

(19)日本国特許庁(JP)

(12)特許公報(B2)

(11)特許番号
特許第7653924号
(P7653924)

(45)発行日 令和7年3月31日(2025.3.31)

(24)登録日 令和7年3月21日(2025.3.21)

(51)国際特許分類		F I	
C 1 2 Q	1/6806(2018.01)	C 1 2 Q	1/6806
C 4 0 B	50/14 (2006.01)	C 4 0 B	50/14
C 1 2 N	15/10 (2006.01)	C 1 2 N	15/10
		C 1 2 N	15/10
			1 0 0 Z
			1 1 0 Z
請求項の数 25 (全139頁)			
(21)出願番号	特願2021-577189(P2021-577189)	(73)特許権者	517286593
(86)(22)出願日	令和2年6月25日(2020.6.25)		ダブテイル ゲノミクス エルエルシー
(65)公表番号	特表2022-541387(P2022-541387 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4年9月26日(2022.9.26)	(74)代理人	100082072
(86)国際出願番号	PCT/US2020/039656		弁理士 清原 義博
(87)国際公開番号	WO2020/264185	(74)復代理人	110003797
(87)国際公開日	令和2年12月30日(2020.12.30)		弁理士法人清原国際特許事務所
審査請求日	令和5年6月23日(2023.6.23)	(72)発明者	ムンディング, エリザベス
(31)優先権主張番号	62/867,463		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
(32)優先日	令和1年6月27日(2019.6.27)	(72)発明者	ブランシェット, マルコ
(33)優先権主張国・地域又は機関	米国(US)		アメリカ合衆国 9 5 0 6 6 カリフォル
(31)優先権主張番号	62/870,297		最終頁に続く
(32)優先日	令和1年7月3日(2019.7.3)		最終頁に続く
	最終頁に続く		

(54)【発明の名称】 近接ライゲーションのための方法および組成物

(57)【特許請求の範囲】

【請求項1】

(a) 核酸結合タンパク質に複合化された核酸分子を含む、架橋された生体サンプルを得る工程、

(b) 前記架橋された生体サンプルを非特異的エンドヌクレアーゼに接触させて、前記核酸分子を複数のセグメントへと切断する工程、

(c) 1つの接合部にて、前記複数のセグメントの第1のセグメントと第2のセグメントを付着させて、産物を得る工程、および、

(d) 前記(c)の産物中の架橋を反転させて、反転させた産物を得る工程、

(e) 前記(d)の反転された産物をサイズ選択にかけて、約100bp〜約600bpのサイズを有する複数の選択されたセグメントを得る工程を含む、方法であって、

前記付着させることは、前記第1のセグメントと前記第2のセグメントを少なくとも1つの架橋のためのオリゴヌクレオチドに付着させることを含む、方法。

【請求項2】

(e)に先立って、前記複数のセグメントから配列決定ライブラリを調整する工程をさらに含む、請求項1に記載の方法。

【請求項3】

サイズ選択されたライブラリを得るために、前記配列決定ライブラリをサイズ選択にかける工程をさらに含む、請求項2に記載の方法。

【請求項4】

前記方法は、1つのQC値を得るために、前記複数の選択されたセグメントを解析する工程をさらに含み、前記QC値は、クロマチン消化効率(CDE)およびクロマチン消化指標(CDI)から選択される、請求項1から3のいずれか1つに記載の方法。

【請求項5】

(b)に続いて、前記複数のセグメントを1つ以上のビーズ、アレイ、またはウェルの表面に結合させる工程をさらに含む、請求項1に記載の方法。

【請求項6】

前記架橋された生体サンプルは、架橋された細胞溶解物、架橋された無傷細胞、または架橋された無傷の核を含むことを特徴とする、請求項1から5のいずれか1つに記載の方法。

【請求項7】

前記架橋された生体サンプルは、前記無傷細胞、または前記無傷の核を含み、(b)は、前記無傷細胞あるいは前記無傷の核の溶解に先立って実施されることを特徴とする、請求項6に記載の方法。

【請求項8】

(c)に先立ち、前記架橋された生体サンプル中の細胞または核を溶解する工程をさらに含む、請求項1に記載の方法。

【請求項9】

前記架橋された生体サンプルは3,000,000未満の細胞を含むことを特徴とする、請求項1から8のいずれか1つに記載の方法。

【請求項10】

前記架橋された生体サンプルは10μg未満のDNAを含むことを特徴とする、請求項1から9のいずれか1つに記載の方法。

【請求項11】

前記非特異的エンドヌクレアーゼは、DNase I、DNase II、およびミクロコッカスヌクレアーゼからなる群から選択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】

前記架橋された生体サンプルは、化学的固定剤または紫外線から選択される橋架剤で処置されていることを特徴とする、請求項1から11のいずれか1つに記載の方法。

【請求項13】

前記化学的固定剤はホルムアルデヒド、ソラレン、グルタル酸ジスクシンイミジル(DSG)、またはエチレングリコールビス(コハク酸スクシンイミジル)(EGS)を含むことを特徴とする、請求項12に記載の方法。

【請求項14】

前記複数の選択されたセグメントを抗体に接触させる工程、および複数のセグメントについて免疫沈降を行なう工程をさらに含む、請求項1から13のいずれか1つに記載の方法。

【請求項15】

前記免疫沈降は前記付着させた後に行なわれることを特徴とする、請求項14に記載の方法。

【請求項16】

前記(c)において前記付着させることは、ビオチンをタグ付けされたヌクレオチドを使用して粘着末端を埋めること、タグ付けされていないヌクレオチドを使用して粘着末端を埋めること、平滑末端をライゲーションすること、オーバーハングを付加すること、または前記第1のセグメントと前記第2のセグメントをバーコードに接触させることを含む、請求項1から15のいずれか1つに記載の方法。

【請求項17】

前記架橋のためのオリゴヌクレオチドは、バーコード配列、またはアフィニティタグを含むことを特徴とする、請求項1に記載の方法。

10

20

30

40

50

【請求項 18】

前記(c)において前記付着させることは、前記第1のセグメントと前記第2のセグメントを複数の架橋のためのオリゴヌクレオチドに連続して接触させることを含む、請求項1に記載の方法。

【請求項 19】

前記架橋のためのオリゴヌクレオチドが、1つの免疫グロブリン結合タンパク質あるいはその断片に連結され、ここで、前記免疫グロブリン結合タンパク質は、プロテインA、プロテインG、プロテインA/G、およびプロテインLから選択されることを特徴とする、請求項1から18のいずれか1つに記載の方法。

【請求項 20】

(e) 第1のリードペアを生成するために、前記接合部の両側で少なくともいくつかの配列を得る工程をさらに含む、請求項1に記載の方法。

【請求項 21】

(f) コンティグのセットに前記第1のリードペアをマッピングする工程、および、
(g) 順序またはゲノムへの配向を表現する前記コンティグのセットにわたる経路を決定する工程をさらに含む、請求項20に記載の方法。

【請求項 22】

(f) コンティグのセットに前記第1のリードペアをマッピングする工程、および、
(g) 前記コンティグのセットから、架橋された生体サンプルにおける構造的変異体の存在またはヘテロ結合性の減少を判定する工程をさらに含む、請求項20に記載の方法。

【請求項 23】

(f) コンティグのセットに前記第1のリードペアをマッピングする工程、および、
(g) 前記コンティグのセットにおける変異体をフェーズに割り当てる工程をさらに含む、請求項20に記載の方法。

【請求項 24】

前記変異体はヒト白血球抗原(HLA)変異体、またはキラー細胞免疫グロブリン様受容体(KIR)変異体であることを特徴とする、請求項23に記載の方法。

【請求項 25】

前記非特異的エンドヌクレアーゼは、免疫グロブリン結合タンパク質またはその断片に連結または融合されることを特徴とする、請求項1から24のいずれか1つに記載の方法。

【発明の詳細な説明】**【技術分野】****【0001】****相互参照**

本出願は、2019年6月27日に出願された米国仮特許出願62/867,463号、2019年11月5日に出願された米国仮特許出願62/931,069号、2020年4月17日に出願された米国仮特許出願63/011,490号、2019年7月3日に
出願された米国仮特許出願62/870,297号、および2020年4月23日に
出願された米国仮特許出願63/014,422号の利益を主張し、その各々は、参照により全体が本明細書に取り込まれる。

【背景技術】**【0002】**

高品質の隣接するゲノム配列を得ることは、多くの場合、特に、配列解析に利用可能な供給源材料が限られている場合に、困難である。生の配列データを得ることはより速くなり、またより低コストで利用可能になったが、データを効率的かつ正確に分析し、および組み立てるための適切な方法は、難題のままである。

【発明の概要】**【0003】**

一態様では、核酸解析の方法が提供される。場合によっては、方法は、(a)少なくとも1つの核酸結合タンパク質に複合化された核酸分子を含む、安定化された生体サンプル

10

20

30

40

50

を得る工程、(b) 核酸分子を複数のセグメントへと切断するために、安定化された生体サンプルを非特異的エンドヌクレアーゼに接触させる工程、(c) 1つの接合部にて、複数のセグメントの第1のセグメントと第2のセグメントを付着させる工程、および、(d) 複数の選択されたセグメントを得るために、複数のセグメントをサイズ選択にかける工程。場合によっては、複数の選択されたセグメントは、約145～約600bpである。場合によっては、複数の選択されたセグメントは約100～約2500bpである。場合によっては、複数の選択されたセグメントは約100～約600bpである。場合によっては、複数の選択されたセグメントは約600～約2500bpである。場合によっては、方法は、工程(d)に先立って、複数のセグメントから配列決定ライブラリを調整する工程をさらに含む。場合によっては、方法は、サイズ選択されたライブラリを得るために、配列決定ライブラリをサイズ選択にかける工程をさらに含む。場合によっては、サイズ選択されたライブラリは、約350bpから1000bpの間のサイズにある。場合によっては、サイズ選択は、ゲル電気泳動、キャピラリー電気泳動、サイズ選択ビーズ、またはゲルろ過カラムを用いて行なわれる。場合によっては、方法は、あるQC値を得るために、複数の選択されたセグメントを解析する工程をさらに含む。場合によっては、QC値は、工程(d)の前の、100から2500bpの間のサイズのセグメントの割合に基づくクロマチン消化効率(CDE)である。場合によっては、方法は、CDE値が少なくとも65%であるときに、さらなる解析のためにサンプルを選択する工程をさらに含む。場合によっては、QC値は、工程(d)の前の、モノヌクレオソーム・サイズのセグメントの数のジヌクレオソーム・サイズのセグメントの数に対する比に基づくクロマチン消化指標(CDI)である。場合によっては、方法は、CDI値が-1.5より大きく、かつ1未満のあるとき、さらなる解析のためにサンプルを選択する工程をさらに含む。場合によっては、方法は、安定化された生体サンプルを非特異的エンドヌクレアーゼに接触させる工程に続いて、複数のセグメントを1つ以上の表面に結合させる工程をさらに含む。場合によっては、1つ以上の表面は、1つ以上のビーズを含む。場合によっては、1つ以上のビーズは、固相可逆固定(SPRI)ビーズである。場合によっては、安定化された生体サンプルは、安定化された細胞溶解物を含む。場合によっては、安定化された生体サンプルは、安定化された無傷細胞を含む。場合によっては、安定化された生体サンプルは、安定化された無傷の核を含む。場合によっては、工程(b)は、無傷細胞あるいは無傷の核の溶解に先立って行なわれる。場合によっては、方法は、工程(c)に先立ち、安定化された生体サンプル中の細胞および／または核を溶解する工程をさらに含む。場合によっては、安定化された生体サンプルは3,000,000未満の細胞を含む。場合によっては、安定化された生体サンプルは1,000,000未満の細胞を含む。場合によっては、安定化された生体サンプルは100,000未満の細胞を含む。場合によっては、安定化された生体サンプルは10μg未満のDNAを含む。場合によっては、安定化された生体サンプルは1μg未満のDNAを含む。場合によっては、非特異的エンドヌクレアーゼはDNase(DNase)である。場合によっては、DNaseはDNase Iである。場合によっては、DNaseはDNase IIである。場合によっては、DNaseはミクロコッカスヌクレアーゼである。場合によっては、DNaseが、DNase I、DNase II、およびミクロコッカスヌクレアーゼの1つ以上から選択される。場合によっては、安定された生体サンプルは、橋架剤で処置されている。場合によっては、橋架剤は化学的固定剤である。場合によっては、化学的固定剤はホルムアルデヒドを含む。場合によっては、化学的固定剤はソラレンを含む。場合によっては、化学的固定剤はグルタル酸ジスクシンイミジル(DSG)を含む。場合によっては、化学的固定剤はエチレンジグリコールビス(コハク酸スクシンイミジル(succinimidyl succinate))(EGS)を含む。場合によっては、化学的な固定剤がdisuccinimidylグルタル酸塩(DSG)とエチレンジグリコールビス(スクシンイミジル・コハク酸塩)(EGS)を含む。場合によっては、橋架剤は紫外線である。場合によっては、安定化された生体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。場合によっては、方法は、選択された複数のセグメントを抗体に接触させる工程をさらに含む。場合

10

20

30

40

50

によっては、方法は、複数のセグメント上で免疫沈降を行なう工程をさらに含む。場合によっては、免疫沈降が付着させる工程の後に続いて行なわれる。場合によっては、付着させる工程は、ビオチンをタグ付けされたヌクレオチドを使用して粘着末端 (sticky end) を埋めることを含む。場合によっては、付着させる工程は、タグ付けされていないヌクレオチドを使用して粘着末端を埋めることを含む。場合によっては、付着させる工程は、平滑末端 (blunt end) をライゲーションすることを含む。場合によっては、付着させる工程は、オーバーハングを付加することを含む。場合によっては、オーバーハングを付加することは、アデニル化を含む。場合によっては、付着させる工程は、少なくとも第1のセグメントと第2のセグメントを少なくとも1つの架橋 (bridge) オリゴヌクレオチドに接触させることを含む。場合によっては、架橋オリゴヌクレオチドは、長さが少なくとも10bpである。場合によっては、架橋オリゴヌクレオチドは、長さが少なくとも12bpである。場合によっては、架橋オリゴヌクレオチドは、長さが12bpである。場合によっては、架橋オリゴヌクレオチドはバーコード配列を含む。いくつかの実施形態では、第1オリゴヌクレオチドは、アフィニティタグを含む。場合によっては、アフィニティタグはビオチンである。場合によっては、付着させる工程は、少なくとも第1のセグメントと第2のセグメントを複数の架橋オリゴヌクレオチドに連続して接触させることを含む。場合によっては、付着させる工程は、安定化された生体サンプルのサンプル、細胞、核、染色体、または核酸分子が架橋オリゴヌクレオチドの固有の配列を受け取ることを結果としてもたらす。場合によっては、少なくとも1つの架橋オリゴヌクレオチドが1つの免疫グロブリン結合タンパク質あるいは1つのその断片に連結される。場合によっては、少なくとも1つの架橋オリゴヌクレオチドが2つ以上の免疫グロブリン結合タンパク質あるいは2つ以上のその断片に連結または融合される。場合によっては、免疫グロブリン結合タンパク質は、プロテインA、プロテインG、プロテインA/G、およびプロテインLから選択される。場合によっては、付着させる工程は、少なくとも第1のセグメントと第2のセグメントをバーコードに接触させることを含む。場合によっては、方法は剪断する工程を含まない。場合によっては、方法は、(e) 第1のリードペアを生成するために、接合部の両側で少なくともいくつかの配列を獲得する工程をさらに含む。場合によっては、方法は、(f) 1セットのコンティグに第1のリードペアをマッピングする工程、および、(g) ゲノムへの順序および／または配向を表わす、コンティグのセットにわたる経路を決定する工程をさらに含む。代替的に、あるいは組み合わせて、方法は、(f) コンティグの1セットに第1のリードペアをマッピングする工程、および、(g) 前記コンティグのセットから、安定化された生体サンプルにおける構造的変異体の存在またはヘテロ結合性の減少を判定する工程をさらに含む。代替的に、あるいは組み合わせて、方法は、(f) 1セットのコンティグに第1のリードペアをマッピングする工程、および、(g) コンティグのセットにおける変異体をフェーズに割り当てる工程をさらに含む。場合によっては、変異体はヒト白血球抗原 (HLA) 変異体である。場合によっては、変異体はキラー細胞免疫グロブリン様受容体 (KIR) 変異体である。代替的に、あるいは組み合わせて、方法は、(f) コンティグの1セットに第1のリードペアをマッピングする工程、(g) 前記コンティグのセットから、安定化された生体における変異体の存在を判定する工程、および、(h) (1) 安定化された生体サンプルについて疾患段階、予後、あるいは処置の方針を確認するステップ、(2) 変異体の存在に基づいて、薬物を選択するステップ、または、(3) 安定化された生体サンプルに対する薬効を確認するステップから選択される1つ以上のステップを行なう工程をさらに含む。場合によっては、DNアーゼは免疫グロブリン結合タンパク質あるいはその断片に連結または融合される。場合によっては、DNaseは2以上の免疫グロブリン結合タンパク質あるいはその断片に連結される。場合によっては、免疫グロブリン結合タンパク質は、プロテインA、プロテインG、プロテインA/G、およびプロテインLから選択される。

【0004】

他の態様では、以下の工程を含む方法が提供される。(a) 少なくとも1つの核酸結合タンパク質に複合化された核酸分子を含む、安定化された生体サンプルを得る工程、(b

10

20

30

40

50

）核酸分子を複数のセグメントへと切断するために、安定化された生体サンプルをマイクロコッカスヌクレアーゼ（M N a s e）に接触させる工程、および、（c）1つの接合部に、複数のセグメントの第1のセグメントと第2のセグメントを付着させる工程。場合によっては、方法は、（d）複数の選択されたセグメントを得るために、複数のセグメントをサイズ選択にかける工程をさらに含む。場合によっては、複数の選択されたセグメントは、約145～約600bpである。場合によっては、複数の選択されたセグメントは約100～約2500bpである。場合によっては、複数の選択されたセグメントは約100～約600bpである。場合によっては、複数の選択されたセグメントは約600～約2500bpである。場合によっては、方法は、工程（d）に先立って、複数のセグメントから配列決定ライブラリを調整する工程をさらに含む。場合によっては、方法は、サイズ選択されたライブラリを得るために、配列決定ライブラリをサイズ選択にかける工程をさらに含む。場合によっては、サイズ選択されたライブラリは、約350bpから1000bpの間のサイズにある。場合によっては、サイズ選択は、ゲル電気泳動、キャピラリー電気泳動、サイズ選択ビーズ、またはゲルろ過カラムを用いて行なわれる。場合によっては、方法は、1つのQC値を得るために、複数の選択されたセグメントを解析する工程をさらに含む。場合によっては、QC値は、工程（d）の前の、100から2500bpの間のサイズのセグメントの割合に基づくクロマチン消化効率（CDE）である。場合によっては、方法は、CDE値が少なくとも65%であるときに、さらなる解析のためにサンプルを選択する工程をさらに含む。場合によっては、QC値は、工程（d）の前の、モノヌクレオソーム・サイズのセグメントの数のジヌクレオソーム・サイズのセグメントの数に対する比に基づくクロマチン消化指標（CDI）である。場合によっては、方法は、CDI値が-1.5より大きく、かつ1未満であるとき、さらなる解析のためにサンプルを選択する工程をさらに含む。場合によっては、方法は、安定化された生体サンプルをM N a s eに接触させる工程に続いて、複数のセグメントを1つ以上の表面に結合する工程をさらに含む。場合によっては、1つ以上の表面は、1つ以上のビーズを含む。場合によっては、1つ以上のビーズは、固相可逆固定（SPRI）ビーズである。場合によっては、安定化された生体サンプルは、安定化された細胞溶解物を含む。場合によっては、安定化された生体サンプルは、安定化された無傷細胞を含む。場合によっては、安定化された生体サンプルは、安定化された無傷の核を含む。場合によっては、工程（b）は、無傷細胞あるいは無傷の核の溶解に先立って行なわれる。場合によっては、方法は、工程（c）に先立ち、安定化された生体サンプル中の細胞および／または核を溶解する工程をさらに含む。場合によっては、安定化された生体サンプルは3,000,000未満の細胞を含む。場合によっては、安定化された生体サンプルは1,000,000未満の細胞を含む。場合によっては、安定化された生体サンプルは100,000未満の細胞を含む。場合によっては、安定化された生体サンプルは10μg未満のDNAを含む。場合によっては、安定化された生体サンプルは1μg未満のDNAを含む。場合によっては、安定された生体サンプルは、DNaseでさらに処置される。場合によっては、DNaseはDNase Iである。場合によっては、DNaseはDNase IIである。場合によっては、DNaseは、DNase I、およびDNase IIの1つ以上から選択される。場合によっては、安定された生体サンプルは、橋架剤で処置されている。場合によっては、橋架剤は化学的固定剤である。場合によっては、化学的固定剤はホルムアルデヒドを含む。場合によっては、化学的固定剤はソラレンを含む。場合によっては、化学的固定剤はグルタル酸ジスクシンイミジル（DSG）を含む。場合によっては、化学的固定剤はエチレングリコールビス（コハク酸スクシンイミジル（succinimidyl succinate））（EGS）を含む。場合によっては、化学的固定剤は、グルタル酸ジスクシンイミジル（disuccinimidyl glutarate）（DSG）とエチレングリコールビス（コハク酸スクシンイミジル）（EGS）を含む。場合によっては、橋架剤は紫外線である。場合によっては、安定化された生体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。場合によっては、方法は、選択された複数のセグメントを抗体に接触させる工程をさらに含む。場合によっては、方法は、複数のセグメント上で免疫

10

20

30

40

50

沈降を行なう工程をさらに含む。場合によっては、免疫沈降が付着させる工程の後に続いて行なわれる。場合によっては、付着させる工程は、ビオチンをタグ付けされたヌクレオチドを使用して粘着末端を埋めることを含む。場合によっては、付着させる工程は、タグ付けされていないヌクレオチドを使用して粘着末端を埋めることを含む。場合によっては、付着させる工程は、平滑末端をライゲーションすることを含む。場合によっては、付着させる工程は、オーバーハングを付加することを含む。場合によっては、オーバーハングを添加することは、アデニル化を含む。場合によっては、付着させる工程は、少なくとも第1のセグメントと第2のセグメントを架橋オリゴヌクレオチドに接触させることを含む。場合によっては、架橋オリゴヌクレオチドは、長さが少なくとも10bpである。場合によっては、架橋オリゴヌクレオチドは、長さが少なくとも12bpである。場合によっては、架橋オリゴヌクレオチドは、長さが12bpである。場合によっては、架橋オリゴヌクレオチドはバーコード配列を含む。いくつかの実施形態では、第1オリゴヌクレオチドは、アフィニティタグを含む。場合によっては、アフィニティタグはビオチンである。場合によっては、付着させる工程は、少なくとも第1のセグメントと第2のセグメントを複数の架橋オリゴヌクレオチドに連続して接触させることを含む。場合によっては、前記付着させる工程は、安定化された生体サンプルのサンプル、細胞、核、染色体、または核酸分子が架橋オリゴヌクレオチドの固有の配列を受け取ることを結果としてもたらす。場合によっては、少なくとも1つの架橋オリゴヌクレオチドが1つの免疫グロブリン結合タンパク質あるいは1つのその断片に連結される。場合によっては、少なくとも1つの架橋オリゴヌクレオチドが2つ以上の免疫グロブリン結合タンパク質あるいは2つ以上のその断片に連結される。場合によっては、免疫グロブリン結合タンパク質は、プロテインA、プロテインG、プロテインA/G、およびプロテインLから選択される。場合によっては、付着させる工程は、少なくとも第1のセグメントと第2のセグメントをバーコードに接触させることを含む。場合によっては、方法は剪断する工程を含まない。場合によっては、方法は、(e) 第1のリードペアを生成するために、接合部の両側で少なくともいくつかの配列を獲得する工程をさらに含む。場合によっては、方法は、(f) 1セットのコンティグに第1のリードペアをマッピングする工程、および、(g) 順序および/またはゲノムへの配向を表現するコンティグのセットにわたる経路を決定する工程をさらに含む。代替的に、あるいは組み合わせて、方法は、(f) コンティグの1セットに第1のリードペアをマッピングする工程、および、(g) 前記コンティグのセットから、安定化された生体サンプルにおける構造的変異体の存在またはヘテロ結合性の減少を判定する工程をさらに含む。代替的に、あるいは組み合わせて、方法は、(f) 1セットのコンティグに第1のリードペアをマッピングする工程、および、(g) コンティグのセットにおける変異体をフェーズに割り当てる工程をさらに含む。場合によっては、変異体はヒト白血球抗原(HLA)変異体である。場合によっては、変異体はキラー細胞免疫グロブリン様受容体(KIR)変異体である。代替的に、あるいは組み合わせて、方法は、(f) コンティグの1セットに第1のリードペアをマッピングする工程、(g) 前記コンティグのセットから、安定化された生体における変異体の存在を判定する工程、および、(h) (1) 安定化された生体サンプルについて疾患段階、予後、あるいは処置の方針を確認するステップ、(2) 変異体の存在に基づいて、薬物を選択するステップ、または、(3) 安定化された生体サンプルに対する薬効を確認するステップから選択される1つ以上のステップを行なう工程をさらに含む。場合によっては、MNaseは免疫グロブリン結合タンパク質に連結または融合される。場合によっては、MNaseは2以上の免疫グロブリン結合タンパク質あるいはその断片に連結または融合される。場合によっては、免疫グロブリン結合タンパク質は、プロテインA、プロテインG、プロテインA/G、およびプロテインLから選択される。

【0005】

付加的な態様では、以下を含む核酸ライブラリが提供される。(a) 複数の第1の細胞ゲノム断片対を含む第1の細胞ゲノムライブラリー構成要素であって、ここで、前記第1の細胞ゲノム断片対の少なくとも1つは、第1の細胞ゲノムライブラリー表示タグを含む

10

20

30

40

50

核酸セグメントを介してテザーされる2つの第1の細胞ゲノムセグメントを含む、第1の細胞ゲノムライブラリー構成要素。および、(b)複数の第2の細胞ゲノム断片対を含む第2の細胞ゲノムライブラリー構成要素であって、ここで、前記第2の細胞ゲノム断片対の少なくとも1つは、第2の細胞ゲノムライブラリー表示タグを含む核酸セグメントを介してテザーされる2つの第2の細胞ゲノムセグメントを含む、第2の細胞ゲノムライブラリー構成要素。場合によっては、第1の細胞ゲノムライブラリー表示タグを含む核酸セグメントを介してテザーされる2つの第1の細胞ゲノムセグメントは、第1の細胞における第1の細胞ゲノム構造を表示する。場合によっては、核酸セグメントを介してテザーされた、2つの第2の細胞ゲノムセグメントは、第2の細胞における第2の細胞ゲノム構造を表示する第2の細胞ゲノムライブラリー表示タグを含み、ここで、第2の細胞ゲノム構造は第1の細胞ゲノム構造とは異なる。場合によっては、第1の細胞ゲノムライブラリー構成要素は単離された真核生物の核から得られる。場合によっては、複数の第1の細胞ゲノム断片対がリコンビナーゼ部位によって両側から挟まれる。場合によっては、リコンビナーゼ部位はインテグラーゼ組み込み部位である。場合によっては、リコンビナーゼ部位はトランスポサース・モザイク末端である。場合によっては、リコンビナーゼ部位の少なくとも1つのリコンビナーゼ部位は、エキソヌクレアーゼ耐性部分を含む。場合によっては、エキソヌクレアーゼ耐性部分は、ホスホロチオエートを含む。場合によっては、第1の細胞ゲノムライブラリー表示タグを含む核酸セグメントは、リコンビナーゼ左境界とリコンビナーゼ右境界をさらに含む。場合によっては、リコンビナーゼはインテグラーゼである。場合によっては、リコンビナーゼはトランスポサースである。場合によっては、第1の細胞ゲノムライブラリー表示タグを含む核酸セグメントは、アフィニティタグを含む。場合によっては、アフィニティタグはビオチンを含む。場合によっては、少なくともいくつかのライブラリー・メンバはクローンコピーである。場合によっては、核酸レファレンスの比較可能な領域にマッピングされるリードペアの共起は、細胞における領域の互いの距離を示す。場合によっては、距離は相対距離である。

【0006】

付加的な態様では、システムが提供され、該システムは、複数の細胞ゲノムアリコートを含み、ここで細胞ゲノムアリコートのうちの少なくともいくつかはゲノム構成要素の位置情報を持ち続ける部分を結合するゲノムを含み、および、前記システムは、複数のリコンビナーゼ核酸のアリコートを含み、ここでリコンビナーゼ核酸のアリコートのうちの少なくともいくつかは、少なくとも1つの他のアリコートに対して識別可能な配列を含む。場合によっては、リコンビナーゼはインテグラーゼである。場合によっては、リコンビナーゼはトランスポサースである。場合によっては、細胞ゲノムアリコートのうちの少なくともいくつかは断片化されたゲノム分子を含む。場合によっては、断片化されたゲノム分子のうちの少なくともいくつかは、組み込み部位末端を含む。場合によっては、細胞ゲノムアリコートは、真核細胞ゲノムアリコートを含む。場合によっては、ゲノム構成要素の位置情報を持ち続けるゲノム結合部分は、クロマチン成分を含む。場合によっては、ゲノム構成要素位置情報を持ち続けるゲノム結合部分は、ヌクレオソームを含む。場合によっては、複数の細胞ゲノムアリコートは、インテグラーゼ酵素を含む。場合によっては、複数のリコンビナーゼ核酸アリコートは、インテグラーゼ組み込み部位があるインテグラーゼ核酸分子を含む。場合によっては、複数の細胞ゲノムアリコートは、トランスポサース酵素を含む。場合によっては、複数のリコンビナーゼ核酸一定分量は、トランスポサース・モザイク末端があるトランスポサース核酸分子を含む。場合によっては、複数の組み込み部位の少なくとも1つの組み込み部位は、エキソヌクレアーゼ耐性部分を含む。場合によっては、複数のモザイク末端の少なくとも1つのモザイク末端が耐エキソヌクレアーゼ抵抗性の部分を含む。場合によっては、エキソヌクレアーゼ耐性部分は、ホスホロチオエートを含む。場合によっては、リコンビナーゼ核酸分子のうちの少なくともいくつかは、リコンビナーゼ左境界とリコンビナーゼ右境界を含む。場合によっては、第1の細胞ゲノムライブラリー表示タグを含む核酸セグメントは、アフィニティタグを含む。場合によっては、アフィニティタグはビオチンを含む。場合によっては、少なくとも1つの他のアリ

10

20

30

40

50

コートに対して識別可能な配列は、アリコートにおいて複数の同一の核酸配列を含む。場合によっては、第1のアリコートは、少なくとも1つの他のアリコートに対して共通の識別可能な配列を有する複数の核酸分子を含む。場合によっては、複数の細胞ゲノムアリコートは単細胞ゲノムアリコートをを含む。

【0007】

付加的な態様では、少なくとも2つの細胞間の染色体立体配座変異についてアッセイの方法が提供される。場合によっては、方法は、染色体の立体配座変異が保存されている2つの細胞由来のゲノムの核酸を得る工程、2つの細胞由来のゲノム核酸の中に内部切断を導入する工程；および、複数のタグ付けされたセグメントのうちの1つを介して、内部切断部に互いに隣接している2つの露出した末端をリンクさせる工程であって、ここで第1の細胞のタグは第2の細胞のタグから識別可能である、工程を含む。場合によっては、2つの細胞由来のゲノム核酸は、前記リンクさせる工程に先立って単離される。場合によっては、2つの細胞由来の染色体切断接合部が識別可能であるように識別可能な配列を含む内部セグメントによってリンクされる染色体切断接合部末端を含むアンプリコンを作り出すために、前記リンクさせる工程の結果として生じた核酸分子を増幅する工程が含まれる。場合によっては、第1のタグ付けされたセグメントによってリンクされた2つの以前は露出していた末端を含む2つの細胞のうちの第1の細胞に由来する断片、および、第2のタグ付けされたセグメントによってリンクされた2つの以前は露出していた末端を含む少なくとも2つの細胞のうちの第2の細胞に由来する断片を、単離する工程が含まれる。場合によっては、得られたペアエンド (paired end) の配列情報は、少なくともいくつかの第1のタグ情報と少なくともいくつかの第2のタグ情報を含む。場合によっては、細胞における共通の近接部に、ペアエンドを割り当てる工程が含まれる。場合によっては、共通の近接部にペアエンドを割り当てる工程は、2つの共通のクラスターにマッピングするペアエンドの発生回数を算定すること、および、発生回数に相対的な近接を関連させることを含む。場合によっては、第1の細胞由来の前記断片と、第2の細胞由来の前記断片とは、共通の容積中に単離される。場合によっては、第1の細胞由来の前記断片と、第2の細胞由来の前記断片とは、共通の容積中で配列決定される。場合によっては、互いに内部切断部に隣接している2つの露出された末端をリンクさせる工程は、前記内部切断を導入する工程の前にはすぐ傍には隣接していなかった2つの露出した端部をリンクさせることを含む。場合によっては、互いに内部切断部に隣接している2つの露出された末端をリンクさせる工程は、前記内部切断を導入する工程の前には共通の核酸分子上で互いに遠隔にあった2つの露出した末端をリンクさせることを含む。場合によっては、互いに内部切断部に隣接している2つの露出された末端をリンクさせる工程は、前記内部切断を導入する工程の前には互いに物理的に隣接していた2つの露出した端部をリンクさせることを含む。場合によっては、少なくとも2つの細胞は、少なくとも2つの細胞集団を含む。

【0008】

さらなる態様では、少なくとも1つの核酸結合タンパク質に複合化された核酸分子を含む、安定化されたサンプルを得る工程、少なくとも第1のセグメントと第2のセグメントを含む複数のセグメントへと核酸分子を切断する工程、第1のセグメントおよび第2のセグメントに第1のリコンビナーゼ部位を含むアダプターを付着させる工程、および、第1のセグメントおよび第2のセグメントを、第2のリコンビナーゼ部位を含むリンカーと、リコンビナーゼの存在下において、接触させる工程であって、それにより、第1のセグメント由来の第1の配列と、リンカー由来のリンカー配列と、第2のセグメント由来の第2の配列とを含むリンクされた核酸を生成する、工程を含む方法が提供される。場合によっては、リコンビナーゼはインテグラーゼである。場合によっては、リコンビナーゼはトランスポゼースである。場合によっては、方法は、リンクされた核酸の少なくとも一部を配列決定する工程をさらに含む。場合によっては、配列決定する工程は、第1の配列の少なくとも一部および第2の配列の少なくとも一部を配列決定することを含む。場合によっては、方法は、第1の配列の少なくとも一部および第2の配列の少なくとも一部をゲノムにマッピングする工程をさらに含む。場合によっては、方法は、配列決定する工程からの情

10

20

30

40

50

報を使用して、三次元ゲノム解析を行なう工程をさらに含む。場合によっては、安定化された生体サンプルは架橋の試料である。場合によっては、安定化されたサンプルを得る工程は、サンプルを得ること、およびサンプルを安定化させることを含む。場合によっては、安定したサンプルを得る工程は、予め安定化されたサンプルを得ることを含む。場合によっては、核酸結合タンパク質は、クロマチンあるいはその成分を含む。場合によっては、切断する工程は、酵素学的な消化を含む。場合によっては、酵素学的な消化は、1つ以上の制限酵素による消化を含む。場合によっては、酵素学的な消化は、1つ以上の非特異的ヌクレアーゼによる消化を含む。場合によっては、1つ以上の非特異的ヌクレアーゼは、DNaseまたはMNaseを含む。場合によっては、第1のリコンビナーゼ部位を付着させる工程は、ライゲーションを含む。場合によっては、第1のリコンビナーゼ部位および第2のリコンビナーゼ部位は、インテグラーゼ部位attPおよびattBを含む。場合によっては、アダプターは、配列決定アダプター領域をさらに含む。場合によっては、配列決定アダプター領域は、Yアダプターを含む。場合によっては、配列決定アダプター領域は、P5および／またはP7アダプターを含む。場合によっては、第1のリコンビナーゼ部位と第2のリコンビナーゼ部位は、トランスポゼース・モザイク末端を含む。場合によっては、リンカー配列は、アフィニティタグを含む。場合によっては、アフィニティタグはビオチンである。場合によっては、リンカー配列はバーコード配列を含む。場合によっては、バーコード配列は、起源となる区画を示す。場合によっては、バーコード配列は、起源となる細胞を示す。場合によっては、バーコード配列は、起源となる細胞集団を示す。場合によっては、バーコード配列は、起源となる生物を示す。場合によっては、バーコード配列は、起源となる種を示す。

【0009】

さらなる態様では、方法が提供され、該方法は、少なくとも1つの核酸結合タンパク質に複合化された核酸分子を含む、安定化された生体サンプルを得る工程、少なくとも第1のセグメントと第2のセグメントを含む複数のセグメントへと核酸分子を切断する工程、第2のセグメントに第1セグメントを付着させる工程であって、それにより、近接ライゲーションされたセグメントを作り出す、工程、近接ライゲーションされたセグメントを回復する工程、および、近接ライゲーションされたセグメントの少なくとも一部を配列決定する工程であって、ここで配列決定アダプターは、近接ライゲーションされたセグメントに回復後に付着されない、工程を含む。場合によっては、付着させる工程は、第1のセグメントを第2のセグメントにライゲーションすることにより行なわれる。場合によっては、付着させる工程は、リコンビナーゼを使用して行なわれる。場合によっては、付着させる工程は、リンカーを介して行なわれる。場合によっては、リンカーは、アフィニティタグを含む。場合によっては、アフィニティタグはビオチンである。場合によっては、方法は、(c)の付着させる工程の前に、第1のセグメントおよび第2のセグメントへのリコンビナーゼ部位を含む組み換えアダプターを付着させる工程をさらに含む。場合によっては、組み換えアダプターは、配列決定アダプターを含む。場合によっては、配列決定アダプターは、Yアダプターを含む。場合によっては、配列決定アダプターは、P5および／またはP7アダプターを含む。

【0010】

本明細書において方法が提供され、該方法は、(a)少なくとも1つの核酸結合タンパク質に複合化された核酸分子を含む、安定化された生体サンプルを得る工程、(b)核酸分子を複数のセグメントへと切断するために、安定化された生体サンプルをNaseに接触させる工程、(c)1つの接合部にて、複数のセグメントの第1のセグメントと第2のセグメントを付着させる工程、および、(d)複数の選択されたセグメントを得るために、複数のセグメントをサイズ選択にかける工程を含む。場合によっては、複数の選択されたセグメントは、約145～約600bpである。場合によっては、複数の選択されたセグメントは約100～約2500bpである。場合によっては、複数の選択されたセグメントは約100～約600bpである。場合によっては、複数の選択されたセグメントは約600～約2500bpである。場合によっては、方法は、工程(d)に先立って、

複数のセグメントから配列決定ライブラリを調整する工程をさらに含む。場合によっては、方法は、サイズ選択されたライブラリを得るために、配列決定ライブラリをサイズ選択にかける工程をさらに含む。場合によっては、サイズ選択されたライブラリは、約350bpから1000bpの間のサイズにある。場合によっては、サイズ選択は、ゲル電気泳動、キャピラリー電気泳動、サイズ選択ビーズ、またはゲルろ過カラムを用いて行なわれる。場合によっては、方法は、1つのQC値を得るために、複数の選択されたセグメントを解析する工程をさらに含む。場合によっては、QC値は、工程(d)の前の、100から2500bpの間のサイズのセグメントの割合に基づくクロマチン消化効率(CDE)である。場合によっては、方法は、CDE値が少なくとも65%であるときに、さらなる解析のためにサンプルを選択する工程をさらに含む。場合によっては、QC値は、工程(d)の前の、モノヌクレオソーム・サイズのセグメントの数のジヌクレオソーム・サイズのセグメントの数に対する比に基づくクロマチン消化指標(CDI)である。場合によっては、方法は、CDI値が-1.5より大きく、かつ1未満であるとき、さらなる解析のためにサンプルを選択する工程をさらに含む。場合によっては、安定化された生体サンプルは、安定化された細胞溶解物を含む。場合によっては、安定化された生体サンプルは、安定化された無傷細胞を含む。場合によっては、安定化された生体サンプルは、安定化された無傷の核を含む。場合によっては、工程(b)は、無傷細胞あるいは無傷の核の溶解に先立って行なわれる。場合によっては、方法は、工程(c)に先立ち、安定化された生体サンプル中の細胞および／または核を溶解する工程をさらに含む。場合によっては、安定化された生体サンプルは3,000,000未満の細胞を含む。場合によっては、安定化された生体サンプルは1,000,000未満の細胞を含む。場合によっては、安定化された生体サンプルは100,000未満の細胞を含む。場合によっては、安定化された生体サンプルは10μg未満のDNAを含む。場合によっては、安定化された生体サンプルは1μg未満のDNAを含む。場合によっては、DNaseはDNase Iである。場合によっては、DNaseはDNase IIである。場合によっては、DNaseはマイクロコッカスヌクレアーゼである。場合によっては、DNaseは、DNase I、DNase II、およびマイクロコッカスヌクレアーゼの1つ以上から選択される。場合によっては、安定された生体サンプルは、橋架剤で処置されている。場合によっては、橋架剤は化学的固定剤である。場合によっては、化学的固定剤はホルムアルデヒドを含む。場合によっては、化学的固定剤はソラレンを含む。場合によっては、化学的固定剤はグルタル酸ジスクシンイミジル(DSG)を含む。場合によっては、化学的固定剤はエチレングリコールビス(コハク酸スクシンイミジル(succinimidyl succinate))(EGS)を含む。場合によっては、橋架剤は紫外線である。場合によっては、安定化された生体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。場合によっては、方法は、選択された複数のセグメントを抗体に接触させる工程をさらに含む。場合によっては、付着させる工程は、ビオチンをタグ付けされたヌクレオチドを使用して粘着末端を埋め、および平滑末端をライゲーションする工程を含む。場合によっては、付着させる工程は、少なくとも第1のセグメントと第2のセグメントを少なくとも1つの架橋オリゴヌクレオチドに接触させることを含む。場合によっては、架橋オリゴヌクレオチドはバーコード配列を含む。場合によっては、付着させる工程は、少なくとも第1のセグメントと第2のセグメントを複数の架橋オリゴヌクレオチドに連続して接触させることを含む。場合によっては、前記付着させる工程は、生体サンプルの、細胞、核、染色体、または安定化された核酸分子が架橋オリゴヌクレオチドの固有の配列を受け取ることを結果としてもたらす。場合によっては、付着させる工程は、少なくとも第1のセグメントと第2のセグメントをバーコードに接触させることを含む。場合によっては、方法は剪断する工程を含まない。場合によっては、方法は、(e)第1のリードペアを生成するために、接合部の両側で少なくともいくつかの配列を獲得する工程をさらに含む。場合によっては、方法は、(f)コンティグの1セットに第1のリードペアをマッピングする工程、および、(g)順序および／またはゲノムへの配向を表現するコンティグのセットにわたる経路を決定する工程をさらに含む。場合によっては、方法は、(f)コンティグの1セットに

10

20

30

40

50

第1のリードペアをマッピングする工程、および、(g) 前記コンティグのセットから、安定化された生体サンプルにおける構造的変異体の存在またはヘテロ結合性の減少を判定する工程をさらに含む。場合によっては、方法は、(f) コンティグの1セットに第1のリードペアをマッピングする工程、および、(g) コンティグのセットにおける変異体をフェーズに割り当てる工程をさらに含む。場合によっては、変異体はヒト白血球抗原 (HLA) 変異体である。場合によっては、変異体はキラー細胞免疫グロブリン様受容体 (KIR) 変異体である。場合によっては、方法は、(f) コンティグの1セットに第1のリードペアをマッピングする工程、(g) コンティグのセットから、安定化された生体における変異体の存在を判定する工程、および、(h) (1) 安定化された生体サンプルについて疾患段階、予後、あるいは処置の方針を確認するステップ、(2) 変異体の存在に基づいて、薬物を選択するステップ、または、(3) 安定化された生体サンプルに対する薬効を確認するステップから選択される1つ以上のステップを行なう工程をさらに含む。

【0011】

付加的な態様では、方法が提供され、該方法は、(a) 少なくとも1つの核酸結合タンパク質に複合化された核酸分子を含む、安定化された生体サンプルを得る工程、(b) 核酸分子を複数のセグメントへと切断するために、安定化された生体サンプルをマイクロコックスヌクレアーゼ (MNase) に接触させる工程、および、(c) 1つの接合部にて、複数のセグメントの第1のセグメントと第2のセグメントを付着させる工程。場合によっては、本明細書における方法は、(d) 複数の選択されたセグメントを得るために、複数のセグメントをサイズ選択にかける工程をさらに含む。場合によっては、複数の選択されたセグメントは、約145~約600bpである。場合によっては、複数の選択されたセグメントは約100~約2500bpである。場合によっては、複数の選択されたセグメントは約600~約2500bpである。場合によっては、本明細書における方法は、工程(d)に先立って、複数のセグメントから配列決定ライブラリを調整する工程をさらに含む。場合によっては、本明細書における方法は、サイズ選択されたライブラリを得るために、配列決定ライブラリをサイズ選択にかける工程をさらに含む。場合によっては、サイズ選択されたライブラリは、約350bpから1000bpの間のサイズにある。場合によっては、サイズ選択は、ゲル電気泳動、キャピラリー電気泳動、サイズ選択ビーズ、またはゲルろ過カラムを用いて行なわれる。場合によっては、方法は、1つのQC値を得るために、複数の選択されたセグメントを解析する工程をさらに含む。場合によっては、QC値は、工程(d)の前の、100から2500bpの間のサイズのセグメントの割合に基づくクロマチン消化効率(CDE)である。場合によっては、方法は、CDE値が少なくとも65%であるときに、さらなる解析のためにサンプルを選択する工程をさらに含む。場合によっては、QC値は、工程(d)の前の、モノヌクレオソーム・サイズのセグメントの数のジヌクレオソーム・サイズのセグメントの数に対する比に基づくクロマチン消化指標(CDI)である。場合によっては、方法は、CDI値が-1.5より大きく、かつ1未満であるとき、さらなる解析のためにサンプルを選択する工程をさらに含む。場合によっては、安定化された生体サンプルは、安定化された細胞溶解物を含む。場合によっては、安定化された生体サンプルは、安定化された無傷細胞を含む。場合によっては、安定化された生体サンプルは、安定化された無傷の核を含む。場合によっては、工程(b)は、無傷細胞あるいは無傷の核の溶解に先立って行なわれる。場合によっては、本明細書における方法は、工程(c)に先立ち、安定化された生体サンプル中の細胞および／または核を溶解する工程をさらに含む。場合によっては、安定化された生体サンプルは3,000,000未満の細胞を含む。場合によっては、安定化された生体サンプルは1,000,000未満の細胞を含む。場合によっては、安定化された生体サンプルは100,000未満の細胞を含む。場合によっては、安定化された生体サンプルは10μg未満のDNAを含む。場合によっては、安定化された生体サンプルは1μg未満のDNAを含む。場合によっては、安定化された生体サンプルは、DNaseでさらに処置されている。場合によっては、DNaseはDNase Iである。場合によっては、DNaseはDNase

10

20

30

40

50

IIである。場合によっては、DNaseは、DNase IおよびDNase IIの1つ以上から選択される。場合によっては、安定された生体サンプルは、橋架剤で処置されている。場合によっては、橋架剤は化学的固定剤である。場合によっては、化学的固定剤はホルムアルデヒドを含む。場合によっては、化学的固定剤はソラレンを含む。場合によっては、化学的固定剤はグルタル酸ジスクシンイミジル（DSG）を含む。場合によっては、化学的固定剤はエチレングリコールビス（コハク酸スクシンイミジル（succinimidyl succinate））（EGS）を含む。場合によっては、橋架剤は紫外線である。いくつかの実施形態では、安定化された生体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。場合によっては、本明細書における方法は、選択された複数のセグメントを抗体に接触させる工程を含む。場合によっては、付着させる工程は、ビオチンをタグ付けされたヌクレオチドを使用して粘着末端を埋め、および平滑末端をライゲーションする工程を含む。場合によっては、付着させる工程は、少なくとも第1のセグメントと第2のセグメントを少なくとも1つの架橋オリゴヌクレオチドに接触させることを含む。場合によっては、架橋オリゴヌクレオチドはバーコード配列を含む。場合によっては、付着させる工程は、少なくとも第1のセグメントと第2のセグメントを複数の架橋オリゴヌクレオチドに連続して接触させることを含む。場合によっては、前記付着させる工程は、生体サンプルの、細胞、核、染色体、または安定化された核酸分子が架橋オリゴヌクレオチドの固有の配列を受け取ることを結果としてもたらす。場合によっては、付着させる工程は、少なくとも第1のセグメントと第2のセグメントをバーコードに接触させることを含む。場合によっては、方法は剪断する工程を含まない。場合によっては、本明細書における方法は、（e）第1のリードペアを生成するために、接合部の両側で少なくともいくつかの配列を獲得する工程をさらに含む。場合によっては、本明細書における方法は、（f）コンティグの1セットに第1のリードペアをマッピングする工程、および、（g）順序および／またはゲノムへの配向を表現するコンティグのセットにわたる経路を決定する工程をさらに含む。場合によっては、本明細書における方法は、（f）コンティグの1セットに第1のリードペアをマッピングする工程、および、（g）前記コンティグのセットから、安定化された生体サンプルにおける構造的変異体の存在またはヘテロ結合性の減少を判定する工程をさらに含む。場合によっては、本明細書における方法は、（f）コンティグの1セットに第1のリードペアをマッピングする工程、および、（g）コンティグのセットにおける変異体をフェーズに割り当てる工程をさらに含む。場合によっては、変異体はヒト白血球抗原（HLA）変異体である。場合によっては、変異体はキラー細胞免疫グロブリン様受容体（KIR）変異体である。場合によっては、本明細書における方法は、（f）コンティグの1セットに第1のリードペアをマッピングする工程、および、（g）コンティグのセットから、安定化された生体における変異体の存在を判定する工程、および、（h）（1）安定化された生体サンプルについて疾患段階、予後、あるいは処置の方針を確認するステップ、（2）変異体の存在に基づいて、薬物を選択するステップ、または、（3）安定化された生体サンプルに対する薬効を確認するステップから選択される1つ以上のステップを行なう工程をさらに含む。

【0012】

引用による組み込み

本明細書で言及されるすべての出版物、特許、および特許出願は、あたかも個々の出版物、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。

【図面の簡単な説明】

【0013】

特許出願のファイルは、色つきで作成された少なくとも1つの図面を含む。カラーの図面を有するこの特許出願のコピーは、必要な料金の請求および支払い後に当該事務局によって提供される。

【0014】

本発明の特徴および利点のより良い理解は、本発明の原理が用いられる例示的实施形態

10

20

30

40

50

を説明する以下の詳細な説明と、以下の添付図面とを引用することによって得られるであろう。

【0015】

【図1A】処理が不十分な安定化された組織サンプルを例示する。

【図1B】処理が十分な安定化された組織サンプルを例示する。

【図2】定量化および断片サイズ解析を例示する。

【図3】本開示の様々な実施形態による例示的なコンピュータシステムの様々な構成要素を説明する。

【図4】本開示の様々な実施形態に関して使用され得る例示的なコンピュータシステムのアーキテクチャを説明するブロック図である。

【図5】本開示の様々な実施形態に関して使用され得る例示的なコンピュータ・ネットワークを説明するダイアグラムである。

【図6】本開示の様々な実施形態に関して使用され得る別の例示的なコンピュータシステムのアーキテクチャを例示するブロック図である。

【図7】MNase-Cと比較された、DNase-Cについてのリードペア分離の分布のグラフを示す。

【図8】染色体1について計算された、MNase-Cと比較された、DNase-Cについてのリンケージ距離の累積分布のグラフを示す。

【図9】MNase消化の様々な時間および条件について高占有率CTCF結合部位のまわりの相対的なリードカバレッジのグラフを示す。

【図10】MNase消化の様々な時間および条件について、MNaseで処置されたサンプルの消化パターン、および、各サンプル中のモノヌクレオソーム対ジヌクレオソームの計算された比率を示す。

【図11】the University of California, Santa Cruz (UCSC) Genome BrowserによるEncyclopedia of DNA Elements (ENCODE)において報告されたピークと比較して、ChIP-seqおよびHiChIPの結果を示す。

【図12】HiChIPサンプルについて、CTCF結合部位のまわりの相対的なリードカバレッジを示す。

【図13】(図11中に示された)標的タンパク質に関するリードのピルアップを示したリードカバレッジのグラフ上、および遺伝子注釈のグラフ上に提示された、リードペアについてのコンタクトマップを示す。

【図14】図11のENCODEピークに対するMNase HiChIPの結果の同じ比較を示しているが、サンプルは同日および後日における複製である。

【図15】本開示の様々な実施形態により各DNAセグメントをリンクさせるために架橋オリゴヌクレオチドを使用する、改善された近接ライゲーションのための例示的なワークフローを示す。

【図16】本開示の様々な実施形態による、スプリッティング(splitting)およびプーリング(pooling)・アプローチを使用する、例示的なワークフローを示す。

【図17】本開示の様々な実施形態による、スプリッティングおよびプーリング・アプローチを使用する、例示的なワークフローを示す。

【図18】本開示の様々な実施形態による、スプリッティングおよびプーリング・アプローチから結果として生じるバーコードと架橋の組み合わせの例を示す。

【図19A】図19A～図19Dは、例示的なインテグラーゼ活性の工程を示し、図19Aでは、phiC31プロファージDNAがストレプトマイセス エスピー中へと組み込まれることが示される。

【図19B】図19A～図19Dは、例示的なインテグラーゼ活性の工程を示し、図19Bでは、インテグラーゼが細菌ゲノム中のファージattB部位およびattP配列へ結合し、および鎖交換をトリガーすることが示される。

10

20

30

40

50

【図 19 C】図 19 A～図 19 D は、例示的なインテグラーゼ活性の工程を示し、図 19 C では、組み込みが 3 塩基短い 2 つの新規な配列 a t t L および a t t R へと解決することが示される。

【図 19 D】図 19 A～図 19 D は、例示的なインテグラーゼ活性の工程を示し、図 19 D では、環状リンカーは組み込みに必要とされないことが示される。

【図 20】組み込み部位の内部露出末端への送達の例として、D N a s e に消化されたクロマチンへのアダプターライゲーションによる a t t P 送達を描写する。

【図 21】例えば、区画を識別するセグメントを使用して、33 塩基の a t t B セグメントが両側にある配列を置き換えることができることを示す。

【図 22】a t t B 線状 DNA の組み込みが集合内ライゲーションを引き起こすことを示す。

10

【図 23 A】図 23 A～図 23 D は、ライブラリ調整アプローチを描写し、図 23 A では、a t t B を含有している分子にビオチンが組み込まれ、一方、a t t P アダプターがホスホチオレート化された (p h o s p h o - t h i o l a t e d) ヌクレオチドを運んでいることが示される。

【図 23 B】図 23 A～図 23 D は、ライブラリ調整アプローチを描写し、図 23 B では、ストレプトアビジンのプルダウンがビオチンを含有している分子だけをプルダウンするために使用されることが示される。

【図 23 C】図 23 A～図 23 D は、ライブラリ調整アプローチを描写し、図 23 C では、組み込まれた分子だけを増幅するために a t t P 特異的な増幅が使用されることが示される。

20

【図 23 D】図 23 A～図 23 D は、ライブラリ調整アプローチを描写し、図 23 D では、交互にあるいは組み合わせて、組み込まれていない分子およびアダプターライゲーションされていない核酸を除去するために、エキソヌクレアーゼ活性が使用されることが示される。

【図 24】近接ライゲーション事象の単核タグ付けのために組み込みを使用することが描写される。

【図 25】リコンビナーゼ（例えば、インテグラーゼ）に基づく近接ライゲーションについての例示的な概要を示し、（図の一番上から一番下へ）断片化（例えば、D N a s e を用いるなど、酵素学的消化）、末端研磨および A テーリング、組み換え部位（例えば、A テール適合 a t t B 部位）のライゲーション、リンカー（例えば、a t t P リンカー）を用いた組み換え、および架橋反転が含まれており、ゲノムの様々な領域から近接ライゲーションされた核酸がもたらされている。

30

【図 26】リコンビナーゼに基づく近接ライゲーションのための例示的な核酸配列を示し、（図の一番上から一番下へ）非組み換え g DNA にライゲーションされた例示的な E P オーバーハング a t t B 部位、非組み換え g DNA にライゲーションされた例示的な a t t B 部位、a t t P 部位を有する例示的な非組み換えビオチン・リンカー、g DNA 上の a t t B 部位に組み換えられたリンカーの 1 つの末端、および g DNA 上の a t t B 部位に組み換えられたリンカーの両方の末端が含まれている。

【図 27】配列決定アダプター部位（例えば、P 7 および P 5 配列決定 Y アダプター）を有する a t t B 部位のための例示的な設計および核酸配列を、単独で（上）および非組み換え g DNA にライゲーションされた状態で（下）示す。

40

【発明を実施するための形態】

【0016】

一態様では、長距離かつ構造的なゲノム情報を含むゲノム配列の判定、細胞における核酸の物理的な立体配座の判定に関連するとともに、当該技術分野において既に開示された方法よりも結果が改善された核酸に対して非常に長距離のリードペアを生成するための、組成物、システム、および方法が本明細書に提供される。本明細書の方法は、D N a s e 消化、ミクロコッカスヌクレアーゼ (M N a s e) 消化、リコンビナーゼ処置、サイズ選択、Q C 制御、全細胞あるいは全核ヌクレアーゼ消化、単細胞解析、低投入量要件を含む

50

がこれらに限定されない技術を利用して最適な結果を達成することができる。本明細書の方法は、さらに、核酸サンプル中の抗体結合部位に対してオリゴヌクレオチド、あるいはオリゴヌクレオチドおよびヌクレアーゼを標的とするために、免疫グロブリン結合タンパク質またはその断片を利用することを含み得る。本明細書において、さらに、HiChIP、HiChIRP、およびメチルHiCの改善された方法が提供される。

【0017】

別の態様において、単細胞あるいは複数の細胞、または細胞集団のための、核酸立体配座評価、核酸配列解析、または核酸フェーズ情報判定に関する実施形態が本明細書で提供される。

【0018】

場合によっては、立体配座が保存された、あるいは立体配座が再構築された核酸サンプルは、断片化され得、および、アリコート識別配列セグメントが添加され得るアリコートまたは区画に分配され得、その結果、サンプルから生成されたペアエンド・ライブラリの解析に際して、ペアエンド (paired end) が、起源となる、区画、または細胞に、割り当てられ得る。こうして、配列および／または三次元の核酸構造における細胞特異的な変異が判定され得る。

【0019】

核酸立体配座評価

本明細書において、第2の細胞または細胞の集団の物理的な立体配座から識別可能な単細胞または細胞集団などの、細胞における核酸の物理的な立体配座の判定に関する組成物、システム、および方法が開示される。本明細書における開示の実施を通じて、三次元の核酸相対位置を示す核酸分子が生成され得、および、複数の分子について、共通の起源となる細胞または細胞集団を識別するためのタグ（例えば、核酸バーコード）と共に、随意に提供され得る。

