

(19) 日本国特許庁(JP)

(12) 公表特許公報(A)

(11) 特許出願公表番号

特表2017-522908

(P2017-522908A)

(43) 公表日 平成29年8月17日(2017.8.17)

(51) Int. Cl.	F I	テーマコード (参考)
C 1 2 Q 1/68 (2006.01)	C 1 2 Q 1/68	Z N A A 4 B O 6 3
G O 1 N 33/53 (2006.01)	C 1 2 Q 1/68	Z
	G O 1 N 33/53	M

審査請求 未請求 予備審査請求 未請求 (全 84 頁)

(21) 出願番号 特願2017-525327 (P2017-525327)
 (86) (22) 出願日 平成27年7月27日 (2015. 7. 27)
 (85) 翻訳文提出日 平成29年3月24日 (2017. 3. 24)
 (86) 国際出願番号 PCT/US2015/042310
 (87) 国際公開番号 W02016/015058
 (87) 国際公開日 平成28年1月28日 (2016. 1. 28)
 (31) 優先権主張番号 62/029, 178
 (32) 優先日 平成26年7月25日 (2014. 7. 25)
 (33) 優先権主張国 米国 (US)
 (31) 優先権主張番号 62/087, 619
 (32) 優先日 平成26年12月4日 (2014. 12. 4)
 (33) 優先権主張国 米国 (US)

(71) 出願人 517025822
 ユニヴァーシティ オブ ワシントン
 アメリカ合衆国 ワシントン州 9810
 5-4721 シアトル ルーズヴェルト
 ウェイ ノースイースト 4545 ス
 イート 400
 (74) 代理人 100094569
 弁理士 田中 伸一郎
 (74) 代理人 100088694
 弁理士 弟子丸 健
 (74) 代理人 100103610
 弁理士 ▲吉▼田 和彦
 (74) 代理人 100084663
 弁理士 箱田 篤

最終頁に続く

(54) 【発明の名称】 セルフリーDNAを生じる組織及び/又は細胞タイプを決定する方法、並びにそれを用いて疾患又は異常を識別する方法

(57) 【要約】

本開示は、対象者からの生物学的サンプル中のセルフリーDNA (“ cfDNA ”) に寄与する1つ以上の組織及び/又は細胞タイプを決定する方法を提供する。いくつかの実施態様では、本開示は、対象者の疾患又は異常を、当該対象者からの生物学的サンプル中のcfDNAに寄与すると決定された1つ以上の組織及び/又は細胞タイプの関数と認定する方法を提供する。

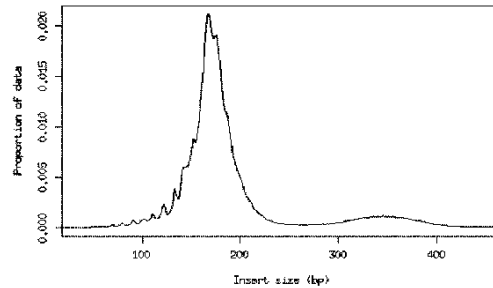


Figure 2

【特許請求の範囲】

【請求項 1】

対象者においてセルフリーDNA (cfDNA) を生じる組織及び/又は細胞タイプを決定する方法であって、

当該対象者からの生物学的サンプルからcfDNAを単離する工程であって、当該単離cfDNAが複数のcfDNAフラグメントを含む、前記工程；

当該複数のcfDNAフラグメントの少なくとも一部分に付随する配列を決定する工程；

当該複数のcfDNAフラグメントの少なくともいくつかのcfDNAフラグメント末端の参照ゲノム内におけるゲノム上の場所を、当該cfDNAフラグメント配列の関数として決定する工程；及び

当該cfDNAフラグメントを生じる組織及び/又は細胞タイプの少なくともいくつかを、当該cfDNAフラグメント末端の少なくともいくつかのゲノム上の場所の関数として決定する工程を含む、

前記方法。

【請求項 2】

cfDNAフラグメントを生じる組織及び/又は細胞タイプの少なくともいくつかを決定する工程が、当該cfDNAフラグメント末端の少なくともいくつかのゲノム上の場所を1つ以上の参照マップと比較する工程を含む、請求項1に記載の方法。

【請求項 3】

cfDNAフラグメントを生じる組織及び/又は細胞タイプの少なくともいくつかを決定する工程が、当該cfDNAフラグメント末端の少なくともいくつかのゲノム上の場所の分布に関して数学的変換を実施する工程を含む、請求項1又は請求項2に記載の方法。

【請求項 4】

数学的変換がフーリエ変換を含む、請求項3に記載の方法。

【請求項 5】

参照ゲノムの少なくともいくつかの座標の各々に対するスコアを決定する工程をさらに含み、ここで当該スコアが少なくとも複数のcfDNAフラグメント末端及びそれらのゲノム上の場所の関数として決定され、さらに観察されるcfDNAフラグメントを生じる組織及び/又は細胞タイプの少なくともいくつかを決定する工程が当該スコアと1つ以上の参照マップを比較する工程を含む、請求項1から4のいずれかに記載の方法。

【請求項 6】

座標に関するスコアが、当該座標がcfDNAフラグメント末端の場所である確率を表しているか又は前記確率と関係性がある、請求項5に記載の方法。

【請求項 7】

参照マップが、少なくとも1つの細胞タイプ又は組織から作製されたDNase I高感受性サイトデータセットを含む、請求項2から6のいずれか1項に記載の方法。

【請求項 8】

参照マップが、少なくとも1つの細胞タイプ又は組織から作製されたRNA発現データセットを含む、請求項2から7のいずれか1項に記載の方法。

【請求項 9】

当該参照マップが、ヒト組織又は細胞を異種移植された動物のcfDNAから作製される、請求項2から8のいずれか1項に記載の方法。

【請求項 10】

参照マップが、少なくとも1つの細胞タイプ又は組織から作製された染色体コンフォーメーションマップを含む、請求項2から9のいずれか1項に記載の方法。

【請求項 11】

参照マップが、少なくとも1つの細胞タイプ又は組織から作製されたクロマチン接近容易性マップを含む、請求項2から10のいずれか1項に記載の方法。

【請求項 12】

参照マップが、少なくとも1人の参照対象者から入手されるサンプルから得られる配列

10

20

30

40

50

データを含む、請求項2から11のいずれか1項に記載の方法。

【請求項13】

参照マップが、疾患又は異常に付随する少なくとも1つの細胞タイプ又は組織に対応する、請求項2から12のいずれか1項に記載の方法。

【請求項14】

参照マップが、組織又は細胞タイプのヌクレオソーム及び/又はクロマトソームの位置又は間隔を含む、請求項2から13のいずれか1項に記載の方法。

【請求項15】

参照マップが、少なくとも1つの細胞タイプ又は組織から入手されたクロマチンを外因性ヌクレアーゼ（例えばマイクロコッカスヌクレアーゼ）で消化することによって作製される、請求項2から14のいずれか1項に記載の方法。

10

【請求項16】

参照マップが、転移による方法（例えばATAC-seq）によって決定されたクロマチン接近容易性データを含む、請求項2から15のいずれか1項に記載の方法。

【請求項17】

参照マップが、ある組織又は細胞タイプのDNA結合及び/又はDNA占拠タンパク質の位置と結びついたデータを含む、請求項2から16のいずれか1項に記載の方法。

【請求項18】

DNA結合及び/又はDNA占拠タンパク質が転写因子である、請求項17に記載の方法。

【請求項19】

位置が、架橋DNA-タンパク質複合体のクロマチン免疫沈澱によって決定される、請求項17又は請求項18に記載の方法。

20

【請求項20】

位置が、当該組織又は細胞タイプに付随するDNAをヌクレアーゼ（例えばDNase-I）で処理することによって決定される、請求項17又は請求項18に記載の方法。

【請求項21】

参照マップが、ある組織又は細胞タイプにおけるヌクレオソーム、クロマトソーム、又は他のDNA結合若しくはDNA占拠タンパク質の位置又は間隔に関連する生物学的特徴を含む、請求項2から20のいずれか1項に記載の方法。

【請求項22】

生物学的特徴が1つ以上の遺伝子の定量的発現である、請求項21に記載の方法。

30

【請求項23】

生物学的特徴が1つ以上のヒストンマークの有無である、請求項21又は請求項22に記載の方法。

【請求項24】

生物学的特徴がヌクレアーゼ切断に対する高感受性である、請求項21から23のいずれか1項に記載の方法。

【請求項25】

参照マップを作製するために用いられる組織又は細胞タイプが、疾患又は異常を有する対象者の一次組織である、請求項2から24のいずれか1項に記載の方法。

40

【請求項26】

疾患又は異常が、癌、正常妊娠、妊娠（例えば異数性妊娠）の合併症、心筋梗塞、炎症性腸疾患、全身性自己免疫疾患、局所性自己免疫疾患、拒絶を伴う同種間移植、拒絶を伴わない同種間移植、卒中、及び局在性組織損傷から成る群から選択される、請求項25に記載の方法。

【請求項27】

参照マップを作製するために用いられる組織又は細胞タイプが、健康な対象者の一次組織である、請求項2から24のいずれか1項に記載の方法。

【請求項28】

参照マップを作製するために用いられる組織又は細胞タイプが不死化細胞株である、請

50

求項2から24のいずれか1項に記載の方法。

【請求項 29】

参照マップを作製するために用いられる組織又は細胞タイプが腫瘍の生検である、請求項2から24のいずれか1項に記載の方法。

【請求項 30】

配列データがcfDNAフラグメント末端の位置を含む、請求項12に記載の方法。

【請求項 31】

参照対象者が健康である、請求項30に記載の方法。

【請求項 32】

参照対象者が疾患又は異常を有する、請求項30に記載の方法。

10

【請求項 33】

疾患又は異常が、癌、正常妊娠、妊娠（例えば異数性妊娠）の合併症、心筋梗塞、炎症性腸疾患、全身性自己免疫疾患、局所性自己免疫疾患、拒絶を伴う同種間移植、拒絶を伴わない同種間移植、卒中、及び局在性組織損傷から成る群から選択される、請求項32に記載の方法。

【請求項 34】

参照マップが、組織又は細胞タイプに付随する参照ゲノムの座標の少なくとも一部分に対する参照スコアを含む、請求項13から33のいずれか1項に記載の方法。

【請求項 35】

参照マップがスコアの数学的変換を含む、請求項34に記載の方法。

20

【請求項 36】

スコアが、組織又は細胞タイプの全ての参照ゲノム座標のサブセットを表す、請求項34に記載の方法。

【請求項 37】

サブセットがヌクレオソーム及び/又はクロマトソームの位置又は間隔と結びついている、請求項36に記載の方法。

【請求項 38】

サブセットが転写開始サイト及び/又は転写終了サイトと結びついている、請求項36又は請求項37に記載の方法。

【請求項 39】

サブセットが少なくとも1つの転写因子の結合サイトと結びついている、請求項36から38のいずれか1項に記載の方法。

30

【請求項 40】

サブセットがヌクレアーゼ高感受性サイトと結びついている、請求項36から39のいずれか1項に記載の方法。

【請求項 41】

サブセットが、さらに加えて少なくとも1つのオルトゴナルな生物学的特徴と結びついている、請求項36から40のいずれか1項に記載の方法。

【請求項 42】

オルトゴナルな生物学的特徴が高発現遺伝子に付随する、請求項41に記載の方法。

40

【請求項 43】

オルトゴナルな生物学的特徴が低発現遺伝子に付随する、請求項41に記載の方法。

【請求項 44】

数学的変換がフーリエ変換を含む、請求項35から43のいずれか1項に記載の方法。

【請求項 45】

複数のスコアの少なくとも1つのサブセットが閾値を超えるスコアを有する、請求項5から44のいずれか1項に記載の方法。

【請求項 46】

cfDNAを生じる組織及び/又は細胞タイプを、当該cfDNAフラグメント末端の少なくともいくつかの複数のゲノム上の場所の関数として決定する工程が、当該cfDNAフラグメント

50

末端の少なくともいくつかの複数のゲノム上の場所のフーリエ変換、又はその数学的変換を、参照マップと比較する工程を含む、請求項1から45のいずれか1項に記載の方法。

【請求項47】

単離cfDNAを生じると決定された組織及び/又は細胞タイプの一覧を含む報告を作製する工程をさらに含む、請求項1から46のいずれか1項に記載の方法。

【請求項48】

対象者における疾患又は異常を識別する方法であって、当該対象者からの生物学的サンプルからセルフリーDNA(cfDNA)を単離する工程であって、当該単離cfDNAが複数のcfDNAフラグメントを含む、前記工程；
当該複数のcfDNAフラグメントの少なくとも一部分に付随する配列を決定する工程；
当該複数のcfDNAフラグメントの少なくともいくつかのcfDNAフラグメント末端に対する参照ゲノム内のゲノム上の場所を当該cfDNAフラグメント配列の関数として決定する工程；
当該cfDNAを生じる組織及び/又は細胞タイプの少なくともいくつかを、当該cfDNAフラグメント末端の少なくともいくつかのゲノム上の場所の関数として決定する工程；及び
当該疾患又は異常を、当該cfDNAを生じると決定された組織及び/又は細胞タイプの関数と認定する工程を含む、
前記方法。

10

【請求項49】

cfDNAフラグメントを生じる組織及び/又は細胞タイプを決定する工程が、当該cfDNAフラグメント末端の少なくともいくつかのゲノム上の場所を1つ以上の参照マップと比較する工程を含む、請求項48に記載の方法。

20

【請求項50】

cfDNAフラグメントを生じる組織及び/又は細胞タイプを決定する工程が、複数のcfDNAフラグメント末端の少なくともいくつかのゲノム上の場所の分布に関して数学的変換を実施する工程を含む、請求項48又は請求項49に記載の方法。

【請求項51】

数学的変換がフーリエ変換を含む、請求項50に記載の方法。

【請求項52】

参照ゲノムの少なくともいくつかの座標の各々に対するスコアを決定する工程をさらに含み、当該スコアが少なくとも複数のcfDNAフラグメント末端及びそれらのゲノム上の場所の関数として決定され、さらに観察されたcfDNAフラグメントを生じる組織及び/又は細胞タイプの少なくともいくつかを決定する工程が当該スコアと1つ以上の参照マップを比較する工程を含む、請求項48から51のいずれかに記載の方法。

30

【請求項53】

座標に対するスコアが、座標がcfDNAフラグメント末端の位置である確率を表しているか又は前記と関係性がある、請求項52に記載の方法。

【請求項54】

参照マップが、DNase I高感受性サイトデータセット、RNA発現データセット、発現データ、染色体コンフォーメーションマップ、クロマチン接近容易性マップ、クロマチン断片化マップ、又は少なくとも1人の参照対象者から入手され、かつある疾患又は異常に付随する少なくとも1つの細胞タイプ又は組織に対応するサンプルから得られる配列データ、及び/又はある組織又は細胞タイプのヌクレオソーム及び/又はクロマトソームの位置又は間隔の取り方を含む、請求項49から53のいずれか1項に記載の方法。

40

【請求項55】

参照マップが、少なくとも1つの細胞タイプ又は組織のクロマチンを外因性ヌクレアーゼ(例えばマイクロコッカスヌクレアーゼ)で消化することによって作製される、請求項49から54のいずれか1項に記載の方法。

【請求項56】

参照マップが、転移による方法(例えばATAC-seq)を少なくとも1つの細胞タイプ又は組織の核又はクロマチンに適用することによって決定されたクロマチン接近容易性データ

50

を含む、請求項54又は請求項55に記載の方法。

【請求項57】

参照マップが、組織又は細胞タイプのDNA結合及び/又はDNA占拠タンパク質の位置と結びついたデータを含む、請求項49から56のいずれか1項に記載の方法。

【請求項58】

DNA結合及び/又はDNA占拠タンパク質が転写因子である、請求項57に記載の方法。

【請求項59】

位置が、架橋DNA-タンパク質複合体のクロマチン免疫沈澱を少なくとも1つの細胞タイプ又は組織に適用することによって決定される、請求項57又は請求項58に記載の方法。

【請求項60】

位置が、組織又は細胞タイプに付随するDNAをヌクレアーゼ（例えばDNase-I）で処理することによって決定される、請求項57又は請求項58に記載の方法。

【請求項61】

当該参照マップが、ある組織又は細胞タイプのヌクレオソーム、クロマトソーム、又は他のDNA結合若しくはDNA占拠タンパク質の位置又は間隔に関連する生物学的特徴を含む、請求項48から60のいずれか1項に記載の方法。

【請求項62】

生物学的特徴が1つ以上の遺伝子の定量的発現である、請求項61に記載の方法。

【請求項63】

生物学的特徴が1つ以上のヒストンマークの有無である、請求項61又は請求項62に記載の方法。

【請求項64】

生物学的特徴がヌクレアーゼ切断に対する高感受性である、請求項61から63のいずれか1項に記載の方法。

【請求項65】

参照マップを作製するために用いられる組織又は細胞タイプが、疾患又は異常を有する対象者の一次組織である、請求項49から64のいずれか1項に記載の方法。

【請求項66】

疾患又は異常が、癌、正常妊娠、妊娠（例えば異数性妊娠）の合併症、心筋梗塞、炎症性腸疾患、全身性自己免疫疾患、局所性自己免疫疾患、拒絶を伴う同種間移植、拒絶を伴わない同種間移植、卒中、及び局在性組織損傷から成る群から選択される、請求項65に記載の方法。

【請求項67】

マップを作製するために用いられる組織又は細胞タイプが、健康な対象者の一次組織である、請求項49から65のいずれか1項に記載の方法。

【請求項68】

参照マップを作製するために用いられる組織又は細胞タイプが不死化細胞株である、請求項49から65のいずれか1項に記載の方法。

【請求項69】

参照マップを作製するために用いられる組織又は細胞タイプが腫瘍の生検である、請求項49から65のいずれか1項に記載の方法。

【請求項70】

少なくとも1人の参照対象者から入手されるサンプルから得られる配列データがcfDNAフラグメント末端の位置の確率を含む、請求項54に記載の方法。

【請求項71】

参照対象者が健康である、請求項70に記載の方法。

【請求項72】

参照対象者が疾患又は異常を有する、請求項70に記載の方法。

【請求項73】

疾患又は異常が、癌、正常妊娠、妊娠（例えば異数性妊娠）の合併症、心筋梗塞、炎症

10

20

30

40

50

性腸疾患、全身性自己免疫疾患、局所性自己免疫疾患、拒絶を伴う同種間移植、拒絶を伴わない同種間移植、卒中、及び局在性組織損傷から成る群から選択される、請求項72に記載の方法。

【請求項74】

参照マップが、当該組織又は細胞タイプと結びついた参照ゲノムの少なくとも一部分についてcfDNAフラグメント末端確率を含む、請求項54から73のいずれか1項に記載の方法。

【請求項75】

参照マップが当該cfDNAフラグメント末端確率の数学的変換を含む、請求項74に記載の方法。

【請求項76】

cfDNAフラグメント末端確率が、当該組織又は細胞タイプの全ての参照ゲノム座標のサブセットを表す、請求項74に記載の方法。

【請求項77】

サブセットがヌクレオソーム及び/又はクロマトソームの位置又は間隔と結びついている、請求項76に記載の方法。

【請求項78】

サブセットが転写開始サイト及び/又は転写終了サイトと結びついている、請求項76又は請求項77に記載の方法。

【請求項79】

サブセットが少なくとも1つの転写因子の結合サイトと結びついている、請求項76から78のいずれか1項に記載の方法。

【請求項80】

サブセットがヌクレアーゼ高感受性サイトと結びついている、請求項76から79のいずれか1項に記載の方法。

【請求項81】

サブセットが、さらに加えて少なくとも1つのオルトゴナルな生物学的特徴と結びついている、請求項76から80のいずれか1項に記載の方法。

【請求項82】

オルトゴナルな生物学的特徴が高発現遺伝子に付随する、請求項81に記載の方法。

【請求項83】

オルトゴナルな生物学的特徴が低発現遺伝子に付随する、請求項81に記載の方法。

【請求項84】

数学的変換がフーリエ変換を含む、請求項75から83のいずれか1項に記載の方法。

【請求項85】

複数のcfDNAフラグメント末端スコアの少なくとも1つのサブセットが各々閾値を超えるスコアを有する、請求項52から84のいずれか1項に記載の方法。

【請求項86】

cfDNAを生じる組織及び/又は細胞タイプを、当該cfDNAフラグメント末端の少なくともいくつかの複数のゲノム上の場所の関数として決定する工程が、当該cfDNAフラグメント末端の少なくともいくつかの複数のゲノム上の場所のフーリエ変換、又はその数学的変換を参照マップと比較する工程を含む、請求項48から85のいずれか1項に記載の方法。

【請求項87】

参照マップが、疾患又は異常に付随する少なくとも1つの組織に対応するDNA又はクロマチン断片化データを含む、請求項48から86のいずれか1項に記載の方法。

【請求項88】

参照ゲノムがヒトに付随する、請求項48から87のいずれか1項に記載の方法。

【請求項89】

疾患又は異常を識別する記述を含む報告を作製する工程をさらに含む、請求項48から88のいずれか1項に記載の方法。

【請求項90】

10

20

30

40

50

報告が、単離cfDNAを生じると決定された組織及び/又は細胞タイプの一覧をさらに含む、請求項89に記載の方法。

【請求項91】

生物学的サンプルが、全血、末梢血血漿、尿又は脳脊髄液を含むか、本質的に前記から成るか、又は前記から成る、請求項1から90のいずれか1項に記載の方法。

【請求項92】

対象者におけるセルフリーDNA (cfDNA) を生じる組織及び/又は細胞タイプを決定する方法であって、

(i) 当該対象者の生物学的サンプルの入手、当該生物学的サンプルのcfDNAの単離、並びにcfDNAのライブラリー構築及び大量パラレルシーケンシングによる下記の分布(a)、(b)及び/又は(c)の測定によってヌクレオソームマップを作製する工程；

(ii) コントロール対象者又は既知疾患を有する対象者の生物学的サンプルの入手、当該生物学的サンプルのcfDNAの単離、cfDNAのライブラリー構築及び大量パラレルシーケンシングによる下記の分布(a)、(b)及び/又は(c)の測定によってヌクレオソームマップの参照セットを作製する工程；並びに

(iii) 当該cfDNAに由来するヌクレオソームマップをヌクレオソームマップの参照セットと比較することによって、当該cfDNAを生じる当該組織及び/又は細胞タイプを決定する工程を含み、

前記(a)、(b)及び(c)が、

(a) ヒトゲノムにおいて任意の特定の塩基対がcfDNAフラグメントの末端に出現する可能性の分布；

(b) ヒトゲノムの塩基対の任意の対がcfDNAフラグメントの末端の対として出現する可能性の分布；及び

(c) ヒトゲノムにおいて任意の特定の塩基対が弁別的なヌクレオソーム占拠の結果としてcfDNAフラグメントに出現する可能性の分布である、

前記方法。

【請求項93】

対象者におけるcfDNAを生じる組織及び/又は細胞タイプを決定する方法であって、

(i) 当該対象者の生物学的サンプルの入手、当該生物学的サンプルのcfDNAの単離、並びにcfDNAのライブラリー構築及び大量パラレルシーケンシングによる下記の分布(a)、(b)及び/又は(c)の測定によってヌクレオソームマップを作製する工程；

(ii) コントロール対象者又は既知疾患を有する対象者の生物学的サンプルの入手、当該生物学的サンプルのcfDNAの単離、マイクロココカスのヌクレアーゼ(MNase)、DNase処理又はATAC-Seqによるクロマチンの消化から誘導したDNAのライブラリー構築及び大量パラレルシーケンシングによる下記の分布(a)、(b)及び/又は(c)の測定によってヌクレオソームマップの参照セットを作製する工程；並びに

(iii) 当該cfDNAに由来するヌクレオソームマップをヌクレオソームマップの参照セットと比較することによって、当該cfDNAを生じる当該組織及び/又は細胞タイプを決定する工程を含み、

前記(a)、(b)及び(c)が、

(a) ヒトゲノムにおいて任意の特定の塩基対が配列決定されたフラグメントの末端に出現する可能性の分布；

(b) ヒトゲノムの塩基対の任意の対が配列決定されたフラグメントの末端の対として出現する可能性の分布；及び

(c) ヒトゲノムにおいて任意の特定の塩基対が弁別的なヌクレオソーム占拠の結果として配列決定されたフラグメントに出現する可能性の分布である、

前記方法。

【請求項94】

対象者において臨床症状を診断する方法であって、

(i) 当該対象者の生物学的サンプルの入手、当該生物学的サンプルのcfDNAの単離、並び

にcfDNAのライブラリー構築及び大量パラレルシーケンシングによる下記の分布(a)、(b)及び/又は(c)の測定によってヌクレオソームマップを作製する工程；

(ii) コントロール対象者又は既知疾患を有する対象者の生物学的サンプルの入手、当該生物学的サンプルのcfDNAの単離、cfDNAのライブラリー構築及び大量パラレルシーケンシングによる下記の分布(a)、(b)及び/又は(c)の測定によってヌクレオソームマップの参照セットを作製する工程；並びに

(iii) 当該cfDNAに由来するヌクレオソームマップをヌクレオソームマップの参照セットと比較することによって、当該臨床症状を決定する工程を含み、前記(a)、(b)及び(c)が、

(a) ヒトゲノムにおいて任意の特定の塩基対がcfDNAフラグメントの末端に出現する可能性の分布；

(b) ヒトゲノムの塩基対の任意の対がcfDNAフラグメントの末端の対として出現する可能性の分布；及び

(c) ヒトゲノムにおいて任意の特定の塩基対が弁別的なヌクレオソーム占拠の結果としてcfDNAフラグメントに出現する可能性の分布である、前記方法。

【請求項95】

対象者において臨床症状を診断する方法であって、

(i) 当該対象者の生物学的サンプルの入手、当該生物学的サンプルのcfDNAの単離、並びにcfDNAのライブラリー構築及び大量パラレルシーケンシングによる下記の分布(a)、(b)及び/又は(c)の測定によってヌクレオソームマップを作製する工程；

(ii) コントロール対象者又は既知疾患を有する対象者の生物学的サンプルの入手、当該生物学的サンプルのcfDNAの単離、マイクロココカスのヌクレアーゼ(MNase)、DNase処理又はATAC-Seqによるクロマチンの消化から誘導したDNAのライブラリー構築及び大量パラレルシーケンシングによる下記の分布(a)、(b)及び/又は(c)の測定によってヌクレオソームマップの参照セットを作製する工程；並びに

(iii) 当該cfDNAに由来するヌクレオソームマップをヌクレオソームマップの参照セットと比較することによって、当該cfDNAの供給源組織の組成を決定する工程を含み、前記(a)、(b)及び(c)が、

(a) ヒトゲノムにおいて任意の特定の塩基対が配列決定されたフラグメントの末端に出現する可能性の分布；

(b) ヒトゲノムの塩基対の任意の対が配列決定されたフラグメントの末端の対として出現する可能性の分布；及び

(c) ヒトゲノムにおいて任意の特定の塩基対が弁別的なヌクレオソーム占拠の結果として配列決定されたフラグメントに出現する可能性の分布である、前記方法。

【請求項96】

ヌクレオソームマップが、

生物学的サンプルから単離されたcfDNAを精製する工程、

アダプター連結及び場合によってPCR増幅によってライブラリーを構築する工程、並びに、

生じたライブラリーを配列決定する工程、

によって作製される、請求項92 - 95のいずれか1項に記載の方法。

【請求項97】

ヌクレオソームマップの参照セットが、

コントロール対象者の生物学的サンプルから単離されたcfDNAを精製する工程、

アダプター連結及び場合によってPCR増幅によってライブラリーを構築する工程、並びに、

生じたライブラリーを配列決定する工程、

によって作製される、請求項92 - 95のいずれか1項に記載の方法。

【請求項98】

10

20

30

40

50

分布 (a)、(b) 若しくは (c) 又はこれらの分布の1つの数学的変換を連続ウィンドウでフーリエ変換に付し、続いてヌクレオソーム占拠に付随した周期レンジについて強度を定量して、ヌクレオソームが各連続ウィンドウ内で構造化された配置を示す範囲を集約する、請求項92 - 95のいずれか1項に記載の方法。

【請求項99】

分布 (a)、(b) 若しくは (c) 又はこれらの分布の1つの数学的変換で、特定の転写因子 (TF) の転写因子結合サイト (TFBS) の直近にシーケンシング読取り開始位置がマップされる参照ヒトゲノムでサイトの分布を定量し (TFがTFBSと結合するときはしばしば当該サイトの直ぐ側にヌクレオソームがフランキングする)、cfDNAに寄与する細胞タイプのTF活性の結果としてのヌクレオソームの配置を集約する、請求項92 - 95のいずれか1項に記載の方法。

10

【請求項100】

ヌクレオソーム占拠シグナルが、他のゲノム上の目印 (例えばDNase I高感受性サイト、転写開始サイト、トポロジカルドメイン、他の後成的な目印、又は他のデータセット (例えば遺伝子発現など) での相関的動態によって規定される全てのそのようなサイトのサブセット) 周辺で、分布 (a)、(b) 及び/又は (c) 又はこれらの分布の1つの数学的変換から、集合化シグナルの任意の1つにしたがって集約される、請求項92 - 95のいずれか1項に記載の方法。

【請求項101】

分布を変換して、当該ゲノムの多様なサブセット内のヌクレオソームの配置の周期的シグナルを集合化又は集約する、例えば、連続ウィンドウでの周期性、或いはまた別に、転写因子結合サイト、遺伝子モデル特徴 (例えば転写開始サイト)、組織発現データ又は他のヌクレオソーム配置に相関するものによって規定されるゲノムの不連続サブセットでの周期性を定量する、請求項92 - 95のいずれか1項に記載の方法。

20

【請求項102】

分布が、組織特異的データ、すなわち組織特異的DNase I高感受性サイト周辺の集合シグナルによって規定される、請求項92 - 95のいずれか1項に記載の方法。

【請求項103】

追加のヌクレオソームマップを当該参照セットと比較するために統計的シグナルをプロセッシングする工程をさらに含む、請求項92 - 95のいずれか1項に記載の方法。

30

【請求項104】

最初に多岐にわたるサンプルセットで当該ゲノムの端から端まで連続ウィンドウ内で長レンジヌクレオソーム順番付けを集約し、続いて主要成分分析 (PCA) を実施して、サンプルをクラスター化するか又は混合割合を概算する、請求項103に記載の方法。

【請求項105】

臨床症状が癌、すなわち悪性腫瘍である、請求項94又は請求項95に記載の方法。

【請求項106】

生物学的サンプルがcfDNAを含む循環血漿であり、そのいくらかの部分が腫瘍に由来する、請求項105に記載の方法。

【請求項107】

臨床症状が、組織損傷、心筋梗塞 (心臓組織の急性損傷)、自己免疫疾患 (多様な組織の慢性損傷)、妊娠、染色体異常 (例えばトリソミー)、及び移植片拒絶から選択される、請求項94又は請求項95に記載の方法。

40

【請求項108】

cfDNAに寄与すると決定された1つ以上の組織又は細胞タイプの各々に割合を割り当てる工程をさらに含む、請求項1から107のいずれか1項に記載の方法。

【請求項109】

1つ以上の決定された組織又は細胞タイプの各々に割り当てられた割合が、少なくとも部分的には、1人又は複数の健康な対象者のcfDNAと対比された相関性の程度又は相関性増加の程度を基準にする、請求項108に記載の方法。

50

【請求項110】

相関性の程度が、少なくとも部分的には、当該生物学的サンプル由来のcfDNAフラグメント末端の分布の数学的変換と当該決定された組織又は細胞タイプに付随する参照マップとの比較に基づき、請求項108又は請求項109に記載の方法。

【請求項111】

1つ以上の決定された組織又は細胞タイプの各々に割り当てられた割合が混合モデルに基づく、請求項108から110のいずれか1項に記載の方法。

【発明の詳細な説明】

【技術分野】

【0001】

10

(優先権の主張)本出願は、米国仮特許出願No.62/029,178(2014年7月25日出願)及びNo.62/087,619(2014年12月4日出願)に対し優先権を主張する(前記の各々の要旨は、あたかも前記が本明細書で完全に説明されたかのように参照により本明細書に含まれる)。

(政府の利益に関する記述)本発明は、アメリカ国立衛生研究所(NIH)のグラント番号1DP1HG007811により政府の支援を受けて達成された。政府は本発明において一定の権利を有する。

(技術分野)

本開示は、セルフリーDNAを生じる1つ以上の組織及び/又は細胞タイプを決定する方法に関する。いくつかの実施態様では、本開示は、ある対象者の疾患又は異常を、当該対象者に由来する生物学的サンプルのセルフリーDNAに付随すると決定された1つ以上の組織及び/又は細胞タイプの関数と認定する方法を提供する。

20

【背景技術】

【0002】

セルフリーDNA("cfDNA")は人間の循環血漿、尿及び他の体液に存在する。cfDNAは二本鎖DNAフラグメントを含む。前記フラグメントは比較的短く(もっぱら200塩基対未満)、通常は低濃度で存在する(例えば血漿中に1-100ng/ μ L)。健康な個体の循環血漿では、cfDNAは主として血液細胞(すなわち造血細胞系列の正常な細胞)のアポトーシスに由来すると考えられている。しかしながら、特殊な状況では、他の組織が、体液(例えば循環血漿)中のcfDNAの組成に実質的に寄与することができる。

cfDNAはある種の専門分野(例えば生殖医療、癌診断及び移植医療)で用いられてきたが、cfDNAに基づく既存の検査は、2つ以上の細胞集団間(例えば母体ゲノムに対して胎児ゲノム、正常ゲノムに対して癌ゲノム、移植片レシピエントゲノムに対してドナーゲノムなど)の遺伝子型(例えば個々の配列の一次配列又は表示コピー数)の相違に着目する。残念ながら、任意の与えられた生物学的サンプルで見いだされるcfDNAフラグメントの圧倒的多数が寄与細胞集団間で配列が同一のゲノム領域に由来するために、既存のcfDNA系検査はその適用範囲が極めて制限される。加えて、多くの疾患及び異常が、cfDNAを生じる組織及び/又は細胞タイプで、当該疾患又は異常が関与する例えば組織損傷又は炎症性プロセスに由来する変化を随伴する。2つのゲノム間で個々の配列の一次配列又は表示コピー数の相違に着目する既存のcfDNA系診断検査は、そのような変化を検出することができない。したがって、強力な非生検型診断方法を提供するcfDNAの潜在能力は計り知れないが、極めて多様な疾患及び異常の診断に適用できるcfDNA系診断方法論に対する要請は今なお続いている。

30

40

【発明の概要】

【0003】

本開示は、対象者からの生物学的サンプルでセルフリーDNA("cfDNA")を生じる1つ以上の組織及び/又は細胞タイプを決定する方法を提供する。いくつかの実施態様では、本開示は、ある対象者の疾患又は異常を、当該対象者に由来する生物学的サンプルのcfDNAと付随すると決定された1つ以上の組織及び/又は細胞タイプの関数と認定する方法を提供する。

いくつかの実施態様では、本開示は、対象者においてセルフリーDNA(cfDNA)を生じる

50

組織及び/又は細胞タイプを決定する方法を提供し、前記方法は以下の工程を含む：当該対象者からの生物学的サンプルからcfDNAを単離する工程（当該単離cfDNAは複数のcfDNAフラグメントを含む）；当該複数のcfDNAフラグメントの少なくとも一部分に付随する配列を決定する工程；当該複数のcfDNAフラグメントの少なくともいくつかのcfDNAフラグメント末端の参照ゲノム内のゲノム上の場所を当該cfDNAフラグメント配列の関数として決定する工程；及び当該cfDNAフラグメントを生じる組織及び/又は細胞タイプの少なくともいくつかを、当該cfDNAフラグメント末端の少なくともいくつかのゲノム上の場所の関数として決定する工程。

他の実施態様では、本開示は対象者における疾患又は異常を識別する方法を提供し、前記方法は以下の工程を含む：当該対象者からの生物学的サンプルからセルフリーDNA(cfDNA A)を単離する工程（当該単離cfDNAは複数のcfDNAフラグメントを含む）；当該複数のcfDNA Aフラグメントの少なくとも一部分に付随する配列を決定する工程；当該複数のcfDNAフラグメントの少なくともいくつかのcfDNAフラグメント末端の参照ゲノム内のゲノム上の場所を当該cfDNAフラグメント配列の関数として決定する工程；当該cfDNAを生じる組織及び/又は細胞タイプの少なくともいくつかを、当該cfDNAフラグメント末端の少なくともいくつかのゲノム上の場所の関数として決定する工程；及び当該疾患又は異常を、当該cfDNAを生じると決定された組織及び/又は細胞タイプの関数として同定する工程。

【0004】

他の実施態様では、本開示は対象者でセルフリーDNA(cfDNA)を生じる組織及び/又は細胞タイプを決定する方法を提供し、前記方法は以下の工程を含む：(i) 当該対象者の生物学的サンプルの入手、当該生物学的サンプルのcfDNAの単離、並びにcfDNAのライブラリー構築及び大量パラレルシーケンシング(massively parallel sequencing)による下記に示す分布(a)、(b)及び/又は(c)の測定によってヌクレオソームマップを作製する工程；(ii) コントロール対象者又は既知疾患を有する対象者の生物学的サンプルの入手、当該生物学的サンプルのcfDNAの単離、cfDNAのライブラリー構築及び大量パラレルシーケンシングによる下記に示す分布(a)、(b)及び/又は(c)の測定によってヌクレオソームマップの参照セットを作製する工程；並びに(iii) 当該生物学的サンプルのcfDNAに由来するヌクレオソームマップをヌクレオソームマップの参照セットと比較することによって、当該生物学的サンプルのcfDNAを生じる組織及び/又は細胞タイプを決定する工程。ここで前記(a)、(b)及び(c)は以下のとおりである：(a) ヒトゲノムにおいて任意の特定の塩基対がcfDNAフラグメントの末端に出現する可能性の分布；(b) ヒトゲノムの塩基対の任意の対がcfDNAフラグメントの末端の対として出現する可能性の分布；及び(c) ヒトゲノムにおいて任意の特定の塩基対が弁別的なヌクレオソーム占拠の結果としてcfDNAフラグメントに出現する可能性の分布。

【0005】

さらに他の実施態様では、本開示は対象者におけるcfDNAを生じる組織及び/又は細胞タイプを決定する方法を提供し、前記方法は以下の工程を含む：(i) 当該対象者からの生物学的サンプルの入手、当該生物学的サンプルのcfDNAの単離、並びにcfDNAのライブラリー構築及び大量パラレルシーケンシングによる下記に示す分布(a)、(b)及び/又は(c)の測定によってヌクレオソームマップを作製する工程；(ii) コントロール対象者又は既知疾患を有する対象者の生物学的サンプルの入手、当該生物学的サンプルのcfDNAの単離、酵素(例えばマイクロココカスのヌクレアーゼ、DNase又はトランスポザーゼ)によるクロマチン断片化から誘導したDNAのライブラリー構築及び大量パラレルシーケンシングによる下記に示す分布(a)、(b)及び/又は(c)の測定によってヌクレオソームマップの参照セットを作製する工程；並びに(iii) 当該生物学的サンプルのcfDNAに由来するヌクレオソームマップをヌクレオソームマップの参照セットと比較することによって、当該生物学的サンプルのcfDNAを生じる組織及び/又は細胞タイプを決定する工程。ここで前記(a)、(b)及び(c)は以下のとおりである：(a) ヒトゲノムにおいて任意の特定の塩基対が配列決定されたフラグメントの末端に出現する可能性の分布；(b) ヒトゲノムの塩基対の任意の対が配列決定されたフラグメントの末端の対として出現する可能性の分

10

20

30

40

50

布；及び(c) ヒトゲノムにおいて任意の特定の塩基対が弁別的なヌクレオソーム占拠の結果として配列決定されたフラグメントに出現する可能性の分布。

【0006】

他の実施態様では、本開示は対象者で臨床症状を診断する方法を提供し、前記方法は以下の工程を含む：(i) 当該対象者の生物学的サンプルの入手、当該生物学的サンプルのcfDNAの単離、並びにcfDNAのライブラリー構築及び大量パラレルシーケンシングによる下記の分布(a)、(b)及び/又は(c)の測定によってヌクレオソームマップを作製する工程；(ii) コントロール対象者又は既知疾患を有する対象者の生物学的サンプルの入手、当該生物学的サンプルのcfDNAの単離、cfDNAのライブラリー構築及び大量パラレルシーケンシングによる下記の分布(a)、(b)及び/又は(c)の測定によってヌクレオソームマップの参照セットを作製する工程；並びに(iii) 当該生物学的サンプルのcfDNAに由来するヌクレオソームマップをヌクレオソームマップの参照セットと比較することによって、当該臨床症状を決定する工程。ここで前記(a)、(b)及び(c)は以下のとおりである：(a) ヒトゲノムにおいて任意の特定の塩基対がcfDNAフラグメントの末端に出現する可能性の分布；(b) ヒトゲノムの塩基対の任意の対がcfDNAフラグメントの末端の対として出現する可能性の分布；及び(c) ヒトゲノムにおいて任意の特定の塩基対が弁別的なヌクレオソーム占拠の結果としてcfDNAフラグメントに出現する可能性の分布。

