

(19)日本国特許庁(JP)

(12)特許公報(B2)

(11)特許番号
特許第7497976号
(P7497976)

(45)発行日 令和6年6月11日(2024.6.11)

(24)登録日 令和6年6月3日(2024.6.3)

(51)国際特許分類		F I		
C 1 2 Q	1/6806(2018.01)	C 1 2 Q	1/6806	Z
C 1 2 Q	1/6869(2018.01)	C 1 2 Q	1/6869	Z
C 1 2 Q	1/6876(2018.01)	C 1 2 Q	1/6876	Z
G 0 1 N	33/50 (2006.01)	G 0 1 N	33/50	P
C 1 2 N	15/10 (2006.01)	C 1 2 N	15/10	Z
請求項の数 12 (全82頁) 最終頁に続く				
(21)出願番号	特願2019-511839(P2019-511839)	(73)特許権者	518399195	
(86)(22)出願日	平成29年5月12日(2017.5.12)		ダブテイル ゲノミクス エルエルシー	
(65)公表番号	特表2019-522487(P2019-522487 A)		アメリカ合衆国 9 5 0 6 6 カリフォルニア州 スコッツバレー エンタープライズ・ウェイ 1 0 0 スイート エー 1 0 1	
(43)公表日	令和1年8月15日(2019.8.15)	(74)代理人	100082072	
(86)国際出願番号	PCT/US2017/032466		弁理士 清原 義博	
(87)国際公開番号	WO2017/197300	(74)復代理人	110003797	
(87)国際公開日	平成29年11月16日(2017.11.16)		弁理士法人清原国際特許事務所	
審査請求日	令和2年5月8日(2020.5.8)	(72)発明者	トロル, クリストファー ジョン	
審査番号	不服2022-4157(P2022-4157/J1)		アメリカ合衆国 9 5 0 6 5 カリフォルニア州 サンタ・クルーズ ルーニー・ストリート 6 3	
審査請求日	令和4年3月18日(2022.3.18)	(72)発明者	パワーズ, マーティン ピー .	
(31)優先権主張番号	62/336,252		アメリカ合衆国 9 4 1 1 6 カリフォルニア州	
(32)優先日	平成28年5月13日(2016.5.13)			最終頁に続く
(33)優先権主張国・地域又は機関	米国(US)			
最終頁に続く				

(54)【発明の名称】 保存されたサンプルからの長距離連鎖情報の回復

(57)【特許請求の範囲】

【請求項 1】

第 1 の DNA セグメント及び第 2 の DNA セグメントがリン酸ジエステル骨格から独立して共に保持されるように、ならびに、第 1 の DNA セグメント及び第 2 の DNA セグメントがそれぞれ、少なくとも 1 つの暴露された核酸末端を有するように、タンパク質 DNA 複合体を含む保存されたサンプルを、タンパク質 DNA 複合体が破壊又は分裂されないように、40 以下の温度でプロテイナーゼを用いて処理する工程、及び第 1 の DNA セグメントと第 2 の DNA セグメントをライゲートすることにより、第 1 の DNA セグメントの暴露された核酸末端及び第 2 の DNA セグメントの暴露された核酸末端にタグを付け、共通のタンパク質 DNA 複合体を示す情報を生成することによりゲノムの構造情報を引き出す工程を含む、

10

ここで、前記保存されたサンプルは、ホルマリン固定パラフィン包埋 (FFPE) サンプルであることを特徴とする方法。

【請求項 2】

前記保存されたサンプルは架橋される、ことを特徴とする請求項 1 に記載の方法。

【請求項 3】

前記保存されたサンプルは、ホルムアルデヒド、ホルマリン、UV 光、マイトマイシン C、ナイトロジェンマスタード、メルファラン、1, 3 - ブタジエンジエポキシド、シスジアミンジクロロプラチナム (II)、及びシクロホスファミドのうち少なくとも 1 つを

20

使用して架橋される、ことを特徴とする請求項 2 に記載の方法。

【請求項 4】

前記保存されたサンプルはその中の核酸に関する位置情報を維持する、ことを特徴とする請求項 1 に記載の方法。

【請求項 5】

前記ゲノムの構造情報は、基準ゲノムに対する逆位、挿入、欠失、及び転座のうち少なくとも 1 つを示す、ことを特徴とする請求項 1 に記載の方法。

【請求項 6】

前記ゲノムの構造情報が、核酸の前記第 1 の DNA セグメント及び前記第 2 の DNA セグメントに関するフェーズ状況を示す情報を含む、請求項 1 に記載の方法。

10

【請求項 7】

前記処理する工程は、前記保存されたサンプルをキシレンとエタノールのうち少なくとも 1 つに接触させることを含む、ことを特徴とする請求項 1 に記載の方法。

【請求項 8】

前記処理する工程は、前記保存されたサンプルをアントラニル酸塩とホスファニル酸塩のうち少なくとも 1 つに接触させることを含む、ことを特徴とする請求項 1 に記載の方法。

【請求項 9】

前記保存されたサンプルは組織中の核酸の位置情報を保存する、ことを特徴とする請求項 1 に記載の方法。

【請求項 10】

20

前記保存されたサンプルは核酸を単離させる前に均質化されない、ことを特徴とする請求項 1 に記載の方法。

【請求項 11】

前記保存されたサンプルは核酸を単離させる前に少なくとも 1 週間保管される、ことを特徴とする請求項 1 に記載の方法。

【請求項 12】

前記保存されたサンプルは核酸を単離させる前に少なくとも 6 か月間保管される、ことを特徴とする請求項 1 に記載の方法。

【発明の詳細な説明】

【技術分野】

30

【0001】

相互参照

本出願は、2016 年 5 月 13 日出願の米国仮特許出願第 62 / 336 , 252 号の利益を主張するものであり、該仮出願は、その全体において引用により本明細書に組み込まれる。更に本出願は、2016 年 10 月 20 日出願の米国仮特許出願第 62 / 410 , 599 号の利益を主張するものであり、該仮出願は全体において引用により明細書に組み込まれる。

【背景技術】

【0002】

高品質であり高度に連続するゲノム配列を生成することは、理論上、及び現実面で困難なままである。ゲノム配列の再生を試み、フェージング情報、或いは他の遺伝子情報が、ホルマリン固定パラフィン包埋 (FFPE) サンプルなどの保存されたサンプルから所望されると、この問題は悪化する。FFPE サンプルは、最も一般的なバンク保存された (banked) 臨床的且つ癌のサンプル型である。しかし、固定及び包埋の工程の他、脱水及び長期保管などの付加的な要因も、DNA の損傷を引き起こすと考えられる。付加的な DNA の損傷及び断片化は、DNA 抽出手順の間に生じる場合があり、この手順は頻繁に、架橋を逆転するために一晚のプロテイナーゼ K 処理及び沸騰を含んでいる。抽出後の典型的な DNA 断片長は、500 未満の塩基対、頻繁には 300 未満の塩基対である。

40

【発明の概要】

【0003】

50

本明細書には、外科的切除に準じて保管されたサンプル又は治験に準じて記録されたサンプルなどの、保存されたサンプルからゲノムの構造情報を得る方法が提供される。幾つかのそのような方法は、被験体から、核酸を含む保存されたサンプルを得る工程；及びサンプル中の核酸を分析することによりゲノムの構造情報を引き出す工程を含む。場合によっては、保存されたサンプルは、例えばホルムアルデヒド、ホルマリン、UV光、マイトマイシンC、ナイトロジェンマスタード、メルファラン、1, 3 - ブタジエンジエポキシド、シスジアミンジクロロプラチナム (II)、及びシクロホスファミドのうち少なくとも1つを使用して架橋される。代替的に、保存されたサンプルはホルマリンを使用して架橋される。頻繁に、保存されたサンプルはその中の核酸に関する位置情報を維持する。随意に、保存されたサンプルは、ホルマリン固定パラフィン包埋 (FFPE) サンプルなどの、包埋されたサンプルである。ゲノムの構造情報は、サンプルのゲノムに存在する場合、基準ゲノムに対する逆位、挿入、欠失、及び転座のうち少なくとも1つを示すのに十分なものである。多くの基準ゲノムは、被験体に共通する種の野生型ゲノム、又は被験体の基準組織から得られたゲノムなどの、本明細書中の開示に一致している。方法は頻繁に、核酸の第1のセグメント及び第2のセグメントに関するフェーズ状況を示す情報を引き出す工程を含む。随意に、前記方法は、物理的連鎖情報を伝えるようにサンプルの暴露された核酸末端にタグを付ける工程を含む。場合によっては、タグを付ける工程は、オリゴヌクレオチドが共通の複合体を示す情報を伝えるように、保存されたサンプルから放たれたDNAタンパク質複合体にオリゴヌクレオチドをライゲートする工程を含む。オリゴヌクレオチドは、複合体に特異的な、又は複合体に特有の塩基配列を含む。代替的に、好ましい実施形態において、タグを付ける工程は、ペアエンド分子を形成するために複合体の第2のセグメントに複合体の第1の核酸セグメントをライゲートする工程を含む。これらの場合、幾つかの方法は、第1の核酸セグメントの一部及び第2の核酸セグメントの一部を配列決定する工程を含む。第1の核酸セグメントの一部に共通するユニーク配列を有するコンティグ、及びに第2の核酸セグメントの一部に共通するユニーク配列を有するコンティグは、核酸アセンブリにおける共通の足場に割り当てられる。幾つかの方法は、蛍光プローブであり又は増幅を支持することが可能であり、且つゲノム構造再編成に係する第1の遺伝子座及び第2の遺伝子座にアニールする、抗体又は核酸プローブなどの、1セットのプローブにペアエンド核酸分子を接触させる工程を含む。頻繁に、第1の遺伝子座及び第2の遺伝子座は、ゲノム構造再編成により影響を受けないゲノムにおいて隣接しない。代替的に、第1の遺伝子座及び第2の遺伝子座は、ゲノム構造再編成により影響を受けないゲノムにおいて隣接する。随意に、前記方法は、1セットのプローブの接触が再編成を示す時にサンプルの核酸を配列決定する工程を含む。幾つかの方法は、核酸プライマーを含む1セットのプローブにペアエンド核酸分子を接触させる工程を含む。場合によっては、1セットの核酸プライマーは、ゲノム構造再編成に係する、第1の遺伝子座及び第2の遺伝子座へとアニールする。これらの場合、第1の遺伝子座及び第2の遺伝子座がライゲートされたペアエンド分子を形成すると、1セットの核酸プライマーは核酸増幅反応においてアンプリコンをもたらす。同様に、場合によっては、第1の遺伝子座及び第2の遺伝子座がライゲートされたペアエンド分子を形成しないと、1セットの核酸プライマーは核酸増幅反応においてアンプリコンをもたらさない。場合によっては、第1の遺伝子座及び第2の遺伝子座は、ゲノム構造再編成により影響を受けないゲノムにおいて隣接しない。代替的に、第1の遺伝子座及び第2の遺伝子座は、ゲノム構造再編成により影響を受けないゲノムにおいて隣接する。幾つかの実施形態は随意に、ペアエンド核酸分子に接触される1セットの核酸プライマーのセットからアンプリコンが生成される時に、サンプルの核酸を配列決定する工程を含む。好ましくは、保存された組織サンプルは、タンパク質DNA複合体が破壊されないように核酸を単離するために処理される。場合によっては、タンパク質DNA複合体は、第1の核酸セグメント及び第2の核酸セグメントがリン酸ジエステル骨格から独立して共に保持されるように単離される。場合によっては、保存された組織サンプルは、保存された組織サンプルをキシレンに接触させることにより処理される。場合によっては、保存された組織サンプルは、保存された組織サンプルをエタノール

10

20

30

40

50

に接触させることにより処理される。場合によっては、保存された組織サンプルは、沸騰状態からサンプルを保護することにより処理される。場合によっては、保存されたサンプルは、保存された組織サンプルをアントラニル酸塩とホスファニル酸塩のうち少なくとも1つに接触させることにより処理される。場合によっては、保存された組織サンプルは40以下の温度で処理される。随意に、DNAタンパク質複合体はクロマチンを含む。場合によっては、保存された組織サンプルは組織中のその配置を反映する位置情報を保存する。頻繁に、保存された組織サンプルは、サンプルから切り取られたDNAタンパク質複合体の位置情報が保存され且つゲノム構造分析の一部として利用可能となるように、保存中又は核酸の単離の前には均質化されない。場合によっては、保存された組織サンプルは核酸を単離させる前に少なくとも1週間保管される。場合によっては、保存された組織サンプルは核酸を単離させる前に少なくとも6か月間保管される。場合によっては、保存された組織サンプルは核酸を単離させる前に収集点から輸送される。場合によっては、保存された組織サンプルは無菌環境で集められる。場合によっては、保存された組織サンプルは核酸を単離させる前に非無菌環境に位置付けられる。

【0004】

本明細書には、架橋されたパラフィン包埋組織サンプルなどの保存されたサンプルからゲノムの構造情報などの長距離配列情報を得る方法が提供される。幾つかのそのような方法は：タンパク質DNA複合体が破壊又は分裂されないように、架橋されたパラフィン包埋組織サンプルから核酸を単離させる工程；第1のDNAセグメント及び第2のDNAセグメントが共通のタンパク質DNA複合体から生じると識別されるように、タンパク質DNA複合体にタグを付ける工程；共通のDNA複合体から第1のDNAセグメント及び第2のDNAセグメントを分離する工程；第1のDNAセグメント及び第2のDNAセグメントから配列情報を生成する工程；及び共通のタンパク質DNA複合体を示すタグ配列を共有する配列情報を共通のゲノムの構造に割り当てる工程を含む。場合によっては、架橋されたパラフィン包埋組織は、核酸を単離させる前に均質化されない。場合によっては、タグ配列は、複合体を識別するオリゴタグを含む。場合によっては、タグ配列は、第1のセグメントから第2のセグメントライゲーションにより生じる。場合によっては、タンパク質DNA複合体が破壊又は分裂されないように、架橋されたパラフィン包埋組織サンプルなどの保存されたサンプルから核酸を単離させる工程は、架橋されたパラフィン包埋組織サンプルをキシレンに接触させる工程を含む。場合によっては、タンパク質DNA複合体が破壊又は分裂されないように、架橋されたパラフィン包埋組織サンプルなどの保存されたサンプルから核酸を単離させる工程は、架橋されたパラフィン包埋組織サンプルをエタノールに接触させる工程を含む。場合によっては、タンパク質DNA複合体が破壊又は分裂されないように、架橋されたパラフィン包埋組織サンプルなどの保存されたサンプルから核酸を単離させる工程は、架橋されたパラフィン包埋組織サンプルをエタノールに接触させる工程を含む。場合によっては、タンパク質DNA複合体が分裂されないように、架橋されたパラフィン包埋組織サンプルなどの保存されたサンプルから核酸を単離させる工程は、沸騰状態からサンプルを保護する工程を含む。場合によっては、共通のDNA複合体から第1のDNAセグメント及び第2のDNAセグメントを分離する工程は、プロテイナーゼK処理を含む。抽出プロセスは随意に、抽出プロセス中の任意の架橋剤の添加を含まない。正しくは、サンプル保存に準拠して生成された複合体は、保存されたサンプル中の核酸を潜在的に害する架橋への暴露の数を最小限にすることなどに、依存される。代替的に、核酸は単離され、橋架剤は核酸単離及びクロマチン再構成の後にのみ添加される。

【0005】

本明細書には、架橋されたパラフィン包埋組織サンプルなどの保存されたサンプルからゲノムの構造情報などの長距離配列情報を得る方法が提供される。幾つかのそのような方法は：50kbを超える核酸断片が再生されるように、架橋されたパラフィン包埋組織サンプルから核酸を単離させる工程；核酸分子の第1のDNAセグメント及び第2のDNAセグメントがそれら共通のリン酸ジエステル骨格から独立して共に保持されるように、少なくとも1つの複合体を形成するために複数の核酸結合部分に核酸を接触させる工程；少

10

20

30

40

50

なくとも1つの複合体の少なくとも1つのリン酸ジエステル骨格を切断する工程；第1のDNAセグメント及び第2のDNAセグメントが共通の複合体から生じると識別されるように、少なくとも1つの複合体にタグを付ける工程；共通の複合体から第1のDNAセグメント及び第2のDNAセグメントを分離する工程；第1のDNAセグメント及び第2のDNAセグメントから配列情報を生成する工程；及び共通のタンパク質DNA複合体を示すタグ配列を共有する配列情報を共通のゲノムの構造に割り当てる工程を含む。場合によっては、架橋されたパラフィン包埋組織サンプルは、核酸を単離させる前に均質化されない。場合によっては、タグ配列は、複合体を識別するオリゴタグを含む。場合によっては、タグ配列は、第1のDNAセグメントのライゲートから第2のDNAセグメントまで生じる。場合によっては、50 kbを超える核酸断片が再生されるように、架橋されたパラフィン包埋組織サンプルなどの保存されたサンプルから核酸を単離させる工程は、架橋されたパラフィン包埋組織サンプルなどの保存されたサンプルをアントラニル酸塩及びホスファニル酸塩のうち少なくとも1つに接触させる工程を含む。場合によっては、単離させる工程は40 以下の温度で行われる。場合によっては、単離させる工程は40 以下の温度で行われる。場合によっては、共通のDNA複合体から第1のDNAセグメント及び第2のDNAセグメントを分離する工程は、プロテイナーゼK処理を含む。場合によっては、複数の核酸結合部分は核タンパク質を含む。場合によっては、複数の核酸結合部分はトランスポサーゼを含む。場合によっては、複数の核酸結合部分はヒストンを含む。場合によっては、複数の核酸結合部分は核酸結合タンパク質を含む。場合によっては、複数の核酸結合部分はナノ粒子を含む。場合によっては、少なくとも1つの複合体の少なくとも1つのリン酸ジエステル骨格を切断する工程は、制限エンドヌクレアーゼに接触させる工程を含む。場合によっては、少なくとも1つの複合体の少なくとも1つのリン酸ジエステル骨格を切断する工程は、非特異的エンドヌクレアーゼに接触させる工程を含む。場合によっては、少なくとも1つの複合体の少なくとも1つのリン酸ジエステル骨格を切断する工程は、DNAを剪断する工程を含む。場合によっては、少なくとも1つの複合体の少なくとも1つのリン酸ジエステル骨格を切断する工程は、トランスポサーゼに接触させる工程を含む。場合によっては、少なくとも1つの複合体の少なくとも1つのリン酸ジエステル骨格を切断する工程は、トポイソメラーゼに接触させる工程を含む。

【0006】

本明細書には、保存された組織サンプルから空間的に分布されたゲノムの構造情報を回復する工程が提供される。幾つかのそのような方法は：組織サンプルを得る工程；固定された三次元のパラフィン包埋組織サンプルなどの前記保存された組織サンプルの第1の部分から一部を抽出する工程；タンパク質DNA複合体が破壊又は分裂されないように、第1の部分の一部から核酸を単離させる工程；第1のDNAセグメント及び第2のDNAセグメントが共通のタンパク質DNA複合体から生じると識別されるように、タンパク質DNA複合体にタグを付ける工程；共通のDNA複合体から第1のDNAセグメント及び第2のDNAセグメントを分離する工程；第1のDNAセグメント及び第2のDNAセグメントから配列情報を生成する工程；共通のタンパク質DNA複合体を示すタグ配列を共有する配列情報を共通のゲノムの構造に割り当てる工程；及び保存された組織サンプルの第1の部分に共通のゲノムの構造を割り当てる工程を含む。場合によっては、保存された組織サンプルは、核酸を単離させる前に均質化されない。場合によっては、組織サンプルは、固定された三次元のパラフィン包埋組織サンプルを含む。場合によっては、タグ配列は、複合体を識別するオリゴタグを含む。場合によっては、タグ配列は、第1のセグメントから第2のセグメントライゲーションにより生じる。場合によっては、タンパク質DNA複合体が破壊又は分裂されないように、架橋されたパラフィン包埋組織サンプルから核酸を単離させる工程は、架橋されたパラフィン包埋組織サンプルをキシレンに接触させる工程を含む。場合によっては、タンパク質DNA複合体が破壊又は分裂されないように、架橋されたパラフィン包埋組織サンプルから核酸を単離させる工程は、架橋されたパラフィン包埋組織サンプルをエタノールに接触させる工程を含む。場合によっては、タンパク質DNA複合体が破壊又は分裂されないように、架橋されたパラフィン包埋組織サンプルが

10

20

30

40

50

ら核酸を単離させる工程は、沸騰状態からサンプルを保護する工程を含む。場合によっては、共通のDNA複合体から第1のDNAセグメント及び第2のDNAセグメントを分離する工程は、プロテイナーゼK処理を含む。場合によっては、組織サンプルは、固定された三次元のパラフィン包埋組織サンプルを含む。

【0007】

本明細書には、処置レジメンの試験結果を再評価する方法が提供される。幾つかのそのような方法は：患者集団の処置レジメン結果に関連するデータを得る工程；前記患者集団の複数の患者から固定された組織サンプルなどの保存された組織サンプルを得る工程；前記固定された組織サンプルから核酸複合体を抽出する工程；前記核酸複合体を複数の前記固定された組織サンプルに使用して、ゲノムの構造情報を判定する工程；及び処置レジメン結果に関連するゲノムの構造情報を識別するように、処置レジメン結果に関連するデータをゲノムの構造情報に関連づける工程を含む。場合によっては、保存された組織サンプルは、核酸を抽出する前に均質化されない。場合によっては、前記固定された組織サンプルから核酸複合体を抽出する工程；及び前記核酸複合体を複数の前記固定された組織サンプルに使用してゲノムの構造情報を判定する工程は、本明細書に開示される方法の何れかを含む。

10

【0008】

本明細書には、ヌクレオチド配列アセンブリの方法が提供される。幾つかのそのような方法は：固定された組織サンプルを提供する工程；前記固定された組織サンプルから架橋されたDNA：タンパク質の複合体を回復する工程；前記架橋されたDNA：タンパク質の複合体のDNAの第1の部分を、前記架橋されたDNA：タンパク質の複合体のDNAの第2の部分にライゲートする工程であって、それによりライゲートされたDNAを形成する、工程；前記架橋されたDNA：タンパク質の複合体からライゲートされたDNAを抽出する工程；前記ライゲートされたDNAを配列決定する工程；及びヌクレオチド配列をアセンブルするために前記配列決定する工程からの情報を使用する工程を含む。場合によっては、前記固定された組織サンプルはホルマリン固定される。場合によっては、固定された組織サンプルは、核酸を単離させる前に均質化されない。場合によっては、前記固定された組織は、ホルマリン固定パラフィン包埋（FFPE）されている。場合によっては、前記架橋されたDNA：タンパク質の複合体はクロマチンを含む。場合によっては、前記ライゲートする工程は平滑末端ライゲーションを含む。場合によっては、本明細書に開示される方法は更に、前記ライゲートする工程の前に、前記架橋されたDNA：タンパク質の複合体からDNAを消化する工程を含む。場合によっては、前記消化する工程は制限酵素消化を含む。場合によっては、本明細書に開示される方法は更に、前記消化する工程の後に、平滑末端を生成するために前記消化する工程から付着末端を充填する工程を含む。場合によっては、前記充填する工程は、ビオチン化ヌクレオチドを使用して行われる。場合によっては、前記回復する工程は、前記架橋されたDNA：タンパク質の複合体のDNAを固形支持体に結合する工程を含む。場合によっては、前記抽出する工程は、前記架橋されたDNA：タンパク質の複合体からタンパク質を消化する工程を含む。場合によっては、前記情報は、2000を超える塩基対（bp）の距離よりも長い距離の情報を含む。場合によっては、前記距離は10,000bpを超える。場合によっては、前記距離は100,000bpを超える。場合によっては、前記距離は200,000bpを超える。場合によっては、本明細書に開示される方法は更に、前記回復する工程の前に、前記固定された組織サンプルの包埋材料を溶かす工程を含む。場合によっては、前記包埋材料はパラフィンを含む。

20

30

40

【0009】

本明細書には、組織サンプル分析の方法が提供される。幾つかのそのような方法は：固定された組織サンプルを提供する工程；前記固定された組織サンプルの第1の部分及び第2の部分を集める工程であって、前記第1の部分及び第2の部分は前記固定された組織サンプルの異なる領域に由来する、工程；前記第1の部分から第1の架橋されたDNA：タンパク質の複合体を、及び前記第2の部分から第2の架橋されたDNA：タンパク質の複

50

合体を回復する工程；(i)前記第1の架橋されたDNA：タンパク質の複合体のDNAの第1の部分を前記第1の架橋されたDNA：タンパク質の複合体のDNAの第2の部分にライゲートする工程であって、それにより第1のライゲートされたDNAを形成する、工程、及び(ii)前記第2の架橋されたDNA：タンパク質の複合体のDNAの第2の部分に前記第2の架橋されたDNA：タンパク質の複合体のDNAの第2の部分にライゲートする工程であって、それにより第2のライゲートされたDNAを形成する、工程；前記第1の架橋されたDNA：タンパク質の複合体から前記第1のライゲートされたDNAを、及び前記第2の架橋されたDNA：タンパク質の複合体から前記第2のライゲートされたDNAを抽出する工程；前記第1のライゲートされたDNA及び第2のライゲートされたDNAを配列決定する工程；及び第1のヌクレオチド配列と第2のヌクレオチド配列をアセンブルするために前記配列決定する工程からの情報を使用する工程を含む。場合によっては、固定された組織サンプルは、核酸を単離させる前に均質化されない。場合によっては、前記固定された組織サンプルはホルマリン固定される。場合によっては、前記固定された組織は、ホルマリン固定パラフィン包埋(F F P E)されている。場合によっては、前記第1の架橋されたDNA：タンパク質の複合体、及び第2の架橋されたDNA：タンパク質の複合体は各々、クロマチンを含む。場合によっては、(d)(i)及び(d)(ii)における前記ライゲートする工程は、平滑末端ライゲーションを含む。場合によっては、本明細書に開示される方法は更に、(d)(i)及び(d)(ii)における前記ライゲートする工程の前に、前記第1の架橋されたDNA：タンパク質の複合体、及び第2の架橋されたDNA：タンパク質の複合体から、DNAを消化する工程を含む。場合によっては、前記消化する工程は制限酵素消化を含む。場合によっては、本明細書に開示される方法は更に、前記消化する工程の後に、平滑末端を生成するために前記消化する工程から付着末端を充填する工程を含む。場合によっては、前記充填する工程は、ビオチン化ヌクレオチドを使用して行われる。場合によっては、前記回復する工程は、前記第1の架橋されたDNA：タンパク質の複合体及び前記第2の架橋されたDNA：タンパク質の複合体のDNAを、固形支持体に結合する工程を含む。場合によっては、前記抽出する工程は、前記第1の架橋されたDNA：タンパク質の複合体及び第2の架橋されたDNA：タンパク質の複合体からタンパク質を消化する工程を含む。場合によっては、前記情報は、2000を超える塩基対(bp)の距離よりも長い距離の情報を含む。場合によっては、前記距離は10,000bpを超える。場合によっては、前記距離は100,000bpを超える。場合によっては、前記距離は200,000bpを超える。場合によっては、本明細書に開示される方法は更に、前記回復する工程の前に、前記固定された組織サンプルの包埋材料を溶かす工程を含む。場合によっては、前記包埋材料はパラフィンを含む。

【0010】

本明細書には、保存されたサンプルからゲノムの構造情報を得るためのキットも提供される。幾つかのそのようなキットは：緩衝液、DNA結合剤、アフィニティタグ結合剤、デオキシリボヌクレオチド、タグ付けされたデオキシリボヌクレオチド、DNA断片化剤(fragmenting agent)、末端修復酵素、リガーゼ、タンパク質除去剤、及び保存されたサンプルからゲノムの構造情報を得る際の使用説明書を含む。随意に、キットは更に、PCRのための試薬、又はPCR試薬と組み合わせたキットの使用説明書を含む。場合によっては、PCRのための試薬は、緩衝液、ヌクレオチド、フォワードプライマー、リバースプライマー、及び熱安定性DNAポリメラーゼを含む。様々な緩衝液は、制限消化緩衝液、末端修復緩衝液、ライゲーション緩衝液、TE緩衝液、洗浄緩衝液、TWB溶液、NTB溶液、LWB溶液、NWB溶液、及び架橋逆転緩衝液のうち少なくとも1つを含む。場合によっては、制限消化緩衝液はDpnII緩衝液を含む。例えば、末端修復緩衝液は頻繁にNEB緩衝液2を含む。ライゲーション緩衝液は頻繁にT4DNAリガーゼ緩衝液、BSA、及びTriton X-100を含む。TE緩衝液は頻繁にトリス及びEDTAを含む。場合によっては、洗浄緩衝液はトリ及び塩化ナトリウムを含む。場合によっては、TWB溶液は、トリス、EDTA、及びTween 20を含む。

10

20

30

40

50

場合によっては、N T B 溶液は、トリス、E D T A、及び塩化ナトリウムを含む。場合によっては、L W B 溶液は、トリス、塩化リチウム、E D T A、及びT w e e n 2 0 を含む。場合によっては、N W B 溶液は、トリス、塩化ナトリウム、E D T A、及びT w e e n 2 0 を含む。場合によっては、架橋逆転緩衝液は、トリス、S D S、及び塩化カルシウムを含む。場合によっては、D N A 結合剤はクロマチンキャプチャビーズを含む。場合によっては、クロマチンキャプチャビーズは、P E G - 8 0 0 粉末、トリス緩衝液、塩化ナトリウム、E D T A、界面活性剤、T E 緩衝液、及びs e r a - m a g ビーズを含む。場合によっては、アフィニティータグ結合剤はストレプトアビジンビーズを含む。場合によっては、ストレプトアビジンビーズはダイナビーズを含む。場合によっては、デオキシリボヌクレオチドは、d A T P、d T T P、d G T P、及びd C T Pのうち少なくとも3つを含む。場合によっては、デオキシリボヌクレオチドは、デオキシd C T P、デオキシd A T P、デオキシd T T P、及びデオキシd G T Pのうち少なくとも3つを含む。場合によっては、D N A 断片化剤は、制限酵素、トランスポサナーゼ、ヌクレアーゼ、音波処理デバイス、流体力学的剪断デバイス（s h e a r i n g d e v i c e）、及び二価金属カチオンのうち少なくとも1つである。場合によっては、制限酵素はD p n I Iを含む。場合によっては、末端修復酵素は、T 4 D N A ポリメラーゼ、クレノウD N A ポリメラーゼ、及びT 4 ポリヌクレオチドキナーゼのうち少なくとも1つを含む。場合によっては、リガーゼはT 4 D N A リガーゼを含む。場合によっては、タンパク質除去剤は、プロテアーゼとフェノールのうち少なくとも1つを含む。場合によっては、プロテアーゼは、プロテイナーゼK、ストレプトマイセス - グリセウスプロテアーゼ、セリンプロテアーゼ、システインプロテアーゼ、トレオニンプロテアーゼ、アスパラギン酸プロテアーゼ、グルタミン酸プロテアーゼ、メタロプロテアーゼ、及びアスパラギンペプチドリアーゼのうち少なくとも1つを含む。場合によっては、キットは、包埋材料の除去のための溶媒を随意に含む。場合によっては、溶媒は、キシレン、ベンゼン、及びトルエンのうち少なくとも1つである。本明細書に列挙されるキットの構成要素、及びその実質的に同等な変形を考慮すると、代替的なキットも考慮され、ここで、少なくとも1つの市販で入手可能なキットの構成要素は除外され、独立して得られた試薬と組み合わせた残りの構成要素の成功的使用のための説明書と取り替えられる。

【0011】

引用による組み込み

本明細書で言及される全ての刊行物、特許、および特許出願は、あたかも個々の刊行物、特許、または特許出願が参照により組み込まれるように具体的かつ個々に指示される程度に、参照により本明細書に組み込まれる。本明細書で言及される全ての刊行物、特許、及び特許出願は、その中で引用されるあらゆる引用と同様に、その全体において引用により本明細書に組み込まれる。

【図面の簡単な説明】

【0012】

【図1A】ホルマリン固定パラフィン包埋（F F P E）組織サンプルの典型的な概略図を表す。

【図1B】クロマチンに基づく次世代配列決定（N G S）ライブラリの調製のためのプロトコルの典型的な概略図を表す。

【図1C】保存されたサンプル（例えばF F P Eサンプル）からのクロマチン抽出及びライブラリ調製（例えばC h i c a g oライブラリ調製）のためのワークフローの典型的な模式図を示す

【図2A】相互転座を見つけ出すために使用され得る典型的な単純なカーネルを表す。

【図2B】相互転座を見つけ出すために使用され得る典型的な単純なカーネルを表す。

【図3】E T V 6 とN T R K 3 との間の相互転座のシグナルを含む画像を表す。

【図4A】3つの異なるサンプルにおいて比較された同じ対の染色体での画像分析に基づく結果を表す。

【図4B】3つの異なるサンプルにおいて比較された同じ対の染色体での画像分析に基づ

く結果を表す。

【図 4 C】3つの異なるサンプルにおいて比較された同じ対の染色体での画像分析に基づく結果を表す。

【図 5 A】染色体 1 対染色体 7 に関する中間の標準化されたリード密度 (10 以上のサンプル) を表す。

【図 5 B】染色体 2 対染色体 5 に関する中間の標準化されたリード密度 (10 以上のサンプル) を表す。

【図 5 C】染色体 1 対染色体 1 に関する中間の標準化されたリード密度 (10 以上のサンプル) を表す。

【図 6 A】様々なピンを取り扱う手法を表す。図 6 A は等しいピンの大きさを示す。

10

【図 6 B】様々なピンを取り扱う手法を表す。図 6 B はピンの補間を示す。

【図 7】全ゲノムスキャニング分析のパイプラインによる分析を表す。

【図 8 A】FFPE に基づく「Chicago」リードペアライブラリに由来する、リードペア距離頻度データを表す。

【図 8 B】古典的な「Chicago」に基づくリードペアライブラリ (図 8 B) に由来する、リードペア距離頻度データを表す。

【図 9 A】リードペアの GRCh38 基準配列上のマッピングされた位置が、GM12878 と基準との間の構造的な差異の付近でプロットされることを例示する。図 9 A は、隣接している 20 kb の反復領域での 80 Kb の逆位に関するデータを表す。

【図 9 B】リードペアの GRCh38 基準配列上のマッピングされた位置が、GM12878 と基準との間の構造的な差異の付近でプロットされることを例示する。図 9 B は、フェージングされた段階的なヘテロ接合性欠失に関するデータを表す。

20

【図 10】本明細書で提供される方法を実施するようにプログラム又は構成される、典型的なコンピュータシステムを示す。

【図 11 A】Hi-C により分析された細胞培養物と比較した、本開示の方法による FFPE 組織及び FFPE 細胞培養物サンプルの分析の結果を示す。

【図 11 B】長距離のゲノム連鎖データを生成するために Ashkenazi father (GM24149) 細胞培養物 FFPE サンプルの分析の結果を示す。

【図 11 C】長距離のゲノム連鎖データを生成するために Ashkenazi father (GM24149) 細胞培養物 FFPE サンプルの分析の結果を示す。

30

【図 11 D】長距離のゲノム連鎖データを生成するために Ashkenazi father (GM24149) 細胞培養物 FFPE サンプルの分析の結果を示す。

【発明を実施するための形態】

【0013】

生体情報の大きなリポジトリが、ホルマリン固定パラフィン包埋 (FFPE) 組織サンプルなどの保存されたサンプルに保管されており、そのようなサンプルは、患者から罹患した又は損傷を受けた組織を切除するための手術などの手術中に慣例的に得られる。しかし、そのようなサンプルの保存中に生じる架橋は、これらサンプルからの DNA 抽出を妨げると考えられていた。保存および保管は技術的に容易であり (straight forward) 且つ経済的であり、その結果、多数の患者サンプルがこの手法を使用して保管されている。その結果、例えば癌治療試験を受ける患者の腫瘍組織からサンプルを獲得且つ保存することが、長い間慣例的であった。

40

【0014】

最近まで、これらサンプルは構造情報へのアクセスにしか有用ではなかった。三次元の組織切片が形態分析のために十分に保存され且つ利用可能であったが、組織保護のプロセスは、保存されたサンプルからゲノムレベル情報にアクセスすることを妨げた。例えば、図 1 A は、保存されたサンプル (例えば FFPE サンプル) の典型的な模式図を表す。細胞 (101) は、それらの三次元分布が保存されるように、固定サンプルの組織 (102) 内に空間的に分布するものとして表される。核酸 (103) は細胞内に存在する。

【0015】

50

これらサンプルから核酸情報を得るための努力が行われてきたが、得られた核酸は短く、局所配列情報のみが獲得可能となるように高度に分解される。従って、再配置に関するゲノムレベル情報は容易に獲得可能ではない。再配置は、限定されないが欠失、重複、挿入、逆位又は逆転、転座、結合、融合、及び分裂を含み得る。

【 0 0 1 6 】

多くの既知の障害において、疾患に関係するものはこれらゲノムスケールの再配置である。遺伝子融合、具体的にゲノム再配置から結果として生じるものは、一部の癌に特に共通するものであり、治療に応じた疾患結果を頻繁に示す。一般的に、これら再配置パターンは、保存されたサンプル中の1つ又は別の形態構造に確実に関連づけられない。正しくは、これらは直接遺伝子型を特定されなければならない。その結果、この情報は、腫瘍サンプル自体が保存され、且つ化学療法又は他の治療に対する腫瘍の反応に関するデータが容易に利用可能であるにもかかわらず、利用できない。

10

【 0 0 1 7 】

本明細書中の方法及び組成物は、上記で考慮されるサンプルなどの保存されたサンプルからのゲノムの構造情報の判定に関連する。幾つかの本明細書中の方法は、保存されたサンプルに含まれるゲノムの構造情報にアクセスするように抽出手法を利用する手法に依存する。タンパク質DNA複合体は、複合体が破壊又は分裂されないようにサンプルから抽出され、且つ、核酸の第1のセグメント及び第2のセグメントがリン酸ジエステル骨格から独立して保持されるという事実を利用する。セグメントは、オリゴを使用して、又はセグメントを互いにライゲートすることによってタグを付けられ、共通の足場へとマッピングされるものにコンティグを割り当ててを可能にする配列情報が得られる。ライゲートされたセグメントの評価により生成されるリードペアの頻度及び型の評価によって、両方の物理的連鎖又はフェーズ情報を推測し、且つ、障害に関係する構造再編成などの特定のゲノムの構造再編成の存在を判定する場合もある。

20

【 0 0 1 8 】

保存された組織の三次元配置も、これらサンプルに保存される。癌性の腫瘍は通常、これらのゲノム構造に関して不均質である。DNA修復欠損、細胞死抑制、腫瘍増殖、及び転移に関連する別個の突然変異により、腫瘍の特性が頻繁に決定づけられる。腫瘍は通常、突然変異の様々な組み合わせを持ち且つ様々な程度 of の健康のリスクがあり、多数の細胞部分集団を含んでいる。頻繁に、これらリスクは局所の形態と関連づけられる。腫瘍細胞集団は、静止性のものから、細胞集団を局所的に複製する良性のもの、比較的高い健康のリスクを表わす転移する細胞集団にまで及ぶ。故に、通常は腫瘍中に与えられたゲノム構造の存在だけでなく、腫瘍サンプル内の空間的に分離された部分集団の局所的ゲノム構造を識別することは、以前の薬物処置の相対的有効性を評価しようと試み、又は未知のリスクの腫瘍を提示する患者に適切な薬物を選択しようと試みる研究者及び従事者にとって、価値がある。具体的に、ゲノム構造を腫瘍中の位置及び腫瘍内の既知の細胞形態と関連づけることは、どのゲノム構造が最高リスクの腫瘍位置及び局所的細胞形態に最も密接に対応するのかを判定するために有益である。

