

【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11月8日(2018.11.8)

【公表番号】特表2018-501539(P2018-501539A)

【公表日】平成30年1月18日(2018.1.18)

【年通号数】公開・登録公報2018-002

【出願番号】特願2017-521153(P2017-521153)

【国際特許分類】

G 06 F 19/22 (2011.01)

C 12 N 15/09 (2006.01)

【F I】

G 06 F 19/22

C 12 N 15/00 A

【手続補正書】

【提出日】平成30年9月28日(2018.9.28)

【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】

基準ゲノム配列に対してゲノムサンプルからバリアントを判定するためのコンピュータ実装方法であって、

少なくとも1つのプロセッサおよびメモリを有する電子デバイスにおいて、

先に配列決定されたサンプルからの配列データのエラーテーブルにアクセスするステップと、

ゲノムサンプルから収集されたリードのセットから可能なハプロタイプのセットを判定するステップと、

前記可能なハプロタイプのセットおよび前記エラーテーブルに基づいて、ディプロタイプのセットを生成するステップであって、前記可能なハプロタイプのセットは、前記エラーテーブルによってフィルタリングされる、ステップと、

前記ディプロタイプのセットをスコアリングするステップと、

前記ディプロタイプのセットをスコアリングするステップに基づいて、バリアントを出力するステップと、を含む、方法。

【請求項2】

収集されたリードのセットからk-merグラフを生成するステップと、

前記生成されたk-merグラフを連続的グラフにまとめるステップと、

前記連続的グラフから前記可能なハプロタイプのセットを生成するステップと、

をさらに含む、請求項1に記載の方法。

【請求項3】

前記ディプロタイプのセットをスコアリングするステップはさらに、ディプロタイプ毎に事後確率を判定することを含む、請求項1または2に記載の方法。

【請求項4】

前記エラーテーブルを生成するステップをさらに含み、前記エラーテーブルを生成するステップは、

リードを基準サンプルに対してアライメントすることと、

リードが前記基準サンプルとミスマッチを有する部位を判定することと、

ミスマッチを有する部位を前記エラーテーブルに追加することと、を含む、請求項 1 ～ 3 のいずれか一項に記載の方法。

【請求項 5】

前記エラーテーブルを生成するステップはさらに、シーケンサエラーと関連付けられない部位を前記エラーテーブルからフィルタリングすることを含む、請求項 4 に記載の方法。

【請求項 6】

前記エラーテーブルを生成するステップはさらに、Hardy - Weinberg 試験、Bayes Factor 試験、または Strand Bias 試験の 1 つまたはそれより多くを使用して、閾値に満たない部位を前記エラーテーブルからフィルタリングすることを含む、請求項 4 または 5 に記載の方法。

【請求項 7】

配列データのエラーテーブルを生成するためのコンピュータ実装方法であって、少なくとも 1 つのプロセッサおよびメモリを有する電子デバイスにおいて、ゲノムサンプルから収集されたリードのセットから可能なハプロタイプのセットを判定するステップと、

前記収集されたリードのセットを基準サンプルに対してアライメントするステップと、前記基準サンプルから前記収集されたリードのセットのリードがミスマッチを有する部位を判定するステップと、

ミスマッチを有する部位をエラーテーブルに追加するステップと、を含む、方法。

【請求項 8】

前記可能なハプロタイプのセットを判定するステップは、
k - m e r グラフを前記収集されたリードのセットから生成することと、
前記生成された k - m e r グラフを連続的グラフにまとめることと、
前記連続的グラフから前記可能なハプロタイプのセットを判定することと、を含む、請求項 7 に記載の方法。

【請求項 9】

非一過性コンピュータ可読記憶媒体であって、
先に配列決定されたサンプルからの配列データのエラーテーブルにアクセスするステップと、
ゲノムサンプルからの収集されたリードのセットから可能なハプロタイプのセットを判定するステップと、
前記可能なハプロタイプのセットおよび前記エラーテーブルに基づいて、ディプロタイプのセットを生成するステップであって、前記可能なハプロタイプのセットは、前記エラーテーブルによってフィルタリングされる、ステップと、
前記ディプロタイプのセットをスコアリングするステップと、
前記ディプロタイプのセットをスコアリングするステップに基づいて、バリアントを出力するステップとのためのコンピュータ実行可能命令を含む、非一過性コンピュータ可読記憶媒体。

