

【公報種別】特許法第17条の2の規定による補正の掲載

【部門区分】第6部門第3区分

【発行日】平成31年3月28日(2019.3.28)

【公表番号】特表2018-513445(P2018-513445A)

【公表日】平成30年5月24日(2018.5.24)

【年通号数】公開・登録公報2018-019

【出願番号】特願2017-541686(P2017-541686)

【国際特許分類】

G 1 6 B	30/00	(2019.01)
C 1 2 Q	1/68	(2018.01)
G 0 1 N	33/50	(2006.01)
A 6 1 K	45/00	(2006.01)
A 6 1 P	3/00	(2006.01)
A 6 1 P	43/00	(2006.01)

【F I】

G 0 6 F	19/22	
C 1 2 Q	1/68	Z
G 0 1 N	33/50	P
A 6 1 K	45/00	
A 6 1 P	3/00	
A 6 1 P	43/00	1 0 5

【手続補正書】

【提出日】平成31年2月7日(2019.2.7)

【手続補正1】

【補正対象書類名】特許請求の範囲

【補正対象項目名】全文

【補正方法】変更

【補正の内容】

【特許請求の範囲】

【請求項1】

一生体試料から取得される試験用核酸において起こる構造変異の尤度を決定する方法であって、

1つまたは複数のプロセッサと、前記1つまたは複数のプロセッサにより実行される1つまたは複数のプログラムを記憶するメモリとを有するコンピュータシステムにおいて、

(A) 前記試験用核酸が断片化される複数の配列決定反応から複数の配列リードを取得することであって、

前記複数の配列リード内の各配列リードは、前記試験用核酸のサブセットに対応する第1部分と、複数のバーコード内の前記各配列リードの各バーコードを符号化する第2部分とを備え、

各バーコードは、前記試験用核酸の前記配列データに依存せず、

前記複数の配列リードは、前記複数のバーコードを集合的に含む、取得することと、

(B) 複数のビンに関するビン情報を取得することであって、

前記複数のビン内の各ビンは、前記試験用核酸の異なる部分を表し、

前記ビン情報は、前記複数のビン内のビンごとに、前記複数の配列リードに含まれる複数の配列リードセット内の配列リードセットを特定し、

前記複数の配列リードセット内の各配列リードセットに含まれる各配列リードの前記各第1部分は、前記試験用核酸のサブセットに対応し、前記各配列リードセットに対応する前記ビンにより表される前記試験用核酸の前記異なる部分と少なくとも部分的にオーバー

ラップする、取得することと、

(C) 前記複数のビンの中から、前記試験用核酸の部分に対応し、かつオーバーラップしない第1ビン及び第2ビンを特定することであって、

前記第1ビンは前記複数の配列リード内の配列リード第1セットにより表され、前記第2ビンは前記複数の配列リード内の配列リード第2セットにより表される、特定することと、

(D) 前記第1セット及び前記第2セットに共通する前記バーコードの数が偶然に起因している数値的可能性または見込みを表す第1値を決定することと、

(E) 前記第1値が所定のカットオフ値を満たすという判定に応じて、前記第1ビン及び前記第2ビンに共通のバーコードごとに断片対を取得することにより、1つまたは複数の断片対を取得することであって、

前記1つまたは複数の断片対内の各断片対は、(i) 前記第1ビン及び前記第2ビンに共通の異なるバーコードに対応し、(ii) 異なる第1計算断片と異なる第2計算断片で構成され、

