

【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 06 F 19/22 (2011.01)

C 12 Q 1/68 (2018.01)

【F I】

G 06 F 19/22

C 12 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ントを検出するための、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断片のシークエ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、前記シークエンシングは前記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ントアレルを有する第一の候補バリエントを同定し、ここで前記第一のバリエントアレルは、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ントアレルに関する第一のバリエント頻度を決定し、

- 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、第一のバリエントアレルの第二のバリエント頻度を決定し、ここで前記第二のバリエント頻度は統計分布を形成し；

- 前記第一のバリエント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ント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ントアレルに関して、第一の試料において第一の候補バリエント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ント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、単一のシークエンシングランにおいてシークエンシングされる、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、前記第二のバリエント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ントアレルを有するバリエントを検出するための、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断片のシークエ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、前記シークエンシングは前記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、前記第一のバリエントアレルが、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ントアレルは、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ントアレルのバリエント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ントアレルの第一のバリエント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ントアレルの第二のバリエント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ントアレルを有するバリエントが発生する確率値を決定し；そして
- 第一のバリエント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ント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ント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 は第一のバリアント数であり、 w_2 は第一の野生型数であり、 a_1 は第二のバリアント数であり、 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 は第一のバリアント数であり、 a_1 は第二のバリアント数であり、 n_2 は第一のバリアント数及び第一の野生型数の合計であり、そして n_1 は第二のバリアント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アント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断片のシークエ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、前記シークエンシングは前記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；

- 複数のバリアントクラス中の1つのバリアントクラス中のバリアント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、前記複数のバリアントクラスの各々は、1又は2以上のバリアントを含むように定義され、各々のバリアントは対応する参照アレルとは異なるバリアントアレルを有し、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、前記バリアントクラス中のバリアントに関するバリアントアレルを有する候補バリアントを同定し、ここで前記バリアントアレルは、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アントの総数は、各試料に関する各位置中のバリアント数であり；

- 前記リード数及びバリアント数に基づいてバリアントクラス中のバリアントのバリアント頻度を決定し、ここで各試料中の各位置に関するバリアント頻度は、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アント頻度は、第一のバリアント頻度であり；
- 前記第一のバリアント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アント頻度の確率値を決定し；そして
- 第一の試料中の候補バリアント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、前記バリアントクラス中のバリアント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。