列を結合しながら区画内の個々の核酸からゲノムDNAを増幅するための標的とされるまたは無作為の/普遍的な増幅プライマー配列、たとえば、配列の存在を特定するためのまたはバーコード付きの核酸を破壊するための配列決定プライマー、ハイブリッド形成配列またはプローブ配列、または多数の他の潜在的な機能的配列のいずれかが挙げられる。再び、試料物質と共にオリゴヌクレオチド及び関連するバーコード及び他の機能的配列を同時に区画化することは、たとえば、US N 14 / 175, 935; 14 / 316, 383; 14 / 316, 398; 14 / 316, 416; 14 / 316, 431; 14 / 316, 447; 及び14 / 316, 463に記載され、あらゆる目的で及び特に核酸を処理することと同様にゲノム物質の配列決定及び他の特徴付けに関して記述された説明、図面及び作業実施例すべてについてその全体が参照によって本明細書に組み入れられる。

10

【0070】

簡潔に言えば、1つの例示的な方法では、それぞれがビーズに遊離可能に結合された多数の上述のオリゴヌクレオチドを含んでもよいビーズが提供され、その際、特定のビーズに結合されるオリゴヌクレオチドのすべては同じ核酸バーコード配列を含んでもよいが、多数の多様なバーコード配列が使用されるビーズの集団にわたって示されてもよい。通常、ビーズの集団は、少なくとも1000の異なるバーコード配列、少なくとも10,000の異なるバーコード配列、少なくとも100,000の異なるバーコード配列、または場合によっては少なくとも1,000,000の異なるバーコード配列を含んでもよい多様なバーコード配列ライブラリを提供してもよい。さらに、各ビーズには通常、結合される多数のオリゴヌクレオチド分子が提供されてもよい。特に、個々のビーズ上でバーコード配列を含むオリゴヌクレオチドの分子の数は、少なくとも約10,000オリゴヌクレオチド、少なくとも約100,000オリゴヌクレオチド分子、少なくとも約1,000,000オリゴヌクレオチド分子、少なくとも約100,000,000オリゴヌクレオチド分子、及び場合によっては、少なくとも10億のオリゴヌクレオチド分子であってもよい。

20

【0071】

オリゴヌクレオチドは特定の刺激がビーズに適用されるとビーズから遊離可能であってもよい。場合によっては、刺激は、たとえば、オリゴヌクレオチドを遊離してもよい光に不安定な結合の切断を介した光刺激であってもよい。場合によっては、熱刺激が使用されてもよく、その際、ビーズ環境の温度の上昇が結合の切断またはビーズからのオリゴヌクレオチドの他の遊離を生じてもよい。場合によっては、オリゴヌクレオチドのビーズに対する結合を切断する、またはさもなければビーズからのオリゴヌクレオチドの遊離を生じてよい化学刺激が使用されてもよい。

30

【0072】

本明細書に記載されている方法及びシステムに従って、結合されたオリゴヌクレオチドを含むビーズは個々の試料と共に同時に区画化されてもよいので、単一のビーズと単一の試料が個々の区画内に含有される。場合によっては、単一のビーズ区画が望ましい場合、占有されているそれらの区画が主として単独で占有されることを保証するために、平均で区画が区画当たり1未満のビーズを含有するように流体の相対流速を制御することが望ましくてもよい。同様に、流速を制御して高い比率の区画が占有されることをもたらし、たとえば、ほんの少ない比率の未占有区画を可能にすることを望んでもよい。好ましい態様では、流れ及び導管の構造は、所望の数の単独で占有された区画、特定のレベル未満の未占有の区画及び特定のレベル未満の多重占有された区画を確保するために制御される。

40

【0073】

図3は試料核酸にバーコードを付け、その後その配列決定をする1つの特定の例示的な方法を説明している。まず、核酸を含む試料を供給源から入手してもよく、300、バー

50

コードを付けたビーズのセットも入手してもよい、310。ビーズは好ましくは、1以上のバーコード配列と同様に、たとえば、無作為N量体または他のプライマーのようなプライマーを含有するオリゴヌクレオチドに連結される。好ましくは、たとえば、バーコードとビーズの間の結合の切断を介して、またはバーコードを遊離する内在するビーズの分解を介して、またはその2つの組み合わせを介して、バーコード配列はバーコードを付けたビーズから遊離可能である。たとえば、特定の好まれる態様では、バーコードを付けたビーズは、たとえば、還元剤のような作用剤によって分解し、または溶解してバーコード配列を遊離することができる。この例では、核酸を含む少量の試料305とバーコードを付けたビーズ315と任意で他の試薬、たとえば、還元剤320とが組み合わせられ、区画化に供される。例として、そのような区画化には、液滴生成システム、たとえば、マイクロ流動装置325に成分を導入することを含んでもよい。マイクロ流動装置325の助けを借りて、油中水エマルション330が形成されてもよく、その際、エマルションは試料核酸305と還元剤320とバーコードを付けたビーズ315とを含有する水性液滴を含有する。還元剤はバーコードを付けたビーズを溶解してもよく、または分解してもよく、それによって液滴335内でビーズからバーコードと無作為N量体を伴ったオリゴヌクレオチドを遊離する。次いで無作為N量体は、試料核酸の異なる領域に刺激を与え、増幅後、試料の増幅されたコピーを生じてもよく、その際、各コピーはバーコード配列でタグ付けされる、340。好ましくは、各液滴は同一のバーコード配列と様々な無作為N量体配列とを含有するオリゴヌクレオチドのセットを含有する。その後、エマルションが破壊され、345、追加の配列（たとえば、特定の配列決定法で役立つ配列、追加のバーコード等）が、たとえば、増幅法（たとえば、PCR）を介して付加されてもよい、350。次いで配列決定が行われてもよく、355、アルゴリズムを適用して配列決定データを解釈してもよい、360。配列決定アルゴリズムは一般にたとえば、バーコードの解析を行って配列決定読み取りを並べ、及び/または特定の配列読み取りが属する試料を特定することができる。加えて、且つ本明細書に記載されているように、これらのアルゴリズムをさらに用いてコピーの配列を元の分子コンテキストに帰属させてもよい。

【0074】

理解されるように、バーコード配列によるタグ付け340に先立ってまたはそれと同時に、本明細書に記載されている方法のいずれかに従って試料を増幅し、ゲノム全体のまたはゲノムの選択された領域の適用範囲を提供することができる。タグ付けされた適用範囲が望ましい実施形態については、タグ付けされた増幅は一般に、ゲノムの他の領域に由来するアンプリコンに比べてゲノムのそれら選択された領域を含有する区画にて核酸（またはその一部）の配列を表すアンプリコンの大きな集団を生じる。その結果、ゲノムの他の領域に由来する区画よりもゲノムの選択された領域に由来する区画の中でバーコード配列340を含有する多数の増幅されたコピーが存在するであろう。ゲノム全体の増幅が望ましい実施形態では、増幅の偏りを出来るだけ抑え、ゲノム全体にわたって堅固なレベルの適用範囲を提供するように設計されたプライマーのライブラリを用いて増幅が実施されてもよい。

【0075】

上記で言及されたように、単一占有が最も所望の状態であってもよい一方で、多重占有された区画または未占有の区画がしばしば存在してもよいことが理解されるであろう。試料とバーコードオリゴヌクレオチドを含むビーズとを同時に区画化するためのマイクロ流体導管構造の例を図4にて模式的に説明する。示されるように、導管接合部412にて導管セグメント402、404、406、408及び410が流体連通で提供される。個々の試料414を含む水性の流れが導管セグメント402を介して導管接合部412に向かって流される。本明細書のどこかに記載されているように、これらの試料は区画化過程に先立って水性流体内で懸濁されてもよい。

【0076】

同時に、バーコードを運ぶビーズ416を含む水性の流れが導管セグメント404を介して導管接合部412に向かって流される。非水性の区画化流体が側面導管406及び4

10

20

30

40

50

08から導管接合部412に導入され、合わせた流れが流出導管410に流される。導管接合部412内では、導管セグメント402及び404からの2つの合わせた水性の流れが合流し、液滴418に区画化されるが、それは同時に区画化された試料414とビーズ416とを含む。前述したように、導管接合部412で合流する流体それぞれの流れの特徴を制御すると共に導管接合部の形状を制御することによって、合流及び区画化を最適化して生成される区画418内でビーズ、試料またはその双方の所望の占有レベルを達成することができる。

【0077】

理解されるように、たとえば、ポリメラーゼ、逆転写酵素、ヌクレオシド三リン酸またはNTPの類似体、プライマー配列及びそのような反応で使用する二価金属のような追加の補因子、リガーゼ酵素やライゲーション配列のようなライゲーション反応試薬、色素、標識または他のタグ付け試薬のような化学刺激、核酸の伸長、転写及び/または増幅の試薬を含む多数の他の試薬が試料及びビーズと共に同時に区画化されてもよい。プライマー配列には、無作為プライマー配列、またはゲノムの選択された領域を増幅することに向けられた標的とされるPCRプライマー、またはそれらの組み合わせが挙げられてもよい。

10

【0078】

いったん同時に区画化されると、ビーズ上に配置されたオリゴヌクレオチドを用いて区画化された試料にバーコードを付け、それを増幅してもよい。試料を増幅し、バーコード付けするのにこれらのバーコードオリゴヌクレオチドを使用する特に洗練された方法は、その完全な開示が全体として参照によって本明細書に組み入れられるUS N 14 / 175, 935; 14 / 316, 383; 14 / 316, 398; 14 / 316, 416; 14 / 316, 431; 14 / 316, 447; 及び14 / 316, 463にて詳細に記載されている。簡潔に言うと、1つの態様では、試料と共に同時に区画化され、試料を伴う区画にビーズから遊離される、ビーズ上に存在するオリゴヌクレオチド。オリゴヌクレオチドは通常、バーコード配列と共に5'末端でプライマー配列を含む。プライマー配列は無作為であってもよく、または構造化されてもよい。無作為プライマー配列は一般に試料の多数の異なる領域に無作為に刺激を与えるように意図される。構造化されたプライマー配列には、試料の特定の標的とされる領域の上流に刺激を与えるように標的とされる定義された配列を含む様々な異なる構造と同様に、限定しないで、ある比率の特定の塩基（たとえば、ある比率のGC N量体）を含有するプライマー、部分的にまたは全体的に縮重配列を含有するプライマー及び/または本明細書の説明のいずれかに従って部分的に無作為であり且つ部分的に構造化される配列を含有するプライマーを含むある種の部分的に定義された構造を有するプライマーを挙げることができる。理解されるように、無作為プライマー及び構造化プライマーの上記の種類の内いずれか1以上が組み合わせでオリゴヌクレオチドに含まれてもよい。

20

30

【0079】

いったん遊離されると、オリゴヌクレオチドのプライマー部分は試料の相補性領域にアニーリングすることができる。試料及びビーズと共に同時に区画化される伸長反応試薬、たとえば、DNAポリメラーゼ、ヌクレオシド三リン酸、補因子（たとえば、Mg²⁺またはMn²⁺等）が次いで、鋳型として試料を用いてプライマー配列を伸長し、プライマーがアニーリングする鋳型の鎖に対する相補性の断片を生じ、相補性の断片はオリゴヌクレオチドとそれに関連するバーコード配列を含む。複数のプライマーの試料の異なる部分に対するアニーリング及び伸長は、それぞれ、それが作り出された区画を示すそれ自体のバーコード配列を持つ、試料の重複する相補性の断片の大きなプールを生じてもよい。場合によっては、区画に存在するオリゴヌクレオチドによって刺激を与えられる鋳型としてこれらの相補性の断片自体を用いて、再びバーコード配列を含む相補体の相補体を作り出してよい。場合によっては、この複製過程は、第1の相補体が複製される場合、それは末端にてまたは末端の近傍で2つの相補性配列を生じ、ヘアピン構造または部分ヘアピン構造の形成を可能にし、それがさらなる反復コピーを作る基礎である分子の能力を低下さ