前記1つまたは複数の断片対内の断片対ごとに、

前記異なる第1計算断片は、前記各断片対に対応する前記バーコードを有する、前記複数の配列リード内の各配列リード第1サブセットで構成され、

前記各配列リード第1サブセット内の各配列リードは、前記各配列リード第1サブセット内の別の配列リードの定義済み遺伝距離内にあり、

前記各断片対の前記異なる第1計算断片は、前記第1ビン内の前記各断片対に対応する前記バーコードを有する第1配列リードに起因し、

前記各配列リード第1サブセット内の各配列リードは 前記第1ビンに由来し、

前記異なる第2計算断片は、前記各断片対に対応する前記バーコードを有する、前記複数の配列リード内の各配列リード第2サブセットで構成され、

前記各配列リード第2サブセット内の各配列リードは、前記各配列リード第2サブセット内の別の配列リードの定義済み遺伝距離内にあり、

前記各断片対の前記異なる第2計算断片は、前記第2ビン内の前記各断片対に対応する前記バーコードを有する第2配列リードに起因し、

前記各配列リード第2サブセット内の各配列リードは前記第2ビンに由来する、取得することと、

(F) 前記1つまたは複数の断片対に関して、第1モデルの発生確率及び第2モデルの発生確率に基づいて各尤度を計算することにより、前記試験用核酸における構造変異の尤度を提供することであって、

(i) 前記第1モデルは、前記1つまたは複数の断片対の前記各第1計算断片及び前記各第2計算断片が、前記対象核酸配列内に構造変異はないと観察され、かつ共通分子の一部であることを明示し、

(ii) 前記第2モデルは、前記1つまたは複数の断片対の前記各第1計算断片及び前記各第2計算断片が、前記対象核酸配列内に構造変異があると観察されることを明示する、提供することと

を含む前記方法。

【請求項2】

前記計算(F)において計算される前記尤度の表現は、前記第1モデルの前記発生確率と前記第2モデルの前記発生確率との比率スコアである、請求項1に記載の方法。

【請求項3】

前記第1ビン及び前記第2ビンは、前記試験用核酸上で少なくとも所定数のキロ塩基分、離れている、請求項1または2に記載の方法。

【請求項4】

前記第1ビン及び前記第2ビンは、前記試験用核酸上で少なくとも50キロ塩基分、離れている、請求項3に記載の方法。

【請求項5】

前記決定(D)は2項検定を用いて前記第1値を計算する、請求項1～4のいずれか1項に記載の方法。

【請求項6】

前記決定(D)は2項検定を用いて、以下の形式の前記第1値を計算し

$$p = 1 - P_{Binom}(n; \frac{n_1 n_2}{B})$$

pはp値と表される前記第1値であり、

nは前記配列リード第1及び第2セットの両方に存在する前記一意的バーコードの数であり、

n_1 は、前記配列リード第1セット内の前記一意的バーコードの数であり、

n_2 は、前記配列リード第2セット内の前記一意的バーコードの数であり、

Bは前記複数のビンにわたる前記一意的バーコードの総数である、

請求項1～4のいずれか1項に記載の方法。

【請求項7】

前記一生体試料はヒトであり、前記試験用核酸は前記生体試料の前記ゲノムであり、前記第1値が 10^{-14} 以下である時に前記第1値は前記所定のカットオフ値を満たす、請求項1～6のいずれか1項に記載の方法。

【請求項8】

前記一生体試料はヒトであり、前記試験用核酸は前記生体試料の前記ゲノムであり、前記第1値が 10^{-15} 以下である時に前記第1値は前記所定のカットオフ値を満たす、請求項1～6のいずれか1項に記載の方法。

【請求項9】

前記構造変異は、前記配列リード第1セットにより表される前記試験用核酸の前記異なる部分に対する、50連続塩基以上の挿入または欠失である、請求項1～8のいずれか1項に記載の方法。

【請求項10】

前記構造変異は、前記配列リード第1セットにより表される前記試験用核酸の前記異なる部分に対する、500連続塩基以上の挿入または欠失である、請求項1～8のいずれか1項に記載の方法。

【請求項11】

前記構造変異は、前記配列リード第1セットにより表される前記試験用核酸の前記異なる部分に対する、5000連続塩基以上の挿入または欠失である、請求項1～8のいずれか1項に記載の方法。

【請求項12】

前記構造変異は、遺伝子疾患に関連する、請求項1～11のいずれか1項に記載の方法。

【請求項13】

前記複数のビン内の各ビンは、前記試験用核酸の少なくとも20キロ塩基、前記試験用核酸の少なくとも50キロ塩基、前記試験用核酸の少なくとも100キロ塩基、前記試験用核酸の少なくとも250キロ塩基、または前記試験用核酸の少なくとも500キロ塩基を表す、請求項1～12のいずれか1項に記載の方法。