【0020】

本明細書における開示された方法の実施を通じて、核酸は、細胞中の三次元構造のすべてまたは少なくともいくらかを保存するように、採取され得る。そのような核酸の、露出された核酸ループは、物理的に近接した末端同士が互いに付着（近接付着）される可能性が高くなるように、無作為に再付着される内部セグメントの末端を露出するように切断され得る。従って、どの露出した末端が互いに付着されるようになるかを判定することにより、生来の細胞構造における末端が隣接した核酸の物理的近接性を示すデータを得ることができる。

【0021】

関連するアプローチは、例えば、2016年9月6日に公開されたDekker et al. によるUS9434985B2において開示され、参照によりその全体が本明細書に組み込まれる。

【0022】

本明細書における開示された方法の実施を通じて、ペアードエンド・ライブラリーの構成要素は、さらにタグ付けされるか、または起源となる細胞を示す配列情報を与えられ、その結果、集団の個別の細胞の間の立体配座の違いが細胞の集団について容易に識別されるか、または、細胞の第1の集団と細胞の第2の集団の間の立体配座の違いが、それらが同時に解析されるときでさえ、容易に識別される。タグは例えば、核酸バーコードを含むことができる。場合によっては、タグは、ゲノム中で隣接しない2つの核酸セグメントの間に接合部を含むことができる。核酸分子は、完全にまたは部分的に配列決定された時、そのゲノム座に個々のゲノムの末端をマッピングするのに十分な少なくともいくらかのゲノム配列が多くの場合で得られるように、および、正確なあるいは有望な、起源となる細胞または細胞集団を同定するのに十分なタグ付け配列またはリンク付け配列がさらに得られるように、生成することができる。このように、ゲノムの2つの領域が互いに物理的に近接していることを示す配列情報が得られ、また一方で、この物理的な立体配座が生じる細胞または細胞集団を示す情報も得られ、その結果、その細胞または細胞集団の中で同時

10

20

30

40

50

に生じる他の物理的な立体配座情報のコンテキストにおいて、その配列情報を評価することができる。

【0023】

細胞中のゲノム核酸または他の核酸は、安定化させることができ、また、真核細胞については、本明細書に組み込まれるか、その他に既知の方法などの、当該技術分野において知られている方法に従って、核は随意に単離することができる。例えば、図1Aと図1Bでは、処理された安定化された組織サンプルが例示される。図1Aは、処理が不十分な組織サンプルを例示する。図1Bは、処理が十分な安定化された組織サンプルを例示する。

【0024】

本明細書の開示と一貫する核酸は、原核生物の初代ゲノムまたはプラスミド核酸、真核生物の核、ミトコンドリアのまたは色素体の核酸などの細胞の核酸、あるいは、場合により、rRNA、mRNA、などの細胞質の核酸、あるいは、ウィルスまたは他の病原体、または他のサンプル外因性の核酸などの、サンプル中の外因性の核酸を、任意の数で含む。

【0025】

安定化された核酸は、場合によっては、少なくともいくつかの核酸が個々の区画へ分配されるように、分配することができる。例示的な区画としてウェル、エマルジョン中の液滴、または表面位置（例えば、アレイ・スポット、ビーズなど）が挙げられ、それらの区画は本明細書に別記されるように差動的に配列決定されたリンカー分子の別個のパッチを含んでいる。当技術分野において既知か、当業者に利用可能であるさらなる区画も企図され、本明細書に開示される方法、組成物、およびシステムに一貫している。

【0026】

安定化された核酸は、後に再結合して特定の細胞についての核酸構造情報を得るために内部切断を露出するように断片化することができる。多くの断片化アプローチが当技術分野において知られており、および本明細書の開示と一貫する。核酸は、制限エンドヌクレアーゼの1以上の集団、ガイドRNAに連結されたCRISPR/Cas分子などのプログラム可能なエンドヌクレアーゼ、非特異的エンドヌクレアーゼ（例えば、DNase）、タグメンテーション（tagmentation）、剪断、音波破碎、加熱、または他の手段を使用して断片化することができる。場合によっては、DNaseは非配列特異的である。場合によっては、DNaseは、一本鎖DNAおよび二本鎖DNAの両方に対して活性がある。場合によっては、DNaseは、二本鎖DNAに対して特異的である。場合によっては、DNaseは、二本鎖DNAに対して優先的である。場合によっては、DNaseは、一本鎖DNAに対して特異的である。場合によっては、DNaseは、一本鎖DNAに対して優先的である。場合によっては、DNaseはDNase Iである。場合によっては、DNaseはDNase IIである。場合によっては、DNaseは、DNase IおよびDNase IIの1つ以上から選択される。場合によっては、DNaseはマイクロコッカスヌクレアーゼである。場合によっては、DNaseは、DNase I、DNase II、およびマイクロコッカスヌクレアーゼの1つ以上から選択される。他の適切なヌクレアーゼもまたこの開示の範囲内にある。

【0027】

特に、2014年8月7日に公開されたGreen et al. のWO2014121091A1の開示は（後に、2015年12月17日にUS20150363550A1として公開され、2018年10月2日にUS10089437B2として発行された）は、その全体が本明細書に組み込まれる。同様に、2016年2月4日に公開されたFields et al. のWO2016019360A1の開示は（後に、2017年11月23日にUS20170335369A1として公開された）、その全体が本明細書に組み込まれる。同様に、2017年8月31日に公開されたGreen et al. のWO2017147279A1の開示は、その全体が本明細書に組み込まれる。

【0028】

核酸は付着の前または後に、表面に結合され得る。例示的な表面は、ビーズ、アレイ、ウェルを含むが、これらに限定されない。場合によっては、表面は、固相可逆的固定（S

10

20

30

40

50

PRI) ビーズなどのSPRI表面である。付着に先立って表面に核酸を結合することは、染色体間のライゲーションまたは付着を減少させる、および、染色体内のライゲーションまたは付着を増加させるなど、下流工程のパフォーマンスを改善することができる。

【0029】

核酸は付着の前または後に、免疫沈降される場合がある。そのような方法は、クロマチンを断片化し、次に断片を、アセチル化ヒストン、具体的にはH3を特異的に認識して結合する抗体に接触させることを含み得る。そのような抗体の例は、限定されないが、Upstate Biotechnology, Lake Placid, N.Y. から入手可能な抗アセチル化ヒストンH3を含む。免疫沈降由来のポリヌクレオチドは、その後免疫沈降から集めることができる。同様の標的化された濃縮方法はまた、限定されないが、アダプター、オリゴヌクレオチド、または他の核酸プローブを含む標的特異的、および、核酸にガイドされるヌクレアーゼ（例えば、触媒現象的に不活発かまたは「死んでいる」ヌクレアーゼを含むCas9などのCasファミリー酵素）な化合物と共に使用される。

【0030】

バーコード、区画特異的配列、または区画識別配列を有しているリンクさせる核酸などの、リンクさせる核酸は、区画特異的または区画識別する配列（例えば、核酸バーコード）をしばしば有する左ゲノムセグメントと、リンクさせる領域、右ゲノムセグメントとを有する核酸セグメントを生成するように露出した内部端末に付着され得、ここで、左ゲノムセグメントと右ゲノムセグメントは、供給源細胞内で物理的に近接しているゲノムセグメントに対してマッピングされる。

【0031】

露出した核酸端末の付着に先立って、端末は処理され得る。そのような処理は、末端研磨（end polishing）または末端平滑化（blunt ending）を含む。末端平滑化された核酸端末は、例えば、他の平滑化された露出した核酸末端、アダプター、またはリンカーに、直接ライゲーションされ得る。そのような処理は、例えば、テーリング（例えば、Aテーリングあるいはアデニル化）によってオーバーハングを生成することを含み得る。一例では、オーバーハングは、1ヌクレオチドのサイズである。一実施例では、オーバーハングは、単一のAヌクレオチドである。テーリングされた露出した核酸端末は、例えば、他のテーリングされた露出した核酸端末に、またはアダプターあるいはリンカーに、直接ライゲーションされ得る。場合によっては、末端平滑化あるいはテーリングは、ビオチン化された核酸などのアフィニティタグ付けされた核酸を組み込み得る。アフィニティタグは、例えば、下流での捕捉または濃縮工程において、使用され得る。他の場合では、末端平滑化あるいはテーリングは、アフィニティタグ付けされた核酸を組み込むことなく（例えば、ビオチン化された核酸なしで）実行され得る。アフィニティタグは、所望の場合、例えば、アダプターあるいはリンカー（例えば、架橋）に、続けて追加することができる。一例では、露出した核酸は、末端研磨され、オーバーハングが生成され、および、露出した末端は、架橋オリゴを介して付着される。

【0032】

付着は、例えば、ライゲーションにより、直接的である。

【0033】

付着は、1つの露出した核酸末端を別のものに接続する1つ以上のリンカーまたは架橋核酸のライゲーションなどの、リンカーまたは架橋による場合がある。

【0034】

付着は、インテグラーゼまたはトランスポサゼ組み込みなどの、リコンビナーゼ組み込みと一貫性があるような、キャッピング核酸アダプター・セグメントの使用による場合がある。リコンビナーゼ部位を有するアダプターは、露出した核酸端末に付加することができ、および、その後、それらの末端は、例えば、組み換えによって、接続され得る。

【0035】

phic31インテグラーゼ・バーコード送達を例に取ると、細胞識別リンカーまたは細胞特異的リンカー（例えば、核酸バーコード）などのリンカーは、以下のように、酵素

10

20

30

40

50

的に付加することができる。

【0036】

内部核酸末端の露出の後に続いて、インテグラーゼ部位は、内部末端またはテロメアが除去された末端などの、露出した線状染色体末端などの露出した核酸末端にライゲーションすることができる。例示的な組み込み部位は、a t t P p h i C 3 1 インテグラーゼ組み込み部位、または a t t P 組み込み部位を含む核酸であるが、他の組み込み部位も本明細書の開示と一貫する。ライゲーションは結果として核酸断片の集団をもたらし、そのうち少なくともいくつかは、a t t P セグメントを含むセグメントなどの、組み込み部位のそばの各末端に接した細胞核酸セグメントを個々に含む。様々な実施形態において、断片化と組み込み部位の付着の一方または両方のどちらかは、区画化に先立って生じ、または、断片化および組み込み部位の付着の1つまたは両方のどちらかは、区画化の後に続く。

10

【0037】

図19A～図19Dは、p h i C 3 1のインテグラーゼに基づく付着アプローチの例示的な概要を示す。図19Aでは、インテグラーゼを介しての、ストレプトマイシズ中への p h i C 3 1 の組み込みの概要が示される。a t t P（破線によって明示される）および a t t B（実線によって明示される）部位を含む核酸が示されるが、a t t B と a t t P 以外の様々な具体化部位、およびインテグラーゼ以外の酵素活性もまた企図され、および本明細書の開示と一貫する。図19Bでは、インテグラーゼおよび関連付けられるタンパク質（円で明示される）が細菌ゲノム中のファージ a t t B 部位および a t t P 配列へ結合し、および鎖交換をトリガーすることが示される。図19Cでは、組み込み事象の結果が示される。組み込みは、a t t B および a t t P が欠如した線状の核酸へと解決するが、a t t B と a t t P の部分のキメラ断片である a t t L および a t t R を有している。a t t L および a t t R 部位は、a t t B および a t t P に比較して、3 b p 短く、および配列において異なる。図19Dでは、環状組み込みまたはリンカーゲノムが必要とされないことが示される。a t t B 部位を包含している線状DNAの組み込みは、DNAを包含している a t t P の切断を引き起こすことになる。

20

【0038】

図20は、本明細書に企図されるような安定した核酸の露出した内部末端へのインテグラーゼ部位の送達を描写する。例えば、a t t P は、D N a s e に消化されたクロマチン（シリンダーによって明示される）の露出した内部末端上へのアダプターライゲーションによって送達することができる。核酸は、ヌクレオソームなどの結合部位との接触を保護するように、場合によっては、フェーズ情報あるいは3次元の物理的位置を保護するように、安定化させることができる。

30

【0039】

図21は、a t t B 部位などのインテグラーゼ部位を使用するリンカー構築物の生成を示す。例えば、最小限の33のヌクレオチドの a t t B 部位は組み込みに十分である。ランキング配列は、バーコードあるいは特別の源（例えば、細胞、液滴、あるいは他の区画、組織）の核酸を指定する他の配列などの選択の配列を使用して置き換えることができる。図22は、a t t B の線状のDNAの組み込みによる集合体内ライゲーションを実証する。結果は、同相だったか、単一ライブラリの構成要素上で接合されている物理的近接中にあった核酸セグメントの露出した内部末端がある線状分子である。ライブラリ構成要素は無傷の組み込み部位（この場合 a t t P）によって制限されるが、内部組み込み部位は a t t R および a t t L の境界によって破壊され、および置換され、その結果、a t t P に関連付けられるプライマーがライブラリ断片を増幅し得る。内部末端に隣接した配列を得ること、およびゲノムあるいはコンティグのセットにそれをマッピングすることによって、細胞内の共通のフェーズまたは共通の三次元位置にコンティグあるいはゲノムのセグメントを割り当てることができる。

40

【0040】

図25は、組み換えに基づく近接ライゲーション・プロトコルの別の例を示す。架橋されたクロマチンを含むゲノムDNAは、例えば、D N a s e で、消化される。露出した端

50

末は、研磨され、例えば、単一のA塩基オーバーハングでAテーリングされる。a t t B部位などのAテールが適合するアダプターを含むリコンビナーゼ部位は、露出したエンドヘライゲーションされる。a t t P部位などの対応するリコンビナーゼ部位を有するリンカーは、サンプルに接触され、および、リコンビナーゼ酵素（例えば、p h i C 3 1）を用いて組み換えが実行され、近接ライゲーションが達成される。リンカーは、下流ブルダウンまたは他の精製あるいは処理を可能にするために、ビオチン（b）などの親和剤を随意に含む。架橋は逆転され、および、近接ライゲーションされた核酸が回収され、例えば、40 b pまでのa t t B部位、次に150 b pまでのゲノムDNA領域1、次に90 b pまでのa t t R部位および親和剤を含むリンカー配列、次に150 b pまでのゲノムDNA領域2、次に40 b pまでのa t t B部位が含まれる。

【0041】

図26は、図25において示されたものに類似するプロトコルを示しており、例示的なアダプターおよびリンカー配列を伴っている。一番上に、EPオーバーハングa t t Bアダプターを有する非組み換えg DNAが示されており、その配列G T G C C A G G G C G T G C C C t t G G G C T C C C C G G G C G C G A T Cは、a t t B部位G C C C T T G G G Cを有し、その相補配列C G C G C C C G G G G A G C C C a a G G G C A C G C C C T G G C A Cは、逆a t t B部位G C C C A A G G G Cを有する。上から2番目には、a t t Bアダプターを有する非組み換えg DNAが示されており、その配列はG T G C C A G G G C G T G C C C t t G G G C T C C C C G G G C G C G T C C C Cであり、および相補配列はG G G G A C G C G C C C G G G G A G C C C a a G G G C A C G C C C T G G C A Cである。上から3番目には、a t t P部位およびビオチンを含む非組み換えリンカーが示されており、その配列g g a g C C C C A A C T G G G G T A A C C T t t G A G T T C T C T C A G T T G G G G a c c a t g g a g a / i B i o d T / c a C C C C A A C T G A G A G A A C T C a a A G G T T A C C C C A G T T G G G G C A C T A Cは、配列A C C T T T G A G Tを有するa t t P部位と、リンカー配列C A T G G A G A T Cとを含む。上から4番目には、a t t B / g DNAと組み換えられた、リンカーの1つの末端が示され、その配列は、g g a g C C C C A A C T G G G G T A A C C T t t G A G T T C T C T C A G T T G G G G a c c a t g g a g a / i B i o d T / c a C C C C A A C T G A G A G A A C T C a a G G G C A C G C C C T G G C A Cである。最も下には、a t t B / g DNAと組み換えられた、リンカーの両方の末端が示されており、その配列G T G C C A G G G C G T G C C C t t G A G T T C T C T C A G T T G G G G a c c a t g g a g a / i B i o d T / c a C C C C A A C T G A G A G A A C T C a a G G G C A C G C C C T G G C A Cは、配列G C C C T T G A G Tを有するa t t R部位と、配列A C T C A A G G G Cを有する逆a t t R部位とを有する。

【0042】

図23Aは、ライブラリ生成を容易にするための例示的なリンカー分子、およびアダプター分子修飾を示す。リンカー分子はアフィニティタグ（この場合、ビオチンであり、円で明示される）を与えられ、一方、アダプターはエキソヌクレアーゼ耐性修飾を備えている（この場合、ホスホチオ化（p h o s p h o t h i o a t i o n）（P S）であり、星印によって明示される）。アフィニティタグは、リンカー分子の単離を、それらが末端に隣接した分子中に組み込まれているかどうかには依存せずに、容易にする。リンカー上のエキソヌクレアーゼ耐性の修飾は、リンカーが付加されなかった核酸分子、および末端に隣接する核酸サンプル分子に組み込まなかったリンカー分子に選択的な分解を容易にする。図23Bは、内部末端に付加されたライゲーション部位中に組み込まれたかどうかには依存しない、タグ付けされた分子のアフィニティー精製（この場合、ストレプトアビジンであり、半円弧によって明示される）を示す。図23Cは、a t t Pに導かれた増幅が、a t t P部位などのインテグラーゼ部位を保存するアフィニティー単離分子を選択的に増幅するために使用される（この場合、プライマーを有するなどの部位を標的とすることによる）。アフィニティタグおよびa t t P部位の存在は、成功した組み込み事象が生じた分子を示す。図23Dは、代替を示しており、それにより、エキソヌクレアーゼ（円形セク

10

20

30

40

50

タあるいは「バックマン」によって明示される)は、エキソヌクレアーゼ耐性修飾を欠いたアフィニティー標識分子を取り除くために使用される。アフィニティタグおよびエキソヌクレアーゼ耐性部位の存在は、成功した組み込み事象が生じた分子を示す。

【0043】

代替的に、Tn3、Tn5、Tn7、あるいはスリーピングビューティートランスポサゼ (sleeping beauty transposase) などのトランスポサゼが、バーコードの送達のために使用され得る。内部核酸末端の露出に続いて、モザイク末端は、内部末端、あるいはテロメアが取り除かれた末端などの露出した線状の染色体末端などの、露出した核酸末端にライゲーションすることができる。例示的なモザイク末端はTn5モザイク末端、またはTn5モザイク末端を含む核酸であるが、他のモザイク末端は本明細書の開示と一貫する。ライゲーションは結果として核酸断片の集団をもたらす、そのうち少なくともいくつかは、Tn5モザイク末端などのモザイク末端のそばの各末端に接した細胞核酸セグメントを個々に含む。

【0044】

リコンビナーゼ・アダプター分子は配列決定アダプター部位、例えば、P5、P7部位をさらに含み得る。図27(上)は、配列決定Yアダプターを有する例示的なattBアダプターを示しており、attBアダプターの配列はGTGCCAGGGCGTGCCCTtGGGCTCCCCGGGCGCGであり、P7の配列は、GATCGGAAGAGCACACGTCTGAACTCCAGTCAであり、および、P5の配列は、ACACCTCTTTCCCTACACGACGCTCTTCCGATCである。図27(下)は、配列決定アダプターを有するリコンビナーゼ・アダプターを有する非組み換えgDNAの概要を示す。配列決定するアダプターは、組み換え後にゲノムDNAと共に留まることになるattB部位の部分に付着されており、それにより、さらなる増幅またはアダプターライゲーションなしで、組み換え後の配列決定が可能になる。配列決定アダプターを含むリコンビナーゼ・アダプターの使用は、図27中に示されるものを含め、しかしこれらに限定されずに、近接ライゲーション生成物を直接配列決定することを可能にし、ここで増幅または別個のアダプター取り込みステップは必要とされない。このことは、結果として生じる配列情報において、増幅バイアスなどのバイアスを抑制し得る。

【0045】

いくつかの実施形態では、様々な実施形態において、断片化と組み込み部位の付着の一方または両方のどちらかは、区画化に先立って生じ、または、断片化および組み込み部位の付着の一方または両方のどちらかは、区画化の後に続いて生じる。図24は、インテグラゼを媒介とした集合体内ライゲーションを使用する、単細胞HiC(あるいは他の近接ライゲーション技術)のための例示的なシステムを描写する。単一の細胞核は、インテグラゼと組み合わせられて、区画の第1のセットにおいてカプセル化される。区画は、この場合、エマルジョン中の液滴である。核は、内部露出した端部を生成し、および局所的な三次元の情報を保存するように、らせん破壊にかけられる。アダプターは露出した内部末端上へとライゲーションされる。アダプターは随意にエキソヌクレアーゼ耐性の末端を含む。この実施形態では、アダプターは区画を識別する情報を伝達しない。区画の第2のセットでは、区画判別配列などの固有の分子識別子(UMI)を有するリンカーは、カプセル化され、および、随意に、増幅、および切断に導かれる線形化にかけられる。区画の第1のセットと第2セットは、ほぼ1:1の比、または、2つの細胞からの核酸が結果として単一の区画へと結合されないであろう条件でマージされる。

【0046】

インテグラゼ部位またはモザイク末端などのリコンビナーゼ部位は、場合により、内部核酸末端上へとライゲーションされることになる未修飾の単鎖または二重鎖の断片で運ばれ得る。代替的に、後の配列決定ライブラリのクリーンアップを容易にするために、attP配列などの、いくつかの単鎖または二重鎖の断片を内部に持つ組み込み部位、または、Tn3、Tn5、Tn7、またはスリーピングビューティートランスポサゼ・モザイク末端などのモザイク末端は、エキソヌクレアーゼまたは他の核酸分解活性に干渉す

る修飾などの、少なくとも1つの修飾を含み得る。例として、組み込み部位を内部に持つ二重鎖の断片が各端末に付加された断片のエキソヌクレアーゼ分解を不可能にするためのチオサルフェート修飾が挙げられる。

【0047】

多くの場合、組み込み部位などのリコンビナーゼ部位またはモザイク端末は、非特異的であり、その中で、そのような組み込み部位、またはattP配列、あるいはTn3、Tn5、Tn7、あるいはスリーピングビューティートランスポサゼ・モザイク端末などの末端の中の配列は、隣接した核酸の細胞供給源を指定するためには使用されない。代替的に、多くの場合、核酸の区画化の後、区画は、組み込み部位またはモザイク端末に隣接している区別できる、具体的な、または細胞を識別させる配列（例えば、核酸バーコード）を有するアダプターを備え、または、区別できる組み込み部位またはモザイク端末を備え、その結果、第1の区画の核酸が第1の識別させるセグメントを有する組み込みセグメントまたはモザイク端末を受け取り、第2の区画の核酸セグメントが第2の識別させるセグメントを有する組み込みセグメントを受け取る。

10

【0048】

インテグラーゼattPセグメントを含む境界などの、リコンビナーゼ境界を有する断片は、その後、共通の溶液中で、attBphiC31組み込み部位などの組み込み部位に接触され得る。一例では、組み込み酵素はphi31インテグラーゼを含む場合があり、組み込み境界はattPセグメントを含む場合があり、および、組み込み部位はattB組み込み部位を含む場合がある。代替的に、断片は、Tn3、Tn5、Tn7、またはスリーピングビューティ・トランスポサゼ・モザイク端末境界などのモザイク端末境界を有する。

20

【0049】

attB組み込み部位、またはTn3、Tn5、Tn7、あるいはスリーピングビューティ・トランスポサゼ・モザイク端末などのリコンビナーゼ部位が、セグメントまたは細胞供給源に固有のものなど（例えば、核酸バーコード）の区画または細胞を同定させる配列を有しているリンクするセグメントの脇に位置するとき、その配列は、特定のまたは共通の細胞供給源または区画から発生したものとして、隣接した細胞核酸を同定し、その結果、それらが配列決定に先立ってまたは配列決定と同時に第2の区画の断片によって大きくなったとしても、共通の細胞識別または区画識別セグメントに接合された共通の細胞由来の複数の露出した末端が、共通の細胞から発生したものとして容易に同定され得る。

30

【0050】

細胞を識別させる配列がリコンビナーゼ部位に隣接した断片を介して送達されるとき、組み込みまたは転位は、区画化の後に続いて、好ましく実行される。それにより、少なくともいくつかの区画の核酸内容は、それ自体のリンカーの、細胞を識別させる配列によって識別することができ、その結果、複数の細胞供給源を形成する核酸が配列決定によって大きなものになったとしても、内部末端ペアを、および、大部分はまたは完全に配列決定されたゲノムにわたり且つそれを含むコンティグのセットにおいてマッピングされた近辺に割り当てられた近接情報を、サンプルの少なくとも1つの他の細胞から識別された共通の細胞に、割り当てることができ、その結果、予想された核酸三次元立体配座における差異が確立され得る。

40

【0051】

組み換え部位に隣接した断片は、細胞または配列を指定する区画（例えば、核酸バーコード）を随意に含むリンカー領域によってリンクされた左境界断片および右境界断片（例えば、attB部位、または、Tn3、Tn5、Tn7、あるいはスリーピングビューティ・トランスポサゼ・モザイク端末）を様々に含む。リンカー領域は、後の単離を容易にするための部分を随意にさらに含む。多くのアフィニティタグまたは修飾塩基は、開示と本明細書に一致している。例示的な部分は、インテグラーゼ処置またはトランスポゼース処置の後にリンカーの物理的もしくは化学的な単離を容易にする。アビジンまたはストレプトアビジンに基づく単離を容易にし得る1つまたは複数のビオチンタグなどの、当

50

業者に知られている任意数のアフィニティタグも、本明細書の開示と一貫する。代替的に、インテグラーゼまたはトランスポサーゼ活性に干渉せずに、単離を容易にするあらゆる抗原、受容体、またはリガンドは、本明細書のいくつかの実施形態のために適している。

【0052】

上述されるように、いくつかのライブラリ生成アプローチは、組み込まれていない試薬を選択的に除去するステップなどの、クリーンアップのステップを含む。例えば、エキソヌクレアーゼ処置は、付着されていないリンカー分子、組み込み部位が付着されていないゲノムの断片、または、付着されていないリンカー分子、組み込み部位が付着されていないゲノムの断片の両方を選択的に除去するために、多くの場合で使用される。チオサルフェート・バックボーンなどの、エキソヌクレアーゼ耐性の修飾を有する組み込み部位断片にライゲーションされたゲノムの断片は、その末端からのエキソヌクレアーゼ分解に耐性があり、および、チオサルフェート・バックボーンなどの、エキソヌクレアーゼ耐性の修飾を有する組み込み部位断片によって両方の末端上で境界される核酸分子は、両方の末端における分解に耐性があり、エキソヌクレアーゼ処理を生き延びることができる。

【0053】

交互にまたは組み合わせで、いくつかのリンカー分子は、attP組み込み部位、またはTn3、Tn5、Tn7、あるいはスリーピングビューティー・トランスポサーゼ・モザイク末端などの組み換え部位の反対側に、カウンター・アフィニティタグ (counter-affinity tag) を含み、その結果、カウンター・アフィニティタグは、組み換え反応の成功に応じて除去される。そのような場合、望まれない試薬は、カウンター・アフィニティタグは、結合パートナーに接触させることにより除去することができる。

【0054】

インテグラーゼ活性は、attB部位およびattP部位などの両方の組み込み部位を、組み込み事象の一部として、部分的に破壊する。従って、ライゲーションされたアダプター部位にattP組み込み部位などをアニールするためのプライマーを設計することによって、単独でまたはリンカーに基づく単離と組み合わせで、少なくとも1つのリンカーにわたるクロンのアンプリコンが生成され得、その結果、細胞またはアリコートを識別させる情報および内部末端の隣接の情報が増幅され、場合によっては、配列決定または他の下流解析が容易にされる。

【0055】

ライブラリの生成および随意にライブラリのクリーンアップに続いて、核酸は、完全にまたは部分的に配列決定することができ、その結果、細胞が識別されたまたは細胞特異的な三次元的な核酸位置評価のための十分な情報が得られる。上述されるように、配列決定は、そのゲノム座にライブラリ要素の個々のゲノム末端をマッピングするのに十分な、少なくともいくつかのゲノムの配列が得られ、および、正確なあるいは有望な起源となる細胞を同定するのに十分なリンクさせる配列がさらに得られるように、好ましく実行される。従って、ゲノムの互いに物理的に近接した2つの領域を示す配列情報が得られ、また一方で、この物理的な立体配座が生じる細胞を示す情報も得られ、その結果、その細胞の中で同時に生じる他の物理的な立体配座情報のコンテキストにおいて、その配列情報を評価することができる。多くの場合、両方のアプローチ他は開示と本明細書に一致しているが、この情報は完全長の配列決定によってではなくペアエンドの配列決定によって得られる。

【0056】

第2の細胞の物理的な立体配座から識別可能な単細胞などの細胞における核酸の物理的な立体配座の判定に関する組成物および方法は、本明細書の開示と一貫する多くのシステム上で実施することができる。いくつかのシステムは、例えば、エマルジョンまたはウェル中の第1の液滴への、固定された細胞核酸物質の、ウェルプレート上における分配を含む。これらの液滴は、エキソヌクレアーゼ耐性に随意に修飾されたインテグラーゼ部位またはモザイク末端などのリコンビナーゼ部位を、並びに、本明細書に記載されるように、インテグラーゼまたはトランスポサーゼ酵素およびリガーゼ酵素をさらに含む。別途、リ

リンカー核酸分子は、エマルジョンの第1の液滴への送達のために構成することができる。リンカー核酸は、随意に第2のエマルジョンまたは第2のウェルの液滴へ分配され得、例えばローリングサークル増幅を使用して、随意に増幅され、および、第2のエマルジョン液滴について所与のリンカー分子の複数のコピーを生成するために処理される。

【0057】

その後、第2のエマルジョン液滴と第1のエマルジョン液滴は一組でマージされ得、その結果、インテグラーゼまたはトランスポサゼをインテグラーゼまたはトランスポサゼ互換性があるリンカーにライゲーションする核酸断片がアセンブルされ、多くの場合、1つの液滴につき様な標識が示される。しかしながら、特に、データ解析が液滴における1つ以上のタイプのタグの存在を示すとき、核酸サンプルにつき2つ以上の識別子がある液滴は、それでもなお、意味のあるデータを産出することができる場合がある。

10

【0058】

一組のマージの代わりとして、場合によっては、インテグラーゼまたはトランスポサゼ適合リンカーは、流れのマージに向かう液滴を介して第1のエマルジョン液滴に接触される試薬の流れの中の固体粒子のコロニーとして、送達することができ、そのことは2017年11月23日に公表されたUS20170335369A1中に記載されたように、参照によってその全体が本明細書に組込まれる。リンカー核酸は、随意に、固体粒子上で、またはゲル中で増幅することができる。第1のエマルジョン液滴は、流れにマージされ得、および、第2のエマルジョン液滴は流れをセグメント化または分割することにより回収することができ、その結果、1:1、1:1以上、または1:1未満などの、リンカー粒子に対する核酸クラスターの所望の比率が得られる。

20

【0059】

交互に、いくつかのシステムおよび方法は、上述のように増幅されないか増幅されたかのどちらの固定された細胞核酸物質を、チップまたはプレートのウェルの中へ分配し、続いて区画の中へリンカー核酸を送達することを含む。

【0060】

交互に、場合によっては、リンカー核酸の送達は、区画化から一時的に分離されない。むしろ、リンカー核酸あるいは酵素活性または酵素活性に必要な因子は、一時的に酵素活性を活性化して、核酸サンプルの露出された末端への、リンカーを介するなどの、リンカーの共有結合をもたらすために、加熱、電磁氣的活性化、または他の投与などの特定の処置まで隔離される。

30

【0061】

多くの組み込み酵素が本明細書の開示と一貫する。ThermoFisherによって商業的に入手可能なものなどのPhiC31インテグラーゼは、本明細書の方法の実施、システムのオペレーション、および組成物における使用のために、多くの利益を発揮する。このインテグラーゼのいくつかの利益は、以下に記載されるとおりである。それは小さな組み込み部位(attB / attP)を使用する。酵素は、それ自体が小さな単一のポリペプチドである。組み込みは、組み込み事象を除去するための別個の酵素を使用しなければ、不可逆的である。活性は高く、および、酵素は活性を変更するために容易に操作される。それにもかかわらず、多くの組み込みシステムが本明細書の開示と一貫するため、その使用は他の酵素を排除することを要求しない。本開示の態様は、PhiC31インテグラーゼに関して記載されるが、あらゆる互換性がある酵素の使用が企図される。多くのトランスポサゼ酵素は、本明細書の開示と一貫する。ThermoFisherによって商業的に入手可能なものなどのTn5トランスポサゼは、本明細書の方法の実施、システムのオペレーション、および組成物における使用のために、多くの利益を発揮する。このトランスポサゼのいくつかの利益は、以下に記載されるとおりである。Tn5は19bpのモザイク末端認識配列を使用し、挿入は偏りをほとんど有しておらず安定しており、および、Tn5は、インヴィヴォの転位のために細胞へと、またはインビトロ反応のために単離した核酸へと、送達することができる。それにもかかわらず、Tn3、Tn7、スリーピングビューティー・トランスポゼースなどの多くのトランスポゼース・システ

40

50

ムは、本明細書の開示と一貫するため、その使用は他の酵素を排除することを要求しない。本開示の態様は、T n 3、T n 5、T n 7、あるいはスリーピングビューティー・トランスポサゼに関して記載される場合があるが、あらゆる互換性がある酵素の使用が企図される。

【0062】

ライブラリ構成要素から得られた配列情報は、インビトロ近接ライゲーションについて、Hi-C, Chicago (登録商標) の、または他の三次元の配座解析の文脈において、当該技術分野で知られるものなどの、多くのアプローチによって評価される。重要なことに、細胞特異的なリードペアの頻度は得ることができ、その結果、末端隣接配列が、ゲノムの特定の領域または特定のコンティグにマッピングする頻度は、細胞特異的な基準で評価することができる。すなわち、有望な3次元立体配座の細胞特異的な発生を評価することができる。場合によっては、三次元の立体配座における細胞特異的な距離に相互関連して、細胞特異的な信号の強さを評価することもでき、その結果、核酸のある領域は、比肩し得るが「より弱い」またはより離れた近接がある第2の細胞に比べて、1の細胞中で比較的近接にあると結論付けることができる。すなわち、三次元構造の質的および定量的評価の両方は、本明細書の開示と一貫する。場合によっては、第2の領域に対する1つの領域の近接は、特に、固有の区画タグなどの共通の区画識別配列を共有するライブラリ構成要素において、第2のクラスターのクラスター構成要素を有するペアエンドのリード中で同時に生じる第1のクラスターのクラスター構成要素の数を集計することによって、少なくとも部分的に評価される。

【0063】

複数のライブラリ構成要素において同一の末端隣接配列の複合の発生によって構造情報を作る必要はない。むしろ、場合によっては、第2の末端隣接配列マッピング部位の近くにマッピングされた末端隣接配列 (共通の「クラスター」に対する) は、両方のクラスターのメンバがゲノムなどの核酸対照の第2の領域上の第2のクラスターの同一ではない領域にマッピングされるとき、三次元立体配座の評価を強化することができる。

【0064】

場合によっては、本明細書に開示される方法は、ポリヌクレオチドあるいはその配列セグメントを標識付けおよび／または関連付けするために、および、そのデータを各種用途に利用するために使用される。場合によっては、本明細書における開示は、約10,000、約20,000、約50,000、約100,000、約200,000、約500,000、約100万、約200万、約500万、約1000万、約2000万、約3000万、約4000万、約5000万、約6000万、約7000万、約8000万、約9000万、約1億、約2億、約3億、約4億、約5億、約6億、約7億、約8億、約9億、約10億未満のリードペアを有する、高度に連続し且つ正確なヒトゲノムアセンブリを生成することができる方法および計算システムを提供する。場合によっては、本開示は、約50%、60%、70%、75%、80%、85%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%、またはそれより高い精度で、ヒトゲノムにおけるヘテロ接合型変異体の約50%、60%、70%、75%、80%、85%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%、またはそれより多くを段階的に実行する、または物理的なリンケージ情報割り当てる、方法を提供する。

【0065】

いくつかの実施形態では、本明細書に記載された組成物と方法は、メタゲノム (例えばヒト腸で見つかったもの) の研究を可能にする。従って、所与の生態学的環境に居住するいくらかまたはすべての生物体の部分的または全体のゲノムの配列を調査することができる。例として、腸全体の微生物、皮膚のある領域で発見された微生物、および、有毒廃棄物用地に住んでいる微生物の無作為の配列決定が挙げられる。これらの環境における微生物集団の組成、ならびにそれぞれのゲノムによってコード化される相互関係にある生化学的側面は、本明細書に記載の組成物と方法を使用して求めることができる。本明細書に記

載された方法は、例えば、2、3、4、5、6、7、8、9、10、12、15、20、25、30、40、50、60、70、80、90、100、125、150、175、200、250、300、400、500、600、700、800、900、1000、5000、あるいは10000以上の生物体および／または生物体変異体を含む、複合的な生物学的環境由来のメタゲノム研究を実現させる場合がある。

【0066】

従って、本明細書に開示される方法は、完全なヒトゲノムDNAサンプルに適用される場合があるが、しかし広く様々な核酸サンプルにも適用され得、逆転写RNAサンプル、循環遊離DNAサンプル、癌組織サンプル、犯罪現場サンプル、考古学サンプル、非ヒトゲノムのサンプル、または、実験室条件の下で容易に培養されない生物体などの、2以上の生物体由来の遺伝子情報を含む環境サンプルなどの環境サンプルなどが挙げられる。

10

【0067】

癌ゲノム配列決定に必要とされる高い精度は、本明細書に記載される方法およ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抗原投与際に塩基呼び出しに困難をもたらしかねない。不均一なサンプルおよび小さな出発物質、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。さらに、ヘテロ接合性の大規模な構造の変異および／または損失の検出は、多くの場合、癌ゲノム配列決定の他、体細胞の変異とベースコーリングにおける誤りとを区別する能力にとっても、重要である。

【0068】

20

本明細書に記載されるシステムおよび方法は、2、3、4、5、6、7、8、9、10、12、15、20、また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、および／また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。いくつかの実施形態において、100ng程度の小さなサンプル、あるいは数百程のゲノム当量の小さな出発サンプルが、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および方法は、変異、大規模な構造変異および再編成の検出を可能とする場合があり、フェージングされた変異体コールが、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、またはそれ以上のヌクレオチドにまたがる長い配列にわたって得られる場合がある。例えば、フェーズの変異体コールは、約1Mbpまたは約2Mbpにまたがる長い配列にわたって得られる場合がある。

30

【0069】

ある態様では、本明細書に開示される方法は、単一のDNA分子に由来する複数のコンティグをアセンブルするために使用される。場合によっては、方法は、多数のナノ粒子に架橋された単一のDNA分子から多数のリードペアを生成する工程、および、およびリードペアを使用して、コンティグを組み立てる工程を含む。ある場合では、単一のDNA分子は細胞の外部で架橋される。場合によっては、リードペアの少なくとも0.1%、0.2%、0.3%、0.4%、0.5%、0.6%、0.7%、0.8%、0.9%、1%、2%、3%、4%、5%、6%、7%、8%、9%、10%、11%、12%、13%、14%、15%、16%、17%、18%、19%、20%、25%、30%、35%、40%、45%、あるいは50%は、単一のDNA分子上で、1kB、2kB、3kB、4kB、5kB、6kB、7kB、8kB、9kB、10kB、15kB、20kB、30kB、40kB、50kB、60kB、70kB、80kB、90kB、100kB、150kB、200kB、250kB、300kB、400kB、500kB、600kB、700kB、800kB、900kB、あるいは1MBより大きい距離にまたがる。ある場合では、リードペアの少なくとも0.5%、0.6%、0.7%、0.8%、0.9%、1%、2%、3%、4%、5%、6%、7%、8%、9%、10%、11%、12%、13%、14%、15%、16%、17%、18%、19%、あるいは20%は、

40

50

単一のDNA分子上で、5 k B 6 k B、7 k B、8 k B、9 k B、10 k B、15 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、150 k B、あるいは200 k Bより大きい距離にまたがる。さらなる場合では、リードペアの少なくとも0.5%、0.6%、0.7%、0.8%、0.9%、1%、2%、3%、4%、あるいは5%は、単一のDNA分子上で、20 k B、30 k B、40 k B、50 k B、60 k B、70 k B、80 k B、90 k B、あるいは100 k Bより大きい距離にまたがる。特定の場合では、リードペアの少なくとも1%あるいは5%は、単一のDNA分子上で、50 k Bあるいは100 k Bより大きい距離にまたがる。場合により、リードペアは、1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、30、40、50、または60日以内に生成される。ある場合では、リードペアは、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、あるいは18日以内に生成される。さらなる場合では、リードペアは、7、8、9、10、11、12、13、あるいは14日以内に生成される。特定の場合では、リードペアは7あるいは14日以内に生成される。

【0070】

本明細書に記載される方法およ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。短い変異体コールは、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、短い変異体コールからの組み合わせられた情報、およ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、セントロメアおよびテロメアの領域、および、反復領域を持つものを含むがこれに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、ある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。

【0071】

サンプル型は、局所的に、ある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、情報のソースが癌また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、一般に、組織の型、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が、これらに限定されない。癌ゲノムと比較した正常ゲノムなどの、完全な、または部分的な比較ゲノム配列が利用可能である場合、サンプルデータと比較ゲノム配列との差異を判定し、そして随意に出力することができる。

【0072】

本開示の方法は、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作用し得るゲノムの領域の遺伝子情報の解析に使用され得る。本明細書に開示されるとおりの増幅方法は、限定されないが、米国特許6449562号、6287766号、7361468号、7414117号、6225109号、および6110709号などの、遺伝子解析の技術において既知の装置、キット、および方法の中で使用され得る。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、または対立遺伝子は、遺伝病などの疾患または疾病に関連付けられ得る。いくつかの他の場合では、多形は、疾患また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、および加齢性の疾病、癌などに関連付けられる。他の場合では、多形は、冠状動脈の健康の増大といった有用な特色、HIVまたはマラリアなどの疾患に対する抵抗性、あるいは骨粗鬆症、アルツハイマー病、また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。

【0073】

本開示の組成物および方法は、診断、予後、治療、患者の層別化、薬物の開発、処置の

選択、およ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生体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。

【0074】

本明細書に提供される方法は、これらの反復的な領域によって提起された実質的な障壁を克服することにより、ゲノミクス分野を大いに進歩させることができ、および、それによりゲノム解析の多くのドメインで重要な進歩を可能にし得る。従来の方を用いてデノボアセンブリを実施するには、多くの小さなスキフォールドへ断片化された集合体を受け入れるか、大きな挿入ライブラリを生成するために、相当な時間および資源を充てるか、または、より連続したアセンブリを生成するために他のアプローチを使用しなければならない。そのようなアプローチは、非常に深い配列決定対象範囲を取得すること、BACまたはフォスミド (fosmid) ライブラリを構築すること、光学マッピング、または、最もあり得る場合として、これらと他の技術との何らかの組み合わせを、含み得る。極度の資源と時間の要求により、そのようなアプローチはほとんどの小さな研究機関にとって手の届かないものとなり、モデルではない生物体を研究することが妨げられている。本明細書に記載された方法は、非常に長距離のリードセットを生成することができるため、デノボアセンブリが単一の配列決定実行で達成される場合がある。このことは、数桁のオーダーでアセンブリ費用を削減し、また数か月または数年必要だった時間を数週間にまで短縮する。場合によっては、本明細書に開示される方法は、多数のリードセットを、14日未満、13日未満、12日未満、11日未満、10日未満、9日未満、8日未満、7日未満、6日未満、5日未満、4日未満、3日未満、2日未満、1日未満で、または、ここに明記された期間のいずれか2つにわたる範囲内で、生成することが可能である。場合によっては、方法は、約10日から14日で多数のリードセットを生成することが可能である。もっともニッチな生物体のためのゲノム構築さえ、ルーチンになり、系統発生のフローサイトメトリー解析は、照合の欠如に悩むことはなくなり、および、10kのゲノムなどのプロジェクトは実現可能になるだろう。

【0075】

本明細書に記載された方法は、染色体またはより短い連続する核酸分子などの物理的なリンケージ群の中へ、予め提供されたか、予め生成されたか、デノボ合成されたコンティグ情報を割り当てることが可能である。同様に、本明細書に開示される方法は、前記コンティグが、互いに対して、物理的な核酸分子に沿った線形順序で配置されることを可能にする。同様に、本明細書に開示される方法は、物理的な核酸分子に沿った線形順序中で互いに対して前記コンティグが配向されることを可能にする。

【0076】

同様に、本明細書に開示される方法は、医療目的のための構造解析およびフェージング解析において進歩をもたらし得る。癌、同じ型の癌を持つ個体の間、あるいは同じ腫瘍の中でさえ、驚くべき異種性が存在する。結果として生じ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およびスループットを必要とする。オーダーメイド医療のドメインにおいて、ゲノムケアのゴールドスタンダードのうち1つは、大小の構造的転位および新規な突然変異を含む、徹底的に特徴づけられ且つフェージングされた全ての変異体を有する、配列決定されたゲノムである。従来技術でこれを達成するためには、現在では、ルーチンの医療処置とするにはあまりに高価かつ重労働である、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。場合によっては、本明細書に開示される方法は、低コストで、完全に正確なゲノムを迅速に作り出し、およびそのことにより、ヒトの疾患の研究と処置において多くの高度に求められる能力をもたらす。

【0077】

さらに、本明細書に開示される方法をフェージングに適用することで、統計的アプローチの便利さを家族の遺伝的解析の正確さと組み合わせることが可能であり、これらの方法

のどちらか一方を単独で使用するよりも、金銭、労働、およびサンプルに関して、大きな節約をもたらす。従来の技術では不可能な高度に望ましいデノボ変異体フェージング解析は、本明細書に開示される方法を使用して、容易に実行することができる。このことは、ヒトの変種の大部分が稀であるため（５％未満のマイナーアレル頻度）、特に重要である。フェージング情報は、リンクされていない遺伝子型に比べて、高度に接続されたハプロタイプ（単一の染色体に割り当てられた変異体の集団）のネットワークから、著しい利益を得る集団遺伝学研究にとって、価値がある。ハプロタイプ情報は、個体数の歴史的変化、部分母集団間の移動、および交流に関するより高い分解能の調査を可能にし得、および、具体的な変異を特定の親および祖父母まで遡ることを可能にする。このことは、次に、疾患に関連付けられる変異体の遺伝的伝達、および、一人の個体中に集約された時の変異体間の相互作用を、明らかにする。さらなる場合では、本開示の方法は、極めて長距離のリードセット（X L R S）または極めて長距離のリードペア（X L R P）のライブラリの調製、配列決定、および解析を可能にする。

【００７８】

本開示のいくつかの実施形態では、被験体の組織あるいはDNAのサンプルが提供され、そして方法は、アセンブルされたゲノム、コールされた変異体（大規模な構造変異体を含む）を伴うアラインメント、フェージングされたバリエーションコール（variant call）、または任意の付加的な解析を返す。他の実施形態において、本明細書に開示される方法は、個体ごとに直接的に、X L R Pライブラリを提供する。

【００７９】

様々な実施形態において、本明細書に開示される方法は、大きな距離で隔てられた極めて長距離のリードペアを生成する。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、２５００、３０００、４０００、５０００kbp、またはそれ以上のゲノム距離におよぶ。場合によっては、リードペアは、最大５００kbpのゲノム距離におよぶ。場合によっては、リードペアは、最大２０００kbpのゲノム距離におよぶ。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、さらに効率、特異性、およびゲノムカバレッジの増大に十分に適している。場合によっては、リードペアは、約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、６０日未満のうちに、あるいは９０日未満のうちに生成される。場合によっては、リードペアは、約１４日未満のうちに生成される。さらなる場合では、リードペアは、約１０日未満のうちに生成される。場合によっては、本開示の方法は、リードペアの約５％、約１０％約１５％、約２０％、約３０％、約４０％、約５０％、約６０％、約７０％、約８０％、約９０％、約９５％、約９９％以上が、あるいは約１００％が、多数のコンティグを正確に順序付けおよび／または配向することにおいて、少なくとも約５０％、約６０％、約７０％、約８０％、約９０％、約９５％、約９９％、または約１００％の精度を有するリードペアを提供する。場合によっては、方法は、多数のコンティグを正確に順序付けしおよび／または配向することにおいて、約９０～１００％の精度を提供する。

【００８０】

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。場合によっては、この方法は、十分に試験され、および／また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。さら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。

【００８１】

本明細書に開示される方法は、広範囲の生物についてのデノボのゲノムアセンブリを劇的に単純化することができる。従来の技術を使用して、このようなアセンブリは、現在、

経済的なメイトペアライブラリーの短いインサートによって制限されている。フォスミドにより利用可能な40～50kbpまでのゲノム距離のリードペアを生成することが可能な場合があるが、これらは高価で手間がかかり、および、ヒトでは300kbpから5Mbpまでのサイズに及ぶセントロメアを含む最長の反復範囲に対しては短すぎる。場合によっては、本明細書に開示される方法は、大きな距離（例えば、メガベースまたはより長い）にまたがることのできるリードペアを提供し、そのことにより、これらのスキャフォールドの完全性の問題を克服する。従って、染色体レベルのアセンブリを生成することは、本明細書に開示される方法を利用することによりルーチンとなり得る。同様に、長距離のフェージング情報を取得することは、集団遺伝学、系統発生学、および疾患の研究に莫大なさらなる力をもたらす。ある場合には、本明細書に開示される方法は、多数の個体についての正確なフェージングを可能にし、それにより、ゲノムを調査する能力の幅と深さを集団の個体数と深い時間のレベルに関して拡大する。

【0082】

個人化医療（personalized medicine）の領域で、本明細書に開示される方法から生成されたXLRSLリードセットは、正確な、低コストの、フェージングされた、および急速に生成される、個人ゲノムのための有意義な進歩を表わす。従来の方法は、長距離の変異体をフェージングする能力において不十分であり、そのことが、複合ヘテロ接合性遺伝子型の表現型の影響の特性評価を妨げている。さらに、ゲノム疾患について実質的な利害がある構造的変異体は、それらのサイズがそれらを研究するために使用されるリードおよびリードインサートと比較して大きいため、従来技術を用いて正確に識別し特性評価するのが困難である。何十キロベースからメガベースまたはそれ以上にまたがるリードセットは、この困難の緩和を助け、そのことにより、構造の変種の高度に並列化されかつ個別化された解析が可能になる。

【0083】

基礎的な進化および生物医学的研究はスループットが高い配列決定における技術的進歩を促進し得る。大規模な多量のDNA配列データを生成することは今や比較的低コストである。しかし、高品質であり高度に連続するゲノム配列を従来技術で生成することは、理論上、および現実面で困難である。さらに、ヒトを含む多くの生物体が二倍体であり、これらの個体はそれぞれ、ゲノムの2つの一倍数のコピーを持っている。ヘテロ結合性（例えば、母親から与えられた対立遺伝子が父親から与えられた対立遺伝子と異なる場合）の部位では、対立遺伝子のどのセットがどの親に由来するかを知るのは難しい（ハプロタイプ・フェージングとして知られる）。この情報は、疾患と形質の関連の研究などの、多くの進化・生物医学的な調査の実行にとって極めて重要になり得る。

【0084】

本開示は、ゲノムアセンブリのための方法を提供し、該方法は、DNA調製のための技術と、所与のゲノム内のクロマチン複合体などの複合体に結合する単一の物理的な核酸分子に由来する配列リードに対応する、短い、中程度の、および長い期間の接続を高スループットで探索するためのタグ付けされた配列リードとを組み合わせる。本開示は、ハプロタイプのフェージングのための、および／またはメタゲノム研究のためのゲノムアセンブリを支援するためにこれらの接続を使用する方法をさらに提供する。被験体のゲノムのアセンブリを決定するために本明細書に提示された方法を使用することができる一方で、ある場合には、本明細書に提示された方法は、染色体、または被験体の様々な長さのクロマチンのアセンブリなどの被験体のゲノムの部分のアセンブリを決定するために使用されることも理解されるべきである。ある症例には、決定するかあるいは染色体に位置しない核酸分子の集合体を指図するために、本明細書に提示された方法が使用されることは、さらに理解されるに違いない。実際に、非反復のコンティグを分離する反復領域の存在によって複雑になっているどのような核酸の配列決定でも、本明細書に開示される方法を使用し、容易にすることができる。

【0085】

さらなる場合では、本明細書に開示される方法は、少量の材料を用いて、遺伝子型アセ

ンブリ、ハプロタイプ・フェージング、およびメタゲノミクスのための正確かつ予測的な結果を可能にする。場合によっては、約100ピコグラム (pg)、約200pg、約300pg、約400pg、約500pg、約600pg、約700pg、約800pg、約900pg、約1.0ナノグラム (ng)、約2.0ng、約3.0ng、約4.0ng、約5.0ng、約6.0ng、約7.0ng、約8.0ng、約9.0ng、約10ng、約15ng、約20ng、約30ng、約40ng、約50ng、約60ng、約70ng、約80ng、約90ng、約100ng、約200ng、約300ng、約400ng、約500ng、約600ng、約700ng、約800ng、約900ng、約1.0マイクログラム (μg)、約1.2μg、約1.4μg、約1.6μg、約1.8μg、約2.0μg、約2.5μg、約3.0μg、約3.5μg、約4.0μg、約4.5μg、約5.0μg、約6.0μg、約7.0μg、約8.0μg、約9.0μg、約10μg、約15μg、約20μg、約30μg、約40μg、約50μg、約60μg、約70μg、約80μg、約90μg、約100μg、約150μg、約200μg、約300μg、約400μg、約500μg、約600μg、約700μg、約800μg、約900μg未満の、または、約1000μg未満のDNAが、本明細書に開示される方法と共に使用される。場合によっては、本明細書に開示される方法において使用されるDNAは、約10,000,000、約5,000,000、約4,000,000、約3,000,000、約2,000,000、約1,000,000、約500,000、約200,000、約100,000、約50,000、約20,000、約10,000、約5,000、約2,000、約1,000、約500、約200、約100、約50、約20未満、または、約10未満の細胞から抽出される。

【0086】

二倍体ゲノムにおいて、どの対立遺伝子の変異体が同じ染色体上に物理的にリンクしているか知ることは、染色体対の相同の位置にマッピングすることよりもむしろ、しばしば重要である。対立遺伝子あるいは他の配列を特定の倍数染色体対の物理的な染色体にマッピングすることは、ハプロタイプ・フェージングとして知られている。高スループットの配列データに由来する短いリードは、特に、対立遺伝子の変異体、最も頻繁にあるような、最長の単一のリードよりも大きな距離によって隔てられている場合には、どの対立遺伝子の変異体がリンクされているかを直接観察することは稀にしか可能ではない。ハプロタイプ・フェージングの計算上の推定は、長距離については信頼性が低くなり得る。本明細書に開示される方法は、リードペア上の対立遺伝子の変異体を使用して、どの変異体が物理的にリンクされているかを決定することを可能にする。

【0087】

様々な場合において、本開示の方法と組成物は、多数の対立遺伝子の変異体に関する二倍体または倍数体ゲノムのハプロタイプ・フェージングを可能にする。本明細書に記載の方法は、このようにして、標識された配列セグメントおよび／またはそれらを用いてアセンブルされたコンティグからの変異体情報に基づいて、リンクされた対立遺伝子変異体を決定することに寄与する。対立遺伝子の変異体の事例は、限定されないが、1000 genomes、UK10K、HapMap、および、ヒトにおける遺伝的変異を発見するための他のプロジェクトによって知られている。場合によっては、疾患の特定遺伝子への関連性は、ハプロタイプ・フェージングのデータを得ることにより、より容易に明らかにされ、例えば、シャルコー・マリー・トゥース神経障害の原因となるSH3TC(2)の両方のコピーにおけるリンクされていない不活性化突然変異(Lupski JR, Reid JG, Gonzaga-Jauregui C, et al. N. Engl. J. Med. 362:1181-91, 2010)、および、高コレステロール血症9の原因となるABCG(5)の両方のコピーにおける不活性化突然変異の発見により、実証されている。

【0088】

ヒトは、平均で1,000部位のうち1部位がヘテロ接合である。場合によっては、高スループットの配列決定法を使用するデータの単一のレーンは、少なくとも約150,0

10

20

30

40

50

00, 000のリードを生成する。さらなる場合では、個々のリードは約100塩基対の長さである。サイズが平均150 kbpのDNA断片を投入して断片当たり100ペアエンドのリードを得るとすると、セット当たり、つまり、100リードペアあたり、30のヘテロ接合部位を観察することが期待できる。セット内にヘテロ接合部位を含有しているすべてのリードペアは、同じセット内の他のすべてのリードペアに対して相同（つまり、分子的なリンク）にある。この特性は、場合によっては、特異なリードのペアとは反対に、セットをフェージングするためのより大きな能力を可能にする。ヒトゲノム中のおよそ30億の塩基があり、および1000個に1個はヘテロ接合であるため、平均的なヒトゲノム中にはおよそ300万のヘテロ接合部位がある。ヘテロ接合の部位を包含している約45,000,000のリードペアに対し、高スループットの配列方法の単一のレーンを使用してフェージングされることになる各ヘテロ接合部位の平均カバレッジは、典型的な高スループット配列決定機械を使用して、約（15X）である。従って、二倍体ヒトゲノムは、本明細書に開示される方法を使用して調製されるサンプル由来の配列変異体に関する高スループット配列データの1つのレーンを用いて、高い信頼性でかつ完全にフェージングすることができる。場合によっては、データのレーンは、1セットのDNA配列リードのデータである。さらなる場合では、データのレーンは、高スループット配列決定機器の一回の実行からの1セットのDNA配列リードのデータである。

【0089】

ヒトゲノムが2つの相同の染色体のセットから成るため、個々の真の遺伝的構造を理解することは、母および父の遺伝子材料のコピーまたはハプロタイプを描写することを必要とする。個体中のハプロタイプを得ることはいくつかの意味で役に立つ。例えば、ハプロタイプは、臓器移植におけるドナーホストマッチングの結果を予測するときに臨床的に有用である。ハプロタイプは疾患関連性を検知するためにますます使用されている。複合ヘテロ接合性を示す遺伝子では、ハプロタイプは、2つの有害な変異体が同じ対立遺伝子上に位置するか（すなわち、「遺伝学術語を使用すると「イン・シス」）、2つの異なる対立遺伝子上に位置するか（「イン・トランス」）どうかに関して情報を提供し、該情報はこれらの変異体の遺伝が有害かどうかの予測に大いに影響し、および、個体が機能的な対立遺伝子と2つの有害の変異体位置を有する単一の機能を果たさない対立遺伝子を持っているのか、それとも、その個体はそれぞれ異なる欠陥を有する2つの機能を果たさない対立遺伝子を持っているのかどうかについての結論を左右する。個体群のハプロタイプは、疫学者と人類学者の両方に関心の母集団構造についての情報をもたらし、および人類の進化の歴史についての知識を与えてきた。加えて、遺伝子発現における広範囲の対立遺伝子の不均衡が報告されており、および、対立遺伝子フェーズ間の遺伝的または後成的な差異が、発現における定量的差異の一因となる場合があることを示唆している。ハプロタイプ構造についての理解は、対立遺伝子の不均衡の一因となる変異体のメカニズムを描写することになる。

