

(19)日本国特許庁(JP)

(12)公表特許公報(A)

(11)公表番号

特表2024-543762

(P2024-543762A)

(43)公表日 令和6年11月26日(2024.11.26)

(51)国際特許分類

G 1 6 B 30/00 (2019.01)

F I

G 1 6 B 30/00

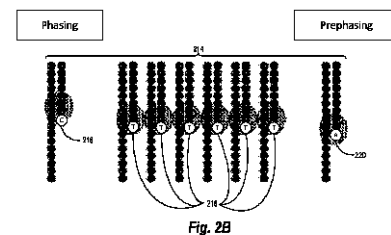
審査請求 未請求 予備審査請求 未請求 (全62頁)

(21)出願番号	特願2023-579819(P2023-579819)	(71)出願人	500358711 イルミナ インコーポレイテッド アメリカ合衆国 カリフォルニア州 9 2 1 2 2 サンディエゴ イルミナ ウェイ 5 2 0 0
(86)(22)出願日	令和4年11月28日(2022.11.28)	(74)代理人	100108453 弁理士 村山 靖彦
(85)翻訳文提出日	令和5年12月26日(2023.12.26)	(74)代理人	100110364 弁理士 実広 信哉
(86)国際出願番号	PCT/US2022/080512	(74)代理人	100133400 弁理士 阿部 達彦
(87)国際公開番号	WO2023/102354	(72)発明者	エリック・ジョン・オジャード アメリカ合衆国・カリフォルニア・9 2 1 2 2・サン・ディエゴ・イルミナ・ウ エイ・5 2 0 0
(87)国際公開日	令和5年6月8日(2023.6.8)		
(31)優先権主張番号	63/285,187		
(32)優先日	令和3年12月2日(2021.12.2)		
(33)優先権主張国・地域又は機関	米国(US)		
(81)指定国・地域	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,A T,BE,BG,CH,CY,CZ,DE,DK,EE,ES,FI,FR ,GB,GR,HR,HU,IE,IS,IT,LT,LU,LV,MC, 最終頁に続く		最終頁に続く

(54)【発明の名称】 ヌクレオチド塩基コールを決定するためのクラスタ固有シグナル補正の生成

(57)【要約】

本開示は、オリゴヌクレオチドの特定のクラスタに対するフェージング及びプリフェージングの効果を正確かつ効率的に推定し、クラスタに対するクラスタ固有のフェージング補正を決定する方法、システム、及び非一時的コンピュータ可読媒体の実施形態を記載する。例えば、開示されたシステムは、フェージング又はプリフェージングを頻繁に引き起こすエラー誘導配列を示すオリゴヌクレオチドのクラスタを動的に特定することができる。開示されたシステムが、そのようなエラー誘導配列に続くリード位置におけるシグナルをサイクル中に検出するとき、開示されたシステムは、クラスタ固有のフェージング係数を生成し、そのようなクラスタ固有のフェージング係数に従ってシグナルを補正することができる。例えば、開示されたシステムは、線形等化器、判定帰還型等化器、又は最尤系列推定器を利用して、クラスタ固有のフェージング係数を生成することができる。



10

20

【特許請求の範囲】

【請求項 1】

非一時的コンピュータ可読記憶媒体であって、少なくとも 1 つのプロセッサによって実行されると、コンピューティング装置に、

オリゴヌクレオチドのクラスタについて、1 つ以上のヌクレオチド断片リード内のエラー誘導配列に続くリード位置を特定させる命令と、

前記リード位置に対応するサイクル中に、前記オリゴヌクレオチドのクラスタ内の標識ヌクレオチド塩基からのシグナルを検出させる命令と、

前記オリゴヌクレオチドのクラスタについて、推定されたフェージング及び推定されたプリフェージングのために前記シグナルを補正するクラスタ固有のフェージング補正を決定させる命令と、

前記クラスタ固有のフェージング補正に基づいて、前記シグナルを調整させる命令と、

調整された前記シグナルに基づいて、前記オリゴヌクレオチドのクラスタに対応する前記リード位置についてのヌクレオチド塩基コールを決定させる命令と、を含む、非一時的コンピュータ可読記憶媒体。

10

【請求項 2】

前記エラー誘導配列が、配列認識モデルによって特定された 1 つ以上の反復ヌクレオチド塩基の配列、配列モチーフ、又はトリガー配列を含む、請求項 1 に記載の非一時的コンピュータ可読記憶媒体。

【請求項 3】

前記 1 つ以上の反復ヌクレオチド塩基の配列又は前記配列モチーフが、同じヌクレオチド塩基のホモポリマー、ニアホモポリマー、グアニン四重鎖、可変数タンデム反復 (V N T R)、ジヌクレオチド反復配列、トリヌクレオチド反復配列、逆方向反復配列、ミニサテライト配列、マイクロサテライト配列、又はパリンドローム配列を含む、請求項 2 に記載の非一時的コンピュータ可読記憶媒体。

20

【請求項 4】

前記少なくとも 1 つのプロセッサによって実行されると、前記コンピューティング装置に、

前記オリゴヌクレオチドのクラスタについて、前のサイクルについてのヌクレオチド塩基に対応するクラスタ固有のフェージング係数、及び後続のサイクルについてのヌクレオチド塩基に対応するクラスタ固有のプリフェージング係数を決定することと、

前記クラスタ固有のフェージング係数及び前記クラスタ固有のプリフェージング係数に基づいて、前記クラスタ固有のフェージング補正を決定することと、によって、前記クラスタ固有のフェージング補正を決定させる命令を更に含む、請求項 1 に記載の非一時的コンピュータ可読記憶媒体。

30

【請求項 5】

前記少なくとも 1 つのプロセッサによって実行されると、前記コンピューティング装置に、

前記クラスタ固有のフェージング係数に基づいて、前記前のサイクルについての前記ヌクレオチド塩基のフェージング効果を推定する前のサイクルの重みを生成することと、

前記クラスタ固有のプリフェージング係数に基づいて、前記後続のサイクルについての前記ヌクレオチド塩基のプリフェージング効果を推定する後続のサイクルの重みを生成することと、

前記クラスタ固有のフェージング係数及び前記クラスタ固有のプリフェージング係数に基づいて、前記サイクルについての前記フェージング効果及び前記プリフェージング効果を推定する現在のサイクルの重みを生成することと、

前記前のサイクルの重み、前記後続のサイクルの重み、及び前記現在のサイクルの重みに基づいて、前記クラスタ固有のフェージング補正を決定することと、によって、前記クラスタ固有のフェージング係数及び前記クラスタ固有のプリフェージング係数に基づいて、前記クラスタ固有のフェージング補正を決定させる命令を更に含む、請求項 4 に記載の

40

50

非一時的コンピュータ可読記憶媒体。

【請求項 6】

前記少なくとも 1 つのプロセッサによって実行されると、前記コンピューティング装置に、前記前のサイクルに対応するシグナル強度、前記サイクルに対応するシグナル強度、及び前記後続のサイクルに対応するシグナル強度に更に基づいて、前記クラスタ固有のフェージング補正を決定させる命令を更に含む、請求項 5 に記載の非一時的コンピュータ可読記憶媒体。

【請求項 7】

前記少なくとも 1 つのプロセッサによって実行されると、前記コンピューティング装置に、

前記オリゴヌクレオチドのクラスタについて、前のサイクルのセットについてのヌクレオチド塩基のセットに対応するクラスタ固有のフェージング係数のセットを決定することと、

前記オリゴヌクレオチドのクラスタについて、後続のサイクルのセットについてのヌクレオチド塩基のセットに対応するクラスタ固有のプリフェージング係数のセットを決定することと、

前記クラスタ固有のフェージング係数のセット及び前記クラスタ固有のプリフェージング係数のセットに基づいて、前記クラスタ固有のフェージング補正を決定することと、によって、前記クラスタ固有のフェージング補正を決定させる命令を更に含む、請求項 1 に記載の非一時的コンピュータ可読記憶媒体。

【請求項 8】

前記少なくとも 1 つのプロセッサによって実行されると、前記コンピューティング装置に、

オリゴヌクレオチドのクラスタのセットについて、推定されたフェージング及び推定されたプリフェージングのために前記クラスタのセットからのシグナルを補正するマルチクラスタフェージング補正を決定させる命令と、

前記クラスタ固有のフェージング補正又は前記マルチクラスタフェージング補正に基づいて、前記シグナルを調整させる命令と、を更に含む、請求項 1 に記載の非一時的コンピュータ可読記憶媒体。

【請求項 9】

前記少なくとも 1 つのプロセッサによって実行されると、前記コンピューティング装置に、前記オリゴヌクレオチドのクラスタ及び後続のリード位置について、後続のサイクルについてのシグナルのフェージング及びプリフェージングのために前記オリゴヌクレオチドのクラスタからの前記後続のサイクルについての前記シグナルを補正する異なるクラスタ固有のフェージング補正を決定させる命令を更に含む、請求項 1 に記載の非一時的コンピュータ可読記憶媒体。

【請求項 10】

前記少なくとも 1 つのプロセッサによって実行されると、前記コンピューティング装置に、

オリゴヌクレオチドの追加のクラスタについて、異なるヌクレオチド断片リード内の前記エラー誘導配列に先行する異なるリード位置を特定させる命令と、

前記異なるリード位置に対応するサイクル中に、前記オリゴヌクレオチドの追加のクラスタ内の標識ヌクレオチド塩基から追加のシグナルを検出させる命令と、

前記オリゴヌクレオチドの追加のクラスタについてのクラスタ固有のフェージング補正のないマルチクラスタフェージング補正に基づいて、前記追加のシグナルを調整させる命令と、を更に含む、請求項 1 に記載の非一時的コンピュータ可読記憶媒体。

【請求項 11】

前記少なくとも 1 つのプロセッサによって実行されると、前記コンピューティング装置に、配列決定装置のプロセッサを利用して、前記クラスタ固有のフェージング補正を決定させる命令を更に含む、請求項 1 に記載の非一時的コンピュータ可読記憶媒体。

【請求項 1 2】

システムであって、

少なくとも 1 つのプロセッサと、

前記少なくとも 1 つのプロセッサによって実行されると、前記システムに、

オリゴヌクレオチドのクラスタについて、1 つ以上のヌクレオチド断片リード内のエラー誘導配列に続くリード位置を特定させる命令と、

前記リード位置に対応するサイクル中に、前記オリゴヌクレオチドのクラスタ内の標識ヌクレオチド塩基からのシグナルを検出させる命令と、

前記オリゴヌクレオチドのクラスタについて、前のサイクルについてのヌクレオチド塩基に対応するクラスタ固有のフェージング係数、及び後続のサイクルについてのヌクレオチド塩基に対応するクラスタ固有のプリフェージング係数を決定させる命令と、

前記クラスタ固有のフェージング係数及び前記クラスタ固有のプリフェージング係数に基づいて、前記シグナルを調整させる命令と、

調整された前記シグナルに基づいて、前記オリゴヌクレオチドのクラスタに対応する前記リード位置についてのヌクレオチド塩基コールを決定させる命令と、を含む、非一時的コンピュータ可読記憶媒体と、を備える、システム。

10

【請求項 1 3】

前記少なくとも 1 つのプロセッサによって実行されると、前記システムに、前記システムの配列決定マシン上で、線形等化器、判定帰還型等化器、最尤系列推定器、順方向－逆方向モデル、又は機械学習モデルを利用して、前記クラスタ固有のフェージング係数及び前記クラスタ固有のプリフェージング係数を決定させる命令を更に含む、請求項 1 2 に記載のシステム。

20

【請求項 1 4】

前記少なくとも 1 つのプロセッサによって実行されると、前記システムに、配列決定実行後に前記クラスタ固有のフェージング係数及び前記クラスタ固有のプリフェージング係数を決定させる命令を更に含む、請求項 1 2 に記載のシステム。

【請求項 1 5】

前記少なくとも 1 つのプロセッサによって実行されると、前記システムに、

オリゴヌクレオチドのクラスタのセットについて、推定されたフェージングのためのマルチクラスタフェージング係数又は推定されたプリフェージングのためのマルチクラスタプリフェージング係数のうちの 1 つ以上を決定させる命令と、

30

前記マルチクラスタフェージング係数、前記クラスタ固有のフェージング係数、前記マルチクラスタプリフェージング係数、又は前記クラスタ固有のプリフェージング係数のうちの 1 つ以上に基づいて、前記シグナルを調整させる命令と、を更に含む、請求項 1 2 に記載のシステム。

【請求項 1 6】

前記少なくとも 1 つのプロセッサによって実行されると、前記システムに、

前記オリゴヌクレオチドのクラスタについて、追加の前のサイクルについての追加のヌクレオチド塩基に対応する追加のクラスタ固有のフェージング係数を決定することと、

前記オリゴヌクレオチドのクラスタについて、追加の後続のサイクルについての追加のヌクレオチド塩基に対応する追加のクラスタ固有のプリフェージング係数を決定することと、

40

前記クラスタ固有のフェージング係数、前記追加のクラスタ固有のフェージング係数、前記クラスタ固有のプリフェージング係数、及び前記追加のクラスタ固有のプリフェージング係数に基づいて、クラスタ固有のフェージング補正を決定することと、によって、前記シグナルを調製させる命令を更に含む、請求項 1 2 に記載のシステム。

【請求項 1 7】

前記少なくとも 1 つのプロセッサによって実行されると、前記システムに、

前記クラスタ固有のフェージング係数に基づいて、前記前のサイクルについての前記ヌクレオチド塩基のフェージング効果を推定する前のサイクルの重みを生成することと、

50

前記クラスタ固有のプリフェージング係数に基づいて、前記後続のサイクルについての前記ヌクレオチド塩基のプリフェージング効果を推定する後続のサイクルの重みを生成することと、

前記クラスタ固有のフェージング係数及び前記クラスタ固有のプリフェージング係数に基づいて、前記サイクルについての前記フェージング効果及び前記プリフェージング効果を推定する現在のサイクルの重みを生成することと、

前記前のサイクルの重み、前記後続のサイクルの重み、及び前記現在のサイクルの重みに基づいて、クラスタ固有のフェージング補正を決定することと、

前記クラスタ固有のフェージング補正を前記シグナルに適用することと、によって、前記クラスタ固有のフェージング係数及び前記クラスタ固有のプリフェージング係数に基づいて、前記シグナルを調整させる命令を更に含む、請求項 12 に記載のシステム。

10

【請求項 18】

コンピュータ実装方法であって、

オリゴヌクレオチドのクラスタについて、1つ以上のヌクレオチド断片リード内のエラー誘導配列に続くリード位置を特定することと、

前記リード位置に対応するサイクル中に、前記オリゴヌクレオチドのクラスタ内の標識ヌクレオチド塩基からのシグナルを検出することと、

前記オリゴヌクレオチドのクラスタについて、フェージング及びプリフェージングのために前記シグナルを補正するクラスタ固有のフェージング補正を決定することと、

前記クラスタ固有のフェージング補正に基づいて、前記シグナルを調整することと、

20

調整された前記シグナルに基づいて、前記オリゴヌクレオチドのクラスタに対応する前記リード位置についてのヌクレオチド塩基コールを決定することと、を含む、コンピュータ実装方法。

【請求項 19】

前記エラー誘導配列が、1つ以上の反復ヌクレオチド塩基の配列又は方向固有の配列モチーフを含む、請求項 18 に記載のコンピュータ実装方法。

【請求項 20】

前記クラスタ固有のフェージング補正を決定することが、

前記オリゴヌクレオチドのクラスタについて、前記サイクルの直前の前のサイクルについてのヌクレオチド塩基に対応するクラスタ固有のフェージング係数、及び前記サイクルの直後の後続のサイクルについてのヌクレオチド塩基に対応するクラスタ固有のプリフェージング係数を決定することと、

30

前記クラスタ固有のフェージング係数及び前記クラスタ固有のプリフェージング係数に基づいて、前記クラスタ固有のフェージング補正を決定することと、を含む、請求項 18 に記載のコンピュータ実装方法。

【請求項 21】

前記クラスタ固有のフェージング補正を決定することが、

前記オリゴヌクレオチドのクラスタについて、前記サイクルの直前の前のサイクルのセットについてのヌクレオチド塩基のセットに対応するクラスタ固有のフェージング係数のセットを決定することと、

40

前記オリゴヌクレオチドのクラスタについて、前記サイクルの直後の後続のサイクルのセットについてのヌクレオチド塩基のセットに対応するクラスタ固有のプリフェージング係数のセットを決定することと、

前記クラスタ固有のフェージング係数の前記セット及び前記クラスタ固有のプリフェージング係数の前記セットに基づいて、前記クラスタ固有のフェージング補正を決定することと、を含む、請求項 18 に記載のコンピュータ実装方法。

【請求項 22】

オリゴヌクレオチドのクラスタのセットについて、フェージング及びプリフェージングのために前記クラスタのセットからのシグナルを補正するマルチクラスタフェージング補正を決定することと、

50

前記クラスタ固有のフェージング補正及び前記マルチクラスタフェージング補正の両方に基づいて、前記シグナルを調整することと、を更に含む、請求項 18 に記載のコンピュータ実装方法。

【発明の詳細な説明】

【技術分野】

【0001】

(関連出願の相互参照)

本出願は、2021年12月2日に出願された「GENERATING CLUSTER-SPECIFIC-SIGNAL CORRECTIONS FOR DETERMINING NUCLEOTIDE-BASE CALLS」と題する米国仮出願第 63 / 285, 187 号の利益及び優先権を主張する。上記出願は、参照によりその全体が本明細書に組み込まれる。

【背景技術】

【0002】

近年、バイオテクノロジー企業及び研究機関は、試料ゲノム又は他の核酸ポリマー中のヌクレオチド塩基の配列を決定するために使用されるハードウェア及びソフトウェアプラットフォームを改善している。例えば、いくつかの既存の核酸配列決定プラットフォームは、従来の Sanger 配列決定又は合成による配列決定 (sequencing-by-synthesis、SBS) を使用することによって、核酸配列の個々のヌクレオチド塩基を決定する。

SBS を使用する場合、既存のプラットフォームは、より正確なヌクレオチド塩基コールを検出するために、クラスタに分類され、並行して合成された数千、数万、又はそれ以上のオリゴヌクレオチドをモニターすることができる。例えば、SBS プラットフォームにおけるカメラは、そのようなクラスタ化され、合成されたオリゴヌクレオチドに組み込まれたヌクレオチド塩基からの照射された蛍光タグの画像を捕捉することができる。画像を捕捉した後、既存の SBS プラットフォームは、配列決定データ分析ソフトウェアを有するコンピューティング装置に画像データを送信して、ゲノム又は他の核酸ポリマーのヌクレオチド塩基配列を決定する。例えば、配列決定データ分析ソフトウェアは、画像データにおいて捕捉された光シグナルに基づいて、所与の画像において照射するタグを有するヌクレオチド塩基を決定することができる。ヌクレオチド塩基をオリゴヌクレオチドに周期的に組み込み、様々な配列決定サイクルにおいて放出された光シグナルの画像を捕捉することによって、SBS プラットフォームは、特定のクラスタに対応するヌクレオチドリードを決定し、全ゲノム試料又は核酸ポリマーの他の試料中に存在するヌクレオチド塩基の配列を決定することができる。

【0003】

これらの最近の進歩にもかかわらず、既存の核酸配列決定プラットフォーム及び配列決定データ分析ソフトウェア (併せて、以下、「既存の配列決定システム」) は、フェージングのためにシグナルを検出及び補正する精度、適用性、及び効率を妨げる技術的制限に悩まされることが多い。既存の核酸配列決定プラットフォームが、様々なクラスタのオリゴヌクレオチドのヌクレオチド塩基を組み込み、検出するサイクルを実行する場合、プラットフォームは、多くの場合、位相がずれたいくつかのヌクレオチド塩基を組み込み、検出する。フェージング及びプリフェージングが起こる場合、核酸配列決定プラットフォームは、それぞれ、前のサイクル (フェージング) に対応するヌクレオチド塩基又は後続のサイクル (プリフェージング) に対応するヌクレオチド塩基を組み込む。フェージング又はプリフェージングのために、核酸配列決定プラットフォームは、現在のサイクルのための組み込まれたヌクレオチド塩基、並びに前又は後続のサイクルに対応する組み込まれたヌクレオチド塩基の混合物を有するクラスタからの光シグナルの画像を捕捉する。既存の配列決定システムは、そのようなフェージング効果及びプリフェージング効果を正確に検出及び補正することができないことが多く、その結果、特定のサイクルにおけるクラスタに対応するヌクレオチドリードに対して不正確なヌクレオチド塩基コールを決定することがある。既存の配列決定システムが正しいヌクレオチド塩基コールを生成する場合であっ

ても、そのようなシステムは、フェージング及びプリフェージングに部分的に起因して、より低いクオリティの配列決定メトリクスを有するリードに対する塩基コールを生成する場合がある。例えば、ある特定の反復ヌクレオチド配列に続くリード位置で混合シグナルを捕捉する既存の配列決定システムは、Phredクオリティスコア（例えば、Q30未満）などのより低いクオリティスコアを有する塩基コールを生成することが多い。

【0004】

既存の配列決定システムは、多くの場合、上述のフェージング及びプリフェージングによって引き起こされる不正確さを回避しようと試みる。しかし、これらのシステムは、しばしば融通が利かず、画一的な手法に依存する。例えば、従来の配列決定システムは、多くの場合、大域的フェージング補正及び大域的プリフェージング補正に依存して、各サイクルに対する強度データの鮮明度を最大化する。鮮明度値は、最も明るい塩基強度を、最も明るい塩基強度と2番目に明るい塩基強度との合計で割った比を示す。大域的フェージング補正及び大域的プリフェージング補正の使用は、スライド（例えば、フローセル）の大きな区画に対するシグナルに対するフェージング補正の有効性を制限する。実際、従来の配列決定システムは、多くの場合、クラスタレベルでの変動性を説明できない。例えば、スライドのセクション（例えば、タイル）内の第1のクラスタは、著しいフェージング効果を示す場合があり、セクション内の第2のクラスタは、著しいプリフェージング効果を示す場合があり、同じセクション内の第3のクラスタは、ほとんど又は全くフェージング又はプリフェージングを示さない場合がある。したがって、大域的フェージング補正及び大域的プリフェージング補正に依存する従来の配列決定システムは、クラスタ内の微妙な変動を説明できないことが多い。

【0005】

更に、従来の配列決定システムは、多くの場合、種々のクラスタの画像データを効率的に捕捉し分析するために、限定された記憶リソース及び他の計算リソースを含む。特に、フェージング補正の適用の一部として、従来の配列決定システムは、配列決定画像データ又は配列決定強度データを頻繁に記憶し分析する。例示すると、従来の配列決定システムは、多くの場合、各サイクルについてシグナルデータを収集し、データを保存し、そしてデータを分析する。そのような画像データをサイクルごとに保存するのに必要な記憶域ロードのために、配列決定マシンのメモリ装置を利用して画像又はシグナルデータを記憶し処理することは、多くの場合、非実用的である。例示すると、従来のシステムは、多くの場合、サイクルごとにシグナルデータを収集し、データを配列決定装置上に記憶し、データをサーバーに転送し、データをサーバー内に記憶し、各サイクルからのデータをサーバー上で処理する。したがって、従来のシステムは、リソースを非効率的に利用するだけでなく、シグナル伝達データを転送し処理することによってかなりの待ち時間をもたらす。

【0006】

これらは、更なる問題及び課題とともに、既存の配列決定システムに存在する。

【発明の概要】

【0007】

本開示は、上記の問題のうちの1つ以上を解決するか、又は当技術分野に勝る他の利点を提供する、システム、方法、及び非一時的コンピュータ可読記憶媒体の1つ以上の実施形態を説明する。特に、開示されたシステムは、オリゴヌクレオチドの特定のクラスタに対するフェージング及びプリフェージングの効果を正確かつ効率的に推定し、クラスタに対するクラスタ固有のフェージング補正を決定することができる。例えば、開示されたシステムは、フェージング又はプリフェージングを頻繁に引き起こすエラー誘導配列を示すオリゴヌクレオチドのクラスタを動的に特定することができる。開示されたシステムが、そのようなエラー誘導配列に続くリード位置におけるシグナルをサイクル中に検出するとき、開示されたシステムは、クラスタ固有のフェージング係数を生成し、そのようなクラスタ固有のフェージング係数に従ってシグナルを補正することができる。例えば、開示されたシステムは、線形等化器、判定帰還型等化器、最尤系列推定器、又は機械学習モデルを利用して、クラスタ固有のフェージング係数を生成することができる。場合によっては

、開示されたシステムは、したがって、エラー誘導配列に続くリード位置を特定し、配列決定装置上でほぼリアルタイムでほとんど又は全くバッファリングせずに、クラスタ固有のフェージング係数を生成することができる。

【0008】

本開示の1つ以上の実施形態の追加の特徴及び利点は、以下の説明に記載され、一部は説明から明らかになるか、又はかかる例示的な実施形態の実施によって習得され得る。

【図面の簡単な説明】

【0009】

詳細な説明により、様々な実施形態が、以下に要約される添付の図面の使用を通じて更に具体的かつ詳細に説明される。

【図1】クラスタ認識塩基コーリングシステム(cluster-aware-base-calling system)が本開示の1つ以上の実施形態に従って動作することができる環境を示す。

【図2A】本開示の1つ以上の実施形態による、クラスタ固有のフェージング補正の前のフェージング及びプリフェージングから生じる不正確な塩基コールを示す、例示的リードパイプアップを示す。

【図2B】本開示の1つ以上の実施形態による、フェージング及びプリフェージングを示す概略図を示す。

【図3】本開示の1つ以上の実施形態による、クラスタ固有のフェージング補正を決定し、クラスタ固有のフェージング補正に基づいてシグナルを調整することに基づいてヌクレオチド塩基コールを決定する、クラスタ認識塩基コーリングシステムの概略図を示す。

【図4】本開示の1つ以上の実施形態による、前のサイクルからのシグナルを分析することに基づいてエラー誘導配列を特定する、クラスタ認識塩基コーリングシステムを示す。

【図5】本開示の1つ以上の実施形態による、クラスタ固有のフェージング係数及びクラスタ固有のプリフェージング係数を決定する、クラスタ認識塩基コーリングシステムを示す。

【図6】本開示の1つ以上の実施形態による、クラスタ固有のフェージング補正を推定するためにクラスタ認識塩基コーリングシステムが利用する、例示的なフェージングモデルを示す。

【図7A】本開示の1つ以上の実施形態による、クラスタ固有のフェージング補正を決定するために、線形等化器、判定帰還型等化器、及び最尤系列推定器を含む様々な受信機タイプを利用する、クラスタ認識塩基コーリングシステムを示す。

【図7B】本開示の1つ以上の実施形態による、クラスタ固有のフェージング補正を決定するために、線形等化器、判定帰還型等化器、及び最尤系列推定器を含む様々な受信機タイプを利用する、クラスタ認識塩基コーリングシステムを示す。

【図7C】本開示の1つ以上の実施形態による、クラスタ固有のフェージング補正を決定するために、線形等化器、判定帰還型等化器、及び最尤系列推定器を含む様々な受信機タイプを利用する、クラスタ認識塩基コーリングシステムを示す。

【図8A】クラスタ認識塩基コーリングシステムが、本開示の1つ以上の実施形態による、クラスタ固有のフェージング補正に基づいてシグナルを調整することによって、塩基コールの精度及び様々な二次配列決定メトリクスを改善することを示すメトリクスを示す、グラフを示す。

【図8B】クラスタ認識塩基コーリングシステムが、本開示の1つ以上の実施形態による、クラスタ固有のフェージング補正に基づいてシグナルを調整することによって、塩基コールの精度及び様々な二次配列決定メトリクスを改善することを示すメトリクスを示す、グラフを示す。

【図9】本開示の1つ以上の実施形態による、クラスタ固有のフェージング補正を決定し、クラスタ固有のフェージング補正に基づいてシグナルを調整することに基づいてヌクレオチド塩基コールを決定するための、一連の動作を示す。