30

【 0 0 1 9 】

当該技術分野の手法を用いてFFPEサンプルなどの保存されたサンプルから抽出されたDNAは頻繁に、長さが300未満の塩基対である。一部のニッキング及び損傷が、保存(例えばFFPE)プロセス、及びその後の脱水と長期保管の間に生じることもある。有意な量の断片化も、抽出プロセス中に生じる場合があり、これは典型的に、架橋を逆転し且つDNAを放つために、一晚のプロテイナーゼK処理、及び沸騰を含んでいる。それにもかかわらず、本明細書中の手法を通じて、そのような核酸分子は、DNAタンパク質複合体の破壊又は分裂なしに切除されたDNAタンパク質複合体に保存される構造情報と組み合わせて、ゲノムの構造再編成に関して有益な情報をもたらす。

40

【 0 0 2 0 】

ネイティブクロマチン及び再構成クロマチン

保存されたサンプルは頻繁に、ネイティブクロマチン又は再構成クロマチンを含み、又

50

は他の場合には複数の点にてタンパク質足場又は非タンパク質足場前に結合した核酸を有し、その結果、第1のセグメント及び第2のセグメントは、架橋剤を接触させる直前にそれら共通のリン酸ジエステル骨格から独立して共に保持される。真核生物において、ゲノムDNAは、核内の染色体としてクロマチンに詰められる。真核生物のネイティブクロマチンの基礎的な構造単位は、ヒストン八量体の周囲に包まれるDNAの146の塩基対(b p)から成るヌクレオソームである。ヒストン八量体は、コアヒストンH2A - H2B二量体及びH3 - H4二量体の各々の2つのコピーから成る。ヌクレオソームは、共通して「ビーズ・オン・ストリング」と称されるものにおけるDNAに沿って規則的に間隔を空けられる。

【0021】

ヌクレオソームへのコアヒストン及びDNAのアセンブリは、シャペロンタンパク質及び関連する集合因子により媒介される。これら因子のほぼ全ては、コアヒストン結合タンパク質である。ヌクレオソームアセンブリタンパク質-1(NAP-1)などのヒストンシャペロンの一部は、ヒストンH3及びH4に結合する嗜好性を示す。新しく合成されたヒストンがアセチル化され、次にクロマチンへのアセンブリ後に引き続き脱アセチル化されることも観察された。それ故、ヒストンのアセチル化又は脱アセチルを媒介する因子は、クロマチンアセンブリプロセスにおいて重要な役割を果たす。

【0022】

一般に、2つのインビトロの方法が、クロマチンを再構成又はアセンブルするために開発されている。1つの方法はATPに依存しないものであるが、他の物はATPに依存するものである。クロマチンを再構成するためのATP非依存性の方法は、ヒストンシャペロンとして作用するようにタンパク質様NAP-1又は塩の何れかを加えたDNA又はコアヒストンを含んでいる。この方法の結果、細胞中のネイティブコアヌクレオソーム粒子を正確に模倣しないDNA上にヒストンの無作為な配置がもたらされる。これらの粒子は、規則的に順序づけられ且つ拡張されたヌクレオソームアレイではなく、使用されるDNA配列が通常は250bpよりも長くないため、頻繁にモノヌクレオソームと呼ばれる(Kundu, T. K. et al., Mol. Cell 6: 551-561, 2000)。より大きな長さのDNA配列上の順序づけられたヌクレオソームの拡張アレイを生成するために、ATP依存性のプロセスを通じてクロマチンをアセンブルしなければならない。

【0023】

過ヨウ素酸ヌクレオソームアレイのATP依存性のアセンブリは、本来のクロマチンに見られるものと同様であり、DNA配列、コアヒストン粒子、シャペロンタンパク質、及びATPを利用するクロマチンアセンブリ因子を必要とする。ACF(ATPを利用するクロマチンアセンブリ因子及びリモデリング因子)又はRSF(リモデリングとスペーシングの因子)は、インビトロでクロマチンへのヌクレオソームの拡張され順序づけられたアレイを生成するために使用される、2つの広く研究されたアセンブリ因子である(Fyodorov, D. V., and Kadonaga, J. T. Method Enzymol. 371: 499-515, 2003; Kundu, T. K. et al. Mol. Cell 6: 551-561, 2000)。

【0024】

特定の実施形態において、本開示の方法は、断片化された二本鎖DNAのあらゆる型に容易に適用され得、これらの型は限定されないが、例えば、血漿、血清、及び/又は尿から単離された遊離DNA; 細胞及び/又は組織からのアポトーシスDNA; インビトロで(例えば、DNase I、トランスポサゼ、及び/又は制限エンドヌクレアーゼにより)酵素学的に断片化されるDNA; 及び/又は機械的な力(水力剪断(hydro-shear)、音波処理、噴霧化など)により断片化されるDNAを含む。

【0025】

再構成クロマチンは、ヌクレオソームや、タンパク質までも含む必要はない。正しくは、広く定義された再構成クロマチンは、第1のセグメント及び第2のセグメントがそれら

10

20

30

40

50

のリン酸ジエステル骨格から独立して共に保持されるように、少なくとも1つの核酸結合を含む。多くの核酸結合部分はクロマチン再構成に適している。例として、ヌクレオソームへと個々にアセンブルされるヒストンなどの核タンパク質の他、転写因子、トランスポゾン、又は核酸結合活性を持つ他のタンパク質などの他の核酸結合タンパク質が挙げられる。オルガネラ核酸結合タンパク質などの非核タンパク質も考慮される。ナノ粒子又は核酸結合表面などの非タンパク質部分も考慮される。

【0026】

保存された抽出核酸中のDNA結合情報の保存

ホルマリン固定パラフィン包埋のサンプルなどの保存されたサンプルは、固定材料及び/又は包埋材料により引き起こされた損傷などの損傷がある核酸を頻繁に含む。DNAを使用する際の関連する構成要素は、DNA傷害物質に晒される単離されたDNAのDNA物理的連鎖情報の完全性を保存している。DNAは比較的安定した分子であるが、DNAの完全性は環境要因、及び具体的に影響を受ける。ヌクレアーゼ汚染、加水分解、酸化、化学物質、物理的且つ機械的な損傷の存在は、DNA保存に対する主な脅威の一部を表す。輸送中にDNAが遭遇した機械的、環境的、及び物理的因子は頻繁に、断片の中に残り、潜在的に長距離の情報を失い、このことはゲノムの分析に重大なものである。DNA情報を保存する既存の方法の多くは、DNAの崩壊を遅らせるが、経時的に、特に断片化が生じると、DNA損傷に対する保護をほとんど提供しない。多くの場合、そのようなDNAの損傷は、長期間の保管を意図したサンプルを固定且つ包埋することにより軽減され得る。例えば、FFPE（ホルマリン固定パラフィン包埋）サンプルは、長時間維持され得る。しかし、保存プロセスは結果としてDNA損傷をもたらしかねない。加えて、後のDNA抽出法は大抵荒いものであり、更なるDNA損傷及び断片化を引き起こす。

【0027】

本明細書には、保存された（例えばFFPE）サンプル（組織ベースの保存されたサンプル及び細胞培養ベースの保存されたサンプルを含む）に保管される架橋クロマチンなどのDNA複合体又はクロマチン凝集体中の核酸分子などの、保存及び/又は保管された核酸分子から長距離ゲノム情報を再生することに関連する方法、組成物、及びキットが提供される。具体的には、方法、組成物、システム、及びキットは、核酸の物理的連鎖情報が保存されるように、これら保存されたサンプルからの核酸サンプルに関連する。物理的連鎖情報は、FFPE抽出プロセスにおける核酸自体の保存により、或いは、抽出プロセスにおいて核酸自体に生じ得る何らかの損傷から独立して物理的連鎖情報が保存されるように核酸複合体を保存することにより、保存される。

【0028】

頻繁に、二本鎖の破壊が、DNA保管中、或いは、FFPEサンプルなどの保存されたサンプルからのDNAの抽出中に生じ、これにより物理的連鎖情報の損失が引き起こされる。物理的連鎖情報の損失は特に不利益なものである。なぜならば、複相生物サンプルにおいて、共通の遺伝子座にマッピングされる突然変異が実際に、同じ対立遺伝子にあるか、又は複相ゲノムの異なる鎖に位置する2つの別個の相同性の対立遺伝子上に存在するかどうかを配列アセンブラが判定することが、この損失により妨げられるからである。ゲノム情報がオーダーメイド医療、或いは、より多くの医薬目的又は治療目的のために使用されると、アセンブルされたコンティグ配列に物理的連鎖情報を割り当てることますます重要となる。

【0029】

DNAの完全性に対するこれらの障害は、ゲノミクス技術がゲノムの世界的な、長期的な、歴史的な、又は大規模の研究のためのプログラムの拡大と共に改善するにつれ、問題となる。そのような研究は、現在のヒト集団及び個人のゲノム、及びヒトの健康に対するそれらの影響を理解する他、更により強力な技術での将来の研究のために現在のゲノムを保存するのに必須である。後者の懸念はまた、後の分析及び識別のためにDNAサンプルを無期限にバンク保存することを求める、法医学の対象（forensic interest）と重複する。

10

20

30

40

50

【 0 0 3 0 】

物理的連鎖の保存

ホルマリン固定パラフィン包埋のサンプルなどの保存されたサンプルは頻繁に、保存されたサンプルから核酸の物理的連鎖情報を判定する際に障害をもたらす。多くの下流分析が、サンプルから物理的連鎖情報を得るために使用され得、従って F F P E サンプル D N A 抽出中のそのような情報の損失により傷つけられ或いは複雑化される。核酸サンプルは頻繁に、例えば関心領域に隣接してアニールすると知られるプライマーを使用するポリメラーゼ連鎖反応 (「 P C R 」) を介した、大きな断片の増幅のための鋳型として意図される。P C R は、多数のアンプリコン核酸分子が生成される鋳型の存在に依存する。増幅は、単一分子上で互いに物理的に結合されている 2 つのアニーリング部位 (又は、アニーリング部位及び別のアニーリング部位の逆補体) に依存する。従って、アニーリング部位間の物理的連鎖の損失は、P C R 増幅を含む分析を複雑にする。

10

【 0 0 3 1 】

同様に、断片が複製され、増幅され、発現され、或いは遺伝子導入で操作されるように断片を細胞宿主へとクローン化することは、出発物質として単一分子を持つことによって大いに容易になる。断片のための物理的連鎖の損失 (即ち、その断片の切断) は、クローン化を複雑にし、断片アセンブリにおける複数の付加的な工程を必要とする。

【 0 0 3 2 】

代替的に、幾つかの分析手法は、物理的近接の保存を必要とするが、核酸の第 1 のセグメント及び第 2 のセグメントがそれらのリン酸ジエステル骨格によって物理的に結合されたままとなることを必要としない。例えば、その 1 つは、プローブが分解していないサンプル中の共通の分子上に存在するかどうかを判定するための第 1 の核酸セグメント及び第 2 の核酸セグメントへのプローブの共局在化のためにアッセイし得る。物理的連鎖の保存はこの分析を容易にするが、そのような分析には必要ではない。第 1 のセグメント及び第 2 のセグメントがそれら共通のリン酸ジエステル骨格から独立して結合されるように分子を再構成クロマチン複合体にアセンブルすることは、そのような分析を容易にする。それら共通のリン酸ジエステル骨格の切断の事象においてさえ、第 1 のセグメント及び第 2 のセグメントの物理的近接情報は、第 1 の断片及び第 2 の断片が元のサンプル中の共通の分子に存在するかどうかを、第 1 及び第 2 のプローブでの複合体のプロービングが示すように、保存されている。

20

30

【 0 0 3 3 】

配列決定は、物理的連鎖情報の保存から利益を得るが、物理的連鎖の又は更に物理的近接の保存を必要としない、別の分析である。物理的連鎖の保存は配列決定を容易にするが、本明細書に開示される及び当業者に既知の他の方法も容易にする。物理的近接の保存は例えば、近接において保持された断片が物理的連鎖情報を伝えるように容易に末端標識されるので、配列決定を容易にする。暴露された内部末端は、隣接した断片配列が共通の分子にマッピングされるのを可能にするオリゴヌクレオチドタグを使用して標識される。代替的に又は組み合わせにおいて、暴露された末端は、リードペアを生成するように無作為に互いにライゲートされ、ここで、印を付けたライゲーション事象の何れかの側の配列は共通の分子にマッピングされる。物理的近接がない場合でさえ、核酸サンプルが物理的近接情報の損失前に物理的近接マーカを加えるように処理される場合、配列分析は容易になる。つまり、核酸分子上の再構成クロマチンのアセンブリ、内部二本鎖末端の暴露、及びクロスライゲーションを介した或いは共通のオリゴヌクレオチドを用いたタグ付けを介したこれら暴露された末端の標識化は、分子のセグメント中の物理的連鎖に危害を及ぼす又はその損失を引き起こしかねない分解にサンプルが晒される前に行われる。

40

【 0 0 3 4 】

以上のような理由から、保存された (例えば F F P E) サンプルの D N A によりコードされた物理的連鎖情報を抽出するための単純で手頃な技術が、当該技術分野のための重大な必需品になった。本明細書に開示される方法は、非限定的な例として、討論学、農業、環境問題研究、再生可能エネルギー、疫学又は疾患発生対応、及び種の保存を含む、多く

50

の分野で有用である。本開示の技術は、腫瘍サンプルなど組織サンプルの異種性のマッピングのために使用される。例えば、組織塊はその用量の全体にわたってサンプリングされ得、本開示の技術はサンプルを分析するために使用され得、それにより組織容量の全体にわたる変化の比較を可能にする。感染も組織容量の全体にわたって分析され得、本開示の技術は、臨床的に重要な領域のフェージング、構造変異種の分析、コピー数変異の分析、擬似遺伝子（例えばSTRC）の解明、癌における新薬の開発につながるような（drugable）構造変異種のための標的パネル、及び他の用途のために使用することができる。

【0035】

本明細書に開示される方法の幾つかの実施形態において、物理的連鎖情報及び／又はサンプル抽出（例えばFFPEサンプルからの抽出）中の物理的連鎖情報の損失は、核酸切断を物理的に防ぐ又は低減することにより回避され、或いは低減される。フェーズ情報及び／又は物理的連鎖情報の損失は、第1のセグメント及び第2のセグメントをそれらのリン酸ジエステル骨格から独立した物理的近接において保持することにより回避され、或いは低減される。代替的に又は組み合わせにおいて、フェーズ情報及び／又は物理的連鎖情報の損失は、共通の又は相互に相補的なタグを使用して第1のセグメント及び第2のセグメントを標識することにより回避され、或いは低減され、この標識化により、物理的近接情報の損失及び共通のリン酸ジエステル骨格テザーの損失の際、第1のセグメント及び第2のセグメントに付けられるタグ情報の配列決定は、元の分解していないサンプル中の共通のフェーズ又は共通の分子を共有するものとして2つのセグメントを識別するのに十分となる。付加的に、又は代替的に、標識化は第2のセグメントへの第1のセグメントのライゲーシオンにより達成され、ここで第2のセグメントは第1のセグメントに隣接していないが、これらは同じ元のDNA分子上で物理的に結合される。

【0036】

核酸分解は多くの多様なソースから生じる。多くのタイプのDNA分解、具体的には、核酸サンプル中の元の共通の分子上の第1のセグメントと第2のセグメントとの物理的連鎖の損失を結果としてもたらすものなどの、二本鎖切断の導入を結果としてもたらすDNA分解からの保護が、本明細書で考慮される。特に有意なものは、保管された核酸サンプルに経時的に生じるか、又は室温で保管されたサンプルに生じるものなど、非酵素的DNA分解である。非酵素的核酸分解は、沸騰、プロテイナーゼ処理、UV放射、酸化、加水分解、剪断又はタングリング（tangling）などの物理的ストレス、又は、分子が切断され或いはラリアットが生じるように核酸分子の内部結合への遊離3'ヒドロキシル基による求核性の付着を含む。非特異的エンドヌクレアーゼ活性、一本鎖ニックング又は二本鎖切断に関するトポイソメラーゼ活性、制限エンドヌクレアーゼ活性、トランスポサゼ活性、DNAミスマッチ修復又は塩基切除などの酵素活性、或いは、フェーズ情報の損失及び／又は物理的連鎖情報の損失などの核酸損傷を結果としてもたらす他の酵素活性から、結果として生じる核酸損傷も、本明細書で考慮される。酵素分解は、場合によっては外因的であり、例えば、離れた場所や、又は、例えば伝染病或いは他の科学資源に対する負荷が原因で無菌条件が容易又は規則的に得られない場所などの「領域における」収集中に遭遇され得るものなどの非無菌環境における、不完全な核酸分離又は初期の分離に起因するものなどである。

【0037】

本明細書中の幾つかの実施形態は、保存された（例えばFFPE）サンプルから抽出される核酸などの、部分的又は完全に単離された核酸へのインビトロでのクロマチンのアセンブルに関連し、それにより、核酸分子の第1のセグメントと第2のセグメントに関連する物理的連鎖情報は、二本鎖切断が第1の核酸分子と第2の核酸分子との間に生じる事象において失われない。再アセンブルされたクロマチンは、場合によっては、別のソースから提供される核酸結合タンパク質を含む。代替的に、場合によっては、そのネイティブクロマチン構成を破壊又は分裂するように、ネイティブヌクレアーゼ活性を不活性化するように、或いはネイティブクロマチンを破壊又は分裂するように、及びネイティブクロマチ

10

20

30

40

50

ン活性を不活性化するように処理される核酸などの、不完全に単離された核酸サンプルは、サンプル中の核酸を安定させるために架橋剤に接触させられる。他の場合、保存されたサンプルの核酸は、サンプル中に保存されるネイティブクロマチン構造を使用して分析される。

【 0 0 3 8 】

二本鎖切断は頻繁に、経時的にDNA保管中に生じる。その結果、DNA分子のフェーシング情報は、変異体が長距離にわたりハプロタイプに明確に関連づけることができないため、獲得するのが困難である。更に、長い反復領域により分離される核酸セグメントは、共通の足場へと結合又はアセンブルすることができない。これらの障害は、FFPE抽出法、沸騰、プロテイナーゼ処理、長期間の保管、室温保管、酵素的分解又は非酵素的分解、或いはヌクレアーゼ活性を持つ組成物での単離中又はその後の汚染から結果として生じる、二本鎖切断の導入によってしか増幅されない。

10

【 0 0 3 9 】

サンプルの分解はデノボアセンブリに有意に影響を及ぼす。本開示は、幾つかの実施形態において、経時的に二本鎖切断を通じてDNA損傷を防ぐことにより、及び随意に、二本鎖切断のフェーズ決定に対する影響を付加的に少なくすることにより、これらの問題に同時に対処する。保存された高いDNA完全性は、数百キロベースもの、及び適切な入力DNAで最大メガベースのゲノム距離に及び、非常に長距離のリードペアデータ(XLRP)を生成する方法を可能にする。

【 0 0 4 0 】

20

そのようなデータは、セントロメアを含むゲノム中の二本鎖切断、DNA断片化、及び大きな反復の領域による物理的連鎖情報の損失による物理的連鎖情報の損失によって提示される実質的な障壁の克服；コスト効率の良いデノボアセンブリを可能にすること；及びゲノムの分析とオーダーメイド医療に十分な完全性及び正確性の再配置決定データの生成には非常に貴重である。

【 0 0 4 1 】

本明細書中の開示は、通常は共通の抽出（例えばFFPE抽出）方法に生じるフェーズ及び/又は物理的連鎖の情報の損失を防ぐことにより、又は代替的に、二本鎖切断から独立したフェーズ及び/又は物理的連鎖の情報を保存することにより、これらの問題に対処し、その結果、物理的連鎖情報はプロテイナーゼ処理の沸騰などの後処理プロセスの際に保存される。物理的連鎖情報は、核酸分子の第1のセグメント及び第2のセグメントを結合し、それによりそれらを共通のリン酸ジエステル骨格から独立して共に保持することにより、物理的に保存され得る。代替的に又は組み合わせで、物理的連鎖情報は、共通の核酸分子の第1のセグメント及び第2のセグメントのタグ付け又は相互の標識化により保存することができ、その結果、セグメント間の二本鎖切断の導入の事象において、第1のセグメント及び隣接する配列、並びに第2のセグメント及び隣接する配列の配列決定により得られるタグ又は他の標識の情報は、共通の核酸分子の共通のフェーズへと第1のセグメント及び第2のセグメントをマッピングするのに十分となる。タグ付けは代替的に、第2のセグメントへの第1のセグメントのライゲーションにより達成され、ここで第2のセグメントは第1のセグメントに隣接していないが、これらは同じ元のDNA分子上で物理的に結合される。例えば、第1のセグメント及び第2のセグメントはDNA分子配列に沿って隣接することはできないが、互いに物理的に近接するか、或いはクロマチンなどの構造での折り重ねにより少なくとも共通の複合体における構成分子となり得る。そのようなセグメントの暴露された端部は共にライゲートされ得る。別の例において、タグ付けは、第1及び第2の両方のセグメントにバーコード（例えばオリゴヌクレオチドバーコード）又は他のタグをライゲートし、それにより第1のセグメント及び第2のセグメントが共通の複合体又は共通の分子へと認識可能にマッピングされることによって達成される。クロマチンの再アセンブリ、或いは核酸の標識化又はタグ付けによる物理的連鎖情報の保存方法は、以前に説明されている（全体が本明細書に組み込まれる、PCT特許出願番号PCT/US2016/024225）。

30

40

50

【 0 0 4 2 】

本明細書の幾つかの実施形態で特に重要なものは、タンパク質又は非タンパク質核酸結合部分を使用してクロマチンが再構成され得るように、F F P E 包埋サンプルなどの保存されたサンプルからの長い核酸の保存である。再構成クロマチンの使用は、D N A の非常に遠く離れているが分子結合されたセグメントの中で会合を形成する際に都合がよい。本開示は、遠く離れたセグメントが共にもたらされ、それら共通のリン酸ジエステル骨格から独立して互いに物理的に結合され、それによって共通のD N A 分子の以前に遠く離れた部分を物理的に接続することを可能にする。結果として、これらの異種の核酸セグメント間の二本鎖連鎖の切断は、フェーズ及び/又は物理的連鎖の情報の損失を結果としてもたらさない。好ましくは、個々の再構成クロマチンユニットにつき1より多くの核酸分子の包含を最小限にする又は予防する条件下でクロマチン再構成が生じるように注意を払う。その後の処理は、関連するセグメントの配列の確認を可能にし、ゲノム上での別離が入力D N A 分子の完全長にまで拡大するリードペアをもたらすのを可能にする。

10

【 0 0 4 3 】

サンプル

本明細書中のサンプルは、例えばホルマリン固定パラフィン包埋サンプルとして保存され、場合によっては、分析の前の相当な期間にわたり保管される。サンプルは、治験に従って得られ、そして、陽性の薬物処置結果に関連する又はそれを予測するゲノムの構造再編成を識別する試みの数年後に試験され得る。そのようなサンプルは、ゲノムの構造情報などの長距離配列情報を判定する際に使用され得る。本明細書に開示される方法により生成された長距離情報は、逆位、欠失、及び重複などの構造変化の検出のために使用され得る。構造変化の検出はまた、いつ活性エンハンサーが腫瘍遺伝子に近接してもたらされるか、或いはいつ抑圧的なシス作用要素が腫瘍抑制因子に近接してもたらされるかを識別するためにも使用され得る。そのようなドライバ (d r i v e r) 事象の識別は、癌研究、具体的には、腫瘍組織が研究の完了後長く保存され且つ腫瘍の様々な細胞亜集団が異なるゲノムの再構成事象を持つ研究に、適用可能である。例えば、新しい構造変異種が検出され、癌型の原因物質であると判定され得る。

20

【 0 0 4 4 】

本明細書中の方法は、患者、研究動物、又は環境サンプルから得られたサンプルなどの保存されたサンプルからゲノムの構造情報を得るために使用される。幾つかのそのようなサンプルは、生検サンプル、外科的サンプル、腫瘍サンプル、臓器全体、及び他のサンプルを含む。これらサンプルは、ホルムアルデヒド、ホルマリン、UV 光、マイトマイシンC、ナイトロジェンマスタード、メルファラン、1 , 3 - ブタジエンジエポキシド、シスジアミンジクロロプラチナム (I I)、又はシクロホスファミドなどの固定液に頻繁に保存される。保存されたサンプルは、場合によっては固定液にサンプルを落とすことにより、直接的に且つ均質化なしに固定される。一旦保存されると、これらサンプルは数か月又は数年にわたり保管され得る。加えて、サンプルの完全な性質は、サンプルの全体にわたって空間的にゲノムの構造情報の分析を可能にするサンプルの位置情報を保存する。例えば、生検サンプルの縁のゲノムの構造情報は、生検サンプルの中心のゲノムの構造情報と比較することができる。

30

40

【 0 0 4 5 】

本明細書に開示される方法に基づいた構造変化の検出は、遺伝子融合のD N A 構造を判定するためにも使用され得る。共通して使用されたF I S H 方法又はR N A - s e q は、D N A 再編成が生じたと判定することができるが、再配置の実際の配列はこれらの手法によって提供されない。他方、対象の遺伝子融合を作り出した構造変異種を判定するための方法が、本明細書で提供される。

【 0 0 4 6 】

三次元のD N A 構造情報を判定する方法が本明細書で提供される。場合によっては、クロマチンの開状態又は閉状態が、これらの方法により検出される。本明細書に開示される方法により集められた構造情報は、インスレーター又はループの存在の有無を判定するた

50

めに、又は、新たなループ或いは他の新たな染色体内又は染色体間の会合を検出するためにも、使用され得る。

【 0 0 4 7 】

本明細書には、組織マッピングの方法が提供される。組織マッピングは、異なる領域にてゲノムの異種性を判定するために、腫瘍などの組織の異なる区域、及び構造又はフェーズの情報が各生検から判定されるプロセスである。

【 0 0 4 8 】

本明細書に開示される方法は、保存された（例えば F F P E ）サンプルから長距離情報を含むリードペアライブラリの生成のために使用され得る。これらのライブラリは、例えば F F P E 組織において無期限の期間にわたり保存されたサンプルから再生され得る。

10

【 0 0 4 9 】

本明細書には、リンパ球の構造及びフェーズの情報を判定する方法が提供される。場合によっては、これら方法は、異種細胞又は受容体亜型を区別するために使用される。

【 0 0 5 0 】

本明細書で提供される方法は、幾つかの実施形態において、長距離データ及びフェーズ情報を含むデータを使用した、構造変異種又はゲノムの再編成の検出のために使用される。これら方法の出発物質は、大半の臨床的サンプル保存に共通するように、ホルマリンに固定され且つパラフィンに包埋されているサンプルである。本明細書で提供される方法を使用して、構造及び長距離の情報がサンプルから得られ；そのような情報は、高レベルの DNA 断片化により現行の方法を使用しても得ることはできない。それ故、本明細書で提供される方法の使用は、臨床調査と創薬の多くの分野においてこの新たなデータを使用する機会を提供する。

20

【 0 0 5 1 】

本明細書で提供される方法の臨床調査用途は、患者サンプルを使用した治療反応又は抵抗性の追跡を含んでいる。ライブラリの調製又は配列決定の変異を軽減するために、サンプルを同時に処理することが有益である。これは、初期の時点のサンプルが F F P E などにより保存されることを必要とする。本明細書で提供される方法は、多数の時点のサンプルを同時に処理且つ分析できるように、これら保存されたサンプルから使用可能なゲノム物質を効率的に抽出する方法を提供する。

【 0 0 5 2 】

30

一例として、サンプル（例えば生検）は患者から得られ、医療処置中に固定液（例えばホルマリン）に入れられる。この固定サンプルは、本開示の技術を使用して後に分析される。例えば、癌に関連する再編成などのゲノムの特徴が識別され得る。腫瘍 / 非腫瘍のフェージングは、体細胞のゲノム情報から癌のゲノム情報を区別するために分析され得る。

【 0 0 5 3 】

更に、本明細書で提供される方法を使用して、有用な長距離のゲノム情報は、そのような抽出法の発明の前に保存されたより古いサンプルからも得ることができる。例えば、腫瘍サンプルバンクは、臨床的に関連する情報のためにこの情報を掘り出すために、本明細書で提供される方法を使用して処理され、患者の既知の結果に関連付けることができる。このように、本明細書で提供される方法は、予後及び診断の関連付けを可能にする。

40

【 0 0 5 4 】

本明細書で提供される方法と組成物は、保存された組織の構造変異プロファイルを判定するために使用され得る。これら構造変異プロファイルは、異なる亜型又は他の集合体を定めるために、他のデータセット、例えば遺伝子発プロファイル、突然変異プロファイル、メチル化プロファイルなどと共に使用され得る。

【 0 0 5 5 】

本明細書で提供される方法により判定される構造変異プロファイルは、経時的に突然変異の構造の進展を判定するためにも使用される。例えば、場合によっては、進行又は退行を通じた、発端からの腫瘍ゲノム構造における構造変異種の進展がモニタリングされ得る。このように、腫瘍悪性腫瘍及び転移が、より良く理解され得る。モニタリングは、サン

50

ブルの利用可能性に依存して、三次元サンプル中の様々な亜群を調べることによって空間的に、且つ保存されたサンプルの時間経過を調べることによって一時的に行われるように利用可能である。

【 0 0 5 6 】

本明細書で提供される方法は、バンク保管され、アーカイブに保管され、或いはそうでなければ長期間保管された遺伝子サンプルに対しても行われ得る。例えば、珍しい又は未知の疾患に悩んでいた、現在は死亡している患者からの保存された組織サンプルのアーカイブは、本明細書で提供される方法によって分析することができ、それにより標準方法を使用して得ることのできない見識を提供する。

【 0 0 5 7 】

本明細書に開示された技術により分析されたサンプルは、分解され得るか、或いは、構造情報を含むDNA又は長距離DNAの情報の保存に有害な条件を含む様々な条件に晒されている。場合によっては、サンプルは酸処理に晒されている。場合によっては、サンプルはホルムアルデヒド又はホルマリンなどの架橋剤に晒されている。場合によっては、サンプルはパラフィン包埋などの包埋に晒されている。場合によっては、サンプルはパラフィン包埋などの包埋に晒されていない。場合によっては、サンプルは熱処理（例えば包埋物質を融解するために）に晒されている。場合によっては、サンプルはキシレンなどの溶媒（例えば、接着剤を溶かすために）に晒されている。

【 0 0 5 8 】

固定サンプルは、固定の後であるが後処理又は分析の前の様々な条件に晒され得る。例えば、固定後の時間は、少なくとも約10分、20分、30分、40分、50分、1時間、1.5時間、2時間、3時間、4時間、5時間、6時間、7時間、8時間、9時間、10時間、11時間、12時間、18時間、1日、2日、3日、4日、5日、6日、1週、2週、3週、4週、1か月、2か月、3か月、4か月、5か月、6か月、7か月、8か月、9か月、10か月、11か月、1年、2年、3年、4年、5年、6年、7年、8年、9年、10年、15年、20年、25年、30年、35年、40年、45年、50年、55年、60年、65年、70年、75年、80年、85年、90年、95年、100年、又はそれ以上経過し得る。固定後、サンプルは、少なくとも約5、10、15、20、25、30、35、40、45、50、55、60、65、70、75、80、85、90、95、100、又はそれ以上の温度上昇に晒され得る。固定後、サンプルは、少なくとも約5、10、15、20、25、30、35、40、45、50、55、60、65、70、75、80、85、90、95、100、又はそれ以上の温度低下に晒され得る。固定後、サンプルは、少なくとも約10パスカル(Pa)、20Pa、30Pa、40Pa、50Pa、60Pa、70Pa、80Pa、90Pa、100Pa、110Pa、120Pa、130Pa、140Pa、150Pa、160Pa、170Pa、180Pa、190Pa、200Pa、210Pa、220Pa、230Pa、240Pa、250Pa、260Pa、270Pa、280Pa、290Pa、300Pa、310Pa、320Pa、330Pa、340Pa、350Pa、360Pa、370Pa、380Pa、390Pa、400Pa、410Pa、420Pa、430Pa、440Pa、450Pa、460Pa、470Pa、480Pa、490Pa、500Pa、550Pa、600Pa、650Pa、700Pa、750Pa、800Pa、850Pa、900Pa、950Pa、1000Pa、2000Pa、3000Pa、4000Pa、5000Pa、6000Pa、7000Pa、8000Pa、9000Pa、10000Pa、20000Pa、30000Pa、40000Pa、50000Pa、60000Pa、70000Pa、80000Pa、90000Pa、100000Pa、101325Pa、又はそれ以上の圧力（例えば大気圧）低下に晒され得る。固定後、サンプルは、少なくとも約10パスカル(Pa)、20Pa、30Pa、40Pa、50Pa、60Pa、70Pa、80Pa、90Pa、100Pa、110Pa、120Pa、130Pa、140Pa、150Pa、160Pa、170Pa、180Pa、190Pa、200Pa、21

10

20

30

40

50

0 Pa、220 Pa、230 Pa、240 Pa、250 Pa、260 Pa、270 Pa、280 Pa、290 Pa、300 Pa、310 Pa、320 Pa、330 Pa、340 Pa、350 Pa、360 Pa、370 Pa、380 Pa、390 Pa、400 Pa、410 Pa、420 Pa、430 Pa、440 Pa、450 Pa、460 Pa、470 Pa、480 Pa、490 Pa、500 Pa、550 Pa、600 Pa、650 Pa、700 Pa、750 Pa、800 Pa、850 Pa、900 Pa、950 Pa、1000 Pa、2000 Pa、3000 Pa、4000 Pa、5000 Pa、6000 Pa、7000 Pa、8000 Pa、9000 Pa、10000 Pa、20000 Pa、30000 Pa、40000 Pa、50000 Pa、60000 Pa、70000 Pa、80000 Pa、90000 Pa、100000 Pa、101325 Pa、又はそれ以上の圧力（例えば大気圧）上昇に晒され得る。固定後、サンプルは、少なくとも約0.1メートル（m）、0.2 m、0.3 m、0.4 m、0.5 m、0.6 m、0.7 m、0.8 m、0.9 m、1 m、2 m、3 m、4 m、5 m、6 m、7 m、8 m、9 m、10 m、11 m、12 m、13 m、14 m、15 m、16 m、17 m、18 m、19 m、20 m、又はそれ以上の高度変化に晒され得る。

【0059】

固定サンプルは、少なくとも約10分、20分、30分、40分、50分、1時間、1.5時間、2時間、3時間、4時間、5時間、6時間、7時間、8時間、9時間、10時間、11時間、12時間、18時間、24時間、又はそれ以上持続する固定反応で固定され得る。場合によっては、固定サンプルは、少なくとも約30分持続する固定反応で固定され得る。場合によっては、固定反応時間は、固定反応がクエンチされる前に経過した時間であり得る。場合によっては、固定サンプルは、クエンチされない固定反応で固定される。

【0060】

本明細書に開示される方法は、選択的なゲノムの関心領域の他、選択的な関心領域と相互に作用し得るゲノムの領域の遺伝子情報の分析に使用され得る。本明細書に開示されるような増幅方法は、限定されないが米国特許第6449562号、第6287766号、第7361468号、第7414117号、第6225109号、及び第6110709号に見出されるものなどの、遺伝子分析の分野で既知のデバイス、キット、及び方法に使用され得る。場合によっては、本開示の増幅方法は、多形の存在の有無を判定するためのDNAハイブリダイゼーション研究のために標的核酸を増幅するために使用され得る。多形、又は対立遺伝子は、遺伝病などの疾患又は疾病に関連付けられ得る。他の場合、多形は、疾患又は疾病に対する感受性に関連付けることができ、例えば、多形は、中毒、退行性及び加齢性の疾病、癌などに関連付けられる。他の場合、多形は、冠状動脈の健康の増大といった有用な特色、HIV又はマラリアなどの疾患に対する抵抗性、或いは骨粗鬆症、アルツハイマー病、又は痴呆などの成人病に対する抵抗性に関連付けることができる。

【0061】

本開示の組成物及び方法は、診断、予後、治療、患者の層別化、薬物の開発、処置の選択、及びスクリーニングの目的のために使用され得る。本開示は、本開示の方法を使用して多くの様々な標的分子が単一の有生分子サンプルから一度に分析され得るという利点を提供する。これにより、例えば、様々な診断試験が1つのサンプル上で行なわれることが可能となる。

【0062】

本開示の組成物及び方法はゲノミクスにおいて使用され得る。本明細書に記載される方法は、この用途に非常に望ましい答えを迅速に示すことができる。本明細書に記載される方法及び組成物は、診断又は予後徴候のために且つ健康と疾患の指標として使用され得るバイオマーカーを見つけ出すプロセスで使用され得る。本明細書に記載される方法及び組成物は、薬物をスクリーニングするために、例えば、薬物の開発、処置の選択、処置の有効性の判定、及び/又は医薬開発の標的の識別のために、使用され得る。薬物に関するスクリーニングアッセイ時に遺伝子発現を試験する能力は、タンパク質が身体において最終

10

20

30

40

50

的な遺伝子産物であるため、非常に重要である。幾つかの実施形態において、本明細書に記載される方法及び組成物は、実行されている特定のスクリーニングに関する最も多くの情報を提供するタンパク質及び遺伝子発現の両方を同時に測定する。

【 0 0 6 3 】

本開示の組成物及び方法は、遺伝子発現分析に使用され得る。本明細書に記載される方法は、ヌクレオチド配列を区別する。標的ヌクレオチド配列間の差は、例えば、単一の核酸塩基の差、核酸欠失、核酸挿入、又は再編成であり得る。1より多くの塩基に関するそのような配列の差も検出され得る。本開示のプロセスは、感染症、遺伝子疾患、及び癌を検出することができる。

【 0 0 6 4 】

本方法は、罹患した細胞型がサンプルに存在するかどうか、疾患の段階、患者の予後、特定の処置にตอบสนองする患者の能力、又は患者にとって最良の処置を判定するために、患者から得られた又は導き出された有生分子サンプルの分析に適用され得る。本方法は、特定の疾患のためのバイオマーカーを識別するためにも適用され得る。

【 0 0 6 5 】

幾つかの実施形態において、本明細書に記載される方法は疾病の診断に使用される。本明細書で使用されるように、用語「診断する」又は疾病の「診断」は、疾病を予測又は診断すること、疾病の素因を判定すること、疾病の処置をモニタリングすること、疾患の治療反応、或いは疾病の予後、疾病の進行、又は疾病の特定の処置に対する反応を診断することを含む。例えば、保存された（例えばFFPE）臨床サンプルは、サンプル中の疾患又は悪性細胞型のマーカーの存在及び/又は量を判定するために、本明細書に記載される方法のうち何れかに従いアッセイされ得る。

【 0 0 6 6 】

幾つかの実施形態において、本明細書に記載される方法及び組成物は、疾病の診断及び予後のために使用される。多数の免疫学的、増殖性、及び悪性の疾患と障害は、本明細書に記載される方法に特に適している。免疫疾患及び障害は、アレルギー性疾患及び障害、免疫機能の障害、並びに自己免疫疾患及び疾病を含む。アレルギー性疾患及び障害は、アレルギー性鼻炎、アレルギー性結膜炎、アレルギー性喘息、アトピー性湿疹、アトピー性皮膚炎、及び食物アレルギーを含むがこれらに限定されない。免疫不全症は、限定されないが、重症複合免疫不全症（SCID）、好酸球増加症候群、慢性肉芽腫症、白血球接着不全症I及びII、高IgE症候群、チェディアック・東、好中球増加症、好中球減少症、無形成症、無ガンマグロブリン血症、高IgM症候群、ディジョージ/軟口蓋帆・心臓・顔症候群、及びびインターフェロン・ガンマ-T_H1経路欠損症を含む。自己免疫性及び免疫調節異常の障害は、限定されないが、関節リウマチ、糖尿病、全身性エリトマトーデス、グレーブス病、グレーブス眠症、クローン病、多発性硬化症、乾癬、全身性硬化症、甲状腺腫及びリンパ腫性甲状腺腫（橋本甲状腺炎、リンパ節様甲状腺腫）、円形脱毛症、自己免疫性心筋炎、硬化性苔癬、自己免疫性ブドウ膜炎、アジソン病、萎縮性胃炎、重症筋無力症、特発性血小板減少性紫斑病、溶血性貧血、原発性胆汁性肝硬変、ウェゲナー肉芽腫症、結節性多発動脈炎、及び炎症性腸疾患、同種異系移植片拒絶反応、及び感染性細菌又は環境抗原に対するアレルギー反応による組織破壊を含む。