40

50

せるように構成される。この一例の模式的な説明を図5に示す。

【0080】

図が示すように、バーコード配列を含むオリゴヌクレオチドが試料核酸504と共に、たとえば、エマルションにおける液滴502で同時に区画化される。本明細書のどこかで言及されているように、オリゴヌクレオチド508は試料核酸504と同時に区画化されるビーズ506上で提供されてもよく、そのオリゴヌクレオチドは好ましくはパネルAで示すようにビーズ506から遊離される。オリゴヌクレオチド508は、1以上の機能的な配列、たとえば、配列510、514及び516に加えてバーコード配列512を含む。たとえば、バーコード配列512と同様に、所与の配列決定システムのための結合または不動態化の配列として機能してもよい配列510、たとえば、Illumina HiSeqまたはMiseqシステムのフローセルでの結合のために使用されるP5配列を含むようなオリゴヌクレオチド508が示される。示されるように、オリゴヌクレオチドは試料核酸504の一部の複製に刺激を与えるための無作為または標的化のN量体を含んでもよいプライマー配列516も含む。オリゴヌクレオチド508内に含まれるのはまた、配列決定システムにおける合成反応によって、ポリメラーゼが介在し、鋳型が向けられる配列決定に刺激を与えるのに使用される「リード1」またはR1プライミング領域のような配列決定プライミング領域を提供してもよい配列514である。多くの場合、バーコード配列512、不動態化配列510及びR1配列514は所与のビーズに結合されるオリゴヌクレオチドのすべてに共通であってもよい。プライマー配列516は無作為N量体のプライマーについて変化してもよく、または特定の標的とされる適用について所与のビーズ上でのオリゴヌクレオチドに共通してもよい。

【0081】

プライマー配列516の存在に基づいて、パネルBで示すようにオリゴヌクレオチドは試料核酸に刺激を与えることができ、それはビーズ506及び試料核酸504と共に同時に区画化されたポリメラーゼ酵素及び他の伸長試薬を用いてオリゴヌクレオチド508及び508aの伸長を可能にする。パネルCで示すように、オリゴヌクレオチドの伸長に続いて、それは、無作為N量体プライマーについて、試料核酸504の複数の異なる領域にアニーリングし、核酸の複数の重複する成分または断片、たとえば、断片518及び520が作り出される。試料核酸の一部に相補性である配列部分、たとえば、配列522及び524を含むが、これらの構築物は一般に本明細書では、結合されたバーコード配列を有する、試料核酸504の断片を含むと見なされる。理解されるように、上述のような鋳型配列の複製された部分は本明細書ではその鋳型配列の「断片」と呼ばれることが多い。しかしながら、前述にかかわらず、用語「断片」は、たとえば、鋳型配列の一部を提供する他のメカニズム、たとえば、配列の所与の分子の実際の断片化によって、たとえば、酵素的な、化学的なまたは機械的な断片化を介して作り出されるものを含む、元の核酸配列、たとえば、鋳型または試料核酸の一部の任意の表現を包含する。しかしながら、好ましい態様では、鋳型配列または試料核酸配列の断片は基本となる配列またはその相補体の複製された部分を示すであろう。

【0082】

バーコードを付けた核酸断片は次いで、たとえば、配列解析を介した特徴付けに供されてもよく、またはそれらはパネルDで示すように過程でさらに増幅されてもよい。たとえば、追加のオリゴヌクレオチド、たとえば、ビーズ506から遊離されるオリゴヌクレオチド508bは断片518及び520に刺激を与えてもよい。特に再び、オリゴヌクレオチド508bにおける無作為N量体プライマー516b（多くの場合、所与の区画における他の無作為N量体、たとえば、プライマー配列516とは異なるであろう）の存在に基づいて、オリゴヌクレオチドは断片518とアニーリングし、伸長して試料核酸配列の一部の複製を含む、配列528を含む断片518の少なくとも一部に対する相補体526を作り出す。オリゴヌクレオチド508bの伸長はそれが断片518のオリゴヌクレオチド部分508を介して複製するまで続く。本明細書のどこかで言及されているように、且つパネルDで説明されているように、オリゴヌクレオチドは、所望の点で、たとえば、断片

5 1 8 内に含まれるオリゴヌクレオチド 5 0 8 の配列 5 1 6 及び 5 1 4 を介した複製の後でポリメラーゼによる複製での停止を促すように構成されてもよい。本明細書に記載されているように、これは、たとえば、使用されるポリメラーゼ酵素によって処理できない異なるヌクレオチド及び/またはヌクレオチド類似体の組込みを含む様々な方法によって達成されてもよい。たとえば、これには、非ウラシル耐性ポリメラーゼを抑えてその領域の複製を止める配列領域 5 1 2 内でのウラシル含有ヌクレオチドの包含が含まれてもよい。その結果、バーコード配列 5 1 2、結合配列 5 1 0、R 1 プライマー領域 5 1 4 及び無作為 N 量体配列 5 1 6 b を含む完全長のオリゴヌクレオチド 5 0 8 b を一方の末端で含む断片 5 2 6 が作り出される。配列の他方の末端では、第 1 のオリゴヌクレオチド 5 0 8 の無作為 N 量体に対する相補体 5 1 6 ' と同様に配列 5 1 4 ' として示される R 1 配列の全部または一部に対する相補体が含まれるであろう。次いで R 1 配列 5 1 4 及びその相補体 5 1 4 ' は一緒にハイブリッド形成して部分的なヘアピン構造 5 2 8 を形成することができる。無作為 N 量体は異なるオリゴヌクレオチド間では異なるので理解されるように、これらの配列及びその相補体はヘアピン形成に加わるようには予想されず、たとえば、無作為 N 量体 5 1 6 に対する相補体である配列 5 1 6 ' は無作為 N 量体 5 1 6 b に対して相補性であるとは予想されない。他の適用、たとえば、N 量体が所与の区画内のオリゴヌクレオチドの間で共通する場合の標的とされるプライマーについてはこうではない。これらの部分的なヘアピン構造を形成することによって、それは、さらなる複製からの試料配列の第 1 レベルの複製の除去を可能にし、たとえば、コピーの反復コピーを防ぐのを可能にする。部分的なヘアピン構造はまた、作り出された断片、たとえば、断片 5 2 6 のその後の処理のための有用な構造も提供する。

【 0 0 8 3 】

次いで、複数の異なる区画に由来する断片のすべてを、本明細書に記載されているような高処理能力シーケンサーで配列決定するためにプールしてもよい。各断片は起源のその区画に関してコードされているので、その断片の配列はバーコードの存在に基づいてその起源に戻って帰属させてもよい。これは図 6 で模式的に説明されている。一例で示すように、第 1 の供給源 6 0 0 (たとえば、個々の染色体、核酸の鎖等)を起源とする核酸 6 0 4 及び異なる染色体 6 0 2 または核酸の鎖に由来する核酸 6 0 6 は、上述のようなバーコードオリゴヌクレオチドのそれ自体のセットと共にそれぞれ区画化される。

【 0 0 8 4 】

各区画内で、各核酸 6 0 4 と 6 0 6 は次いで処理されて第 1 の断片 (複数可) の第 2 の断片の重複するセット、たとえば、第 2 の断片セット 6 0 8 及び 6 1 0 を別々に提供する。この処理は、特定の第 1 の断片に由来する第 2 の断片のそれぞれについて同一であるバーコード配列を伴った第 2 の断片も提供する。示されるように、第 2 の断片セット 6 0 8 についてのバーコード配列は「 1 」で示される一方で、断片セット 6 1 0 についてのバーコード配列は「 2 」で示される。バーコードの多様なライブラリを用いて多数の異なる断片セットに差別的にバーコード付けをしてもよい。しかしながら、異なる第 1 の断片に由来するあらゆる第 2 の断片セットに異なるバーコード配列でバーコード付けする必要はない。実際、多くの場合、複数の異なる第 1 の断片は同時に処理されて同一のバーコード配列を含んでもよい。多様なバーコードライブラリは本明細書のどこかで詳細に記載されている。

【 0 0 8 5 】

次いで、たとえば断片セット 6 0 8 及び 6 1 0 に由来するバーコード断片が、たとえば、Illumina or Ion Torrent division of Theermo Fisher, Inc., 等から入手可能な合成技術による配列を用いた配列決定のためにプールされてもよい。いったん配列決定されると、少なくともある程度含まれるバーコードに基づいて、且つ任意で好ましくはある程度断片自体の配列に基づいて、プールされた断片 6 1 2 からの配列読み取りは、たとえば、集約読み取り 6 1 4 及び 6 1 6 で示されるようなその各断片セットに帰属させることができる。加えて、配列読み取りは、これらの読み取りが原試料内で空間的に密に近接した他の核酸分子と関連して導き出さ

れる核酸の相対的な位置の構造コンテキストに帰属させることができる。各断片セットについて帰属させた配列読み取りを次いで集合させて各試料断片について集合した配列、たとえば、配列 6 1 8 及び 6 2 0 を提供し、それは次に、それぞれの元の染色体または供給源核酸分子 (6 0 0 及び 6 0 2) に戻ってさらに帰属させてもよい。ゲノム配列を集合させる方法及びシステムは、たとえば、その完全な開示が全体として及び特にゲノム配列の集合に関する教示すべてについて参照によって本明細書に組み入れられる、2015年6月26日に出願された米国特許出願第 1 4 / 7 5 2 , 7 7 3 号に記載されている。

【 0 0 8 6 】

III . 構造コンテキストを保持するための方法及び組成物

本開示は遺伝物質の特徴付けのための方法、組成物及びシステムを提供する。一般に、本明細書に記載されている方法、組成物及びシステムは、成分が元々試料に存在したときのそれら成分の構造コンテキストと同様に分子コンテキストに関する情報を保持しながら、試料の成分を解析する方法を提供する。言い換えれば、本明細書の記載は一般に、たとえば、ホルマリンで固定されパラフィンに包埋された試料のような、当該技術分野で既知の方法を用いて固定されているまたは固定される組織試料を含む試料における核酸の空間的な検出に関する。理解されるように、このセクションに記載されている方法のいずれかは、「概説」及び「作業の流れの概説」と題するセクションで上述された方法のいずれかと同様に本明細書の後に続くセクションに記載されている核酸の配列決定法と組み合わせることができる。

【 0 0 8 7 】

一般に、本明細書で開示されている方法は、ゲノム、特に試料の全体的なゲノムを含む試料における核酸を測定すること及び/または解析することに関する。本明細書に記載されている方法は、試料内の空間的コンテキストが保持されている試料における核酸配列 (ゲノム配列を含む) の分布、位置または発現を定量的にまたは定性的に解析する能力を提供する。本明細書で開示されている方法は、配列読み取りのために試料を処理するのに先立って特定の分子標的 (たとえば、特定の遺伝子または他の核酸配列) の特定を必要とすることなく、構造コンテキストに関する情報が高度な処理能力の処理法にて保持されるので、試料にて核酸を地理的にコードする従来の方法に対する利点を提供する。加えて、少量の核酸を必要とし、それは入力核酸、特に DNA が断片化されているまたは低濃度で存在することが多い F F P E 試料のような試料で特に有利である。

【 0 0 8 8 】

本明細書の議論の多くが核酸の解析に関するものであるが、本明細書で議論されている方法及びシステムは、タンパク質及び他の分子を含む試料の他の成分に適用するように適応させることができることが理解されるであろう。

【 0 0 8 9 】

上述したように、本明細書では地理コンテキストを維持すること及び地理をコードすることとも言われる構造コンテキストを維持することは、試料内のそれらの配列読み取りの元の三次元の相対的な位置に帰属させることができる複数の配列読み取りまたは配列読み取りの複数の部分を得ることを可能にする方法を使用することを意味する。言い換えれば、配列読み取りは、試料における近隣の核酸 (及び一部の状況では関連するタンパク質) に関するその試料内の相対的な位置に関連付けることができる。それら近隣の核酸が単一の元の核酸分子の線状配列内に物理的に位置していないとしてもこの空間的な情報は利用可能である。

【 0 0 9 0 】

一般に、本明細書で開示されている方法は、核酸を含有する試料が提供され、その際、核酸は三次元構造を含有する解析を含む。試料の一部は核酸の三次元構造の一部も個別区画に分離されるように個別区画に分離され 互いに空間的に近接し合う核酸配列は同一区画に分離される傾向があるので、後で得られる配列読み取りが同じ個体の元の核酸分子上に元々なかった配列に由来するとしても、その空間的近接性の三次元情報を保持する。図 1 を参照して、核酸分子 1 0 2 及び 1 0 3 及び 1 0 6 を含有する試料 1 0 1 が、試料のサ

ブセットが異なる個別区画に割り当てられるように個別区画に分離される場合、核酸分子 106 と 102 と 103 の間での物理的距離のため、核酸分子 106 よりも核酸分子 102 と 103 が互いに同じ区画に置かれる可能性が高い。従って、同一の個別区画内の核酸分子は原試料にて互いに空間的に近接しあっていたものである。従って、個別区画内での核酸から得られる配列情報は、たとえば、核酸の配列決定を介して核酸を解析し、それらの配列読み取りを元の核酸分子の構造コンテキストに帰属させ戻す方法を提供する。