【0090】

ある実施形態では、本明細書に開示される方法は、長距離のリンケージとフェージングのために求められるようなゲノムの離れた領域間の関連を固定し捕捉するための、インビトロの技術を含む。場合によっては、方法は、ゲノム上で非常に離れたリードペアを送達するための1つ以上のリードセットを構築し、および配列決定する工程を含む。さらなる場合では、リードセットは各々、共通のバーコードによって標識される2つ以上のリードを含み、該リードは共通のポリヌクレオチド由来の2つ以上の配列セグメントを表わし得る。場合によっては、相互作用は、主として、単一のポリヌクレオチド内の確率的な関連性から発生する。場合によっては、ポリヌクレオチドにおける互いに近い配列セグメントは、より頻繁に、および高い確率で相互作用する一方で、分子の離れた部分間の相互作用はそれほど頻繁でないため、配列セグメント間のゲノムの距離が推量される。結果的に、投入されたDNAの上で2つの遺伝子座を接続するペアの数とそれらの近接との間に系統的な関係がある。

【0091】

いくつかの態様では、本開示は、フェージングの極めて高い精度を達成するためのデータを生み出す方法および組成物を提供する。従来の方法と比較して、本明細書に記載された方法は、変異体のより高い割合をフェージングすることができる。場合によっては、フェージングは高水準の正確さを維持しながら達成される。さらなる場合では、このフェージング情報はより長い範囲に延長され、例えば、約200kbp、約300kbp、約400kbp、約500kbp、約600kbp、約700kbp、約800kbp、約900kbp、約1Mbp、約2Mbp、約3Mbp、約4Mbp、約5Mbpよりも大きな、または、約10Mbpよりも大きな、または、約10Mbpを超えて染色体の全長までの長さにもわたる。いくつかの実施形態では、ヒト・サンプルについてのヘテロ接合SNPの90%以上は、99%を超える正確さでフェージングされ、このとき、例えば、Illumina HiSeqデータのたった1つのレーンを使用することにより、約2億5000万未満のリードが使用される。他の場合では、ヒト・サンプルについてのヘテロ接合SNPの約40%、50%、60%、70%、80%、90%、95%超、または約99%超が、約70%、80%、90%、95%超の、または99%超の高い精度でフェージングされ、このとき、例えば、Illumina HiSeqデータのたった1つか2つのレーンを使用することにより、約2億5000万未満、または約5億未満のリードが使用される。場合によっては、ヒト・サンプルについてのヘテロ接合SNPの95%超、または99%超は、約95%または99%よりも高い精度でフェージングされ、このとき約2億5000万未満または約5億未満のリードが使用される。さらなる場合では、付加的な変異体は、約200bp、250bp、300bp、350bp、400bp、450bp、500bp、600bp、800bp、1000bp、1500bp、2kbp、3kbp、4kbp、5kbp、10kbp、20kbp、50kbpまでの、または約100kbpまで、リードの長さを増加させることにより、さらなる変異体を捕捉する。

【0092】

本開示の組成物および方法は、遺伝子発現解析に使用され得る。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、また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、および癌を検出することができる。また、それは環境モニタリング、法医学、および食品科学においても役立つ。核酸について実行することができる遺伝子解析の例として、例えば、SNP検出、STR検出、RNA発現解析、プロモーターメチル化、遺伝子発現、ウィルス検出、ウィルスのサブタイピング、および薬物抵抗性が挙げられる。

【0093】

本発明の方法は、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、特定の処置に応答する患者の能力、または患者にとって最良の処置を判定するために、患者から得られたまたは導き出された生体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。

【0094】

いくつかの実施形態では、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」または疾病の「診断」は、疾病を予測または診断すること、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、また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、血液サンプルは、サンプル中の疾患または悪性細胞型のマーカーの存在および／また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ってアッセイされ得る。

【0095】

いくつかの実施形態では、本明細書に記載される方法および組成物は、疾病の診断および予後のために使用される。

【0096】

多数の免疫学的、増殖性、およ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。免疫疾患および障害は、アレルギー性疾患および障害、免疫機能の障害、ならびに自己免疫疾患および疾病を含む。アレルギー性疾患および障害は、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、およ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および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、およびインターフェロン・ガンマーTH1経路欠損症を含む。自己免疫性およ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、甲状腺腫およ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、および炎症性腸疾患、同種異系移植片拒絶反応、および感染性細菌また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。

【0097】

本開示の方法により評価され得る増殖性疾患および障害は、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、線維囊胞症（例えば乳房または子宮の）、サルコイドーシス、ペロニーおよ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、および血管再狭窄を含む。

【0098】

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

【0099】

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、および他のリンパ腫、急性または慢性白血病、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、および形質細胞障害を含む。

【0100】

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、およ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。

【0101】

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、および他の骨癌、睾丸および卵巣の癌腫、頭頸部腫瘍、および子宮頸部の新生物を含む。

【0102】

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、染色体および遺伝子の異常、または遺伝子疾患に関する出生前また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、ダウン症候群または他のトリソミー、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نفelter症候群、ハンチントン病、自己免疫疾患、リポドーシス、肥満欠損症（obesity defect）、血友病、先天性代謝異常症、および糖尿病を含む。

【0103】

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

【0104】

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、ウィルス、寄生生物、およ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。

【0105】

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、トレポネーマ・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 (*Actinomyces*) を含む。

10

【0106】

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・デルマトフィティディス、ヒストプラスマ・カプスラーツム、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 (リゾプス)、スポトリックス・シェンキー、クロモミコシス、およびマズラミコシスを含む。

20

【0107】

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

30

【0108】

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 medinesis*)、吸虫、広節裂頭条虫、テニア種、ニューモシスチス・カリニ、およびアメリカ鉤虫 (*Necator americanus*) を含む。

40

【0109】

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・ツベルクロシス、および耐AZT性ヒト免疫不全ウィルスは全て、本開示で識別され得る。

【0110】

故に、本開示の組成物およ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)、あるいは細菌またはウィルスマーカーなどの、異物による感染のマ-

50

カーの何れかであり得る。

【0111】

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

【0112】

いくつかの実施形態において、本開示の方法および組成物は、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、また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、および亜臨床的な状態の識別に有用である。

【0113】

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。いく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。いく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またはその一部であり、その非限定的な例には、動物、真菌類、原生生物、モネラ、ウィルス、ミトコンドリア、および葉緑体が含まれる。サンプルのポリヌクレオチドは、例えば、培養細胞株、生検、血液サンプル、または細胞を包含している流体サンプルを含め、細胞サンプル、組織サンプル、またはそこから引き出される臓器サンプルなどの被験体から分離され得る。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、通常、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。いくつかの実施形態において、サンプルはDNAを含む。いくつかの実施形態において、サンプルはゲノムDNAを含む。いく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、酵母人工染色体、オリゴヌクレオチドタグ、またはそれらの組み合わせを含む。いく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、およ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、逆転写の産物は相補的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、およ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。

【0114】

いくつかの実施形態において、核酸の鋳型分子（例えば、DNAまたはRNA）は、タンパク質、脂質、および非鋳型核酸などの、様々な他の成分を含有する生体サンプルから単離することができる。核酸鋳型分子は、任意の細胞材料から得られ、動物、植物、細菌、真菌類、または他の細胞生物から得られ得る。本開示での使用のための生体サンプルは、ウィルスの粒子または調製物を含む。核酸の鋳型分子は、生物から直接、或いは、生物から得た生体サンプル、例えば血液、尿、脳脊髄液、精液、唾液、痰、便、および組織から得ることができる。任意の組織または体液の標本が、本開示の使用のために、核酸のソースとして使用されてもよい。核酸の鋳型分子はまた、初代細胞培養物または細胞株などの培養細胞から単離され得る。鋳型核酸が得られる細胞または組織は、ウィルスまたは他の細胞内病原体に感染され得る。サンプルはまた、生体標本、cDNAライブラリ、ウィルスDNA、またはゲノムDNAから抽出された総体的なRNAであり得る。サンプルはまた、細胞構造がない起源から単離されたDNA、例えば、冷凍装置から増幅／単離されたDNAであり得る。

【0115】

核酸の抽出および精製の方法は、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、あるいはTRIzolおよびTri Reagent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の使用を伴うまたは伴わない、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、エタノール沈降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）；および（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))。核酸の単離および／または精製の別の例は、磁性粒子の使用を含み、核酸は特異的また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）。いく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また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号を参照されたい。望ましい場合、RNase阻害剤を、溶解緩衝液に追加することができる。特定の細胞また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、またはその両方を単離することを目的とし得る。抽出手順の間またはその後にDNAとRNAの両方が一緒に単離されると、さらなる工程を利用して、一方また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、または他の物理的若しくは化学的特性による精製により、抽出した核酸のサブ断片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、過剰なまたは不要な試薬、反応物、また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。

【0116】

核酸の鑄型分子は、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) によって記載された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、核酸からさらに除去することができる。

【0117】

他の実施形態では、本開示は、例えば、組織、細胞培養物、体液、動物組織、植物、細菌、真菌、ウィルスなどから単離されるDNAを含めた、高分子量二本鎖DNAへと容易に適用され得る。

【0118】

サイズ選択を含むHi-C法

本明細書において、方法が提供され、該方法は、少なくとも1つの核酸結合タンパク質に複合化された核酸分子を含む、安定化された生体サンプルを得る工程、核酸分子を複数のセグメントへと切断するために、安定化された生体サンプルをDNaseに接触させる工程、1つの接合部にて、複数のセグメントの第1のセグメントと第2のセグメントを付着させる工程、および、複数の選択されたセグメントを得るために、複数のセグメントをサイズ選択にかける工程を含む。場合によっては、複数の選択されたセグメントは、約145～約600bpである。場合によっては、複数の選択されたセグメントは約100～約2500bpである。場合によっては、複数の選択されたセグメントは約100～約600bpである。場合によっては、複数の選択されたセグメントは、約100bp～約600

b pの間、約100bp～約700bpの間、約100bp～約800bpの間、約100bp～約900bpの間、約100bp～約1000bpの間、約100bp～約1100bpの間、約100bp～約1200bpの間、約100bp～約1300bpの間、約100bp～約1400bpの間、約100bp～約1500bpの間、約100bp～約1600bpの間、約100bp～約1700bpの間、約100bp～約1800bpの間、約100bp～約1900bpの間、約100bp～約2000bpの間、約100bp～約2100bpの間、約100bp～約2200bpの間、約100bp～約2300bpの間、約100bp～約2400bpの間、または、約100bp～約2500bpの間である。

【0119】

本明細書において提供されるサイズ選択のステップを伴う方法の別の態様では、方法は、サイズ選択のステップに先立ち、複数のセグメントから配列決定ライブラリを調製する工程をさらに含む。いくつかの実施形態では、方法は、サイズ選択されたライブラリを得るために、配列決定ライブラリをサイズ選択にかける工程をさらに含む。場合によっては、サイズ選択されたライブラリは、約350bpから約1000bpの間のサイズにある。場合によっては、サイズ選択されたライブラリは、約100bpから2500bpの間のサイズにあり、例えば、約100bp～約350bpの間、約350bp～約500bpの間、約500bp～約1000bpの間、約1000bp～約1500bp～約2000bpの間、約2000bp～約2500bpの間、約350bp～約1000bpの間、約350bp～約1500bpの間、約350bp～約2000bpの間、約350bp～約2500bpの間、約500bp～約1500bpの間、約500bp～約2000bpの間、約500bp～約3500bpの間、約1000bp～約1500bpの間、約1000bp～約2000bpの間、約1000bp～約2500bpの間、約1500bp～約2000bpの間、約1500bp～約2500bpの間、または、約100bp～約700bpの間のサイズにある。

【0120】

本明細書に提供されるサイズ選択のステップを伴う方法において利用されるサイズ選択は、ゲル電気泳動、キャピラリー電気泳動、サイズ選択ビーズ、ゲル濾過カラム、他の適切な方法、またはその組み合わせによって行なわれる。

【0121】

別の態様では、本明細書に提供されるサイズ選択のステップを伴う方法は、1つのQC値を得るために、複数の選択されたセグメントを解析する工程をさらに含む。場合によっては、QC値は、クロマチン消化効率(CDE)およびクロマチン消化指標(CDI)から選択される。CDEは所望の長さを有するセグメントの割合として計算される。例えば、場合によっては、CDEは、サイズ選択の前にサイズが100～2500bpの間にあるセグメントの割合として計算される。場合によっては、CDE値が少なくとも65%である場合、サンプルはさらなる解析のために選択される。場合によっては、サンプルは、CDE値が少なくとも約50%、少なくとも約55%、少なくとも約60%、少なくとも約65%、少なくとも約70%、少なくとも約75%、少なくとも約80%、少なくとも約85%、少なくとも約90%、または、少なくとも約95%であるとき、さらなる解析のために選択される。CDIは、サイズ選択の前の、モノヌクレオソームのサイズのセグメントの数の、ジヌクレオソームのサイズのセグメントの数に対する比として計算される。例えば、CDIは、600～2500bpのサイズを有する断片対100～600bpのサイズを有する断片の比の対数として計算され得る。場合によっては、CDI値が-1.5を超えかつ1未満であるとき、サンプルはさらなる解析のために選択される。場合によっては、サンプルは、CDI値が約-2を超えかつ約1.5未満であるとき、約-1.9を超えかつ約1.5未満であるとき、約-1.8を超えかつ約1.5未満であるとき、約-1.7を超えかつ約1.5未満であるとき、約-1.6を超えかつ約1.5未満であるとき、約-1.5を超えかつ約1.5未満であるとき、約-1.4を超えかつ約1.5未満であるとき、約-1.3を超えかつ約1.5未満であるとき、約-1.2を超えかつ

10

20

30

40

50

約 1.5 未満であるとき、約 -1.1 を超えかつ約 1.5 未満であるとき、約 -2 を超えかつ約 1.5 未満であるとき、約 -1 を超えかつ約 1.5 未満であるとき、約 -0.9 を超えかつ約 1.5 未満であるとき、約 -0.8 を超えかつ約 1.5 未満であるとき、約 -0.7 を超えかつ約 1.5 未満であるとき、約 -0.6 を超えかつ約 1.5 未満であるとき、約 -0.5 を超えかつ約 1.5 未満であるとき、約 -2 を超えかつ約 1.4 未満であるとき、約 -2 を超えかつ約 1.3 未満であるとき、約 -2 を超えかつ約 1.2 未満であるとき、約 -2 を超えかつ約 1.1 未満であるとき、約 -2 を超えかつ約 1 未満であるとき、約 -2 を超えかつ約 0.9 未満であるとき、約 -2 を超えかつ約 0.8 未満であるとき、約 -2 を超えかつ約 0.7 未満であるとき、約 -2 を超えかつ約 0.6 未満であるとき、または、約 -2 を超えかつ約 0.5 未満であるとき、さらなる解析のために選択される。

10

【0122】

別の態様では、本明細書にサイズ選択のステップを伴う方法において使用される安定化された生体サンプルは、安定化剤を用いた処置された生体試料を含む。場合によっては、安定化された生体サンプルは、安定化された細胞溶解物を含む。あるいは、安定化された生体サンプルは、安定化された無傷細胞を含む。あるいは、安定化された生体サンプルは、安定化された無傷の核を含む。場合によっては、安定化された無傷細胞または無傷の核サンプルを DNase に接触させる工程は、無傷細胞または無傷の核の溶解に先立って行なわれる。場合によっては、細胞および／または核は、複数のセグメントの第1のセグメントと第2のセグメントとを接合部にて付着させる前に溶解される。

20

【0123】

別の態様では、本明細書における、サイズ選択のステップを伴う方法は、少数の細胞または少量の核酸しか包含していない小さなサンプルに対して行なわれる。例えば、場合によっては、安定化された生体サンプルは 3,000,000 未満の細胞を含む。場合によっては、安定化された生体サンプルは 2,000,000 未満の細胞を含む。場合によっては、安定化された生体サンプルは 1,000,000 未満の細胞を含む。場合によっては、安定化された生体サンプルは 500,000 未満の細胞を含む。場合によっては、安定化された生体サンプルは 400,000 未満の細胞を含む。場合によっては、安定化された生体サンプルは 300,000 未満の細胞を含む。場合によっては、安定化された生体サンプルは 200,000 未満の細胞を含む。場合によっては、安定化された生体サンプルは 100,000 未満の細胞を含む。場合によっては、安定化された生体サンプルは 10 μ g 未満の DNA を含む。場合によっては、安定化された生体サンプルは 9 μ g 未満の DNA を含む。場合によっては、安定化された生体サンプルは 8 μ g 未満の DNA を含む。場合によっては、安定化された生体サンプルは 7 μ g 未満の DNA を含む。場合によっては、安定化された生体サンプルは 6 μ g 未満の DNA を含む。場合によっては、安定化された生体サンプルは 5 μ g 未満の DNA を含む。場合によっては、安定化された生体サンプルは 4 μ g 未満の DNA を含む。場合によっては、安定化された生体サンプルは 3 μ g 未満の DNA を含む。場合によっては、安定化された生体サンプルは 2 μ g 未満の DNA を含む。場合によっては、安定化された生体サンプルは 1 μ g 未満の DNA を含む。場合によっては、安定化された生体サンプルは 0.5 μ g 未満の DNA を含む。

30

40

【0124】

別の態様では、本明細書におけるサイズ選択のステップを伴う方法は、個々の細胞または単細胞に対して行なうことができる。例えば、本明細書における方法は、個別の区画へと分配された細胞に対して行なわれる場合がある。典型的な区画として、限定されないが、ウェル、エマルジョン中の液滴、または表面位置（例えば、アレイ・スポット、ビーズなど）が挙げられ、それらの区画は本明細書に別記されるように差動的に配列決定されたリンカー分子の別個のパッチを含んでいる。付加的な区画もまた企図され、そして本明細書に開示される方法、組成物、およびシステムと一貫する。

【0125】

付加的な態様では、本明細書におけるサイズ選択のステップを伴う方法において使用さ

50

れる安定化された生体サンプルは、DNAの断片を作り出すために、DNaseなどのヌクレアーゼを用いて処置される。場合によっては、DNaseは非配列特異性である。場合によっては、DNaseは、一本鎖DNAおよび二本鎖DNAの両方に対して活性がある。場合によっては、DNaseは、二本鎖DNAに対して特異的である。場合によっては、DNaseは、二本鎖DNAを優先的に切断する。場合によっては、DNaseは、一本鎖DNAに対して特異的である。場合によっては、DNaseは、一本鎖DNAを優先的に切断する。場合によっては、DNaseはDNase Iである。場合によっては、DNaseはDNase IIである。場合によっては、DNaseは、DNase IおよびDNase IIの1つ以上から選択される。場合によっては、DNaseはミクロコッカスヌクレアーゼである。場合によっては、DNaseは、DNase I、DNase II、およびミクロコッカスヌクレアーゼの1つ以上から選択される。場合によっては、DNアーゼは免疫グロブリン結合タンパク質あるいはその断片に結合または融合され得る。免疫グロブリン結合タンパク質は、例えば、プロテインA、プロテインG、プロテインA/G、またはプロテインLであり得る。いくつかの実施形態では、DNアーゼは、2つ以上の免疫グロブリン結合タンパク質および/またはその断片を含む融合タンパク質に連結され得る。他の適切なヌクレアーゼもまたこの開示の範囲内にある。

【0126】

付加的な態様では、サイズ選択のステップを伴う方法において使用するための、本明細書において提供されるような、安定化された生体サンプルは、1つ以上の橋架剤で処置される。場合によっては、橋架剤は化学的固定剤である。場合によっては、化学的固定剤は、約2.3〜2.7オングストローム(A)のスペーサーアーム長を有するホルムアルデヒドを含む。場合によっては、化学的固定剤は、長いスペーサーアーム長を有する橋架剤を含み、例えば、橋架剤は、少なくとも長さ約3A、4A、5A、6A、7A、8A、9A、10A、11A、12A、13A、14A、15A、16A、17A、18A、19A、または20Aのスペーサー長を有し得る。化学的固定剤は、エチレングリコールビス(コハク酸スクシンイミジル)(EGS)を含み得、長さが約16.1Aのスペーサーアームを有する。化学的固定剤は、グルタル酸ジスクシンイミジル(DSG)を含み得、長さが約7.7Aのスペーサーアームを有する。場合によっては、化学的固定剤は、ホルムアルデヒドとEGS、ホルムアルデヒドとDSG、または、ホルムアルデヒドとEGSとDSGを含む。複数の化学的固定剤が利用される場合によっては、各化学的固定剤は、順序的に使用され、他の場合では、複数の化学的な固定剤のうちのいくつかあるいはすべては、サンプルに同時に適用される。長いスペーサーアームを有する架橋剤の使用は、大きな(例えば、>1kb)リードペア分離距離を有するリードペアの断片を増加させ得る。例えば、図7は、結果として生じるライブラリ(DNアーゼの消化とMNaseの消化の両方)について、ホルムアルデヒド単独で架橋されたものと、ホルムアルデヒドにDSGまたはEGSを加えたもので架橋されたものとの比較を示す。DSGは両端部にNHSエステル反応基を有し、およびアミノ基(例えば、一級アミン)に対して反応性であり得る。DSGは、膜浸透性であり、細胞内の架橋を可能にしている。DSGは、スベリン酸ジスクシンイミジル(DSS)と比較して、いくつかの応用において、架橋効率を増大させることができる。EGSは両端部にNHSエステル反応基を有し、およびアミノ基(例えば、一級アミン)に対して反応性であり得る。EGSは、膜浸透性であり、細胞内の架橋を可能にしている。EGS架橋は、例えば、pH 8.5にて3〜6時間ヒドロキシルアミンで処置することにより、逆行され得、一例では、ラクトースデヒドロゲナーゼは、EGSによる可逆的な架橋の後に、その活性の60%を保持した。場合によっては、化学的固定剤はソラレンを含む。場合によっては、橋架剤は紫外線である。いくつかの実施形態では、安定化された生体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。

【0127】

さらなる態様では、本明細書において提供されるサイズ選択のステップを伴う方法は、複数の選択されたセグメントを抗体に接触させる工程を含む。

【0128】

付加的な態様では、本明細書に提供されるサイズ選択ステップを伴う方法は、1つの接合部にて、複数のセグメントの第1のセグメントと第2のセグメントを付着させる工程を含む。場合によっては、付着させる工程は、ビオチンをタグ付けされたヌクレオチドを使用して粘着末端を埋め、および平滑末端をライゲーションする工程を含む。場合によっては、付着させる工程は、少なくとも第1のセグメントと第2のセグメントを架橋オリゴヌクレオチドに接触させることを含む。場合によっては、付着させる工程は、少なくとも第1のセグメントと第2のセグメントをバーコードに接触させることを含む。いくつかの実施形態では、本明細書における架橋オリゴヌクレオチドは、長さが少なくとも約5ヌクレオチドから約50ヌクレオチドまでであり得る。いくつかの実施形態では、本明細書における架橋オリゴヌクレオチドは、長さが少なくとも約15ヌクレオチドから約18ヌクレオチドまでであり得る。いくつかの実施形態では、架橋オリゴヌクレオチドは、長さが約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35、約40、約45、または約50ヌクレオチドであり得る。いくつかの実施形態では、本明細書における架橋オリゴヌクレオチドは、バーコードを含む場合がある。いくつかの実施形態では、架橋オリゴヌクレオチドは、複数のバーコードを含み得る。いくつかの実施形態では、架橋オリゴヌクレオチドは、相互に接続している複数の架橋オリゴヌクレオチドを含む。いくつかの実施形態では、架橋オリゴヌクレオチドは、プロテインA、Gタンパク質、タンパク質A/G、またはタンパク質Lなどの免疫グロブリン結合タンパク質またはそれらの断片に連結されるかリンクされる場合がある。場合によっては、連結された架橋オリゴヌクレオチドは、サンプルの核酸における抗体が結合される位置に送達され得る。

【0129】

スプリッティングとプーリングのアプローチは、固有のバーコードを有する架橋オリゴヌクレオチドを生み出すために利用することができる。サンプルの集団は複数の群へと分割することができ、架橋オリゴヌクレオチドは、架橋オリゴヌクレオチドのバーコードが群の間で異なるが、群内では同じであるように、サンプルに付着させることができ、サンプルの群は、再び一緒にプールすることができ、および、このプロセスは複数回繰り返すことができる。このプロセスを繰り返すことは、固有の一連の架橋オリゴヌクレオチド・バーコードを有する集団における各サンプルを結果として生じ、単一のサンプル（例えば、単細胞、単一の核、単一の染色体）の解析を可能にする場合がある。1つの実証的な例において、ビーズの固体支持体に付着された、架橋され消化された核のサンプルは、8本の管にわたって分離され、各々、ライゲーションされる二本鎖DNA（dsDNA）アダプターを含む第1のアダプター群（第1の反復）の8つの固有のメンバのうちの1つをそれぞれ包含する。8つのアダプターの各々は、核中の架橋クロマチン集合体の核酸末端へのライゲーションのための同じ5'オーバーハング配列を有する場合があるが、そうでなければ、固有のdsDNA配列を有する。第1のアダプター群がライゲーションされた後、前記核は再び一緒にプールされ、および、ライゲーション反応成分を除去するために洗浄され得る。分配し、ライゲーションし、プールするスキームは、さらに2回反復され得る（2つの反復）。各アダプター群のメンバのライゲーションの後、架橋クロマチン集合体は、一連の複数のバーコードに付着され得る。場合によっては、複数のアダプター群（反復）の複数のメンバの連続するライゲーションは、結果としてバーコードの組み合わせを生じる。利用可能なバーコードの組み合わせの数は、1つの反復当たりの群の数、および使用されたバーコード・オリゴヌクレオチドの合計数に依存する。例えば、8つのメンバを含む3つの反復は、83の可能性がある組み合わせを個々に持ち得る。場合によっては、バーコードの組み合わせは固有である。場合によっては、バーコードの組み合わせには冗長性がある。バーコードの組み合わせの総数は、固有のバーコードを受け取る群の数を増加あるいは減少させること、および／または、反復の数を増加あるいは減少させることにより調整することができる。1つを超えるアダプター群が使用される場合、分配し、付随させ、プールするスキームが、反復するアダプター付着のために使用され得る。場合によっては、分配し、付着させ、プールするスキームは、少なくとも3、4、5、6、7、

8、9、または10回、追加で反復され得る。場合によっては、最後のアダプター群のメンバは、例えば、PCR増幅による配列決定ライブラリの調製の間における、後の、アダプター付着DNAの濃縮のための配列を含む。

【0130】

さらなる態様では、本明細書におけるサイズ選択のステップを伴う方法は、剪断のステップを含まない（例えば、核酸は剪断されない）。

【0131】

本明細書におけるサイズ選択のステップを伴う方法のさらなる態様では、方法は、第1のリードペアを生成するために、接合部の両側で少なくともいくつかの配列を獲得する工程を含む。例えば、方法は、第1のリードペア生成するために、接合部の各側に、少なくとも約50bp、少なくとも約100bp、少なくとも約150bp、少なくとも約200bp、少なくとも約250bp、または少なくとも約300bpの配列を得る工程を含み得る。

10

【0132】

本明細書におけるサイズ選択のステップを伴う方法の付加的な態様では、方法は、コンティグのセットに第1のリードペアをマッピングする工程、および、ゲノムへの順序および／または配向を表現するコンティグのセットにわたる経路を決定する工程をさらに含む。

【0133】

本明細書におけるサイズ選択のステップを伴う方法のさらなる態様では、方法は、コンティグのセットに第1のリードペアをマッピングする工程、および、前記コンティグのセットから、安定化された生体サンプルにおける構造的変異体の存在またはヘテロ結合性の減少を判定する工程をさらに含む。

20

【0134】

本明細書におけるサイズ選択のステップを伴う方法の付加的な態様では、方法は、1セットのコンティグに第1のリードペアをマッピングする工程、およびコンティグのセットにおける変異体をフェーズに割り当てる工程を含む。

【0135】

本明細書におけるサイズ選択のステップを伴う方法のさらなる態様では、方法は、コンティグのセットに第1のリードペアをマッピングする工程、コンティグのセットから、コンティグのセットにおける変異体の存在を判定する工程、および、（1）安定化された生体サンプルについて疾患段階、予後、あるいは処置の方針を確認するステップ、（2）変異体の存在に基づいて、薬物を選択するステップ、または、（3）安定化された生体サンプルに対する薬効を確認するステップから選択される1つ以上のステップを行なう工程をさらに含む。

30

【0136】

QC計算を含むHi-C法

付加的に、本明細書において方法が提供され、該方法は、少なくとも1つの核酸結合タンパク質に複合化された核酸分子を含む、安定化された生体サンプルを得る工程、核酸分子を複数のセグメントへと切断するために、安定化された生体サンプルをDNaseに接触させる工程、1つの接合部にて、複数のセグメントの第1のセグメントと第2のセグメントを付着させる工程、および、QC値を決定するために複数のセグメントを解析する工程を含む。場合によっては、QC値は、クロマチン消化効率(CDE)およびクロマチン消化指標(CDI)から選択される。CDEは、所望の長さを有するセグメントの割合として計算される。例えば、場合によっては、CDEは、サイズ選択の前にサイズが100~2500bpの間にあるセグメントの割合として計算される。場合によっては、CDE値が少なくとも65%である場合、サンプルはさらなる解析のために選択される。いくつかの実施形態では、サンプルは、CDE値が少なくとも約50%、少なくとも約55%、少なくとも約60%、少なくとも約65%、少なくとも約70%、少なくとも約75%、少なくとも約80%、少なくとも約85%、少なくとも約90%、または少なくとも約95%であるとき、さらなる解析のために選択される。CDIは、サイズ選択の前の、モノ

40

50

ヌクレオソームのサイズのセグメントの数の、ジヌクレオソームのサイズのセグメントの数の比として計算される。例えば、CDIは、600～2500bpのサイズを有する断片対100～600bpのサイズを有する断片の比の対数として計算され得る。場合によっては、サンプルは、CDI値が-1.5を超えかつ1未満であるとき、さらなる解析のために選択される。場合によっては、サンプルは、CDI値が約-2を超えかつ約1.5未満であるとき、約-1.9を超えかつ約1.5未満であるとき、約-1.8を超えかつ約1.5未満であるとき、約-1.7を超えかつ約1.5未満であるとき、約-1.6を超えかつ約1.5未満であるとき、約-1.5を超えかつ約1.5未満であるとき、約-1.4を超えかつ約1.5未満であるとき、約-1.3を超えかつ約1.5未満であるとき、約-1.2を超えかつ約1.5未満であるとき、約-1.1を超えかつ約1.5未満であるとき、約-2を超えかつ約1.5未満であるとき、約-1を超えかつ約1.5未満であるとき、約-0.9を超えかつ約1.5未満であるとき、約-0.8を超えかつ約1.5未満であるとき、約-0.7を超えかつ約1.5未満であるとき、約-0.6を超えかつ約1.5未満であるとき、約-0.5を超えかつ約1.5未満であるとき、約-2を超えかつ約1.4未満であるとき、約-2を超えかつ約1.3未満であるとき、約-2を超えかつ約1.2未満であるとき、約-2を超えかつ約1.1未満であるとき、約-2を超えかつ約1未満であるとき、約-2を超えかつ約0.9未満であるとき、約-2を超えかつ約0.8未満であるとき、約-2を超えかつ約0.7未満であるとき、約-2を超えかつ約0.6未満であるとき、または、約-2を超えかつ約0.5未満であるとき、さらなる解析のために選択される。

【0137】

他の態様では、本明細書におけるQC定量のステップを伴う方法は、複数の選択されたセグメントを得るために、複数のセグメントをサイズ選択にかける工程を含み得る。場合によっては、複数の選択されたセグメントは、約145～約600bpである。場合によっては、複数の選択されたセグメントは約100～約2500bpである。場合によっては、複数の選択されたセグメントは約100～約600bpである。場合によっては、複数の選択されたセグメントは約600～約2500bpである。場合によっては、場合によっては、複数の選択されたセグメントは、約100～約600bpの間、約100bp～約700bpの間、約100bp～約800bpの間、約100bp～約900bpの間、約100bp～約1000bpの間、約100bp～約1100bpの間、約100bp～約1200bpの間、約100bp～約1300bpの間、約100bp～約1400bpの間、約100bp～約1500bpの間、約100bp～約1600bpの間、約100bp～約1700bpの間、約100bp～約1800bpの間、約100bp～約1900bpの間、約100bp～約2000bpの間、約100bp～約2100bpの間、約100bp～約2200bpの間、約100bp～約2300bpの間、約100bp～約2400bpの間、または、約100bp～約2500bpの間である。

【0138】

本明細書において提供されるQC定量のステップを伴う方法の別の態様では、方法は、サイズ選択のステップに先立ち、複数のセグメントから配列決定ライブラリを調製する工程をさらに含み得る。いくつかの実施形態では、方法は、サイズ選択されたライブラリを得るために、配列決定ライブラリをサイズ選択にかける工程をさらに含む。場合によっては、サイズ選択されたライブラリは、約350bpから約1000bpの間のサイズにある。場合によっては、サイズ選択されたライブラリは、約100bpから約2500bpまでの間のサイズにあり、例えば、約100bp～約350bpの間、約350bp～約500bpの間、約500bp～約1000bpの間、約1000bp～約1500bpの間、約2000bpの間、約2000bp～約2500bpの間、約350bp～約1000bpの間、約350bp～約1500bpの間、約350bp～約2000bpの間、約350bp～約2500bpの間、約500bp～約1500bpの間、約500bp～約2000bpの間、約500bp～約3500bpの間、約1000bp～約1500bpの間、約1000bp～約2000bpの間、約1000bp～約2500bpの間、

間、約1500bp～約2000bpの間、約1500bp～約2500bpの間、または、約100bp～約700bpの間のサイズにある。

【0139】

本明細書における、QC定量のステップを伴う方法において利用されるサイズ選択は、ゲル電気泳動、キャピラリー電気泳動、サイズ選択ビーズ、ゲル濾過カラム、またはそれらの組み合わせを用いて行なわれる場合がある。他の適切なサイズ選択の方法もまたこの開示の範囲内にある。

【0140】

他の態様では、本明細書における、QC定量のステップを伴うことにおいて使用される安定化された生体サンプルは、安定化剤を用いて処置された生体材料を含む。場合によっては、安定化された生体サンプルは、安定化された細胞溶解物を含む。あるいは、安定化された生体サンプルは、安定化された無傷細胞を含む。あるいは、安定化された生体サンプルは、安定化された無傷の核を含む。場合によっては、安定化された無傷細胞または無傷の核サンプルをDNaseに接触させる工程は、無傷細胞または無傷の核の溶解に先立って行なわれる。場合によっては、細胞および／または核は、複数のセグメントの第1のセグメントと第2のセグメントとを接合部にて付着させる前に溶解される。

【0141】

他の態様では、本明細書における、QC定量のステップを伴う方法は、少数の細胞または小量の核酸しか包含していない小さなサンプルに対して行なわれる。場合によっては、安定化された生体サンプルは3,000,000未満の細胞を含む。場合によっては、安定化された生体サンプルは2,000,000未満の細胞を含む。場合によっては、安定化された生体サンプルは1,000,000未満の細胞を含む。場合によっては、安定化された生体サンプルは500,000未満の細胞を含む。場合によっては、安定化された生体サンプルは400,000未満の細胞を含む。場合によっては、安定化された生体サンプルは300,000未満の細胞を含む。場合によっては、安定化された生体サンプルは200,000未満の細胞を含む。場合によっては、安定化された生体サンプルは100,000未満の細胞を含む。場合によっては、安定化された生体サンプルは10μg未満のDNAを含む。場合によっては、安定化された生体サンプルは9μg未満のDNAを含む。場合によっては、安定化された生体サンプルは8μg未満のDNAを含む。場合によっては、安定化された生体サンプルは7μg未満のDNAを含む。場合によっては、安定化された生体サンプルは6μg未満のDNAを含む。場合によっては、安定化された生体サンプルは5μg未満のDNAを含む。場合によっては、安定化された生体サンプルは4μg未満のDNAを含む。場合によっては、安定化された生体サンプルは3μg未満のDNAを含む。場合によっては、安定化された生体サンプルは2μg未満のDNAを含む。場合によっては、安定化された生体サンプルは1μg未満のDNAを含む。場合によっては、安定化された生体サンプルは0.5μg未満のDNAを含む。

【0142】

他の態様では、本明細書におけるQC定量のステップを伴う方法は、個別の細胞または単細胞について行なうことができる。例えば、本明細書における方法は、個別の区画へと分配された細胞について行なわれる場合がある。典型的な区画として、限定されないが、ウェル、エマルジョン中の液滴、または表面位置（例えば、アレイ・スポット、ビーズなど）が挙げられ、それらの区画は本明細書に別記されるように差動的に配列決定されたリンカー分子の別個のパッチを含んでいる。付加的な区画も考えられ、および本明細書に開示される方法、組成物、およびシステムと一貫する。

【0143】

一態様において、本明細書における、QC定量のステップを伴うことにおいて使用される安定化された生体サンプルは、DNAの断片を作り出すために、DNaseなどのヌクレアーゼで処置される。場合によっては、DNaseは非配列特異性である。場合によっては、DNaseは、一本鎖DNAおよび二本鎖DNAの両方に対して活性がある。場合によっては、DNaseは、二本鎖DNAに対して特異的である。場合によっては、DN

10

20

30

40

50

a s e は、二本鎖 DNA を優先的に切断する。場合によっては、D N a s e は、一本鎖 DNA に対して特異的である。場合によっては、D N a s e は、一本鎖 DNA を優先的に切断する。場合によっては、D N a s e は D N a s e I である。場合によっては、D N a s e は D N a s e II である。場合によっては、D N a s e は、D N a s e I および D N a s e II の 1 つ以上から選択される。場合によっては、D N a s e はミクロコッカスヌクレアーゼである。場合によっては、D N a s e は、D N a s e I、D N a s e II、およびミクロコッカスヌクレアーゼの 1 つ以上から選択される。場合によっては、D N a s e は、架橋オリゴヌクレオチドは、プロテイン A、G タンパク質、タンパク質 A / G、またはタンパク質 L などの免疫グロブリン結合タンパク質またはそれらの断片に連結または融合され得る。他の適切なヌクレアーゼもまたこの開示の範囲内にある。

10

【0144】

付加的な態様では、本明細書における QC 定量ステップを伴う方法において使用される安定化された生体サンプルは、橋架剤で処置される。場合によっては、橋架剤は化学的固定剤である。場合によっては、化学的固定剤は、約 2.3 ~ 2.7 オングストローム (Å) のスペーサーアーム長を有するホルムアルデヒドを含む。場合によっては、化学的固定剤は、長いスペーサーアーム長を有する橋架剤を含み、例えば、橋架剤は、少なくとも長さ約 3 Å、4 Å、5 Å、6 Å、7 Å、8 Å、9 Å、10 Å、11 Å、12 Å、13 Å、14 Å、15 Å、16 Å、17 Å、18 Å、19 Å、または 20 Å のスペーサー長を有し得る。化学的固定剤は、エチレングリコールビス(コハク酸スクシンイミジル) (EGS) を含み、長さが約 16.1 Å のスペーサーアームを有する。化学的固定剤は、グルタル酸ジスクシンイミジル (DSG) を含み得、長さが約 7.7 Å のスペーサーアームを有する。場合によっては、化学的固定剤は、ホルムアルデヒドと EGS、ホルムアルデヒドと DSG、または、ホルムアルデヒドと EGS と DSG を含む。複数の化学的固定剤が利用されるいくつかの場合では、各化学的固定剤は順序的に使用され、他の場合では、複数の化学的な固定剤のうちのいくつかあるいはすべては、サンプルに同時に適用される。長いスペーサーアームを有する架橋剤の使用は、大きな(例えば、> 1 kb) リードペア分離距離を有するリードペアの断片を増加させ得る。例えば、図 7 は、結果として生じるライブラリ (DNアーゼの消化と MNase の消化の両方) について、ホルムアルデヒド単独で架橋されたものと、ホルムアルデヒドに DSG または EGS を加えたもので架橋されたものの比較を示す。DSG は両端部に NHSEステル反応基を有し、およびアミノ基(例えば、一級アミン) に対して反応性であり得る。DSG は、膜浸透性であり、細胞内の架橋を可能にしている。DSG は、スベリン酸ジサクシンイミジル (DSS) と比較して、いくつかの応用において、架橋効率を増大させることができる。EGS は両端部に NHSEステル反応基を有し、およびアミノ基(例えば、一級アミン) に対して反応性であり得る。EGS は、膜浸透性であり、細胞内の架橋を可能にしている。EGS 架橋は、例えば、pH 8.5 にて 3 ~ 6 時間ヒドロキシルアミンで処置することにより、逆行され得、一例では、ラクトースデヒドロゲナーゼは、EGS による可逆的な架橋の後に、その活性の 60 % を保持した。場合によっては、化学的固定剤はソラレンを含む。場合によっては、橋架剤は紫外線である。いくつかの実施形態では、安定化された生体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。

20

30

40

【0145】

さらなる実施形態では、本明細書において提供される QC 定量のステップを伴う方法は、複数の選択されたセグメントを抗体に接触させる工程を含む。

【0146】

付加的な態様において、本明細書における QC 定量のステップを伴う方法は、1 つの接合部にて、複数のセグメントの第 1 のセグメントと第 2 のセグメントを付着させる工程を含む。場合によっては、付着させる工程は、ビオチンをタグ付けされたヌクレオチドを使用して粘着末端を埋め、および平滑末端をライゲーションする工程を含む。場合によっては、付着させる工程は、少なくとも第 1 のセグメントと第 2 のセグメントを架橋オリゴヌクレオチドに接触させることを含む。場合によっては、付着させる工程は、少なくとも第

50

1のセグメントと第2のセグメントをバーコードに接触させることを含む。いくつかの実施形態では、本明細書における架橋オリゴヌクレオチドは、長さが少なくとも約5ヌクレオチドから約50ヌクレオチドまでであり得る。いくつかの実施形態では、本明細書における架橋オリゴヌクレオチドは、長さが少なくとも約15ヌクレオチドから約18ヌクレオチドまでであり得る。いくつかの実施形態では、架橋オリゴヌクレオチドは、長さが約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35、約40、約45、または約50ヌクレオチドである。いくつかの実施形態において、本明細書における架橋オリゴヌクレオチドは、バーコードを含み得る。いくつかの実施形態において、架橋オリゴヌクレオチドは、複数のバーコードを含み得る。いくつかの実施形態では、架橋オリゴヌクレオチドは、相互に接続している複数の架橋オリゴヌクレオチドを含む。いくつかの実施形態では、架橋オリゴヌクレオチドは、プロテインA、Gタンパク質、タンパク質A/G、またはタンパク質Lなどの免疫グロブリン結合タンパク質またはそれらの断片に連結されるかリンクされる場合がある。場合によっては、連結された架橋オリゴヌクレオチドは、サンプルの核酸における抗体が結合される位置に送達され得る。

【0147】

付加的な態様において、本明細書におけるQC定量のステップを伴う方法は、剪断するステップを含まない。

【0148】

本明細書におけるQC定量のステップを伴う方法のさらなる態様において、方法は、第1のリードペアを生成するために、接合部の両側で少なくともいくつかの配列を獲得する工程を含む。例えば、方法は、第1のリードペア生成するために、接合部の各側に、少なくとも約50bp、少なくとも約100bp、少なくとも約150bp、少なくとも約200bp、少なくとも約250bp、または少なくとも約300bpの配列を得る工程を含み得る。

【0149】

本明細書におけるQC定量のステップを伴う方法の付加的な態様において、方法は、1セットのコンティグに第1のリードペアをマッピングする工程、および、ゲノムへの順序および／または配向を表現するコンティグのセットにわたる経路を決定する工程をさらに含む。

【0150】

本明細書におけるQC定量のステップを伴う方法のさらなる態様において、方法は、コンティグのセットに第1のリードペアをマッピングする工程、および、前記コンティグのセットから、安定化された生体サンプルにおける構造的変異体の存在またはヘテロ結合性の減少を判定する工程を含み得る。

【0151】

本明細書におけるQC定量のステップを伴う方法の付加的な態様において、方法は、1セットのコンティグに第1のリードペアをマッピングする工程、およびコンティグのセットにおける変異体をフェーズに割り当てる工程を含む。

【0152】

本明細書におけるQC定量のステップを伴う方法のさらなる態様において、方法は、コンティグのセットに第1のリードペアをマッピングする工程、コンティグのセットから、コンティグのセットにおける変異体の存在を判定する工程、および、(1)安定化された生体サンプルについて疾患段階、予後、あるいは処置の方針を確認するステップ、(2)変異体の存在に基づいて、薬物を選択するステップ、または、(3)安定化された生体サンプルに対する薬効を確認するステップから選択される1つ以上のステップを行なう工程を含む。

【0153】

細胞全体または核全体の消化を含む、Hi-C法

本明細書において、方法がさらに提供され、該方法は、少なくとも1つの核酸結合タン

パク質に複合化された核酸分子を含む、安定化された生体サンプルを得る工程、核酸分子を複数のセグメントへと切断するために、安定化された生体サンプルをDNaseに接触させる工程、および、1つの接合部にて、複数のセグメントの第1のセグメントと第2のセグメントを付着させる工程、を含み、ここで、安定化された生体サンプルは、無傷細胞および／または完全な核を含む。場合によっては、安定化された生体サンプルは、安定化された無傷細胞を含む。代替的に、または組み合わせて、安定化された生体サンプルは、安定化された無傷の核を含む。場合によっては、安定化された無傷細胞または無傷の核サンプルをDNaseに接触させる工程は、無傷細胞または無傷の核の溶解に先立って行なわれる。場合によっては、細胞および／または核は、複数のセグメントの第1のセグメントと第2のセグメントとを接合部にて付着させる前に溶解される。

10

【0154】

他の態様では、本明細書における、細胞全体または核全体の消化を伴う方法は、複数の選択されたセグメントを得るために、複数のセグメントをサイズ選択にかけることを含み得る。場合によっては、複数の選択されたセグメントは、約145～約600bpである。場合によっては、複数の選択されたセグメントは約100～約2500bpである。場合によっては、複数の選択されたセグメントは約100～約600bpである。場合によっては、複数の選択されたセグメントは約600～約2500bpである。場合によっては、複数の選択されたセグメントは、約100bp～約600bpの間、約100bp～約700bpの間、約100bp～約800bpの間、約100bp～約900bpの間、約100bp～約1000bpの間、約100bp～約1100bpの間、約100bp～約1200bpの間、約100bp～約1300bpの間、約100bp～約1400bpの間、約100bp～約1500bpの間、約100bp～約1600bpの間、約100bp～約1700bpの間、約100bp～約1800bpの間、約100bp～約1900bpの間、約100bp～約2000bpの間、約100bp～約2100bpの間、約100bp～約2200bpの間、約100bp～約2300bpの間、約100bp～約2400bpの間、または、約100bp～約2500bpの間である。

20

【0155】

本明細書において提供される、細胞全体または核全体の消化を伴う方法の別の態様では、方法は、サイズ選択ステップに先立ち、複数のセグメントから配列決定ライブラリを調製する工程をさらに含む。いくつかの実施形態では、方法は、サイズ選択されたライブラリを得るために、配列決定ライブラリをサイズ選択にかける工程をさらに含む。場合によっては、サイズ選択されたライブラリは、約350bpから約1000bpの間のサイズにある。場合によっては、サイズ選択されたライブラリは、約100bp～約2500bpの間のサイズにあり、例えば、約100bp～約350bpの間、約350bp～約500bpの間、約500bp～約1000bpの間、約1000bp～約1500bp～約2000bpの間、約2000bp～約2500bpの間、約350bp～約1000bpの間、約350bp～約1500bpの間、約350bp～約2000bpの間、約350bp～約2500bpの間、約500bp～約1500bpの間、約500bp～約2000bpの間、約500bp～約3500bpの間、約1000bp～約1500bpの間、約1000bp～約2000bpの間、約1000bp～約2500bpの間、約1500bp～約2000bpの間、約1500bp～約2500bpの間、または、約100bp～約700bpの間のサイズにある。

30

40

【0156】

本明細書における、細胞全体または核全体の消化を伴う方法において利用されるサイズ選択は、ゲル電気泳動、キャピラリー電気泳動、サイズ選択ビーズ、ゲル濾過カラム、またはそれらの組み合わせにより行なわれ得る。

【0157】

他の態様では、本明細書における、細胞全体または核全体の消化を伴う方法は、1つのQC値を得るために、複数の選択されたセグメントを解析する工程をさらに含み得る。場合によっては、QC値は、クロマチン消化効率(CDE)およびクロマチン消化指標(C

50

D I) から選択される。C D E は、所望の長さを有するセグメントの割合として計算される。例えば、場合によっては、C D E は、サイズ選択の前にサイズが 1 0 0 ~ 2 5 0 0 b p の間にあるセグメントの割合として計算される。場合によっては、C D E 値が少なくとも 6 5 % である場合、サンプルはさらなる解析のために選択される。場合によっては、サンプルは、C D E 値が少なくとも約 5 0 %、少なくとも約 5 5 %、少なくとも約 6 0 %、少なくとも約 6 5 %、少なくとも約 7 0 %、少なくとも約 7 5 %、少なくとも約 8 0 %、少なくとも約 8 5 %、少なくとも約 9 0 %、または少なくとも約 9 5 % であるとき、さらなる解析のために選択される。C D I は、サイズ選択の前の、モノヌクレオソームのサイズのセグメントの数の、ジヌクレオソームのサイズのセグメントの数に対する比として計算される。例えば、C D I は、6 0 0 ~ 2 5 0 0 b p のサイズを有する断片対 1 0 0 ~ 6 0 0 b p のサイズを有する断片の比の対数として計算され得る。場合によっては、サンプルは、C D I 値が - 1 . 5 を超えかつ 1 未満であるとき、さらなる解析のために選択される。場合によっては、サンプルは、C D I 値が、約 - 2 を超えかつ約 1 . 5 未満であるとき、約 - 1 . 9 を超えかつ約 1 . 5 未満であるとき、約 - 1 . 8 を超えかつ約 1 . 5 未満であるとき、約 - 1 . 7 を超えかつ約 1 . 5 未満であるとき、約 - 1 . 6 を超えかつ約 1 . 5 未満であるとき、約 - 1 . 5 を超えかつ約 1 . 5 未満であるとき、約 - 1 . 4 を超えかつ約 1 . 5 未満であるとき、約 - 1 . 3 を超えかつ約 1 . 5 未満であるとき、約 - 1 . 2 を超えかつ約 1 . 5 未満であるとき、約 - 1 . 1 を超えかつ約 1 . 5 未満であるとき、約 - 2 を超えかつ約 1 . 5 未満であるとき、約 - 1 を超えかつ約 1 . 5 未満であるとき、約 - 0 . 9 を超えかつ約 1 . 5 未満であるとき、約 - 0 . 8 を超えかつ約 1 . 5 未満であるとき、約 - 0 . 7 を超えかつ約 1 . 5 未満であるとき、約 - 0 . 6 を超えかつ約 1 . 5 未満であるとき、約 - 0 . 5 を超えかつ約 1 . 5 未満であるとき、約 - 2 を超えかつ約 1 . 4 未満であるとき、約 - 2 を超えかつ約 1 . 3 未満であるとき、約 - 2 を超えかつ約 1 . 2 未満であるとき、約 - 2 を超えかつ約 1 . 1 未満であるとき、約 - 2 を超えかつ約 1 未満であるとき、約 - 2 を超えかつ約 0 . 9 未満であるとき、約 - 2 を超えかつ約 0 . 8 未満であるとき、約 - 2 を超えかつ約 0 . 7 未満であるとき、約 - 2 を超えかつ約 0 . 6 未満であるとき、または、約 - 2 を超えかつ約 0 . 5 未満であるとき、さらなる解析のために選択される。

【 0 1 5 8 】

他の態様では、本明細書における、細胞全体または核全体の消化を伴う方法は、少数の細胞または少量の核酸しか包含していない小さなサンプルに対して行なわれる。場合によっては、安定化された生体サンプルは 3 , 0 0 0 , 0 0 0 未満の細胞を含む。場合によっては、安定化された生体サンプルは 2 , 0 0 0 , 0 0 0 未満の細胞を含む。場合によっては、安定化された生体サンプルは 1 , 0 0 0 , 0 0 0 未満の細胞を含む。場合によっては、安定化された生体サンプルは 5 0 0 , 0 0 0 未満の細胞を含む。場合によっては、安定化された生体サンプルは 4 0 0 , 0 0 0 未満の細胞を含む。場合によっては、安定化された生体サンプルは 3 0 0 , 0 0 0 未満の細胞を含む。場合によっては、安定化された生体サンプルは 2 0 0 , 0 0 0 未満の細胞を含む。場合によっては、安定化された生体サンプルは 1 0 0 , 0 0 0 未満の細胞を含む。場合によっては、安定化された生体サンプルは 1 0 μ g 未満の D N A を含む。場合によっては、安定化された生体サンプルは 9 μ g 未満の D N A を含む。場合によっては、安定化された生体サンプルは 8 μ g 未満の D N A を含む。場合によっては、安定化された生体サンプルは 7 μ g 未満の D N A を含む。場合によっては、安定化された生体サンプルは 6 μ g 未満の D N A を含む。場合によっては、安定化された生体サンプルは 5 μ g 未満の D N A を含む。場合によっては、安定化された生体サンプルは 4 μ g 未満の D N A を含む。場合によっては、安定化された生体サンプルは 3 μ g 未満の D N A を含む。場合によっては、安定化された生体サンプルは 2 μ g 未満の D N A を含む。場合によっては、安定化された生体サンプルは 1 μ g 未満の D N A を含む。場合によっては、安定化された生体サンプルは 0 . 5 μ g 未満の D N A を含む。

【 0 1 5 9 】

他の態様では、本明細書における、細胞全体または核全体の消化を伴う方法は、個別の

細胞または単細胞について行なわれ得る。例えば、本明細書における方法は、個別の区画へと分配された細胞について行なわれる場合がある。典型的な区画として、限定されないが、ウェル、エマルジョン中の液滴、または表面位置（例えば、アレイ・スポット、ビーズなど）が挙げられ、それらの区画は本明細書に別記されるように差動的に配列決定されたリンカー分子の別個のパッチを含んでいる。付加的な区画も考えられ、および、本明細書に開示される方法、組成物、およびシステムと一貫している。

【0160】

付加的な態様において、本明細書における、細胞全体または核全体の消化を伴う方法において利用される安定化された生体サンプルは、DNAの断片を作り出すために、DNaseなどのヌクレアーゼで処置される。場合によっては、DNaseは非配列特異性である。場合によっては、DNaseは、一本鎖DNAおよび二本鎖DNAの両方に対して活性がある。場合によっては、DNaseは、二本鎖DNAに対して特異的である。場合によっては、DNaseは、二本鎖DNAを優先的に切断する。場合によっては、DNaseは、一本鎖DNAに対して特異的である。場合によっては、DNaseは、一本鎖DNAを優先的に切断する。場合によっては、DNaseはDNase Iである。場合によっては、DNaseはDNase IIである。場合によっては、DNaseは、DNase IおよびDNase IIの1つ以上から選択される。場合によっては、DNaseはミクロコッカスヌクレアーゼである。場合によっては、DNaseは、DNase I、DNase II、およびミクロコッカスヌクレアーゼの1つ以上から選択される。場合によっては、DNaseは、プロテインA、Gタンパク質、タンパク質A/G、またはタンパク質Lなどの免疫グロブリン結合タンパク質またはそれらの断片に連結されるか融合される場合がある。他の適切なヌクレアーゼもまたこの開示の範囲内にある。

【0161】

付加的な態様において、本明細書における、細胞全体または核全体の消化を伴う方法において利用される安定化されたサンプルは、橋架剤で処置される。場合によっては、橋架剤は化学的固定剤である。場合によっては、化学的固定剤は、約2.3〜2.7オングストローム（Å）のスペーサーアーム長を有するホルムアルデヒドを含む。場合によっては、化学的固定剤は、長いスペーサーアーム長を有する橋架剤を含み、例えば、橋架剤は、少なくとも長さ約3 Å、4 Å、5 Å、6 Å、7 Å、8 Å、9 Å、10 Å、11 Å、12 Å、13 Å、14 Å、15 Å、16 Å、17 Å、18 Å、19 Å、または20 Åのスペーサー長を有する場合がある。化学的固定剤は、エチレングリコールビス（コハク酸スクシンイミジル）（EGS）を含み、長さが約16.1 Åのスペーサーアームを有する。化学的固定剤は、グルタル酸ジスクシンイミジル（DSG）を含み得、長さが約7.7 Åのスペーサーアームを有する。場合によっては、化学的固定剤は、ホルムアルデヒドとEGS、ホルムアルデヒドとDSG、または、ホルムアルデヒドとEGSとDSGを含む。複数の化学的固定剤が利用されるいくつかの場合では、各化学的固定剤は順序的に使用され、他の場合では、複数の化学的な固定剤のうちのいくつかあるいはすべては、サンプルに同時に適用される。長いスペーサーアームを有する架橋剤の使用は、大きな（例えば、>1 kb）リードペア分離距離を有するリードペアの断片を増加させ得る。例えば、図7は、結果として生じるライブラリ（DNアーゼの消化とMNaseの消化の両方）について、ホルムアルデヒド単独で架橋されたものと、ホルムアルデヒドにDSGまたはEGSを加えたもので架橋されたものとの比較を示す。DSGは両端部にNHSEステル反応基を有し、およびアミノ基（例えば、一級アミン）に対して反応性であり得る。DSGは、膜浸透性であり、細胞内の架橋を可能にしている。DSGは、スベリン酸ジスクシンイミジル（DSS）と比較して、いくつかの応用において、架橋効率を増大させることができる。EGSは両端部にNHSEステル反応基を有し、およびアミノ基（例えば、一級アミン）に対して反応性であり得る。EGSは、膜浸透性であり、細胞内の架橋を可能にしている。EGS架橋は、例えば、pH 8.5にて3〜6時間ヒドロキシルアミンで処置することにより、逆行され得、一例では、ラクトースデヒドロゲナーゼは、EGSによる可逆的な架橋の後に、その活性の60%を保持した。場合によっては、化学的固定剤はソラレンを

10

20

30

40

50

含む。場合によっては、橋架剤は紫外線である。いくつかの実施形態では、安定化された生体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。

【0162】

さらなる実施形態では、本明細書における、細胞全体または核全体の消化を伴う方法は、複数の選択されたセグメントを抗体に接触させる工程を含む。

【0163】

付加的な態様において、本明細書において提供される、細胞全体のまたは核全体の消化を伴う方法は、1つの接合部にて、複数のセグメントの第1のセグメントと第2のセグメントを付着させる工程を含む。場合によっては、付着させる工程は、ビオチンをタグ付けされたヌクレオチドを使用して粘着末端を埋め、および平滑末端をライゲーションする工程を含む。場合によっては、付着させる工程は、少なくとも第1のセグメントと第2のセグメントを架橋オリゴヌクレオチドに接触させることを含む。場合によっては、付着させる工程は、少なくとも第1のセグメントと第2のセグメントをバーコードに接触させることを含む。いくつかの実施形態では、本明細書における架橋オリゴヌクレオチドは、長さが少なくとも約5ヌクレオチドから約50ヌクレオチドまでであり得る。いくつかの実施形態では、本明細書における架橋オリゴヌクレオチドは、長さが少なくとも約15ヌクレオチドから約18ヌクレオチドまでであり得る。いくつかの実施形態において、架橋オリゴヌクレオチドは、長さが、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35、約40、約45、または約50ヌクレオチドである。いくつかの実施形態において、本明細書における架橋オリゴヌクレオチドは、バーコードを含み得る。いくつかの実施形態において、架橋オリゴヌクレオチドは、複数のバーコードを含み得る。いくつかの実施形態では、架橋オリゴヌクレオチドは、相互に接続している複数の架橋オリゴヌクレオチドを含む。いくつかの実施形態では、架橋オリゴヌクレオチドは、プロテインA、Gタンパク質、タンパク質A/G、またはタンパク質Lなどの免疫グロブリン結合タンパク質またはそれらの断片に連結されるかリンクされる場合がある。場合によっては、連結された架橋オリゴヌクレオチドは、サンプルの核酸における抗体が結合される位置に送達され得る。