【図10】本開示の1つ以上の実施形態による、例示的なコンピューティング装置のブロック図を示す。

10

20

30

40

50

【発明を実施するための形態】

【0010】

本開示は、クラスタごとにフェージング誤差を推定する、クラスタ認識塩基コーリングシステムの1つ以上の実施形態を説明する。特に、クラスタ認識塩基コーリングシステムは、シグナル劣化を頻繁に誘導する配列を特定する。例えば、クラスタ認識塩基コーリングシステムは、オリゴヌクレオチドのクラスタに対応するヌクレオチド断片リード内のホモポリマー配列、G-四重鎖配列、又は他のエラー誘導配列を特定することができる。クラスタ認識塩基コーリングシステムは、現在のサイクルからのヌクレオチド塩基についてのシグナルに対するフェージング及びプリフェージングの効果を推定する係数を更に決定することができる。クラスタ認識塩基コーリングシステムは、クラスタ固有のフェージング係数を利用して、ヌクレオチド塩基コールが行われるシグナル強度を補正する。推定されたフェージング又はプリフェージングをクラスタごとに補正することによって、クラスタ認識塩基コーリングシステムは、補正されたシグナル強度を分析して、より正確なヌクレオチド塩基コールを生成することができる。

10

【0011】

例示すると、1つ以上の実施形態では、クラスタ認識塩基コーリングシステムは、オリゴヌクレオチドのクラスタについて、1つ以上のヌクレオチド断片リード内のエラー誘導配列に続くリード位置を特定する。クラスタ認識塩基コーリングシステムは、リード位置に対応するサイクル中にオリゴヌクレオチドのクラスタ内の標識ヌクレオチド塩基からのシグナルを更に検出することができる。同じクラスタについて、クラスタ認識塩基コーリングシステムは、推定されたフェージング及び推定されたプリフェージングに対してシグナルを補正するクラスタ固有のフェージング補正を決定する。次いで、クラスタ認識塩基コーリングシステムは、クラスタ固有のフェージング補正に基づいてシグナルを調整することができる。調整されたシグナルに基づいて、クラスタ認識塩基コーリングシステムは、オリゴヌクレオチドのクラスタに対応するリード位置についてのヌクレオチド塩基コールを決定することができる。

20

【0012】

言及したように、場合によっては、クラスタ認識塩基コーリングシステムは、オリゴヌクレオチドのクラスタに対応する1つ以上のヌクレオチド断片リード内のエラー誘導配列に続くリード位置を特定する。そのようなエラー誘導配列は、配列決定ランのクオリティ及び精度に悪影響を及ぼす、系統的な配列決定エラーをトリガーし得る。クラスタ固有のフェージング補正が決定されるクラスタの数を低減するために、いくつかの実施形態では、クラスタ認識塩基コーリングシステムは、エラー誘導配列に続くクラスタのリード位置に対してのみそのようなクラスタ固有のフェージング補正を決定することによって、フェージング補正に使用されるコンピューティングリソースを制限する。エラー誘導配列の例としては、ホモポリマーなどの1つ以上の反復ヌクレオチド塩基、又はグアニン四重鎖などの配列モチーフを挙げることができる。クラスタ認識塩基コーリングシステムは、前の配列決定サイクルからのオリゴヌクレオチドのクラスタからのシグナルを分析して、クラスタに対応するヌクレオチド断片リード内のエラー誘導配列の存在を決定することができる。

30

40

【0013】

オリゴヌクレオチドのクラスタに対応するエラー誘導配列を特定した後又は特定している間、クラスタ認識塩基コーリングシステムは、リード位置に対応するサイクル中にオリゴヌクレオチドのクラスタ内の標識ヌクレオチド塩基からのシグナルを検出することができる。言及したように、SBS配列決定システムは、標識ヌクレオチド塩基がクラスタのオリゴヌクレオチドに反復して組み込まれる際に、標識ヌクレオチド塩基から照射された蛍光タグの画像を捕捉する。クラスタ認識塩基コーリングシステムは、エラー誘導配列に続く1つ以上のリード位置に対応するサイクルに固有に、標識ヌクレオチド塩基からのシグナルを検出し、そのようなシグナルをクラスタ固有のフェージング補正の標的として特定することができる。

50

【0014】

エラー誘導配列に続く関連するリード位置に対応するシグナルを特定した後、クラスタ認識塩基コーリングシステムは、クラスタ固有のフェージング補正を決定して、推定されたフェージング及び推定されたプリフェージングについてシグナルを補正することができる。言及したように、系統的配列決定エラーは、ヌクレオチド塩基がそれぞれ後期又は初期に組み込まれる、フェージング及びプリフェージングを含み得る。いくつかの実施形態では、クラスタ認識塩基コーリングシステムは、(i) 1つ以上の前のサイクルについてのヌクレオチド塩基に対応する1つ以上のクラスタ固有のフェージング係数、及び(ii) 1つ以上の後続のサイクルについてのヌクレオチド塩基に対応する1つ以上のクラスタ固有のプリフェージング係数を決定することによって、クラスタ固有のフェージング補正を決定する。クラスタ認識塩基コーリングシステムは、クラスタ固有のフェージング係数及びクラスタ固有のプリフェージング係数に基づいて、クラスタ固有のフェージング補正を更に決定することができる。

10

【0015】

そのようなクラスタ固有のフェージング及びプリフェージング係数を決定するために、クラスタ認識塩基コーリングシステムは、いくつかのモデル又はアルゴリズムを利用することができる。例えば、場合によっては、クラスタ認識塩基コーリングシステムは、リアルタイム線形等化器を利用して、クラスタ固有のフェージング係数及びクラスタ固有のプリフェージング係数を推定する。線形等化器は、計算上効率的であり、代替係数アルゴリズムと比較してバッファリングをほとんど又は全く必要としない。したがって、クラスタ認識塩基コーリングシステムは、配列決定装置上で線形等化器を実装して、クラスタ固有のフェージング補正をリアルタイムで推定することができる。代替として、いくつかの実施形態では、クラスタ認識塩基コーリングシステムは、クラスタ固有のフェージング補正を推定するために、線形等化器の代わりに、又はそれに加えて、判定帰還型等化器、最尤等化器、又は機械学習モデルを利用する。

20

【0016】

クラスタ固有のフェージング補正を決定した後、クラスタ認識塩基コーリングシステムは、クラスタ固有のフェージング補正に基づいてシグナルを調整することができる。特に、クラスタ認識塩基コーリングシステムは、エラー誘導配列を有するクラスタに対するクラスタ固有のフェージング補正を推定し、クラスタ固有のフェージング補正をクラスタからのシグナルに適用する。いくつかの実施形態では、クラスタ認識塩基コーリングシステムはまた、クラスタのセットについて、クラスタのセットにわたる配列決定エラーを補正するためにマルチクラスタフェージング補正を決定する。そのようなマルチクラスタフェージング補正は、例えば、フローセルのタイル内のクラスタのための大域的フェージング補正の一部として、大域的フェージング係数及び大域的プリフェージング係数を含んでもよい。クラスタ認識塩基コーリングシステムはまた、クラスタ固有のフェージング補正とマルチクラスタフェージング補正との組み合わせに基づいて、クラスタのシグナルを調整することができる。

30

【0017】

クラスタ認識塩基コーリングシステムは、既存の配列決定システムに対していくつかの技術的利点を提供する。特に、クラスタ認識塩基コーリングシステムは、既存の配列決定システムに対して、フェージング補正の精度、調整された適用性、及び効率を改善することができる。言及したように、クラスタ認識塩基コーリングシステムは、既存の配列決定システムよりも高い精度で、そのような補正されたシグナルに基づいて、シグナル及びヌクレオチド塩基コールについてのフェージング補正の両方を決定する。クラスタに対応するある特定のリード位置についてのシグナルに対してクラスタ固有のフェージング補正を決定し適用することによって、クラスタ認識塩基コーリングシステムは、予測されるヌクレオチド塩基コールの精度に対するホモポリマー配列、G-四重鎖配列、又は他のエラー誘導配列の負の影響を低減することができる。更に、推定されたフェージング及びプリフェージングのためのシグナルをクラスタごとに調整することによって、クラスタ認識塩基

40

50

コーリングシステムは、オリゴヌクレオチドの特定のクラスタの組み込まれたヌクレオチド塩基からのシグナルにおけるフェージング又はプリフェージング効果によって引き起こされるノイズの量を低減することができる。簡単に言えば、クラスタ認識塩基コーリングシステムは、既存の配列決定システムよりも良好に、特定のクラスタについてのフェージング効果及びプリフェージング効果を特定し補正することができる。

【0018】

以下に更に示されるように、ヌクレオチド塩基コールを生成するために使用されるシグナルを補正することによって、クラスタ認識塩基コーリングシステムはまた、塩基コールデータのためのより良好なクオリティメトリクスなどの二次配列決定メトリクスを改善し、例えば、シグナル対ノイズ比 (signal to noise ratio、SNR) メトリクスを改善することによって、配列決定装置のためのメトリクスを推定又は較正するためのベースラインを改善する。クラスタ固有のフェージング補正は、ヌクレオチド塩基コールを生成するために使用されるシグナルを改善するため、クラスタ認識塩基コーリングシステムはまた、コール生成モデル (例えば、DRAGEN) のマップー及びアラインメント構成要素又はコール生成モデルのバリエーションコーラー構成要素などの下流ヌクレオチド塩基コーリングツールの性能に悪影響を及ぼすように次々に複合する相関エラー誘導配列 (例えば、系統的配列決定エラーをトリガーする配列) の影響を低減することができる。

【0019】

より正確であることに加えて、クラスタ認識塩基コーリングシステムは、既存の配列決定システムよりもクラスタ固有の配列決定エラーに合わせて調整されたフェージング補正を生成する。オリゴヌクレオチドのクラスタ群又は全てのクラスタにわたってフェージング補正を適用する既存のシステムとは対照的に、クラスタ認識塩基コーリングシステムは、クラスタ固有のフェージング係数を決定する。実際、場合によっては、クラスタ認識塩基コーリングシステムは、選択的に決定し、クラスタ固有のフェージング補正をある特定のクラスタについてのポストエラー誘導配列リード位置におけるシグナルに適用し、マルチクラスタフェージング補正 (クラスタ固有のフェージング補正を伴わない) をそのようなエラー誘導配列を欠くある特定の他のクラスタについてのリード位置におけるシグナルに適用する。したがって、配列決定が進行するにつれてクラスタがより問題になり得る場合であっても (配列決定ラン中にフェージング効果及びプリフェージング効果が増加する傾向がある場合)、クラスタ認識塩基コーリングシステムは、クラスタ固有のフェージング補正を調整して、ヌクレオチド塩基コールに対する対応する調整を行う。

【0020】

上記で示されるように、いくつかの実施形態では、クラスタ認識塩基コーリングシステムは、フェージング補正のための代替計算モデルに対して、フェージング効果及びプリフェージング効果のためにシグナルを補正するコンピューティング効率を改善することができる。全てのサイクルにわたって各クラスタのフェージング及びプリフェージングを処理し補正する計算モデルとは対照的に、クラスタ認識塩基コーリングシステムは、エラー誘導配列に続く標識ヌクレオチド塩基からのシグナルを処理し補正することによって利用されるコンピューティングリソースの量を低減する。言及したように、いくつかの実施形態では、クラスタ認識塩基コーリングシステムは、エラー誘導配列に続くクラスタのリード位置についてのみクラスタ固有のフェージング補正を決定することによって、フェージング補正に使用されるコンピューティングリソースを制限する。

【0021】

更に、線形等化器ベースの手法を利用してフェージング補正を決定することによって、場合によっては、クラスタ認識塩基コーリングシステムは、配列決定装置上でリアルタイム (又はほぼリアルタイム) でクラスタ固有のフェージング補正を推定することができる。いくつかの既存の配列決定システムは、配列決定ラン全体のために全てのクラスタのシグナルの画像データを保存し、配列決定ランが終了した後にのみフェージング補正を決定することによって、配列決定マシン (又は他のコンピューティング装置) 上で著しく多くのコンピューティングメモリを消費する。対照的に、ある特定の実施形態では、クラスタ

認識塩基コーリングシステムは、クラスタ固有のフェージング補正及び／又はマルチクラスタフェージング補正を適用した後に、シグナルのデータを破棄する。少なくとも1つの実施形態では、配列決定装置上のフェージング効果及びプリフェージング効果のためにシグナルを処理し補正することによって、クラスタ認識塩基コーリングシステムは、データを中央位置に通信し、データを処理し、結果を通信するために典型的に必要とされる記憶、通信、及びコンピューティングリソースの量を低減することができる。

【0022】

前述の議論によって示されるように、本開示は、クラスタ認識塩基コーリングシステムの特徴及び利点を説明するために、種々の用語を利用する。ここで、かかる用語の意味に関して更なる詳細を提供する。例えば、本明細書で使用される場合、「クラスタ」という用語は、ヌクレオチド試料スライド上に組織化された試料ゲノム由来のオリゴヌクレオチド又は核酸セグメントの群を指す。特に、クラスタは、クローン化された又は同じDNA又はRNAセグメントの数十、数百、数千、又はそれ以上のコピーを含む。例えば、1つ以上の実施形態では、クラスタは、ヌクレオチド試料スライド（例えば、フローセル）のセクション（切片）に固定化されたオリゴヌクレオチドの分類を含む。いくつかの実施形態では、クラスタは、パターン化されたヌクレオチド試料スライド内で、均一に間隔を置かれるか、又は系統的構造に組織化される。対照的に、場合によっては、クラスタは、パターン化されていないヌクレオチド試料スライド内でランダムに組織化される。

【0023】

本明細書で使用される場合、「オリゴヌクレオチド」という用語は、ヌクレオチド又は模倣物のオリゴマー又は他のポリマーを指す。特に、オリゴヌクレオチドは、ヌクレオチド中のペントースの3'位と隣接するヌクレオチド中のペントースの5'位との間の修飾ホスホジエステル又はホスホジエステル結合によって形成された共有結合ヌクレオチドの配列を含む合成分子又は天然分子を含み得る。例えば、オリゴヌクレオチドは、SBS配列決定の一部として分析又は配列決定される一本鎖ポリヌクレオチドにアニールされた短いDNA又はRNA分子を含むことができる。

【0024】

本明細書で更に使用される場合、「ヌクレオチド試料スライド」という用語は、試料ゲノム又は他の試料核酸ポリマーのヌクレオチドセグメントを配列決定するためのオリゴヌクレオチドを含むプレート又はスライドを指す。特に、ヌクレオチド試料スライドは、配列決定の一部として試薬及び緩衝液が移動することができる流体チャネルを含有するスライドを指す。例えば、1つ以上の実施形態では、ヌクレオチド試料スライドは、小さな流体チャネル及びアダプター配列に相補的な短いオリゴヌクレオチドを含むフローセル（例えば、パターン化されたフローセル又はパターン化されていないフローセル）を含む。上記のように、ヌクレオチド試料スライドは、オリゴヌクレオチドのクラスタを含むウェル（例えば、ナノウェル）を含むことができる。

【0025】

本明細書で使用される場合、フローセル又は他のヌクレオチド試料スライドは、(i) 反応構造の上に延びて、反応構造の複数の反応部位と通信しているフローチャネルをそれらの間に形成する蓋を有する装置を含むことができ、(ii) 反応部位で、又は反応部位の近くで起こる指定された反応を検出するように構成された検出装置を含むことができる。フローセル又は他のヌクレオチド試料スライドは、電荷結合素子 (Charge-Coupled Device、CCD) 又は相補型金属酸化膜半導体 (Complementary Metal-Oxide Semiconductor、CMOS) (光) 検出装置などの、固体光検出装置又は「撮像」装置を含み得る。1つの具体的な例として、フローセルは、流体的に構成し、バイオアッセイシステムに流体的及び／又は電気的に結合するように構成することができるカートリッジ (一体型ポンプを有する) に電気的に結合することができる。カートリッジ及び／又はバイオアッセイシステムは、所定のプロトコル (例えば、合成による配列決定) に従ってフローセルの反応部位に反応溶液を送達し、複数の撮像事象を実行することができる。例えば、カートリッジ及び／又はバイオアッセイシステムは、フローセルのフローチャネルを通

10

20

30

40

50

して、それによって反応部位に沿って、1つ以上の反応溶液を導くことができる。反応溶液のうちの少なくとも1つは、同じ又は異なる蛍光標識を有する4種類のヌクレオチドを含んでもよい。ヌクレオチドは、反応部位の対応するオリゴヌクレオチドなどの、フローセルの反応部位に結合させることができる。次いで、カートリッジ及び／又はバイオアクセスシステムは、励起光源（例えば、発光ダイオード（light-emitting diode、LED）などの固体光源）を使用して、反応部位を照明することができる。励起光は、フローセルの光センサーによって検出され得る発光シグナル（例えば、励起光とは異なる、かつ潜在的には互いに異なる、1つ又は複数の波長の光）を提供することができる。

【0026】

本明細書で使用される場合、「リード位置」という用語は、ヌクレオチド断片リード上の位置又は座標を指す。特に、リード位置は、標識ヌクレオチドが付加されたヌクレオチド断片リードに沿った位置を含む。例えば、リード位置は、カメラがヌクレオチド試料スライド又はヌクレオチド試料スライドの切片の画像を捕捉するときに、クラスタ内の対応するオリゴヌクレオチドに最も最近追加された標識ヌクレオチドがあるヌクレオチド断片リード内の位置を示すことができる。

【0027】

本明細書で使用される場合、「ヌクレオチド断片リード」という用語は、試料ヌクレオチド配列の全部又は一部からの1つ以上のヌクレオチド塩基（又は核酸塩基対）の推定配列を指す。特に、ヌクレオチド断片リードは、ゲノム試料に対応する配列決定ライブラリからのヌクレオチド断片（又はモノクローナルヌクレオチド断片の群）に対するヌクレオチド塩基コールの決定又は予測された配列を含む。例えば、場合によっては、配列決定装置は、ヌクレオチド試料スライドのナノポアを通過した、蛍光タグ付けを介して決定した、又はフローセル内のクラスタから決定されたヌクレオチド塩基に対するヌクレオチド塩基コールを生成することによって、ヌクレオチド断片リードを決定する。

【0028】

本明細書で使用される場合、「エラー誘導配列」という用語は、配列決定エラーを誘導又はトリガーするヌクレオチド塩基配列又は対応する化学構造を指す。特に、エラー誘導配列は、SBS配列決定中に系統的配列決定エラー（systematic sequencing errors、SSE）をトリガーするヌクレオチド塩基配列を指す。例えば、エラー誘導配列は、誤ったサイクルで不正確な標識ヌクレオチド塩基を付加又は組み込むように配列決定装置を誘導することによって、ディフェージングを引き起こし得る。例えば、エラー誘導配列は、同じヌクレオチド塩基のホモポリマー、グアニン四重鎖、可変数タンデム反復（variable number tandem repeat、VNTR）、ジヌクレオチド反復配列、トリヌクレオチド反復配列、逆方向反復配列、ミニサテライト配列、マイクロサテライト配列、パンドローム配列、又は他の配列を含み得る。

【0029】

本明細書で使用される場合、「シグナル」という用語は、標識ヌクレオチド塩基又は標識ヌクレオチド塩基の群（例えば、オリゴヌクレオチドのクラスタに付加された標識ヌクレオチド塩基）から放出されるか、反射されるか、又は別様に通信されるシグナルを指す。特に、シグナルは、ヌクレオチド塩基の型を示すシグナルを指すことができる。例えば、シグナルは、ヌクレオチド塩基の蛍光タグ又はオリゴヌクレオチドに組み込まれた複数のヌクレオチド塩基の蛍光タグから放出又は反射される光シグナルを含むことができる。いくつかの実装形態では、クラスタ認識塩基コーリングシステムは、レーザー又は他の光源などの外部刺激を介してシグナルをトリガーする。場合によっては、クラスタ認識塩基コーリングシステムは、いくつかの内部刺激を介してシグナルをトリガーする。更に、いくつかの実施形態では、クラスタ認識塩基コーリングシステムは、ヌクレオチド試料スライド（例えば、ヌクレオチド試料スライドの切片）の画像を捕捉するときに適用されるフィルタを使用してシグナルを観察する。上記で示唆したように、ある特定の例において、シグナルは、オリゴヌクレオチドのクラスタ中の個々のオリゴヌクレオチドに付加された各標識ヌクレオチド塩基によって提供されるシグナルの集合体を含む。

10

20

30

40

50

【0030】

本明細書で使用される場合、「標識ヌクレオチド塩基」という用語は、ヌクレオチド塩基の分類の蛍光又は光ベースの指標を有するヌクレオチド塩基を指す。特に、標識ヌクレオチド塩基は、ヌクレオチド塩基の型（例えば、アデニン、シトシン、チミン、又はグアニン）を特定するための蛍光又は光ベースのインジケータを組み込むヌクレオチド塩基を指すことができる。例えば、1つ以上の実施形態では、標識ヌクレオチド塩基は、ヌクレオチド塩基型を特定するシグナルを放出する蛍光タグを有するヌクレオチド塩基を含む。

【0031】

本明細書で使用される場合、「配列決定サイクル」（又は「サイクル」）という用語は、オリゴヌクレオチドにヌクレオチド塩基を付加若しくは組み込む反復、又は並行してオリゴヌクレオチドにヌクレオチド塩基を付加若しくは組み込む反復を指す。特に、サイクルは、オリゴヌクレオチドに、又は並行してオリゴヌクレオチドに付加又は組み込まれた個々のヌクレオチド塩基を示すデータを用いて1つ以上の画像を分析する反復を含むことができる。したがって、核酸ポリマー（例えば、試料ゲノム）の配列決定の一部としてサイクルを繰り返すことができる。例えば、1つ以上の実施形態では、各配列決定サイクルは、DNA鎖若しくはRNA鎖が単一方向のみで読み取られる単一ヌクレオチド断片リード、又はDNA鎖若しくはRNA鎖が両端から読み取られるペアエンドリードのいずれかを伴う。更に、ある特定の場において、各配列決定サイクルは、特定のオリゴヌクレオチドに付加又は組み込まれた特定のヌクレオチド塩基を決定するための画像データを生成するために、ヌクレオチド試料スライド又はヌクレオチド試料スライドの複数のセクションの画像を撮影するカメラを伴う。画像撮影段階に続いて、配列決定システムは、組み込まれたヌクレオチド塩基からある特定の蛍光標識を除去し、核酸ポリマーが完全に配列決定されるまで別の配列決定サイクルを実施することができる。1つ以上の実施形態では、配列決定サイクルは、合成による配列決定（SBS）ラン内のサイクルを含む。

【0032】

本明細書で使用される場合、「クラスタ固有のフェージング補正」という用語は、適用されたときに、オリゴヌクレオチドの特定のクラスタ内の標識ヌクレオチド塩基からのシグナルを調整して、推定されたフェージング又はプリフェージングを補正するプロセス又は機能を指す。特に、クラスタ固有のフェージング補正は、フーリエ変換を使用して推定されたフェージング又はプリフェージングの推定された効果を補正するためにクラスタからのシグナルが調整されるアルゴリズム又は機能を含むことができる。

【0033】

本明細書で使用される場合、「フェージング」という用語は、標識ヌクレオチド塩基が特定の配列決定サイクルの後に組み込まれるインスタンス（又はその速度）を指す。フェージングは、クラスタ内の標識ヌクレオチド塩基が、特定の配列決定サイクルについてクラスタ内の他の標識ヌクレオチド塩基の後ろに非同期的に組み込まれるインスタンス（又はその速度）を含む。特に、SBSの間、クラスタ中の各DNA鎖は、1サイクル当たり1ヌクレオチド塩基だけ組み込みを延長する。クラスタ内の1つ以上のオリゴヌクレオチド鎖は、現在のサイクルと位相がずれる可能性がある。フェージングは、クラスタ内の1つ以上のオリゴヌクレオチドのヌクレオチド塩基が、1つ以上の組み込みサイクルに遅れる場合に起こる。例えば、第1の位置から第3の位置までのヌクレオチド配列は、CTAであり得る。この例では、Cヌクレオチドは、第1のサイクルで組み込まれ、Tは、第2のサイクルで組み込まれ、Aは、第3のサイクルで組み込まれるべきである。フェージングが第2の配列決定サイクル中に起こる場合、1つ以上の標識Cヌクレオチドが、Tヌクレオチドの代わりに組み込まれる。関連して、本明細書で使用される場合、「プリフェージング」という用語は、1つ以上のヌクレオチド塩基が特定のサイクルの前に組み込まれるインスタンス（又はその速度）を指す。プリフェージングは、クラスタ内の標識ヌクレオチド塩基が、特定の配列決定サイクルについてクラスタ内の他の標識ヌクレオチド塩基の後ろに非同期的に組み込まれるインスタンス（又はその速度）を含む。例示すると、上記の例において第2の配列決定サイクルの間にプリフェージングが起こる場合、1つ以

上の標識 A ヌクレオチドが、T ヌクレオチドの代わりに組み込まれる。

【0034】

本明細書で使用される場合、「クラスタ固有のフェージング係数」という用語は、クラスタのシグナルに対するクラスタ固有のフェージングを推定又は測定する係数又は値を指す。特に、クラスタ固有のフェージング係数は、所与の配列決定サイクル内のクラスタに対するフェージングの効果を推定する。例えば、クラスタ固有のフェージング係数は、前のサイクルについてのヌクレオチド塩基が、現在のサイクルについての標識ヌクレオチド塩基からのシグナルに対して有する効果を示すことができる。例示すると、上記の例において、クラスタ固有のフェージング係数は、第2の配列決定サイクル中にTヌクレオチドの代わりに組み込まれるCヌクレオチドからフェージングの効果を推定することができる。

10

【0035】

関連して、「クラスタ固有のプリフェージング係数」という用語は、クラスタのシグナルに対するクラスタ固有のプリフェージングを推定又は測定する係数又は値を指す。特に、クラスタ固有のプリフェージング係数は、所与の配列決定サイクル内のクラスタに対するプリフェージングの効果を推定する。例えば、クラスタ固有のプリフェージング係数は、後続のサイクルについてのヌクレオチド塩基が、現在のサイクルについての標識ヌクレオチド塩基からのシグナルに対して有する効果を示すことができる。例示すると、上記の例において、クラスタ固有のプリフェージング係数は、第2の配列決定サイクル中にTヌクレオチドの代わりに組み込まれるAヌクレオチドからプリフェージングの効果を推定する。

20

【0036】

本明細書で使用される場合、「ヌクレオチド塩基コール」（又は単に「塩基コール」）という用語は、配列決定サイクル中の試料ゲノムのゲノム座標又はオリゴヌクレオチドについての特定のヌクレオチド塩基（又はヌクレオチド塩基対）の決定又は予測を指す。特に、ヌクレオチド塩基コールは、(i)ヌクレオチド試料スライド上のオリゴヌクレオチド内に組み込まれたヌクレオチド塩基の型の決定若しくは予測（例えば、リードベースのヌクレオチド塩基コール）、又は(ii)デジタル出力ファイルにおけるバリエーションコール若しくは非バリエーションコールを含む、ゲノム内のゲノム座標若しくは領域に存在するヌクレオチド塩基の型の決定若しくは予測を示すことができる。場合によっては、ヌクレオチド断片リードについて、ヌクレオチド塩基コールは、ヌクレオチド試料スライド（例えば、フローセルのクラスタ内）のオリゴヌクレオチドに付加された蛍光タグ付きヌクレオチドから得られる強度値に基づくヌクレオチド塩基の決定又は予測を含む。あるいは、ヌクレオチド塩基コールは、ヌクレオチド試料スライドのナノポアを通過するヌクレオチドから生じるクロマトグラムピーク又は電流変化からのヌクレオチド塩基の決定又は予測を含む。対照的に、ヌクレオチド塩基コールはまた、ゲノム座標に対応するヌクレオチド断片リードに基づく、バリエーションコールファイル又は他の塩基コール出力ファイルのための、試料ゲノムのゲノム座標におけるヌクレオチド塩基の最終予測を含み得る。したがって、ヌクレオチド塩基コールは、ゲノム座標及び参照ゲノムに対応する塩基コール、例えば、参照ゲノムに対応する特定の位置におけるバリエーション又は非バリエーションの表示を含むことができる。実際、ヌクレオチド塩基コールは、単一ヌクレオチドバリエーション(single nucleotide variant、SNV)、挿入若しくは欠失(インデル)、又は構造バリエーションの一部である塩基コールを含むがこれらに限定されないバリエーションコールを指すことができる。上記で示唆したように、単一ヌクレオチド塩基コールは、アデニン(A)コール、シトシン(C)コール、グアニン(G)コール、又はチミン(T)コールであり得る。