【0091】

一部の例では、試料の空間的なまたは地理的なコード化のためにタグのライブラリが試料に適用される。特定の実施形態では、タグはオリゴヌクレオチドタグ（「オリゴヌクレオチドのバーコード」及び「DNAバーコード」を含むことができる）であるが、理解されるように、限定しないで粒子、ビーズ、色素、分子反転プローブ（MIP）等を含む、試料に付加することができる任意の種類のタグを使用することができる。タグのライブラリは、試料の拡散を介して、または、たとえば、組織培養もしくは細胞培養の試料内での細胞過程のような能動的な過程を介して試料に適用することができる。細胞の輸送過程には、限定しないで浸透作用、細胞輸送タンパク質の関与を介した促進拡散、受動輸送、及び細胞輸送タンパク質と ATP のような分子に由来するエネルギーの入力の関与を介して能動輸送が挙げられる。一般に、タグは、試料内での異なる空間的な / 地理的な位置が異なるタグ及び / またはタグの異なる集団を受け取るように適用される。試料のさらなる処理及び試料内の核酸の解析は、タグの特定を介して特定の空間的コンテキストに帰属させることができる。たとえば、図 1 を参照して、試料 101 へのタグのライブラリの付加は、核酸 106 よりもタグのライブラリの異なる部分または集団に対して空間的な近接性を有する核酸 102 及び 103 を生じることになる。本明細書に記載されている作業の流れに従った試料のさらなる処理は、次いでタグの同一部分 / 集団に関連する核酸 102 及び 103 を生じることになるので、それらのタグの特定は核酸 102 及び 103 が原試料 101 にて互いに空間的に近接し合っていたことを示すことになる。タグの異なる部分 / 集団による核酸 106 の特定は、核酸 106 が原試料にて核酸 102 及び 103 とは異なる空間的位置にあったことを示すことになる。

【0092】

さらなる例では、区画に専用のバーコードが採用されるので、得られる任意の配列読み取りは元の核酸分子が位置していた区画に帰属させ戻すことができる。上述したように、配列読み取りを特定の区画に関連付けることは、原試料の地理にて互いに空間的に近接し合っていた核酸分子を特定する。図 2 で描いたもののような作業の流れのさらなる使用も、配列読み取りの分子コンテキストに関する情報を提供するので、個々の配列読み取りはそれらが起源とした個々の核酸分子に帰属させることができる。

【0093】

試料のタグ付けを可能にするには、オリゴヌクレオチドタグまたは他の標識のような外来性の分子の適用を可能にする当該技術分野で既知の方法を用いて試料を処理してもよい。たとえば、FFPE 試料が使用される実施形態では、試料を加熱して試料の中へのタグを埋め込ませることによってタグを試料に適用させることができ、次いで試料を冷却してもよく、さらに、個別区画への分割及び、試料中の核酸の配列及びそれら配列読み取りに対して空間的に密に近接するタグを特定するさらなる解析を含む、本明細書に記載されている方法のいずれかに従って処理されてもよいことから、それらの配列読み取りの構造コンテキストが保持される。他の試料の処理方法には、分子及びタンパク質の要素を保持しながら、細胞外マトリクス及び / または他の構造的な障害を取り除く組織処理方法が挙げられる。そのような方法には一部の非限定例では、たとえば、Tomer, et al., VOL. 9, NO. 7, 2014, Nature Protocols; Kerschull, et al., Neuron, Volume 91, Issue 5, 7, September, 2016, Pages, 975 - 987; Chung, K. et al. Structural and molecular interrogation of intact biological systems. Nature, 497,

332-337(2013); Susaki, E. A. et al. Whole-brain imaging with single-cell resolution using chemical cocktails and computational analysis. Cell 157, 726-739 (2014); 及び Lee, et al., ACT-RESTO: Rapid and consistent tissue clearing and labeling method for 3-dimensional (3D) imaging, Scientific Reports, 2016/01/11/オンライン; Vol. 6, p. 18631に記載されるものを含む、他の組織一掃法及び標識法の使用と同様にCLARITY法が挙げられ、そのそれぞれがあらゆる目的で、特に構造上及び分子上の照会法での使用のために試料を処理することに關するすべての教示について全体として参照によって本明細書に組み入れられる。

10

【0094】

特定の実施形態では、本明細書に記載されている方法は、特に、たとえば、FFPE試料のようなスライドに固定化される試料について、試料内でのタグの空間的な位置を特定する画像化法と組み合わせて使用される。そのような画像化法は、配列読み取りのスライド上の特定の位置との相関を可能にしてもよく、それはそれらの試料で実施されていてもよい他の病理/画像化試験との相関を可能にする。たとえば、画像化法を用いて病態の予備的な特定を提供してもよい。構造コンテキストを維持しながら配列読み取りをさらに提供する本明細書に記載されている配列決定法は、そのような画像解析と組み合わせて配列読み取りを構造コンテキストと相関させ、病態のその予備的な特定に関するさらなる情報を裏付けるまたは提供する。加えて、画像化法は光学特性を持つタグと組み合わせて使用されてもよいので、特定のタグが画像化された試料の特定の領域に関連付けられる。それらの特定されたタグと相関する配列読み取りは次いでそれらタグの位置のおかげで画像化された試料の領域とさらに相関してもよい。しかしながら、本明細書に記載されている方法はそのような画像法とは無関係であり、構造コンテキストを保持する能力は試料における核酸の空間的情報を決定するのに画像化法を使用することとは無関係である。

20

【0095】

例示的な態様の1つでは、試料にてオリゴヌクレオチドの勾配が生成され、配列決定を介したその後の処理を介して解読することができる座標系を提供する。そのような勾配はオリゴヌクレオチドまたはオリゴヌクレオチドの集団によって試料における細胞及び/または核酸にタグ付けすることを可能にし、それは原試料内での物理的な位置にマッピングすることができる。この座標系は、オリゴヌクレオチドのライブラリを試料の中に拡散させることによって、及び/または試料の特定の領域にオリゴヌクレオチドを注入することによって開発することができる。拡散を使用する場合、拡散動態の標準的な計算によってオリゴヌクレオチドタグの集団と原試料におけるその空間的な位置との間で相関が提供されるであろう。従って、オリゴヌクレオチドタグのその集団で特定された他の核酸が次に試料の特定の地理的領域に相関することができる。

30

【0096】

さらなる例示的な実施形態では、方法には、試料の異なる地理的領域が異なるタグを受け取るようにタグのライブラリが試料に適用される構造コンテキストを維持しながら核酸を解析する過程が含まれる。今や付加したタグと同様に元の核酸を含有する試料の一部を次いで個別区画に分離するので、試料内の地理的位置で互いに近接し合うタグのライブラリの一部と核酸の一部は最終的に同じ個別区画に落ち着く。本明細書で詳細に記載されているもののような配列決定の過程を用いて個別区画における核酸の配列読み取りを提供する。それら配列決定の過程の前に、その後でまたはそれと同時にタグも特定することができる。配列読み取りの特定のタグ(またはタグの濃度勾配が使用される実施形態におけるタグの集団)との相関はそれによって配列読み取りの空間的コンテキストを提供するのに役立つ。上述したように、空間的なコード化に使用されるタグが区画専用のバーコードと併せて使用される実施形態はさらに配列読み取りについての構造コンテキスト及び分子コ

40

50

ンテキストを提供する。

【0097】

IV. 方法及びシステムの核酸配列決定への適用

本明細書に記載されている方法、組成物及びシステムは核酸の配列決定法で特に使用し易い。短い読み取り及び長い読み取りの配列決定法を含むそのような配列決定法には当該技術分野で既知の技法を挙げることができる。特定の態様では、本明細書に記載されている方法、組成物及びシステムは短い読み取りで精度の高い配列決定法にて使用される。

【0098】

一般に、本明細書に記載されている方法及びシステムは、短い読み取りの配列決定技術の極端に低い配列決定誤差率と高い処理能力の利点を有する方法を用いてゲノム配列決定を達成する。前述したように、本明細書に記載されている方法及びシステムの利点は、それらが普遍的に利用できる短い読み取りの配列決定技術の使用を介して所望の結果を達成できることである。そのような技術は、利用し易く且つ研究社会に広く普及しているという利点を有し、十分に特徴付けられ、高度に効果的であるプロトコール及び試薬のシステムを伴う。これらの短い読み取りの配列決定技術には、たとえば、Illumina, Inc. (GAIIx, NextSeq, MiSeq, HiSeq, X10)、Ion Torrent division of Thermo-Fisher (Ion Proton and Ion PGM) から入手できるもの、ピロ配列決定等が挙げられる。

【0099】

特に有利なことは、本明細書に記載されている方法及びシステムがこれらの短い読み取りの配列決定技術を利用するので、それに関連する低い誤差率で済ませることである。特に、本明細書に記載されている方法及びシステムは、上述されているように所望の個々の分子の読み取り長さまたはコンテキストを達成するが、個々の配列読み取りは1000bpより短い、500bpより短い、300bpより短い、200bpより短い、150bpより短い、またはさらに短いメイトペアの伸長を排除し、そのような個々の分子読み取り長さについての配列決定誤差率は5%未満、1%未満、0.5%未満、0.1%未満、0.05%未満、0.01%未満、0.005%未満、またはさらに0.001%未満である。

【0100】

本出願に記載されている方法及びシステムに従って核酸を処理し、配列決定する方法は、US N 14 / 316, 383; 14 / 316, 398; 14 / 316, 416; 14 / 316, 431; 14 / 316, 447; 及び14 / 316, 463にもさらに詳細に記載され、あらゆる目的で、特に核酸を処理すること及び配列決定及びゲノム物質の他の特徴付けに関して記述された説明、図面及び作業実施例すべてについてその全体が参照によって本明細書に組み入れられる。

【0101】

一部の実施形態では、構造コンテキスト及び分子コンテキストの双方を保持しながら配列情報を入手するための本明細書に記載されている方法及びシステムは全ゲノム配列決定に使用される。一部の実施形態では、本明細書に記載されている方法はゲノムの標的とされる領域の配列決定に使用される。さらなる実施形態では、本明細書に記載されている配列決定法は選択される領域の深い適用範囲と、ゲノムのさらに長い範囲にわたる低レベルの連結した読み取りとの組み合わせを含む。理解されるように、デノボ配列決定及び再配列決定のこの組み合わせは、ゲノム全体及び/またはゲノムの大きな部分を配列決定する効率的な方法を提供する。不十分にしか特徴付けられていない領域及び/または高度に多形である領域の標的とされる適用範囲はさらに、デノボ配列集合に必要な量の核酸物質を提供するのに対して、ゲノムの他の領域にわたる連結されたゲノム配列決定はゲノムの残りの高い処理能力の配列決定を維持する。本明細書に記載されている方法及び組成物は、同一の配列決定基盤が適用範囲の双方の種類に使用され得るので、デノボ読み取り配列決定及び連結された読み取り配列決定のこの組み合わせを可能にし易い。本明細書に記載されている方法に従って配列決定される核酸及び/または核酸の断片の集団は、デノボ配列

決定のためのゲノム領域及び再配列決定のためのゲノム領域の双方に由来する配列を含有することができる。

【0102】

具体例では、本明細書に記載されている方法は、配列決定に先立ってゲノムの全体領域または選択される領域が増幅されるステップを含む。一般に当該技術分野で既知の方法（限定しないでPCR増幅を含む）を用いて実施されるこの増幅は、ゲノムの全体領域または選択される領域の少なくとも1x、2x、3x、4x、5x、6x、7x、8x、9x、10x、11x、12x、13x、14x、15x、16x、17x、18x、19x、または20xの適用範囲を提供する。さらなる実施形態では、増幅はゲノムの全体領域または選択される領域の少なくとも1x~30x、2x~25x、3x~20x、4x~15x、または5x~10xの適用範囲を提供する。

10

【0103】

ゲノム全体及び/またはゲノムの選択した標的とされる領域の適用範囲についての増幅は一般にゲノムの選択された領域内またはその近傍の配列に対して相補性のプライマーの伸長を介して実施される。場合によっては、対象とするゲノム領域をうめるように設計されているプライマーのライブラリが使用され 言い換えれば、プライマーのライブラリは、これが選択された領域にわたるか、全ゲノムにわたるかにかかわらず、ゲノムに沿った特定の距離での領域を増幅するように設計される。一部の例では、選択的な増幅はゲノムの選択された領域に沿って10、15、20、25、50、100、200、250、500、750、1000または10000の塩基ごとに相補性であるプライマーを利用する。別のさらなる例では、プライマーのタイル状のライブラリは距離の混合を捕捉するように設計され その混合は距離の無作為な混合であることができ、または選択された領域の特定の部分もしくは比率が異なるプライマー対によって増幅されるように賢く設計することができる。さらなる実施形態では、プライマー対は、各対がゲノムの選択された部分の連続する領域の約1~5%、2~10%、3~15%、4~20%、5~25%、6~30%、7~35%、8~40%、9~45%、または10~50%を増幅するように設計される。