10

【0007】

他の実施態様では、本開示は対象者で臨床症状を診断する方法を提供し、前記方法は以下の工程を含む：(i) 当該対象者の生物学的サンプルの入手、当該生物学的サンプルのcfDNAの単離、並びにcfDNAのライブラリー構築及び大量パラレルシーケンシングによる下記の分布(a)、(b)及び/又は(c)の測定によってヌクレオソームマップを作製する工程；(ii) コントロール対象者又は既知疾患を有する対象者の生物学的サンプルの入手、当該生物学的サンプルのcfDNAの単離、酵素(例えばマイクロコッカスのヌクレアーゼ(MNase)、DNase又はトランスポザーゼ)によるクロマチン断片化から誘導したDNAのライブラリー構築及び大量パラレルシーケンシングによる下記の分布(a)、(b)及び/又は(c)の測定によってヌクレオソームマップの参照セットを作製する工程；並びに(iii) 当該生物学的サンプルのcfDNAに由来するヌクレオソームマップをヌクレオソームマップの参照セットと比較することによって、当該生物学的サンプルの供給源組織(tissue-of-origin)の組成を決定する工程。ここで前記(a)、(b)及び(c)は以下のとおりである：(a) ヒトゲノムにおいて任意の特定の塩基対が配列決定されたフラグメントの末端に出現する可能性の分布；(b) ヒトゲノムの塩基対の任意の対が配列決定されたフラグメントの末端の対として出現する可能性の分布；及び(c) ヒトゲノムにおいて任意の特定の塩基対が弁別的なヌクレオソーム占拠の結果として配列決定されたフラグメントに出現する可能性の分布。

20

30

これらの実施態様及び他の実施態様は下記ではるかに詳細に記載される。

【図面の簡単な説明】

【0008】

【図1A - C】cfDNA断片化パターンをヌクレオソーム占拠に関連付ける3つのタイプの情報を示す(小さなゲノム領域について例示されている)。これらと同じタイプの情報はまた、酵素(例えばマイクロコッカスのヌクレアーゼ(MNase)、DNase又はトランスポザーゼ)によるクロマチンの断片化によってももたらされうる。図1Aは、ヒトゲノムにおいて任意の特定の塩基対が配列決定されたフラグメントの末端(すなわち断片化点)に出現する可能性の分布を示す。図1Bは、ヒトゲノムの塩基対の任意の対が配列決定されたフラグメントの末端の対として出現する可能性の分布を示す(すなわち、個々の分子を生じる断片化点の連続する対)。図1Cは、ヒトゲノムにおいて任意の特定の塩基対が、配列決定されたフラグメント内で、弁別的なヌクレオソーム占拠の結果として出現する可能性の分布である(すなわち相対的カバレッジ)。

40

【図2】典型的なcfDNAシーケンシングライブラリーの挿入サイズ分布を示し、ここでは人数が明らかでない健康な個体からの提供を含むヒト血漿に由来するcfDNAプールについ

50

て示されている（バルクcfDNA）。

【図3A】全cfDNAサンプル（血漿）、腫瘍患者サンプルのcfDNA（腫瘍）、妊婦個体のcfDNA（妊娠）、種々のヒト細胞株のヒトMNase（細胞株）及びヒトDNAショットガンシーケンスライブラリー（ショットガン）にわたって、第一の（chr1）ヒト常染色体にマッピングした読取り開始座標の高速フーリエ変換（FFT）の平均ピリオドグラム強度を示す。

【図3B】全cfDNAサンプル（血漿）、腫瘍患者サンプルのcfDNA（腫瘍）、妊婦個体のcfDNA（妊娠）、種々のヒト細胞株のヒトMNase（細胞株）及びヒトDNAショットガンシーケンスライブラリー（ショットガン）にわたって、最後の（chr22）ヒト常染色体にマッピングした読取り開始座標の高速フーリエ変換（FFT）の平均ピリオドグラム強度を示す。

10

【図4A】全常染色体の10キロ塩基対（kbp）ブロックでの196塩基対（bp）周期性における強度の最初の3主要成分（PC）を示す。図4AはPC2対PC1を示し、図4BはPC3対PC2を示す。

【図4B】全常染色体の10キロ塩基対（kbp）ブロックでの196塩基対（bp）周期性における強度の最初の3主要成分（PC）を示す。図4AはPC2対PC1を示し、図4BはPC3対PC2を示す。

【図5】全常染色体を通して10kbpブロックで196bp周期性により測定された強度のユークリッド距離の階層クラスター化樹形図を示す。

【図6A】全常染色体を通して10kbpブロックでの181bpから202bp周期性における強度の最初の3主要成分を示す。図6AはPC2対PC1を示し、図6BはPC3対PC2を示す。

20

【図6B】全常染色体を通して10kbpブロックでの181bpから202bp周期性における強度の最初の3主要成分を示す。図6AはPC2対PC1を示し、図6BはPC3対PC2を示す。

【図7】全常染色体を通して10kbpブロックで181bpから202bp周期性により測定された強度のユークリッド距離の階層クラスター化樹形図を示す。

【図8A - D】cfDNAデータセットのための全常染色体を通して10kbpブロックでの181bpから202bp周期性における強度の主要成分分析（10PCの最初の7つ）を示す。図8AはPC2対PC1を示し、図8BはPC3対PC2を示し、図8CはPC4対PC3を示し、図8DはPC5対PC4を示す。

【図8E - F】cfDNAデータセットのための全常染色体を通して10kbpブロックでの181bpから202bp周期性における強度の主要成分分析（10PCの最初の7つ）を示す。図8EはPC6対PC5を示し、図8FはPC7対PC6を示す。

30

【図9A - B】MNaseデータセットのための全常染色体を通して10kbpブロックでの181bpから202bp周期性における強度の主要成分分析を示す。図9AはPC2対PC1を示し、図9BはPC3対PC2を示す。

【図9C - E】MNaseデータセットのための全常染色体を通して10kbpブロックでの181bpから202bp周期性における強度の主要成分分析を示す。図9CはPC4対PC3を示し、図9DはPC5対PC4を示し、図9EはPC6対PC5を示す。

【図10】全ての合成cfDNA及びMNaseデータセット混合物の代表的ヒト常染色体（chr11）の平均ピリオドグラム強度を示す。

【図11】合成MNaseデータセット混合物について全常染色体を通して10kbpブロックでの181bpから202bp周期性における強度の最初の2主要成分を示す。

40

【図12】合成cfDNAデータセット混合物について全常染色体を通して10kbpブロックでの181bpから202bp周期性における強度の最初の2主要成分を示す。

【図13】合成MNase及びcfDNA混合物データセットのための全常染色体の10kbpブロックでの181bpから202bp周期性における強度のユークリッド距離の階層クラスター化樹形図を示す。

【図14】少なくとも100Mの読取り（リード）を有するサンプルセットの23,666 TCF結合サイト周辺の1kbpウィンドウにおける読取り開始点密度を示す。

【図15】少なくとも100Mの読取りを有するサンプルセットの5,644 c-Jun結合サイト周辺の1kbpウィンドウにおける読取り開始点密度を示す。

50

【図16】少なくとも100Mの読取りを有するサンプルセットの4,417 NF-YB結合サイト周辺の1kbpウィンドウにおける読取り開始点密度を示す。

【図17】cfDNAフラグメント末端を生じる過程の模式図を示す。アポトーシス及び/又は壊死による細胞死はネイティブなクロマチンのほぼ完全な消化をもたらす。タンパク質結合DNAフラグメント(典型的にはヒストン又は転写因子と結合している)は、優先的に消化に耐え循環中に放出されるが、裸のDNAは消失する。フラグメントは、プロテアーゼ処理に続いて末梢血の血漿から回収できる。健康な個体では、cfDNAは主として骨髄系及びリンパ系細胞系列に由来するが、ある種の医学的症状では1つ以上の追加の組織の寄与が存在しうる。

【図18】伝統的シーケンシングライブラリー調製で観察されるcfDNAのフラグメントの長さを示す。長さはペア形成末端シーケンシング読取りのアラインメントから推論される。フラグメントの長さが167塩基対(bp)(緑色破線)の再現性を有するピークはクロマトソームとの結合と一致する。付加的ピークは~10.4bpの周期性を立証し、ヌクレオソームコアにおけるDNAのらせんピッチと一致する。ライブラリー調製時の酵素による末端修復は5'及び3'オーバーハングを除去し、真の切断サイトを曖昧にする可能性がある。

【図19】伝統的ライブラリーの167bpフラグメント及びフランキングするゲノム配列のジヌクレオチド組成を示す。BH01ライブラリーで観察されるジヌクレオチド頻度をシミュレーションしたフラグメントの予想される周期と比較した(切断及びアダプター連結の優先性の両方から生じる末端偏向と適合する)。

【図20】cfDNAフラグメントの一本鎖ライブラリー調製プロトコルの模式図を示す。

【図21】一本鎖シーケンシングライブラリー調製で観察されるcfDNAのフラグメントの長さを示す。ライブラリー調製時に鋳型分子に対して酵素による末端修復は実施されない。50-120bpの短いフラグメントが伝統的ライブラリーと比較して高度に濃縮されている。~10.4bp周期性が存続するが、その相は~3bpシフトしている。

【図22】一本鎖ライブラリーの167bpフラグメント及びフランキングするゲノム配列のジヌクレオチド組成を示す。IH02ライブラリーで観察されるジヌクレオチド頻度をシミュレーションしたフラグメントの予想される頻度と比較したところ、再び末端偏向と適合する。BH01とIH02との間の偏向のバックグラウンドレベルにおける見かけの相違は、真のライブラリーというよりはシミュレーションしたライブラリー間の相違と関係する(データは示されていない)。

【図23A-B】伝統的プロトコルで調製された代表的なcfDNAシーケンシングライブラリーのゲル画像を示す。

【図24A】cfDNAフラグメントのモノヌクレオチド切断偏向を示す。

【図24B】cfDNAフラグメントのジヌクレオチド切断偏向を示す。

【図25】ヌクレオソーム配置の推論の模式図を示す。塩基毎のウィンドウ化保護スコア(per-base windowed protection score, WPS)は、120bpウィンドウ内部にあるフラグメント末端の数を、完全にこのウィンドウにまたがるフラグメントの数から差し引くことによって計算される。高WPS値は消化からのDNAの保護の増加を示し、低い値はDNAが保護されないことを示す。ピークコールは上昇WPSの連続領域を識別する。

【図26】よく研究されたアルファサテライトアレイで強力的に位置決めされたヌクレオソームを示す。サンプルCH01のカパレッジ、フラグメント末端及びWPS値が、染色体12の動原体周辺遺伝子座の長フラグメント(120bpウィンドウ; 120-180bp読取り)又は短フラグメント(16bpウィンドウ; 35-80bp読取り)bin(ピン)について示されている。CH01のヌクレオソームコール(中央、青色の棒)は遺伝子座を通して規則的な間隔を有する。公表された2つの研究のMNase消化に基づくヌクレオソームコール(中央、紫色及び黒色の棒)もまた示されている。遺伝子座は、注釈を有するアルファサテライトアレイとオーバーラップする。

【図27】染色体9のDNase I高感受性サイト(DHS)周辺の推論されるヌクレオソーム配置を示す。サンプルCH01のカパレッジ、フラグメント末端及びWPS値が長及び短フラグメントピンについて示されている。高感受性領域(灰色で強調されている)は、長フラグメ

10

20

30

40

50

ントbinのカバレッジの低下によって特徴づけられる。DHSに隣接するCH01のヌクレオソームコール（中央、青色枠）は典型的な隣接対よりも間隔が広く、調節タンパク質（転写因子を含む）に対する介在配列の接近容易性と一致する。より短いフラグメント（前記はそのようなタンパク質と結合できる）のカバレッジはDHSで増加し、前記は、いくつかの注釈付き転写因子結合サイトとオーバーラップする（データは示されていない）。公表された2つの研究のMNase消化に基づくヌクレオソームコールが、図26のように示されている。

【図28】本発明のある実施態様のピークコール及びスコア付けの模式図を示す。

【図29】GC含有量によるCH01ピーク密度を示す。

【図30】サンプルごとの隣接ピーク間の距離のヒストグラムを示す。ピークコールから隣接コールの距離が測定される。

【図31】サンプル間のピークコールの比較を示す。サンプルの各対について、ピークがより少ないサンプルの各ピークコールと他のサンプルの最も近いピークコールとの間の距離を計算し、1のbinサイズのヒストグラムとして可視化する。負の数は最も近いピークが上流にあり、正の数は最も近いピークが下流にあることを示す。

【図32】サンプル間でのピークコールの比較を示す。図32AはIH01対BH01を示し、図32BはIH02対BH01を示し、図32CはIH02対IH01を示す。

【図33A】真のピーク対シミュレーションピークのヌクレオソームスコアを示す。

【図33B】あるスコアbin内の当該スコアbinの関数としてのピークオフセットの中央値（左y軸）及び各スコアbinのピークの数（右y軸）を示す。

【図34A-C】サンプルと適合シミュレーション結果間でのピークコールの比較を示す。図34AはBH01シミュレーション対BH01実在物を示し、図34BはIH01シミュレーション対BH01実在物を示し、図34CはIH02シミュレーション対IH01実在物を示す。

【図35】サンプルCH01の隣接ピーク間の距離を示す。黒色の点線は分布の態様を示す（185bp）。

【図36】22,626の転写開始サイト（TSS）周辺の集合、調整ウインドウ化保護スコア（WPS；120bpウインドウ）を示す。TSSは、転写の鎖及び方向について調整した後0位でアラインメントされる。集合WPSは、真のデータ及びシミュレーションデータの両方について中心TSSに対する各位置のTSS毎のWPSを合計することによって作表される。プロットされた値は、真の集合WPSとシミュレーション集合WPS間の相違を表す（前記は下記でより詳しく述べるようにさらに調整される）。より高いWPS値は切断からの優先的な保護を示す。

【図37】22,626の転写開始サイト（TSS）周辺の集合、調整ウインドウ化保護スコア（WPS；120bpウインドウ）を示す。TSSは、転写の鎖及び方向について調整した後0位でアラインメントされる。集合WPSは、真のデータ及びシミュレーションデータの両方について中心TSSに対する各位置のTSS毎のWPSを合計することによって作表される。プロットされた値は、真の集合WPSとシミュレーション集合WPS間の相違を表す（前記は下記でより詳しく述べるようにさらに調整される）。より高いWPS値は切断からの優先的な保護を示す。図37は、22,626開始コドン周辺の集合、調整WPSを示す。

【図38】22,626の転写開始サイト（TSS）周辺の集合、調整ウインドウ化保護スコア（WPS；120bpウインドウ）を示す。TSSは、転写の鎖及び方向について調整した後0位でアラインメントされる。集合WPSは、真のデータ及びシミュレーションデータの両方について中心TSSに対する各位置のTSS毎のWPSを合計することによって作表される。プロットされた値は、真の集合WPSとシミュレーション集合WPS間の相違を表す（前記は下記でより詳しく述べるようにさらに調整される）。より高いWPS値は切断からの優先的な保護を示す。図38は、224,910スプライシングドナーサイト周辺の集合、調整WPSを示す。

【図39】22,626の転写開始サイト（TSS）周辺の集合、調整ウインドウ化保護スコア（WPS；120bpウインドウ）を示す。TSSは、転写の鎖及び方向について調整した後0位でアラインメントされる。集合WPSは、真のデータ及びシミュレーションデータの両方について中心TSSに対する各位置のTSS毎のWPSを合計することによって作表される。プロットされた値は、真の集合WPSとシミュレーション集合WPS間の相違を表す（前記は下記でより詳しく

10

20

30

40

50

く述べるようにさらに調整される)。より高いWPS値は切断からの優先的な保護を示す。
 図39は、224,910スプライシングアクセプターサイト周辺の集合、調整WPSを示す。

【図40】図40は、CH01のデータ(真のデータ、適合シミュレーションデータ及びそれらの相違を含む)に関する多様な遺伝的特徴周辺の集合、調整WPSを示す。

【図41】A/B区画におけるヌクレオソームの間隔を示す。非オーバーラップ100キロ塩基(kb) bin(各々~500ヌクレオソームコールを含む)におけるヌクレオソームの間隔の中央値をゲノムワイドで計算する。GM12878についてのA/B区画予測(前記もまた100kb解像度による)は公開情報源による。区画Aはオープンクロマチンと結合し、区画Bは閉鎖クロマチンと結合する。

【図42】染色体7及び11のヌクレオソームの間隔及びA/B区画を示す。A/Bセグメント化(赤色及び青色の棒線)は主として染色体G-バンド形成を概括する(表意記号、灰色棒線)。ヌクレオソームの間隔の取り方の中央値(黒色ドット)は100kb binで計算され、A/Bセグメントの上部にプロットされている。

【図43】長画分(上)及び短画分(下)について93,550CTCFサイトの集合、調整WPSを示す。

【図44】CTCFサイトの短画分cfDNAの集合、調整WPSの拡大図を示す。淡赤色棒線(及びプロット内の対応する陰影)は、公知の52bp CTCF結合モチーフの位置を示す。この棒線の暗赤色部分は、FIMOモチーフ検索に用いられた17bpモチーフの場所を示す。

【図45】クラスター化FIMO予測CTCFサイト(純粋にモチーフを根拠とする:518,632サイト)、これらの予測のうちENCODE ChIP-seqピークとオーバーラップするサブセット(93,530サイト)、及び19細胞株にわたって活性であることが実験的に観察されたさらに別のサブセット(23,723サイト)に由来するCTCFサイト周辺で計算された-1から+1のヌクレオソームの間隔の取り方を示す。CTCFサイトのもっともストリンジェンシーが低いセットは、もっぱらゲノムワイド平均(~190bp)とほぼ同じ距離で隔てられている。しかしながら、最高のストリンジェンシーでは、ほとんどのCTCFサイトがはるかに広い距離(~260bp)で隔てられ、活性なCTCF結合及び隣接ヌクレオソーム再配置と一致する。

【図46】ヌクレオソームにフランキングするCTCF占拠再配置を示す。図46は、FIMOによって予測される518,632のCTCF結合サイトについての3つの最も近い上流のピークコール及び3つの最も近い下流のピークコールのピーク間距離を示す。図47は、図46のように、FIMOによって予測される518,632のCTCF結合サイトについての3つの最も近い上流のピークコール及び3つの最も近い下流のピークコールのピーク間距離を示すが、ただしここで同じCTCFサイトセットがENCODE ChIP-seqピークとのオーバーラップに基づいてフィルター処理されて、93,530サイトが残された。図48は、図47のように、FIMOによって予測される93,530のCTCF結合サイトについての3つの最も近い上流のピークコール及び3つの最も近い下流のピークコールのピーク間距離を示すが、ただしここでCTCFサイトセットは19細胞株にわたって実験的に観察される活性なCTCFサイトセットとのオーバーラップに基づいてフィルター処理されて、23,732サイトが残された。

【図47】ヌクレオソームにフランキングするCTCF占拠再配置を示す。図46は、FIMOによって予測される518,632のCTCF結合サイトについての3つの最も近い上流のピークコール及び3つの最も近い下流のピークコールのピーク間距離を示す。図47は、図46のように、FIMOによって予測される518,632のCTCF結合サイトについての3つの最も近い上流のピークコール及び3つの最も近い下流のピークコールのピーク間距離を示すが、ただしここで同じCTCFサイトセットがENCODE ChIP-seqピークとのオーバーラップに基づいてフィルター処理されて、93,530サイトが残された。図48は、図47のように、FIMOによって予測される93,530のCTCF結合サイトについての3つの最も近い上流のピークコール及び3つの最も近い下流のピークコールのピーク間距離を示すが、ただしここでCTCFサイトセットは19細胞株にわたって実験的に観察される活性なCTCFサイトセットとのオーバーラップに基づいてフィルター処理されて、23,732サイトが残された。

【図48】ヌクレオソームにフランキングするCTCF占拠再配置を示す。図46は、FIMOによって予測される518,632のCTCF結合サイトについての3つの最も近い上流のピークコール及

10

20

30

40

50

び3つの最も近い下流のピークコールのピーク間距離を示す。図47は、図46のように、FIMOによって予測される518,632のCTCF結合サイトについての3つの最も近い上流のピークコール及び3つの最も近い下流のピークコールのピーク間距離を示すが、ただしここで同じCTCFサイトセットがENCODE ChIP-seqピークとのオーバーラップに基づいてフィルター処理されて、93,530サイトが残された。図48は、図47のように、FIMOによって予測される93,530のCTCF結合サイトについての3つの最も近い上流のピークコール及び3つの最も近い下流のピークコールのピーク間距離を示すが、ただしここでCTCFサイトセットは19細胞株にわたって実験的に観察される活性なCTCFサイトセットとのオーバーラップに基づいてフィルター処理されて、23,732サイトが残された。

【図49】広い間隔(230-270bp)が置かれたフランキングヌクレオソームを有する推定的CTCFサイトサブセットについて、図49は、長画分(上)及び短画分(下)の両方が、ストリンジェンシーが増加するCTCFサイトサブセットでより強い配置シグナルを提示することを示す。要となる限界着色線については図45を参照されたい。

【図50】ヌクレオソームにフランキングするCTCF占拠再配置を示す。図50は、当該518,632サイトについて平均短画分WPS(上パネル)及び平均長画分WPS(下パネル)を示し、前記は、各サイトについてフランキングする+1及び-1ヌクレオソームコールを隔てる塩基対の数を示す距離ピンに分けられている。図51は、図50の518,632サイトについて平均短画分WPS(上パネル)及び平均長画分WPS(下パネル)を示すが、ただしここで同じCTCFサイトセットはENCODE ChIP-seqピークとのオーバーラップに基づいてフィルター処理された。図52は、図51のサイトについて平均短画分WPS(上パネル)及び平均長画分WPS(下パネル)を示すが、ただしここで同じサイトセットは19細胞株にわたって実験的に観察される活性なCTCFサイトセットとのオーバーラップに基づいてさらにフィルター処理された。図50の要となる限界着色線は図51及び図52と同じである。

【図51】ヌクレオソームにフランキングするCTCF占拠再配置を示す。図50は、当該518,632サイトについて平均短画分WPS(上パネル)及び平均長画分WPS(下パネル)を示し、前記は、各サイトについてフランキングする+1及び-1ヌクレオソームコールを隔てる塩基対の数を示す距離ピンに分けられている。図51は、図50の518,632サイトについて平均短画分WPS(上パネル)及び平均長画分WPS(下パネル)を示すが、ただしここで同じCTCFサイトセットはENCODE ChIP-seqピークとのオーバーラップに基づいてフィルター処理された。図52は、図51のサイトについて平均短画分WPS(上パネル)及び平均長画分WPS(下パネル)を示すが、ただしここで同じサイトセットは19細胞株にわたって実験的に観察される活性なCTCFサイトセットとのオーバーラップに基づいてさらにフィルター処理された。図50の要となる限界着色線は図51及び図52と同じである。

【図52】ヌクレオソームにフランキングするCTCF占拠再配置を示す。図50は、当該518,632サイトについて平均短画分WPS(上パネル)及び平均長画分WPS(下パネル)を示し、前記は、各サイトについてフランキングする+1及び-1ヌクレオソームコールを隔てる塩基対の数を示す距離ピンに分けられている。図51は、図50の518,632サイトについて平均短画分WPS(上パネル)及び平均長画分WPS(下パネル)を示すが、ただしここで同じCTCFサイトセットはENCODE ChIP-seqピークとのオーバーラップに基づいてフィルター処理された。図52は、図51のサイトについて平均短画分WPS(上パネル)及び平均長画分WPS(下パネル)を示すが、ただしここで同じサイトセットは19細胞株にわたって実験的に観察される活性なCTCFサイトセットとのオーバーラップに基づいてさらにフィルター処理された。図50の要となる限界着色線は図51及び図52と同じである。

【図53A-D】短cfDNA及び長cfDNAフラグメントに由来する転写因子結合サイトのフットプリントを示す。クラスター化FIMO結合サイト予測をENCODE ChIP-seqデータでインターセクト(intersect)して、追加因子セットのために転写因子(TF)結合サイトの信頼セットを入手した。得られたTF結合サイトセットにフランキングする領域に対する集合、調整WPSを、cfDNAフラグメントの長画分及び短画分の両方について示す。より高いWPS値は、それぞれヌクレオソーム又はTF占拠のより高い可能性を示す。図53A: AP-2、図53B: E2F-2、図53C: EBOX-TF、図53D: IRF。

10

20

30

40

50

【図53E-H】短cfDNA及び長cfDNAフラグメントに由来する転写因子結合サイトのフットプリントを示す。クラスター化FIMO結合サイト予測をENCODE ChIP-seqデータでインターセクト(intersect)して、追加因子セットのために転写因子(TF)結合サイトの信頼セットを入手した。得られたTF結合サイトセットにランキングする領域に対する集合、調整WPSを、cfDNAフラグメントの長画分及び短画分の両方について示す。より高いWPS値は、それぞれヌクレオソーム又はTF占拠のより高い可能性を示す。図53E: MYC-MAX、図53F: PAX5-2、図53G: RUNX-AML、図53H: YY1。

【図54】図54は、転写因子ETS(210,798サイト)に対する集合、調整WPSを示す。長cfDNA画分(上)及び短cfDNA画分(下)の両方から計算したWPSが示されている。周辺のヌクレオソーム(長画分)の組織化による結合サイトそのもの(短画分)におけるTF保護に一致するシグナルが観察される。追加のTFについての同様な分析は図53A-Hに示されている。

【図55】図55は、転写因子MAFK(32,159サイト)に対する集合、調整WPSを示す。長cfDNA画分(上)及び短cfDNA画分(下)の両方から計算したWPSが示されている。周辺のヌクレオソーム(長画分)の組織化による結合サイトそのもの(短画分)におけるTF保護に一致するシグナルが観察される。追加のTFについての同様な分析は図53A-Hに示されている。

【図56】DNase高感受性(DHS)サイトに基づく、セルフリーDNAに寄与する細胞タイプ混合物の推論を示す。116の多岐にわたる生物学的サンプルに由来するDHSサイトのヌクレオソームコールのピークからピークまでの間隔の取り方の度数分布は二峰性分布を示し、第二峰は、おそらく活性なDHSサイトにおける介在転写因子の結合のために広げられた(〜190bp - >260bp)ヌクレオソームの間隔の取り方に一致する。リンパ系又は骨髄系サンプルで識別されるDHSサイトは、ヌクレオソームの間隔の取り方が広がったDHSサイトの最大割合を示し、健康な個体では造血細胞死が優先的なcfDNA供給源であることと一致する。

【図57】転写開始サイト(TSS)周辺の調整WPSスコアをNB-4(急性前骨髄細胞白血病細胞株)について識別された5遺伝子発現bin(五分位数)にどのように分けられるかによって、ヌクレオソームの間隔の取り方及び配置の相違が明らかにされることを示す。高度に発現される遺伝子は転写本体内のヌクレオソームの強力な位相整合を示す。TSSの上流で、-1ヌクレオソームは発現binを通して好位置に存在するが、-2及び-3ヌクレオソームは中等度から高度発現遺伝子に対してのみ好位置に存在する。

【図58】中等度から高度発現遺伝子について、短フラグメントピークはTSSと-1ヌクレオソームの間で観察され、転写活性遺伝子の転写の前開始複合体又はその何らかの成分のフットプリントと一致することを示す。

【図59】転写本体におけるヌクレオソーム距離の中央値はNB-4細胞株(= -0.17、n = 19,677遺伝子)について測定される遺伝子発現と負の相関関係を有することを示す。ほとんど又は全く遺伝子発現のない遺伝子は193bpというヌクレオソーム距離の中央値を示すが、発現遺伝子ではこの範囲は186 - 193bpの間である。この負の相関関係は、より多くのヌクレオソームコールがより正確な距離の中央値の決定に用いられるときにより強くなる(例えば少なくとも60ヌクレオソーム、 = -0.50、n = 12,344遺伝子を必要とする)。

【図60】多重的寄与をデコンボリューションするため、どのように高速フーリエ変換(FFT)を用いて、各TSSで始まる遺伝子本体の最初の10kbのための長フラグメントWPSで特定の周期寄与(強度)の多さを定量するかを示す。種々の周期でこれらの強度を有する76細胞株及び一次組織のRNA発現間の相関関係の曲線が示される。NB-4細胞株は太い黒線で示される。相関関係は193 - 199bpの周期レンジで強度の規模が最大である。

【図61】健康状態及び癌におけるセルフリーDNAに寄与する細胞タイプの推論を示す。上段パネルは、多様なcfDNAライブラリーについて、193 - 199bp周期レンジで平均強度を有する76のRNA発現データセットに対する相関関係ランクを示す(前記cfDNAライブラリーはタイプによって分類され、最高ランク(一番上の列)から最低ランク(一番下の列)に記入されている)。相関値及び完全な細胞株名又は組織名は表3に提供されている。3つの

10

20

30

40

50

全健康サンプル（BH01、IH01及びIH02；最初の3カラム）で最高の相関関係はいずれも、リンパ系及び骨髄系細胞株とともに骨髄である。対照的に、第IV期癌患者から入手されたcfDNAサンプル（IC15、IC17、IC20、IC35、IC37；最後の5カラム）は、多様な癌細胞株と上位相関関係を示す。例えば、IC17（肝細胞癌（HCC））はHepG2（肝細胞癌腫細胞株）と、IC35（乳管癌（DC））はMCF7（転移性乳腺癌細胞株）と最高の相関関係を示す。癌サンプルで観察された細胞株/組織ランクを3つの健康なサンプルの各々と比較しランク変化を平均したとき（下段パネル）、最大ランク変化は、3つの健康サンプルを互に比較しランク変化を平均して観察されたもの（コントロール）の2倍を超える。例えば、IC15（小細胞肺癌（SCLC））については、SCLC-21H（小細胞肺癌細胞株）のランクは平均31位増加し、IC20（扁平上皮細胞肺癌（SCC））については、SK-BR-3（転移性乳腺癌細胞株）で平均ランク21に増加し、IC37（結腸直腸腺癌（AC））については、HepG2で24ランクに増加した。

10

【図62A】循環腫瘍DNA高負荷サンプルの選別のための、カバレッジ（図62A）を基準にする異数性定量を示す。図62Aは各染色体のZスコアの合計を示し、前記スコアは、異数性がないと仮定するシミュレーションサンプル（赤色ドット）と比較した各サンプル（黒色ドット）のシーケンシング読取りの観察数と予測数を基準にして算出された。

【図62B】循環腫瘍DNA高負荷サンプルの選別のための、対立遺伝子バランス（図62B）を基準にする異数性定量を示す。図62Bは、追加のシーケンシングのために選択されたサンプルのサブセットに対する、染色体毎に評価された48,800の共通SNPの各々における対立遺伝子バランスを示す。

20

【図63A】ピークコールと公表されたヌクレオソームコールセットとの比較を示す。図63Aは、3つの公表されたデータセット（Gaffney et al. 2012；JS Pedersen et al. 2014；及びA Schep et al. 2015）全体のヌクレオソームピークコールとともに本明細書で作製したコール（CA01の適合するシミュレーションを含む）との間の距離を示す。以前に公表されたデータセットは、正規の～185bpヌクレオソーム距離では規定された1つのモードを示さず、これはおそらく標本の乏しさ及び広いコール範囲のためであろう。対照的に、cfDNAに由来する全てのヌクレオソームコールは1つの良好に規定されたモードを示す。適合するシミュレーションデータセットはより短いモード（166bp）及びより広い分布を有する。さらにまた、コールの作製に用いられるcfDNAデータセットのカバレッジが高ければ高いほど、当該分布のモードによって表されるコールの割合は高くなる。

30

【図63B】ピークコールと公表されたヌクレオソームコールセットとの比較を示す。図63Bは、図63Aと同じセット一覧の各々についてヌクレオソームの数を示す。このcfDNAヌクレオソームコールは約13Mヌクレオソームピークコールを有するもっとも包括的なコールセットを提示する。図63Cは、IH01 cfDNAサンプルの各ピークコールと以前に公表されたデータセットの最も近いピークコールとの間の距離を示す。図63Dは、IH02 cfDNAサンプルの各ピークコールと以前に公表されたデータセットの最も近いピークコールとの間の距離を示す。図63Eは、BH01 cfDNAサンプルの各ピークコールと以前に公表されたデータセットの最も近いピークコールとの間の距離を示す。図63Fは、CH01 cfDNAサンプルの各ピークコールと以前に公表されたデータセットの最も近いピークコールとの間の距離を示す。図63Gは、CA01 cfDNAサンプルの各ピークコールと以前に公表されたデータセットの最も近いピークコールとの間の距離を示す。負の数は最も近いピークが上流にあり、正の数は最も近いピークが下流にあることを示す。cfDNAのカバレッジが増加するにつれ、以前に公表されたコールがより高い割合で当該決められたヌクレオソームコールにより接近して見いだされる。最高の一致は、Gaffneyら（Gaffney et al., PLoS Genet., vol. 8, e1003036 (2012) and A Schep et al. 2015）及びA.Schepら（2015）が作製したコールで見いだされた。図63Hは、各ピークコールと以前に公表されたデータセットの最も近いピークコールとの間の距離を示すが、この場合CA01の適合シミュレーションに対するものである。最も近い真のヌクレオソームの位置は、以下の研究者ら（Gaffney et al., PLoS Genet., vol. 8, e1003036, 2012；及びJS Pedersen et al., Genome Research, vol. 24, pp. 454-466, 2014）のコールに対するシミュレーションにおいてコールされたピークが

40

50

ら離れる傾向にある。A.Schepら（2015）によって作製されたコールはシミュレーションコールとある程度のオーバーラップを示すように思われる。

【図63C-E】ピークコールと公表されたヌクレオソームコールセットとの比較を示す。図63Cは、IH01 cfDNAサンプルの各ピークコールと以前に公表されたデータセットの最も近いピークコールとの間の距離を示す。図63Dは、IH02 cfDNAサンプルの各ピークコールと以前に公表されたデータセットの最も近いピークコールとの間の距離を示す。図63Eは、BH01 cfDNAサンプルの各ピークコールと以前に公表されたデータセットの最も近いピークコールとの間の距離を示す。

【図63F-H】ピークコールと公表されたヌクレオソームコールセットとの比較を示す。図63Fは、CH01 cfDNAサンプルの各ピークコールと以前に公表されたデータセットの最も近いピークコールとの間の距離を示す。図63Gは、CA01 cfDNAサンプルの各ピークコールと以前に公表されたデータセットの最も近いピークコールとの間の距離を示す。負の数は最も近いピークが上流にあり、正の数は最も近いピークが下流にあることを示す。cfDNAのカバレッジが増加するにつれ、以前に公表されたコールがより高い割合で当該決められたヌクレオソームコールにより接近して見いだされる。最高の一致は、Gaffneyら（Gaffney et al., PLoS Genet., vol. 8, e1003036 (2012) and A Schep et al. 2015）及びA.Schepら（2015）が作製したコールで見いだされた。図63Hは、各ピークコールと以前に公表されたデータセットの最も近いピークコールとの間の距離を示すが、この場合CA01の適合シミュレーションに対するものである。最も近い真のヌクレオソームの位置は、以下の研究者ら（Gaffney et al., PLoS Genet., vol. 8, e1003036, 2012；及びJS Pedersen et al., Genome Research, vol. 24, pp. 454-466, 2014）のコールに対するシミュレーションにおいてコールされたピークから離れる傾向にある。A.Schepら（2015）によって作製されたコールはシミュレーションコールとある程度のオーバーラップを示すように思われる。

【発明を実施するための形態】

【0009】

本開示は、対象者からの生物学的サンプル中のセルフリーDNAを生じる1つ以上の組織及び/又は細胞タイプを決定する方法を提供する。いくつかの実施態様では、本開示は、ある対象者の疾患又は異常を、当該対象者に由来する生物学的サンプル中のcfDNAに付随すると決定された1つ以上の組織及び/又は細胞タイプの関数と認定する方法を提供する。

本開示は、異なる細胞タイプ又は組織を起源とするcfDNA分子は以下の(a)-(c)に関して異なるという予測に基づく：(a) ヒトゲノムにおいて任意の特定の塩基対がcfDNAフラグメントの末端（すなわち断片化点）に出現する可能性の分布；(b) ヒトゲノムの塩基対の任意の対がcfDNAフラグメントの末端の対（すなわち、個々のcfDNA分子を生じる断片化点の連続する対）として出現する可能性の分布；及び(c) ヒトゲノムにおいて任意の特定の塩基対が弁別的なヌクレオソーム占拠の結果としてcfDNAフラグメント内に出現する可能性の分布（すなわち相対的カバレッジ）。これらは下記では分布(a)、(b)及び(c)と称されるか、又は包括的に“ヌクレオソーム依存切断確率マップ”、“切断接近容易性マップ”又は“ヌクレオソームマップ”と称される（図1）。注目すべきことに、ヌクレオソームマップはまた、酵素（例えばマイクロコッカスヌクレアーゼ（MNase）、DNase又はトランスポザーゼ）又はヌクレオソーム若しくはクロマトソームの間で或いはそれらの境界でゲノムDNAを優先的に断片化する同等の手順でクロマチンを断片化することによって誘導されるフラグメントのシーケンシングを介して測定することが可能である。

【0010】

健康な個体では、cfDNAは圧倒的に血液細胞（すなわち造血細胞系列の細胞）のアポトーシスに由来する。これらの細胞はプログラムされた細胞死を経るので、それらのゲノムDNAは切断され循環中に放出される（循環中で前記DNAはヌクレアーゼによって分解され続ける）。cfDNAの長さの分布は、約10.5塩基対（bp）の周期で変動し（これはヌクレオソームの周りに巻き付いているDNAのらせんピッチと一致する）、さらに顕著なピークを167

bp周辺（リンカー結合モノヌクレオソームと結合するDNAの長さとも一致する）に有する（図2）。前記実証から、cfDNAのヌクレオソームとの結合は、循環中での完全で急速な分解からcfDNAを保護するという仮説に至った。また別の可能性は、長さ分布はアポトーシスそのものの最中のDNA切断パターンから単純に生じるということである（前記はヌクレオソームの配置によって直接影響を受ける）。それにもかかわらず、cfDNAの長さの分布は、cfDNAを生じる断片化プロセスはヌクレオソームの配置によって影響を受けるという証拠を提供する。

【0011】

いくつかの実施態様では、本開示は、ヌクレオソームマップを、体液由来cfDNA又はクロマチンの断片化から誘導されるDNAのライブラリー構築及び大量パラレルシーケンシングによる分布（a）、（b）及び/又は（c）の測定と規定する（前記クロマチンの断片化は、酵素（例えばマイクロコッカスヌクレアーゼ（MNase）、DNase又はトランスポザーゼ）、又はヌクレオソーム若しくはクロマトソームの間で或いはそれらの境界でゲノムDNAを優先的に断片化する同等の手順による）。下記で述べるように、これらの分布を“変換”して、ゲノムの多様なサブセット内のヌクレオソームの配置取りによる周期的シグナルを、例えば連続ウィンドウ中或いはまたゲノムの不連続サブセット中（転写因子結合サイト、遺伝子モデル特徴（例えば転写因子開始サイト又は遺伝子本体）、トポロジカルに結合するドメイン、組織発現データ又はヌクレオソームの配置と相関関係を有する他のものによって定される）で周期性を定量し、集合化又は集約することができる。さらに、これらは組織特異的データによって規定することができるかもしれない。例えば、組織特異的DNase I高感受性サイトの近傍のシグナルを集合化又は集約することができよう。

本開示は、血漿由来cfDNAフラグメントから推論されるin vivoヌクレオソーム保護に関する密でゲノムワイドのマップを提供する。健康な個体のcfDNAから誘導されたCH01マップは、マッピング可能なヒト参照ゲノムのほぼ大半に及び、13M近い、均一な間隔で配置されたヌクレオソーム保護の極大値を含む。CH01ではピークの数には本質的に飽和状態であるが、他の品質測定基準はシーケンシング深度さの関数であることを維持した（図33A - B）。したがって、本研究及び他の実験のために、追加のゲノムワイドのヌクレオソームマップ（このマップは本発明者らが今日まで実施してきたcfDNAシーケンシングのほぼ全てに基づく）を同一方法で構築した（‘CA01’、14.5億（G）フラグメント；カバレッジ700重；13.0Mピーク）。このマップはより均一な間隔の取り方及びより高度に支持されるピークコールを示すが（図33A - B、63A - H）、我々は前記が健康個体及び非健康個体の両方に由来するcfDNAに基づくことに注意を喚起する（表1、5）。