【請求項14】

前記複数の配列リード内の各配列リードセットに含まれる各配列リードは、前記試験用核酸のサブセットに対応する各第1部分を有し、前記各第1部分は、前記各配列リードセットに対応する前記ビンにより表される前記試験用核酸の前記異なる部分と完全にオーバーラップする、請求項1～12のいずれか1項に記載の方法。

【請求項15】

前記複数のビンは、10,000個以上のビンを含む、請求項1～14のいずれか1項に記載の方法。

【請求項 1 6】

前記複数のビンは、100,000個以上のビンを含む、請求項1～14のいずれか1項に記載の方法。

【請求項 1 7】

前記複数のビンは、1,000,000個以上のビンを含む、請求項1～14のいずれか1項に記載の方法。

【請求項 1 8】

前記生体試料は複数染色体生物種に由来し、前記試験用核酸は、前記複数染色体生物種に含まれる複数の染色体を集合的に表す複数の核酸を含む、請求項1～17のいずれか1項に記載の方法。

【請求項 1 9】

前記複数の配列リード内の各配列リードの前記第2部分における前記バーコードは、前記セット{1, . . . , 1024}から選択される、前記セット{1, . . . , 4096}から選択される、前記セット{1, . . . , 16384}から選択される、前記セット{1, . . . , 65536}から選択される、前記セット{1, . . . , 262144}から選択される、前記セット{1, . . . , 1048576}から選択される、前記セット{1, . . . , 4194304}から選択される、前記セット{1, . . . , 16777216}から選択される、前記セット{1, . . . , 67108864}から選択される、または前記セット{1, . . . , 1×10¹²}から選択される一意的な所定値を符号化する、請求項1～18のいずれか1項に記載の方法。

【請求項 2 0】

前記複数の配列リード内の各配列リードの前記第2部分における前記バーコードは、前記各配列リード内のオリゴヌクレオチド連続セットに局所化される、請求項1～19のいずれか1項に記載の方法。

【請求項 2 1】

前記オリゴヌクレオチド連続セットはN塩基長であり、Nは前記セット{4, . . . , 20}から選択される整数である、請求項2 0に記載の方法。

【請求項 2 2】

前記複数の配列リード内の配列リードの前記第2部分における前記バーコードは、前記配列リード内の所定のヌクレオチド非連続セットに局所化される、請求項1～10のいずれか1項に記載の方法。

【請求項 2 3】

前記所定のヌクレオチド非連続セットは合わせてN個のヌクレオチドで構成され、Nは前記セット{4, . . . , 20}内の整数である、請求項2 2に記載の方法。

【請求項 2 4】

前記第1配列リードは、10キロ塩基より大きい前記試験用核酸の第1サブセットに対応する、請求項1～2 3のいずれか1項に記載の方法。

【請求項 2 5】

前記第1配列リードは、20キロ塩基より大きい前記試験用核酸の第1サブセットに対応する、請求項1～2 3のいずれか1項に記載の方法。

【請求項 2 6】

前記構造変異が起こったとみなされると、前記構造変異に応じる治療方式で前記生体試料の由来する対象者を治療することをさらに含む請求項1～2 5のいずれか1項に記載の方法。