30

40

【0037】

ここで、クラスタ認識塩基コーリングシステムの例示的な実施形態及び実装形態を示す例示的な図に関連して、クラスタ認識塩基コーリングシステムに関する更なる詳細を提供する。例えば、図1は、クラスタ認識塩基コーリングシステム106が1つ以上の実施形態に従って動作するシステム環境（又は「環境」）100の概略図を示す。図示されるよ

50

うに、環境 100 は、ネットワーク 112 を介してユーザクライアント装置 108 及び配列決定装置 114 に接続された 1 つ以上のサーバ装置 102 を含む。図 1 はクラスタ認識塩基コーリングシステム 106 の一実施形態を示すが、代替的な実施形態及び構成が可能である。

【0038】

図 1 に更に示されるように、サーバ装置 102、ユーザクライアント装置 108、及び配列決定装置 114 は、ネットワーク 112 を介して接続される。環境 100 の構成要素の各々は、ネットワーク 112 を介して通信することができる。ネットワーク 112 は、コンピューティング装置が通信することができる任意の適切なネットワークを含む。例示的なネットワークを、図 10 に関連して以下で更に詳細に説明する。

10

【0039】

図 1 に示されるように、環境 100 は、配列決定装置 114 を含む。配列決定装置 114 は、全ゲノム又は他の核酸ポリマーを配列決定するための装置を含む。いくつかの実施形態では、配列決定装置 114 は、配列決定装置 114 上で直接的又は間接的のいずれかで本明細書に記載されるコンピュータ実装方法及びシステムを利用して、試料を分析してデータを生成する。1 つ以上の実施形態では、配列決定装置 114 は、合成による配列決定 (SBS) を利用して、全ゲノム又は他の核酸ポリマーを配列決定する。示されるように、いくつかの実施形態では、配列決定装置 114 は、ネットワーク 112 をバイパスし、ユーザクライアント装置 108 と直接通信する。

【0040】

図 1 によって更に示されるように、環境 100 は、サーバ装置 102 を含む。サーバ装置 102 は、核酸ポリマーを配列決定するためのデータなどの電子データを生成、受信、分析、記憶、受信、及び転送することができる。サーバ装置 102 は、配列決定装置 114 からデータを受信することができる。例えば、サーバ装置 102 は、ヌクレオチド塩基コールデータ、クオリティデータ、及び核酸ポリマーの配列決定に関連する他のデータを含む配列決定データを収集及び／又は受信することができる。サーバ装置 102 は、ユーザクライアント装置 108 とも通信することができる。特に、サーバ装置 102 は、核酸ポリマー配列、エラーデータ、及び他の情報をユーザクライアント装置 108 に送信することができる。いくつかの実施形態では、サーバ装置 102 は、分散型サーバを含み、サーバ装置 102 は、ネットワーク 112 にわたって分散され、異なる物理的場所に位置する、いくつかのサーバ装置を含む。サーバ装置 102 は、コンテンツサーバ、アプリケーションサーバ、通信サーバ、ウェブホスティングサーバ、又は別のタイプのサーバを含むことができる。

20

30

【0041】

図 1 に更に示されるように、サーバ装置 102 は、配列決定システム 104 を含むことができる。概して、配列決定システム 104 は、配列決定装置 114 から受信した配列決定データを分析して、全ゲノム又は核酸ポリマーについてのヌクレオチド配列を決定する。例えば、配列決定システム 104 は、配列決定装置 114 から生データ (例えば、ヌクレオチド断片リードの塩基コールデータ) を受信し、試料ゲノムの核酸配列を決定することができる。例示すると、配列決定システム 104 は、配列決定装置 114 からヌクレオチド断片リードを受信することができ、配列決定システム 104 は、ヌクレオチド断片リードから試料ゲノムに対するヌクレオチド塩基コールを生成する。いくつかの実施形態では、配列決定システム 104 は、DNA 及び／又は RNA 中のヌクレオチド塩基の配列を決定する。核酸ポリマーについての配列を処理及び決定することに加えて、配列決定システム 104 はまた、配列決定データを分析して、個々の又は複数の配列決定サイクルにおける不規則性を検出する。

40

【0042】

図 1 に図示されるように、配列決定装置 114 は、クラスタ認識塩基コーリングシステム 106 を含む。概して、クラスタ認識塩基コーリングシステム 106 は、推定されたフェージング及びプリフェージングのためにシグナルを補正するクラスタ固有のフェージン

50

グ補正を推定する。より具体的には、いくつかの実施形態では、クラスタ認識塩基コーリングシステム106は、1つ以上のヌクレオチド断片リード内のエラー誘導配列に続くリード位置を特定する。クラスタ認識塩基コーリングシステム106は、リード位置に対応するサイクル中にオリゴヌクレオチドのクラスタ内の標識ヌクレオチド塩基からのシグナルを更に検出する。クラスタ認識塩基コーリングシステム106は、推定されたフェージング及び推定されたプリフェージングのためにシグナルを補正するクラスタ固有のフェージング補正を決定する。クラスタ認識塩基コーリングシステム106は、クラスタ固有のフェージング補正に基づいてシグナルを調整し、調整されたシグナルに基づいてオリゴヌクレオチドのクラスタに対応するリード位置についてのヌクレオチド塩基コールを決定する。

10

【0043】

図1に図示される環境100は、ユーザクライアント装置108を更に含む。ユーザクライアント装置108は、デジタルデータを生成、記憶、受信、及び送信することができる。特に、ユーザクライアント装置108は、配列決定装置114から配列決定データを受信することができる。更に、ユーザクライアント装置108は、サーバー装置102と通信して、ヌクレオチド塩基コール、ヌクレオチド配列、及び配列決定ラン内の不規則性の報告を受信することができる。ユーザクライアント装置108は、ユーザクライアント装置108に関連付けられたユーザに配列決定データを提示することができる。

【0044】

図1に図示するユーザクライアント装置108は、様々なタイプのクライアント装置を含むことができる。例えば、いくつかの実施形態では、ユーザクライアント装置108は、デスクトップコンピュータ若しくはサーバー、又は他のタイプのクライアント装置等の非モバイル装置を含む。更に他の実施形態では、ユーザクライアント装置108は、例えば、ラップトップ、タブレット、携帯電話、スマートフォン等のモバイル装置を含む。ユーザクライアント装置108に関する更なる詳細は、図10に関して以下で説明される。

20

【0045】

図1に更に図示されるように、ユーザクライアント装置108は、配列決定アプリケーション110を含む。配列決定アプリケーション110は、ユーザクライアント装置108上のウェブアプリケーション又はネイティブアプリケーション（例えば、モバイルアプリケーション、デスクトップアプリケーション等）であってもよい。配列決定アプリケーション110は、（実行されると）ユーザクライアント装置108に、クラスタ認識塩基コーリングシステム106からデータを受信又は要求させ、配列決定データを提示させる命令を含むことができる。更に、配列決定アプリケーション110は、（実行されると）ユーザクライアント装置108に、試料ゲノムについてのリードパイルアップ又はリードアライメントのグラフィカル視覚化を提供させる命令を含むことができる。

30

【0046】

図1に更に図示されるように、クラスタ認識塩基コーリングシステム106は、配列決定アプリケーション110の一部として、ユーザクライアント装置108上に位置してもよい。図示されるように、いくつかの実施形態では、クラスタ認識塩基コーリングシステム106は、ユーザクライアント装置108上に（例えば、完全に又は部分的に位置して）実装される。更に他の実施形態では、クラスタ認識塩基コーリングシステム106は、環境100の1つ以上の他の構成要素によって実装される。特に、クラスタ認識塩基コーリングシステム106は、サーバー装置102、ユーザクライアント装置108、及び配列決定装置114にわたって様々な異なる方法で実装することができる。一実施例では、クラスタ認識塩基コーリングシステム106は、配列決定装置114上に部分的に位置し、サーバー装置102上にも位置する。特に、クラスタ認識塩基コーリングシステム106は、配列決定装置114上のクラスタ固有のフェージング補正に基づいてシグナルを調整し、サーバー装置102の一部としての、調整されたシグナルに基づいてオリゴヌクレオチドのクラスタに対応するリード位置についてのヌクレオチド塩基コールを決定することができる。

40

50

【0047】

図1は、ネットワーク112を介して通信する環境100の構成要素を図示しているが、いくつかの実装形態では、環境100の構成要素は、ネットワークを迂回して互いに直接通信する。例えば、前述したように、ユーザクライアント装置108は、配列決定装置114と直接通信することができる。更に、ユーザクライアント装置108は、ネットワーク112を迂回して、クラスタ認識塩基コーリングシステム106と直接通信することができる。更に、クラスタ認識塩基コーリングシステム106は、サーバー装置102又は環境100内の他の場所に収容された1つ以上のデータベースにアクセスすることができる。

【0048】

前述したように、クラスタ認識塩基コーリングシステム106は、推定されたフェージング及び推定されたプリフェージングのためにシグナルを補正するクラスタ固有のフェージング補正を決定することができる。以下の図及び説明は、いくつかの実装形態に従って、クラスタ認識塩基コーリングシステム106がクラスタ固有のフェージング補正をどのように推定するかに関する更なる詳細を提供する。特に、図2Aは、1つ以上の実施形態による、エラー誘導配列によるフェージング及びプリフェージングの効果を実証する、いくつかのヌクレオチド断片リードを含む例示的なリードパイルアップを図示する。対照的に、図2Bは、1つ以上の実施形態による、フェージング及びプリフェージングが分子レベルでどのように起こるかを図示する。

【0049】

言及したように、図2Aは、1つ以上の実施形態による、塩基コール精度及び二次配列決定メトリクスに対するエラー誘導配列の効果を反映する例示的なリードパイルアップを図示する。特に、図2Aは、ホモポリマー206を有する参照ゲノム212についてのヌクレオチド断片リード202を含むリードパイルアップ200を図示する。図2Aはまた、リードパイルアップ200のヌクレオチド断片リード202に対応する塩基クオリティ204、塩基深度208、及びエラータイプカウンタ210を示す。

【0050】

言及したように、リードパイルアップ200は、いくつかの配列決定サイクルに関するデータを反映する。特に、塩基深度208は、ヌクレオチド断片リード202内のいくつかのリードが各塩基をカバーするかを反映する。例えば、塩基深度208は、順方向ヌクレオチド断片リードと逆方向ヌクレオチド断片リード202との間で最大の重複を有する塩基をカバーするリードの数がより多いことを示す、薄灰色のバーを含む。例示すると、リードパイルアップ200の中心の塩基は、最大数のリードに対応する。

【0051】

図2Aに図示されるように、リードパイルアップ200は、ヌクレオチド断片リード202を含む。概して、ヌクレオチド断片リード202は、ゲノム内の様々なDNA断片の配列を示す。前述したように、いくつかの実装形態では、クラスタ認識塩基コーリングシステム106は、配列決定装置114を利用して、ヌクレオチド断片リード202を生成することができる。そのような配列決定の間、クラスタ認識塩基コーリングシステム106は、それぞれのクラスタのオリゴヌクレオチドに組み込まれた標識ヌクレオチド塩基に基づいて、ヌクレオチド断片リード202の各々を決定することができる。クラスタ認識塩基コーリングシステム106は更に、参照ゲノム212に沿ってヌクレオチド断片リード202を整列させて、参照ゲノム212に対するヌクレオチド塩基コールを決定する。

【0052】

図2Aに更に図示されるように、リードパイルアップ200は、ヌクレオチド断片リード202のリード方向及びエラーを示す。例えば、ヌクレオチド断片リード202の末端の矢印によって図示されるように、1～10と標識されたヌクレオチド断片リード202は、逆方向のサイクルによって付加される標識ヌクレオチド塩基を含む。11～20と標識されたヌクレオチド断片リード202は、順方向のサイクルによって付加される標識ヌクレオチド塩基を含む。ヌクレオチド断片リード202と重複する縦の灰色線又は陰影は

10

20

30

40

50

、正しいヌクレオチド塩基コールを示す。より具体的には、正しいヌクレオチド塩基コールは、参照ゲノムのヌクレオチド塩基と一致する。ヌクレオチド断片リード202内の文字は、参照ゲノム212からの塩基と一致しない不正確なヌクレオチド塩基コールを示す。

【0053】

図2Aに図示されるように、リードパイルアップ200は、塩基クオリティ204を含む。塩基クオリティ204は、ヌクレオチド断片リード202の各々の塩基クオリティを反映する。概して、正確なヌクレオチド塩基コールのより多くの発生は、より高い塩基クオリティに対応し、不正確なヌクレオチド塩基コールは、より低い塩基クオリティに対応する。例えば、いくつかの実施形態では、塩基クオリティ204は、ヌクレオチド断片リード202のうちの1つ内の塩基コールが誤っている確率を推定するPhredスコア(Q30)を反映する。対照的に、エラータイプカウンタ210は、様々なゲノム座標における色分けされたバー又は灰色スケール陰影付きバーを使用して、不正確な塩基コールの各タイプのエラー数を示す。例えば、いくつかの実施形態では、エラータイプカウンタ210は、不正確なヌクレオチド塩基コールを示す色分けされた棒グラフを含む。

10

【0054】

不正確なヌクレオチド塩基コールが図2Aに示されるように、参照ゲノム212は、エラー誘導配列を含む。特に、参照ゲノム212は、ホモポリマー206を含有する。ホモポリマー206は、連続するAヌクレオチドを有する配列を含む。図2Aに示されるように、不正確なヌクレオチド塩基コールの数は、ホモポリマー206に続く様々なリード位置で増加する。例えば、ヌクレオチド断片リード2について、エラーの数は、ホモポリマー206の後のヌクレオチド塩基について増加する。同様に、ヌクレオチド断片リード13についても、ホモポリマー206の後にエラーが増加する。しかし、不正確なヌクレオチド塩基コールは、ヌクレオチド断片リード1~10内の同じリード位置で異なる。そのようなエラー分散は、エラー誘導配列(ここでは、ホモポリマー206)が、エラー誘導配列に続くリード位置に対応するシグナルに対してフェージング又はプリフェージング効果を示すことを示す。

20

【0055】

図2Aに示されるように、不正確なヌクレオチド塩基コールは、ヌクレオチド断片リードの方向と一致するエラー誘導配列に従う。特に、ヌクレオチド断片リード202に対するヌクレオチド塩基コールは、しばしば正確であり、エラー誘導配列の前の高い塩基クオリティに対応する。エラー誘導配列に遭遇すると、SBSポリメラーゼは、スリップするか、そうでなければ追加の標識ヌクレオチド塩基を正確に組み込むことができない可能性がある。例示すると、前述したように、ヌクレオチド断片リード1~10は逆方向リードであり、ヌクレオチド断片リード11~20は順方向リードである。図2Aに図示されるように、エラーの数は、ヌクレオチド断片リードの方向と一致してホモポリマー206の後に増加する。したがって、いくつかの実施形態では、クラスタ認識塩基コーリングシステム106は、リード位置が、ヌクレオチド断片リードの方向と一致するエラー誘導配列に続くことを決定する。

30

【0056】

図2Aに更に示されるように、エラータイプカウンタ210は、ヌクレオチド断片リード202内の塩基コールエラーの位置及び大きさを示す。図2Aに図示されるように、エラータイプカウンタ210はまた、ホモポリマー206を取り囲む塩基コールエラーの発生の増加を示す。

40

【0057】

図2Aに示されるように、エラー誘導配列は、エラー誘導配列に続くリード位置におけるオリゴヌクレオチドのクラスタについてのシグナルにおけるフェージング効果及びプリフェージング効果を引き起こし得る。言及したように、図2Bは、1つ以上の実施形態によるフェージング及びプリフェージングを実証するためのクラスタ内の例示的なオリゴヌクレオチドを図示する。特に、図2Bは、配列決定サイクル中の特定のクラスタ内のオリ

50

ゴヌクレオチド 214 を図示する。概して、サイクルのための標識ヌクレオチド塩基 218 は、サイクル中に光シグナルに応答して蛍光を発する標識ヌクレオチド塩基を含む。例えば、標識 T ヌクレオチド塩基は、図 2 B に図示される所定のサイクルの間、大部分のオリゴヌクレオチドに付加されている。

【0058】

図 2 B はまた、フェージング及びプリフェージングを図示する。フェージングの一実施例では、図 2 B は、現在のサイクルに対応する標識ヌクレオチド塩基 218（ここでは「T」）のうちの 1 つではなく、前のサイクルに対応する標識ヌクレオチド塩基 216（ここでは「C」）をオリゴヌクレオチドに組み込む配列決定装置を図示する。したがって、前のサイクルの標識ヌクレオチド塩基 216 は、1 サイクル遅れて組み込まれる。プリフェージングの一実施例では、図 2 B は、現在のサイクルに対応する標識ヌクレオチド塩基 218（ここでは「T」）のうちの 1 つではなく、後続のサイクルに対応する標識ヌクレオチド塩基 220（ここでは「A」）を異なるオリゴヌクレオチドに組み込む配列決定装置を図示する。したがって、後続のサイクルのための標識ヌクレオチド塩基 220 は、1 サイクル早く組み込まれる。

【0059】

図 2 B によって示唆されるように、フェージング及びプリフェージングの両方が、クラスタ内の標識ヌクレオチド塩基からのシグナルに影響を与える。特に、現在のサイクルのための標識ヌクレオチド塩基 218 によって放出される光を含む純粋なシグナルを検出する代わりに、クラスタ認識塩基コーリングシステム 106 は、前のサイクルのための標識ヌクレオチド塩基 216 及び後続のサイクルのための標識ヌクレオチド塩基 220 からの蛍光を含む混合シグナルを検出する。以下の図及び段落は、クラスタ認識塩基コーリングシステム 106 がどのようにクラスタ固有のフェージング補正を生成してシグナルを調整し、フェージングされたヌクレオチド塩基及びプリフェージングされたヌクレオチド塩基を説明するかを更に説明する。

【0060】

図 3 は、クラスタ固有のフェージング補正を生成し、シグナルを調整して、特定のクラスタに対応する正確なヌクレオチド塩基コールを決定する、クラスタ認識塩基コーリングシステム 106 の概要を提供する。図 3 の概要として、クラスタ認識塩基コーリングシステム 106 は、エラー誘導配列に続くリード位置を特定する動作 302、リード位置に対応する標識ヌクレオチド塩基からのシグナルを検出する動作 304、クラスタ固有のフェージング補正を決定する動作 306、クラスタ固有のフェージング補正に基づいてシグナルを調整する動作 308、及びヌクレオチド塩基コールを決定する動作 310 を含む、一連の動作 300 を実行する。

【0061】

上述したように、図 3 は、エラー誘導配列に続くリード位置を特定する動作 302 を図示する。上述したように、いくつかの実施形態では、クラスタ認識塩基コーリングシステム 106 は、クラスタ固有のフェージング補正を、特定されたエラー誘導配列に続くリード位置のシグナルに制限することによって、クラスタのシグナルを補正するのに必要なコンピューティングリソースを部分的に制限する。図 3 に図示されるように、いくつかの実施形態では、クラスタ認識塩基コーリングシステム 106 は、前のサイクルからのシグナルに対するヌクレオチド塩基コールに基づいて、ホモポリマー、グアニン四重鎖、VNT R、又は他のエラー誘導配列を特定することによって、エラー誘導配列 312 を特定する。一実施例では、クラスタ認識塩基コーリングシステム 106 は、前のサイクルからのシグナルを分析し、閾値数の前のサイクルからのシグナルが同じヌクレオチド塩基を示すことを決定する。したがって、クラスタ認識塩基コーリングシステム 106 は、エラー誘導配列であるホモポリマーの存在を決定する。図 4 及び対応する説明は、エラー誘導配列の更なる詳細及び実施例を提供する。

【0062】

動作 302 の一部として、クラスタ認識塩基コーリングシステム 106 は、エラー誘導

配列に続くリード位置を特定する。図 3 に図示されるように、例えば、クラスタ認識塩基コーリングシステム 106 は、エラー誘導配列 312 に続くリード位置 314 を特定する。いくつかの実施形態では、クラスタ認識塩基コーリングシステム 106 は、エラー誘導配列 312 の特定された末端の後のリード位置 314 を特定する。例えば、エラー誘導配列 312 が、閾値類似性以内のシグナルを放出するヌクレオチド塩基を有するホモポリマーを含む場合、クラスタ認識塩基コーリングシステム 106 は、標識ヌクレオチド塩基が異なるシグナルを放出する第 1 の位置又は第 2 の位置でリード位置 314 を特定することができる。更に、又は代替として、クラスタ認識塩基コーリングシステム 106 は、(i) ヌクレオチド断片リードの最後の位置までエラー誘導配列に続く、又は (ii) エラー誘導配列 312 に続く閾値数のリード位置以内（例えば、エラー誘導配列に続く 200 又は 300 ヌクレオチド塩基以内）の 1 つ以上のリード位置を特定する。

10

【0063】

そのようなリード位置を特定した後、クラスタ認識塩基コーリングシステム 106 は、リード位置に対応する標識ヌクレオチド塩基からのシグナルを検出する動作 304 を実行する。特に、動作 304 を実行するとき、クラスタ認識塩基コーリングシステム 106 は、リード位置に対応するサイクル中にオリゴヌクレオチドのクラスタ内の標識ヌクレオチド塩基からのシグナルを検出する。したがって、動作 304 を実行することの一部として、クラスタ認識塩基コーリングシステム 106 は、標識ヌクレオチド塩基がリード位置 314 においてオリゴヌクレオチド内に組み込まれるサイクルを特定することによって、リード位置 314 に対応するサイクルを特定する。一実施例では、クラスタ認識塩基コーリングシステム 106 は、エラー誘導配列 312 に対応する前のサイクルの直後又は閾値数以内（例えば、そこから 2 サイクル以内）に続くサイクルを特定する。

20

【0064】

図 3 に更に図示されるように、動作 304 を実行するとき、クラスタ認識塩基コーリングシステム 106 は、クラスタ 320 の画像 316 を捕捉することができる。いくつかの実施形態では、クラスタ認識塩基コーリングシステム 106 は、配列決定装置のカメラを利用して、ヌクレオチド試料スライドの少なくとも 1 つのセクションの画像 316 を捕捉する。この実施例では、画像 316 は、ヌクレオチド試料スライドのタイル内のいくつかのクラスタを表す。追加の実施形態では、クラスタ認識塩基コーリングシステム 106 は、ヌクレオチド試料スライドのサブセクション、タイル、チャネル、又は他の部分など、ヌクレオチド試料スライドの他の部分の 1 つ以上の画像を捕捉する。更に示されるように、画像 316 は、クラスタ 320 から放出されたシグナル 318 を表す。シグナル 318 は、サイクル中にオリゴヌクレオチドのクラスタ内に組み込まれた標識ヌクレオチド塩基から放出される光シグナルを含む。

30

【0065】

関連するクラスタ内の標識ヌクレオチド塩基からそのようなシグナルを検出した後、クラスタ認識塩基コーリングシステム 106 は、クラスタ固有のフェージング補正を決定する動作 306 を実行する。特に、動作 306 を実行するとき、クラスタ認識塩基コーリングシステム 106 は、オリゴヌクレオチドのクラスタについて、推定されたフェージング及び推定されたプリフェージングのためにシグナルを補正するためのクラスタ固有のフェージング補正を決定する。より具体的には、いくつかの実施形態では、クラスタ認識塩基コーリングシステム 106 は、(i) 前のサイクルについてのヌクレオチド塩基に対応するクラスタ固有のフェージング係数、及び (ii) 後続のサイクルについてのヌクレオチド塩基に対応するクラスタ固有のプリフェージング係数を決定する。例えば、図 3 に図示されるように、係数 a はクラスタ固有のフェージング係数を表し、係数 b はクラスタ固有のプリフェージング係数を表す。クラスタ認識塩基コーリングシステム 106 は更に、クラスタ固有のフェージング補正を決定するためのアルゴリズム又は機能の一部として係数を利用することができる。例えば、いくつかの実施形態では、クラスタ認識塩基コーリングシステム 106 は、有限インパルス応答 (Finite Impulse Response、FIR) フィルタ内のクラスタ固有のフェージング係数及びクラスタ固有のプリフェージング係数を

40

50

利用する。

【0066】

図3は、単一のクラスタ固有のフェージング係数及び単一のクラスタ固有のプリフェージング係数を決定することを図示するが、いくつかの実施形態では、クラスタ認識塩基コーリングシステム106は、より多くの前のサイクル（例えば、2つ、3つ、4つ等の前のサイクル）及び／又はより多くの後続のサイクル（例えば、2つ、3つ、4つ等の後続のサイクル）に対応する複数の追加の係数を決定する。図5及び対応する段落は、1つ以上の実施形態による、クラスタ認識塩基コーリングシステム106がクラスタ固有のフェージング係数 a 及びクラスタ固有のプリフェージング係数 b をどのように決定するかを更に詳述する。

10

【0067】

クラスタ認識塩基コーリングシステム106は、クラスタ固有のフェージング補正を決定する動作306を実行する一部として、いくつかのモデルを利用することができる。例えば、クラスタ認識塩基コーリングシステム106は、線形等化器（Linear Equalizer、LE）、判定帰還型等化器（Decision Feedback Equalizer、DFE）、又は最尤系列推定器（Maximum Likelihood Sequence Estimator、MLSE）を利用して、クラスタ固有のフェージング係数及びクラスタ固有のプリフェージング係数を決定することができる。図7A～図7C及び付随する議論は、これらのモデルの各々に関する更なる詳細を提供する。

【0068】

20

いくつかの実施形態では、動作306を実行することの一部として、クラスタ認識塩基コーリングシステム106は、クラスタ固有のフェージング係数 a 及びクラスタ固有のプリフェージング係数 b を利用して、前のサイクル（ w_{-1} ）、現在のサイクル（ w_0 ）、及び後続のサイクル（ w_1 ）に対応する重みを決定する。いくつかの実施形態では、重みは、クラスタ認識塩基コーリングシステム106がシグナルを調整するために利用する等化器係数を表す。図3は、前のサイクル、現在のサイクル、及び後続のサイクルに対応する3つの重みのウィンドウを図示するが、クラスタ認識塩基コーリングシステム106は、上で示したように、より多くの重みを生成することができる。例えば、クラスタ認識塩基コーリングシステム106は、5つの重みを生成することができる。例示すると、5つの重みのうち、クラスタ認識塩基コーリングシステム106は、前のサイクルに先行するサイクル（ w_{-2} ）、前のサイクル（ w_{-1} ）、現在のサイクル（ w_0 ）、後続のサイクル（ w_1 ）、及び後続のサイクルに続くサイクル（ w_2 ）に対応する重みを決定する。したがって、クラスタ認識塩基コーリングシステム106は、特定された重みの数を7つ、9つ、又は任意の関連ウィンドウに拡張することができる。

30

【0069】

クラスタ固有のフェージング補正を決定した後、クラスタ認識塩基コーリングシステム106は、クラスタ固有のフェージング補正に基づいてシグナルを調整する動作308を実行する。概して、クラスタ認識塩基コーリングシステム106は、クラスタ固有のフェージング係数（ a ）及びクラスタ固有のプリフェージング係数（ b ）に基づいてシグナルを調整する。いくつかの実施形態では、クラスタ認識塩基コーリングシステム106は、オリゴヌクレオチドのクラスタからのシグナルに上述の重みを適用することによって動作308を実行する。例えば、図3は、前のサイクル、サイクル、及び後続のサイクルのシグナルを $\{x_{-1}, x_0, x_1\}$ として表す。クラスタ認識塩基コーリングシステム106は、前のサイクル、現在のサイクル、及び後続のサイクル $\{w_{-1}, w_0, w_1\}$ に対する重みを適用して、前のサイクル、サイクル、及び後続のサイクル