【 0 0 6 7 】

本開示の方法により評価され得る増殖性疾患及び障害は、限定されないが、新生児の血管腫症；二次性進行型多発性硬化症；慢性進行性骨髄変性疾患；神経線維腫症；神経節神経腫症；ケロイド形成；骨のパジエット病；線維嚢胞症（例えば乳房又は子宮の）；サルコイドーシス；ペロニー及びデュピュイトランの繊維症、硬変、アテローム性動脈硬化、及び血管再狭窄を含む。

【 0 0 6 8 】

本開示の方法により評価され得る悪性疾患及び障害は、血液悪性腫瘍及び固形腫瘍の両方を含む。

【 0 0 6 9 】

10

20

30

40

50

血液悪性腫瘍は特に、そのような悪性腫瘍が血液感染性の細胞の変化に関与するため、サンプルが血液サンプルである時に本開示の方法に適している。そのような悪性腫瘍は、非ホジキンリンパ腫、ホジキンリンパ腫、非B細胞リンパ腫、及び他のリンパ腫、急性又は慢性白血病、多血球血症、血小板血症、多発性骨髄腫、骨髄異形成障害、骨髄増殖性障害、脊髄脳炎、異型免疫リンパ球増殖、及び形質細胞障害を含む。

【0070】

本開示の方法により評価され得る形質細胞疾患は、多発性骨髄腫、アミロイドーシス、及びヴァルデンストレームマクログロブリン血症を含んでいる。

【0071】

固形腫瘍の例は、限定されないが、結腸癌、乳癌、肺癌、前立腺癌、脳腫瘍、中枢神経系腫瘍、膀胱腫瘍、黒色腫、肝臓癌、骨肉腫、及び他の骨癌、睾丸及び卵巣の癌腫、頭頸部腫瘍、及び子宮頸部の新生物を含む。

10

【0072】

遺伝子障害も、本開示のプロセスにより検出することができる。これは、染色体及び遺伝子の異常、又は遺伝子疾患に関する出生前又は出生後のスクリーニングにより実行され得る。検出可能な遺伝子疾患の例は、21ヒドロキシラーゼ欠損、嚢胞性繊維症、脆弱X症候群、ターナー症候群、デュシェンヌ型筋ジストロフィー、ダウン症候群又は他のトリソミー、心疾患、単一の遺伝子疾患、HLA型判定、フェニルケトン尿症、鎌状赤血球貧血、テイ・ザックス病、サラセミア、クラインフェルター症候群、ハンチントン病、自己免疫疾患、リビドーシス、肥満欠損症 (obesity defect)、血友病、先天性代謝異常症、及び糖尿病を含む。

20

【0073】

本明細書に記載される方法は、サンプル中の細菌又はウイルスそれぞれのマーカーの存在及び/又は量を判定することにより、病原体感染、例えば細胞内細菌及びウイルスによる感染を診断するために使用され得る。

【0074】

種々様々な感染症が、本開示のプロセスにより検出することができる。感染症は、細菌、ウイルス、寄生物、及び真菌の感染因子により引き起こされ得る。薬物に対する様々な感染因子の抵抗性も、本開示を使用して判定され得る。

【0075】

30

本開示により検出され得る細菌感染因子は、エシェリキア・コリ、サルモネラ、シゲラ、クレブシエラ (Klebsiella)、シュードモナス、リステリア・モノサイトゲネス、マイコバクテリウム・ツベルクローシス、マイコバクテリウム・アビウム・イントラセルラーレ、エルシニア、フランシセラ、パスツレラ、ブルセラ、クロストリジウム、ボルデテラ・ペルツシス、バクテロイデス、スタフィロコッカス・アウレウス、ストレプトコッカス・ニューモニエ、B溶血性連鎖球菌 (B-Hemolytic strep.)、コリネバクテリア、レジオネラ、マイコプラズマ、ウレアプラズマ、クラミジア、ナイセリア・ゴノレア、髄膜炎菌、ヘモフィルス・インフルエンザ、エンテロコッカス・フェカーリス、プロテウス・ブルガリス、プロテウス・ミラビリス、ヘリコバクター・ピロリ、トレポネマ・パラジウム、ボレリア・ブルグドルフェリ、ボレリア・レカレンチス、リケッチア病原微生物、ノカルジア、及び放射菌 (Actinomycetes) を含む。

40

【0076】

本開示により検出され得る真菌感染因子は、クリプトコッカス・ネオフォルマンس、ブラストミセス・デルマトיתיディス、ヒストプラスマ・カブスラーツム、コクシジオイデス・イミチス、パラコクシジオイデス・ブラジリエンシス、カンジダ・アルビカンس、アスペルギルス・フミガーツス (Aspergillus fumigatus)、藻菌類 (リゾープス)、スポトリックス・シェンキー、クロモミコーシス、及びマズラミコーシスを含む。

【0077】

50

本開示により検出されるウイルス感染因子は、ヒト免疫不全ウイルス、ヒトT細胞リンパ性細胞栄養性ウイルス(human T-cell lymphocytotropic virus)、肝炎ウイルス(例えばB型肝炎ウイルス及びC型肝炎ウイルス)、エプスタイン-バーウイルス、サイトメガロウイルス、ヒトパピローマウイルス、オルソミクソウイルス、パラミクソウイルス、アデノウイルス、コロナウイルス、ラブドウイルス、ポリオウイルス、トーガウイルス、ブンヤウイルス、アレナウイルス、風疹ウイルス、及びレオウイルスを含む。

【0078】

本開示により検出され得る寄生生物因子は、熱帯熱マラリア原虫、四日熱マラリア原虫、三日熱マラリア原虫、卵型マラリア原虫、回旋糸状虫(Onchocerca volvulus)、リーシュマニア、トリパノゾーマ種、住血吸虫種、エントアメーバ-ヒストリチカ、クリプトスポリジウム(Cryptosporidium)、ジアルジア種、トリコモナス(Trichomonas)種、大腸バランチジウム(Balantidium Coli)、バンクロフト糸状虫、トキソプラズマ種、蟯虫、回虫、鞭虫、メジナ虫(Dracunculus medinensis)、吸虫、広節裂頭条虫、テニア種、ニューモシスチス-カリニ、及びアメリカ鉤虫(Necator americanus)を含む。

【0079】

本開示は、感染因子による薬物抵抗性の検出にも有用である。例えば、バイコマイシン耐性エンテロコッカス・フェシウム、メチシリン耐性スタフィロコッカス-アウレウス、ペニシリン耐性ストレプトコッカス-ニューモニエ、多剤耐性マイコバクテリウム-ツベルクローシス、及び耐AZT性ヒト免疫不全ウイルスは全て、本開示で識別され得る。

【0080】

故に、本開示の組成物及び方法を使用して検出された標的分子は、患者のマーカー(癌マーカーなど)、又は細菌或いはウイルスマーカーなどの異物による感染のマーカーの何れかであり得る。

【0081】

本開示の組成物及び方法は、標的分子を識別及び/又は定量化するために使用され得、標的分子の存在量は、生物学的状態又は疾患条件、たとえb、疾患状態の結果としてアップレギュレート又はダウンレギュレートされる血液マーカーを示す。

【0082】

幾つかの実施形態において、本開示の方法及び組成物は、サイトカイン発現のために使用され得る。本明細書に記載される方法の低感受性は、例えば癌などの疾患の状態、診断、又は予後のバイオマーカーとしてのサイトカインの早期発見、及び亜臨床的な状態の識別に有用である。

【0083】

標的ポリヌクレオチドが由来する様々なサンプルは、同じ個体の複数のサンプル、異なる個体のサンプル、或いはそれらの組み合わせを含み得る。幾つかの実施形態において、サンプルは、1人の個体の複数のポリヌクレオチドを含む。幾つかの実施形態において、サンプルは、2人以上の個体の複数のポリヌクレオチドを含む。個体は、標的ポリヌクレオチドが由来する生物体又はその一部であり、その非限定的な例には、動物、真菌類、原生生物、モネラ、ウイルス、ミトコンドリア、及び葉緑体が含まれる。例えば組織又は腫瘍生検を含む、保存された(例えばFFPE)細胞サンプル、保存された(例えばFFPE)組織サンプル、又はそこから由来する臓器サンプルなどの、サンプルのポリヌクレオチドは被験体から単離することができる。被験体は、ウシ、ブタ、マウス、ラット、ニワトリ、ネコ、イヌなどの動物を含むがこれらに限定されない動物であり、場合によってはヒトなどの哺乳動物である。サンプルはまた、化学合成などにより人工的に引き出すことができる。幾つかの実施形態において、サンプルはDNAを含む。幾つかの実施形態において、サンプルはゲノムDNAを含む。幾つかの実施形態において、サンプルは、ミトコンドリアDNA、葉緑体DNA、プラスミドDNA、バクテリア人工染色体、酵母人工染

10

20

30

40

50

色体、オリゴヌクレオチドタグ、又はそれらの組み合わせを含む。幾つかの実施形態において、サンプルは、ポリメラーゼ連鎖反応（PCR）、逆転写、及びそれらの組み合わせを含むがこれらに限定されない、プライマーとDNAポリメラーゼとの適切な組み合わせを使用したプライマー伸張反応により生成されたDNAを含む。プライマー伸張反応のための鋳型がRNAである場合、逆転写の生成物は相補的DNA（cDNA）と呼ばれる。プライマー伸張反応に有用なプライマーは、1つ以上の標的に特異的な配列、ランダムシーケンス、部分的ランダムシーケンス、及びそれらの組み合わせを含み得る。プライマー伸張反応に適した反応条件は当該技術分野で知られている。一般に、サンプルのポリヌクレオチドは、サンプル中にポリヌクレオチドを含み、これは標的ポリヌクレオチドを含むこともあれば、含まない場合もある。

10

【0084】

核酸の抽出及び精製の方法は、当該技術分野において周知である。例えば、核酸は、フェノール、フェノール/クロロホルム/イソアミルアルコール、或いはTRIzol及びTriReagentを含む同様の製剤での有機抽出により精製され得る。抽出技術の他の限定されない例は、以下を含む：（1）自動核酸抽出器、例えばApplied Biosystems（Foster City, Calif.）から入手可能なModel 341 DNA Extractorの使用を伴う又は伴わない、例えばフェノール/クロロホルムの有機試薬（Ausubel et al., 1993）を使用する、エタノール沈殿が後続する有機抽出；（2）固定相吸着法（米国特許第5,234,809号；Walsh et al., 1991）；及び（3）典型的に「塩析」方法と称される沈澱法などの、塩で誘導された核酸沈澱法（Miller et al., (1988)）。核酸の単離及び/又は精製の別の例は磁性粒子の使用を含み、核酸は特異的又は非特異的に磁性粒子に結合し、その後磁石を使用してビーズを単離し、洗浄し、そしてビーズから核酸を溶出することができる（例えば米国特許第5,705,628号を参照）。幾つかの実施形態において、上記の単離方法は、サンプルから不要なタンパク質を取り除くのに役立つ酵素消化工程、例えばプロテイナーゼK又は他のプロテアーゼによる消化より始められてもよい。例えば米国特許第7,001,724号を参照。望ましい場合、RNase阻害剤を、溶解緩衝液に添加することができる。特定の細胞又はサンプル型について、前記プロトコルにタンパク質変性/消化の工程を加えることが望ましい場合もある。精製方法は、DNA、RNA、又はその両方を単離することを目的とし得る。抽出手順の間又はその後にDNAとRNAの両方が一緒に単離されると、更なる工程を利用して、一方又は両方を他とは別々に精製することができる。例えば、サイズ、配列、又は他の物理的若しくは化学的特性による精製により、抽出した核酸の細画分を生成することもできる。最初の核酸単離工程に加えて、本開示の方法における工程の後に、過剰な又は不要な試薬、反応物、又は生成物を除去するなどのために、核酸の精製を実施することができる。核酸の鋳型分子は、2003年10月9日に公開された米国特許出願公開第2002/0190663号A1に記載の通り得ることができる。通常、核酸は、Maniatis, et al., Molecular Cloning: A Laboratory Manual, Cold Spring Harbor, N.Y., pp. 280-281 (1982)に記載されるものなどの様々な技術によって生体サンプルから抽出することができる。場合によっては、核酸は、最初に生体サンプルから抽出され、次いでインビトロで架橋され得る。場合によっては、天然の会合タンパク質（例えばヒストン）を、核酸から更に除去することができる。

20

30

40

【0085】

ネイティブクロマチンの抽出及び回復

本明細書には、保存されたサンプル（例えばFFPEサンプル）から長い断片長及び/又はフェーズ情報を含有する断片を抽出する方法が提供される。場合によっては、これら方法は、保存されたサンプル（例えばFFPEサンプル）中に既に存在するクロマチン構造を保存するために、保存された細胞（例えばFFPE細胞）の核を優しく処理する工程を含む。

50

【 0 0 8 6 】

本明細書には、長距離 DNA 断片及び / 又はフェーズ情報を含有する断片の保存のために、抽出及びインサイツでのライブラリ調製を行なう方法が開示される。その後、放たれた DNA は、リードペアライブラリを生成するために使用されるなど、分析のために更に処理され得る。

【 0 0 8 7 】

保存されたサンプル (F F P E サンプルなど) は、包埋材料 (例えばパラフィン) を溶かすための溶解剤で処理され得る。場合によっては、溶解剤はキシレンなどの溶媒である。適切な溶媒剤 (s o l v e n t a g e n t) の他の例は、限定されないが、キシレン、トルエン、及びベンゼンなどの有機溶媒の他、各々の適切な異性体を含む。組成物は、包埋材料が溶解剤中で溶かされるように混合され得る。場合によっては、混合は、ボルテックス、高速振盪、又は攪拌を含む。代替的に、場合によっては優しい攪拌が使用される。サンプルは、サンプルを小球にするのに十分な速度での遠心分離などにより、溶媒からサンプル及び溶解した包埋材料を分離するために処理される。十分な速度は、限定されないが、1分につき14000回転など、テーブルトップ遠心機の最大速度を含む。その後、溶かされた包埋材料を含む溶解剤は、小球を妨害しないように頻繁に優しく取り除かれ得る。その後、過剰な溶解剤は、洗浄試薬で取り除かれ得る。幾つかの例において、洗浄剤はエタノール、例えば100%のエタノールである。サンプルは、保持容器の内壁からサンプル小球を除去するために混合され、ボルテックス処理され、又は攪拌される。サンプルは随意に、再び小球にするために再度遠心分離され得る。その後、残りの液体は、保持容器から除去され、サンプルが乾燥される。代表的な乾燥技術は、空気乾燥、真空乾燥、又は当該技術分野で周知の他の乾燥技術を含む。乾燥後、溶解緩衝液などの緩衝液がサンプルに加えられる。溶解緩衝液は、トリス、塩化ナトリウムなどの塩、硫酸ドデシルナトリウム (S D S) などの1つ以上の洗浄剤、トリトン、E D T A などのキレート剤、及びそれらの任意の組み合わせなどの緩衝剤を含み得る。代表的な溶解緩衝液は、50 mM のトリス pH 8、50 mM の N a C l、1% の S D S、0.15% のトリトン、1 mM の E D T A を含むが、当業者は、この組成物に対する変形が容易に生成され得ることを理解している。他の包埋剤を除去するため適切なプロトコルが、利用され得る。

【 0 0 8 8 】

サンプルは、十分な量の時間にわたりインキュベートし (例えば37℃で)、その一方で随意に振盪又は優しく攪拌すること等により、再水和され得る。その後、溶解緩衝液中で小球を崩壊させる及び再懸濁するために、サンプルは攪拌され、ピペットで移され、又はそうでなければ混合され得る。その後、残りの非可溶性の残屑は、十分な速度での遠心分離などによって溶解緩衝液から分離され得る。DNA タンパク質複合体は、核酸断片にタグを付ける技術などの下流技術を使用して再生且つ評価され得る

【 0 0 8 9 】

ネイティブ DNA : タンパク質複合体 (例えばクロマチン) は、核酸ではなく複合体が無傷に保存されるように、保存されたサンプル (例えば F F P E サンプル) から単離され得る。これら手法において、核酸の物理的連鎖情報は、必ずしも核酸リン酸ジエステル骨格の保存によってではないが、リン酸ジエステル骨格状況から独立して連鎖情報を保存することによって保存することができ、その結果、複合体の共通してタグを付けられた断片は、元のサンプル中の構造的又は物理的な結合配置を持つと推測することができる。

【 0 0 9 0 】

クロマチンの可溶化は、F F P E サンプルなどの保存されたサンプルからネイティブ DNA : タンパク質の複合体を単離し及び長距離連鎖情報を抽出する際に重要な工程であり得る。クロマチン複合体は、プロテイナーゼ消化及び音波処理を含むがこれらに限定されない様々な方法により可溶化され得る。そのような可溶化方法は、可溶性のクロマチンを放つために組織とクロマチンを妨害することができる。

【 0 0 9 1 】

プロテイナーゼ消化を介した可溶化は、様々なプロテイナーゼ酵素 (ペプチダーゼ又は

10

20

30

40

50

プロテアーゼ酵素としても知られる)を利用することができ、これには、限定されないが、プロテイナーゼK、エンドプロテイナーゼトリプシン、キモトリプシン、エンドプロテイナーゼA s p - N、エンドプロテイナーゼA r g - C、エンドプロテイナーゼG l u - C、エンドプロテイナーゼL y s - C、テルモリジン、パパイン、サブチリシン、クロストリパイン、カルボキシペプチダーゼB、カルボキシペプチダーゼP、カルボキシペプチダーゼY、カテプシンC、アシルアミノ酸を放出する酵素、及びピログルタメートアミノペプチダーゼのうち1つ以上が含まれる。プロテイナーゼ酵素は、セリンプロテアーゼ、システインプロテアーゼ、トレオニンプロテアーゼ、アスパラギン酸プロテアーゼ、グルタミン酸プロテアーゼ、メタロプロテアーゼ、又はアスパラギンペプチドリアーゼであり得る。

10

【0092】

プロテイナーゼ消化を介する可溶化のための典型的なプロトコルは、包埋材料(例えばパラフィン)の除去、プロテイナーゼ消化、可溶化クロマチンの回復(例えば、S P R I ピーズなどのカルボキシル化ビーズによる)、及びライブラリ調製物の配列決定を含み得る。例えば、最初に、組織材料はチューブ(例えば1.5 mLのE p p e n d o r f チューブ)に入れることができる。その後、包埋材料(例えばパラフィン)は、キシレン、H e m o - D e、又はリモネンなどの溶媒を使用して溶かすことができる。エタノール(例えば100%のE t O H)は溶媒を除去するために使用され、サンプルはエタノールを除去するために乾燥され得る。その後、サンプルはプロテイナーゼ酵素(例えばプロテイナーゼK)で消化され得る。この結果、可溶化されている組織サンプルの大部分又は全てが

20

【0093】

超音波処理を介する可溶化のための典型的なプロトコルは、包埋材料(例えばパラフィン)の除去、溶解、均質化、超音波処理、可溶化クロマチンの回復(例えば、S P R I ピーズなどのカルボキシル化ビーズによる)、及びライブラリ調製物の配列決定を含み得る。例えば、最初に、包埋材料(例えばパラフィン)は、キシレン、H e m o - D e、又はリモネンなどの溶媒を使用して溶かすことができる。その後、組織標本は、例えば100%のエタノールから純水までの異なるエタノール濃度の連続洗浄において再水和され得る。その後、組織材料はチューブに入れられ、(例えば1時間にわたり)溶解緩衝液中でインキュベートされ得る。その後、組織は、消化緩衝液(例えばM N a s e 消化緩衝液)などの緩衝液の中で再懸濁され得る。その後、サンプルは、D o u n c e 均質化を含むがこれに限定されない方法によって均質化され得る。その後、サンプルは、超音波処理緩衝液中で超音波処理され且つ再懸濁され得る。その後、超音波処理サイクル(例えば最高の力で30秒)は、十分な可溶化クロマチンを得るのに必要とされるのと同じくらいのサイクル(例えば10サイクル、20サイクル、30サイクル、40サイクル)にわたり繰り返され得る。その後、可溶性分画が再生され得る。

30

【0094】

可溶化の後、サンプルは、可溶化クロマチンの回復(例えば、固相可逆的固定化(S P R I)ビーズへの結合による)、本明細書に記載されるC h i c a g o ライブラリなどの配列決定ライブラリの調製(例えば、核酸の切断、タグ付け、及びライゲーション)、配列決定(例えば、長距離情報を含む)、及び配列のアセンブリなどの、本明細書で議論された方法に従って更に処理され得る。

40

【0095】

サイズ選択

保存された(例えばF F P E)生物学的サンプルから得た核酸は、分析に適切な断片を産生するために断片化され得る。鋳型核酸は、様々な機械的、化学的、及び/又は酵素的な方法を使用して、望ましい長さに断片化又は切断され得る。DNAは、超音波処理、例えばC o v a r i s 方法、D N a s e への簡単な暴露、又は1つ以上の制限酵素、或いは

50

トランスポサゼ又はニッキング酵素の使用を介して無作為に切断され得る。RNAは、RNAse、熱、そしてマグネシウムへの簡単な暴露、又は切断によって断片化され得る。RNAはcDNAへと変換され得る。断片化が利用される場合、RNAは断片化の前又は後に、cDNAへと変換されてもよい。幾つかの実施形態において、生物学的サンプルの核酸は超音波処理により断片化される。他の実施形態では、核酸は、ハイドロシェア機器により断片化される。一般的に、個々の核酸鋳型分子は約2 kbから約40 kbの塩基であり得る。様々な実施形態において、核酸は約6 kb - 10 kbの断片であり得る。核酸分子は、一本鎖、二本鎖、又は、一本鎖領域を含む二本鎖（例えばステム構造とループ構造）でもよい。

【0096】

幾つかの実施形態において、架橋されたDNA分子は、サイズ選択工程に晒され得る。核酸のサイズ選択は、一定のサイズより下又は上の架橋されたDNA分子に対して行われてもよい。サイズ選択は、架橋の頻度及び/又は断片化方法により、例えば、フリーケンシー又はレアカッター制限酵素の選択により、更に影響を受けることもある。幾つかの実施形態において、組成物は、約1 kb ~ 5 Mb、約5 kb ~ 5 Mb、約5 kb ~ 2 Mb、約10 kb ~ 2 Mb、約10 kb ~ 1 Mb、約20 kb ~ 1 Mb、約20 kb ~ 500 kb、約50 kb ~ 500 kb、約50 kb ~ 200 kb、約60 kb ~ 200 kb、約60 kb ~ 150 kb、約80 kb ~ 150 kb、約80 kb ~ 120 kb、又は約100 kb ~ 120 kb、或いはこれらの値の何れかにより境界を付けた任意の範囲（例えば約150 kb ~ 1 Mb）の範囲で、DNA分子を架橋して調製され得る。

【0097】

幾つかの実施形態において、サンプルのポリヌクレオチドは、1以上の特定の大きさの範囲の断片化DNA分子の集団へと断片化される。幾つかの実施形態において、断片は、少なくとも約1、約2、約5、約10、約20、約50、約100、約200、約500、約1000、約2000、約5000、約10,000、約20,000、約50,000、約100,000、約200,000、約500,000、約1,000,000、約2,000,000、約5,000,000、約10,000,000、又はそれ以上の、出発DNAのゲノム等価物から生成され得る。断片化は、化学的、酵素的、及び機械的な断片化を含む、当該技術分野で既知の方法により遂行され得る。幾つかの実施形態において、断片は、平均の長さが約10 ~ 約10,000、約20,000、約30,000、約40,000、約50,000、約60,000、約70,000、約80,000、約90,000、約100,000、約150,000、約200,000、約300,000、約400,000、約500,000、約600,000、約700,000、約800,000、約900,000、約1,000,000、約2,000,000、約5,000,000、約10,000,000、又はそれ以上であるヌクレオチドを有する。幾つかの実施形態において、断片は、約1 kb ~ 約10 Mbの平均長さを有している。幾つかの実施形態において、断片は、約1 kb ~ 5 Mb、約5 kb ~ 5 Mb、約5 kb ~ 2 Mb、約10 kb ~ 2 Mb、約10 kb ~ 1 Mb、約20 kb ~ 1 Mb、約20 kb ~ 500 kb、約50 kb ~ 500 kb、約50 kb ~ 200 kb、約60 kb ~ 200 kb、約60 kb ~ 150 kb、約80 kb ~ 150 kb、約80 kb ~ 120 kb、又は約100 kb ~ 120 kb、或いはこれらの値の何れかにより境界を付けた任意の範囲（例えば約60 ~ 120 kb）の平均長さを有している。幾つかの実施形態において、断片は、約10 Mb未満、約5 Mb未満、約1 Mb未満、約500 kb未満、約200 kb未満、約100 kb未満、又は約50 kb未満の平均長さを有している。他の実施形態において、断片は、約5 kbより長い、約10 kbより長い、約50 kbより長い、約100 kbより長い、約200 kbより長い、約500 kbより長い、約1 Mbより長い、約5 Mbより長い、又は約10 Mbより長い平均長さを有している。幾つかの実施形態において、断片化は、サンプルのDNA分子の音響超音波処理への暴露を機械的に含んで、遂行される。幾つかの実施形態において、断片化は、二本鎖核酸切断を生成するために、1以上の酵素に適切な条件下で1以上の酵素によりサンプルDNA分子を処理する

10

20

30

40

50

工程を含む。DNA断片の生成に有用な酵素の例は、配列に特異的な及び配列に特異的でないヌクレアーゼを含んでいる。ヌクレアーゼの非限定的な例は、DNase I、フラグメンターゼ、制限エンドヌクレアーゼ、それらの変異体、及びそれらの組み合わせを含む。例えば、DNase Iでの消化は、 Mg^{++} が無い状態及び Mn^{++} がある状態で、DNA中のランダムな二本鎖切断を誘導することができる。幾つかの実施形態において、断片化は、1以上の制限エンドヌクレアーゼによりサンプルDNA分子を処理する工程を含む。断片化は、5'オーバーハング、3'オーバーハング、平滑末端、又はそれらの組み合わせを持つ断片を産生することができる。幾つかの実施形態において、断片化が1以上の制限エンドヌクレアーゼの使用を含むなどの場合に、サンプルDNA分子の切断は、オーバーハングに予測可能な配列を持たせる。幾つかの実施形態において、前記方法は、アガロースゲルからのカラム精製又は単離などの標準の方法を介して、断片のサイズを選択する工程を含む。

10

【0098】

配列決定ライブラリの調製

図1Bは、クロマチンベースの次世代配列決定(NGS)ライブラリ調製(例えば「Chicago」)の典型的な模式図を示す。第1の工程(111)において、クロマチンヌクレアーゼ(青色の円)は、架橋され(赤線)、クロマチン凝集体を形成する。第2の工程(112)において、クロマチン凝集体は制限エンドヌクレアーゼで切断される。第3の工程(113)において、切断末端は、平滑末端化され、ライゲートされ、及び(例えばピオチンで)印をつけられる(小さな緑色の円)。第4の工程(114)において、平滑末端は無作為にライゲートされ、短距離、中距離、及び長距離の会合を形成する(赤色のアスタリスクはライゲーション事象を示す)。第5の工程(115)において、架橋は逆転され、DNAは精製され、情報のライゲーション含有断片はマーカースのプルダウンのために選択される。その後、従来の配列決定ライブラリ調製が行われ得る。結果として生じるリードペアは、最大サイズの入力DNAにまでゲノム距離をまたぐことができる。そのようなライブラリは、染色体スケールの超足場(chromosome-scale super-scaffolds)を備えた高度に連続するゲノムアセンブリを構築するために使用され得る。

20

【0099】

図1Cは、保存されたサンプル(例えばFFPEサンプル)からのクロマチン抽出及びライブラリ調製(例えばChicagoライブラリ調製)のためのワークフローの典型的な模式図を示す。保存されたサンプルは、長距離ゲノム連鎖情報を生成及び配列決定する方法を後に受けることができる、固定されたクロマチンを抽出するために処理され得る。例えば、保存されたサンプル(121)は、抽出されたクロマチン(122)及び断片化されたクロマチン(例えばDpnIIなどの制限酵素で)を有し得る。クロマチンは架橋(123)を含み得る。オーバーハング(例えば、4bpの5'オーバーハング)は、ピオチン化ヌクレオチド(124)を含むヌクレオチド混合物で満たされ得る。その後、平滑末端はライゲートされ(125)、及びマーカース(例えばピオチン)は(例えばストレプトアビジンを使用して)プルダウンされ得る(126)。その後、印を付けていない(例えば、非ピオチン化)平滑末端は除去され、配列決定アダプター(例えば、イルミナ配列決定アダプター、Pacific Biosciences配列決定アダプター、ナノ細孔配列決定アダプター)が付けられ、配列決定ライブラリ(127)が調製され得る。ライブラリはピオチン化されライゲートされた結合を含む分子のために富化され、(例えばPCRにより)増幅され、(例えば、MiSeq又はHiSeqなどのイルミナシークエンサーを使用して、Pacific Biosciencesのロングリードシークエンサーを使用して、Oxford Nanopore又はGeniaなどのナノ細孔シークエンサーを使用して)配列決定され得る。場合によっては、Pacific Biosciences又はナノ細孔シークエンサーのようなロングリードシークエンサーを使用した時、多数の分子が配列決定前により長い分子に結合(例えば、ライゲート)され得る。

30

40

【0100】

50

富化は、対象の遺伝領域のために、標識化ヌクレオチド（例えばビオチン化ヌクレオチド、後成的に修飾したヌクレオチド）の富化の代わりに、又はそれに加えて行われ得る。例えば、サンプル又はライブラリは、融合遺伝子の既知の関連する半分を標的とするなどにより、融合遺伝子のために富化され得る。本明細書で議論されるような他の遺伝子及びゲノム特徴も、富化のために標的とされ得る。

【0101】

多くの場合、固定液剤は、精製工程の一部として以前に得られたサンプル（FFPEサンプルなど）に加えられない。むしろ、元のサンプル保存プロセスに準じて以前に生成された架橋は、本明細書中の単離されるDNAタンパク質（例えばクロマチン）複合体を安定させるために依存することができ、抽出プロセスは、相当な量の新たな複合体を生成するのではなく、結合された複合体を保存する。その後、溶解緩衝液中で可溶化されたサンプルの分画は、本明細書に開示される方法の何れかによって処理される。

【0102】

代替的に、幾つかの実施形態において、インビトロの近接ライゲーション（例えば、Chicagoのインビトロの近接ライゲーション）又は他のタンパク質DNA複合体のタグ付け方法が、DNAを含む保存されたサンプル（FFPEの保存されたサンプルなど）から抽出された高品質の核酸から生成された再構成クロマチンからリードペアライブラリを生成するために、使用される。例えば、保存されたサンプル（例えばFFPEサンプル）は、抽出プロセスにてDNAの損傷を最小限にするようにDNAなどの核酸を抽出するために処理され得る。場合によっては、ボルテックス、剪断、沸騰、高温のインキュベーション、又はDNase関連の酵素処理のうち1つ以上が、単離されたネイキッドDNAに対する損傷を減らすように、核酸抽出プロトコルから除外される。回復された単離DNAは、物理的連鎖、フェーズ、又はゲノム構造情報を保存するのに十分な品質となり得る。抽出された核酸は、再構成クロマチンを生成するために希釈及び使用され得（例えば、その全体において引用により本明細書に組み込まれる、2014年8月7日公開のPCT公報WO2014/121091、その全体において引用により本明細書に組み込まれる、2016年2月4日公開のPCT公報WO2016/019360にて教示されるものといった方法を使用する）、その結果、DNA：タンパク質の複合体は単一のDNA分子及び少なくとも1つのDNA結合部分を含む。再構成クロマチンは、それらの共通のリン酸ジエステル骨格から独立して、同じDNA分子内のDNA配列の近位の情報を保存するために、ホルムアルデヒドなどで架橋され得る。重要なことに、架橋は、保存されたサンプル（FFPEサンプルなど）からの単離後に、保存されたサンプルから抽出されたDNA二対し行われ得る。DNAタンパク質複合体の単離に関して上記で議論されるように、多くの場合、架橋剤は遮断プロセス中に加えられない。これら架橋され再構成された複合体は、ビオチン、メチル化、スルフィル化（sulfylation）、アセチル化、又は他の塩基修飾などで標識され、その後、ビオチン標識化の場合にストレプトアビジンビーズなどで単離され得る。その後、単離された複合体は、ビオチン化ヌクレオチド又は言及されるような他のヌクレオチドなどで標識されたヌクレオチドで後に満たされる、遊離付着末端を生成するために制限酵素で消化され得る。

【0103】

DNA：タンパク質の複合体における暴露されたDNA末端は、（例えば保存されたサンプルの分解から）先在していても、或いは本明細書に開示されるプロトコルの結果（例えば酵素的又は物理的な切断）でも、同じDNA分子の内のDNA配列の間のペアエンドを生成するためにライゲートされ得る。これらライゲートされたペアエンドは大抵、元々はDNA分子上で互いに隣接していない場合がある。ペアエンドは、場合によっては、付着末端への充填の結果として平滑末端化され得る。

【0104】

代替的に又は付加的に、暴露された核酸複合体末端は、本明細書で議論されるような終点オリゴヌクレオチドを通じて互いにライゲートされ、或いは、核酸断片が共通のDNAタンパク質複合体へと識別可能にマッピングされるようにオリゴヌクレオチドタグの集団

10

20

30

40

50

を使用してタグ付けされ得る。場合によっては、ペアエンドリードは、直接ライゲートされるDNA複合体の切断末端ではなく、共通の終点オリゴヌクレオチドに結合される切断末端から生成される。終点オリゴヌクレオチドは、フェーズを保存する再編成を受けるサンプル分子の2つの切断された内部末端を架橋するように、標的ポリヌクレオチドに結合され得るオリゴヌクレオチドを含んでいる。終点オリゴヌクレオチドは、DNA、RNA、ヌクレオチドアナログ、非標準のヌクレオチド、標識されたヌクレオチド、修飾されたヌクレオチド、又はそれらの組み合わせを含み得る。多くの例において、二本鎖終点オリゴヌクレオチドは、互いにハイブリダイズされる2つの別個のオリゴヌクレオチド（「オリゴヌクレオチドデュプレックス」とも称される）を含み、ハイブリダイゼーションは、1つ以上の平滑末端、1つ以上の3'オーバーハング、1つ以上の5'オーバーハング、不適正及び/又は非対合のヌクレオチドから結果として生じる1つ以上のバルジ、又はこれらの任意の組み合わせを残し得る。幾つかの例において、異なる終点オリゴヌクレオチドが、連続的な反応において又は同時に、ポリヌクレオチドを標的とするために結合される。例えば、第1及び第2の終点オリゴヌクレオチドが同じ反応に加えられ得る。代替的に、終点オリゴの集団は場合によっては均一である。ゲノムの構造及び近接情報を保存及び判定するのに使用される終点分子及び方法は、以前に記載されている（米国仮特許出願第62/298906号、第62/298966号、及び第62/305957号、これらの3つ全てはそれらの全体において本明細書に組み込まれる）。幾つかの終点オリゴヌクレオチドは、終点オリゴヌクレオチドを含むライブラリの断片が容易に単離されるように、ビオチンタグなどの分離を容易にするためのタグ又は標識を含む。代替的なタグは、メチル化、アセチル化、又は他の塩基修飾を含むがこれらに限定されない。一般的に、終点オリゴヌクレオチドは、暴露された核酸末端にライゲートされるが、ライブラリに終点オリゴヌクレオチドを組み込む代替的な手法も考慮される。

【0105】

付着末端に充填するために使用されるものなどのヌクレオチドが、標識され得る。標識されたヌクレオチドは、ビオチン化され、硫酸化され、フルオロフォアに結合され、脱リン酸化され、又は他のあらゆるヌクレオチド修飾を受ける場合がある。ヌクレオチド修飾はまた、メチル化（例えば5-mC、5-hmC、5-fC、5-caC、4-mC、6-mA、8-oxoG、8-oxoA）などの後成的修飾を含み得る。標識又は修飾は、ナノ細孔配列決定によって検知可能な後成的修飾などの、配列決定中に検知可能なものから選択することができ；このように、ライゲーション結合の位置は配列決定中に検出することができる。これらの標識又は修飾はまた、結合又は富化のために標的とされ得；例えば、メチル-シトシンを標的とする抗体は、メチル-シントンで充填された平滑末端を捕捉し、標的化し、結合し、又は標識するために使用され得る。非天然ヌクレオチド、非標準ヌクレオチド、又は修飾されたヌクレオチド、及び核酸アナログも、平滑末端の充填の位置を標識するために使用され得る。非標準ヌクレオチド又は修飾されたヌクレオチドは、プソイドウリジン（ ）、ジヒドロウリジン（D）、イノシン（I）、7-メチルグアノシン（m7G）、キサンチン、ヒポキサンチン、プリン、2,6-ジアミノプリン、及び6,8-ジアミノプリンを含み得る。核酸アナログは、ペプチド核酸（PNA）、モルホリノ及びロックド核酸（LNA）、グリコール核酸（GNA）、及びトレオース核酸（TNA）を含み得る。場合によっては、オーバーハングは、ビオチンの無いdNTPなどの、標識されていないdNTPで充填される。場合によっては、トランスポゾンでの切断など、充填を必要としない平滑末端が生成される。トランスポサゼが2つの結合されていない終点オリゴヌクレオチドを挿入すると、このような遊離平滑末端が生成される。しかし、終点オリゴヌクレオチドは、望ましいものとして付着末端又は平滑末端を持つように合成され得る。ヒストンなどのサンプル核酸に関連付けられるタンパク質も、修飾され得る。例えば、ヒストンは、アセチル化（例えばリジン残基にて）及び/又はメチル化（例えばリジンとアルギニン残基にて）され得る。

【0106】

幾つかの実施形態において、Hi-C又は他のライゲーション或いはタグ付けにより媒

10

20

30

40

50

介される方法は、架橋されるクロマチン、例えばサンプル保存に準じて架橋されるクロマチンの自然発生からリードペアライブラリを生成するために使用され得る。DNAは、保存プロセス中にネイティブクロマチン構造を保存するためにホルムアルデヒドなどで架橋され得る。架橋されたDNA-タンパク質構造を妨壊することなく、パラフィンなどのサンプル防腐剤又は固定液からこれらDNA-タンパク質構造を分離し、それにより、リン酸ジエステル骨格から独立したDNA分子間の近接情報を保存するために、上記のように抽出を行うことができる。これら橋架構造は、ビオチンで標識したヌクレオチドなどのタグ付けされたヌクレオチドで後に充填される、遊離付着末端を生成するために、制限酵素で消化され得る。結果として生じる平滑末端は、DNA断片のペアエンドを生成するために共にライゲートされ得る。これらペアエンドは、クロマチン構造中で互いに近接するDNA分子を表す。Hi-C方法及びその変形は、当該技術分野で既知である（全体において本明細書に組み込まれる、Lieberman-Aiden et al., 2009, Science 326, 289; 全体において本明細書に組み込まれる、US 20130096009）。

【0107】

ペアエンドは、（例えばプロテイナーゼKなどのプロテイナーゼでの）酵素消化などにより、クロマチンタンパク質から放たれ得る。放たれたペアエンドは、標識されたヌクレオチドのみがライゲートされたペアエンド間に存在するように、残りの遊離末端から標識されたヌクレオチドを取り除くためにエキソヌクレアーゼで処理され得る。その後、これらペアエンドは、ビオチン標識の場合にストレプトアビジンビーズなどで精製され得る。精製はまた、SPRIビーズ（例えば、カルボキシル化ビーズ）などで、又は電気泳動法（例えばゲル電気泳動、キャピラリー電気泳動）を介した、他の手段により処理され得る。その後、ペアエンドは配列決定のために調製され得る。例えば、ペアエンドは、配列決定アダプターに付けられ、その後、リードペアライブラリを生成するために配列決定され得る。Chicagoのインビトロの近接ライゲーション方法が、以前に記載されている（例えば、全体において引用により本明細書に組み込まれる米国特許出願公開第20140220587号；全体において引用により本明細書に組み込まれる米国特許出願公開第20150363550号）。