【0164】

スプリッティングとプーリングのアプローチは、固有のバーコードを有する架橋オリゴヌクレオチドを生み出すために利用することができる。サンプルの集団は複数の群へと分割することができ、架橋オリゴヌクレオチドは、架橋オリゴヌクレオチドのバーコードが群の間で異なるが、群内では同じであるように、サンプルに付着させることができ、サンプルの群は、再び一緒にプールすることができ、および、このプロセスは複数回繰り返すことができる。このプロセスを繰り返すことは、固有の一連の架橋オリゴヌクレオチド・バーコードを有する集団における各サンプルを結果として生じ、単一のサンプル（例えば、単細胞、単一の核、単一の染色体）の解析を可能にする場合がある。1つの実証的な例では、1つの実証的な例において、ビーズの固体支持体に付着された、架橋され消化された核のサンプルは、8本の管にわたって分離され、各々、ライゲーションされる二本鎖DNA (dsDNA) アダプターを含む第1のアダプター群（第1の反復）の8つの固有のメンバのうちの1つをそれぞれ包含する。8つのアダプターの各々は、核中の架橋クロマチン集合体の核酸末端へのライゲーションのための同じ5'オーバーハング配列を有する場合があるが、そうでなければ、固有のdsDNA配列を有する。第1のアダプター群がライゲーションされた後、前記核は再び一緒にプールされ、および、ライゲーション反応成分を除去するために洗浄され得る。分配し、ライゲーションし、プールするスキームは、さらに2回反復され得る（2つの反復）。各アダプター群のメンバのライゲーションの後、架橋クロマチン集合体は、一連の複数のバーコードに付着され得る。場合によっては、複数のアダプター群（反復）の複数のメンバの連続するライゲーションは、結果としてバーコードの組み合わせを生じる。利用可能なバーコードの組み合わせの数は、1つの反復当たりの群の数、および使用されたバーコード・オリゴヌクレオチドの合計数に依存する

10

20

30

40

50

。例えば、8つのメンバを含む3つの反復は、83の可能性がある組み合わせを個々に持ち得る。場合によっては、バーコードの組み合わせは固有である。場合によっては、バーコードの組み合わせには冗長性がある。バーコードの組み合わせの総数は、固有のバーコードを受け取る群の数を増加あるいは減少させること、および／または、反復の数を増加あるいは減少させることにより調整することができる。1つを超えるアダプター群が使用される場合、分配し、付随させ、プールするスキームが、反復するアダプター付着のために使用され得る。場合によっては、分配し、付着させ、プールするスキームは、少なくとも3、4、5、6、7、8、9、または10回、追加で反復され得る。場合によっては、最後のアダプター群のメンバは、例えば、PCR増幅による配列決定ライブラリの調製の間における、後の、アダプター付着DNAの濃縮のための配列を含む。

10

【0165】

付加的な態様では、本明細書における、細胞全体または核全体の消化を伴う方法は、剪断するステップを含まない。

【0166】

本明細書において提供される、細胞全体または核全体の消化を伴う方法のさらなる態様では、方法は、第1のリードペアを生成するために、接合部の両側で少なくともいくつかの配列を獲得する工程を含む。例えば、方法は、第1のリードペア生成するために、接合部の各側に、少なくとも約50bp、少なくとも約100bp、少なくとも約150bp、少なくとも約200bp、少なくとも約250bp、または少なくとも約300bpの配列を得る工程を含み得る。

20

【0167】

本明細書において提供される、細胞全体または核全体の消化を伴う方法の付加的な態様では、方法は、1セットのコンティグに第1のリードペアをマッピングする工程、および、ゲノムへの順序および／または配向を表現するコンティグのセットにわたる経路を決定する工程をさらに含む。

【0168】

本明細書において提供される、細胞全体または核全体の消化を伴う方法のさらなる態様では、方法は、コンティグのセットに第1のリードペアをマッピングする工程、および、前記コンティグのセットから、安定化された生体サンプルにおける構造的変異体の存在またはヘテロ結合性の減少を判定する工程を含む。

30

【0169】

本明細書において提供される、細胞全体または核全体の消化を伴う方法の付加的な態様では、方法は、1セットのコンティグに第1のリードペアをマッピングする工程、およびコンティグのセットにおける変異体をフェーズに割り当てる工程を含む。

【0170】

本明細書において提供される、細胞全体または核全体の消化を伴う方法のさらなる態様では、方法は、コンティグのセットに第1のリードペアをマッピングする工程、コンティグのセットから、コンティグのセットにおける変異体の存在を判定する工程、および、(1) 安定化された生体サンプルについて疾患段階、予後、あるいは処置の方針を確認するステップ、(2) 変異体の存在に基づいて、薬物を選択するステップ、または、(3) 安定化された生体サンプルに対する薬効を確認するステップから選択される1つ以上のステップを行なう工程をさらに含む。

40

【0171】

低核酸投入量要件を有するHi-C方法

加えて、方法が本明細書に提供され、当該方法は、少なくとも1つの核酸結合タンパク質に複合体化された核酸分子を含む安定化された生体サンプルを得る工程と、核酸分子を複数のセグメントに切断するために、安定化された生体サンプルをDNaseに接触させる工程と、接合部で、複数のセグメントのうちの第1のセグメントと第2のセグメントとを付着させる工程であって、ここで、安定化された生体サンプルは、3,000,000未満の細胞または10μg未満のDNAを含む、工程と、を含む。場合によっては、安定

50

化された生体サンプルは、3,000,000未満の細胞を含む。場合によっては、安定化された生体サンプルは、2,000,000未満の細胞を含む。場合によっては、安定化された生体サンプルは、1,000,000未満の細胞を含む。場合によっては、安定化された生体サンプルは、500,000未満の細胞を含む。場合によっては、安定化された生体サンプルは、400,000未満の細胞を含む。場合によっては、安定化された生体サンプルは、300,000未満の細胞を含む。場合によっては、安定化された生体サンプルは、200,000未満の細胞を含む。場合によっては、安定化された生体サンプルは、100,000未満の細胞を含む。場合によっては、安定化された生体サンプルは、10 μ g未満のDNAを含む。場合によっては、安定化された生体サンプルは、9 μ g未満のDNAを含む。場合によっては、安定化された生体サンプルは、8 μ g未満のDNAを含む。場合によっては、安定化された生体サンプルは、7 μ g未満のDNAを含む。場合によっては、安定化された生体サンプルは、6 μ g未満のDNAを含む。場合によっては、安定化された生体サンプルは、5 μ g未満のDNAを含む。場合によっては、安定化された生体サンプルは、4 μ g未満のDNAを含む。場合によっては、安定化された生体サンプルは、3 μ g未満のDNAを含む。場合によっては、安定化された生体サンプルは、2 μ g未満のDNAを含む。場合によっては、安定化された生体サンプルは、1 μ g未満のDNAを含む。場合によっては、安定化された生体サンプルは、0.5 μ g未満のDNAを含む。

【0172】

別の態様では、本明細書における低核酸投入量要件を有する方法が、個々の細胞または単一の細胞上で実行されてもよい。例えば、本明細書における方法は、個々の区画に分配された細胞上で実行されてもよい。例示的な区画の例としては、ウェル、エマルジョン中の液滴、または本明細書に別記される通りの異なる配列のリンカー分子の個別のパッチを含む面位置（例えばアレイスポット、ビーズなど）を含むがこれらに限定されない。追加的な区画も企図され、それは、本明細書に開示される方法、組成物、およびシステムと一致している。

【0173】

別の態様では、本明細書における低核酸投入量要件を有する方法は、複数の選択されたセグメントを得るために、複数のセグメントをサイズ選択にかける工程を含む。場合によっては、複数の選択されたセグメントは、約145～約600bpである。場合によっては、複数の選択されたセグメントは、約100～約2500bpである。場合によっては、複数の選択されたセグメントは、約100～約600bpである。場合によっては、複数の選択されたセグメントは、約600～約2500bpである。場合によっては、複数の選択されたセグメントは、約100bp～約600bp、約100bp～約700bp、約100bp～約800bp、約100bp～約900bp、約100bp～約1000bp、約100～約1100bp、約100bp～約1200bp、約100bp～約1300bp、約100bp～約1400bp、約100bp～約1500bp、約100bp～約1600bp、約100bp～約1700bp、約100bp～約1800bp、約100bp～約1900bp、約100bp～約2000bp、約100bp～約2100bp、約100bp～約2200bp、約100bp～約2300bp、約100bp～約2400bp、または約100bp～約2500bpである。

【0174】

本明細書に提供される、低核酸投入量要件を有する方法の別の態様では、方法は、サイズ選択工程前に、複数のセグメントから配列決定ライブラリを調製する工程を、さらに含む。いくつかの実施形態では、方法は、サイズ選択されたライブラリを得るために、配列決定ライブラリをサイズ選択にかける工程を、さらに含む。場合によっては、サイズ選択されたライブラリは、約350bp～約1000bpのサイズである。場合によっては、サイズ選択されたライブラリは、約100bp～約2500bpのサイズであり、例えば、約100bp～約350bp、約350bp～約500bp、約500bp～約1000bp、約1000～約1500bp～約2000bp、約2000～約2500bp、

約350bp～約1000bp、約350bp～約1500bp、約350bp～約2000bp、約350bp～約2500bp、約500bp～約1500bp、約500bp～約2000bp、約500bp～約3500bp、約1000bp～約1500bp、約1000bp～約2000bp、約1000bp～約2500bp、約1500bp～約2000bp、約1500bp～約2500bp、または約2000bp～約2500bpである。

【0175】

本明細書における低核酸投入量要件を有する方法で利用されるサイズ選択は、ゲル電気泳動、キャピラリー電気泳動、サイズ選択ビーズ、ゲル濾過カラム、またはそれらの組み合わせで、しばしば実行される。

【0176】

別の態様では、本明細書における低核酸投入量要件を有する方法は、QC値を得るために、複数の選択されたセグメントを解析する工程を、さらに含んでもよい。場合によっては、QC値は、クロマチン消化効率(CDE)とクロマチン消化指標(CDI)とから選択される。CDEは、所望の長さを有するセグメントの割合として計算される。例えば、場合によっては、CDEは、サイズ選択前の、100～2500bpのサイズのセグメントの割合として計算される。場合によっては、CDE値が少なくとも65%である時、サンプルはさらなる解析用に選択される。場合によっては、CDE値が少なくとも約50%、少なくとも約55%、少なくとも約60%、少なくとも約65%、少なくとも約70%、少なくとも約75%、少なくとも約80%、少なくとも約85%、少なくとも約90%、または少なくとも約95%である時、サンプルはさらなる解析用に選択される。CDIは、サイズ選択前の、モノヌクレオソーム・サイズのセグメントの数対ジヌクレオソーム・サイズのセグメントの数の割合として、計算される。例えば、CDIは、600～2500bpのサイズを有する断片対100～600bpのサイズを有する断片の割合の対数として、計算されてもよい。場合によっては、CDI値が-1.5を超え且つ1未満である時、サンプルはさらなる解析用に選択される。場合によっては、CDI値が約-2を超え且つ約1.5未満、約-1.9を超え且つ約1.5未満、約-1.8を超え且つ約1.5未満、約-1.7を超え且つ約1.5未満、約-1.6を超え且つ約1.5未満、約-1.5を超え且つ約1.5未満、約-1.4を超え且つ約1.5未満、約-1.3を超え且つ約1.5未満、約-1.2を超え且つ約1.5未満、約-1.1を超え且つ約1.5未満、約-2を超え且つ約1.5未満、約-1を超え且つ約1.5未満、約-0.9を超え且つ約1.5未満、約-0.8を超え且つ約1.5未満、約-0.7を超え且つ約1.5未満、約-0.6を超え且つ約1.5未満、約-0.5を超え且つ約1.5未満、約-2を超え且つ約1.4未満より、約-2を超え且つ約1.3未満より、約-2を超え且つ約1.2未満、約-2を超え且つ約1.1未満、約-2を超え且つ約1未満、約-2を超え且つ約0.9未満、約-2を超え且つ約0.8未満、約-2を超え且つ約0.7未満、約-2を超え且つ約0.6未満、約-2を超え且つ約0.5未満である時、サンプルはさらなる解析用に選択される。

【0177】

別の態様では、本明細書における低核酸投入量要件を有する方法で使用される安定化された生体サンプルは、安定剤で処理された生体物質を含む。場合によっては、安定化された生体サンプルは、安定された細胞可溶化物を含む。あるいは、安定化された生体サンプルは、安定化された無傷の細胞を含む。あるいは、安定化された生体サンプルは、安定化された無傷の核を含む。場合によっては、無傷の細胞または無傷の核の溶解前に、安定化された無傷の細胞または無傷の核のサンプルをDNaseに接触させる工程が、実行される。場合によっては、細胞および／または核は、接合部で複数のセグメントのうちの第1のセグメントと第2のセグメントとを付着させる工程の前に、溶解される。

【0178】

追加的な態様では、本明細書における低核酸投入量要件を有する方法で使用される安定化された生体サンプルは、DNAの断片を作り出すために、DNaseなどのヌクレアー

10

20

30

40

50

ぜで処理される。場合によっては、DNaseは、配列特異的ではない。場合によっては、DNaseは、一本鎖DNAと二本鎖DNAの両方に対して活性である。場合によっては、DNaseは、二本鎖DNAに対して特異的である。場合によっては、DNaseは、優先的に二本鎖DNAを切断する。場合によっては、DNaseは、一本鎖DNAに対して特異的である。場合によっては、DNaseは、優先的に一本鎖DNAを切断する。場合によっては、DNaseは、DNase Iである。場合によっては、DNaseは、DNase IIである。場合によっては、DNaseは、DNase IとDNase IIの1以上から選択される。場合によっては、DNaseは、小球菌ヌクレアーゼである。場合によっては、DNaseは、DNase I、DNase II、および小球菌ヌクレアーゼの1以上から選択される。場合によっては、DNaseは、タンパク質A、タンパク質G、タンパク質A/G、またはタンパク質Lなどの、免疫グロブリン結合タンパク質またはその断片に、連結されてもよく、または融合されてもよい。他の適切なヌクレアーゼも、本開示の範囲内である。

【0179】

追加的な態様では、本明細書における低核酸投入量要件を有する方法で使用される安定化された生体サンプルは、架橋剤で処理される。場合によっては、架橋剤は、化学固定液である。場合によっては、化学固定液は、ホルムアルデヒドを含み、約2.3〜2.7オングストローム(A)のスペーサーアーム長を有する。場合によっては、化学固定液は、長いスペーサーアーム長を有する架橋剤を含み、例えば、架橋剤は、少なくとも約3A、4A、5A、6A、7A、8A、9A、10A、11A、12A、13A、14A、15A、16A、17A、18A、19A、または20Aのスペーサー長を有し得る。化学固定液は、エチレングリコールビス(コハク酸スクシンイミジル)(EGS)を含むことができ、これは、約16.1Aの長さを有するスペーサーアームを有する。化学固定液は、グルタル酸ジスクシンイミジル(DSG)を含むことができ、これは、約7.7Aの長さを有するスペーサーアームを有する。場合によっては、化学固定液は、ホルムアルデヒドとEGS、ホルムアルデヒドとDSG、または、ホルムアルデヒド、EGS、およびDSGを含む。複数の化学固定液が使用されるいくつかの場合には、各化学固定液は、連続的に使用される；他の場合では、複数の化学固定液のいくつかまたは全ては、サンプルに同時に適用される。長いスペーサーアームを有する架橋剤の使用は、大きな(例えば>1kbの)リードペアの分離距離を有するリードペアの画分を増加させることができる。例えば、図7は、ホルムアルデヒドのみで架橋された(消化されたDNaseとMNaseの両方)結果のライブラリと、ホルムアルデヒドとDSGまたはEGSとで架橋された結果のライブラリとの、比較を示す。DSGは、両端にNHSエステル反応基を有し、そしてアミノ基(例えば、一級アミン)に対して反応性を有し得る。DSGは、膜透過性を有し、細胞内架橋を可能にする。DSGは、いくつかの用途において、スベリン酸ジスクシンイミジル(DSS)と比較して、架橋効率を増加させることができる。EGSは、両端にNHSエステル反応基を有し、そしてアミノ基(例えば、一級アミン)に対して反応性を有し得る。EGSは、膜透過性を有し、細胞内架橋を可能にする。例えばpH8.5で3〜6時間ヒドロキシルアミンで処理することによって、EGS架橋を逆転させることができる；一例では、乳酸脱水素酵素は、EGSとの可逆的な架橋後に、その活性の60%を保持した。場合によっては、化学固定液は、ソラレンを含む。場合によっては、架橋剤は、紫外線である。場合によっては、安定化された生体サンプルは、架橋結合されたパラフィン包埋組織サンプルである。

【0180】

さらなる態様では、本明細書に提供される方法は、複数の選択されたセグメントを抗体に接触させる工程を含む。

【0181】

追加的な態様では、本明細書に提供される低核酸投入量要件を有する方法は、接合部で複数のセグメントのうちの第1のセグメントと第2のセグメントとを付着させる工程を含む。場合によっては、付着させる工程は、ビオチンタグ付きヌクレオチドを使用して粘着

10

20

30

40

50

性末を埋めること、および平滑末端をライゲーションすることを含む。場合によっては、付着させる工程は、少なくとも第1のセグメントと第2のセグメントとを、架橋オリゴヌクレオチドに接触させることを含む。場合によっては、付着させる工程は、少なくとも第1のセグメントと第2のセグメントとを、バーコードに接触させることを含む。いくつかの実施形態では、本明細書における架橋オリゴヌクレオチドは、少なくとも約5ヌクレオチドの長さから約50ヌクレオチドの長さであってもよい。いくつかの実施形態では、本明細書における架橋オリゴヌクレオチドは、約15から約18ヌクレオチドの長さであってもよい。いくつかの実施形態では、架橋オリゴヌクレオチドは、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35、約40、約45、または約50ヌクレオチドの長さであってもよい。いくつかの実施形態では、本明細書における架橋オリゴヌクレオチドは、バーコードを含んでもよい。いくつかの実施形態では、架橋オリゴヌクレオチドは、複数のバーコードを含み得る。いくつかの実施形態では、架橋オリゴヌクレオチドは、ともに接続している複数の架橋オリゴヌクレオチドを含む。いくつかの実施形態では、架橋オリゴヌクレオチドは、タンパク質A、タンパク質G、タンパク質A/G、またはタンパク質Lなどの、免疫グロブリン結合タンパク質またはその断片に、連結されてもよく、またはリンクされてもよい。場合によっては、連結された架橋オリゴヌクレオチドは、サンプル核酸中の位置に送達されてもよく、そこで抗体が結合される。

【0182】

スプリッティングとプーリングのアプローチは、固有のバーコードを有する架橋オリゴヌクレオチドを生成するために使用され得る。サンプルの母集団は複数のグループに分割され得、架橋オリゴヌクレオチドは、架橋ヌクレオチドバーコードがグループ間では異なるが、1つのグループ内では同じになるようにサンプルに付着され得、サンプルのグループは再び共にプーリングされ得、そしてこの処理が複数回反復され得る。この処理を繰り返すことで、最終的に、集団内の各サンプルが固有の一連の架橋オリゴヌクレオチド・バーコードを有するようになり、単一サンプル（例えば単一の細胞、単一の核、単一の染色体）の解析が可能になる。1つの例示的な例では、ビーズの固形サポートに取り付けられた、架橋結合された消化された核のサンプルは8つの管にわたって分割され、それぞれが、ライゲーションされる二本鎖DNA（dsDNA）アダプターを含む、第1のアダプタグループ（第1の繰り返し）8つの固有のメンバのうちの1つを含有する。8つのアダプターの各々は、核中の架橋結合されたクロマチン集合体の核酸末端にライゲーションするための、同じ5'オーバーハング配列を有することができるが、それ以外の場合には、固有のdsDNA配列を有する。第1のアダプタグループがライゲーションされた後、核は再度共にプーリングされ得、ライゲーション反応成分を取り除くために洗浄され得る。分配、ライゲーション、およびプーリングのスキームは、2回追加的に反復（2回の繰り返し）され得る。各アダプタグループからのメンバのライゲーション後に、架橋されたクロマチン集合体は、複数のバーコードに順次取り付けられ得る。場合によっては、複数のアダプタグループの複数のメンバの連続的なライゲーション（繰り返し）は、結果的に、バーコードの組み合わせをもたらす。利用可能なバーコードの組み合わせの数は、1回の繰り返し当たりのグループの数と、使用されたバーコード・オリゴヌクレオチドの総数とに依存する。例えば、各々8つのメンバを含む3回の繰り返しは、83の可能な組み合わせを有し得る。場合によっては、バーコードの組み合わせは固有である。場合によっては、バーコードの組み合わせは冗長である。バーコードの組み合わせの総数は、固有のバーコードを受け取るグループの数を増加または減少させることによって、および／または、繰り返しの数を増加または減少させることによって、調節され得る。1を超えるアダプタグループが使用される時、分配、付着、およびプーリングのスキームは、繰り返しのアダプター付着に使用され得る。場合によっては、分配、付着、およびプーリングのスキームは、少なくとも3回、4回、5回、6回、7回、8回、9回、または10回追加的に反復され得る。場合によっては、最後のアダプタグループのメンバは、その後の、例えばPCR増幅を通じた配列決定ライブラリの調製中の、アダプター付着DNAの濃縮のための、配列

10

20

30

40

50

を含む。

【0183】

追加的な態様では、本明細書における低核酸投入量要件を有する方法は、剪断工程を含まない。

【0184】

本明細書における低核酸投入量要件を有する方法のさらなる態様では、方法は、第1のリードペアを生成するために、接合部の各側面で、少なくとも何らかの配列を得る工程を含む。例えば、方法は、第1のリードペアを生成するために、接合部の各側面で、少なくとも約50bp、少なくとも約100bp、少なくとも約150bp、少なくとも約200bp、少なくとも約250bp、または少なくとも約300bpの配列を得る工程を含んでもよい。

10

【0185】

本明細書における低核酸投入量要件を有する方法の追加的な態様では、方法は、第1のリードペアをコンティグのセットにマッピングする工程と、順序および／または配向を表すコンティグのセットを通してゲノムに至る経路を判定する工程とを含む。

【0186】

本明細書における低核酸投入量要件を有する方法のさらなる態様では、方法は、第1のリードペアをコンティグのセットにマッピングする工程と、コンティグのセットから、安定化された生体サンプルにおける構造変異体の存在またはヘテロ接合性の喪失を判定する工程とを含む。

20

【0187】

本明細書における低核酸投入量要件を有する方法の追加的な態様では、方法は、第1のリードペアをコンティグのセットにマッピングする工程と、コンティグのセット中の変異体をフェーズに割り当てる工程とを含む。

【0188】

本明細書における低核酸投入量要件を有する方法のさらなる態様では、方法は、第1のリードペアをコンティグのセットにマッピングする工程と、コンティグのセットから、コンティグのセット中の変異体の存在を判定する工程と、(1)安定化された生体サンプルに対して、疾患段階、予後、または処置の方針を同定すること、(2)変異体の存在に基づいて、薬物を選択すること、または、(3)安定化された生体サンプルに対する、薬物有効性を同定すること、の1以上から選択されるステップを実施する工程と、を含む。

30

【0189】

小球菌ヌクレアーゼ(MNase)を使用するHi-C方法

加えて、方法が本明細書に提供され、当該方法は、少なくとも1つの核酸結合タンパク質に複合体化された核酸分子を含む安定化された生体サンプルを得る工程と、核酸分子を複数のセグメントに切断するために、安定化された生体サンプルを小球菌ヌクレアーゼ(MNase)に接触させる工程と、接合部で複数のセグメントのうちの第1のセグメントと第2のセグメントとを付着させる工程と、を含んでもよい。本明細書における方法におけるMNaseの使用は、例えば、MNaseがDNA結合タンパク質に結合していないすべての塩基のペアを切断し得るために、どこで、DNA結合タンパク質が最大で単一の塩基のペアの分解能でクロマチンに結合しているのかに関する情報を、提供してもよい。加えて、MNase消化の使用は、三次元のクロマチン構造情報を解読するために、接触マップとトポロジ的に関連するドメインとを作り出すことを可能にしてもよい。場合によっては、MNaseは、タンパク質A、タンパク質G、タンパク質A/G、またはタンパク質Lなどの、免疫グロブリン結合タンパク質またはその断片に、連結されてもよく、または融合されてもよい。

40

【0190】

例えば、MNase Hi-C方法は、約1bp、2bp、3bp、4bp、5bp、6bp、7bp、8bp、9bp、10bp、20bp、30bp、40bp、50bp、60bp、70bp、80bp、90bp、100bp、200bp、300bp、4

50

00bp、500bp、600bp、700bp、800bp、900bp、1000bp、2000bp、3000bp、4000bp、5000bp、6000bp、7000bp、8000bp、9000bp、10kb、20kb、30kb、40kb、50kb、60kb、70kb、80kb、90kb、または100kb未満、あるいはそれらに等しい分解能で、タンパク質結合またはゲノム接触相互作用の位置を提供することができる。場合によっては、タンパク質結合部位、タンパク質フットプリント、接触相互作用、または他の特徴は、1000bp内、900bp内、800bp内、700bp内、600bp内、500bp内、400bp内、300bp内、200bp内、190bp内、180bp内、170bp内、160bp内、150bp内、140bp内、130bp内、120bp内、110bp内、100bp内、90bp内、80bp内、70bp内、60bp内、50bp内、40bp内、30bp内、20bp内、10bp内、9bp内、8bp内、7bp内、6bp内、5bp内、4bp内、3bp内、2bp内、1bp内でマッピングされ得る。

【0191】

ある態様では、本明細書における低核酸MNase消化工程を含む方法は、複数の選択されたセグメントを得るために、複数のセグメントをサイズ選択にかける工程を、さらに含んでもよい。場合によっては、複数の選択されたセグメントは、約145～約600bpであり得る。場合によっては、複数の選択されたセグメントは、約100～約2500bpであり得る。場合によっては、複数の選択されたセグメントは、約100～約600bpであり得る。場合によっては、複数の選択されたセグメントは、約600～約2500bpであり得る。場合によっては、複数の選択されたセグメントは、約100bp～約600bp、約100bp～約700bp、約100bp～約800bp、約100bp～約900bp、約100bp～約1000bp、約100～約1100bp、約100bp～約1200bp、約100bp～約1300bp、約100bp～約1400bp、約100bp～約1500bp、約100bp～約1600bp、約100bp～約1700bp、約100bp～約1800bp、約100bp～約1900bp、約100bp～約2000bp、約100bp～約2100bp、約100bp～約2200bp、約100bp～約2300bp、約100bp～約2400bp、または約100bp～約2500bpであり得る。

【0192】

本明細書に提供される通りのMNase消化工程を含む方法の別の態様では、方法は、複数のセグメントからの配列決定ライブラリを調製する工程を、さらに含んでもよい。いくつかの実施形態では、方法は、サイズ選択されたライブラリを得るために、配列決定ライブラリをサイズ選択にかける工程を、さらに含んでもよい。場合によっては、サイズ選択されたライブラリは、約350bp～約1000bpのサイズであってもよい。場合によっては、サイズ選択されたライブラリは、約100bp～約2500bpのサイズであってもよく、例えば、約100bp～約350bp、約350bp～約500bp、約500bp～約1000bp、約1000～約1500bp、約2000～約2500bp、約350bp～約1000bp、約350bp～約1500bp、約350bp～約2000bp、約350bp～約2500bp、約500bp～約1500bp、約500bp～約2000bp、約500bp～約3500bp、約1000bp～約1500bp、約1000bp～約2000bp、約1000bp～約2500bp、約1500bp～約2000bp、約1500bp～約2500bp、または約2000bp～約2500bpであってもよい。

【0193】

別の態様では、本明細書に提供される通りのMNase消化工程を含む方法は、QC値を得るために、複数のセグメントを解析する工程を、さらに含むことができる。場合によっては、QC値は、クロマチン消化効率(CDE)とクロマチン消化指標(CDI)とから選択されてもよい。CDEは、所望の長さを有するセグメントの割合として計算され得る。例えば、場合によっては、CDEは、サイズ選択前の、100bp～2500bpの

10

20

30

40

50

サイズのセグメントの割合として計算され得る。場合によっては、C D E 値が少なくとも 6 5 % である時、サンプルはさらなる解析用に選択されてもよい。場合によっては、C D E 値が少なくとも約 5 0 %、少なくとも約 5 5 %、少なくとも約 6 0 %、少なくとも約 6 5 %、少なくとも約 7 0 %、少なくとも約 7 5 %、少なくとも約 8 0 %、少なくとも約 8 5 %、少なくとも約 9 0 %、または少なくとも約 9 5 % である時、サンプルはさらなる解析用に選択されてもよい。

【0194】

C D I は、サイズ選択前の、モノヌクレオソーム・サイズのセグメントの数対ジヌクレオソーム・サイズのセグメントの数の割合として、計算され得る。例えば、C D I は、6 0 0 - 2 5 0 0 b p のサイズを有する断片対 1 0 0 - 6 0 0 b p のサイズを有する断片の割合の対数として、計算されてもよい。場合によっては、C D I 値が - 1 . 5 を超え且つ 1 未満である時、サンプルはさらなる解析用に選択されてもよい。場合によっては、C D I 値が約 - 2 を超え且つ約 1 . 5 未満、約 - 1 . 9 を超え且つ約 1 . 5 未満、約 - 1 . 8 を超え且つ約 1 . 5 未満、約 - 1 . 7 を超え且つ約 1 . 5 未満、約 - 1 . 6 を超え且つ約 1 . 5 未満、約 - 1 . 5 を超え且つ約 1 . 5 未満、約 - 1 . 4 を超え且つ約 1 . 5 未満、約 - 1 . 3 を超え且つ約 1 . 5 未満、約 - 1 . 2 を超え且つ約 1 . 5 未満、約 - 1 . 1 を超え且つ約 1 . 5 未満、約 - 2 を超え且つ約 1 . 5 未満、約 - 1 を超え且つ約 1 . 5 未満、約 - 0 . 9 を超え且つ約 1 . 5 未満、約 - 0 . 8 を超え且つ約 1 . 5 未満、約 - 0 . 7 を超え且つ約 1 . 5 未満、約 - 0 . 6 を超え且つ約 1 . 5 未満、約 - 0 . 5 を超え且つ約 1 . 5 未満、約 - 2 を超え且つ約 1 . 4 未満より、約 - 2 を超え且つ約 1 . 3 未満より、約 - 2 を超え且つ約 1 . 2 未満、約 - 2 を超え且つ約 1 . 1 未満、約 - 2 を超え且つ約 1 未満、約 - 2 を超え且つ約 0 . 9 未満、約 - 2 を超え且つ約 0 . 8 未満、約 - 2 を超え且つ約 0 . 7 未満、約 - 2 を超え且つ約 0 . 6 未満、約 - 2 を超え且つ約 0 . 5 未満である時、サンプルはさらなる解析用に選択されてもよい。

【0195】

別の態様では、本明細書に提供される通りの M N a s e 消化工程を有する方法で使用される安定化された生体サンプルは、安定剤で処理された生体物質を含んでもよい。場合によっては、安定化された生体サンプルは、細胞可溶化物を含んでもよい。あるいは、安定化された生体サンプルは、安定化された無傷の細胞を含んでもよい。あるいは、安定化された生体サンプルは、安定化された無傷の核を含んでもよい。場合によっては、無傷の細胞または無傷の核の溶解前に、安定化された無傷の細胞または無傷の核のサンプルを M N a s e に接触させる工程が、実行されてもよい。場合によっては、細胞および／または核は、接合部で複数のセグメントのうちの第 1 のセグメントと第 2 のセグメントとを付着させる工程の前に、溶解されてもよい。

【0196】

別の態様では、本明細書に提供される通りの M N a s e 消化工程を含む方法は、細胞をほとんど含んでいないか、または少量の核酸を含む、小サンプル上で実行されてもよい。例えば、場合によっては、安定化された生体サンプルは、3 , 0 0 0 , 0 0 0 未満の細胞を含んでもよい。場合によっては、安定化された生体サンプルは、2 , 0 0 0 , 0 0 0 未満の細胞を含んでもよい。場合によっては、安定化された生体サンプルは、1 , 0 0 0 , 0 0 0 未満の細胞を含んでもよい。場合によっては、安定化された生体サンプルは、5 0 0 , 0 0 0 未満の細胞を含んでもよい。場合によっては、安定化された生体サンプルは、4 0 0 , 0 0 0 未満の細胞を含んでもよい。場合によっては、安定化された生体サンプルは、3 0 0 , 0 0 0 未満の細胞を含んでもよい。場合によっては、安定化された生体サンプルは、2 0 0 , 0 0 0 未満の細胞を含んでもよい。場合によっては、安定化された生体サンプルは、1 0 0 , 0 0 0 未満の細胞を含んでもよい。場合によっては、安定化された生体サンプルは、1 0 μ g 未満の D N A を含んでもよい。場合によっては、安定化された生体サンプルは、9 μ g 未満の D N A を含んでもよい。場合によっては、安定化された生体サンプルは、8 μ g 未満の D N A を含んでもよい。場合によっては、安定化された生体サンプルは、7 μ g 未満の D N A を含んでもよい。場合によっては、安定化された生体サ

ンプルは、 $6\text{ }\mu\text{g}$ 未満のDNAを含んでもよい。場合によっては、安定化された生体サンプルは、 $5\text{ }\mu\text{g}$ 未満のDNAを含んでもよい。場合によっては、安定化された生体サンプルは、 $4\text{ }\mu\text{g}$ 未満のDNAを含んでもよい。場合によっては、安定化された生体サンプルは、 $3\text{ }\mu\text{g}$ 未満のDNAを含んでもよい。場合によっては、安定化された生体サンプルは、 $2\text{ }\mu\text{g}$ 未満のDNAを含んでもよい。場合によっては、安定化された生体サンプルは、 $1\text{ }\mu\text{g}$ 未満のDNA含む。場合によっては、安定化された生体サンプルは、 $0.5\text{ }\mu\text{g}$ 未満のDNA含む。

【0197】

別の態様では、本明細書におけるMNase消化工程を含む方法が、個々の細胞または単一の細胞上で実行されてもよい。例えば、本明細書における方法は、個々の区画に分配された細胞上で実行されてもよい。例示的な区画の例としては、ウェル、エマルジョン中の液滴、または本明細書に別記される通りの異なる配列のリンカー分子の個別のパッチを含む面位置（例えばアレイスポット、ビーズなど）を含むがこれらに限定されない。追加的な区画も企図され、それは、本明細書に開示される方法、組成物、およびシステムと一致している。

【0198】

追加的な態様では、本明細書におけるMNase消化工程を含む方法で使用される安定化された生体サンプルは、DNAの断片を作り出すために、DNaseなどの追加的なヌクレアーゼでさらに処理されてもよい。場合によっては、DNaseは、配列特異的でなくてもよい。場合によっては、DNaseは、一本鎖DNAと二本鎖DNAの両方に対して活性であってもよい。場合によっては、DNaseは、二本鎖DNAに対して特異的であってもよい。場合によっては、DNaseは、優先的に二本鎖DNAを切断してもよい。場合によっては、DNaseは、一本鎖DNAに対して特異的であってもよい。場合によっては、DNaseは、優先的に一本鎖DNAを切断してもよい。場合によっては、DNaseはDNase Iであり得る。場合によっては、DNaseはDNase IIであり得る。場合によっては、DNaseは、DNase IとDNase IIの1以上から選択されてもよい。場合によっては、DNaseは、タンパク質A、タンパク質G、タンパク質A/G、またはタンパク質Lなどの、免疫グロブリン結合タンパク質またはその断片に、連結されてもよく、または融合されてもよい。他の適切なヌクレアーゼも、本開示の範囲内である。

【0199】

追加的な態様では、MNase消化工程を含む方法で使用するための、本明細書に提供される通りの安定化された生体サンプルは、架橋剤で処理され得る。場合によっては、架橋剤は、化学固定液であってもよい。場合によっては、化学固定液は、ホルムアルデヒドを含み、約2.3〜2.7オングストローム（Å）のスペーサーアーム長を有する。場合によっては、化学固定液は、長いスペーサーアーム長を有する架橋剤を含み、例えば、架橋剤は、少なくとも約3Å、4Å、5Å、6Å、7Å、8Å、9Å、10Å、11Å、12Å、13Å、14Å、15Å、16Å、17Å、18Å、19Å、または20Åのスペーサー長を有することができる。化学固定液は、エチレングリコールビス（コハク酸スクシンイミジル）（EGS）を含むことができ、これは、約16.1Åの長さを有するスペーサーアームを有する。化学固定液は、グルタル酸ジスクシンイミジル（DSG）を含むことができ、これは、約7.7Åの長さを有するスペーサーアームを有する。場合によっては、化学固定液は、ホルムアルデヒドとEGS、ホルムアルデヒドとDSG、または、ホルムアルデヒド、EGS、および7DSGを含む。複数の化学固定液が使用されるいくつかの場合には、各化学固定液は、連続的に使用される；他の場合では、複数の化学固定液のいくつかまたは全ては、サンプルに同時に適用される。長いスペーサーアームを有する架橋剤の使用は、大きな（例えば $>1\text{ kb}$ の）リードペアの分離距離を有するリードペアの画分を増加させることができる。例えば、図7は、ホルムアルデヒドのみで架橋された（消化されたDNaseとMNaseの両方）結果のライブラリと、ホルムアルデヒドとDSGまたはEGSとで架橋された結果のライブラリとの、比較を示す。DSGは、

10

20

30

40

50

両端にNHSEステル反応基を有し、そしてアミノ基（例えば、一級アミン）に対して反応性を有することができる。DSGは、膜透過性を有し、細胞内架橋を可能にする。DSGは、いくつかの用途において、スベリン酸ジサクシンイミジル（DSS）と比較して、架橋効率を増加させることができる。EGSは、両端にNHSEステル反応基を有し、そしてアミノ基（例えば、一級アミン）に対して反応性を有することができる。EGSは、膜透過性を有し、細胞内架橋を可能にする。例えばpH8.5で3～6時間ヒドロキシルアミンで処理することによって、EGS架橋を逆転させることができる；一例では、乳酸脱水素酵素は、EGSとの可逆的な架橋後に、その活性の60%を保持した。場合によっては、化学固定液は、ソラレンを含んでもよい。場合によっては、架橋剤は、紫外線であってもよい。場合によっては、安定化された生体サンプルは、架橋結合されたパラフィン包埋組織サンプルであってもよい。

10

【0200】

さらなる態様では、本明細書に提供されるMNase消化工程を含む方法は、複数の選択されたセグメントを抗体に接触させる工程を含んでもよい。場合によっては、オリゴヌクレオチド・アダプタにテザーされた免疫グロブリン結合タンパク質またはその断片は、複数の選択されたセグメントに結合された抗体を標的にしてもよい。

【0201】

追加的な態様では、本明細書に提供されるMNase消化工程を含む方法は、接合部で複数のセグメントのうちの第1のセグメントと第2のセグメントとを付着させる工程を含んでもよい。場合によっては、付着させる工程は、ビオチンタグ付きヌクレオチドを使用して粘着末端を埋めること、および平滑末端をライゲーションすることを含んでもよい。場合によっては、付着させる工程は、少なくとも第1のセグメントと第2のセグメントとを、架橋オリゴヌクレオチドに接触させることを含んでもよい。場合によっては、付着させる工程は、少なくとも第1のセグメントと第2のセグメントとを、バーコードに接触させることを含んでもよい。いくつかの実施形態では、本明細書における架橋オリゴヌクレオチドは、少なくとも約5ヌクレオチドから約50ヌクレオチドの長さであってもよい。いくつかの実施形態では、本明細書における架橋オリゴヌクレオチドは、約15～約18ヌクレオチドの長さであってもよい。いくつかの実施形態では、架橋オリゴヌクレオチドは、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35、約40、約45、または約50ヌクレオチドの長さであってもよい。いくつかの実施形態では、本明細書における架橋オリゴヌクレオチドは、バーコードを含んでもよい。

20

30

【0202】

本明細書におけるMNase消化工程を含む方法のさらなる態様では、方法は、第1のリードペアを生成するために、接合部の各側面で、少なくとも何らかの配列を得る工程を含み得る。例えば、方法は、第1のリードペアを生成するために、接合部の各側面で、少なくとも約50bp、少なくとも約100bp、少なくとも約150bp、少なくとも約200bp、少なくとも約250bp、または少なくとも約300bpの配列を得る工程を含んでもよい。

【0203】

本明細書におけるMNase消化工程を含む方法の追加的な態様では、方法は、第1のリードペアをコンティグのセットにマッピングする工程と、順序および／または配向を表すコンティグのセットを通してゲノムに至る経路を判定する工程とを含むことができる。

40

【0204】

本明細書におけるMNase消化工程を含む方法のさらなる態様では、方法は、第1のリードペアをコンティグのセットにマッピングする工程と、コンティグのセットから、安定化された生体サンプルにおける構造変異体の存在またはヘテロ接合性の喪失を判定する工程と、を含み得る。

【0205】

本明細書におけるMNase消化工程を含む方法の追加的な態様では、方法は、第1の

50

リードペアをコンティグのセットにマッピングする工程と、コンティグのセット中の変異体をフェーズに割り当てる工程とを含み得る。

【0206】

本明細書におけるMNase消化工程を含む方法のさらなる態様では、方法は、第1のリードペアをコンティグのセットにマッピングする工程と、コンティグのセットから、コンティグのセット中の変異体の存在を判定する工程と、(1)安定化された生体サンプルに対して、疾患段階、予後、または処置の方針を同定すること、(2)変異体の存在に基づいて、薬物を選択すること、または、(3)安定化された生体サンプルに対する、薬物有効性を同定すること、の1以上から選択されるステップを実施する工程と、を含むことができる。

【0207】

HiChIP、HiChIRP、およびメチルHiCのための改善された方法

HiChIPは、HiCの方法をクロマチン免疫沈降の方法と組み合わせるアプローチであり、1以上の関心タンパク質を含む相互作用の標的解析を可能にする。近接ライゲーションされた核酸が調製され得、そして、標的化された領域がさらなる解析のために免疫沈殿され得る。HiChIRP、つまり関連するアプローチは、RNA精製(ChIRP)濃縮によるクロマチン単離を、HiC方法と組み合わせて使用し、長い非コーディングRNA(lncRNA)のスクフォールド機能などのRNAの照会を可能にする。メチル-HiCは、メチル化解析をHiC方法と組み合わせて、染色体立体配座とDNAメチローム情報の同時キャプチャを可能にする。メチル-HiCは、核中で空間的に近接にある遠位ゲノムセグメント間の協調的なDNAメチル化状態を明らかにし、混合集団におけるクロマチンアーキテクチャとDNAメチロームの両方の異種性を描写し、そして複雑な組織中の細胞型に特異的なクロマチン組織とエピゲノムの同時特性評価を可能にする。これらの方法および他の方法は、サイズ選択工程、表面結合工程(例えば、SPRIBーズなどのビーズへの結合)、近接ライゲーションを実施するための架橋オリゴヌクレオチドの使用、近接ライゲーションを実施するための組み換えの使用などを含むがこれらに限定されない、本開示の技術の使用によって改善され得る。

【0208】

追加的な態様では、HiChIP、HiChIRP、およびメチルHiCのための改善された方法が本明細書に提供され、当該方法は、例えば、核酸結合タンパク質に結合した核酸の免疫沈降によって、またはメチル化された核酸の免疫沈降によって、少なくとも1つの核酸結合タンパク質に複合体化された核酸分子を含む安定化された生体サンプルを得る工程と、核酸分子を複数のセグメントに切断するために、安定化された生体サンプルをDNaseに接触させる工程と、接合部で複数のセグメントのうちの第1のセグメントと第2のセグメントとを付着させる工程と、複数の選択されたセグメントを得るために、複数のセグメントをサイズ選択にかける工程と、を含むことができる。あるいは、または組み合わせで、本明細書における方法は、例えば、核酸結合タンパク質に結合した核酸の免疫沈降によって、またはメチル化された核酸の免疫沈降によって、少なくとも1つの核酸結合タンパク質に複合体化された核酸分子を含む安定化された生体サンプルを得る工程と、核酸分子を複数のセグメントに切断するために、安定化された生体サンプルを小球菌ヌクレアーゼ(MNase)に接触させる工程と、接合部で複数のセグメントのうちの第1のセグメントと第2のセグメントとを付着させる工程と、を含むことができる。

【0209】

本明細書におけるHiChIP、HiChIRP、およびメチルHiCのための改善された方法のいくつかの態様では、安定化された生体サンプルは、無傷の細胞および/または無傷の核を含むことができる。場合によっては、安定化された生体サンプルは、安定化された無傷の細胞を含むことができる。あるいは、または組み合わせで、安定化された生体サンプルは、安定化された無傷の核を含むことができる。場合によっては、無傷の細胞または無傷の核の溶解前に、安定化された無傷の細胞または無傷の核のサンプルをDNaseに接触させる工程が、実行されてもよい。場合によっては、細胞および/または核は

、接合部で複数のセグメントのうちの第1のセグメントと第2のセグメントとを付着させる工程の前に、溶解されてもよい。

【0210】

別の態様では、本明細書におけるHiChIP、HiChIRP、およびメチルHiCのための改善された方法は、複数の選択されたセグメントを得るために、複数のセグメントをサイズ選択にかける工程を含むことができる。場合によっては、複数の選択されたセグメントは、約145～約600bpであってもよい。場合によっては、複数の選択されたセグメントは、約100～約2500bpであってもよい。場合によっては、複数の選択されたセグメントは、約100～約600bpであってもよい。場合によっては、複数の選択されたセグメントは、約600～約2500bpであってもよい。場合によっては、複数の選択されたセグメントは、約100bp～約600bp、約100bp～約700bp、約100bp～約800bp、約100bp～約900bp、約100bp～約1000bp、約100～約1100bp、約100bp～約1200bp、約100bp～約1300bp、約100bp～約1400bp、約100bp～約1500bp、約100bp～約1600bp、約100bp～約1700bp、約100bp～約1800bp、約100bp～約1900bp、約100bp～約2000bp、約100bp～約2100bp、約100bp～約2200bp、約100bp～約2300bp、約100bp～約2400bp、または約100bp～約2500bpであってもよい。

【0211】

本明細書におけるHiChIP、HiChIRP、およびメチルHiCのための改善された方法を含む方法の別の態様では、方法は、サイズ選択工程前に、複数のセグメントから配列決定ライブラリを調製する工程を、さらに含んでもよい。いくつかの実施形態では、方法は、サイズ選択されたライブラリを得るために、配列決定ライブラリをサイズ選択にかける工程を、さらに含んでもよい。場合によっては、サイズ選択されたライブラリは、約350bp～約1000bpのサイズであってもよい。場合によっては、サイズ選択されたライブラリは、約100bp～約2500bpのサイズであってもよく、例えば、約100bp～約350bp、約350bp～約500bp、約500bp～約1000bp、約1000～約1500bp、約2000～約2500bp、約350bp～約1000bp、約350bp～約1500bp、約350bp～約2000bp、約350bp～約2500bp、約500bp～約1500bp、約500bp～約2000bp、約500bp～約3500bp、約1000bp～約1500bp、約1000bp～約2000bp、約1000bp～約2500bp、約1500bp～約2000bp、約1500bp～約2500bp、または約2000bp～約2500bpであってもよい。

【0212】

本明細書におけるHiChIP、HiChIRP、およびメチルHiCのための改善された方法を含む方法で利用されるサイズ選択は、ゲル電気泳動、キャピラリー電気泳動、サイズ選択ビーズ、ゲル濾過カラム、その組み合わせ、または任意の他の適切な方法で実行され得る。

【0213】

別の態様では、HiChIP、HiChIRP、およびメチルHiCのための改善された方法を含む方法は、QC値を得るために、複数の選択されたセグメントをさらに解析する工程を、含んでもよい。場合によっては、QC値は、クロマチン消化効率(CDE)とクロマチン消化指標(CDI)とから選択されてもよい。CDEは、所望の長さを有するセグメントの割合として計算され得る。例えば、場合によっては、CDEは、サイズ選択前の、100～2500bpのサイズのセグメントの割合として計算され得る。場合によっては、CDE値が少なくとも65%である時、サンプルはさらなる解析用に選択されてもよい。場合によっては、CDE値が少なくとも約50%、少なくとも約55%、少なくとも約60%、少なくとも約65%、少なくとも約70%、少なくとも約75%、少なくとも約80%、少なくとも約85%、少なくとも約90%、または少なくとも約95%で

ある時に、サンプルはさらなる解析用に選択されてもよい。

【0214】

C D I は、サイズ選択前の、モノヌクレオソーム・サイズのセグメントの数対ジヌクレオソーム・サイズのセグメントの数の割合として、計算され得る。例えば、C D I は、600-2500bpのサイズを有する断片対100-600bpのサイズを有する断片の割合の対数として、計算されてもよい。場合によっては、C D I 値が-1.5を超え且つ1未満である時、サンプルはさらなる解析用に選択されてもよい。場合によっては、C D I 値が約-2を超え且つ約1.5未満、約-1.9を超え且つ約1.5未満、約-1.8を超え且つ約1.5未満、約-1.7を超え且つ約1.5未満、約-1.6を超え且つ約1.5未満、約-1.5を超え且つ約1.5未満、約-1.4を超え且つ約1.5未満、約-1.3を超え且つ約1.5未満、約-1.2を超え且つ約1.5未満、約-1.1を超え且つ約1.5未満、約-2を超え且つ約1.5未満、約-1を超え且つ約1.5未満、約-0.9を超え且つ約1.5未満、約-0.8を超え且つ約1.5未満、約-0.7を超え且つ約1.5未満、約-0.6を超え且つ約1.5未満、約-0.5を超え且つ約1.5未満、約-2を超え且つ約1.4未満より、約-2を超え且つ約1.3未満より、約-2を超え且つ約1.2未満、約-2を超え且つ約1.1未満、約-2を超え且つ約1未満、約-2を超え且つ約0.9未満、約-2を超え且つ約0.8未満、約-2を超え且つ約0.7未満、約-2を超え且つ約0.6未満、約-2を超え且つ約0.5未満である時、サンプルはさらなる解析用に選択されてもよい。

【0215】

別の態様では、本明細書におけるH i C h I P、H i C h I R P、およびメチルH i Cのための改善された方法は、細胞をほとんど含んでいないか、または少量の核酸を含む、小サンプル上で実行され得る。場合によっては、安定化された生体サンプルは、3,000,000未満の細胞を含んでもよい。場合によっては、安定化された生体サンプルは、2,000,000未満の細胞を含んでもよい。場合によっては、安定化された生体サンプルは、1,000,000未満の細胞を含んでもよい。場合によっては、安定化された生体サンプルは、500,000未満の細胞を含んでもよい。場合によっては、安定化された生体サンプルは、400,000未満の細胞を含んでもよい。場合によっては、安定化された生体サンプルは、300,000未満の細胞を含んでもよい。場合によっては、安定化された生体サンプルは、200,000未満の細胞を含んでもよい。場合によっては、安定化された生体サンプルは、100,000未満の細胞を含んでもよい。場合によっては、安定化された生体サンプルは、10μg未満のDNAを含んでもよい。場合によっては、安定化された生体サンプルは、9μg未満のDNAを含んでもよい。場合によっては、安定化された生体サンプルは、8μg未満のDNAを含んでもよい。場合によっては、安定化された生体サンプルは、7μg未満のDNAを含んでもよい。場合によっては、安定化された生体サンプルは、6μg未満のDNAを含んでもよい。場合によっては、安定化された生体サンプルは、5μg未満のDNAを含んでもよい。場合によっては、安定化された生体サンプルは、4μg未満のDNAを含んでもよい。場合によっては、安定化された生体サンプルは、3μg未満のDNAを含んでもよい。場合によっては、安定化された生体サンプルは、2μg未満のDNAを含んでもよい。場合によっては、安定化された生体サンプルは、1μg未満のDNAを含んでもよい。場合によっては、安定化された生体サンプルは、0.5μg未満のDNAを含んでもよい。

【0216】

別の態様では、本明細書におけるH i C h I P、H i C h I R P、およびメチルH i Cのための改善された方法を含む方法は、個々の細胞または単一の細胞上で実行されてもよい。例えば、本明細書における方法は、個々の区画に分配された細胞上で実行されてもよい。例示的な区画の例としては、ウェル、エマルジョン中の液滴、または本明細書に別記される通りの異なる配列のリンカー分子の個別のパッチを含む面位置（例えばアレイスポット、ビーズなど）を含むがこれらに限定されない。追加的な区画も企図され、それは、本明細書に開示される方法、組成物、およびシステムと一致している。

【0217】

追加的な態様では、本明細書におけるH i C h I P、H i C h I R P、およびメチルH i Cのための改善された方法を含む方法で使用される安定化された生体サンプルは、DNAの断片を作り出すために、DNaseなどのヌクレアーゼで処理され得る。場合によっては、DNaseは、配列特異的でなくてもよい。場合によっては、DNaseは、一本鎖DNAと二本鎖DNAの両方に対して活性であってもよい。場合によっては、DNaseは、二本鎖DNAに対して特異的であってもよい。場合によっては、DNaseは、優先的に二本鎖DNAを切断してもよい。場合によっては、DNaseは、一本鎖DNAに対して特異的であってもよい。場合によっては、DNaseは、優先的に一本鎖DNAを切断してもよい。場合によっては、DNaseは、DNase Iであってもよい。場合によっては、DNaseは、DNase IIであってもよい。場合によっては、DNaseは、DNase IとDNase IIの1以上から選択されてもよい。場合によっては、DNaseは、小球菌ヌクレアーゼであってもよい。場合によっては、DNaseは、DNase I、DNase II、および小球菌ヌクレアーゼの1以上から選択されてもよい。場合によっては、DNaseは、タンパク質A、タンパク質G、タンパク質A/G、またはタンパク質Lなどの、免疫グロブリン結合タンパク質またはその断片に、連結されてもよく、または融合されてもよい。他の適切なヌクレアーゼも、本開示の範囲内である。

10

【0218】

追加的な態様では、本明細書におけるH i C h I P、H i C h I R P、およびメチルH i Cのための改善された方法を含む方法で使用される安定化された生体サンプルは、架橋剤で処理されてもよい。場合によっては、架橋剤は、化学固定液であってもよい。場合によっては、化学固定液は、ホルムアルデヒドを含み、約2.3〜2.7オングストローム(A)のスペーサーアーム長を有する。場合によっては、化学固定液は、長いスペーサーアーム長を有する架橋剤を含み、例えば、架橋剤は、少なくとも約3A、4A、5A、6A、7A、8A、9A、10A、11A、12A、13A、14A、15A、16A、17A、18A、19A、または20Aのスペーサー長を有することができる。化学固定液は、エチレングリコールビス(コハク酸スクシンイミジル)(EGS)を含むことができ、これは、約16.1Aの長さを有するスペーサーアームを有する。化学固定液は、グルタル酸ジスクシンイミジル(DSG)を含むことができ、これは、約7.7Aの長さを有するスペーサーアームを有する。場合によっては、化学固定液は、ホルムアルデヒドとEGS、ホルムアルデヒドとDSG、またはホルムアルデヒド、EGS、およびDSGを含む。複数の化学固定液が使用されるいくつかの場合には、各化学固定液は、連続的に使用される；他の場合では、複数の化学固定液のいくつかまたは全ては、サンプルに同時に適用される。長いスペーサーアームを有する架橋剤の使用は、大きな(例えば>1kbの)リードペアの分離距離を有するリードペアの画分を増加させることができる。例えば、図7は、ホルムアルデヒドのみで架橋された(消化されたDNaseとMNaseの両方)結果のライブラリと、ホルムアルデヒドとDSGまたはEGSとで架橋された結果のライブラリとの、比較を示す。DSGは、両端にNHSエステル反応基を有し、そしてアミノ基(例えば、一級アミン)に対して反応性を有することができる。DSGは、膜透過性を有し、細胞内架橋を可能にする。DSGは、いくつかの用途において、スベリン酸ジスクシンイミジル(DSS)と比較して、架橋効率を増加させることができる。EGSは、両端にNHSエステル反応基を有し、そしてアミノ基(例えば、一級アミン)に対して反応性を有することができる。EGSは、膜透過性を有し、細胞内架橋を可能にする。例えばpH8.5で3〜6時間ヒドロキシルアミンで処理することによって、EGS架橋を逆転させることができる；一例では、乳酸脱水素酵素は、EGSとの可逆的な架橋後に、その活性の60%を保持した。場合によっては、化学固定液は、ソラレンを含んでもよい。場合によっては、架橋剤は、紫外線であってもよい。場合によっては、安定化された生体サンプルは、架橋結合されたパラフィン包埋組織サンプルであってもよい。

20

30

40

【0219】

50

追加的な態様では、本明細書におけるH i C h I P、H i C h I R P、およびメチルH i Cのための改善された方法を含む方法は、接合部で複数のセグメントのうちの第1のセグメントと第2のセグメントとを付着させる工程を含んでもよい。場合によっては、付着させる工程は、ビオチンタグ付きヌクレオチドを使用して粘着末端を埋めること、および平滑末端をライゲーションすることを含み得る。場合によっては、付着させる工程は、少なくとも第1のセグメントと第2のセグメントとを、架橋オリゴヌクレオチドに接触させることを含み得る。場合によっては、付着させる工程は、少なくとも第1のセグメントと第2のセグメントとを、バーコードに接触させることを含み得る。いくつかの実施形態では、本明細書における架橋オリゴヌクレオチドは、少なくとも約5ヌクレオチドから約50ヌクレオチドの長さであってもよい。いくつかの実施形態では、本明細書における架橋オリゴヌクレオチドは、約15～約18ヌクレオチドの長さであってもよい。いくつかの実施形態では、架橋オリゴヌクレオチドは、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35、約40、約45、または約50ヌクレオチドの長さであってもよい。いくつかの実施形態では、本明細書における架橋オリゴヌクレオチドは、バーコードを含んでもよい。

【0220】

追加的な態様では、本明細書におけるH i C h I P、H i C h I R P、およびメチルH i Cのための改善された方法を含む方法は、剪断する工程を含まなくてもよい。

【0221】

本明細書におけるH i C h I P、H i C h I R P、およびメチルH i Cのための改善された方法を含む方法のさらなる態様では、方法は、第1のリードペアを生成するために、接合部の各側面で、少なくとも何らかの配列を得る工程を含んでもよい。例えば、方法は、第1のリードペアを生成するために、接合部の各側面で、少なくとも約50bp、少なくとも約100bp、少なくとも約150bp、少なくとも約200bp、少なくとも約250bp、または少なくとも約300bpの配列を得る工程を含んでもよい。