20

【0104】

特定の実施形態では、且つ上述のいずれかに従って、増幅は少なくとも3メガ塩基対(Mb)の長さであるゲノムの領域にわたって生じる。さらなる実施形態では、ゲノムの選択された領域は本明細書に記載されている方法のいずれかに従って選択的に増幅され、その選択された領域は少なくとも3.5、4、4.5、5、5.5、6、6.5、7、7.5、8、8.5、9、9.5、または10Mbの長さである。別のさらなる実施形態では、ゲノムの選択された領域は約2~20、3~18、4~16、5~14、6~12、または7~10Mbの長さである。増幅は、これらの領域の末端でのまたは末端の近傍での配列に相補性の単一のプライマー対を用いてこれらの領域にわたって生じ得る。他の実施形態では、増幅は領域の長さをうめるプライマー対のライブラリで実施されるので、領域に沿った規則的なセグメント、無作為のセグメント、または異なるセグメント距離のいくつかの組み合わせが上の記載に従って適用範囲の程度によって増幅される。

30

【0105】

一部の実施形態では、ゲノムの選択された領域の選択的な増幅で使用するプライマーはプライマー自体が増幅されないようにウラシルを含有する。

40

【0106】

使用される配列決定の基盤にかかわらず、一般に、且つ本明細書に記載されている方法のいずれかに従って、核酸の配列決定は通常、配列読み取りまたは配列読み取りの一部の構造コンテキスト及び分子コンテキストを維持する方法で実施される。これは、複数の配列読み取りまたは配列読み取りの複数の部分が他の核酸に比べて原試料内で相対的な空間的位置（構造コンテキスト）に、及び/または核酸の単一の元の分子の線状配列内での位置（分子コンテキスト）に帰属可能であってもよいことを意味する。

【0107】

50

理解されるように、核酸の単一の元の分子が種々の長さのいずれかであってもよい一方で、好ましい態様では、それは相対的に長い分子であり、長距離の分子コンテクストの維持を可能にする。特に、単一の元の分子は好ましくは、典型的な短い読み取り長さより実質的に長く、たとえば、200塩基より長く、少なくとも1000塩基以上長く、5000塩基以上長く、10,000塩基以上長く、20,000塩基以上長く、30,000塩基以上長く、40,000塩基以上長く、50,000塩基以上長く、60,000塩基以上長く、70,000塩基以上長く、80,000塩基以上長く、90,000塩基以上長く、または100,000塩基以上長いことが多く、場合によっては、1メガ塩基以上長い。

【0108】

一般に、本発明の方法は、図2で説明されているような工程を含み、それは本明細書でさらに詳細に議論されている本発明の方法の図式的概観を提供する。理解されるように、図2で要点が述べられている方法は必要に応じて且つ本明細書に記載されているように変更されてもよくまたは改変されてもよい例示的な実施形態である。

【0109】

図2で示すように、本明細書に記載されている方法はほとんどの例で試料が区画化されるステップ(202)を含む。その区画化するステップに先立って、試料における核酸が連結されて、互いに空間的に密に近接する配列領域を結合する任意のステップ(201)があってもよい。一般に、対象とするゲノム領域に由来する核酸を含有する各区画はある種の断片化過程を経験し、断片の元の分子コンテクストは一般にそれらが含有される区画に専用である断片にバーコードを付けることによって普通に保持されるであろう(203)。各区画は一部の例では1を超える核酸を含んでもよく、一部の例では、数百の核酸分子を含有するであろうし、複数の核酸が区画内にある状況では、ゲノムの特定の遺伝子座が一般にバーコード付けに先立って単一の個体の核酸によって表されるであろう。上述したように、ステップ203のバーコード付き断片は当該技術分野で既知の方法を用いて生成することができ、一部の例では、オリゴヌクレオチドが個別区画内での試料である。そのようなオリゴヌクレオチドは試料の多数の異なる領域に刺激を与えるように意図される無作為な配列を含んでもよく、またはそれらは試料の標的とされる領域の上流に刺激を与えるように標的とされる特定のプライマー配列を含んでもよい。さらなる例では、これらのオリゴヌクレオチドはバーコード配列も含有するので、複製過程も原試料核酸の得られる複製された断片にバーコードを付ける。区画にも含有される伸長反応試薬、たとえば、DNAポリメラーゼ、ヌクレオシド三リン酸、補因子(たとえば、 Mg^{2+} または Mn^{2+} 等)が次いで、試料を鋳型として用いてプライマー配列を伸長させ、プライマーがアニーリングされる鋳型の鎖に相補性の断片を生じ、相補性の断片はオリゴヌクレオチドとそれに関連するバーコード配列とを含む。試料の異なる部分への複数のプライマーのアニーリングと伸長は試料の重複する相補性断片の大きなプールを生じることができ、各々それが作り出された区画を示すそれ自体のバーコード配列を処理する。場合によっては、これらの相補性断片はそれ自体が区画に存在するオリゴヌクレオチドによって刺激を与えられる鋳型として使用されて、再びバーコード配列を含む相補体の相補体を生じてよい。さらなる例では、この複製過程は、第1の相補体が複製されるとき、それは末端でまたは末端の近傍で2つの相補性配列を生じてヘアピン構造または部分ヘアピン構造の形成を可能にし、それがさらなる反復コピーを生じる基礎である分子の能力を低下させるように構成される。

【0110】

図2で例示されている方法に戻って、いったん区画専用のバーコードがコピーされた断片に結合されると、バーコード付き断片は任意で次いでプールされ得る(204)。プールされた断片を次いで配列決定し(205)、断片の配列を元の分子コンテクストに帰属させる(206)ので、対象とする標的とされる領域が特定され、その元の分子コンテクストとも結び付けられる。本明細書に記載されている方法及びシステムの利点は、標的とされるゲノム領域について断片を濃縮するのに先立って区画または試料に専用のバーコー

10

20

30

40

50

ドをコピーされた断片に結合することが、それら標的とされる領域の元の分子コンテキストを維持し、それらを元の区画、従ってその元の試料核酸に帰属させることである。

【 0 1 1 1 】

上記の作業の流れに加えて、標的とされるゲノム領域はさらに、チップに基づく及び溶液に基づく捕捉方法の双方を含む方法を用いてさらなる解析、特に配列決定のために濃縮され、単離され、または分離され、すなわち、「壊されて」もよい。そのような方法は対象とするゲノム領域に、または対象とするゲノム領域の近傍もしくは隣接する領域に対して相補性であるプローブを利用する。たとえば、ハイブリッド（またはチップに基づく）捕捉では、対象とする領域に総合して及ぶ配列と共に捕捉プローブ（普通、一本鎖オリゴヌクレオチド）を含有するマイクロアレイが表面に固定される。ゲノムDNAは断片化され、さらに、たとえば、平滑末端を生じる末端修飾及び／または汎用プライミング配列のような追加の特徴の付加のような処理を受けてもよい。これらの断片はマイクロアレイ上のプローブとハイブリッド形成する。ハイブリッド形成しなかった断片は洗い流され、所望の断片は溶出されるか、またはさもなければ配列決定もしくは他の解析のために表面で処理されるので、表面上に残っている断片の集団は対象とする標的とされる領域（たとえば、捕捉プローブに含有されるものに相補性である配列を含む領域）を含有する断片について濃縮される。断片の濃縮された集団は当該技術分野で既知の増幅法を用いてさらに増幅されてもよい。そのような標的とされるプルダウン濃縮法についての例示的な方法は、2014年10月29日に出願されたUS 2014/0207216 A1に記載されており、それは、あらゆる目的で及び特に、記述された説明、図面及び実施例を含む、標的とされるプルダウン濃縮法及び配列決定法に関する教示すべてについてその全体が参照によって本明細書に組み入れられる。

【 0 1 1 2 】

一部の例では、ゲノム全体の配列決定ではなく、ゲノムの選択された領域に焦点を当てることが望ましい。ゲノムのサブセットが大きな線状の距離であるが、原試料の三次元コンテキストでは潜在的にほぼ近接している場合でさえ、ゲノムのそれらサブセットを標的とする能力はこれらの方法の有利な特徴であるので、本明細書に記載されている方法はそのような解析に特に適している。一部の態様では、ゲノムの選択された領域の適用範囲についての方法には、それらの選択された領域に由来する核酸分子及び／またはその断片を含有する個別区画自体がさらなる処理のために選別される方法が含まれる。理解されるように、個別区画のこの選別は、選択的な増幅の他の方法及び／または本明細書に記載されている対象とするゲノム領域の標的とされる破壊との組み合わせで、特に上述されている作業の流れのステップとの組み合わせで行われてもよい。

【 0 1 1 3 】

一般に、個別区画を選別する方法には、ゲノムの1以上の選択された部分の少なくとも一部を含有する区画がゲノムのそれらの部分に由来する配列を含有しない区画から分離されるステップが含まれる。これらの方法には、ゲノムのそれらの部分に由来する配列を含有する個別区画内でゲノムの1以上の選択された部分の少なくとも一部を含む断片の配列について濃縮された集団を提供するステップが含まれる。そのような濃縮は一般に、ゲノムの1以上の選択された部分の少なくとも一部を含んで集団を生じる個別区画内で断片の指向性PCR増幅を使用することを介して達成される。従って、この指向性PCR増幅はゲノムの1以上の選択された部分の少なくとも一部を含むアンプリコンを生じる。特定の実施形態では、これらのアンプリコンは、一部の非限定の実施形態では、蛍光分子を含んでもよい検出可能な標識に結合される。一般に、そのような結合は、ゲノムの1以上の選択された部分を含有する断片から生成されるそれらのアンプリコンのみが検出可能な標識に結合されるように生じる。一部の実施形態では、検出可能な標識の結合はゲノムの1以上の選択された部分の選択的な増幅の間に生じる。そのような検出可能な標識には、さらなる実施形態では、限定しないで蛍光標識、電気化学的標識、磁気ビーズ及びナノ粒子が挙げられる。検出可能な標識のこの結合は当該技術分野で既知の方法を用いて達成することができる。別のさらなる実施形態では、ゲノムの1以上の選択された部分の少なくとも

一部を含む断片を含有する個別区画はそれらの区画内のアンブリコンの結合された検出可能な標識から発せられるシグナルに基づいて選別される。

【0114】

さらなる実施形態では、ゲノムの選択された部分を含有する個別区画をそのような配列を含有しないものから選別するステップには、(a) 出発ゲノム物質を提供するステップと；(b) 各個別区画が第1の個々の核酸分子を含有するように出発ゲノム物質から個々の核酸分子を個別区画に分配するステップと；(c) 個別区画の少なくとも一部の中で、ゲノムの1以上の選択された部分の少なくとも一部を含む断片の配列について濃縮されている集団を提供するステップと；(d) 断片のそれぞれがそれが含有されていた個別区画に帰属可能であるように各個別区画内で断片に共通のバーコード配列を結合するステップと；(e) ゲノムの1以上の選択された部分を含む断片を含有しない個別区画からゲノムの1以上の選択された部分の少なくとも一部を含む断片を含有する個別区画を分離するステップと；(f) ゲノムの1以上の選択された部分の少なくとも一部を含む断片から配列情報を入手し、それによって分子コンテキストを保持しながら、ゲノム試料の1以上の標的とされる部分を配列決定するステップとが含まれる。理解されるように、そのような方法のステップ(a)は1を超える個々の核酸分子を含むことができる。

10

【0115】

さらなる実施形態では、且つ上記のいずれかに従って、断片から配列情報を入手することに先立って、個別区画が組み合わせられ、断片と一緒にプールされる。さらなる実施形態では、断片から配列情報を入手するステップは断片の配列の構造コンテキスト及び分子コンテキストを維持するような方法で実施されるので、特定することはさらに、原試料内で物理的に密に近接して位置する及び/または同じ第1の個々の核酸分子に位置する核酸に由来する断片を特定することを含む。別のさらなる実施形態では、配列情報のこの入手には、短い読み取り長さの配列決定反応及び長い読み取り長さの配列決定反応から成る群から選択される配列決定反応が含まれる。別のさらなる実施形態では、配列決定反応は短い読み取りで精度の高い配列決定反応である。

20

【0116】

別のさらなる実施形態では、且つ上記のいずれかに従って、個別区画はエマルションにおける液滴を含む。さらなる実施形態では、個別区画内でのバーコード付き断片はゲノムの1以上の選択された部分の約 $1 \times \sim 10 \times$ の適用範囲を表す。別のさらなる実施形態では、個別区画内でのバーコード付き断片はゲノムの1以上の選択された部分の約 $2 \times \sim 5 \times$ の適用範囲を表す。その上さらなる実施形態では、個別区画内でのバーコード付き断片はゲノムの1以上の選択された部分の少なくとも $1 \times$ の適用範囲を表す。別のさらなる実施形態では、個別区画内でのバーコード付き断片はゲノムの1以上の選択された部分の少なくとも $2 \times$ または $5 \times$ の適用範囲を表す。