40

【0070】

【数1】

$$\{\hat{x}_{-1}, \hat{x}_0, \hat{x}_1\}$$

50

に対する調整されたシグナルを生成する。

いくつかの実施形態では、クラスタ認識塩基コーリングシステム106は、前のステップにおいて決定された重みの数に基づいて、追加のサイクルのための調整されたシグナルを生成する。

【0071】

シグナルを調整した後、クラスタ認識塩基コーリングシステム106は、ヌクレオチド塩基コールを決定する動作310を実行する。特に、動作310を実行するとき、クラスタ認識塩基コーリングシステム106は、調整されたシグナルに基づいて、オリゴヌクレオチドのクラスタに対応するリード位置についてのヌクレオチド塩基コールを決定する。例えば、図3に図示されるように、クラスタ認識塩基コーリングシステム106は、調整されたシグナルに基づいて、リード位置314におけるヌクレオチド塩基の同一性がチミン（T）であることを決定する。概して、クラスタ認識塩基コーリングシステム106は、配列決定システム104を利用して、クラスタ内のヌクレオチド塩基の同一性を示すヌクレオチド塩基コールを生成して、ヌクレオチド断片リードを決定することができる。クラスタ認識塩基コーリングシステム106は、他の核酸ポリマーの試料ゲノムの配列を示すために、調整されたシグナルの分析から得られたヌクレオチド断片リードを更に整列させることができる。

【0072】

図3は、クラスタ認識塩基コーリングシステム106が、配列決定サイクルにおいて、又は配列決定サイクル中に、所与のクラスタからのシグナルに対して、クラスタ固有のフェージング係数及びクラスタ固有のプリフェージング係数を決定し、そのような係数に基づいてシグナルを調整することを示すが、いくつかの実施形態では、クラスタ認識塩基コーリングシステム106は、配列決定サイクルが継続するにつれて、所与のクラスタからのシグナルに対してそのような係数を決定及び再決定することができる。例えば、いくつかの実施形態では、クラスタ認識塩基コーリングシステム106は、配列決定サイクルにおけるオリゴヌクレオチドの所与のクラスタについてのクラスタ固有のフェージング係数及びクラスタ固有のプリフェージング係数（及び対応する重み）を決定し、次いで、後続の配列決定サイクルにおけるオリゴヌクレオチドの所与のクラスタについての更新されたクラスタ固有のフェージング係数及び更新されたクラスタ固有のプリフェージング係数（及び対応する重み）を決定することができる。各後続のサイクルについても同様である。したがって、クラスタ認識塩基コーリングシステム106は、所与のクラスタに対応するヌクレオチド断片リードに対するヌクレオチド塩基コールを決定する過程にわたって、オリゴヌクレオチドの所与のクラスタに対するクラスタ固有のフェージング係数及びクラスタ固有のプリフェージング係数を再決定し、変更する。

【0073】

図3は、1つ以上の実施形態による、推定フェージング及びプリフェージングのために調整されたシグナルからヌクレオチド塩基コールを決定することの一部として、クラスタ認識塩基コーリングシステム106によって実行される動作の概要を提供する。図4は、1つ以上の実施形態による、エラー誘導配列を特定するためにクラスタ認識塩基コーリングシステム106によって実行される一連の動作を図示する。概して、クラスタ認識塩基コーリングシステム106は、クラスタ固有のフェージング補正を選択的に決定し、クラスタ固有のフェージング補正に従って、エラー誘導配列に続く特定のサイクルからのシグナルを調整する。図4の一連の動作400によって示されるように、クラスタ認識塩基コーリングシステム106は、複数のサイクルからのシグナルを分析する動作402、シグナルからヌクレオチド塩基コールを決定する動作403、及びエラー誘導配列を特定する動作404を実行することによって、エラー誘導配列を特定する。

【0074】

図4に図示されるように、クラスタ認識塩基コーリングシステム106は、複数のサイクルからのシグナルを分析する動作402を実行する。概して、クラスタ認識塩基コーリングシステム106は、クラスタの1つ以上の画像を撮影することによって、クラスタか

らの標識ヌクレオチド塩基からのシグナルを検出する。より具体的には、クラスタ認識塩基コーリングシステム106は、複数のクラスタを含有するヌクレオチド試料スライド（例えば、フローセルのタイル）のセクションの1つ以上の画像を捕捉する。画像は、クラスタから放出されるシグナルを捕捉する。クラスタ認識塩基コーリングシステム106は、シグナル406a～406cを検出するために画像を分析する。シグナル406a～406cは、異なるサイクルについてクラスタ内の標識ヌクレオチド塩基から放出されるシグナルを含む。例えば、クラスタ認識塩基コーリングシステム106は、第1のサイクルについてシグナル406aを記録し、第2のサイクルについてシグナル406bを記録し、第3のサイクルについてシグナル406cを記録する。

【0075】

いくつかの実施形態では、シグナル406a～406cは、異なる検出チャンネルから得られた画像から導出される。例えば、シグナル406a～406cは、2チャンネル又は4チャンネル配列決定から得られた画像に基づいて生成することができる。各ヌクレオチド塩基は、異なるシグナルと関連付けられる。例示すると、2チャンネルSSBでは、緑色クラスタはCヌクレオチド塩基に対応し、赤色クラスタはTヌクレオチド塩基に対応し、赤色及び緑色の両方で観察されるクラスタはAヌクレオチド塩基としてフラグが立てられ、非標識クラスタはGヌクレオチド塩基に対応する。対照的に、1つ以上の実施形態では、クラスタ認識塩基コーリングシステム106は、単一の検出チャンネルからのシグナルを検出する。例えば、シグナル406a～406cは、1チャンネル配列決定から得られた画像に基づいて生成される。

【0076】

いくつかの実施形態では、複数のサイクルからのシグナルを分析する動作402を実行することの一部として、クラスタ認識塩基コーリングシステム106は、フェージング／フレージング及びノイズのためにシグナル406a～406cを調整する。特に、クラスタ認識塩基コーリングシステム106は、推定されたフェージング及び／又は推定されたプリフェージングのためにシグナル406a～406cを補正するために、クラスタ固有のフェージング補正を決定することができる。一実施例では、クラスタ認識塩基コーリングシステム106は、シグナル406a～406cを調整してノイズを低減することによって、複数のサイクルからのシグナルを更に分析する。例えば、いくつかの実施形態では、クラスタ認識塩基コーリングシステム106は、ノイズを除去するためのデノイザー（de-noiser）又はアルゴリズムを利用する。実際、場合によっては、ノイズはシグナルの一部であり、観察された母集団における分布をもたらす（又は反映する）シグナル変動を含む。シグナル変動は、ヌクレオチド試料スライド（例えば、フローセル）又は配列決定装置の構成要素又は内容物の化学的又は物理的特性、例えば、オリゴヌクレオチド長、フェージング若しくはプリフェージング、又はカメラ若しくは他のセンサーの視野に対するオリゴヌクレオチドのクラスタの位置に起因するシグナル変動に由来し得る。ノイズを除去することに加えて、クラスタ認識塩基コーリングシステム106は、他のメトリクスを改善するためにシグナル406a～406cを更に精緻化することができる。例えば、いくつかの実施形態では、クラスタ認識塩基コーリングシステム106は、シグナル406a～406cの強度値に対応するオフセット及びスケーリング係数に基づいて、シグナル406a～406cを調整する。

【0077】

更に、複数のサイクルからのシグナルを分析する動作402を実行することの一部として、クラスタ認識塩基コーリングシステム106は、調整されたシグナルの強度値を強度値境界のセットと比較する。概して、強度値境界は、シグナルに対するヌクレオチド塩基コールを生成する際に使用される決定境界を指す。特に、強度値境界は、シグナルの1つ以上の強度値に基づいてヌクレオチド塩基を分類する決定境界を指すことができる。例示すると、強度値境界は、ヌクレオチド塩基の各々に対応するヌクレオチドクラウドの境界を定義するか、又は別様に示すことができる。特に、クラスタ認識塩基コーリングシステム106は、各可能なヌクレオチド塩基（例えば、A、T、C、又はG）に対応する強度

10

20

30

40

50

値境界のセットを特定する。いくつかの実施形態では、クラスタ認識塩基コーリングシステム106は、強度値境界のセットのうちの1つの外側の強度値を有する調整されたシグナルを破棄する。例えば、クラスタについて調整されたシグナルが強度値境界のセットのうちの1つの外側の強度値を有すると決定することに基づいて、クラスタ認識塩基コーリングシステム106は、クラスタについてヌクレオチド塩基コールを生成しないことを決定する。

【0078】

図4に更に図示されるように、一連の動作400は、シグナルからヌクレオチド塩基コールを決定する動作403を含む。特に、クラスタ認識塩基コーリングシステム106は、強度値境界のセットのうちの1つを利用して、シグナルに対するヌクレオチド塩基コールを生成することができる。特に、クラスタ認識塩基コーリングシステム106は、強度値境界のセットを利用してヌクレオチド塩基コールを生成することができる。概して、強度値境界のセットとシグナル406aとの間の相関を決定することに基づいて、クラスタ認識塩基コーリングシステム106は、シグナル406aの調整されたバージョン（すなわち、調整されたシグナル）に対応するサイクルについてのヌクレオチド塩基コールを決定する。例えば、シグナル406aの調整されたバージョン（すなわち、調整されたシグナル）に対応する強度値が、Aヌクレオチド塩基に対応する強度値境界のセット内に入ると決定することに基づいて、クラスタ認識塩基コーリングシステム106は、Aヌクレオチド塩基コールを決定する。

【0079】

いくつかの実施形態では、クラスタ認識塩基コーリングシステム106は、ヌクレオチド塩基コールを決定した後にシグナルデータを破棄する。クラスタ固有のフェージング補正を推定するために必要とされる記憶域ロードを低減するために、クラスタ認識塩基コーリングシステム106は、シグナルデータを周期的に削除又は破棄することができる。例えば、いくつかの実施形態では、クラスタ認識塩基コーリングシステム106は、閾値数のサイクル内のシグナルデータを破棄する。例えば、クラスタ認識塩基コーリングシステム106は、特定のサイクルについてヌクレオチド塩基コールを決定する閾値数のサイクル（例えば、3、5、10等）内のシグナルデータを削除することができる。前述したように、クラスタ認識塩基コーリングシステム106は、エラー誘導配列に続くリード位置に対応するサイクルのシグナルを選択的に訂正する。したがって、場合によっては、クラスタ認識塩基コーリングシステム106は、エラー誘導配列によって影響されないサイクルのシグナルデータを削除する。いくつかの実施形態では、所与のクラスタについて、クラスタ認識塩基コーリングシステム106は、エラー誘導配列によって影響されないサイクルを特定し、対応するシグナルデータを破棄する。例えば、クラスタ認識塩基コーリングシステム106は、前のサイクルに対するヌクレオチド塩基コールが特定可能なエラー誘導配列を示さないと決定することができる。この決定に基づいて、クラスタ認識塩基コーリングシステム106は、そのサイクルのシグナル伝達データを破棄する。

【0080】

図4に更に図示されるように、クラスタ認識塩基コーリングシステム106は、複数のサイクルにわたって動作403を繰り返す。特に、クラスタ認識塩基コーリングシステム106は、複数のサイクルからのシグナルに対するヌクレオチド塩基コールを決定する。クラスタに対する各サイクルで得られるヌクレオチド塩基コールの配列は、クラスタに対するヌクレオチド断片リードとなる。例えば、図4に図示されるように、クラスタ認識塩基コーリングシステム106は、配列「CTGTAAAAA」を有するヌクレオチド断片リードを生成する。

【0081】

図4に更に図示されるように、クラスタ認識塩基コーリングシステム106は、エラー誘導配列を特定する動作404を実行する。概して、クラスタ認識塩基コーリングシステム106は、ヌクレオチド断片リードからのヌクレオチド塩基（先行するサイクルに対応する）の配列を分析して、エラー誘導配列の存在を検出する。例えば、特定のサイクルに

ついて特定のヌクレオチド塩基コールを決定した後、クラスタ認識塩基コーリングシステム106は、成長ヌクレオチド断片リードからのヌクレオチド塩基コールの配列を、可能性のあるエラー誘導配列のデータベースと比較することができる。そのようなエラー誘導配列のデータベースを使用することによって、クラスタ認識塩基コーリングシステム106は、ヌクレオチド断片リードがエラー誘導配列を含むかどうかを決定するために、ヌクレオチド塩基コールの配列を分析することができる。そのようなヌクレオチド断片リードからのヌクレオチド塩基コールの配列が特定のエラー誘導配列と一致する（又はそれから閾値数のヌクレオチド塩基内にある）場合、クラスタ認識塩基コーリングシステム106は、ヌクレオチド断片リード内のエラー誘導配列を特定する。

【0082】

概して、エラー誘導配列は、1つ以上の反復ヌクレオチド塩基又は配列モチーフの配列を含む。配列モチーフは、ゲノム内で起こるヌクレオチドパターンを含み得る。いくつかの実施例では、配列モチーフは、生物学的機能に関連する。図4は、1つ以上の実施形態によるいくつかの例示的なエラー誘導配列を図示する。以下の段落は、クラスタ認識塩基コーリングシステム106によって特定されたエラー誘導配列の様々な例を説明する。いくつかの実施形態では、配列認識モデルは、エラー誘導配列のトリガーを特定する。例えば、配列認識モデルは、塩基コーリングエラーを引き起こすヌクレオチド塩基配列を特定又は予測するように訓練された機械学習モデルを含むことができる。更に、又は代替として、エラー誘導配列は、配列内の塩基のブロック又は群の塩基数に基づいて特定可能である。

【0083】

図4に図示されるように、ホモポリマーは、エラー誘導配列であり得る。概して、ホモポリマーは、同一のモノマー単位からなるか又は同一のモノマー単位を含むポリマーを含む。特に、ホモポリマーは、単一の反復ヌクレオチド塩基を有する配列を含む。例えば、ホモポリマーは、15個以上の反復Aヌクレオチドのセグメントを含み得る。ホモポリマーは、多くの場合、クラスタ化の間にポリメラーゼスリップを引き起こすことによってエラーを誘導する。ポリメラーゼスリップは、ポリメラーゼがオリゴヌクレオチドから一時的に解離し、異なる位置に再付着する場合に起こる。そのようなポリメラーゼスリップは、多くの場合、不均一な長さのフィラメントを生成し、これは下流の急性フェージング又はプリフェージングエラーとして現れる。ホモポリマーは、A、T、G、又はCのホモポリマーを含む、任意のヌクレオチド塩基の反復配列を含むことができる。いくつかの実施形態では、ニアホモポリマーもエラー誘導配列とみなされる。特に、ニアホモポリマーは、いくつかを除いて全てのモノマーが同じであるポリマーを含む。例えば、ニアホモポリマーは、単一の異なる塩基によって中断された反復塩基（例えば、20）の鎖を含むことができる。

【0084】

図4に図示されるエラー誘導配列の別の例は、グアニン四重鎖（G-四重鎖）を含む。G-四重鎖は、グアニンに富む配列によって形成される安定な二次構造である。特に、G-四重鎖は、SBS中に鑄型オリゴヌクレオチド上に鎖内二次構造を形成する。G-四重鎖は、SBSポリメラーゼを遮断することによってSBSにおけるエラーを誘導することができる。より具体的には、配列決定サイクル後に洗い流されるポリメラーゼは、再付着の効率が低いことが多く、壊滅的なフェージングを引き起こす。クラスタ認識塩基コーリングシステム106は、グアニンが豊富な配列を特定することによってG-四重鎖を特定し得る。いくつかの実施形態では、クラスタ認識塩基コーリングシステム106は、G-四重鎖配列モチーフを計算的に予測することができる。例えば、クラスタ認識塩基コーリングシステム106は、配列ベースの計算モデルなどの機械学習モデルを利用して、G-四重鎖の形成を予測することができる。

【0085】

いくつかのエラー誘導配列（例えば、G-四重鎖）は、ホモポリマーを含む他のエラー誘導配列よりも特定が困難である。例えば、クラスタ認識塩基コーリングシステム106

10

20

30

40

50

は、G－四重鎖の存在を誤って検出し、したがって、クラスタ固有のフェージング補正の決定に進む可能性がある。このタイプの早すぎる決定は、性能に悪影響を与えないが、追加のリソースを消費する。いくつかの実施形態では、クラスタ認識塩基コーリングシステム106は、エラー誘導配列が容易に特定可能なヌクレオチド配列（例えば、ホモポリマー及びニアホモポリマー）でない限り、クラスタ固有のフェージング補正を決定しない。

【0086】

図4に更に図示されるように、可変タンデム反復（VNTR）は、エラー誘導配列の別の例である。VNTRは、短いヌクレオチド配列（20～100塩基対）がタンデム反復として組織化されるゲノム中の位置を含み得る。例えば、VNTRは、6つの反復AGTCGGTAAG配列又は様々な他の数の反復サブ配列で構成される配列を含み得る。VNTRは、下流のフェージング及びプリフェージングにつながるポリメラーゼスリップを引き起こすことによって、SBSにおけるエラーを引き起こし得る。

10

【0087】

VNTRの他の例としては、ミニサテライト配列及びマイクロサテライト配列が挙げられる。ミニサテライト配列は、ある特定のDNAモチーフ（10～60塩基対の長さの範囲）が典型的には5～50回繰り返される、反復DNAのトラクトを指す。マイクロサテライト配列は、ある特定のDNAモチーフ（1～6塩基対又はそれ以上の長さの範囲）が典型的には5～50回繰り返される、反復DNAのトラクトである。

【0088】

図4に更に図示されるように、エラー誘導配列はまた、ジヌクレオチド反復配列及びトリヌクレオチド反復配列を含み得る。ジヌクレオチド反復配列は、正確に2つのヌクレオチドが反復される場合に起こる。ATAAT配列は、ジヌクレオチド反復配列の一例である。同様に、トリヌクレオチド反復配列は、正確に3つのヌクレオチドが反復される場合に起こる。例えば、DNA配列CAGCAGCAGCAGは、4つのCAG反復を含有する。ジヌクレオチド反復配列及びトリヌクレオチド反復配列は、ポリメラーゼスリップを引き起こすことによってSBSに悪影響を及ぼす。更に、いくつかの実施例では、ジヌクレオチド反復配列及びトリヌクレオチド反復配列はまた、SBSのPCR調製ステップに負の影響を与え得る。

20

【0089】

図4に図示されるエラー誘導配列の別の例は、逆方向反復配列である。逆方向反復配列は、ヌクレオチドの一本鎖配列と、下流に続くその逆相補鎖とを含む。最初の配列と逆相補鎖との間のヌクレオチドの介在配列は、0を含む任意の長さであり得る。例えば、TTACGnnnnCGTAは、逆方向反復配列である。逆方向反復配列は、多くの場合、鎖間ヘアピン又は鎖内ハイブリダイゼーションを引き起こし得る。得られた二次構造は、多くの場合、SBS中にSBSポリメラーゼがオリゴヌクレオチドに再付着するのを遮断する。

30

【0090】

パリンドローム配列は、クラスタ認識塩基コーリングシステム106によって特定可能なエラー誘導配列の別の例を表す。パリンドローム配列は、ヌクレオチド塩基の第1のランと、それに続く相補的塩基の第2のランを逆の順序で含む。GGATCCは、パリンドローム配列の一例である。パリンドローム配列は、クラスタ内で鎖内及び鎖間ハイブリダイゼーションを引き起こすため、SBS中に問題となり得る。例えば、パリンドローム配列は、モチーフ自体内でハイブリダイゼーションを引き起こすことができる。パリンドローム配列はまた、1つのオリゴヌクレオチド上の配列が第2のオリゴヌクレオチド上の配列とハイブリダイズする、鎖間ハイブリダイゼーションを引き起こし得る。相互作用の両方の形態は、SBS中にポリメラーゼを遮断する。

40

【0091】

いくつかの実施形態では、クラスタ認識塩基コーリングシステム106は、方向固有の配列モチーフを特定する。特に、クラスタ認識塩基コーリングシステム106は、配列モチーフが特定の方向にあると決定することに基づいて、配列モチーフにエラー誘導配列と

50

してフラグを立てることができる。クラスタ認識塩基コーリングシステム106は、反対方向の同じ配列モチーフがエラー誘導配列を含まないことを決定することができる。一実施例では、順鎖上のG-四重鎖は、SBS中に鎖内二次構造を生成し、配列決定リードに悪影響を及ぼす可能性がある。対照的に、G-四重鎖の逆鎖又は相補鎖は、通常、鎖内二次構造を生成しない（逆方向もG-四重鎖を含む場合を除く）。鎖内二次構造を形成する傾向がある他のエラー誘導配列もまた、方向固有の配列モチーフであり得る。

【0092】

図4及び上記の付随する議論は、1つ以上の実施形態による、ヌクレオチド断片リード内のエラー誘導配列を特定するクラスタ認識塩基コーリングシステム106を説明する。前述したように、クラスタ認識塩基コーリングシステム106はまた、エラー誘導配列に続くリード位置を特定する。クラスタ認識塩基コーリングシステム106は、リード位置に対応するサイクル中に標識ヌクレオチド塩基からのシグナルを更に処理する。シグナルを処理することの一部として、クラスタ認識塩基コーリングシステム106は、シグナルを補正するためにクラスタ固有のフェージング補正を決定する。特に、クラスタ認識塩基コーリングシステム106は、クラスタ固有のフェージング係数及びクラスタ固有のプリフェージング係数に基づいて、クラスタ固有のフェージング補正を決定することができる。図5及び対応する段落は、1つ以上の実施形態による、クラスタ固有のフェージング係数を決定し、クラスタ固有のプリフェージング係数を決定するための一連の動作500を説明する。

【0093】

図5に示されるように、クラスタ認識塩基コーリングシステム106は、クラスタ固有のフェージング係数を決定する動作502を実行する。特に、動作502の一部として、クラスタ認識塩基コーリングシステム106は、オリゴヌクレオチドのクラスタについて、前のサイクルについてのヌクレオチド塩基に対応するクラスタ固有のフェージング係数を決定する。

【0094】

図5は、オリゴヌクレオチドのクラスタ内の標識ヌクレオチド塩基から放出されるシグナルを図示する。例えば、図5は、サイクルについての単一クラスタ内の標識ヌクレオチド塩基からの現在のサイクルシグナル508、及び前のサイクルについてのクラスタ内の標識ヌクレオチド塩基からの前のサイクルシグナル506を図示する。クラスタのオリゴヌクレオチドに組み込まれた他の標識ヌクレオチド塩基（図示せず）とともに、クラスタは、画像によって捕捉された集合的シグナルを放出する。説明を容易にするために、本開示では、前のサイクルシグナル506、現在のサイクルシグナル508、及び後続のサイクルシグナル510を、所与のサイクルのクラスタの集合的シグナルを構成するシグナルの集合と呼ぶ。示されるように、各円は、クラスタ内の単一の標識ヌクレオチド塩基によって放出されたシグナルを表す。図示されるように、現在のサイクルシグナル508は、緑色光を放出する2つの標識ヌクレオチド塩基、赤色光を放出する標識ヌクレオチド塩基上、並びに緑色及び赤色の両方を放出する1つの標識ヌクレオチド塩基を含む。

【0095】

いくつかの実施形態では、クラスタ認識塩基コーリングシステム106は、現在のサイクルの直前の前のサイクルについてのヌクレオチド塩基に対応するクラスタ固有のフェージング係数を決定する。言及したように、フェージングは、クラスタ内の1つ以上のオリゴヌクレオチドが、ヌクレオチド塩基の組み込みに遅れる場合に起こる。例えば、図5に図示されるように、クラスタ認識塩基コーリングシステム106は、前のサイクルシグナル506を特定する。前のサイクルシグナル506は、前のサイクル中にクラスタ内でオリゴヌクレオチドに付加された標識ヌクレオチドが、赤色シグナルを放出することを示す。現在のサイクルシグナル508は、そのサイクル中にフェージングが起こったことを示す。より具体的には、現在のサイクルシグナル508は、赤色光を放出する1つの標識ヌクレオチド塩基を含み、これは前のサイクルシグナル506の赤色光に対応する。以下で更に説明されるように、クラスタ認識塩基コーリングシステム106は、前のサイクルに

ついで、ヌクレオチド塩基に対応するクラスタ固有のフェージング係数を決定する。

【0096】

図5に更に図示されるように、クラスタ認識塩基コーリングシステム106はまた、クラスタ固有のプリフェージング係数を決定する動作504を実行する。特に、クラスタ認識塩基コーリングシステム106は、オリゴヌクレオチドのクラスタについて、そのサイクルの直後の後続のサイクルについてのヌクレオチド塩基に対応するクラスタ固有のプリフェージング係数を決定する。言及したように、プリフェージングは、1つ以上のオリゴヌクレオチドが1サイクル以上早くヌクレオチド塩基を組み込む場合に起こる。図5に図示されるように、現在のサイクルシグナル508は、緑色光と赤色光との組み合わせを放出する標識ヌクレオチド塩基を含む。クラスタ内の標識ヌクレオチドによって放出された緑色及び赤色（green and red、G/R）光は、後続のサイクルシグナル510からのG/R標識ヌクレオチドに対応する。以下で更に説明されるように、動作504を実行することの一部として、クラスタ認識塩基コーリングシステム106は、後続のサイクルからのG/Rヌクレオチド塩基に対応するクラスタ固有のプリフェージング係数を決定する。

10

【0097】

いくつかの実施形態では、クラスタ認識塩基コーリングシステム106は、入力シグナル、所望の出力シグナル、及び様々なパラメータに基づいて、クラスタ固有のプリフェージング係数及びクラスタ固有のフェージング係数を決定する。特に、クラスタ認識塩基コーリングシステム106が3タップ線形等化器を利用する1つ以上の実装形態では、クラスタ認識塩基コーリングシステム106は、入力シグナル（ v ）、所望の出力シグナル（ d ）、並びに分布の平均（ μ ）及び標準偏差（ σ ）を含むパラメータに基づいて、3タップ線形等化器のためのクラスタ固有のプリフェージング係数及びクラスタ固有のフェージング係数を生成する。概して、クラスタ認識塩基コーリングシステム106は、判定指向型の適応を利用する。特に、クラスタ認識塩基コーリングシステム106は、所望の出力シグナル（ d ）を塩基コールのクラウドの中心に設定し、所望の出力シグナル（ d ）を使用して、分布の平均（ μ ）及び標準偏差（ σ ）を含むパラメータを更新する。クラスタ認識塩基コーリングシステム106がクラスタ固有のフェージング係数及びクラスタ固有のプリフェージング係数をどのように決定するかの特定の例は、図7Aに付随する段落において以下に提供される。

20

30

【0098】

図5は、クラスタ固有のフェージング係数及びクラスタ固有のプリフェージング係数を決定するクラスタ認識塩基コーリングシステム106を図示するが、いくつかの実施形態では、クラスタ認識塩基コーリングシステム106は、追加のクラスタ固有のフェージング係数及び追加のクラスタ固有のプリフェージング係数を決定する。フェージングは、ヌクレオチド塩基が1サイクル遅れて付加される場合を指すことができ、プリフェージングは、ヌクレオチド塩基が1サイクル早く付加される場合を指すことができる。しかしながら、フェージング及びプリフェージングは、それぞれ2サイクル以上遅れて、及び2サイクル以上早く付加されたヌクレオチド塩基を指すこともできる。したがって、いくつかの実施形態では、クラスタ認識塩基コーリングシステム106は、追加の前のサイクル（すなわち、サイクルの2サイクル前）の追加のヌクレオチド塩基に対応する追加のクラスタ固有のフェージング係数を決定する。クラスタ認識塩基コーリングシステム106はまた、追加の後続のサイクル（すなわち、サイクルの2サイクル後）の追加のヌクレオチド塩基に対応する追加のクラスタ固有のプリフェージング係数を決定することができる。

40

【0099】

クラスタ認識塩基コーリングシステム106はまた、サイクルの直前の前のサイクルのセットについてのヌクレオチド塩基のセットに対応するクラスタ固有のフェージング係数のセットを決定することができる。そのような前のサイクルのセットは、任意の数の先行サイクルを含むことができる。同様に、クラスタ認識塩基コーリングシステム106はまた、サイクルの直後の後続のサイクルのセットに対応するクラスタ固有のプリフェージン