【請求項 2 7】

前記治療方式は、食事改善を含む、請求項2 6に記載の方法。

【請求項 2 8】

前記治療方式は、前記構造変異に関連する生物学的経路を抑制または増強する医薬組成物の適用を含む、請求項2 7に記載の方法。

【請求項 2 9】

前記第1及び第2ピンの識別は、疎行列乗算を使用する前記特定(C)により決定される、請求項1～28のいずれか1項に記載の方法。

【請求項30】

前記第1及び第2ピンの識別は、以下の形式の疎行列乗算を使用する前記特定(C)により決定される、

$$V = A_1^T A_2$$

A_1 は前記第1ピンを含む第1 $B \times N_1$ バーコード行列であり、

A_2 は前記第2ピンを含む第2 $B \times N_2$ バーコード行列であり、

B は前記複数のピンにわたる前記一意的バーコードの数であり、

N_1 は A_1 内の前記ピンの数であり、

N_2 は A_2 内の前記ピンの数であり、

$$A_1^T$$

は前記行列 A_1 の転置行列である、

請求項1～28のいずれか1項に記載の方法。

【請求項31】

前記第1ピンは前記生体試料の第1染色体に対応付けられ、

前記第2ピンは前記生体試料の第2染色体に対応付けられ、

N_1 は前記第1染色体に対応付けられた前記ピンの数であり、

N_2 は前記第2染色体に対応付けられた前記ピンの数である、

請求項30に記載の方法。

【請求項32】

前記第1及び第2ピンは両方とも、前記生体試料の第1染色体に対応付けられ、

N_1 は前記第1染色体に対応付けられた前記ピンの数であり、

N_2 は N_1 と等しい、

請求項30に記載の方法。

【請求項33】

前記試験用核酸の複数のブラックアウト領域を含むブラックアウトリストを保持することをさらに含む前記方法であって、前記特定(C)の前に、

配列リードの前記第1部分が前記複数のブラックアウト領域内のブラックアウト領域とオーバーラップする場合、前記複数の配列リードから、前記配列リードを除くことをさらに含む請求項1～32のいずれか1項に記載の方法。

【請求項34】

前記計算(F)において前記計算尤度は、以下のように計算され

$$LR = \frac{P(\text{observed fragments} | SV)}{P(\text{observed fragments} | \text{no } SV)}$$

LRは複数の項の積に等しく、前記複数の項内の各項は(i)前記1つまたは複数の断片対内の各断片対を表し、(ii)以下の形式を有し

$$= \frac{P(r_1, r_2, l_1, l_2, d | SV; a_b)}{P(r_1, r_2, l_1, l_2, d | \text{no } SV; a_b)}$$

r_1 は、前記各断片対の前記第1計算断片における前記各配列リード第1サブセット内の配列リードの数であり、

l_1 は、前記各断片対の前記配列リード第1サブセットにより決定される前記第1計算断片の長さであり、

r_2 は、前記各断片対の前記第2計算断片における前記各配列リード第2サブセット内のリードの数であり、

l_2 は、前記各断片対の前記配列リード第2サブセットにより決定される前記第2計算断片の長さであり、

d は前記試験用核酸内の前記各断片対の前記第1計算断片と前記第2計算断片との間の距離であり、

a_b は前記複数の配列リードにわたる前記第1バーコードのリードレートであり、

SV は前記第1モデルに従って前記第1計算断片及び前記第2計算断片が観察されたことを示し、

$no\ SV$ は前記第2モデルに従って前記第1計算断片及び前記第2計算断片が観察されたことを示す、

請求項1に記載の方法。

【請求項3-5】

$$P(r_1, r_2, l_1, l_2, d | no\ SV; a_b) = \\ P(r_1, r_2, l_1, l_2, d | SM, no\ SV; a_b)P(SM | no\ SV) + \\ P(r_1, r_2, l_1, l_2, d | DM, no\ SV; a_b)P(DM | no\ SV),$$

であり、

SM は、前記複数の配列決定反応において前記第1計算分子及び前記第2計算分子が前記試験用核酸の同一断片に由来しているという前記仮定であり、

DM は、前記複数の配列決定反応において前記第1計算分子及び前記第2計算分子が前記試験用核酸の異なる断片に由来しているという前記仮定であり、

$$P(r_1, r_2, l_1, l_2, d | DM, no\ SV; a_b) = P_{frag}(r_1, l_1; a_b)P_{frag}(r_2, l_2; a_b)$$