【請求項 10】

収集されたリードのセットから k - m e r グラフを生成するステップと、
前記生成された k - m e r グラフを連続的グラフにまとめるステップと、
前記連続的グラフから前記可能なハプロタイプのセットを生成するステップとをさらに含む、請求項 9 に記載の非一過性コンピュータ可読記憶媒体。

【請求項 11】

前記ディプロタイプのセットをスコアリングするステップはさらに、ディプロタイプ毎に事後確率を判定することを含む、請求項 9 または 10 に記載の非一過性コンピュータ可読記憶媒体。

【請求項 12】

前記エラーテーブルを生成するステップをさらに含み、前記エラーテーブルを生成するステップは、

リードを基準サンプルに対してアライメントすることと、
リードが前記基準サンプルとミスマッチを有する部位を判定することと、
ミスマッチを有する部位を前記エラーテーブルに追加することとを含む、請求項 9 ~ 11 のいずれか一項に記載の非一過性コンピュータ可読記憶媒体。

【請求項 13】

前記エラーテーブルを生成するステップはさらに、シーケンサエラーと関連付けられない部位を前記エラーテーブルからフィルタリングすることを含む、請求項 12 に記載の非一過性コンピュータ可読記憶媒体。

【請求項 14】

前記エラーテーブルを生成するステップはさらに、Hardy - Weinberg 試験、Bayes Factor 試験、または Strand Bias 試験の 1 つまたはそれより多くを使用して、閾値に満たない部位を前記エラーテーブルからフィルタリングすることを含む、請求項 12 または 13 に記載の非一過性コンピュータ可読記憶媒体。

【請求項 15】

システムであって、
1 つまたはそれより多くのプロセッサと、
メモリと、

1 つまたはそれより多くのプログラムであって、前記 1 つまたはそれより多くのプログラムは、前記メモリ内に記憶され、前記 1 つまたはそれより多くのプロセッサによって実行されるように構成され、

先に配列決定されたサンプルからの配列データのエラーテーブルにアクセスするステップと、

ゲノムサンプルから収集されたリードのセットから可能なハプロタイプのセットを判定するステップと、

前記可能なハプロタイプのセットおよび前記エラーテーブルに基づいて、ディプロタイプのセットを生成するステップであって、前記可能なハプロタイプのセットは、前記エラーテーブルによってフィルタリングされる、ステップと、

前記ディプロタイプのセットをスコアリングするステップと、

前記ディプロタイプのセットをスコアリングするステップに基づいて、バリアントを出力するステップとのための命令を含む、1 つまたはそれより多くのプログラムと、を備える、システム。

【請求項 16】

収集されたリードのセットから k - mer グラフを生成するステップと、
前記生成された k - mer グラフを連続的グラフにまとめるステップと、
前記連続的グラフから前記可能なハプロタイプのセットを生成するステップと、
をさらに含む、請求項 15 に記載のシステム。

【請求項 17】

前記ディプロタイプのセットをスコアリングするステップはさらに、ディプロタイプ毎に事後確率を判定することを含む、請求項 15 または 16 に記載のシステム。

【請求項 18】

前記エラーテーブルを生成するステップをさらに含み、前記エラーテーブルを生成するステップは、

リードを基準サンプルに対してアライメントすることと、
リードが前記基準サンプルとミスマッチを有する部位を判定することと、
ミスマッチを有する部位を前記エラーテーブルに追加することとを含む、請求項 15 ~ 17 のいずれか一項に記載のシステム。

【請求項 19】

前記エラーテーブルを生成するステップはさらに、シーケンサエラーと関連付けられない部位を前記エラーテーブルからフィルタリングすることを含む、請求項 18 に記載のシステム。

