

【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30年6月21日(2018.6.21)

【公表番号】特表2017-520821(P2017-520821A)

【公表日】平成29年7月27日(2017.7.27)

【年通号数】公開・登録公報2017-028

【出願番号】特願2016-565058(P2016-565058)

【国際特許分類】

G 0 6 F 19/22 (2011.01)

C 1 2 Q 1/68 (2018.01)

【F I】

G 0 6 F 19/22

C 1 2 Q 1/68 Z

【手続補正書】

【提出日】平成30年5月11日(2018.5.11)

【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】

第一の試料中の標的領域における低頻度バリエーションを検出するための、コンピュータで実行される方法であって、コンピュータシステムにおいて、

- 1 又は 2 以上の試料からの DNA 断片のシーケンシングから得られる複数の配列リードを受け取り、ここで前記 1 又は 2 以上の試料は第一の試料を含み、前記シーケンシングは前記 DNA 断片中の標的領域を標的化することを含み；
- 前記複数の配列リードを、参照配列の標的領域にアラインメントし；
- 第一の試料の配列リードに基づいて、標的領域の第一の位置において第一のバリエーションアレルを有する第一の候補バリエーションを同定し、ここで前記第一のバリエーションアレルは、前記参照配列の第一の位置の参照アレルとは異なり；
- 前記参照配列の第一の位置にアラインメントする第一の試料の配列リードに基づいて、第一の位置における第一のバリエーションアレルに関する第一のバリエーション頻度を決定し、
- 前記参照アレルを有する前記参照配列の標的領域中の 1 セットの第二の位置を同定し、ここで前記 1 セットの第二の位置は前記第一の位置を含み；

前記 1 セットの第二の位置の各々の位置において、かつ前記 1 又は 2 以上の試料の各々に関して：

- 前記参照配列の 1 セットの第二の位置の各々の位置にアラインメントする試料の配列リードに基づいて、第一のバリエーションアレルの第二のバリエーション頻度を決定し、ここで前記第二のバリエーション頻度は統計分布を形成し；
- 前記第一のバリエーション頻度を前記統計分布の統計値と比較して、前記統計分布の統計値に対する第一のバリエーション頻度の確率値を決定し；そして
- 前記第一の位置における第一のバリエーションアレルに関して、第一の試料において第一の候補バリエーションが真陽性であるか否かを決定する一部として、前記確率値を閾値と比較すること、ここで前記閾値は前記第一のバリエーションアレルに関して偽陽性と真陽性とを区別する、

含む、方法。

【請求項 2】

前記参照配列は、正常細胞から決定されるとおりのコンセンサス配列に対応する、請求項 1 に記載の方法。

【請求項 3】

前記 1 又は 2 以上の試料は無細胞 DNA 断片由来である、請求項 1 又は 2 に記載の方法。

【請求項 4】

前記 1 又は 2 以上の試料は生物試料の RNA 由来である、請求項 1 又は 2 に記載の方法。

【請求項 5】

複数の試料は、単一のシーケンシングランにおいてシーケンシングされる、請求項 1 ~ 4 のいずれか 1 項に記載の方法。

【請求項 6】

前記確率値は、 z スコア、修正された z スコア、累積確率、フレッド (Phred) クオリティスコア又は修正されたフレッドクオリティスコアである、請求項 1 ~ 5 のいずれか 1 項に記載の方法。

【請求項 7】

前記統計分布は、前記第二のバリエーション頻度の対数変換の統計分布である、請求項 1 ~ 6 のいずれか 1 項に記載の方法。

【請求項 8】

前記閾値は、既知の真陽性と偽陽性とを有するトレーニングデータに基づいて、サポートベクターマシン分類器を用いて決定される、請求項 1 ~ 7 のいずれか 1 項に記載の方法。

【請求項 9】

第一の試料中の標的領域における第一の位置に第一のバリエーションアレルを有するバリエーションを検出するための、コンピュータで実行される方法であって、コンピュータシステムにおいて、

- 少なくとも 2 つの試料からの DNA 断片のシーケンシングから得られる複数の配列リードを受け取り、ここで前記少なくとも 2 つの試料は第一の試料を含み、前記シーケンシングは前記 DNA 断片中の標的領域を標的化することを含み；
- 前記複数の配列リードを、参照配列の標的領域にアラインメントし；
- 第一の位置において各試料のアラインメントされた配列リードに基づいて、前記第一のバリエーションアレルが、前記少なくとも 2 つの試料の各々の試料中の第一の位置に存在するか否かを同定し、ここで前記第一のバリエーションアレルは、前記参照配列の第一の位置における参照アレルとは異なり；
- 前記少なくとも 2 つの試料の各試料に関する第一の位置にアラインメントされた配列リードを用いて、前記第一の位置における第一のバリエーションアレルのバリエーション数及び前記第一の位置における参照アレルの野生型数を決定し；
- 前記少なくとも 2 つの試料から、1 つ又は 2 つ以上の試料を 1 つ又は 2 つ以上の参照試料として選択し；
- 第一の試料に関する第一の位置における第一のバリエーションアレルの第一のバリエーション数及び第一の位置における参照アレルの第一の野生型数と、前記 1 つ又は 2 つ以上の参照試料に関する第一の位置における第一のバリエーションアレルの第二のバリエーション数及び第一の位置における参照アレルの第二の野生型数とを比較して、第一の試料に関する第一の位置に第一のバリエーションアレルを有するバリエーションが発生する確率値を決定し；そして
- 第一のバリエーションアレルに関して第一の試料中の第一の位置における第一のバリエーションアレルが真陽性を含むか否かを決定する一部として、前記確率値を閾値と比較すること、ここで前記閾値は第一の位置の第一のバリエーションアレルに関して偽陽性と真陽性とを区別する、