であり、

$$P_{frag}(r_1, l_1; a_b)$$

は、長さが不明の第1分子から、観察された長さ l_1 に及ぶ r_1 個のリードを観察する前記確率であり、

$$P_{frag}(r_2, l_2; a_b)$$

は、長さが不明の第2分子から、観察された長さ l_2 に及ぶ r_2 個のリードを観察する前記確率である、

請求項3-4に記載の方法。

【請求項3-6】

$$P_{frag}(r_1, l_1; a_b)$$

及び

$$P_{frag}(r_2, l_2; a_b)$$

はそれぞれ以下のように計算され

$$\begin{aligned}
 & \sum_{m:m \geq l} \left(r(r-1) \left(\frac{l}{m} \right)^{r-2} \frac{m-l}{m^2} \right) P_p(r; ma_b) P_L(m) \\
 = & \sum_{m:m \geq l} (m-l) P_p(r-2; a_b l) P_p(0; a_b(m-l)) a_b^2 P_L(m) \\
 & P_p(r; b)
 \end{aligned}$$

はパラメータ b によるポアソン分布の前記確率質量関数であり、
 $P_L(m)$

は前記各分子の前記実分子長が m である前記（予測）確率である、

請求項 3_5 に記載の方法。

【請求項 3_7】

$$P(r_1, r_2, l_1, l_2, d | SM, \text{no } SV; a_b)$$

は、以下のように計算され

$$\begin{aligned}
 & \sum_{m:m \geq l_1+l_2+d} (m-l_1-l_2-d) P_p(r_1-2; a_b l_1) P_p(r_2 \\
 & \quad - 2; a_b l_2) P_p(0; a_b(m-l_1-l_2)) a_b^4 P_L(m)
 \end{aligned}$$

m は前記実分子長の前記長さであり、
 $P_p(r_1-2; a_b l_1)$

は r_1 に対するパラメータ b によるポアソン分布の確率質量関数であり、
 $P_p(r_2-2; a_b l_2)$

は r_2 に対するパラメータ b によるポアソン分布の確率質量関数であり、
 $P_p(0; a_b(m-l_1-l_2))$

はパラメータ b によるポアソン分布の確率質量関数であり、
 $P_L(m)$

は前記実の共通分子長が m である予測確率である、

請求項 3_5 に記載の方法。

【請求項 3_8】

$$P(r_1, r_2, l_1, l_2, d | SV; a_b) =$$

$$\begin{aligned}
 & P(r_1, r_2, l_1, l_2, 2d' | SM, \text{no } SV; a_b) P(SM | \text{no } SV) + \\
 & P(r_1, r_2, l_1, l_2, 2d' | DM, \text{no } SV; a_b) P(DM | \text{no } SV),
 \end{aligned}$$

であり、

S M は、前記複数の配列決定反応において前記第 1 計算分子及び前記第 2 計算分子が前記試験用核酸の同一断片に由来しているという前記仮定であり、

D M は、前記複数の配列決定反応において前記第 1 計算分子及び前記第 2 計算分子が前記試験用核酸の異なる断片に由来しているという前記仮定であり、

$$P(r_1, r_2, l_1, l_2, 2d' | DM, SV; a_b) = P_{frag}(r_1, l_1; a_b) P_{frag}(r_2, l_2; a_b)$$

であり、
 $P_{frag}(r_1, l_1; a_b)$

は、長さが不明の第1分子から、観察された長さ l_1 に及ぶ r_1 個のリードを観察する前記確率であり、

$$P_{frag}(r_2, l_2; a_b)$$

は、長さが不明の第2分子から、観察された長さ l_2 に及ぶ r_2 個のリードを観察する前記確率であり、

2 d' は、前記第1計算分子及び前記第2計算分子に対応付けられた構造変異の前記ブレークポイントの推定を考慮した、前記試験用核酸内の前記各断片対の前記第1計算断片と前記第2計算断片との間の距離である、

請求項3-4に記載の方法。

【請求項3-9】
 $P_{frag}(r_1, l_1; a_b)$

及び
 $P_{frag}(r_2, l_2; a_b)$

はそれぞれ以下のように計算され

$$\begin{aligned} & \sum_{m:m \geq l} \left(r(r-1) \left(\frac{l}{m} \right)^{r-2} \frac{m-l}{m^2} \right) P_p(r; ma_b) P_L(m) \\ &= \sum_{m:m \geq l} (m-l) P_p(r-2; a_b l) P_p(0; a_b(m-l)) a_b^2 P_L(m) \\ & P_p(r; b) \end{aligned}$$