本明細書に開示する密でゲノムワイドのヌクレオソーム保護のマップは、ヒト参照ゲノムのマッピング可能部分の飽和に近づき、ヒトゲノムワイドのヌクレオソーム配置又は保護マップの作製のために実施された以前の労作よりも、顕著に均一で期待されたヌクレオソームリピートの長さとも一致するピークからピークまでの間隔の取り方を有する（図63A - H）。ほぼ全ての以前の労作とは対照的に、本明細書で観察されたフラグメントは内因性の生理学的プロセスによって作製され、したがってin vitroのマイクロコッカスヌクレアーゼ消化に随伴する技術的変動を被る恐れは少ない。この参照マップで考えられるcfDNAを生じる細胞タイプは必然的に不均一である（例えば健康な個体ではリンパ系及び骨髄系細胞タイプの混合物）。それにもかかわらず、このマップの相対的完全性は、ヒト細胞におけるヌクレオソームの配置及び間隔、および、ヌクレオソームと後成的調節、転写産物及び核構造との相互作用を指令するプロセスのより深い理解を促進することができる。

【0012】

対象者からの生物学的サンプルにおいてcfDNAの起源を決定する方法

上記で一般的に考察したように、さらに下記の実施例でより具体的に示すように、本技術を用いて、対象者からの生物学的サンプル中のcfDNAに寄与する組織及び/又は細胞タイプを決定することができる。

したがって、いくつかの実施態様では、本開示は、対象者においてセルフリーDNA（cfDNA）を生じる組織及び/又は細胞タイプを決定する方法を提供し、前記方法は、対象者由

10

20

30

40

50

来の生物学的サンプルからcfDNAを単離する工程（当該単離cfDNAは複数のcfDNAフラグメントを含む）；当該複数のcfDNAフラグメントの少なくとも一部分に付随する配列を決定する工程；当該複数のcfDNAフラグメントの少なくともいくつかのcfDNAフラグメント末端の参照ゲノム内のゲノム上の場所を、当該cfDNAフラグメント配列の関数として決定する工程；及び当該cfDNAフラグメントを生じる組織及び/又は細胞タイプの少なくともいくつかを、当該cfDNAフラグメント末端の少なくともいくつかのゲノム上の場所の関数として決定する工程を含む。

いくつかの実施態様では、生物学的サンプルは、全血、末梢血血漿、尿又は脳脊髄液を含むか、本質的にそれらから成るか、又はそれら全血、末梢血血漿、尿又は脳脊髄液から成る。

【0013】

いくつかの実施態様では、cfDNAフラグメントを生じる組織及び/又は細胞タイプの少なくともいくつかを決定する工程は、当該cfDNAフラグメント末端の少なくともいくつかのゲノム上の場所又はそれらの分布の数学的変換と1つ以上の参照マップとの比較を含む。本明細書で用いられるように、“参照マップ”という用語は任意のタイプ又は形態のデータを指し、前記は、cfDNA配列がアラインメントされるゲノム（例えば参照ゲノム）内の座標の関数として、対象者からの生物学的サンプル中のcfDNAの属性と相関関係を有するか又は前記の匹敵することができる。参照マップは、任意の適切な手段によって対象者からの生物学的サンプル中のcfDNAの属性と相関させるか又は比較することができる。例えば言えばかつ非限定的に言えば、相関又は比較は、対象者からの生物学的サンプル中のcfDNA末端の周期を、参照ゲノムの同等の座標について規定された数値又は任意の他の状態の観点で、直接的に又は参照ゲノム内のウィンドウ全体にわたってそれらの分布について数学的変換を実施した後で、参照マップで分析することによって達成できる。別の非限定的な例では、相関又は比較は、対象者からの生物学的サンプルのcfDNAに基づいて決定したヌクレオソームの間隔を、当該決定したヌクレオソームの間隔の取り方又は当該ヌクレオソームの間隔と相関関係を有する別の特性の観点から参照マップで分析することによって達成できる。

【0014】

参照マップは、任意の適切なデータ源に由来するか又はそれらから導ける。前記データ源には例えば、ゲノム情報の公開データベース、公表データ、又は各々共通の属性（例えば疾患の状態）を有しうる参照対象者の具体的集団について作製されたデータが含まれる。いくつかの実施態様では、参照マップはDNase I高感受性データセットを含む。いくつかの実施態様では、参照マップはRNA発現データセットを含む。いくつかの実施態様では、参照マップは染色体コンフォーメーションマップを含む。いくつかの実施態様では、参照マップはクロマチン接近容易性マップを含む。いくつかの実施態様では、参照マップは、ある疾患又は異常に付随する少なくとも1つの組織又は細胞タイプから作製される。いくつかの実施態様では、参照マップは、ある組織又は細胞タイプにおけるヌクレオソーム及び/又はクロマトソームの位置を含む。いくつかの実施態様では、参照マップは、外因性ヌクレアーゼ（例えばマイクロコッカスヌクレアーゼ）によるクロマチンの消化を含む手順によって作製される。いくつかの実施態様では、参照マップは、転移による方法（例えばATAC-seq）によって決定されるクロマチン接近容易性データを含む。いくつかの実施態様では、参照マップは、ある組織又は細胞タイプのDNA結合及び/又はDNA占拠タンパク質の位置に付随するデータを含む。いくつかの実施態様では、DNA結合及び/又はDNA占拠タンパク質は転写因子である。いくつかの実施態様では、位置は、架橋したDNA-タンパク質複合体のクロマチン免疫沈澱を含む手順によって決定される。いくつかの実施態様では、位置は、当該組織又は細胞タイプに付随するDNAのヌクレアーゼ（例えばDNase I）による処理を含む手順によって決定される。いくつかの実施態様では、参照マップは、既知の疾患を有する1人以上の個体に由来する生物学的サンプルのcfDNAフラグメントのシーケンシングによって作製される。いくつかの実施態様では、この生物学的サンプル（前記から参照マップが作製される）は、ヒト細胞又は組織が異種移植された動物から収集される。

10

20

30

40

50

いくつかの実施態様では、参照マップは、ある組織又は細胞タイプのDNA結合又はDNA占拠タンパク質の位置に対応する生物学的特徴を含む。いくつかの実施態様では、参照マップは、1つ以上の遺伝子の定量的なRNA発現に対応する生物学的特徴を含む。いくつかの実施態様では、参照マップは、1つ以上のヒストンマークの有無に対応する生物学的特徴を含む。いくつかの実施態様では、参照マップは、ヌクレアーゼ切断に対する高感受性に対応する生物学的特徴を含む。

【0015】

cfDNA末端の少なくともいくつかのゲノム上の場所を1つ以上の参照マップと比較する工程は多様な方法で達成できる。いくつかの実施態様では、生物学的サンプルから作製されるcfDNAデータ（例えばcfDNAフラグメントのゲノム上の場所、それらの末端、それらの末端の周期、及び/又はそれらの分布から推論されるヌクレオソームの間隔の取り方）が1つ以上の参照マップと比較される。そのような実施態様では、生物学的サンプル中のcfDNAと最高の相関関係を有する参照マップと関係する組織又は細胞タイプが寄与しているときみなされる。例えて言えばかつ非限定的に言えば、cfDNAデータが、ありそうに思われるcfDNAの末端及び参照ゲノム内のそれらの場所の一覧を含むならば、cfDNA末端及びそれらの参照ゲノム内の場所のもっとも類似する一覧を有する参照マップが寄与していると考えられることができる。別の非限定的な例として、生物学的サンプルに由来するcfDNAフラグメント末端の分布の数学的変換と最高の相関関係を有する（又は健康な対象者由来のcfDNAと比較して相関関係が増加する）参照マップが寄与していると考えられることができる。寄与していると考えられるそれら参照マップに一致する組織タイプ及び/又は細胞タイプは、したがって当該生物学的サンプルから単離されたcfDNAの潜在的な供給源と考えられる。

いくつかの実施態様では、cfDNAを生じる組織及び/又は細胞タイプの少なくともいくつかを決定する工程は、cfDNAフラグメント末端の少なくともいくつかのゲノム上の場所の分布の数学的変換を実施する工程を含む。本技術に係る使用に適切な数学的変換の非限定的な一例はフーリエ変換、例えば高速フーリエ変換（“FFT”）である。

【0016】

いくつかの実施態様では、当該方法はさらに、参照ゲノムの少なくともいくつかの座標の各々についてスコアを決定する工程を含み、ここで当該スコアは少なくとも複数のcfDNAフラグメント末端およびその末端のゲノム上の位置の関数として決定され、さらに、観察されるcfDNAフラグメントを生じる組織及び/又は細胞タイプの少なくともいくつかを決定する工程は当該スコアを1つ以上の参照マップと比較する工程を含む。スコアは、参照ゲノムの座標に相対的値又は絶対的値を割り当てるために用いることができる任意の測量基準（たとえば数値的ランク付けまたは確率）でありうる。例えば、スコアは、ある確率（例えば座標がcfDNAフラグメント末端の場所を表す確率、又は座標がヌクレオソーム若しくはタンパク質結合によってヌクレアーゼ切断から優先的に保護されるゲノムの場所を表す確率）から成るか、又は前記確率と関係性を有することができる。別の例として、スコアは、ゲノムの特定の領域内におけるcfDNAフラグメント末端の分布の数学的変換によって決定されるように、当該領域内のヌクレオソームの間隔の取り方と関係を有することができる。そのようなスコアは、以下を含む任意の適切な手段によって当該座標に割り当てることができる：例えば特定の座標と関連する絶対的又は相対的事象（例えばcfDNA末端の数）を計測するか、又は当該領域若しくはゲノム座標のそのような計測値に関して数学的変換を実施する。いくつかの実施態様では、ある座標についてのスコアは、当該座標がcfDNAフラグメント末端の場所であるという確率と結びついている。他の実施態様では、ある座標のスコアは、当該座標が、ヌクレオソーム又はタンパク質結合によってヌクレアーゼ切断から優先的に保護されるゲノムの場所を表すという確率と結びついている。いくつかの実施態様では、スコアは、当該座標の当該ゲノム領域でのヌクレオソームの間隔の取り方と結びついている。

【0017】

本明細書に記載する方法で言及する組織及び/又は細胞タイプは、cfDNAを生じる任意の組織又は細胞タイプでありうる。いくつかの実施態様では、組織又は細胞タイプはある疾

10

20

30

40

50

患又は異常を有する対象者の一次組織である。いくつかの実施態様では、疾患又は異常は以下から成る群から選択される：癌、正常妊娠、妊娠（例えば異数性妊娠）の合併症、心筋梗塞、炎症性腸疾患、全身性自己免疫疾患、局所性自己免疫疾患、拒絶を伴う同種間移植、拒絶を伴わない同種間移植、卒中、及び局在性組織損傷。

いくつかの実施態様では、組織又は細胞タイプは健康な対象者の一次組織である。

いくつかの実施態様では、組織又は細胞タイプは不死化細胞株である。

いくつかの実施態様では、組織又は細胞タイプは腫瘍の生検である。

いくつかの実施態様では、参照マップは、少なくとも1人の参照対象者から入手したサンプルから得られる配列に基づく。いくつかの実施態様では、この配列データは、例えば参照マップが既知疾患を有する対象者のcfDNAのシーケンシングによって作製される場合は、参照ゲノム内のcfDNAフラグメント末端の場所を規定する。他の実施態様では、この配列データ（参照マップは前記データに基づく）は以下の任意の1つ以上を含むことができる：DNase I高感受性サイトデータセット、RNA発現データセット、染色体コンフォーメーションマップ、クロマチン接近容易性マップ、又はマイクロコッカスヌクレアーゼによるクロマチン消化によって作製されるヌクレオソームの間隔の取り方マップ。

いくつかの実施態様では、参照対象者は健康である。いくつかの実施態様では、参照対象者は、場合によって以下から成る群から選択される疾患又は異常を有する：癌、正常妊娠、妊娠（例えば異数性妊娠）の合併症、心筋梗塞、炎症性腸疾患、全身性自己免疫疾患、局所性自己免疫疾患、拒絶を伴う同種間移植、拒絶を伴わない同種間移植、卒中、及び局在性組織損傷。

【0018】

いくつかの実施態様では、参照マップは、当該組織又は細胞タイプと結びついた参照ゲノムの座標の少なくとも一部分に対するスコアを含む。いくつかの実施態様では、参照マップは、スコアの数学的変換（例えばスコアのフーリエ変換）を含む。いくつかの実施態様では、スコアは、当該組織又は細胞タイプに対する参照ゲノム座標の注釈に基づく。いくつかの実施態様では、スコアはヌクレオソーム及び/又はクロマトソームの位置に基づく。いくつかの実施態様では、スコアは転写開始サイト及び/又は転写終了サイトに基づく。いくつかの実施態様では、スコアは少なくとも1つの転写因子の予測される結合サイトに基づく。いくつかの実施態様では、スコアは予測されるヌクレアーゼ高感受性サイトに基づく。いくつかの実施態様では、スコアは予測されるヌクレオソームの間隔に基づく。

いくつかの実施態様では、スコアは少なくとも1つのオルトゴナルな生物学的特徴と結びついている。いくつかの実施態様では、オルトゴナルな生物学的特徴は高度に発現される遺伝子と結びついている。いくつかの実施態様では、オルトゴナルな生物学的特徴は低発現遺伝子と結びついている。

いくつかの実施態様では、複数のスコアの少なくともいくつかは閾（最小）値を超える値を有する。そのような実施態様では、閾（最小）値より低いスコアは、参照マップとのスコアの比較工程から除外される。いくつかの実施態様では、閾値は、cfDNAを生じる組織及び/又は細胞タイプの決定前に決定される。他の実施態様では、閾値は、cfDNAを生じる組織及び/又は細胞タイプの決定後に決定される。

【0019】

いくつかの実施態様では、cfDNAを生じる組織及び/又は細胞タイプを、当該cfDNAフラグメント末端の少なくともいくつかの複数のゲノム上の場所の関数として決定する工程は、サンプルのcfDNAフラグメント末端の少なくともいくつかのゲノム上の場所の分布の数学的変換を1つ以上の参照マップの1つ以上の特徴と比較する工程を含む。この目的に適切な数学的変換の非限定的な一例は、フーリエ変換（例えば高速フーリエ変換（“FFT”））である。

本明細書に記載されるいずれかの実施態様では、当該方法はさらに、単離cfDNAを生じると決定された組織及び/又は細胞タイプの一覧を含む報告を作製する工程を含む。当該報告は場合によってさらに以下についての任意の情報を含む：サンプル及び/又は対象者

、生物学的サンプルのタイプ、生物学的サンプルを対象者から入手した日付、cfDNA単離工程を実施した日付、及び/又は生物学的サンプルから単離されたいずれのcfDNAもおそらく生じなかった組織及び/又は細胞タイプ。

いくつかの実施態様では、当該報告はさらに推奨される処理プロトコルを含む。前記プロトコルは、例えて言えばかつ非限定的に言えば、対象者から追加の診断検査を入手することの勧め、治療レジメンの開始の勧め、既存の治療レジメンの対象者に関する変更の勧め、及び/又は既存の治療レジメンの一時停止又は中止の勧めを含む。

【0020】

対象者における疾患又は異常を識別する方法

上記で一般的に考察し、さらに下記実施例でより具体的に示すように、本開示技術を用いて、疾患若しくは異常又は疾患若しくは異常が存在しないことを、対象者からの生物学的サンプルに寄与する組織及び/又は細胞タイプに少なくとも部分的に基づいて決定（例えば予測）することができる。

したがって、いくつかの実施態様では、本開示は対象者の疾患又は異常を識別する方法を提供し、前記方法は以下の工程を含む：対象者からの生物学的サンプルからセルフリーDNA（cfDNA）を単離する工程（当該単離cfDNAは複数のcfDNAを含む）；当該複数のcfDNAフラグメントの少なくとも一部分に付随する配列を決定する工程；当該複数のcfDNAフラグメントの少なくともいくつかのcfDNAフラグメントの末端の参照ゲノム内のゲノム上の場所を、当該cfDNAフラグメント配列の関数として決定する工程；当該cfDNAを生じる組織及び/又は細胞タイプの少なくともいくつかを、当該cfDNAフラグメント末端の少なくともいくつかのゲノム上の場所の関数として決定する工程；及び当該疾患又は異常を、当該cfDNAを生じると決定された組織及び/又は細胞タイプの関数として同定する工程。

いくつかの実施態様では、生物学的サンプルは全血、末梢血血漿、尿又は脳脊髄液を含むか、本質的に前記から成るか、又は前記から成る。

【0021】

いくつかの実施態様では、cfDNAフラグメントを生じる組織及び/又は細胞タイプを決定する工程は、当該cfDNAフラグメント末端の少なくともいくつかのゲノム上の場所又はそれらの分布の数学的変換を1つ以上の参照マップと比較する工程を含む。これらの実施態様と関連して用いられる“参照マップ”という用語は、対象者からの生物学的サンプル中のcfDNAを生じる組織及び/又は細胞タイプを決定する方法に関して上記に記載した意味と同じ意味を有することができる。いくつかの実施態様では、参照マップは以下の任意の1つ以上を含むことができる：DNase I高感受性サイトデータセット、RNA発現データセット、染色体コンフォーメーションマップ、クロマチン接近容易性マップ、少なくとも1人の参照対象者から入手されるサンプルから作製される配列データ、疾患又は異常に付随する少なくとも1つの組織に対応する酵素媒介断片化データ、及び/又はある組織又は細胞タイプのヌクレオソーム及び/又はクロマトソームの位置。いくつかの実施態様では、参照マップは、既知の疾患を有する1人以上の個体に由来する生物学的サンプルのcfDNAフラグメントのシーケンシングによって作製される。いくつかの実施態様では、この生物学的サンプル（前記から参照マップが作製される）は、ヒト細胞又は組織が異種移植された動物から収集される。

【0022】

いくつかの実施態様では、参照マップは、外因性ヌクレアーゼ（例えばマイクロコッカスヌクレアーゼ）によるクロマチンの消化によって作製される。いくつかの実施態様では、参照マップは、転移による方法（例えばATAC-seq）によって決定されるクロマチン接近容易性データを含む。いくつかの実施態様では、参照マップは、ある組織又は細胞タイプのDNA結合及び/又はDNA占拠タンパク質の位置に付随するデータを含む。いくつかの実施態様では、DNA結合及び/又はDNA占拠タンパク質は転写因子である。いくつかの実施態様では、位置は、架橋したDNA-タンパク質複合体のクロマチン免疫沈澱によって決定される。いくつかの実施態様では、位置は、当該組織又は細胞タイプに付随するDNAをヌクレアーゼ（例えばDNase I）で処理することによって決定される。

いくつかの実施態様では、参照マップは、ある組織又は細胞タイプのDNA結合又はDNA占拠タンパク質の位置に対応する生物学的特徴を含む。いくつかの実施態様では、参照マップは、1つ以上の遺伝子の定量的な発現に対応する生物学的特徴を含む。いくつかの実施態様では、参照マップは、1つ以上のヒストンマークの有無に対応する生物学的特徴を含む。いくつかの実施態様では、参照マップは、ヌクレアーゼ切断に対する高感受性に対応する生物学的特徴を含む。

いくつかの実施態様では、当該cfDNAを生じる組織及び/又は細胞タイプを決定する工程は、当該複数のcfDNAフラグメント末端の少なくともいくつかのゲノム上の場所の分布に関して数学的変換を実施する工程を含む。いくつかの実施態様では、数学的変換はフーリエ変換を含む。

【0023】

いくつかの実施態様では、当該方法はさらに、参照ゲノムの少なくともいくつかの座標の各々についてスコアを決定する工程を含み、ここで前記スコアは当該少なくとも複数のcfDNAフラグメント末端及びそれらのゲノム上の場所の関数として決定され、さらに、観察されるcfDNAフラグメントを生じる組織及び/又は細胞タイプの少なくともいくつかを決定する工程は、当該スコアを1つ以上のマップと比較する工程を含む。スコアは、参照ゲノムの座標に相対的又は絶対的値を割り当てるために用いることができる任意の測量基準（例えば数値的ランキング又は確率）でありうる。例えば、スコアは、ある確率（例えば座標がcfDNAフラグメント末端の場所を表す確率、又は座標がヌクレオソーム若しくはタンパク質結合によってヌクレアーゼ切断から優先的に保護されるゲノムの場所を表す確率）から成るか、又は前記確率と関係性を有することができる。別の例として、スコアは、ゲノムの特定の領域内におけるcfDNAフラグメント末端の分布の数学的変換によって決定されるように、当該領域内のヌクレオソームの間隔と関係性を有することができる。そのようなスコアは、以下を含む任意の適切な手段によって当該座標に割り当てることができる：例えば特定の座標と結びついた絶対的又は相対的事象（例えばcfDNA末端の数）を計測するか、又は当該領域若しくはゲノム座標のそのような計測値に関して数学的変換を実施する。いくつかの実施態様では、ある座標についてのスコアは、当該座標がcfDNAフラグメント末端の場所であるという確率と結びついている。他の実施態様では、ある座標のスコアは、当該座標が、ヌクレオソーム又はタンパク質結合によってヌクレアーゼ切断から優先的に保護されるゲノムの場所を表す確率と結びついている。いくつかの実施態様では、スコアは、当該座標の当該ゲノム領域でのヌクレオソームの間隔と結びついている。

【0024】

これらの実施態様と関連して用いられる“スコア”という用語は、対象者の生物学的サンプルのcfDNAを生じる組織及び/又は細胞タイプを決定する方法に関して上記で述べた意味と同じ意味を有することができる。いくつかの実施態様では、ある座標のスコアは、当該座標がcfDNAフラグメント末端の場所であるという確率と結びついている。他の実施態様では、ある座標のスコアは、当該座標が、ヌクレオソーム又はタンパク質結合によってヌクレアーゼ切断から優先的に保護されるゲノムの場所を表す確率と結びついている。いくつかの実施態様では、スコアは、当該座標の当該ゲノム領域でのヌクレオソームの間隔の取り方と結びついている。

いくつかの実施態様では、参照マップの作製に用いられる組織又は細胞タイプはある疾患又は異常を有する対象者の一次組織である。いくつかの実施態様では、疾患又は異常は以下から成る群から選択される：癌、正常妊娠、妊娠（例えば異数性妊娠）の合併症、心筋梗塞、全身性自己免疫疾患、局所性自己免疫疾患、炎症性腸疾患、拒絶を伴う同種間移植、拒絶を伴わない同種間移植、卒中、及び局在性組織損傷。

いくつかの実施態様では、組織又は細胞タイプは健康な対象者の一次組織である。

いくつかの実施態様では、組織又は細胞タイプは不死化細胞株である。

いくつかの実施態様では、組織又は細胞タイプは腫瘍の生検である。

【0025】

いくつかの実施態様では、参照マップは、少なくとも1人の参照対象者から入手したサ

10

20

30

40

50

ンプルから得られる配列に基づく。いくつかの実施態様では、この配列データは、例えば参照マップが既知疾患を有する対象者のcfDNAのシーケンシングによって作製される場合は、参照ゲノム内のcfDNAフラグメント末端の場所を規定する。他の実施態様では、この配列データ（参照マップは前記データに基づく）は以下の任意の1つ以上を含むことができる：DNase I高感受性サイトデータセット、RNA発現データセット、染色体コンフォーメーションマップ、クロマチン接近容易性マップ、又はマイクロコッカスヌクレアーゼによる消化によって作製されるヌクレオソームの配置マップ。いくつかの実施態様では、参照対象者は健康である。いくつかの実施態様では、参照対象者は疾患又は異常を有する。いくつかの実施態様では、当該疾患又は異常は以下から成る群から選択される：癌、正常妊娠、妊娠（例えば異数性妊娠）の合併症、心筋梗塞、全身性自己免疫疾患、炎症性腸疾患、局所性自己免疫疾患、拒絶を伴う同種間移植、拒絶を伴わない同種間移植、卒中、及び局在性組織損傷。

10

【0026】

いくつかの実施態様では、参照マップは、当該組織又は細胞タイプに結びついている参照ゲノムの少なくとも一部分に対するcfDNAフラグメント末端確率、又はそのような確率と相関性を有する量を含む。いくつかの実施態様では、参照マップは、当該cfDNAフラグメント末端確率又はそのような確率と相関性を有する量の数学的変換を含む。

いくつかの実施態様では、参照マップは、当該組織又は細胞タイプに結びついた参照ゲノムの座標の少なくとも一部分に対するスコアを含む。いくつかの実施態様では、参照マップは、スコアの数学的変換（例えばスコアのフーリエ変換）を含む。いくつかの実施態様では、スコアは当該組織又は細胞タイプに対する参照ゲノム座標の注釈に基づく。いくつかの実施態様では、スコアはヌクレオソーム及び/又はクロマトソームの位置に基づく。いくつかの実施態様では、スコアは転写開始サイト及び/又は転写終了サイトに基づく。いくつかの実施態様では、スコアは少なくとも1つの転写因子の予測される結合サイトに基づく。いくつかの実施態様では、スコアは予測されるヌクレアーゼ高感受性サイトに基づく。

20

【0027】

いくつかの実施態様では、スコアは少なくとも1つのオルトゴナルな生物学的特徴と結びついている。いくつかの実施態様では、オルトゴナルな生物学的特徴は高発現遺伝子と結びついている。いくつかの実施態様では、オルトゴナルな生物学的特徴は低発現遺伝子と結びついている。

30

いくつかの実施態様では、当該複数のスコアの少なくともいくつかは各々閾値を超えるスコアを有する。そのような実施態様では、閾（最小）値より低いスコアは、当該スコアを参照マップと比較する工程から除外される。いくつかの実施態様では、閾値は、cfDNAを生じる組織及び/又は細胞タイプの決定前に決定される。他の実施態様では、閾値は、cfDNAを生じる組織及び/又は細胞タイプの決定後に決定される。

いくつかの実施態様では、cfDNAを生じる組織及び/又は細胞タイプを、当該cfDNAフラグメント末端の少なくともいくつかの複数のゲノム上の場所の関数として決定する工程は、サンプルのcfDNAフラグメント末端の少なくともいくつかのゲノム上の場所の分布の数学的変換を1つ以上の参照マップの1つ以上の特徴と比較する工程を含む。

40

いくつかの実施態様では、この数学的変換はフーリエ変換を含む。

いくつかの実施態様では、参照マップは、当該疾患又は異常と結びついた少なくとも1つの組織に対応する酵素媒介断片化データを含む。

いくつかの実施態様では、参照ゲノムはヒトに付随する。

【0028】

本発明のある特徴では、体液中のcfDNAの分析による悪性疾患の検出、モニタリング並びの供給源組織及び/又は細胞タイプの判断のために、本明細書に記載の方法が用いられる。悪性疾患を有する患者で、体液（例えば循環血漿）中のcfDNAの一部が腫瘍に由来しうるということは今では文書で十分に裏付けられている。本明細書に記載の方法を用いて、この腫瘍由来部分を検出及び定量できる可能性は高い。さらにまた、ヌクレオソーム

50

占拠マップは細胞タイプ特異的であるので、本明細書に記載の方法を用いて、悪性疾患の供給源組織及び/又は細胞タイプを決定できる可能性は高い。さらにまた上記で特記したように、癌では循環血漿cfDNAの濃度に大きな増加があり、前記は腫瘍そのものの寄与に対して潜在的に不均衡であることが観察されている。これは、癌では他の組織（例えば間質系、免疫系）が循環血漿cfDNAに寄与する可能性を示唆している。そのような他の組織のcfDNAへの寄与があるタイプの癌について患者間で一貫しているかぎりにおいては、上記に記載の方法は、癌細胞それ自体ではなくこれらの他の組織のシグナルに基づいて癌検出、癌モニタリング、及び/又は供給源組織及び/又は細胞タイプの割り当てを可能にする。

本発明の別の特徴では、組織の損傷の検出、モニタリング並びに供給源組織及び/又は細胞タイプの判断のために、本明細書に記載の方法が用いられる。多くの病理学的プロセスが損傷組織に由来する体液（例えば循環血漿）中のcfDNAの一部を生じるであろうということは予測できる。本明細書に記載の方法を用いて、組織損傷に由来するcfDNAを検出及び定量できる可能性は高い（対応する供給源組織及び/又は細胞タイプの識別を含む）。これは、病理学的プロセス（心筋梗塞（心臓組織の急性損傷）、自己免疫疾患（多岐にわたる組織の慢性損傷）及び急性又は慢性組織損傷を伴う多くの他のものを含む）の診断及び/又はモニタリングを可能にする。

【0029】

本発明の別の特徴では、妊娠におけるcfDNAの胎児画分の推定及び/又は染色体若しくは他の遺伝子異常の検出強化のために、本明細書に記載の方法が用いられる。上記に記載のヌクレオソームマップと一緒にすれば、母体血漿運搬DNAフラグメントの相対的に低深度のシーケンシングは、費用効率が高く迅速な男女の両胎児妊娠で胎児部分の推定を可能にすることができる。さらにまた、母体ゲノム又は胎児ゲノムに由来したかもしれないという可能性に関して、不均一な確率を個々のシーケンシング読取り（リード）に割り当てることを可能にすることによって、これらの方法はまた、母体の体液中のcfDNAの分析を介して染色体異常（例えばトリソミー）を検出することを目的とする検査のパフォーマンスを強化することができる。

本発明の別の特徴では、cfDNAへの移植（自家移植又は同種移植片）の寄与を定量するために、本明細書に記載の方法が用いられる。急性同種移植片拒絶の早期及び非侵襲性検出の従来の方法は、血漿運搬DNAのシーケンシング及びドナーゲノム由来フラグメントの濃度増加の認定を必要とする。このアプローチは、例えば5 - 10%のドナー画分を検出するために、このプールフラグメントの相対的に高深度シーケンシング（deep sequencing）を必要とする。前記に代わって提供器官のヌクレオソームマップに基づくアプローチは、深度のより低いシーケンシングで同様な推定を、又は同等量のシーケンシングでより感度の高い推定を可能にすることができる。癌と同様に、移植片そのもの以外の細胞タイプが移植片拒絶時にcfDNA組成に寄与することもまたありうる。そのような他の組織のcfDNAへの寄与が移植片拒絶時の患者間で一貫しているかぎり、上記に記載の方法は、移植ドナー細胞それ以外のこれら他の組織に由来するシグナルに基づいて移植片拒絶をモニターすることを可能にする。

【0030】

本開示の追加の実施態様

本開示はまた、ある既知の疾患又は異常を有する対象者から作製したヌクレオソーム参照マップを用いてある疾患又は異常を診断する方法を提供する。いくつかのそのような実施態様では、前記方法は以下の工程を含む：（1）ヌクレオソームマップの参照セットを作製する工程（ここで各ヌクレオソームマップは以下のいずれかから誘導される：所定の臨床症状（例えば正常、妊娠、癌タイプA、癌タイプBなど）を有する個体の体液に由来するcfDNA、及び/又は特定の組織及び/又は細胞タイプのクロマチンの消化に由来するDNA）；（2）個体のcfDNAに由来するヌクレオソームマップを当該ヌクレオソームマップの参照セットと比較することによって、当該個体の体液から臨床症状及び/又はcfDNAの供給源組織/細胞タイプ組成を予測する工程。

10

20

30

40

50

工程1：ヌクレオソームマップの参照セットの作製、及びヌクレオソーム配置に由来するシグナルの集合化又は集約。

ヌクレオソームマップの作製のために好ましい方法は、DNAの精製、ライブラリーの構築（アダプター連結及びことによるとPCR増幅による）、及び体液由来cfDNAの大量平行シーケンシングを含む。ヌクレオソームマップのまた別の供給源（前記は参照点として本発明の関係で又は変型物の主要成分の識別のために有用である）は、マイクロコッカスヌクレアーゼ（MNase）、DNase処理、ATAC-Seq又は他の関連方法によるクロマチンの消化から誘導されるDNAであり、ここでヌクレオソームの配置についての情報は分布（a）、（b）又は（c）で捕捉される。これらの分布（a）、（b）及び（c）の記述は上記段落[0009]で提供され、さらに図1のグラフで示されている。

10

【0031】

原則として、そのようなライブラリーの非常に高深度のシーケンシングを用いて、cfDNAに寄与する集合細胞タイプで当該ゲノムの特定の座標におけるヌクレオソーム占拠を定量することができるが、しかし今日これは非常に費用がかかる。しかしながら、ヌクレオソーム占拠パターンに付随するシグナルは、当該ゲノムの連続又は不連続領域にわたって集約又は集合化することができる。例えば、本明細書に提供する実施例1又は2では、シーケンシング読取り開始サイト位置がマップされる参照ヒトゲノムのサイト分布（すなわち分布（a））が、10キロ塩基対（kbp）の連続ウィンドウでフーリエ変換に付され、続いてヌクレオソーム占拠に付随する周期範囲について強度の定量が実施される。これは、ヌクレオソームが構造化された配置を10kbpウィンドウ内で示す範囲を効率的に集約する。本明細書で提供される実施例3では、我々は、特定の転写因子（TF）の転写因子結合サイト（TFBS）の直近にシーケンシング読取り開始サイト位置がマップされる参照ヒトゲノムにおいてサイトの分布（すなわち分布（a））を定量する（TFがTFBSと結合するときしばしば当該サイトの直ぐ側にヌクレオソームがフランキングする）。これは、cfDNAに寄与する細胞タイプのTF活性の結果としてヌクレオソームの配置を効率的に集約する。重要なことに、ヌクレオソーム占拠シグナルを有意義に集約することができる多くの関連する方法が存在する。これら方法には、他のゲノム上の目印（例えばDNase I高感受性サイト、転写開始サイト、トポロジカルドメイン、他の後成的な目印又は、他のデータセット（例えば遺伝子発現など）での相関的動態によって規定される全てのそのようなサイト）周辺で分布（a）、（b）及び/又は（c）のシグナルの集合化が含まれる。シーケンシングの費用がさらに低下するので、未知のcfDNAサンプルと比較するために、ヌクレオソーム占拠のマップ（既知疾患に付随するcfDNAサンプルから作製されるマップを含む）を参照マップとして、直接（すなわちシグナルを集合化することなく）使用することもまた可能であろう。いくつかの実施態様では、この（ヌクレオソーム占拠の参照マップが作製される）生物学的サンプルは、（ヒト細胞又は組織が異種移植される）動物から収集される。この利点は、造血細胞系列と一緒に問題の細胞/組織に由来するcfDNA混合物が提示されるのとは対照的に、ヒトゲノムをマッピングする配列決定cfDNAフラグメントは、例外なく当該異種移植された細胞又は組織に由来する。

20

30

【0032】

工程2：1つ以上の新規な個体/サンプルのcfDNA誘導ヌクレオソームマップをヌクレオソームマップの参照セットと、直接又は各マップの数学的変換後に比較することに基づく病理、臨床症状、及び/又は供給源組織/細胞タイプの予測。

40

いったんヌクレオソームマップの参照セットを作製したら、追加のヌクレオソームマップを参照セットと比較する多様な統計的シグナルのプロセッシング方法が存在する。実施例1及び2では、まず初めに多岐にわたるサンプルセット中の当該ゲノムに沿って10kbpウィンドウでロングレンジヌクレオソーム順番付けを集約し、続いて主要成分分析（PCA）を実施して、サンプルをクラスター化するか（実施例1）又は混合割合を推定する（実施例2）。我々は、これらの実施例で用いられるcfDNAサンプルの全ての臨床症状及び全ての細胞株サンプルの供給源組織/細胞タイプを承知しているが、これらサンプルの一切は“未知のもの”であり、かつ臨床症状の有無又はその供給源組織/細胞タイプの予測に用い

50

られるPCA分析におけるその動態（前記は他の全てのヌクレオソームマップと対比されるPCA分析におけるその動態に基づく）は不明である。

当該未知のサンプルでは、参照セットの1+メンバーと1:1の態様で厳密に適合することは必ずしも要求されない。逆に、互いに対するその類似性を定量することができ（実施例1）、又はそのヌクレオソームマップを参照セットに由来する2+サンプルの不均一混合物のモデルにすることができる（実施例2）。

【0033】

本発明の方法の成功がするために、各サンプル中のcfDNAの供給源組織/細胞タイプの組成を予測すること又は最終的に知ることは必要とされない。寧ろ、本明細書に記載の方法は、個々の病理又は臨床症状の関係でcfDNAの供給源組織及び/又は細胞タイプの組成の一貫性を必要とする。しかしながら、クロマチンの消化に由来するDNAを分析しこれらをヌクレオソームマップに加えて多数の組織及び/又は細胞タイプのヌクレオソームマップを直接調査することによって、未知のcfDNA由来サンプルに寄与する組織及び/又は細胞タイプを推測することが可能となる。

本明細書に記載する任意の実施態様で、当該方法はさらに、疾患又は異常を識別する記述を含む報告を作製する工程を含むことができる。いくつかの実施態様では、当該報告はさらに、当該単離cfDNAを生じたと決定される組織及び/又は細胞タイプの一覧を含むことができる。いくつかの実施態様では、当該報告はさらに、当該対象者と関係することがおそれない疾患及び/又は異常の一覧を含む。当該報告は場合によってはさらに、当該サンプル及び/又は対象者、生物学的サンプルのタイプ、生物学的サンプルが対象者から入手された日付、cfDNA単離工程が実施された日付、及び/又は当該生物学的サンプルから単離されたcfDNAのいずれもおそらくそれらから生じたのではないであろうと思われる組織及び/又は細胞タイプについての任意の他の情報を含むことができる。

いくつかの実施態様では、当該報告はさらに以下を含む推奨治療プロトコルを含む：例えて言えばかつ非限定的に言えば、対象者の追加の診断検査の入手の助言、治療レジメン開始の助言、対象者に関する既存レジメンの改変の助言、及び/又は既存の治療レジメンの一時停止又は停止の助言。

【実施例1】

【0034】

セルフリーDNAヌクレオソームマップの主要構成要素分析

cfDNA抽出物及びMNase消化実験から得られたシーケンシングデータにおいて読取り開始位置の分布を調べ、ヌクレオソーム配置に関連するシグナルの存在を評価する。この目的のために、プールcfDNAサンプル（数が明らかでない健康な個体から提供されたものを含むヒト血漿；バルクcfDNA）、単一の健康な男性コントロール個体由来のcfDNAサンプル（MC2.cfDNA）、頭蓋内腫瘍を有する患者に由来する4つのcfDNAサンプル（腫瘍2349、腫瘍2350、腫瘍2351、腫瘍2353）、5つの異なるヒト細胞株由来の6つのMNase消化実験（Hap1.MNase、HeLa.MNase、HEK.MNase、NA12878.MNase、HeLaS3、MCF.7）及び異なる妊婦個体に由来する7つのcfDNAサンプル（gm1matplas、gm2matplas、im1matplas、fgs002、fgs00、fgs004、fgs005）を分析し、女性リンパ芽球系細胞株（NA12878）から抽出したDNAの通常ショットガンシーケンシングデータセットと対比した。プールcfDNA（バルクcfDNA_部分、26%）及び健康男性コントロール個体（MC2.cfDNA_部分、18%）のサブセットも別個のサンプルとして加え、シーケンシング深度の影響を精査した。

読取り開始座標を抽出し、方法のセクションに記載したように高速フーリエ変換（FFT）を用いてピリオドグラムを作製した。この分析は、読取り開始サイトの分布においてどれくらい多くの非均一性が特定の周期（周波数）/周期性のシグナルによって説明できるかを決定する。我々は120 - 250bpのレンジに焦点を当てた。このレンジは、ヌクレオソームリンカー配列（10 - 80bp）の追加配列とともに単一ヌクレオソームに巻き付くDNAの長さの範囲（120 - 250bp）を含む。図3は、ヒト染色体1及びヒト染色体22の全ブロックに及び各周期についての平均強度を示す。cfDNAサンプルと同様にMNase消化実験は200bp周期性より低い明瞭なピークを示す。そのようなピークはヒトショットガンデータでは観察さ

れない。これらの分析は、cfDNAのフラグメント境界の分布におけるヌクレオソーム配置の主要な効果と一致する。

【0035】

サンプル間の正確なピーク周期の変動もまた観察された。これは、各細胞タイプでリンカー配列の長さの分布が異なることの結果かもしれない。ピークはヌクレオソーム結合DNAパターン+リンカー配列から生じるということは、ピーク周辺にフランキングするものは対称的ではないという観察、及びピークより低い周期と比較してピークより高い周期の強度は低いという観察によって支持される。これは、図3に提示したプロットに類似するプロットを用いて、cfDNA及びMNaseシーケンシングデータの品質管理を実施することができること示唆している。cfDNA及び通常（ショットガン）DNAによるMNaseのランダムな断片化又は夾雑は希釈を生じ、極端な場合にはこれらの特徴的な強度パターンの完全な消失をもたらすであろう。