50

グ係数のセットを決定することができる。そのような後続のサイクルのセットは、任意の数の後続のサイクルを含むことができる。

【0100】

いくつかの実施形態では、クラスタ認識塩基コーリングシステム106は、前のサイクルの非対称セット及び後続のサイクルのセットからのシグナルを分析する。例えば、クラスタ認識塩基コーリングシステム106は、(i)シグナルを処理して、単一の先行サイクルについてのクラスタ固有のフェージング係数を決定し、(ii)複数のシグナルを処理して、複数の後続のサイクル(例えば、2つ又は3つの後続のサイクル)についてのクラスタ固有のプリフェージング係数を決定することができる。更なる実施例として、クラスタ認識塩基コーリングシステム106は、(i)複数のシグナルを処理して、複数の先行するサイクル(例えば、2つ又は3つの先行するサイクル)についてのクラスタ固有のフェージング係数を決定し、(ii)単一のシグナルを処理して、単一の後続のサイクルについてのクラスタ固有のプリフェージング係数を決定することができる。更に、又は代替として、クラスタ認識塩基コーリングシステム106は、非連続サイクルからのシグナルを処理することができる。例示すると、クラスタ認識塩基コーリングシステム106は、前のサイクルに先行するサイクル、現在のサイクル、及び後続のサイクルからのシグナルについてのクラスタ固有の係数を分析し決定することができる。この実施例では、クラスタ認識塩基コーリングシステム106は、前のサイクルからのシグナルを分析しないことを決定するが、現在のサイクルの前又は後の別の非連続サイクルを選択することができる。

10

20

【0101】

説明したように、図5は、1つ以上の実施形態によるクラスタ固有のフェージング補正を決定することの一部として、クラスタ固有のフェージング係数及びクラスタ固有のプリフェージング係数を決定するクラスタ認識塩基コーリングシステム106を図示する。いくつかの実施形態では、クラスタ認識塩基コーリングシステム106は、様々なアルゴリズムとともにクラスタ固有のフェージング補正を決定する。図6は、1つ以上の実施形態による、フェージング補正を決定するための例示的なフェージングモデルを図示する。概して、クラスタ認識塩基コーリングシステム106は、オリゴヌクレオチドのクラスタからのシグナルを補正するためのクラスタ固有のフェージング補正、並びにクラスタからのシグナル及びクラスタのセットからのシグナルを補正するためのマルチクラスタフェージング補正を決定することができる。図6は、連続した2つの畳み込み演算としてモデル化されたクラスタ固有の係数演算606及びマルチクラスタ係数演算608を図示する。

30

【0102】

具体的には、図6は、クラスタ固有のフェージング補正及びマルチクラスタフェージング補正を生成することの一部として様々な係数を推定するためのフェージングモデル600を図示する。フェージングモデル600は、シーケンサ602又は他の配列決定マシン上で行われる演算、並びにシグナル処理604中に行われる演算を含む。例えば、いくつかの実施形態では、クラスタ認識塩基コーリングシステム106は、クラスタ固有の係数演算606を実行してクラスタ固有のフェージング係数を推定し、マルチクラスタ係数演算608を実行してマルチクラスタフェージング係数を推定する。クラスタ認識塩基コーリングシステム106は、シグナル処理604の一部として、クラスタ固有のフェージング係数及びマルチクラスタフェージング係数を更に利用することができる。より具体的には、クラスタ認識塩基コーリングシステム106は、マルチクラスタフェージング補正610を実行して、マルチクラスタフェージング係数に基づいてシグナルを調整する。更に、クラスタ認識塩基コーリングシステム106は、クラスタ固有のフェージング補正及び塩基コーリング612を実行して、クラスタ固有のフェージング係数に基づいてシグナルを調整し、調整されたシグナルに基づいてヌクレオチド塩基コールを生成する。

40

【0103】

フェージングモデル600は、リアルタイム(又はほぼリアルタイム)コンピューティングアーキテクチャ又はバッファ型コンピューティングアーキテクチャを含むことができ

50

る。概して、リアルタイムコンピューティングアーキテクチャを利用することによって、クラスタ認識塩基コーリングシステム106は、シーケンサ602のプロセッサ（例えば、配列決定装置114）を利用して、図6に図示される全ての演算を実行する。対照的に、クラスタ認識塩基コーリングシステム106はまた、配列決定マシン及び1つ以上のサーバー（例えば、サーバー装置102）の両方を含む、バッファ型コンピューティングアーキテクチャを採用し得る。一実施例では、クラスタ認識塩基コーリングシステム106は、シーケンサ602においてクラスタ固有の係数演算606及びマルチクラスタ係数演算608を実行しながら、1つ以上のサーバー装置においてシグナル処理604を実行する。より具体的には、クラスタ認識塩基コーリングシステム106は、サーバー装置のプロセッサにおいて、(i) マルチクラスタフェージング補正610、並びに(ii) クラスタ固有のフェージング補正及び塩基コーリング612を実行することができる。

10

【0104】

概して、また前述したように、フェージング及びプリフェージングは、それぞれ1つ以上の前又は後のサイクルに対応するヌクレオチド塩基を組み込むことによって、クラスタ中のオリゴヌクレオチドの画分が前方又は後方にシフトする現象を指す。クラスタ認識塩基コーリングシステム106は、クラスタのためのシグナル（入力シグナル x ）及びクラスタ固有のフェージング係数（入力係数 h ）の畳み込みに基づいて、補正されたシグナル（出力シグナル y ）を生成することができる。より詳細には、クラスタ固有のフェージング係数（ h ）は、クラスタ固有のプリフェージング係数とクラスタ固有のフェージング係数の両方を含む。補正されたシグナルは、畳み込み演算 $y_c = \sum_i h_i x_{c-i}$ としてモデル化することができ、これは $y = x * h$ として書かれる。シグナル減衰がないと仮定すると、クラスタ固有の係数 h は、 $\sum_i h_i = 1$ 、 $h_i \geq 0$ によって制約される。シグナル処理及び通信システムの文献では、 D 変換表記を使用することが一般的であり、 D^k は、 k サイクルの遅延を示す： $h(D) = \dots + h_{-2}D^{-2} + h_{-1}D^{-1} + h_0 + h_1D + h_2D^2 + \dots$ 。書かれるように、 $h_{-2}D^{-2} + h_{-1}D^{-1}$ は、現在のサイクルの2サイクル及び1サイクル前のヌクレオチド塩基に対応するフェージング係数を表す。 $h_1D + h_2D^2$ は、現在のサイクルから1サイクル及び2サイクル後のヌクレオチド塩基に対応するプリフェージング係数を表す。

20

【0105】

図6に図示されるように、クラスタ認識塩基コーリングシステム106は、クラスタ固有の係数演算606を実行して、エラー誘導配列に続くリード位置を有する各クラスタについて、クラスタ固有のフェージング係数及びクラスタ固有のプリフェージング係数を決定する。例示すると、クラスタ認識塩基コーリングシステム106は、前のサイクル（ h_{-1} ）、現在のサイクル（ h_0 ）、及び後続のサイクル（ h_1 ）に対応する様々なクラスタ固有のフェージング係数（ h ）を決定する。クラスタ固有のフェージング係数は、クラスタにわたって独立して変動し、いくつかのクラスタに対して決定されなくてもよい（例えば、エラー誘導配列に先行する、又はエラー誘導配列内のリード位置において）。推定されたフェージング又はプリフェージングによって影響を受けない大部分のクラスタは、値 $h = [0 \ 1 \ 0]$ を有する。しかしながら、クラスタ認識塩基コーリングシステム106は、クラスタ固有のフェージング係数が、ホモポリマーなどのエラー誘導配列の後にランダムかつ突然に変化することを決定することができる。いくつかの実施形態では、クラスタ固有のフェージング係数は、合計して1になり、関数 $\sum_i h_i(c) = 1$ 、 $h_i \geq 0$ によって表されるように負ではない。

30

40

【0106】

図6に更に図示されるように、クラスタ認識塩基コーリングシステム106は、マルチクラスタ係数演算608を実行して、マルチクラスタフェージング係数を決定する。クラスタ認識塩基コーリングシステム106は、ヌクレオチド試料スライドの特定のセクション（例えば、フローセルのタイル）内のクラスタにわたるマルチクラスタフェージング係数を利用することができる。マルチクラスタフェージング係数値は、サイクルごとに徐々に変化し得る。これらの値は、統計値を何百万ものクラスタにわたって平均化することが

50

できるため、クラスタ固有のフェージング係数よりも正確に推定するのが簡単である。

【0107】

図6に示されるように、例えば、クラスタ認識塩基コーリングシステム106は、前のサイクル(g_{-1})、現在のサイクル(g_0)、及び後続のサイクル(g_1)に対応する様々なマルチクラスタフェージング係数(g)を計算する。クラスタ固有のフェージング係数と同様に、マルチクラスタフェージング係数(g)は、合計して1になり、関数 $\sum_i g_i(c) = 1$ 、 $g_i \geq 0$ によって表されるように負ではない。図6に図示されるように、クラスタ認識塩基コーリングシステム106は、クラスタ固有のフェージング補正(クラスタ固有のフェージング係数を含む)及びマルチクラスタフェージング補正(マルチクラスタフェージング係数を含む)の両方に基づいてシグナルを調整する。

10

【0108】

いくつかの実施形態では、クラスタ認識塩基コーリングシステム106は、クラスタ固有の係数演算606及びマルチクラスタ係数演算608の両方をクラスタに適用する。更に、又は代替として、クラスタ認識塩基コーリングシステム106は、いくつかのクラスタに、マルチクラスタ係数演算608を適用するが、クラスタ固有の係数演算606は適用しない。特に、いくつかの実施形態では、クラスタ認識塩基コーリングシステム106は、クラスタ固有のフェージング補正を伴わずに、マルチクラスタフェージング補正に基づいて、1つ以上のクラスタからのシグナルを調整する。例えば、前述したように、エラー誘導配列に先行するヌクレオチド塩基に対するシグナルは、シグナルがエラー誘導配列によって影響を受けていないため、クラスタ固有のフェージング補正を必要としない場合がある。したがって、いくつかの実施形態では、クラスタ認識塩基コーリングシステム106は、オリゴヌクレオチドの追加のクラスタについて、異なるヌクレオチド断片リード内のエラー誘導配列に先行する異なるリード位置を特定する。クラスタ認識塩基コーリングシステム106は、異なるリード位置に対応するサイクル中に、オリゴヌクレオチドの追加のクラスタ内の標識ヌクレオチド塩基から追加のシグナルを更に検出する。次いで、クラスタ認識塩基コーリングシステム106は、オリゴヌクレオチドの追加のクラスタについてのクラスタ固有のフェージング補正を伴わずに、マルチクラスタフェージング補正に基づいて更なるシグナルを調整する。

20

【0109】

更に他の実施形態では、クラスタ認識塩基コーリングシステム106は、マルチクラスタ係数演算608を実行することなく、クラスタ固有の係数演算606を所与のクラスタのシグナルに適用する。例えば、場合によっては、クラスタ認識塩基コーリングシステム106は、マルチクラスタ係数演算から生じるパラメータを適用することなく、所与のクラスタについてのクラスタ固有のフェージング係数及びクラスタ固有のプリフェージング係数(又は他のパラメータ)を所与のクラスタについてのシグナルに適用する。したがって、ヌクレオチド試料スライド内のクラスタを処理する場合、クラスタ認識塩基コーリングシステム106は、クラスタ固有のフェージング補正(マルチクラスタフェージング補正なし)を所与のクラスタについてのシグナルに適用することができるが、クラスタ固有のフェージング補正及びマルチクラスタフェージング補正を異なるクラスタについてのシグナルに適用することができる。

30

40

【0110】

前述したように、クラスタ認識塩基コーリングシステム106は、シグナル処理604の一部として、クラスタ固有のフェージング係数及びマルチクラスタフェージング係数に基づいてシグナルを調整する。特に、図6に図示されるように、クラスタ認識塩基コーリングシステム106は、シグナル処理604の一部としてマルチクラスタフェージング補正610を実行する。クラスタ認識塩基コーリングシステム106は、アルゴリズム(FIRアルゴリズムなど)とともにマルチクラスタ係数演算608から生成されたマルチクラスタフェージング係数を利用して、マルチクラスタフェージング補正610を実行する。例えば、クラスタ認識塩基コーリングシステム106は、前のサイクル(γ_{-1})、現在のサイクル(γ_0)、及び後続のサイクル(γ_1)に対応する補正(γ)に基づいてシ

50

グナルを調整する。

【0111】

図6に更に図示されるように、クラスタ認識塩基コーリングシステム106は、シグナル処理604の一部としてクラスタ固有のフェージング補正及び塩基コーリング612を実行する。特に、クラスタ固有のフェージング補正及び塩基コーリング612の一部として、クラスタ認識塩基コーリングシステム106は、クラスタ固有の係数演算606の一部として生成されたクラスタ固有のフェージング係数を利用して、クラスタ固有のフェージング補正を推定し、シグナルに適用する。いくつかの実施形態では、クラスタ認識塩基コーリングシステム106は、FIRアルゴリズムなどのアルゴリズムとともにクラスタ固有のフェージング係数を利用して、クラスタ固有のフェージング補正を実行する。更に、図6に図示されるように、クラスタ認識塩基コーリングシステム106はまた、塩基コーリングを実行する。特に、クラスタ認識塩基コーリングシステム106は、調整されたシグナルに基づいてヌクレオチド塩基コールを生成する。

10

【0112】

前述したように、クラスタ認識塩基コーリングシステム106は、いくつかのモデル又はアルゴリズムを利用して、クラスタ固有のフェージング係数及びクラスタ固有のプリフェージング係数を決定することができる。より具体的には、クラスタ認識塩基コーリングシステム106は、様々なモデルを利用して、クラスタ固有の係数演算606を実行することができる。特に、クラスタ認識塩基コーリングシステム106は、線形等化器(LE)、判定帰還型等化器(DFE)、最尤系列推定器(MLSE)、又は順方向-逆方向モデルを利用して、クラスタ固有のフェージング係数及びクラスタ固有のプリフェージング係数を決定することができる。更に、クラスタ認識塩基コーリングシステム106は、多層パーセプトロンなどの機械学習モデルを利用して、係数を決定することができる。

20

【0113】

図7A~図7C及び対応する段落は、1つ以上の実施形態によるクラスタ認識塩基コーリングシステム106がLE、DFE、又はMLSEをどのように利用するかを詳述する。概して、クラスタ認識塩基コーリングシステム106は、様々な受信機タイプ及びコンピューティングアーキテクチャを使用して、クラスタ固有のフェージング係数及びクラスタ固有のプリフェージング係数を推定することができる。より具体的には、クラスタ認識塩基コーリングシステム106は、配列決定ランの過程で経時的に係数を生成し更新することができる。上述したように、クラスタ認識塩基コーリングシステム106は、受信機として、LE、DFE、及びMLSEの3つのモデル又はアルゴリズムのうちの少なくとも1つを利用することができる。いくつかの実施形態では、クラスタ認識塩基コーリングシステム106は、順方向-逆方向モデル及び／又は機械学習モデルを利用して、クラスタ固有のフェージング係数及びクラスタ固有のプリフェージング係数を推定する。更に、いくつかの実施形態では、クラスタ認識塩基コーリングシステム106は、最小二乗誤差又は他の最適化を使用して、クラスタ固有のフェージング係数及びクラスタ固有のプリフェージング係数を導出する。

30

【0114】

クラスタ認識塩基コーリングシステム106は、リアルタイム(又はほぼリアルタイム)コンピューティングアーキテクチャ又はバッファ型コンピューティングアーキテクチャを更に利用することができる。クラスタ認識塩基コーリングシステム106は、リアルタイムコンピューティングアーキテクチャを利用して、全ての将来のサイクルデータにアクセスすることなく、各サイクルにおいて最終塩基コールを出力する。例えば、いくつかの実施形態では、クラスタ認識塩基コーリングシステム106は、リアルタイムコンピューティングアーキテクチャを利用するために限られたシグナルデータのみを必要とする。更に、又は代替として、クラスタ認識塩基コーリングシステム106は、バッファ型コンピューティングアーキテクチャを利用する。クラスタ認識塩基コーリングシステム106は、最終的な塩基コールを行う前に全てのサイクルからのシグナルデータを利用することによって、バッファ型コンピューティングアーキテクチャを利用する。例えば、クラスタ認

40

50

識塩基コーリングシステム106は、バッファ型コンピューティングアーキテクチャを利用して、全ての前のサイクル及び後続のサイクルからのシグナルデータに基づいてクラスタに対するクラスタ固有のフェージング補正を生成することができる。クラスタ認識塩基コーリングシステム106は、異なる受信機タイプを異なるコンピューターアーキテクチャと組み合わせることができる。例えば、クラスタ認識塩基コーリングシステム106は、単純なリアルタイム線形等化器又は最も複雑なバッファ型MLSEを利用することができる。

【0115】

概して、リアルタイムコンピューティングアーキテクチャは、リアルタイム（又はほぼリアルタイム）情報のみを使用することによってコンピューティングの複雑さを制限する。例示すると、クラスタ認識塩基コーリングシステム106がリアルタイムコンピューティングアーキテクチャを利用するとき、クラスタ認識塩基コーリングシステム106は、1つ以上の前のサイクル、現在のサイクル、及び1つ以上の後続のサイクルについてのシグナルデータのみを必要とする。いくつかの実施形態では、クラスタ認識塩基コーリングシステム106は、前のサイクルからのシグナル伝達データのセットと、後続のデータからのシグナル伝達データのセットとを利用する。リアルタイムコンピューティングアーキテクチャは、計算効率がより高いため、クラスタ認識塩基コーリングシステム106は、配列決定装置114などの配列決定マシン又は装置のプロセスを利用するリアルタイムコンピューティングアーキテクチャを利用して演算を実行することができる。

【0116】

対照的に、いくつかの実施形態では、クラスタ認識塩基コーリングシステム106は、配列決定装置がヌクレオチド試料スライド上のオリゴヌクレオチドのクラスタに対するヌクレオチド断片リードを決定した後に、クラスタ固有のフェージング補正をオフラインで決定する。例えば、MLSE又は機械学習モデルを使用するいくつかの場合において、クラスタ認識塩基コーリングシステム106は、配列決定装置が所与のクラスタについてのヌクレオチド断片リードを決定した後、異なるコンピューティング装置上で、所与のクラスタについてのクラスタ固有のフェージング係数及びクラスタ固有のプリフェージング係数を決定し、所与のクラスタに対応するシグナルを調整する。

【0117】

対照的に、バッファ型コンピューティングアーキテクチャは、より多くのコンピューティングリソースを必要とする傾向がある。しかしながら、クラスタ認識塩基コーリングシステム106は、バッファ型コンピューティングアーキテクチャを利用することによって、より正確な結果を生成し得る。例示すると、バッファ型コンピューティングアーキテクチャを利用することによって、クラスタ認識塩基コーリングシステム106は、多数のクラスタ及びサイクルを並列に処理する。このタイプの処理は、クラスタごとのフェージング及びプリフェージング推定のために、大量の記憶、通信、及びコンピューティングリソースを必要とする。しかしながら、バッファ型コンピューティングアーキテクチャを利用することはまた、クラスタ認識塩基コーリングシステム106が全てのサイクルについてシグナル伝達データを処理するため、より正確な結果をもたらし得る。いくつかの実施形態では、クラスタ認識塩基コーリングシステム106は、配列決定マシン又は装置がオンラインであり、中央処理システムとアクティブに通信しているときに、バッファ型コンピューティングを実行する。

【0118】

言及したように、図7Aは、クラスタ固有のフェージング係数及びクラスタ固有のプリフェージング係数を決定するために線形等化器（LE）を利用するクラスタ認識塩基コーリングシステム106を図示する。概して、LEは、シンボル間干渉（intersymbol interference、ISI）を抑制するか、又はノイズを除去するように設計又は最適化することができる線形フィルタである。ISIは、1つのシンボルが後続のシンボルに干渉するシグナルの歪みの形態を指す。他のシンボルの効果は、ノイズと同様の効果を有する可能性があり、したがって、通信の信頼性を低下させる。クラスタ認識塩基コーリングシ

システム 106 は、LE を最適化して、ISI を抑制することとノイズ増幅を最小化することとの間の適切なトレードオフを見出すことができる。いくつかの実施形態では、クラスタ認識塩基コーリングシステム 106 は、FIR フィルタとして実装される線形等化器を利用する。そのような等化器を利用して、クラスタ認識塩基コーリングシステム 106 は、フィルタ係数によって入力信号の現在の値及び前の値を線形に重み付けする。例えば、いくつかの実施形態では、現在の値及び前の値は、クラスタからの現在の信号及び前の信号を含む。クラスタ認識塩基コーリングシステム 106 は更に、重み付けされた現在の値と前の値とを合計して、調整された信号を生成する。

【0119】

図 7A は、1 つ以上の実施形態による線形等化器アーキテクチャ 700 を図示する。概して、クラスタ認識塩基コーリングシステム 106 は、入力信号 x を線形等化器アーキテクチャ 700 に入力して、調整された信号

【0120】

【数 2】

$\hat{x}$

を生成する。前述したように、 h は、クラスタ固有のフェージング係数を表す。したがって、 $h(D)$ は、第 1 のフィルタを表す。加法性ノイズは、 $n \sim CN(0, \sigma^2)$ によって表される。図 7A に更に図示されるように、 w は重みを表し、 $w(D)$ は第 2 のフィルタを表す。クラスタ認識塩基コーリングシステム 106 は、判定装置 702 を更に利用して信号を処理し、調整された信号

【0121】

【数 3】

$\hat{x}$

を生成する。

【0122】

図 7A に示される LE 構造における h を決定するために、 $S(f)$ を周波数ドメイン SINR とし、

【0123】

【数 4】

$$S(f) = \frac{|F(h)|^2}{\sigma^2}$$

ここで、 $F(h)$ は、 $h(D)$ のフーリエ変換を表す。クラスタ認識塩基コーリングシステム 106 は、信号対干渉 + ノイズ比 (Signal to Interference plus Noise Ratio、SINR) を決定することによって信号クオリティの尺度を生成することができる。ガウスノイズを仮定すると、SINR 比を使用して、バイナリ信号又は他の変調タイプのエラー率を導出することができる。理想的な無限長の不偏最小平均二乗誤差線形等化器 (unbiased minimum-mean-squared-error linear equalizer、UMMSE-LE) に対して、以下のように示すことができる

【0124】

【数 5】

$$SINR_{U-MMSE-LE} = \left(\int_{-0.5}^{0.5} (1 + S(f))^{-1} \right) - 1$$

10

20

30

40

50

エラー率は、以下によって近似させることができ、

【0125】

【数6】

$$P_{error} \approx 2Q\left(\sqrt{SINR_{U-MMSE-LE}}\right), \text{ ここで } Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^{\infty} e^{-x^2/2} dx$$

ここで、 P_{error} は、エラーの転送パワーを表す。図7A及び対応する関数によって示唆されるように、周波数帯域にわたるシグナル及びノイズレベルが与えられると、クラスタ認識塩基コーリングシステム106は、受信機処理後に総SNRを計算し、その後、SNRをエラー率推定値に変換する。

10

【0126】

いくつかの実施形態では、クラスタ認識塩基コーリングシステム106は、3タップLEを利用して、前のサイクルの重み、後続のサイクルの重み、及び現在のサイクルの重みを生成する。特に、クラスタ認識塩基コーリングシステム106は、クラスタ固有のフェージング係数に基づいて、前のサイクルについてのヌクレオチド塩基のフェージング効果を推定する前のサイクルの重みを生成する。クラスタ認識塩基コーリングシステム106はまた、クラスタ固有のプリフェージング係数に基づいて、後続のサイクルについてのヌクレオチド塩基のプリフェージング効果を推定する後続のサイクルの重みを生成する。更に、クラスタ認識塩基コーリングシステム106はまた、クラスタ固有のフェージング係数及びクラスタ固有のプリフェージング係数に基づいて、フェージング効果及びプリフェージング効果を推定する現在のサイクルの重みを生成する。

20

【0127】

いくつかの実施形態では、クラスタ認識塩基コーリングシステム106は、前のサイクルの重み (w_{-1})、現在のサイクルの重み (w_0)、及び後続のサイクルの重み (w_1) を決定する。概して、クラスタ認識塩基コーリングシステム106は、最小二乗誤差又は別の最適化アルゴリズムなどの最適化アルゴリズムを使用してパラメータを最適化することができる。例えば、クラスタ認識塩基コーリングシステム106は、判定指向型の最小の最小二乗推定値を生成することができる。

【0128】

判定指向型の最小の最小二乗推定値を生成した後、又は別様にパラメータを最適化した後、クラスタ認識塩基コーリングシステム106は、次いで、中間統計を使用して、クラスタ固有のフェージング係数 (a) 及びクラスタ固有のプリフェージング係数 (b) を計算してもよい。特に、クラスタ認識塩基コーリングシステム106は、いくつかのサイクルにわたって、及び1つ以上のチャンネルにわたって二乗誤差を最小化することの一部である中間統計を利用する。チャンネルごとのサイクルごとに全ての値を維持する代わりに、クラスタ認識塩基コーリングシステム106は、実行統計を効率的に蓄積する。

30

【0129】

クラスタ固有のフェージング係数 (a) 及びクラスタ固有のプリフェージング係数 (b) に基づいて、クラスタ認識塩基コーリングシステム106は、次いで、前のサイクルの重み (w_{-1})、現在のサイクルの重み (w_0)、及び後続のサイクルの重み (w_1) を決定する。クラスタ認識塩基コーリングシステム106は、推定された重みの各々を各クラスタからのシグナルに適用する。いくつかの実施形態では、クラスタ認識塩基コーリングシステム106は、以下のように重み (w) を推定する。

40

$$\{w_{-1}, w_0, w_1\} = \{-a, 1+a+b, -b\}$$

【0130】

上記の関数及び本明細書中の他の関数が示唆するように、いくつかの実施形態では、クラスタ認識塩基コーリングシステム106は、配列決定サイクルにおけるオリゴヌクレオチドの所与のクラスタについてのクラスタ固有のフェージング係数及びクラスタ固有のプリフェージング係数 (及び対応する重み) を決定し、次いで、後続の配列決定サイクルに

50

おけるオリゴヌクレオチドの所与のクラスタについての更新されたクラスタ固有のフェージング係数及び更新されたクラスタ固有のプリフェージング係数（及び対応する重み）を決定することができ、各後続のサイクルについても同様である。実際、クラスタ認識塩基コーリングシステム 106 は、所与のクラスタに対応するヌクレオチド断片リードに対するヌクレオチド塩基コールを決定する過程にわたって、オリゴヌクレオチドの所与のクラスタに対するクラスタ固有のフェージング係数及びクラスタ固有のプリフェージング係数を再決定し、変更することができる。したがって、場合によっては、クラスタ認識塩基コーリングシステム 106 は、単に所与のクラスタについてクラスタ固有のフェージング係数及びクラスタ固有のプリフェージング係数を 1 回決定するのではなく、配列決定サイクルが進行するにつれて、所与のクラスタについてそのようなクラスタ固有のフェージング係数及びクラスタ固有のプリフェージング係数を繰り返し決定し、更新する。

10

【0131】

前に説明したように、クラスタ認識塩基コーリングシステム 106 は、クラスタ固有のフェージング係数及びクラスタ固有のプリフェージング係数を決定するために、判定帰還型等化器（DFE）を利用することもできる。図 7B 及び対応する段落は、1 つ以上の実施形態による、クラスタ認識塩基コーリングシステム 106 が DFE 及び判定帰還型等化器アーキテクチャ 706 をどのように利用するかを図示する。概して、DFE は、現在のシグナルを補正するために前のシグナルのレベルに関する決定に依存する非線形等化の形態である。特に、クラスタ認識塩基コーリングシステム 106 は、DFE を利用して、以前の決定をトレーニング配列として採用する。これにより、クラスタ認識塩基コーリングシステム 106 は、以前のシグナルによって引き起こされた現在のシグナルにおける歪みを考慮することができる。いくつかの実施形態では、DFE は、フィードフォワードフィルタ（feed forward filter、FFF）及びフィードバックフィルタ（feedback filter、FBF）を備える。FFF は、出力が判定装置に与えられる線形等化器を備えることができる。FBF は、判定装置の出力によって駆動される。