【0108】

典型的な実施形態において、ライブラリは、1つの切片につき約 3×10^5 の細胞を持つ厚さ15 - 20ミクロンの切片において、以前にFFPEに包埋された細胞から作られる。代替的に、FFPEに包埋された細胞は、1つのセクションにつき約 10^3 、 10^4 、 10^5 、 10^6 、又は 10^7 の細胞を持つ、厚さ1 - 5、5 - 10、10 - 15、15 - 20、25 - 30、又は45 - 50のセクションで提供される。場合によっては、サンプルは、AJGIAB（「Genome In A Bottle」）サンプルGM24149（父）及びGM24385（子）である。セクションは、包埋材料、例えばキシレン、トルエン、ベンゼンを除去するために溶媒で洗浄される。溶媒は、エタノール溶媒でセクションを洗浄することによって除去され、場合によっては100%のエタノールがセクションを洗浄するために使用される。その後、パラフィンの無い組織サンプルは、緩衝液、例えば洗浄性の緩衝液中で可溶化される。その後、サンプル中の核酸は、エンドヌクレアーゼ、例えばMboIなどの制限酵素で消化される。平滑末端は、ビオチン化dNTPなどのDNAポリメラーゼ及びヌクレオチドを使用する制限酵素消化から結果として生じるオーバーハングに充填することにより、消化された核酸の中で作られる。平滑末端は、平滑末端ライゲーションを好む反応においてDNAリガーゼ、例えばT4 DNAリガーゼを使用して共にライゲートされ、その結果、DNAのビオチン化断片がもたらされる。これら断片は配列決定反応における使用のために調製される。

【0109】

配列決定

本明細書にはまた、物理的連鎖情報などのゲノムの構造情報を持つ核酸配列決定ライブラリを生成するための方法及び組成物も、開示される。DNA複合体は、FFPE由来の

10

20

30

40

50

核酸サンプルなどの保存されたサンプルから生成される。ペアエンド、ライゲーション結合、終端末端、又は共通してタグ付けされた末端は、第1のセグメント及び第2のセグメントがリン酸ジエステル骨格結合から独立して共に保持され、暴露された末端がタグ付けされ、タグ結合が単離されるように、結合された核酸複合体の単離を介して生成される。タグ付けは様々に、別の暴露された末端を直接使用して1つの暴露された末端をタグ付けする工程を含み、その結果、結合は、結合の何れかの側での配列が、ゲノム足場上の遠位位置に相当するコンティグへとマッピングされ、足場形成されず (unscaffolded)、又は配置されていないゲノムにおける異なる染色体にマッピングされるという事実から、識別可能となる。代替的に、タグ付けは、終点オリゴを使用して暴露された末端を結合する工程、或いは、複合体の暴露された末端に共通のオリゴタグを加える工程を含み、その結果、タグ付けした末端に隣接する配列は共通のDNA複合体に確実にマッピングされ、それ故、DNA複合体が生成されたソースとなる核酸の共通のフェーズにマッピングされる。

【0110】

ペアエンド、鎖状体化された (concatamerized) ペアエンド、又は終点化された分子は、適切なショートリード又はロングリードの配列決定技術のプラットフォームを使用して配列決定され、配列リードはその後分析される。

【0111】

場合によっては、複数のペアエンド分子が本明細書に記載されるように生成され、その後、ショートリード配列決定技術を使用して配列決定される。これらの場合、ペアエンドライゲーション結合にわたる何れかの短い配列リードが生成され、又は、ペアエンド断片の各末端からショートリードが、リードペアを作るために生成される。第1及び第2の核酸セグメントからの配列が単一の配列リード又はリードペアにおいて検出される場合、第1及び第2の核酸セグメントが入力DNAサンプル中の同じDNA分子上で同フェーズにあることが、判定される。そのような場合、生成された配列ライブラリは、DNAセグメントのためのフェーズ及び構造情報をもたらす。

【0112】

与えられた終点化された分子配列リード又はリードペアについて、終点要素により局所的に中断されない配列セグメントが観察される。これらセグメント中の配列は、同フェーズにあり、且つ局所的に正確に順序づけられ且つ配向されると推定される。セグメントは、終点オリゴにより分離されると観察される。終点オリゴの何れかの側のセグメントは、共通のサンプル核酸分子上で同フェーズにあるが、終点分子上で互いに対して正確に順序付けられ且つ配向されないと推測される。再編成の利点は、互いに遠く離れて位置付けられたセグメントが時々近接することであり、その結果これらは、サンプル分子においてフェーズ配列と同一で困難な長距離で分離されたとしても、共通のリードにて読み取られ、確実に共通のフェーズへと割り当てられる。別の利点は、セグメント配列自体が元のサンプル配列の大半、略全て、又は全てを含むことであり、その結果、フェーズ情報に加えて、場合によっては、デノボ配列アセンブリを行うのに十分なコンティグ配列が判定される。このデノボ配列は、新たな足場又はコンティグセットを生成し、或いは、以前に又は独立して生成されたコンティグ又は足場配列セットを増大させるために、随意に使用される。

【0113】

場合によっては、複数の終点化されたDNA分子が、本明細書に開示されるように生成され、単一の長い核酸分子に鎖状体化され、或いは、単一の再編成された長い分子として剪断又は切断無しに保存され、及びその後、ロングリード配列決定技術を使用して配列決定される。終点化された分子はそれぞれ配列決定され、配列リードが分析される。好ましい例において、配列リードは、配列反応について平均10kbとなる。他の例において、配列リードは、平均約5kb、6kb、7kb、8kb、9kb、10kb、11kb、12kb、13kb、14kb、15kb、16kb、17kb、18kb、19kb、20kb、21kb、22kb、25kb、30kb、35kb、40kb、又はそれ以上となる。好ましい例において、配列リードは、終点オリゴ配列に結合される、少なくとも

10

20

30

40

50

も500の塩基の第1のセグメント、及び500の塩基の第2のセグメントを含むと識別される。他の例において、配列リードは、少なくとも約100の塩基、200の塩基、300の塩基、400の塩基、500の塩基、600の塩基、700の塩基、800の塩基、900の塩基、1000の塩基、又はそれ以上の第1のDNAセグメント、及び、少なくとも約100の塩基、200の塩基、300の塩基、400の塩基、500の塩基、600の塩基、700の塩基、800の塩基、900の塩基、1000の塩基、又はそれ以上の第2のDNAセグメントを含む。幾つかの例において、第1及び第2のセグメントの配列は足場ゲノムにマッピングされ、少なくとも100kb分離されるコンティグにマッピングされると見出される。他の例において、分離距離は、8kb、9kb、10kb、12.5kb、15kb、17.5kb、20kb、25kb、30kb、35kb、40kb、45kb、50kb、60kb、70kb、80kb、90kb、100kb、125kb、150kb、200kb、300kb、400kb、500kb、600kb、700kb、800kb、900kb、1Mb、又はそれ以上である。大半の場合、第1のコンティグ及び第2のコンティグはそれぞれ、単一のヘテロ接合位置を含み、そのフェーズは足場では判定されない。好ましい例において、第1のコンティグのヘテロ接合位置にはロングリードの第1のセグメントが跨り、第2のコンティグのヘテロ接合位置にはロングリードの第2のセグメントが跨る。そのような場合、各リードが、それらのコンティグの各ヘテロ接合領域を跨り、リードセグメントの配列は、第1のコンティグの第1の対立遺伝子及び第2のコンティグの第1の対立遺伝子が同フェーズにあることを示している。第1及び第2の核酸セグメントからの配列が単一の長い配列リードにおいて検出される場合、第1及び第2の核酸セグメントが入力DNAサンプル中の同じDNA分子に含まれることが、判定される。これら実施形態において、本明細書に開示された方法及び組成物により生成された核酸配列ライブラリは、ゲノム足場上で互いに遠く離れて位置付けられるコンティグに関するフェーズ情報を提供する。

【0114】

代替的に、複数のペアエンド分子が本明細書に記載されるように生成され、その後、ロングリード配列決定技術を使用して配列決定される。場合によっては、ライブラリの平均リード長は、約1kbであると判定される。他の場合、ライブラリの平均リード長は、約100bp、200bp、300bp、400bp、500bp、600bp、700bp、800bp、900bp、1kb、1.1kb、1.2kb、1.3kb、1.4kb、1.5kb、又はそれ以上である。大半の例において、ペアエンド分子は、入力DNAサンプル内で、同フェーズにあり且つ10kbより長い距離だけ分離される、第1のDNAセグメント及び第2のDNAセグメントを含む。幾つかの例において、2つのそのようなDNAセグメント間の分離距離は、約5kb、6kb、7kb、8kb、9kb、10kb、11kb、12kb、13kb、14kb、15kb、20kb、23kb、25kb、30kb、32kb、35kb、40kb、50kb、60kb、75kb、100kb、200kb、300kb、400kb、500kb、750kb、1Mbよりも長い。大半の場合、配列リードはペアエンド分子から生成され、その幾つかは、第1の核酸セグメントから少なくとも300の塩基の配列、及び第2核酸セグメントから少なくとも300の塩基の配列を含む。他の例において、配列リードは、少なくとも約50の塩基、100の塩基、150の塩基、200の塩基、250の塩基、300の塩基、350の塩基、400の塩基、450の塩基、500の塩基、550の塩基、600の塩基、650の塩基、700の塩基、750の塩基、800の塩基、又はそれ以上の第1のDNAセグメント、及び、少なくとも約50の塩基、100の塩基、150の塩基、200の塩基、250の塩基、300の塩基、350の塩基、400の塩基、450の塩基、500の塩基、550の塩基、600の塩基、650の塩基、700の塩基、750の塩基、800の塩基、又はそれ以上の第2のDNAセグメントを含む。第1及び第2の核酸セグメントからの配列が単一の配列リード又はリードペアにおいて検出される場合、第1及び第2の核酸セグメントが入力DNAサンプル中の同じDNA分子上で同フェーズにあることが、判定される。そのような場合、生成された配列ライブラリは、DNAセグメントを配

10

20

30

40

50

列決定ために使用される配列決定技術のリード長よりも長い長さを核酸サンプル中で分離される、DNAセグメントのフェーズ情報をもたらす。

【0115】

様々な実施形態において、本明細書に記載される又は当該技術分野で既知の適切な配列決定方法が、サンプル内の核酸分子から配列情報を得るために使用される。配列決定は、当該技術分野で周知の古典的なサンガー配列決定方法を通じて遂行され得る。配列決定 (Sequence) はまた、ハイスループットシステムを使用して遂行され、その一部は、増大する鎖への組み込みの直後又はその際の配列決定されたヌクレオチドの検出、即ち、リアルタイム又はほぼリアルタイムでの配列の検出を可能にする。場合によっては、ハイスループット配列決定は、1時間につき少なくとも1,000、少なくとも5,000、少なくとも10,000、少なくとも20,000、少なくとも30,000、少なくとも40,000、少なくとも50,000、少なくとも100,000、又は少なくとも500,000の配列リードを生成し；ここで、配列決定リードは、1つのリードにつき少なくとも約50、約60、約70、約80、約90、約100、約120、約150、約180、約210、約240、約270、約300、約350、約400、約450、約500、約600、約700、約800、又は約1000の塩基であり得る。

10

【0116】

幾つかの実施形態において、ハイスループット配列決定は、HiSeq 2500、HiSeq 1500、HiSeq 2000、又はHiSeq 1000の機械を使用するものなど、Illumina's Genome Analyzer IIX、MiSeq パーソナルシーケンサー、又はHiSeqシステムにより利用可能な技術の使用を含む。これらの機械は、合成化学による可逆的なターミネーターベースの配列を使用する。これらの機械は、8日で2000億以上のDNAのリードを行うことができる。より小さなシステムが、3、2、1日以下の時間内での実行のために利用され得る。

20

【0117】

幾つかの実施形態において、ハイスループット配列決定は、ABI Solid Systemにより利用可能な技術の使用を含む。クローン的に増幅されたDNA断片の超並列配列決定を可能にする、この遺伝子解析プラットフォームは、ビーズに連結した。配列決定の方法論は、染料で標識したオリゴヌクレオチドでの連続的なライゲーシオンに基づく。

【0118】

次世代の配列決定は、(例えば、Life Technologies (Ion Torrent) の技術を使用する) イオン半導体配列決定を含み得る。イオン半導体配列決定は、ヌクレオチドがDNAの鎖へ組み込まれる時にイオンが放出され得るという事実を利用することができる。イオン半導体配列決定を行なうために、微細加工されたウェルの高密度アレイを形成することができる。各ウェルは、単一のDNA鋳型を保持することができる。ウェルの下はイオン感受性の層であり、イオン感受性の層の下はイオンセンサーであり得る。ヌクレオチドがDNAに加えられると、H⁺が放出され、pHの変化として測定され得る。H⁺イオンは電圧に変換され、半導体センサーにより記録され得る。アレイチップは、1つのヌクレオチドで連続して氾濫させられる (flooded) 場合がある。走査、光、又はカメラは必要ではない。場合によっては、ION PROTON (商標) シーケンサーが核酸を配列決定するために使用される。場合によっては、ION PGM (商標) シーケンサーが使用される。Ion Torrent Personal Genome Machine (PGM)。PGMは2時間で1000万のリードを行うことができる。

30

40

【0119】

幾つかの実施形態において、ハイスループット配列決定は、合成による単一分子配列決定 (SMSS) の方法など、Helicos BioSciences Corporation (Cambridge, Massachusetts) により利用可能な技術の使用を含む。SMSSは、最大24時間で全体のヒトゲノムの配列決定を可能にするため、固有なものである。最終的に、SMSSは、米国特許出願公開第2006002471

50

1号；第20060024678号；第20060012793号；第20060012784号；及び第20050100932号に部分的に記載されている。

【0120】

幾つかの実施形態において、ハイスループット配列決定は、PicoTiterPlate装置などの454 Lifesciences, Inc. (Branford, Connecticut)により利用可能な技術の使用を含み、これは、機器におけるCCDカメラにより記録される配列決定反応により生成された、化学発光シグナルを伝達するファイバオプティックプレートを含む。このファイバオプティクスの使用は、4.5時間で最低2000万の塩基対の検出を可能にする。

【0121】

ビーズの増幅、その後でファイバオプティクスの検出を使用する方法は、Marguiles, M., et al. "Genome sequencing in microfabricated high-density picolitre reactors", Nature, doi:10.1038/nature03959; 及び同様に、米国特許出願公開第20020012930号；第20030068629号；第20030100102号；第20030148344号；第20040248161号；第20050079510号；第20050124022号；及び第20060078909号に記載されている。

【0122】

幾つかの実施形態において、ハイスループット配列決定は、Clonal Single Molecule Array (Solexa, Inc.)、又は可逆的なターミネーター化学を利用する、合成による配列決定 (SBS) を使用して行なわれる。これらの技術は、米国特許第6,969,488号；第6,897,023号；第6,833,246号；第6,787,308号；及び、米国特許出願公開第20040106110号；第20030064398号；第20030022207号；及びConstans, A., The Scientist 2003, 17(13):36に部分的に記載されている。

【0123】

次世代の配列決定技術は、Pacific Biosciencesによるリアルタイム (SMRT (商標)) 技術を含み得る。SMRTにおいて、4つのDNA塩基の各々は、4つの異なる蛍光染料の1つに付けられ得る。これらの染料はホスホ連結 (phospho linked) され得る。単一のDNAポリメラーゼは、ゼロモード導波路 (ZMW) の底部にある鋳型的一本鎖DNAの単一分子で固定され得る。ZMWは、(マイクロ秒で) ZMWの中及び外で急速に拡散することができる蛍光ヌクレオチドのバックグラウンドに対する、DNAポリメラーゼによる単一のヌクレオチドの組み込みの観察を可能にする、制限構造となり得る。増大している鎖にヌクレオチドを組み込むのに数ミリ秒かかる場合がある。この間に、蛍光標識は励起され、蛍光シグナルを生成することができ、蛍光タグは切断され得る。ZMWは下から照らすことができる。励起ビームからの減じられた光は、より低い20-30nmのZMWそれぞれに浸透し得る。20zeptoliter (10⁻²¹ リットル) の検出限界を持つ顕微鏡が作成され得る。小さな検出量は、バックグラウンドノイズの減少において1000倍の改善をもたらし得る。染料の対応する蛍光の検出は、どの塩基が組み込まれるかを示すことができる。このプロセスは繰り返すことができる。

【0124】

場合によっては、次世代の配列決定はナノ細孔配列決定である (例えば、Soni GV and Meller A. (2007) Clin Chem 53: 1996-2001を参照)。ナノ細孔は、直径約1ナノメートルの小さな穴であり得る。導電性流体におけるナノ細孔の浸漬、及びそれをわたる電位の適用は、結果として、ナノ細孔を介したイオンの伝導が原因で、僅かな電流をもたらし得る。流れる電流の量はナノ細孔の大きさに敏感であり得る。DNA分子がナノ細孔を通ると、DNA分子上のヌクレオチドはそれぞ

10

20

30

40

50

れ、異なる程度にまでナノ細孔を塞ぐことができる。故に、DNA分子がナノ細孔を通る時の、ナノ細孔を通る電流の変化は、DNA配列のリードを表すことができる。ナノ細孔配列決定技術は、Oxford Nanopore Technologies；例えばGridIONシステムに由来し得る。単一のナノ細孔は、マイクロウェルの上部にわたる高分子膜に挿入され得る。マイクロウェルはそれぞれ、個々の感知のための電極を有し得る。マイクロウェルは、1つのチップ当たり100,000以上のマイクロウェル（例えば200,000、300,000、400,000、500,000、600,000、700,000、800,000、900,000、又は1,000,000より多く）で、アレイチップへと組み立てられ得る。機器（又はノード）がチップを解析するために使用され得る。データはリアルタイムで解析され得る。1つ以上の機器は一度に操作され得る。ナノ細孔は、タンパク質ナノ細孔、例えばタンパク質アルファ溶血素、七量体タンパク質細孔であり得る。ナノ細孔は、ソリッドステートナノ細孔で出来ており、例えば、合成膜（例えばSiNx、又はSiO₂）に形成されるナノメートルサイズの穴であり得る。ナノ細孔は、ハイブリッド細孔（例えば、ソリッドステート膜へのタンパク質細孔の統合）であり得る。ナノ細孔は、集積センサー（例えば、トンネル電極検出器、容量検出器、又はグラフェン性ナノギャップ若しくはエッジ状態検出器（例えば、Gara et al. (2010) Nature vol. 67, doi: 10.1038/nature09379を参照））を持つナノ細孔であり得る。ナノ細孔は、特定の型の分子を解析するために官能化することができる（例えば、DNA、RNA又はタンパク質）。ナノ細孔配列決定は、「鎖配列決定」を含むことができ、そこでは完全なDNAポリマーは、DNAが細孔を転移させる際にリアルタイムで配列決定されながらタンパク質ナノ細孔を通過することができる。酵素は二本鎖DNAの鎖を分離することができ、ナノ細孔を通じて鎖を与えることができる。DNAは一方の末端でヘアピンを有することができる、システムは両方の鎖を読み取ることができる。場合によっては、ナノ細孔配列決定は、「エキソヌクレアーゼ配列決定」であり、そこでは個々のヌクレオチドは前進性エキソヌクレアーゼによってDNA鎖から切断されることができ、そのヌクレオチドはタンパク質ナノ細孔を通過することができる。ヌクレオチドは、細孔内の分子（例えば、シクロデキストラン）に一過的に結合することができる。電流の特性的断絶を使用して、塩基を同定することができる。

【0125】

GENIA製のナノ細孔配列決定技術を使用できる。操作したタンパク質細孔を、脂質二重層の膜に包埋することができる。「能動的制御」技術を使用して、効果的なナノ細孔-膜アセンブリ及びチャネルを通るDNA運動の制御を可能にする。場合によっては、ナノ細孔配列決定技術は、NABsys製である。ゲノムDNAは、平均長約100kbの鎖に断片化され得る。100kbの断片を一本鎖にし、その後6量体のプローブでハイブリダイズさせることができる。プローブを持つゲノム断片は、ナノ細孔を通り抜けることができ、電流対時間の追跡を作り出すことができる。電流追跡は、各ゲノム断片上のプローブ位置を提供することができる。ゲノム断片を並べて、ゲノムに対するプローブマップを作り出すことができる。そのプロセスは、プローブのライブラリに対して並列に行うことができる。各プローブに対するゲノム長のプローブマップを生成することができる。誤りは、「移動窓配列決定バイハイブリダイゼーション(Sequencing By Hybridization)(mwsBH)」と称されるプロセスで直すことができる。場合によっては、ナノ細孔配列決定技術は、IBM/Roché製である。電子ビームを使用して、マイクロチップにナノ細孔サイズの開口部を作製することができる。電界を使用して、ナノ細孔を通じてDNAを引き寄せるとはねじ込むことができる。ナノ細孔におけるDNAトランジスタ装置は、金属と誘電体が交互になったナノメートルサイズの層を含むことができる。DNA骨格中の別個の電荷を、電界によってDNAナノ細孔の内部に閉じ込めることができる。ゲート電圧をオフ/オンすることにより、DNA配列を読み取ることができる。

【0126】

次世代配列決定は、場合によってはDNAナノボール配列決定を含むことができる（例えば、Complete Genomicsによって実施されるように；例えば、Drmanac et al. (2010) Science 327: 78-81を参照）。DNAを、単離し、断片化し、サイズ選択することができる。例えば、DNAは、約500bpの平均長へと（例えば、超音波処理によって）断片化することができる。アダプター（Ad1）を、断片の末端に結合させることができる。アダプターを使用して、配列決定反応のためのアンカーにハイブリダイズさせることができる。各末端に結合したアダプターを持つDNAを、PCR増幅することができる。アダプター配列を修飾することができ、それにより相補的一本鎖末端が互いに結合して環状DNAを形成する。DNAをメチル化して、その後の工程において使用されるIIS型制限酵素による切断からこれを保護することができる。アダプター（例えば、右のアダプター）は、制限認識部位を有することができ、制限認識部位は非メチル化されたままであり得る。アダプター中にある非メチル化制限認識部位は、制限酵素（例えば、AclI）によって認識され得、DNAは、AclIによって右アダプターの右側13bpで切断されて、線状二本鎖DNAを形成することができる。右及び左のアダプターの第2巡（Ad2）を、線状DNAのいずれかの末端にライゲートすることができ、両方のアダプターが結合しているDNAは全て、PCR増幅することができる（例えば、PCRによって）。Ad2配列を修飾して、それらが互いに結合して環状DNAを形成することを可能にする。DNAはメチル化することができるが、制限酵素認識部位は、左Ad1アダプターにおいて非メチル化のままであり得る。制限酵素（例えば、AclI）を適用することができ、DNAは、Ad1の左側13bpで切断され、線状DNA断片を形成することができる。右及び左のアダプターの第3巡（Ad3）を、線状DNAの右及び左側面にライゲートすることができ、その結果生じる断片はPCR増幅することができる。アダプターは修飾することができ、それにより互いに結合し、環状DNAを形成することができる。IIE型制限酵素（例えばEcoPI5）を加えることができ；EcoPI5は、Ad3の左側に対してDNA26bpを、Ad2の右側に対して26bpを切断することができる。この切断は、DNAの大きなセグメントを除去し、DNAを再び線状化することができる。右及び左のアダプターの第4巡（Ad4）を、DNAにライゲートすることができ、DNAを（例えば、PCRによって）増幅し且つ修飾することができ、それによりそれらは互いに結合し、完成した環状DNA鑄型を形成する。

【0127】

ローリングサークル複製（例えばPhi29 DNAポリメラーゼを使用する）を使用して、DNAの小さな断片を増幅することができる。4つのアダプター配列は、ハイブリダイズできるバリンδροーム配列を含有することができ、一本鎖は、それ自体の上へと折りたたまれて、平均で直径およそ200～300ナノメートルとなり得るDNAナノボール（DNB（商標））を形成することができる。DNAナノボールは、マイクロアレイ（配列決定フローセル）に付けることができる（例えば、吸着により）。フローセルは、二酸化ケイ素、チタン、及びヘキサメチルジシラザン（HMDS）、並びにフォトレジスト材料でコーティングされたシリコンウエハであり得る。配列決定は、DNAに蛍光プローブをライゲートすることによる連鎖しない配列決定によって実施することができる。問い合わせられた位置の蛍光の色は、高解像度カメラによって可視化することができる。アダプター配列間のヌクレオチド配列の同一性を決定することができる。

【0128】

幾つかの実施形態において、ハイスループット配列決定は、AnyDot.chips（Genovox, Germany）を使用して行うことができる。特に、AnyDot.chipsは、ヌクレオチド蛍光シグナル検出を10x-50x増強できる。AnyDot.chip、及びそれらを使用するための方法は、国際出願公開WO02088382、WO03020968、WO03031947、WO2005044836、PCT/EP05/05657、PCT/EP05/05655；及びドイツ特許出願DE10149786、DE10214395、DE10356837、DE102004

10

20

30

40

50

009 704、DE10 2004 025 696、DE10 2004 025 746、DE10 2004 025 694、DE10 2004 025 695、DE10 2004 025 744、DE10 2004 025 745、及びDE10 2005 012 301に部分的に記載されている。

【0129】

他のハイスループット配列決定システムは、2001年2月16日のVenter, J., et al. Science; 2000年3月24日のAdams, M. et al. Science; 及び2003年1月のM. J. Levene, et al. Science 299: 682 - 686; 加えて米国特許出願公開第20030044781号及び第2006/0078937号に開示されるものを含む。そのようなシステム全体は、配列決定される鋳型核酸分子上の核酸重合酵素の活性がリアルタイムで追跡されるなど、核酸分子上で測定される重合反応による塩基の経時的な付加によって、複数の塩基を有する標的核酸分子を配列決定することを含む。次いで配列の塩基付加の各工程で、どの塩基が、核酸重合酵素の触媒活性により標的核酸の成長相補鎖に組み込まれているかについて同定することにより、配列を推論することができる。標的核酸分子複合体上のポリメラーゼは、標的核酸分子に沿って移動し、活性部位でオリゴヌクレオチドプライマーを伸長するのに適した位置に提供される。それぞれ識別可能な型のヌクレオチドアナログが標的核酸配列中の異なるヌクレオチドに対して相補的である、複数の標識型のヌクレオチドアナログが活性部位のすぐ近くに提供される。成長核酸鎖は、ポリメラーゼを使用して活性部位にて核酸鎖にヌクレオチドアナログを付加することによって伸長され、付加されるヌクレオチドアナログは、活性部位にて標的核酸のヌクレオチドに相補的である。重合工程の結果としてオリゴヌクレオチドプライマーに付加されたヌクレオチドアナログが、同定される。標識したヌクレオチドアナログを提供する工程と、成長核酸鎖を重合させる工程と、付加されたヌクレオチドアナログを同定する工程は繰り返され、それにより核酸鎖が更に伸長され、標的核酸の配列が決定される。

【0130】

配列決定の前に、核酸分子はバーコーディングされ、又はそうでなければ標識され得る。バーコーディングは、配列リードのより容易なグルーピングを可能にし得る。例えば、バーコードは、同じ核酸分子又はDNAタンパク質複合体から生じる配列を識別するために使用され得る。バーコードはまた、個々の結合を一意に識別するために使用され得る。例えば、各結合は、結合を一意に識別することができる固有の（例えば、無作為に生成された）バーコードで印をつけられ得る。同じ核酸分子又はDNAタンパク質複合体から生じる配列を識別するための第1のバーコード、又は個々の結合を一意に識別するための第2のバーコードなど、複数のバーコードを共に使用することができる。

【0131】

バーコーディングは、多数の技術を介して達成できる。場合によっては、バーコードは、終点オリゴヌクレオチド内の配列として含まれ得る。他の場合、核酸分子は、少なくとも2つのセグメントを含むオリゴヌクレオチドに接触させることができ：1つのセグメントはバーコードを含み、別のセグメントは終点配列に相補的な配列を含む。終点配列へとアニールした後、バーコーディングされたオリゴヌクレオチドは、ポリメラーゼで拡張されて、同じ終点化された核酸分子からバーコーディングされた分子を産出することができる。終点化された核酸分子は、中でフェーズ情報が保存されている入力核酸分子の再編成されたバージョンであるため、生成されたバーコーディングされた分子はまた、同じ入力核酸分子から由来する。これらバーコーディングされた分子は、バーコード配列、終点相補的配列、及びゲノム配列を含む。

【0132】

終点を含む又は含まない核酸分子（例えば、DNAタンパク質複合体の一部又はそこから回復される核酸）について、分子は他の手段によりバーコーディングされ得る。例えば、核酸分子は、核酸分子からの配列を組み込むために拡張され得る、バーコーディングされたオリゴヌクレオチドに接触させることができる。バーコードは、終点配列に、制限酵

素認識部位に、対象の部位（例えばゲノム関心領域）に、又はランダム部位（例えばバーコードオリゴヌクレオチド上の無作為の n - 量体配列による）に、ハイブリダイズすることができる。核酸分子は、サンプル中の他の核酸分子から適切な濃縮及び／又は分離（例えば空間的又は時間的な分離）を使用してバーコードに接触させることができ、その結果、複数の核酸分子は与えられず、同じバーコード配列が与えられる。例えば、核酸分子を含む溶液は、1つの核酸分子又は1つのDNAタンパク質複合体のみが、バーコード又は与えられたバーコード配列を含むバーコードの群に接触させられるような濃度に希釈され得る。バーコードは、遊離溶液中で、流体区分中で（例えば液滴又はウェル）、或いはアレイ上で（例えば特定のアレイスポットにて）、核酸分子に接触され得る。

【0133】

バーコーディングされた核酸分子（例えば拡張生成物）は、例えば、ショートリード配列決定機械上で配列決定することができ、配列情報は、共通の位置合わせ、フェーズ、又は他の群への、同じバーコードを持つ配列リードのグルーピングによって判定される。このように、合成のロングリードは、ショートリード配列決定を介して達成することができる。代替的に、配列決定の前に、バーコーディングされた生成物は、例えばロングリード配列決定技術を使用して配列決定される長い分子を生成するために、例えばバルクライゲーション（bulk ligation）を介して共に結合され得る。これらの場合、包埋されたリードペアは、増幅アダプター及び終点配列を介して識別可能である。更なるフェーズ情報は、リードペアのバーコード配列から得られる。

【0134】

代替的に、場合によっては、本明細書に記載されるように生成されるライブラリ分子は、終点オリゴ挿入なしで連結される。これら分子は、それにもかかわらず、5 kb、10 kb、20 kb又はそれ以上の長さのリードを生成するのに商業上利用可能なロングリード化学を使用して配列決定を行うのに適切である。これらの場合、連結結合は、配列解析を介して容易に同定される。

【0135】

ロングリード（例えば合成又は実際のロングリード）は、ショートリードから決定するのが困難又は富可能な場合がある、フェージング情報などの情報を得るために使用することができる。フェージング情報は、母／父のフェージングの他、腫瘍／非腫瘍フェージング情報も含む。腫瘍／非腫瘍のフェージングは、体細胞のゲノム情報から癌のゲノム情報を区別するために使用され得る。

【0136】

一例において、上述のように、FFPEサンプルから作成されたライブラリなどのライブラリの断片は、配列決定された末端である。各末端がマッピングされるコンティグがサンプル中の共通の核酸分子上で物理的に結合されることを示す、リードペアが観察される。結果として生じるライブラリは、単離された配列の位置をゲノムアセンブリと比較することにより回復された断片のペアエンド間の距離を決定するために、配列決定によって更に解析される。FFPEサンプル中の長距離のリードペアの頻度は、非FFPEサンプルの長距離のリードペアの頻度と比較される。上記のライブラリなどの典型的なライブラリにおいて、配列決定は、FFPE - Chica go方法が、非FFPEサンプル上で行なわれたChica go方法に匹敵する（> 200 kbpの挿入）又はそれよりも大きな（100 kbp - 200 kbpの挿入）長距離のリードペアの頻度を結果としてもたらすことを、明らかにする、FFPE - Chica goライブラリの複雑性及び生の配列決定能力バレッジも決定される。ライブラリの複雑性は、ライブラリ内の異なる分子の多様性を指す。

【0137】

遺伝子情報

フェージング情報、染色体立体配座、配列アセンブリ、及び、構造変動（SV）、コピー数変異（CNV）、ヘテロ接合性の損失（LOH）、単一のヌクレオチド変異（SNV）、一塩基変異多型（SNP）、染色体転座、遺伝子融合、及び挿入と欠失（INDEL

10

20

30

40

50

）を含むがこれらに限定されない遺伝子特徴が、本明細書に開示される方法により産出された配列リードデータの解析によって決定することができる。遺伝子特徴の解析のための他の入力は、基準ゲノム（例えばアノテーションを伴う）、ゲノムマスキング情報、及び、候補遺伝子、遺伝子対、及び／又は対象の座標のリストを含み得る。設定パラメータ及びゲノムマスキング情報はカスタマイズすることができ、又は、デフォルトパラメータ及びゲノムマスキングが使用され得る。一例において、リードペアはゲノムにマッピングされ、次いで各ペアは、リードペアのリード1及びリード2それぞれの連結された基準染色体上のマッピング位置に等しいx及びyの座標を持つ平面において点として表わされる。x-y平面は、重複しない二乗ビン（square bin）に分けることができ、各ビンにマッピングされるリードペアの数を表にすることができる。ビンの数は、ピンをピクセルに対応させた画像（例えばヒートマップ）として視覚化することができる。画像処理技術などの様々な解析技術を使用して、異なる再編成などの遺伝子特徴のシグネチャを識別することができる。例えば、カーネルコンボリユーション・フィルタリングを使用して、融合されるゲノムの遺伝子座のペアに対応する画像中の点を見つけ出すことができる。図2A及び図2Bは、図3に示されるものなどの、相互転座を見つけ出すために使用され得る典型的な単純なカーネルを示す。図3は、ETV6とNTRK3との間の相互転座の信号を含む画像を示す。右上部分及び左下部分の「蝶ネクタイ」形状の特徴は、相互転座のゲノム特徴のこれらの2つの領域間の相互作用を示す。

10

【0138】

配列リードデータなどの入力は、適切なファイル形式でフォーマット化され得る。例えば、配列リードデータは、FASTAファイル、FASTQファイル、BAMファイル、SAMファイル、又は他のファイル形式で含むことができる。入力配列リードデータは整列されない場合がある。入力配列リードデータは整列される場合がある。

20

【0139】

配列リードデータを解析のために調製することができる。例えば、リードは品質のために整えることができる。必要ならば、リードは配列決定アダプターを除去するためにも整えることができる。

【0140】

配列リードデータは整列することができる。例えば、リードペアは、特定の基準ゲノムに整列され得る。場合によっては、基準ゲノムはCRCh38である。整列は、SNAP、Burrows-Wheelerアライナー（例えばbwa-sw、bwa-mem、bwa-aln）、Bowtie2、Novoalign、及びそれらの改良又は変形により、行うことができる。

30

【0141】

解析の品質管理（QC）レポートも生成することができる。QCレポートを使用して、より深い配列決定を実行する前に失敗したライブラリを同定することができる。そのような品質管理レポートは様々なメトリックを含み得る。QCメトリックは、限定されないが、合計のリードペア、重複の割合（例えばPCR重複）、マッピングされてないリードの割合、低いマップ品質のリードの割合（例えば、 $Q < 20$ ）、異なる染色体にマッピングされたリードペアの割合、0～1kbpの（マッピング位置間の距離等の）リードペア挿入の割合、1kbp～100kbpのリードペア挿入の割合、100kbp～1Mbpのリードペア挿入の割合、1Mbpより上のリードペア挿入の割合、ライゲーション結合を含むリードペアの割合、制限断片末端への近接、リードペア分離プロット、及びライブラリの複雑性の推定を含み得る。QCメトリックを使用して、解析を最適化し、且つ、試薬、サンプル、及びユーザーにおける品質問題を同定することができる。配列アラインメントは、QCメトリックのうち1つ以上に基づいてフィルタリングすることができる。重複リードも、例えば非常に対応する位置でのリードの比較に基づいてフィルタリングすることができる。

40

【0142】

配列リードの解析結果は、結合密度の結果を含み得る。結合密度の結果は、結合密度の

50

結果の全ゲノム、1つの遺伝子座、及び2つの遺伝子座の見解を含み得る。結合密度の結果をデータセットとして出力することができる。結合密度の結果は、染色体又はゲノムの領域間の相互作用（例えば接触）のヒートマップなどの、連鎖密度プロット（LDP）として提示することができる。連鎖密度の結果は、品質スコアなどのスコアに関連づけることができる。場合によっては、連鎖密度の可視化が、スコア閾値を越える結果のために出力される。一例において、可視化は、全ゲノムのために、スコア閾値を越えるデノボコール（*calls*）のために、スコア閾値を越える片側の候補コールのために、及び、負として分類されるものを含む全ての両側の候補のために、含まれる。結合密度の可視化は、スケール（例えばカラースケール）、長さスケールバー、遺伝子ネーム標識、遺伝子のためのエキソン/イントロン構造グリフ（*structure glyphs*）、及び検出された再編成の強調表示を含み得る。

10

【0143】

結合情報は、カバレッジ、断片のマッピング可能性、断片GC含量、及び断片長などの効果及び偏重（*biases*）を制御するために標準化することができる。標準化は、マトリックス平衡化、又は他の要因に依存しない方法により行うことができる。マトリックス平衡化は、Sinkhorn-Knoppのアルゴリズム又はナイト-ルイスの標準化などのアルゴリズムを利用することができる。標準化は、偽陽性につながりかねないバックグラウンド信号を補正するためにも行うことができる。例えば、図4A、図4B、及び図4Cは、3つの異なるサンプルにおいて比較された同じ対の染色体での画像分析に基づく結果を示す。様々な「ヒット」（図中、円で囲まれている）が、複数のサンプルにわたる同じ位置で見出され、これらが偽陽性であるという疑いを高めている。サンプル（例えば10のサンプル）のプールにわたる中間の標準化されたリード密度などによる正規化を使用して、例えば中間のピクセルでサンプルピクセルを割ることにより個々のサンプルデータを補正することができる。図5A、図5B、図5Cは、染色体1対染色体7（図5A）、染色体2対染色体5（図5B）、及び染色体1対染色体1（図5C）に関する中間の標準化された読み取り密度（10以上のサンプル）を示す。標準化は、図6Aに示されるように等しいピンのサイズを含む様々なピンを取り扱う手法で、及び、図6Bに示されるようにピンの補間で行うことができる。場合によっては、ピンの補間は、等しいピンのサイズと比較して減少したバックグラウンドノイズをもたらし、結果としてより急激に分解された特徴をもたらすことができる。

20

30

【0144】

整列された配列データは、全ゲノムによる再編成及び特定の2つの遺伝子座（又は両側の）の候補遺伝子での再編成を含む再編成のために解析され得る。解析は、接触、融合、及び結合の同定も含み得る。（例えばBAMファイル又は他の適切なフォーマットでの）配列リードデータの整列は、解析に入力され得る。ゲノムマスキング情報は同様に入力することができる、或いは、デフォルトゲノムマスキング情報を解析に使用することができる。解析はゲノム全体にわたって行うことができる。付加的に又は代替的に、解析は、両側の候補融合のリストのために行うことができる。場合によっては、候補融合のリストのために行われた解析は、全ゲノムのために行われた解析よりも敏感である。両側の候補融合の解析は、比較的ゲノムスキャンによって見落とされ得るDNAの比較的短いセグメントの転座を含む融合を検出することができる。