30

【0117】

ゲノムの選択された領域から配列情報を入手する能力を提供することに加えて、本明細書に記載されている方法及びシステムは、US 2014/0316,383；2014/0316,398；2014/0316,416；2014/0316,431；2014/0316,447；及び2014/0316,463に詳細に記載されているような、ハプロタイプの位相化、構造上の変動の特定、及びコピー数の変動を特定することを限定しないで含む、ゲノム物質の他の特徴付けも提供することができ、あらゆる目的で、特にゲノム物質の特徴付けに向けられた記述された説明、図面及び作業実施例すべてについてその全体が参照によって本明細書に組み入れられる。

40

【0118】

1つの態様では、且つ本明細書で上述及び後述されている方法のいずれかと併せて、本明細書に記載されている方法及びシステムは、試料核酸またはその断片の個別の区分または区画（好ましくは相互交換可能に区画と呼ばれる）への区分化、配置または区画化を提供し、その際、各区画は他の区画の内容物からのそれ自体の内容物の分離を維持する。独特の識別子、たとえば、バーコードは、特徴、たとえば、核酸の配列情報の特定の区分内

50

に含まれる試料核酸への、特に元々区画に入れられてもよい連続的な試料核酸の相対的に長い区間へのその後の帰属を可能にするために、区分化されたまたは区画化された試料核酸を保持する区画に予め、続いてまたは同時に送達されてもよい。

【0119】

本明細書に記載されている方法で利用される試料核酸は通常、解析される試料全体、たとえば、染色体全体、エキソームまたは他の大きなゲノム部分の多数の重複する部分を表す。これらの試料核酸はゲノム全体、個々の染色体、エキソーム、アンブリコンまたは対象とする様々な異なる核酸を含んでもよい。試料核酸は通常、連続的な核酸分子の相対的に長い断片または区間で区画に存在するように区画化される。通常、試料核酸のこれらの断片は、1 kb より長く、5 kb より長く、10 kb より長く、15 kb より長く、20 kb より長く、30 kb より長く、40 kb より長く、50 kb より長く、60 kb より長く、70 kb より長く、80 kb より長く、90 kb より長くまたはさらに100 kb より長くてもよく、それは、上述されているさらに長距離の分子コンテキストを可能にする。

【0120】

試料核酸はまた通常、それによって所与の区画が出発試料核酸の2つの重複する断片を含む非常に低い確率を有するレベルでも区画化される。これは通常、区画化過程の間に少ない入力量及び/または低い濃度で試料核酸を提供することによって達成される。その結果、好まれる場合では、所与の区画は出発試料核酸の多数の長い、重複しない断片を含んでもよい。次いで異なる区画における試料核酸を独特の識別子に関連付け、その際、任意の所与の区画については、それに含有される核酸は同一の独特の識別子を持つが、異なる区画は異なる独特の識別子を含んでもよい。さらに、区画化ステップは試料成分を非常に小さな容量の区画または液滴に割り当てるので、上記で述べられているような所望の割り当てを達成するために、大きな容量の過程、たとえば、試験管、またはマルチウェルプレートのウェルにて必要とされるような、試料の実質的な希釈を行う必要がない。さらに、本明細書に記載されているシステムはそのような高レベルのバーコード多様性を採用するので、上記で提供されているような多数のゲノム同等物の間で多様なバーコードを割り当てることができる。特に、前述のように、マルチウェルプレートのやり方（たとえば、米国公開出願第2013-0079231号及び第2013-0157870号）は通常、百から数百の異なるバーコード配列で操作するにすぎず、バーコードを異なる細胞/核酸に帰属させ得るためには試料の限界希釈法を採用する。従って、それらは一般に100よりはるかに少ない細胞で作動するであろうし、それは通常、1:10、確実には1:100を優に超える桁でゲノム: (バーコードの種類) の比を提供する。本明細書に記載されているシステムは一方、たとえば10,000、100,000、500,000、600,000、700,000等を超える高レベルのバーコード多様性のゆえに、1:50以下、1:100以下、1:1000以下またはさらに小さな比の桁であるゲノム: (バーコードの種類) の比で操作することができる一方で、ゲノム当たりはるかに改善されたバーコード多様性を依然として提供しながら、多数のゲノム（たとえば、アッセイ当たり100ゲノムを超える、アッセイ当たり500ゲノムを超える、アッセイ当たり1000ゲノムを超える、またはそれ以上の桁で）を負荷することも可能にする。

【0121】

大抵、試料は区画化ステップに先立って遊離可能にビーズに結合されるオリゴヌクレオチドタグのセットと組み合わせられる。核酸にバーコードを付ける方法は当該技術分野で既知であり、本明細書に記載されている。一部の例では、方法は、Amini, et al, 2014, Nature Genetics, Advance Online Publication)に記載されているように利用され、それはあらゆる目的で及び特に核酸にバーコードまたは他のオリゴヌクレオチドタグを結合することに関する教示すべてについてその全体が参照によって本明細書に組み入れられる。さらなる例では、オリゴヌクレオチドは少なくとも第1と第2の領域を含んでもよい。第1の領域は、所与の区画内のオリゴヌクレオチド間に関して実質的に同一のバーコード配列であってもよいが、異

なる区画間に関して異なるバーコード配列であってもよく、ほとんどの場合、異なるバーコード配列であるバーコード領域であってもよい。第2の領域は、区画内での試料内で核酸に刺激を与えるのに使用することができるN量体（無作為N量体または特定の配列を標的とするように設計されたN量体）であってもよい。場合によっては、N量体が特定の配列を標的とするように設計される場合、それは特定の染色体（たとえば、第1、13、18または21染色体）または染色体の領域、たとえば、エキソームまたは他の標的とされる領域を標的とするように設計されてもよい。本明細書で議論されているように、N量体も、不十分にしか特徴付けられない傾向がある、または参照配列から高度に多形であるもしくはは分岐するゲノムの選択された領域であるように設計されてもよい。場合によってはN量体は特定の遺伝子または遺伝領域、たとえば、疾患または障害（たとえば、癌）に関連する遺伝子または領域を標的とするように設計されてもよい。区画内では、第2のN量体を用いて増幅反応を行って核酸の長さの沿って異なる場所で核酸試料に刺激を与えてもよい。増幅の結果、各区画は、同一のまたはほぼ同一のバーコードに結合され、各区画にて核酸の重複するさらに小さい断片を表してもよい、核酸の増幅産物を含有してもよい。バーコードは、核酸のセットが同一区画を起源とし、従って潜在的に核酸の同一鎖も起源としたことを示すマーカーとしての役割を果たすことができる。増幅に続いて、核酸をプールし、配列決定し、配列アルゴリズムを用いて並べてもよい。さらに短い配列読み取りは、それに関連するバーコード配列のおかげで試料核酸の単一の長い断片に対して並べられ、帰属させられるので、その配列上の特定される変異のすべては単一の元の断片及び単一の元の染色体に帰属させることができる。さらに、複数の長い断片にわたって複数の同時に位置する変異を並べることによってその染色体の寄与をさらに特徴付けることができる。従って、特定の遺伝的変異の位相化に関する結論は、ゲノム配列の長距離にわたって解析できるように、たとえば、ゲノムの不十分にしか特徴付けされない領域の区間にわたる配列情報の特定が、次いで引き出されてもよい。そのような情報は、一般に同一核酸鎖または異なる核酸鎖に存在する遺伝的変異の特定されたセットであるハプロタイプを特定するのににも有用であってもよい。コピー数の変動もこの方法で特定されてもよい。

【0122】

記載されている方法及びシステムは、現在の核酸の配列決定技術及びそれに関連する試料調製方法を超える有意な利点を提供する。集団試料調製及び配列決定の方法は、試料における主要な構成成分を主として特定しやすく、特徴付けやすく、且つ、少数構成成分、たとえば、1本の染色体が寄与する、不十分にしか特徴付けられないまたは高度に多形のゲノムの領域に由来する遺伝物質、または抽出された試料にて全DNAの小さな比率を構成する、血流で循環している細胞1つもしくはは数個の細胞、もしくは断片化した腫瘍細胞のDNA分子を特定し、特徴付けるようには設計されていない。本明細書に記載されている方法はこれらの少数構成成分に由来する遺伝物質を増やす選択的増幅法を含み、この遺伝物質の分子コンテキストを保持する能力はこれらの構成成分の遺伝的な特徴付けをさらに提供する。記載されている方法及びシステムはまたさらに大きな試料に存在する集団を検出するための有意な利点も提供する。従って、それらは、ハプロタイプ及びコピー数の変動を評価するのに特に有用であり、本明細書で開示されている方法は、それらの配列が由来する原試料及び元の核酸分子の三次元空間内で互いに空間的に近接して位置した配列についての配列情報を提供するのにも有用である。

【0123】

本明細書で開示されているバーコード付け技法の使用は、ゲノムの配列及び領域についての個々の構造コンテキスト及び分子コンテキストを提供する独特の能力を付与する。ゲノムのそのような領域は遺伝子マーカーの所与のセットを含んでもよく、すなわち、遺伝子マーカー（単一マーカーの対語として）の所与のセットを個々の試料核酸分子に帰属させることを含んで、及び変異が協調する組み立てを介して複数の試料核酸の間で、及び/または特定の染色体に対して、さらに広範囲のまたはさらに長距離の推論された個々の分子コンテキストを提供してもよい。これらの遺伝子マーカーは特定の遺伝子座、たとえば、SNPのような変異体を含んでもよく、またはそれらは短い配列を含んでもよい。さら

に、バーコード付けの使用は、たとえば、血流で循環している腫瘍DNAの検出及び特徴付けのための、試料から抽出された全核酸集団の少数構成成分と主要構成成分との間を区別する、且つまた任意の増幅ステップの間で増幅の偏りを減らすまたは排除する能力を促進する追加の利点を付与する。加えて、マイクロ流体工学の形式の実施は、DNAの極端に少ない容量及び少ない入力量で作業する能力と同様に多数の試料区画（液滴）を処理してゲノム規模でのタグ付けを円滑にする能力を付与する。

【0124】

上記で言及されたように、本明細書に記載されている方法及びシステムはさらに長い核酸の短い配列読み取りについて個々の構造コンテキスト及び分子コンテキストを提供する。本明細書で使用されるとき、構造コンテキストは、原試料内での元の核酸分子の三次元空間内における配列の位置を指す。上述したように、ゲノムは線状であると考えられることが多いが、染色体は剛性ではなく、2つのゲノム遺伝子座間の空間的な距離はゲノムに沿ったその距離とは必ずしも相関せず 線状の配列に沿って幾つかのメガ塩基によって分離されるゲノムの領域は三次元空間では互いにすぐ近接していてもよい。配列読み取りの元の空間的な近接性に関する情報を保持することによって、本明細書に記載されている方法及び組成物は配列読み取りを長距離のゲノム相互作用に帰属させる方法を提供する。

【0125】

同様に、本明細書に記載されている方法によって可能である個々の分子コンテキストの保持は、特定の配列読み取りを越えた配列コンテキスト、たとえば、配列読み取り自体には含まれない隣接するまたは近接する配列との関係を提供し、そのようなものとして、通常、それらは、短い配列読み取り、たとえば、対合読み取りについて約150塩基または300塩基の読み取りに全体でまたは部分的に含まれないようなものであろう。特に好ましい態様では、方法及びシステムは短い配列読み取りについて長距離の配列コンテキストを提供する。そのような長距離のコンテキストには、1kbより長い、5kbより長い、10kbより長い、15kbより長い、20kbより長い、30kbより長い、40kbより長い、50kbより長い、60kbより長い、70kbより長い、80kbより長い、90kbより長いまたはさらに100kbより長い、またはそれ以上長い互いの距離の範囲内である配列読み取りに対する所与の配列読み取りの関係または繋がりが含まれる。さらに長距離の個々の分子コンテキストを提供することによって、本発明の方法及びシステムははるかに長い推論された分子コンテキストも提供する。本明細書に記載されているような配列コンテキストは、たとえば、短い配列読み取りを個々の長い分子または結合した分子のコンティグにマッピングすることからの低い分解能のコンテキストと同様に、たとえば、個々の分子の連続的な決定された配列を有するさらに長い個々の分子の大きな部分の長距離の配列決定からの高い分解能の配列コンテキストを含むことができ、その際、そのような決定された配列は1kbより長い、5kbより長い、10kbより長い、15kbより長い、20kbより長い、30kbより長い、40kbより長い、50kbより長い、60kbより長い、70kbより長い、80kbより長い、90kbより長いまたはさらに100kbより長い。配列コンテキストと同様に、短い配列のさらに長い核酸への帰属、たとえば、個々の長い核酸分子双方または連結した核酸分子もしくはコンティグの収集への帰属は、長い核酸区間に対して短い配列をマッピングして高いレベルの配列コンテキストを提供することと同様にこれらの長い核酸を介して短い配列から組み立てた配列を提供することの双方を含んでもよい。