20

【0132】

特に、図 7B に図示されるように、クラスタ認識塩基コーリングシステム 106 は、入力シグナル x を判定帰還型等化器アーキテクチャ 706 に入力して、調整されたシグナル

【0133】

【数 7】

30

$\hat{x}$

を生成する。図示されるように、判定帰還型等化器アーキテクチャ 706 は、クラスタ固有のフェージング係数 h に対応するフィードフォワードフィルタ $h(D)$ を含む。シグナル x に対する加法性ノイズは、 $n \sim CN(0, \sigma^2)$ によって表される。判定帰還型等化器アーキテクチャ 706 は、シグナルを処理する判定装置 708 を更にも含む。概して、判定装置 708 は、ノイズが所定の値を超えるか否かを決定する。判定帰還型等化器アーキテクチャ 706 は、フィードバックフィルタ $b(D)$ を更にも含む。

40

【0134】

無限長の不偏最小平均二乗誤差判定帰還型等化器（U-MMSE-DFE）に対して、以下のように示すことができ、

【0135】

【数 8】

$$SINR_{U-MMSE-DFE} = \exp\left(\int_{-0.5}^{0.5} \log(1 + S(f)) df\right) - 1$$

50

正しい（ジニーによる）判定を仮定する。 $S(f)$ は、 (i) チャネルのフーリエ変換の二乗された大きさの、 (ii) 周波数帯域にわたるノイズパワーに対する比を表す。 $S(f)$ が与えられると、クラスタ認識塩基コーリングシステム106は、スライサにおいて、又はスライサを使用してSINRを計算することができ、クラスタ認識塩基コーリングシステム106は、スライサを利用して、バイナリシグナルのビットエラー率を推定する。前述したように、クラスタ認識塩基コーリングシステム106は、シグナル対干渉＋ノイズ比（SINR）を決定することによってシグナルクオリティの尺度を生成することができる。この式はシャノン限界（Shannon Limit）に関連することが分かる。

【0136】

【数9】

$$C = \int_{-0.5}^{0.5} \log(1 + S(f)) df = \log(1 + SINR_{U-MMSE-DFE})$$

10

【0137】

チャネル容量（C）は、加法的白色ガウスノイズの影響を受けるアナログ通信チャネルを介して平均受信シグナル電力（S）を使用して任意の低いエラー率で通信することができるデータの情報理論的に最も厳しい上限を表す。実世界の通信システムでは、シャノン限界は、強い符号、ガウスコンステレーション整形、及びプレコーディングを組み合わせることによって近づくことができる。符号化されていないQPSKの場合、エラー伝搬は不可避であり、エラー率は以下の式によってより低く制限される：

20

【0138】

【数10】

$$P_{error} > \sim 2Q(\sqrt{SINR_{U-MMSE-DFE}})$$

ここで、 P_{error} は、エラーの転送パワーを表す。

【0139】

更に他の実施形態では、クラスタ認識塩基コーリングシステム106は、第3のタイプの受信機である最尤系列推定器（MLSE）を利用して、クラスタ固有のフェージング係数及びクラスタ固有のプリフェージング係数を決定する。図7Cは、1つ以上の実施形態による最尤系列推定器アーキテクチャ710を図示する。MLSEは、等化フィルタをMLSE推定で置き換える非線形推定技法である。概して、クラスタ認識塩基コーリングシステム106は、MLSEを利用して、（各受信シグナルをそれ自体で復号するのではなく）全ての可能なデータ配列を試験し、最大確率を有する出力シグナルを出力として選択する。MLSEは、ビタビ復号器712を使用して、全ての可能な転送配列の確率を決定する。図7Cに図示されるように、クラスタ認識塩基コーリングシステム106は、入力シグナルxを最尤系列推定器アーキテクチャ710に入力して、調整されたシグナル

30

【0140】

【数11】

$\hat{x}$

40

を生成する。最尤系列推定器アーキテクチャ710は、クラスタ固有のフェージング係数hに対応するフィルタh(D)を含む。シグナルxに対する加法的ノイズは、 $n \sim CN(0, \sigma^2)$ によって表される。

【0141】

図7Cに図示されるように、エラー率は、以下のように、整合フィルタ限界（Matched Filter Bound、MFB）によって制限される：

50

【 0 1 4 2 】

【 数 1 2 】

$$P_{error} \geq 2Q(\sqrt{SNR_{MFB}})$$

$$SNR_{MFB} = \frac{\sum_i |h_i|^2}{\sigma^2} = \int_{-0.5}^{0.5} S(f) df$$

10

ここで、 SNR はシグナル対ノイズ比を表し、 P_{error} はエラーの転送パワーを表す。概して、 SNR は、所望のシグナルのレベルをバックグラウンドノイズのレベルと比較する。図 7 C 及び対応する関数によって示されるように、クラスタ認識塩基コーリングシステム 106 は、パーセバルの定理を利用して、時間ドメインにおける応答を合計することによって総シグナルパワーを決定する。総シグナルパワーは、周波数ドメインにおける総パワーと同一又は同等であり得る。クラスタ認識塩基コーリングシステム 106 が SNR を決定すると、クラスタ認識塩基コーリングシステム 106 はエラー限界を計算する。図 7 C に対応する上記の関数では、状態の数は $N_{長さ(h)} - 1$ によって与えられ、ここで N はコンステレーション点の数である。無相関ノイズを有する正方形コンステレーションの場合、2つの SBS チャンネルを独立して処理することができ、状態の数を低減する。

20

【 0 1 4 3 】

上述したように、クラスタ認識塩基コーリングシステム 106 は、図 7 A ~ 7 C に図示される受信機 LE 、 $D FE$ 、及び $ML SE$ に加えて、他のモデルを利用することができる。より具体的には、クラスタ認識塩基コーリングシステム 106 は、上に列挙したものに加えて、他の隠れマルコフモデル (Hidden Markov Model、 HMM) を利用することができる。例えば、いくつかの実施形態では、クラスタ認識塩基コーリングシステム 106 は、順方向-逆方向モデルを利用して、最大事後確率 (maximum a posteriori probability、 $M A P$) 推定値を生成することができる。順方向-逆方向モデルは、所与の時間における各状態の事後最大パス確率を計算する。概して、順方向-逆方向モデルは、動的計画法の原理を利用して、2つのパスで事後限界分布を得るのに必要な値を計算する。第 1 のパスは時間的に順方向に進み、第 2 のパスは時間的に逆方向に進む。

30

【 0 1 4 4 】

上に列挙したモデルに加えて、クラスタ認識塩基コーリングシステム 106 は、機械学習モデルを利用して、クラスタ固有のフェージング係数及びクラスタ固有のプリフェージング係数を決定することができる。概して、クラスタ認識塩基コーリングシステム 106 は、機械学習モデルを使用して、クラスタ固有のフェージング係数及びクラスタ固有のプリフェージング係数を推定し、得られたシグナルを調整するか、又はヌクレオチド塩基コールを直接調整することができる。例示すると、いくつかの実施形態では、クラスタ認識塩基コーリングシステム 106 は、畳み込み層に基づく配列間機械学習モデルを利用する。更に、又は代替として、クラスタ認識塩基コーリングシステム 106 は、長・短期記憶 (Long Short-Term Memory、 $LSTM$) などの回帰型ニューラルネットワーク (Recurrent Neural Network、 RNN) を利用して、クラスタ固有のフェージング係数及びクラスタ固有のプリフェージング係数を推定することができる。更に他の実施形態では、クラスタ認識塩基コーリングシステム 106 は、アテンションベースモデルを利用する。

40

【 0 1 4 5 】

図 7 A ~ 図 7 C は、クラスタ認識塩基コーリングシステム 106 が、1つ以上の実施形態によるクラスタ固有のフェージング補正を決定するために利用する、異なる受信機を図

50

示する。図 8 A～図 8 B は、1 つ以上の実施形態によるリアルタイム L E 及びバッファ型 M L S E を利用するクラスタ認識塩基コーリングシステム 1 0 6 から生じる技術的改善を図示する。特に、図 8 A は、補正なし、リアルタイム L E、及びバッファ型 M L S E に対応する例示的なリードパイルアップを図示する。図 8 B は、クラスタ固有のフェージング補正からの二次配列決定メトリクスにおいて大きな利得を示すクラスタを図示する。

【 0 1 4 6 】

言及したように、図 8 A は、補正なし、リアルタイム L E、及びバッファ型 M L S E に対応する 3 つのリードパイルアップを図示する。特に、図 8 A は、補正されていないリードパイルアップ 8 0 2、リアルタイム線形等化器によるクラスタ固有のフェージング補正を使用して調整されたシグナルからのヌクレオチド塩基コールを有するリードパイルアップ 8 0 4、及びバッファ型 M L S E によるクラスタ固有のフェージング補正を使用して調整されたシグナルからのヌクレオチド塩基コールを有するリードパイルアップ 8 0 6 を図示する。補正されていないリードパイルアップ 8 0 2 は、図 2 A に図示されたリードパイルアップ 2 0 0 と同様である。特に、補正されていないリードパイルアップ 8 0 2 は、塩基コール精度がエラー誘導配列の後に低下することを反映する。例示すると、図 8 A において、補正されていないエラータイプカウンタ 8 0 8 は、エラー誘導配列を取り囲む塩基コールエラーの発生の増加を示す。

【 0 1 4 7 】

図 8 A は更に、リアルタイム線形等化器を使用することによって、クラスタ認識塩基コーリングシステム 1 0 6 が塩基コールエラーの発生を減少させることを図示する。特に、リアルタイム線形等化器によるクラスタ固有のフェージング補正を使用して調整されたシグナルからのヌクレオチド塩基コールを有するリードパイルアップ 8 0 4 は、補正されていないリードパイルアップ 8 0 2 よりも、エラー誘導配列を取り囲む場合であっても、より少ない塩基コールエラーを示す。例えば、補正されていないエラータイプカウンタ 8 0 8 と比較すると、線形等化器エラータイプカウンタ 8 1 0 は、より少なく、かつより短いバーを含む。図 8 A に図示されるように、リアルタイム L E を使用してクラスタ固有のフェージング補正を決定することによって、クラスタ認識塩基コーリングシステム 1 0 6 は、補正されていないリードパイルアップ 8 0 2 においてエラー（又は不正確なヌクレオチド塩基コール）として示されるヌクレオチド塩基コールの約 7 0 % を正確に決定する。しかしながら、エラー誘導配列と高度に相関するいくつかの塩基コールエラーが依然として存在する。例えば、リードパイルアップ 8 0 4 は、エラー誘導配列のすぐ周囲の塩基にいくつかの塩基コールエラーを依然として含む。

【 0 1 4 8 】

前述したように、計算効率が低いことが多いが、クラスタ認識塩基コーリングシステム 1 0 6 は、バッファ型 M L S E を使用することによって、リアルタイム線形等化器を使用する場合と比較しても、ヌクレオチド塩基コールの精度を改善することができる。図 8 A は更に、バッファ型 M L S E エラータイプカウンタ 8 1 2 を有するリードパイルアップ 8 0 6 を図示する。バッファ型 M L S E エラータイプカウンタ 8 1 2 は、バッファ型 M L S E を使用してクラスタ固有のフェージング補正を決定することによって、クラスタ認識塩基コーリングシステム 1 0 6 が、補正されていないリードパイルアップ 8 0 2 においてエラー（又は不正確なヌクレオチド塩基コール）として示されるヌクレオチド塩基コールの約 8 5 % を正確に決定することを示す。

【 0 1 4 9 】

図 8 A は、クラスタ固有のフェージング補正に従ってシグナルを調整することに基づくヌクレオチド塩基コール精度の改善を図示するが、図 8 B は、1 つ以上の実施形態によるクラスタ固有のフェージング補正に従ってシグナルを調整することに基づく二次配列決定メトリクスの改善を図示する。特に、図 8 B は、補正されていないシグナル、及び L E を利用するクラスタ固有のフェージング補正によって補正されたシグナルから得られる様々な二次配列決定メトリクスの比較を図示する。例えば、図 8 B は、補正されていない強度に対応する二次配列決定メトリクスを図示する。特に、図 8 B は、補正されていないグラ

フ 8 1 4、補正されていない強度分布 8 1 8、補正されていない S N R グラフ 8 2 0、及び補正されていないクオリティスコアグラフ 8 2 4を含む。図 8 B はまた、L E を利用するクラスタ固有のフェージング補正によって調整されたシグナルからの二次配列決定メトリクスを図示する。特に、図 8 B は、調整されたグラフ 8 1 6、調整された強度分布 8 2 6、調整された S N R グラフ 8 2 8、及び調整されたクオリティスコアグラフ 8 3 0を含む。

【 0 1 5 0 】

図 8 B に示されるように、L E の利用は、クラスタ認識塩基コーリングシステム 1 0 6 が、以前の配列決定システムよりも強度値境界に対してより良好な鮮明度を有するヌクレオチド塩基コールのためのシグナルを生成することを可能にする。特に、図 8 B は、補正されていない強度値境界 8 3 2 を含む補正されていないグラフ 8 1 4 と、調整された強度値境界 8 3 4 を含む調整されたグラフ 8 1 6 とを含む。前述したように、強度値境界は、各可能なヌクレオチド塩基（例えば、A、T、C、又は G）に対応する。図 8 B に図示されるように、クラスタ認識塩基コーリングシステム 1 0 6 は、補正されていないグラフ 8 1 4 よりも、調整されたグラフ 8 1 6 において、強度値境界に関してより良好な鮮明度値を有するヌクレオチド塩基コールのためのシグナルを生成する。図 8 B に図示されるように、調整されたグラフ 8 1 6 は、鮮明度フィルタを通過しない値を有するより少ない調整されたシグナルを示す。特に、フェージング及びプリフェージングを考慮してシグナルを調整した結果として、クラスタ認識塩基コーリングシステム 1 0 6 は、鮮明度フィルタに不合格となる値を有するシグナルの数を低減する。対照的に、補正されていないグラフ 8 1 4 は、補正されていない強度値境界 8 3 2 の外側に位置する三角形の数が、調整されたグラフ 8 1 6 内の調整された強度値境界 8 3 4 の外側の三角形の数よりも多いため、鮮明度フィルタに不合格となる値を有するノイズ又はシグナルのより高い発生を示す。

【 0 1 5 1 】

図 8 B の補正されていない強度分布 8 1 8 及び調整された強度分布 8 2 6 は、クラスタ認識塩基コーリングシステム 1 0 6 が、クラスタ固有のフェージング補正に基づいてシグナルを調整することによって、どのようにシグナル強度を明確にするかを図示する。概して、強度分布は、2つのチャネルの強度を1つの軸上に重ね合わせるように変換する。理想的には、2つのチャネルからのシグナルは良好な分離を有するべきであり、これはシグナルの明瞭さを示す。図 8 B に図示されるように、補正されていない強度分布 8 1 8 は、エラー誘導配列後のシグナル強度が混乱していることを示す。対照的に、調整された強度分布 8 2 6 は、エラー誘導配列の後であっても、シグナルのより明確な描写を示す。

【 0 1 5 2 】

図 8 B に更に図示されるように、クラスタ認識塩基コーリングシステム 1 0 6 はまた、L E を利用してシグナルを調整するためのクラスタ固有のフェージング補正を決定することによって、S N R メトリクスを改善する。特に、補正されていない S N R グラフ 8 2 0 は、リード位置 1 5 0 の直後のエラー誘導配列に続く S N R メトリクスの劇的な低下を示す。対照的に、調整された S N R グラフ 8 2 8 は、リード位置 1 5 0 の直後のエラー誘導配列の後であっても、S N R メトリクスのより小さい減少を示す。したがって、L E を利用することによって、クラスタ認識塩基コーリングシステム 1 0 6 は、S N R メトリクスを改善することができる。

【 0 1 5 3 】

図 8 B はまた、L E を利用してシグナルを調整するためのクラスタ固有のフェージング補正を決定することに基づいて、エラー誘導配列に続くサイクルにおけるクオリティスコアの改善を図示する。図示されるように、補正されていないクオリティスコアグラフ 8 2 4 は、クオリティスコアの劇的な低下を含む。いくつかの実施形態では、クラスタ認識塩基コーリングシステム 1 0 6 は、P h r e d (Q 3 0) クオリティスコアを測定する。エラー誘導配列に続くサイクルにおける偶発的なクオリティスコアピークを示す補正されていないクオリティスコアグラフ 8 2 4 とは対照的に、調整されたクオリティスコアグラフ 8 3 0 は、エラー誘導配列に続くサイクルにおける偶発的なディップを伴い、一貫してよ

り高いクオリティスコアを示す。

【0154】

図1～図8B、対応する本文、及び実施例は、クラスタ認識塩基コーリングシステム106のいくつかの異なる方法、システム、装置、及び非一時的コンピュータ可読媒体を提供する。上記に加えて、1つ以上の実施形態はまた、図9に示される動作のフローチャートなどの、特定の結果を達成するための動作を含むフローチャートに関して説明することができる。更に、本明細書で説明される動作は、互いに並行して、又は同じ若しくは同様の動作の異なる出現と並行して、繰り返されるか、又は実行され得る。

【0155】

図9は、クラスタ固有のフェージング補正に基づいてヌクレオチド塩基コールを決定するための一連の動作900のフローチャートを図示する。図9は、一実施形態による動作を図示するが、代替実施形態は、図9に示される動作のいずれかを省略、追加、再配列、及び／又は修正してもよい。図9の動作は、方法の一部として実行することができる。あるいは、非一時的コンピュータ可読媒体は、1つ以上のプロセッサによって実行されると、コンピューティング装置に図9の動作を実行させる命令を含むことができる。いくつかの実施形態では、システムは、図9の動作を実行することができる。

【0156】

1つ以上の実施形態では、一連の動作900は、図10に図示されるコンピューティング装置などの1つ以上のコンピューティング装置上で実施される。加えて、いくつかの実施形態では、一連の動作900は、核酸ポリマーを配列決定するためのデジタル環境において実施される。図9に図示されるように、一連の動作900は、エラー誘導配列に続くリード位置を特定する動作902、標識ヌクレオチド塩基からのシグナルを検出する動作904、クラスタ固有のフェージング補正を決定する動作906、シグナルを調整する動作908、及びヌクレオチド塩基コールを決定する動作910を含む。

【0157】

図9に図示される一連の動作900は、エラー誘導配列に続くリード位置を特定する動作902を含む。特に、動作902は、オリゴヌクレオチドのクラスタについて、1つ以上のヌクレオチド断片リード内のエラー誘導配列に続くリード位置を特定することを含む。1つ以上の実施形態では、エラー誘導配列は、1つ以上の反復ヌクレオチド塩基の配列又は配列モチーフを含む。更に、いくつかの実施形態では、1つ以上の反復ヌクレオチド塩基の配列又は配列モチーフは、同じヌクレオチド塩基のホモポリマー、ニアホモポリマー、グアニン四重鎖、可変数タンデム反復(VNTR)、ジヌクレオチド反復配列、トリヌクレオチド反復配列、逆方向反復配列、ミニサテライト配列、マイクロサテライト配列、又はパンドローム配列を含む。1つ以上の実施形態では、エラー誘導配列は、1つ以上の反復ヌクレオチド塩基の配列又は方向固有の配列モチーフを含む。

【0158】

図9は更に、標識ヌクレオチド塩基からのシグナルを検出する動作904を図示する。特に、動作904は、リード位置に対応するサイクル中にオリゴヌクレオチドのクラスタ内の標識ヌクレオチド塩基からのシグナルを検出することを含む。

【0159】

図9に図示される一連の動作900は、クラスタ固有のフェージング補正を決定する動作906を更に含む。特に、動作906は、オリゴヌクレオチドのクラスタに対して、推定されたフェージング及び推定されたプリフェージングのためにシグナルを補正するためのクラスタ固有のフェージング補正を決定することを含む。いくつかの実施形態では、動作906は、オリゴヌクレオチドのクラスタについて、前のサイクルについてのヌクレオチド塩基に対応するクラスタ固有のフェージング係数、及び後続のサイクルについてのヌクレオチド塩基に対応するクラスタ固有のプリフェージング係数を決定することを含む。いくつかの実施形態では、動作906は、オリゴヌクレオチドのクラスタについて、フェージング及びプリフェージングのためにシグナルを補正するクラスタ固有のフェージング補正を決定することを含む。1つ以上の実施形態では、クラスタ固有のフェージング補正

を決定するステップは、オリゴヌクレオチドのクラスタについて、サイクルの直前の前のサイクルについてのヌクレオチド塩基に対応するクラスタ固有のフェージング係数及びサイクルの直後の後続のサイクルについてのヌクレオチド塩基に対応するクラスタ固有のプリフェージング係数を決定することと、クラスタ固有のフェージング係数及びクラスタ固有のプリフェージング係数に基づいてクラスタ固有のフェージング補正を決定することとを含む。

【0160】

いくつかの実施形態では、動作906は、オリゴヌクレオチドのクラスタについて、前のサイクルについてのヌクレオチド塩基に対応するクラスタ固有のフェージング係数及び後続のサイクルについてのヌクレオチド塩基に対応するクラスタ固有のプリフェージング係数を決定することと、クラスタ固有のフェージング係数及びクラスタ固有のプリフェージング係数に基づいて、クラスタ固有のフェージング補正を決定することと、によって、クラスタ固有のフェージング補正を決定することを更に含む。更に、いくつかの実施形態では、動作906は、クラスタ固有のフェージング係数に基づいて前のサイクルの重みを生成し、前のサイクルに対するヌクレオチド塩基のフェージング効果を推定することと、クラスタ固有のプリフェージング係数に基づいて、後続のサイクルについてのヌクレオチド塩基のプリフェージング効果を推定する後続のサイクルの重みを生成することと、クラスタ固有のフェージング係数及びクラスタ固有のプリフェージング係数に基づいて、サイクルについてのフェージング効果及びプリフェージング効果を推定する現在のサイクルの重みを生成することと、前のサイクルの重み、後続のサイクルの重み、及び現在のサイクルの重みに基づいて、クラスタ固有のフェージング補正を決定することと、によって、クラスタ固有のフェージング係数及びクラスタ固有のプリフェージング係数に基づいてクラスタ固有のフェージング補正を決定することを更に含む。場合によっては、クラスタ固有のフェージング補正を決定することは更に、前のサイクルに対応するシグナル強度、現在のサイクルに対応するシグナル強度、及び後続のサイクルに対応するシグナル強度に基づく。

【0161】

同様に、いくつかの実施形態では、動作906は、クラスタ固有のフェージング係数に基づいて前のサイクルの重みを生成し、前のサイクルに対するヌクレオチド塩基のフェージング効果を推定することと、クラスタ固有のプリフェージング係数に基づいて、後続のサイクルについてのヌクレオチド塩基のプリフェージング効果を推定する後続のサイクルの重みを生成することと、クラスタ固有のフェージング係数及びクラスタ固有のプリフェージング係数に基づいて、サイクルについてのフェージング効果及びプリフェージング効果を推定する現在のサイクルの重みを生成することと、前のサイクルの重み、後続のサイクルの重み、及び現在のサイクルの重みに基づいて、クラスタ固有のフェージング補正を決定することと、クラスタ固有のフェージング補正をシグナルに適用することと、によって、クラスタ固有のフェージング係数及びクラスタ固有のプリフェージング係数に基づいてシグナルを調整することを更に含む。

【0162】

更に、いくつかの実施形態では、動作906は、オリゴヌクレオチドのクラスタについて、前のサイクルのセットについてのヌクレオチド塩基のセットに対応するクラスタ固有のフェージング係数のセットを決定することと、オリゴヌクレオチドのクラスタについて、後続のサイクルのセットについてのヌクレオチド塩基のセットに対応するクラスタ固有のプリフェージング係数のセットを決定することと、クラスタ固有のフェージング係数のセット及びクラスタ固有のプリフェージング係数のセットに基づいて、クラスタ固有のフェージング補正を決定することと、によって、クラスタ固有のフェージング補正を決定することを更に含む。いくつかの実施形態では、動作906は、配列決定装置のプロセッサを利用してクラスタ固有のフェージング補正を決定することを更に含む。

【0163】

いくつかの実施形態では、動作906は、システムの配列決定マシン上で、線形等化器

、判定帰還型等化器、最尤系列推定器、順方向－逆方向モデル、又は機械学習モデルを利用して、クラスタ固有のフェージング係数及びクラスタ固有のプリフェージング係数を決定することを更に含む。更に、いくつかの実施形態では、動作 906 は、配列決定ラン後にクラスタ固有のフェージング係数及びクラスタ固有のプリフェージング係数を決定することを更に含む。

【0164】

更に、1つ以上の実施形態では、動作 906 は、オリゴヌクレオチドのクラスタについて、サイクルの直前の前のサイクルのセットについてのヌクレオチド塩基のセットに対応するクラスタ固有のフェージング係数のセットを決定することと、オリゴヌクレオチドのクラスタについて、サイクルの直後の後続のサイクルのセットについてのヌクレオチド塩基のセットに対応するクラスタ固有のプリフェージング係数のセットを決定することと、クラスタ固有のフェージング係数のセット及びクラスタ固有のプリフェージング係数のセットに基づいて、クラスタ固有のフェージング補正を決定することと、を更に含む。

【0165】

図 9 に図示されるように、一連の動作 900 は、シグナルを調整する動作 908 を含む。特に、動作 908 は、クラスタ固有のフェージング補正に基づいてシグナルを調整することを含む。いくつかの実施形態では、動作 908 は、クラスタ固有のフェージング係数及びクラスタ固有のプリフェージング係数に基づいてシグナルを調整することを含む。更に、いくつかの実施形態では、動作 908 は、オリゴヌクレオチドのクラスタについて、追加の以前のサイクルの追加のヌクレオチド塩基に対応する追加のクラスタ固有のフェージング係数を決定することと、オリゴヌクレオチドのクラスタについて、追加の後続のサイクルのための追加のヌクレオチド塩基に対応する追加のクラスタ固有のプリフェージング係数を決定することと、クラスタ固有のフェージング係数、追加のクラスタ固有のフェージング係数、クラスタ固有のプリフェージング係数、及び追加のクラスタ固有のプリフェージング係数に基づいて、クラスタ固有のフェージング補正を決定することと、によって、シグナルを調整することを更に含む。

【0166】

一連の動作 900 はまた、ヌクレオチド塩基コールを決定する動作 910 を含む。特に、動作 910 は、調整されたシグナルに基づいて、オリゴヌクレオチドのクラスタに対応するリード位置についてのヌクレオチド塩基コールを決定することを含む。

【0167】

1つ以上の実施形態では、一連の動作 900 は、オリゴヌクレオチドのクラスタのセットについて、推定されたフェージング及び推定されたプリフェージングのためにクラスタのセットからのシグナルを補正するためのマルチクラスタフェージング補正を決定する追加の動作と、クラスタ固有のフェージング補正又はマルチクラスタフェージング補正に基づいてシグナルを調整する追加の動作と、を含む。いくつかの実施形態では、一連の動作 900 は、オリゴヌクレオチドのクラスタのセットについて、推定されたフェージングのためのマルチクラスタフェージング係数又は推定されたプリフェージングのためのマルチクラスタプリフェージング係数のうちの1つ以上を決定する追加の動作と、マルチクラスタフェージング係数、クラスタ固有のフェージング係数、マルチクラスタプリフェージング係数、又はクラスタ固有のプリフェージング係数のうちの1つ以上に基づいてシグナルを調整する追加の動作と、を含む。いくつかの実施形態では、一連の動作 900 は、オリゴヌクレオチドのクラスタのセットについて、マルチクラスタフェージング補正を決定して、フェージング及びプリフェージングのためにクラスタのセットからのシグナルを補正する動作と、クラスタ固有のフェージング補正及びマルチクラスタフェージング補正の両方に基づいてシグナルを調整する動作と、を更に含む。