40

【0145】

接触及び再編成などの特徴（欠失、重複、挿入、逆位又は逆転、転座、結合、融合、及び分裂を含むがこれらに限定されない）及び他の相互作用を識別する解析は、様々な技術を用いて行なうことができる。解析技術は、統計解析及び確率解析、フーリエ解析、コンピュータビジョン、及び他の画像処理を含む信号処理、言語処理（例えば自然言語処理）、並びに機械学習を含み得る。例えば、接触マトリクスなどの相互作用プロットが、特徴を示す特徴のために解析され得る。場合によっては、フィルタは、プロット又は他のデータに適用することができる。フィルタは、平滑化フィルタ（例えば、カーネル平滑化又はSavitzky-Golayのフィルタ、ガウシアンぼかし）を含むがこれらに限定

50

されない、コンボリレーションフィルタであり得る。

【0146】

幾つかの実施形態は、ゲノム構造決定の構成要素として機械学習を含み、従って幾つかのコンピュータシステムは、機械学習能力を持つモジュールを含むように構成される。機械学習モジュールは、機械学習機能性を構成するように以下に列挙された様式のうち少なくとも1つを含む。

【0147】

機械学習を構成する様式は様々に、自動化された質量分析データスポットの検出及び要求を実行することができるような、データフィルタリング能力を実証する。この様式は、場合によっては、逆位、挿入、欠失、又は転座などの様々なゲノムの構造変化を示す、予測されたパターンの存在により容易にされる。

10

【0148】

機械学習を構成する様式は様々に、下リードペアの頻度を下流分析に役立つ形態にするように、データ処置又はデータ処理の能力を実証する。データ処置の例は、下流解析に導かれる形態でデータを与えるような、対数変換、換算比の割当、又は作成された特徴へのデータのマッピングを含むが、必ずしもこれらに限定されない。

【0149】

本明細書に開示されるような機械学習データ解析の構成要素は、1～10,000の特徴、又は2～300,000の特徴、或いはこれら範囲のうち何れかの中の特徴の数、又はこれら範囲のうち何れかよりも高い特徴の数などの、リードペアのデータセットにおける広範囲の特徴を規則的に処理する。場合によっては、データ解析は、少なくとも1k、2k、3k、4k、5k、6k、7k、8k、9k、10k、20k、30k、40k、50k、60k、70k、80k、90k、100k、120k、140k、160k、180k、200k、220k、2240k、260k、280k、300k、又は300k以上の特徴を含む。

20

【0150】

リードペア分布パターンは、本明細書の開示と一致するあらゆる数の手法を使用して同定される。場合によっては、リードペア分布パターンの選択は、弾性ネット (elastic net)、情報利得、ランダムフォレストのインビューティング、或いは、本明細書の開示と一致し且つ当業者に精通する他の特徴選択手法を含む。

30

【0151】

選択されたリードペア分布パターンは、本明細書の開示と一致するあらゆる数の手法を再び使用して、ゲノムの構造変化を示す予測されたパターンと対抗させられる。場合によっては、リードペアパターンの検出は、ロジスティック回帰、SVM、ランダムフォレスト、KNN、或いは、本明細書の開示と一致し且つ当業者に精通する他のクラシファイヤー手法を含む。

【0152】

機械学習を適用すること、或いは、本明細書に開示される解析のために構成されたコンピュータに機械学習モジュールを提供することで、進行中のモニタリング手順の一部として無症候性の疾患検出又は早期検出のための、関連するゲノムの構造変化の検出が可能となり、それにより、症状の進行の前に、或いは、介入がより容易に達成される又はおそらく成功の結果を引き起こす可能性が高い間に、疾患又は障害を識別することが可能となる。

40

【0153】

機械学習を適用すること、或いは、本明細書に開示される解析のために構成されたコンピュータに機械学習モジュールを提供することで、例えば治験の一部として薬物処置に晒される個体の構造再編成の同定も可能となり、その結果、個体又は集団に関する試験の結果は、薬物の効果と正又は負に一致する特定のゲノムの構造的事象を同定するように、同時に又は遡及的に関連づけられ得る。

【0154】

機械学習を適用すること、或いは、本明細書に開示される解析のために構成されたコン

50

コンピュータに機械学習モジュールを提供することで、サンプル中の位置情報を保存するように均質化なしで集められた腫瘍組織サンプルなどの遺伝的に不均一なサンプルの特定の領域に一致する、構造再編成の同定を可能にする。幾つかの腫瘍領域は、転移又は腫瘍の伝播に特に熟達した (a d e p t) 細胞集団に対応すると知られるため、そのような細胞集団と関連するゲノム再編成又は他のフェーズ情報の同定は、これらの特に危険な細胞集団を標的とするための処置レジメンを選択することを補助する。

【 0 1 5 5 】

モニタリングは、発症又は進行のサインがモニタリングされる勝敗に関する遺伝性素因を示す遺伝的評価と組み合わせ、又はそれを支持して頻繁に実行されるが、必ずしもそうであるというわけではない。同様に、場合によっては、機械学習は、処置レジメンの処置効果のモニタリング又はその評価を容易にするために使用され、その結果、処置レジメンは、進行中のプロテオミクスで媒介されるモニタリングにより示されるように、経時的に修正し、継続し、又は分解することができる。

10

【 0 1 5 6 】

機械学習手法、及び、機械学習アルゴリズムを実行するように構成されたモジュールを持つコンピュータシステムは、様々な複雑性のデータセットにおけるフェーズ情報又はゲノム再編成の同定を容易にする。場合によっては、フェーズ情報又はゲノム再編成は、多数の時点で1人の個体から、対象の疾病に関する既知の状況或いは既知の最終的な処置結果又は反応の複数の個体などの複数の個体から得たサンプルから、又は複数の時点且つ複数の個体から得られたデータなどの、大量の質量分析データを含む標的とされていないデータベースから同定される。

20

【 0 1 5 7 】

代替的に、場合によっては、機械学習は、例えば、個体の健康状態がその時点で分かっているときに複数の時点にわたり1人の個体からゲノム再編成又はフェーズ情報を集めることにより、対象の疾病に関する既知の状況の複数の個体から配列情報を集めることにより、又は複数の時点で複数の個体から配列情報を集めることにより、ゲノムの再編成又はフェーズ情報に標的とされたデータベースの解析を介して、ゲノムの再編成又はフェーズ情報の強化を容易にする。容易に明らかとなるように、場合によっては、配列情報の収集は、手術に準じて集められた架橋サンプル、又は治験に準じて集められた F F P E サンプルなどの保存されたサンプルの使用により容易にされる。

30

【 0 1 5 8 】

故に、配列情報は、単独で、又は治験結果或いは外科的介入結果の情報と組み合わせで収集される。配列データは、単独で又は1つ以上の追加のマーカと組み合わせで健康状態信号を構成するゲノムの再編成に対応するパターンを示すリードペアの亜群を同定するように、例えば本明細書に開示されるように構成されるコンピュータシステム上で、機械学習に晒される。故に、機械学習は場合によっては、個体の健康状態に個々に有益な、配列、DNA又はRNA何れかの配列、或いはゲノム再構成の同定を容易にする。

【 0 1 5 9 】

検知可能な再編成の限界点間の最小距離は、2 b p、3 b p、4 b p、5 b p、6 b p、7 b p、8 b p、9 b p、10 b p、20 b p、30 b p、40 b p、50 b p、60 b p、70 b p、80 b p、90 b p、100 b p、200 b p、300 b p、400 b p、500 b p、600 b p、700 b p、800 b p、900 b p、1 k b、2 k b、3 k b、4 k b、5 k b、6 k b、7 k b、8 k b、9 k b、10 k b、20 k b、30 k b、40 k b、50 k b、60 k b、70 k b、80 k b、90 k b、100 k b、200 k b、300 k b、400 k b、500 k b、600 k b、700 k b、800 k b、900 k b、1 M b、2 M b、3 M b、4 M b、5 M b、6 M b、7 M b、8 M b、9 M b、10 M b、20 M b、30 M b、40 M b、50 M b、60 M b、70 M b、80 M b、90 M b、100 M b、200 M b、300 M b、400 M b、500 M b、600 M b、700 M b、800 M b、900 M b、又は1 G bを含む核酸長のリストから選択された2つの数により定義された範囲の中の数、そのおよその数、又はその数未満であ

40

50

り得る。

【0160】

再編成解析は、被験体ゲノム中で結合されると認められる限界点のペアのリストを生成することができる。限界点座標のペアのリストは、限界点座標ペアに関する統計的有意差又はコンフィデンスメトリック（例えばp値）も含み得る。限界点のこれらのペアは、ブラウザ拡張可能なデータ（BED）又はBED - PEなど適切なフォーマットで出力され得る。

【0161】

染色体立体配座の解析も、本明細書に開示された技術を使用して行うことができる。例えば、ドメイン（TAD）及びTAD境界をフェーズ的に関連づけることが、決定され得る。他のフェーズ幾何学ドメイン及び境界も判定することができ、ラミナ関連ドメイン（LAD）、複製時間領域、及び大きな組織化されたクロマチンK9修飾（LOCK）ドメインを含むが、これらに限定されない。

10

【0162】

図7は、全ゲノムスキニング分析のパイプラインによる分析を示す。分析的パイプラインにより行われたサンプルコール（Sample calls）を、白色の円で囲んで示す。図7は、250kのピンでの、染色体3対染色体6のプロットを示す。

【0163】

典型的な実施形態において、配列決定データを使用して、出発FFPEサンプルにあると知られる多形に関するフェージング情報を判定する。例えば、配列決定データを使用して、SNPなどの特定の多形が同じ又は異なるDNA分子に存在したかどうかを判定する。この方法を使用して判定したフェージングの制度は、GIABサンプルの配列などの既知の配列と比較することにより測定される。例えば、場合によっては、0 - 10,000の間で、132,796のSNPsが見出され、正確なフェーズには99.059%があったことが分かる。高い一致（>95%）は、約1.5MBまで見られる（13のうち1つが一致しない70-80kbのピン、及び15のうち2つが一致しない1.1-1.3MBのピンを除く）。1.7-1.9MBの範囲において、7つのSNPペアフェーズのうちの7つが適切に要求された。これらのデータから、低レベルの偽性の連鎖にもかかわらず、適切な長距離の情報は、最大でメガベースまでFFPE-Chicago方法を使用して決定されると、結論付けられる。重要なことに、これらの「一致」予測割合は、偶然から予測する50%の成功率よりも著しく高く、95%以上である。

20

30

【0164】

構造のフェージング情報

現在、構造及びフェージングの解析（例えば医療目的のための）は、困難なままである。例えば、癌、同じ型の癌を持つ個体、或いは同じ腫瘍の中でさえ、驚くべき異種性が存在する。結果として生ずる効果から原因となるものを引き出すことは、低いサンプルごとのコストで非常に高い精度及びスループットを必要とし得る。オーダーメイド医療のドメインにおいて、ゲノムケアの金本位のうち1つは、大きい且つ小さな構造再編成及び新しい突然変異を含む、徹底的に特徴付けられ且つフェージングされた全ての変異体を持つ、配列決定されたゲノムである。以前の技術でこれを達成するためには、現在ではあまりに高価であり且つ慣例的な医療処置を要求する、デノボアセンブリに必要なものと同様の労力を必要とする。

40

【0165】

フェージング情報は、母/父のフェージングの他、腫瘍/非腫瘍フェージング情報も含む。腫瘍/非腫瘍のフェージングは、体細胞のゲノム情報から癌のゲノム情報を区別するために使用され得る。

【0166】

本開示の幾つかの実施形態において、被験体の保存された組織（例えばFFPE組織）を提供することができ、方法は、アセンブルされたゲノム、要求された変異体を伴う整理（大きな構造変異種及びコピー数変異体を含む）、フェージングされたバリエントコール

50

(variant calls)、又は任意の付加的な解析に戻ることができる。他の実施形態において、本明細書に開示される方法は、個体に直接、長距離リードペアライブラリを提供することができる。

【0167】

本開示の様々な実施形態において、本明細書に開示される方法は、長い距離分離された長距離のリードペアを生成することができる。この距離の上限は、大きなサイズのDNAサンプルを集める能力により改善され得る。場合によっては、リードペアは、最大50、60、70、80、90、100、125、150、175、200、225、250、300、400、500、600、700、800、900、1000、1500、2000年、2500、3000、4000、5000 kbp、又はそれ以上のゲノム距離におよぶ。幾つかの例において、リードペアは、最大500 kbpのゲノム距離におよぶことができる。他の例において、リードペアは、最大2000 kbpのゲノム距離におよぶことができる。本明細書に開示される方法は、分子生物学の標準技術に基づいて一体化し且つ構築することができ、更に効率、特異性、及びゲノムカバレッジの増大に十分に適している。

10

【0168】

他の実施形態において、本明細書に開示される方法は、現在利用される配列決定技術と共に使用することができる。例えば、この方法は、十分に試験され及び/又は広く配備された配列決定機器と組み合わせて使用することができる。更なる実施形態において、本明細書に開示される方法は、現在利用される配列決定技術に由来する技術と手法と共に使用することができる。

20

【0169】

様々な実施形態において、本開示は、本明細書に開示される1つ以上の方法を提供し、該方法は、保存された(例えばFFPE)サンプル又は細胞内の染色体の物理的なレイアウトを調査する工程を含む。配列決定を通じて染色体の物理的な設計を調査する技術の例は、染色体立体配座キャプチャ(「3C」)、環状染色体立体配座キャプチャ(「4C」)、カーボン・コピー染色体キャプチャ(「5C」)、及びHi-Cベースの方法などの技術の「C」ファミリー;及び、ChIP-ループ、ChIP-PETなどのChIPベースの方法を含む。これら技術は、核の中の空間的関係を固めるために生細胞のクロマチンの固定を利用する。生成物の後の処理及び配列決定により、研究者は、ゲノムの領域中の近似の会合のマトリクスを回復することが可能になる。更なる解析により、これら会合を使用して、保存された(例えばFFPE)サンプル中で物理的に染色体が配置されると、染色体の三次元の幾何学的なマップを作ることができる。そのような技術は、染色体の離散的な空間的機構を説明し、且つ、染色体の遺伝子座の中の機能的相補作用の正確な見解を提供する。

30

【0170】

幾つかの実施形態において、染色体内相互作用が、染色体の結合と関連づけられる。場合によっては、染色体内データがゲノムアセンブリを補助することができる。場合によっては、クロマチンはインビトロで再構成される。このことは、クロマチン、具体的にはヒストン、クロマチンの主要タンパク質構成要素が、配列決定を通じてクロマチンの立体配座及び構造を検出するための技術の最も一般的な「C」ファミリー:3C、4C、5C、及びHi-Cの下での固定に重要であるので、都合が良い場合がある。クロマチンは、配列に関して非常に非特異的であり、通常はゲノムにわたって均一にアセンブルする。場合によっては、クロマチンを使用しない種のゲノムは、再構成されたクロマチン上でアセンブルすることができ、それにより、生命の全てのドメインへの開示のための範囲(horizon)を拡大することができる。

40

【0171】

リードペアデータは、クロマチン立体配座キャプチャ技術から得ることができる。幾つかの例において、ライゲーション又は他のタグ付けは、近くの物理的近接状態にあるゲノム領域に印を付けるように遂行される。タンパク質(ヒストンなど)がクロマチン内で、

50

DNA分子、例えばゲノムDNAとの複合体において安定して結合されるような、複合体の架橋は、本明細書の他の場所で更に詳しく記載され又は当該技術分野で既知の適切な方法に従って、遂行され得る。場合によっては、サンプル保存（例えば固定）から生じた架橋は、プロテイナーゼK処理の排除などによりDNA-タンパク質複合体が分解されないような条件下で、このような複合体を抽出することにより利用される。例えば、ゲノム配列に沿って近接していないヌクレオチドセグメントは、クロマチンなどの構造の一部である場合に物理的に近接し得る。そのようなヌクレオチドセグメントは共にライゲートされ、その後、本開示の方法に従って解析することができる。例えば、ライゲートされたヌクレオチドセグメントは配列決定され、2つのライゲートされたセグメントの配列決定された末端間の距離（挿入距離）を分析することができる。図8Aは、本開示の技術により解析される保存されたサンプル（例えばFFPEサンプル）のための塩基対（bp）における挿入距離の関数としての特定の範囲での挿入の可能性のグラフを示す。図8Bは、Chicago法を使用して解析されたサンプルの同様のグラフを示す。両方のグラフにおいて、X軸は0から300,000の挿入距離（bp）を示し、一方でY軸は、軸の頂部にて 10^0 から軸の底部にて 10^{-8} までの、その距離の挿入の可能性（対数）を示している。

10

【0172】

場合によっては、2つ以上のヌクレオチド配列が、1つ以上のヌクレオチド配列に結合されたタンパク質を介して架橋され得る。1つの手法は、紫外線照射にクロマチンを晒すことである（Gilmour et al., Proc. Nat'l. Acad. Sci. USA 81:4275-4279, 1984）。ポリヌクレオチドセグメントの架橋も、化学的又は物理的（例えば光学的）な架橋などの他の手法を利用して行われてもよい。適切な化学架橋剤は、限定されないが、ホルムアルデヒド及びソラレンを含む（Solomon et al., Proc. Natl. Acad. Sci. USA 82:6470-6474, 1985; Solomon et al., Cell 53:937-947, 1988）。例えば、架橋は、DNA分子及びクロマチンタンパク質を含む混合物に2%のホルムアルデヒドを加えることにより行われ得る。DNAを架橋するために使用され得る薬剤の他の例は、UV光、マイトマイシンC、ナイトロジェンマスタード、メルファラン、1,3-ブタジエンジエポキシド、シス-ジアミンジクロロ白金（II）、及びシクロホスファミドを含むがこれらに限定されない。適切なものとして、架橋剤は、約2 nmなどの比較的短い距離に架橋する架橋を形成し、それにより逆転され得る密接な相互作用を選択する。

20

30

【0173】

普遍的に、Hi-Cベースの技術などの染色体の物理的なレイアウトを調べるための手順は、培養細胞又は初代組織から単離されたクロマチンなどの、細胞/生物体内に形成されるクロマチンを利用する。Chicagoベースの方法は、細胞/生物体から単離されたクロマチンだけでなく、再構成クロマチンを用いたそのような技術の使用を提供する。再構成クロマチンは、様々な特徴にわたり細胞/生物体内に形成されたクロマチンから分化される。まず、多くのサンプルについて、ネイキッドDNAサンプルの収集は、体液を集めること、頬又は直腸の領域を拭き取ること、上皮サンプルを得ることなど、様々な非侵襲的な方法から侵襲的な方法までを使用することによって、達成することができる。第2に、クロマチンの再構成は、ゲノムアセンブリ及びハプロタイプフェージングのための人工物を生成する、染色体間の及び他の長距離の相互作用の形成を実質的に妨げる。場合によっては、サンプルには、約20、15、12、11、10、9、8、7、6、5、4、3、2、1、0.5、0.4、0.3、0.2、0.1%未満、又はそれ以下の、本開示の方法及び組成物に従った染色体間又は分子間の架橋があってもよい。幾つかの例において、サンプルには、約5%未満の染色体間又は分子間の架橋があってもよい。幾つかの例において、サンプルには、約3%未満の染色体間又は分子間の架橋があってもよい。更なる例において、サンプルには、約1%未満の染色体間又は分子間の架橋があってもよい。第3に、架橋が可能な部位の頻度、従ってポリヌクレオチド内の分子内架橋の頻度は、

40

50

調整することができる。例えば、DNA対ヒストンの比率は変動する場合があり、それにより、ヌクレオソーム密度を望ましい値に調整することができる。場合によっては、ヌクレオソーム密度は、生理的レベルより下に減らされる。従って、架橋の分布は、より長い範囲の相互作用を支持するように変更することができる。幾つかの実施形態において、様々な架橋を持つサブサンプルが、短距離及び長距離両方の会合を覆うように調製され得る。例えば、架橋条件は、架橋の少なくとも約1%、約2%、約3%、約4%、約5%、約6%、約7%、約8%、約9%、約10%、約11%、約12%、約13%、約14%、約15%、約16%、約17%、約18%、約19%、約20%、約25%、約30%、約40%、約45%、約50%、約60%、約70%、約80%、約90%、約95%、又は約100%が、サンプルDNA分子上で少なくとも約50kb、約60kb、約70kb、約80kb、約90kb、約100kb、約110kb、約120kb、約130kb、約140kb、約150kb、約160kb、約180kb、約200kb、約250kb、約300kb、約350kb、約400kb、約450kb、又は約500kb離れているDNAセグメント間に生じるように、調整され得る。

【0174】

癌ゲノム配列決定により必要とされる高い精度は、本明細書に記載される方法及びシステムを使用して達成することができる。不正確な基準ゲノムは、癌ゲノムを配列決定する際に塩基呼び出しの困難をもたらしかねない。不均一なサンプル及び小さな出発物質、例えば生検により得られたサンプルが、更なる困難をもたらす。更に、ヘテロ接合性の大規模な構造の変異及び/又は損失の検出は大抵、癌ゲノム配列決定の他、体細胞の変異と塩基要求における誤りとを区別する能力にも、重要である。

【0175】

本明細書に記載されるシステム及び方法は、2、3、4、5、6、7、8、9、10、12、15、20、又はそれ以上の様々なゲノムを含む複合サンプルから正確な長い配列を生成し得る。通常の、良性の、及び/又は腫瘍由来の混合サンプルは、随意に通常の対照を必要とすることなく、解析されてもよい。幾つかの実施形態において、100ngもの小さな、或いは数百程のゲノム当量の小さな出発サンプルは、正確な長い配列を生成するために利用される。本明細書に記載されるシステム及び方法は、コピー数変異、大規模な構造変異及び再編成の検出が可能となる場合があり、フェージングされたバリエーションが、約1kbp、約2kbp、約5kbp、約10kbp、20kbp、約50kbp、約100kbp、約200kbp、約500kbp、約1Mbp、約2Mbp、約5Mbp、約10Mbp、約20Mbp、約50Mbp、約100Mbp、又はそれ以上のヌクレオチドに跨る長い配列にわたって得られる場合がある。例えば、フェーズのバリエーションは、約1Mbp又は約2Mbpに跨る長い配列にわたって得られる場合がある。

【0176】

サンプルは、様々な容量及び表面積の組織切片を含み得る。場合によっては、サンプルは、約5 μ m~10 μ mの厚みの組織切片を含む。場合によっては、サンプルは、約1 μ m、2 μ m、3 μ m、4 μ m、5 μ m、6 μ m、7 μ m、8 μ m、9 μ m、10 μ m、11 μ m、12 μ m、13 μ m、14 μ m、15 μ m、16 μ m、17 μ m、18 μ m、19 μ m、20 μ m、25 μ m、30 μ m、35 μ m、40 μ m、45 μ m、50 μ m、55 μ m、60 μ m、65 μ m、70 μ m、75 μ m、80 μ m、85 μ m、90 μ m、95 μ m、100 μ m、150 μ m、200 μ m、250 μ m、300 μ m、350 μ m、400 μ m、450 μ m、500 μ m、550 μ m、600 μ m、650 μ m、700 μ m、750 μ m、800 μ m、850 μ m、900 μ m、950 μ m、1000 μ m、又はそれ以上の厚みの組織切片を含む。場合によっては、サンプルは、少なくとも約1 μ m、2 μ m、3 μ m、4 μ m、5 μ m、6 μ m、7 μ m、8 μ m、9 μ m、10 μ m、11 μ m、12 μ m、13 μ m、14 μ m、15 μ m、16 μ m、17 μ m、18 μ m、19 μ m、20 μ m、25 μ m、30 μ m、35 μ m、40 μ m、45 μ m、50 μ m、55 μ m、60 μ m、65 μ m、70 μ m、75 μ m、80 μ m、85 μ m、90 μ m、95 μ m、100 μ m、150 μ m、200 μ m、250 μ m、300 μ m、350 μ m、4

0 0 μm 、4 5 0 μm 、5 0 0 μm 、5 5 0 μm 、6 0 0 μm 、6 5 0 μm 、7 0 0 μm 、7 5 0 μm 、8 0 0 μm 、8 5 0 μm 、9 0 0 μm 、9 5 0 μm 、1 0 0 0 μm 、又はそれ以上の厚みの組織切片を含む。場合によっては、サンプルは、多くとも約 1 μm 、2 μm 、3 μm 、4 μm 、5 μm 、6 μm 、7 μm 、8 μm 、9 μm 、1 0 μm 、1 1 μm 、1 2 μm 、1 3 μm 、1 4 μm 、1 5 μm 、1 6 μm 、1 7 μm 、1 8 μm 、1 9 μm 、2 0 μm 、2 5 μm 、3 0 μm 、3 5 μm 、4 0 μm 、4 5 μm 、5 0 μm 、5 5 μm 、6 0 μm 、6 5 μm 、7 0 μm 、7 5 μm 、8 0 μm 、8 5 μm 、9 0 μm 、9 5 μm 、1 0 0 μm 、1 5 0 μm 、2 0 0 μm 、2 5 0 μm 、3 0 0 μm 、3 5 0 μm 、4 0 0 μm 、4 5 0 μm 、5 0 0 μm 、5 5 0 μm 、6 0 0 μm 、6 5 0 μm 、7 0 0 μm 、7 5 0 μm 、8 0 0 μm 、8 5 0 μm 、9 0 0 μm 、9 5 0 μm 、1 0 0 0 μm 、又はそれ以上の厚みの組織切片を含む。場合によっては、サンプルは、約 1 0 0 ~ 3 0 0 mm^2 の表面積を持つ組織切片を含む。場合によっては、サンプルは、約 1 0 mm^2 、2 0 mm^2 、3 0 mm^2 、4 0 mm^2 、5 0 mm^2 、6 0 mm^2 、7 0 mm^2 、8 0 mm^2 、9 0 mm^2 、1 0 0 mm^2 、2 0 0 mm^2 、3 0 0 mm^2 、4 0 0 mm^2 、5 0 0 mm^2 、6 0 0 mm^2 、7 0 0 mm^2 、8 0 0 mm^2 、9 0 0 mm^2 、1 0 0 0 mm^2 、又はそれ以上の表面積を持つ組織切片を含む。場合によっては、サンプルは、少なくとも約 1 0 mm^2 、2 0 mm^2 、3 0 mm^2 、4 0 mm^2 、5 0 mm^2 、6 0 mm^2 、7 0 mm^2 、8 0 mm^2 、9 0 mm^2 、1 0 0 mm^2 、2 0 0 mm^2 、3 0 0 mm^2 、4 0 0 mm^2 、5 0 0 mm^2 、6 0 0 mm^2 、7 0 0 mm^2 、8 0 0 mm^2 、9 0 0 mm^2 、1 0 0 0 mm^2 、又はそれ以上の表面積を持つ組織切片を含む。場合によっては、サンプルは、多くとも約 1 0 mm^2 、2 0 mm^2 、3 0 mm^2 、4 0 mm^2 、5 0 mm^2 、6 0 mm^2 、7 0 mm^2 、8 0 mm^2 、9 0 mm^2 、1 0 0 mm^2 、2 0 0 mm^2 、3 0 0 mm^2 、4 0 0 mm^2 、5 0 0 mm^2 、6 0 0 mm^2 、7 0 0 mm^2 、8 0 0 mm^2 、9 0 0 mm^2 、1 0 0 0 mm^2 、又はそれ以上の表面積を持つ組織切片を含む。

【0177】

本明細書に記載される方法及びシステムを使用して決定されるハプロタイプは、コンピュータリソース、例えば、クラウドシステムなどのネットワーク上のコンピュータリソースに割り当てられてもよい。短いバリエーションコールは、必要であれば、コンピュータリソースに保管される関連情報を使用して修正することができる。構造変異は、短いバリエーションコールからの組み合わせられた情報、及びコンピュータリソースに保管された情報に基づいて検出することができる。分節重複、構造変動の傾向がある領域、高度に可変的であり且つ医学的に関連する MHC 領域、セントロメア及びテロメアの領域、及び、反復領域を持つものを含むがこれに限定されない他の異質染色質領域、低い配列精度、高い変異の割合、ALU 反復、分節重複、或いは当該技術分野で既知の他の関連する問題となる部分などの、ゲノムの問題となる部分が、精度の向上ために再アセンブルされ得る。

【0178】

サンプル型は、局所的に、或いはクラウドなどのネットワーク接続されたコンピュータリソースにおいて、配列情報に割り当てることができる。情報のソースが分かっている場合、例えば、情報のソースが癌又は正常組織由来である時、このソースはサンプル型の一部としてサンプルに割り当てることができる。他のサンプル型の例は通常、限定されないが、組織型、サンプル収集方法、感染の存在、感染の型、処理方法、サンプルのサイズなどを含む。癌ゲノムとの比較における通常のゲノムのなどの、完全又は部分的な比較ゲノム配列が利用可能な場合、サンプルデータと比較ゲノム配列との差が判定され、随意に出力され得る。

【0179】

ハプロタイプフェージングの方法

本明細書に開示される方法により生成されたリードペアは通常、染色体間接触から由来するため、ヘテロ接合性の部位を含むあらゆるリードペアは、それらのフェージングに関する情報も運ぶ。この情報を使用して、短い、中間の、及び更には長い（メガベース）距離にわたる確実なフェージングが、迅速且つ正確に行うことができる。1 0 0 0 のゲノム

のトリオ (trios) (母 / 父 / 子のゲノムのセット) のうち 1 つからのデータをフェーシングするよう設計された実験は、確実にフェーシングを推測した。加えて、Selvaraj et al. (Nature Biotechnology 31 : 1111 - 1118 (2013)) と同様の近接ライゲーションを用いたハプロタイプ再構成も、本明細書に開示されるハプロタイプフェーシング方法と共に使用することができる。

【 0180 】

例えば、近接ライゲーションベースの方法を使用するハプロタイプ再構成も、ゲノムのフェーシングにおいて本明細書に開示される方法に使用することができる。近接ライゲーションベースの方法を使用するハプロタイプ再構成は、近接ライゲーション及び DNA 配列決定を、ハプロタイプアセンブリのための確率的アルゴリズムと組み合わせる。第 1 に、近接ライゲーション配列決定は、Hi-C プロトコルなどの染色体キャプチャプロトコルを使用して行なわれる。これらの方法は、三次元空間の中で共にループした、2 つの離れたゲノムの遺伝子座から DNA 断片をキャプチャすることができる。結果として生じる DNA ライブラリのショットガン DNA 配列決定の後、ペアエンド配列決定リードには、数百から数千万もの塩基対に及ぶ「挿入サイズ」がある。故に、Hi-C 実験で生成された短い DNA 断片は、小さなハプロタイプブロックをもたらすことができ、長い断片は最終的に、共にこれらの小さなブロックを結合することができる。十分な配列決定カバレッジで、この手法には、不連続のブロックにおいて変異体を結合し、且つ単一のハプロタイプに全てのそのようなブロックをアセンブルする可能性がある。その後、このデータは、ハプロタイプアセンブリのための確率的アルゴリズムと組み合わせられる。確率的アルゴリズムは、ノードがヘテロ接合の変異に相当し、縁が変異に結合し得る重複する配列断片に相当するグラフを利用する。このグラフは、配列決定の誤り又はトランス相互作用から結果として生じる、偽性の縁を含み得る。その後、最大限に切断されたアルゴリズムを使用して、入力配列決定リードのセットによって提供されるハプロタイプ情報と最大限に一致している、簡潔な解決法を予測する。近接ライゲーションが従来のゲノム配列決定又はメイトペア配列決定よりも大きなグラフを生成するので、コンピューティング時間及び反復の回数は、妥当な速度及び高精度でハプロタイプを予測することができるように修正される。その後、結果として生じるデータを使用して、Beagle ソフトウェア及びゲノム計画の配列決定データを用いて局所的なフェーシングをガイドして、高い分解能及び精度を持つ染色体スパンニングハプロタイプ (chromosome - spanning haplotype) を生成することができる。

【 0181 】

ペアエンドでのフェーズ情報の判定

本明細書には更に、FFPE - サンプル由来のペアエンドからフェーズ情報を判定するための方法及び組成物が提供される。ペアエンドは、開示された方法の何れか又は提供された実施例に更に例示されるものによって生成され得る。例えば、後に切断される固体表面に結合される DNA 分子の場合、遊離末端の再ライゲーションの後、再びライゲートされた DNA セグメントは、例えば制限消化によって固相が結合した DNA 分子から放たれる。この放出の結果、複数のペアエンド断片がもたらされる。場合によっては、ペアエンドは、増幅アダプターにライゲートされ、増幅され、且つ短い読み取り技術で配列決定される。これらの場合、多数の異なる固相に結合した DNA 分子からのペアエンドは、配列決定されたサンプル内にある。しかし、ペアエンドの結合の何れかの側について、結合隣接配列は共通の分子の共通のフェーズから由来することが、明確に結論付けられる。ペアエンドが終点オリゴヌクレオチドに結合される場合、配列決定リードにおけるペアエンドの結合は、終点オリゴヌクレオチド配列によって識別される。他の場合、ペアエンドは修飾されたヌクレオチドによって結合され、これは、使用される修飾されたヌクレオチドの配列に基づいて識別され得る。

【 0182 】

代替的に、ペアエンドの放出後、遊離ペアエンドは、増幅アダプターにライゲートされ、増幅され得る。これらの場合、複数のペアエンドはその後、ロングリード配列決定技術

10

20

30

40

50

を使用して読み取られる長い分子を生成するために、共にバルクライゲートされる。他の例において、放出されたペアエンドは、増幅工程を導入させることなく互いにバルクライゲートされる。何れの場合も、包埋されたリードペアは、終点配列又は修飾されたヌクレオチドなどの、結合配列に隣接している天然DNA配列を介して識別可能である。連結されたペアエンドは、長い配列デバイス上で読み取られ、多数の結合のための配列情報が得られる。ペアエンドは複数の異なる固相に結合したDNA分子から由来するので、増幅アダプター配列に隣接するものなどの2つの個々のペアエンドに及ぶ配列は、多数の異なるDNA分子にマッピングされると見出される。しかし、ペアエンドの結合の何れかの側について、結合隣接配列は共通の分子の共通のフェーズから由来することが、明確に結論付けられる。例えば、終点化された分子から由来するペアエンドの場合、終点配列に隣接する配列は、共通のDNA分子へと明確に割り当てられる。好ましい場合、本明細書に開示される方法及び組成物を使用して個々のペアエンドが連結されるので、単一リードにおいて複数のペアエンドを配列決定することができる。

10

【0183】

本明細書に記載される方法及び組成物を使用して生成された配列決定データは、好ましい実施形態において、フェージングされたデノボ配列アセンブリを生成し、フェーズ情報を判定し、及び/又は構造変異を同定するために使用される。

【0184】

構造変異及び他の遺伝子特徴の判定

図9Aと図9Bを参照すると、例は、再アセンブルされたクロマチンからのDNAの近接ライゲーションから生成されたリードペアの参照配列、例えばGRCh38上のマッピングされた位置が、GM12878と参照との構造差の付近にプロットされる例が、提供されている。生成されたリードペアはそれぞれ、対角線の上と下両方に表される。対角線上で、陰影は、示されている規模でのマップ品質スコアを示し；対角線下で、陰影は、フェージングされたSNPとの重複に基づいて生成されたリードペアの推測されたハプロタイプフェーズを示す。幾つかの実施形態において、図9Bに示されるように、生成されたプロットは、反復領域に隣接している逆位を表す。幾つかの実施形態において、図9Bに示されるように、生成されたプロットは、フェージングされたヘテロ接合の欠失のデータを表す。

20

【0185】

基準に対して1つの個体からのペアになった配列リードのマッピングが、逆位、欠失、及び重複のようなコンティグ核酸又はゲノムの構造における差を同定する、最も共通して使用される配列ベースの方法である(Tuzun et al., 2005)。図9Aと図9Bは、ヒト基準ゲノムGRCh38にマッピングされたGM12878から再びアセンブルされたクロマチンからのDNAの近接ライゲーションによって生成されたリードペアが、どのようにして2つのそのような構造差を明らかにするのかを示す。構造差の識別のためにリードペアデータの感度及び特異性を推定するために、ヘテロ接合の逆位の効果をシミュレートするように構築される、シミュレートされたデータセット上の最大尤度の弁別が試験された。試験データは、GRCh38基準配列に生成されたNA12878リードのマッピングから定義された長さLの間隔を無作為に選択して、独立して無作為に生成されたリードペアそれぞれを逆転された又は基準のハプロタイプに割り当てることにより、及び、マッピングされた座標を適宜編集することにより、構築された。非対立性の相同組換えは、ヒトゲノムに観察された構造変異の大半の原因であり、その結果、反復配列の長いブロックに生じる多くの変異の切断点がもたらされる(Kidd et al., 2008)。逆位切断点を取り囲む反復配列の長さを変える効果は、それらの距離Wの内にマッピングされた全てのリードを除去することによりシミュレートされた。逆位切断点に反復配列が無い場合、1Kbp、2Kbp、及び5Kbpそれぞれの逆位について、感度(特異性)はそれぞれ0.76(0.88)、0.89(0.89)、及び0.97(0.94)であった。逆位切断点の反復(マッピングされない)配列の1Kbp領域がシミュレーションに使用されると、5Kbpの逆位の感度(特異性)は0.81(0.76

30

40

50

）であった。

【 0 1 8 6 】

実行

本明細書に開示される技術で行われる解析を、高精度で行うことができる。解析は、少なくとも約 5 0 %、6 0 %、7 0 %、7 5 %、8 0 %、8 5 %、9 0 %、9 5 %、9 6 %、9 7 %、9 8 %、9 9 %、9 9 . 9 %、9 9 . 9 9 %、9 9 . 9 9 9 %、又はそれ以上の精度で行うことができる。解析は、少なくとも 7 0 % の精度で行うことができる。解析は、少なくとも 8 0 % の精度で行うことができる。解析は、少なくとも 9 0 % の精度で行うことができる。

【 0 1 8 7 】

本明細書に開示される技術で行われる解析を、高特異性で行うことができる。解析は、少なくとも約 5 0 %、6 0 %、7 0 %、7 5 %、8 0 %、8 5 %、9 0 %、9 5 %、9 6 %、9 7 %、9 8 %、9 9 %、9 9 . 9 %、9 9 . 9 9 %、9 9 . 9 9 9 %、又はそれ以上の特異性で行うことができる。解析は、少なくとも 7 0 % の特殊性で行うことができる。解析は、少なくとも 8 0 % の特異性で行うことができる。解析は、少なくとも 9 0 % の特異性で行うことができる。

【 0 1 8 8 】

本明細書に開示される技術で行われる解析を、高感度で行うことができる。解析は、少なくとも約 5 0 %、6 0 %、7 0 %、7 5 %、8 0 %、8 5 %、9 0 %、9 5 %、9 6 %、9 7 %、9 8 %、9 9 %、9 9 . 9 %、9 9 . 9 9 %、9 9 . 9 9 9 %、又はそれ以上の感度で行うことができる。解析は、少なくとも 7 0 % の感度で行うことができる。解析は、少なくとも 8 0 % の感度で行うことができる。解析は、少なくとも 9 0 % の感度で行うことができる。

【 0 1 8 9 】

本開示の技術の使用は、それらが実施されるコンピュータシステムの機能化を改善することができる。例えば、この技術は、与えられた解析の処理時間を、少なくとも約 5 %、1 0 %、1 5 %、2 0 %、2 5 %、3 0 %、3 5 %、4 0 %、4 5 %、5 0 %、5 5 %、6 0 %、6 5 %、7 0 %、7 5 %、8 0 %、8 5 %、9 0 %、9 5 %、又はそれ以上減らすことができる。この技術は、与えられた解析に必要なメモリを、少なくとも約 5 %、1 0 %、1 5 %、2 0 %、2 5 %、3 0 %、3 5 %、4 0 %、4 5 %、5 0 %、5 5 %、6 0 %、6 5 %、7 0 %、7 5 %、8 0 %、8 5 %、9 0 %、9 5 %、又はそれ以上減らすことができる。