はパラメータ b によるポアソン分布の前記確率質量関数であり、

$$P_L(m)$$

は前記各分子の前記実分子長が m である前記（予測）確率である、

請求項3-8に記載の方法。

【請求項4-0】
 $P(r_1, r_2, l_1, l_2, 2d' | SM, SV; a_b)$

は、以下のように計算され

$$\begin{aligned} & \sum_{m:m \geq l_1 + l_2 + d} (m - l_1 - l_2 - 2d') P_p(r_1 - 2; a_b l_1) P_p(r_2 \\ & - 2; a_b l_2) P_p(0; a_b(m - l_1 - l_2)) a_b^4 P_L(m) \end{aligned}$$

m は前記実分子長の前記長さであり、
 $P_p(r_1 - 2; a_b l_1)$

は r_1 に対するパラメータ b によるポアソン分布の確率質量関数であり、
 $P_p(r_2 - 2; a_b l_2)$

は r_2 に対するパラメータ b によるポアソン分布の確率質量関数であり、
 $P_p(0; a_b(m - l_1 - l_2))$

はパラメータ b によるポアソン分布の確率質量関数であり、
 $P_L(m)$

は前記実の共通分子長が m である予測確率である、
 請求項 3 8 に記載の方法。

【請求項 4 1】

d' は、

$$P_p(0; a_b d') \geq 0.75$$

となる前記最大限 d' を計算することにより推定される、請求項 3 8 に記載の方法。

【請求項 4 2】

前記複数の配列リードは、全ゲノム配列データを表す、請求項 1 ~ 4 1 のいずれか 1 項に記載の方法。

【請求項 4 3】

前記複数の配列リードは、ゲノムのサブセットの標的配列を表し、

前記複数の配列リードの第 1 サブセットは、前記ゲノムの前記サブセット内に由来し、
 a_b

の第 1 リードレートを有し、

前記複数の配列リードの第 2 サブセットは、前記ゲノムの前記サブセット外に由来し、
 \tilde{a}_b

の第 1 リードレートを有し、

\tilde{a}_b

は

a_b

とは異なり、前記計算 (F) の前記尤度により前記複数の配列リードの前記それぞれの第 1 及び第 2 サブセットの前記異なるリードレートは補正される、

請求項 1 ~ 4 1 のいずれか 1 項に記載の方法。

【請求項 4 4】

複数の第 1 及び第 2 ビンに関して前記 (D) ~ (F) が計算されることにより、前記試験用核酸内の 1 つまたは複数の構造変異がコールされ、前記方法はさらに、前記複数の配列リードと前記 1 つまたは複数の構造変異のコールを使用して、前記試験用核酸内のブレークポイント位置を精製することを含む、請求項 1 ~ 4 3 のいずれか 1 項に記載の方法。

【請求項 4 5】

前記複数のバーコードは、1 0 0 0 個以上のバーコードを含む、請求項 1 ~ 4 4 のいず

れか 1 項に記載の方法。

【請求項 4 6】

前記複数のバーコードは、10,000 個以上のバーコードを含む、請求項 1 ~ 4 4 のいずれか 1 項に記載の方法。

【請求項 4 7】

前記複数のバーコードは、100,000 個以上のバーコードを含む、請求項 1 ~ 4 4 のいずれか 1 項に記載の方法。

【請求項 4 8】

前記複数のバーコードは、 1×10^6 個以上のバーコードを含む、請求項 1 ~ 4 4 のいずれか 1 項に記載の方法。

【請求項 4 9】

1 つまたは複数のプロセッサと、

前記 1 つまたは複数のプロセッサにより実行される 1 つまたは複数のプログラムを記憶するメモリと

を備えるコンピューティングシステムであって、

前記 1 つまたは複数のプログラムは、

(A) 前記試験用核酸が断片化される複数の配列決定反応から複数の配列リードを取得する命令であって、

前記複数の配列リード内の各配列リードは、前記試験用核酸のサブセットに対応する第 1 部分と、複数のバーコード内の前記各配列リードの各バーコードを符号化する第 2 部分とを備え、