【0222】

本明細書におけるH i C h I P、H i C h I R P、およびメチルH i Cのための改善された方法を含む方法の追加的な態様では、方法は、第1のリードペアをコンティグのセットにマッピングする工程と、順序および／または配向を表すコンティグのセットを通してゲノムに至る経路を判定する工程とを含んでもよい。

【0223】

本明細書におけるH i C h I P、H i C h I R P、およびメチルH i Cのための改善された方法を含む方法のさらなる態様では、方法は、第1のリードペアをコンティグのセットにマッピングする工程と、コンティグのセットから、安定化された生体サンプルにおける構造変異体の存在またはヘテロ接合性の喪失を判定する工程と、を含んでもよい。

【0224】

本明細書における細胞全体または核全体の消化を含む方法の追加的な態様では、方法は、第1のリードペアをコンティグのセットにマッピングする工程と、コンティグのセット中の変異体をフェーズに割り当てる工程と、を含んでもよい。

【0225】

本明細書におけるH i C h I P、H i C h I R P、およびメチルH i Cのための改善された方法を含む方法のさらなる態様では、方法は、第1のリードペアをコンティグのセットにマッピングする工程と、コンティグのセットから、コンティグのセット中の変異体の存在を判定する工程と、(1) 安定化された生体サンプルに対して、疾患段階、予後、または処置の方針を同定すること、(2) 変異体の存在に基づいて、薬物を選択すること、または、(3) 安定化された生体サンプルに対する、薬物有効性を同定すること、の1以上から選択されるステップを実施する工程と、を含んでもよい。

【0226】

長距離リードペアを生成すること

本開示は、超長距離リードペアを生成し、そのデータを前述のすべての追求の進歩のために利用するための方法を提供する。いくつかの実施形態では、本開示は、わずか〜3億のリードペアで、非常に近接しており、かつ非常に正確なヒトゲノムアセンブリを生成する方法を提供する。他の実施形態では、本開示は、99%以上の精度で、ヒトゲノム中のヘテロ接合性変異体の90%以上をフェージングする方法を提供する。さらに、本開示によって生成されたリードペアの範囲は、はるかに大きなゲノム距離におよぶように拡張され得る。アセンブリは、超長距離のリードペアライブラリに加えて、標準ショットガンライブラリから生成される。さらに他の実施形態では、本開示は、配列決定データのこれらのセットの両方を利用することができるソフトウェアを提供する。フェージングされた変異体は、単一の長距離リードペアライブラリで生成され、そこからのリードはリファレンスゲノム (reference genome) にマッピングされ、そして変異体を個体の2つの親染色体の1つに割り当てるために使用される。最終的に、本開示は、例外的な長いリードを生成するように、既知の技術を使用してさらに大きなDNA断片の抽出を提供する。

【0227】

これらの反復がアセンブリおよび整列の処理を妨害するメカニズムはかなり単純であり、そして、結局は曖昧さの結果である。大きな反復領域の場合には、困難はスパンの1つであり得る。リードまたはリードペアが反復領域におよぶほど十分に長くない場合、反復要素を境界する領域を信頼性を有しながら接続することができないかもしれない。より小さな反復要素の場合には、問題は主として配置であり得る。領域が、ゲノムに共通する2つの反復要素によって隣接される場合、その正確な配置を判定することは不可能ではないにしても困難になり、その理由は、隣接する要素がそれらのクラスの他のすべてと類似しているからである。どちらの場合でも、識別、従って特定の反復の配置を困難にするのは、反復中の識別情報の欠如である。必要なのは、反復領域によって縁取られた、または分離された、固有のセグメント間の接続を、実験的に確立する能力である。

【0228】

本開示の方法は、これらの反復領域によってもたらされる実質的な障壁を克服することによって、ゲノムの分野を進歩させ、そして、それによって、ゲノム解析の多くのドメインにおける重要な進歩を可能にし得る。以前の技術でデノボアセンブリを実行するためには、より連続的なアセンブリを生成するために、多くの小さなスキフォールドに断片化されたアセンブリに落ち着くか、または、大規模挿入ライブラリを作成すること、あるいは他のアプローチを使用することによりかなりの時間とリソースを費やすか、のいずれかが必要である。そのようなアプローチは、非常に深い配列決定カバレッジを取得すること、BACまたは fosmid (fosmid) ライブラリを構築すること、光学マッピング、またはこれらの技術および／または他の技術のいくつかの組み合わせを含んでもよい。膨大なリソースと時間の要件により、そのようなアプローチは、ほとんどの小規模なラボでは手の届かないものになり、非モデル生物の研究が妨げられる。本明細書に記載される方法が非常に長距離のリードペアを生成し得るため、デノボアセンブリは、単一の配列決定の実行で達成され得る。これによって、アセンブリコストが桁違いに削減され、そして、必要な時間が数か月または数年から、数週間に短縮されるだろう。場合によっては、本明細書に開示される方法は、14日未満、13日未満、12日未満、11日未満、10日未満、9日未満、8日未満、7日未満、6日未満、5日未満、4日未満、または先に規定された期間の任意の2つの間の範囲で、複数のリードペアを生成することを可能にする。例えば、方法は、約10日〜14日で、複数のリードペアを生成することを可能にし得る。最もニッチな生物に対してでさえゲノムを構築することはルーチン的となり、系統発生解析は比較物の欠如に悩まされることはなくなり、Genome 10kなどのプロジェクトを実現することができるだろう。

【0229】

同様に、医療目的のための構造的解析とフェージング解析はさらに挑戦し続ける。癌同士で、同じ型の癌を持つ個体同士で、あるいは同じ腫瘍内でさえ、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、サンプル当たりの低コストで非常に高い精度とスループットとを必要とする。個別化医療（personalized medicine）の領域において、ゲノムケアの絶対的基準の1つは、大きな構造的再編成と小さな構造的再編成、および新しい突然変異を含む、徹底的に特徴評価された、かつフェージングされた、すべての変異体を有する、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、デノボアセンブリに必要なものと同様の労力が必要であり、これは、現在、ルーチン的な医療処置としては費用と労力がかかりすぎている。開示される方法は、完全で正確なゲノムを低コストで迅速に生成することができ、それによって、ヒト疾患の研究および処置において、多くの非常に求められている能力を生み出し得る。

【0230】

本明細書に開示される方法をフェージングに適用することで、統計的アプローチの利便性をファミリー解析の精度と組み合わせることができ、どちらかの方法を単独で使用するよりも、一銭、労力、サンプルの一節約を提供する。デノボ変異体フェージングは、以前の技術では禁止されている非常に望ましいフェージング解析であり、本明細書に開示される方法を使用して容易に実行され得る。ヒト変異の大部分はまれである（マイナー対立遺伝子頻度5%未満）ため、このことは特に重要である。フェージング情報は、リンクされていない遺伝子型と比較して、高度に接続されたハプロタイプ（単一の染色体に割り当てられた変異体のコレクション）のネットワークから著しい利点を得る集団遺伝研究にとって、価値あるものである。ハプロタイプ情報は、集団サイズ、移動、およびサブ集団間の交換における歴史的変化に関するより高分解能の研究を可能にし得、そして、特定の変異体を特定の親および祖父母に遡ることを可能にする。これは次に、疾患に会合された変異体の遺伝的伝達と、単一の個体で共にまとめられた場合の変異体間の相互作用とを明らかにする。本開示の方法は、最終的に、超長距離リードペア（XL RP）ライブラリの調製、配列決定、および解析を可能にし得る。

【0231】

本開示のいくつかの実施形態では、被験体からの組織またはDNAサンプルが提供され得、そして、方法は、アセンブルされたゲノム、コールされた（called）変異体（大きな構造変異体を含む）との整列、フェージングされた変異体コール、または任意の追加的な解析を返し得る。他の実施形態では、本明細書に開示される方法は、個体に直接、XL RPライブラリを提供し得る。

【0232】

超長距離リードペア

本開示の様々な実施形態では、本明細書に開示される方法は、長距離で分離された超長距離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って改善され得る。場合によっては、リードペアは、最大で、50、60、70、80、90、100、125、150、175、200、225、250、300、400、500、600、700、800、900、1000、1500、2000、2500、3000、4000、5000 kbp、またはそれ以上のゲノム距離におよぶ。いくつかの例では、リードペアは、最大500 kbpのゲノム距離におよび得る。他の例では、リードペアは、最大2000 kbpのゲノム距離におよび得る。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かつ構築することができ、さらに効率、特異性、およびゲノムカバレッジの増大に十分に適している。場合によっては、リードペアは、約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、60、または90日未満で生成され得る。いくつかの例では、リードペアは、約14日未満で生成され得る。さらなる例では、リードペアは、約10日未満で生成され得る。場合によっては、本開示の方法は、約5%、約10%、約15%、約20%、約30%、約40%、約50%、約60%、約70%、約80%、約90%、約95%、約99%、または約100%を上回るリードペアを提供することができ、そ

れは、複数のコンティグを正しく順序付けるおよび／または配向する際に、少なくとも約 50%、約 60%、約 70%、約 80%、約 90%、約 95%、約 99%、または約 100%の精度を有する。例えば、方法は複数のコンティグを正しく順序付けるおよび／または配向する際に、約 90～100%の精度を提供することができる。

【0233】

他の実施形態では、本明細書に開示される方法は、現在利用されてい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、十分に試験されおよび／また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。さらなる実施形態では、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。

【0234】

本開示の方法は、広範囲の有機体のためのデノボゲノムアセンブリを劇的に単純化する。以前の技術を使用すると、そのようなアセンブリは、経済的なメイトーペアライブラリの短い挿入によって、現在制限されている。フォスミドでアクセス可能な最大 40–50 kbp のゲノム距離でリードペアを生成することは可能かもしれないが、一方、これらは高価で扱いにくく、セントロメア内のものを含む最長の反復ストレッチにおよぶには短すぎて、一ヒトにおいては、300 kbp～5 Mbp のサイズの範囲になり得る。本明細書に開示される方法は、大きな距離（例えば、メガベース、またはより長い）におよぶことができるリードペアを提供することができ、そしてそのため、これらのスキャフォールドの完全性についての難点を克服することができる。従って、染色体レベルのアセンブリを生成することは、本開示の方法を利用することによって、ルーチンとなり得る。アセンブリのためのより骨の折れる手段、一現在は、研究室にとって信じられないほどの時間とお金がかかり、膨大なゲノムカタログを禁止している）が、不要になるかもしれず、リソースをより意義深い解析のために解放することができるかもしれない。同様に、長距離フェージング情報の獲得は、集団ゲノム化学、系統発生学、および疾患に関する研究に、途方もない追加的パワーを提供することができる。本明細書に開示される方法は、多数の個々の正確なフェージングを可能にし、従って、集団および深層時間レベルでゲノムを調査するための我々の能力の幅と深さを拡張する。

【0235】

個別化医療 (personalized medicine) の分野では、本明細書に開示される方法から生成された XLRP リードペアは、正確で、低コストで、フェージング化されており、かつ迅速に生成される個人化ゲノムに向けた有意義な進歩を見せている。現在の方法は、長距離で変異体をフェージングする能力が不十分であり、それによって、複合ヘテロ接合性遺伝子型の表現型の影響の特徴評価の妨げとなっている。追加的に、ゲノム疾患にとっての実質的な関心構造変異体は、それらを研究するために使用されるリードおよびリードペア挿入と比較してサイズが大きいため、現在の技術で正確に同定および特徴評価することが困難である。数十キロベースからメガベース以上におよぶリードペアは、この困難を軽減するのに役立ち得、それによって、構造変位の高度に並列化された個別化解析が可能になる。

【0236】

基礎的な進化的および生物医学的研究は、ハイスループット配列決定の技術的進歩によって推進されている。以前は、ゲノム配列決定およびアセンブリの全体が、大規模なゲノム配列決定のセンターの起源であったが、現在は、市販されているシーケンサーが十分に安価であり、ほとんどの研究大学がこれらの機械の 1 以上を有しているほどである。現在、大量の DNA 配列データを生成することは比較的安価である。しかし、高品質であり高度に連続するゲノム配列を現在の技術で生成することは、理論上、および現実面で困難なままである。さらに、人間を含め、解析したいと思うほとんどの生物は二倍体であるため、各個体はゲノムの 2 つのハプロイドコピーを有している。ヘテロ接合性の部位（例えば、母親によって与えられた対立遺伝子が父親によって与えられた対立遺伝子と異なる場合）では、どのセットの対立遺伝子がどちらの親に由来するのかわかること（ハプロタイプ

・フェージングとして知られている)は困難である。この情報は、疾患や形質関連研究など、多くの進化的および生物医学的研究を実行するために使用することができる。

【0237】

様々な実施形態では、本開示は、所与のゲノム内の短期、中期、および長期の接続についてのハイスループットな発見のためのペア化末端配列決定 (paired-end sequencing) を使用するDNA調製のための技術を組み合わせるゲノムアセンブリのための方法を提供する。本開示は、さらに、これらの接続を使用して、ハプロタイプ・フェージングおよび／またはメタゲノム研究のためのゲノムアセンブリを支援するための方法を提供する。本明細書に提示される方法が被験体のゲノムのアセンブリを判定するために使用することができる一方で、本明細書に提示される方法は、染色体などの被験体のゲノムの一部のアセンブリ、または様々な長さの被験体のクロマチンのアセンブリを判定するために使用することができる。

【0238】

いくつかの実施形態では、本開示は本明細書に開示される1以上の方法を提供し、当該方法は、被験体から得られた標的DNAの断片を配列決定することから複数のコンティグを生成する工程を含む。標的DNAの長いストレッチは、1以上のヌクレアーゼ (DNase I、DNase II、小球菌ヌクレアーゼなど) でDNAをカットすることによって、断片化され得る。結果として生じる断片は、複数の配列決定リードを得るために、高スループット配列決定方法を使用して配列決定され得る。本開示の方法と共に使用され得るハイスループット配列決定方法の例としては、Roche Diagnosticsが開発した454パイロシーケンシング方法、Illuminaによって開発された「クラスター」シーケンシング方法、Life Technologiesによって開発されたSOLiDおよびイオン半導体シーケンシング方法、およびComplete Genomicsによって開発されたDNAナノボールシーケンシング方法が挙げられるが、これらに限定されない。その後、異なる配列決定リードのオーバーラップする末端をアセンブルして、コンティグを形成することができる。あるいは、断片化された標的DNAを、ベクターにクローン化することができる。その後、細胞または生物は、ライブラリを形成するためにDNAベクターでトランスフェクトされる。トランスフェクトされた細胞または生物を複製した後、ベクターを単離および配列決定して、複数の配列決定リードを生成する。その後、異なる配列決定リードのオーバーラップする末端をアセンブルして、コンティグを形成することができる。

【0239】

特にハイスループット配列決定技術を使用するゲノムアセンブリは、問題を引き起こす場合がある。しばしば、アセンブリは、数千または数万の短いコンティグからなる。これらのコンティグの順序と配向は、一般的に不明であり、ゲノムアセンブリの有用性を制限している。これらのスキファールドを順序付けし、そして配向するための技術は存在するが、それらは一般に高価であり、労働集約的であり、そしてしばしば非常に長距離の相互作用の発見に失敗する。

【0240】

コンティグを生成するために使用される標的DNAを含むサンプルは、体液 (例えば、血液、尿、血清、リンパ液、唾液、口腔内からのぬぐい液、肛門および膣分泌物、発汗および精液) を採取すること、組織を取ること、または細胞／生物を収集することを含む任意の数の手段によって、被験体から得ることができる。得られたサンプルは、単一の型の細胞／生物で構成されている場合もあれば、複数の型の細胞／生物で構成されている場合もある。DNAは、被験体のサンプルから抽出し、そして調製することができる。例えば、サンプルは、既知の溶解緩衝液、超音波処理技術、エレクトロポレーションなどを使用して、ポリヌクレオチドを含む細胞を溶解するように処理されてもよい。標的DNAは、さらに精製され、そしてアルコール抽出、セシウム勾配、および／またはカラムクロマトグラフィーを使用することによって、タンパク質などの汚染物質が取り除かれてもよい。

【0241】

開示の他の実施形態では、非常に高い分子量のDNAを抽出するための方法が提供される。場合によっては、XLRPライブラリからのデータは、入力DNAの断片サイズを大きくすることによって改善され得る。いくつかの例では、細胞からメガベースサイズのDNAの断片を抽出することで、ゲノム中にメガベースで分離されたリードペアを生成することができる。場合によっては、生成されたリードペアは、約10kB、約50kB、約100kB、約200kB、約500kB、約1Mb、約2Mb、約5Mb、約10Mb、約100Mbを超えるスパンにわたる配列情報を提供することができる。いくつかの例では、リードペアは、約500kBを超えるスパンにわたる配列情報を提供することができる。さらなる例では、リードペアは、約2Mbを超えるスパンにわたる配列情報を提供することができる。場合によっては、非常に高い分子量のDNAが、非常に穏やかな細胞溶解 (Teague, B. 等 (2010) Proc. Nat. Acad. Sci. USA 107 (24), 10848-53)、およびアガロースプラグ (Schwartz, D. C., & Cantor, C. R. (1984) Cell, 37 (1), 67-75) によって抽出され得る。他の場合には、最大メガベースの長さのDNA分子を精製することができる市販の機械を使用して、非常に高い分子量のDNAを抽出することができる。

【0242】

染色体の物理的レイアウトの調査

様々な実施形態では、本開示は、本明細書に開示される1以上の方法を提供し、当該方法は、生細胞内の染色体の物理的レイアウトを調査する工程を含む、シーケンシングを通じて染色体の物理的レイアウトを調査する技術の例としては、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」) などの「C」ファミリーの技術およびHi-Cベースの方法、ならびに、ChIP-ループ、ChIA-PET、およびHiChIPなどのChIPベースの方法が挙げられる。これらの技術は、生細胞中のクロマチンの固定を利用して、核中の空間的関係を固定する。産物の後の処理および配列決定により、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。さらなる解析では、これらの会合を使用して、染色体が生体の核に物理的に配置されている時の、染色体の3次元幾何学マップを作成することができる。そのような技術は、生細胞中の染色体の離散的な空間的編成を説明し、そして、染色体の遺伝子座中の機能的相互作用の正確な見解を提供する。これらの機能的な研究を悩ませた1つの問題は、非特異的な相互作用、会合の存在であり、これは、染色体の近接性にのみ起因するデータに存在していた。本開示では、これらの非特異的染色体内部相互作用は、アセンブリに関する貴重な情報を提供するために、本明細書に提示される方法によってキャプチャされる。

【0243】

いくつかの実施形態では、染色体内部相互作用は、染色体の接続性と相関を有する。場合によっては、染色体内部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築される。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成分—は、3C、4C、5C、およびHi-Cの配列決定を通して、クロマチンの立体配座と構造とを検出するための最も一般的な「C」ファミリーの技術を下にした固定にとっては重要である。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、再構築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。

【0244】

クロマチン立体配座キャプチャ技術が要約される。簡潔に言えば、架橋は、物理的に近接にあるゲノム領域間で作成される。タンパク質 (ヒストンなど) のクロマチン内における、DNA分子、例えばゲノムDNAとの架橋は、本明細書の他の場所でさらに詳細に記載される、また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、2以上のヌクレオチド配列が、1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つのアプローチは、紫外線照射にクロマチンを曝露することである (Gilmour等, 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、化学的または物理的 (例えば光学的) な架橋などの他のアプローチを利用して実行されてもよい。適切な化学架橋剤は、限定されないが、ホルムアルデヒドおよびソラレンを含む (Solomon等, Proc. Nat'l. Acad. Sci. USA 82:6470-6474, 1985; Solomon et al., Cell 53:937-947, 1988)。例えば、架橋結合は、DNA分子およ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って実行され得る。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)、およびシクロホスファミドを含むがこれらに限定されない。好適に、架橋剤は、約2 Åなどの比較的短距離を架橋する架橋を形成し、それによって、反転され得る密接な相互作用を選択する。

【0245】

いくつかの実施形態では、DNA分子は、架橋前または架橋後に、免疫沈降されてもよい。場合によっては、DNA分子は、断片化され得る。断片は、アセチル化ヒストン、例えばH3に対して特異的に認識および結合する抗体などの、接触パートナーと接触されてもよい。そのような抗体の例は、Upstate Biotechnology, Lake Placid, N.Y. から入手可能な、抗アセチル化ヒストンH3を含むが、これに限定されない。免疫沈降物からのポリヌクレオチドは、その後に免疫沈降物から収集され得る。クロマチンを断片化する前に、アセチル化ヒストンは、隣接するポリヌクレオチド配列と架橋することができる。その後、その混合物は、混合物中のポリヌクレオチドを分留するために処理される。本明細書における分留技術は、デオキシリボヌクレアーゼ (DNase) 酵素の使用を含む。本明細書の方法に適したDNaseは、DNase I、DNase II、および小球菌ヌクレアーゼを含むが、これらに限定されない。結果として生じる断片は、様々なサイズであり得る。結果として生じる断片はまた、5' または3' 末端で一本鎖オーバーハンド (overhand) を含んでもよい。

【0246】

いくつかの実施形態では、約145 bp~約600 bpの断片を得ることができる。あるいは、約100 bp~約2500 bpの断片、約100 bp~約600 bpの断片、約600 bp~約2500 bpの断片を得ることができる。サンプルは、架橋された連結配列セグメントの配列決定用に調製することができる。場合によっては、例えば、分子内で架橋された2つの配列セグメントをライゲーションすることによって、ポリヌクレオチドの単一の短いストレッチを作成することができる。配列情報は、本明細書にさらに詳細に別記されている任意の適切な配列決定技術、またはハイスループット配列決定方法などの他の適切な方法を使用して、サンプルから得られてもよい。例えば、ライゲーション産物は、断片の各末端から配列情報を得るペア化末端配列決定にかけることができる。配列セグメントのペアは、得られた配列情報で表すことができ、ポリヌクレオチドに沿って2つの配列セグメントを分離する直線距離にわたってハプロタイピング情報を関連付ける。

【0247】

Hi-Cによって生成されたデータの1つの特徴は、ほとんどのリードペアが、ゲノムに再マッピングされる時に、ほぼ線形に近接していることが確認されることである。すなわち、ほとんどのリードペアは、ゲノム中で互いに接近していることが確認される。結果として得られるデータセットでは、染色体が別個の領域を占有する場合に予想されるように、染色体間接触の確率は平均して染色体間接触の確率よりもはるかに高くなる。さらに、相互作用の確率は直線距離で急速に減衰するが、同じ染色体上で>200 Mbで分離されている遺伝子座でさえ、異なる染色体上の遺伝子座よりも、相互作用する可能性が高くなる。長距離の染色体間接触、特に染色体間接触を検出する際、短距離および中距離の染色体間接触のこの「バックグラウンド」は、Hi-C解析を使用して除外されるバックグ

ラウンドノイズである。

【0248】

特に、真核生物中のHi-C実験は、種特異的および細胞型特異的なクロマチン相互作用に加えて、2つのカノニカルな相互作用パターンを示す。1つのパターンである距離依存減衰（DDD）は、ゲノム距離の関数としての相互作用頻度における減衰の一般的な傾向である。第2のパターンであるシーストランス比（CTR）では、異なる染色体上の遺伝子座と比較して、数十メガベースの配列で分離されている時ですら、同じ染色体上に位置する遺伝子座間の相互作用頻度が大幅に高くなる。これらのパターンは、一般的な高分子ダイナミクスを反映してもよく、ここでは、近位遺伝子座がランダムに相互作用する可能性、ならびに、染色体領域の形成、フェーズ間染色体がほとんど混合せずに核中の別個の容積を占有する傾向を有するという現象などの、特定の核組織の特徴の可能性が、高くなる。これらの2つのパターンの正確な詳細は、種、細胞型、および細胞状態間で異なる場合があるが、それらは遍在的であり目立つ存在である。これらのパターンは非常に強力で一貫しているため、実験の質を評価するために使用され、そして通常、詳細な相互作用を明らかにするためにデータから正規化される。しかし、本明細書に開示される方法では、ゲノムアセンブリは、ゲノムの三次元構造を利用することができる。カノニカルなHi-C相互作用パターンを特定のループ相互作用の解析の妨げにってしまう特徴、つまり、それらの遍在性、強度、および一貫性は、コンティグのゲノムポジションを推定するための強力なツールとして使用され得る。

【0249】

特定の実装では、染色体内リードペア間の物理的距離の検討は、ゲノムアセンブリに関するデータのいくつかの有用な特徴を示す。第1に、短距離の相互作用は、長距離の相互作用よりも一般的である。つまり、リードペアの各リードは、遠くにある領域よりも、実際のゲノム中の近い領域とメイティングする可能性が高くなる。第2に、中距離および長距離の相互作用の長いテール部（tail）が存在する。すなわち、リードペアは、キロベース（kB）の距離で、またはメガベース（Mb）の距離ですら、染色体内配列に関する情報を伝達する。例えば、リードペアは、約10kB、約50kB、約100kB、約200kB、約500kB、約1Mb、約2Mb、約5Mb、約10Mb、または約100Mbを超えるスパンにわたる配列情報を提供することができる。データのこれらの特徴は、同じ染色体上で近接しているゲノムの領域が物理的に近接している可能性がより高いことを示しており、—このことは、それらがDNAバックボーンを通じて互いに化学的にリンクされているために、予測される結果である。Hi-Cによって生成されたものなどの、ゲノム全体のクロマチン相互作用データセットは、染色体全体に沿った配列のグループ化と線形編成に関する長距離情報を提供すると、推測されていた。

【0250】

Hi-Cのための実験方法は単純で比較的低コストであるが、ゲノムアセンブリとハプロタイピングの現在のプロトコルは、300–500万個の細胞をと必要とし、かなり大量の材料は、特に、特定のヒト患者サンプルからは取得できない可能性がある。対照的に、本明細書に開示される方法は、細胞からの材料が著しく少ない、遺伝子型アセンブリ、ハプロタイプ・フェージング、およびメタゲノミクスの、正確かつ予測的な結果を可能にする方法を含む。例えば、約0.1μg、約0.2μg、約0.3μg、約0.4μg、約0.5μg、約0.6μg、約0.7μg、約0.8μg、約0.9μg、約1.0μg、約1.2μg、約1.4μg、約1.6μg、約1.8μg、約2.0μg、約2.5μg、約3.0μg、約3.5μg、約4.0μg、約4.5μg、約5.0μg、約6.0μg、約7.0μg、約8.0μg、約9.0μg、約10μg、約15μg、約20μg、約30μg、約40μg、約50μg、約60μg、約70μg、約80μg、約90μg、約100μg、約150μg、約200μg、約300μg、約400μg、約500μg、約600μg、約700μg、約800μg、約900μg、約1000μg、約1200μg、約1400μg、約1600μg、約1800μg、約2000μg、約2200μg、約2400μg、約2600μg、約2800μg、約30

00 μg 、約3200 μg 、約3400 μg 、約3600 μg 、約3800 μg 、約4000 μg 、約4200 μg 、約4400 μg 、約4600 μg 、約4800 μg 、約5000 μg 、約5200 μg 、約5400 μg 、約5600 μg 、約5800 μg 、約6000 μg 、約6200 μg 、約6400 μg 、約6600 μg 、約6800 μg 、約7000 μg 、約7200 μg 、約7400 μg 、約7600 μg 、約7800 μg 、約8000 μg 、約8200 μg 、約8400 μg 、約8600 μg 、約8800 μg 、約9000 μg 、約9200 μg 、約9400 μg 、約9600 μg 、約9800 μg 、約10,000 μg のDNAを、本明細書で開示される方法で使用することができる。いくつかの例では、本明細書に開示される方法で使用されるDNAを、約3,000,000、約2,500,000、約2,000,000、約1,500,000、約1,000,000、約500,000、約100,000、約50,000、約10,000、約5,000、約1,000、約500、または約100未満の細胞から抽出することができる。

【0251】

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査するための手順は、培養細胞または初代組織から単離されたクロマチンなどの、細胞／生物内に形成されるクロマチンを利用する。本開示は、細胞／生物から単離されたクロマチンだけでなく、再構築されたクロマチンを用いたそのような技術の使用を提供する。再構築クロマチンは、様々な特徴にわたり細胞／生物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、頬または直腸の領域をぬぐうこと、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、クロマチンの再構築は、ゲノムアセンブリおよびハプロタイプ・フェージングのための人工物を生成する、染色体間の、およ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、またはそれを下回る、本開示の方法および組成物に従った染色体間または分子間の架橋を有してもよい。いくつかの例では、サンプルは、約5%未満の染色体間または分子間の架橋を有してもよい。いくつかの例では、サンプルは、約3%未満の染色体間または分子間の架橋を有してもよい。さらなる例では、サンプルは、約1%未満の染色体間または分子間の架橋を有してもよい。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、調節することができる。例えば、DNA対ヒストンの比率は変動する場合があります、それによって、ヌクレオソーム密度を望ましい値に調節することができる。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改変することができる。いくつかの実施形態では、架橋密度が変動するサブサンプルが、短いおよび長い範囲両方の会合をカバーするように調節され得る。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%、または約100%が、サンプルDNA分子上で、少なくとも約50 kb、約60 kb、約70 kb、約80 kb、約90 kb、約100 kb、約110 kb、約120 kb、約130 kb、約140 kb、約150 kb、約160 kb、約180 kb、約200 kb、約250 kb、約300 kb、約350 kb、約400 kb、約450 kb、約500 kb離れているDNAセグメント間に生じるように、調節され得る。

【0252】

接触マッピングおよびトポロジー

本開示の方法によって生成されたリードペアは、ゲノム、およびその中の染色体と核酸分子の、三次元構造を解析するために使用され得る。本明細書で議論される通り、リードペア中の各リードは、ゲノム中の異なる領域にマッピングされ得る。所与のリードペアについては、それらがマッピングされる、ゲノム中の2つの異なる領域は、共にライゲーシ

ヨンされることができるようにするために、互いに空間的に近接していたであろうと推測することができる。リードペア中の両方のリードの座標によるサンプルからのリードペアをプロットすることによって、接触マップをサンプルのために作り出すことができる。例示的な接触マップは図13に確認することができ、ここで、接触マップに上の各ポイントは、そのリードペアのマッピングされた位置に応じてプロットされたリードペアを表わす。

【0253】

サンプルの全体にわたる接触の解析は、染色体およびゲノムの構造の解析を可能にすることができる。ゲノムの、AおよびBのコンパートメント、活性および非活性のコンパートメント、染色体コンパートメント、ユークロマチンおよびヘテロクロマチン、TADサブタイプを含むトポロジカル関連ドメイン（TAD）、および他の構造への編成は、キロベース、またはメガベーススケールで解析され得る。接触マップの解析はまた、再配置、転位置、コピー数変異、反転、欠失、および挿入などの、構造変異体などのゲノムの特徴の検出を可能にし得る。

【0254】

本開示の方法は、約1bp、2bp、3bp、4bp、5bp、6bp、7bp、8bp、9bp、10bp、20bp、30bp、40bp、50bp、60bp、70bp、80bp、90bp、100bp、200bp、300bp、400bp、500bp、600bp、700bp、800bp、900bp、1000bp、2000bp、3000bp、4000bp、5000bp、6000bp、7000bp、8000bp、9000bp、10kb、20kb、30kb、40kb、50kb、60kb、70kb、80kb、90kb、または100kb未満、あるいはそれらに等しい分解能で、タンパク質結合、構造変異、またはゲノム接触相互作用の位置を提供することができる。場合によっては、タンパク質結合部位、タンパク質フットプリント、接触相互作用、または他の特徴は、1000bp内、900bp内、800bp内、700bp内、600bp内、500bp内、400bp内、300bp内、200bp内、190bp内、180bp内、170bp内、160bp内、150bp内、140bp内、130bp内、120bp内、110bp内、100bp内、90bp内、80bp内、70bp内、60bp内、50bp内、40bp内、30bp内、20bp内、10bp内、9bp内、8bp内、7bp内、6bp内、5bp内、4bp内、3bp内、2bp内、1bp内でマッピングされ得る。一例では、本開示の方法は、ゲノム上で互いに10,000bp、5,000bp、2,000bp、または1,000bp内にある部位（例えばCTCF部位などのタンパク質結合部位）の分解能を可能にし得る。場合によっては、保護されていない核酸（例えば、結合タンパク質のフットプリント内にはない核酸）を分解するMNaseまたは他のエンドヌクレアーゼの使用によって、改善された分解能またはマッピングを達成することができ、これによって、結果として、保護領域の端で発生する近接ライゲーション事象がもたらされる（例えば、タンパク質フットプリント）。

【0255】

コンティグマッピング

様々な実施形態では、本開示は、複数のコンティグへの複数のリードペアのマッピングを可能にする、様々な方法を提供する。リードをコンティグ配列にマッピングするための公的に利用可能なコンピュータプログラムがいくつか存在する。これらのリードマッピング・プログラムデータはまた、特定のリードマッピングがゲノム内でどれほど固有であるかを説明するデータを提供する。コンティグ内で高い信頼性を有しながら固有にマッピングするリードの集団から、各リードペアのリード間の距離の分布を推論することができる。異なるコンティグに信頼性を有しながらマッピングするリードのリードペアについては、このマッピングデータは、問題の2つのコンティグ間の接続を示唆する。それはまた、上記の解析から学習した距離の分布に比例する2つのコンティグ間の距離も示唆する。従って、異なるコンティグに信頼性を有しながらマッピングするリードの各リードペアは、正確なアセンブリ中のそれらの2つのコンティグ間の接続を示唆する。このようなマッピングされたすべてのリードペアから推測される接続は隣接行列に要約することができ、こ

ここで、各コンティグは行と列の両方によって表わされる。コンティグを接続するリードペアは、リードペア中のリードがマッピングされていたコンティグを示す対応する行と列で、非0値としてマーキングされる。ほとんどのリードペアは、コンティグ内でマッピングするであろう、そしてそこから、リードペア間の距離の分布が学習され得、そしてそこから、コンティグの隣接行列が、異なるコンティグにマッピングするリードペアを使用して構築され得る。

【0256】

様々な実施形態では、本開示は、リードペアデータからのリードマッピングデータを使用して、コンティグの隣接行列を構築する工程を含む方法を提供する。いくつかの実施形態では、隣接行列は、長距離相互作用よりも短距離相互作用の傾向を組み込むリードペアに対して、重み付けスキームを使用する。短距離におよぶリードペアは、通常、より長い距離におよぶリードペアより一般的である。特定の距離の確率を説明する関数は、この分布を学習するために単一のコンティグにマッピングするリードペアデータを使用して、適合することができる。従って、異なるコンティグにマッピングするリードペアの1つの重要な特徴は、それらがマッピングを行うコンティグ上のポジションである。両方がコンティグの一方の末端の近くにマッピングを行うリードペアの場合、これらのコンティグ間の推定距離は短くなる可能性があり、そしてそのため、結合されたリード間の距離は小さくなる可能性がある。リードペア間のより短い距離がより長い距離より一般的であるため、この構成は、コンティグの端から遠く離れたリードマッピングよりも、これら2つのコンティグが隣接しているという強力な証拠を提供する。従って、隣接行列中の接続は、リードの、コンティグの端までの距離によってさらに重み付けされる。さらなる実施形態では、隣接行列をさらに再スケーリングして、ゲノムの無差別領域を表すいくつかのコンティグ上の多数の接触をダウンウェイト (down-weight) することができる。ゲノムのこれらの領域は、それらへのリードマッピングの高割合を有することによって同定可能であり、アセンブリに誤った情報を与え得る偽のリードマッピングを、演繹的に含有する可能性がより高くなる。またさらなる実施形態では、このスケーリングは、転写抑制因子CTCF、内分泌受容体、コヒーシン、または共有的に修飾されたヒストンなどの、クロマチンのスクヤフォールド相互作用を調節する1以上の薬剤に対して、1以上の保存された結合部位をサーチすることによって指示され得る。

【0257】

いくつかの実施形態では、本開示は、本明細書に開示される1以上の方法を提供し、当該方法は、隣接行列を解析する工程であって、それによってゲノムへのそれらの順序および／または配向を表すコンティグを通る経路を判定する、工程を含む。他の実施形態では、コンティグを通る経路は、各コンティグが正確に1度訪れられるように、選択され得る。さらなる実施形態では、コンティグを通る経路は、隣接行列を通る経路が訪れられる橋の重み付けの合計を最大化するように、選択される。このようにして、最も恐らくは、コンティグ接続は、正確なアセンブリに提案される。またさらなる実施形態では、コンティグを通る経路は、各コンティグが正確に一度訪れられ、そして隣接行列の端の重み付けが最大化されるように、選択され得る。

【0258】

ハプロタイプ・フェージング

二倍体ゲノムでは、どの対立遺伝子変異体が同じ染色体上でリンクされているかを知ることが、しばしば重要である。これは、ハプロタイプ・フェージングとして知られている。ハイスループット配列データからの短いリードでは、どの対立遺伝子変異体がリンクされているかを直接観察できることはめったにない。ハプロタイプ・フェージングの計算上の推論は、長距離で信頼性を損ない得る。本開示は、どの対立遺伝子変異体が発見されたリードペア上の対立遺伝子変異体を使用してリンクされるかを判定することを可能にする、1以上の方法を提供する。場合によっては、本開示の方法によるフェージングは、インピュテーション (imputation) なしで実行される。

【0259】

様々な実施形態では、本開示の方法および組成物は、複数の対立遺伝子変異体に関して二倍体または倍数体ゲノムのハプロタイプ・フェージングを可能にする。従って、本明細書に記載される方法は、リードペアおよび／またはそれを使用してアセンブルされたコンティグからの変異体情報に基づいてリンクされる、リンクされた対立遺伝子変異体の判定を提供することができる。対立遺伝子変異体の例は、1000人ゲノムプロジェクト（1000 genomes）、UK10K、HapMap、およびヒト間の遺伝的変異を発見するための他のプロジェクトによって周知であるものを含むが、これらに限定されない。特定の遺伝子との疾患の関連性は、例えば、シャルコー・マリー・トゥース神経障害（Charcot-Marie-Tooth neuropathy）につながるSH3TC2の両コピー中に、リンクされていない不活性な変異が確認されたこと（Lupski JR, Reid JG, Gonzaga-Jauregui C等 N. Engl. J. Med. 362:1181-91, 2010）、および、高コレステロール血症9につながるABCG5の両コピー中に、リンクされていない不活性な変異が確認されたこと（Rios J, Stein E, Shendure J等 Hum. Mol. Genet. 19:4313-18, 2010）、によって実証される通り、ハプロタイプフェージングデータを有することによって、より容易に明らかにされ得る。

【0260】

ヒトは、平均的に、1,000中1部位でヘテロ接合性を有する。場合によっては、ハイスループット配列決定方法を使用する単一のレーンのデータは、少なくとも約150,000,000のリードペアを生成することができる。リードペアは、約100の塩基対の長さであり得る。これらのパラメータから、ヒト・サンプルからのすべてのリードの10分の1が、ヘテロ接合性部位をカバーすると推定される。従って、平均で、ヒト・サンプルからのすべてのリードペアの100分の1は、ヘテロ接合性部位のペアをカバーすると推定される。従って、約1,500,000のリードペア（150,000,000の100分の1）が、単一のレーンを使用して、フェージングデータを提供する。ヒトゲノムは約30億の塩基を有し、そして1000分の1がヘテロ接合性であるため、平均的なヒトゲノムには、約300万のヘテロ接合性部位が存在する。ヘテロ接合性部位のペアを表す約1,500,000のリードペアを使用して、ハイスループット配列決定方法の単一のレーンを使用してフェージングされる各ヘテロ接合性部位の平均カバレッジは、一般的なハイスループット配列決定の機械を使用して、約（1X）である。従って、二倍体ヒトゲノムは、本明細書に開示される方法を使用して調製されたサンプルからの配列変異体に関連する高スループット配列データの1レーンで、確実にかつ完全にフェージングされ得る。いくつかの例では、データのレーンは、DNA配列リードデータのセットであり得る。さらなる例では、データのレーンは、ハイスループット配列決定機器の単一の実行からのDNA配列リードデータのセットであり得る。

【0261】

ヒトゲノムは染色体の2つの相同的なセットからなるため、個体の真の遺伝的組立を理解するには、遺伝物質の母方および父方のコピーまたはハプロタイプの描写が必要である。個体のハプロタイプを取得することは、いくつかの点で有用である。第一に、ハプロタイプは、臨床的に臓器移植におけるドナー・ホストマッチングの結果を予測する際に有用であり、そして、疾患の関連性を検出するための手段としてますます使用されている。第二に、複合ヘテロ接合性を示す遺伝子において、ハプロタイプは、2つの有害な変異体と同じ対立遺伝子に位置するかい否かについての情報を提供し、これらの変異体の遺伝が有害であるか否かの予測に大きく影響を与える。第三に、個人の群からのハプロタイプは、人口構造と人類の進化の歴史に関する情報を提供している。最後に、最近説明された遺伝子発現における広範な対立遺伝子の不均衡は、対立遺伝子間の遺伝的差異または後成的差異が、発現の量的差異に寄与し得ることを示唆している。ハプロタイプ構造の理解は、対立遺伝子の不均衡に寄与する変異体のメカニズムを描写する。

【0262】

特定の実施形態では、本明細書に開示される方法は、長距離リンケージおよびフェージ

ングのために必要とされるような、ゲノムの離れた領域間の会合を固定およびキャプチャするための、インビトロ技術を含む。場合によっては、方法は、ゲノム的に非常に離れたリードペアを提供するために、XLRPライブラリを構築し、そして配列決定する工程を含む。場合によっては、相互作用は主に、単一のDNA断片内のランダム会合から生じる。いくつかの例では、セグメント間のゲノム距離を推測することができ、その理由は、DNA分子中で互いに近いセグメントはより頻繁に、かつより高い確率で相互作用する一方で、分子の離れた部分間の相互作用はそれほど頻繁ではないからである。結果的に、2つの遺伝子座を接続するペアの数と入力DNA上のそれらの近接性との間には、体系的な関係が存在する。本開示は、抽出に際して最大のDNA断片におよぶことができるリードペアを生成することができる。このライブラリの入力DNAの最大長は150kbpであり、これは配列決定データから観察された、意味のある最長のリードペアである。このことは、より大きな入力DNA断片が提供される場合に、本方法がさらにゲノム的に離れた遺伝子座をリンクすることができることを示唆している。本方法によって生成されたデータの型を処理するように特別に適合された改善されたアセンブリソフトウェアツールを適用することによって、完全なゲノムアセンブリが可能となり得る。

【0263】

非常に高いフェージング精度は、本開示の方法および組成物を使用して生成されたデータによって達成され得る。以前の方法との比較では、本明細書に記載される方法は、より高割合の変異体をフェージングすることができる。フェージングは、ハイレベルな精度を維持しつつ達成され得る。本明細書における技術は、約70%、80%、90%、95%、96%、97%、98%、99%、99.9%、99.99%、または99.999%を超える精度でフェージングを可能にすることができる。本明細書における技術は、500x未満の配列決定深度、450x未満の配列決定深度、400x未満の配列決定深度、350x未満の配列決定深度、300x未満の配列決定深度、250x未満の配列決定深度、200x未満の配列決定深度、150x未満の配列決定深度、100x未満の配列決定深度、または50x未満の配列決定深度を有する正確なフェージングを可能にすることができる。このフェーズ情報は、例えば、約200kbp、約300kbp、約400kbp、約500kbp、約600kbp、約700kbp、約800kbp、約900kbp、約1Mbp、約2Mbp、約3Mbp、約4Mbp、約5Mbp、または約10Mbpを超える長距離に拡張され得る。いくつかの実施形態では、90%を超えるヒト・サンプルのヘテロ接合性SNPが、例えば、Illumina HiSeqのデータの1レーンのみを使用することによって、約2億5000万未満のリードまたはリードペアを使用して、99%を超える精度でフェージングされ得る。他の場合では、約40%、50%、60%、70%、80%、90%、95%、または99%を超える、ヒト・サンプルのためのヘテロ接合性SNPが、約70%、80%、90%、95%、96%、97%、98%、99%、99.9%、99.99%、または99.999%を超える精度で、約2億5000万または約5億未満のリードおよびリードペアを使用して、例えばIllumina HiSeqデータの1レーンまたは2レーンのみを使用して、フェージングされ得る。例えば、95%または99%を超えるヒト・サンプルのヘテロ接合性SNPは、約95%または99%を超える精度で、約2億5000万または約5億のリードを使用して、フェージングされ得る。さらなる場合では、追加的変異体は、リード長さを、約200bp、250bp、300bp、350bp、400bp、450bp、500bp、600bp、800bp、1000bp、1500bp、2kbp、3kbp、4kbp、5kbp、10kbp、20kbp、50kbp、または100kbpに増加させることによって、キャプチャされ得る。

【0264】

本開示の他の実施形態では、XLRPライブラリからのデータが、長距離のリードペアのフェージング能力を確認するために使用され得る。これらの結果の精度は、以前に利用可能であった最良技術と同等であるが、有意により長い距離にさらに拡張されている。特定の配列決定方法の現在のサンプル調製プロトコルは、フェージングの標的部位のリード

10

20

30

40

50

長さ、例えば150bp内に位置する変異体を認識する。一例では、アセンブリのベンチマークサンプルであるNA12878用に構築されたXLRPライブラリから、存在する1,703,909個のヘテロ接合性SNPの44%が、99%を超える精度でフェージングされた。場合によっては、この比率は、酵素または消化条件の賢明な選択によって、ほぼすべての可変部位に拡張され得る。

【0265】

ハプロタイプ・フェージングは、ヒト白血球抗原（HLA）領域（例えば、クラスIのHLA-A、B、およびC；クラスIIのHLA-DRB1/3/4/5、HLA-DQA1、HLA-DQB1、HLA-DPA1、HLA-DPB1）をフェージングすることを含み得る。ゲノムのHLA領域は、密に多形型であり、そして標準的な配列決定アプローチで配列決定すること、またはフェージングすることが困難であり得る。本開示の技術は、ゲノムのHLA領域の改善された配列決定、および改善されたフェージングを提供することができる。本開示の技術を使用して、ゲノムのHLA領域は、より大きな領域（例えば、染色体アーム、染色体、ゲノム全体）のフェージングの一部として、またはそれ自体で（例えば、ハイブリッドキャプチャなどの標的濃縮によって）、正確にフェージングされ得る。一例では、HLA領域自体は、約300xの配列決定深度で、正確にフェージングされた。これらの技術は、長距離PCRなどの、HLA解析のための従来のアプローチに勝る利点を提供することができ、例えば、長距離PCRは、複合体プロトコルや、多数の別の反応を含むことができる。本明細書でさらに議論される通り、サンプルは、例えば、架橋オリゴヌクレオチドまたは他の場所にサンプル識別バーコードを含むことによって、および、バーコードに基づいて配列情報を逆多重化することによって、配列決定解析のため多重化され得る。一例では、複数のサンプルが近接ライゲーションにかけられ、サンプル識別バーコードで（例えば架橋オリゴヌクレオチド中で）バーコード化され、HLA領域が（例えば、ハイブリッドキャプチャによって）標的化され、多重配列決定が実行されて、複数のサンプルのためのHLAのフェージングが可能になる。場合によっては、HLA領域のフェージングは、インピュテーションなしで実行される。

【0266】

ハプロタイプ・フェージングは、キラー細胞免疫グロブリン様受容体（KIR）領域をフェージングすることを含み得る。ゲノムのKIR領域は、トランスポゾン媒介組み換えのために相同性が高く、かつ構造的に動的であり、そして、標準的な配列決定アプローチを使用した配列決定またはフェージングが困難になり得る。本開示の技術は、ゲノムのKIR領域の改善された配列決定、および改善されたフェージング精度を提供することができる。本開示の技術を使用して、ゲノムのKIR領域は、より大きな領域（例えば、染色体アーム、染色体、ゲノム全体）のフェージングの一部として、またはそれ自体で（例えば、ハイブリッドキャプチャなどの標的濃縮によって）、正確にフェージングされ得る。これらの技術は、長距離PCRなどの、HLA解析のための従来のアプローチに勝る利点を提供することができ、例えば、長距離PCRは、複合体プロトコルや、多数の別の反応を含むことができる。本明細書でさらに議論される通り、サンプルは、例えば、架橋オリゴヌクレオチドまたは他の場所にサンプル識別バーコードを含むことによって、および、バーコードに基づいて配列情報を逆多重化することによって、配列決定解析のため多重化され得る。一例では、複数のサンプルが近接ライゲーションにかけられ、サンプル識別バーコードで（例えば架橋オリゴヌクレオチド中で）バーコード化され、KIR領域が（例えば、ハイブリッドキャプチャによって）標的化され、多重配列決定が実行されて、複数のサンプルのためのKIR領域のフェージングが可能になる。少なくとも約8、9、10、11、12、13、14、15、16、17、またはそれを超える遺伝子および／または偽遺伝子が、フェージングされ得る。場合によっては、KIR領域のフェージングは、インピュテーションなしで実行される。

【0267】

メタゲノミクス解析

いくつかの実施形態では、本明細書に記載される組成物および方法は、例えばヒトの腸

10

20

30

40

50

で確認されるものなどの、メタゲノムの調査を可能にする。従って、所与の生態学的環境に生息するいくつかまたはすべての生物の部分的または全体的なゲノム配列を、調査することができる。例としては、すべての腸内微生物、皮膚の特定のエリアで確認される微生物、および有毒な廃棄物サイトに生息する微生物の、ランダムシーケンシングが含まれる。これらの環境における微生物集団の組成は、本明細書に記載の組成物および方法、ならびにそれらのそれぞれのゲノムによってコードされた、相互に関連する生化学的態様を使用して、決定することができる。本明細書に記載される方法は、複雑な生物学的環境からのメタゲノム研究を可能にし、例えば、2、3、4、5、6、7、8、9、10、12、15、20、25、30、40、50、60、70、80、90、100、125、150、175、200、250、300、400、500、600、700、800、900、1000、5000、または10000を超える生物および／または生物の変異体を含むものなどが挙げられる。

10

【0268】

癌ゲノムシーケンシングによって必要とされる高い精度は、本明細書に記載される方法およびシステムを使用して達成され得る。不正確な基準ゲノムは、癌ゲノムを配列決定する際に、ベースコールの困難をもたらしかねない。不均一なサンプルおよび小さな出発材料、例えば生検によって得られたサンプルが、さらなる困難をもたらす。さらに、ヘテロ接合性の大規模な構造の変異および／または損失の検出は、大抵、癌ゲノムシーケンシング、ならびに体細胞変異とベースコールにおけるエラーとを区別する能力にとって、重要である。

20

【0269】

改善された配列決定精度

本明細書に記載されるシステムおよび方法は、2、3、4、5、6、7、8、9、10、12、15、20、また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、および／また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。いくつかの実施形態では、100 ngもの小さな、あるいは数百程のゲノム当量ですらある小さな出発サンプルが、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および方法は、大規模な構造変異および再配置の検出を可能にしてもよく、フェージングされた変異体コールが、約1 kbp、約2 kbp、約5 kbp、約10 kbp、20 kbp、約50 kbp、約100 kbp、約200 kbp、約500 kbp、約1 Mbp、約2 Mbp、約5 Mbp、約10 Mbp、約20 Mbp、約50 Mbp、約100 Mbp、またはそれを超えるヌクレオチドにおよぶ長い配列にわたって得られる場合がある。例えば、フェーズの変異体コールは、約1 Mbpまたは約2 Mbpにおよぶ長い配列にわたって得られる場合がある。

30

【0270】

本明細書に記載される方法およびシステムを使用して判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。短い変異体コールは、必要であれば、コンピュータリソースに記憶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、短い変異体コールからの組み合わせられた情報、およびコンピュータリソースに記憶された情報に基づいて検出することができる。セグメント重複、構造変異の傾向がある領域、高度に可変的であり、かつ医学的に関連するMHC領域、セントロメアおよびテロメアの領域、および、反復領域を持つものを含むがこれら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、セグメント重複、ある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、精度の向上のために再アセンブルされ得る。

40

【0271】

サンプル型は、ローカルで、また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のいずれかで、配列情報に割り当てられ得る。情報のソースが分かっ

50

ている場合、例えば、情報のソースが癌また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られ得る。他のサンプル型の例は、一般に、組織の型、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が、これらに限定されない。癌ゲノムと比較した正常ゲノムなどの、完全な、または部分的な比較ゲノム配列が利用可能である場合、サンプルデータと比較ゲノム配列との差異を判定し、そして随意に出力することができる。

【0272】

臨床的応用

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解析に使用され得る。本明細書に開示される増幅方法は、限定されないが、米国特許第6,449,562号、6,287,766号、7,361,468号、7,414,117号、6,225,109号、および6,110,709号などに確認できるものなどの、遺伝子解析について、当技術分野で周知のデバイス、キット、および方法でを使用することができる。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、または対立遺伝子は、遺伝的疾患などの疾患または疾病に関連付けられ得る。他の場合、多形は、疾患また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、および年齢関連疾病、癌などに関連付けることができる。他の場合、多形は、冠動脈の健康の増大といった有用な特色、HIVまたはマラリアなどの疾患に対する抵抗性、あるいは骨粗鬆症、アルツハイマー病、また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。

【0273】

本開示の組成物および方法は、診断、予後、治療、患者の層別化、薬物の開発、処置の選択、およびスクリーニングの目的のために使用することができる。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解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。

【0274】

本開示の組成物および方法は、ゲノミクスにおいて使用することができる。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および組成物は、診断また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および組成物は、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、および／または医薬開発の標的の同定のために、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的な遺伝子産物であるため、非常に重要である。いくつかの実施形態では、本明細書に記載される方法および組成物は、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および遺伝子発現の両方を同時に測定する。

【0275】

本開示の組成物および方法は、遺伝子発現解析に使用することができる。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、または再配置であり得る。1を超える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、および癌を検出することができる。また、それは環境モニタリング、科学捜査、および食品科学に有用である。核酸上で実行できる遺伝子解析の例には、例えば、SNP検出、STR検出、RNA発現解析、プロモーターメチル化、遺伝子発現、ウィルス検出、ウィルスサブタイピング、および薬剤耐性が含まれる。

【0276】

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、

特定の処置に応答する患者の能力、または患者にとっての最良の処置を判定するために、患者から得られたまたは導き出された有生分子サンプルの解析に適用することができる。本方法は、特定の疾患のためのバイオマーカーを同定するためにも適用することができる。

【0277】

いくつかの実施形態では、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」または疾病の「診断」は、疾病を予測または診断すること、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、あるいは疾病の予後、疾病の進行、また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、血液サンプルは、サンプル中の疾患または悪性細胞型のマーカーの存在および／また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ってアッセイされ得る。

10

【0278】

いくつかの実施形態では、本明細書に記載される方法および組成物は、疾病の診断および予後のために使用される。

【0279】

多数の免疫性、増殖性、およ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。免疫疾患および障害は、アレルギー性疾患および障害、免疫機能の障害、ならびに自己免疫疾患および疾病を含む。アレルギー性疾患および障害は、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、およ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およびII、高IgE症候群、チェディアック・東症候群、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（DiGeorge/Velocardial-facial syndromes）、およびインターフェロン・ガンマーTH1経路欠損症を含む。自己免疫性および免疫性調節異常障害は、限定されないが、関節リウマチ、糖尿病、全身性エリテマトーデス、グレーブス病、グレーブス眼症、クローン病、多発性硬化症、乾癬、全身性硬化症、甲状腺腫とリンパ性甲状腺腫（橋本甲状腺炎、リンパ節様甲状腺腫）、円形脱毛症（alopecia areata）、自己免疫性心筋炎、苔癬硬化症、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、および炎症性腸疾患、同種異系移植片拒絶反応、および感染性細菌また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。

20

【0280】

本開示の方法によって評価され得る増殖性疾患および障害は、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（例えば乳房または子宮の）；サルコイドーシス；ペロニーおよ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、および血管再狭窄を含む。

【0281】

40

本開示の方法によって評価され得る悪性疾患および障害は、血液悪性腫瘍および固形腫瘍の両方を含む。

【0282】

血液悪性腫瘍は、そのような悪性腫瘍が血液由来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特に適している。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、および他のリンパ腫、急性または慢性白血病、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、および形質細胞障害を含む。

【0283】

本開示の方法により評価され得る形質細胞障害は、多発性骨髄腫、アミロイドーシス、

50

およびヴァルデンストレームマクログロブリン血症を含む。

【0284】

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、および他の骨癌、睾丸および卵巣の癌腫、頭頸部腫瘍、および子宮頸部の新生物を含む。

【0285】

遺伝子疾患も、本開示のプロセスにより検出することができる。これは、染色体および遺伝子の異常、または遺伝子疾患に関する出生前また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、ダウン症候群または他のトリソミー、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、リポドーシス、肥満欠損症 (obesity defect)、血友病、先天性代謝異常症、および糖尿病を含む。

【0286】

本開示の方法は、遺伝的疾患に関連する遺伝的またはゲノム的特徴を検出するために使用することができ、遺伝子融合、構造変異、再配置、およびトポロジーの変化、例えばTAD境界の喪失または改変、TADサブタイプの変化、区画の変化、クロマチンタイプの変化、およびメチル化状態 (例えば、CpGメチル化、H3K4me3、H3K27me3、または他のヒストンメチル化) などの修飾状態の変化を含むがこれらに限定されない。

【0287】

本明細書に記載される方法は、サンプル中の細菌またはウイルスそれぞれのマーカーの存在および／または量を判定することによって、病原菌感染、例えば細胞内細菌およびウイルスによる感染を診断するために使用され得る。

【0288】

種々様々な感染症疾患が、本開示のプロセスにより検出され得る。感染症は、細菌、ウイルス、寄生生物、およ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。

【0289】

本開示によって検出することができる細菌感染因子は、大腸菌、サルモネラ、赤痢菌、クレブシエラ、シュドモナス、リステリア菌、結核菌、マイコバクテリウム・アビウム・イントラセルラーレ (Mycobacterium avium intracellulare)、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、百日咳菌、バクテロイド、黄色ブドウ球菌、連鎖球菌肺炎、B溶血性連鎖球菌、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、ナイセリア淋病、ナイセリア髄膜炎疾病群、ヘモフィルス属インフルエンザ、エンテロコッカス・フェカリス、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菌、トレポネーマ・パラジウム、ボレリア・ブルグドルフェリ、ボレリア・リカレンチス、リケッチア性病原菌、ノカルジア、また放線菌 (Actinomycetes) を含む。

【0290】

本開示によって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・デルマトフィティディス、ヒストプラスマ・カプスラーツム、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、カンジダーアルビカンス、アスペルギルス・フミガウタス (Aspergillus fumigatus)、藻菌類 (リゾプス)、スポロトリックス・シェンキー、クロモミコーシス、およびマズラミコーシスを含む。

【0291】

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型肝炎ウイルスおよびC型肝炎ウイルス)、

10

20

30

40

50

エプスタインバーウィルス、サイトメガロウィルス、ヒトパピローマウィルス、オルソミクソウィルス、パラミクソウィルス、アデノウィルス、コロナウィルス、ラブドウィルス、ポリオウィルス、トogaウィルス、ブンヤウィルス、アレナウィルス、風疹ウィルス、およびレオウィルスを含む。