【 0 1 9 0 】

本開示の技術の使用は、以前は可能でなかった解析を行うことができる。例えば、特定の遺伝子特徴は、本開示の方法無しに配列情報から検出することができない、配列情報から検出することができる。

【 0 1 9 1 】

コンピュータシステム

図 1 0 は、本明細書で提供される方法を実施するようにプログラム又は構成される、コンピュータシステム (1 0 0 1) を示す。コンピュータシステム (1 0 0 1) はユーザー又はコンピュータシステムの電子デバイスであり、ユーザー又はコンピュータシステムは電子デバイスに対して遠隔に位置付けられる。電子デバイスはモバイル電子デバイスでもよい。

【 0 1 9 2 】

コンピュータシステム (1 0 0 1) は、中央処理装置 (C P U、本明細書では「プロセッサ」及び「コンピュータプロセッサ」とも称される) (1 0 0 5) を含み、これは、シングルコア又はマルチコアのプロセッサ、或いは並行処理のための複数のプロセッサであり得る。コンピュータシステム (1 0 0 1) はまた、メモリ又はメモリ位置 (1 0 1 0) (例えば、ランダムアクセスメモリ、読み取り専用メモリ、フラッシュメモリ)、電子記憶装置 (1 0 1 5) (例えばハードディスク)、1 つ以上の他のシステムと通信す

10

20

30

40

50

るための通信インターフェース(1020)(例えばネットワークアダプタ)、及び、キャッシュ、他のメモリ、データ記憶装置、及び/又は電子ディスプレイアダプタなどの周辺機器(1025)を含む。メモリ(1010)、記憶装置(1015)、インターフェース(1020)、及び周辺機器(1025)は、マザーボードなどの通信バス(実線)を通じてCPU(1005)と通信状態にある。記憶装置(1015)は、データを保存するためのデータ記憶装置(又はデータレポジトリ)であり得る。コンピュータシステム(1001)は、通信インターフェース(1020)の助けによってコンピュータネットワーク(「ネットワーク」)(1030)に動作可能に連結され得る。ネットワーク(1030)は、インターネット及び/又はエクストラネット、インターネットと通信状態にあるイントラネット及び/又はエクストラネットであり得る。場合によっては、ネットワーク(1030)は、電気通信及び/又はデータのネットワークである。ネットワーク(1030)は、1つ以上のコンピュータサーバーを含むことができ、これはクラウドコンピューティングなどの分散コンピューティングを可能にすることができる。ネットワーク(1030)は、場合によってはコンピュータシステム(1001)の助けにより、ピアツーピア・ネットワークを実施することができ、これは、コンピュータシステム(1001)に連結されたデバイスが、クライアント又はサーバーとして動くことを可能にし得る。

10

【0193】

CPU(1005)は、プログラム又はソフトウェアに統合され得る一連の機械可読命令を実行することができる。この命令は、メモリ(1010)などのメモリ位置に保存され得る。この命令は、CPU(1005)に向けることができ、これは後に、本開示の方法を実施するようにCPU(1005)をプログラム又は構成することができる。CPU(1005)により実行される動作の例は、フェッチ、デコード、実行、及びライトバックを含み得る。

20

【0194】

CPU(1005)は、集積回路など回路の一部であり得る。システム(1001)の1つ以上の他のコンポーネントを回路に含めることができる。場合によっては、回路は特定用途向け集積回路(ASIC)である。

【0195】

記憶装置(1015)は、ドライバー、ライブラリ、及びセーブされたプログラムなどのファイルを記憶することができる。記憶装置(1015)は、ユーザーデータ、例えばユーザーの嗜好性及びユーザーのプログラムを記憶することができる。コンピュータシステム(1001)は、場合によっては、イントラネット又はインターネットを通じてコンピュータシステム(1001)と通信状態にあるリモートサーバー上に位置付けられるなど、コンピュータシステム(1001)の外側にある1つ以上の追加のデータ記憶装置を含み得る。

30

【0196】

コンピュータシステム(1001)は、ネットワーク(1030)を通じて1つ以上のリモートコンピュータシステムと通信することができる。例えば、コンピュータシステム(1001)は、ユーザー(例えばサービス提供者)のリモートコンピュータシステムと通信することができる。リモートコンピュータシステムの例は、パーソナルコンピュータ(例えば、持ち運び可能なPC)、スレート又はタブレットPC(例えばApple(登録商標) iPad(登録商標)、Samsung(登録商標) Galaxy Tab)、電話、スマートフォン(例えばApple(登録商標) iPhone(登録商標)、Android-enabledデバイス、Blackberry(登録商標))、又は携帯情報端末を含む。ユーザーは、ネットワーク(1030)を介してコンピュータシステム(1001)にアクセスすることができる。

40

【0197】

本明細書に記載されるような方法は、例えばメモリ(1010)又は電子記憶装置(1015)の上など、コンピュータシステム(1001)の電子記憶装置の位置に記憶された機械(例えばコンピュータプロセッサ)実行可能コードによって実施することができる

50

。機械実行可能又は機械可読コードは、ソフトウェアの形で提供することができる。

【 0 1 9 8 】

使用中、コードはプロセッサ (1 0 0 5) により実行され得る。場合によっては、コードは、記憶装置 (1 0 1 5) から検索され、且つプロセッサ (1 0 0 5) による即時のアクセスのためにメモリ (1 0 1 0) に記憶することができる。幾つかの状況において、電子記憶装置 (1 0 1 5) は排除することができ、機械実行可能命令はメモリ (1 0 1 0) に記憶される。

【 0 1 9 9 】

コードは、コードを実行するのに適したプロセッサを持つ機械との使用のために予めコンパイルされ且つ構成することができ、或いは、実行時間中にコンパイルすることができる。コードは、予めコンパイルされた又はアズコンパイルされた (a s - c o m p i l e d) 様式でコードが実行を可能にするために選択することができる、プログラミング言語で供給され得る。

【 0 2 0 0 】

コンピュータシステム (1 0 0 1) などの、本明細書で提供されるシステムと方法の態様は、プログラミングの際に統合することができる。この技術の様々な態様は、典型的に一種の機械可読媒体上で運ばれる又はそれに統合される機械 (又はプロセッサ) 実行可能コード及び / 又は関連データの形で、「製品」又は「製造用品」として考慮され得る。機械実行可能コードは、メモリ (例えば、読み取り専用メモリ、ランダムアクセスメモリ、フラッシュメモリ) 又はハードディスクなどの電子記憶装置に記憶することができる。「記憶」型の媒体は、様々な半導体メモリ、テープドライブ、ディスクドライブなどの、コンピュータやプロセッサの有形メモリ、或いはその関連するモジュールの何れか又は全てを含むことができ、これは、ソフトウェアのプログラミングのためにいかなる時も非一時的な記憶を提供し得る。ソフトウェアの全て又は一部は、インターネット又は様々な他の電気通信ネットワークを介して時々通信される。そのような通信は、例えば、1つのコンピュータ又はプロセッサから別のものへの、例えば、管理サーバー又はホストコンピュータからアプリケーションサーバーのコンピュータプラットフォームへのソフトウェアのローディングを可能にし得る。故に、ソフトウェア要素を持ち得る別のタイプの媒体は、有線及び光地上通信線ネットワークを通じた、及び様々なエアリンク (a i r - l i n k s) 上での、ローカルデバイス間の物理インターフェースにわたって使用されるものなどの、光波、電波、及び電磁波を含む。有線又は無線リンク、光リンクなどの、そのような波を運ぶ物理要素はまた、ソフトウェアを持つ媒体と考慮され得る。本明細書で使用されるように、非一時的で有形の「記憶」媒体に制限されなければ、コンピュータ又は機械「可読媒体」などの用語は、実行のためのプロセッサに命令を提供することに関与する媒体を指す。

【 0 2 0 1 】

従って、コンピュータ実行可能コードなどの機械可読媒体は、有形記憶媒体、搬送波媒体、又は物理送信媒体を含むがこれらに限定されない、多くの形態をとってもよい。不揮発性記憶媒体は、例えば、図面に示されるデータベースなどを実施するために使用され得るものなど、コンピュータなどにおける記憶デバイスの何れかといった、光ディスク又は磁気ディスクを含む。揮発性記憶媒体は、そのようなコンピュータプラットフォームのメインメモリなどのダイナミックメモリを含む。有形送信媒体は、同軸ケーブル；コンピュータシステム内のバスを含むワイヤーを含む、銅線及び光ファイバーを含んでいる。搬送波送信媒体は、無線周波 (R F) 及び赤外線 (I R) データ通信中に生成されたものなどの、電気信号又は電磁気信号、或いは音波又は光波の形態をとり得る。それ故、コンピュータ可読媒体の共通の形態は、例えば：フロッピーディスク、フレキシブルディスク、ハードディスク、磁気テープ、他の磁気媒体、C D - R O M、D V D、又はD V D - R O M、他の光学媒体、パンチカードペーパーテープ、ホールのパターンを備えた (w i t h p a t t e r n s o f h o l e s) 他の物理的記憶媒体、R A M、R O M、P R O M、及びE P R O M、F L A S H - E P R O M、他のメモリチップ又はカートリッジ、データ

10

20

30

40

50

又は命令を運ぶ搬送波、そのような搬送波を運ぶケーブル又はリンク、又は、コンピュータがプログラミングコード及び／又はデータを読む他の媒体を含む。コンピュータ可読媒体のこれらの形態の多くは、実行のためにプロセッサに１つ以上の命令の１つ以上のシーケンスを運ぶことに関係し得る。

【０２０２】

コンピュータシステム（１００１）は、例えば、訓練されたアルゴリズムの出力又は読み出しを提供するためにユーザインターフェース（ＵＩ）（１０４０）を含む電子ディスプレイ（１０３５）を含むか、又はそれと通信状態にあってもよい。ＵＩの例は、限定されないが、グラフィカルユーザインターフェース（ＧＵＩ）及びウェブベースのユーザインターフェースを含む。

10

【０２０３】

本開示の方法とシステムは、１つ以上のアルゴリズムによって実施することができる。アルゴリズムは、中央処理装置（１００５）による実行の後にソフトウェアによって実施することができる。

【０２０４】

本明細書中のコンピュータシステムは、場合によっては、本明細書に開示される又は当業者に既知のものといった、機械学習動作を実行するように構成される。

【０２０５】

配列決定ベースのアッセイ

ハイブリダイゼーション（例えば、標識、アレイハイブリダイゼーション、ＦＩＳＨなどの蛍光プローブハイブリダイゼーション、抗体ハイブリダイゼーション）又は増幅（例えばＰＣＲ）などの、非配列決定ベースのアッセイを利用して、ＤＮＡタンパク質複合体（例えばクロマチン）又は他の結合ＤＮＡ複合体（例えばビーズ或いは他の基質と複合されるＤＮＡ）上で遺伝子特徴（例えば遺伝子再配置）を検出することができる。

20

【０２０６】

ＤＮＡ複合体（例えばクロマチン又は他の結合ＤＮＡ複合体などのＤＮＡタンパク質複合体）は、本明細書で議論される技術を使用して集めることができる。例えば、ＤＮＡ複合体は、保存されたサンプル（例えばＦＦＰＥサンプル）から回復され、又は、単離されたＤＮＡから再構成することができる。一例において、クロマチンは、熱処理とタンパク質分解により、保存されたサンプル（例えばＦＦＰＥサンプル）から解放することができる。

30

【０２０７】

ＤＮＡ複合体はキャプチャ又は精製され得る。例えば、ＤＮＡ複合体（例えばクロマチン）は、固相上でキャプチャすることができる。場合によっては、固相は、カルボキシル化された常磁性粒子などのカルボキシル化された基質を含む。

【０２０８】

ＤＮＡ複合体は、酵素的（例えば、制限酵素、フラグメンターゼ、トランスポサアーゼ）、熱的、及び物理的な断片化を含むがこれらに限定されない、本明細書に開示される方法によって断片化及びライゲートすることができる。ライゲーションの前に平滑末端化が行われてもよい。

40

【０２０９】

ＤＮＡ複合体を更なる解析のために分割することができる。例えば、ＤＮＡ複合体（例えばクロマチン）は、液滴（例えば微小流体液滴）、ウェル、アレイスポット、又は他の区分に分割することができる。

【０２１０】

ＤＮＡ複合体を様々な手段により解析することができる。増幅（例えばＰＣＲ）は、様々な切断点を標的として（例えば、プライマーペアで標的として）、（例えば液滴ＰＣＲなどの区分において）処理することができる。蛍光オリゴヌクレオチドプローブなどでのハイブリダイゼーションアッセイを使用して、様々な切断点を標的とすることができる。再編成は、近くの遺伝子座の近接ライゲーションの可能性の変化によるシグナルの変化に

50

より、検出され得る。場合によっては、Taq-Manプローブを使用することができる。場合によっては、SYBRプローブを使用することができる。そのような解析は、例えば液滴、ウェル、アレイスポット、又は他の区分において多重化され得る。

【0211】

一例において、クロマチンは、緩やかな熱処理とタンパク質分解により、保存されたサンプル（例えばFFPE）から解放される。解放されたクロマチンは、常磁性のカルボキシル化されたポリスチレンビーズを含む固相上で、キャプチャされる。キャプチャされたクロマチンに結合されたDNAは断片化され（例えば酵素学的に）、断片化された末端は平滑末端化される。クロマチンに関連づけられる平滑末端化されたDNAは、他の近くのDNAにライゲートされる。染色体間変異体の存在は、液滴ベースのPCR又は蛍光オリゴヌクレオチドプローブのハイブリダイゼーションなどにより定量化される。欠失及び逆位は、近くの遺伝子座の近接ライゲーションの可能性の変化（例えば増加）により、信号を変化（例えば増加）させる。

10

【0212】

再編成アッセイは、再編成の配列決定ベースのアッセイを含む、本明細書に記載されるものなどの配列決定ベースのアッセイと組み合わせることができる。例えば、PCR又はハイブリダイゼーションのアッセイの後、クロマチンは本明細書に開示されるように配列決定され且つ解析され得る。

【0213】

キット

20

本明細書には、本明細書に開示される技術を実行するためのキットが開示される。キットは箱などの包装材料のなかに含めることができ、特定数の反応のための材料が包装材料の各ユニット中にある。場合によっては、キットは、1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50、又はそれ以上の反応のための試薬を含む。

【0214】

本明細書に開示されるようなキットは、本明細書に開示される方法を実行し、且つ本明細書に開示される組成物を生成又は解析するのに必要な試薬を幾つか又は全て含む。場合によっては、キットは、本明細書に開示される方法を実行し且つ本明細書に開示される組成物を生成又は解析するのに必要な試薬の亜群を含み、随意に、キットには含まれないが試薬の供給業者から容易に入手可能な、試薬に関する指示書を含む。

30

【0215】

本明細書に開示される幾つかのキットは、緩衝液、DNA結合剤、アフィニティータグ結合剤、デオキシリボヌクレオチド、タグ付けされたデオキシリボヌクレオチド、DNA断片化剤、末端修復酵素、リガーゼ、タンパク質除去剤、及び保存されたサンプルからゲノムの構造情報を得る際の使用説明書を含む。キットは随意に、緩衝液、ヌクレオチド、フォワードプライマー、リバースプライマー、及び熱安定性DNAポリメラーゼなどの、PCRのための試薬を含む。

40

【0216】

一部のキットの中の緩衝液は、制限消化緩衝液、末端修復緩衝液、ライゲーション緩衝液、TE緩衝液、洗浄緩衝液、TWB溶液、NTB溶液、LWB溶液、NWB溶液、及び架橋逆転緩衝液のうち少なくとも1つを含む。代表的な消化緩衝液は、DpnII緩衝液、又は、NEB緩衝液2などの商用の緩衝液、或いはNEB緩衝液2と機能的に類似している商用の緩衝液である。典型的なライゲーション緩衝液は、T4DNAリガーゼ緩衝液、BSA、及びTriton X-100を含む。

【0217】

キットに含まれる、又はキットの試薬と組み合わせて説明書において言及される、他の典型的な試薬は、トリスとEDTAを含むTE緩衝液、トリスと塩化ナトリウムを含む洗

50

浄緩衝液、トリス、EDTA、及びTween 20のうち1つ以上を含むTWB溶液、トリスの、EDTA、及び塩化ナトリウムのうち1つ以上を含むNTB溶液、トリス、塩化リチウム、EDTA、及びTween 20のうち1つ以上を含むLWB溶液、トリス、塩化ナトリウム、EDTA、及びTween 20のうち少なくとも1つを含むNWB溶液、及び、トリス、SDS、及び塩化カルシウムのうち1つ以上を含む架橋逆転緩衝液を含む。

【0218】

幾つかのキットは、ストレプトアビジンビーズ、例えばdynamabeadsなどのアフィニティタグ結合剤を含むか、又はそれに適合するように構成される。

【0219】

キットは、dATP、dCTP、dGTP、及びdTTPなどのヌクレオチド、場合によっては、ヌクレオチドのビオチン化バージョンを含むか、それらに適合する。

10

【0220】

本明細書中のキットに含まれる又はそれに適合するDNA断片化剤は、DpnI、トランスポサゼ、ヌクレアーゼ、音波処理デバイス、流体力学的剪断デバイス、及び二価金属カチオンなどの制限酵素のうち少なくとも1つを含む。

【0221】

本明細書中のキットに含まれる又はそれに適合する末端修復酵素は、T4 DNAポリメラーゼ、クレノウDNAポリメラーゼ、及びT4ポリヌクレオチドキナーゼのうち少なくとも1つを含む。

【0222】

20

本明細書中のキットに含まれる又はそれに適合する典型的なりガーゼは、T4リガーゼを含む。

【0223】

本明細書中のキットに含まれる又はそれと組み合わせて使用されるタンパク質除去試薬は、プロテイナーゼK、ストレプトマイセス-グリセウスプロテアーゼ、セリンプロテアーゼ、システインプロテアーゼ、トレオニンプロテアーゼ、アスパラギン酸プロテアーゼ、グルタミン酸プロテアーゼ、メタロプロテアーゼ、及びアスパラギンペプチドリアーゼなどの、フェノールおよびプロテイナーゼを含む。

【0224】

キットは随意に、パラフィンなどの包埋材料を除去するために使用される溶媒などの溶媒を含むか、又はそれに適合する。

30

【0225】

定義

本明細書及び添付の請求項に使用されるように、単数形「a」、「an」、及び「the」は、文脈が明確に別段に指定していない限り、複数の指示対象を含む。故に、例えば、「コンティグ」に対する言及は、そのようなコンティグを複数含み、「染色体の物理レイアウトをプローブする」に対する言及は、当業者に既知の染色体及びその同等物の物理レイアウトをプローブする1以上の方法に対する言及を含んでいる。

【0226】

また、「及び」の使用は、特に明記しない限り、「及び/又は」を意味する。同様に、「含む(comprise)」、「含む(comprises)」、「含むこと(comprising)」、「含む(include)」、「含む(includes)」、及び「含むこと(including)」は互換的なものであり、限定的なものとしては意図されていない。

40

【0227】

様々な実施形態の記載が用語「含むこと(comprising)」を使用する場合、当業者は、幾つかの特定の例において、言語「～から実質的に成る」又は「～から成る」を使用して実施形態が代替的に記載され得ることを理解すると、更に理解されたい。

【0228】

用語「配列決定リード」は、本明細書で使用されるように、配列が判定されたDNAの

50

断片を指す。

【0229】

用語「コンティグ」は、本明細書で使用されるように、DNA配列の隣接する領域を指す。「コンティグ」は、当該技術分野で既知である任意の数の方法、例えば、配列を重複させるために配列決定リードを比較すること、及び/又は、どの配列決定リードが隣接する可能性が高いかを識別するために既知の配列のデータベースに対して配列決定リードを比較することにより判定され得る。

【0230】

本明細書で使用されるような用語「被験体」は、真核生物又は原核生物を指す場合がある。

10

【0231】

用語「ネイキッドDNA」は、本明細書で使用されるように、複合化されたタンパク質が実質的に無いDNAを指す場合がある。例えば、それは、約50%、約40%、約30%、約20%、約10%、約5%、又は約1%未満の、細胞核に見出される内因性タンパク質と共に複合化されたDNAを指す場合がある。

【0232】

用語「再構成クロマチン」は、本明細書で使用されるように、ネイキッドDNAなどの核酸に対する核酸結合部分を複合化することにより形成されたクロマチンを指す場合がある。場合によっては、これら部分は、核タンパク質又はヒストンなどの核酸タンパク質であるが、ナノ粒子などの他の部分も考慮される。

20

【0233】

本明細書で使用されるような用語「リードペア(read pair)」又は「リードペア(read - pair)」は、配列情報を提供するために結合される2つ以上の要素を指す場合がある。場合によっては、リードペアの数は、マッピング可能なリードペアの数を指す場合がある。他の場合において、リードペアの数は、生成されたリードペアの総数を指す場合がある。

【0234】

本明細書で使用されるような「組織サンプル」は、個体、又は核酸を潜在的に含む環境の、生物学的サンプルを指す。腫瘍は例えば、組織であると考慮され、腫瘍から得られたサンプルは組織サンプルを構築するが、場合によっては、この用語は、胃又は腸の切片などの異機種環境から得られたサンプル、或いは、互いに対して空間的に分配された複数のソースからの核酸を含む環境サンプルを指す。

30

【0235】

数に関して本明細書で使用されるような「約」は、その数の+/-10%の数を指す。範囲に関して使用されるように、「約」は、範囲の示された下限の10%未満の下限、及び範囲の示された上限の10%を超える上限を持つ範囲を指す。

【0236】

本明細書で使用されるような「プローブ」は、標的への結合を通じて情報を伝える分子を指す。典型的なプローブは、オリゴヌクレオチド(oligonucleotide)分子及び抗体を含む。オリゴヌクレオチド分子は、標的にアニールし、且つ蛍光特性を変えることで情報を伝えることにより、或いは代替的に、標的にアニールし、且つ標的存在を示すアンプリコン等の生成物の合成を容易にすることにより、プローブとして作用し得る。即ち、本明細書で使用されるような用語「プローブ」は様々に、抗体プローブおよび他の小分子プローブの他、オリゴ核酸分子も考慮し、これらは例えば、蛍光状況の変化に通じる標的へのハイブリダイゼーションを直接介して信号を生成することにより作用し、或いは、標的存在を示すアンプリコンの合成を容易にすることにより作用する。

40

【0237】

本明細書で使用されるように、タンパク質及び核酸が複合体を形成するようにこれ以上アセンブルされないと、DNAタンパク質複合体は破壊又は分裂される。場合によっては、複合体は、完全に変性され又は分解され、その結果、タンパク質DNA結合は残らない

50

。代替的に、場合によっては、DNAタンパク質複合体は、第1の核酸セグメント及び第2の核酸セグメントがリン酸ジエステル結合から独立してこれ以上共に保持されないと、実質的に破壊される。

【0238】

他に定義されない限り、本明細書で使用される全ての技術用語及び科学用語は、本開示が属する分野における当業者により共通して理解されるものと同じ意味を有する。本明細書に記載されるものと同様又は同等の方法及び試薬が、開示された方法及び組成物の実施に使用され得るが、ここでは例示的な方法及び材料が記載されている。

【0239】

本明細書の開示は、以下のように番号を付けた実施形態の部分的なリストを参照することで更に明確にされる。1. 保存された組織サンプルからゲノムの構造情報を得る方法であって、該方法は：タンパク質DNA複合体が分裂されないように、保存された組織サンプルから核酸を単離させる工程；第1のDNAセグメント及び第2のDNAセグメントが共通のタンパク質DNA複合体から生じると識別されるように、タンパク質DNA複合体にタグを付ける工程；共通のDNA複合体から第1のDNAセグメント及び第2のDNAセグメントを分離する工程；第1のDNAセグメント及び第2のDNAセグメントから配列情報を生成する工程；及び共通のタンパク質DNA複合体を示すタグ配列を共有する配列情報を共通のゲノムの構造に割り当てる工程を含む。2. 保存された組織サンプルは、架橋されたパラフィン包埋サンプルである、ことを特徴とする実施形態1などの上記実施形態の何れか1つに記載の方法。3. タグ配列は、複合体を識別するオリゴタグを含む、ことを特徴とする実施形態1などの上記実施形態の何れか1つに記載の方法。4. タグ配列は、第1のセグメントから第2のセグメントライゲーションにより生じる、ことを特徴とする実施形態1などの上記実施形態の何れか1つに記載の方法。5. タンパク質DNA複合体が分裂されないように、保存された組織サンプルから核酸を単離させる工程は、架橋されたパラフィン包埋組織サンプルをキシレンに接触させる工程を含む、ことを特徴とする実施形態1などの上記実施形態の何れか1つに記載の方法。6. タンパク質DNA複合体が分裂されないように、保存された組織サンプルから核酸を単離させる工程は、保存された組織サンプルをエタノールに接触させる工程を含む、ことを特徴とする実施形態1などの上記実施形態の何れか1つに記載の方法。7. タンパク質DNA複合体が分裂されないように、保存された組織サンプルから核酸を単離させる工程は、沸騰状態からサンプルを保護する工程を含む、ことを特徴とする実施形態1などの上記実施形態の何れか1つに記載の方法。8. 共通のDNA複合体から第1のDNAセグメント及び第2のDNAセグメントを分離する工程は、プロテイナーゼK処理を含む、ことを特徴とする実施形態1などの上記実施形態の何れか1つに記載の方法。9. 保存された組織サンプルは組織中のその配置を反映する位置情報を保存する、ことを特徴とする実施形態1などの上記実施形態の何れか1つに記載の方法。10. 保存された組織サンプルは核酸を単離させる前に均質化されない、ことを特徴とする実施形態1などの上記実施形態の何れか1つに記載の方法。11. 保存された組織サンプルは核酸を単離させる前に少なくとも1週間保管される、ことを特徴とする実施形態1などの上記実施形態の何れか1つに記載の方法。12. 保存された組織サンプルは核酸を単離させる前に少なくとも6か月間保管される、ことを特徴とする実施形態1などの上記実施形態の何れか1つに記載の方法。13. 保存された組織サンプルは核酸を単離させる前に収集点から輸送される、ことを特徴とする実施形態1などの上記実施形態の何れか1つに記載の方法。14. 保存された組織サンプルは無菌環境で集められる、ことを特徴とする実施形態1などの上記実施形態の何れか1つに記載の方法。15. 保存された組織サンプルは核酸を単離させる前に非無菌環境に位置付けられる、ことを特徴とする実施形態1などの上記実施形態の何れか1つに記載の方法。16. 保存された組織サンプルからゲノムの構造情報を得る方法であって、該方法は：50 kbを超える核酸断片が再生されるように、保存された組織サンプルから核酸を単離させる工程；核酸分子の第1のDNAセグメント及び第2のDNAセグメントがそれら共通のリン酸ジエステル骨格から独立して共に保持されるように、少なくとも1つの複合体を形成す

10

20

30

40

50

るために複数の核酸結合部分に核酸を接触させる工程；少なくとも1つの複合体の少なくとも1つのリン酸ジエステル骨格を切断する工程；第1のDNAセグメント及び第2のDNAセグメントが共通の複合体から生じると識別されるように、少なくとも1つの複合体にタグを付ける工程；共通の複合体から第1のDNAセグメント及び第2のDNAセグメントを分離する工程；第1のDNAセグメント及び第2のDNAセグメントから配列情報を生成する工程；及び共通のタンパク質DNA複合体を示すタグ配列を共有する配列情報を共通のゲノムの構造に割り当てる工程を含む。17．保存された組織サンプルは、架橋されたパラフィン包埋組織サンプルである、ことを特徴とする実施形態16などの上記実施形態の何れか1つに記載の方法。18．タグ配列は、複合体を識別するオリゴタグを含む、ことを特徴とする実施形態16などの上記実施形態の何れか1つに記載の方法。19．タグ配列は、第1のDNAセグメントのライゲートから第2のDNAセグメントまで生じる、ことを特徴とする実施形態16などの上記実施形態の何れか1つに記載の方法。20．50 kb より多くの核酸断片が回復されるように、保存された組織サンプルから核酸を単離させる工程は、アントラニル酸塩とホスファニル酸塩のうち少なくとも1つに保存された組織サンプルを接触させる工程を含む、ことを特徴とする実施形態16などの上記実施形態の何れか1つに記載の方法。21．単離させる工程は40 以下の温度で行われる、ことを特徴とする実施形態20などの上記実施形態の何れか1つに記載の方法。22．単離させる工程は40 以下の温度で行われる、ことを特徴とする実施形態16などの上記実施形態の何れか1つに記載の方法。23．共通のDNA複合体から第1のDNAセグメント及び第2のDNAセグメントを分離する工程は、プロテイナーゼK処理を含む、ことを特徴とする実施形態16などの上記実施形態の何れか1つに記載の方法。24．複数の核酸結合部分は核タンパク質を含む、ことを特徴とする実施形態16などの上記実施形態の何れか1つに記載の方法。25．複数の核酸結合部分はトランスポサーゼを含む、ことを特徴とする実施形態16などの上記実施形態の何れか1つに記載の方法。26．複数の核酸結合部分はヒストンを含む、ことを特徴とする実施形態16などの上記実施形態の何れか1つに記載の方法。27．複数の核酸結合部分は核酸結合タンパク質を含む、ことを特徴とする実施形態16などの上記実施形態の何れか1つに記載の方法。28．複数の核酸結合部分はナノ粒子を含む、ことを特徴とする実施形態16などの上記実施形態の何れか1つに記載の方法。29．少なくとも1つの複合体の少なくとも1つのリン酸ジエステル骨格を切断する工程は、制限エンドヌクレアーゼに接触させる工程を含む、ことを特徴とする実施形態16などの上記実施形態の何れか1つに記載の方法。30．少なくとも1つの複合体の少なくとも1つのリン酸ジエステル骨格を切断する工程は、非特異的エンドヌクレアーゼに接触させる工程を含む、ことを特徴とする実施形態16などの上記実施形態の何れか1つに記載の方法。31．少なくとも1つの複合体の少なくとも1つのリン酸ジエステル骨格を切断する工程は、DNAを剪断する工程を含む、ことを特徴とする実施形態16などの上記実施形態の何れか1つに記載の方法。32．少なくとも1つの複合体の少なくとも1つのリン酸ジエステル骨格を切断する工程は、トランスポサーゼに接触させる工程を含む、ことを特徴とする実施形態16などの上記実施形態の何れか1つに記載の方法。33．少なくとも1つの複合体の少なくとも1つのリン酸ジエステル骨格を切断する工程は、トポイソメラーゼに接触させる工程を含む、ことを特徴とする実施形態16などの上記実施形態の何れか1つに記載の方法。34．保存された組織サンプルは組織中のその配置を反映する位置情報を保存する、ことを特徴とする実施形態16などの上記実施形態の何れか1つに記載の方法。35．保存された組織サンプルは核酸を単離させる前に均質化されない、ことを特徴とする実施形態16などの上記実施形態の何れか1つに記載の方法。36．保存された組織サンプルは核酸を単離させる前に少なくとも1週間保管される、ことを特徴とする実施形態16などの上記実施形態の何れか1つに記載の方法。37．保存された組織サンプルは核酸を単離させる前に少なくとも6か月間保管される、ことを特徴とする実施形態16などの上記実施形態の何れか1つに記載の方法。38．保存された組織サンプルは核酸を単離させる前に収集点から輸送される、ことを特徴とする実施形態16などの上記実施形態の何れか1つに記載の方法。39．保存された組織

10

20

30

40

50

サンプルは無菌環境で集められる、ことを特徴とする実施形態 16 などの上記実施形態の何れか 1 つに記載の方法。40. 保存された組織サンプルは核酸を単離させる前に非無菌環境に位置付けられる、ことを特徴とする実施形態 16 などの上記実施形態の何れか 1 つに記載の方法。41. 空間的に分布されたゲノムの構造情報を回復する方法であって、該方法は：組織サンプルを得る工程；固定された三次元のパラフィン包埋組織サンプルの第 1 の部分から一部を抽出する工程；タンパク質 DNA 複合体が分裂されないように、第 1 の部分の一部から核酸を単離させる工程；第 1 の DNA セグメント及び第 2 の DNA セグメントが共通のタンパク質 DNA 複合体から生じると識別されるように、タンパク質 DNA 複合体にタグを付ける工程；共通の DNA 複合体から第 1 の DNA セグメント及び第 2 の DNA セグメントを分離する工程；第 1 の DNA セグメント及び第 2 の DNA セグメントから配列情報を生成する工程；共通のタンパク質 DNA 複合体を示すタグ配列を共有する配列情報を共通のゲノムの構造に割り当てる工程；及び組織サンプルの第 1 の部分に共通のゲノムの構造を割り当てる工程を含む。42. 組織サンプルは、固定された三次元のパラフィン包埋サンプルである、ことを特徴とする実施形態 41 などの上記実施形態の何れか 1 つに記載の方法。43. 架橋されたパラフィン包埋組織サンプルは組織中のその配置を反映する位置情報を保存する、ことを特徴とする実施形態 41 などの上記実施形態の何れか 1 つに記載の方法。44. 架橋されたパラフィン包埋組織サンプルは核酸を単離させる前に均質化されない、ことを特徴とする実施形態 41 などの上記実施形態の何れか 1 つに記載の方法。45. 架橋されたパラフィン包埋組織サンプルは核酸を単離させる前に少なくとも 1 週間保管される、ことを特徴とする実施形態 41 などの上記実施形態の何れか 1 つに記載の方法。46. 架橋されたパラフィン包埋組織サンプルは核酸を単離させる前に少なくとも 6 か月間保管される、ことを特徴とする実施形態 41 などの上記実施形態の何れか 1 つに記載の方法。47. 架橋されたパラフィン包埋組織サンプルは核酸を単離させる前に収集点から輸送される、ことを特徴とする実施形態 41 などの上記実施形態の何れか 1 つに記載の方法。48. 架橋されたパラフィン包埋組織サンプルは無菌環境で集められる、ことを特徴とする実施形態 41 などの上記実施形態の何れか 1 つに記載の方法。49. 架橋されたパラフィン包埋組織サンプルは核酸を単離させる前に非無菌環境に位置付けられる、ことを特徴とする実施形態 41 などの上記実施形態の何れか 1 つに記載の方法。50. タグ配列は、複合体を識別するオリゴタグを含む、ことを特徴とする実施形態 41 などの上記実施形態の何れか 1 つに記載の方法。51. タグ配列は、第 1 のセグメントから第 2 のセグメントライゲーションにより生じる、ことを特徴とする実施形態 41 などの上記実施形態の何れか 1 つに記載の方法。52. タンパク質 DNA 複合体が分裂されないように、架橋されたパラフィン包埋組織サンプルから核酸を単離させる工程は、架橋されたパラフィン包埋組織サンプルをキシレンに接触させる工程を含む、ことを特徴とする実施形態 41 などの上記実施形態の何れか 1 つに記載の方法。53. タンパク質 DNA 複合体が分裂されないように、架橋されたパラフィン包埋組織サンプルから核酸を単離させる工程は、架橋されたパラフィン包埋組織サンプルをエタノールに接触させる工程を含む、ことを特徴とする実施形態 41 などの上記実施形態の何れか 1 つに記載の方法。54. タンパク質 DNA 複合体が分裂されないように、架橋されたパラフィン包埋組織サンプルから核酸を単離させる工程は、沸騰状態からサンプルを保護する工程を含む、ことを特徴とする実施形態 41 などの上記実施形態の何れか 1 つに記載の方法。55. 共通の DNA 複合体から第 1 の DNA セグメント及び第 2 の DNA セグメントを分離する工程は、プロテイナーゼ K 処理を含む、ことを特徴とする実施形態 41 などの上記実施形態の何れか 1 つに記載の方法。56. 組織サンプルは、固定された三次元のパラフィン包埋サンプルである、ことを特徴とする実施形態 41 などの上記実施形態の何れか 1 つに記載の方法。57. 処置レジメンの試験結果を再評価する方法であって、該方法は：患者集団の処置レジメン結果に関連するデータを得る工程；前記患者集団の複数の患者から固定された組織サンプルを得る工程；前記固定された組織サンプルから核酸複合体を抽出する工程；前記核酸複合体を複数の前記固定された組織サンプルに使用して、ゲノムの構造情報を判定する工程；及び処置レジメン結果に関連するゲノムの構造情報を識別するよ

10

20

30

40

50

うに、処置レジメン結果に関連するデータをゲノムの構造情報に関連づける工程を含む。

58．前記固定された組織サンプルから核酸複合体を抽出する工程；及び前記核酸複合体を複数の前記固定された組織サンプルに使用してゲノムの構造情報を判定する工程は、実施形態1乃至15の何れか1つに記載の方法を含む、ことを特徴とする実施形態57などの上記実施形態の何れか1つに記載の方法。

59．前記固定された組織サンプルから核酸複合体を抽出する工程；及び前記核酸複合体を複数の前記固定された組織サンプルに使用してゲノムの構造情報を判定する工程は、実施形態16などの上記実施形態の何れか1つに記載の方法を含む、ことを特徴とする実施形態57などの上記実施形態の何れか1つに記載の方法。

60．前記固定された組織サンプルから核酸複合体を抽出する工程；及び前記核酸複合体を複数の前記固定された組織サンプルに使用してゲノムの構造情報を判定する工程は、実施形態41などの上記実施形態の何れか1つに記載の方法を含む、ことを特徴とする実施形態57などの上記実施形態の何れか1つに記載の方法。

61．ヌクレオチド配列アセンブリの方法であって、該方法は：(a)固定された組織サンプルを提供する工程；(b)前記固定された組織サンプルから架橋されたDNA：タンパク質の複合体を回復する工程；(c)前記架橋されたDNA：タンパク質の複合体のDNAの第1の部分を、前記架橋されたDNA：タンパク質の複合体のDNAの第2の部分にライゲートする工程であって、それによりライゲートされたDNAを形成する、工程；(d)前記架橋されたDNA：タンパク質の複合体からライゲートされたDNAを抽出する工程；(e)前記ライゲートされたDNAのライゲーション結合の何れかの側で少なくとも一部を配列決定する工程；及び(f)ヌクレオチド配列をアセンブルするために前記配列決定する工程からの情報を使用する工程を含む。

62．前記固定された組織サンプルはホルマリン固定される、ことを特徴とする実施形態61などの上記実施形態の何れか1つに記載の方法。

63．前記固定された組織は、ホルマリン固定パラフィン包埋(FFP E)される、ことを特徴とする実施形態61などの上記実施形態の何れか1つに記載の方法。

64．前記架橋されたDNA：タンパク質の複合体はクロマチンを含む、ことを特徴とする実施形態61などの上記実施形態の何れか1つに記載の方法。

65．前記ライゲートする工程は平滑末端ライゲーションを含む、ことを特徴とする実施形態61などの上記実施形態の何れか1つに記載の方法。

66．前記ライゲートする工程の前に、前記架橋されたDNA：タンパク質の複合体からDNAを消化する工程を更に含む、実施形態61などの上記実施形態の何れか1つに記載の方法。

67．前記消化する工程は制限酵素消化を含む、ことを特徴とする実施形態66などの上記実施形態の何れか1つに記載の方法。

68．前記消化する工程の後に、平滑末端を産生するために前記消化する工程から接着末端を充填する工程を更に含む、請求項66などの上記実施形態の何れか1つに記載の方法。

69．前記充填する工程はビオチン化ヌクレオチドを使用して行われる、ことを特徴とする実施形態68などの上記実施形態の何れか1つに記載の方法。

70．前記回復する工程は、前記架橋されたDNA：タンパク質の複合体のDNAを固形支持体に結合する工程を含む、ことを特徴とする実施形態61などの上記実施形態の何れか1つに記載の方法。

71．前記抽出する工程は、前記架橋されたDNA：タンパク質の複合体からタンパク質を消化する工程を含む、ことを特徴とする実施形態61などの上記実施形態の何れか1つに記載の方法。