以下では、データは、181bpから202bpの周期（周波数）レンジについて決定された全強度および196bpの周期性で測定された強度に基づいて分析された。より高度の解析を提供するためにより広い周波数（周期）レンジを選択した（なぜならば、より広範囲のリンカーの長さが捕捉されているからである）。これらの強度はここでは純粹に計算的理由から焦点を当てて選択されている。関連する実施態様では異なる周期レンジを用いることができる。図4及び5は、ヒト常染色体の完全長を覆う連続的でオーバーラップしない10kbpブロックにわたって196bpでのピリオドグラム強度の可視化を追求した（詳細については方法の項を参照されたい）。図4は、データの主要成分分析（PCA）及び最初の3成分に対する推定を示す。主要成分1（PC1）（28.1%の分散）は図3で観察される強度の強さにおける相違を捕捉し、それによってMNase及びcfDNAサンプルをゲノムショットガンデータから切り離す。対照的に、PC2（9.7%の分散）はMNaseとcfDNAサンプルとの間の相違を捕捉する。PC3（6.4%分散）は個々のサンプル間の相違を捕捉する。図5はこのデータの階層クラスタ化樹形図を示し、前記樹形図は強度ベクターのユークリッド距離に基づく。我々は、これら2つのHeLa S3実験が、（データが異なる研究室で作製され異なる実験プロトコルに従っているにもかかわらず）PCA及び樹形図において堅固なクラスターを形成していることに注目している。“正常”cfDNAサンプル、腫瘍cfDNAサンプル及び細胞株MNaseサンプルグループもまたクラスターを形成した。具体的には、同じ腫瘍タイプ（多形性神経膠芽腫）に由来する3つの腫瘍サンプルは、腫瘍2351（異なる腫瘍タイプに由来する（表1参照））とは別個のクラスターを形成するように思われる。GM1及びIM1サンプルは、妊婦から入手した他のcfDNAサンプルとは別個のクラスターを形成する。これは、これらサンプルのピークより低い周期について観察されるより高い強度と一致する（すなわち図3のより明瞭な左肩）。これは、2つのサンプルセット間のcfDNAの調製における微妙な相違、又は制御されなかった生物学的相違（例えば妊娠期間）を示しているのかもしれない。

【0036】

図6及び7は同等の分析の結果を示すが、ただし181bpから202bpの周期レンジに基づく。これらのプロットを比較するとき、それらの結果は、より広い周期レンジについて大いに安定しているが、周期の追加はより微細スケールによる分析で感度を改善することができる。供給源細胞タイプ特異的パターンをさらに追及するために、cfDNA及びMNaseデータセットを、この周期レンジに対する強度のPCAを用いて別個に分析した。以下の分析セットでは、妊婦の5つのcfDNAサンプルは除外した（前記サンプルは図3で明瞭な左肩を示す）。図8はcfDNAデータの最初の7つの主要成分を示し、図9は6つのMNaseデータセットについて6つの主要成分の全てを示す。関連するサンプルのクラスタ化が存在するが、各サンプルをその残りのサンプルと分離する顕著な変動（生物学的及び技術的変動）もまた存在する。例えば、バルクcfDNAとバルクcfDNA_部分との分離、或いはMC2.cfDNAとMC2.cfDNA_部分との分離から分かるように、シーケンシングの深度の影響が認められた。この技術的な混乱を修正するために読取りサンプリングを用いることができる。

この実験におけるいくつかの重要な観察には以下が含まれる：

1) cfDNAシーケンシングデータの読取り開始座標はヌクレオソーム配置の強力なシグナル

を捕捉する。

2) ゲノムのサブセット (例えば連続10kbpウィンドウ) にわたって集められた、ヌクレオソーム配置シグナルにおける相違は、サンプルの起源と相関関係を有する。

【実施例 2】

【0037】

ヌクレオソームマップの混合割合の推定

実施例1では、新規作成されたか又は公開データベースからダウンロードされたサンプルの基礎的クラスタリングが調べられた。この分析は、これらのデータセットの読取り開始座標は、(2千万配列から10億を超える配列から入手されたある範囲のシーケンシング深度にわたって)ヌクレオソーム配置の強力なシグナルを捕捉すること、及びサンプルの起源はこのシグナルと相関関係を有することを示した。この方法の目標のために、既知細胞タイプの混合物を識別し、さらにこのシグナルから各細胞タイプの寄与をある程度定量する能力はまた有用であろう。この目的のために、本実施例では、2つのサンプルの(すなわち配列の読取りに基づく)合成混合物を調べた。我々は、2つのMNaseデータセット(MCF.7及びNA12878.MNase)及び2つのcfDNAデータセット(腫瘍2349及びバルクcfDNA)について、5:95、10:90、15:85、20:80、30:70、40:60、50:50、60:40、30:70、80:20、90:10及び95:5の割合でシーケンシング読取りを混合した。合成MNase混合データセットは、2セットの169.9百万のアラインメントされた読取り(各々は元のサンプルの1つに由来する)から引き出され、合成cfDNA混合データセットは、2セットの181.1百万のアラインメントされた読取り(各々は元のサンプルの1つに由来する)から引き出された。

図10は、図3と同様に染色体11についての平均強度を示すが、ただしこれらの合成混合物に対するものである。異なるサンプルの寄与がどのように全体的周期の強度パターンでシフトを引き起こすかが、図10から理解できる。このシグナルを利用して合成混合物の割合を推論することができる。図11はMNaseデータセット混合物の最初の2つの主要成分を示し、図12はcfDNAデータセット混合物の最初の2つの主要成分を示す。両事例において、最初のPCは混合データセットの組成を直接捕捉する。したがって、適切な参照セットが与えられ、例えば回帰モデルを用いるならば、2つ及びことによると3つ以上の細胞タイプについてどれくらいの混合割合かを周期の強度データの変換から推論できることは容易に考えられる。図13は両データセットの樹形図を示し、cfDNAとMNaseサンプルの分離と同様に、類似するサンプル割合に由来する混合サンプルの全体的な類似性が確認される。

本実施例の重要な観察の1つは、未知サンプルに対する多様なサンプルタイプ(cfDNA又は細胞/組織タイプ)の混合割合はヌクレオソーム占拠パターンのモデリングによって推定できるということである。

【実施例 3】

【0038】

cfDNAシーケンシングデータを用いる転写因子結合サイトに対するヌクレオソーム占拠の測定

以前の実施例は、ヌクレオソーム配置シグナルは、連続でオーバーラップしない10kbpウィンドウにゲノムを分割することによって入手できることを示すが、オルトゴナルな方法を用いて切断接近容易性マップを作製することもまた可能であり、この方法はウィンドウサイズ及び境界に起因する人工物を生じにくい。本実施例でいくらか詳細に究明されるそのような1つの方法は、転写因子(TF)結合サイト周辺の読取り開始点で観察される周期性によるヌクレオソーム配置の推測である。

局所的ヌクレオソーム配置は近くのTF占拠によって影響されることは明瞭に確立されている。クロマチンの局所再構築における影響、したがって近くのヌクレオソームの安定な配置における影響は、TFセットにわたって均一ではない。あるTFの占拠はヌクレオソーム配置に局所的な影響を有することができ、前記はもっぱら結合サイトの5'又は3'側であり、特定の細胞タイプで長いあるいは短いゲノム距離に及ぶ。さらにまた、かつ本開示の目的のために重要なことであるが、個々の細胞においてin vivoで占拠されるTF結合サイトセットは組織及び細胞タイプ間で変動し、したがって、もし問題の組織又は細胞タイプ

のTF結合サイト占拠マップを同定し、さらにこのプロセスを1つ以上のTFについて反復できるならば、1つ以上の細胞タイプ特異的又は組織特異的TF結合サイト占拠プロフィールの濃縮又は枯渇を明らかにすることによって、cfDNAの集団に寄与する細胞タイプ及び組織の混合物の成分を同定することができよう。

この着想を実証するために、TF結合サイトの近隣の読取り開始点を用いて、優先的な局所ヌクレオソーム配置を反映する切断偏向を視覚的に確認した。ChIP-seq転写因子(TF)ピークをDNAエレメントエンサイクロペディア(“ENCODE”)プロジェクト(National Human Genome Research Institute, National Institutes of Health, Bethesda, MD)から入手した。これらのピークのゲノム間隔は広いので(平均して200から400bp)、保存的p-値カットオフ(1×10^{-5} 、詳細については方法の項を参照されたい)を用い対応する結合モチーフについてゲノムをインフォーマティブにスキャンすることによって、これら間隔内の活性化結合サイトを見つけた。続いて、独立して導かれたこれら2つの予想されるTF結合サイトのセットのインターセクションを下流分析に持ち込んだ。

【0039】

各候補TF結合サイトの500bp内の各位置における読取り開始点の数を、少なくとも1億の配列を有するサンプルで計算した。各サンプルで全ての読取り開始点を各位置で合計し、TF認識配列の長さに応じて、TFにつきサンプル当たり合計1,014から1,019の位置を得た。

図14は、多様な種々のサンプルでヒトゲノムの24,666のCTCF結合サイト周辺の、当該結合サイトそのものの周辺を中心とする読取り開始点の分布を示す。CTCFはインスレーター結合タンパク質であり、転写抑制で主要な役割を果たす。以前の研究は、CTCF結合サイトは局所ヌクレオソーム配置の拠り所であり、したがって少なくとも20ヌクレオソームがある結合サイトの周辺で約185bpの周期で対称的にかつ規則的に間隔をおいて配置されることを示唆した。図14のサンプルのほぼ全てで共通の1つの驚くべき特色は、結合サイトの上流及び下流の両方におけるヌクレオソーム配置の明瞭な周期性であり、CTCFの*in vivo*結合の局所的かつ大いに対称的な影響が多様なcfDNA及びMNase消化サンプルで反復されることを示唆している。興味深いことに、上流及び下流ピークの周期性はサンプルセット全体で均一ではなく、MNase消化サンプルは、結合サイトに対してピークのわずかに広い間隔の取り方を示し、ピークの強度だけでなくそれらの周期の有用性を示唆している。

図15は5,644のc-Jun結合サイト周辺の読取り開始点の分布を示す。この図では、よく知られている周期性がまたいくつかのサンプルについて視覚的に識別できるが、その影響は均一ではない。注目すべきことに、MNase消化サンプルのうち3つ(Hap1.MNase、HEK.MNase、及びNA12878.MNase)は非常に平坦な分布を有し、これは、これらの細胞ではc-Jun結合サイトは強く占拠されていないか、又はこれらの細胞タイプでは局所クロマチン再構築におけるc-Jun結合の影響はあまり明瞭ではないことを示しているのかもしれない。内在するメカニズムが何であれ、局所の読取り開始点近くにおける偏向がTF毎にさらにサンプルタイプ間で変動するという観察は、cfDNAサンプルの供給源組織組成に相関させ又は前記を解きほぐすために読取り開始点に基づいてヌクレオソーム占拠を推測するというその潜在的役割を強化する。

図16は4,417のNF-YB結合サイト周辺の読取り開始点の分布を示す。これらのTF結合サイト近くの開始サイト分布が対称性から外れていることが示されている。すなわち下流の影響(各図の右側)は、cfDNAサンプルのわずかに上向きの曲線によって証明されるように、上流の影響よりも強いように思われる。さらにまた注目すべきことは、MNase消化サンプルとcfDNAサンプル間の相違である。すなわち、前者は、ピークの識別が困難な平均してより平坦なプロフィールを示し、一方、後者は、より明瞭に識別しうる周期性及びより識別しやすいピークを有する。

【0040】

実施例1-3の方法

臨床サンプル及びコントロールサンプル

妊娠第三期の日常的胎児検診で妊婦fgs002、fgs003、fgs004及びfgs005から全血を採取し、EDTA(BD)を含むバキュテイナー(Vacutainer)チューブで一時的に保存した。妊婦

10

20

30

40

50

IM1、GM1及びGM2の全血は、それぞれ妊娠18、13及び10週に入手され、EDTA (BD) を含むバキュテイナーチューブで一時的に保存した。神経膠腫患者2349、2350、2351、及び2353の全血を脳外科手術手順の一部として収集し、EDTA (BD) を含むバキュテイナーチューブで3時間以内保存した。男性コントロール2 (MC2) (健康な成人男性) の全血はEDTA (BD) を含むバキュテイナーチューブで収集された。各個体について4から10mLの血液が利用可能であった。遠心分離 (1,000 x g、10分、4) によって全血から血漿を分離し、その後で上清を収集し、再び遠心分離した (2,000 x g、15分、4) 。精製血漿は1mLアリコットで-80 にて使用まで保存された。

バルクヒト血漿 (数が明らかでない健康個体の提供物を含む) をSTEMCELLテクノロジーズ (Vancouver, British Columbia, Canada) から入手し、2mLアリコットで-80 にて使用まで保存した。

10

血漿サンプルの処理

凍結された血漿アリコットを使用直前にベンチトップで融解した。循環cfDNAは、QiaAMP 循環核酸キット (Qiagen, Venlo, Netherlands) を用い製造業者のプロトコルに従って各血漿サンプルの2mLから精製された。DNAは、キュビットフルオロメーター (Qubit fluorometer; Invitrogen, Carlsbad, California) 及びカスタムqPCRアッセイ (ヒトAlu配列を標的とする) で定量した。

【0041】

MNase消化

標準的方法を用いて各細胞株 (GM12878、HeLa S3、HEK、Hap1) で約5千万の細胞を増殖させた。増殖培地を吸引しPBSで細胞を洗浄した。細胞をトリプシン処理し、2x 体積のCSS培地で中和し、遠心分離 (1,300rpm、5分、4) により円錐チューブ中でペレット状にした。細胞ペレットを12mLの氷冷PBSに再懸濁し1x プロテアーゼ阻害剤カクテルを加えて計測し、続いて遠心分離 (1,300rpm、5分、4) によりペレット状にした。細胞ペレットをRSB緩衝液 (10mM Tris-HCl、10mM NaCl、3mM MgCl₂、0.5mM スペルミジン、0.02% NP-40、1x プロテアーゼ阻害剤カクテル) で3百万細胞/mLの濃度に再懸濁し、氷上で穏やかに上下逆転させながら10分間インキュベートした。核を遠心分離 (1,300rpm、5分、4) によりペレット状にした。ペレット化した核をNSB緩衝液 (25%グリセロール、5mM MgAc₂、5mM HEPES、0.08mM EDTA、0.5mM スペルミジン、1mM DTT、1x プロテアーゼ阻害剤カクテル) で15M/mLの最終濃度に再懸濁させた。核を再び遠心分離 (1,300rpm、5分、4) によりペレット状にし、MN緩衝液 (500mM Tris-HCl、10mM NaCl、3mM MgCl₂、1mM CaCl₂、1x プロテアーゼ阻害剤カクテル) に30M/mLの最終濃度に再懸濁した。核を200 µLアリコットに分割し、4Uのマイクロコッカスヌクレアーゼ (Worthington Biochemical Corp., Lakewood, NJ, USA) により37 で5分間消化した。氷上で85 µLのMNSTOP緩衝液 (500mM NaCl、50mM EDTA、0.07% NP-40、1x プロテアーゼ阻害剤カクテル) を添加して反応を停止させ、続いて穏やかに上下逆転させながら4 で90分間インキュベートした。フェノール：クロロホルム：イソアミルアルコール抽出を用いてDNAを精製した。標準的方法を用い2%アガロースゲル電気泳動によりモノヌクレオソームフラグメントをサイズ選別し、ナノドロップ分光光度計 (Thermo Fisher Scientific Inc., Waltham, MA, USA) で定量した。

20

30

シーケンシングライブラリーの調製

全サンプルについてバーコード付加シーケンシングライブラリーをThruPLEX-FD又はThruPLEX DNA-seq 48Dキット (Rubicon Genomics, Ann Arbor, Michigan) (末端修復、連結及び増幅反応の独占的シリーズを含む) により調製した。全ての臨床サンプルライブラリーについて3.0から10.0ngのDNAをインプットとして用いた。2つのバルク血漿cfDNAライブラリーは各ライブラリーについて30ngで構築した (各ライブラリーは別個にバーコードを付加された)。MC2の2つのライブラリーは各ライブラリーについて2ngのインプットで構築し、各ライブラリーは別個にバーコードを付加された。MNase消化細胞株の各々のライブラリーは20ngのサイズ選別インプットDNAを用いて構築された。全てのサンプルのライブラリー増幅をリアルタイムPCRによってモニターし、過剰増幅を回避した。

40

【0042】

50

シーケンシング

9bpのインデックス読取りを含むペア形成末端101bp読取りを用いて、HiSeq2000装置（Illumina, Inc., San Diego, CA, USA）で全てのライブラリーをシーケンシングした。プールサンプルfgs002、fgs003、fgs004及びfgs005について1レーンのシーケンシングを実施し、各サンプルにつき合計して約 4.5×10^7 の読取りペアを得た。サンプルIM1、GM1及びGM2をいくつかのレーンにわたってシーケンシングし、それぞれ 1.2×10^9 、 8.4×10^8 及び 7.6×10^7 の読取りペアを生じた。1レーンのシーケンシングをサンプル2349、2350、2351及び2353の各々について実施し、各サンプル当たり約 2.0×10^8 の読取りペアを各サンプルについて得た。4つの細胞株のMNase消化ライブラリーの各々について1レーンのシーケンシングを実施し、各ライブラリーにつき約 2.0×10^8 の読取りペアを得た。2つの複製MC2ライブラリーの1つについて4レーンのシーケンシングを、さらに2つの複製バルク血漿ライブラリーの1つについて3レーンのシーケンシングを実施し、それぞれ合計 10.6×10^9 及び 7.8×10^8 の読取りペアを得た。

10

cfDNAシーケンシングデータの処理

cfDNA及びMNaseライブラリーの両方についてDNA挿入物サイズは短い傾向があり（大半のデータで80bpから240bp）、したがって、いくつかの分子の読取り末端においてアダプター配列が期待された。読取り末端から開始するアダプター配列をトリミングし、短い元の分子に関するペア形成末端（“PE”）データのフォワード及びリバース読取りを単一読取り（“SR”）に収斂させた。少なくとも11bpの読取りでオーバーラップするPE読取りをSRに収斂させた。30bpより短いか、又は5塩基を超えて10未満の品質スコアを示すSRは廃棄した。高速アラインメントツール（BWA-ALN又はBWA-MEM）を用いて、残りのPE及びSRデータをヒト参照ゲノム（GRCh37、1000Gレリースv2）とアラインメントさせた。得られたSAM（Sequence Alignment/Map（配列アラインメント/マップ））様式を分類済みBAM（Binary Sequence Alignment/Map（二進法配列アラインメント/マップ））様式）にSAMツールを用いて変換した。

20

【 0 0 4 3 】

追加される公的に利用可能なデータ

HeLa-S3 MNase実験（アクセッションSRR633612、SRR633613）及びMCF-7 MNase実験（アクセッションSRR999659-SRR999662）の公的に利用可能なPEデータをダウンロードして、上記に記載したようにプロセッシングした。

30

イルミナケンブリッジ社（Illumina Cambridge Ltd., Essex, UK）によって作製されたCEPH家系146個体NA12878の公的に利用可能なゲノムショットガンシーケンシングデータを欧州ヌクレオチドアーカイブ（ENA、アクセッションERR174324-ERR174329）から入手した。このデータをイルミナHiSeqプラットフォームで 2×101 bp読取りによりPEについて配列を調べ、さらにシーケンシング前にライブラリーをより長い挿入物サイズについて選別した。したがって、読取り末端のアダプター配列は必要とされず、このデータはしたがってBWA-MEMを用いて直接アラインメントした。

読取り末端の情報の抽出

PEデータは、シーケンシングライブラリーの調製で用いられるDNA分子の2つの物理的末端についての情報を提供する。この情報をSAMtoolsアプリケーションプログラミングインターフェース（API）を用いてBAMファイルから抽出した。PEデータの両方の外側アラインメント座標が用いられた（両方の読取りが同じ染色体とアラインメントされ、そこでは読取りは反対の向きを有する）。トリミングされていないSRデータについては、1つの読取り末端だけが元のDNA分子の物理的末端についての情報を提供する。読取りが参照ゲノムのプラス鎖とアラインメントされる場合、もっとも左の座標が用いられた。読取りがリバース鎖とアラインメントされる場合、そのもっとも右の座標が代わりに用いられた。PEデータがアダプタートリミングによって単一読取りデータに変換される事例では、両方の末端座標が考慮された。少なくとも5つのアダプター塩基がSRシーケンシング実験からトリミングされる場合には、両方の末端座標がまた考慮された。

40

ヒト参照配列の全常染色体（染色体1から染色体22）については、読取り末端の数及び

50

全ての位置のカバレッジが10,000塩基のウインドウ（ブロック）中に抽出された。あるブロック中にアラインメントされる読取りが存在しない場合、当該ブロックは当該特定のサンプルについて空であるとみなされた。

【0044】

平滑ピリオドグラム

読取り開始点とカバレッジとの比率を、各サンプルの空ではない各ブロックについて計算した。カバレッジが0である場合、比率は0と設定された。これらの比率を用いて、1/500塩基から1/100塩基の間の周波数で高速フーリエ変換（FFT、R統計プログラミング環境でspec.pgram）を用いて各ブロックのピリオドグラムを計算した。場合によって、平滑化へのパラメーター（3bpダニエルスムーサー；半分の重みを末端値に与える移動平均）及びデータのトレンド除去（例えばシリーズの平均を差し引く、さらに線形トレンドを除去する）を用いた。各ブロックについて周期レンジ120 - 250bpに対する強度をセーブした。

10

平均染色体強度

1セットのサンプルについて、全サンプルを通して空ではないブロックが同定された。ある特定の周期に対する強度が各常染色体について各サンプルの全ブロックを通して平均された。

主要成分分析及び樹形図

サンプルを通して空ではないブロックを収集した。主要成分分析（PCA；R統計プログラミング環境ではprcomp）を用いて、データの次元削減を実施し前記を二次元空間でプロットした。PCAは、データの大半の変動を捕捉して直交次元を構築する次元を識別し、当該データの変動量の減少を説明する。

20

サンプル強度間のペア毎のユークリッド距離を計算し、樹形図として可視化する（R統計プログラミング環境ではstatsライブラリー）。

転写因子結合サイト予測

推定的転写因子結合サイト（多数の細胞タイプを通して作製されたChIP-seqデータの分析から得られる）を、ENCODEプロジェクトから入手した。

MEMEソフトウェアパッケージ（バージョン4.10.0_1）のプログラムfimoを用いヒト参照ゲノム（GRCh37, 1000Gリリースv2）をスキャンすることによって、転写因子結合サイト候補の独立したセットを入手した。位置重み行列を用いてスキャンを実施した（前記行列は、JASPAR_CORE_2014_脊椎動物データベースから“--verbosity 1 --thresh 1e-5”のオプションを用いて入手した）。用いた転写因子モチーフ識別子はMA0139.1、MA0502.1及びMA0489.1であった。

30

ベッドツールv2.17.0を用いて予測されるサイトの両セットに由来する染色体座標をインターセクトした。プロット中の一切の非対称性を保存するために、“+”鎖上の予測結合サイトのみを用いた。読取り開始点が予測結合サイトのどちらかの末端から500bp内であれば、各サンプルについて当該読取り開始点を記録し、全てのそのようなサイトにわたってその位置によってサンプル内で合計した。合計して少なくとも1億の読取りを有するサンプルのみをこの分析に用いた。

【実施例4】

【0045】

40

cfDNAからの正常/健康起源組織の決定

単一個体のcfDNAで観察される断片化パターンが、寄与細胞タイプ間で遺伝子型に相違が存在しないときでも、これらのフラグメントを生じた細胞の（したがってcfDNA分子集団の供給源組織の）ゲノム編成の証拠を含んでいる可能性を判定するために、cfDNAを高深度で配列を決定してcfDNAを生じるプロセスのより一層の理解に供した。得られたデータを用いてヌクレオソーム占拠のゲノムワイドマップを構築した。前記マップは他の研究者による以前の研究に基づいて構築されたが、実質的にはより包括的である。ライブラリー調製プロトコルを最適化して短いフラグメントを回収することによって、転写因子（例えばCTCF）のin vivo占拠もまたcfDNAによって直接フットプリントが得られることが見出された。最後に、調節エレメント及び遺伝子本体におけるヌクレオソームの間隔の取り方

50

は、健康な個体のcfDNAシーケンシングによって明らかにされたように、リンパ系及び骨髄系細胞株におけるDNase高感受性並びに遺伝子発現と極めて強く相関性を示すことが見出された。

cfDNAフラグメントはクロマトソームに対応し実質的なDNA損傷を含む

数が明らかでない健康個体（“BH01”）のプール血漿又は単一個体（“IH01”）の血漿から精製されたcfDNAフラグメントの末端修復及びアダプター連結によって、通常的なシーケンシングライブラリーを調製した（図17、表1）。

表1：血漿サンプルのシーケンシング統計

【0046】

【表1】

サンプルの名称	ライブラリーのタイプ	読取り	配列決定フラグメント	アラインメント	アインメント (Q30)	カバレッジ	複製物の概算%	35-80bp	120-180bp
BH01	DSP	2x101	1489569204	97.20%	88.85%	96.32	6.00%	0.65%	57.64%
IH01	DSP	2x101	1572050374	98.58%	90.60%	104.92	21.00%	0.77%	47.83%
IH02	SSP	2x50, 43/42	779794090	93.19%	75.27%	30.08	20.05%	21.83%	44.00%
CH01	-	-	3841413668	96.95%	86.81%	231.32	14.99%	5.00%	50.85%

SSP（一本鎖ライブラリー調製プロトコル）、DSP（二本鎖ライブラリー調製プロトコル）

【0047】

各サンプルについて、シーケンシング関連統計値を集計した。前記には以下が含まれる：配列を決定したフラグメントの総数、読取り（リード）の長さ、あるマッピング品質閾値で又はそのような閾値無しに参照に対してアラインメントを示したフラグメントのパーセンテージ、平均カバレッジ、重複率、及び2つの長さのbinにおける配列決定フラグメントの割合。フラグメント長はペアとなった末端の読取りのアラインメントから推定した。読取りの長さが短いために、カバレッジは完全なフラグメントが読み取られたと仮定することによって計算された。重複フラグメントの概数はフラグメント末端を基準にし、前記は高度に定型化された切断の存在下では真の重複率を過大評価する可能性がある。SSPは一本鎖ライブラリー調製プロトコルであり、DSPは二本鎖ライブラリー調製プロトコルである。

ライブラリー-BH01及びIH01は、それぞれ96重及び105重カバレッジで配列決定された（1.5G及び1.6Gフラグメント）。フラグメントの長さの分布（ペア形成末端読取りのアラインメントから推測）は、~167bp（クロマトソームに結合するDNAの長さとも一致）に支配的ピークを有し、100 - 160bpの長さのレンジで~10.4bpの周期性を有する（図18）。これらの分布は、cfDNAが、細胞死前及び細胞死後の両方でタンパク質（この事例ではヌクレオソームコア粒子及びリンカーヒストン）と結合することによってヌクレアーゼから優先的に保護されるモデル（ただしこの場合にはヌクレオソーム結合DNAのらせんピッチとの関係である程度の追加のニック及び切断が生じる）と一致する。さらにこのモデルを支持するのは、これら167bpフラグメントのジヌクレオチド組成であり、前記は、MNaseにより誘導されるヌクレオソーム結合フラグメントに関する以前の研究の重要な特色を概括し（例えば2個1組でA/Tジヌクレオチドに対する偏向）、さらにヌクレオソームコア粒子はクロマトソームに対して対称的に配置されるという見解を支持する（図19）。

cfDNAオントロジーにおけるこのモデルの予測は、広範なDNA損傷（例えば一本鎖ニックとともに5'及び3'オーバーハング）である。伝統的なライブラリー調製の最中に、ニックの付いた鎖は増幅されず、オーバーハングは末端修復によって平滑化され、短い二本鎖DNA（“dsDNA”）分子は単純に回復が困難である（前記分子は全cfDNAの実質的部分を占める）。これに対処するために、追加の健康な個体に由来する血漿運搬cfDNAの一本鎖シーケンシングライブラリー（‘IH02’）を、Gansaugeらによる古いDNAの研究（広範囲のDNA損傷及びヌクレオソーム周辺のヌクレアーゼ切断が報告されている）から改変したプロトコルを用いて調製した。略記すれば、cfDNAを変性させ、得られたフラグメントにビオチン結合一本鎖アダプターを連結させた。続いて、この連結フラグメントを第二の鎖

10

20

30

40

50

の合成、末端修復及び第二のアダプター連結に付し、その間このフラグメントをストレプトアビジンビーズに固定した。最後に、最少PCR増幅を実施してアダプター保有分子を濃縮し、一方サンプルインデックスもまた添付した(図20、表2)。

【0048】

表2: 一本鎖シーケンシングライブラリーの調製で用いられた合成オリゴ

【表2】

オリゴの名称	配列 (5'-3')	付記
CL9	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT	HPLC精製
アダプター-2.1	CGACGCTCTTCCGATC/ddT/	HPLC精製
アダプター-2.2	/5Phos/AGATCGGAAGAGCGTCGTGTAGGGAAAGAG*T*G*T*A	HPLC精製
CL78	/5Phos/AGATCGGAAG/iSpC3/iSpC3/iSpC3/iSpC3/iSpC3/iSpC3/iSpC3/iSpC3/iSpC3/3BioTEG/	二重HPLC精製

10

20

30

40

50

【0049】

IH02について、得られたライブラリーを30重カバレッジで配列を決定した(779Mフラグメント)。フラグメントの長さの分布は、クロマトソームに対応して再度~167bpに支配的ピークを示したが、通常のライブラリー調製と比較してより短いフラグメントが顕著に濃縮されていた(図21、22、23A-B、24A-B)。全てのライブラリーは~10.4bp周期性を示すが、フラグメントサイズはこの2つの方法について3bp差し引きされ、その真の末端が一本鎖ライブラリーでより忠実に表される、損傷した又はノンフラッシュインプット分子と一致する。

【0050】

高深度cfDNAシーケンシングに基づくin vivoヌクレオソーム保護のゲノムワイドマップ cfDNAに寄与する組織のヒトゲノム全体に対するヌクレオソームの優先的な局所的位置を、アラインメントされたフラグメント末端の分布又はその数学的変換を1つ以上の参照マップと比較することによって推測することが可能か否かを評価するために、ウインドウ化保護スコア("WPS")を開発した。具体的には、cfDNAフラグメント末端はヌクレオソーム境界に隣接してクラスターを形成し、一方ヌクレオソームそれ自体の上では枯渇するはずであると予想された。これを定量化するために、WPSを開発した。WPSは、与えられたゲノム座標を中心とする120bpウインドウを完全にわたるDNAフラグメントの数から同じウインドウ内部に末端を有するフラグメントの数を差し引いたものを表す(図25)。意図したように、WPSの値は、in vitroの方法又は古いDNAを用いた他のグループによってマッピングされたように、強力的に位置決めされたアレイ内のヌクレオソームの場所と相関性を有する(図26)。他のサイトでは、WPSは、ゲノム上の特徴(例えばDNase I高感受性(DHS)サイト)と相関関係を有する(例えば遠位の調節エレメントにフランキングするヌクレオソームの再配置と一致する)(図27)。

ヒューリスティックアルゴリズムをBH01、IH01及びIH02データセットのゲノムワイドWPSに適用して、それぞれ12.6M、11.9M、及び9.7Mの極大ヌクレオソーム保護を認定した(図25-31)。各サンプルで、隣接ピーク間の距離の分布の様子は185bpで分散は少なく、一般的にヒト又はマウス細胞におけるヌクレオソームリピートの以前の分析と一致した。

ピークコールの位置がサンプルを通して同様であるか否かを決定するために、サンプルの各ピークについて他のサンプルの各々の最も近いピークに対するゲノム距離を計算した。高い一致が観察された(図31、図32A-C)。BH01ピークコールから最も近い近傍のIH01ピークコールまでの距離の中央値(絶対値)は全体として23bpであったが、もっとも高いスコアを与えられたピークについては10bp未満であった(図33A-B)。

ヌクレアーゼ特異性によって又はライブラリー調製中に導入される偏向はヌクレオソーム保護のシグナルに人為的寄与を与える可能性があるため、フラグメント末端もシミュレ

ーションし、各サンプルの深度、サイズ分布及び末端ジヌクレオチド頻度についてマッチングさせた。続いてゲノムワイドWPSを計算し、それぞれBH01、IH01及びIH02と適合させたシミュレーションデータセットについて、10.3M、10.2M及び8.0Mが同じヒューリスティックアルゴリズムによって極大値としてコールされた。シミュレーションデータセットのピークは真のデータセットのピークよりも低いスコアを伴っていた(図33A - B)。さらにまた、真のデータベースからコールされたピークの相対的に再現性を有する場所(図31、図32A - C)は、シミュレーションデータセットからコールされたピークの場所とは良好なアラインメントを示さなかった(図31、図34A - C)。

【0051】

ゲノムワイドのヌクレオソームマップの正確さ及び完全さを改善するために、BH01、IH01及びIH02のcfDNAシーケンシングデータをプールし、1つにまとめた231重カバレッジについて再分析した(‘CH01’; 3.8Bフラグメント; 表1)。WPSを計算し、この合体サンプルについて12.9Mピークがコールされた。このピークコールセットはより高いスコアと関係し、ピークの数に関して飽和に近づいていた(図33A - B)。500bp未満の全てのピークからピークまでの距離を考慮すると(図35)、CH01ピークセットはヒト参照ゲノムの2.53ギガ塩基にわたる。

ヌクレオソームは、遺伝子調節の目印(例えば転写開始サイト及びエクソン-イントロン境界)に対応して適切に配置されていることが知られている。そのような理解と一致して、このデータでも同様に、転写、翻訳及びスプライシングの目印に対応して同じような配置が観察された(図36 - 40)。転写活性を有するヌクレオソームの間隔の取り方とクロマチンマークとの間の相関性に関する過去の観察に基づいて、リンパ芽球系細胞株のロングレンジ相互作用(in situ Hi-C)を基準にして区画A(オープンクロマチンに豊富)又は区画B(閉鎖クロマチンに豊富)に割り当てた100キロベース(kb)ウィンドウ内のピークからピークまでの間隔の中央値を調べた。区画Aのヌクレオソームは、区画Bのヌクレオソームよりも密な間隔の取り方を示し(Aの中央値は187bpに対しBの中央値は190bp)、ある種のサブ区画間においては更なる相違を示した(図41)。染色体の全長にわたって、一般的パターンは認められなかったが、ただしヌクレオソームの間隔の取り方の中央値はセントロメア周辺領域で急激に下降し、アルファサテライトのアレイにわたる強力な位置決めによって駆動されていた(171bpのモノマー長さ; 図42、図26)。

【0052】

短cfDNAはCTCF及び他の転写因子のフットプリントを直接検証する

DNase I切断パターンに関する以前の研究は2つの支配的なフラグメントクラスを識別した(すなわち、ヌクレオソーム間の切断と関連する、より長いフラグメント、及び転写因子結合サイト(TFBS)に隣接する切断と関連する、より短いフラグメント)。in vivo由来cfDNAフラグメントもまた2つのヌクレオソーム感受性クラスから生じるのか否かを査定するために、推測されるフラグメントの長さを基準にして配列の読取り(CH01)を分けし、さらに、長フラグメント(120 - 180bp; 120bpウィンドウ; ヌクレオソームコールのために上記に記載したWPSと効果としては同じ)又は短フラグメント(35 - 80bp; 16bpウィンドウ)を別々に用いてWPSを再度計算した(図26 - 27)。我々のデータ中で活性的に結合されるサイトが濃縮された明確なTFBSのセットを得るために、各TFについてクラスター化FIMO予想をENCODE(TfbsClusteredV3)由来のChIP-seqピークの統合セットとインターセクトした。

長画分WPSはCTCF結合サイトの近傍でヌクレオソームの強力な編成を支援する(図43)。しかしながら、CTCF結合サイトそのものと一致する、短画分WPSの強力なシグナルもまた観察された(図44 - 45)。CTCF結合サイトをそれらがin vivoで結合されるという仮定に基づいて層別化した(全FIMO予想v.s.ENCODE ChIP-seqとインターセクトするサブセットv.s.19細胞株を通して利用されると思われるものとインターセクトする更なるサブセット)。実験的に明瞭に支持されるCTCFサイトは、長画分WPSに基づけばランキングする-1と+1ヌクレオソームとの間でかなり広い間隔を示し、CTCF結合時のそれらの再配置と一致した(~190bp - > ~260bp; 図45 - 48)。さらにまた、実験的に明瞭に支持されるCTCF

10

20

30

40

50

サイトは、CTCF結合サイトそのものよりもはるかに強いシグナルを短画分WPSについて示す(図49-52)。

同様な分析をFIMO予想及びENCODE Chip-seqデータの両方が入手可能であった追加のTFについて実施した(図53A-H)。これらのTF(例えばETS及びMAFK(図54-55))の多くについて、短画分のフットプリントが観察され、長画分WPSに周期性シグナルを伴った。これは、結合TFBSを取り囲むヌクレオソームの強力な配置決めと一致する。総合すれば、これらのデータは、短いcfDNAフラグメント(前記は一本鎖プロトコルによってはるかに良好に回収される(図18、図21))はDNA結合転写因子(CTCF及び他のものを含む)の*in vivo*占拠の直接のフットプリントであるという見解を支持する。

【0053】

ヌクレオソームの間隔の取り方はcfDNAの供給源組織の情報を提供する

cfDNAシーケンシングを介して測定される*in vivo*ヌクレオソーム保護を用いて、健康な個体のcfDNAに寄与する細胞タイプを推測できるか否かを決定するために、116の多岐にわたる生物学的サンプルで明らかにされるDHSサイト内のヌクレオソームコールのピークからピークまでの間隔を調べた。調節エレメントにおいて-1と+1ヌクレオソーム間で間隔が広がるのが以前に観察された(例えば部分的にはDHSサイトで(図27)又は包括的には結合CTCFサイトで(図45))。結合されたCTCFサイトと同様に、かなり広い間隔がDHSサイトのサブセット内でヌクレオソームペアについて観察され、おそらくcfDNAを生じる細胞タイプにおいて、転写因子結合介入によりヌクレオソームが再配置されるサイトと一致すると思われる(〜190bp ->〜260bp; 図56)。実際、拡大されるヌクレオソームの間隔(〜260bp)の割合は、どの細胞タイプのDHSサイトが用いられるかにしたがって顕著に変動する。しかしながら、この割合が最も高い供給源細胞タイプはいずれもリンパ系又は骨髄系である(例えば図56のCD3_CB-DS17706など)。これは、健康な個体のcfDNAの支配的な供給源としての造血細胞死と一致する。

次に、転写開始サイトの近傍のヌクレオソーム保護シグナルを再度調べた(図36)。シグナルを類リンパ細胞系列の細胞株(NB-4)の遺伝子発現に基づいて階層化したとき、TSSに対応してヌクレオソーム保護の場所及び強度に強い相違が観察され、高度に発現される遺伝子で強く、発現が弱い遺伝子で低かった(図57)。さらにまた、短画分WPSはTSSのすぐ上流に明瞭なフットプリントを示し、その強度はまた発現レベルと強く相関した(図58)。これは、おそらく転写が活性な遺伝子における転写前開始複合体(又はその何らかの成分)のフットプリントを反映している。

これらのデータは、cfDNA断片化パターンは、実際のところcfDNAを生じる組織又は細胞タイプの予測に用いることが可能なシグナルを確かに含んでいることを示している。

しかしながら、問題は、ゲノムワイドのcfDNAライブラリーでは比較的わずかな読取りしかDHSサイト及び転写開始サイトと直接オーバーラップしないということである。

【0054】

ヌクレオソームの間隔は、細胞間で、かつクロマチン状態及び遺伝子発現の関数として変動する。一般的には、オープンクロマチン及び転写はより短いヌクレオソームリピートの長さに関係し、本実施例の区画A対Bの分析と一致する(図41)。本実施例のピークコールデータはまた、遺伝子本体を通してのヌクレオソームの間隔の取り方とそれらの発現レベルとの間の相関性を示し、より密な間隔がより高い発現と結びついている(図59; $r = -0.17$; $n = 19,677$ 遺伝子)。この相関性は、隣接領域と比較して遺伝子本体そのものについて最も高い(上流10kb $r = -0.08$; 下流10kb $r = -0.01$)。分析が少なくとも60ヌクレオソームコールに広がる遺伝子本体に限定される場合、ヌクレオソームの間隔の取り方のより密なものが遺伝子発現とさらに強い相関性を示す($r = -0.50$; $n = 12,344$ 遺伝子)。

例えば遺伝子本体又は他のドメインにおけるヌクレオソームの間隔の取り方のようなシグナルを探索する1つの利点は、はるかに大きな割合のcfDNAが情報を提供するであろうということである。別の潜在的な利点は、cfDNAに寄与する複数の細胞タイプから生じるシグナル混合物を検出できる可能性である。これを試験するために、更なる数学的変換(高速フーリエ変換(FFT))を、遺伝子本体の最初の10kbにわたる長フラグメントWPSで、及