各バーコードは、前記試験用核酸の前記配列データに依存せず、

前記複数の配列リードは、前記複数のバーコードを集合的に含む、命令と、

(B) 複数のビンに関するビン情報を取得する命令であって、

前記複数のビン内の各ビンは、前記試験用核酸の異なる部分を表し、

前記ビン情報は、前記複数のビン内のビンごとに、前記複数の配列リードに含まれる複数の配列リードセット内の配列リードセットを特定し、

前記複数の配列リードセット内の各配列リードセットに含まれる各配列リードの前記各第 1 部分は、前記試験用核酸のサブセットに対応し、前記各配列リードセットに対応する前記ビンにより表される前記試験用核酸の前記異なる部分と少なくとも部分的にオーバーラップする、命令と、

(C) 前記複数のビンの中から、前記試験用核酸の部分に対応し、かつオーバーラップしない第 1 ビン及び第 2 ビンを特定する命令であって、

前記第 1 ビンは前記複数の配列リード内の配列リード第 1 セットにより表され、前記第 2 ビンは前記複数の配列リード内の配列リード第 2 セットにより表される、命令と、

(D) 前記第 1 セット及び前記第 2 セットに共通する前記バーコードの数が偶然に起因している数値的可能性または見込みを表す第 1 値を決定する命令と、

(E) 前記第 1 値が所定のカットオフ値を満たすという判定に応じて、前記第 1 ビン及び前記第 2 ビンに共通のバーコードごとに断片対を取得することにより、1 つまたは複数の断片対を取得する命令であって、

前記 1 つまたは複数の断片対内の各断片対は、(i) 前記第 1 ビン及び前記第 2 ビンに共通の異なるバーコードに対応し、(i i) 異なる第 1 計算断片と異なる第 2 計算断片で構成され、

前記 1 つまたは複数の断片対内の断片対ごとに、

前記異なる第 1 計算断片は、前記各断片対に対応する前記バーコードを有する、前記複数の配列リード内の各配列リード第 1 サブセットで構成され、

前記各配列リード第 1 サブセット内の各配列リードは、前記各配列リード第 1 サブセット内の別の配列リードの定義済み遺伝距離内にあり、

前記各断片対の前記異なる第 1 計算断片は、前記第 1 ビン内の前記各断片対に対応する前記バーコードを有する第 1 配列リードに起因し、

前記各配列リード第1サブセット内の各配列リードは、前記第1ピンに由来し、前記異なる第2計算断片は、前記各断片対に対応する前記バーコードを有する、前記複数の配列リード内の各配列リード第2サブセットで構成され、

前記各配列リード第2サブセット内の各配列リードは、前記各配列リード第2サブセット内の別の配列リードの定義済み遺伝距離内にあり、

前記各断片対の前記異なる第2計算断片は、前記第2ピン内の前記各断片対に対応する前記バーコードを有する第2配列リードに起因し、

前記各配列リード第2サブセット内の各配列リードは前記第2ピンに由来する、命令と、

(F) 前記1つまたは複数の断片対に関して、第1モデルの発生確率及び第2モデルの発生確率に基づいて各尤度を計算することにより、前記試験用核酸における構造変異の尤度を提供する命令であって、

(i) 前記第1モデルは、前記1つまたは複数の断片対の前記各第1計算断片及び前記各第2計算断片が、前記対象核酸配列内に構造変異はないと観察され、かつ共通分子の一部であることを明示し、

(ii) 前記第2モデルは、前記1つまたは複数の断片対の前記各第1計算断片及び前記各第2計算断片が、前記対象核酸配列内に構造変異があると観察されることを明示する、命令と

を含む前記コンピューティングシステム。

【請求項50】

前記複数のバーコードは、1000個以上のバーコード、10,000個以上のバーコード、100,000個以上のバーコード、または1×10⁶個以上のバーコードを含む、請求項49に記載のコンピューティングシステム。