【0292】

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 (*Onchoverva volvulus*)、リーシュマニア、トリパノゾーマ種、住血吸虫種、エントアメーバーヒストリチカ、クリプトスポリジウム (*Cryptosporidium*)、ジアルジア種、トリコモナス (*Trichimonas*) 種、大腸バランチジウム (*Balantidium Colli*)、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 medinesis*)、吸虫、広節裂頭条虫、テニア種、ニューモシスチスーカリニ、およびアメリカ鉤虫 (*Necator americanis*) を含む。

【0293】

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シス、および耐AZT性ヒト免疫不全ウィルスは全て、本開示で同定され得る。

【0294】

従って、本開示の組成物およ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）、または細菌マーカーあるいはウィルスマーカーなどの異物による感染のマーカーのいずれかであり得る。

【0295】

本開示の組成物および方法は、標的分子を識別および／または定量化するために使用することができ、当該標的分子の存在量は、生物学的状態または疾患条件、例えば、疾患状態の結果としてアップレギュレートまたはダウンレギュレートされる血液マーカーを示す。

【0296】

いくつかの実施形態では、本開示の方法および組成物は、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、また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、および亜臨床的な状態の識別に有用である。

【0297】

本開示の方法は、癌に関連する遺伝的またはゲノムの特徴を検出するために使用することができ、遺伝子融合、構造変異、再配置、およびトポロジーの変化、例えばTAD境界の喪失または改変、TADサブタイプの変化、区画の変化、クロマチンタイプの変化、およびメチル化状態（例えば、CpGメチル化、H3K4me3、H3K27me3、または他のヒストンメチル化）などの修飾状態の変化を含むがこれらに限定されない。

【0298】

サンプル

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、またはそれらの組み合わせを含み得る。いくつかの実施形態では、サンプルは、単一の個体の複数のポリヌクレオチドを含む。いくつかの実施形態では、サンプルは、2つ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またはその一部であり、その非限定的な例には、動物、真菌類、原生生物、モネラ、ウィルス、ミトコンドリア、および葉緑体が含まれる。サンプルのポリヌクレオチドは、例えば、培養細胞株、生検、血液サンプル、細胞を含有している流体サンプルを含む、細胞サンプル、組織サンプル、またはそこから由来する臓器サンプルなどの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、通常はヒトなどの哺乳

10

20

30

40

50

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。いくつかの実施形態では、サンプルはDNAを含む。いくつかの実施形態では、サンプルはゲノムDNAを含む。いくつかの実施形態では、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、酵母人工染色体、オリゴヌクレオチドタグ、またはそれらの組み合わせを含む。いくつかの実施形態では、サンプルは、ポリメラーゼ連鎖反応（PCR）、逆転写、およ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する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、逆転写の産物は相補的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、ランダム配列、部分的ランダム配列、およ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、サンプル中に存在する任意の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。

【0299】

いくつかの実施形態では、核酸の鋳型分子（例えばDNAまたはRNA）は、タンパク質、脂質、および非鋳型核酸などの、様々な他の成分を含有している生体サンプルから単離することができる。核酸の鋳型分子は任意の細胞材料から得られ、動物、植物、細菌、真菌類、または他の細胞生物から得ることができる。本開示での使用のための生体サンプルは、ウィルスの粒子または調製物を含む。核酸の鋳型分子は、生物から直接、あるいは、生物から得た生体サンプル、例えば血液、尿、脳脊髄液、精液、唾液、痰、便、および組織から得ることができる。任意の組織または体液の標本が、本開示で使用する核酸のためのソースであり得る。核酸の鋳型分子はまた、初代細胞培養物または細胞株などの培養細胞から単離され得る。鋳型核酸が得られる細胞または組織は、ウィルスまたは他の細胞内病原体に感染され得る。サンプルはまた、生体標本、cDNAライブラリ、ウィルスDNA、またはゲノムDNAから抽出された総体的なRNAであり得る。サンプルはまた、非細胞起源から単離されたDNA、例えば冷凍装置から増幅／単離されたDNAを含み得る。

【0300】

核酸の抽出および精製の方法は、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、またはトリゾールおよびトリリジェント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の使用を伴うまたは伴わない、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、エタノール沈降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）；および（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）。核酸の単離および／または精製の別の例は磁性粒子の使用を含み、核酸は特異的また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）。いくつかの実施形態では、例えば、プロテイナーゼKまたは他の同様のプロテアーゼを用いる消化などの酵素消化工程は、サンプルから望ましくないタンパク質を除去するのを支援するために上記の単離方法に先行し得る（例えば米国特許第7,001,724を参照）。望ましい場合、RNase阻害剤を、溶解緩衝液に加え得る。特定の細胞また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、またはその両方を単離することを目的とし得る。抽出手順の間またはその後DNAとRNAの両方が一緒に単離される時、さらなる工程を利用して、一方または両方を他とは別々に精製する場合もある。例えば、サイズ、配列

10

20

30

40

50

、または他の物理的あるいは化学的特性による精製により、抽出した核酸のサブ断片を生成することもできる。最初の核酸単離ステップに加えて、本開示の方法における任意のステップの後に、過剰なまたは不要な試薬、反応物、または生成物を除去するなどのために、核酸の精製を実行することができる。

【0301】

核酸の鑄型分子は、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。いくつかの場合では、核酸は、最初に生体サンプルから抽出され、その後インビトロで架橋され得る。いくつかの場合では、天然の会合タンパク質（例えばヒストン）を、核酸から更に除去し得る。

【0302】

いくつかの実施形態では、本開示は、例えば、組織、細胞培養物、体液、動物組織、植物、細菌、真菌、ウィルスなどから単離されるDNAを含んでいる高分子量二本鎖DNAへと容易に適用し得る。

【0303】

いくつかの実施形態では、複数の独立したサンプルの各々は、少なくとも約1 ng、2 ng、5 ng、10 ng、20 ng、30 ng、40 ng、50 ng、75 ng、100 ng、150 ng、200 ng、250 ng、300 ng、400 ng、500 ng、1 μg、1.5 μg、2 μg、5 μg、10 μg、20 μg、50 μg、100 μg、200 μg、500 μg、または1000 μg、もしくはそれ以上の核酸物質を独立して含み得る。いくつかの実施形態では、複数の独立したサンプルの各々は、約1 ng、2 ng、5 ng、10 ng、20 ng、30 ng、40 ng、50 ng、75 ng、100 ng、150 ng、200 ng、250 ng、300 ng、400 ng、500 ng、1 μg、1.5 μg、2 μg、5 μg、10 μg、20 μg、50 μg、100 μg、200 μg、500 μg、または1000 μg未満またはより多くの核酸を独立して含み得る。

【0304】

いくつかの実施形態では、末端修復はEpicentre Biotechnologies (Madison, WI) から入手可能な市販のキットを使用して平滑末端5'リン酸化核酸末端を生成するために実行される。

【0305】

アダプター

アダプターオリゴヌクレオチドは、標的ポリヌクレオチドに接合可能な、その少なくとも一部は既知である配列、を有する任意のオリゴヌクレオチドを含む。アダプター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、またはそれらの組み合わせを含み得る。アダプターオリゴヌクレオチドは、一本鎖、二本鎖、または部分的に二重であり得る。一般に、部分的に二重のアダプターは、1つ以上の一本鎖領域および1つ以上の二本鎖領域を含む。二本鎖アダプターは、互いにハイブリダイズされた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）を含み得、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、ミスマッチおよび／または非対合のヌクレオチドに由来する1つ以上のバルジ、またはこれらの任意の組み合わせを残し得る。いくつかの実施形態では、一本鎖アダプターは、互いとハイブリダイズすることができる2つ以上の配列を含む。2つのそのようなハイブリダイズ可能な配列が一本鎖アダプターに含有される時、ハイブリダイゼーションはヘアピン構造（ヘアピンアダプター）をとる。アダプターの2つのハイブリダイズされた領域がハイブリダイズされていない領域によって互いに分離される時、結果として「バブル」構造をとる。バブル構造を含むアダプターは、内部のハイブリダイゼーションを

含む単一のアダプターオリゴヌクレオチドから成るか、または互いにハイブリダイズされた2つ以上のアダプターオリゴヌクレオチドを含み得る。アダプターにおける2つのハイブリダイズ可能な配列間など、内部配列ハイブリダイゼーションは、一本鎖アダプターオリゴヌクレオチド中に二本鎖構造を産生することができる。異なる種類のアダプターは、ヘアピンアダプターと二本鎖アダプター、または異なる配列のアダプターなどを組み合わせて使用することができる。ヘアピンアダプター中のハイブリダイズ可能な配列はオリゴヌクレオチドの1末端または両末端を含むこともあれば、含まない場合もある。どの末端もハイブリダイズ可能な配列に含有されない時、両末端は「フリー (free)」または「オーバーハング」である。1末端のみがアダプター中で別の配列にハイブリダイズされる時、一方の末端は3'オーバーハングまたは5'オーバーハングなどのオーバーハングを形成する。5'ターミナルヌクレオチドおよび3'ターミナルヌクレオチドの両方がハイブリダイズ可能な配列に含有され、それにより5'ターミナルヌクレオチドおよび3'ターミナルヌクレオチドが相補的になり且つ互いとハイブリダイズするようになり、末端は「平滑」と称される。異なるアダプターは、連続反応または同時に、標的ポリヌクレオチドに接合することができる。例えば、第1のアダプターおよび第2のアダプターが同じ反応に加えられ得る。アダプターは、標的ポリヌクレオチドと接合する前に、操作可能である。例えば、ターミナルリン酸が付加または除去され得る。

【0306】

アダプターは、配列またはその相補をアニールする1つ以上の増幅プライマー、配列またはその相補をアニールする配列決定するプライマー、1つ以上のバーコード配列、複数の異なるアダプターまたは異なるアダプターのサブセット間で共有される1つ以上の共通配列、1つ以上の制限酵素の認識部位、1つ以上の標的ポリヌクレオチドのオーバーハングに相補的な1つ以上のオーバーハング、1つ以上のプローブ結合部位（例えば、Illumina, Inc.によって開発されるものなど、大規模の平行配列決定のためのフローセルなど、配列決定プラットフォームへの付着のため）、1つ以上のランダム配列またはランダムに近い配列（例えば、1つ以上の位置で2つ以上の異なるヌクレオチドからランダムに選択された1つ以上のヌクレオチドであり、異なるヌクレオチドの各々はランダム配列を含むアダプターのプールで表された1つ以上の位置で選択される）、およびそれらの組み合わせを含むが、これらに限定されない1つ以上の様々な配列因子を含有することができる。2つ以上の配列因子は、互いに隣接しておらず（例えば、1つ以上のヌクレオチドにより分離される）、互いに隣接し、部分的に重複し、または完全に重複し得る。例えば、配列をアニールする増幅プライマーは、配列をアニールする配列決定プライマーとしても役立つことができる。配列因子を、3'末端またはその付近に、5'末端またはその付近に、あるいはアダプターオリゴヌクレオチドの内部に位置付けることができる。アダプターオリゴヌクレオチドがヘアピンなどの二次構造を形成することができる時、配列因子は、二次構造の外部に部分的または完全に、二次構造の内部に部分的または完全に、あるいは二次構造に関与する配列間に位置付けることができる。例えば、アダプターオリゴヌクレオチドがヘアピン構造を含む時、配列因子は、ハイブリダイズ可能な配列（「ループ」）間の配列の中で含有するハイブリダイズ可能な配列（「ステム」）の内部または外部に部分的あるいは完全に位置付けることができる。いくつかの実施形態では、異なるバーコード配列を有する複数の第1のアダプターオリゴヌクレオチドにおける第1のアダプターオリゴヌクレオチドは、複数における全ての第1のアダプターオリゴヌクレオチド中で共通の配列因子を含む。いくつかの実施形態では、全ての第2のアダプターオリゴヌクレオチドは、第1のアダプターオリゴヌクレオチドにより共有される共通配列要素とは異なる、全ての第2のアダプターオリゴヌクレオチド中で共通の配列因子を含む。配列因子の差異は、異なるアダプターの少なくとも一部が、例えば、配列長さの変化、1つ以上のヌクレオチドの損失または挿入、あるいは1つ以上のヌクレオチド位置でのヌクレオチド組成の変化（塩基の変化または塩基の修飾など）が原因で完全には整列しないようなものであり得る。いくつかの実施形態では、アダプターオリゴヌクレオチドは、1つ以上の標的ポリヌクレオチドに相補的である、5'オーバーハング、3'オーバーハング、また

10

20

30

40

50

はその両方を含む。相補的なオーバーハングは、長さが1、2、3、4、5、6、7、8、9、10、11、12、13、14、15以上のヌクレオチドが含有されるが、これらに限定されない長さの1以上のヌクレオチドであり得る。例えば、相補的なオーバーハングは、長さが約1、2、3、4、5、または6のヌクレオチドであり得る。相補的なオーバーハングは固定配列を含み得る。相補的なオーバーハングは、1つ以上のヌクレオチドのランダム配列を含み得、1つ以上のヌクレオチドが1つ以上の位置で2つ以上の異なるヌクレオチドのセットからランダムに選択されるようになり、異なるヌクレオチドの各々は、ランダム配列を含む相補的なオーバーハングを備えたアダプターのプールで表された1つ以上の位置で選択される。いくつかの実施形態では、アダプターオーバーハングはアデニンまたはチミンからなる。

10

【0307】

アダプターオリゴヌクレオチドは、少なくともそれらが構成される1つ以上の配列因子を収容するのに十分な、任意の適切な長さを呈し得る。いくつかの実施形態では、アダプターは、長さが約10、15、20、25、30、35、40、45、50、55、60、65、70、75、80、90、100、200、それ未満またはそれ以上のヌクレオチドである。いくつかの例では、アダプターは長さが約10から約50のヌクレオチドであり得る。更なる例では、アダプターは長さが約20から約40ヌクレオチドであり得る。

【0308】

本明細書で使用されるように、用語「バーコード」は、バーコードが識別されるために関連するポリヌクレオチドのいくつかの特徴を可能にする既知の核酸配列を指す。いくつかの実施形態において、識別されるポリヌクレオチドの特徴はポリヌクレオチドが由来するサンプルである。いくつかの実施形態において、バーコードは、長さが少なくとも3、4、5、6、7、8、9、10、11、12、13、14、15、またはそれ以上のヌクレオチドであり得る。例えば、バーコードは、長さが少なくとも10、11、12、13、14、または15のヌクレオチドであり得る。いくつかの実施形態において、バーコードは、長さが10、9、8、7、6、5、または4のヌクレオチドより短くなり得る。例えば、バーコードは、長さが10より短いヌクレオチドであり得る。いくつかの実施形態において、いくつかのポリヌクレオチドに関連したバーコードは、バーコードが他のポリヌクレオチドに関連するバーコードとは異なる長さのものである。一般的に、バーコードは十分な長さで、それらが関連するバーコードに基づいてサンプルを識別することができるほど十分に異なる配列を含む。いくつかの実施形態において、バーコード、およびそれが関連するサンプルソースは、1、2、3、4、5、6、7、8、9、10またはそれ以上のヌクレオチドの突然変異、挿入または損失などのバーコード配列の1つ以上のヌクレオチドの突然変異、挿入または損失の後に正確に識別することができる。いくつかの例において、1、2または3のヌクレオチドが突然変異、挿入および／または損失され得る。いくつかの実施形態において、複数のバーコードの各バーコードは、少なくとも2、3、4、5、6、7、8、9、10またはそれ以上の位置など、複数の少なくとも2つのヌクレオチドの位置で、他のすべてのバーコードとは異なる。いくつかの例において、各バーコードは、少なくとも2、3、4または5の位置で、すべてのバーコードと異なり得る。いくつかの実施形態において、第1の部位と第2の部位は両方、複数のバーコード配列の少なくとも1つを含む。いくつかの実施形態において、第2の部位のバーコードは、第1のアダプターオリゴヌクレオチドのためにバーコードから独立して選択される。いくつかの実施形態において、バーコードを有する第1の部位と第2の部位はペアになり、ペアの配列は同じか異なる1つ以上のバーコードを含むようになる。いくつかの実施形態において、開示の方法はさらに、標的ポリヌクレオチドが接合されるバーコード配列に基づいて標的ポリヌクレオチドが由来するサンプルを識別する工程を含む。一般的に、バーコードは、標的ポリヌクレオチドに接合された時に、標的ポリヌクレオチドが由来するサンプルの識別子として役立つ核酸配列を含み得る。

20

30

40

【0309】

アダプターオリゴヌクレオチドは、免疫グロブリンまたは免疫グロブリン結合タンパク

50

質あるいはその断片に連結されるか、リンクされるか、または繋がれ得る。例えば、MNaseなどのDNaseを用いて架橋されたサンプルのインサイツゲノムを消化した後、1つ以上の抗体は、メチル化された部位または転写因子結合部位などで、消化されたクロマチンを結合するサンプルに加えられ得る。次に、タンパク質A、タンパク質G、タンパク質A/G、またはタンパク質Lなどの、免疫グロブリン結合タンパク質またはその断片に連結されるか、リンクされるか、または繋がれたビオチン化されたアダプターオリゴヌクレオチドは、クロマチン中の1つ以上の特異部位にアダプターを標的とするためにサンプルに加えられ得る。その後、サンプルは近接ライゲーションを引き起こすためにリガーゼを用いて処置され得る。さらに、ストレプトアビジンはアダプターにライゲートされたDNAを単離するために使用され得る。その後、架橋はPCRと配列決定を使用して、サンプルを増幅する前に反転し得る。あるいは、オリゴヌクレオチドにリンクされたアダプターは、クリック化学を使用して、精製試薬にリンクすることができる修飾されたヌクレオチドを含み得る。

【0310】

架橋オリゴヌクレオチド

本明細書にもたらされた方法は接合部で複数のセグメントの第1のセグメントと第2のセグメントを付着する工程を含むことができる。いくつかの場合では、付着させる工程は、ビオチンがタグ付けされたヌクレオチドを使用して付着末端を充填し、平滑末端をライゲーションすることを含むことができる。ある場合では、付着させる工程は、少なくとも第1のセグメントと第2のセグメントを架橋オリゴヌクレオチドに接触させることを含む。図15は、核酸が第1のセグメントと第2のセグメントを形成するためにインサイツで消化される、第1のセグメントと第2のセグメントを結合するため、架橋オリゴヌクレオチドを使用して例示的なワークフローを図示する。末端は、架橋オリゴヌクレオチドを第1のセグメントと第2のセグメントの各々にライゲーションする前に、研磨されてポリアデニル化される。第1のセグメントと第2のセグメントは、その後、架橋オリゴヌクレオチドを含む接合部を作成するため、ライゲーションされる。様々な場合では、付着させる工程は、少なくとも第1のセグメントと第2のセグメントをバーコードに接触させることを含むことができる。

【0311】

いくつかの実施形態では、本明細書にもたらされるような架橋オリゴヌクレオチドは、長さが少なくとも約5のヌクレオチドから長さが約50のヌクレオチドまでであり得る。ある実施形態では、架橋オリゴヌクレオチドは長さが約15のヌクレオチドから長さが約18のヌクレオチドであり得る。様々な実施形態では、架橋オリゴヌクレオチドは、長さが少なくとも5、少なくとも6、少なくとも7、少なくとも8、少なくとも9、少なくとも10、少なくとも11、少なくとも12、少なくとも13、少なくとも14、少なくとも15、少なくとも16、少なくとも17、少なくとも18、少なくとも19、少なくとも20、少なくとも25、少なくとも30、少なくとも35、少なくとも40、少なくとも45、少なくとも50、またはそれ以上の長さのヌクレオチドであり得る。一例では、架橋オリゴヌクレオチドは、長さが少なくとも10のヌクレオチドである。別の例では、架橋オリゴヌクレオチドは長さが12のヌクレオチドまたは長さが約12のヌクレオチドである。いくつかの場合では、少なくとも10bpの架橋オリゴヌクレオチドは、安定性を増大させて、短いインサート、染色体間のライゲーション、非特異的なライゲーション、および架橋セルフライゲーションなどの、有害な近接ライゲーション事象を減らすことができる。

【0312】

スプリティングとプーリング・アプローチを使用して、特有のバーコードを有する架橋オリゴヌクレオチドを産生することができる。サンプルの集団は複数の群へ分割することができ、架橋オリゴヌクレオチドはサンプルに付着させることができ、その結果、架橋オリゴヌクレオチド・バーコードは群の間では異なるが群内では同じであるようになり、サンプルの群はまたともにプーリングすることができ、このプロセスを複数回繰り返すこ

とができる。例えば、ポリヌクレオチドの集団は群Aと群Bに分割することができる。第1の架橋オリゴヌクレオチドは群Aでポリヌクレオチドに付着させることができ、第2の架橋オリゴヌクレオチドは群Bでポリヌクレオチドに付着させることができる。適切に、架橋オリゴヌクレオチド・バーコードは群A内で同じであるが、架橋オリゴヌクレオチドは群Aと群Bとで異なる。このプロセスを繰り返すことによって、架橋オリゴヌクレオチド・バーコードの特有のシリーズを有する集団中の各サンプルをもたらすことができ、単一サンプル（例えば、単細胞、単核、単染色体）解析を可能にする。1つの例示的な例では、ビーズの固体支持体に付着した架橋された消化核のサンプルは、8つの管にわたって分割することができ、各々は、ライゲートされる二本鎖DNA（dsDNA）アダプターを含む第1のアダプター群（第1の反復）の8つの特有のメンバのうち1つを含有する。8つのアダプターの各々は、核中の架橋されたクロマチン集合体の核酸末端にライゲーションするために同じ5'オーバーハング配列を有することができるが、そうでなければ、特有のdsDNA配列を有する。第1のアダプター群がライゲートされた後、核は再びプーリングされ、ライゲーション反応成分を除去するために洗浄することができる。分配、ライゲートティング、およびプーリングスキームは、さらに2回（2反復）繰り返すことができる。各アダプター群のメンバのライゲーション後、架橋されたクロマチン集合体を複数のバーコードに連続して付着させることができる。いくつかの場合では、複数のアダプター群（反復）の複数のメンバの連続するライゲーションは、結果としてバーコードの組み合わせが生じる。利用可能なバーコードの組み合わせの数は、1つの反復当たりの群の数と、使用されるバーコード・オリゴヌクレオチドの合計数次第である。例えば、8つのメンバを含む3つの反復はそれぞれ83の組み合わせの可能性があり得る。いくつかの場合では、バーコードの組み合わせは特有である。いくつかの場合では、バーコードの組み合わせは冗長である。バーコードの組み合わせの合計数は、特有のバーコードを受け入れる群の数を増加または減少させること、および／または反復の数を増加または減少させることによって、調節することができる。1つ以上のアダプター群が使用される場合、分配、付着、およびプーリングスキームは、反復性のアダプターの付着のために使用することができる。いくつかの場合では、分配、付着、およびプーリングスキームは、さらに少なくとも3、4、5、6、7、8、9、または10回繰り返すことができる。いくつかの場合では、最後のアダプター群のメンバは、例えばPCR増幅による配列決定ライブラリ調製の間、アダプターが付着したDNAの次の濃縮のための配列を含む。

【0313】

このプロセス（スプリッティングとプーリングの）を繰り返すことは、最終的に結果として架橋オリゴヌクレオチド・バーコードの特有のシリーズを有する集団中の各サンプルをもたらすことができ、単細胞（例えば、単細胞、単核、および単染色体）解析を可能にする。図16および図17は、スプリッティングとプーリング・アプローチを使用して例示的なワークフローを示し、核酸はインサイツで消化され、その後末端は研磨されてポリアデニル化される。単細胞は分注され、バーコードは各細胞中（例えば、バーコードbc1）に存在する末端にライゲートされる。細胞はプーリングされ、単細胞はその後単離され、第2のバーコードは各細胞中（例えば、バーコードbc2）に存在する末端にライゲートされる。細胞は再びプーリングされて架橋アダプター（例えば、バイオブリッジ（Bio-Bridge））をライゲートする前に単細胞へと分離し、バーコードの特有の組み合わせを有する2つのセグメントと、接合部が由来した細胞を識別するアダプター（バーコードbc1とbc2）との間の接合部を形成する別のDNAセグメントにライゲートすることができる。架橋アダプターは、後のブルダウンまたは他の精製のために、ビオチンなどの1つ以上の親和性試薬を含むことができる。図18は、バーコードの組み合わせの例と、スプリッティングとプーリング・アプローチから生じる架橋を示す。

【0314】

別の例示的な例では、ビーズの固体支持体に付着した架橋された消化核のサンプルは、8つの管にわたって分割することができ、各々は、ライゲートされる二本鎖DNA（dsDNA）アダプターを含む第1のアダプター群（第1の反復）の8つの特有のメンバのう

ち1つを含有する。8つのアダプターの各々は、核中の架橋されたクロマチン集合体の核酸末端にライゲーションするために同じ5'オーバーハング配列を有することができるが、そうでなければ、特有のdsDNA配列を有する。第1のアダプター群がライゲートされた後、核は再びプーリングされ、ライゲーション反応成分を除去するために洗浄することができる。分配、ライゲーション、およびプーリングスキームは、さらに2回（2反復）繰り返すことができる。各アダプター群のメンバのライゲーション後、架橋されたクロマチン集合体を複数のバーコードに連続して付着させることができる。

【0315】

いくつかの場合では、複数のアダプター群（反復）の複数のメンバの連続するライゲーションによって、バーコードの組み合わせが生じ得る。利用可能なバーコードの組み合わせの数は、1つの反復当たりの群の数と、使用されるバーコード・オリゴヌクレオチドの合計数次第であり得る。例えば、8つのメンバを含む3つの反復はそれぞれ83の組み合わせの可能性がある。いくつかの場合では、バーコードの組み合わせは特有である。いくつかの場合では、バーコードの組み合わせは冗長である。バーコードの組み合わせの合計数は、特有のバーコードを受け入れる群の数を増加または減少させること、および／または反復の数を増加または減少させることによって調節することができる。1つ以上のアダプター群が使用される場合、分配、付着、およびプーリングスキームは、反復性のアダプターの付着のために使用することができる。様々な場合では、分配、付着、およびプーリングスキームは、さらに少なくとも3、4、5、6、7、8、9、10回、またはそれ以上の回数で繰り返すことができる。いくつかの場合では、最後のアダプター群のメンバは、例えばPCR増幅によるライブラリ調製物の配列決定の間、アダプターが付着したDNAの次の濃縮のための配列を含み得る。

【0316】

いくつかの場合では、3つのオリゴ設計が使用され得、8つの異なるビオチン化オリゴと組み合わせられる2つの96-ウェルプレートが使用され得るスプリッティングプーリングストラテジーを可能にし、73, 728の異なる分子を明確なバーコーディングを可能にする。ある場合では、8つのオリゴの第1の2つのセットはビオチン化されず、8つのオリゴの第3のセットはビオチン化される。様々な場合では、各バーコードされたオリゴヌクレオチドは方向性があり、1つのオリゴのみが各ラウンドで加えられることを可能にする。架橋オリゴヌクレオチドはそれに対応する末端と調和することを可能にする配列を有することができる。

【0317】

ある場合では、バーコードとアダプターが、十分にライゲートされた架橋によって得られた配列間隔の量を減らすために、より短い配列を有し得る。様々な場合では、架橋は30bpの配列間隔を占めることがある。いくつかの場合では、架橋は54bpの配列間隔を占めるが、特有分子の識別子（UMI）のための更なる位置を提示し得る。ある場合では、UMIは、73, 728の異なる組み合わせを有する単細胞識別を可能にし得る。様々な場合では、第1の2つのオリゴセットは未修飾であり、第3のオリゴセットがビオチン化される。

【0318】

架橋アダプター中のバーコード配列はサンプルの多重配列決定を可能にするために使用することができる。例えば、近接ライゲーションは、いくつかの異なるサンプルで行うことができ、各サンプルは異なるバーコード配列を有する架橋オリゴヌクレオチドを使用する。その後、サンプルは多重配列決定解析のためにプーリングすることができ、配列情報はバーコード配列に基づいて個体サンプルに脱多重化することができる。

【0319】

核酸

真核生物において、ゲノムDNAは、核内の染色体として成り立つためクロマチンに充填される。クロマチンの基本的な構造単位はヌクレオソームであり、これはヒストン八量体の周りを包むDNAの146の塩基対（bp）から成る。ヒストン八量体は、コアヒス

10

20

30

40

50

トンH2A-H2B二量体および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、一般に「数珠玉構造」と呼ばれるものにおいて、DNAに沿って、規則的に間隔が空いている。

【0320】

ヌクレオソームへのコアヒストンおよびDNAのアセンブリは、シャペロンタンパク質および関連するアセンブリ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および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また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。

10

【0321】

一般に、2つのインビトロの方法が、クロマチンを再構成また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、第2のものは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または塩のいずれかを加えたDNAおよび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上にヒストンのランダム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か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)。より長い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することができる。

20

【0322】

過ヨウ素ヌクレオソームアレイのATP依存性のアセンブリは、これはネイティブ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、および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およびリモデリング因子)または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)。

30

【0323】

特定の実施形態では、本開示の方法は、例えば、血漿、血清、および／または尿から単離された遊離DNA；細胞および／または組織からのアポトーシスDNA；および／またはインビトロの酵素的に断片化されたDNA(例えば、DNase Iによって)を含むが、これらに限定されない断片化された二本鎖DNAの任意の型に容易に適用され得る。

【0324】

生体サンプルから得られた核酸は解析に適切な断片を産生するために断片化することができる。様々な酵素法を使用して、鋳型核酸は所望の長さに断片化され得る。DNAはDNaseへの短い暴露でランダムに切断され得る。RNAは、RNase、熱、そしてマグネシウムへの短い暴露、また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は断片化の前または後に、cDNAへと変換されてもよい。核酸分子は、一本鎖、二本鎖、また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。

40

【0325】

いくつかの実施形態では、架橋されたDNA分子は、サイズ選択のステップに晒され得る。核酸のサイズ選択は、一定のサイズより下または上の架橋されたDNA分子に対して

50

行われてもよい。サイズ選択は、さらに架橋の頻度および／または断片化方法によって影響を受けることがある。いくつかの実施形態では、組成物は約145bpから約600bp、約100bpから約2500bp、約600bpから約2500bp、約350bpから約1000bpの範囲、またはこれらの値のいずれか（例えば、約100bpから約2500bp）によって境界される範囲で、DNA分子を架橋することを含み、調製され得る。

【0326】

いくつかの実施形態では、サンプルのポリヌクレオチドは、1以上の特定のサイズの範囲の断片化されたDNA分子の集団へと断片化される。いくつかの実施形態では、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、またはそれ以上の、出発DNAのゲノム等価物から生成され得る。断片化はDNase処置によって達成され得る。いくつかの実施形態では、断片は、平均の長さが、約10から約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、またはそれ以上のヌクレオチドを有する。いくつかの実施形態では、断片は、約145bpから約600bp、約100bpから約2500bp、約600から約2500bp、約350bpから約1000bp、またはこれらの値のいずれか（例えば、約100bpから約2500bp）によって境界される任意の範囲の平均の長さを有する。いくつかの実施形態では、断片は、約2500bp未満、約1200bp未満、約1000bp未満、約800bp未満、約600bp未満、約350bp未満、または約200bp未満の平均の長さを有する。他の実施形態では、断片は、約100bp以上、約350bp以上、約600bp以上、約800bp以上、約1000bp以上、約1200bp以上、または約2000bp以上の平均の長さを有する。DNaseの非限定的な例は、DNase I、DNase II、マイクロコッカスヌクレアーゼ、その変異体、およびそれらの組み合わせを含む。例えば、DNase Iでの消化は、Mg++が無い状態および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。断片化は、5'オーバーハング、3'オーバーハング、平滑末端、またはそれらの組み合わせを有する断片を生成することができる。いくつかの実施形態では、前記方法は、アガロースゲルからのカラム精製また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ステップを含む。

【0327】

標的とされたヌクレアーゼ酵素

本明細書にもたらされるような断片化されたDNAは、ヌクレアーゼ（例えば制限エンドヌクレアーゼ）またはDNase（例えばMnase）の任意の数を用いるインサイツ消化などの消化によって作成されるか生成されることがある。いくつかの場合では、酵素が所望の消化または断片化を達成するために組み合わせて使用されてもよい。様々な場合では、ヌクレアーゼ（またはドメインもしくはその断片）は1つ以上の抗体を使用して、あるゲノムの部位へ標的とされることがある。例えば、架橋されたサンプルは、ヒストン結合部位、転写因子結合部位またはメチル化されたDNA部位などのDNAのある領域へ結合する抗体に接触することがある。タンパク質A、タンパク質G、タンパク質A/G、またはタンパク質Lなど免疫グロブリン結合タンパク質、またはその断片にリンクされたまたは融合されたヌクレアーゼは、その後、サンプルに加えられ得、ヌクレアーゼは抗体が結合した領域のみでDNAを消化し得る。これは、例えば、第1の抗体がDNAサンプルに結合される場合に組み合わせて行われ得、その後、ヌクレアーゼは第1の抗体に標的

とされ、その後、第2の抗体はDNAサンプルに結合され、ヌクレアーゼは第2の抗体を標的とし、所望の消化パターンを達成する。

【0328】

ライゲーション

いくつかの実施形態では、断片化DNAの5'および/または3'の末端ヌクレオチド配列は、ライゲーション前に修飾されない。例えば、予測可能な平滑末端を残す酵素による切断を行い、その後、平滑末端を含む、アダプター、オリゴヌクレオチド、またはポリヌクレオチドなどの核酸への、平滑末端化されたDNA断片のライゲーションが行われ得る。いくつかの実施形態では、断片化DNA分子は、アダプターに接合される前に平滑末端を有するDNA断片を産生するために、研磨された平滑末端（または「修復された末端」）である。平滑末端を研磨するステップは、3'から5'のエキソヌクレアーゼ活性および5'から3'のポリメラーゼ活性を有するDNAポリメラーゼ、例えばT4ポリメラーゼなどの適切な酵素でのインキュベーションにより達成され得る。いくつかの実施形態では、末端の修復は、オーバーハングを生成するために、1つ以上のアデニン、1つ以上のチミン、1つ以上のグアニン、または1つ以上のシトシンなどの1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、またはそれ以上のヌクレオチドの付加後に行われ得る。例えば、末端の修復は、1、2、3、4、5、または6つのヌクレオチドの付加後に行われ得る。オーバーハングを有するDNA断片は、ライゲーション反応などにおいて、相補的なオーバーハングを有するオリゴヌクレオチド、アダプターオリゴヌクレオチド、またはポリヌクレオチドなどの1つ以上の核酸に接合され得る。例えば、単一アデニンは、鋳型の独立したポリメラーゼを使用して末端が修復されたDNA断片の3'末端に加えられ、その後、3'末端で各々がチミンを有する1つ以上のアダプターへのライゲーションが行われ得る。いくつかの実施形態では、オリゴヌクレオチドまたはポリヌクレオチドなどの核酸は、1つ以上のヌクレオチドによる3'末端の伸長とそれに続く5'リン酸化により修飾された平滑末端の二本鎖DNA分子に接合され得る。いくつかの場合では、3'末端の伸長は、マグネシウムを含有する適切な緩衝液中の1つ以上のdNTPの存在下で、クレノーポリメラーゼまたは本明細書で提供される適切なポリメラーゼのいずれかなどのポリメラーゼ、もしくは、ターミナルデオキシヌクレオチドトランスフェラーゼの使用により、実行され得る。いくつかの実施形態では、平滑末端を有する標的ポリヌクレオチドは、平滑末端を含む1つ以上のアダプターに接合される。DNA断片の分子の5'末端のリン酸化は、例えば、ATPおよびマグネシウムを含有する適切な緩衝液中のT4ポリヌクレオチドキナーゼにより実行され得る。断片化DNA分子は随意に、例えばホスファターゼなどの当該技術で既知の酵素の使用により、5'末端または3'末端を脱リン酸化するために処理され得る。

【0329】

本明細書で使用されるような用語「結合すること (connecting)」、「接合すること (joining)」、および「ライゲーション」は、アダプターオリゴヌクレオチドおよび標的ポリヌクレオチドなどの2つのポリヌクレオチドに関して、近接するバックボーンを有する単一のより大きなポリヌクレオチドを産生するために、2つの別個のDNAセグメントの共有結合を指す。2つのDNAセグメントを接合する方法は、当該技術で知られており、限定されないが酵素的および非酵素的（例えば化学的）な方法を含む。非酵素学的なライゲーション反応の例は、文献によって本明細書に取り込まれ、米国特許第5,780,613号および第5,476,930号に記載された非酵素学的なライゲーション技術を含む。いくつかの実施形態では、アダプターオリゴヌクレオチドは、リガーゼ、例えばDNAリガーゼまたはRNAリガーゼによる標的ポリヌクレオチドに接合される。各々が特徴化された反応条件を有する複数のリガーゼは、当該技術で知られており、限定されないが、tRNAリガーゼ、Taq DNAリガーゼ、好熱菌DNAリガーゼ、大腸菌DNAリガーゼ、Tth DNAリガーゼ、Thermus scotoductus DNAリガーゼ (IとII)、熱安定リガーゼ、Ampligase 熱安定DNAリガーゼ、VanC型リガーゼ、9°N DNAリガーゼ、Tsp DNAリガーゼ、およ

びバイオプロスペクティングにより発見された新規なりガーゼを含む、NAD⁺依存性リガーゼ、T4 RNAリガーゼ、T4 DNAリガーゼ、T3 DNAリガーゼ、T7 DNAリガーゼ、Pfu DNAリガーゼ、DNAリガーゼ1、DNAリガーゼIII、DNAリガーゼIV、およびバイオプロスペクティングにより発見された新規なりガーゼを含む、ATP依存性リガーゼ、およびそれらの野生型、突然変異体アイソフォーム、および遺伝的に設計された変異体が挙げられる。

【0330】

ライゲーションは、相補的なオーバーハングなどのハイブリダイズ可能な配列を有するDNAセグメント間で行われ得る。ライゲーションは2つの平滑末端間でも行われ得る。通常、5'リン酸がライゲーション反応に利用される。5'リン酸は標的ポリヌクレオチド、アダプターオリゴヌクレオチドまたは両方によって提供され得る。5'リン酸は、必要に応じて、接合されるDNAセグメントに加えられ、またはそこから除去され得る。5'リン酸の付加または除去のための方法は、当該技術分野では知られており、制限なしで酵素および化学プロセスを含む。5'リン酸の付加および／または除去に役立つ酵素は、キナーゼ、ホスファターゼおよびポリメラーゼを含む。いくつかの実施形態では、ライゲーション反応（例えばアダプター末端、および標的ポリヌクレオチド末端）における接合された2つの末端の両方とも、5'リン酸を提供し、2つの末端を接合する際に、2つの共有結合が作られるようになる。いくつかの実施形態では、ライゲーション反応（例えばアダプター末端および標的ポリヌクレオチド末端のうちの1つのみ）における接合された2つの末端の1つのみが、5'リン酸を提供し、2つの末端を接合する際に、1つの共有結合のみが作られる。

【0331】

いくつかの実施形態では、標的ポリヌクレオチドの1末端または両末端での一鎖のみが、アダプターオリゴヌクレオチドに接合される。いくつかの実施形態では、標的ポリヌクレオチドの1末端または両末端での両鎖が、アダプターオリゴヌクレオチドに接合される。いくつかの実施形態では、3'リン酸はライゲーションに先立って除去される。いくつかの実施形態では、アダプターオリゴヌクレオチドは、標的ポリヌクレオチドの両末端に付加され、各末端の一鎖または両鎖が、1つ以上のアダプターオリゴヌクレオチドに接合されている。両末端の両方の鎖がアダプターオリゴヌクレオチドに接合されている場合、接合に続いて切断反応が起き、それは対応する3'末端の延長のための鋳型としての役割を果たすことができる5'末端オーバーハングを残し、その3'末端はアダプターオリゴヌクレオチドに由来する1つ以上のヌクレオチドを含み得る。いくつかの実施形態では、標的ポリヌクレオチドは、一方の末端の第1のアダプターオリゴヌクレオチドと他方の末端の第2のアダプターオリゴヌクレオチドに接合される。いくつかの実施形態では、標的ポリヌクレオチドの2つの末端は、単一アダプターオリゴヌクレオチドの逆末端に接合される。いくつかの実施形態では、標的ポリヌクレオチドおよびそれが接合されるアダプターオリゴヌクレオチドは、平滑末端を含む。いくつかの実施形態では、別個のライゲーション反応は、少なくとも1つのバーコード配列を含む異なる第1のアダプターオリゴヌクレオチドを使用して各サンプルのために実行され得、バーコード配列は1つより多くのサンプルの標的ポリヌクレオチドに接合されないようになる。DNAセグメント、またはそれに接合されたアダプターオリゴヌクレオチドを有する標的ポリヌクレオチドは、接合されたアダプターにより「タグ付けされた」と考慮される。

【0332】

いくつかの場合では、ライゲーション反応は、約0.1 ng/μL、約0.2 ng/μL、約0.3 ng/μL、約0.4 ng/μL、約0.5 ng/μL、約0.6 ng/μL、約0.7 ng/μL、約0.8 ng/μL、約0.9 ng/μL、約1.0 ng/μL、約1.2 ng/μL、約1.4 ng/μL、約1.6 ng/μL、約1.8 ng/μL、約2.0 ng/μL、約2.5 ng/μL、約3.0 ng/μL、約3.5 ng/μL、約4.0 ng/μL、約4.5 ng/μL、約5.0 ng/μL、約6.0 ng/μL、約7.0 ng/μL、約8.0 ng/μL、約9.0 ng/μL、約10 ng/μL

、約15 ng/μL、約20 ng/μL、約30 ng/μL、約40 ng/μL、約50 ng/μL、約60 ng/μL、約70 ng/μL、約80 ng/μL、約90 ng/μL、約100 ng/μL、約150 ng/μL、約200 ng/μL、約300 ng/μL、約400 ng/μL、約500 ng/μL、約600 ng/μL、約800 ng/μL、または約1000 ng/μLのDNAセグメントまたは標的ポリヌクレオチドの濃度で実行され得る。例えば、ライゲーションは、約100 ng/μL、約150 ng/μL、約200 ng/μL、約300 ng/μL、約400 ng/μL、または約500 ng/μLのDNAセグメントまたは標的ポリヌクレオチドの濃度で実行され得る。

【0333】

いくつかの場合では、ライゲーション反応は、約0.1から1000 ng/μL、約1から1000 ng/μL、約1から800 ng/μL、約10から800 ng/μL、約10から600 ng/μL、約100から600 ng/μL、または約100から500 ng/μLのDNAセグメントまたは標的ポリヌクレオチドの濃度で実行され得る。

【0334】

いくつかの場合では、ライゲーション反応は、約5分、約10分、約20分、約30分、約40分、約50分、約60分、約90分、約2時間、約3時間、約4時間、約5時間、約6時間、約8時間、約10時間、約12時間、約18時間、約24時間、約36時間、約48時間、または約96時間以上にわたり、実行され得る。他の場合、ライゲーション反応は、約5分、約10分、約20分、約30分、約40分、約50分、約60分、約90分、約2時間、約3時間、約4時間、約5時間、約6時間、約8時間、約10時間、約12時間、約18時間、約24時間、約36時間、約48時間、または約96時間未満にわたり実行され得る。例えば、ライゲーション反応は約30分から約90分の間、実行され得る。いくつかの実施形態では、標的ポリヌクレオチドへのアダプターの接合は、アダプター由来のヌクレオチド配列を含む3'オーバーハングを有する接合された産物のポリヌクレオチドを産生する。

【0335】

いくつかの実施形態では、標的ポリヌクレオチドに少なくとも1つのアダプターオリゴヌクレオチドを接合した後、1つ以上の標的ポリヌクレオチドの3'末端は、鑄型として1つ以上の接合されたアダプターオリゴヌクレオチドを使用して伸長される。例えば、標的ポリヌクレオチドの5'末端のみに接合される、2つのハイブリダイズされたオリゴヌクレオチドを含むアダプターは、鑄型としてアダプターの接合された鎖を使用して、標的の接合していない3'末端の伸長を可能にし、それと同時にまたはその後、接合していない鎖の変位を可能にする。2つのハイブリダイズされたオリゴヌクレオチドを含むアダプターの両鎖は、標的ポリヌクレオチドに接合され得、接合された産物は5'オーバーハングを有し、相補的な3'末端は鑄型として5'オーバーハングを使用することで伸長され得るようになる。更なる例として、ヘアピンアダプターオリゴヌクレオチドは、標的ポリヌクレオチドの5'末端に接合され得る。いくつかの実施形態では、伸長される標的ポリヌクレオチドの3'末端は、アダプターオリゴヌクレオチドから1つ以上のヌクレオチドを含む。アダプターが両末端上で接合される標的ポリヌクレオチドについて、伸長は、5'オーバーハングを有する二本鎖標的ポリヌクレオチドの3'末端の両方に対し行われ得る。この3'末端の伸長、または「充填 (fill-in)」反応は、鑄型へハイブリダイズされるアダプターオリゴヌクレオチドの鑄型に、相補配列または「補体」を生成し、それにより、二本鎖配列領域を産生するために5'オーバーハングを充填する。二本鎖標的ポリヌクレオチドの両末端が相補鎖の3'末端の伸長により充填される5'オーバーハングを有する場合、産物は完全に二本鎖である。伸長は、DNAポリメラーゼなどの当該技術分野で既知の任意の適切なポリメラーゼにより行なわれ、それらの多くは市販で入手可能である。DNAポリメラーゼは、DNA依存性DNAポリメラーゼ活性、RNA依存性DNAポリメラーゼ活性、またはDNA依存性かつRNA依存性DNAポリメラーゼ活性を含み得る。DNAポリメラーゼは熱安定性または非熱安定性であり得る。DNAポリメラーゼの例は、限定されないが、Taqポリメラーゼ、Tthポリメラーゼ、Tliポリメラーゼ、Pfuポ

10

20

30

40

50

リメラーゼ、P f u t u b oポリメラーゼ、P y r o b e s tポリメラーゼ、P w oポリメラーゼ、K O Dポリメラーゼ、B s tポリメラーゼ、S a cポリメラーゼ、S s oポリメラーゼ、P o cポリメラーゼ、P a bポリメラーゼ、M t hポリメラーゼ、P h oポリメラーゼ、E S 4ポリメラーゼ、V E N Tポリメラーゼ、D E E P V E N Tポリメラーゼ、E X - T a qポリメラーゼ、L A - T a qポリメラーゼ、E x p a n dポリメラーゼ、P l a t i n u m T a qポリメラーゼ、H i - F iポリメラーゼ、T b rポリメラーゼ、T f lポリメラーゼ、T r uポリメラーゼ、T a cポリメラーゼ、T n eポリメラーゼ、T m aポリメラーゼ、T i hポリメラーゼ、T f iポリメラーゼ、クレノー断片、およびそれらの変異体、修飾した産物、および誘導体を含み、3'末端の伸長は、独立したサンプルから標的ポリヌクレオチドのプーリング前またはプーリング後に実行され得る。

10

【0336】

標的濃縮

ある実施形態では、本開示は、標的核酸の濃縮および標的核酸の解析のための方法を提供する。いくつかの場合では、濃縮の方法は、溶液ベースのフォーマットである。いくつかの場合では、標的核酸は標識剤で標識され得る。他の場合では、標的核酸は、標識剤で標識される1つ以上の会合分子に架橋され得る。標識剤の例は、限定されないが、ビオチン、ポリヒスチジンタグ、および化学タグ（例えば、クリックケミストリー方法に使用されるアルキンおよびアジドの誘導体）を含む。更に、標識された標的核酸が捕捉され、それにより捕捉剤の使用によって濃縮され得る。捕捉剤は、ストレプトアビジンおよび／またはアビジン、抗体、化学部分（例えばアルキン、アジド）、および当該技術分野で既知のアフィニティー精製に使用される任意の生物学的、化学的、物理的、または酵素的な薬剤であり得る。

20

【0337】

いくつかの場合では、固定されたまたは固定されない核酸プローブが、標的核酸を捕捉するために使用され得る。例えば、標的核酸は、固体支持体上で、または溶液中でのプローブへのハイブリダイゼーションにより、サンプルから濃縮され得る。いくつかの例では、サンプルはゲノムサンプルであり得る。いくつかの例では、プローブはアンプリコンであり得る。アンプリコンは予め定められた配列を含み得る。更に、ハイブリダイズされた標的核酸は、プローブから洗浄および／または溶出され得る。標的核酸はDNA、RNA、cDNA、またはmRNAの分子であり得る。

30

【0338】

いくつかの場合では、濃縮方法は、標的核酸を含むサンプルをプローブに接触させる工程、および標的核酸を固体支持体に結合させる工程を含み得る。いくつかの場合では、サンプルは、標的核酸を得るために、酵素的な方法を使用して断片化され得る。いくつかの場合では、プローブは、標的核酸へと特異的にハイブリダイズされ得る。いくつかの場合では、標的核酸は約145bpから約600bp、約100bpから約2500bp、約600から約2500bp、または約350bpから約1000bpの平均サイズを有し得る。標的核酸は更に、サンプル中の結合されていない核酸から分離され得る。固体支持体は、濃縮された標的核酸を提供するために洗浄および／または溶出され得る。いくつかの例では、濃縮ステップは約1、2、3、4、5、6、7、8、9、または10回繰り返され得る。例えば、濃縮ステップは約1、2、または3回繰り返され得る。

40

【0339】

いくつかの場合では、濃縮方法は、プローブ由来のアンプリコンを提供する工程を含み、増幅のための前記プローブは固体支持体に付着される。固体支持体は、サンプルから特異標的核酸を捕捉するために支持体に固定された核酸プローブを含み得る。プローブ由来のアンプリコンは、標的核酸にハイブリダイズされ得る。プローブアンプリコンへのハイブリダイゼーションの後、サンプル中の標的核酸は、捕捉されたプローブからハイブリダイズされた標的核酸を捕捉し（例えばビオチン、抗体などとしての捕捉剤を介して）、かつ洗浄および／または溶出することにより、濃縮され得る。標的核酸配列は更に、例えば濃縮されたPCR産物の増幅されたプールを産生するためのPCR方法を使用して、増幅

50

され得る。

【0340】

いくつかの場合では、固体支持体は、マイクロアレイ、スライド、チップ、マイクロウェル、カラム、管、粒子、またはビーズであり得る。いくつかの例では、固体支持体は、ストレプトアビジンおよび／またはアビジンで覆われ得る。他の例では、固体支持体は抗体で覆われ得る。更に、固体支持体は、ガラス、金属、セラミック、または高分子材料を含み得る。いくつかの実施形態では、固体支持体は核酸マイクロアレイ（例えばDNAマイクロアレイ）であり得る。他の実施形態では、固形支持体は常磁性ビーズであり得る。

【0341】

特定の実施形態では、本開示は、濃縮されたDNAを増幅する方法を提供する。いくつかの場合では、濃縮されたDNAはリードペアである。リードペアは、本開示の方法によって得ることができる。

【0342】

いくつかの実施形態では、1以上の増幅および／または複製のステップが、配列決定されるべきライブラリの調製のために使用される。当該技術分野で既知の任意の増幅方法が使用されてもよい。使用され得る増幅技術の例は、限定されないが、定量PCR、定量蛍光PCR（QF-PCR）、多重蛍光PCR（MF-PCR）、リアルタイムPCR（RT-PCR）、単細胞PCR、制限断片長多型PCR（PCR-RFLP）、PCK-RFLP、IRTP-PCR-IRFLP、ホットスタートPCR、ネステッドPCR、インサイツ polonony PCR、インサイツローリングサークル増幅（RCA）、架橋PCR、ライゲーションを媒介するPCR、Qbレプリカーゼ増幅、逆PCR、picotiter PCR、およびエマルジョンPCRを含み得る。他の適切な増幅方法は、リガーゼ連鎖反応（LCR）、転写増幅、自立配列複製、標的ポリヌクレオチド配列の選択的な増幅、コンセンサス配列プライムポリメラーゼ連鎖反応（CP-PCR）、任意のプライムポリメラーゼ連鎖反応（AP-PCR）、縮重オリゴヌクレオチドプライムPCR（DOP-PCR）、および核酸塩基配列増幅（NABSA）を含む。本明細書で使用され得る他の増幅方法は、米国特許第5,242,794号、第5,494,810号、第4,988,617号、および第6,582,938号に記載されるものを含み得る。

【0343】

特定の実施形態では、PCRは、DNA分子が個々の区分へ分注された後でDNA分子を増幅するために使用される。いくつかの場合では、増幅アダプター内の1つ以上の特異プライミング配列が、PCR増幅のために利用される。増幅アダプターは、個々の区分への分注前または分注後に、断片化DNA分子にライゲートされ得る。両末端上で適切なプライミング配列を有する増幅アダプターを含むポリヌクレオチドは、指数関数的に増幅されたPCRであり得る。適切なプライミング配列の1つのみを有するポリヌクレオチドは、例えば、プライミング配列を含む増幅アダプターの不完全なライゲーション効果が原因で、直線的な増幅を経るのみの場合がある。更に、適切なプライミング配列を含むアダプターがライゲートされない場合、ポリヌクレオチドは、増幅、例えばPCR増幅から一斉に取り除かれ得る。いくつかの実施形態では、PCRサイクルの数は10から30の間で変動するが、9、8、7、6、5、4、3、2、またはそれ以下にまで低くなるか、もしくは40、45、50、55、60、またはそれ以上にまで高くなり得る。その結果、適切なプライミング配列を有する増幅アダプターを運ぶ指数関数的に増幅可能な断片は、PCR増幅の後、直線的に増幅可能または増幅可能ではない断片と比較して、はるかに高い（1000倍以上）濃度で存在し得る。PCRの利点は、全ゲノム増幅技術（ランダム化されたプライマーを用いる増幅またはphi29ポリメラーゼを使用する多置換増幅など）と比較して、限定されないが、より一様な相対配列被覆を含み、一各断片は最大で1サイクル当たり1回コピーすることができ、増幅は熱サイクリングプログラムによって制御されるため、例えば、MDA（Lasken et al., 2007, BMC Biotechnology）より、キメラ分子の形成率が実質的に低くなり、一キメラ分子はアセンブリグラフ中に非生物学的配列を表すことによって、正確な配列アセンブリにと

10

20

30

40

50

って重要な課題を提示し、結果としてミスアセンブリまたは非常に曖昧で断片化されたアセンブリがより高確率で生じるため、MDAで一般的に使用されるランダム化されたプライマーの結合から生じ得る配列の特異的なバイアスが、特異配列を有する特異プライミング部位の使用に対して減少し、最終的に増幅されたDNA産物の量における高い再現性は、PCRサイクル数の選択によって制御することができ、かつ当該技術分野で既知の共通の全ゲノム増幅技術と比較して、PCRにおいて一般的に使用されるポリメラーゼでの複製の忠実性はより高くなる。

【0344】

いくつかの実施形態では、充填反応は、第1のプライマーおよび第2のプライマーを使用した1つ以上の標的ポリヌクレオチドの増幅の後、またはその一部として実行され、第1のプライマーは、第1のアダプターオリゴヌクレオチドの1つ以上の補体の少なくとも一部にハイブリダイズされる配列を含み、さらに第2のプライマーは、第2のアダプターオリゴヌクレオチドの1つ以上の補体の少なくとも一部にハイブリダイズされる配列を含む。第1のプライマーと第2のプライマーの各々は、約10、15、20、25、30、35、40、45、50、55、60、65、70、75、80、90、100以上、もしくはそれら未満またはそれらより多くのヌクレオチドなど、任意の適切な長さであり得、その一部または全てが対応する標的配列に相補的であり得る（例えば、約5、10、15、20、25、30、35、40、45、50またはそれ以上のおおよそ、未満、またはそれ以上のヌクレオチド）。例えば、約10から50のヌクレオチドが対応する標的配列に相補的であり得る。

【0345】

「増幅」は、標的配列のコピー数が増大される任意のプロセスを指す。いくつかの場合では、複製反応は、ポリヌクレオチドの単一の相補的なコピー／レプリカのみを産生し得る。標的ポリヌクレオチドの、プライマーに向けられた増幅のための方法は当該技術分野で知られており、限定されないがポリメラーゼ連鎖反応（PCR）に基づく方法を含む。当該技術分野で既知のPCRによる標的配列の増幅に好ましい条件は、プロセスにおける様々なステップにて最適化され、かつ、標的のタイプ、標的の濃度、増幅される配列の長さ、標的および／または1つ以上のプライマーの配列、プライマーの長さ、プライマーの濃度、使用するポリメラーゼ、反応量、1つ以上の要素と1つ以上の他の要素との比率、およびその他などの反応における要素の特徴に依存し、それらの一部または全ては変更され得る。一般に、PCRは、（二本鎖の場合に）増幅される標的の変性、1つ以上のプライマーの標的へのハイブリダイゼーション、およびDNAポリメラーゼによるプライマーの伸長のステップを含み、ステップは、標的配列を増幅するために繰り返される（または「サイクルされる」）。このプロセスにおける工程は、収量を増強するために、擬似的な産物の形成を減らすために、および／またはプライマーのアニーリングの特異性を増加または減少させるためになど、様々な結果について最適化され得る。最適化の方法は当該技術分野で周知であり、および、増幅反応における要素の型または数、および／または、特定のステップでの温度、特定のステップの持続時間、および／またはサイクル数など、プロセスにおける所与のステップの条件に対する調整を含む。

【0346】

いくつかの実施形態では、増幅反応は、少なくとも約5、10、15、20、25、30、35、40、50、60、70、80、90、100、150、200、またはそれ以上のサイクルを含み得る。いくつかの例では、増幅反応は、少なくとも約20、25、30、35、または40のサイクルを含み得る。いくつかの実施形態では、増幅反応は、ほんの約5、10、15、20、25、35、40、50、60、70、80、90、100、150、200、またはそれ以上のサイクルを含む。サイクルは、1、2、3、4、5、6、7、8、9、10、またはそれ以上の工程など、任意の数のステップを含み得る。ステップは、限定されないが、3'末端伸長（例えばアダプター充填）、プライマーアニーリング、プライマー伸長および鎖変性を含む、所与のステップの目的を達成するのに適切な、任意の温度または温度の勾配を含み得る。ステップの任意の持続時間は、限定さ

れないが、約 1、5、10、15、20、25、30、35、40、45、50、55、60、70、80、90、100、120、180、240、300、360、420、480、540、600、1200、1800 秒以上、もしくはそれ未満またはそれより多く、手動的に中断されるまで無制限であることを含み得る。異なるステップを含む任意の数のサイクルは、任意の順で組み合わせられ得る。いくつかの実施形態では、異なるステップを含む異なるサイクルが組み合わせられ、その結果、組み合わせにおけるサイクルの総数は、約 5、10、15、20、25、30、35、40、50、60、70、80、90、100、150、または 200 以上、もしくはそれら未満またはそれらより多くのサイクルである。いくつかの実施形態では、増幅は充填反応の後に行なわれる。

【0347】

いくつかの実施形態では、増幅反応は、少なくとも約 1、2、3、4、5、6、7、8、9、10、12、14、16、18、20、25、30、40、50、100、200、300、400、500、600、800、1000 ng の標的 DNA 分子上で行なわれ得る。他の実施形態では、増幅反応は、約 1、2、3、4、5、6、7、8、9、10、12、14、16、18、20、25、30、40、50、100、200、300、400、500、600、800、1000 ng 未満の標的 DNA 分子上で行なわれ得る。