10

20

30

40

50

び遺伝子ごとに実施した。FFTシグナルの強度は特異的な周波数レンジで遺伝子発現と相関し、正の相関性については177 - 180bpで最大で、負の相関性では～199bpで最小であった(図60)。ヒト細胞株及び一次組織の76の発現データセットに対してこの分析を実施したとき、もっとも強い相関性は造血細胞系列に関するものであった(図60)。例えば、3つの健康なサンプル(BH01、IH01、IH02)の各々について193 - 199bpの周期レンジで平均強度を有する最高ランクの負の相関性は、いずれもリンパ系細胞株、骨髄系細胞株又は骨髄組織に対するものであった(図61、表3)。

【 0 0 5 5 】

表3：WPS FFT強度と遺伝子発現データセットとの相関性

【 0 0 5 6 】

【表 3】

RNA名称	カテゴリ	タイプ	詳細	相関関係								健康	ランク相違					
				IC15	IC20	IC47	IC37	IC35	IC45	IC20	IC47		IC37	IC35				
A.431	皮膚	皮膚癌	類表皮癌 (扁平上皮細胞)	-0.296	-0.186	-0.149	-0.200	-0.140	-0.178	-0.195	-0.176	2	3	-9	-9	-12	-21	
A549	肺	肺癌	肺癌細胞株	-0.289	-0.185	-0.144	-0.202	-0.139	-0.172	-0.166	-0.170	3	-14	-12	-9	-2	-13	
	脂肪組織	一次組織	脂肪組織	一次組織	-0.270	-0.169	-0.137	-0.168	-0.121	-0.153	-0.166	-0.148	1	12	5	0	14	12
	副腎	一次組織	副腎	一次組織	-0.257	-0.158	-0.131	-0.173	-0.118	-0.145	-0.161	-0.136	-2	-11	-5	1	5	8
AN3.CA	乳房/ 女性生殖器	子宮癌	転移性 子宮内膜腺 癌細胞株	-0.303	-0.194	-0.157	-0.213	-0.147	-0.183	-0.195	-0.171	-4	-16	-13	-15	-8	-2	
虫垂	一次組織	虫垂	一次組織	-0.287	-0.185	-0.137	-0.168	-0.118	-0.148	-0.171	-0.152	6	24	20	23	8	9	
BEWO	その他	子宮癌	転移性絨毛 癌細胞株	-0.284	-0.184	-0.147	-0.193	-0.139	-0.173	-0.194	-0.173	-5	3	-12	-15	-19	-27	
	骨髓	一次組織	骨髓	一次組織	-0.343	-0.230	-0.165	-0.192	-0.142	-0.167	-0.193	-0.165	2	40	9	30	16	28
CACO.2	腹部	結腸腺癌	結腸腺 癌細胞株	-0.281	-0.177	-0.137	-0.192	-0.128	-0.169	-0.184	-0.164	5	-5	-5	-14	-10	-9	
CAPAN.2	腹部	膵臓腺癌	膵臓腺 癌細胞株	-0.291	-0.187	-0.145	-0.202	-0.136	-0.176	-0.195	-0.175	3	-12	-2	-18	-19	-25	
	大脳皮質	一次組織	大脳皮質	一次組織	-0.225	-0.136	-0.120	-0.168	-0.106	-0.134	-0.142	-0.125	-1	-9	-3	0	0	0
	結腸	一次組織	結腸	一次組織	-0.261	-0.162	-0.124	-0.164	-0.111	-0.145	-0.166	-0.146	7	8	8	6	-7	1
Daudi	リンバ系	ヒトT-キラー リンパ腫	ヒトT-キラー リンパ腫 細胞株	-0.321	-0.206	-0.153	-0.195	-0.133	-0.165	-0.189	-0.160	4	17	19	19	13	24	
	十二指腸	一次組織	十二指腸	一次組織	-0.261	-0.164	-0.122	-0.159	-0.109	-0.144	-0.166	-0.144	10	10	10	7	-4	7
EFO.21	乳房/ 女性生殖器	卵巣癌	転移性 卵巣漿液性 嚢胞腺 癌細胞株	-0.267	-0.186	-0.149	-0.201	-0.140	-0.176	-0.168	-0.169	-7	-9	-14	-20	-1	-8	
	子宮内膜	一次組織	子宮内膜	一次組織	-0.257	-0.158	-0.132	-0.178	-0.119	-0.151	-0.166	-0.151	-3	-11	-4	-8	-3	-12
	食道	一次組織	食道	一次組織	-0.237	-0.147	-0.124	-0.156	-0.116	-0.141	-0.158	-0.145	-3	1	-7	0	0	-7
	卵管	一次組織	卵管	一次組織	-0.247	-0.157	-0.129	-0.171	-0.114	-0.145	-0.161	-0.145	-4	-13	-2	-3	3	-2
	胆嚢	一次組織	胆嚢	一次組織	-0.249	-0.156	-0.119	-0.153	-0.103	-0.138	-0.154	-0.141	4	4	4	3	4	1
HsCaT	皮膚	メラノサイト 細胞株	メラノサイト 細胞株	-0.290	-0.186	-0.149	-0.194	-0.142	-0.173	-0.193	-0.173	-5	7	-18	-8	-8	-17	
HDLM.2	リンバ系	ホジキン リンパ腫	ホジキン リンパ腫 細胞株	-0.316	-0.200	-0.154	-0.201	-0.136	-0.173	-0.195	-0.171	1	6	11	1	-5	-5	
	心筋	一次組織	心筋	一次組織	-0.246	-0.149	-0.126	-0.166	-0.113	-0.141	-0.155	-0.140	-3	-3	-3	0	3	2
HEK_293	その他	腎臓 副腎 前駆 細胞株	77c/ウルス 5型によって 形質転換 された胎児 腎細胞株	-0.292	-0.187	-0.150	-0.209	-0.139	-0.172	-0.189	-0.168	-4	-17	-4	-2	3	0	
HEL	骨髓系	赤白血病	赤白血病 細胞株 (ホジキン病の 治療後再発 したAML M6)	-0.324	-0.205	-0.161	-0.210	-0.140	-0.172	-0.194	-0.168	-1	-5	4	12	5	14	
HeLa	乳房/ 女性生殖器	子宮頸癌	子宮頸 上皮腺癌 細胞株	-0.296	-0.186	-0.149	-0.203	-0.139	-0.172	-0.190	-0.171	1	-10	-5	-3	1	-8	
Hep_G2	腹部	肝細胞癌	肝細胞癌 細胞株	-0.294	-0.186	-0.152	-0.202	-0.145	-0.166	-0.196	-0.167	-4	-6	-18	-24	-17	2	
HL_60	骨髓系	前骨髄 細胞性 白血病	急性前骨髄 細胞性白血病 (APL) 細胞株	-0.332	-0.208	-0.161	-0.202	-0.137	-0.171	-0.197	-0.170	2	8	18	18	-1	11	
HMC.1	骨髓系	マスト細胞 白血病	マスト細胞 白血病細胞株	-0.337	-0.228	-0.165	-0.212	-0.149	-0.181	-0.199	-0.180	0	-1	-2	3	0	-2	
K.562	リンバ系	白血病	慢性骨髄性 白血病 (CML) 細胞株	-0.317	-0.202	-0.158	-0.211	-0.143	-0.178	-0.195	-0.166	-3	-9	-5	-6	-1	13	
Karpas.707	リンバ系	多発性 ミエローマ	多発性ミエロ-マ 細胞株	-0.325	-0.210	-0.155	-0.195	-0.136	-0.167	-0.188	-0.164	4	20	18	22	21	22	
	腎臓	一次組織	腎臓	一次組織	-0.245	-0.150	-0.130	-0.168	-0.119	-0.153	-0.171	-0.147	-7	-4	-12	-19	-21	-6
	肝臓	一次組織	肝臓	一次組織	-0.248	-0.148	-0.122	-0.150	-0.110	-0.150	-0.164	-0.138	1	4	-1	-13	-4	3
	肺臓	一次組織	肺臓	一次組織	-0.264	-0.170	-0.133	-0.170	-0.121	-0.148	-0.167	-0.149	3	4	0	7	3	6

10

20

30

40

RNA名称	カテゴリー	タイプ	詳細	相関関係							健康	ランク相違					
				IC15	IC20	IC17	IC37	IC35	IC15	IC20		IC17	IC37	IC35			
リンパ節	一次組織	リンパ節	一次組織	-0.308	-0.195	-0.148	-0.182	-0.128	-0.155	-0.181	-0.156	7	24	17	25	14	22
MCF7	乳房/ 女性生殖器	乳癌	転移性乳房腺 癌細胞株	-0.298	-0.195	-0.154	-0.207	-0.145	-0.183	-0.198	-0.181	-3	-9	-12	-18	-11	-19
MOLT.4	リンパ系	白血病 (ALL)	急性リンパ 芽球性白血病 (T-ALL) 細胞株	-0.323	-0.204	-0.163	-0.212	-0.144	-0.177	-0.197	-0.173	-3	-7	-2	-1	-3	-1
NB.4	骨髄系	前骨髄 細胞性 白血病	急性前骨髄 細胞性白血病 (APL) 細胞株	-0.348	-0.228	-0.172	-0.211	-0.148	-0.182	-0.202	-0.171	0	4	3	5	2	13
NTERA.2	泌尿器/ 男性生殖器	泌尿器 癌	転移性胎生期 癌細胞株 (TERA-2から クローニング)	-0.268	-0.170	-0.137	-0.193	-0.117	-0.157	-0.169	-0.153	-2	-8	16	-2	5	0
卵巣	一次組織	卵巣	一次組織	-0.268	-0.162	-0.135	-0.181	-0.120	-0.152	-0.166	-0.151	1	-7	2	-2	6	-1
膵臓	一次組織	膵臓	一次組織	-0.250	-0.159	-0.132	-0.170	-0.116	-0.150	-0.166	-0.150	-5	-5	1	-6	-5	-7
PC.3	泌尿器/ 男性生殖器	前立腺 癌	急性 低分化性 前立腺 腺癌細胞株	-0.295	-0.190	-0.151	-0.204	-0.138	-0.174	-0.188	-0.173	-3	-10	2	-6	8	-12
胎盤	一次組織	胎盤	一次組織	-0.266	-0.166	-0.134	-0.168	-0.126	-0.151	-0.166	-0.150	3	10	-7	1	9	6
前立腺	一次組織	前立腺	一次組織	-0.248	-0.161	-0.133	-0.175	-0.123	-0.150	-0.165	-0.151	-8	-10	-11	-6	1	-12
直腸	一次組織	直腸	一次組織	-0.255	-0.154	-0.117	-0.159	-0.102	-0.138	-0.161	-0.142	6	0	5	4	-2	0
REH	リンパ系	白血病 (ALL)	前B細胞白血病 細胞株 (ALL、 最初の再発)	-0.330	-0.216	-0.165	-0.214	-0.150	-0.182	-0.204	-0.174	-2	-5	-5	-2	-4	1
RH.30	肉腫	横紋 筋肉腫	転移性横紋 筋肉腫細胞株	-0.280	-0.165	-0.137	-0.194	-0.125	-0.158	-0.175	-0.158	2	-14	-3	-7	-7	-7
RPML8226	リンパ系	多発性 ミエロマ	多発性ミエロマ 細胞株	-0.322	-0.207	-0.155	-0.198	-0.138	-0.169	-0.190	-0.164	1	16	11	19	14	22
RT4	泌尿器/ 男性生殖器	膀胱癌	膀胱移行 上皮 癌細胞株	-0.282	-0.168	-0.145	-0.192	-0.136	-0.170	-0.191	-0.171	-5	-1	-12	-16	-19	-25
唾液腺	一次組織	唾液腺	一次組織	-0.262	-0.166	-0.138	-0.177	-0.128	-0.154	-0.172	-0.155	-7	2	-9	-2	-5	-5
SCLC.21H	肺臓	小細胞 肺癌	小細胞 肺癌細胞株	-0.259	-0.160	-0.138	-0.201	-0.123	-0.157	-0.172	-0.148	-11	-31	-5	-12	-10	8
SH.5Y5Y	脳	神経芽腫	転移性神経 芽腫、神経 上皮腫細胞株 SK-N-SH700- の亜株	-0.271	-0.170	-0.137	-0.201	-0.124	-0.157	-0.170	-0.151	2	-25	2	-5	1	6
SiHa	乳房/ 女性生殖器	子宮頸癌	子宮頸扁平 上皮細胞癌 細胞株、 HPV16の 1-2コピー 組込み	-0.288	-0.180	-0.148	-0.201	-0.139	-0.176	-0.193	-0.175	-2	-7	-15	-19	-11	-27
SK_BR.3	乳房/ 女性生殖器	乳癌	転移性乳房 腺癌細胞株	-0.288	-0.176	-0.148	-0.195	-0.140	-0.176	-0.191	-0.169	-3	-4	-21	-22	-12	-11
SK_MEL.30	一次組織	メラノーマ	転移性悪性 メラノーマ細胞株	-0.301	-0.187	-0.154	-0.208	-0.141	-0.174	-0.193	-0.171	-2	-12	-8	-3	-1	-6
骨格筋	一次組織	骨格筋	一次組織	-0.281	-0.166	-0.134	-0.179	-0.125	-0.150	-0.164	-0.145	-1	-7	-7	0	9	11
皮膚	皮膚	皮膚	一次組織	-0.259	-0.166	-0.134	-0.168	-0.127	-0.148	-0.167	-0.151	-4	8	-14	5	-1	-4
小腸	一次組織	小腸	一次組織	-0.260	-0.164	-0.121	-0.158	-0.107	-0.141	-0.166	-0.142	9	10	11	9	0	7
平滑筋	一次組織	平滑筋	一次組織	-0.259	-0.158	-0.127	-0.169	-0.113	-0.144	-0.181	-0.149	2	-6	3	4	4	-5
脾臓	一次組織	脾臓	一次組織	-0.308	-0.202	-0.148	-0.180	-0.130	-0.155	-0.177	-0.154	7	27	15	25	20	25
胃	一次組織	胃	一次組織	-0.264	-0.170	-0.131	-0.170	-0.117	-0.149	-0.189	-0.151	6	3	9	6	0	2
精巣	一次組織	精巣	一次組織	-0.215	-0.142	-0.109	-0.147	-0.093	-0.126	-0.133	-0.123	0	0	0	0	0	0
THP.1	骨髄系	単球性 白血病	急性単球性 白血病 (AML) 細胞株	-0.338	-0.218	-0.168	-0.208	-0.149	-0.182	-0.204	-0.176	-1	8	-1	1	-3	0

10

20

30

40

RNA名称	カテゴリー	タイプ	詳細	相関関係								健康					ランク相違				
				IH01	IH01	IH02	IC15	IC20	IC17	IC17	IC35	IC45	IC20	IC17	IC17	IC35					
甲状腺	一次組織	甲状腺	一次組織	-0.261	-0.158	-0.136	-0.178	-0.121	-0.153	-0.170	-0.161	-2	-7	-2	-6	-6	-19				
TIME	その他	微小血管上皮細胞株	テロメラーゼ不活化ヒト微小血管上皮細胞株 (プールされたもの)	-0.296	-0.180	-0.147	-0.198	-0.134	-0.170	-0.186	-0.170	5	-3	3	-1	3	-11				
扁桃	一次組織	扁桃	一次組織	-0.282	-0.179	-0.141	-0.169	-0.125	-0.147	-0.173	-0.152	-1	20	8	23	4	9				
U.138_MG	脳	神経膠芽腫	神経膠芽腫細胞株	-0.288	-0.177	-0.144	-0.191	-0.126	-0.162	-0.177	-0.161	1	6	7	0	2	2				
U.2_05	肉腫	骨肉腫	骨肉腫細胞株	-0.275	-0.175	-0.139	-0.192	-0.134	-0.159	-0.170	-0.160	-2	0	-11	-3	6	-3				
U.2197	肉腫	肉腫	悪性線維性組織球腫細胞株	-0.290	-0.181	-0.146	-0.195	-0.129	-0.164	-0.160	-0.165	2	1	5	3	4	0				
U.251_MG	脳	神経膠芽腫	神経膠芽腫細胞株	-0.292	-0.178	-0.140	-0.197	-0.125	-0.160	-0.177	-0.165	9	-6	11	4	4	-4				
U.266.70	リンパ系	多発性ミエローマ	多発性ミエローマ細胞株 (1970, IL-6-依存性)	-0.320	-0.207	-0.157	-0.202	-0.135	-0.170	-0.191	-0.165	-1	4	19	15	12	17				
U.266.84	リンパ系	多発性ミエローマ	多発性ミエローマ細胞株 (984, in vitro分化)	-0.326	-0.212	-0.162	-0.207	-0.139	-0.175	-0.194	-0.169	-1	2	11	8	10	14				
U.698	リンパ系	B細胞リンパ腫	B細胞リンパ腫細胞株 (リンパ芽球性リンパ肉腫)	-0.328	-0.212	-0.159	-0.203	-0.137	-0.170	-0.194	-0.166	2	5	16	20	6	20				
U.87_MG	脳	神経膠芽腫、星状細胞腫	神経膠芽腫、星状細胞腫細胞株	-0.285	-0.175	-0.143	-0.192	-0.127	-0.160	-0.174	-0.162	1	0	2	-2	2	-4				
U.937	骨髄系	骨髄単球性組織球性リンパ腫	骨髄単球性組織球性リンパ腫細胞株	-0.346	-0.224	-0.167	-0.201	-0.146	-0.180	-0.199	-0.173	1	18	3	5	2	6				
膀胱	一次組織	膀胱	一次組織	-0.260	-0.158	-0.130	-0.165	-0.118	-0.146	-0.164	-0.150	3	5	-2	1	3	-6				
WM.115	皮膚	メラノーマ	悪性メラノーマ細胞株	-0.284	-0.175	-0.144	-0.193	-0.130	-0.160	-0.178	-0.157	-1	-4	-4	-3	-3	2				

10

20

【0057】

転写開始サイトの最初の10kb下流の193 - 199bp周期についての平均FFT (高速フーリエ変換) とHuman Protein Atlasによる44のヒト細胞株及び32の一次組織の19,378のアンサンブル遺伝子識別因子について測定されたFPKM発現値との相関値。表3はまた、Protein Atlasによって提供される発現サンプルの各々についての簡単な説明を、IH01、IH02及びBH01サンプルに対するランク変換及びランク相違とともに含む。

30

【実施例5】

【0058】

cfDNAから健康ではない供給源組織を決定する

健康ではない状態にあるさらに別の寄与組織の推論が可能か否かを試験するために、5人の後期癌患者から入手したcfDNAを配列決定した。これらのサンプルのヌクレオソームの間隔の取り方のパターンは、非造血性組織又は細胞株と極めて強く相関する、cfDNAに対するさらに別の寄与を示し、前記パターンは患者の癌の解剖学的起源としばしば適合する。

癌患者のcfDNAのヌクレオソームの間隔の取り方は非造血性細胞の寄与を明らかにする

健康ではない状態の循環cfDNAに寄与する非造血性細胞系列のシグナチャーの検出が可能か否かを決定するために、多様な第IV期癌の臨床的診断を有する個体の44の血漿サンプルを、cfDNAから調製した一本鎖ライブラリーの低深度シーケンシングによりスクリーニングした (表4; 中央値は2.2重カバレッジ)。

40

【0059】

表4: 癌パネルのための臨床診断とcfDNA収量

【表 4】

サンプル 識別番号	臨床診断	病期	c.fDNA収量 (ng/mL)	患者性別
IC01 †	腎癌 (移行細胞)	IV	242	F
IC02	卵巣癌 (未同定)	IV	22.5	F
IC03	皮膚癌 (メラノーマ)	IV	12.0	M
IC04	乳癌 (侵襲性/浸潤性腺管性)	IV	12.6	F
IC05	肺癌 (腺癌)	IV	5.4	M
IC06	肺癌 (中皮腫)	IV	11.4	M
IC07 †	胃癌 (未同定)	IV	52.2	M
IC08	子宮癌 (未同定)	IV	15.0	F
IC09	卵巣癌 (漿液性腫瘍)	IV	8.4	F
IC10	肺癌 (腺癌)	IV	11.4	F
IC11	結腸直腸癌 (未同定)	IV	11.4	M
IC12	乳癌 (侵襲性/浸潤性小葉性)	IV	12.0	F
IC13	前立腺癌 (未同定)	IV	12.3	M
IC14	頭頸部癌 (未同定)	IV	27.0	M
IC15 §	肺癌 (小細胞)	IV	22.5	M
IC16	膀胱癌 (未同定)	IV	14.1	M
IC17 §	肝癌 (肝細胞癌)	IV	39.0	M
IC18	腎癌 (明細胞)	IV	10.5	F
IC19	精巣癌 (精上皮腫)	IV	9.6	M
IC20 §	肺癌 (扁平上皮細胞癌)	IV	21.9	M
IC21	膵癌 (腺管性腺癌)	IV	35.4	M
IC22	肺癌 (腺癌)	IV	11.4	F
IC23	肝癌 (肝細胞癌)	IV	17.1	M
IC24	膵癌 (腺管性腺癌)	IV	37.2	M
IC25	膵癌 (腺管性腺癌)	IV	27.9	M
IC26	前立腺癌 (腺癌)	IV	24.6	M
IC27	子宮癌 (未同定)	IV	19.2	F
IC28	肺癌 (扁平上皮細胞癌)	IV	33.3	M

10

20

30

40

50

IC29	頭頸部癌 (未同定)	IV	14.4	M
IC30	食道癌 (未同定)	IV	10.5	M
IC31 †	卵巣癌 (未同定)	IV	334.8	F
IC32	肺癌 (小細胞)	IV	9.6	F
IC33	結腸直腸癌 (腺癌)	IV	13.8	M
IC34	乳癌 (侵襲性/浸潤性小葉性)	IV	33.6	F
IC35 §	乳癌 (in situ腺管癌)	IV	16.2	F
IC36	肝癌 (未同定)	IV	26.4	M
IC37 §	結腸直腸癌 (腺癌)	IV	15.9	F
IC38	膀胱癌 (未同定)	IV	6.6	M
IC39	腎癌 (未同定)	IV	39.0	M
IC40	前立腺癌 (腺癌)	IV	13.8	M
IC41	精巣癌 (精上皮腫)	IV	16.5	M
IC42	肺癌 (腺癌)	IV	11.4	F
IC43	皮膚癌 (メラノーマ)	IV	21.9	F
IC44	食道癌 (未同定)	IV	25.8	F
IC45 †	結腸直腸癌 (腺癌)	IV	3.0	M
IC46 **	乳癌 (in situ腺管癌)	IV	36.6	F
IC47	膵癌 (腺管性腺癌)	IV	19.2	F
IC48 **	乳癌 (侵襲性/浸潤性小葉性)	IV	13.8	F

§ : サンプルは追加のシーケンシングのために選別された

** : このサンプルでは0.5mLの血漿のみが利用可能であった

† : サンプルはQCをパスせず、更なる分析には用いられなかった

【 0 0 6 0 】

表4は、高い腫瘍組織量の証拠についてその血漿運搬cfDNAがスクリーニングされた48人の患者の臨床的及び組織学的診断とともに、各個体の血漿1.0mLの全cfDNA収量及び関連する臨床的共変量を示す。これら48人のうち、44人はQCをパスし、十分な材料を有していた。これら44人のうち、5人をより深度の高いシーケンシングのために選別した。cfDNA収量はキュビットフルオロメーター (Qubit Fluorometer) 2.0 (Life Technologies) によって決定した。

これらのサンプルは、実施例4のIH02と同じプロトコルで、さらに多くは同じバッチで調製された。第IV期癌の臨床診断を有する52個体（表4）のヒト末梢血血漿をコンバーサント・バイオ（Conversant Bio）又はプラズマラブ・インターナショナル（PlasmaLab International）（Everett, Washington, USA）から入手し、0.5mL又は1mLアリコットで使用まで-80 で保存した。全身性エリテマトーデスの臨床診断を有する4個体のヒト末梢血血漿はコンバーサント・バイオから入手し、0.5mLアリコットで使用まで-80 で保存した。凍結血漿アリコットを使用直前にベンチトップで融解した。循環セルフリーDNAは、Qia AMP循環核酸キット（Qiagen）を用い製造業者のプロトコルに従って各血漿サンプルの2mLから精製した。DNAはキュビットフルオロメーター（Invitrogen）で定量した。サンプルのサブセットでcfDNA収量を検証するために、精製DNAをさらに、マルチコピーヒトAlu配列を標的とするカスタムqPCRアッセイで定量した。2つの概数が合致することが見出された。

10

適合腫瘍遺伝子型を入手できなかったので、高い割合で腫瘍由来cfDNAを含む蓋然性が高いサブセットを同定するために、異数性に関する2つの測定基準に基づいて各サンプルにスコアを付与した。すなわち第一は、各染色体から得られる読取りの期待される割合の偏差（図26A）、及び第二は、共通の単一ヌクレオチド多形性パネルに関する各染色体の対立遺伝子バランスプロフィール（図62B）である。これらの測定基準に基づいて、5個体（小細胞肺癌、扁平上皮細胞肺癌、結腸直腸腺癌、肝細胞癌、及び乳管癌in situ乳癌を有する個体）に由来する一本鎖ライブラリーを実施例4のIH02の深度と同様な深度で配列決定した（表5；中央値は30重カバレッジ）。

20

【 0 0 6 1 】

表5：CA01セットに含まれる追加サンプルについてのシーケンシング統計

【 0 0 6 2 】

【表5】

サンプル の名称	ライブラリ のタイプ	読取り	配列決定 フラグメント	アライン メント	アラインメント (Q30)	カバ レッジ	複製物の 概算%	35-80bp	120- 180bp
IH03	SSP	2x39	53292855	92.66%	72.37%	2.29	15.46%	11.05%	52.34%
IP01†	DSP	2x101, 2x102	1214536629	97.22%	86.38%	76.11	0.55%	0.08%	62.77%
IP02†	DSP	2x101, 2x102	855040273	97.16%	87.72%	52.46	0.83%	0.07%	68.10%
IA01	SSP	2x39	53934607	87.42%	68.30%	2.02	22.70%	15.20%	49.77%
IA02	SSP	2x39	42496222	95.42%	76.61%	1.95	4.74%	12.28%	59.00%
IA03	SSP	2x39	51278489	93.12%	71.33%	2.05	25.68%	14.27%	52.57%
IA04	SSP	2x39	50768476	90.30%	70.51%	2.14	7.83%	17.80%	36.76%
IA05	DSP	2x101	194985271	98.80%	90.61%	11.09	12.05%	2.24%	71.67%
IA06	DSP	2x101	171670054	98.90%	90.88%	9.90	5.41%	1.93%	71.26%
IA07	DSP	2x101	208609489	98.67%	90.34%	11.69	11.45%	2.59%	74.84%
IA08	DSP	2x101	193729556	98.81%	90.70%	10.84	11.96%	2.58%	76.24%
IC02	SSP	2x39	67913606	95.07%	75.57%	2.59	5.40%	12.98%	60.00%
IC03	SSP	2x39	63862631	95.78%	75.66%	2.79	8.32%	13.25%	62.20%
IC04	SSP	2x39	55239248	95.47%	76.26%	2.57	8.28%	10.98%	58.48%
IC05	SSP	2x39	39623850	89.80%	69.92%	1.60	9.24%	14.63%	50.33%
IC06	SSP	2x39	59679981	95.57%	74.90%	2.11	3.93%	24.30%	41.46%
IC08	SSP	2x39	46933688	94.38%	74.21%	1.92	5.92%	15.04%	45.25%
IC09	SSP	2x42	59639583	91.22%	71.15%	2.13	6.69%	21.39%	43.50%
IC10	SSP	2x42	53994406	93.73%	73.40%	1.83	2.00%	27.08%	37.62%
IC11	SSP	2x42	59225460	93.25%	72.51%	2.15	5.26%	21.30%	43.33%
IC12	SSP	2x42	57884742	93.52%	74.33%	2.34	2.66%	18.28%	46.58%
IC13	SSP	2x42	71946779	92.94%	72.47%	2.52	2.18%	23.51%	43.97%
IC14	SSP	2x42	61649203	94.54%	73.47%	2.20	3.23%	22.26%	43.37%
IC15	SSP	2x50, 43/42	908512803	95.49%	76.83%	29.77	10.66%	25.42%	38.47%
IC16	SSP	2x42	62739733	92.81%	72.85%	2.47	2.77%	17.71%	48.04%
IC17	SSP	2x50, 2x39	1072374044	96.02%	76.42%	42.08	12.16%	17.08%	50.02%
IC18	SSP	2x39	59976914	87.91%	68.67%	2.24	4.39%	18.85%	44.44%
IC19	SSP	2x39	51447149	89.38%	69.39%	2.02	8.24%	17.30%	46.33%
IC20	SSP	2x50, 2x39	640838540	96.30%	79.11%	23.38	12.43%	25.72%	39.87%
IC21	SSP	2x39	53000579	94.64%	74.57%	1.79	37.39%	29.89%	43.81%
IC22	SSP	2x39	58102606	94.08%	74.08%	2.51	6.24%	13.65%	58.41%
IC23	SSP	2x39	65859970	95.67%	75.67%	2.94	5.34%	11.09%	60.85%
IC24	SSP	43/42	66344431	94.63%	74.46%	2.48	2.00%	22.46%	46.31%
IC25	SSP	43/42	75066833	93.75%	73.66%	2.86	2.24%	21.30%	46.19%
IC26	SSP	43/42	79180860	92.59%	72.32%	2.97	2.93%	22.34%	40.42%
IC27	SSP	43/42	78037377	88.81%	67.04%	2.20	1.50%	31.31%	30.59%
IC28	SSP	43/42	61402081	95.24%	75.74%	2.60	2.46%	18.71%	46.44%
IC29	SSP	2x39	49989522	94.46%	73.36%	1.75	3.03%	25.82%	36.23%
IC30	SSP	2x39	58439504	93.52%	71.19%	1.75	17.35%	29.58%	30.47%
IC32	SSP	43/42	78233981	87.86%	66.80%	2.25	1.79%	30.12%	31.20%
IC33	SSP	43/42	62196185	87.26%	66.71%	1.93	1.93%	27.44%	36.92%
IC34	SSP	43/42	63572169	95.42%	76.74%	2.53	2.35%	19.64%	48.55%

10

20

30

40

サンプルの名称	ライブラリーのタイプ	読取り	配列決定フラグメント	アラインメント	アラインメント (Q30)	カバレッジ	複製物の概算%	35-80bp	120-180bp
IC35	SSP	43/42	618554393	86.47%	65.90%	18.22	5.23%	28.18%	35.24%
IC36	SSP	43/42	54402943	94.62%	74.73%	2.21	3.32%	17.02%	52.42%
IC37	SSP	2x50, 43/42	1175553677	93.00%	74.46%	38.22	10.15%	28.47%	35.11%
IC38	SSP	43/42	47981963	89.35%	69.45%	1.78	6.47%	18.59%	43.03%
IC39	SSP	43/42	61968854	95.29%	75.57%	2.62	2.54%	14.42%	57.28%
IC40	SSP	2x39	53228209	93.54%	71.69%	1.81	8.85%	24.88%	34.95%
IC41	SSP	43/42	78081655	87.11%	65.25%	2.26	1.61%	27.94%	35.21%
IC42	SSP	2x39	53017317	93.59%	74.33%	2.02	10.74%	19.04%	44.12%
IC43	SSP	43/42	76395478	88.41%	67.21%	2.40	1.56%	26.68%	37.76%
IC44	SSP	43/42	61354307	95.15%	74.88%	2.45	4.34%	19.10%	46.39%
IC46	SSP	2x39	60123123	94.51%	72.23%	2.13	10.37%	15.46%	50.93%
IC47	SSP	2x39	59438172	95.58%	73.84%	2.07	9.33%	21.67%	43.34%
IC48	SSP	43/42	55704417	91.35%	72.79%	2.01	13.87%	22.56%	38.68%
IC49	DSP	2x101	170489015	99.02%	90.53%	11.19	5.93%	2.41%	59.93%
IC50	DSP	2x101	203828224	98.72%	90.28%	10.82	2.83%	4.81%	66.23%
IC51	DSP	2x101	200454421	98.63%	90.53%	11.77	9.50%	2.58%	67.04%
IC52	DSP	2x101	186975845	98.97%	91.25%	11.37	2.57%	0.83%	68.96%

10

SSP : 一本鎖ライブラリー調製プロトコル DSP : 二本鎖ライブラリー調製プロトコル
 † : サンプルは以前に公表された
 (J. O. Kitzman et al., Science Translational Medicine, 2012)

20

【0063】

表5はシーケンシング関連統計値の表であり、各サンプルについて配列を決定したフラグメントの総数、読取りの長さ、あるマッピング品質閾値で又はそのような閾値無しに参照に対してアラインメントを示したフラグメントのパーセンテージ、平均カバレッジ、重複率、及び2つの長さのbinにおける配列決定フラグメントの割合を含む。フラグメントの長さはペア形成末端の読取りのアラインメントから推論した。短い読取りの長さのために、カバレッジは全フラグメントが読み取られたと仮定することによって計算した。重複フラグメントの概算数はフラグメント末端を基準にし、前記は高度に定型化された切断の存在下では真の重複率を過大評価する可能性がある。

30

上記に記載したように、FFTは遺伝子本体を通して長フラグメントWPS値について実施し、ヒト細胞株及び一次組織の同じ76発現データセットに対して193 - 199bp周波数レンジにおける平均強度と相関した。実施例4(上位10の全て及び上位20のほぼ全ての相関関係はリンパ系細胞系列又は骨髄系細胞系列に対するものであった)の健康な個体に由来する3つのサンプルと対照的に、最高ランクの細胞株又は組織の多くは非造血性細胞系列であり、いくつかの事例では癌タイプと整合性を示した(図61、表3)。例えば、IC17については(前記では患者は肝癌を有していた)、上位ランクの相関性はHepG2(肝細胞癌細胞株)とのものであった。IC35については(患者は乳管癌in situ乳癌を有していた)、上位ランク相関性はMCF7(転移性乳房腺癌細胞株)とのものであった。他の事例では、相関性ランクで最大の変化を示す細胞株又は一次組織は癌タイプとの整合性を示した。例えば、IC15については(前記では患者は小細胞肺癌を有していた)、相関性ランクにおける最大の変化(-31)は小細胞肺癌細胞株(SCLC-2H)との整合性であった。IC20(肺扁平上皮細胞癌)及びIC35(結腸直腸腺癌)については、相関性ランクに関して多くの非造血性癌細胞株がリンパ系/骨髄系細胞株と入れ替わったが、これらと特異的な癌タイプとの整合性はあまり明瞭ではなかった。これらの癌の特異的な分子プロファイルはこの76発現データセットでは良好に象徴されてなかった可能性がある(例えばこれらデータセットのいずれも肺扁平上皮細胞癌ではなく、CACO-2は結腸直腸腺癌由来細胞株であるが高度に不均質性であることが知られている)。

40

【0064】

50

忍耐を要する反復アプローチを用いて、生物学的サンプルに由来するcfDNAに寄与する多様な細胞タイプ及び/又は組織の割合を概算した。第一に、その参照マップ（ここでは76RNA発現データセットによって規定される）が、ある与えられたcfDNAサンプルについて遺伝子本体全体にわたってWPS長フラグメント値の193 - 199bp周期で平均FFT強度と最高の相関性を有する、細胞タイプ又は組織を認定した。次に、一連の“2組織”線形混合モデルを適合させた。前記は、最高の相関性を有する細胞タイプ又は組織とともに、参照マップの完全なセットに由来する他の残りの細胞タイプ又は組織も含んでいた。後者のセットのうち、最高係数を有する細胞タイプ又は組織は、係数が1%未満でないかぎり寄与するものとして維持され、1%未満の場合には当該手順を終了し、この最後の組織又は細胞タイプは加えられなかった。この手順は、混合モデルによって1%未満の寄与と概算される新規付加組織に基づいて終了まで繰り返される（すなわち“3組織”、“4組織”など）。混合モデルは以下の様式をとる：

$$\operatorname{argmax}_{\{a,b,c,\dots\}} \operatorname{cor}(\text{平均_FFT強度}_{193-199}, a \cdot \log_2 \text{実験組織} + b \cdot \log_2 \text{組織}2 + c \cdot \log_2 \text{組織}3 + \dots + (1 - a - b - c - \dots) \cdot \log_2 \text{実験組織}N).$$

例えば、IC17について（進行した肝細胞癌を有する患者に由来するcfDNA）、この手順は9つの寄与しうる細胞タイプを予測し、前記細胞タイプには、Hep_G2（28.6%）、HMC.1（14.3%）、REH（14.0%）、MCF7（12.6%）、AN3.CA（10.7%）、THP.1（7.4%）、NB.4（5.5%）、U.266.84（4.5%）及びU.937（2.4%）が含まれていた。BH01については（健康な個体の混合物に一致するcfDNA）、この手順は7つの寄与しうる細胞タイプ又は組織を予測し、前記には、骨髄（30.0%）、NB.4（19.6%）、HMC.1（13.9%）、U.937（13.4%）、U.266.84（12.5%）、Karpas.707（6.5%）及びREH（4.2%）が含まれていた。注目すべきことに、IC17については（癌患者由来サンプル）、予測される寄与の最高の割合は、このcfDNAが由来した患者に存在する癌タイプと極めて関係する細胞株と一致する（Hep_G2と肝細胞癌）。対照的に、BH01については、このアプローチは、もっぱら造血と結びついた組織又は細胞タイプ（健康個体の血漿cfDNAの主要な供給源）にのみ一致する寄与を予測する。

【実施例6】

【0065】

実施例4-5のための一般的方法

サンプル

バルクヒト末梢血血漿（人数が明らかでない健康な個体からの提供を含む）をSTEMCELLテクノロジーズ（Vancouver, British Columbia, Canada）から入手し、2mLアリコットで使用まで-80 で保存した。匿名の健康なドナーの個々のヒト末梢血血漿をコンバーサント・バイオ（Huntsville, Alabama, USA）から入手し、0.5mLアリコットで使用まで-80 で保存した。

妊婦IP01及びIP02の全血を妊娠18及び13週でそれぞれ入手し、先に述べたようにプロセッシングした。

第IV期癌（付録表4）と診断された52個体のヒト末梢血血漿をコンバーサント・バイオ又はプラズマ・インターナショナル（Everett, Washington, USA）から入手し、0.5mL又は1mLアリコットで使用まで-80 で保存した。全身性エリテマトーデスの臨床診断を有する4個体のヒト末梢血血漿をコンバーサント・バイオから入手し、0.5mLアリコットで使用まで-80 で保存した。

【0066】

血漿サンプルのプロセッシング

凍結血漿アリコットを使用直前にベンチトップで融解した。循環セルフリーDNAは、QiaAMP循環核酸キット（Qiagen）を用い製造業者のプロトコルに従って各血漿サンプルの2mLから精製した。DNAはキュビットフルオロメーター（Invitrogen）で定量した。サンプルのサブセットでcfDNA収量を検証するために、精製DNAをさらに、マルチコピーヒトAlu配列を標的とするカスタムqPCRアッセイで定量した。2つの概数が合致することが見出された。

二本鎖シーケンシングライブラリーの調製

10

20

30

40

50

バーコード付加シーケンシングライブラリーをThruPLEX-FD又はThruPLEX DNA-seq 48Dキット (Rubicon Genomics) (末端修復、連結及び増幅反応の独占的シリーズを含む)を用いて調製した。全ての臨床サンプルライブラリーについてインプットとして0.5ngから30.0ngのcfDNAを用いた。全てのサンプルについてライブラリー増幅をリアルタイムPCRによってモニターして過剰増幅を回避し、典型的には4-6サイクル後に終了させた。