【0168】

1つ以上の実施形態では、一連の動作 900 は、オリゴヌクレオチドのクラスタ及び後続のリード位置について、異なるクラスタ固有のフェージング補正を決定して、後続のサイクルについてのシグナルのフェージング及びプリフェージングのためにオリゴヌクレオ

チドのクラスタからの後続のサイクルについてのシグナルを補正する追加の動作を含む。

【0169】

いくつかの実施形態では、図9に図示される一連の動作900は、オリゴヌクレオチドの追加のクラスタについて、異なるヌクレオチド断片リード内のエラー誘導配列に先行する異なるリード位置を特定する追加の動作と、異なるリード位置に対応するサイクル中に、オリゴヌクレオチドの追加のクラスタ内の標識ヌクレオチド塩基から追加のシグナルを検出する追加の動作と、オリゴヌクレオチドの追加のクラスタについてのクラスタ固有のフェージング補正なしに、マルチクラスタフェージング補正に基づいて追加のシグナルを調整する追加の動作と、を含む。

【0170】

本明細書に記載の方法は、様々な核酸配列決定技術と併せて使用することができる。特に適用可能な技術は、核酸を、それらの相対的位置が変化しないようにアレイ内の固定位置に付着させ、アレイが繰り返し撮像されるものである。例えば、1つのヌクレオチド塩基型を別のヌクレオチド塩基型と区別するために使用される異なる標識と一致する異なる色チャンネルで画像が得られる実施形態は、特に適用可能である。いくつかの実施形態では、標的核酸のヌクレオチド配列を決定するプロセスは、自動化プロセスであり得る。好ましい実施形態は、合成による配列決定 (sequencing-by-synthesis「SBS」) 技術を含む。

【0171】

SBS技術は、一般に、鋳型鎖に対するヌクレオチドの反復的付加による、新生核酸鎖の酵素的伸長を伴う。SBSの従来方法では、単一のヌクレオチドモノマーが、各送達においてポリメラーゼの存在下で標的ヌクレオチドに提供され得る。しかしながら、本明細書に記載の方法では、送達中のポリメラーゼの存在下で、複数の種類のヌクレオチドモノマーを標的核酸に提供することができる。

【0172】

以下に記載されるSBS技術は、シングルリード配列決定又はペアエンド配列決定を利用することができる。シングルリード配列決定において、配列決定装置は、塩基対の配列を生成するために、一方の末端から他方の末端まで断片を読み取る。対照的に、ペアエンド配列決定中、配列決定装置は、1つの読み取りで開始し、同じ方向で指定された読み取り長さの読み取りを終了し、断片の反対端から別の読み取りを開始する。

【0173】

SBSは、ターミネーター部分を有するヌクレオチドモノマー、又は任意のターミネーター部分を欠くヌクレオチドモノマーを利用することができる。ターミネーターを欠くヌクレオチドモノマーを利用する方法としては、例えば、以下に更に詳細に記載されるように、γ-リン酸標識ヌクレオチドを使用するピロ配列決定及び配列決定が挙げられる。ターミネーターを含まないヌクレオチドモノマーを使用する方法では、各サイクルに添加されるヌクレオチドの数は、概ね可変であり、鋳型配列及びヌクレオチド送達のモードに依存する。ターミネーター部分を有するヌクレオチドモノマーを利用するSBS技術では、ターミネーターは、ジデオキシヌクレオチドを利用する従来Sanger配列決定の場合のように使用される配列決定条件下で有効に不可逆的であり得るか、又はターミネーターは、Solexa (現Illumina) によって開発された配列決定方法の場合のように可逆的であり得る。

【0174】

SBS技術は、標識部分を有するヌクレオチドモノマー、又は標識部分を欠くヌクレオチドモノマーを使用することができる。したがって、標識の蛍光などの標識の特性、分子量又は電荷などのヌクレオチドモノマーの特性、ピロリン酸の放出などのヌクレオチドの組み込みの副生成物などに基づいて、組み込みイベントを検出することができる。2つ以上の異なるヌクレオチドが配列決定試薬中に存在する実施形態では、異なるヌクレオチドは、互いに区別可能であり得るか、又は代替的に、2つ以上の異なる標識は、使用される検出技術の下で区別可能であり得る。例えば、配列決定試薬中に存在する異なるヌクレオ

10

20

30

40

50

チドは、異なる標識を有することができ、それらは、S o l e x a (現 I l l u m i n a) によって開発された配列決定方法によって例示される適切な光学系を使用して区別することができる。

【0175】

好ましい実施形態としては、パイロシーケンシング (パイロ配列決定) 技術が挙げられる。パイロシーケンシングは、特定のヌクレオチドが新生鎖に組み込まれるときに無機ピロリン酸塩 (inorganic pyrophosphate、P P i) の放出を検出する (R o n a g h i , M . , K a r a m o h a m e d , S . , P e t t e r s s o n , B . , U h l e n , M . a n d N y r e n , P . (1 9 9 6) 「R e a l - t i m e D N A s e q u e n c i n g u s i n g d e t e c t i o n o f p y r o p h o s p h a t e r e l e a s e . 」 A n a l y t i c a l B i o c h e m i s t r y 2 4 2 (1) , 8 4 - 9 、 R o n a g h i , M . (2 0 0 1) 「P y r o s e q u e n c i n g s h e d s l i g h t o n D N A s e q u e n c i n g . 」 G e n o m e R e s . 1 1 (1) , 3 - 1 1 、 R o n a g h i , M . , U h l e n , M . a n d N y r e n , P . (1 9 9 8) 「A s e q u e n c i n g m e t h o d b a s e d o n r e a l - t i m e p y r o p h o s p h a t e 」 , S c i e n c e 2 8 1 (5 3 7 5) , 3 6 3 、 米国特許第 6 , 2 1 0 , 8 9 1 号、同第 6 , 2 5 8 , 5 6 8 号及び同第 6 , 2 7 4 , 3 2 0 号、参照によりその開示の全体が本明細書に組み込まれる)。パイロシーケンシングにおいて、放出された P P i は、A T P スルフラゼによってアデノシン三リン酸 (adenosine triphosphate、A T P) に即座に変換されることによって検出することができ、生成された A T P のレベルはルシフェラーゼで生成された光子を介して検出される。配列決定される核酸は、アレイ中の特徴に結合させることができ、アレイは、アレイの特徴にヌクレオチドを組み込むことにより生成される化学発光シグナルを捕捉するために画像化することができる。アレイを特定のヌクレオチド型 (例えば、T、C、又は G) で処理した後に、画像を得ることができる。各ヌクレオチド型の添加後に得られる画像は、アレイ内のどの特徴が検出されるかに関して異なる。画像内のこれらの差異は、アレイ上の特徴の異なる配列コンテンツを反映する。しかしながら、各特徴の相対的な位置は、画像内で変わらないままである。画像は、本明細書に記載の方法を使用して記憶、処理、及び分析することができる。例えば、アレイを各異なるヌクレオチド型で処理した後に得られる画像は、可逆的ターミネーターベースの配列決定方法についての異なる検出チャンネルから得られる画像について、本明細書に例示されるものと同じ方法で処理することができる。

【0176】

別の例示的な種類の S B S では、サイクル配列決定は、例えば、その開示が参照により組み込まれる、国際公開第 0 4 / 0 1 8 4 9 7 号及び米国特許第 7 , 0 5 7 , 0 2 6 号に記載されているような切断可能な又は光漂白可能な色素標識を含む可逆的ターミネーターヌクレオチドを段階的に付加することによって達成される。この手法は、S o l e x a (n o w I l l u m i n a I n c .) によって商品化されており、国際公開第 9 1 / 0 6 6 7 8 号及び同第 0 7 / 1 2 3 , 7 4 4 号にも記載されており、これらのそれぞれは、参照により本明細書に組み込まれる。終端の両方を逆転させることができ、蛍光標識が開裂された蛍光標識ターミネーターの可用性は、効率的な循環可逆的終端 (cyclic reversible termination、C R T) 配列決定を容易にする。ポリメラーゼはまた、これらの修飾されたヌクレオチドを効率的に組み込み、かつそこから伸長するように共操作することもできる。

【0177】

好ましくは、可逆的ターミネーターベースの配列決定実施形態では、標識は、S B S 反応条件下での伸長を実質的に阻害しない。しかしながら、検出標識は、例えば、開裂又は分解によって除去可能であり得る。画像は、アレイ化された核酸特徴への標識の組み込み後に撮影することができる。特定の実施形態では、各サイクルは、アレイへの 4 つの異なるヌクレオチド型の同時送達を伴い、各ヌクレオチド型は、スペクトル的に異なる標識を

有する。次に、4つの異なる標識の1つに選択的な検出チャンネルをそれぞれ使用して、4つの画像を得ることができる。代替的に、異なるヌクレオチド型を順次追加することができる。各追加ステップの間にアレイの画像を得ることができる。このような実施形態では、各画像は、特定の型のヌクレオチドを組み込んだ核酸特徴を示す。各特徴の配列コンテンツが異なるため、様々な画像に様々な特徴が存在するか、存在しない。しかしながら、特徴の相対的な位置は、画像内で変わらないままである。そのような可逆的ターミネーターSBS方法から得られる画像は、本明細書に記載されるように保存、処理、及び分析することができる。画像撮影ステップに続いて、標識を除去することができ、その後のヌクレオチド添加及び検出のサイクルについて可逆的ターミネーター部分を除去することができる。特定のサイクルで検出された後、及び後続のサイクルの前に標識を除去すると、サイクル間のバックグラウンドシグナル及びクロストークを低減できるという利点がある。有用な標識及び除去方法の例を以下に記載する。

10

【0178】

特定の実施形態では、ヌクレオチドモノマーの一部又は全ては、可逆的ターミネーターを含むことができる。このような実施形態では、可逆的ターミネーター／開裂可能なフルオロフォア (fluor) は、3' エステル結合を介してリボース部分に結合したフルオロフォア (fluor) を含むことができる (Metzker, Genome Res. 15: 1767-1776 (2005)、これは参照により本明細書に組み込まれる)。他の手法は、ターミネーターの化学を蛍光標識の切断から分離している (参照によりその全体が本明細書に組み込まれる、Ruparel et al., Proc Natl Acad Sci USA 102: 5932-7 (2005))。Ruparelらは、少量の3' アリル基を使用して伸長をブロックするが、パラジウム触媒で短時間処理することにより容易に脱ブロックすることができる可逆的ターミネーターの開発について説明している。フルオロフォアは、長波長UV光への30秒の曝露によって容易に開裂することができる光開裂可能リンカーを介して基に付着された。したがって、ジスルフィド還元又は光開裂のいずれかを開裂可能なリンカーとして使用することができる。可逆的終端への別の手法は、dNTP上に嵩高な染料を配置した後に続く自然終端の使用である。dNTP上の帯電した嵩高な染料の存在は、立体障害及び／又は静電障害を介して効果的なターミネーターとして作用することができる。1つの組み込みイベントの存在は、染料が除去されない限り、それ以上の結合を防止する。染料の開裂は、フルオロフォア (fluor) を除去し、終端を効果的に逆転させる。修飾ヌクレオチドの例はまた、米国特許第7,427,673号及び米国特許第7,057,026号に記載されており、これらの開示は、参照によりそれらの全体が本明細書に組み込まれる。

20

30

【0179】

本明細書に記載の方法及びシステムとともに利用することができる追加の例示的なSBSシステム及び方法は、米国特許出願公開第2007/0166705号、米国特許出願公開第2006/0188901号、米国特許第7,057,026号、米国特許出願公開第2006/0240439号、米国特許出願公開第2006/0281109号、国際公開第05/065814号、米国特許出願公開第2005/0100900号、国際公開第06/064199号、国際公開第07/010,251号、米国特許出願公開第2012/0270305号、及び米国特許出願公開第2013/0260372号に記載されており、これらの開示は、参照によりその全体が本明細書に組み込まれる。

40

【0180】

いくつかの実施形態は、4つ未満の異なる標識を使用する4つの異なるヌクレオチドの検出を利用することができる。例えば、SBSは、組み込まれた資料である米国特許出願公開第2013/0079232号に記載される方法及びシステムを使用して実施することができる。第1の例として、ヌクレオチド型の対は、同じ波長で検出することができるが、対のうちの1つのメンバーに対する強度の差に基づいて、又は、対の他の部材について検出されたシグナルと比較して明らかなシグナルを出現又は消失させる、対の1つのメンバーへの変化 (例えば、化学修飾、光化学修飾、又は物理的改質を行うことを介して)

50

に基づいて区別され得る。第2の例として、4つの異なるヌクレオチド型のうちの3つを特定の条件下で検出することができ、一方、第4のヌクレオチド型は、それらの条件下で検出可能な標識がないか、又はそれらの条件下で最小限に検出される（例えば、バックグラウンド蛍光による最小限の検出など）。最初の3つのヌクレオチド型を核酸に組み込むことは、それらの対応するシグナルの存在に基づいて決定することができ、第4のヌクレオチド型を核酸に組み込むことは、任意のシグナルの不在又は最小限の検出に基づいて決定することができる。第3の例として、1つのヌクレオチド型は、2つの異なるチャンネルで検出される標識を含むことができ、一方、他のヌクレオチド型は、チャンネルのうちの1つ以下で検出される。前述の3つの例示的な構成は、相互に排他的であるとはみなされず、様々な組み合わせで使用することができる。3つ全ての例を組み合わせた例示的な実施形態は、第1のチャンネルで検出される第1のヌクレオチド型（例えば、第1の励起波長によって励起されたときに第1のチャンネルで検出される標識を有するdATP）、第2のチャンネルで検出される第2のヌクレオチド型（例えば、第2の励起波長によって励起されたときに第2のチャンネルで検出される標識を有するdCTP）、第1及び第2のチャンネルの両方において検出される第3のヌクレオチド型（例えば、第1及び／又は第2の励起波長によって励起されたときに両方のチャンネルで検出される少なくとも1つの標識を有するdTTP）、及びいずれのチャンネルでも検出されないか、又は最小限に検出される標識を欠く第4のヌクレオチド型（例えば、標識のないdGTP）を使用する蛍光ベースのSBS方法である。

10

20

【0181】

更に、組み込まれた資料である米国特許出願公開第2013／0079232号に記載のように、配列決定データは、単一のチャンネルを使用して得ることができる。このようないわゆる1つの染料配列決定方法では、第1のヌクレオチド型は標識されるが、第1の画像が生成された後に標識が除去され、第2のヌクレオチド型は、第1の画像が生成された後にのみ標識される。第3のヌクレオチド型は、第1及び第2の画像の両方においてその標識を保持し、第4のヌクレオチド型は、両方の画像において標識されていないままである。

30

40

【0182】

いくつかの実施形態は、ライゲーション技術による配列決定を利用することができる。このような技術は、DNAリガーゼを利用してオリゴヌクレオチドを組み込み、そのようなオリゴヌクレオチドの組み込みを特定する。オリゴヌクレオチドは、典型的には、オリゴヌクレオチドがハイブリダイズする配列中の特定のヌクレオチドの同一性と相関する異なる標識を有する。他のSBS方法と同様に、標識された配列決定試薬で核酸特徴のアレイを処理した後、画像を得ることができる。各画像は、特定の型の標識を組み込んだ核酸特徴を示す。各特徴の配列コンテンツが異なるため、様々な画像に様々な特徴が存在するか、存在しないが、特徴の相対的な位置は、画像内で変わらないままである。ライゲーションベースの配列決定方法から得られる画像は、本明細書に記載されるように保存、処理、及び分析することができる。本明細書に記載の方法及びシステムとともに利用することができる例示的なSBSシステム及び方法は、米国特許第6,969,488号、米国特許第6,172,218号、及び米国特許第6,306,597号に記載されており、これらの開示は、参照によりそれらの全体が本明細書に組み込まれる。

【0183】

いくつかの実施形態は、ナノ細孔配列決定を利用することができる（Deamer, D. W. & Akeson, M. 「Nanopores and nucleic acids: prospects for ultrarapid sequencing.」 Trends Biotechnol. 18, 147-151 (2000)、Deamer, D. and D. Branton, 「Characterization of nucleic acids by nanopore analysis」. Acc. Chem. Res. 35: 817-825 (2002)、Li, J., M. Gershow, D. Stein, E. Brandin, and J. A. Golovchenko, 「

50

DNA molecules and configurations in a solid-state nanopore microscope」Nat. Mater. 2: 611-615 (2003)、これらの開示は、参照によりそれらの全体が本明細書に組み込まれる)。そのような実施形態では、標的核酸はナノ細孔を通過する。ナノ細孔は、 α -ヘモリジンなどの合成孔又は生体膜タンパク質であり得る。標的核酸がナノ細孔を通過するとき、各塩基対は、細孔の電気コンダクタンスの変動を測定することによって特定することができる。(米国特許第7,001,792号、Soni, G. V. & Meller, 「A. Progress toward ultrafast DNA sequencing using solid-state nanopores.」Clin. Chem. 53, 1996-2001 (2007)、Healy, K. 「Nanopore-based single-molecule DNA analysis.」Nanomed. 2, 459-481 (2007)、Cockroft, S. L., Chu, J., Amorin, M. & Ghadiri, M. R. 「A single-molecule nanopore device detects DNA polymerase activity with single-nucleotide resolution.」J. Am. Chem. Soc. 130, 818-820 (2008)、これらの開示は、参照によりそれらの全体が本明細書に組み込まれる)。ナノ細孔配列決定から得られるデータは、本明細書に記載されるように、保存、処理、及び分析することができる。具体的には、データは、本明細書に記載される光学画像及び他の画像の例示的な処理に従って、画像として処理することができる。

10

20

【0184】

いくつかの実施形態は、DNAポリメラーゼ活性のリアルタイムモニタリングを伴う方法を利用することができる。ヌクレオチドの組み込みは、例えば、各々が参照により本明細書に組み込まれる、米国特許第7,329,492号及び米国特許第7,211,414号に記載されているようなフルオロフォア含有ポリメラーゼと γ -ホスフェート標識ヌクレオチドとの間の蛍光共鳴エネルギー移動(FRET)相互作用を介して検出することができ、又はヌクレオチドの組み込みは、例えば、参照により本明細書に組み込まれる米国特許第7,315,019号に記載されているようなゼロモード導波路、並びに、例えば、各々が参照により本明細書に組み込まれる、米国特許第7,405,281号及び米国特許出願公開第2008/0108082号に記載されているような蛍光ヌクレオチド類似体及び操作ポリメラーゼを使用して検出することができる。照明は、蛍光標識されたヌクレオチドの組み込みが低バックグラウンドで観察され得るように、表面繫留ポリメラーゼの周囲のzeptorリットルスケールの体積に制限することができる(Levene, M. J. et al. 「Zero-mode waveguides for single-molecule analysis at high concentrations.」Science, 299, 682-686 (2003)、Lundquist, P. M. et al. 「Parallel confocal detection of single molecules in real time.」Opt. Lett. 33, 1026-1028 (2008)、Korlach, J. et al. 「Selective aluminum passivation for targeted immobilization of single DNA polymerase molecules in zero-mode waveguide nanostructures.」Proc. Natl. Acad. Sci. USA 105, 1176-1181 (2008)、これらの開示は、参照によりそれらの全体が本明細書に組み込まれる)。このような方法から得られる画像は、本明細書に記載されるように、記憶、処理、及び分析することができる。

30

40

【0185】

いくつかのSBS実施形態は、伸長産物へのヌクレオチドの組み込み時に放出されるプロトンの検出を含む。例えば、放出されたプロトンの検出に基づく配列決定は、Ion Torrent (Guilford, CT、Life Technologiesの子会

50

社) から市販されている電気検出器及び関連技術、又は、米国特許出願公開第 2009/0026082 (A1) 号、同第 2009/0127589 (A1) 号、同第 2010/0137143 (A1) 号、若しくは同第 2010/0282617 (A1) 号 (これらの各々は、参照により本明細書に組み込まれる) に記載されている配列決定方法及びシステムを使用することができる。動力学的除外を使用して標的核酸を増幅するための本明細書に記載の方法は、プロトンを検出するために使用される基質に容易に適用することができる。より具体的には、本明細書に記載の方法を使用して、プロトンを検出するために使用されるアンプリコンのクローン集団を生成することができる。

【0186】

上記の SBS 方法は、複数の異なる標的核酸が同時に操作されるように、多重形式で有利に実施することができる。特定の実施形態では、異なる標的核酸は、共通の反応容器又は特定の基質の表面上で処理することができる。これにより、配列決定試薬の簡便な送達、未反応試薬の除去、及び取り込み事象の検出が多重方式で可能になる。表面結合された標的核酸を使用する実施形態では、標的核酸は、アレイ形式であり得る。アレイ形式では、標的核酸は、典型的には、空間的に区別可能な方式で表面に結合され得る。標的核酸は、直接共有結合、ビーズ若しくは他の粒子への結合、又は表面に結合したポリマーゼ若しくは他の分子への結合によって結合され得る。アレイは、各部位 (特徴とも称される) における標的核酸の単一コピーを含むことができ、又は同じ配列を有する複数のコピーは、各部位若しくは特徴に存在することができる。複数のコピーは、以下で更に詳細に記載されるブリッジ増幅又はエマルジョン PCR などの増幅方法によって生成することができる。

【0187】

本明細書に記載の方法は、例えば、少なくとも約 10 個の特徴 / cm^2 、100 個の特徴 / cm^2 、500 個の特徴 / cm^2 、1,000 個の特徴 / cm^2 、5,000 個の特徴 / cm^2 、10,000 個の特徴 / cm^2 、50,000 個の特徴 / cm^2 、100,000 個の特徴 / cm^2 、1,000,000 個の特徴 / cm^2 、又はそれ以上を含む、様々な密度のいずれかの特徴を有するアレイを使用することができる。

【0188】

本明細書に記載の方法の利点は、複数の標的核酸の迅速かつ効率的な検出を並行して提供することである。したがって、本開示は、上記で例示されるものなどの当該技術分野において既知の技術を使用して核酸を調製及び検出することができる統合システムを提供する。したがって、本開示の統合システムは、増幅試薬及び / 又は配列決定試薬を 1 つ以上の固定化された DNA 断片に送達することができる流体コンポーネントを含むことができ、システムは、ポンプ、弁、リザーバ、流体ラインなどのコンポーネントを含む。フローセルは、標的核酸を検出するための統合システムで構成及び / 又は使用することができる。例示的なフローセルは、例えば、米国特許第 2010/0111768 (A1) 号及び米国特許出願第 13/273,666 号に記載され、これらの各々は、参照により本明細書に組み込まれる。フローセルについて例示されるように、統合システムの流体コンポーネントの 1 つ以上を増幅方法及び検出方法に使用することができる。核酸配列決定の実施形態を一例として取ると、統合システムの流体コンポーネントの 1 つ以上を、本明細書に記載の増幅方法、及び上記に例示したような配列決定方法における配列決定試薬の送達に使用することができる。代替的に、統合システムは、増幅方法を実施し、検出方法を実施するための別々の流体システムを含み得る。増幅された核酸を作成し、また核酸の配列を決定することができる統合配列決定システムの例としては、MiSeq (商標) プラットフォーム (Illumina Inc., San Diego, CA)、及び参照により本明細書に組み込まれる、米国特許出願第 13/273,666 号に記載の装置が挙げられるが、これらに限定されない。

【0189】

上記の配列決定システムは、配列決定装置によって受け取られた試料中に存在する核酸

ポリマーを配列決定する。本明細書で定義されるように、「試料」及びその派生語は、最も広い意味で使用され、標的を含むことが疑われる任意の試料、培養物などを含む。いくつかの実施形態では、試料は、DNA、RNA、PNA、LNA、キメラ又はハイブリッド形態の核酸を含む。試料は、1以上の核酸を含有する任意の生物学的試料、臨床試料、外科試料、農業試料、大気試料又は水試料を含むことができる。この用語はまた、任意の単離された核酸試料、例えば、ゲノムDNA、新鮮凍結又はホルマリン固定パラフィン包埋核酸試料を含む。試料は、単一個体、遺伝的に関連するメンバーからの核酸試料のコレクション、遺伝的に関連しないメンバーからの核酸試料、腫瘍試料及び正常組織試料のような単一個体からの核酸試料（適合）、又は母体被験体から得られた母体及び胎児DNAのような遺伝物質の2つの異なる形態を含む単一供給源からの試料、又は植物又は動物DNAを含む試料中の混入細菌DNAの存在に由来し得ることも想定される。いくつかの実施形態では、核酸物質の供給源は、例えば新生児スクリーニングに典型的に使用されるような新生児から得られた核酸を含むことができる。

10

【0190】

核酸試料は、ゲノムDNA（genomic DNA、gDNA）などの高分子量物質を含むことができる。試料は、FFPE又は保管されたDNA試料から得られた核酸分子などの低分子量物質を含むことができる。別の実施形態では、低分子量物質は、酵素的又は機械的に断片化されたDNAを含む。試料は、無細胞循環DNAを含むことができる。いくつかの実施形態では、試料は、生検、腫瘍、擦過物、スワブ、血液、粘液、尿、血漿、精液、毛髪、レーザー捕捉顕微解剖、外科的切除、及び他の臨床的又は実験室で得られた試料から得られた核酸分子を含むことができる。いくつかの実施形態では、試料は、疫学、農業、法医学又は病原性の試料であり得る。いくつかの実施形態では、試料は、ヒト又は哺乳動物源などの動物から得られた核酸分子を含むことができる。別の実施形態では、試料は、植物、細菌、ウイルス又は真菌などの非哺乳類源から得られた核酸分子を含むことができる。いくつかの実施形態では、核酸分子の供給源は、保存された又は絶滅した試料若しくは種であり得る。

20

【0191】

更に、本明細書中に開示される方法及び組成物は、法医学試料からの分解及び／又は断片化されたゲノムDNAなどの低品質核酸分子を有する核酸試料を増幅するのに有用であり得る。一実施形態では、法医学試料は、犯罪現場から得られた核酸、行方不明者DNAデータベースから得られた核酸、法医学調査と関連した研究所から得られた核酸を含むことができ、又は法執行機関、1つ以上のミリタリーサービス若しくはそのような隊員によって得られた法医学試料を含むことができる。核酸試料は、例えば、口腔スワブ、紙、布、又は唾液、血液、若しくは他の体液で含浸され得る他の基質に由来する、精製された試料又は溶解物を含む粗DNAであり得る。したがって、いくつかの実施形態では、核酸試料は、ゲノムDNAなどの、少量のDNA又は断片化されたDNAの部分を含むことができる。いくつかの実施形態では、標的配列は、限定されるものではないが、血液、痰、血漿、精液、尿及び血清を含む1つ以上の体液に存在し得る。いくつかの実施形態では、標的配列は、犠牲者の毛髪、皮膚、組織試料、剖検又は遺体から得ることができる。いくつかの実施形態では、1つ以上の標的配列を含む核酸は、死亡した動物又はヒトから得ることができる。いくつかの実施形態では、標的配列は、微生物、植物又は昆虫学的DNAなど非ヒトDNAから得られた核酸を含むことができる。いくつかの実施形態では、標的配列又は増幅された標的配列は、ヒト同定を目的とする。いくつかの実施形態では、本開示は、概して、法医学試料の特徴を同定するための方法に関する。いくつかの実施形態では、本開示は、概して、本明細書に開示された1つ以上の標的特異的プライマー、又は本明細書に概説されたプライマー設計基準を用いて設計された1以上の標的特異的プライマーを使用するヒト同定方法に関する。一実施形態では、少なくとも1つの標的配列を含む法医学試料又はヒト同定試料は、本明細書に開示された標的特異的プライマーのいずれか1つ以上を用いて、又は本明細書に概説されたプライマー基準を用いて増幅することができる。