【0126】

本明細書に記載されている方法、組成物及びシステムは、ゲノムにわたる長距離の相互作用の特徴付けと同様に試料内の関連するタンパク質及び他の分子の特徴付けを可能にする。タンパク質の高レベルの組織化のように、DNA及びクロマチンの折り曲げ及び折り畳みは多種多様な規模で機能的に有意な構造を作り出す。小さな規模では、DNAがヒストンのようなタンパク質の周りに巻き付いてヌクレオソームとして知られる構造を作り出すことが多いことは周知である。これらのヌクレオソームはさらに大きな「クロマチン線維」に詰め込まれ、詰込みパターンは転写のような細胞過程によって影響を受けるとされ

10

20

30

40

50

ている。機能的な構造は大きな規模でも存在し：ゲノムの線状配列の長さの多数のメガ塩基によって分離された領域は3次元空間ではすぐそばで隣接し得る。ゲノム遺伝子座間のそのような長距離の相互作用は機能的な特徴で役割を担ってもよく：たとえば、遺伝子のエンハンサー、サイレンサー及びインシュレーターの要素はすべて広いゲノムの距離にわたって機能してもよく、それらの主要な作用様式には標的遺伝子、非コーディングRNA及び/または調節要素との直接的な物理的関連性が関与し得る。長距離の相互作用は、シスで位置する、すなわち、同一染色体に沿って位置する要素に限定されないが、トランスで位置する、すなわち、異なる染色体に位置するゲノム遺伝子座間でも生じ得る。長距離の相互作用の存在は、相互作用する調節性要素が標的遺伝子から大きなゲノム距離にあり、別の染色体にすらあり得るため、細胞過程を調節する経路を理解する努力を複雑にし得る。癌遺伝子及び他の疾患関連の遺伝子の場合、長距離の遺伝子調節因子の特定は病態及び病態がもたらされる過程に関与する遺伝的変異を特定するのに多用することができる。従って、本明細書に記載されている方法に従って構造コンテキスト及び分子コンテキストを保持する能力は長距離のゲノム相互作用を特定し、同様に関連するタンパク質を特徴付ける方法を提供する。

10

【0127】

本明細書に記載されている方法は、過去のFFPE組織試料を含むFFPE組織試料に由来する核酸の特徴付けに特に有用である。核酸は断片化され、さもなければ分解されることが多く、それは従来する方法を用いて得ることができる情報の量を制限し得るので、FFPE試料は一般に核酸の特徴付けに対して難題を提示する。本明細書に記載されている方法にて保持されている構造コンテキスト及び分子コンテキストの情報は、長距離情報が短い読み取りの配列決定技術を通じてアクセスできるので、コンテキスト情報が分解された試料についてさえ長距離ゲノム相互作用の特徴付けを提供することができるので、独特の機会にそのような試料を提供する。FFPEの核酸の特徴付けの適用は、1以上の過去の試料に由来する配列の、対象、たとえば、癌患者の試料に由来する配列との比較を含んで、診断上または予後診断上の情報を提供する。たとえば、過去試料における1以上の分子マーカーの状態は1以上の治療成果に相関することができ、1以上の過去試料における治療成果の分子マーカーの状態との相関を用いて対象、たとえば、癌患者についての治療成果を予測することができる。これらの予測は対象に対して薬剤治療の選択肢を推奨されるかどうかを決定する基礎であり得る。

20

30

【0128】

V. 試料

理解されるように、本明細書で議論されている方法及びシステムを用いて任意の種類のゲノム物質から配列情報を得ることができる。そのようなゲノム物質は患者から採取された試料から得られてもよい。本明細書で議論されている方法及びシステムで使用される例示的な試料及びゲノム物質の種類には限定しないで、ポリヌクレオチド、核酸、オリゴヌクレオチド、循環している無細胞核酸、循環している腫瘍細胞(CTC)、核酸断片、ヌクレオチド、DNA、RNA、ペプチドポリヌクレオチド、相補性のDNA(cDNA)、二本鎖DNA(dsDNA)、一本鎖DNA(ssDNA)、プラスミドDNA、コスミドDNA、染色体DNA、ゲノムDNA(gDNA)、ウイルスDNA、細菌DNA、mtDNA(ミトコンドリアDNA)、リボソームRNA、無細胞DNA、無細胞胎児性DNA(cffDNA)、mRNA、rRNA、tRNA、nRNA、siRNA、snRNA、snoRNA、scaRNA、マイクロRNA、dsRNA、ウイルスRNA、等が挙げられる。要約すれば、使用される試料は特定の処理の必要性に応じて変化してもよい。

40

【0129】

特定の態様では、本発明で使用する試料には試料分解のリスクが高い他の試料の種類を含めてホルマリン固定し、パラフィン包埋した(FFPE)細胞及び組織の試料等が挙げられる。固定された試料の他の種類には限定しないで、アクロレイン、グリオキサール、四酸化オスミウム、カルボジイミド、塩化第二水銀、亜鉛塩、ピクリン酸、重クロム酸カ

50

リウム、エタノール、メタノール、アセトン及び／または酢酸を用いて固定した試料が挙げられる。

【 0 1 3 0 】

さらなる実施形態では、本明細書に記載されている方法及びシステムで使用される試料は核基質を含む。「核基質」は、核酸とタンパク質を含む任意の組成物を指す。核酸は染色体に構造化されてもよく、その際、タンパク質（すなわち、たとえば、ヒストン）は調節性の機能を有する染色体に関わるようになる。

【 0 1 3 1 】

本明細書で提供されている方法及びシステムは、出発核酸（たとえば、DNA、mRNA等）もしくは出発標的核酸が少量で存在する、または解析について標的とされる核酸が試料内の全核酸のうち相対的に低い比率で存在する核酸の配列決定の適用に特に有用である。1つの態様では、本開示は、入力核酸分子が50ナノグラム（ng）未満の量で存在する核酸を解析する方法を提供する。さらなる実施形態では、核酸分子は40ng未満の入力量である。一部の実施形態では、量は20ng未満である。一部の実施形態では、量は10ng未満である。一部の実施形態では、量は5ng未満である。一部の実施形態では、量は1ng未満である。一部の実施形態では、量は0.1ng未満である。出発入力量が少量である核酸を単離し、解析する方法は、たとえば、2015年6月26日に出版されたUS 9,147,602にてさらに記載されており、それは、あらゆる目的で及び特に核酸が少量で存在する試料に由来する核酸の単離及び特徴付けに関する教示すべてについてその全体が参照によって本明細書に組み入れられる。

【 0 1 3 2 】

理解されるように、本明細書に記載されている方法の間のどこかの任意の時点で当該技術分野で既知の方法を用いて試料を処理することができる。たとえば、区画化に先立って、または試料が個別区画に区画化された後で、試料を処理することができる。

【 0 1 3 3 】

特定の実施形態では、試料を処理して、さらに長い核酸の鎖が確実に保持されるようにする。FFPE試料が使用される実施形態では、そのような試料を処理に供してホルムアルデヒド付加体を取り除き、核酸の収率を改善してもよい。そのような処理方法には非限定の一例では、Karmakar, et al., (2015), Nature Chemistry, DOI: 10.1038/NCHEM.2307に記載されるようなRNA及びDNAの塩基からのホルムアルデヒド付加体の破壊を促進する水溶性有機触媒の使用が挙げられてもよく、その全体が、特にFFPE試料の処置及び処理に関する教示すべてについて参照によって本明細書に組み入れられる。

【 0 1 3 4 】

核酸を含むいかなる物質も試料の供給源であってもよい。物質は流体、たとえば、生体液であってもよい。流体物質には、血液、臍帯血、唾液、尿、汗、血清、精液、膣液、胃液及び消化液、脊髄液、胎盤液、空洞液、眼液、血清、母乳、リンパ液、またはそれらの組み合わせが挙げられてもよいが、これらに限定されない。物質は固形物、たとえば、生体組織であってもよい。物質は、正常で健常な組織、病変組織または健常な組織と病変組織の混合を含んでもよい。場合によっては、物質は腫瘍を含んでもよい。腫瘍は良性（非癌）または悪性（癌）であってもよい。腫瘍の非限定例には、線維肉腫、粘液肉腫、脂肪肉腫、軟骨肉腫、骨肉腫、軟骨腫、血管肉腫、内皮肉腫、リンパ管肉腫、リンパ管内皮肉腫、滑液腫瘍、中皮腫、ユーイング肉腫、平滑筋肉腫、横紋筋肉腫、消化器系の癌、結腸癌、膵臓癌、乳癌、生殖泌尿器系の癌、卵巣癌、前立腺癌、扁平上皮癌、基底細胞癌、腺癌、汗腺癌、皮脂腺癌、乳頭癌、乳頭腺癌、嚢胞腺癌、髄様癌、気管支癌、腎細胞癌、肝癌、胆管癌、絨毛癌、精上皮腫、胚性癌、ウィルム腫瘍、子宮頸癌、内分泌系の癌、精巣腫瘍、肺癌、小細胞肺癌、非小細胞肺癌、膀胱癌、上皮癌、神経膠腫、星状細胞種、髄芽細胞種、頭蓋咽頭腫、上衣細胞種、松果体腫、血管芽細胞種、聴神経腫、乏突起膠腫、髄膜腫、黒色腫、神経芽細胞種、網膜芽細胞種、またはそれら組合せが挙げられてもよい。物質は様々な種類の臓器に関連してもよい。臓器の非限定例には、脳、肝臓、肺、腎臓、

前立腺、卵巣、脾臓、リンパ節（扁桃を含む）、甲状腺、膵臓、心臓、骨格筋、腸管、喉頭、食道、胃、またはそれらの組み合わせが挙げられてもよい。場合によっては、物質は、真核細胞、原核細胞、真菌細胞、心臓細胞、肺細胞、腎臓細胞、肝臓細胞、膵臓細胞、生殖細胞、幹細胞、人工多能性幹細胞、消化器細胞、血液細胞、癌細胞、細菌細胞、ヒト微生物叢試料から単離された細菌細胞等を含むが、これらに限定されない種々の細胞を含んでもよい。場合によっては、物質は細胞の内容物、たとえば、単一細胞の内容物または複数細胞の内容物を含んでもよい。個々の細胞を解析するための方法及びシステムは、たとえば、その完全な開示が全体として参照によって本明細書に組み入れられる2015年6月26日に出版されたUS N 14 / 752, 641にて提供されている。

【0135】

10

試料は種々の対象から得られてもよい。対象は生きている対象であってもよいし、または死んだ対象であってもよい。対象の例には、ヒト、哺乳類、非ヒト哺乳類、齧歯類、両生類、爬虫類、イヌ、ネコ、ウシ、ウマ、ヤギ、ヒツジ、めんどり、鳥類、マウス、ウサギ、昆虫、ナメクジ、微生物、細菌、寄生虫、または魚類が挙げられてもよいが、これらに限定されない。場合によっては、対象は、疾患または障害を有している、有することが疑われる、またはそれを発症するリスクがある患者であってもよい。場合によっては、対象は妊娠女性であってもよい。場合によっては、対象は正常で健常な妊娠女性であってもよい。場合によっては、対象は特定の先天性欠損症を伴う赤ん坊を身ごもるリスクがある妊娠女性であってもよい。

【0136】

20

試料は当該技術分野で既知の手段によって対象から得られてもよい。たとえば、試料は、循環系にアクセスすること（たとえば、注射器または他の装置を介して静脈内または動脈内で）、分泌された生体試料（たとえば、唾液、痰、尿、便等）を回収すること、外科的に生体試料を取得すること（たとえば、術中試料、術後試料等）、スワビング（たとえば、頬側スワブ、口腔咽頭スワブ）、またはピペット操作を介して対象から得られてもよい。

【0137】

VI. 実施形態

一部の態様では、本開示は構造コンテキストを維持しながら核酸を解析する方法を提供する。そのような方法には、（a）核酸が三次元構造を含む核酸を含有する試料を提供するステップと、（b）核酸の三次元構造の一部も個別区画に分離されるように試料の一部を個別区画に分離するステップと、（c）核酸から配列情報を入手し、それによって構造コンテキストを維持しながら核酸を解析するステップとが含まれる。