【0067】

一本鎖シーケンシングライブラリーの調製

アダプター2を以下によって調製した：4.5 µLのTE (pH 8)、0.5 µLの1M NaCl、10 µLのオリゴアダプター2.1 (500 µM) 及び10 µLのアダプター2.2 (500 µM) を一緒にし、95 °Cで10秒間インキュベートし、0.1 °C/sの速度で14 °Cに温度を下げる。精製cfDNAフラグメントを以下によって脱リン酸化した：2×CircLigase II緩衝液 (Epicentre)、5mM MnCl₂ 及び1UのFastAPアルカリホスファターゼ (Thermo Fisher) を0.5 - 10ngのフラグメントと20 µLの反応体積中で一緒にし、37 °Cで30分間インキュベートする。続いてフラグメントを95 °Cで3分間加熱して変性させ、直ちに氷浴へ移した。前記反応物にビオチン結合アダプターオリゴCL78 (5pmol)、20% PEG-6000 (w/v) 及び200U CircLigase II (Epicentre) を補充し (総体積40 µL)、回転させながら一晩60 °Cでインキュベートし、95 °Cで3分間加熱し、さらに氷浴に置いた。各サンプルについて、20 µLのMyOne C1ビーズ (Life Technologies) をビーズ結合緩衝液 (BBB) (10mM Tris-HCl (pH8)、1M NaCl、1mM EDTA (pH8)、0.05% Tween-20及び0.5% SDS) 中で2回洗浄し、250 µLのBBBに再懸濁させた。室温で60分間回転させてアダプター連結フラグメントをビーズに結合させた。ビーズを磁性ラック上で収集し、上清を廃棄した。ビーズを500 µLの洗浄緩衝液A (WBA) (10mM Tris-HCl (pH8)、1mM EDTA (pH8)、0.05% Tween-20、100mM NaCl、0.5% SDS) で1回、500 µLの洗浄緩衝液B (WBB) (10mM Tris-HCl (pH8)、1mM EDTA (pH8)、0.05% Tween-20、100mM NaCl) で1回洗浄した。ビーズを50 µLの反応体積中で1×イソサーマル増幅緩衝液 (NEB)、2.5 µMオリゴCL9、それぞれ250 µMのdNTP、及び24UのBst 2.0 DNAポリメラーゼ (NEB) と一緒にし、穏やかに振盪させながら15 °Cから37 °Cまで1 °C/分で温度を上昇させることによってインキュベートし、さらに10分間37 °Cで維持した。磁性ラック上で収集した後、ビーズを200 µLのWBAで1回洗浄し、200 µLのストリンジェンシー洗浄緩衝液 (SWB) (0.1×SSC、0.1% SDS) に再懸濁し、さらに45 °Cで3分間インキュベートした。ビーズを再度収集し、200 µLのWBBで1回洗浄した。続いて、ビーズを1×CutSmart緩衝液 (NEB)、0.025% Tween-20、それぞれ100 µMのdNTP及び5UのT4 DNAポリメラーゼ (NEB) と一緒にし、穏やかに振盪しながら室温で30分間インキュベートした。ビーズをWBA、SWB、及びWBBの各々で1回上記に記載したように洗浄した。続いて、ビーズを1×CutSmart緩衝液 (NEB)、5% PEG-6000、0.025% Tween-20、2 µMの二本鎖アダプター2、及び10U T4 DNAリガーゼ (NEB) と混合し、穏やかに振盪しながら室温で2時間インキュベートした。ビーズをWBA、SWB、及びWBBの各々で1回上記に記載したように洗浄し、25 µLのTET緩衝液 (10mM Tris-HCl (pH8)、1mM EDTA (pH8)、0.05% Tween-20) に再懸濁した。95 °Cで加熱することによって第二の鎖を溶出させ、磁性ラック上でビーズを収集し、上清を新しいチューブに移した。全てのサンプルについて実験室増幅はリアルタイムPCRによってモニターし過剰増幅を回避し、ライブラリーにつき平均して4から6サイクルを要した。

【0068】

シーケンシング

全てのライブラリーをHiSeq2000又はNextSeq500装置 (Illumina) で配列を決定した。

一次シーケンシングデータのプロセッシング

バーコード付加ペア形成末端 (PE) イルミナシーケンシングデータを、バーコード内の1つまでの置換を可能にしつつ分割した。読取りの長さより短いか又はそれと等しい読取りをコンセンサスコールし、アダプターをトリミングした。残りのコンセンサス単一末端読取り (SR) 及び個々のPEの読取りをヒト参照ゲノム配列 (GRCh37、ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/technical/reference/phase2_reference_assembly_sequence/) からダウンロードされた1000ゲノムフェーズ2テクニカルレファレンス) に対して、BWA v

10

20

30

40

50

0.7.10で実行されるALNアルゴリズムを用いアラインメントした。PEの読取りをBWA SAMPL Eでさらにプロセッシングして、読取りペアの曖昧な配置を解決するか、又は1つの配置 済みの読取り末端の場所の周辺でさらに感度の高いアラインメント工程によって失われた アラインメントを回復させた。SAMtools APIを用いて、アラインメントされたSR及びPEデ ータを分類BAM様式に直接変換した。サンプルのBAMファイルをレーン及びシーケンシング ランを通して合併させた。

品質管理はFastQC (v0.11.2)を用いて実施し、ライブラリーの複雑度の概数を入手し (Picard tools v1.113)、アダプターダイマーの割合を決定し、推測されるライブラリ ー挿入物サイズ、外側読取り末端におけるヌクレオチド及びジヌクレオチド周期の分析と ともに各ライブラリーのマッピング品質分布をチェックした。

10

シミュレーションした読取りデータセット

アラインメントされたシーケンシングデータをヒト参照物 (GRC37h) の全ての主要な染 色体についてシミュレーションした (45bpより短い場合はSR、そうでなければPE 45bp) 。この目的のために、ジヌクレオチド頻度を両方の読取り末端及び両方の鎖の向きにつ いて真のデータから決定した。ジヌクレオチド頻度はまた両方の鎖について参照ゲノムに ついて記録した。さらにまた、真のデータの挿入物サイズ分布を1 - 500bpレンジについて 抽出した。主要な参照染色体の配列の端から端まで反復適用することによって読取りをシ ミュレーションした。各工程で (すなわち所望されるカバレッジに応じて各位置で1回以 上)、(1) 鎖をランダムに選択し、(2) 参照配列のジヌクレオチド頻度を超える真のデ ータのジヌクレオチド頻度の比率を用いて、開始ヌクレオチドが考慮されるか否かをラン 20

20

【0069】

カバレッジ、読取り開始点及びウィンドウ保護スコア

本開示のデータは、シーケンシングライブラリーの調製で用いられるDNA分子の2つの物 理的末端についての情報を提供する。我々は、この情報をBAMファイルからSAMtoolsアプ リケーションプログラミングインターフェース (API) を用いて抽出する。読取り開始点 として、我々はPEデータの両方の外側アラインメント座標を用いる (前記に対して両方の 30

30

ヌクレオソームピークコール

ヌクレオソーム保護の極大は長画分WPSからコールされ、我々は、Savitzky-Golayフ ィルター (ウィンドウサイズ21、二次多項式) を用いてこれを局所的に移動中央値0 (1kbウ インドウ) に調製し、平滑化した。続いてWPSトラックを、0を超える領域にセグメント化 する (0未満の5つまでの連続位置を許容する)。得られた領域が50 - 150bp長の場合、我 々は、当該領域の中央値を同定し、当該中央値を超える最大合計連続ウィンドウを検索す る。我々は、このウィンドウの開始、末端及び中心座標を報告する。ピークからピークま での距離などをこの中心座標から計算する。コールのスコアは、当該ウィンドウ中の最大 値と当該領域の近隣の2つの隣接するWPS最小値の平均との間の距離として決定される。 同定された領域が150 - 450bp長である場合、我々は上述したのと同じ中央値を超える連続ウ インドウアプローチを適用するが、サイズが50 - 150bpのウィンドウのみを報告する。 40

40

50

- 450bp領域に由来する複数のウインドウのスコア計算については、我々は領域内の近接最小値を0と仮定する。50bpより短い領域及び450bpより長い領域は廃棄する。

【0070】

167bpフラグメントのジヌクレオチド組成

フラグメントサイズ分布の支配的なピークと一致する正確に167bpの予測された長さを有するフラグメントをサンプル内でフィルター処理して重複物を除去した。ジヌクレオチド頻度は、2bpスライドウインドウ及び各位置の参照対立遺伝子を用い、1つのフラグメントの端点の50bp上流で開始し、他方の端点の50bp下流で終了させて、鎖認識態様で計算された。各位置の観察されたジヌクレオチド頻度を、ライブラリー特異的態様で計算した同じ切断偏向を反映するシミュレーション読取りセットから決定した期待されるジヌクレオチド頻度と比較した（詳細については上記参照）。

10

転写因子結合サイト及びゲノム上の特徴周辺のプロフィール

分析は、コンピュータ予測転写因子結合サイトのセットを規定するクラスター化FIMO（モチーフ依拠）の間隔の初期セットを用いて開始した。クラスター化転写因子のサブセット（AP-2-2、AP-2、CTCF_Core-2、E2F-2、EBF1、Ebox-CACCTG、Ebox、ESR1、ETS、IRF-2、IRF-3、IRF、MAFK、MEF2A-2、MEF2A、MYC-MAX、PAX5-2、RUNX2、RUNX-AML、STAF-2、TCF-LEF、YY1）については、当該サイトのセットは、実験データに基づいて実際に結合する転写因子結合サイトのさらに信頼性のあるセットに精練された。この目的のために、公的に利用可能なENCODEデータ（UCSCからダウンロードされたTfbsClusteredV3セット）からChIP-seq実験によって明らかにされたピークとオーバーラップする予測結合サイトのみを維持した。

20

これらサイト周囲のウインドウ化保護スコアがCH01サンプルとその対応するシミュレーションの両方について抽出された。各サイト/特徴の保護スコアが、各結合サイト及び集合物の開始座標に対して各位置で計算された。CTCFの公知の52bp結合フットプリントの中心がx軸上に0座標になるように、CTCF結合サイトのプロットをシフトさせた。続いて、5kb抽出WPSシグナルの最初と最後の500bp（前記はもっぱら平坦で平均オフセットを表す）の平均を元のシグナルから差し引いた。長フラグメントシグナルについては、200bpウインドウを用いてスライドウインドウ平均を計算し、元のシグナルから差し引いた。最後に、シミュレーションについての修正WPSプロフィールをCH01についての修正WPSプロフィールから差し引いて、フラグメントの長さ及びライゲーション偏向の産物であるシグナルについて修正した。この最後のプロフィールをプロットし、“調整WPS”と称した。

30

ゲノム上の特徴（例えば転写開始サイト、転写終了サイト、開始コドン、スプライシングドナー、及びスプライシングアクセプターサイト）をEnsembl Buildバージョン75から得た。これら特徴周辺の調整WPSを計算し、転写因子結合サイトについて上記に記載したようにプロットした。

【0071】

CTCF結合サイト周辺のヌクレオソームの間隔の取り方及び対応するWPSの分析

この分析に用いられたCTCFサイトは最初CTCF結合サイトのクラスター化FIMO予測（モチーフによりコンピュータ予測された）を含んでいた。続いて、我々はこのセットの2つの追加サブセットを作製した：1) ENCODE TfbsClusteredV3（上記参照）から利用可能なCTCF ChIP-seqピークのセットとのインターセクション、及び2) 19組織を通して活性であることが実験的に観察されたCTCFサイトのセットとのインターセクション。

40

この結合サイトのどちらかの側の10ヌクレオソームの位置を各サイトについて抽出した。我々は、全ての隣接するヌクレオソーム間の距離を計算し、サイトの各セットについてヌクレオソーム間距離の分布を得た。-1から+1ヌクレオソームの間隔の分布はかなり変化し、特に230 - 270bpレンジでより大きな間隔にシフトした。これは、真に活性なCTCFサイトは-1から+1ヌクレオソームの間でより広い間隔に大きくシフトすること、したがって、長及び短読取り画分の両方についてWPSにおける相違が明らかでありうることを示唆した。したがって、CTCFサイトの中心に対して各位置の平均短及び長フラグメントWPSも計算した。ヌクレオソーム間隔の影響を調べるために、この平均を160未満、160-200、200-23

50

0、230-270、270-420、420-460、及び420bpより大きい-1から+1ヌクレオソーム間隔のbin内で取得した。これらの間隔は、関心のある間隔（例えば支配的なピーク及びより信頼性のある活性なサイトについて230 - 270bpで出現するピーク）をほぼ捕らえた。

DNase I高感受性サイト (DHS) の分析

MauranoらによるBED様式の349の一次組織及び細胞株のDHSピークをワシントン大学のEncodeデータベースからダウンロードした (Maurano et al., Science, vol. 337(6099), p. 1190-95, 2012; “all_fdr0.05_hot” ファイル、2012年2月13日に最後の修正)。胎児組織由来サンプル (これらのピークセットの233を含む) は、組織タイプ内で態様が一貫しなかったため分析から除去した (おそらく各組織サンプル内での複数の細胞タイプの不均等な提示のためであろう)。多様な細胞系列を示す116のサンプルが分析のために維持された。個々のセットの各DHSピークの中心点について、CH01コールセットの最も近くの上流及び下流のコールを確認し、それら2つのコールの中心間のゲノム距離を計算した。各DHSピークコールセットについて、0から500bpの間の距離について計算した平滑化密度の概数を用いて全てのそのような距離の分布を視覚化した。

10

【0072】

遺伝子発現分析

本実験ではFRKM発現値 (Human Protein Atlas (“ma.csv” ファイル) による44のヒト細胞株及び32の一次組織で20,344のアンサンプル (Ensembl) 遺伝子識別子について測定) を用いた。組織横断的分析について、3非ゼロ発現値未満の遺伝子は除外した (19,378遺伝子がこのフィルターを通過する)。発現データセットはFPKM値について小数点一桁で提供される。したがってゼロ発現値 (0.0) は0と0.05未満の値との間の発現を指す。特段の記載がなければ、最小発現値は、発現値の \log_2 変換前の値で0.04FPKMに設定された。

20

平滑ピリオドグラム及び軌跡の平滑化

長フラグメントWPSを用いてゲノム領域のピリオドグラムを計算し、1/500塩基から1/100塩基の周期で高速フーリエ変換 (FFT、R統計プログラミング環境でspec.pgram) を用いた。平滑へのパラメーター (3bpダニエルスムーサー; 半分の重みを末端値に与える移動平均) 及びデータのトレンド除去 (すなわちシリーズの平均を差し引く、さらに線形トレンドを除去する) を場合によってさらに加えて用いた。

特に表示した場合には、R統計プログラミング環境で実装される帰納的時系列フィルターを用いて軌跡から高い周波数変動を除去した。24フィルター周波数 (1/seq(5,100,4)) を用い、さらに初期値として軌跡の最初の24個の値を用いた。得られた軌跡のこの24値-シフトについての調整を、軌跡の最後の24個の値を反復することによって実施した。

30

FFT強度と発現値の相関性

120 - 280bpレンジの遺伝子発現で平滑化ピリオドグラム (FFT) から決定される強度値を分析した。遺伝子発現値と主要なヌクレオソーム間距離ピークとの間のS字形ピアソン相関性が観察された。明白な負の相関性が193 - 199bpレンジで観察された。結果として、この周期レンジにおける強度は平均して \log_2 変換発現値と相関性を有する。

【0073】

〔更なる実施例〕

実施例7: 対象者でセルフリーDNA (cfDNA) を生じる組織及び/又は細胞タイプを決定する方法であって、以下の工程を含む方法:

40

当該対象者からの生物学的サンプルからcfDNAを単離する工程であって、当該単離cfDNAは複数のcfDNAフラグメントを含む、前記工程;

当該複数のcfDNAフラグメントの少なくとも一部分に付随する配列を決定する工程;

当該複数のcfDNAフラグメントの少なくともいくつかのcfDNAフラグメント末端について、参照ゲノム内のゲノム上の場所を当該cfDNAフラグメント配列の関数として決定する工程;

当該cfDNAフラグメントを生じる組織及び/又は細胞タイプの少なくともいくつかを、当該cfDNAフラグメント末端の少なくともいくつかのゲノム上の場所の関数として決定する工程。

50

実施例8：当該cfDNAフラグメントを生じる組織及び/又は細胞タイプの少なくともいくつかを決定する工程が、当該cfDNAフラグメント末端の少なくともいくつかのゲノム上の場所を1つ以上の参照マップと比較する工程を含む、実施例7の方法。

実施例9：当該cfDNAフラグメントを生じる組織及び/又は細胞タイプの少なくともいくつかを決定する工程が、当該cfDNAフラグメント末端の少なくともいくつかのゲノム上の場所の分布に関して数学的変換を実施する工程を含む、実施例7または実施例8の方法。

実施例10：当該数学的変換がフーリエ変換を含む、実施例9の方法。

実施例11：当該参照ゲノムの少なくともいくつかの座標の各々についてスコアを決定する工程をさらに含み、ここで当該スコアが少なくとも複数のcfDNAフラグメント末端及びそれらのゲノム上の場所の関数として決定され、さらに観察されるcfDNAフラグメントを生じる組織及び/又は細胞タイプの少なくともいくつかを決定する工程が当該スコアと1つ以上の参照マップを比較する工程を含む、先行するいずれかの実施例の方法。

10

【0074】

実施例12：座標についてのスコアが、当該座標がcfDNAフラグメント末端の場所である確率を表しているか又は前記確率と関係性がある、実施例11の方法。

実施例13：参照マップが、少なくとも1つの細胞タイプ又は組織から作製されたDNase I 高感受性サイトマップを含む、実施例8から12のいずれか1つの方法。

実施例14：参照マップが、少なくとも1つの細胞タイプ又は組織から作製されたRNA発現マップを含む、実施例8から13のいずれか1つの方法。

実施例15：参照マップが、ヒト組織又は細胞を異種移植された動物のcfDNAから作製される、実施例8から14のいずれか1つの方法。

20

実施例16：参照マップが、少なくとも1つの細胞タイプ又は組織から作製された染色体コンフォーメーションマップを含む、実施例8から15のいずれか1つの方法。

実施例17：参照マップが、少なくとも1つの細胞タイプ又は組織から作製されたクロマチン接近容易性マップを含む、実施例8から16のいずれか1つの方法。

実施例18：参照マップが、少なくとも1人の参照対象者から入手されるサンプルから得られる配列データを含む、実施例8から17のいずれか1つの方法。

実施例19：参照マップが、ある疾患又は異常に伴う少なくとも1つの細胞タイプ又は組織に対応する、実施例8から18のいずれか1つの方法。

実施例20：参照マップが、ある組織又は細胞タイプのヌクレオソーム及び/又はクロマトソームの位置又は間隔を含む、実施例8から19のいずれか1つの方法。

30

実施例21：参照マップが、少なくとも1つの細胞タイプ又は組織から入手されたクロマチンを外因性ヌクレアーゼ（例えばマイクロコッカスヌクレアーゼ）で消化することによって作製される、実施例8から20のいずれか1つの方法。

実施例22：当該参照マップが、少なくとも1つの細胞タイプ又は組織から転移による方法（例えばATAC-seq）によって決定されたクロマチン接近容易性データを含む、実施例8から21のいずれか1つの方法。

【0075】

実施例23：参照マップが、ある組織又は細胞タイプのDNA結合及び/又はDNA占拠タンパク質の位置と結びついたデータを含む、実施例8から22のいずれか1つの方法。

40

実施例24：DNA結合及び/又はDNA占拠タンパク質が転写因子である、実施例23の方法。

実施例25：位置が、架橋DNAタンパク質複合体のクロマチン免疫沈澱によって決定される、実施例23又は実施例24の方法。

実施例26：位置が、当該組織又は細胞タイプに付随するDNAをヌクレアーゼ（例えばDNase-I）で処理することによって決定される、実施例23又は実施例24の方法。

実施例27：参照マップが、ある組織又は細胞タイプ内のヌクレオソーム、クロマトソーム、又は他のDNA結合若しくはDNA占拠タンパク質の位置又は間隔の取り方に関連する生物学的特徴を含む、実施例8から26のいずれか1つの方法。

実施例28：生物学的特徴が1つ以上の遺伝子の定量的発現である、実施例27の方法。

実施例29：生物学的特徴が1つ以上のヒストンマークの有無である、実施例27又は実施

50

例28の方法。

実施例30：生物学的特徴がヌクレアーゼ切断に対する高感受性である、実施例27から29のいずれか1つの方法。

実施例31：参照マップを作製するために用いられる組織又は細胞タイプが、ある疾患又は異常を有する対象者の一次組織である、実施例8から30のいずれか1つの方法。

実施例32：当該疾患又は異常が、癌、正常妊娠、妊娠（例えば異数性妊娠）の合併症、心筋梗塞、炎症性腸疾患、全身性自己免疫疾患、局所性自己免疫疾患、拒絶を伴う同種間移植、拒絶を伴わない同種間移植、卒中、及び局在性組織損傷から成る群から選択される、実施例31の方法。

実施例33：参照マップを作製するために用いられる組織又は細胞タイプが、健康な対象者の一次組織である、実施例8から30のいずれか1つの方法。

実施例34：参照マップを作製するために用いられる組織又は細胞タイプが不死化細胞株である、実施例8から31のいずれか1つの方法。

【0076】

実施例35：参照マップを作製するために用いられる組織又は細胞タイプが腫瘍の生検である、実施例8から30のいずれか1つの方法。

実施例36：配列データがcfDNAフラグメント末端の位置を含む、実施例18の方法。

実施例37：参照対象者が健康である、実施例36の方法。

実施例38：参照対象者が疾患又は異常を有する、実施例36の方法。

実施例39：疾患又は異常が、癌、正常妊娠、妊娠（例えば異数性妊娠）の合併症、心筋梗塞、炎症性腸疾患、全身性自己免疫疾患、局所性自己免疫疾患、拒絶を伴う同種間移植、拒絶を伴わない同種間移植、卒中、及び局在性組織損傷から成る群から選択される、実施例38の方法。

実施例40：参照マップが、当該組織又は細胞タイプと結びついた参照ゲノムの座標の少なくとも一部分の参照スコアを含む、実施例19から39のいずれか1つの方法。

実施例41：参照マップが当該スコアの数学的変換を含む、実施例40の方法。

実施例42：スコアが、当該組織又は細胞タイプのための全ての参照ゲノム座標のサブセットを表す、実施例40の方法。

実施例43：サブセットがヌクレオソーム及び/又はクロマトソームの位置又は間隔の取り方と結びついている、実施例42の方法。

実施例44：サブセットが転写開始サイト及び/又は転写終了サイトと結びついている、実施例42又は実施例43の方法。

実施例45：サブセットが少なくとも1つの転写因子の結合サイトと結びついている、実施例42から44のいずれか1つの方法。

実施例46：サブセットがヌクレアーゼ高感受性サイトと結びついている、実施例42から45のいずれか1つの方法。

【0077】

実施例47：サブセットが、さらに加えて少なくとも1つのオルトゴナルな生物学的特徴と結びついている、実施例40から46のいずれか1つの方法。

実施例48：オルトゴナルな生物学的特徴が高発現遺伝子と結びついている、実施例47の方法。

実施例49：オルトゴナルな生物学的特徴が低発現遺伝子と結びついている、実施例47の方法。

実施例50：数学的変換がフーリエ変換を含む、実施例41から49のいずれか1つの方法。

実施例51：複数のスコアの少なくとも1つのサブセットが閾値を超えるスコアを有する、実施例11から50のいずれか1つの方法。

実施例52：cfDNAを生じる組織及び/又は細胞タイプを、当該cfDNAフラグメント末端の少なくともいくつかの複数のゲノム上の場所の関数として決定する工程が、当該cfDNAフラグメント末端の少なくともいくつかの複数のゲノム上の場所のフーリエ変換、又はその数学的変換を参照マップと比較する工程を含む、実施例7から51のいずれか1つの方法。

10

20

30

40

50

実施例53：単離cfDNAを生じると決定された組織及び/又は細胞タイプの一覧を含む報告を作製する工程をさらに含む、先行する実施例のいずれかの方法。

実施例54：対象者で疾患又は異常を識別する方法であって、以下の工程を含む方法：当該対象者からの生物学的サンプルからセルフリーDNA(cfDNA)を単離する工程であって、当該単離cfDNAが複数のcfDNAフラグメントを含む、前記工程；当該複数のcfDNAフラグメントの少なくとも一部分に付随する配列を決定する工程；当該複数のcfDNAフラグメントの少なくともいくつかのcfDNAフラグメント末端の参照ゲノム内のゲノム上の場所を当該cfDNAフラグメント配列の関数として決定する工程；当該cfDNAを生じる組織及び/又は細胞タイプの少なくともいくつかを、当該cfDNAフラグメント末端の少なくともいくつかのゲノム上の場所の関数として決定する工程；及び当該疾患又は異常を、当該cfDNAを生じると決定された組織及び/又は細胞タイプの関数と認定する工程。

10

【0078】

実施例55：cfDNAフラグメントを生じる組織及び/又は細胞タイプを決定する工程が、当該cfDNAフラグメント末端の少なくともいくつかのゲノム上の場所を1つ以上の参照マップと比較する工程を含む、実施例54の方法。

実施例56：cfDNAフラグメントを生じる組織及び/又は細胞タイプを決定する工程が、当該複数のcfDNAフラグメント末端の少なくともいくつかのゲノム上の場所の分布に対して数学的変換を実施する工程を含む、実施例54又は実施例55の方法。

実施例57：数学的変換がフーリエ変換を含む、実施例56の方法。

20

実施例58：参照ゲノムの少なくともいくつかの座標の各々についてスコアを決定する工程をさらに含み、ここで当該スコアは、当該少なくとも複数のcfDNAフラグメント末端及びそれらのゲノム上の場所の関数として決定され、さらに当該観察されるcfDNAフラグメントを生じる組織及び/又は細胞タイプの少なくともいくつかを決定する工程が、当該スコアと1つ以上の参照マップを比較する工程を含む、実施例54から57のいずれか1つの方法。

実施例59：座標についてのスコアが、当該座標がcfDNAフラグメント末端の位置である確率を表しているか又は前記と関係性がある、実施例58の方法。

実施例60：参照マップが、DNase I高感受性サイトマップ、RNA発現マップ、発現データ、染色体コンフォーメーションマップ、クロマチン接近容易性マップ、クロマチン断片化マップ、又は少なくとも1人の参照対象者から入手され、かつある疾患又は異常と結びついている少なくとも1つの細胞タイプ又は組織に対応するサンプルから得られる配列データ、及び/又はある組織又は細胞タイプのヌクレオソーム及び/又はクロマトソームの位置又は間隔を含む、実施例55から59のいずれか1つの方法。

30

実施例61：参照マップが、少なくとも1つの細胞タイプ又は組織からのクロマチンを外因性ヌクレアーゼ（例えばマイクロコッカスヌクレアーゼ）で消化することによって作製される、実施例55から60のいずれか1つの方法。

実施例62：参照マップが、転移による方法（例えばATAC-seq）を少なくとも1つの細胞タイプ又は組織の核又はクロマチンに適用することによって決定されたクロマチン接近容易性データを含む、実施例60又は実施例61の方法。

40

【0079】

実施例63：参照マップが、組織又は細胞タイプのDNA結合及び/又はDNA占拠タンパク質の位置と結びついたデータを含む、実施例55から62のいずれか1つの方法。

実施例64：DNA結合及び/又はDNA占拠タンパク質が転写因子である、実施例63の方法。

実施例65：位置が、架橋DNA-タンパク質複合体のクロマチン免疫沈澱を少なくとも1つの細胞タイプ又は組織に適用することによって決定される、実施例63又は実施例64の方法。

実施例66：位置が、当該組織又は細胞タイプに付随するDNAをヌクレアーゼ（例えばDNase-I）で処理することによって決定される、実施例63又は実施例64の方法。

実施例67：参照マップが、ある組織又は細胞タイプ内のヌクレオソーム、クロマトソーム

50

ム、又は他のDNA結合若しくはDNA占拠タンパク質の位置又は間隔の取り方に関連する生物学的特徴を含む、実施例54から66のいずれか1つの方法。

実施例68：生物学的特徴が1つ以上の遺伝子の定量的発現である、実施例67の方法。

実施例69：生物学的特徴が1つ以上のヒストンマークの有無である、実施例67又は実施例68の方法。

実施例70：生物学的特徴がヌクレアーゼ切断に対する高感受性である、実施例67から69のいずれか1つの方法。

実施例71：参照マップを作製するために用いられる組織又は細胞タイプが、ある疾患又は異常を有する対象者の一次組織である、実施例55から70のいずれか1つの方法。

実施例72：疾患又は異常が、癌、正常妊娠、妊娠（例えば異数性妊娠）の合併症、心筋梗塞、炎症性腸疾患、全身性自己免疫疾患、局所性自己免疫疾患、拒絶を伴う同種間移植、拒絶を伴わない同種間移植、卒中、及び局在性組織損傷から成る群から選択される、実施例71の方法。

実施例73：参照マップを作製するために用いられる組織又は細胞タイプが、健康な対象者の一次組織である、実施例55から70のいずれか1つの方法。

実施例74：参照マップを作製するために用いられる組織又は細胞タイプが不死化細胞株である、実施例55から70のいずれか1つの方法。

実施例75：参照マップを作製するために用いられる組織又は細胞タイプが腫瘍の生検である、実施例55から70のいずれか1つの方法。

【0080】

実施例76：少なくとも1人の参照対象者から入手されるサンプルから得られる配列データがcfDNAフラグメント末端確率の位置を含む、実施例60の方法。

実施例77：参照対象者が健康である、実施例76の方法。

実施例78：参照対象者が疾患又は異常を有する、実施例76の方法。

実施例79：疾患又は異常が、癌、正常妊娠、妊娠（例えば異数性妊娠）の合併症、心筋梗塞、炎症性腸疾患、全身性自己免疫疾患、局所性自己免疫疾患、拒絶を伴う同種間移植、拒絶を伴わない同種間移植、卒中、及び局在性組織損傷から成る群から選択される、実施例78の方法。

実施例80：参照マップが、当該組織又は細胞タイプに付随する参照ゲノムの少なくとも一部分についてcfDNAフラグメント末端確率を含む、実施例60から79のいずれか1つの方法。

実施例81：参照マップが当該cfDNAフラグメント末端確率の数学的変換を含む、実施例80の方法。

実施例82：cfDNAフラグメント末端確率が、当該組織又は細胞タイプの全ての参照ゲノム座標のサブセットを表す、実施例80の方法。

実施例83：サブセットがヌクレオソーム及び/又はクロマトソームの位置又は間隔と結びついている、実施例82の方法。

実施例84：サブセットが転写開始サイト及び/又は転写終了サイトと結びついている、実施例82又は実施例83の方法。

実施例85：サブセットが少なくとも1つの転写因子の結合サイトと結びついている、実施例82から84のいずれか1つの方法。

実施例86：サブセットがヌクレアーゼ高感受性サイトと結びついている、実施例82から85のいずれか1つの方法。

【0081】

実施例87：サブセットが、さらに加えて少なくとも1つのオルトゴナルな生物学的特徴と結びついている、実施例82から86のいずれか1つの方法。

実施例88：オルトゴナルな生物学的特徴が高発現遺伝子に付随する、実施例87の方法。

実施例89：オルトゴナルな生物学的特徴が低発現遺伝子に付随する、実施例87の方法。

実施例90：数学的変換がフーリエ変換を含む、実施例81から89のいずれか1つの方法。

実施例91：複数のcfDNAフラグメント末端スコアの少なくとも1つのサブセットが各々関

10

20

30

40

50

値を超えるスコアを有する、実施例58から90のいずれか1つの方法。

実施例92：cfDNAの組織及び/又は細胞タイプを、当該cfDNAフラグメント末端の少なくともいくつかの複数のゲノム上の場所の関数として決定する工程が、当該cfDNAフラグメント末端の少なくともいくつかの複数のゲノム上の場所のフリー変換、又はその数学的変換を参照マップと比較する工程を含む、実施例54から91のいずれか1つの方法。

実施例93：参照マップが、当該疾患又は異常に付随する少なくとも1つの組織に対応するDNA又はクロマチン断片化データを含む、実施例54から92のいずれか1つの方法。

実施例94：参照ゲノムがヒトに付随する、実施例54から93のいずれか1つの方法。

実施例95：疾患又は異常を識別する記述を含む報告を作製する工程をさらに含む、実施例54から94のいずれか1つの方法。

実施例96：報告が、決定された当該単離cfDNAの組織及び/又は細胞タイプの一覧をさらに含む、実施例95の方法。

実施例97：生物学的サンプルが、全血、末梢血血漿、尿又は脳脊髄液を含むか、本質的に前記から成るか、又は前記から成る、先行する実施例のいずれかの方法。

【0082】

実施例98：対象者でセルフリーDNA (cfDNA) を生じる組織及び/又は細胞タイプを決定する方法であって、

(i) 当該対象者の生物学的サンプルの入手、当該生物学的サンプルのcfDNAの単離、並びにcfDNAのライブラリー構築及び大量パラレルシーケンシングによる下記の分布(a)、(b)及び/又は(c)の測定によってヌクレオソームマップを作製する工程；

(ii) コントロール対象者又は既知疾患を有する対象者の生物学的サンプルの入手、当該生物学的サンプルのcfDNAの単離、cfDNAのライブラリー構築及び大量パラレルシーケンシングによる下記の分布(a)、(b)及び/又は(c)の測定によってヌクレオソームマップの参照セットを作製する工程；並びに

(iii) 当該cfDNAに由来するヌクレオソームマップをヌクレオソームマップの参照セットと比較することによって、当該cfDNAを生じる当該組織及び/又は細胞タイプを決定する工程を含み、

ここで前記(a)、(b)及び(c)は、

(a) ヒトゲノムにおいて任意の特定の塩基対がcfDNAフラグメントの末端に出現する可能性の分布；

(b) ヒトゲノムの塩基対の任意の対がcfDNAフラグメントの末端の対として出現する可能性の分布；及び

(c) ヒトゲノムにおいて任意の特定の塩基対が弁別的なヌクレオソーム占拠の結果としてcfDNAフラグメントに出現する可能性の分布である、前記方法。

【0083】

実施例99：対象者においてセルフリーDNAを生じる組織及び/又は細胞タイプを決定する方法であって、

(i) 当該対象者の生物学的サンプルの入手、当該生物学的サンプルのcfDNAの単離、並びにcfDNAのライブラリー構築及び大量パラレルシーケンシングによる下記の分布(a)、(b)及び/又は(c)の測定によってヌクレオソームマップを作製する工程；

(ii) コントロール対象者又は既知疾患を有する対象者の生物学的サンプルの入手、当該生物学的サンプルのcfDNAの単離、マイクロココカスのヌクレアーゼ(MNase)、DNase処理又はATAC-Seqによるクロマチンの消化から誘導したDNAのライブラリー構築及び大量パラレルシーケンシングによる下記の分布(a)、(b)及び/又は(c)の測定によってヌクレオソームマップの参照セットを作製する工程；並びに

(iii) 当該cfDNAに由来するヌクレオソームマップをヌクレオソームマップの参照セットと比較することによって、当該cfDNAを生じる当該組織及び/又は細胞タイプを決定する工程を含み、

ここで前記(a)、(b)及び(c)は、

(a) ヒトゲノムにおいて任意の特定の塩基対が配列決定されたフラグメントの末端に出

10

20

30

40

50

現する可能性の分布；

(b) ヒトゲノムの塩基対の任意の対が配列決定されたフラグメントの末端の対として出現する可能性の分布；及び

(c) ヒトゲノムにおいて任意の特定の塩基対が弁別的なヌクレオソーム占拠の結果として配列決定されたフラグメントに出現する可能性の分布である、前記方法。

【0084】

実施例100：対象者における臨床症状を診断する方法であって、

(i) 当該対象者の生物学的サンプルの入手、当該生物学的サンプルのcfDNAの単離、並びにcfDNAのライブラリー構築及び大量パラレルシーケンシングによる下記の分布(a)、(b)及び/又は(c)の測定によってヌクレオソームマップを作製する工程；

10

(ii) コントロール対象者又は既知疾患を有する対象者の生物学的サンプルの入手、当該生物学的サンプルのcfDNAの単離、cfDNAのライブラリー構築及び大量パラレルシーケンシングによる下記の分布(a)、(b)及び/又は(c)の測定によってヌクレオソームマップの参照セットを作製する工程；並びに

(iii) 当該cfDNAに由来するヌクレオソームマップをヌクレオソームマップの参照セットと比較することによって、当該臨床症状を決定する工程を含み、ここで前記(a)、(b)及び(c)は、

(a) ヒトゲノムにおいて任意の特定の塩基対がcfDNAフラグメントの末端に出現する可能性の分布；

(b) ヒトゲノムの塩基対の任意の対がcfDNAフラグメントの末端の対として出現する可能性の分布；及び

20

(c) ヒトゲノムにおいて任意の特定の塩基対が弁別的なヌクレオソーム占拠の結果としてcfDNAフラグメントに出現する可能性の分布である、前記方法。

【0085】

実施例101：対象者における臨床症状を診断する方法であって、

(i) 当該対象者の生物学的サンプルの入手、当該生物学的サンプルのcfDNAの単離、並びにcfDNAのライブラリー構築及び大量パラレルシーケンシングによる下記の分布(a)、(b)及び/又は(c)の測定によってヌクレオソームマップを作製する工程；

30

(ii) コントロール対象者又は既知疾患を有する対象者の生物学的サンプルの入手、当該生物学的サンプルのcfDNAの単離、マイクロココカスのヌクレアーゼ(MNase)、DNase処理又はATAC-Seqによるクロマチンの消化から誘導したDNAのライブラリー構築及び大量パラレルシーケンシングによる下記の分布(a)、(b)及び/又は(c)の測定によってヌクレオソームマップの参照セットを作製する工程；並びに

(iii) 当該cfDNAに由来するヌクレオソームマップをヌクレオソームマップの参照セットと比較することによって、当該cfDNAの供給源組織の組成を決定する工程を含み、ここで前記(a)、(b)及び(c)は、

(a) ヒトゲノムにおいて任意の特定の塩基対が配列決定されたフラグメントの末端に出現する可能性の分布；

(b) ヒトゲノムの塩基対の任意の対が配列決定されたフラグメントの末端の対として出現する可能性の分布；及び

40

(c) ヒトゲノムにおいて任意の特定の塩基対が弁別的なヌクレオソーム占拠の結果として配列決定されたフラグメントに出現する可能性の分布である、前記方法。

【0086】

実施例102：ヌクレオソームマップが、

当該生物学的サンプルから単離されたcfDNAを精製する工程、

アダプター連結及び場合によってPCR増幅によってライブラリーを構築する工程、並びに、

生じたライブラリーを配列決定する工程、

によって作製される、実施例98 - 101のいずれか1つの方法。

実施例103：ヌクレオソームマップの参照セットが、

50

コントロール対象者からの生物学的サンプルから単離されたcfDNAを精製する工程、アダプター連結及び場合によってPCR増幅によってライブラリーを構築する工程、並びに、

生じたライブラリーを配列決定する工程、
によって作製される、実施例98 - 101のいずれか1つの方法。

実施例104：分布(a)、(b)若しくは(c)又はこれらの分布の1つの数学的変換を連続ウインドウ内でフーリエ変換に付し、続いてヌクレオソーム占拠に付随した周期範囲について強度を定量して、ヌクレオソームが各連続ウインドウ内で構造化された配置を示す範囲を集約する、実施例98 - 101のいずれか1つの方法。

実施例105：分布(a)、(b)若しくは(c)又はこれらの分布の1つの数学的変換において、特定の転写因子(TF)の転写因子結合サイト(TFBS)の直近にシーケンシング読取り開始サイトが位置する参照ヒトゲノムでサイトの分布を定量し(TFがTFBSと結合するときはしばしば当該サイトの直ぐ側にヌクレオソームがフランキングする)、cfDNAに寄与する細胞タイプのTF活性の結果としてのヌクレオソームの配置を集約する、実施例98 - 101のいずれか1つの方法。

【0087】

実施例106：ヌクレオソーム占拠シグナルが、他のゲノム上の目印(例えばDNase I高感受性サイト、転写開始サイト、トポロジカルドメイン、他の後成的な目印、又は他のデータセット(例えば遺伝子発現など)での相関的動態によって規定される全てのそのようなサイトのサブセット)周辺で、分布(a)、(b)及び/又は(c)又はこれらの分布の1つの数学的変換から、集合化シグナルの任意の1つにしたがって集約される、実施例98 - 101のいずれか1つの方法。

実施例107：分布を変換して、当該ゲノムの多様なサブセット内のヌクレオソームの配置の周期的シグナルを集合化又は集約する、例えば、連続ウインドウでの周期性、或いはまた別に、転写因子結合サイト、遺伝子モデル特徴(例えば転写開始サイト)、組織発現データ又は他のヌクレオソーム配置に相関するものによって規定されるゲノムの不連続サブセットでの周期性を定量する、実施例98 - 101のいずれか1つの方法。

実施例108：分布が、組織特異的データ、すなわち組織特異的DNase I高感受性サイト近傍の集合シグナルによって規定される、実施例98 - 101のいずれか1つの方法。

実施例109：追加のヌクレオソームマップを当該参照セットと比較するために統計的シグナルをプロセッシングする工程をさらに含む、実施例98 - 101のいずれか1つの方法。

実施例110：最初に多岐にわたるサンプルセットで当該ゲノムの端から端まで連続ウインドウ内で長レンジヌクレオソーム順番付けを集約し、続いて主要成分分析(PCA)を実施して、サンプルをクラスター化するか又は混合割合を概算する、実施例109の方法。

実施例111：臨床症状が癌、すなわち悪性腫瘍である、実施例100又は実施例101の方法。

実施例112：生物学的サンプルがcfDNAを含む循環血漿であり、そのいくらかの部分が腫瘍に由来する、実施例111の方法。

【0088】

実施例113：臨床症状が、組織損傷、心筋梗塞(心臓組織の急性損傷)、自己免疫疾患(多岐にわたる組織の慢性損傷)、妊娠、染色体異常(例えばトリソミー)、及び移植片拒絶から選択される、実施例100又は実施例101の方法。