【0348】

増幅は、独立したサンプルから標的ポリヌクレオチドをプーリングする前またはプーリングした後に実行され得る。

【0349】

本開示の方法は、サンプルに存在する増幅可能な核酸の量を判定する工程を含む。任意の既知の方法は増幅可能な核酸を定量化するために使用され典型的な方法はポリメラーゼ連鎖反応 (PCR)、具体的に定量的ポリメラーゼ連鎖反応 (qPCR) である。qPCR はポリメラーゼ連鎖反応に基づく技術であり、標的とされた核酸分子を増幅かつ同時に定量化するために使用される。qPCR は、DNA サンプル中の特異配列の検出および定量化 (DNA 入力または追加の標準化遺伝子へと標準化された時のコピーまたは相対量の絶対数として) の両方を可能にする。手順は、増幅された DNA が各増幅サイクルの後にリアルタイムでの反応において蓄積すると定量化される付加的な特徴と共に、ポリメラーゼ連鎖反応の一般的な原理に従う。QPCR は、例えば、Kurnit et al. (米国特許第 6,033,854 号)、Wang et al. (米国特許第 5,567,583 号および第 5,348,853 号)、Ma et al. (The Journal of American Science, 2(3), 2006)、Heid et al. (Genome Research 986-994, 1996)、Sambrook および Russell (Quantitative PCR, Cold Spring Harbor Protocols, 2006)、および Higuchi (米国特許第 6,171,785 号および第 5,994,056 号) に記載されている。これらの内容は、全体において本明細書での引用により組み込まれる。

【0350】

定量化の他の方法は、二本鎖 DNA で挿入する蛍光染料、および、相補的 DNA でハイブリダイズされた時に蛍光を発する修飾された DNA オリゴヌクレオチドプローブの使用を含む。これらの方法は広く使用することができるが、具体的には一例として更に詳細に記載されるようなリアルタイム PCR にも適している。第 1 の方法において、DNA 結合染料は PCR において全ての二本鎖 (ds) DNA に結合し、その結果、染料の蛍光が生じる。それ故、PCR 中の DNA 産物の増加は、蛍光強度の増加を引き起こし、各サイクルにて測定され、故に DNA 濃度の定量化を可能にする。反応は、蛍光性の (ds) DNA 染料の付加により、標準 PCR 反応へと同様に調製される。反応はサーモサイクラー中で実行される、各サイクルの後、蛍光のレベルが検出器で比較され、(ds) DNA (すなわち PCR 産物) に結合された時、染料は蛍光を発するのみである。標準稀釈液に関して、PCR における (ds) DNA 濃度が測定され得る。他のリアルタイム PCR 方法のように、得られた値は、その値に関連する絶対単位を有さない。測定された DNA/RN

10

20

30

40

50

Aサンプルと標準希釈液との比較は、標準に相対的なサンプルの断片または比率をもたらして、異なる組織または実験条件の間の相対的な比較を可能にする。標的遺伝子の定量化および／または発現における正確性を確保することは、安定して発現された遺伝子に関して標準化され得る。未知の遺伝子のコピー数は、既知のコピー数の遺伝子に対して同様に標準化され得る。

【0351】

第2の方法は、プローブ配列を含有するDNAのみを定量化するために配列特異的なRNAまたはDNAに基づくプローブを使用し、それ故、レポータープローブの使用は、特異性を著しく増大させ、いくつかの非特異的なDNA増幅の存在下でさえも定量化を可能にする。これにより、多重化、すなわち、異なるように色を付けられた標識を伴う特異的なプローブにより同じ反応における様々な遺伝子についてのアッセイを可能にするが、ただし、全ての遺伝子は同様の効果で増幅されるものとする。

【0352】

この方法は、プローブの1末端では蛍光レポーター（例えば6-カルボキシフルオレセイン）、およびプローブの逆末端では蛍光のクエンチャー（例えば6-カルボキシテトラメチルローダミン）を有するDNAベースのプローブで一般的に実行される。クエンチャーに対するレポーターの近接性は、その蛍光の検出を妨げる。ポリメラーゼ（例えばTaqポリメラーゼ）の5'から3'のエキソヌクレアーゼ活性によるプローブの破壊は、レポーター-クエンチャーの近接性を破壊し、故に検出され得る蛍光のクエンチされていない発光を可能にする。各PCRサイクルにてレポータープローブにより標的とされる産物の増加の結果、プローブの破壊およびレポーターの放出に起因した蛍光の比例的な増加が生じる。反応は標準PCR反応へと同様に調製され、レポータープローブが加えられる。反応が始まると、PCRのアニール段階中に、プローブとプライマーの両方がDNA標的へとアニールされる。新しいDNA鎖の重合はプライマーから始まり、一旦ポリメラーゼがプローブに到達すると、その5'-3'-エキソヌクレアーゼは、プローブを分解し、クエンチャーから蛍光レポーターを物理的に分離して、結果として蛍光の増加をもたらす。蛍光はリアルタイムPCRサーモサイクラーにおいて検出かつ測定され、産物の指数関数的な増加に対応する蛍光の幾何学的な増加は、各反応における閾値サイクルを判定するために使用される。

【0353】

反応の指数関数的なフェーズ中に存在するDNAの相対濃度は、対数目盛りでのサイクル数に対する蛍光をプロットすることにより判定される（そのため、指数関数的に増大する量は直線をもたらす）。バックグラウンドの上の蛍光の検出のための閾値が判定される。サンプルからの蛍光が閾値にわたるサイクルは、サイクル閾値、C_tと称される。DNAの量は指数関数的なフェーズ中に全てのサイクルごとに2倍になるため、DNAの相対量を計算することができ、例えば、別のものより早い3つのサイクルのC_tを有するサンプルは、 $2^3 = 8$ 倍以上の濃度を有する。その後、核酸（例えばRNAまたはDNA）の量は、結果を、既知量の核酸の連続希釈のリアルタイムPCRにより産生された標準曲線（例えば、希釈無し、1:4、1:16、1:64）と比較することにより判定される。

【0354】

特定の実施形態では、qPCR反応は、蛍光共鳴エネルギー転移（FRET）、例えばLIGHTCYCLERハイブリダイゼーションプローブを利用する2重のフルオロフォア方法を含んでいて、2つのオリゴヌクレオチドプローブがアンプリコンへとアニールされる（例えば、米国特許第6,174,670号）。オリゴヌクレオチドは、効果的なエネルギー転移に適合する距離で分離されたフルオロフォアで頭-尾の配向においてハイブリダイズするように設計される。核酸に結合されたまたは伸長産物に組み込まれた時にシグナルを発するように構造化される標識化オリゴヌクレオチドの他の例は、SCORPIONSプローブ（例えば、Whitcombe et al., Nature Biotechnology 17:804-807, 1999、および米国特許第6,326,145号）、Sunrise（またはAMPLIFLOUR）プライマー（例えば、Na

10

20

30

40

50

zarenko et al., Nuc. Acids Res. 25:2516-2521, 1997、および米国特許第6,117,635号)、およびLUXプライマー並びにMOLECULAR BEACONSプローブ(例えば、Tyagi et al., Nature Biotechnology 14:303-308, 1996および米国特許第5,989,823号)を含む。

【0355】

他の実施形態では、qPCR反応は、蛍光Taqmanの方法、およびリアルタイムで蛍光を測定することが可能な機器(例えば、ABI Prism 7700 Sequence Detector)を用いる。Taqman反応は、2つの異なる蛍光染料で標識されるハイブリダイゼーションプローブを使用する。一方の染料はレポーター染料(6-カルボキシフルオレセイン)であり、他方はクエンチ染料(6-カルボキシテトラメチルローダミン)である。プローブが無傷の場合、蛍光エネルギー転移が生じ、レポーター染料の蛍光的な発光がクエンチ染料により吸収される。PCRサイクルの伸長段階中に、蛍光ハイブリダイゼーションプローブは、DNAポリメラーゼの5'-3'の核酸分解活性により切断される。プローブの切断時に、レポーター染料の発光はこれ以上クエンチ染料へと効率的に転移されず、その結果レポーター染料の蛍光発光スペクトルが生じる。リアルタイムの方法または単点の検出方法を含む、任意の核酸定量化方法は、サンプル中の核酸の量を定量化するために使用され得る。検出は、様々な異なる方法(例えば、染色、標識化プローブでのハイブリダイゼーション、ビオチン化プライマーの組み込み、その後のアビジン-酵素の結合の検出、増幅されたセグメントへの、dCTPまたはdATPなどの32P標識化デオキシヌクレオチド3リン酸の組み込み)に加えて、核酸定量化のための当該技術分野で既知の任意の他の適切な検出方法で実行され得る。定量化は増幅ステップを含むこともあれば、含まないこともある。

【0356】

いくつかの実施形態では、本開示は、リンクされたDNAセグメントを識別または定量化するための標識を提供する。いくつかの場合では、リンクされたDNAセグメントは、アレイのハイブリダイゼーションなどの下流の適用を補助するために標識され得る。例えば、リンクされたDNAセグメントは、ランダムなプライミングまたはニックトランスレーションを使用して標識され得る。

【0357】

多種多様な標識(例えば、レポーター)は、本明細書に記載されるヌクレオチド配列を標識するために使用されてもよく、限定されないが増幅ステップ中を含む。適切な標識は、放射性核種、酵素、蛍光、化学発光、または色原体の薬剤に加え、リガンド、補助因子、阻害剤、磁気微粒子などを含む。そのような標識の例は、米国特許第3,817,837号、米国特許第3,850,752号、米国特許第3,939,350号、米国特許第3,996,345号、米国特許第4,277,437号、米国特許4,275,149号、および米国特許第4,366,241号に含まれるが、これらは、その全体が引用によって本明細書に組み込まれる。

【0358】

追加の標識は、限定されないが、β-ガラクトシダーゼ、インベルターゼ、緑色蛍光タンパク質、ルシフェラーゼ、クロラムフェニコール、アセチルトランスフェラーゼ、β-グルクロニダーゼ、exoglycanaseおよびグルコアミラーゼを含む。蛍光標識も、特定の化学的性質で特異的に合成された蛍光試薬と同様に使用されてもよい。蛍光を測定する多種多様な方法が利用可能である。例えば、いくつかの蛍光標識は、励起または発光のスペクトルの変化を示し、いくつかの蛍光標識は1つの蛍光レポーターが蛍光を失う共鳴エネルギー転移を示すが、第2の蛍光標識は蛍光を増加させ、いくつかの蛍光標識は蛍光の消失(クエンチ)または出現を示し、いくつかの蛍光標識は回転動作を報告する。

【0359】

更に、標識化のための十分な材料を得るために、多数の増幅が、1つの反応ごとの増幅サイクルの回数を増やす代わりにプーリングされ得る。代替的に、標識されたヌクレオチ

10

20

30

40

50

ドは、増幅反応の最後のサイクルにおいて組み込まれ得、例えば、30サイクルのPCR（標識なし）に10サイクルのPCR（標識あり）が加えられる。

【0360】

特定の実施形態では、本開示は、リンクされたDNAセグメントに付着し得るプローブを提供する。本明細書で使用されるように、用語「プローブ」は、対象の別の分子（例えば、別のオリゴヌクレオチド）にハイブリダイズ可能な分子（例えば、精製された制限消化物などにおいて自然に生じ、または合成的に、組み換え的に、もしくはPCR増幅により産生されたかどうかにかかわらず、オリゴヌクレオチド）を指す。プローブは、オリゴヌクレオチドであると、一本鎖または二本鎖であり得る。プローブは、特定の標的（例えば遺伝子配列）の検出、識別、および単離に役立つ。いくつかの場合では、プローブは、任意の検出システムにおいて検出可能となるような標識に関連付けられ、限定されないが、酵素（例えば、ELISAに加え、酵素に基づく組織化学アッセイ）、蛍光、放射性、および発光のシステムを含む。

【0361】

アレイおよびマイクロアレイに関して、用語「プローブ」は、前記プローブへとハイブリダイズしたヌクレオチド配列を検出するためにアレイに固定される、任意のハイブリダイズ可能な材料を指すために使用される。いくつかの場合では、プローブは、約10bpから500bp、約10bpから250bp、約20bpから250bp、約25bpから200bp、約25bpから100bp、約30bpから100bp、または約30bpから80bpであり得る。いくつかの場合では、プローブは、約10bp、約20bp、約30bp、約40bp、約50bp、約60bp、約70bp、約80bp、約90bp、約100bp、約150bp、約200bp、約250bp、約300bp、約400bp、または約500bpの長さであり得る。例えば、プローブは約20から約50bpの長さであり得る。プローブ設計の例および論理的根拠は、WO95/11995、EP717,113、およびWO97/29212にて見出すことができる。

【0362】

プローブ、プローブのアレイ、またはプローブのセットは、支持体上で固定され得る。支持体（例えば固体支持体）は、ガラス、シリカ、プラスチック、ナイロン、またはニトロセルロースなどの様々な材料で作られ得る。支持体は剛性であり、平らな表面を有し得る。支持体は、約1から10,000,000の分解された遺伝子座を有し得る。例えば、支持体は、約10から10,000,000、約10から5,000,000、約100から5,000,000、約100から4,000,000、約1000から4,000,000、約1000から3,000,000、約10,000から3,000,000、約10,000から2,000,000、約100,000から2,000,000、または約100,000から1,000,000の分解された遺伝子座を有し得る。分解された遺伝子座の密度は、平方センチメートル内で少なくとも約10、約100、約1000、約10,000、約100,000、または約1,000,000の分解された遺伝子座であり得る。いくつかの場合では、分解された遺伝子座の各々は、単一のタイプのオリゴヌクレオチドの>95%まで占有され得る。他の場合では、分解された遺伝子座の各々は、プローブのプーリングされた混合物またはプローブのセットにより占有され得る。

更なる場合では、いくつかの分解された遺伝子座は、プローブのプーリングされた混合物またはプローブのセットにより占有され、他の分解された遺伝子座は単一のタイプのオリゴヌクレオチドの>95%により占有される。

【0363】

いくつかの場合では、アレイ上の所与のヌクレオチド配列のためのプローブの数は、そのようなアレイにハイブリダイズされるDNAサンプルに対し非常に過剰な場合がある。例えば、アレイは、約10、約100、約1000、約10,000、約100,000、約1,000,000、または約100,000,000倍の、投入サンプルにおけるDNAの量に対するプローブの数を有し得る。

【0364】

いくつかの場合では、アレイは、約10、約100、約1000、約10,000、約100,000、約1,000,000、約10,000,000、または約1,000,000,000のプローブを有し得る。

【0365】

プローブのアレイまたはプローブのセットは、支持体上で段階ごとの様式で合成され、または予め合成された形態で付着され得る。合成方法の1つは、高密度の小型化したアレイにおけるオリゴヌクレオチドプローブの合成に向けるための光の使用を伴う、VLSIPS（商標）（米国特許第5,143,854号およびEP476,014に記載される）である。合成サイクル数を減らすためのマスクのデザインのためのアルゴリズムは、米国特許第5,571,639号および米国特許第5,593,839号に記載される。アレイはまた、EP624,059に記載されるように、機械的に制限された流路により支持体の細胞にモノマーを送達することにより、組み合わせの様式で合成され得る。アレイはまた、インクジェットプリンターを使用して支持体上へと試薬を付ける（spotting）ことにより合成され得る（例えば、EP728,520を参照）。

【0366】

いくつかの実施形態では、本開示は、アレイ上へとリンクされたDNAセグメントをハイブリダイズする方法を提供する。「基質」または「アレイ」は、合成的または生合成的のいずれかで調製され、かつ様々な異なるフォーマット（例えば可溶性分子のライブラリ、および、樹脂ビーズ、シリカチップ、または他の固体支持体に拘束されたオリゴヌクレオチドのライブラリ）で生物活性についてスクリーンされ得る核酸の、故意に作り出された集まりである。加えて、用語「アレイ」は、基質上へと実質的に任意の長さ（例えば、1から約1000の長さのヌクレオチドモノマー）の核酸を付けることにより調製され得る核酸のそのようなライブラリを含んでいる。

【0367】

アレイの技術、および様々な関連技術、並びに適用は、一般的に多数のテキストブックおよび文書で記載されている。例えば、これらはLemieux et al., 1998, Molecular Breeding 4, 277-289、Schena and Davis, Parallel Analysis with Biological Chips. in PCR Methods Manual (eds. M. Innis, D. Gelfand, J. Sninsky)、Schena and Davis, 1999, Genes, Genomes and Chips. In DNA Microarrays: A Practical Approach (ed. M. Schena), Oxford University Press, Oxford, UK, 1999)、The Chipping Forecast (Nature Genetics special issue; January 1999 Supplement)、Mark Schena (Ed.), Microarray Biochip Technology, (Eaton Publishing Company); Cortes, 2000, The Scientist 14 [17]: 25、Gwynn and Page, Microarray analysis: the next revolution in molecular biology, Science, 1999 Aug. 6、および Eakins and Chu, 1999, Trends in Biotechnology, 17, 217-218を含む。

【0368】

通常、任意のライブラリは、ライブラリのメンバを空間的に分離することにより、アレイへと順序正しい方式で配置され得る。配置に適切なライブラリの例は、とりわけ、リガンドライブラリなどの任意の分子を含むライブラリと同様に、核酸ライブラリ（DNAおよびcDNA、オリゴヌクレオチドなどのライブラリを含む）、ペプチド、ポリペプチド、およびタンパク質のライブラリを含む。

【0369】

ライブラリは、メンバの拡散および混合を制限するために、固相（例えば固体の基質）上へと定着または固定され得る。いくつかの場合では、DNA結合リガンドのライブラリが調製され得る。特にライブラリは、膜、およびプラスチックやガラスなどの非多孔基材を含む、実質的に平らな固相に固定され得る。更に、ライブラリは、インデキシング（すなわち、特定のメンバへの言及またはアクセス）が容易になるような方法で配置され得る。いくつかの例では、ライブラリのメンバは、格子形態のスポットとして適用され得る。一般的なアッセイ系はこの目的に適し得る。例えば、アレイは、ウェルに複数のメンバを有するか、または各ウェルに単一のメンバのどちらかを有するマイクロプレートの表面上で固定され得る。更に、固体の基質は、ニトロセルロースまたはナイロンの膜（例えば、実験をプロットするのに使用される膜）などの膜でもよい。代替的な基質は、ガラス、またはシリカ基質を含む。故に、ライブラリは、当該技術分野で既知の適切な方法、例えば、電荷相互作用、または、ウェルの壁または底、もしくは膜の表面への化学結合により、固定され得る。配置および固定の他の手段は、例えば、ピペット、点滴薬タッチ、圧電手段、インクジェットおよびバブルジェット技術、静電気適用などを使用し得る。シリコンチップの場合、フォトリソグラフィーはチップ上のライブラリを配置し固定するために利用し得る。

10

【0370】

ライブラリは、固体の基質上に「付ける（spotted）」ことにより配置され、これは、手により、またはメンバを堆積するためにロボット工学を利用することにより行われてもよい。一般に、アレイはマクロアレイまたはマイクロアレイと記載されてもよく、違いはスポットの大きさである。マクロアレイは、約300ミクロン以上のスポットサイズを含有し、既存のゲルおよびプロットスキャナーにより容易に撮像され得る。マイクロアレイにおけるスポットサイズは、直径200ミクロン未満であり、これらアレイは通常、何千ものスポットを含有する。したがって、マイクロアレイは特定のロボット工学と撮像装置を必要とすることがあり、これは注文仕様の必要があることがある。計測手段はCortese, 2000, The Scientist 14 [11]: 26によってレビューで一般的に記載される。

20

【0371】

DNA分子の固定されたライブラリを産生するための技術は当該技術分野で記載されてきた。一般的に、大半の先行技術の方法は、例えば、固体の基質上の様々な別個の位置で配列の様々な並べ替えを構築するためにマスキング技術を使用して、どのように一本鎖核酸分子ライブラリを合成するかを記載している。米国特許第5,837,832号は、超大規模集積回路技術に基づいてシリコン基質に固定されたDNAアレイを産生するための改善された方法を記載している。特に、米国特許第5,837,832号は、本開示の固定されたDNAライブラリを産生するために使用され得る基質上の空間的に定められた場所でプローブの特定のセットを合成するために「タイリング」と呼ばれる戦略について記載する。米国特許第5,837,832号は、さらに使用されることがある初期の技術のための文献をもたず。他の場合では、アレイはまた、光析出化学（photo deposition chemistry）を使用して構築され得る。

30

【0372】

ペプチド（またはペプチド模倣体）のアレイも、アレイにおける別個の予め定められた場所で異なるライブラリのメンバ（例えば、特有のペプチド配列）をそれぞれ配する様式で、表面上に合成され得る。ライブラリのメンバそれぞれの同一性は、アレイにおけるその空間的な場所により判定される。予め定めた分子（例えば標的またはプローブ）と反応的なライブラリのメンバとの間の結合相互作用が生じる、アレイにおける場所が判定され、それにより、空間的な場所に基づき反応的なライブラリのメンバの配列の識別が行われる。方法は、米国特許第5,143,854号、WO90/15070とWO92/10092、Fodor et al. (1991) Science, 251: 767、Dower and Fodor (1991) Ann. Rep. Med. Chem.,

40

50

26: 271で記載される。

【0373】

検出を支援するために、標識は、任意の容易に検出可能なレポーター、例えば蛍光性、生物発光性、リン光性、放射性などのレポーターを使用することができる（上記で議論されるように）。そのようなレポーター、それらの検出、標的／プローブへの結合などは、本文書の他の場所で議論される。プローブおよび標的の標識化はまた、Shalton et al., 1996, Genome Res 6 (7): 639-45に開示されている。

【0374】

いくつかの市販で入手可能なマイクロアレイのフォーマットの例は、Marshall and Hodgson, 1998, Nature Biotechnology, 16 (1), 27-31に述べられている。

【0375】

アレイベースのアッセイからデータを生成するために、シグナルは、プローブとヌクレオチド配列との間のハイブリダイゼーションの存在または不在を示すために検出され得る。更に、直接および間接的な標識化技術も利用され得る。例えば、直接的な標識化は、アレイに関連したプローブへとハイブリダイズするヌクレオチド配列へ直接、蛍光染料を組み込む（例えば、染料は、標識化ヌクレオチドまたはPCRプライマーの存在下で酵素的合成によりヌクレオチド配列に組み込まれる）。直接的な標識化のスキームは、例えば同様の化学構造および特徴を有する蛍光染料のファミリーの使用により強固なハイブリダイゼーションシグナルをもたらす、かつ簡単に実施することができる。核酸の直接的な標識化を含む場合、シアニンまたはアレクサ (alex a) のアナログが、複数の蛍光比較的アレイ解析 (multiple-fluor comparative array analyses) に利用され得る。他の実施形態では、間接的な標識化のスキームは、マイクロアレイプローブへのハイブリダイゼーションの前または後に、核酸にエピトープを組み込むために利用され得る。1つ以上の染色の手順および試薬は、ハイブリダイズされた複合体（例えば、エピトープに結合し、それによりハイブリダイズされた種のエピトープへの色素分子の結合によって蛍光シグナルをもたらす、蛍光分子）を標識するために使用され得る。

【0376】

配列決定

様々な実施形態では、本明細書に記載されたまた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またはその際の配列決定されたヌクレオチドの検出、すなわち、リアルタイムまたはほぼリアルタイムでの配列の検出を可能にする。いくつかの場合では、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、または少なくとも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、または約1000の塩基であり得る。

【0377】

配列決定は特定の関心領域の濃縮の有無に関わらず全ゲノムであり得る。配列決定は、ゲノムの特定の部位を標的とし得る。濃縮または標的とされ得るゲノムの領域は、限定されないが、単一遺伝子（またはその領域）、遺伝子パネル、遺伝子融合、ヒト白血球抗原 (HLA) 遺伝子座（例えば、クラスI HLA-A、B、およびC、クラスII HLA-DRB1/3/4/5、HLA-DQA1、HLA-DQB1、HLA-DPA1、H

L A - D P B 1)、エキソン領域、エキソーム、および他の遺伝子座を含む。ゲノム領域は、免疫応答、免疫レパートリー、免疫細胞の多様性、転写（例えばエクソーム）、癌（例えば B R C A 1、B R C A 2、遺伝子または、ホットスポット領域などのその領域、体性の変異体、S N V、増幅、融合、腫瘍遺伝子変異量（T M B）、マイクロサテライト不安定性（M S I）、心疾患、遺伝病、および他の疾患または疾病に関連し得る。様々な方法は、限定されないが、配列キャプチャを含み、関心領域を濃縮または標的とするために使用され得る。いくつかの場合では、キャプチャ H i - C（C H i - C）プロトコルまたは C H i - C のようなプロトコルは、ライブラリ調製の前または後の配列キャプチャーステップ（例えば、標的濃縮アレイによって）を使用して、使用される。

【0378】

いくつかの実施形態では、ハイスループット配列決定は、H i S e q 2 5 0 0、H i S e q 1 5 0 0、H i S e q 2 0 0 0、または H i S e q 1 0 0 0 の機械を使用するものなど、I l l u m i n a ' s G e n o m e A n a l y z e r I I X、M i S e q パーソナルシーケンサー、または H i S e q 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。

【0379】

いくつかの実施形態では、ハイスループット配列決定は、A B I S o l i d システム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、ビーズにリンクした。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ョンに基づく。

【0380】

次世代の配列決定は、イオン半導体配列決定（例えば、L i f e T e c h n o l o g i e s（I o n T o r r e n t）の技術を使用する）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、p H 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（f l o o d e d）場合がある。走査、光、またはカメラは必要ではない。いくつかの場合では、I O N P R O T O N（商標）シーケンサーが核酸を配列するために使用される。いくつかの場合では、I O N P G M（商標）シーケンサーが使用される。I o n T o r r e n t P e r s o n a l G e n o m e M a c h i n e（P G M）。P G Mは2時間で1000万のリードを行うことができる。

【0381】

いくつかの実施形態では、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（S M S S）方法など、H e l i c o s B i o S c i e n c e s C o r p o r a t i o n（C a m b r i d g e，M a s s a c h u s e t t s）により利用可能な技術の使用を含む。S M S S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、S M S Sは、米国特許出願公開第20060024711号、第20060024678号、第20060012793号、第20060012784号、および第20050100932号に部分的に記載されている。

【0382】

いくつかの実施形態では、ハイスループット配列決定は、P i c o T i t e r P l a t e 装置などの454 L i f e s c i e n c e s，I n c.（B r a n f o r d，C o n n e c t i c u t）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C C D 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

10

20

30

40

50

アイバオプティックプレートを含む。このファイバー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。

【0383】

ファイバーオプティクスの検出後のビーズの増幅を使用する方法は、Marguile s, M., et al. "Genome sequencing in microfab ricated high-density pricolitre reactor s," Nature, doi:10.1038/nature03959、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、および第20060078909号に記載されている。

10

【0384】

いくつかの実施形態では、ハイスループット配列決定は、Clonal Single Molecule Array (Solexa, Inc.)、または可逆的なターミネ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、および米国特許出願公開第20040106110号、第20030064398号、第20030022207号、およびConstans, A., The Scientist 2003, 17(13):36に記載されている。

【0385】

次世代シーケンシング技術は、Pacific Biosciences.によってリアルタイム (SMRT (商標)) の技術を含み得る。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つに付着され得る。これらの染料はホスホリンク (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) の底部にある鋳型の一本鎖DNAの単一分子で固定され得る。ZMWは、(マイクロ秒で) ZMWの中およ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。20zeptolitre (20x10-21リットル) の検出限界を有する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。

20

30

【0386】

いくつかの場合では、次世代の配列決定はナノ細孔配列決定である (例えば、Sonig V 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、およ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のものであり得；例えばGridIONシステムであり得る。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、または1,

40

50

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(またはノード)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、または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、集積化したセンサーを有するナノ細孔であり得る(例えば、トンネル電極検出器、容量検出器またはグラフェン性ナノギャップもしくはエッジ状態検出器(例えば、Gara j e t a l. (2010) Nature vol. 67, doi: 10.1038/nature09379を参照))。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またはタンパク質)。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、システムは両方の鎖を読み取ることができる。いくつかの場合では、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、塩基を識別することができる。

【0387】

GENIA製のナノ細孔配列決定技術を使用し得る。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-膜アセンブリおよびチャネルを通るDNA運動の制御を可能にする。いくつかの場合では、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。プローブを有するゲノム断片は、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。いくつかの場合では、ナノ細孔配列決定技術は、IBM/Roche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を引き寄せるまた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。

【0388】

次世代配列決定は、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(Adl)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有する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

10

20

30

40

50

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、その後のステップ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。右およ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アダプター上で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。右および左のアダプターの第3巡（Ad3）を、線状DNAの右およ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。III型制限酵素（例えばEcoPI5）を加えることができる；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。右およ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）増幅しつ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。

【0389】

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、ハイブリダイズすることができるパンドローム配列を含有することができ、平均の直径でおよそ200-300ナノメートルであり得るDNAナノボール（DNB（商標））を形成するように、一本鎖がそれ自体の上に折りたたまれ得る。DNAナノボールは、マイクロアレイ（配列決定フローセル）に付着させることができる（例えば、吸着により）。フローセルは、二酸化ケイ素、チタン、およ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。調べ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。

【0390】

いくつかの実施形態では、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10×-50×増強することを可能にする。それらを使用するためのAnyDot.chipsおよび方法は、国際公開第WO 02088382号、第WO03020968号、第WO03031947号、第WO2005044836号、PCT/EP 05/05657、PCT/EP 05/05655；およびドイツ特許出願DE 101 49 786、DE 102 14 395、DE 103 56 837、DE 10 2004 009 704、DE 10 2004 025 696、DE 10 2004 025 746、DE 10 2004 025 694、DE 10 2004 025 695、DE 10 2004 025 744、DE 10 2004 025 745、および DE 10 2005 012 301に部分的に記載されている。

【0391】

他のハイスループット配列決定システムは、Venter, J., et al. Science 16 February 2001、Adams, M. et al.

10

20

30

40

50

Science 24 March 2000、および M. J. Levene, et al. Science 299:682-686, January 2003と米国特許出願公開第20030044781号および第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、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み、すなわち、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。次いで配列の塩基付加の各ステップで、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識別することにより、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。重合ステップ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、識別される。標識されたヌクレオチドアナログを提供するステップと、成長核酸鎖を重合させるステップと、付加されたヌクレオチドアナログを識別するステップは繰り返され、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。

【0392】

キット

特定の実施形態では、本開示は、本開示の1つ以上の成分を含むキットを更に提供する。キットは、上述のものを含めて、当業者に明らかな任意の用途に使用することができる。キットは、例えば、複数の会合分子、固定剤、ヌクレアーゼ、リガーゼおよび／またはそれらの組み合わせを含み得る。いくつかの場合では、会合分子は、例えばヒストンを含むタンパク質であり得る。いくつかの場合では、固定剤は、ホルムアルデヒド、またはDSG、EGS、またはDSSを含む他の任意のDNA架橋材であり得る。

【0393】

いくつかの場合では、キットは複数のビーズをさらに含むことができる。ビーズは、常磁性でありおよび／または捕捉剤でコーティングすることができる。例えば、ビーズは、ストレプトアビジンおよび／または抗体でコーティングすることができる。

【0394】

いくつかの場合では、キットは、アダプターオリゴヌクレオチドおよび／または配列決定プライマーを含むことができる。更に、キットは、アダプターオリゴヌクレオチドおよび／または配列決定プライマーを使用してリードペアを増幅することが可能な装置を含むことができる。

【0395】

いくつかの場合では、キットは、溶解緩衝液、ライゲーション試薬（例えばdNTP、ポリメラーゼ、ポリヌクレオチドキナーゼおよび／またはリガーゼ緩衝液など）、およびPCR試薬（例えばdNTP、ポリメラーゼおよび／またはPCR緩衝液など）を含むが、これらに限定されない他の試薬を含むこともできる。

【0396】

キットは、キットの成分を使用するためのおよび／またはリードペアを生成するための命令を含むこともできる。

【0397】

コンピュータおよびシステム

図3に例示されるコンピュータシステム(500)は、媒体(511)および／またはネットワークポート(505)から命令を読み出すことが可能な、論理的な装置として理解され、固定された媒体(512)を有するサーバー(509)に随意に接続され得る。図3に示されるものなどのシステムは、CPU(501)、ディスクドライブ(503)

、キーボード（５１５）および／またはマウス（５１６）などの随意の入力装置、および随意のモニター（５０７）を含み得る。データ通信は指示された通信媒体を介して局所位置または遠隔位置のサーバーまで達成され得る。通信媒体は、データを送信および／または受信する任意の手段を含むことができる。例えば、通信媒体は、ネットワーク接続、無線接続、またはインターネット接続であってもよい。そのような接続は、ワールドワイドウェブ上での通信を提供することができる。本開示に関するデータは、図３に例示されるように当事者（５２２）による受信および／またはレビューのためにそのようなネットワークまたは接続によって伝達され得ることが想定される。

【０３９８】

図４は、本開示の例となる実施形態に関連して使用され得るコンピュータシステム（１００）の第１の例となるアーキテクチャを例示するブロック図である。図４に表されるように、例示的コンピュータシステムは、命令を処理するためのプロセッサ（１０２）を含み得る。プロセッサの非限定的な例は、Intel Xeon（商標）プロセッサ、AMD Opteron（商標）プロセッサ、Samsung 32-bit RISC ARM 1176JZ（F）-S v1.0（商標）プロセッサ、ARM Cortex-A8 Samsung S5PC100（商標）プロセッサ、ARM Cortex-A8 Apple A4（商標）プロセッサ、Marvell PXA 930（商標）プロセッサ、または同等の機能をもつプロセッサを含む。実行の複数のスレッドが並列処理に使用可能である。いくつかの実施形態では、単一コンピュータシステム、クラスターであろうと、または複数のコンピュータ、携帯電話、および／またはパーソナルデータアシスタントデバイスを含むネットワーク上でシステムにわたって分配されようと、複数のコアを備える複数のプロセッサが使用され得る。

【０３９９】

図４に例示されるように、高速キャッシュ（１０４）は、プロセッサ（１０２）に接続するか、または組み込まれることで、プロセッサ（１０２）により近年使用されてきたまたは頻繁に使用されている命令またはデータのための高速メモリを提供することができる。プロセッサ（１０２）は、プロセッサバス（１０８）によりノースブリッジ（１０６）に接続される。ノースブリッジ（１０６）は、メモリバス（１１２）によりランダムアクセスメモリ（RAM）（１１０）に接続され、プロセッサ（１０２）によりRAM（１１０）へのアクセスを管理する。ノースブリッジ（１０６）はまた、チップセットバス（１１６）によりサウスブリッジ（１１４）に接続される。サウスブリッジ（１１４）は次に、ペリフェラルバス（１１８）に接続される。ペリフェラルバスは、例えば、PCI、PCI-X、PCI Express、または他のペリフェラルバスであってもよい。ノースブリッジおよびサウスブリッジは頻繁に、プロセッサチップセットと称され、ペリフェラルバス（１１８）上でプロセッサと、RAMと、周辺コンポーネントとの間のデータ転送を管理する。いくつかの代替的なアーキテクチャでは、ノースブリッジの機能性は、別のノースブリッジチップを使用する代わりにプロセッサに組み込まれることができる。

【０４００】

いくつかの実施形態では、システム（１００）は、ペリフェラルバス（１１８）に接続されるアクセラレータカード（１２２）を含み得る。アクセラレータは、フィールドプログラマブルゲートアレイ（FPGA）または特定の処理を促進するための他のハードウェアを含んでもよい。例えば、アクセラレータは、適応データの再構成のために、または拡張された設定処理に使用される代数式を評価するために使用され得る。

【０４０１】

ソフトウェアとデータは、外部記憶装置（１２４）に記憶され、プロセッサにより使用されるRAM（１１０）および／またはキャッシュ（１０４）へとロードすることができる。システム（１００）は、システムリソースの管理のためのオペレーティングシステムを含み、オペレーティングシステムの非限定的な例は、Linux、Windows（商標）、MACOS（商標）、BlackBerry OS（商標）、iOS（商標）、および他の機能的に同等なオペレーティングシステムの他に、本開示の例となる実施形態に

10

20

30

40

50

従ってデータの記憶および最適化を管理するためのオペレーティングシステム上で実行するアプリケーションソフトウェアを含む。

【0402】

この例では、システム（100）はまた、ネットワーク接続ストレージ（NAS）などの外部記憶装置、および分散並列処理に使用され得る他のコンピュータシステムに対するネットワークインターフェースを提供するために、ペリフェラルバスに接続されるネットワークインターフェースカード（NIC）（120）および（121）を含む。

【0403】

図5は、複数のコンピュータシステム（202a）および（202b）、複数の携帯電話および個人用携帯情報端末（202c）、並びにネットワーク接続ストレージ（NAS）（204a）および（204b）を含むネットワーク（200）を示す略図である。実施形態の例では、システム（202a）、（202b）、および（202c）は、データ記憶を管理し、ネットワーク接続ストレージ（NAS）（204a）および（204b）に記憶されたデータに対するデータアクセスを最適化することができる。数学モデルはこのデータに対して使用され、コンピュータシステム（202a）および（202b）、並びに携帯電話および個人用携帯情報端末システム（202c）にわたって分散並列処理を使用して評価され得る。コンピュータシステム（202a）および（202b）、並びに携帯電話および個人用携帯情報端末システム（202c）はまた、ネットワーク接続ストレージ（NAS）（204a）および（204b）に記憶されたデータの適応データ再構築に対して並列処理を提供することができる。図5は一例のみを例示しており、様々な他のコンピュータのアーキテクチャおよびシステムは、本開示の様々な実施形態と共に使用され得る。例えば、ブレードサーバーは並列処理を提供するために使用することができる。プロセッサブレードは、並列処理を提供するためにバックプレーンを介して接続可能である。ストレージも、バックプレーンに接続することができるか、または別のネットワークインターフェースを介してネットワーク接続ストレージ（NAS）として接続可能である。

【0404】

いくつかの実施形態の例では、プロセッサは、別のメモリ空間を維持することができ、ネットワークインターフェース、バックプレーン、または他のプロセッサによる並列処理のための他のコネクタを介してデータを送信することができる。他の実施形態では、プロセッサのいくつかまたはすべては、共有の仮想アドレスメモリ空間を使用することができる。

【0405】

図6は、例となる実施形態による共有の仮想アドレスメモリ空間を使用して、マルチプロセッサコンピュータシステム（300）のブロック図である。システムは、共有メモリサブシステム（304）にアクセス可能な複数のプロセッサ（302a-f）を含む。システムは、メモリサブシステム（304）に複数のプログラマブルハードウェアのメモリアルゴリズムプロセッサ（MAP）（306a-f）を組み込む。MAP（306a-f）は各々、メモリ（308a-f）および1つ以上のフィールドプログラマブルゲートアレイ（FPGA）（310a-f）を含み得る。MAPは構成可能な機能ユニットを提供し、特定のアルゴリズムまたはアルゴリズムの一部は、各プロセッサと密接に協働して処理を行うためにFPGA（310a-f）に提供され得る。例えば、MAPは、データモデルに関する代数式を評価するおよび例となる実施形態において適応データの再構成を実行するために使用され得る。この例では、各MAPは、このような目的のためにプロセッサすべてによって世界中からアクセス可能である。1つの構成において、MAPは各々、関連するメモリ（308a-f）にアクセスするためにダイレクトメモリアccess（DMA）を使用することができ、それにより、各マイクロプロセッサ（302a-f）とは別個に、かつこれらから非同期的にタスクを実行することが可能となる。この構成では、MAPは、アルゴリズムのパイプライン処理（pipelining）および並列実行のために別のMAPに直接結果を供給することができる。

【0406】

上記のコンピュータアーキテクチャおよびシステムは、単に例であり、汎用のプロセッサ、コプロセッサ、FPGAおよび他のプログラマブルロジックデバイス、システムオンチップ（SOC）、特定用途向け集積回路（ASIC）、並びに他の処理素子および論理素子のあらゆる組み合わせを使用するシステムを含む、多種多様な他のコンピュータ、携帯電話、パーソナルデータアシスタントのアーキテクチャおよびシステムが、例となる実施形態に関連して使用され得る。いくつかの実施形態では、コンピュータシステムのすべてまたは一部は、ソフトウェアまたはハードウェアにおいて実施され得る。ランダムアクセスメモリ、ハードドライブ、フラッシュメモリ、テープドライブ、ディスクアレイ、ネットワーク接続ストレージ（NAS）、他のローカルまたは配信されたデータストレージデバイスおよびシステムを含む、あらゆる種類のデータストレージ媒体も、例となる実施形態に関連して使用され得る。

10

【0407】

例となる実施形態では、コンピュータシステムは、上記または他のコンピュータアーキテクチャおよびシステムのいずれかで実行するソフトウェアモジュールを使用して実施され得る。他の実施形態では、システムの機能は、ファームウェア、図11で言及されるようなフィールドプログラマブルゲートアレイ（FPGA）などのプログラム可能論理回路、システムオンチップ（SOC）、特定用途向け集積回路（ASIC）、または他の処理要素および論理素子において部分的または完全に実施され得る。例えば、セットプロセッサおよびオブティマイザは、図4に例示されるアクセラレータカード（122）などのハードウェアアクセラレータカードの使用によるハードウェアアクセラレーションで実施され得る。

20

【0408】

定義

他に定義されない限り、本明細書で用いる全ての技術的および科学的用語は、本開示の属する技術分野における当業者によって一般に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または同等の方法および試薬が、開示された方法およ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および材料が記載されている。

【0409】

本明細書と添付の請求項において使用されるように、単数形「1つ（a）」、「1つ（an）」、および「その（the）」は、文脈が明確に別のことを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およびその同等物の物理レイアウトをプローブする1つ以上の方法に対する言及を含んでいる。

30

【0410】

また、「および」の使用は、特に明記しない限り、「および／または」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）」、および「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

【0411】

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、いくつかの特定の例において、言語「～から実質的に成る」または「～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。

【0412】

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の断片を指す。

【0413】

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

50

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、および／または、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。

【0414】

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

【0415】

用語「リードペア (read pair)」または「リードペア (read-pair)」は、本明細書で使用されるように、配列情報を提供するためにリンクされる2つ以上の要素を指す場合がある。いくつかの場合では、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。他の場合では、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。

10

【0416】

用語「約」は、本明細書で使用されるように、別段の定めがない限り、数字をその数字の10%プラスまたはマイナスされた数字を含む値の範囲として、記載する場合がある。

【0417】

本明細書で使用されるように、「核酸の暴露された内部末端」は、フェーズまたはローカルな3次元構造的な情報を促進するために、末端に隣接した核酸配列情報にアクセスするように導入されたものなど、安定した核酸または安定していない核酸へ導入された切断部位の生成を通じて生成された暴露した末端を指す場合がある。

20

【0418】

本明細書で使用されるように、用語「約」の番号は、その数字の+/-10%に及ぶ範囲を指すが、「約」の範囲は、記述された範囲制限より10%低い範囲から10%大きい範囲までを指す。

【0419】

本明細書で使用されるように、リンカーまたは別の配列セグメントは、その配列の特定によって特定の第1の区画または由来の細胞に割り当てて、第2の区画または由来の細胞を除外することが容易になる場合、区画指定、または細胞指定である。特徴的な配列は、いくつかの場合では区画または細胞に特有であり、その結果、それは他の全ての細胞とは区別され、これが技術的に実行可能である場合、特有のタグは下流の解析を容易にする。しかしながら、特有の配列はすべての場合に必要とされるわけではない。いくつかの場合では、冗長なバーコーディングは計算的に下流で解決され、その結果、特有でないタグがマッピングされるが、それにもかかわらず、第2の区画または細胞の核酸と第1の区画または細胞の核酸とを区別するのに十分であるようになる。

30

【0420】

本明細書に使用されるように、クラスターは、複数の異なる末端隣接配列または配列タグがマッピングするものに関する核酸の領域である。いくつかの場合では、第2の領域に対する1つの領域の近接は、第2のクラスターのクラスター構成要素とペアになった末端リードで同時に生じる第1のクラスターのクラスター構成要素の数を数えることにより、少なくとも部分的に評価される。

40

【0421】

実施例

以下の実施例は、本発明の様々な実施形態を例示する目的で与えられ、いかなる様式でも本発明を制限するようには意図されていない。本明細書に記載される方法とともに、本実施例は、好ましい実施形態を代表するものであり、例示的なものであり、および、本発明の範囲を限定することを意図したものではない。請求項の範囲によって定義される本発明の精神内に包含されるその変化および他の使用が、当業者に想定される。

【0422】

実施例1：サンプル調製

50

サンプル型に依存するサンプル調製のための2つの別個のプロトコル：細胞または組織がある。

溶解物定量化ステップは両方のサンプル型に対して同じである。サンプル調製には2時間かかる。

【0423】

注釈：10X HiC洗浄緩衝液、10X架橋反転緩衝液、および20%のSDSは保管中に沈降した可能性がある。沈降物が見えなくなるまで、溶液を37℃で15分間インキュベートする。ボルテックスをし、使用する前に攪拌する。超純水を用いて10X HiC洗浄緩衝液を1Xまで薄くする。室温で保管する。約15mLの1X HiC洗浄緩衝液は1サンプル当たり必要である。1X HiC洗浄緩衝液も、プロトコルの残り全体にわたって使用することができる。1X HiC洗浄緩衝液は、室温で2ヶ月間安定する。超純水を用いて10X架橋反転緩衝液を1Xまで薄くする。室温で保管する。約1mLの1X架橋反転緩衝液は1サンプル当たり必要である。1X架橋反転緩衝液も近接ライゲーション・プロトコルに使用することができる。1X架橋反転緩衝液は室温で2ヶ月間安定する。攪拌サーマルミキサーを、1.5mL管の場合、1250rpmに設定しなければならない。氷上の緩衝液を解凍することと使用前にボルテックスすることを含む、優良試験所規範を使用する。

【0424】

細胞のためのプロトコル

注釈：洗浄中の損失を考慮して出発材料として 10×10^6 の細胞を使用することが、推奨される。 10×10^6 未満の細胞が利用可能な場合、低い投入量プロトコルを指す。開始前に、新鮮な1Xヌクレアーゼ消化緩衝液を調製し、室温で保管する。1Xヌクレアーゼ消化緩衝液は室温で1日安定する。1Xヌクレアーゼ消化緩衝液を調製するためには、140μLの超純水；20μLの10Xヌクレアーゼ消化緩衝液；20μLの100mM MnCl₂；20μLの10%のトリトンを攪拌する。

【0425】

細胞を収穫し、1XPBSで洗浄する。細胞を数えて 10×10^6 の細胞を等分し、2000xgで5分間遠心分離する。上澄みを注意深く除去する。ペレットを5mLの1XPBSと135μLの37%のホルムアルデヒドで再懸濁する。サンプルを5mL管に転送し、細胞が定着しないように、あるスピードで、室温で10分間回転させる。管を2000xgで5分間遠心分離する。細胞ペレットが緩み得るため、上澄みを注意深く慎重に除去する。ペレットをHiC洗浄緩衝液で洗浄し、初めに塊を解体するために200μLを用いて、それから残りの4.8mLを加え、ペレットを完全に再懸濁するために上下にピペットする。管を2000xgで5分間スピンスさせ上澄みを注意深く除去する。洗浄ステップを繰り返して合計2回洗浄する。第2の洗浄剤を除去した後に、ペレットを1mLの1XHiC洗浄緩衝液中で再懸濁し、ペレットを再懸濁する。細胞を数えて、 1×10^6 の細胞を3つの別個の管に加え、残りの細胞を-80℃で凍結ペレットとして保管する。3つの管を2000xgで5分間遠心分離して、上澄みを除去する。各管中のペレットを50μLの1Xヌクレアーゼ消化緩衝液（新鮮に調製された）で再懸濁する。管を1250rpmにある攪拌サーマルミキサーで2分間30℃に予め温める。7.5μLのヌクレアーゼ酵素混合物を有する新鮮な1.5mL管を、1250rpmにある攪拌サーマルミキサーで2分間30℃で予め温める。予め温められたヌクレアーゼ酵素混合物を、次のように、0.5μLを第1の管に、2.0μLを第2の管に、および4.0μLを第3の管に転送する。管を1250rpmにある攪拌サーマルミキサーでちょうど30分間30℃でインキュベートする。ヌクレアーゼ反応を5μLの0.5MEDTAを加えて混ぜることにより停止させる。3μLの20%のSDSを、細胞を溶解するために加えて、細胞を1250rpmにある攪拌サーマルミキサーで5分間30℃でインキュベートする。

【0426】

組織のためのプロトコル

注釈：60 mgの組織を使用することが推奨される。60 mg未満の組織が利用可能な場合、低い入力プロトコルを指す。開始前に、新鮮な1 Xヌクレアーゼ消化緩衝液を調製して室温で保管する。1 Xヌクレアーゼ消化緩衝液は室温で1日安定する。1 Xヌクレアーゼ消化緩衝液を調製するためには、140 μ lの超純水；20 μ lの10 Xヌクレアーゼ消化緩衝液；20 μ lの100 mM MnCl₂；20 μ lの10%のトリトンを撹拌する。

【0427】

少なくとも60 mgの凍結組織を量り分けて、図1 Aと図1 Bに例示された稠度まで液体窒素中で乳鉢と乳棒を用いて微粉に砕かれ、図1 Aは不十分な組織粉碎を示し、図1 Bは十分な組織粉碎を示す。破壊された組織を、5 mlの1 X PBSと135 μ lの37%のホルムアルデヒドを用いて5 ml管に転送する。管を室温で10分間回転させる。管を2000 x gで5分間遠心分離して上澄みを注意深く除去する。組織がペレットしない事象で、管を5分間最高速度でスピンのさせる。ペレットを200 μ lの洗浄緩衝液上で再懸濁し、その後、4.8 mlの1 X HiC洗浄緩衝液を加える。管を2000 x gで5分間遠心分離して上澄みを除去する。洗浄ステップは2回実施され、最終ペレットを1 mlの1 X HiC洗浄緩衝液中で再懸濁する。再懸濁した細胞を200 μ mフィルターを通して新鮮な5 ml管へと通過させ、必要ならばフィルターを交換する。さらに2 mlの1 X HiC洗浄緩衝液を200 μ mフィルターを通して通過させる。サンプルを3つに分けて、20 mgの組織に対応する各一定分量を備えた3つの別個の管に1 mlずつ等分する。過剰な組織は造粒され、-80 °Cで保管され得る。3つの管を2000 x gで5分間遠心分離して上澄みを除去する。各管中のペレットを50 μ lの1 Xヌクレアーゼ消化緩衝液（新鮮に調製された）で再懸濁させる。管を1250 rpmにある撹拌サーマルミキサーで2分間30 °Cに予め温める。7.5 μ lのヌクレアーゼ酵素混合物を有する新鮮な1.5 ml管を1250 rpmにある撹拌サーマルミキサーで2分間30 °Cで予め温める。予め温められたヌクレアーゼ酵素混合物を、次のように、0.5 μ lを第1の管に、2.0 μ lを第2の管に、および4.0 μ lを第3の管に転送する。管を1250 rpmにある撹拌サーマルミキサーでちょうど30分間30 °Cでインキュベートする。ヌクレアーゼ反応を5 μ lの0.5 M EDTAを加えて混合することにより停止させる。3 μ lの20%のSDSを細胞を溶解するために加え、細胞を1250 rpmにある撹拌サーマルミキサーで5分間30 °Cでインキュベートする。

【0428】

実施例2：溶解物定量化

注釈：溶解物定量化は2時間かかる。80%のエタノールを、SPRIselectビーズを使用して、DNA精製のために新鮮に調製する。定量化ステップは2つの目的を有する：近接ライゲーションステップで使用するためにサンプルの量を決定すること、およびサンプル調製で得られた3つの管のどれを近接ライゲーションで使用するかを決定すること。

【0429】

各溶解物を18 μ lの1 X HiC洗浄緩衝液を備えた各管からの2 μ lの溶解物を混ぜることにより1：10に薄める。薄められた溶解物を-80 °Cで保管する。各溶解物の2.5 μ lを、50 μ lの架橋反転緩衝液および1.5 μ lのプロテイナーゼKを有する管に転送する。混合物をピペットによって混合し、1250 rpmにある撹拌サーマルミキサーで55 °Cで15分間インキュベートした後、68 °Cで45分間インキュベートする。100 μ lのSPRIselectビーズを各管に加え、懸濁させるためにボルテックスし、沈降させて磁石から遠ざけて室温で5分間インキュベートする。管を5分間または溶液が透明になるまで磁石に置き、ビーズを完全に分離した。上澄みを除去し、磁石から管を取り除くことなく、ビーズを1分間80%のエタノールを用いて2回洗浄する。2回目の洗浄後、管を遠心沈降にかけ、1分間磁石に置く。残りのエタノールをピペットを用いて除去する。ビーズを、エタノールがなくなるまで、しかし過度に乾燥させることなく、磁石上で5分間風乾させる。管を磁石から取り外し、ビーズをpH 8.0で1

0 μ l の T E 緩衝液で再懸濁させる。管をボルテックスし、遠心沈降にかけ、磁石に 1 分間置く。8 μ l の上澄みを新鮮な管に転送する。サンプルをキュービット蛍光光度計とキュービット dsDNA HS キットを使用して定量化する。濃度をスプレッドシートに記録する。断片サイズ分布を Tape Station D5000 または D5000 HS ScreenTape を使用して決定する。D5000 HS ScreenTape を使用する場合、サンプルを 1 ng / μ l に薄めなければならない。領域を Tape Station システムで解析する：領域 1 100 ~ 2500 bp；領域 2 100 ~ 600 bp、および領域 3 600 ~ 2500 bp。合計のパーセントを計算し記録する。データは、図 2 に示されるように、合計が 100 bp から 2500 bp については 76.33%、合計が 100 bp から 600 bp については 28.82%、および合計が 600 bp から 2500 bp については 47.82% となっている。

10

【0430】

1000 ng に対応するサンプルの量を計算する。クロマチン消化効率 (CDE) およびクロマチン消化インデックス (CI) を計算する。どのサンプルが QC メトリックを通過するかを決定する。

【0431】

実施例 3：近接ライゲーション

注釈：近接ライゲーションには 5.5 時間かかる。攪拌サーマルミキサーを 1.5 ml 管に対して 1250 rpm に設定する。磁石にサンプルを置く場合、上澄みを除去する前に溶液が完全に透明になるまで時間を置く。新鮮な架橋ライゲーション混合物 (50 μ l) を調製し使用前に氷上で保管する。50 μ l の架橋ライゲーション混合物を調製するために、10 μ l の 5X ライゲーション緩衝液、5 μ l 架橋、および 35 μ l の超純水を混合する。80% のエタノールを、SPRI select ビーズを用いる DNA 精製のために新鮮に調製する。

20

【0432】

クロマチンのクロマチンキャプチャビーズへの結合

クロマチンキャプチャビーズを室温にすることを可能にし、使用前にボルテックスする。100 μ l のクロマチンキャプチャビーズを、計算されたように、例 1 の 1000 ng のサンプルを有する 1.5 ml 管に転送する。サンプルをピペットによって混合し、磁石から遠ざけて室温でインキュベートする。管を磁石に 5 分間または溶液が透明になりビーズが分離するまで置く。上澄みを除去する。管を磁石ラックから取り除き、ビーズを 150 μ l の 1X Hi-C 洗浄緩衝液で洗浄し、ピペットによって 10 回混合し、上澄みを除去する前に 1 分間磁石に置く。洗浄ステップを 1 回反復する。

30

【0433】

末端研磨

管を磁石ラックから取り除き、50 μ l の末端研磨緩衝液と 3.5 μ l の末端研磨酵素混合物を加える。サンプルをピペットによって混合し、1250 rpm にある攪拌サーマルミキサーで 30 分間 22 °C でインキュベートし、その後 65 °C で 30 分間インキュベートする。管が室温に到達することを可能にし、その後 1 分間または溶液が透明およびビーズが分離するまで磁石ラックに置く。上澄みを除去する。管を磁石ラックから取り除き、ビーズを 150 μ l の 1X Hi-C 洗浄緩衝液で一度洗浄し、ピペットによって混合する。その後、管を磁石ラックに 1 分間戻し、上澄みを除去する。

40

【0434】

架橋ライゲーション

管を磁石ラックから取り除き、50 μ l の架橋ライゲーション混合物に 1 μ l の T4 DNA リガーゼを加える (新鮮に作られた)。サンプルをピペットによって混合し、22 °C で 30 分間インキュベートする。管を磁石ラックに 1 分間または溶液が透明になるまで置く。上澄みを除去し、管を磁石ラックから取り除く。ビーズを 150 μ l の 1X Hi-C 洗浄緩衝液を用いて再懸濁し、ピペットによって混合する。管を 1 分間磁石に置いて上澄みを除去する。

50

【0435】

集合体内ライゲーション

管を磁石ラックから取り除き、50 μ lの集合体内ライゲーション緩衝液と2 μ lの集合体内ライゲーション酵素混合物をビーズに加える。サンプルをピペットによって混合し、攪拌サーマルミキサーで22 $^{\circ}$ Cで1時間インキュベートする。管を磁石ラックに1分間または溶液が透明になりビーズが分離するまで置く。上澄みを除去する。

【0436】

架橋反転

管を磁石ラックから取り除き、50 μ lプラス1.5 μ lのプロテイナーゼKをビーズに加える。サンプルをピペットによって混合し、攪拌サーマルミキサーで55 $^{\circ}$ Cで15分間インキュベートし、その後68 $^{\circ}$ Cで45分間インキュベートする。S

【0437】

PRISelectビーズ上のDNA精製

SPRISelectビーズを再懸濁するために30秒間ボルテックスする。35 μ lの再懸濁されたビーズを1.5mlのサンプル管に加える。サンプルをボルテックスし、磁石から離れて5分間室温でインキュベートして遠心沈降させる。管を磁石に5分間または溶液が透明になりビーズが分離するまで置く。上澄みを除去する。150 μ lの80%のエタノールの2回の洗浄のために、管を磁石上に残す。ビーズをこれら洗浄剤で再懸濁させ、エタノールを加え、1分間インキュベートしてエタノールを除去する。2回目の洗浄後、管を遠心沈降させ、磁石に1分間置き、ピペットを使用してエタノールの最後を除去する。エタノールがなくなるまで、しかしビーズを過度に乾燥させることなく、ビーズを磁石上で5分間風乾させる。サンプルを磁石から取り外し、52 μ lのTE緩衝液 pH 8.0を加える。サンプルをボルテックスし、遠心沈降させて磁石から外し、室温で5分間インキュベートする。管を遠心沈降させ、磁石に1分間置く。50 μ lの上澄みを新鮮な1.5 ml管に転送する。ビーズを破棄する。サンプルをキュービット蛍光光度計とキュービットdsDNA HSキットを使用して定量化する。200 ngはライブラリ調製ステップに進むために必要である。精製DNAを最大6か月間20 $^{\circ}$ Cで保管する。

【0438】

実施例4：ライブラリ調製

注釈：ライブラリ調製プロトコルは、断片化を必要とせず、約2時間かかる。

【0439】

末端の修復

注釈：末端修復緩衝液は、保存中に沈降することがあり、目に見える沈降物がなくなるまで37 $^{\circ}$ Cで少なくとも10分間インキュベートする必要がある。250 mMのDTTを、使用の前に完全に混合するために分注することによって混合する。