含む、方法。

【請求項 10】

前記確率値は、以下の式：

【数 1】

$$\chi^2 = n \times (a_1 \times w_2 - a_2 \times w_1)^2 / (n_1 \times n_2 \times a \times w)$$

[式中、

a_2 は第一のバリエーション数であり、 w_2 は第一の野生型数であり、 a_1 は第二のバリエーション数であり、 w_1 は第二の野生型数であり、 a は a_1 及び a_2 の合計であり、 w は w_1 及び w_2 の合計であり、 n_1 は a_1 及び w_1 の合計であり、 n_2 は a_2 及び w_2 の合計であり、 n は n_1 及び n_2 の合計である]

によって定義されるカイ二乗統計値に基づくカイ二乗累積分布関数を用いて決定される、請求項 9 に記載の方法。

【請求項 11】

前記確率値は、2つの比率である p_1 及び p_2 に基づいて決定され、ここで $p_1 = a_1 / n_1$ であり、 $p_2 = a_2 / n_2$ であり、 a_2 は第一のバリエーション数であり、 a_1 は第二のバリエーション数であり、 n_2 は第一のバリエーション数及び第一の野生型数の合計であり、そして n_1 は第二のバリエーション数及び第二の野生型数の合計である、請求項 9 に記載の方法。

【請求項 12】

前記確率値は、 z スコア、修正された z スコア、 p 値、カイ二乗値、累積確率値及び信頼水準を示すクオリティスコアの 1 つ又は 2 つ以上である、請求項 9 に記載の方法。

【請求項 13】

前記クオリティスコアはルックアップテーブルを用いて決定され、ここで前記ルックアップテーブルの入力は、カイ二乗値又は正規クオントイル値の少なくとも 1 つである、請求項 12 に記載の方法。

【請求項 14】

前記閾値は、既知の真陽性と偽陽性とを有するトレーニングデータに基づいて、サポートベクターマシン分類器を用いて決定される、請求項 9 ~ 13 のいずれか 1 項に記載の方法。

【請求項 15】

実施されるときにコンピュータシステムを制御して第一の試料の標的領域における真のバリエーションを検出する複数の指示を記憶する、非一過性のコンピュータ可読媒体を含むコンピュータ製品であって、前記指示は、

- 1 又は 2 以上の試料からの DNA 断片のシーケンシングから得られる複数の配列リードを受け取り、ここで前記 1 又は 2 以上の試料は第一の試料を含み、前記シーケンシングは前記 DNA 断片中の標的領域を標的化することを含み；
- 前記複数の配列リードを、参照配列の標的領域にアラインメントし；
- 複数のバリエーションクラス中の 1 つのバリエーションクラス中のバリエーションの参照アレルを各々が有する参照配列の標的領域における 1 セットの配列位置を同定し、ここで、前記複数のバリエーションクラスの各々は、1 又は 2 以上のバリエーションを含むように定義され、各々のバリエーションは対応する参照アレルとは異なるバリエーションアレルを有し、そして前記 1 セットの配列位置は第一の位置を含み、

前記 1 セットの配列位置の各々の位置において、かつ前記 1 又は 2 以上の試料の各々の試料に関して：

- 各試料に関する各位置におけるリード数を決定し；
- 各試料の配列リードに基づいて、前記バリエーションクラス中のバリエーションに関するバリエーションアレルを有する候補バリエーションを同定し、ここで前記バリエーションアレルは、前記参照配列の同一の位置における参照アレルとは異なり、そして各試料中の各位置における候補バリエーションの総数は、各試料に関する各位置中のバリエーション数であり；

- 前記リード数及びバリエーション数に基づいてバリエーションクラス中のバリエーションのバリエーション頻度を決定し、ここで各試料中の各位置に関するバリエーション頻度は、統計分布を形成し、第一の試料に関する前記1セットの配列位置中の第一の位置におけるバリエーション頻度は、第一のバリエーション頻度であり；
- 前記第一のバリエーション頻度を統計分布の値と比較して、前記統計分布の値に対する前記第一のバリエーション頻度の確率値を決定し；そして
- 第一の試料中の候補バリエーションが真陽性であるか否かを決定する一部として、前記確率値を閾値と比較すること、ここで前記閾値は、前記バリエーションクラス中のバリエーションに関して偽陽性と真陽性とを区別する、を含む、コンピュータ製品。