実施例114：cfDNAに寄与すると決定された1つ以上の組織又は細胞タイプの各々に割合を割り当てる工程をさらに含む、先行する実施例のいずれかの方法。

実施例115：1つ以上の決定された組織又は細胞タイプの各々に割り当てられた割合が、少なくとも部分的には、1人又は複数の健康な対象者のcfDNAと対比された相関性の程度又は相関性増加の程度を基準にする、実施例114の方法。

実施例116：相関性の程度が、少なくとも部分的には、当該生物学的サンプル由来のcfDNAフラグメント末端の分布の数学的変換と当該決定された組織又は細胞タイプに付随する参照マップとの比較に基づく、実施例114又は実施例115の方法。

10

20

30

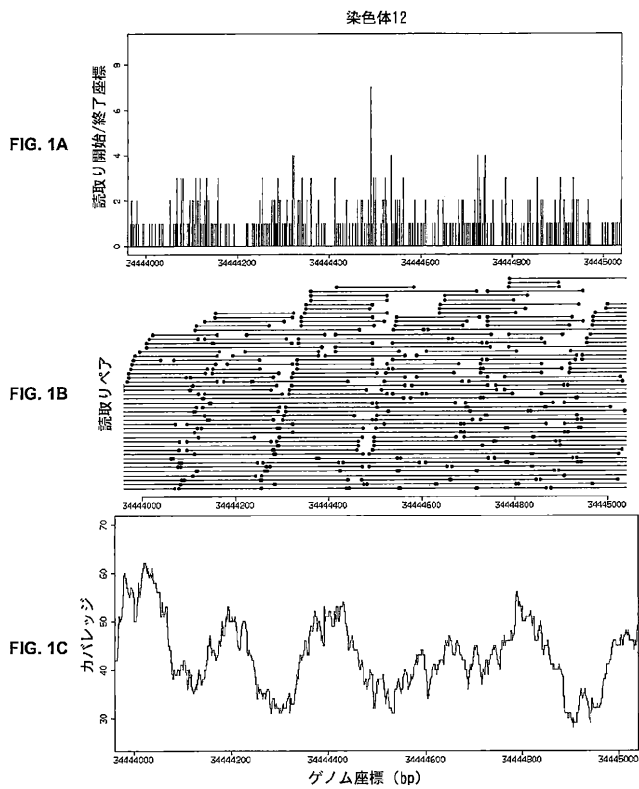
40

50

実施例117：1つ以上の決定された組織又は細胞タイプの各々に割り当てられた割合が混合モデルに基づく、実施例114から116の方法。

上述したように、本発明の具体的な実施態様をこれまで記載してきたが、多様な改変が本発明の範囲を逸脱することなく実施可能であることは理解されるであろう。したがって、本発明は特許請求の範囲の記載によるものを除いて限定されない。

【 図 1 A - C 】



【 図 2 】

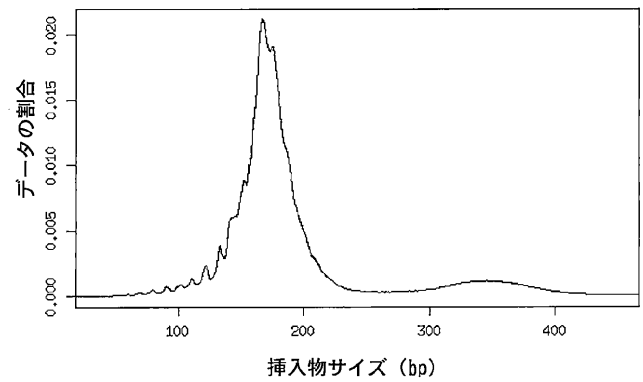


FIG. 2

【 図 5 】

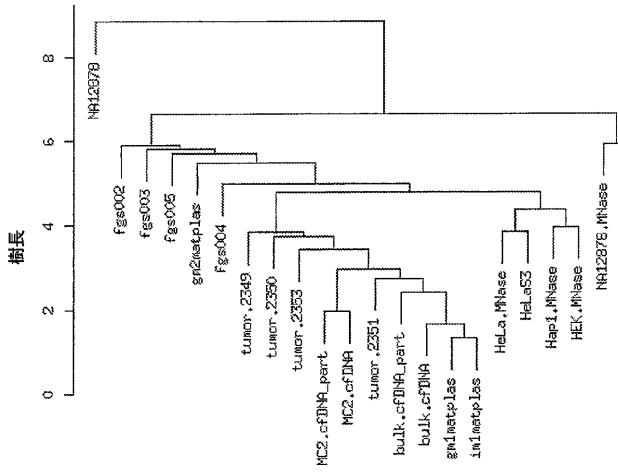


FIG. 5

【 図 6 A 】

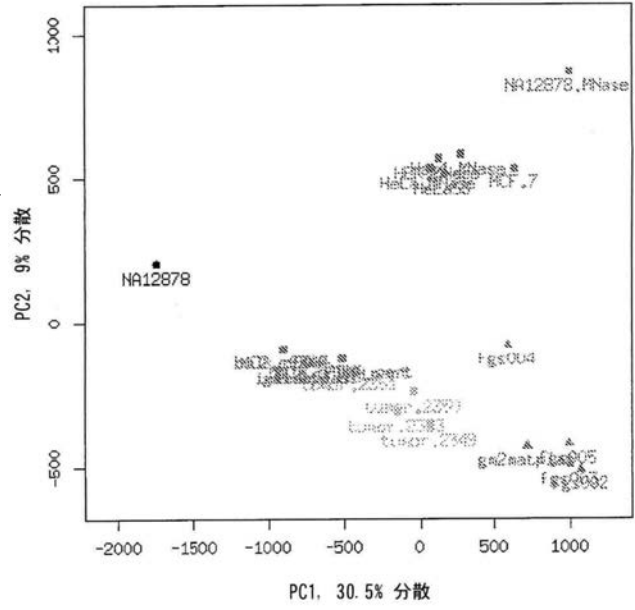


FIG. 6A

【 図 6 B 】

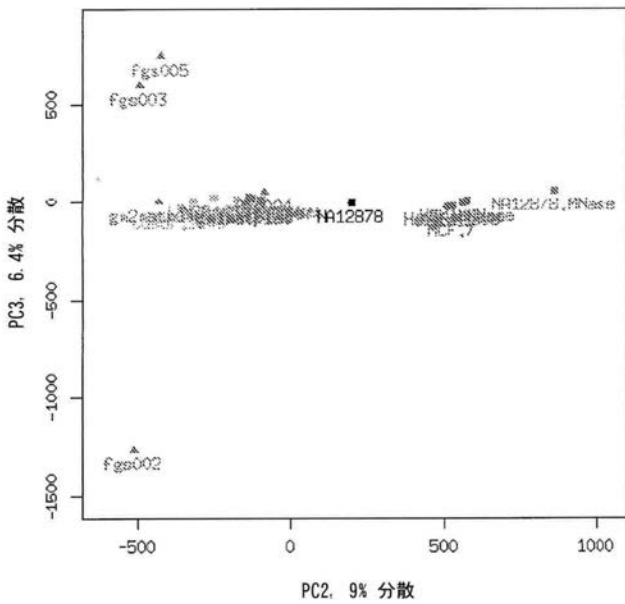


FIG. 6B

【 図 7 】

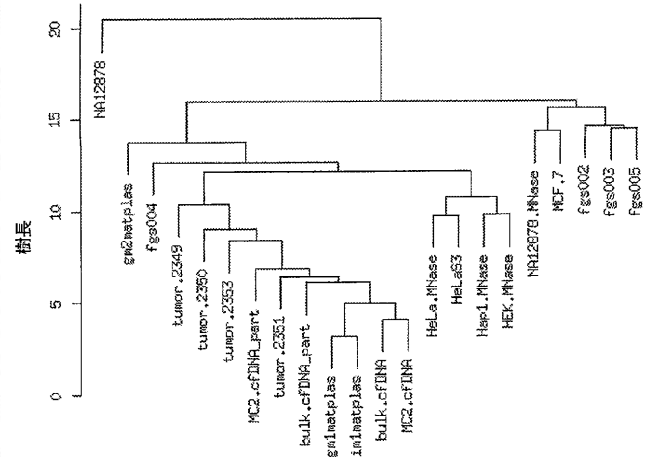
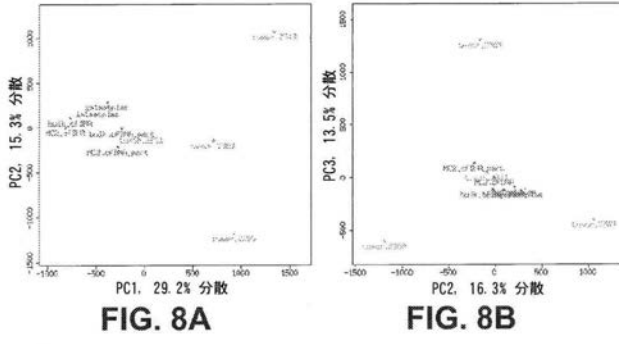
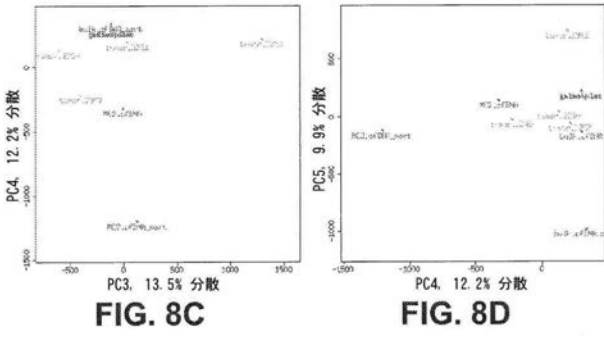
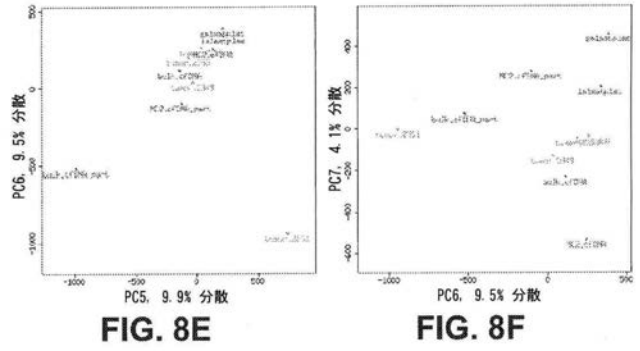


FIG. 7

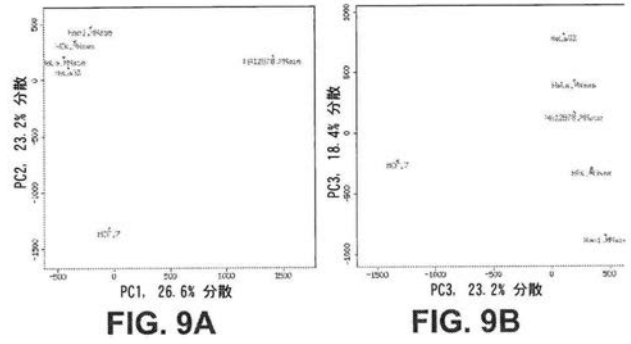
【 図 8 A - D 】



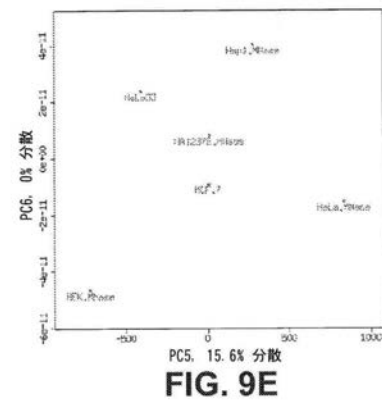
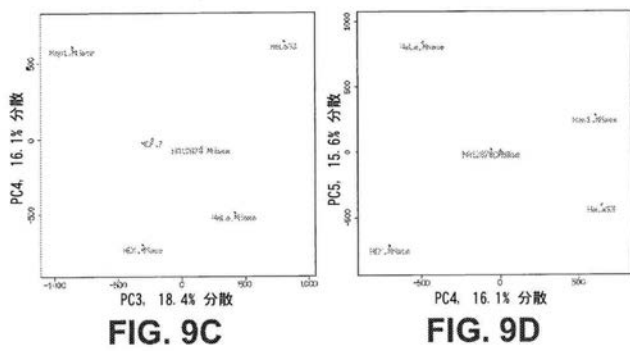
【 図 8 E - F 】



【 図 9 A - B 】



【 図 9 C - E 】



【 図 1 0 】

開始位置のピリオドグラムの強度についてchr11ブロックの平均

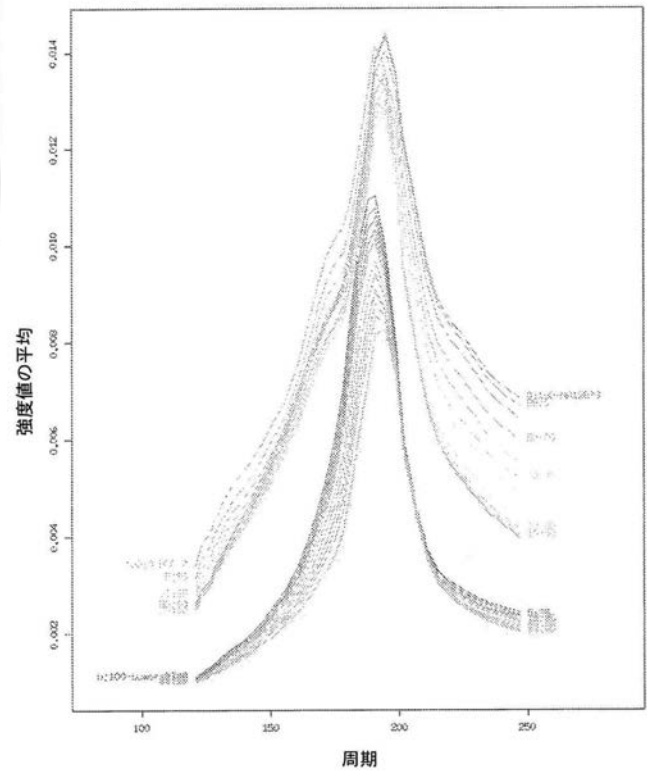


FIG. 10

【 図 1 1 】

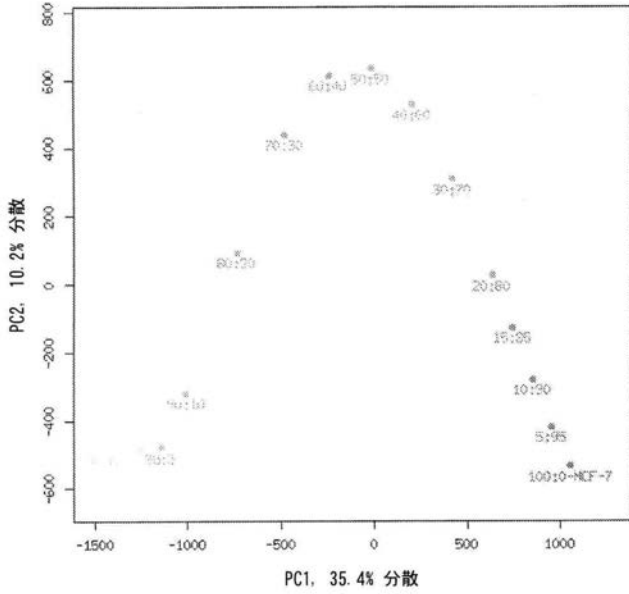


FIG. 11

【 図 1 2 】

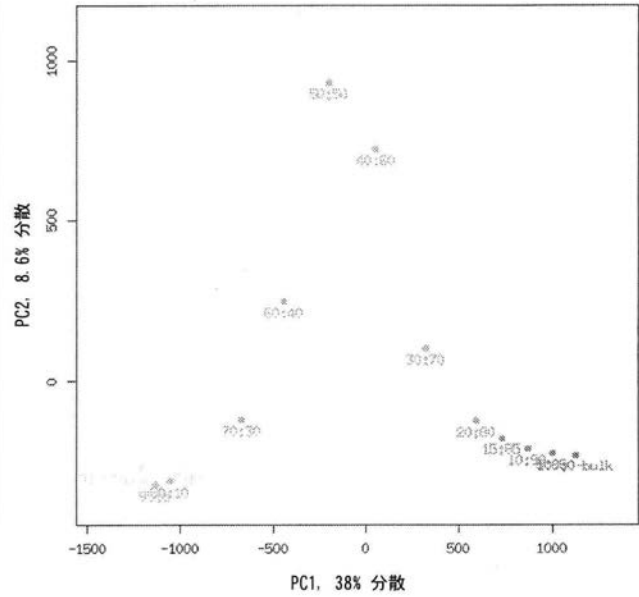


FIG. 12

【 図 1 3 】

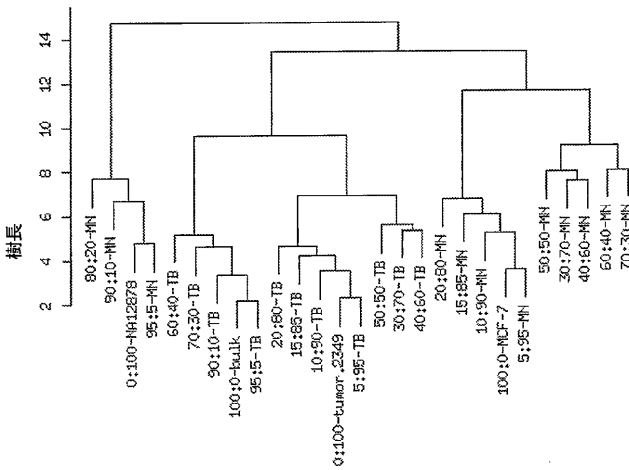


FIG. 13

【 図 1 4 】

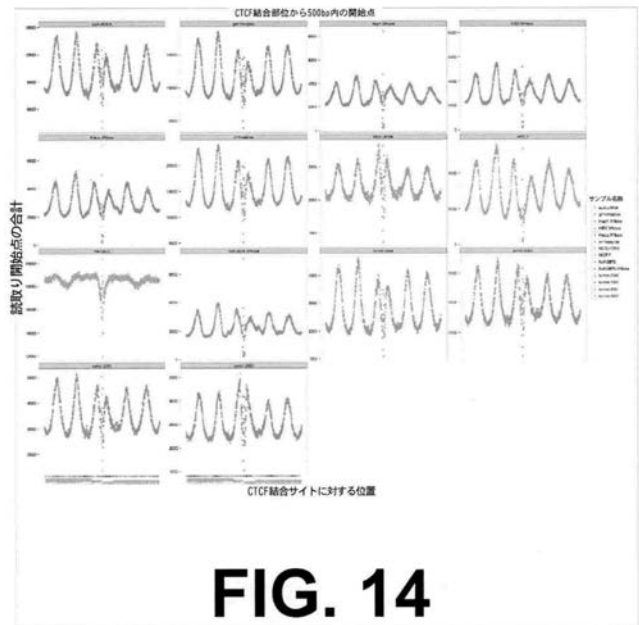


FIG. 14

【 図 1 5 】

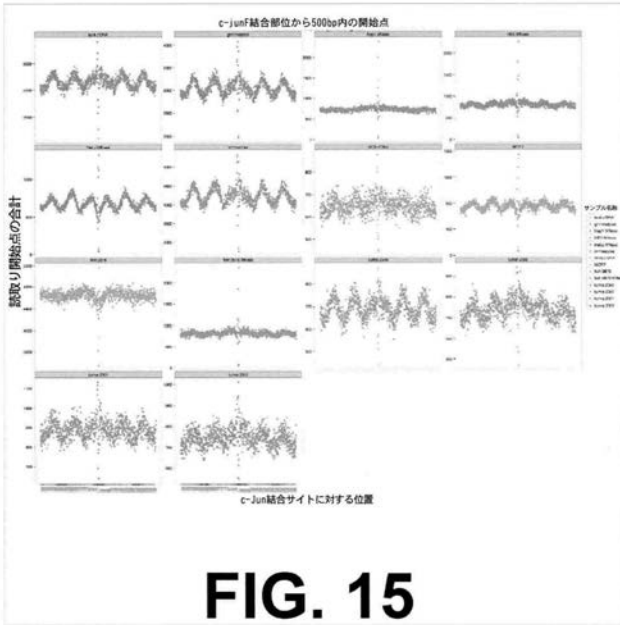


FIG. 15

【 図 1 6 】

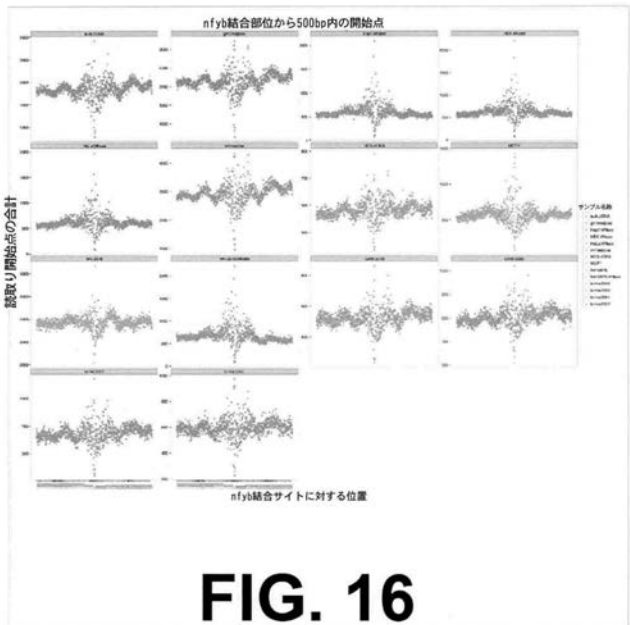


FIG. 16

【 図 1 7 】

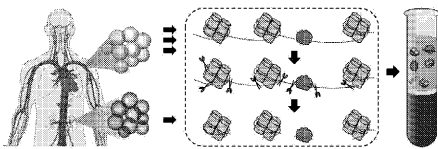


FIG. 17

【 図 1 9 】

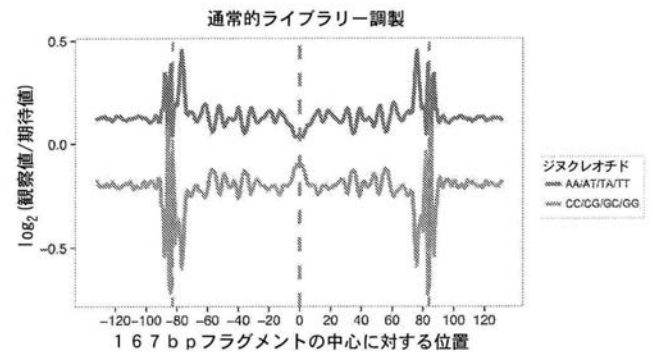


FIG. 19

【 図 1 8 】

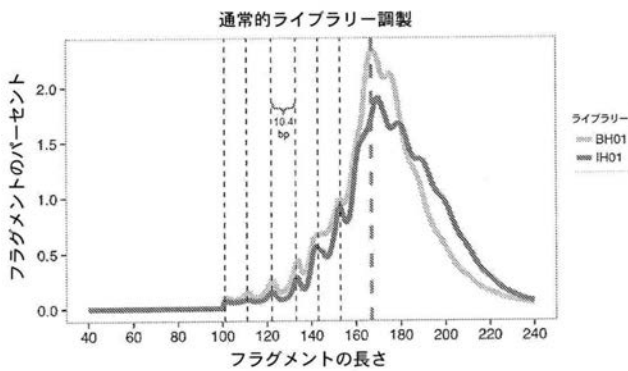


FIG. 18

【 図 2 4 A 】

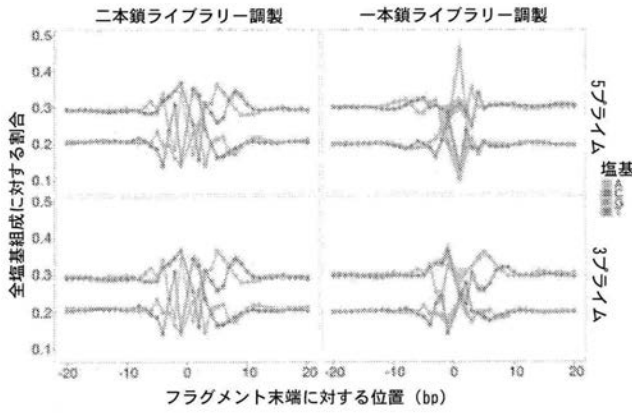


FIG. 24A

【 図 2 4 B 】

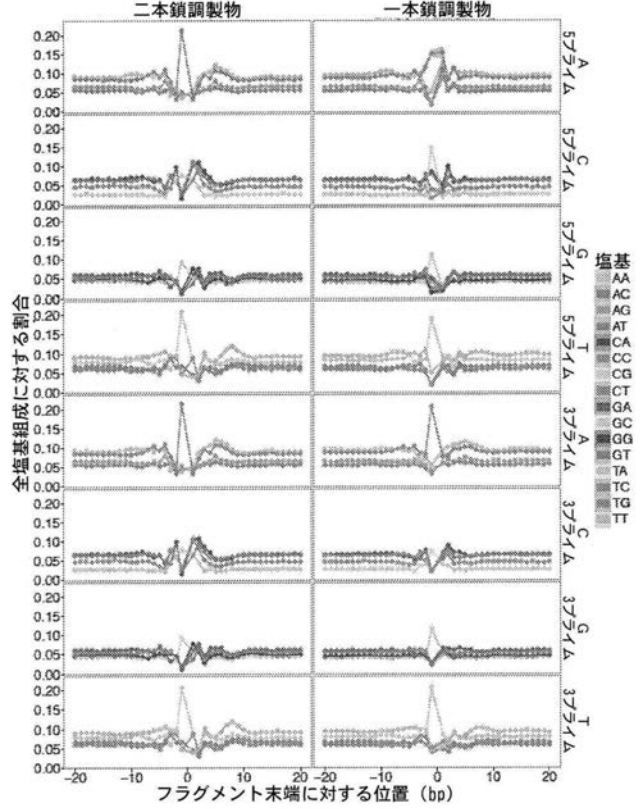


FIG. 24B

【 図 2 5 】

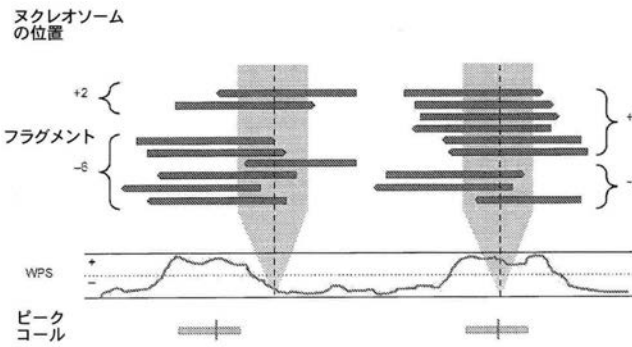


FIG. 25

【 図 2 6 】

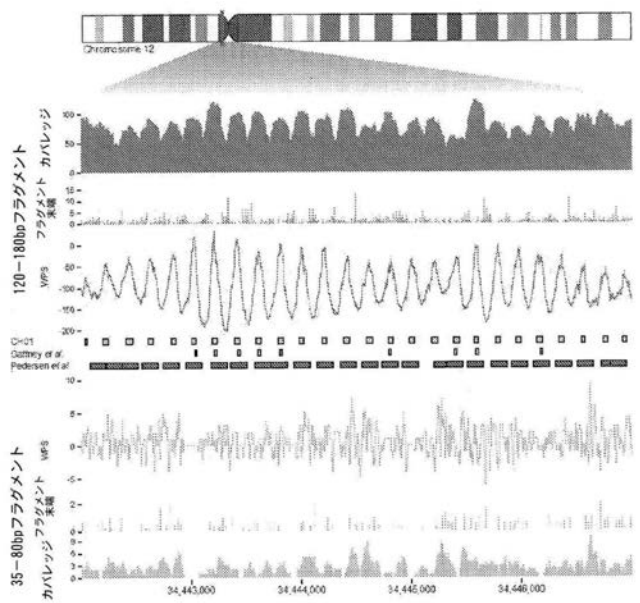


FIG. 26

【 図 2 7 】

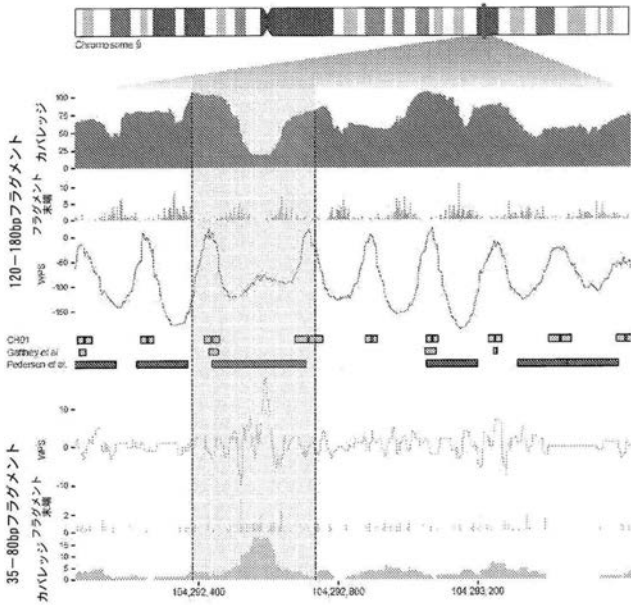


FIG. 27

【 図 2 8 】

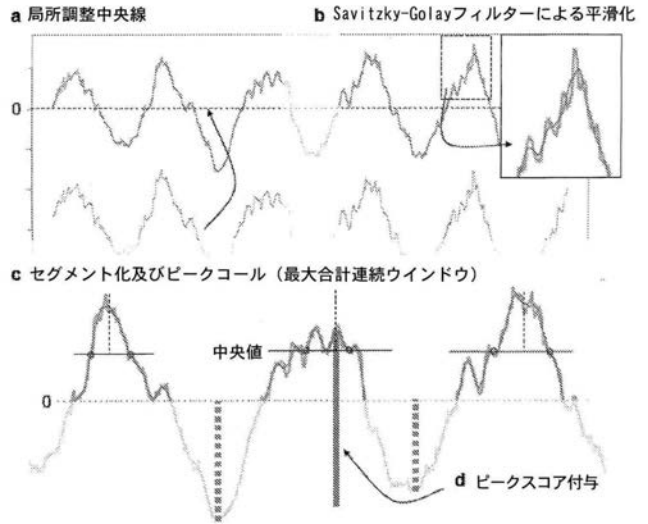


FIG. 28

【 図 2 9 】

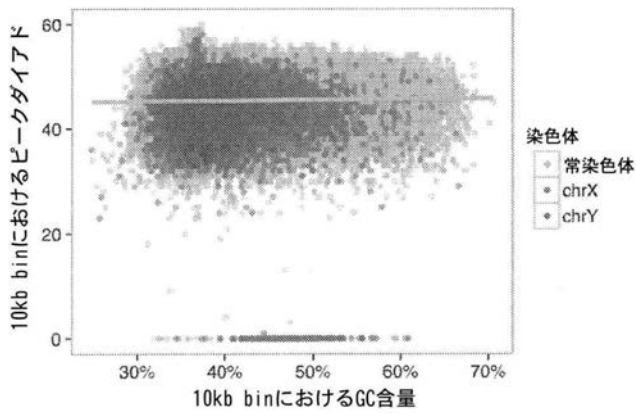


FIG. 29

【 図 3 0 】

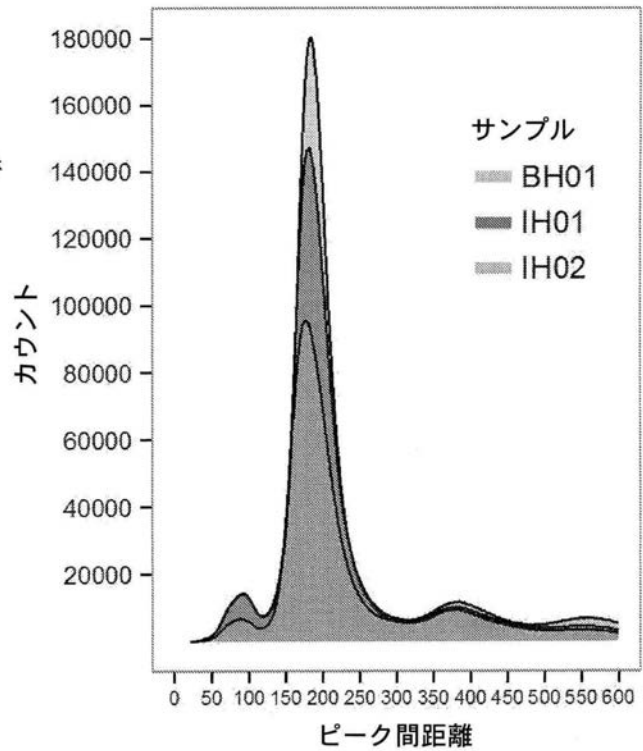


FIG. 30

【 図 3 1 】

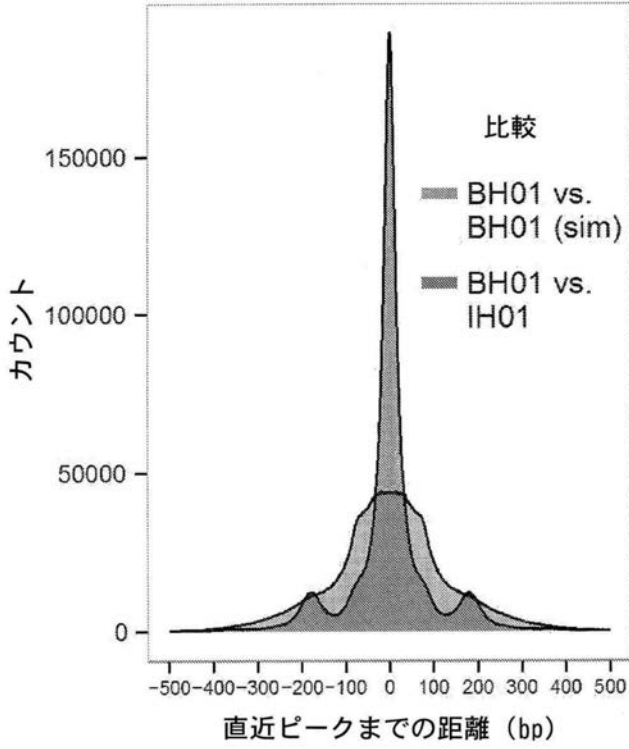


FIG. 31

【 図 3 2 】

FIG. 32A

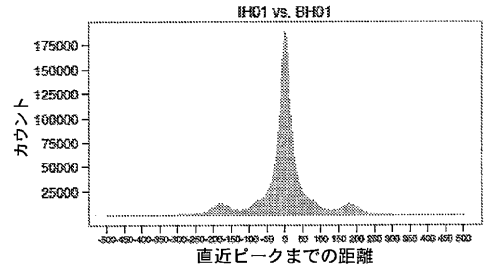


FIG. 32B

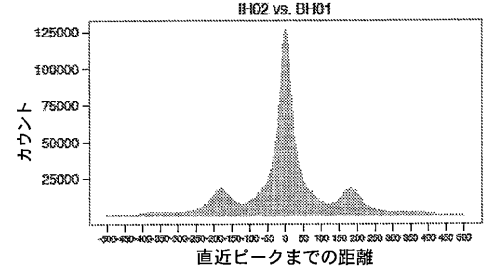
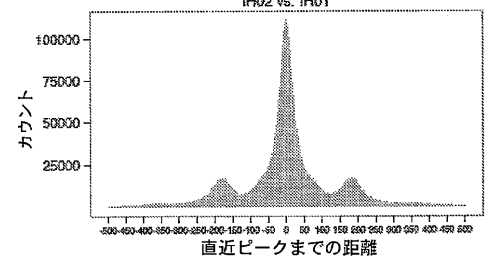


FIG. 32C



【 図 3 3 A 】

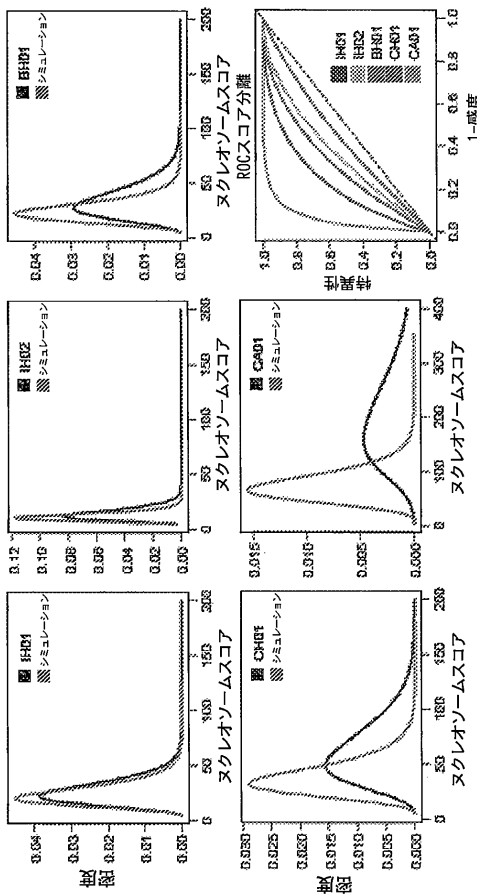


FIG. 33A

【 図 3 3 B 】

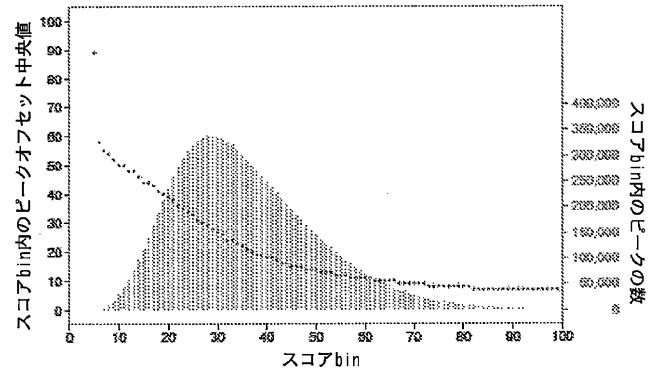


FIG. 33B

【 図 3 4 A - C 】

FIG. 34A

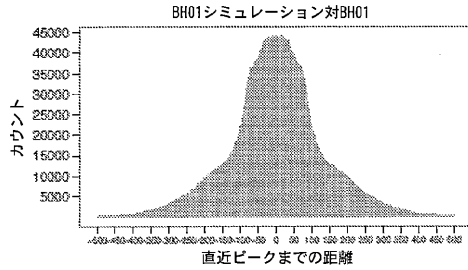


FIG. 34B

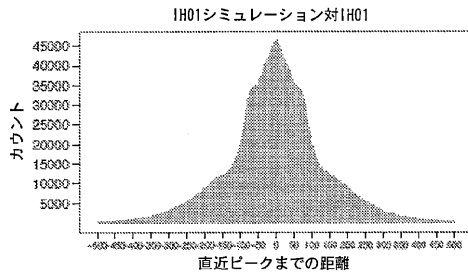
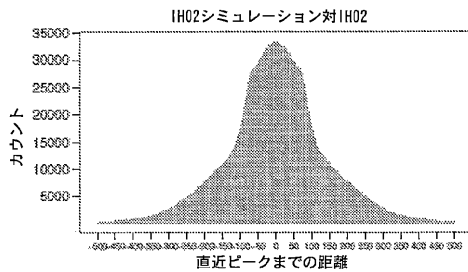


FIG. 34C



【 図 3 5 】

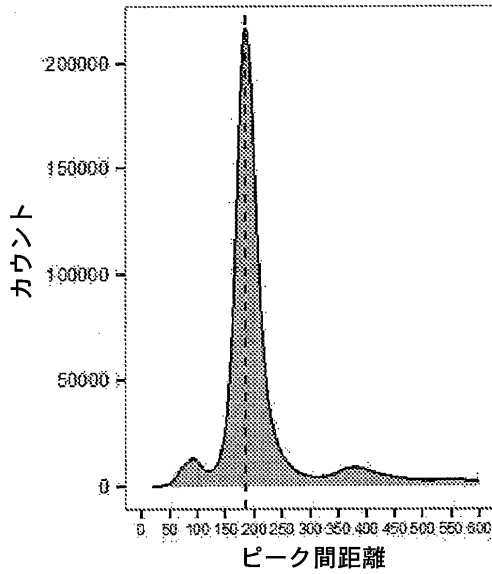
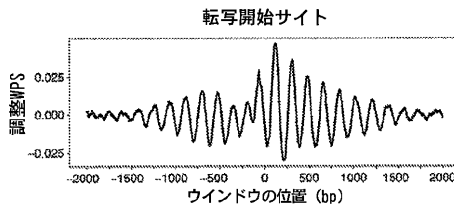