【0440】

0.2 mlのPCR管を、48 μ lの精製されたサンプル、7 μ lの末端修復緩衝液、3 μ lの末端修復酵素ミックス、および0.5 μ lの250 mMのDTTで調製する。その混合物を分注によって混合し、遠沈する。サンプルを、20 $^{\circ}$ Cで30分間インキュベートし、その後サーマルサイクラー中で65 $^{\circ}$ Cで30分間続く。サンプルを12 $^{\circ}$ Cで保持する。

【0441】

アダプターライゲーションおよびUSERの消化

サンプルを含有している0.2 mlのPCR管を、Illuminaのための2.5 μ lのアダプター、1 μ lのライゲーション・エンハンサー、および30 μ lのライゲーション酵素ミックスと混合する。サンプルを分注によって混合し、および管を遠沈する。サンプルをサーマルサイクラーで20 $^{\circ}$ Cで15分間インキュベートし、そして12 $^{\circ}$ Cで保持する。3 μ lのUSER酵素ミックスをPCR管に添加する。サンプルを分注によって混合し、遠沈する。サンプルをサーマルサイクラーで37 $^{\circ}$ Cで15分間インキュベートし、そして12 $^{\circ}$ Cで保持する。

【0442】

DNA精製

SPRIselectビーズを、再懸濁するために30秒間ボルテックスする。80 μ lの再懸濁されたビーズを、PCR管に添加する。サンプルを、ボルテックスし、磁石から離れて室温で5分間インキュベートして遠沈する。管を、5分間、あるいは溶液が透明に見える、かつビーズは離間されるまで磁石の上に置く。上澄みを除去する。管を、150 μ lの80%エタノールの2つの洗浄のために磁石の上に残す。ビーズをこれらの洗浄において再懸濁せずに、エタノールを添加し、1分間インキュベートし、エタノールを除去する。2回目の洗浄後に、管を遠沈し、1分間磁石の上に置き、ピペットを使用してエタノールの最後を除去する。エタノールが残らないまで、ビーズを磁石の上で5分間風乾するが、ビーズを乾燥し過ぎない。サンプルを磁石から取り除き、100 μ lのTE緩衝液pH8.0を添加する。サンプルをボルテックスし、遠沈し、磁石から離れて室温で5分間インキュベートする。管を遠沈し、磁石の上に1分間置く。95 μ lの上澄みを新鮮な1.5mlの管に転送する。ビーズを廃棄する。精製されたDNAを、20°Cで夜通し保存する。

10

【0443】

実施例5：ライゲーション・キャプチャーおよび増幅

【0444】

注釈：ライゲーション・キャプチャーおよび増幅プロトコルは、2時間かかる。

【0445】

20

ストレプトアビジンビーズの調製

注釈：このステップは、いずれかのDNAサンプルを関与していない。

【0446】

ストレプトアビジンビーズを再懸濁するためにボルテックスする。25mlの再懸濁されたストレプトアビジンビーズを、1.5mlの管に転送する。ストレプトアビジンビーズを含有している管を、磁石の上に5分間置き、そして上澄みを除去する。管を磁気ラックから取り除き、そしてストレプトアビジンビーズを200 μ lのTWB（赤ラベル）で洗浄し、分注によって混合する。サンプルを、磁気ラックの上に1分間置き、上澄みを除去する。洗浄ステップを1回繰り返す。その後、ストレプトアビジンビーズを100 μ lの2X NTP（黄色ラベル）で再懸濁し、分注によって混合する。

30

【0447】

ライゲーション・キャプチャー

95 μ lの精製されたDNAを、100 μ lの2X NTPを再懸濁されたストレプトアビジンビーズを有する1.5mlの管に転送する。管を10秒間ボルテックスし、遠沈する。混合物を、攪拌サーマルミキサーで25°Cで30分間インキュベートする。

【0448】

ストレプトアビジンビーズ上の洗浄サンプル

注釈：各々の洗浄のために、管を、磁気ラックから取り除き、示された緩衝液をビーズに添加し、ビーズを再懸濁し、その後、管を磁石の上に1分間置き、および上澄みを、各洗浄の間に全ての上澄みを除去するように注意しながら、除去する。

40

【0449】

管を遠沈し、その後、磁石の上に1分間置き、および上澄みを除去する。ビーズを、200 μ lのLWBで1回洗浄する。ビーズを、200 μ lのNWBで2回洗浄する。ビーズを200 μ lの1X HiC洗浄緩衝液で2回洗浄する。

【0450】

PCR指標

注釈：すべてのPCR酵素およびマスターミックスはストレプトアビジンビーズの存在下で増幅に適合するとは限らないため、供給されるPCRレディミックスを使用する。

【0451】

最後の洗浄剤が除去された後、管を磁気ラックから取り除き、および25 μ lのHot

50

Start PCRレディミック、5 μ lのユニバーサルPCRプライマー、5 μ lの指標プライマー（各サンプルに特有のもの）、および15 μ lのDNaseとRNaseを含まない蒸留水をビーズに添加する。その混合物を分注によって混合し、0.2 mlのPCR管に転送する。管を遠沈し、以下のようなプログラムを実行するためにサーモサイクラーに置く：98 °Cで3分間、（98 °Cで20秒間、65 °Cで30秒間、72 °Cで30秒間）の12サイクル、72 °Cで1分間、12 °Cで保持。

【0452】

サイズ選択

PCR管を遠沈し、磁石の上に1分間置く。47 μ lの上澄みを、1.5 mlの管に転送し、およびビーズを廃棄する。全容積を100 μ lにするために、53 μ lのTE緩衝液pH 8.0を管に添加する。SPRIselectビーズを30秒間ボルテックスし、および45 μ lの再懸濁されたSPRIselectビーズを、サンプルを含有している1.5 mlの管に添加する。その混合物は再懸濁するためにボルテックスし、遠沈し、磁石から離して室温で10分間インキュベートする。管を遠沈し、磁石の上に5分間置く。145 μ lの上澄みを新規な1.5 mlの管に転送し、ビーズを廃棄する。35 μ lのSPRIselectビーズを、1.5 mlの管に添加し、再懸濁するためにボルテックスし、遠沈し、磁石から離れて室温で10分間インキュベートする。管を遠沈し、磁石の上に5分間置く。上澄みを除去する。管を磁石の上に残し、およびビーズを200 μ lの80%エタノールで2回洗浄する。ビーズを、これらの洗浄の間は再懸濁しない。管を遠沈し、磁石の上に1分間置く。10 μ lのピペット先端をエタノールの残留を除去するために使用する。残存するエタノールが残らないまで、ただし、乾燥しすぎないように、ビーズを磁石の上で5分間風乾する。ビーズを分注によって混合しながら、30 μ lのTE緩衝液pH 8.0において再懸濁する。管を遠沈し、磁石から離れて室温で2分間インキュベートする。管を遠沈し、磁石の上に1分間置く。28 μ lの上澄みを新しい1.5 mlの管に転送する。この管はライブラリを含有している。サイズ選択されたライブラリの量をキュービット蛍光測定器およびキュービットdsDNA HSキットを使用して計る。少なくとも60 ngのDNAが回復される。60 ng未満のDNAが回復される場合、ライブラリを廃棄する。TapeStationまたはBioanalyzerを、サイズ選択されたライブラリのサイズ分布を証明するために使用し、およびライブラリのサイズ範囲は、350 bpと1000 bpの間である。ライブラリを、-20 °Cで最大6ヶ月保存する。

【0453】

実施例6：低投入量サンプル調製

【0454】

勧められた投入量が利用可能でない場合、これを使用する。より低いサンプル投入量は、最終ライブラリのより低い複雑度をもたらす。

【0455】

細胞

利用可能な細胞の数を使用し、およびヌクレアーゼステップまで細胞用の実施例1の方法を使用する。ヌクレアーゼステップでは、0.1 μ lの事前に暖められたヌクレアーゼ酵素ミックスを第1の管に添加し、0.5 μ lの事前に暖められたヌクレアーゼ酵素ミックスを第2の管に添加し、そして2.0 μ lの事前に暖められたヌクレアーゼ酵素ミックスを第3の管に添加する。

【0456】

組織

利用可能な（少なくとも5 mg）組織の量を使用し、および組織用の実施例1の方法をヌクレアーゼステップまで使用する。ヌクレアーゼステップでは、0.1 μ lの事前に暖められたヌクレアーゼ酵素ミックスを第1の管に添加し、0.5 μ lの事前に暖められたヌクレアーゼ酵素ミックスを第2の管に添加し、そして2.0 μ lの事前に暖められたヌクレアーゼ酵素ミックスを第3の管に添加する。

【0457】

実施例7：指標プライマー

【0458】

以下の指標プライマーを使用する：

【0459】

【表1】

表1：指標プライマー	
指標プライマー	配列
指標プライマー2	CGATGT
指標プライマー4	TGACCA
指標プライマー5	ACAGTG
指標プライマー6	GCCAAT
指標プライマー7	CAGATC
指標プライマー8	ACTTGA
指標プライマー12	CTTGTA
指標プライマー19	GTGAAA

10

20

30

【0460】

指標プライマーを以下のスキームに応じて選択する：

【0461】

40

50

【表 2】

表 2：指標プライマー選択	
ライブラリの数	指標プライマー組み合わせ
2	6 および 1 2
2	5 および 1 9
3	2、7、および 1 9
3	10 の 2 乗オプションのいずれかと他の任意の指標プライマー
4	5、6、1 2、および 1 9
4	10 の 3 乗オプションのいずれかと他の任意の指標プライマー

【0 4 6 2】

実施例 8：サンプル調製－M N a s e

【0 4 6 3】

サンプルの種類、細胞または組織に応じて、サンプル調製には 2 つの別々のプロトコルがある。溶解物定量化ステップは、両方のサンプルの種類には同じである。サンプル調製は 2 時間かかる。

【0 4 6 4】

注釈：1 0 X H i C 洗浄緩衝液、1 0 X 架橋反転緩衝液、および 2 0 % の S D S は、保存場中に沈降しているかもしれない。溶液を、沈降物が見えなくなるまで、3 7 °C で 1 5 分間インキュベートする。使用の前に混合するためにボルテックスする。1 0 X H i C 洗浄緩衝液を 1 X まで U l t r a P u r e 水で希釈する。室温で保存する。1 つのサンプルあたりに約 1 5 m L の 1 X H i C 洗浄緩衝液が必要である。1 X H i C 洗浄緩衝液は、プロトコルの残り全体にわたっても使用することができる。1 X H i C 洗浄緩衝液は、室温で 2 ヶ月安定している。1 0 X 架橋反転緩衝液を 1 X まで U l t r a P u r e 水で希釈する。室温で保存する。1 つのサンプルあたりに約 1 m L の 1 X H i C 架橋反転緩衝液が必要である。1 X 架橋反転緩衝液は、近接ライゲーション・プロトコルにも使用することができる。1 X 架橋反転緩衝液は、室温で 2 ヶ月安定している。攪拌サーマルミキサーを、1 . 5 m L の管のために 1 2 5 0 r p m でセットするべきである。緩衝液を氷の上で解凍する、使用の前にボルテックスすることを含む優良試験所基準を使用する。

【0 4 6 5】

細胞のためのプロトコル

洗浄中の損失を考慮して、出発材料として $1 0 \times 1 0^6$ 個の細胞を使用することをお勧め

めする。10×10⁶個未満の細胞が利用可能な場合、低投入量プロトコルを参照する。開始の前に、新鮮な1X MNase消化緩衝液を調製し、室温で保存する。1X MNase消化緩衝液は室温で1日間安定している。1X MNase消化緩衝液を調製するために、以下を混合する：140μlのUltraPure水、20μlの10X MNase消化緩衝液、20μlの100mM MnCl₂、20μlの10%トリトン。

【0466】

細胞を採取し、1X PBSを洗浄する。細胞を計算し、10×10⁶個の細胞を分注し、2000×gで5分間遠心分離する。上澄みを注意深く除去する。ペレットを、5mlの1X PBSおよび135μlの37%ホルムアルデヒドにおいて再懸濁する。サンプルを5mlの管に転送し、細胞が沈降しない速度で室温で10分間回転する。管を2000×gで5分間遠心分離する。細胞ペレットが緩くなり得るから、上澄みを注意深く注意して除去する。ペレットをHiC洗浄緩衝液で洗浄し、最初に凝集塊を粉碎するために200μlで、その後、残りの4.8mlを加え、上下にピペティングしてペレットを完全に再懸濁する。管を2000×gで5分間回し、上澄みを注意深く除去する。洗浄ステップを、合計2回の洗浄繰り返す。2回目の洗浄剤を除去した後、ペレットを1mlの1X HiC洗浄緩衝液において再懸濁し、ペレットを再懸濁する。細胞を集計し、1×10⁶個の細胞を3つの別個の管に添加し、残りの細胞を、-80℃で冷凍されたペレットとして保存する。3つの管を2000×gで5分間遠心分離し、上澄みを除去する。各管のペレットを、50μlの1X MNase消化緩衝液（新鮮に調製された）において再懸濁する。管を、1250rpmで攪拌サーマルミキサーで30℃まで2分間あらかじめ暖める。7.5μlのMNase酵素ミックスを有する新鮮な1.5mlの管を、1250rpmで攪拌サーマルミキサーで30℃で2分間あらかじめ暖める。あらかじめ暖められたMNase酵素ミックス、各々のあらかじめ暖められた管に以下のように転送する、第1の管に0.5μl、第2の管に2.0μl、および第3の管に4.0μl。管を1250rpmで攪拌サーマルミキサーで30℃で30分間ちょうどインキュベートする。MNase反応を、5μlの0.5MのEDTAを添加し、および混合することによって止める。3μlの20%SDSを細胞を溶解するために添加し、そして、細胞を、1250rpmで攪拌サーマルミキサーで30℃で5分間インキュベートする。

【0467】

組織のためのプロトコル

注釈：60mgの組織を使用することが推奨される。60mg未満の組織が利用可能な場合、低投入量プロトコルを参照する。開始の前に、新鮮な1X MNase消化緩衝液を調製し、室温で保存する。1X MNase消化緩衝液は、室温で1日間安定している。1X MNase消化緩衝液を調製するために、以下を混合する：140μlのUltraPure水、20μlの10X MNase消化緩衝液、20μlの100mM MnCl₂、20μlの10%トリトン。少なくとも60mgの凍結組織を量り分け、液体窒素中で乳鉢と乳棒を用いて、図1Aおよび図1Bに例示される粘度に微粉碎し、ここで、図1Aは不十分な組織の粉碎を示し、図1Bは十分な組織の粉碎を示している。破裂された組織を、5mlの1X PBSおよび135μlの37%ホルムアルデヒドを有する5mlの管に転送する。管を室温で10分間回転させる。管を2000×gで5分間遠心分離し、上澄みを注意深く除去する。組織がペレット化しない事象で、管を最大速度で5分間回転される。ペレットを200μlの洗浄剤緩衝液上で再懸濁し、その後、4.8mlの1X HiC洗浄緩衝液を添加する。管を2000×gで5分間遠心分離し、上澄みを注意深く除去する。洗浄ステップを2回実施し、最終ペレットを1mlの1X HiC洗浄緩衝液において再懸濁する。必要に応じてフィルターを交換しながら、再懸濁された細胞を200μmのフィルターを通して新鮮な5mlの管へと通過させる。追加の2mlの1X HiC洗浄緩衝液を200μmのフィルターに通過させる。サンプルを、3つの別個の管で1mlの一定分量に分け、各々の一定分量は20mgの組織に対応する。過剰な組織をペレット化し、-80℃で保存することができる。3つの管を2000×gで5分間遠心分離し、上澄みを除去する。各管のペレットを、50μlの1X MNase消

10

20

30

40

50

化緩衝液（新鮮に調製された）において再懸濁する。管を、1250rpmで攪拌サーマルミキサーで30℃で2分間あらかじめ暖める。7.5μlのMNase酵素ミックスを有する新鮮な1.5mlの管を、1250rpmで攪拌サーマルミキサーで30℃で2分間あらかじめ暖める。あらかじめ暖められたMNase酵素ミックス、各々のあらかじめ暖められた管に以下のように転送する、第1の管に0.5μl、第2の管に2.0μl、および第3の管に4.0μl。管を1250rpmで攪拌サーマルミキサーで30℃で30分間ちょうどインキュベートする。MNase反応を、5μlの0.5MのEDTAを添加し、および混合することによって止める。3μlの20%SDSを細胞を溶解するために添加し、そして、細胞を、1250rpmで攪拌サーマルミキサーで30℃で5分間インキュベートする。

10

【0468】

実施例9：MNase-Cライブラリからの結果

【0469】

近接ライゲーションライブラリを、本明細書の方法を使用して調製し、長距離情報の測定値を決定するために配列決定した。図7は、MNase調製ライブラリ（「MNase-C」）、ならびに様々な架橋剤と比較した、DNase調製ライブラリ（「DNase-C」）、を使用したリード分離を示す。各バーについては、一番下のセグメントは、ゲノム内で1kb以上分離されているリードペアのパーセンテージを示し、中間のセグメントは、ゲノム内で1kb未満分離されているリードペアのパーセンテージを示し、一番上のセグメントは、染色体間相互作用のリードペアを示す。バーは左から右に、1）ホルムアルデヒド架橋剤を含むDNase、2）DSG架橋剤を含むDNase、3）10分間反応させたEGS架橋剤を含むMNase、4）30分間反応させたEGS架橋剤を含むMNase、5）10分間反応させたDSG架橋剤を含むMNase、6）30分間反応させたDSG架橋剤を含むMNaseの結果を示す。図8は、DNase-Cで調製したライブラリ（DSGとホルムアルデヒドの両方の架橋剤を含む）と比較した、MNase-Cで調製したライブラリ（DSGとEGSの両方の架橋剤を含む）の染色体1について計算したリンケージ距離の分布を示す。DSGのより長いスパーサーアームで調製されたDNase-Cライブラリは、図7にも反映されているように、ホルムアルデヒドで調製されたDNase-Cライブラリと比較して、より大きなリンケージ距離でより多くのリード率を示す。

20

【0470】

ゲノム幅のヌクレオソームマッピングの測定値も求められた。図9は、さまざまな量のMNaseおよびさまざまな消化時間、凡例に示されているように、30分で0.05ユニット、30分で0.5ユニット、20分で2.5ユニット、60分で2.5ユニット、で調製されたライブラリの高占有CTCF結合部位周辺の相対的なリードカバレッジを示す。図10は、（左から右へ）30'で0.05U、30'で0.5U、20'で2.5U、および60'で2.5Uにおける、調整されたライブラリにおいて発見されたジヌクレオソーム対モノヌクレオソームの比率を示し、右端でハシゴがある。ジヌクレオソーム対モノヌクレオソームの比率は（左から右へ）0.96、1.51、2.39、および4.86であった。

30

【0471】

実施例10：MNase消化されたサンプルでのHiC解析

ユーザーは、生体サンプル中の核酸結合タンパク質の位置を高精度で決定したい。生体サンプルを、核酸結合タンパク質をそれらが結合している核酸に架橋するために、化学固定剤を使用して架橋する。その後、固定されたサンプルを、タンパク質に結合していないすべての核酸を消化するマイクロコッカスヌクレアーゼ（MNase）で消化する。その後、MNaseで処置された核酸をDNAリガーゼで処置し、近接ライゲーション産物を取得する。核酸を精製し、配列ライブラリを作製する。配列ライブラリを、MNase消化サンプルのリードペアを取得するために配列決定する。リードペアから、核酸結合タンパク質に結合した核酸からの配列のみが取得されるため、生体サンプル中の核酸結合タンパ

40

50

ク質の局在が高精度で求められる。

【0472】

実施例11：MNase消化されたサンプルでのHiChIP

【0473】

MNaseHiChIP分析を以下のように実行する。細胞のサンプルをホルムアルデヒドとDSG架橋剤で架橋し、MNaseでインサイツ消化し、RIPA溶解緩衝液で溶解した。CCCTC結合因子(CTCF)の抗体およびH3K4me3修飾ヒストンをサンプルに接触させ、磁気ビーズを介してプルダウンした。エンドポリッシング、ブリッジライゲーション、および骨材内ライゲーションを含む近接ライゲーションを実施した。その後、架橋を反転し、DNAをクリーンアップし、および配列ライブラリを調製し、2,000~4,000万の2x150bpリードで配列決定した。

【0474】

比較のために、ChIP-seq分析は、同じ抗体を使用したサンプルの並列セットで実行された。手順は、架橋、消化、および近接ライゲーションを除いて、同様であった。

【0475】

図11は、University of California, Santa Cruz (UCSC) Genome BrowserのDNA要素(ENCODE)の百科事典に報告されたピークと比較されたChIP-seqおよびHiChIPの結果を示す。陰性対照ライブラリは全ゲノムにわたってカバレッジを示すが、CTCFとH3K4me3のライブラリは、それら結合部位のENCODEピークの位置に一致する遺伝子座で配列決定リードのパイルアップを示す。これは、選択された抗体に対応するタンパク質が結合すると予想される領域からほぼ完全に読み取られていて、これらライブラリがChIP-seqシグナルを伝えることを示す。

【0476】

図12は、HiChIPサンプルのCTCF結合部位周辺の相対的なリードカバレッジを示す。相対的なリードカバレッジは、およそ146bpの周期性を示し、これはヌクレオソーム中のヒストンの存在によってMNase消化から保護されたDNAのエリアと一致している。これは、これらライブラリがMNase-Cライブラリに期待される保護プロファイルを有することを示す。

【0477】

表3-5では、inter_chrは、染色体間のリードペア（ペア中の各リードは異なる染色体にマップする）の割合を表わし、<1kbは1kb未満のゲノム上の分離距離に及ぶリードペアの割合を表わし、>1kbは1kbを越えるゲノム上の分離距離に及ぶリードペアの割合を表わし、fracMappedはマップされたリードペアの割合を表し、preSeqAt300Mは300万のリードの集団からの特有のリードの数を表わす。

【0478】

表3は、MNase-Cライブラリのための典型ライブラリの品質管理(QC)メトリックを示す。染色体間のリードペアの割合は33%と35%の間であり、1kb未満の分離距離に及ぶリードペアの割合は7%と12%の間であり、1kbを越える分離距離に及ぶリードペアの割合は55%と58%の間であり、マップされた断片は67%と79%の間にあり、300万のうちの特有のリードの数は168万と238万の間である。表4に示されるように、MNaseHiChIPライブラリのQCメトリックは、比較的にMNase-Cライブラリのものであり、染色体間リードペアは21.6%と43.4%の間であり、1kb未満に及ぶリードペアは18.2%と31.2%の間であり、1kbを越えるリードペアは25.4%と56.2%の間であり、特有のリードは300万のうち272万と279万の間である。比較すると、ChIP-seqライブラリのためのQCメトリックはMNase-Cライブラリ（または他の近接ライゲーションライブラリ）とは異なり、染色体間リードは0.46%と1.65%の間であり、1kb未満に及ぶリードペアは98.64%と99.45%の間であり、1kbを越えるリードペアは0.

0.9%と0.16%の間である。

【0479】

【表3】

表3：MNase-Cのための典型ライブラリのQCメトリック					
サンプルID	inter.chr	<1kb	>1kb	fracMapped	preSeqAt300M
mdev3A	35%	7%	58%	67%	232,517,526
mdev3B	35%	8%	57%	68%	237,654,318
mdev3C	34%	11%	55%	77%	218,231,018
mdev3D	33%	12%	55%	77%	217,283,442
mdev3E	33%	11%	56%	78%	171,109,061
mdev3F	33%	11%	56%	79%	168,589,170

【0480】

【表4】

表4：MNase HiChIP ライブラリのためのQCメトリック				
	inter.chr	<1kb	>1kb	preSeqAt300M
陰性対照	21.6%	22.2%	56.2%	270,540,612
CTCF	40.3%	18.2%	41.5%	278,669,408
H3K36me3	43.4%	31.2%	25.4%	272,358,079

【0481】

【表5】

表5：ChIP-seq ライブラリのためのQCメトリック				
	inter.chr	<1kb	>1kb	preSeqAt300M
陰性対照	1.23%	98.64%	0.13%	NA
CTCF	1.65%	98.19%	0.16%	NA
H3K36me3	0.46%	99.45%	0.09%	NA

【0482】

図13は、標的タンパク質（図11に示されるように）と関連するリードのパイルアップを示すリードカバレッジのグラフ、および遺伝子アノテーションのグラフで表示されたリードペアのコンタクトマップを示す。H3K4me3修飾は近くの遺伝子の転写と関連する。コンタクト（1301）の高密度領域は、リードカバレッジのH3K4me3に関連されたピーク（1302）で始まり、左方向に続き；下部に示されるように、この領域（1303）はまた同じ方向のリードである遺伝子を含むものとしても注釈が付けら

れる。同様に、(1304)で、コンタクトの高密度領域は、リードカバレッジの別のH3K4me3に関連されたピーク(1305)で始まり、右方向に続き；下部に示されるように、この領域(1306)はまた同じ方向のリードであるFABP5遺伝子を含むものとしても注釈が付けられる。CTCFはクロマチンループの形成に関連し、2つのCTCFタンパク質は一体となる異なる遺伝子座に結合され、それらの間のDNAはループを形成する。例えば、トロポロジ関連ドメイン(TAD)を表すコンタクト密度の三角形が見られ、例えば、ピーク(1311)、(1314)、(1317)、および(1320)を有するものである。これら三角形の左端と右端をたどると、これら領域の境界はCTCFに関連されたピークと一致することが分かり(例えば、CTCFピーク(1312)および(1313)を有する(1311)、CTCFピーク(1315)および(1316)を有する(1314)、CTCFピーク(1318)および(1319)を有する(1317)、およびCTCFピーク(1321)および(1322)を有する(1320))、どの特定のCTCF部位が一体となってどのループとドメインを形成するかの識別を可能にする。

【0483】

図14は、ENCODEピークとMNase HiChIP結果の同じ比較を示すが、サンプルは同日および後日における複製である。これは、プロトコルの一貫性と再現性を示す。

【0484】

全体的に見ると、これら実験は、MNase HiChIPライブラリがChIP-seq特徴、MNase特性、Hi-C特性を有し、タンパク質ピーク間のHi-C相互作用を示すこと、およびプロトコルは堅牢であり高い再現性を有することを実証する。

【0485】

実施例12：スプリットプール標識アプローチを使用する近接ライゲーション

架橋剤で固定された細胞を含む安定した生体サンプルを得る。サンプルをDNaseで処置し、インサイツで細胞のDNAを消化する。その後、サンプルを酵素で処置し、DNAの末端を研磨し、DNA末端をポリアデニル化する。その後、細胞を1つのウェルあたりに1つの細胞を用いて、96ウェルプレートのウェルへ分注する。バーコードを各ウェルに加え、各ウェル中のDNA末端にライゲートする。その後、細胞をプーリングし、各ウェル中の単細胞で再び分注する。第2のバーコードを加え、各ウェル中の第1のバーコードにライゲートする。その後、細胞を再びプーリングして1つのウェルあたりに1つの単細胞を用いて再び分注し、架橋アダプターを第2の架橋アダプターと適合するオーバーハングでライゲートする。このアプローチは図16に図示される。その後、近位端をライゲートし、結果として、各末端は2つのバーコードと他方の末端上の別の架橋にリンクする架橋を有する、図17に図示される分子をもたらす。図18は、バーコードの組み合わせの例、およびスプリティングとプーリング・アプローチから生じる架橋を示す。架橋はその後反転され、核酸は精製および配列決定されて、配列情報を得る。

【0486】

実施例13：標的とされたアダプターを使用する近接ライゲーション

架橋剤で固定された細胞を含む安定した生体サンプルを得る。サンプルをDNaseで処置して、インサイツで細胞のDNAを消化する。その後、サンプルを酵素で処置して、DNAの末端を研磨し、DNA末端をポリアデニル化する。その後、サンプルをDNA中のヒストンに結合する抗体と接触させ、その後DNA末端をライゲートする前に、サンプルを複数のタンパク質Aの繋がれたビオチン化されたライブラリアダプターと接触させる。アダプターを、ヒストン結合抗体が結合された近接端間でライゲートする。ストレプトアビジンを使用してビオチン化したアダプターを引き下ろし、架橋は結果として生じる精製されたサンプル上で反転する。その後、増幅とPCRは配列情報を得るために実行される。

【0487】

実施例14：集合体3次元核酸構成の決定は細胞の特異的な情報を失うことがある。

細胞集団を3次元核酸構成解析のために採取する。細胞の集団は染色体1および染色体2の領域を近接させる核酸構成を共有する一方で、染色体3および染色体4の領域は細胞のすべてではなく細胞のいくつかの近接にある。安定した核を、区画し、内部末端を暴露するために断片化し、a t t B タグを付け、その後、p h i C 3 1 インテグラーゼの存在下での細胞の識別情報を欠いた a t t P リンク核酸の集団に接触させる。

【0488】

ライブラリ構成要素を末端配列決定する。共通分子上の染色体1および染色体2にマッピングするリードペアが、バックグラウンド上に差動的に観察されることが観察される。共通分子上の染色体3および染色体4にマッピングするリードペアが、より低い頻度でバックグラウンド上に差動的に観察されることが観察される。

【0489】

染色体3および染色体4は近接であるが染色体1および染色体2よりも互いに遠く離れているかどうか、または細胞集団のメンバ間の構成変動があるかどうかを区別することはできない。

【0490】

実施例15：細胞特異的な3次元核酸構成情報は、本明細書の方法を通して保護される。

細胞集団を3次元核酸構成解析のために採取する。細胞の集団は染色体1および染色体2の領域を近接させる核酸構成を共有する一方で、染色体3および染色体4の領域は細胞のすべてではなく細胞のいくつかの近接にある。安定した核を、区画し、内部末端を暴露するために断片化し、a t t B タグを付け、その後、p h i C 3 1 インテグラーゼの存在下での細胞の識別情報を有する a t t P リンク核酸の集団に接触させる。

【0491】

ライブラリ構成要素を末端配列決定することで、内部末端の隣接する配列とリンカーの区画を識別する配列を得るようになる。共通分子上の染色体1および染色体2にマッピングするリードペアが、区画を識別する配列とは無関係のバックグラウンド上に差動的に観察されることが観察される。共通分子上の染色体3および染色体4にマッピングするリードペアが、細胞識別化ライブラリ構成要素の第1の集団では染色体1および染色体2のものと匹敵するレベルで観察されるが、細胞識別化ライブラリ構成要素の第2の集団のバックグラウンド上で観察されないことが観察される。染色体3および染色体4の関連するセグメントは集団内の立体構造の変動を示し、いくつかの細胞は染色体1および染色体2のセグメントに対して染色体3および染色体4のセグメントの3次元近接を示す一方で、他の細胞は染色体1および染色体2のセグメントに対して染色体3および染色体4のセグメントの3次元近接を示さないようになる。

【0492】

実施例16：細胞特異的な3次元核酸構成情報は定量的に測定される。

細胞集団を3次元核酸構成解析のために採取する。細胞の集団は染色体1および染色体2の領域を近接させる核酸構成を共有する一方で、染色体3および4の領域は細胞間でそれらの近接性が量的に変動する。安定した核を、区画し、内部末端を暴露するために断片化し、a t t B タグを付け、その後、p h i C 3 1 インテグラーゼの存在下で細胞識別情報を有する a t t P リンク核酸の集団に接触させる。

【0493】

ライブラリ構成要素を末端配列決定することで、内部末端の隣接する配列とリンカーの区画を識別する配列を得るようになる。共通分子上の染色体1および染色体2にマッピングするリードペアが、区画を識別する配列とは無関係のバックグラウンド上に差動的に観察されることが観察される。共通分子上の染色体3および染色体4にマッピングするリードペアが、リードペアの頻度と組み合わせた区画を識別する配列情報によって示されるように、細胞によって変動するレベルで観察される。染色体1および染色体2の示されたセグメントは細胞の集団全体にわたって近接にあることが結論付けられる。染色体3および染色体4の関連するセグメントは集団内の立体構造の変動を示し、染色体3および4の示されたセグメントの近接性に関して、連続体全体で量的な変動があることが結論付けられ

10

20

30

40

50

る。

【0494】

本発明の好ましい実施形態が本明細書中で示され記載されてきたが、このような実施形態はほんの一例として提供されているに過ぎないことが当業者に明らかであろう。当業者であれば、多くの変更、変化、および置換が、本発明から逸脱することなく思いつくだろう。本明細書に記載される実施形態の様々な代案が利用され得ることを理解されたい。以下の請求項は本発明の範囲を定義するものであり、この請求項とその均等物の範囲内の方法、および構造体がそれによって包含されるものであるということが意図されている。

10

20

30

40

50

【図面】

【図 1 A】

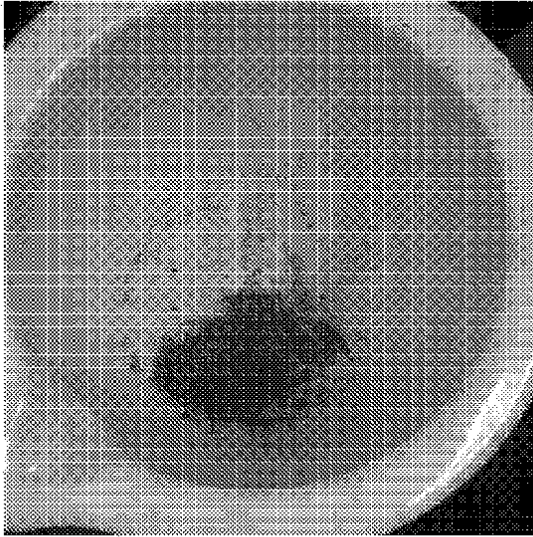


FIG. 1A

【図 1 B】

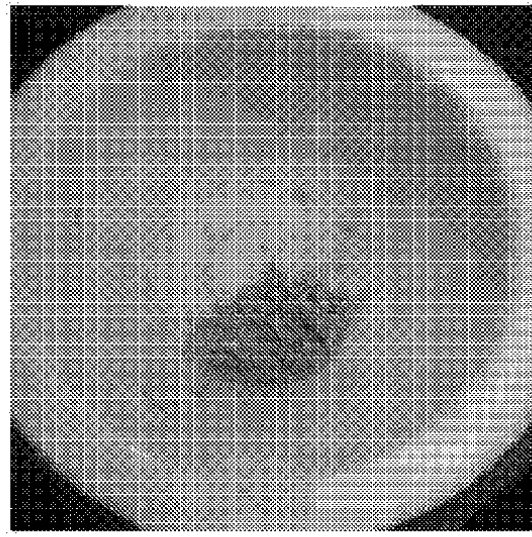


FIG. 1B

【図 2】

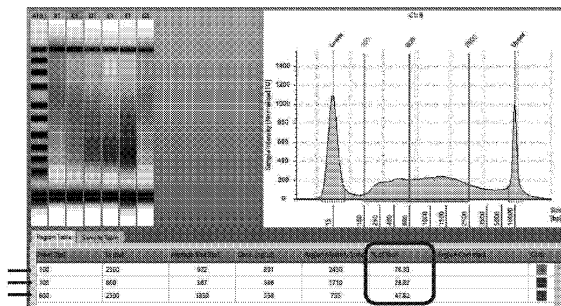
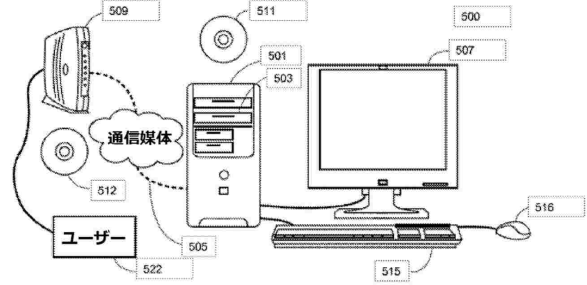


FIG. 2

【図 3】



10

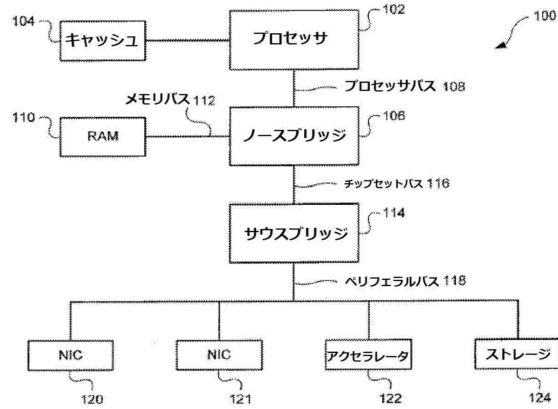
20

30

40

50

【図 4】



【図 5】

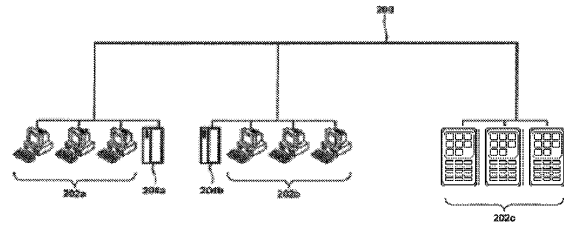
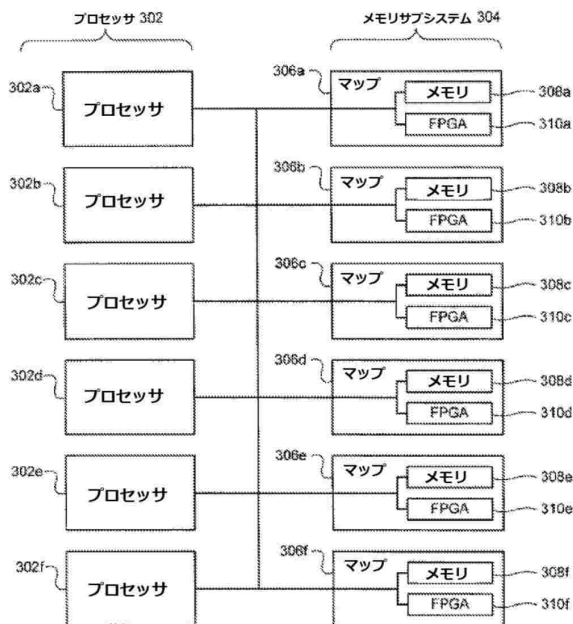
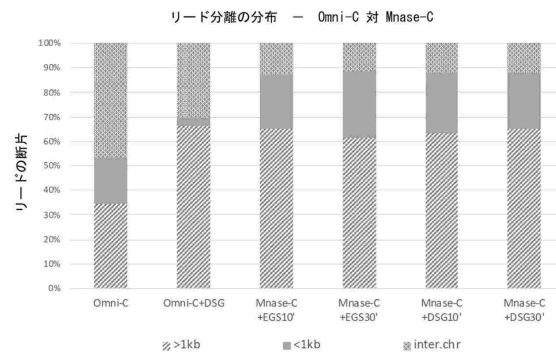


FIG. 5

【図 6】



【図 7】



10

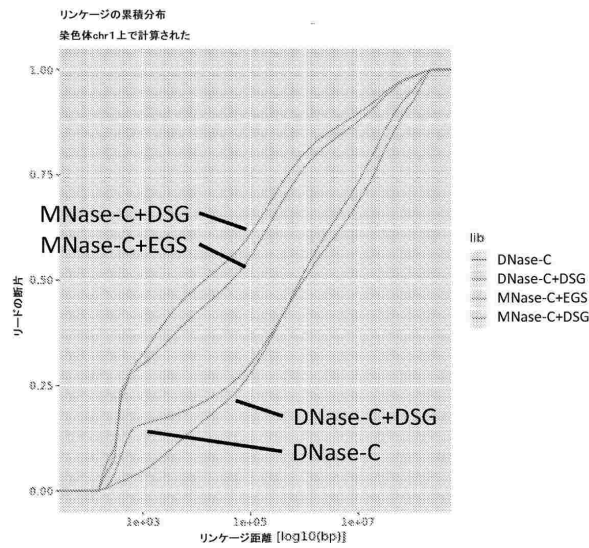
20

30

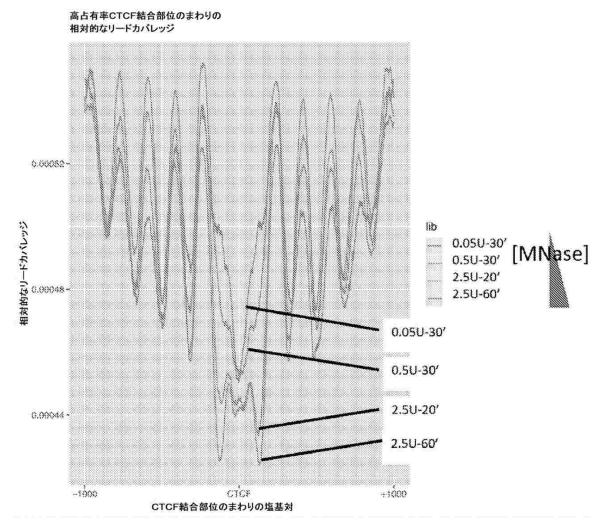
40

50

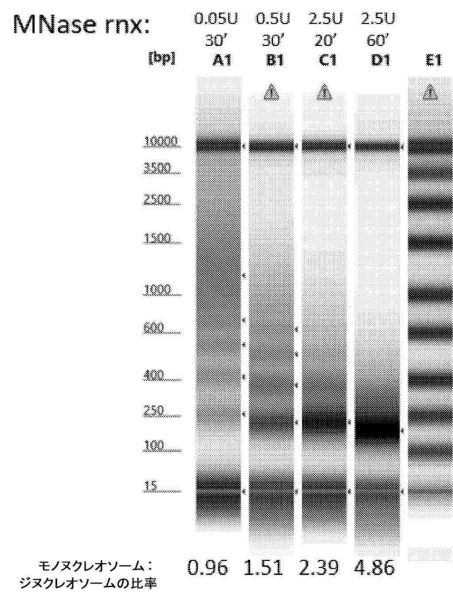
【図 8】



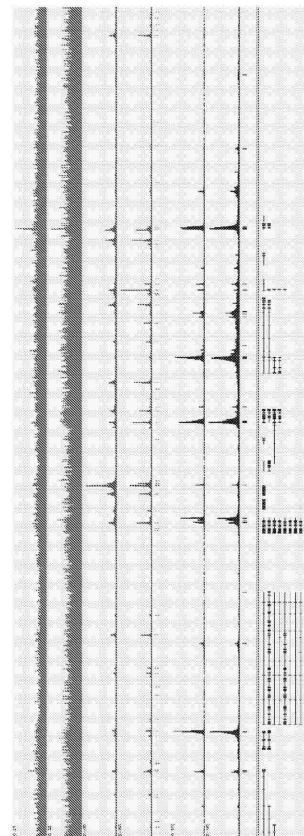
【図 9】



【図 10】



【図 11】



10

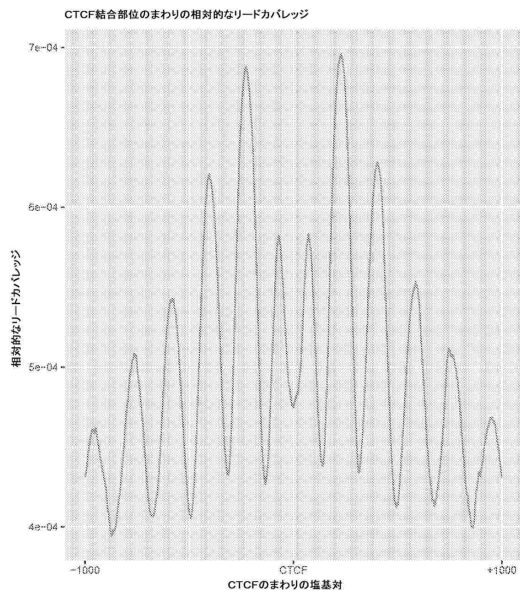
20

30

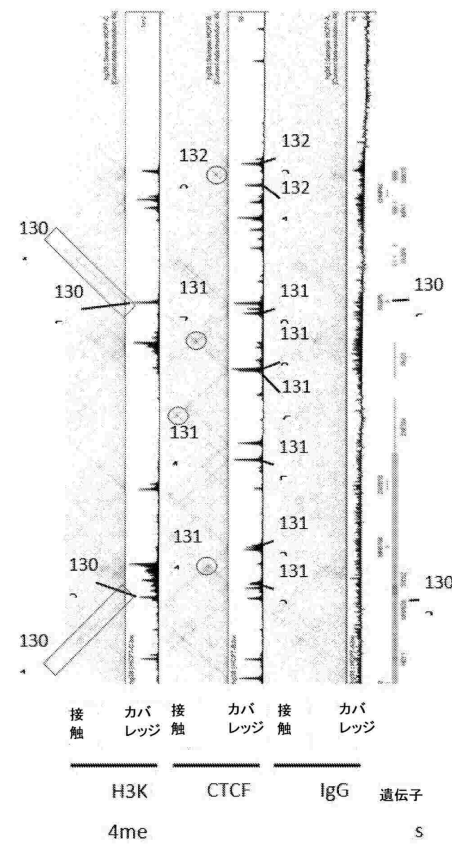
40

50

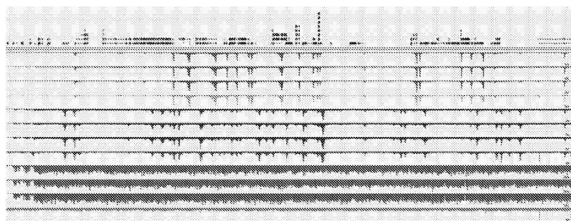
【図 1 2】



【図 1 3】

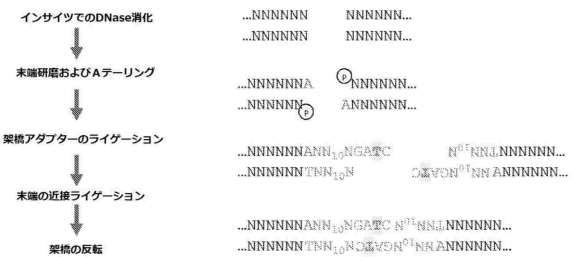


【図 1 4】



rep1	1	2	3	1	2	3	1	2	3	遺伝子
日	2	1	2	1	2	1	2			
	陰性	CTCF				H3K				
	対照					4me				

【図 1 5】



10

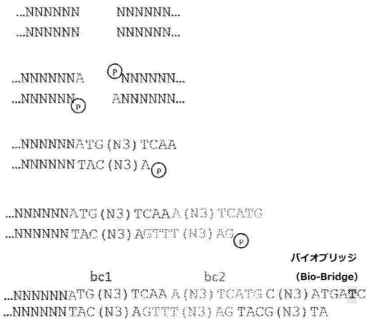
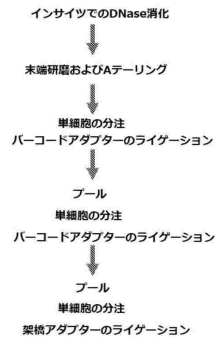
20

30

40

50

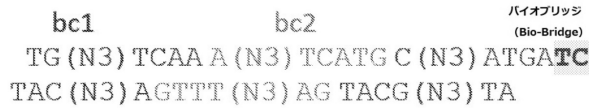
【図 16】



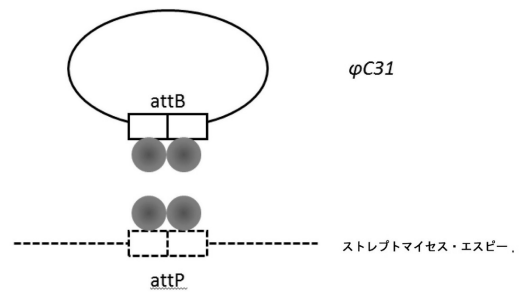
【図 17】



【図 18】



【図 19 A】



10

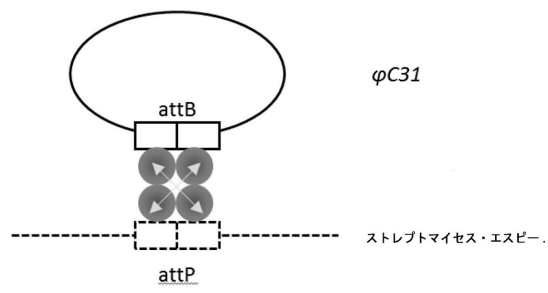
20

30

40

50

【図 19 B】



【図 19 C】



10

20

【図 19 D】

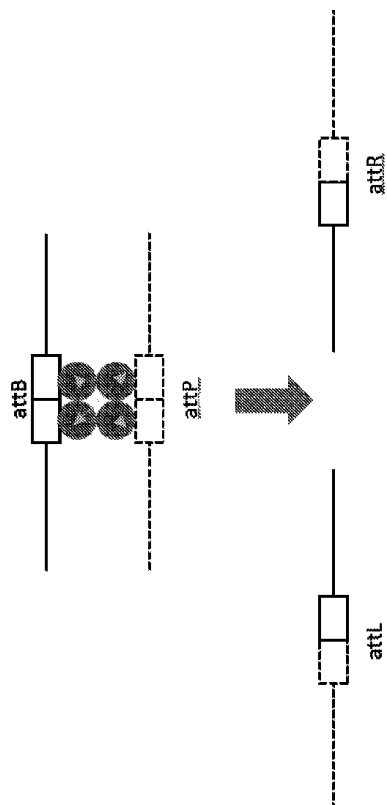
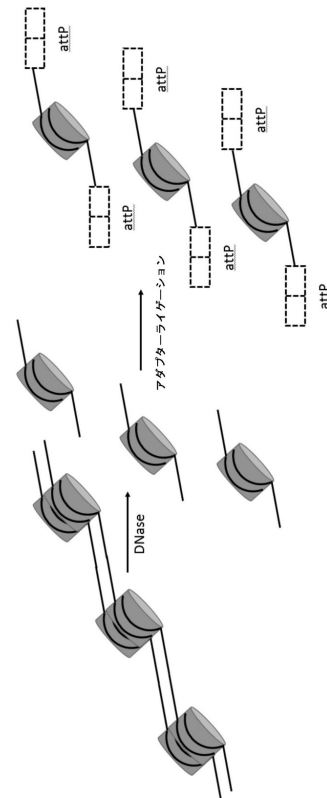


FIG. 19D

【図 20】



30

40

50

【図 2 1】

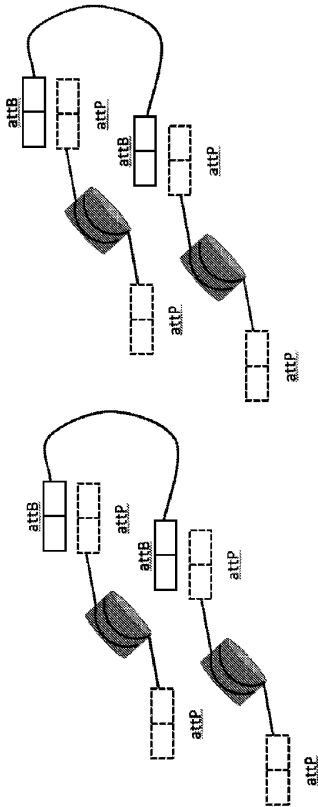


FIG. 21

【図 2 2】

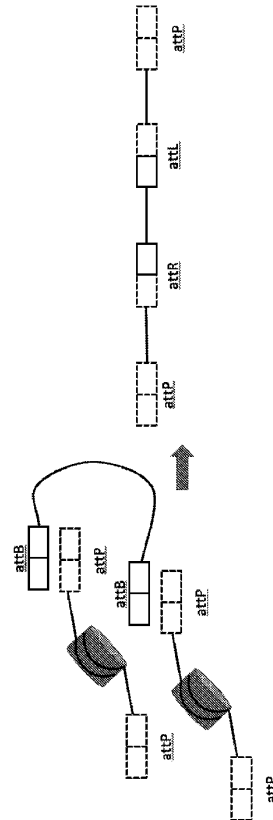
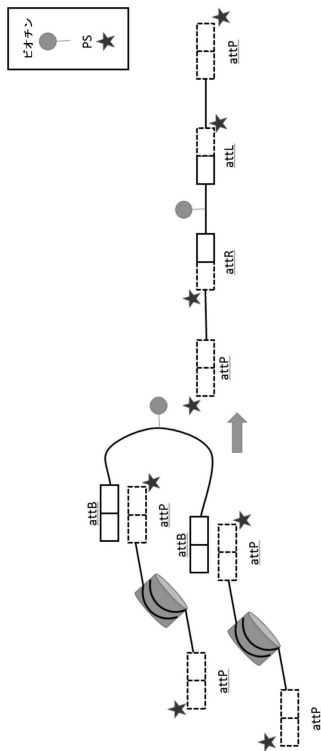
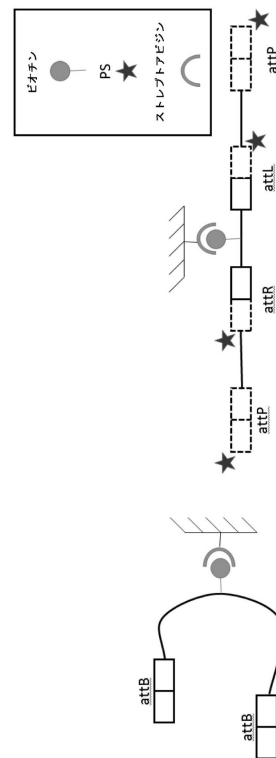


FIG. 22

【図 2 3 A】



【図 2 3 B】



10

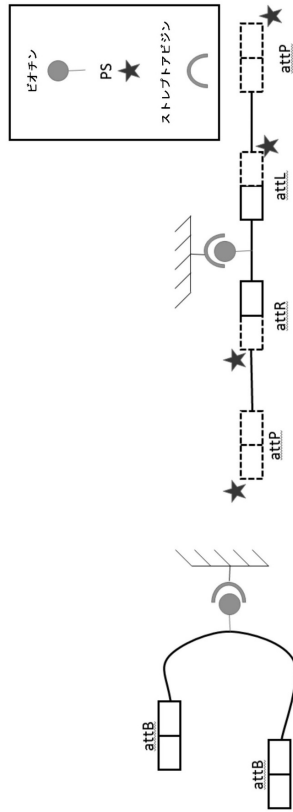
20

30

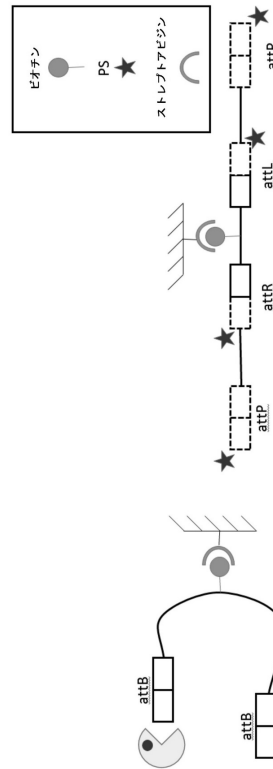
40

50

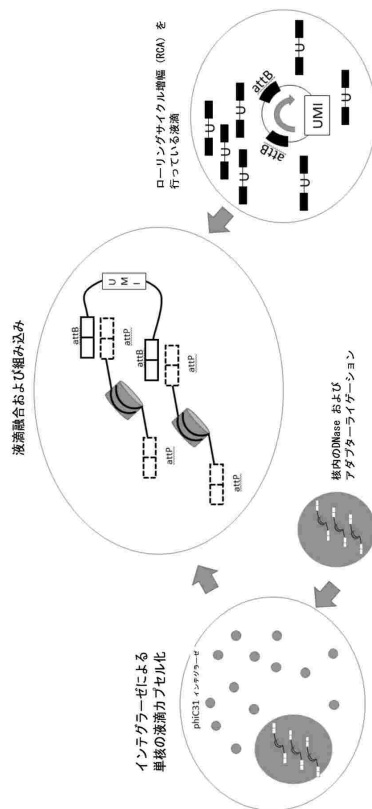
【図 2 3 C】



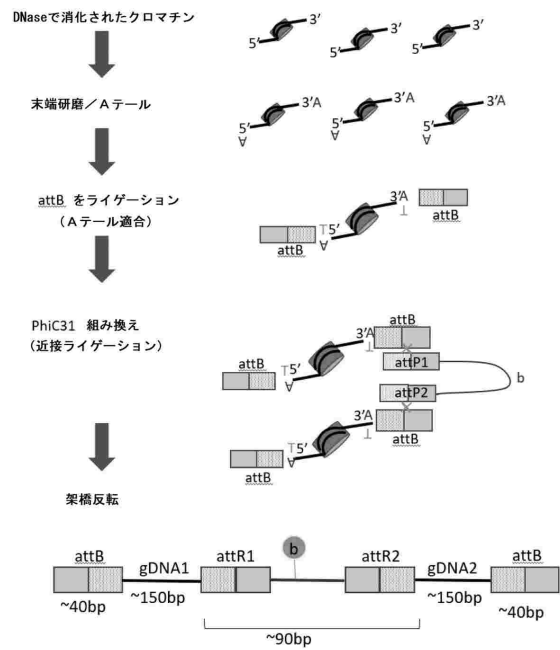
【図 2 3 D】



【図 2 4】



【図 2 5】



10

20

30

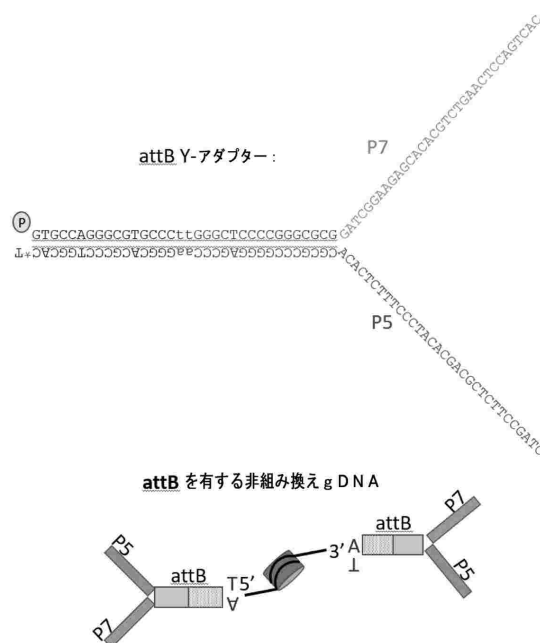
40

50

【圖 26】



【圖 27】



【配列表】

0007653924000001.app

フロントページの続き

(33)優先権主張国・地域又は機関

米国(US)

(31)優先権主張番号 62/931,069

(32)優先日 令和1年11月5日(2019.11.5)

(33)優先権主張国・地域又は機関

米国(US)

(31)優先権主張番号 63/011,490

(32)優先日 令和2年4月17日(2020.4.17)

(33)優先権主張国・地域又は機関

米国(US)

(31)優先権主張番号 63/014,422

(32)優先日 令和2年4月23日(2020.4.23)

(33)優先権主張国・地域又は機関

米国(US)

ニア州 スコッツバレー エンタープライズ・ウェイ 100 スイート エー101

審査官 福間 信子

(56)参考文献 特表2019-500009 (JP, A)

BMC Mol Biol, 2012年, 13:15

Trends Genet, 2015年, vol.7, pp.357-372

(58)調査した分野 (Int.Cl., DB名)

C12N

C A p l u s / M E D L I N E / E M B A S E / B I O S I S (S T N)