【0138】

一部の実施形態では、入手するステップ（c）に由来する配列情報には互いに空間的に近接し合う核酸の特定が含まれる。

【0139】

任意の実施形態では、入手するステップ（c）はゲノム遺伝子座間の染色体内の及び／または染色体間の相互作用に関する情報を提供する。

【0140】

40

任意の実施形態では、入手するステップ（c）は染色体の高次構造に関する情報を提供する。

【0141】

任意の実施形態では、分離するステップ（b）に先立って、三次元構造の少なくとも一部が処理されて三次元構造内で互いに近接し合う核酸の異なる部分を連結する。

【0142】

任意の実施形態では、試料はホルマリンで固定されたパラフィン試料である。

【0143】

任意の実施形態では、分離するステップ（b）に先立って、核酸は試料から単離されない。

50

【 0 1 4 4 】

任意の実施形態では、個別区画はビーズを含む。

【 0 1 4 5 】

任意の実施形態では、ビーズはゲルビーズである。

【 0 1 4 6 】

任意の実施形態では、入手するステップ (c) に先立って、個別区画内での核酸にバーコードを付けて複数のバーコード付き断片を形成し、その際、所与の個別区画内での断片はそれぞれ共通のバーコードを含むので、バーコードによって所与の区画に由来する核酸が特定される。

【 0 1 4 7 】

任意の実施形態では、入手するステップ (c) は短い読み取り長さの配列決定反応及び長い読み取り長さの配列決定反応から成る群から選択される配列決定反応を含む。

【 0 1 4 8 】

任意の実施形態では、試料は腫瘍試料を含む。

【 0 1 4 9 】

任意の実施形態では、試料は腫瘍細胞と正常細胞の混合物を含む。

【 0 1 5 0 】

任意の実施形態では、試料は核基質を含む。

【 0 1 5 1 】

任意の実施形態では、核酸は R N A を含む。

【 0 1 5 2 】

任意の実施形態では、試料における核酸の量は、5、10、15、20、25、30、35、40、45、または50 ng / ml 未満である。

【 0 1 5 3 】

一部の態様では、本開示は、(a) 空間的に隣接する核酸セグメントが連結されるように試料内で連結された核酸を形成するステップと、(b) 連結された核酸を処理して複数のライゲーション産物を作り出し、ライゲーション産物が空間的に隣接する核酸セグメントの一部を含有するステップと、(c) 複数のライゲーション産物を個別区画に入れるステップと、(d) 個別区画内のライゲーション産物にバーコードを付けて複数のバーコード付き断片を形成し、その際、所与の個別区画内の断片がそれぞれ共通のバーコードを含み、それによって各断片をそれが送達される連結された核酸に関連付けるステップと、(e) 複数のバーコード付き断片から配列情報を入手し、それによって、構造コンテキストを維持しながら、試料に由来する核酸を解析するステップとを含む、構造コンテキストを維持しながら、核酸を解析する方法を提供する。

【 0 1 5 4 】

さらなる実施形態では、処理するステップ (b) は分子内ライゲーションに都合が良い条件下での平滑末端ライゲーションを含むので、空間的に隣接する核酸セグメントは同じ分子内でライゲーションされる。

【 0 1 5 5 】

任意の実施形態では、分子内ライゲーションに都合が良い条件は試料を希釈して核酸の濃度を10 ng / μ l 未満に低下させることを含む。

【 0 1 5 6 】

任意の実施形態では、ステップ (a) に先立って核酸は試料から単離されない。

【 0 1 5 7 】

任意の実施形態では、形成するステップ (a) に先立って、関連する D N A 結合タンパク質が核酸に結合したままであるように核酸を免疫沈降させる。

【 0 1 5 8 】

任意の実施形態では、区画はビーズを含む。

【 0 1 5 9 】

任意の実施形態では、ビーズはゲルビーズである。

10

20

30

40

50

【 0 1 6 0 】

任意の実施形態では、試料は腫瘍試料を含む。

【 0 1 6 1 】

任意の実施形態では、試料は腫瘍細胞と正常細胞の混合物を含む。

【 0 1 6 2 】

任意の実施形態では、処理するステップには、ライゲーション産物を形成することによく連結の破棄が含まれる。

【 0 1 6 3 】

任意の実施形態では、入手するステップ (e) はゲノムの遺伝子座間の染色体内の及び / または染色体間の相互作用に関する情報を提供する。

10

【 0 1 6 4 】

任意の実施形態では、入手するステップ (e) は染色体の高次構造に関する情報を提供する。

【 0 1 6 5 】

任意の実施形態では、染色体の高次構造は病態に関連する。

【 0 1 6 6 】

任意の実施形態では、処理するステップは、試料にて元々空間的に密に近接し合っていた核酸を含むライゲーション産物を生じる。

【 0 1 6 7 】

任意の実施形態では、入手するステップ (e) は短い読み取り長さの配列決定反応及び長い読み取り長さの配列決定反応から成る群から選択される配列決定反応を含む。

20

【 0 1 6 8 】

任意の実施形態では、配列決定反応は短い読み取りで精度の高い配列決定反応である。

【 0 1 6 9 】

任意の実施形態では、形成するステップ (a) は試料にて核酸を架橋することを含む。

【 0 1 7 0 】

任意の実施形態では、形成するステップ (a) は空間的に隣接する核酸セグメント間で共有結合を生じる。

【 0 1 7 1 】

一部の態様では、本開示は、 (a) 空間的に隣接する核酸セグメントが連結されるように試料内で連結された核酸を形成するステップと、 (b) 連結された核酸を個別区画に入れるステップと、 (c) 連結された核酸を処理して複数のライゲーション産物を作り出し、ライゲーション産物は空間的に隣接する核酸セグメントの一部を含有するステップと、 (d) 個別区画内でのライゲーション産物にバーコードを付けて複数のバーコード付き断片を形成し、その際、所与の個別区画内での断片はそれぞれ共通のバーコードを含み、それによって各断片をそれが由来する連結された核酸に関連付けるステップと、 (e) 複数のバーコード付き断片から配列情報入手し、それによって、構造コンテキストを維持しながら核酸を解析するステップとを含む、構造コンテキストを維持しながら核酸を解析する方法を提供する。

30

【 0 1 7 2 】

さらなる実施形態では、処理するステップ (c) は分子内ライゲーションに都合が良い条件下での平滑末端ライゲーションを含むので、空間的に隣接する核酸セグメントは同じ分子内でライゲーションされる。

40

【 0 1 7 3 】

任意の実施形態では、試料はホルマリン固定したパラフィン試料である。

【 0 1 7 4 】

任意の実施形態では、試料は核基質を含む。

【 0 1 7 5 】

任意の実施形態では、核酸は R N A を含む。

【 0 1 7 6 】

50

任意の実施形態では、核酸はステップ (a) に先立って試料から単離されない。

【 0 1 7 7 】

任意の実施形態では、形成するステップ (a) に先立って、関連する D N A 結合タンパク質が核酸に結合したままであるように核酸を免疫沈降させる。

【 0 1 7 8 】

任意の実施形態では、区画はビーズを含む。

【 0 1 7 9 】

任意の実施形態では、ビーズはゲルビーズである。

【 0 1 8 0 】

任意の実施形態では、試料は腫瘍試料を含む。

10

【 0 1 8 1 】

任意の実施形態では、試料は腫瘍細胞と正常細胞の混合物を含む。

【 0 1 8 2 】

任意の実施形態では、処理するステップ (c) は試料にて元々空間的に密に近接し合っていた核酸を含むライゲーション産物を生じる。

【 0 1 8 3 】

任意の実施形態では、入手するステップ (e) はゲノムの遺伝子座間の染色体内の及び / または染色体間の相互作用に関する情報を提供する。

【 0 1 8 4 】

任意の実施形態では、入手するステップ (e) は短い読み取り長さの配列決定反応及び長い読み取り長さの配列決定反応から成る群から選択される配列決定反応を含む。

20

【 0 1 8 5 】

任意の実施形態では、配列決定反応は短い読み取りで精度の高い配列決定反応である。

【 0 1 8 6 】

一部の態様では、本開示は、 (a) 試料内で核酸を架橋して架橋された核酸を形成し、架橋が空間的に隣接する核酸セグメント間で共有結合を形成するステップと、 (b) 架橋された核酸を個別区画に入れるステップと、 (c) 架橋された核酸を処理して複数のライゲーション産物を作り出し、ライゲーション産物が空間的に隣接する核酸セグメントの一部を含有するステップと、 (d) 複数のライゲーション産物から配列情報を入手し、それによって、構造コンテキストを維持しながら試料に由来する核酸を解析するステップとを含む、構造コンテキストを維持しながら核酸を解析する方法を提供する。

30

【 0 1 8 7 】

さらなる実施形態では、処理するステップ (b) は分子内ライゲーションに都合が良い条件下での平滑末端ライゲーションを含むので、空間的に隣接する核酸セグメントは同じ分子内でライゲーションされる。

【 0 1 8 8 】

任意の実施形態では、試料はホルマリン固定したパラフィン試料である。

【 0 1 8 9 】

任意の実施形態では、試料は核基質を含む。

【 0 1 9 0 】

任意の実施形態では、核酸は R N A を含む。

40

【 0 1 9 1 】

任意の実施形態では、架橋するステップ (a) に先立って核酸は試料から単離されない。

【 0 1 9 2 】

任意の実施形態では、試料における核酸の量は、 5 、 1 0 、 1 5 、 2 0 、 2 5 、 3 0 、 3 5 、 4 0 、 4 5 、 または 5 0 n g / m l 未満である。

【 0 1 9 3 】

任意の実施形態では、架橋するステップ (a) に先立って、関連する D N A 結合タンパク質が核酸に結合したままであるように核酸を免疫沈降させる。

50

【0194】

任意の実施形態では、入手するステップ(d)に先立って、ライゲーション産物をバーコードに関連付ける。

【0195】

任意の実施形態では、同一区画内のライゲーション産物は共通のバーコードを受け取るので、バーコードによって所与の区画に由来するライゲーション産物が特定される。

【0196】

任意の実施形態では、入手するステップ(d)は短い読み取り長さの配列決定反応及び長い読み取り長さの配列決定反応から成る群から選択される配列決定反応を含む。

【0197】

本発明の好まれる実施形態が本明細書で示され、記載されてきたが、そのような実施形態は例示のみの目的で提供されていることは当業者に明白であろう。本発明から逸脱することなく多数の変動、変化及び置き換えが今や当業者に発生するであろう。本明細書に記載されている本発明の実施形態に対する種々の代替が本発明を实践することにおいて採用されてもよいことが理解されるべきである。以下のクレームが本発明の範囲を定義し、これらのクレームの範囲内での方法及び構造ならびにそれらの同等物はそれによって網羅されることが意図される。

【実施例】

【0198】

実施例1：試料の調製

試料の調製方法を改変してFFPE試料から長いDNA分子を提供した。図7は例示的な作業の流れを説明しており、ゲノム全体の配列決定(WGS)及びエキソーム全体の配列決定(WES)双方のためのFFPE試料を調製するための改変が示されている。たとえば、DNA抽出の後、標準の熱サイクルプロトコールは701で改変されて98度の変性ステップを各サイクルの終了から開始に移した。加えて、70度の保持を各サイクルの終了時に2分間加えた。

【0199】

サイクル後の精製702及びWESライブラリの調製及び標的濃縮ステップ704と705の間、正常なプロトコール全体にわたって1.8×固相可逆性不働化(SPRI)ビーズを使用した。

【0200】

別の改変は剪断ステップ703の間に条件を変えることを含み、その際、50のピーク入射電力を持つ標準的な超音波破碎装置とは対照的に、約450のピーク入射電力を持つ超音波破碎装置を使用した。

【0201】

特定の状況で使用されてもよい追加の改変は、たとえば、Karmakar, et al., (2015), Nature Chemistry, DOI: 10.1038/NCHEM.2307に記載されるように、ホルムアルデヒド付加体を取り除くためにまず、FFPE試料を有機触媒で処理することである。そのようなプロトコールには30mM pH7のトリス緩衝液中の5mMの有機触媒を試料に加えて付加体の破棄を達成することが含まれる。有効な有機触媒には、限定しないで、たとえば、Karmakar, et al.に記載されるアントラニレート及びホスファニレート触媒のような水溶性二官能性の触媒が挙げられる。付加体の破棄は試料からの核酸収量の収率を改善する効果を有する。

【0202】

実施例2：FFPE試料のバーコード付け

FFPE試料(スライド上のFFPE試料を含むことができる)は、たとえば、DNAのマイクロアレイプリンティングで使用されるもののような空間的に十分に定義されたパターンで適用されるDNAバーコードでタグ付けすることができる。DNAバーコード(以後バーコード1と呼ぶ)は、それがその後のステップで拡散してしまわないように、ま