FIG. 35

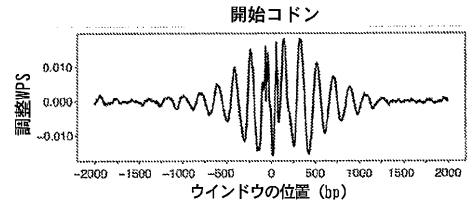
【 図 3 6 】

FIG. 36



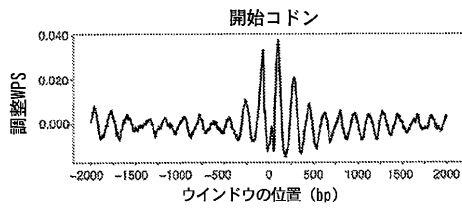
【 図 3 9 】

FIG. 39



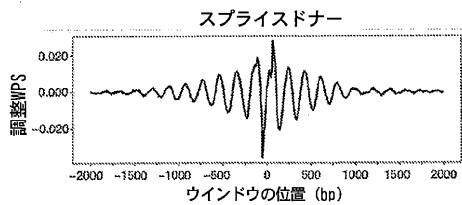
【 図 3 7 】

FIG. 37



【 図 3 8 】

FIG. 38



【 図 4 0 】

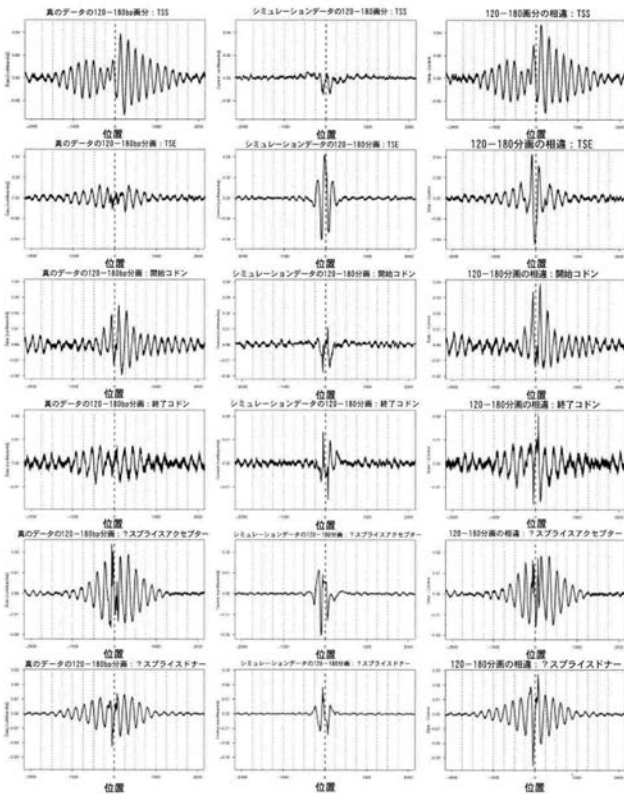


FIG. 40

【 図 4 1 】

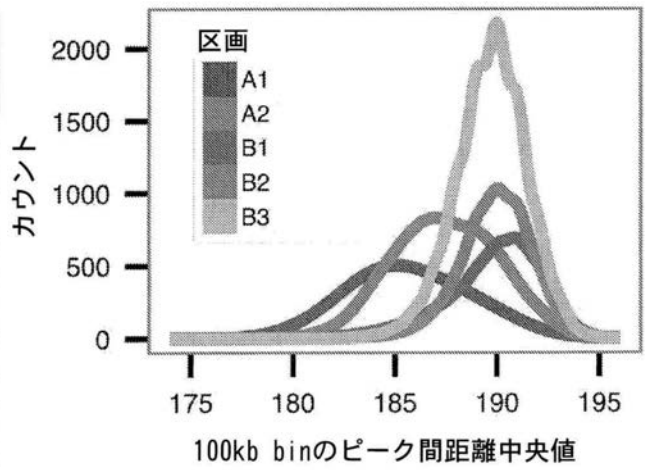


FIG. 41

【 図 4 2 】

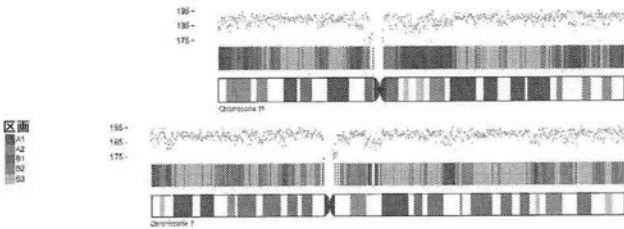


FIG. 42

【 図 4 3 】

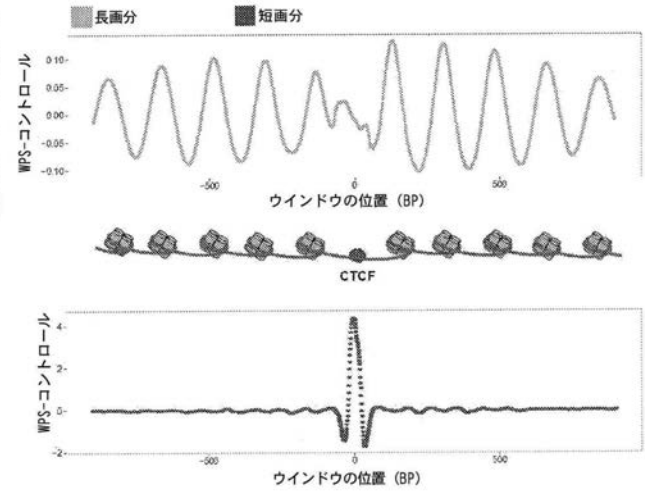


FIG. 43

【 図 4 4 】

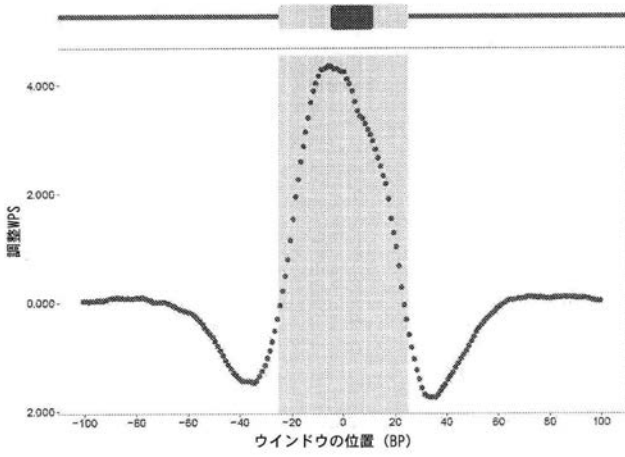


FIG. 44

【 図 4 5 】

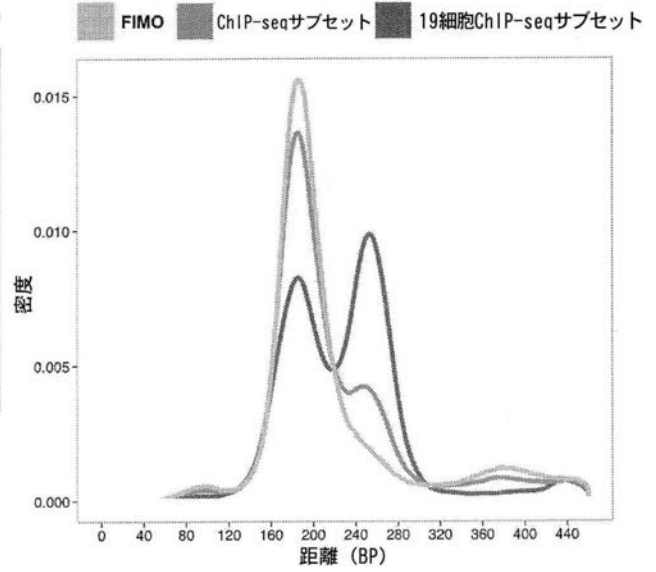


FIG. 45

【 図 4 6 】

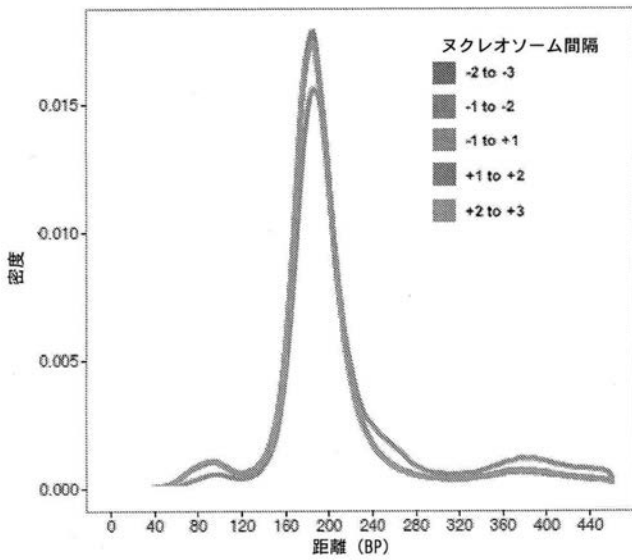


FIG. 46

【 図 4 7 】

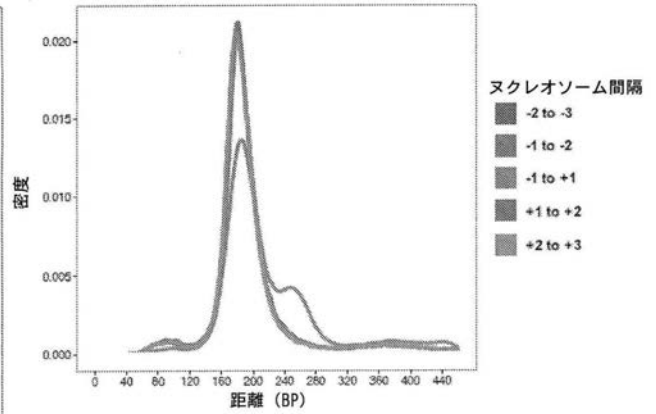


FIG. 47

【 図 4 8 】

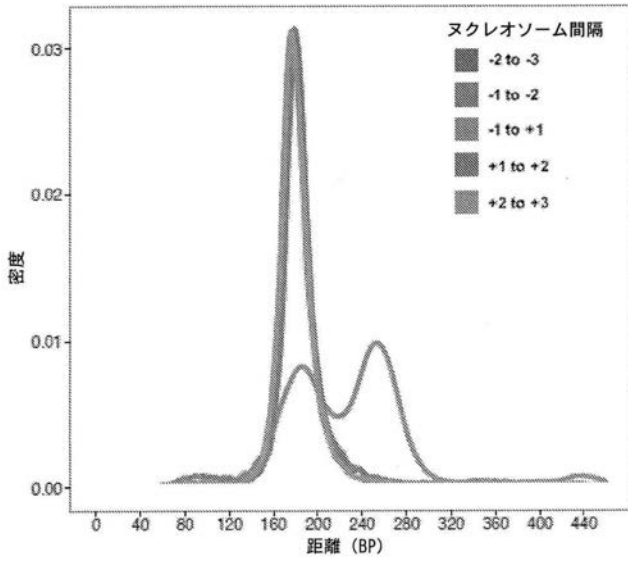


FIG. 48

【 図 4 9 】

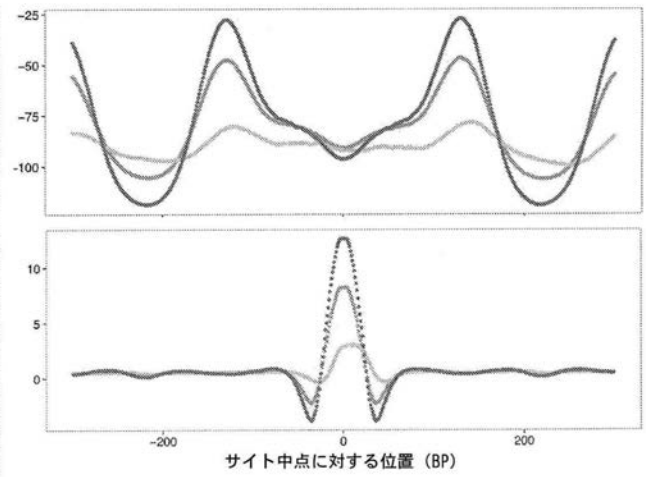


FIG. 49

【 図 5 0 】

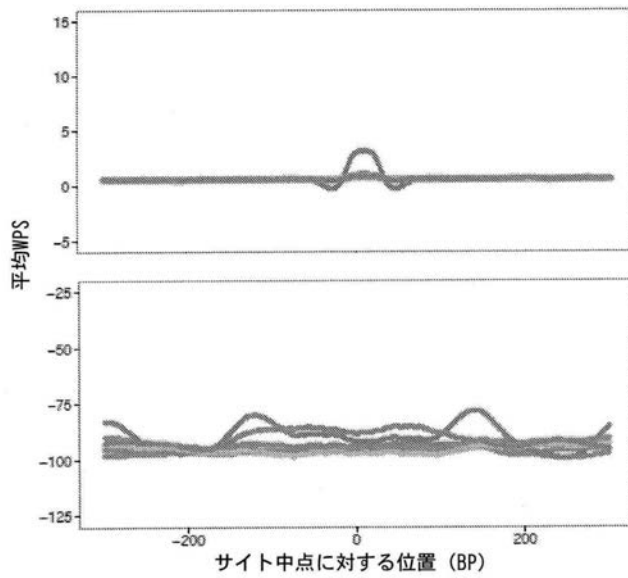


FIG. 50

【 図 5 1 】

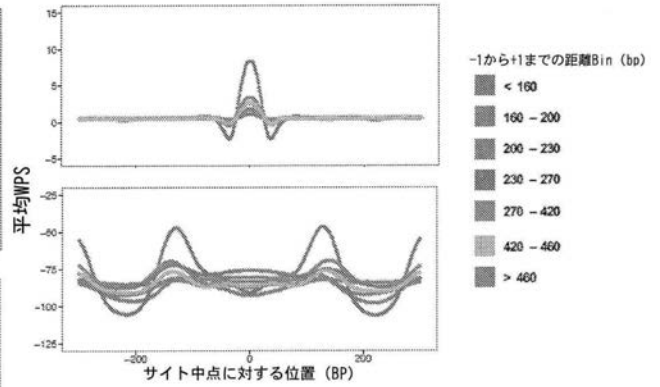


FIG. 51

【 図 5 2 】

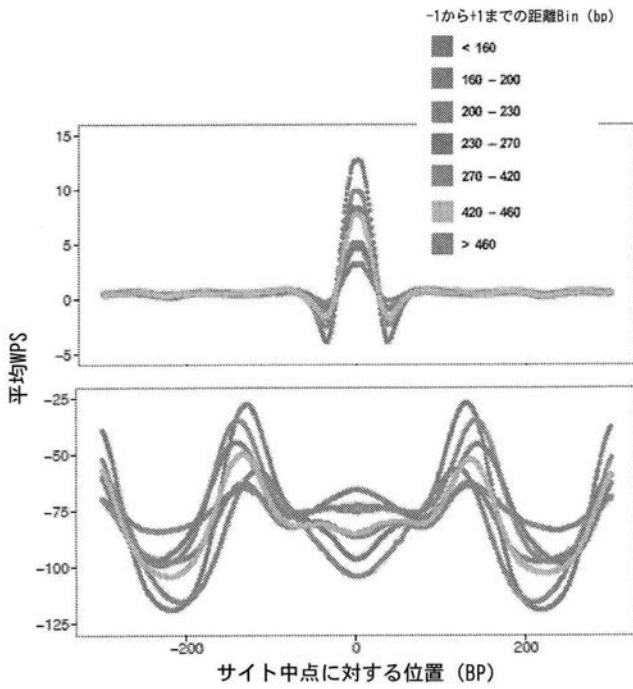


FIG. 52

【 図 5 3 A - D 】

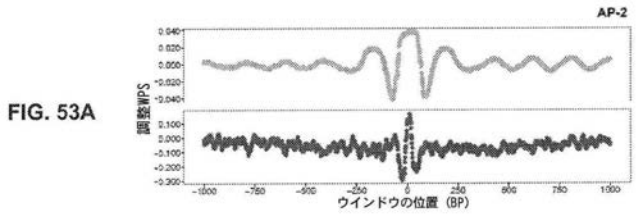


FIG. 53A

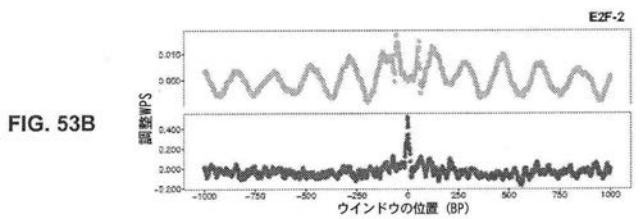


FIG. 53B

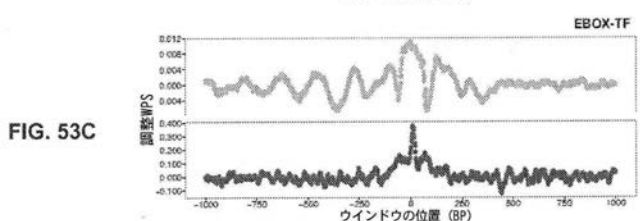


FIG. 53C

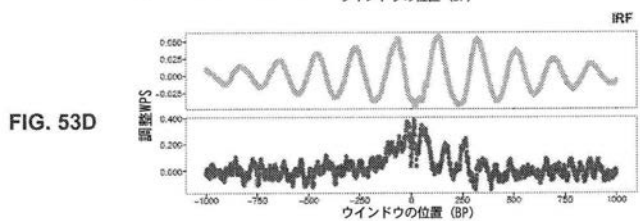


FIG. 53D

【 図 5 3 E - H 】

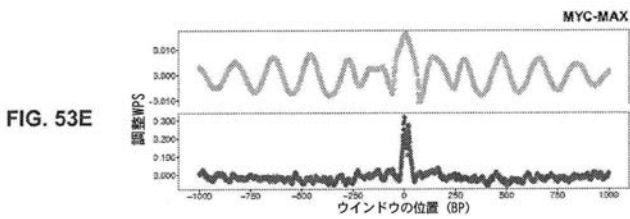


FIG. 53E

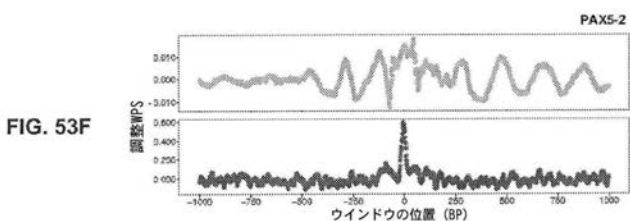


FIG. 53F

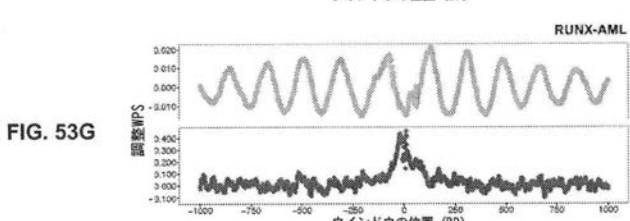


FIG. 53G

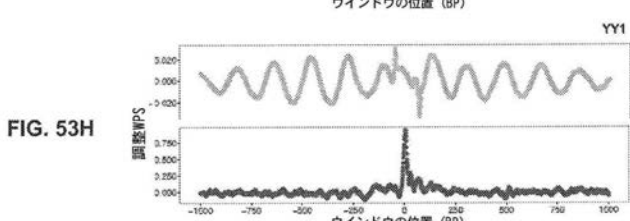


FIG. 53H

【 図 5 4 】

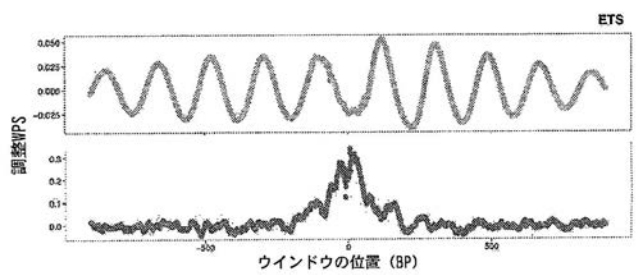


FIG. 54

【 図 5 5 】

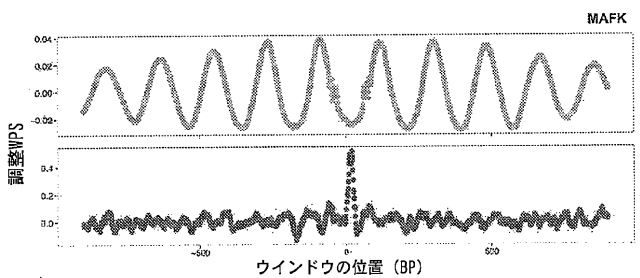


FIG. 55

【 図 5 6 】

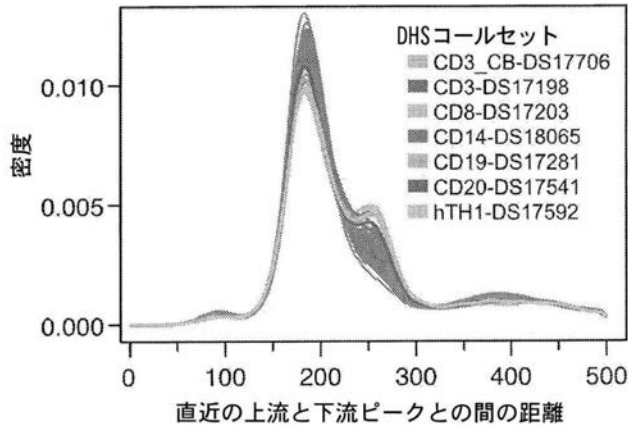


FIG. 56

【 図 5 7 】

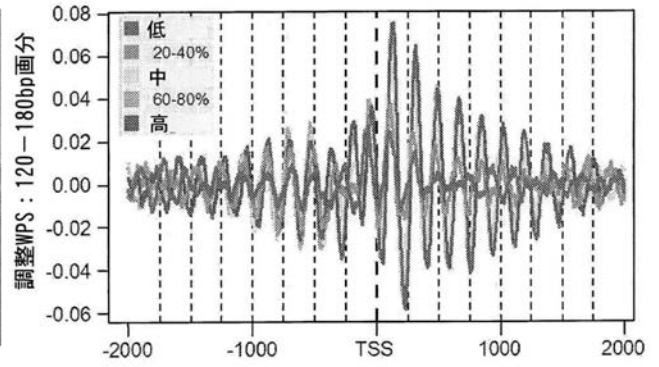


FIG. 57

【 図 5 8 】

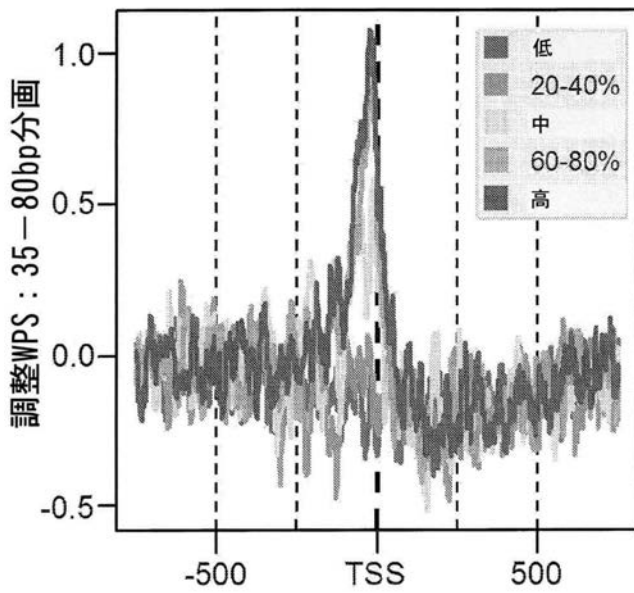


FIG. 58

【 図 5 9 】

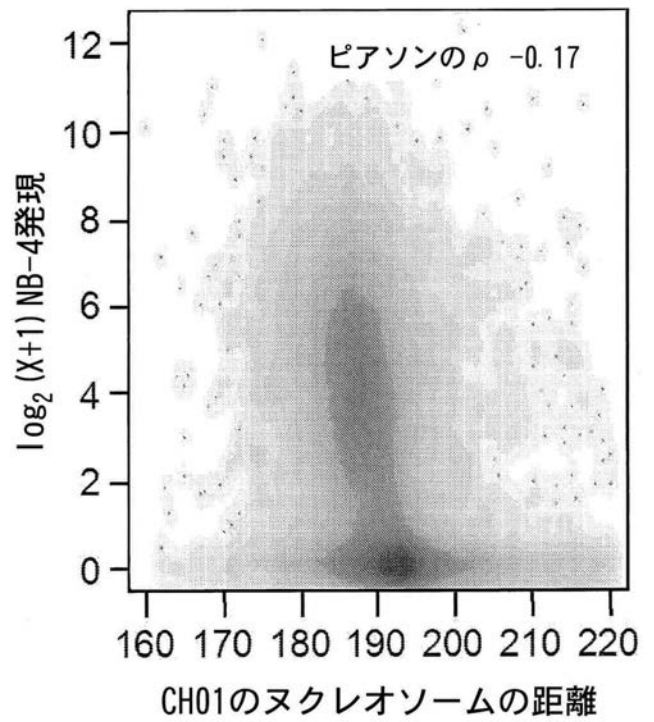


FIG. 59

【 図 6 0 】

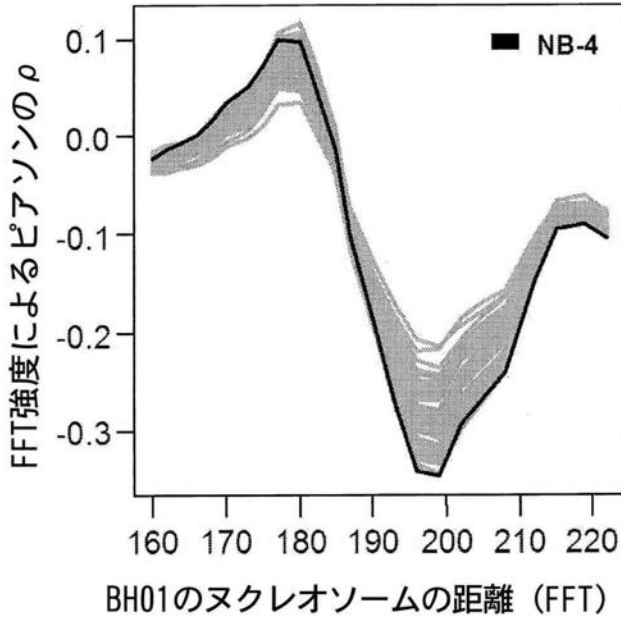


FIG. 60

【 図 6 1 】

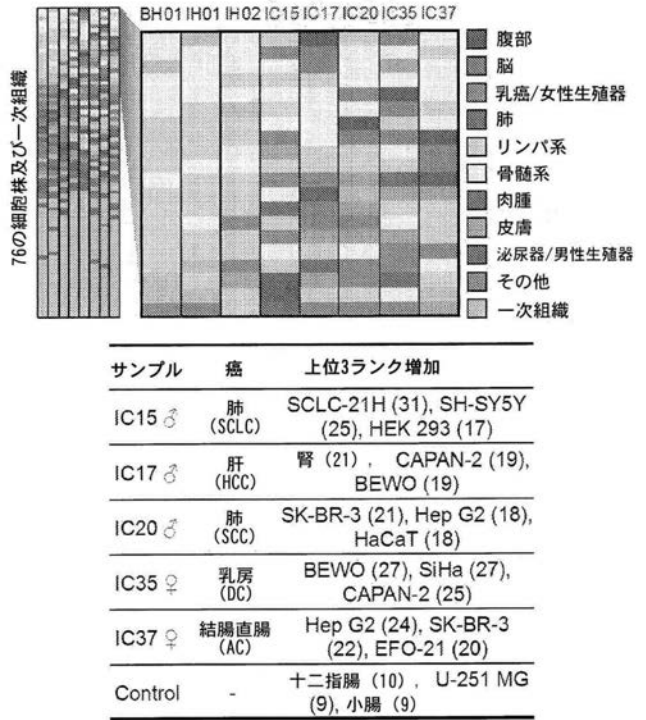


FIG. 61

【 図 6 2 A 】

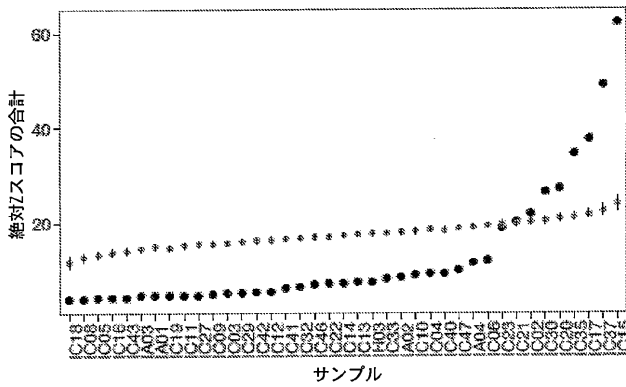


FIG. 62A

【 図 6 2 B 】

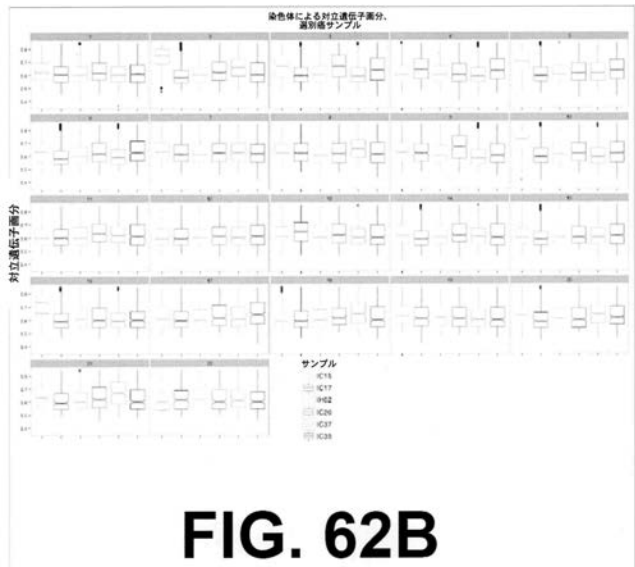
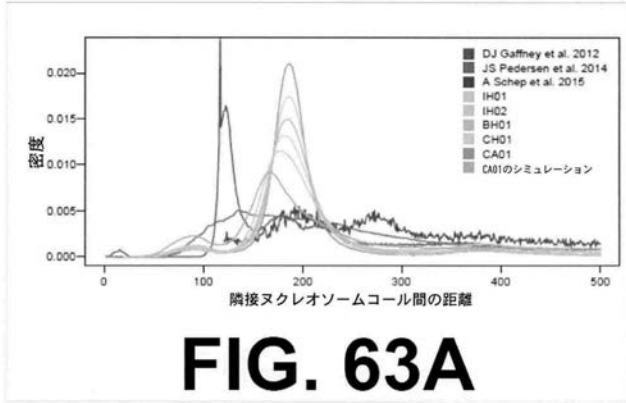
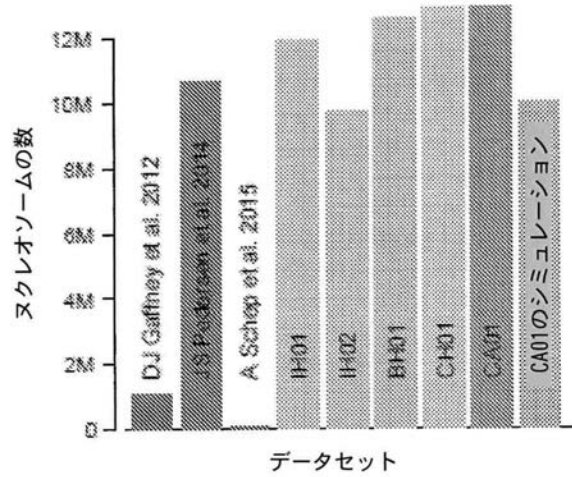


FIG. 62B

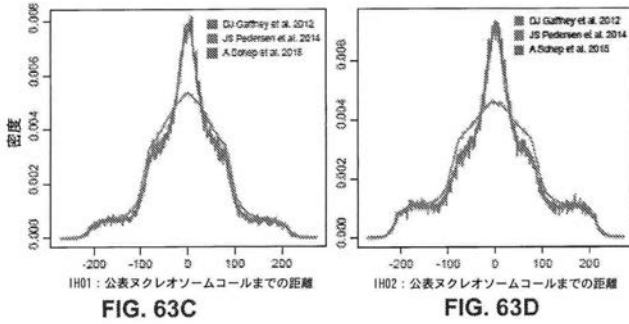
【 図 6 3 A 】



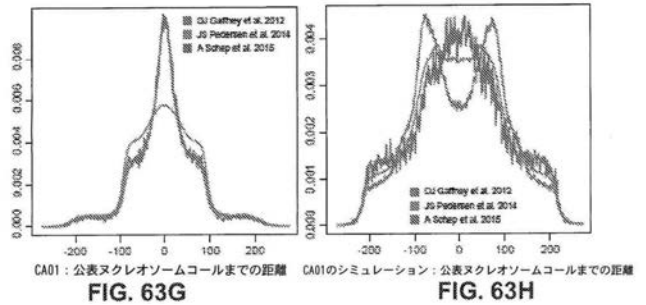
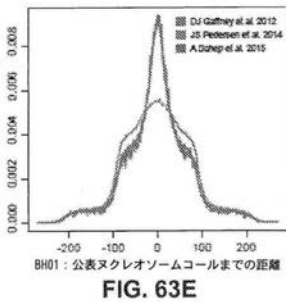
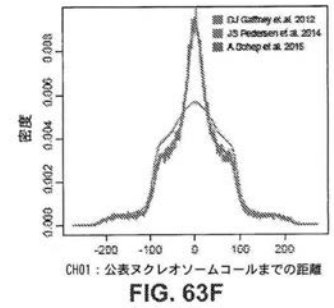
【 図 6 3 B 】



【 図 6 3 C - E 】



【 図 6 3 F - H 】



【配列表】

2017522908000001.app

【 国際調査報告 】

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US2015/042310

A. CLASSIFICATION OF SUBJECT MATTER IPC(8) - C12Q 1/68 (2015.01) CPC - C12Q 1/6809 (2015.10) According to International Patent Classification (IPC) or to both national classification and IPC		
B. FIELDS SEARCHED Minimum documentation searched (classification system followed by classification symbols) IPC(8) - C12Q 1/68; G01N 33/53; G06F 19/22 (2015.01) CPC - C12Q 1/6809, 1/6883; G06F 19/22, 19/24 (2015.10) Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched CPC - C12Q 1/6809, 1/6883; G06F 19/22, 19/24 (2015.10) (keyword delimited) US Classes - 435/6.1, 6.12, 7.1 Electronic data base consulted during the international search (name of data base and, where practicable, search terms used) PatBase, Google Patents, Google, PubMed Search terms used: cfDNA, cell free DNA, nucleosome, Fourier transform, distribution, mapping, locus, heatmap,		
C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X -- Y	WO 2013/177581 A2 (UNIVERSITY OF WASHINGTON THROUGH ITS CENTER FOR COMMERCIALIZATION et al) 28 November 2013 (28.11.2013) entire document	1-3, 48-50 4, 51
X -- Y	US 2013/0230858 A1 (SEQUENOM et al) 05 September 2013 (05.09.2013) entire document	92-97, 102, 103, 105-107 98
Y	KATOH et al. "MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform," Nucleic Acids Research, 15 July 2002 (15.07.2002), Vol. 30, No. 14, Pgs. 3059-3066. entire document	4, 51, 98
P, X	WO 2014/190286 A2 (SEQUENOM et al) 27 November 2014 (27.11.2014) entire document	1-4, 48-51, 92-97, 102, 103, 105-107
P, X	WO 2014/194113 A2 (CHRONIX BIOMEDICAL et al) 04 December 2014 (04.12.2014) entire document	1-3, 48-50, 92-97, 102, 103, 105-107
A	HANDSTAD et al. "A ChIP-Seq Benchmark Shows That Sequence Conservation Mainly Improves Detection of Strong Transcription Factor Binding Sites," PLoS One, 14 April 2011 (14.04.2011), Vol. 6, Iss. 4, Pgs. 1-9. entire document	1-4, 48-51, 92-107
<input type="checkbox"/> Further documents are listed in the continuation of Box C. <input type="checkbox"/> See patent family annex.		
* Special categories of cited documents: "A" document defining the general state of the art which is not considered to be of particular relevance "E" earlier application or patent but published on or after the international filing date "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) "O" document referring to an oral disclosure, use, exhibition or other means "P" document published prior to the international filing date but later than the priority date claimed "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art "&" document member of the same patent family		
Date of the actual completion of the international search 29 October 2015	Date of mailing of the international search report 12 JAN 2016	
Name and mailing address of the ISA/ Mail Stop PCT, Attn: ISA/US, Commissioner for Patents P.O. Box 1450, Alexandria, Virginia 22313-1450 Facsimile No. 571-273-8300	Authorized officer Blaine Copenheaver PCT Helpdesk: 571-272-4300 PCT OSP: 571-272-7774	

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

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US2015/042310

Box No. II Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. Claims Nos.: 5-47, 52-91, 108-111
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. As all searchable claims could be searched without effort justifying additional fees, this Authority did not invite payment of additional fees.
3. As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

- Remark on Protest**
- The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.
- The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.
- No protest accompanied the payment of additional search fees.

フロントページの続き

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

(74)代理人 100093300

弁理士 浅井 賢治

(74)代理人 100119013

弁理士 山崎 一夫

(74)代理人 100123777

弁理士 市川 さつき

(74)代理人 100111796

弁理士 服部 博信

(74)代理人 100111501

弁理士 滝澤 敏雄

(72)発明者 シェンデュア ジェイ

アメリカ合衆国 ワシントン州 98105-4721 シアトル ルーズヴェルト ウェイ ノースイースト 4545 スイート 400 ユニヴァーシティ オブ ワシントン内

(72)発明者 スナイダー マシュー

アメリカ合衆国 ワシントン州 98105-4721 シアトル ルーズヴェルト ウェイ ノースイースト 4545 スイート 400 ユニヴァーシティ オブ ワシントン内

(72)発明者 キルヒャー マーティン

アメリカ合衆国 ワシントン州 98105-4721 シアトル ルーズヴェルト ウェイ ノースイースト 4545 スイート 400 ユニヴァーシティ オブ ワシントン内

Fターム(参考) 4B063 QA01 QQ03 QQ34 QQ42 QR32 QR40 QR72 QS39 QX10

专利名称(译)	用于确定产生无细胞DNA的组织 and/或细胞类型的方法，以及使用其测定疾病或异常的方法		
公开(公告)号	JP2017522908A	公开(公告)日	2017-08-17
申请号	JP2017525327	申请日	2015-07-27
[标]申请(专利权)人(译)	华盛顿大学		
申请(专利权)人(译)	盐湖城华盛顿		
[标]发明人	シエンデュアジェイ スナイダーマシュー キルヒャーマーティン		
发明人	シエンデュア ジェイ スナイダー マシュー キルヒャー マーティン		
IPC分类号	C12Q1/68 G01N33/53 G16B45/00		
CPC分类号	C12Q1/6881 G16B20/00 G16B40/00 G16H50/20 C12Q1/6883 G16B45/00		
FI分类号	C12Q1/68.ZNA.A C12Q1/68.Z G01N33/53.M		
F-TERM分类号	4B063/QA01 4B063/QQ03 4B063/QQ34 4B063/QQ42 4B063/QR32 4B063/QR40 4B063/QR72 4B063/QS39 4B063/QX10		
代理人(译)	田中真一郎 ▲▼吉尔场和彦 山崎 一夫 服部博信		
优先权	62/029178 2014-07-25 US 62/087619 2014-12-04 US		
其他公开文献	JP2017522908A5		
外部链接	Espacenet		

摘要(译)

本公开提供了用于确定有助于来自受试者的生物样品中的无细胞DNA (“cfDNA”) 的一种或多种组织和/或细胞类型的方法。在一些实施方案中，本公开内容根据一种或多种组织和/或细胞类型的函数提供受试者的疾病或病症，所述一种或多种组织和/或细胞类型被确定为对来自该受试者的生物样品中的cfDNA有贡献。提供一种认证方式。

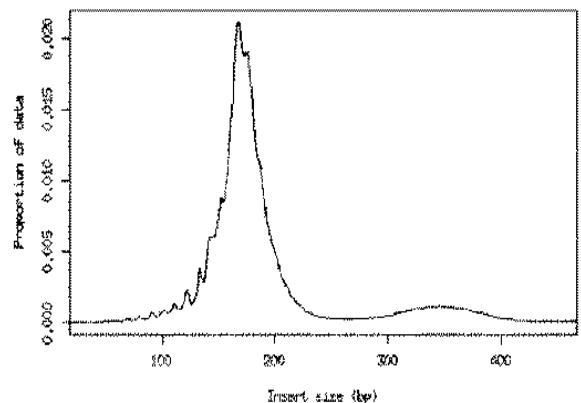


Figure 2