72．前記情報は、2000を超える塩基対(bp)の距離にわたる長距離情報を含む、ことを特徴とする実施形態61などの上記実施形態の何れか1つに記載の方法。

73．前記距離は10000bpを超える、ことを特徴とする実施形態72などの上記実施形態の何れか1つに記載の方法。

74．前記距離は100000bpを超える、ことを特徴とする実施形態73などの上記実施形態の何れか1つに記載の方法。

75．前記距離は200000bpを超える、ことを特徴とする実施形態74などの上記実施形態の何れか1つに記載の方法。

76．前記回復する工程の前に、前記固定された組織サンプルの包埋材料を溶かす工程を更に含む、実施形態61などの上記実施形態の何れか1つに記載の方法。

77．前記包埋材料はパラフィンを含む、ことを特徴とする実施形態76などの上記実施形態の何れか1つに記載の方法。

78．架橋されたパラフィン包埋組織サンプルは組織中のその配置を反映する位置情報を保存する、ことを特徴とする実施形態61などの上記実施形態

10

20

30

40

50

の何れか 1 つに記載の方法。79．架橋されたパラフィン包埋組織サンプルは核酸を単離させる前に均質化されない、ことを特徴とする実施形態 61 などの上記実施形態の何れか 1 つに記載の方法。80．架橋されたパラフィン包埋組織サンプルは核酸を単離させる前に少なくとも 1 週間保管される、ことを特徴とする実施形態 61 などの上記実施形態の何れか 1 つに記載の方法。81．架橋されたパラフィン包埋組織サンプルは核酸を単離させる前に少なくとも 6 か月間保管される、ことを特徴とする実施形態 61 などの上記実施形態の何れか 1 つに記載の方法。82．架橋されたパラフィン包埋組織サンプルは核酸を単離させる前に収集点から輸送される、ことを特徴とする実施形態 61 などの上記実施形態の何れか 1 つに記載の方法。83．架橋されたパラフィン包埋組織サンプルは無菌環境で集められる、ことを特徴とする実施形態 61 などの上記実施形態の何れか 1 つに記載の方法。84．架橋されたパラフィン包埋組織サンプルは核酸を単離させる前に非無菌環境に位置付けられる、ことを特徴とする実施形態 61 などの上記実施形態の何れか 1 つに記載の方法。85．組織サンプル解析の方法であって、該方法は：(a) 固定された組織サンプルを提供する工程；(b) 前記固定された組織サンプルの第 1 の部分及び第 2 の部分を集める工程であって、前記第 1 の部分及び第 2 の部分は前記固定された組織サンプルの異なる領域に由来する、工程；(c) 前記第 1 の部分から第 1 の架橋された DNA：タンパク質の複合体を、及び前記第 2 の部分から第 2 の架橋された DNA：タンパク質の複合体を回復する工程；(d) (i) 前記第 1 の架橋された DNA：タンパク質の複合体の DNA の第 1 の部分を前記第 1 の架橋された DNA：タンパク質の複合体の DNA の第 2 の部分にライゲートする工程であって、それにより第 1 のライゲートされた DNA を形成する、工程、及び (ii) 前記第 2 の架橋された DNA：タンパク質の複合体の DNA の第 2 の部分を前記第 2 の架橋された DNA：タンパク質の複合体の DNA の第 2 の部分にライゲートする工程であって、それにより第 2 のライゲートされた DNA を形成する、工程；(e) 前記第 1 の架橋された DNA：タンパク質の複合体から前記第 1 のライゲートされた DNA を、及び前記第 2 の架橋された DNA：タンパク質の複合体から前記第 2 のライゲートされた DNA を抽出する工程；(f) 前記第 1 のライゲートされた DNA 及び第 2 のライゲートされた DNA を配列決定する工程；及び (g) 第 1 のヌクレオチド配列と第 2 のヌクレオチド配列をアセンブルするために前記配列決定する工程からの情報を使用する工程を含む。86．前記固定された組織サンプルはホルマリン固定される、ことを特徴とする実施形態 85 などの上記実施形態の何れか 1 つに記載の方法。87．前記固定された組織は、ホルマリン固定パラフィン包埋 (FFPE) される、ことを特徴とする実施形態 85 などの上記実施形態の何れか 1 つに記載の方法。88．前記第 1 の架橋された DNA：タンパク質の複合体及び前記第 2 の架橋された DNA：タンパク質の複合体はクロマチンを含む、ことを特徴とする実施形態 85 などの上記実施形態の何れか 1 つに記載の方法。89．(d) (i) 及び (d) (ii) における前記ライゲートする工程は、平滑末端ライゲーションを含む、ことを特徴とする実施形態 85 などの上記実施形態の何れか 1 つに記載の方法。90．(d) (i) 及び (d) (ii) における前記ライゲートする工程の前に、前記第 1 の架橋された DNA：タンパク質の複合体、及び第 2 の架橋された DNA：タンパク質の複合体から、DNA を消化する工程を更に含む、実施形態 85 などの上記実施形態の何れか 1 つに記載の方法。91．前記消化する工程は制限酵素消化を含む、ことを特徴とする実施形態 90 などの上記実施形態の何れか 1 つに記載の方法。92．前記消化する工程の後に、平滑末端を産生するために前記消化する工程から接着末端を充填する工程を更に含む、請求項 90 などの上記実施形態の何れか 1 つに記載の方法。93．前記充填する工程はピオチン化ヌクレオチドを使用して行われる、ことを特徴とする実施形態 92 などの上記実施形態の何れか 1 つに記載の方法。94．前記回復する工程は、前記第 1 の架橋された DNA：タンパク質の複合体及び前記第 2 の架橋された DNA：タンパク質の複合体の DNA を、固形支持体に結合する工程を含む、ことを特徴とする実施形態 85 などの上記実施形態の何れか 1 つに記載の方法。95．前記抽出する工程は、前記第 1 の架橋された DNA：タンパク質の複合体及び前記第 2 の架橋された DNA：タンパク質の複合体からタンパク質を消化する工程を含む、ことを特徴とする実施形態 85 など

10

20

30

40

50

の上記実施形態の何れか1つに記載の方法。96．前記情報は、2000を超える塩基対（bp）の距離にわたる長距離情報を含む、ことを特徴とする実施形態85などの上記実施形態の何れか1つに記載の方法。97．前記距離は10000bpを超える、ことを特徴とする実施形態96などの上記実施形態の何れか1つに記載の方法。98．前記距離は100000bpを超える、ことを特徴とする実施形態97などの上記実施形態の何れか1つに記載の方法。99．前記距離は200000bpを超える、ことを特徴とする実施形態98などの上記実施形態の何れか1つに記載の方法。100．前記回復する工程の前に、前記固定された組織サンプルの包埋材料を溶かす工程を更に含む、実施形態85などの上記実施形態の何れか1つに記載の方法。101．前記包埋材料はパラフィンを含む、ことを特徴とする実施形態100などの上記実施形態の何れか1つに記載の方法。102．固定された組織サンプルは組織中のその配置を反映する位置情報を保存する、ことを特徴とする実施形態85などの上記実施形態の何れか1つに記載の方法。103．固定された組織サンプルは核酸を単離させる前に均質化されない、ことを特徴とする実施形態85などの上記実施形態の何れか1つに記載の方法。104．固定された組織サンプルは核酸を単離させる前に少なくとも1週間保管される、ことを特徴とする実施形態85などの上記実施形態の何れか1つに記載の方法。105．固定された組織サンプルは核酸を単離させる前に少なくとも6か月間保管される、ことを特徴とする実施形態85などの上記実施形態の何れか1つに記載の方法。106．固定された組織サンプルは核酸を単離させる前に収集点から輸送される、ことを特徴とする実施形態85などの上記実施形態の何れか1つに記載の方法。107．固定された組織サンプルは無菌環境で集められる、ことを特徴とする実施形態85などの上記実施形態の何れか1つに記載の方法。108．固定された組織サンプルは核酸を単離させる前に非無菌環境に位置付けられる、ことを特徴とする実施形態85などの上記実施形態の何れか1つに記載の方法。109．保存された組織サンプルからゲノム再編成を検出する方法であって、該方法は：タンパク質DNA複合体が破壊されないように、保存された組織サンプルからタンパク質DNA複合体を単離させる工程；少なくとも1つのペアエンドライゲーション生成物を形成するために複合体の晒されたDNA末端をライゲートする工程；1対のプロープに少なくとも1つのペアエンドライゲーション生成物を接触させる工程であって、プロープの対は、細胞型において再編成された第1の領域及び第2の領域に結合する、工程を含む。110．タンパク質DNA複合体は、第1のセグメント及び第2セグメントがリン酸ジエステル骨格から独立して共に保持されるように単離される、ことを特徴とする実施形態109などの上記実施形態の何れか1つに記載の方法。111．保存されたサンプルは架橋される、ことを特徴とする実施形態109などの上記実施形態の何れか1つに記載の方法。112．プロープの対は標識される、ことを特徴とする実施形態109などの上記実施形態の何れか1つに記載の方法。113．プロープの対はフルオロフォアを含む、ことを特徴とする実施形態109などの上記実施形態の何れか1つに記載の方法。114．プロープの対はオリゴヌクレオチドプロープをプロープ含む、ことを特徴とする実施形態109などの上記実施形態の何れか1つに記載の方法。115．共通のペアエンドライゲーション生成物上にオリゴ核酸の対をアニールするためのアッセイを行う工程を更に含む、実施形態110などの上記実施形態の何れか1つに記載の方法。116．単離された核酸の少なくとも幾つかを配列決定する工程を更に含む、実施形態115などの上記実施形態の何れか1つに記載の方法。117．プロープの対はフォワードプライマー及びリバースプライマーを含み、フォワードプライマー及びリバースプライマーのうち少なくとも1つは、再編成に関与するDNAセグメントにアニールする、ことを特徴とする実施形態109などの上記実施形態の何れか1つに記載の方法。118．フォワードプライマーとリバースプライマーを使用して核酸増幅を行なう工程を更に含む、実施形態117などの上記実施形態の何れか1つに記載の方法。118．単離された核酸の少なくとも幾つかを配列決定する工程を含む、実施形態118などの上記実施形態の何れか1つに記載の方法。120．ゲノム再編成は、逆位、挿入、欠失、及び転座から選択される、ことを特徴とする実施形態109などの上記実施形態の何れか1つに記載の方法。121．保存された組織サンプルはホルマリン固定される、

10

20

30

40

50

ことを特徴とする実施形態 109 などの上記実施形態の何れか 1 つに記載の方法。122 . 保存された組織は、ホルマリン固定パラフィン包埋 (FFPE) される、ことを特徴とする実施形態 109 などの上記実施形態の何れか 1 つに記載の方法。123 . 単離する工程の前に、固定された組織サンプルの包埋材料を除去する工程を更に含む、実施形態 109 乃至 122 などの上記実施形態の何れか 1 つに記載の方法。124 . 前記包埋材料はパラフィンを含む、ことを特徴とする実施形態 123 などの上記実施形態の何れか 1 つに記載の方法。125 . 単離する工程は、保存された組織サンプルをキシレンに接触させる工程を含む、ことを特徴とする実施形態 109 などの上記実施形態の何れか 1 つに記載の方法。126 . 単離する工程は、保存された組織サンプルをエタノールに接触させる工程を含む、ことを特徴とする実施形態 109 などの上記実施形態の何れか 1 つに記載の方法。127 . 単離する工程は、サンプルを沸騰状態から保護する工程を含む、ことを特徴とする実施形態 109 などの上記実施形態の何れか 1 つに記載の方法。128 . 単離する工程は、架橋された組織サンプルを、アントラニル酸塩及びホスファニル酸塩のうち少なくとも 1 つに接触させる工程を含む、ことを特徴とする実施形態 109 などの上記実施形態の何れか 1 つに記載の方法。129 . 単離させる工程は 40 以下の温度で行われる、ことを特徴とする実施形態 109 などの上記実施形態の何れか 1 つに記載の方法。130 . 前記架橋された DNA : タンパク質の複合体はクロマチンを含む、ことを特徴とする実施形態 109 などの上記実施形態の何れか 1 つに記載の方法。131 . 単離する工程は、架橋された DNA : タンパク質の複合体の DNA を固形支持体に結合する工程を含む、ことを特徴とする実施形態 109 などの上記実施形態の何れか 1 つに記載の方法。132 . DNA セグメントにおけるゲノム再編成を検出する方法であって、該方法は：DNA セグメントのためにゲノム遺伝子座相互作用の情報を得る工程；及びゲノム遺伝子座相互作用の情報の観察された分布を、ゲノム遺伝子座相互作用の情報の予期される分布と比較する工程を含む。133 . 観察された分布と予期される分布との差は、DNA セグメントの再編成を示す、ことを特徴とする実施形態 132 などの上記実施形態の何れか 1 つに記載の方法。134 . ゲノム遺伝子座相互作用の情報は、DNA セグメントのライゲートされた垂群のペアエンドのリードペア情報を含む、ことを特徴とする実施形態 132 などの上記実施形態の何れか 1 つに記載の方法。135 . ゲノム再編成は、逆位、挿入、欠失、及び転座から選択される、ことを特徴とする実施形態 132 などの上記実施形態の何れか 1 つに記載の方法。136 . 観察された分布の相互作用頻度は、予期される分布の相互作用頻度より大きく、ゲノム再編成は逆位を含む、ことを特徴とする実施形態 132 などの上記実施形態の何れか 1 つに記載の方法。137 . 観察された分布の相互作用頻度は、予期される分布の相互作用頻度未満であり、ゲノム再編成は欠失を含む、ことを特徴とする実施形態 132 などの上記実施形態の何れか 1 つに記載の方法。138 . DNA セグメントは、架橋された組織サンプルから得られる、ことを特徴とする実施形態 132 などの上記実施形態の何れか 1 つに記載の方法。139 . 架橋された組織サンプルはホルマリン固定される、ことを特徴とする実施形態 138 などの上記実施形態の何れか 1 つに記載の方法。140 . 架橋された組織サンプルは、ホルマリン固定パラフィン包埋 (FFPE) される、ことを特徴とする実施形態 138 などの上記実施形態の何れか 1 つに記載の方法。141 . 架橋された組織サンプルは、タンパク質 DNA 複合体が破壊されないように、架橋された組織サンプルから核酸を単離するために処理される、ことを特徴とする実施形態 138 などの上記実施形態の何れか 1 つに記載の方法。142 . タンパク質 DNA 複合体は、第 1 のセグメント及び第 2 セグメントがリン酸ジエステル骨格から独立して共に保持されるように単離される、ことを特徴とする実施形態 141 などの上記実施形態の何れか 1 つに記載の方法。143 . 処理の前に、固定された組織サンプルの包埋材料が溶かされる、ことを特徴とする実施形態 141 などの上記実施形態の何れか 1 つに記載の方法。144 . 包埋材料はパラフィンを含む、ことを特徴とする実施形態 142 などの上記実施形態の何れか 1 つに記載の方法。145 . 処置は、架橋されたパラフィン包埋組織サンプルをキシレンに接触させる工程を含む、ことを特徴とする実施形態 141 などの上記実施形態の何れか 1 つに記載の方法。146 . 処置は、架橋されたパラフィン包埋組織サンプルをエタノール

10

20

30

40

50

に接触させる工程を含む、ことを特徴とする実施形態 1 4 1 などの上記実施形態の何れか 1 つに記載の方法。1 4 7 . 処置は、サンプルを沸騰状態から保護する工程を含む、ことを特徴とする実施形態 1 4 1 などの上記実施形態の何れか 1 つに記載の方法。1 4 8 . 処置は、架橋された組織サンプルを、アントラニル酸塩及びホスファニル酸塩のうち少なくとも 1 つに接触させる工程を含む、ことを特徴とする実施形態 1 4 1 などの上記実施形態の何れか 1 つに記載の方法。1 4 9 . 処置は 4 0 以下の温度で行われる、ことを特徴とする実施形態 1 4 1 などの上記実施形態の何れか 1 つに記載の方法。1 5 0 . DNA タンパク質複合体はクロマチンを含む、ことを特徴とする実施形態 1 4 1 などの上記実施形態の何れか 1 つに記載の方法。1 5 1 . 共通の保存されたサンプルに由来する第 1 の DNA タンパク質複合体及び第 2 の DNA タンパク質複合体を含む組成物であって、ここで、第 1 の DNA タンパク質複合体は、タグ付けされた DNA セグメントを含み、それによりこのセグメントは共通の複合体から生じると同定され、及び、第 1 の DNA タンパク質複合体は共通の保存されたサンプルの第 1 の位置に割り当て可能であり、第 2 の DNA タンパク質複合体は共通の保存されたサンプルの第 2 の位置に割り当て可能である。1 5 2 . タグ付けされた DNA セグメントは、共通の複合体を示す配列を持つオリゴヌクレオチドを使用してタグ付けされる、ことを特徴とする実施形態 1 5 1 などの上記実施形態の何れか 1 つに記載の組成物。1 5 3 . タグ付けされた DNA セグメントは、ライゲーション結合の何れかの側での固有の配列が共通の複合体に割り当てられるように、ペアエンドを形成するためのライゲーションによりタグ付けされる、ことを特徴とする実施形態 1 5 1 などの上記実施形態の何れか 1 つに記載の組成物。1 5 4 . 共通の保存されたサンプルは架橋剤に接触させられる、ことを特徴とする実施形態 1 5 1 などの上記実施形態の何れか 1 つに記載の組成物。1 5 5 . 架橋剤はホルムアルデヒド又はホルマリンのうち少なくとも 1 つを含む、ことを特徴とする実施形態 1 5 1 などの上記実施形態の何れか 1 つに記載の組成物。1 5 6 . 橋架剤は、UV 光、マイトマイシン C、ナイトロジェンマスタード、メルファラン、1, 3 - ブタジエンジエポキシド、シスジアミンジクロロプラチナム (II)、及びシクロホスファミドのうち少なくとも 1 つを含む、ことを特徴とする実施形態 1 5 1 などの上記実施形態の何れか 1 つに記載の組成物。1 5 7 . 保存されたサンプルは、ホルマリン固定パラフィン包埋 (FFPE) される、ことを特徴とする実施形態 1 5 1 などの上記実施形態の何れか 1 つに記載の組成物。1 5 8 . 保存された組織サンプルは、タンパク質 DNA 複合体が破壊されないように、保存された組織サンプルから核酸を単離するために処理される、ことを特徴とする実施形態 1 5 1 などの上記実施形態の何れか 1 つに記載の組成物。1 5 9 . タンパク質 DNA 複合体は、第 1 のセグメント及び第 2 セグメントがリン酸ジエステル骨格から独立して共に保持されるように単離される、ことを特徴とする実施形態 1 5 8 などの上記実施形態の何れか 1 つに記載の方法。1 6 0 . 処理の前に、保存された組織サンプルの包埋材料を溶かす工程を更に含む、実施形態 1 5 8 などの上記実施形態の何れか 1 つに記載の組成物。1 6 1 . 包埋材料はパラフィンを含む、ことを特徴とする実施形態 1 5 9 などの上記実施形態の何れか 1 つに記載の組成物。1 6 2 . 処置は、架橋されたパラフィン包埋組織サンプルをキシレンに接触させる工程を含む、ことを特徴とする実施形態 1 5 1 などの上記実施形態の何れか 1 つに記載の組成物。1 6 3 . 処置は、架橋されたパラフィン包埋組織サンプルをエタノールに接触させる工程を含む、ことを特徴とする実施形態 1 5 1 などの上記実施形態の何れか 1 つに記載の組成物。1 6 4 . 処置は、サンプルを沸騰状態から保護する工程を含む、ことを特徴とする実施形態 1 5 1 などの上記実施形態の何れか 1 つに記載の組成物。1 6 5 . 処置は、架橋された組織サンプルを、アントラニル酸塩及びホスファニル酸塩のうち少なくとも 1 つに接触させる工程を含む、ことを特徴とする実施形態 1 5 1 などの上記実施形態の何れか 1 つに記載の組成物。1 6 6 . 処置は 4 0 以下の温度で行われる、ことを特徴とする実施形態 1 5 1 などの上記実施形態の何れか 1 つに記載の組成物。1 6 7 . 第 1 の DNA タンパク質複合体又は第 2 の DNA タンパク質複合体はクロマチンを含む、ことを特徴とする実施形態 1 5 1 などの上記実施形態の何れか 1 つに記載の組成物。1 6 8 . 被験体から、核酸を含む保存されたサンプルを得る工程；及びサンプル中の核酸を分

10

20

30

40

50

析することによりゲノムの構造情報を引き出す工程を含む、方法。169．保存されたサンプルは架橋される、ことを特徴とする実施形態168などの上記実施形態の何れか1つに記載の方法。170．保存されたサンプルは、ホルムアルデヒド、ホルマリン、UV光、マイトマイシンC、ナイトロジェンマスタード、メルファラン、1,3-ブタジエンジエポキシド、シスジアミンジクロロプラチナム(II)、及びシクロホスファミドのうち少なくとも1つを使用して架橋される、ことを特徴とする実施形態169などの上記実施形態の何れか1つに記載の方法。171．保存されたサンプルはホルマリンを使用して架橋される、ことを特徴とする実施形態169などの上記実施形態の何れか1つに記載の方法。172．保存されたサンプルは、その中の核酸に関する位置情報を維持する、ことを特徴とする実施形態168などの上記実施形態の何れか1つに記載の方法。173．保存されたサンプルは包埋サンプルである、ことを特徴とする実施形態168などの上記実施形態の何れか1つに記載の方法。174．保存されたサンプルは、ホルマリン固定パラフィン包埋(FFPE)サンプルである、ことを特徴とする実施形態168などの上記実施形態の何れか1つに記載の方法。175．ゲノムの構造情報は、基準ゲノムに対する、逆位、挿入、欠失、及び転座を示す、ことを特徴とする実施形態168などの上記実施形態の何れか1つに記載の方法。176．基準ゲノムは、被験体に共通する種の野生型ゲノムである、ことを特徴とする実施形態175などの上記実施形態の何れか1つに記載の方法。177．基準ゲノムは、被験体の基準組織から得られる、ことを特徴とする実施形態175などの上記実施形態の何れか1つに記載の方法。178．核酸の第1のセグメント及び第2のセグメントに関するフェーズ状況を示す情報を引き出す工程を含む、実施形態168乃至177などの上記実施形態の何れか1つに記載の方法。179．物理的連鎖情報を伝えるようにサンプルの暴露された核酸末端にタグ付けする工程を含む、ことを特徴とする実施形態168などの上記実施形態の何れか1つに記載の方法。180．タグを付ける工程は、オリゴヌクレオチドが共通の複合体を示す情報を伝えるように、保存されたサンプルから放たれたDNAタンパク質複合体にオリゴヌクレオチドをライゲートする工程を含む、ことを特徴とする実施形態179などの上記実施形態の何れか1つに記載の方法。181．オリゴヌクレオチドは複合体に特異的な塩基配列を含む、ことを特徴とする実施形態180などの上記実施形態の何れか1つに記載の方法。182．オリゴヌクレオチドは複合体に固有の塩基配列を含む、ことを特徴とする実施形態180などの上記実施形態の何れか1つに記載の方法。183．タグ付けする工程は、ペアエンド分子を形成するために複合体の第2のセグメントに複合体の第1の核酸セグメントをライゲートする工程を含む、ことを特徴とする実施形態179などの上記実施形態の何れか1つに記載の方法。184．第1の核酸セグメントの一部及び第2の核酸セグメントの一部を配列決定する工程を含む、実施形態183などの上記実施形態の何れか1つに記載の方法。185．第1の核酸セグメントの一部に共通するユニーク配列を有するコンティグ、及びに第2の核酸セグメントの一部に共通するユニーク配列を有するコンティグを、核酸アセンブリにおける共通の足場に割り当てる工程を含む、実施形態184などの上記実施形態の何れか1つに記載の方法。186．ペアエンド核酸分子を核酸分子プローブのセットに接触させる工程を含む、実施形態183などの上記実施形態の何れか1つに記載の方法。187．核酸プローブのセットは蛍光プローブである、ことを特徴とする実施形態186などの上記実施形態の何れか1つに記載の方法。188．核酸プローブのセットは、ゲノム構造再編成に参与する第1の遺伝子座及び第2の遺伝子座にアニールする、ことを特徴とする実施形態186などの上記実施形態の何れか1つに記載の方法。189．第1の遺伝子座及び第2の遺伝子座は、ゲノム構造再編成に影響されないゲノムにおいて隣接しない、ことを特徴とする実施形態188などの上記実施形態の何れか1つに記載の方法。190．第1の遺伝子座及び第2の遺伝子座は、ゲノム構造再編成に影響されないゲノムにおいて隣接する、ことを特徴とする実施形態188などの上記実施形態の何れか1つに記載の方法。191．1セットの核酸プローブの接触が再編成を示す時にサンプルの核酸を配列決定する工程を含む、実施形態186などの上記実施形態の何れか1つに記載の方法。192．ペアエンド核酸分子を核酸分子プライマーのセットに接触させる工程を含む、実施形態1

10

20

30

40

50

83 などの上記実施形態の何れか1つに記載の方法。193. 核酸プライマーのセットは、ゲノム構造再編成に關与する第1の遺伝子座及び第2の遺伝子座にアニールする、ことを特徴とする実施形態192などの上記実施形態の何れか1つに記載の方法。194. 第1の遺伝子座及び第2の遺伝子座がライゲートされたペアエンド分子を形成すると、1セットの核酸プライマーは核酸増幅反応においてアンプリコンをもたらす、ことを特徴とする実施形態193などの上記実施形態の何れか1つに記載の方法。195. 第1の遺伝子座及び第2の遺伝子座がライゲートされたペアエンド分子を形成しないと、1セットの核酸プライマーは核酸増幅反応においてアンプリコンをもたらさない、ことを特徴とする実施形態193などの上記実施形態の何れか1つに記載の方法。196. 第1の遺伝子座及び第2の遺伝子座は、ゲノム構造再編成に影響されないゲノムにおいて隣接しない、ことを特徴とする実施形態188などの上記実施形態の何れか1つに記載の方法。197. 第1の遺伝子座及び第2の遺伝子座は、ゲノム構造再編成に影響されないゲノムにおいて隣接する、ことを特徴とする実施形態188などの上記実施形態の何れか1つに記載の方法。198. ペアエンド核酸分子に接触される1セットの核酸プライマーのセットからアンプリコンが生成される時に、サンプルの核酸を配列決定する工程を含む、実施形態192などの上記実施形態の何れか1つに記載の方法。199. 保存された組織サンプルは、タンパク質DNA複合体が破壊されないように、核酸を単離するために処理される、ことを特徴とする実施形態169などの上記実施形態の何れか1つに記載の方法。200. タンパク質DNA複合体は、第1のセグメント及び第2セグメントがリン酸ジエステル骨格から独立して共に保持されるように単離される、ことを特徴とする実施形態199などの上記実施形態の何れか1つに記載の方法。201. 保存された組織サンプルは、保存された組織サンプルをキシレンに接触させることにより処理される、ことを特徴とする実施形態199などの上記実施形態の何れか1つに記載の方法。202. 保存された組織サンプルは、保存された組織サンプルをエタノールに接触させることにより処理される、ことを特徴とする実施形態199に記載の方法。203. 保存された組織サンプルは、沸騰状態からサンプルを保護することにより処理される、ことを特徴とする実施形態199などの上記実施形態の何れか1つに記載の方法。204. 保存された組織サンプルは、保存された組織サンプルをアントラニル酸塩とホスファニル酸塩のうち少なくとも1つに接触させることにより処理される、ことを特徴とする実施形態199などの上記実施形態の何れか1つに記載の方法。205. 保存された組織サンプルは40 以下の温度で処理される、ことを特徴とする実施形態199などの上記実施形態の何れか1つに記載の方法。206. DNAタンパク質複合体はクロマチンを含む、ことを特徴とする実施形態199などの上記実施形態の何れか1つに記載の方法。207. 保存された組織サンプルは組織中のその配置を反映する位置情報を保存する、ことを特徴とする実施形態168などの上記実施形態の何れか1つに記載の方法。208. 保存された組織サンプルは核酸を単離させる前に均質化されない、ことを特徴とする実施形態168などの上記実施形態の何れか1つに記載の方法。209. 保存された組織サンプルは核酸を単離させる前に少なくとも1週間保管される、ことを特徴とする実施形態168などの上記実施形態の何れか1つに記載の方法。210. 保存された組織サンプルは核酸を単離させる前に少なくとも6か月間保管される、ことを特徴とする実施形態168などの上記実施形態の何れか1つに記載の方法。211. 保存された組織サンプルは核酸を単離させる前に収集点から輸送される、ことを特徴とする実施形態168などの上記実施形態の何れか1つに記載の方法。205. 保存された組織サンプルは無菌環境で集められる、ことを特徴とする実施形態168などの上記実施形態の何れか1つに記載の方法。213. 保存された組織サンプルは核酸を単離させる前に非無菌環境に位置付けられる、ことを特徴とする実施形態168などの上記実施形態の何れか1つに記載の方法。214. 保存されたサンプルからゲノムの構造情報を得るためのキットであって、該キットは：緩衝液、DNA結合剤、アフィニティータグ薬剤、デオキシリボヌクレオチド、タグ付けされたデオキシリボヌクレオチド、DNA断片化剤、末端修復酵素、リガーゼ、タンパク質除去剤、及び保存されたサンプルからゲノムの構造情報を得る際の使用説明書を含む。215. PCRのための試薬を更に含む、実施形

10

20

30

40

50

態 2 1 4 などの上記実施形態の何れか 1 つに記載のキット。2 1 6 . P C R のための試薬は、緩衝液、ヌクレオチド、フォワードプライマー、リバースプライマー、及び熱安定性 D N A ポリメラーゼを含む、ことを特徴とする実施形態 2 1 5 などの上記実施形態の何れか 1 つに記載のキット。2 1 7 . 緩衝液は、制限消化緩衝液、末端修復緩衝液、ライゲーション緩衝液、T E 緩衝液、洗浄緩衝液、T W B 溶液、N T B 溶液、L W B 溶液、N W B 溶液、及び架橋逆転緩衝液のうち少なくとも 1 つを含む、ことを特徴とする実施形態 2 1 4 などの上記実施形態の何れか 1 つに記載のキット。2 1 8 . 制限消化緩衝液は D p n I I 緩衝液を含む、ことを特徴とする実施形態 2 1 7 などの上記実施形態の何れか 1 つに記載のキット。2 1 9 . 末端修復緩衝液は N E B 緩衝液 2 を含む、ことを特徴とする実施形態 2 1 7 などの上記実施形態の何れか 1 つに記載のキット。2 2 0 . ライゲーション緩衝液は、T 4 D N A リガーゼ緩衝液、B S A、及び T r i t o n X - 1 0 0 を含む、ことを特徴とする実施形態 2 1 7 などの上記実施形態の何れか 1 つに記載のキット。2 2 1 . T E 緩衝液はトリス及び E D T A を含む、ことを特徴とする実施形態 2 1 7 などの上記実施形態の何れか 1 つに記載のキット。2 2 2 . 洗浄緩衝液はトリス及び塩化ナトリウムを含む、ことを特徴とする実施形態 2 1 7 などの上記実施形態の何れか 1 つに記載のキット。2 2 3 . T W B 溶液はトリス、E D T A、及び T w e e n 2 0 を含む、ことを特徴とする実施形態 2 1 7 などの上記実施形態の何れか 1 つに記載のキット。2 2 4 . N T B 溶液はトリス、E D T A、及び塩化ナトリウムを含む、ことを特徴とする実施形態 2 1 7 などの上記実施形態の何れか 1 つに記載のキット。2 2 5 . L W B 溶液は、トリス、塩化リチウム、E D T A、及び T w e e n 2 0 を含む、ことを特徴とする実施形態 2 1 7 などの上記実施形態の何れか 1 つに記載のキット。2 2 6 . N W B 溶液は、トリス、塩化ナトリウム、E D T A、及び T w e e n 2 0 を含む、ことを特徴とする実施形態 2 1 7 などの上記実施形態の何れか 1 つに記載のキット。2 2 7 . 架橋逆転緩衝液は、トリス、S D S、及び塩化カルシウムを含む、ことを特徴とする実施形態 2 1 7 などの上記実施形態の何れか 1 つに記載のキット。2 2 8 . D N A 結合剤は、クロマチンキャプチャビーズを含む、ことを特徴とする実施形態 2 1 4 などの上記実施形態の何れか 1 つに記載のキット。2 2 9 . クロマチンキャプチャビーズは、P E G - 8 0 0 粉末、トリス緩衝液、塩化ナトリウム、E D T A、界面活性剤、T E 緩衝液、及び s e r a - m a g ビーズを含む、ことを特徴とする実施形態 2 2 8 などの上記実施形態の何れか 1 つに記載のキット。2 3 0 . アフィニティータグ結合剤は、ストレプトアビジンビーズを含む、ことを特徴とする実施形態 2 1 4 などの上記実施形態の何れか 1 つに記載のキット。2 3 1 . ストレプトアビジンビーズは d y n a b e a d s を含む、ことを特徴とする実施形態 2 3 0 などの上記実施形態の何れか 1 つに記載のキット。2 3 2 . デオキシリボヌクレオチドは、d A T P、d T T P、d G T P、及び d C T P のうち少なくとも 3 つを含む、ことを特徴とする実施形態 2 1 4 などの上記実施形態の何れか 1 つに記載のキット。2 3 3 . ピオチン化デオキシリボヌクレオチドは、ピオチン化 d C T P、ピオチン化 d A T P、ピオチン化 d T T P、及びピオチン化 d G T P のうち少なくとも 1 つを含む、ことを特徴とする実施形態 2 1 4 などの上記実施形態の何れか 1 つに記載のキット。2 3 4 . D N A 断片化剤は、制限酵素、トランスポサーゼ、ヌクレアーゼ、音波処理デバイス、流体力学的剪断デバイス、及び二価金属カチオンのうち少なくとも 1 つである、ことを特徴とする実施形態 2 1 4 などの上記実施形態の何れか 1 つに記載のキット。2 3 5 . 制限酵素は D p n I I を含む、ことを特徴とする実施形態 2 3 4 などの上記実施形態の何れか 1 つに記載のキット。2 3 6 . 末端修復酵素は、T 4 D N A ポリメラーゼ、クレノウ D N A ポリメラーゼ、及び T 4 ポリヌクレオチドキナーゼのうち少なくとも 1 つを含む、ことを特徴とする実施形態 2 1 4 などの上記実施形態の何れか 1 つに記載のキット。2 3 7 . リガーゼは T 4 D N A リガーゼを含む、ことを特徴とする実施形態 2 1 4 などの上記実施形態の何れか 1 つに記載のキット。2 3 8 . タンパク質除去剤は、プロテアーゼ及びフェノールのうち少なくとも 1 つを含む、ことを特徴とする実施形態 2 1 4 などの上記実施形態の何れか 1 つに記載のキット。2 3 9 . プロテアーゼは、プロテイナーゼ K、ストレプトマイセス・グリセウスプロテアーゼ、セリンプロテアーゼ、システインプロテアーゼ、トレ

10

20

30

40

50

オニンプロテアーゼ、アスパラギン酸プロテアーゼ、グルタミン酸プロテアーゼ、メタロプロテアーゼ、及びアスパラギンペプチドリアーゼのうち少なくとも1つを含む、ことを特徴とする実施形態238などの上記実施形態の何れか1つに記載のキット。240．包埋材料を除去するための溶媒を更に含む、実施形態214などの上記実施形態の何れか1つに記載のキット。241．溶媒は、キシレン、ベンゼン、及びトルエンのうち少なくとも1つを含む、ことを特徴とする実施形態240などの上記実施形態の何れか1つに記載のキット。

【0240】

以下の実施例は、本開示を例示するが、これを限定しないように意図されている。それらは使用され得るものの典型である一方で、当業者に既知の他の手順が代替的に使用されてもよい。

【実施例】

【0241】

実施例1．FFPEサンプルからのリードペアライブラリの生成

AJ GIAB (「Genome In A Bottle」) サンプルGM24149 (父) 及びGM24385 (子) を、Horizon Discovery から獲得した。細胞株は以前にFFPEに埋め込まれている。1つの切片当たり約 3×10^5 の細胞を含む、厚さおよそ15 - 20ミクロンの切片を、この実験に使用した。切片をキシレンで洗浄し、パラフィンワックスを除去した。エタノールで切片を洗浄することによりキシレンを除去した。その後、放たれた組織サンプルを洗浄性緩衝液の中で再懸濁した。その後、核酸を含むサンプルを末端ライゲーションに晒し、末端ライゲーションは、制限酵素、この実施例においてMboIでDNAを消化する工程、次いでビオチン化ヌクレオチドで結果として生じるオーバーハングを充填する工程を含む。平滑末端を共にライゲートし、その後、ライゲートされた末端を放出した。ビオチン化断片を得て、末端配列決定を行い、リードペアを得ることで、それぞれマッピングされたコンティグがサンプル中の共通の核酸分子に物理的に結合されたことを示した。

【0242】

単離された配列の位置をゲノムアセンブリと比較することにより回復された断片のペアエンド間の距離を決定するために、配列決定を行った。その結果、FFPE - Chicago法 (表1 - GIABの列) は結果として、非FFPEサンプル上で行ったChicago法 (表1 - Chicagoの列) に匹敵する ($> 200 \text{ kbp}$ の挿入) 又はそれよりも大きな ($100 \text{ kbp} - 200 \text{ kbp}$ の挿入) 長距離リードペアの頻度をもたらすことを明らかにした。これらデータも解析し、FFPE - Chicagoライブラリの複雑性及び生の配列決定カバレッジを決定した (表2)。ライブラリの複雑性は、ライブラリ内の異なる分子の多様性を指す。

【0243】

【表1】

表1．挿入長さの頻度

	GIAB	CHICAGO	CHICAGO
0 < 挿入 ≤ 2kbp	48.078%	20.731%	9.92%
2kbp < 挿入 ≤ 10kbp	0.458%	6.045%	1.811%
10 kbp < 挿入 ≤ 100kbp	0.553%	5.356%	1.884%
100kbp < 挿入 ≤ 200kbp	0.171%	0.022%	0.044%
200kbp < 挿入	1.49%	1.828%	1.499%

10

20

30

40

50

【 0 2 4 4 】

【表 2】

表 2. 複雑性及び生の物理カバレッジ

	FFPE	CHICAGO
ライブラリの複雑性 (Poisson)	229,196,982	1,013,303,912
150M のリードペアにスケールを合わせた生の物理カバレッジ	5.622 X	66.343 X

10

【 0 2 4 5 】

実施例 2 . F F P E - C h i c a g o ライブラリからのフェーズ決定

実施例 1 で生成され配列決定データを使用して、出発 G I A B サンプル中にあると知られる S N P のセットのフェージング情報を決定した。要するに、配列決定データを使用して、S N P のセットが同じ又は異なる D N A 分子に存在したかどうかを判定した。その後、フェーズ要求 (p h a s e - c a l l i n g) の正確性を決定するために、これらのデータを G I A B サンプルの既知の配列と比較した。

【 0 2 4 6 】

20

表 3 におけるビンの各々は、見出された S N P の数、及び、どれが次のビンのサイズまで一致しているかを示している。例えば、第 1 の線は、0 - 1 0 , 0 0 0 の間で、1 3 2 , 7 9 6 の S N P S が見出され、正確なフェーズには 9 9 . 0 5 9 % があったことを示している。高い一致 (> 9 5 %) は、約 1 . 5 M B まで見られる (1 3 のうち 1 つが見逃される 7 0 - 8 0 k b のビン、及び 1 5 のうち 2 つが見逃される 1 . 1 - 1 . 3 M B のビンを除く) 。 1 . 7 - 1 . 9 M B の範囲において、7 つの S N P ペアフェーズのうちの 7 つが適切に要求された。

【 0 2 4 7 】

これらのデータから、低レベルの偽性の連鎖にもかかわらず、適切な長距離の情報は、最大でメガベースまで F F P E - C h i c a g o 方法を使用して決定されると、結論付けられる。重要なことに、これらの「一致」予測割合は、偶然から予測する 5 0 % の成功率よりも著しく高く、9 5 % 以上である。