10

20

30

40

50

たは F F P E 試料に共有結合で適用されるように長い。バーコード付け D N A が F F P E スライドに埋め込まれるのを可能にするために、試料を加熱し、次いでバーコードを加える。異なるバーコードがスライドの異なる部分で提供されるように、バーコードは一般にバーコードのライブラリである。バーコードを異なる集団でスライドの異なる部分に加えて地理的コード化に役立ててもよく、その状況では、バーコードのライブラリは同一のバーコードまたは異なるバーコードを含んでもよい。バーコードを加えた後、次いでスライドを冷却し、一般にたとえば、レーザー微細切断、機械的 / 音響的手段等を用いた方法での切断を介して小分けする。バーコードの代わりに蛍光色素分子または Q d o t s を使用してもよいが、バーコード付けは、局所の空間的情報（たとえば、腫瘍対正常細胞）を保持しながら、試料部分の大規模に平行な無作為封入を可能にする。

10

【 0 2 0 3 】

次いでバーコードを含有する試料の部分を、たとえば、1 0 X G e n o m i c s C h r o m i u m (商 標) システムのような液滴に基づくシステムを含む配列決定システムに入れることができるので、単一のバーコード付き部分が液滴ごとに封入される。

【 0 2 0 4 】

加熱することによって液滴にて試料の脱パラフィン化を行うことができる。パラフィン は水とは非混和性であるが、特定の油には可溶性であるので、チップ上の液滴を加熱する際、液滴からパラフィンを容易に取り除くことができる。液体 / 液体の抽出過程でキシレンを使用して試料部分を脱パラフィン化し、さらなる処理のために核酸内容を準備してもよい。

20

【 0 2 0 5 】

さらなるステップには脱パラフィン化した試料のメチレン架橋を脱架橋することが含まれる。このステップについては、特殊化した化学的手段を用いて架橋を外すことができ、それによって、本明細書で議論されている核酸のバーコード付け、増幅及びライブラリ調製のステップを含むその後の処理のために含有される核酸へのアクセスを可能にする（たとえば、図 2 を参照のこと）。空間的バーコード付き D N A も液滴に封入されることに留意のこと。個々の核酸の第 2 のバーコード付けステップは、空間的に試料をコードするのに使用される核酸とバーコードとにバーコードを付けるのに役立つであろう。次いで配列読み取りと一緒にまとめて、試料における元の空間的位置とその後比較することができるので病理データに関連付けることができる情報を提供することができる。

30

【 0 2 0 6 】

この空間的なコード化の作業の流れの代替の型では、脱架橋するステップが先ず、液滴の中で実施され、次いでゲノム D N A と同様に空間的コード化のバーコードを含む試料における核酸が粒子に結合され、またはさもなければ試料から単離される。核酸は次いで再封入され、図 2 で描かれたものを含む本明細書に記載されている方法におけるバーコード付け及び配列決定の作業の流れに供される。

【 0 2 0 7 】

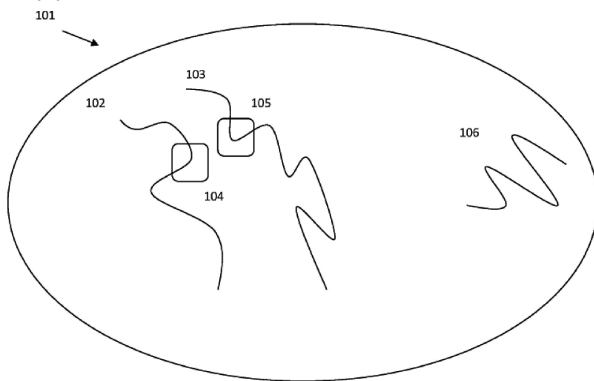
本明細書は、現在記載されている技術の例示態様における方法論、システム及び / または構造及びその使用の完全な説明を提供している。この技術の種々の態様がある程度の特殊性と共にまたは 1 以上の個々の態様を参照して上に記載されているが、当業者は、この技術の精神及び範囲から逸脱することなく、開示された態様に対して多数の変更を行ってもよい。多数の態様は現在記載されている技術の精神及び範囲から逸脱することなく行うことができるので、以後に添付されるクレームには適当な範囲が存在する。従って他の態様が熟考される。さらに、明白に請求されない限りまたは特定の順序がクレーム言語によって本質的に必要とされない限り、任意の操作は任意の順序で行われてもよいことが理解されるべきである。上記の説明に含有され、添付の図面で示される事柄はすべて特定の態様の説明に役立つとしてのみ解釈されるべきであり、示される実施形態に限定されるものではない。文脈から明瞭でない限り、または明白に述べられない限り、本明細書で提供されている濃度値は一般に、混合物の特定の成分の添加の際、または添加ののち生じる変換を考慮することなく混合物の値または比率という点で与えられる。すでに明白に本明細書

40

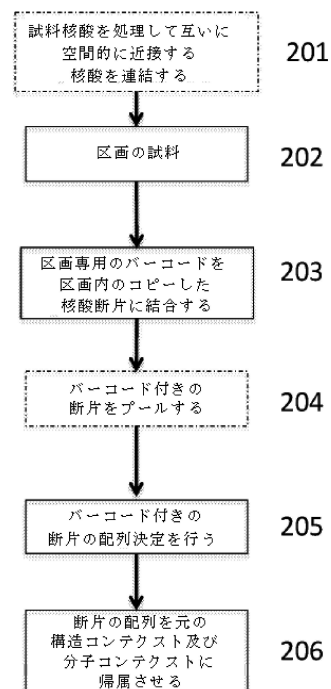
50

に組み込まれていない範囲について、本開示で引用されているすべての公開された参考文献及び特許文書はあらゆる目的でその全体が参照によって本明細書に組み入れられる。詳細または構造における変更は、以下のクレームで定義されるような本技術の基本要素から逸脱することなく行われてもよい。

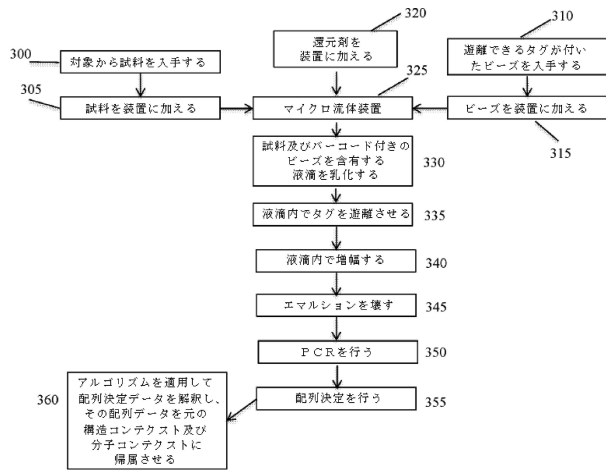
【図 1】



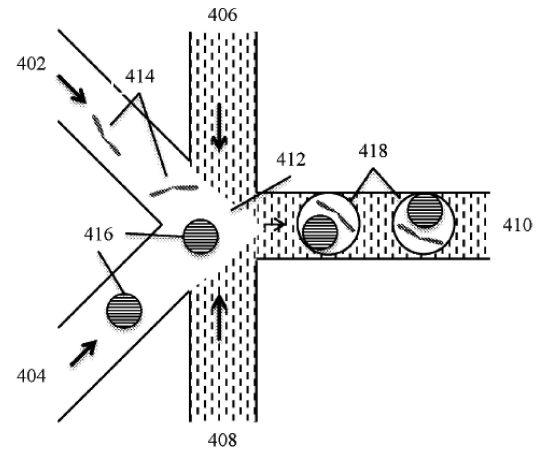
【図 2】



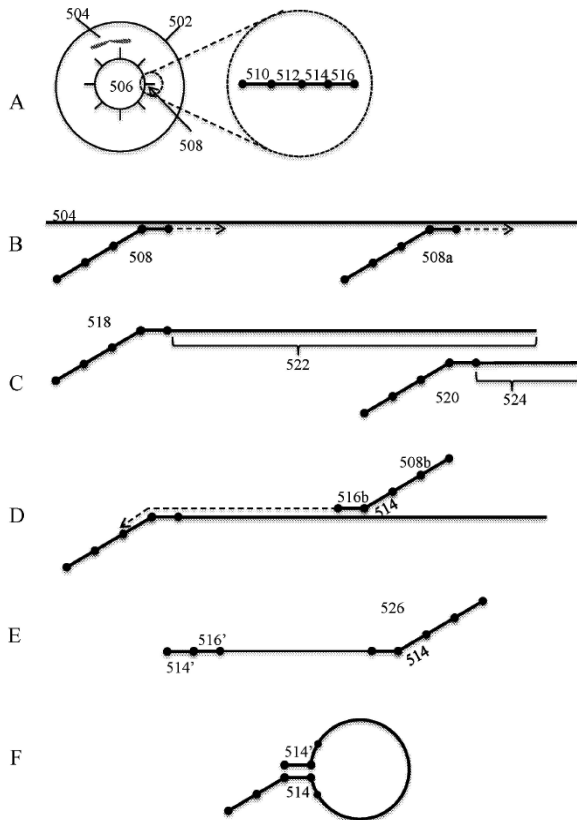
【図 3】



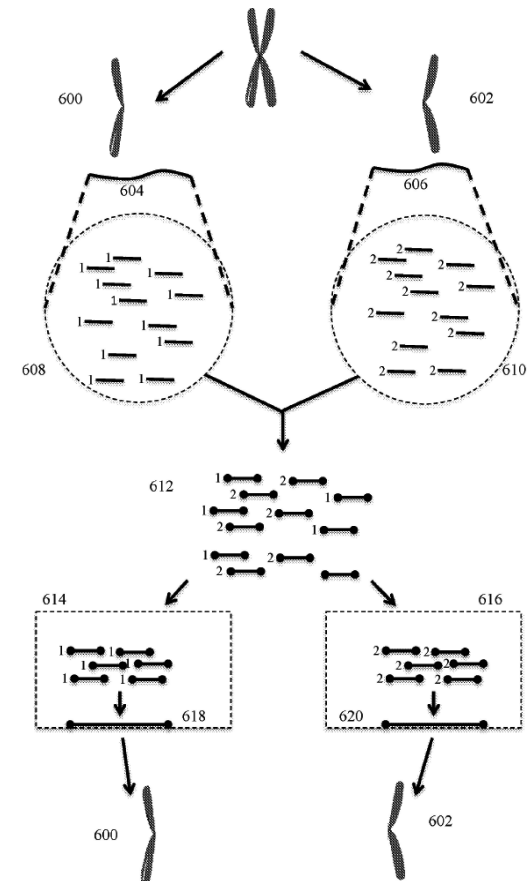
【図 4】



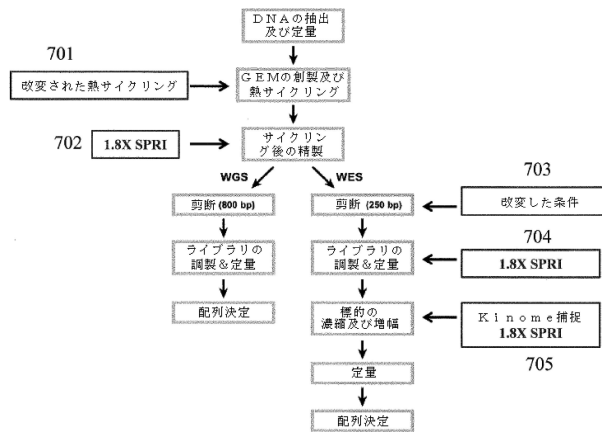
【図 5】



【図 6】



【図 7】



フロントページの続き

- (72)発明者 ジョン, シンイン
アメリカ合衆国, カリフォルニア州 9 4 5 6 6 , プレザントン, コール センター パークウェイ 7 0 6 8 , スイート 4 0 1
- (72)発明者 サクソノブ, サージ
アメリカ合衆国, カリフォルニア州 9 4 5 6 6 , プレザントン, コール センター パークウェイ 7 0 6 8 , スイート 4 0 1
- (72)発明者 シュナル-レヴィン, マイケル
アメリカ合衆国, カリフォルニア州 9 4 5 6 6 , プレザントン, コール センター パークウェイ 7 0 6 8 , スイート 4 0 1
- (72)発明者 ネス, ケヴィン
アメリカ合衆国, カリフォルニア州 9 4 5 6 6 , プレザントン, コール センター パークウェイ 7 0 6 8 , スイート 4 0 1
- (72)発明者 バーラドワージ, ラジーヴ
アメリカ合衆国, カリフォルニア州 9 4 5 6 6 , プレザントン, コール センター パークウェイ 7 0 6 8 , スイート 4 0 1

審査官 中野 あい

- (56)参考文献 米国特許出願公開第2012/0220494 (US, A1)
特表2014-512826 (JP, A)
特表2015-528283 (JP, A)
国際公開第2013/134261 (WO, A1)
国際公開第2015/071748 (WO, A1)

- (58)調査した分野(Int.Cl., DB名)
C12Q 1/00 - 3/00
JSTPlus/JMEDPlus/JST7580(JDreamIII)
CAplus/MEDLINE/EMBASE/BIOSIS(STN)