【請求項 20】

前記エラーテーブルを生成するステップはさらに、
Hardy-Weinberg試験、Bayes Factor試験、または Strand Bias試験の1つまたはそれより多くを使用して、閾値に満たない部位を前記エラーテーブルからフィルタリングすることを含む、請求項18または19に記載のシステム。

【手続補正2】

【補正対象書類名】明細書

【補正対象項目名】0006

【補正方法】変更

【補正の内容】

【0006】

加えて、バリエントコーラーのためと、エラーテーブルを生成するためのシステム、電子デバイス、グラフィカルユーザインターフェース、および非一過性コンピュータ可読記憶媒体（説明される1つまたはそれより多くのプロセスを実行するためのプログラムおよび命令を含む、記憶媒体）が、説明される。

本発明の実施形態において、例えば以下の項目が提供される。

(項目1)

基準ゲノム配列に対してゲノムサンプルからバリエントを判定するためのコンピュータ実装方法であって、

少なくとも1つのプロセッサおよびメモリを有する電子デバイスにおいて、

先に配列決定されたサンプルからの配列データのエラーテーブルにアクセスするステップと、

ゲノムサンプルから収集されたリードのセットから可能なハプロタイプのセットを判定するステップと、

前記可能なハプロタイプのセットおよび前記エラーテーブルに基づいて、ディプロタイプのセットを生成するステップであって、前記可能なハプロタイプのセットは、前記エラーテーブルによってフィルタリングされる、ステップと、

前記ディプロタイプのセットをスコアリングするステップと、

前記ディプロタイプのセットをスコアリングするステップに基づいて、バリエントを出力するステップと、を含む、方法。

(項目2)

収集されたリードのセットからk-merグラフを生成するステップと、

前記生成されたk-merグラフを連続的グラフにまとめるステップと、

前記連続的グラフから前記可能なハプロタイプのセットを生成するステップと、

をさらに含む、項目1に記載の方法。

(項目3)

前記ディプロタイプのセットをスコアリングするステップはさらに、ディプロタイプ毎に事後確率を判定することを含む、項目1に記載の方法。

(項目4)

前記エラーテーブルを生成するステップをさらに含み、前記エラーテーブルを生成するステップは、

リードを基準サンプルに対してアライメントすることと、

リードが前記基準サンプルとミスマッチを有する部位を判定することと、

ミスマッチを有する部位を前記エラーテーブルに追加することと、を含む、項目1に記載の方法。

(項目5)

前記エラーテーブルを生成するステップはさらに、シーケンサエラーと関連付けられない部位を前記エラーテーブルからフィルタリングすることを含む、項目4に記載の方法。

(項目6)

前記エラーテーブルを生成するステップはさらに、Hardy-Weinberg試験、Bayes Factor試験、またはStrand Bias試験の1つまたはそれより多くを使用して、閾値に満たない部位を前記エラーテーブルからフィルタリングすることを含む、項目4に記載の方法。

(項目7)

配列データのエラーテーブルを生成するためのコンピュータ実装方法であって、少なくとも1つのプロセッサおよびメモリを有する電子デバイスにおいて、ゲノムサンプルから収集されたリードのセットから可能なハプロタイプのセットを判定するステップと、

前記収集されたリードのセットを基準サンプルに対してアライメントするステップと、前記基準サンプルから前記収集されたリードのセットのリードがミスマッチを有する部位を判定するステップと、

ミスマッチを有する部位をエラーテーブルに追加するステップと、を含む、方法。

(項目8)

前記可能なハプロタイプのセットを判定するステップは、

k - m e rグラフを前記収集されたリードのセットから生成することと、

前記生成された k - m e rグラフを連続的グラフにまとめることと、

前記連続的グラフから前記可能なハプロタイプのセットを判定することと、を含む、項目7に記載の方法。

(項目9)

非一過性コンピュータ可読記憶媒体であって、

先に配列決定されたサンプルからの配列データのエラーテーブルにアクセスするステップと、

ゲノムサンプルからの収集されたリードのセットから可能なハプロタイプのセットを判定するステップと、

前記可能なハプロタイプのセットおよび前記エラーテーブルに基づいて、ディプロタイプのセットを生成するステップであって、前記可能なハプロタイプのセットは、前記エラーテーブルによってフィルタリングされる、ステップと、