【請求項51】

コンピュータにより実行されるように構成される1つまたは複数のプログラムを記憶する非一時的コンピュータ可読記憶媒体であって、前記1つまたは複数のプログラムは、

(A) 前記試験用核酸が断片化される複数の配列決定反応から複数の配列リードを取得する命令であって、

前記複数の配列リード内の各配列リードは、前記試験用核酸のサブセットに対応する第1部分と、複数のバーコード内の前記各配列リードの各バーコードを符号化する第2部分とを備え、

各バーコードは、前記試験用核酸の前記配列データに依存せず、

前記複数の配列リードは、前記複数のバーコードを集合的に含む、命令と、

(B) 複数のピンに関するピン情報を取得する命令であって、

前記複数のピン内の各ピンは、前記試験用核酸の異なる部分を表し、

前記ピン情報は、前記複数のピン内のピンごとに、前記複数の配列リードに含まれる複数の配列リードセット内の配列リードセットを特定し、

前記複数の配列リードセット内の各配列リードセットに含まれる各配列リードの前記各第1部分は、前記試験用核酸のサブセットに対応し、前記各配列リードセットに対応する前記ピンにより表される前記試験用核酸の前記異なる部分と少なくとも部分的にオーバーラップする、命令と、

(C) 前記複数のピンの中から、前記試験用核酸の部分に対応し、かつオーバーラップしない第1ピン及び第2ピンを特定する命令であって、

前記第1ピンは前記複数の配列リード内の配列リード第1セットにより表され、前記第2ピンは前記複数の配列リード内の配列リード第2セットにより表される、命令と、

(D) 前記第1セット及び前記第2セットに共通する前記バーコードの数が偶然に起している数値的可能性または見込みを表す第1値を決定する命令と、

(E) 前記第1値が所定のカットオフ値を満たすという判定に応じて、前記第1ピン及び前記第2ピンに共通のバーコードごとに断片対を取得することにより、1つまたは複数の断片対を取得する命令であって、

前記 1 つまたは複数の断片対内の各断片対は、(i) 前記第 1 ピン及び前記第 2 ピンに共通の異なるバーコードに対応し、(i i) 異なる第 1 計算断片と異なる第 2 計算断片とで構成され、

前記 1 つまたは複数の断片対内の断片対ごとに、

前記異なる第 1 計算断片は、前記各断片対に対応する前記バーコードを有する、前記複数の配列リード内の各配列リード第 1 サブセットで構成され、

前記各配列リード第 1 サブセット内の各配列リードは、前記各配列リード第 1 サブセット内の別の配列リードの定義済み遺伝距離内にあり、

前記各断片対の前記異なる第 1 計算断片は、前記第 1 ピン内の前記各断片対に対応する前記バーコードを有する第 1 配列リードに起因し、

前記各配列リード第 1 サブセット内の各配列リードは 前記第 1 ピンに由来し、

前記異なる第 2 計算断片は、前記各断片対に対応する前記バーコードを有する、前記複数の配列リード内の各配列リード第 2 サブセットで構成され、

前記各配列リード第 2 サブセット内の各配列リードは、前記各配列リード第 2 サブセット内の別の配列リードの定義済み遺伝距離内にあり、

前記各断片対の前記異なる第 2 計算断片は、前記第 2 ピン内の前記各断片対に対応する前記バーコードを有する第 2 配列リードに起因し、

前記各配列リード第 2 サブセット内の各配列リードは前記第 2 ピンに由来する、命令と、

(F) 前記 1 つまたは複数の断片対に関して、第 1 モデルの発生確率及び第 2 モデルの発生確率に基づいて各尤度を計算することにより、前記試験用核酸における構造変異の尤度を提供する命令であって、

(i) 前記第 1 モデルは、前記 1 つまたは複数の断片対の前記各第 1 計算断片及び前記各第 2 計算断片が、前記対象核酸配列内に構造変異はないと観察され、かつ共通分子の一部であることを明示し、

(i i) 前記第 2 モデルは、前記 1 つまたは複数の断片対の前記各第 1 計算断片及び前記各第 2 計算断片が、前記対象核酸配列内に構造変異があると観察されることを明示する、命令と、

を含む、前記非一時的コンピュータ可読記憶媒体。