30

【 0 2 4 8 】

40

50

【表 3】

表 3. 各ビンの SNP

ビン	一致	一致した n	不一致の n	合計のリードペア
0	99.059	131547	1249	132796
10000	99.346	152	1	153
20000	100	60	0	60
30000	97.619	41	1	42
40000	97.222	35	1	36
50000	100	26	0	26
60000	100	26	0	26
70000	92.308	12	1	13
80000	100	18	0	18
90000	100	8	0	8
100000	98.148	159	3	162
300000	95.238	80	4	84
500000	98	49	1	50
700000	100	28	0	28
900000	96.552	28	1	29
1100000	86.667	13	2	15
1300000	100	16	0	16
1500000	78.571	11	3	14
1700000	100	7	0	7
1900000	85.714	6	1	7
2000000	87.097	27	4	31
3000000	72.222	26	10	36
4000000	84	21	4	25
5000000	69.565	16	7	23
6000000	52.941	9	8	17
7000000	77.778	7	2	9
8000000	61.111	11	7	18
10000000	64.183	267	149	416

【 0 2 4 9 】

実施例 3 . D N A 抽出の改善

洗浄性緩衝液を、S D S 含有緩衝液から、トリトン X 含有緩衝液に変更し、実施例 1 に記載される小球の視覚化の結果、D N A 抽出の増加がもたらされた。その後のライブラリ解析により、実施例 1 と 2 に記載されるライブラリと比較した時、このライブラリは複雑性を増加させつつ、高レベルのロングリードを維持することが明らかになった。結果を表 4 に示す。

【 0 2 5 0 】

ヒトサンプル 1 のデータを、実施例 1 に記載されるように処理された G I A B サンプル

10

20

30

40

50

から集めた（FFPEサンプル上で行なわれる平滑末端ライゲーション）。サンプルのDNAを全て、ライブラリ調製に使用した。

【0251】

ヒトサンプル2のデータを、実施例1に記載されるように処理された別のGIABサンプルから集めた（FFPEサンプル上で行なわれる平滑末端ライゲーション）。サンプルのDNAを全て、ライブラリ調製に使用した。

【0252】

ヒトサンプル3のデータを、実施例1に記載されるように処理された第3のGIABサンプルから集めた（FFPEサンプル上で行なわれる平滑末端ライゲーション）。サンプルのおよそ500 ngのDNAを、ライブラリ調製に使用した。

10

【0253】

ヒトサンプル4のデータを、実施例1に記載されるように処理された第3のGIABサンプル（ヒトサンプル3と同じサンプル）から集めた（FFPEサンプル上で行なわれる平滑末端ライゲーション）。サンプルのおよそ50 ngのDNAを、ライブラリ調製に使用した。

【0254】

ヒトサンプル5のデータを、実施例1に記載されるように処理された第3のGIABサンプル（ヒトサンプル3及び4と同じサンプル）から集めた（FFPEサンプル上で行なわれる平滑末端ライゲーション4）。サンプルのおよそ10 ngのDNAを、ライブラリ調製に使用した。

20

【0255】

30

40

50

【表 4】

表 4. DNA抽出が改善された結果

計画	ヒト (1)	ヒト (2)	ヒト (3)	ヒト (4)	ヒト (5)
ライブラリ ID	DPH593_ chicago_ miseq	DPH594_ chicago_ miseq	DPH595_ chicago_ miseq	DPH596_ chicago_ miseq	DPH597_ chicago_ miseq
PCR/光学複製	0.166%	0.17%	0.179%	0.546%	1.717%
マッピングされていない	8.157%	8.364%	8.263%	8.559%	8.358%
低いマップ品質	10.12%	10.134%	9.99%	9.809%	9.628%
異なる足場	16.481%	16.374%	13.779%	10.576%	10.557%
0 < 挿入 <= 2 kbp	57.001%	56.383%	60.844%	65.109%	64.924%
2 kbp < 挿入 <= 10 kbp	1.661%	1.794%	1.456%	1.154%	1.001%
10 kbp < 挿入 <= 100 kbp	1.438%	1.57%	1.245%	0.979%	0.859%
100 kbp < 挿入 <= 200 kbp	0.44%	0.476%	0.382%	0.303%	0.266%
200 kbp < 挿入	4.536%	4.735%	3.861%	2.965%	2.69%
ライブラリの複雑性 (Poisson)	1,295,157,213	1,144,409,461	1,321,625,959	497,115,139	107,132,825
150M のリードペアにスケールを合わせた生の物理カバレッジ	15.426 X	16.808 X	13.372 X	10.616 X	9.447 X

10

20

30

【 0 2 5 6 】

実施例 4 . F F P E サンプルからの DNA 抽出の失敗

B A 腫瘍サンプルは、癌患者から生検が行われ、パラフィンに包埋する前にホルマリンで固定された。その後、F F P E サンプルを保管する。6 か月後、患者は、腫瘍進行を追跡しつつ新たな化合物で処置することを目的とした臨床研究に入る。処置中に、F F P E 腫瘍生検サンプルを数週ごとに調製し、保管した。患者は処置に非常によく反応し、臨床チームは、患者の特異的な癌亜型についてより多く研究することに興味を抱いている。研究の各段階で腫瘍に存在する構造変異を判定するために、臨床チームは F F P E 腫瘍サンプルから DNA を抽出することを試みた。不運にも、回復された DNA は高度に断片化され、短い断片リードのみが回復される。これら短い断片リードは構造変異を判定するには不適切であり、そのため、重大な臨床情報が失われる。

40

【 0 2 5 7 】

実施例 5 . F F P E サンプル中のネイティブクロマチンからの長距離データの成功

実施例 4 の F F P E 腫瘍サンプルを、ネイティブ DNA タンパク質複合体を保存するために緩やかな方法で処理する。DNA 抽出を、キシレンでパラフィンワックスを除去するために F F P E サンプルを洗浄することによって行う。エタノールで洗浄することによりキシレンを除去した。その後、サンプルを、Hi - C 処理を受ける前に洗浄性緩衝液の中で再懸濁する。F F P R サンプルから単離された固定 DNA タンパク質複合体を消化して、ビオチン標識したヌクレオチドで充填される粘着性のオーバーハングを生成する。結果として生じる平滑末端を共にライゲートして、同じ DNA タンパク質複合体から生じる D

50

N A 配列のペアエンドを生成する。ペアエンドは、D N A 剪断によってD N A タンパク質複合体から放たれ、ストレプトアビジンビーズを使用して単離される。回復されたペアエンドを配列決定アダプターにライゲートし、配列決定して、リードペアライブラリを生成する。

【 0 2 5 8 】

臨床チームは、研究の 6 か月前に得られるサンプルを含む、患者の腫瘍の経時的な構造変異を判定するために、リードペアライブラリを解析することができる。これらのデータを使用して、癌の亜型を判定し、且つ同じ癌亜型を持つ他の患者の処置の予後を通知する。

【 0 2 5 9 】

実施例 6 . F F P E サンプルから再構成クロマチンからの長距離データの成功

10

実施例 5 に記載されるようにD N A をF F P E サンプルから抽出する。ネイキッドD N A を単離し、長さ5 0 k b 以上の断片のためにサイズ選択する。再構成されたクロマチンを、サイズ選択されたD N A を精製されたクロマチンタンパク質に結合することにより生成し、その結果、各D N A タンパク質複合体は単一のD N A 分子を含む。その後、これらD N A タンパク質はホルムアルデヒドを使用して架橋される。その後、架橋された複合体を消化し、処理して、同じD N A 分子から生じるD N A 配列からペアエンドを生成する。ペアエンドを配列決定し、リードペアライブラリを生成する。リードペアライブラリからのデータにより、フェージング及び構造変異の情報を判定するために使用される長距離配列情報は、上述の患者の腫瘍サンプルを特徴付けるのに有用であることが明らかになる。

【 0 2 6 0 】

20

実施例 7 : F F P E サンプルからのゲノム異種性の判定

実施例 4 のF F P E サンプルを研究に使用して、腫瘍の異なる領域のゲノム異種性を判定する。パンチ生検をF F P E 腫瘍サンプルの異なるセグメントから得て、次いで実施例 5 に記載されるように処理する。生成されたデータを使用して、腫瘍の成長している縁を判定し、且つ、実施例 5 に記載される新たな化合物での処置により腫瘍増殖又は退行中に、突然変異と構造変動がどのように進行し、及び蓄積し又は消失するのかを学習する。

【 0 2 6 1 】

実施例 8 : F F P E の可溶化及びサンプル溶解

1 ミリリットルのキシレンをF F P E サンプルに加え、パラフィンが溶けるまでボルテックス処理する。サンプルを 2 分間、1 分につき 1 4 , 0 0 0 の回転数で遠心分離する。キシレンを優しく除去する。1 ミリリットルの 1 0 0 % エタノールを加え、サンプルをボルテックス処理して、チューブの内壁から細胞小球を分離する。サンプルを 2 分間、最大速度で再び遠心分離し、次にエタノールを除去する。小球を大気乾燥する。一旦小球が完全に乾燥すると、5 0 マイクロリットルの溶解緩衝液 (5 0 m M のトリス p H 8、5 0 m M のN a C l、1 % のS D S、0 . 1 5 % のトリトン、1 m M のE D T A) をサンプルに加える。サンプルを軽く振りながら、3 7 ° C で1 5 分間インキュベートする。その後、サンプル全体を1 . 5 m L のチューブに移す。サンプルを繰り返しピペットで移し、細胞小球を崩壊させる。その後、サンプルに1 0 0 μ L のS P R I (固相可逆的固定化) ビーズを、2 : 1 の比率のS P R I ビーズ対可溶性クロマチンの中で加えて、その後、室温で1 0 分間インキュベーションを行う。その後、S P R I ビーズを2 回洗浄する。その後、S P R I ビーズを単離したサンプルを、C h i c a g o 又はH i - C などの下流技術のために使用する。

30

40

【 0 2 6 2 】

実施例 9 : F F P E サンプルは長距離ゲノム連鎖情報を保存する。

F F P E サンプルを得て、本開示の方法に従い処理して、ゲノム連鎖データを抽出した。図 1 1 A は、3 つのサンプルの解析の結果を示す。ヒト細胞培養物 (赤色、1 1 0 3) 及び脾臓組織 (緑色、1 1 0 2) F F P E サンプルを得て、本開示の方法に従い処理して、ゲノム連鎖データを抽出した。ペアエンドをh g 1 9 基準にマッピングし、各リードペアのリード間の物理的距離を計算した。これらのデータを、H i - C 方法を用いた細胞培養サンプル (青色、1 1 0 1) を使用して調製されたデータと比較した。X 軸は、リード

50

間で物理的距離 (m b) によりピンングされたリードペアを示す (軸数は左から右まで 0 . 0 1、0 . 1、1、1 0、及び 1 0 0 である)。Y 軸は、リードペアの分画を示す (軸数は上部から下部まで 0 . 0 1、0 0 0 1、 $1 0^{-4}$ 、 $1 0^{-5}$ 、 $1 0^{-6}$ 、 $1 0^{-7}$ 、 $1 0^{-8}$ 、 $1 0^{-9}$ 、 $1 0^{-10}$ 、 $1 0^{-11}$ 、及び $1 0^{-12}$ である)。

【 0 2 6 3 】

実施例 1 0 : 長距離ゲノム連鎖情報を抽出するために処理された F F P E サンプルにおける S N P の一致

図 1 1 B は、本開示の方法に従い長距離ゲノム連鎖データを生成するために、A s h k e n a z i f a t h e r (G M 2 4 1 4 9) 細胞培養 F F P E サンプルの解析の結果が処理されたことを示す。これらのデータを、両方のペアエンドリードに存在する高信頼 S N P のためにフィルタ処理した。このフィルタ処理したデータセットを、2 つのリード間の物理的距離に基づいてピンに組織化し (X 軸)、一致における S N P ペアの割合を各ピンについて計算した (Y 軸)。上部の赤線 (1 1 1 1) は一致した S N P を示し、下部の青線 (1 1 1 2) は基準に関する無作為の一致を示す。

【 0 2 6 4 】

実施例 1 1 : F F P E サンプルは、構造変異 (S V) 識別を可能にする、長距離ゲノム連鎖情報を保存する。

データはまた、A s h k e n a z i f a t h e r (G M 2 4 1 4 9) 細胞培養 F F P E サンプルから抽出され、h g 1 9 基準に対してリードペアをマッピングすることにより構造変異の存在について解析された。ペアになったリードの中点を、図 1 1 C 及び図 1 1 D の X 軸上でプロットし、対応する物理的分離を Y 軸上でプロットした。マップ品質スコアを、説明文に示されるように各データポイントのグレイスケールにより示す。

【 0 2 6 5 】

図 1 1 C は、 $\sim 1 0 0$ k b の染色体 1 の欠失が、欠けているゲノムセグメントに対応する中点を持つ低密度のリードペアに基づいて明白であることを示す。

【 0 2 6 6 】

図 1 1 D は、 ~ 4 M b の染色体 8 の逆位が、アスタリスクの下に予期された密度のリードよりも高いものに基づいて明白であることを示す。逆位は、一般的に低いマップ品質スコアをもたらす反復領域によって共通して隣接される。

【 0 2 6 7 】

実施例 1 2 : サンプル収集、その後の解析、及び処理の選択

患者は、組織を除去するために手術を受ける。組織を無菌環境で切除し、ホルマリンに漬ける。組織の均質化は収集に準じて生じない。

【 0 2 6 8 】

組織を保存し、患者をモニタリングする。患者は、切除部位での再成長を受けると観察される。組織を、研究所設定での解析に晒し、この解析は、保存された組織の内部及び周囲を含む位置からの核酸タンパク質複合体の切除を含む。

【 0 2 6 9 】

ゲノム情報を、保存された組織から得た核酸タンパク質複合体から得る。ゲノム再編成は、腫瘍転移に關与する特定のゲノム配置を示す周囲組織から識別される。

【 0 2 7 0 】

化学療法処置は、腫瘍転移に關与するゲノム配置に対する既知の有効性に基づいて選択される。患者に化学療法処置を施し、腫瘍は再成長を止めると観察される。

【 0 2 7 1 】

実施例 1 3 : 治験の再評価

共通の腫瘍型を持つ個体に治験を行う。腫瘍サンプルは、治験に付随して得られる。処置された個体の亜群は処置に対し正に反応したが、この処置は全体的に、薬物の開発を是認するのに十分な有効性を持たないと観察される。

【 0 2 7 2 】

処置された集団のサンプルをショットガンゲノム配列決定に晒す。ショートリード配列

10

20

30

40

50

情報を得たが、実質的なゲノム構造情報は得られない。一塩基多型情報などの個々の配列情報は、処置の有効性に関連づけられないと観察される。

【0273】

相当な時間の経過後、サンプルを再評価する。サンプルを、複合体の完全性が保存されるように核酸タンパク質複合体の切除に晒して、本明細書に開示されるような解析に晒す。

【0274】

複合体を単離し、暴露された核酸末端をライゲートして、ペアエンド断片を形成する。ペアエンド断片を、ライゲーション部位に導入されるビオチン化された塩基を使用して単離する。

【0275】

リードペアを配列決定し、ライゲーション結合の何れかの側で配列情報を得る。リードペア情報を解析し、サンプルの亜群は、ショットガン配列決定の解析から明らかにならないゲノム再編成を含むと観察される。

【0276】

薬物反応をゲノム構造情報に照らして再評価し、特定の再編成が処置の有効性に関連付けられることを観察する。処置の有効性に関連付けられるゲノム再編成は、反応者を識別するためのマーカーとして開発され、薬物は、障害を処置するためのマーカーに関する試験と組み合わせて使用される。

【0277】

実施例 14：配列の無い再編成の検出

ペアエンドライブラリを、複数の保存されたサンプルから生成する。癌に關与するゲノムの転座中にフェーズにもたらされると知られているゲノムの領域にアニールするプライマーを使用して、ライブラリを調べる。

【0278】

ライブラリは、サンプルの亜群について高い頻度で転座されたセグメントの中の物理的連鎖を示すアンプリコンを生成すると観察される。アンプリコンをもたらすライブラリを、配列決定及びペアエンド解析に晒すと、癌に關与している疑いのある転座を独立して有していることが分かった。転座は同一ではなく、転座されたセグメントの配向及び近接の位置において異なるものであり、その結果、ゲノムの直接的なPCR解析は恐らく、大多数の転座を検出しない可能性がある。しかし、ライゲートされたペアエンドライブラリの生成を通じて、オリゴヌクレオチドプライマーは、転座の存在についてサンプルを調べるのに有効である。この手法により、下流の配列解析のためにサンプルの亜群からライブラリを選択し、それによりリソースを保存することが可能となる。

10

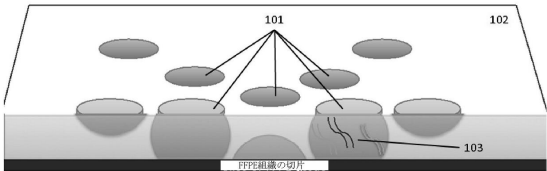
20

30

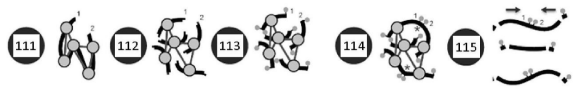
40

50

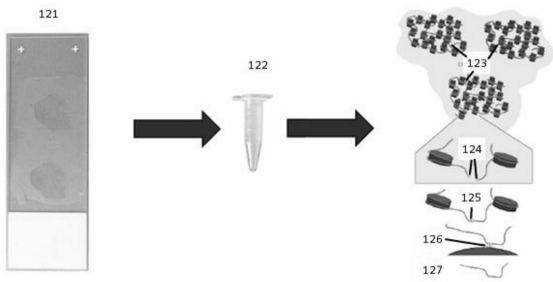
【図面】
【図 1 A】



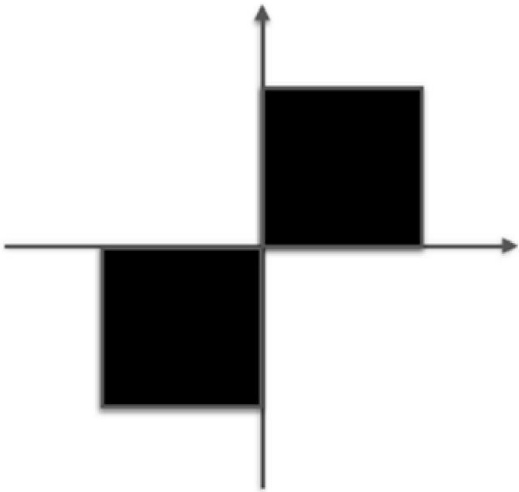
【図 1 B】



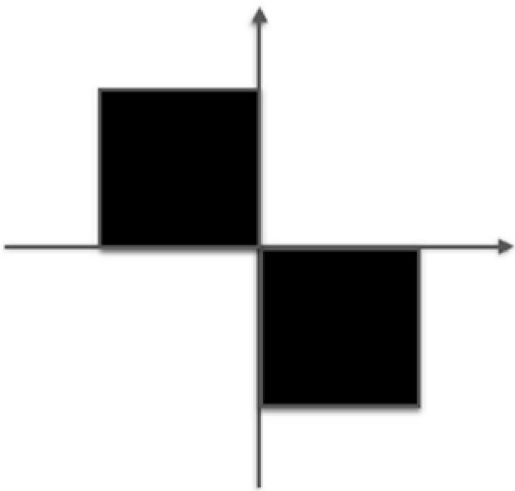
【図 1 C】



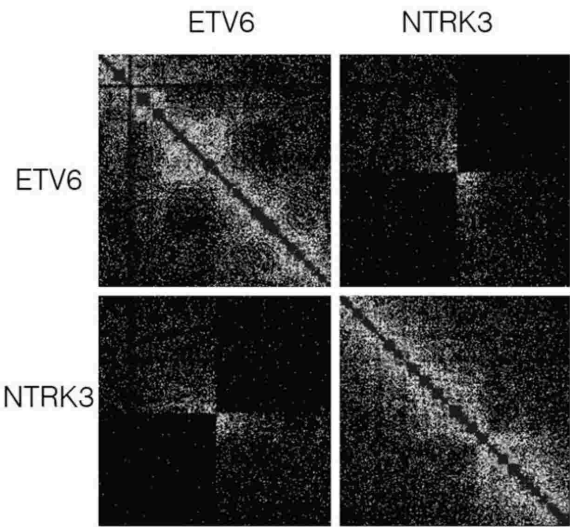
【図 2 A】



【図 2 B】



【図 3】



10

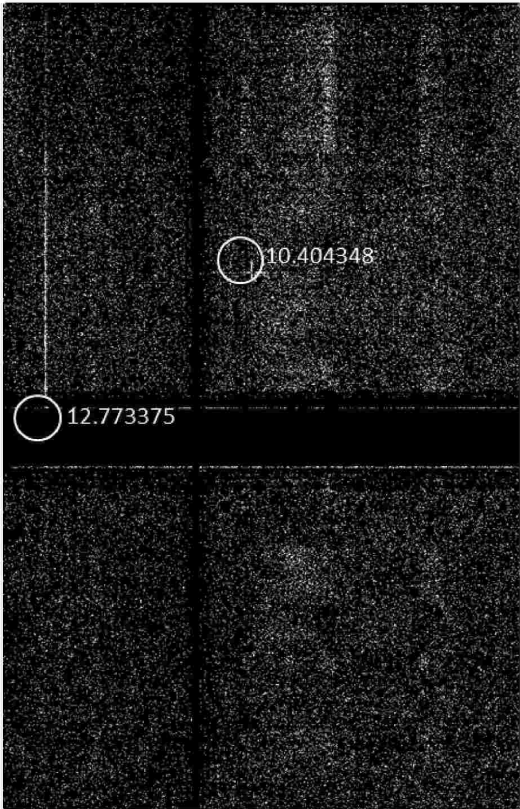
20

30

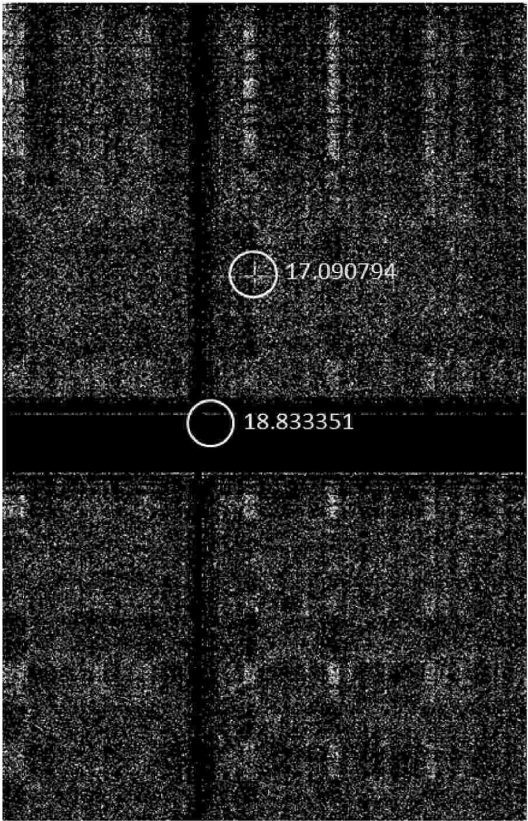
40

50

【図 4 A】



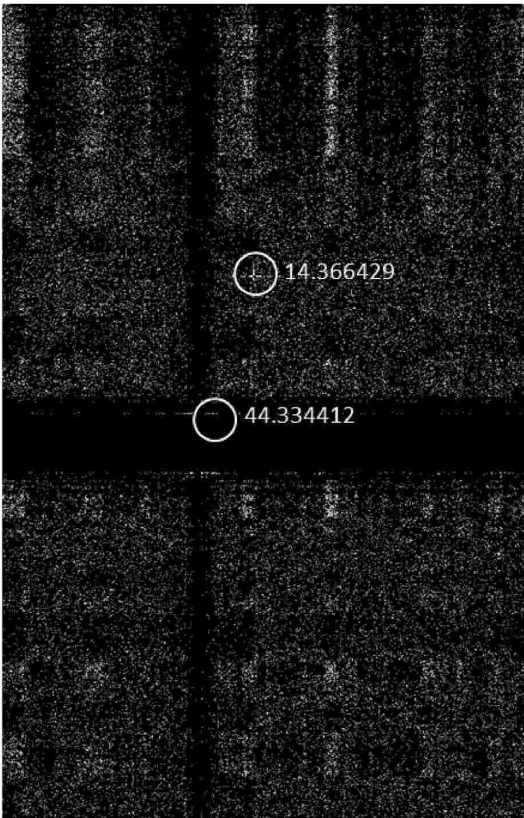
【図 4 B】



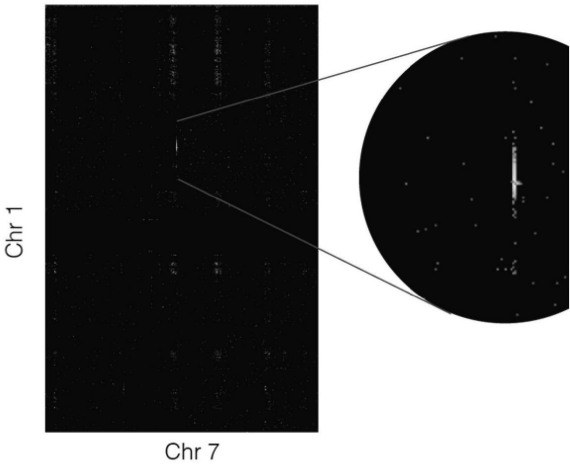
10

20

【図 4 C】



【図 5 A】

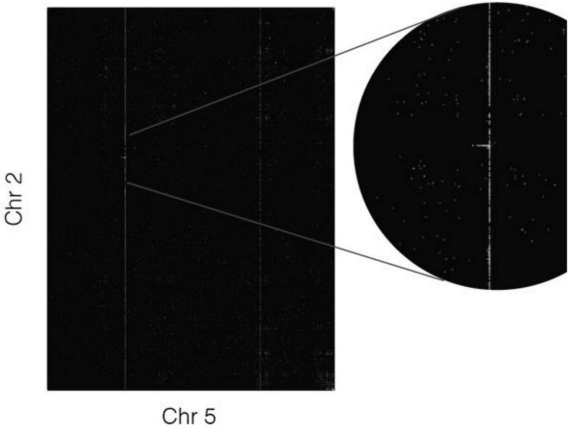


30

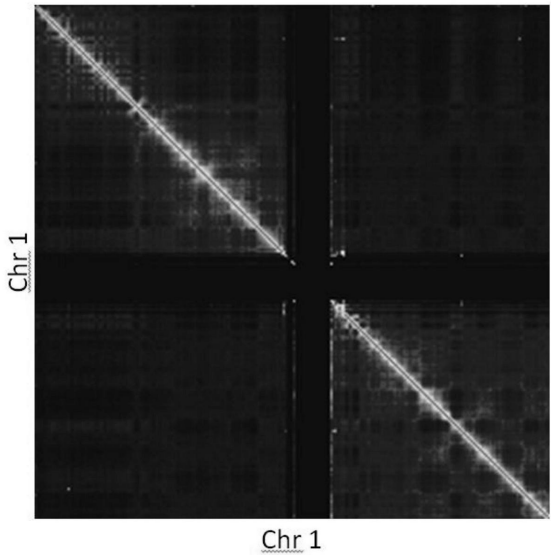
40

50

【図 5 B】

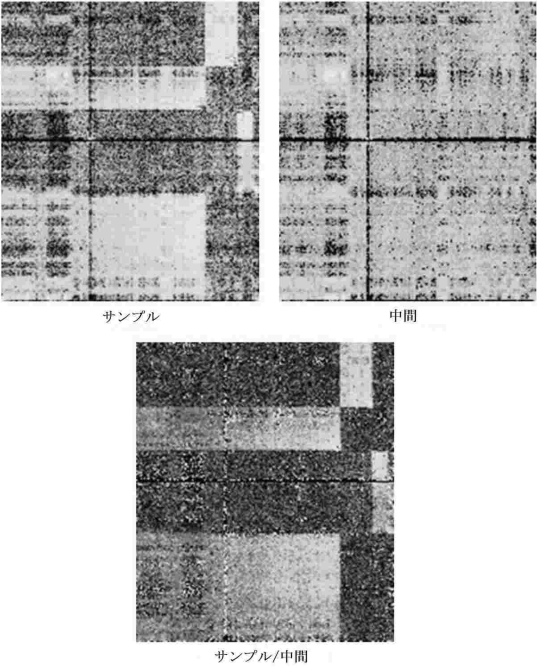


【図 5 C】

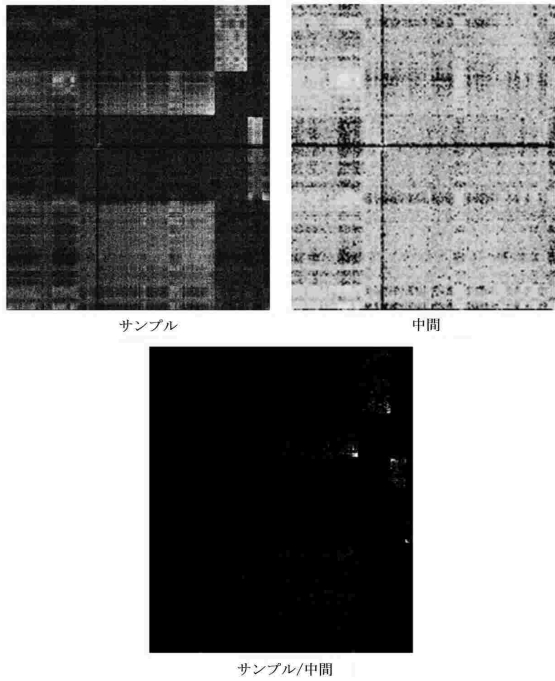


10

【図 6 A】



【図 6 B】



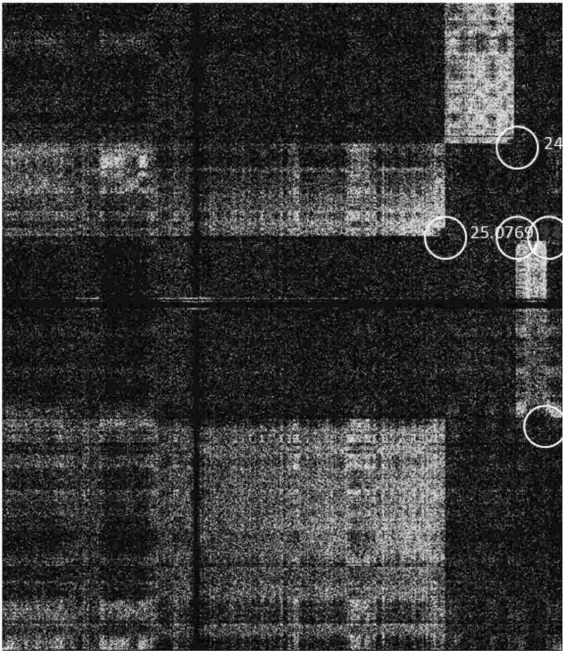
20

30

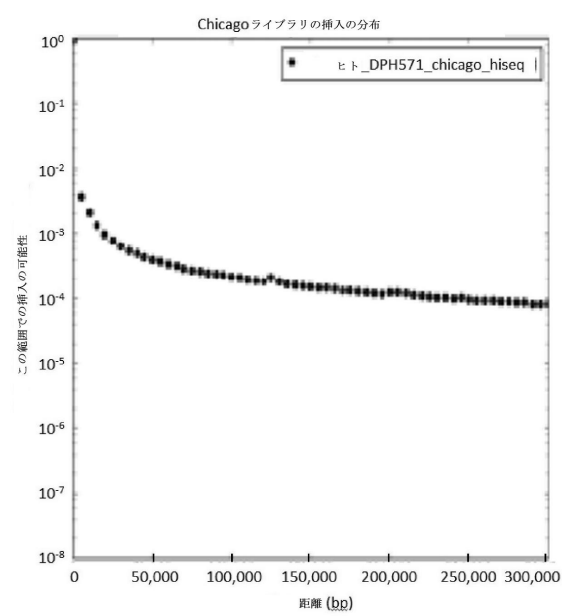
40

50

【図 7】

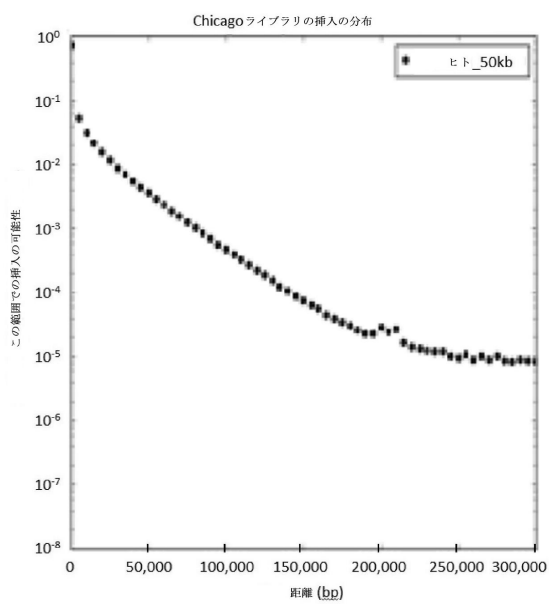


【図 8 A】

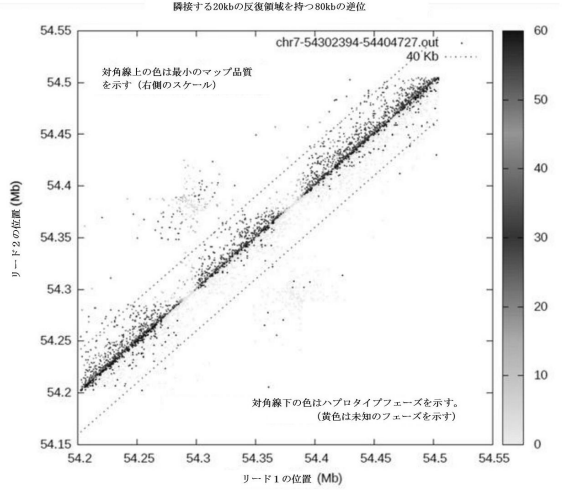


10

【図 8 B】



【図 9 A】

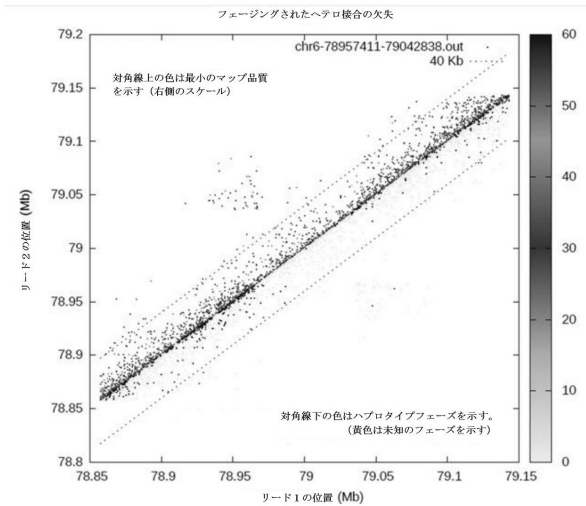


30

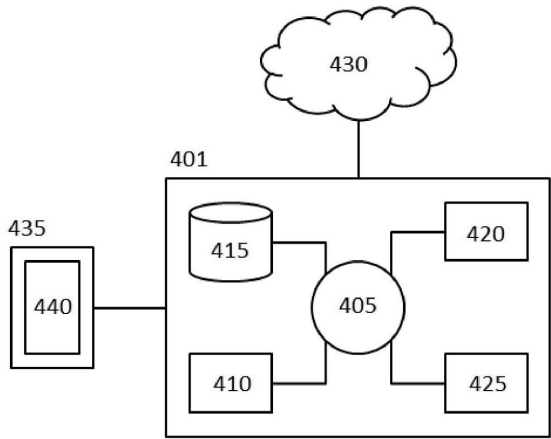
40

50

【図 9 B】

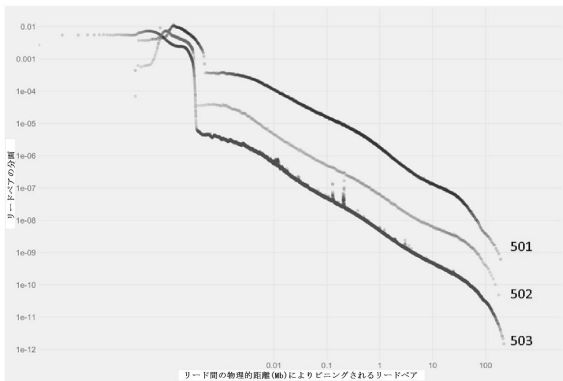


【図 1 0】

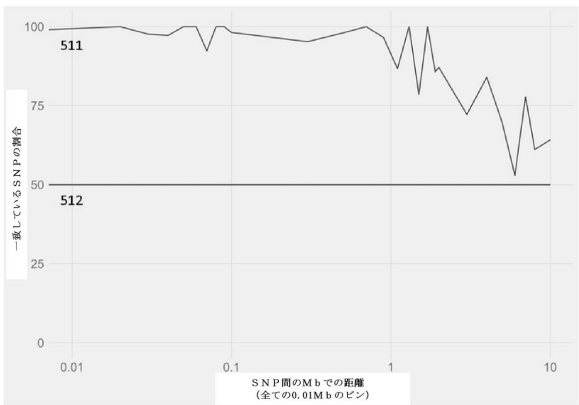


10

【図 1 1 A】



【図 1 1 B】




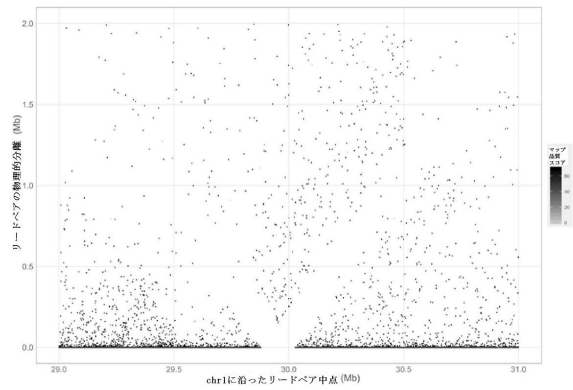
20


30

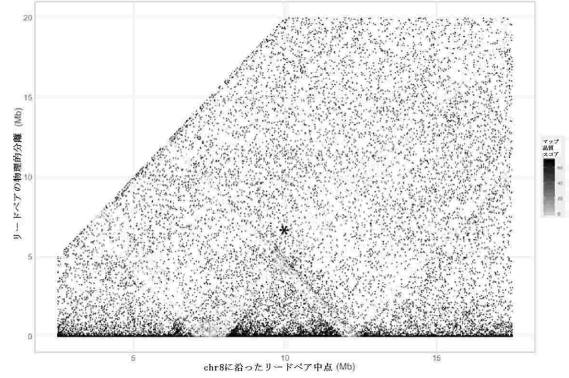
40

50

【 1 1 C】



【 1 1 D】



10

20

30

40

50

フロントページの続き

(51)国際特許分類 F I
G 1 6 B 30/20 (2019.01) G 1 6 B 30/20

(31)優先権主張番号 62/410,599

(32)優先日 平成28年10月20日(2016.10.20)

(33)優先権主張国・地域又は機関

米国(US)

ニア州 サンフランシスコ 22番アベニュー 2654

(72)発明者 バトナム, ニコラス エイチ.

アメリカ合衆国 02453 マサチューセッツ州 ウォルサム リバー・ストリート 36 アパートメント 362

(72)発明者 ブランシェット, マルコ

アメリカ合衆国 95060 カリフォルニア州 サンタクルーズ ペルトン・アベニュー 707 アpartment 307

(72)発明者 ハートレー, ポール

アメリカ合衆国 95136 カリフォルニア州 サンノゼ マウントキャッスル・ウェイ 4265

合議体

審判長 上條 肇

審判官 福井 悟

審判官 深草 亜子