30

40

50

【0192】

クラスタ認識塩基コーリングシステム106の構成要素は、ソフトウェア、ハードウェア、又はその両方を含むことができる。例えば、クラスタ認識塩基コーリングシステム106の構成要素は、非一時的コンピュータ可読記憶媒体上に記憶され、1つ以上のコンピューティング装置（例えば、ユーザクライアント装置108）のプロセッサによって実行可能な1つ以上の命令を含むことができる。1つ以上のプロセッサによって実行されると、クラスタ認識塩基コーリングシステム106のコンピュータ実行可能命令は、コンピューティング装置に、本明細書で説明される障害ソース特定方法を実行させることができる。あるいは、クラスタ認識塩基コーリングシステム106の構成要素は、ある特定の機能又は機能群を実行するための専用処理装置などのハードウェアを含むことができる。更に、又は代替として、クラスタ認識塩基コーリングシステム106の構成要素は、コンピュータ実行可能命令及びハードウェアの組み合わせを含むことができる。

10

【0193】

更に、クラスタ認識塩基コーリングシステム106に関して本明細書で説明される機能を実行するクラスタ認識塩基コーリングシステム106の構成要素は、例えば、スタンドアロンアプリケーションの一部として、アプリケーションのモジュールとして、アプリケーションのプラグインとして、他のアプリケーションによってコーリングされ得るライブラリ関数（複数可）として、及び／又はクラウドコンピューティングモデルとして実装され得る。したがって、クラスタ認識塩基コーリングシステム106の構成要素は、パーソナルコンピューティング装置又はモバイル装置上のスタンドアロンアプリケーションの一部として実装され得る。更に、又は代替として、クラスタ認識塩基コーリングシステム106の構成要素は、限定するものではないが、Illumina BaseSpace、Illumina DRAGEN、又はIllumina TruSightソフトウェアを含む、配列決定サービスを提供する任意のアプリケーションにおいて実装されてもよい。「Illumina」、「BaseSpace」、「DRAGEN」、及び「TruSight」は、米国及び／又は他の国におけるIllumina, Inc.の登録商標又は商標である。

20

【0194】

本開示の実施形態は、以下でより詳細に論じられるように、例えば、1つ以上のプロセッサ及びシステムメモリ等のコンピュータハードウェアを含む、専用又は汎用コンピュータを含み、又は利用してもよい。本開示の範囲内の実施形態はまた、コンピュータ実行可能命令及び／又はデータ構造を搬送又は記憶するための物理的及び他のコンピュータ可読媒体を含む。特に、本明細書で説明されるプロセスのうちの1つ以上は、非一時的コンピュータ可読媒体において具現化され、1つ以上のコンピューティング装置（例えば、本明細書で説明されるメディアコンテンツアクセス装置のうちのいずれか）によって実行可能な命令として少なくとも部分的に実装されてもよい。概して、プロセッサ（例えば、マイクロプロセッサ）は、非一時的コンピュータ可読媒体（例えば、メモリ等）から命令を受信し、それらの命令を実行し、それによって、本明細書で説明されるプロセスのうちの1つ以上を含む、1つ以上のプロセスを実行する。

30

【0195】

コンピュータ可読媒体は、汎用コンピュータシステム又は専用コンピュータシステムによってアクセスされ得る任意の利用可能な媒体であり得る。コンピュータ実行可能命令を記憶するコンピュータ可読媒体は、非一時的コンピュータ可読記憶媒体（装置）である。コンピュータ実行可能命令を搬送するコンピュータ可読媒体は、伝送媒体である。したがって、限定ではなく例として、本開示の実施形態は、少なくとも2つの明確に異なる種類のコンピュータ可読媒体、すなわち非一時的コンピュータ可読記憶媒体（装置）及び伝送媒体を含むことができる。

40

【0196】

非一時的コンピュータ可読記憶媒体（装置）は、RAM、ROM、EEPROM、CD-ROM、（例えば、RAMに基づく）ソリッドステートドライブ（SSD）、フラッシュ

50

メモリ、相変化メモリ（PCM）、他のタイプのメモリ、他の光ディスクストレージ、磁気ディスクストレージ若しくは他の磁気ストレージ装置、又はコンピュータ実行可能命令若しくはデータ構造の形態で所望のプログラムコード手段を記憶するために使用することができ、汎用若しくは専用コンピュータによってアクセスすることができる任意の他の媒体を含む。

【0197】

「ネットワーク」は、コンピュータシステム及び／又はモジュール及び／又は他の電子装置間の電子データの移送を可能にする1つ以上のデータリンクとして定義される。情報が、ネットワーク又は別の通信接続（ハードワイヤード、ワイヤレス、又はハードワイヤード若しくはワイヤレスの組み合わせのいずれか）を介してコンピュータに転送又は提供されるとき、コンピュータは、その接続を伝送媒体として適切に認識する。伝送媒体は、コンピュータ実行可能命令又はデータ構造の形態で所望のプログラムコード手段を搬送するために使用することができ、汎用又は専用コンピュータによってアクセスすることができるネットワーク及び／又はデータリンクを含むことができる。上記の組み合わせも、コンピュータ可読媒体の範囲内に含まれるべきである。

10

【0198】

更に、様々なコンピュータシステム構成要素に到達すると、コンピュータ実行可能命令又はデータ構造の形態のプログラムコード手段は、伝送媒体から非一時的コンピュータ可読記憶媒体（装置）に（又はその逆に）自動的に転送され得る。例えば、ネットワーク又はデータリンクを介して受信されたコンピュータ実行可能命令又はデータ構造は、ネットワークインターフェースモジュール（例えば、NIC）内のRAMにバッファリングされ、次いで、最終的に、コンピュータシステムRAM及び／又はコンピュータシステムにおけるより揮発性の低いコンピュータ記憶媒体（装置）に転送され得る。したがって、非一時的コンピュータ可読記憶媒体（装置）は、伝送媒体も（又は更に主に）利用するコンピュータシステム構成要素に含まれ得ることを理解されたい。

20

【0199】

コンピュータ実行可能命令は、例えば、プロセッサで実行されると、汎用コンピュータ、専用コンピュータ、又は専用処理装置に、ある機能又は機能群を実行させる命令及びデータを含む。いくつかの実施形態では、コンピュータ実行可能命令は、汎用コンピュータ上で実行され、汎用コンピュータを、本開示の要素を実装する専用コンピュータに変える。コンピュータ実行可能命令は、例えば、バイナリ、アセンブリ言語などの中間フォーマット命令、又は更にソースコードであってもよい。主題は、構造的特徴及び／又は方法論的動作に特有の言語で説明されているが、添付の特許請求の範囲において定義される主題は、説明された特徴又は上述の動作に必ずしも限定されないことを理解されたい。むしろ、説明された特徴及び動作は、特許請求の範囲を実装する例示的な形態として開示される。

30

【0200】

当業者は、本開示が、パーソナルコンピュータ、デスクトップコンピュータ、ラップトップコンピュータ、メッセージプロセッサ、ハンドヘルド装置、マルチプロセッサシステム、マイクロプロセッサベース又はプログラム可能な家庭用電化製品、ネットワークPC、ミニコンピュータ、メインフレームコンピュータ、携帯電話、PDA、タブレット、ページャ、ルータ、スイッチなどを含む、多くのタイプのコンピュータシステム構成を有するネットワークコンピューティング環境で実施され得ることを理解するであろう。本開示はまた、ネットワークを介して（ハードワイヤードデータリンク、ワイヤレスデータリンク、又はハードワイヤード及びワイヤレスデータリンクの組み合わせのいずれかによって）リンクされたローカル及びリモートコンピュータシステムが両方ともタスクを実行する分散システム環境において実施され得る。分散システム環境では、プログラムモジュールは、ローカルメモリストレージ装置及びリモートメモリストレージ装置の両方に位置することができる。

40

【0201】

50

本開示の実施形態は、クラウドコンピューティング環境において実装することもできる。本明細書では、「クラウドコンピューティング」は、構成可能なコンピューティングリソースの共有プールへのオンデマンドネットワークアクセスを可能にするためのモデルとして定義される。例えば、クラウドコンピューティングは、構成可能なコンピューティングリソースの共有プールへのユビキタスで便利なオンデマンドアクセスを提供するために、市場で使用され得る。構成可能なコンピューティングリソースの共有プールは、仮想化を介して迅速に設定され、低い管理労力又はサービスプロバイダ対話で公開され、次いで、それに応じて拡大縮小され得る。

【0202】

クラウドコンピューティングモデルは、例えば、オンデマンドセルフサービス、広域ネットワークアクセス、リソースプーリング、迅速な弾力性、測定されたサービス等の種々の特性から構成することができる。クラウドコンピューティングモデルはまた、例えば、Software as a Service (SaaS)、Platform as a Service (PaaS)、及びInfrastructure as a Service (IaaS)などの様々なサービスモデルを公開することができる。クラウドコンピューティングモデルは、プライベートクラウド、コミュニティクラウド、パブリッククラウド、ハイブリッドクラウドなどの異なる展開モデルを使用して展開することもできる。本明細書及び特許請求の範囲において、「クラウドコンピューティング環境」は、クラウドコンピューティングが採用される環境である。

【0203】

図10は、上記で説明したプロセスのうちの1つ以上を実行するように構成され得るコンピューティング装置1000のブロック図を示す。コンピューティング装置1000などの1つ以上のコンピューティング装置が、クラスタ認識塩基コーリングシステム106及び配列決定システム104を実装することができることが理解されよう。図10によって示されるように、コンピューティング装置1000は、プロセッサ1002、メモリ1004、ストレージ（記憶）装置1006、I/Oインターフェース1008、及び通信インターフェース1010を含むことができ、これらは、通信インフラストラクチャ1012によって通信可能に結合され得る。ある特定の実施形態では、コンピューティング装置1000は、図10に示されるものよりも少ない又は多い構成要素を含むことができる。以下の段落は、図10に示されるコンピューティング装置1000の構成要素を更に詳細に説明する。

【0204】

1つ以上の実施形態では、プロセッサ1002は、コンピュータプログラムを構成する命令などの命令を実行するためのハードウェアを含む。限定ではなく、例として、ワークフローを動的に修正するための命令を実行するために、プロセッサ1002は、内部レジスタ、内部キャッシュ、メモリ1004、又はストレージ装置1006から命令を取り出し（又はフェッチし）、それらを復号して実行することができる。メモリ1004は、データ、メタデータ、及びプロセッサによる実行のためのプログラムを記憶するために使用される揮発性又は不揮発性メモリであってもよい。ストレージ装置1006は、本明細書に記載の方法を実行するためのデータ又は命令を記憶するための、ハードディスク、フラッシュディスクドライブ、又は他のデジタルストレージ装置などのストレージを含む。

【0205】

I/Oインターフェース1008は、ユーザが、コンピューティング装置1000に入力を提供し、コンピューティング装置1000から出力を受信し、そうでなければコンピューティング装置1000にデータを転送し、コンピューティング装置1000からデータを受信することを可能にする。I/Oインターフェース1008は、マウス、キーパッド若しくはキーボード、タッチスクリーン、カメラ、光学スキャナ、ネットワークインターフェース、モデム、他の既知のI/O装置、又はかかるI/Oインターフェースの組み合わせを含むことができる。I/Oインターフェース1008は、限定されないが、グラフィックスエンジン、ディスプレイ（例えば、ディスプレイスクリーン）、1つ以上の出

カドライバ（例えば、ディスプレイドライバ）、1つ以上のオーディオスピーカ、及び1つ以上のオーディオドライバを含む、ユーザに出力を提示するための1つ以上の装置を含むことができる。ある特定の実施形態では、I/Oインターフェース1008は、ユーザに提示するためにグラフィカルデータをディスプレイに提供するように構成される。グラフィカルデータは、1つ以上のグラフィカルユーザインターフェース及び／又は特定の実装に役立ち得る任意の他のグラフィカルコンテンツを表してもよい。

【0206】

通信インターフェース1010は、ハードウェア、ソフトウェア、又はその両方を含むことができる。いずれにしても、通信インターフェース1010は、コンピューティング装置1000と1つ以上の他のコンピューティング装置又はネットワークとの間の通信（例えば、パケットベースの通信など）のための1つ以上のインターフェースを提供することができる。限定ではなく例として、通信インターフェース1010は、Ethernet（イーサネット）若しくは他の有線ベースのネットワークと通信するためのネットワークインターフェースコントローラ（network interface controller、NIC）若しくはネットワークアダプタ、又はWiFiなどのワイヤレスネットワークと通信するためのワイヤレスNIC（WNIC）若しくはワイヤレスアダプタを含むことができる。

10

【0207】

更に、通信インターフェース1010は、様々なタイプの有線又は無線ネットワークとの通信を容易にすることができる。通信インターフェース1010はまた、様々な通信プロトコルを使用して、通信を容易にすることもできる。通信インフラストラクチャ1012はまた、コンピューティング装置1000の構成要素を互いに結合するハードウェア、ソフトウェア、又はその両方を含むことができる。例えば、通信インターフェース1010は、1つ以上のネットワーク及び／又はプロトコルを使用して、特定のインフラストラクチャによって接続された複数のコンピューティング装置が互いに通信して、本明細書で説明されるプロセスの1つ以上の態様を実行することを可能にすることができる。例示すると、配列決定プロセスは、複数の装置（例えば、クライアント装置、配列決定装置、及びサーバー装置）が配列決定データ及びエラー通知などの情報を交換することを可能にすることができる。

20

【0208】

前述の明細書において、本開示は、その特定の例示的な実施形態を参照して説明された。本開示の様々な実施形態及び態様は、本明細書で論じられる詳細を参照して説明され、添付の図面は様々な実施形態を図示する。上記の説明及び図面は、本開示の例示であり、本開示を限定するものとして解釈されるべきではない。本開示の様々な実施形態の完全な理解を提供するために、多数の特定の詳細が説明される。

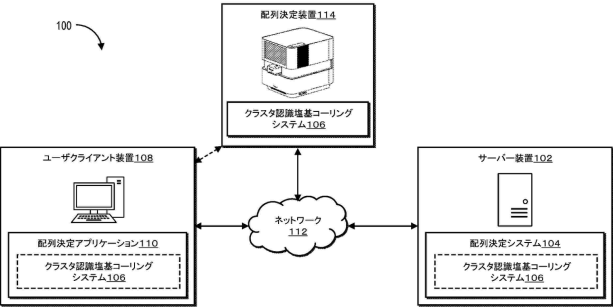
30

【0209】

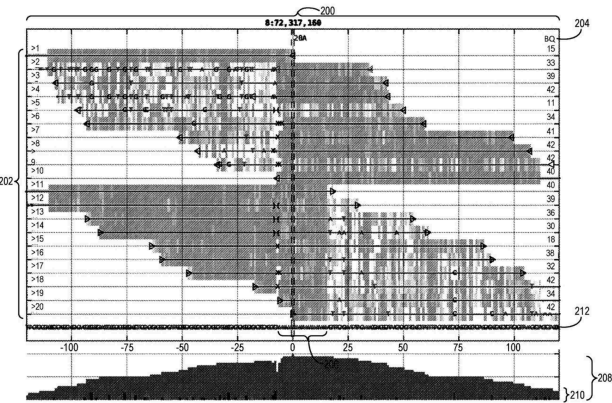
本開示は、その趣旨又は本質的な特徴から逸脱することなく、その他の特定の形態で具現化されてもよい。記載された実施形態は、全ての点において、例示的なものに過ぎず、限定的ではないとみなされるべきである。例えば、本明細書で説明される方法は、より少ない又はより多いステップ／動作を用いて行われてもよく、又はステップ／動作は、異なる順序で行われてもよい。更に、本明細書で説明されるステップ／動作は、互いに並行して、又は同じ若しくは同様の動作の異なる出現と並行して、繰り返されるか、又は実行され得る。したがって、本願の範囲は、前述の説明ではなく、添付の特許請求の範囲によって示される。特許請求の範囲の意味及び均等範囲内に含まれる全ての変更は、それらの範囲内に包含されるものである。

40

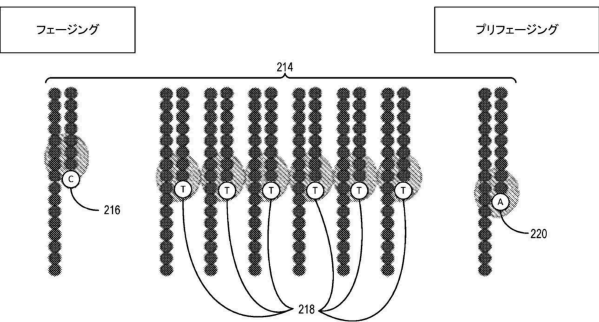
【図面】
【図 1】



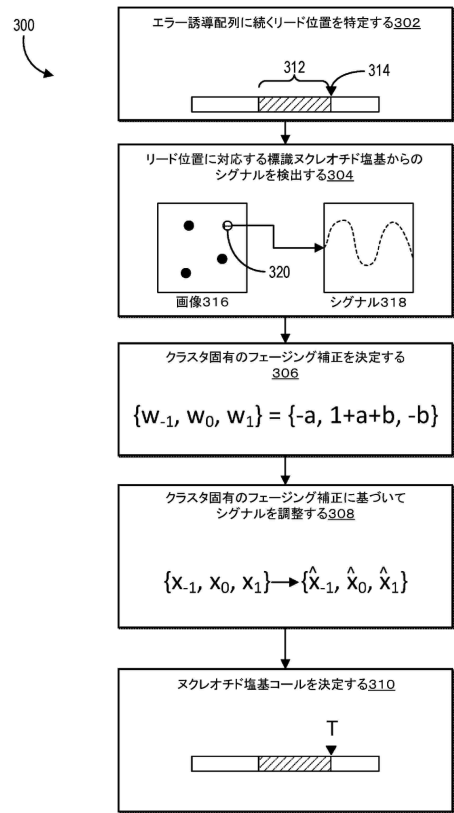
【図 2 A】



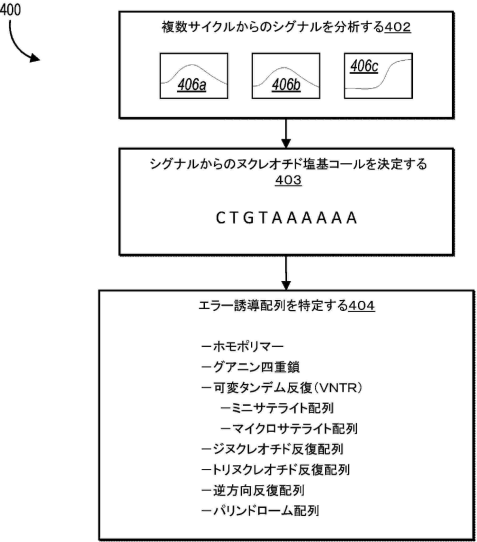
【図 2 B】



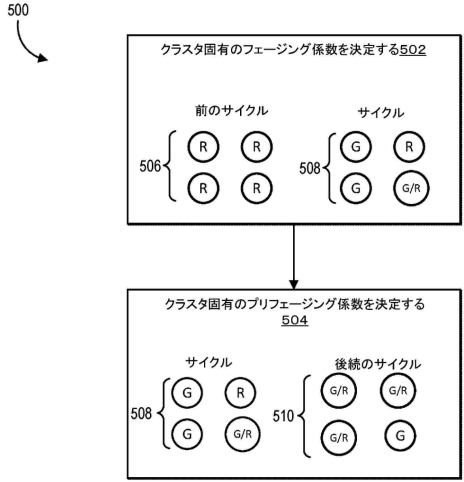
【図 3】



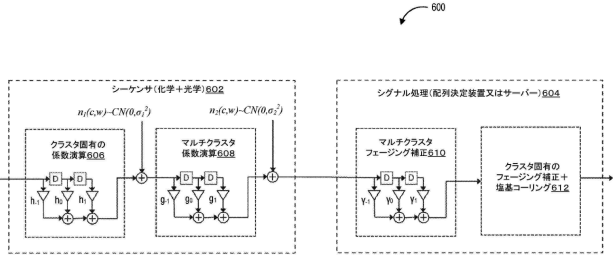
【図 4】



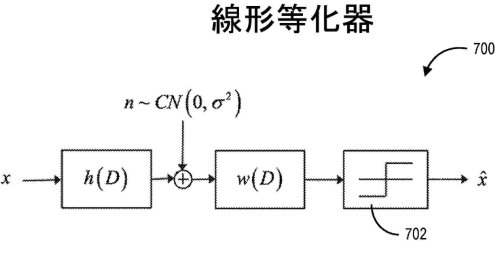
【図 5】



【図 6】

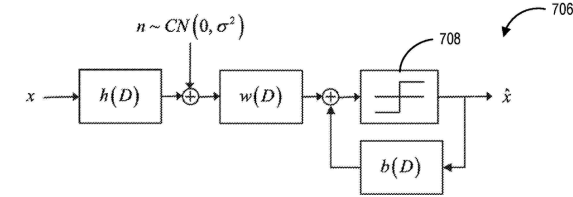


【図 7 A】



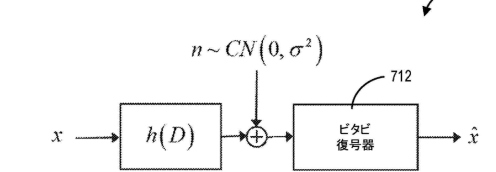
【図 7 B】

判定帰還型等化器(DFE)



【図 7 C】

(MLSE)



10

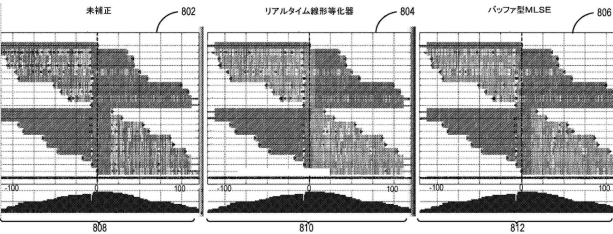
20

30

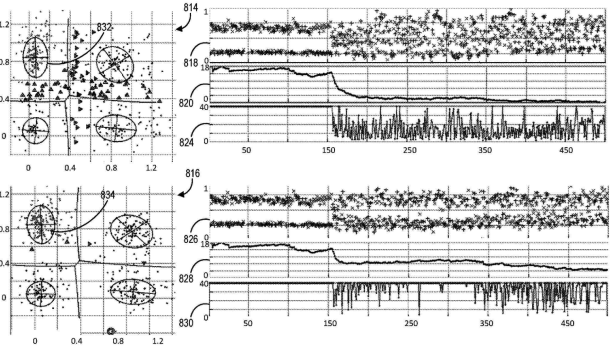
40

50

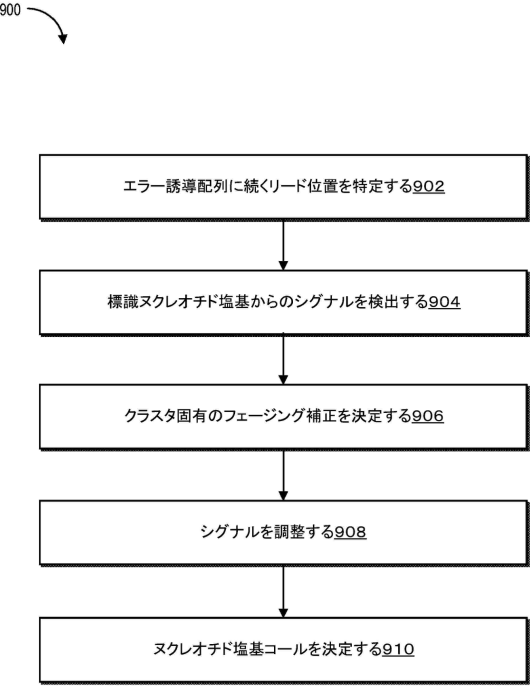
【図 8 A】



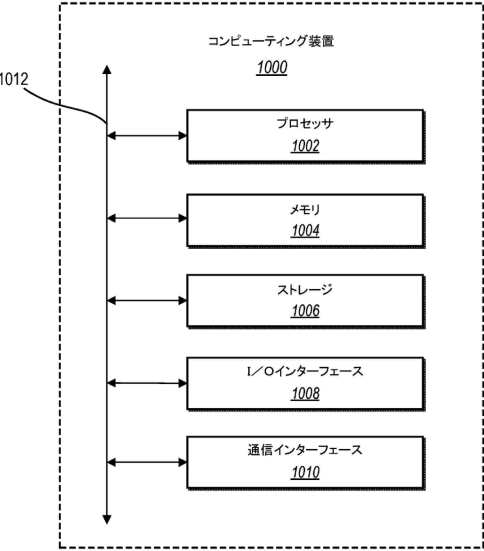
【図 8 B】



【図 9】



【図 10】



10

20

30

40

50

【国際調査報告】

INTERNATIONAL SEARCH REPORT

International application No
PCT/US2022/080512

A. CLASSIFICATION OF SUBJECT MATTER INV. G16B30/00 G16B40/10 ADD. 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) G16B Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched Electronic data base consulted during the international search (name of data base and, where practicable, search terms used) EPO-Internal		
C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	W.-C. KAO ET AL: "BayesCall: A model-based base-calling algorithm for high-throughput short-read sequencing", GENOME RESEARCH, vol. 19, no. 10, 6 August 2009 (2009-08-06), pages 1884-1895, XP055543900, US ISSN: 1088-9051, DOI: 10.1101/gr.095299.109 Abstract; page 1886, left, last paragraph; page 1888, right, first full paragraph -----	1-22
A	EP 3 715 467 A1 (ILLUMINA INC [US]) 30 September 2020 (2020-09-30) paragraphs [0038], [0083], [0091] ----- -/-	1-22
☒ Further documents are listed in the continuation of Box C. ☒ 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 17 February 2023		Date of mailing of the international search report 09/03/2023
Name and mailing address of the ISA/ European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel: (+31-70) 340-2040, Fax: (+31-70) 340-3016		Authorized officer Schmidt, Karsten

International application No
PCT/US2022/080512

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
E	US 2023/018469 A1 (KAGALWALLA ABDE ALI HUNAID [US] ET AL) 19 January 2023 (2023-01-19) paragraphs [0126], [0151] -----	1-22

50

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No

PCT/US2022/080512

Patent document cited in search report		Publication date	Patent family member(s)	Publication date
EP 3715467	A1	30-09-2020	AU 2014360530 A1	28-04-2016
			AU 2020277261 A1	07-01-2021
			CA 2928209 A1	11-06-2015
			CA 3181696 A1	11-06-2015
			DK 3077943 T3	07-09-2020
			EP 3077943 A2	12-10-2016
			EP 3715467 A1	30-09-2020
			EP 3940082 A1	19-01-2022
			ES 2808824 T3	02-03-2021
			HR P20201397 T1	27-11-2020
			HU E050641 T2	28-12-2020
			LT 3077943 T	12-10-2020
			PL 3077943 T3	30-11-2020
			PT 3077943 T	21-08-2020
			SI 3077943 T1	30-10-2020
			US 2018274023 A1	27-09-2018
			US 2020377938 A1	03-12-2020
			US 2021310065 A1	07-10-2021
			WO 2015084985 A2	11-06-2015

US 2023018469	A1	19-01-2023	NONE	

フロントページの続き

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,N
E,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,
CV,CV,CZ,DE,DJ,DK,DM,DO,DZ,EC,EE,EG,ES,FI,GB,GD,GE,GH,GM,GT,HN,HR,HU,ID,IL,IN,IQ,IR,IS,I
T,JM,JO,JP,KE,KG,KH,KN,KP,KR,KW,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,ST,SV,
SY,TH,TJ,TM,TN,TR,TT,TZ,UA,UG,US,UZ,VC,VN,WS,ZA,ZM,ZW

- (72)発明者 ジョン・エス・ヴィエチェリ
アメリカ合衆国・カリフォルニア・ 9 2 1 2 2 ・サン・ディエゴ・イルミナ・ウェイ・ 5 2 0 0
- (72)発明者 ギャヴィン・デレク・パーナビー
アメリカ合衆国・カリフォルニア・ 9 2 1 2 2 ・サン・ディエゴ・イルミナ・ウェイ・ 5 2 0 0
- (72)発明者 ボ・ル
アメリカ合衆国・カリフォルニア・ 9 2 1 2 2 ・サン・ディエゴ・イルミナ・ウェイ・ 5 2 0 0
- (72)発明者 ラミ・メヒオ
アメリカ合衆国・カリフォルニア・ 9 2 1 2 2 ・サン・ディエゴ・イルミナ・ウェイ・ 5 2 0 0