前記ディプロタイプのセットをスコアリングするステップと、

前記ディプロタイプのセットをスコアリングするステップに基づいて、バリエントを出力するステップとのためのコンピュータ実行可能命令を含む、非一過性コンピュータ可読記憶媒体。

(項目10)

収集されたリードのセットから k - m e rグラフを生成するステップと、

前記生成された k - m e rグラフを連続的グラフにまとめるステップと、

前記連続的グラフから前記可能なハプロタイプのセットを生成するステップとをさらに含む、項目9に記載の非一過性コンピュータ可読記憶媒体。

(項目11)

前記ディプロタイプのセットをスコアリングするステップはさらに、ディプロタイプ毎に事後確率を判定することを含む、項目9に記載の非一過性コンピュータ可読記憶媒体。

(項目12)

前記エラーテーブルを生成するステップをさらに含み、前記エラーテーブルを生成するステップは、

リードを基準サンプルに対してアライメントすることと、

リードが前記基準サンプルとミスマッチを有する部位を判定することと、

ミスマッチを有する部位を前記エラーテーブルに追加することとを含む、項目9に記載の非一過性コンピュータ可読記憶媒体。

(項目13)

前記エラーテーブルを生成するステップはさらに、シーケンサエラーと関連付けられない

部位を前記エラーテーブルからフィルタリングすることを含む、項目12に記載の非一過性コンピュータ可読記憶媒体。

(項目14)

前記エラーテーブルを生成するステップはさらに、Hardy-Weinberg試験、Bayes Factor試験、またはStrand Bias試験の1つまたはそれより多くを使用して、閾値に満たない部位を前記エラーテーブルからフィルタリングすることを含む、項目12に記載の非一過性コンピュータ可読記憶媒体。

(項目15)

システムであって、

1つまたはそれより多くのプロセッサと、

メモリと、

1つまたはそれより多くのプログラムであって、前記1つまたはそれより多くのプログラムは、前記メモリ内に記憶され、前記1つまたはそれより多くのプロセッサによって実行されるように構成され、

先に配列決定されたサンプルからの配列データのエラーテーブルにアクセスするステップと、

ゲノムサンプルから収集されたリードのセットから可能なハプロタイプのセットを判定するステップと、

前記可能なハプロタイプのセットおよび前記エラーテーブルに基づいて、ディプロタイプのセットを生成するステップであって、前記可能なハプロタイプのセットは、前記エラーテーブルによってフィルタリングされる、ステップと、

前記ディプロタイプのセットをスコアリングするステップと、

前記ディプロタイプのセットをスコアリングするステップに基づいて、バリアントを出力するステップとのための命令を含む、1つまたはそれより多くのプログラムと、を備える、システム。

(項目16)

収集されたリードのセットからk-merグラフを生成するステップと、

前記生成されたk-merグラフを連続的グラフにまとめるステップと、

前記連続的グラフから前記可能なハプロタイプのセットを生成するステップと、

をさらに含む、項目9に記載のシステム。

(項目17)

前記ディプロタイプのセットをスコアリングするステップはさらに、ディプロタイプ毎に事後確率を判定することを含む、項目9に記載のシステム。

(項目18)

前記エラーテーブルを生成するステップをさらに含み、前記エラーテーブルを生成するステップは、

リードを基準サンプルに対してアライメントすることと、

リードが前記基準サンプルとミスマッチを有する部位を判定することと、

ミスマッチを有する部位を前記エラーテーブルに追加することとを含む、項目9に記載のシステム。

(項目19)

前記エラーテーブルを生成するステップはさらに、シーケンサエラーと関連付けられない部位を前記エラーテーブルからフィルタリングすることを含む、項目18に記載のシステム。

(項目20)

前記エラーテーブルを生成するステップはさらに、

Hardy-Weinberg試験、Bayes Factor試験、またはStrand Bias試験の1つまたはそれより多くを使用して、閾値に満たない部位を前記エラーテーブルからフィルタリングすることを含む、項目18に記載のシステム。