

(19)日本国特許庁(JP)

(12)公表特許公報(A)

(11)公表番号

特表2023-547298

(P2023-547298A)

(43)公表日 令和5年11月10日(2023.11.10)

(51)国際特許分類

G 1 6 B 40/10 (2019.01)

F I

G 1 6 B 40/10

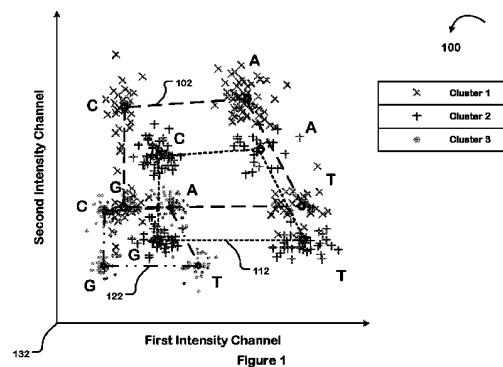
審査請求 未請求 予備審査請求 有 (全57頁)

(21)出願番号	特願2022-580968(P2022-580968)	(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3年10月26日(2021.10.26)	(74)代理人	100108453 弁理士 村山 靖彦
(85)翻訳文提出日	令和4年12月27日(2022.12.27)	(74)代理人	100110364 弁理士 実広 信哉
(86)国際出願番号	PCT/US2021/056711	(74)代理人	100133400 弁理士 阿部 達彦
(87)国際公開番号	WO2022/093865	(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4年5月5日(2022.5.5)		
(31)優先権主張番号	63/106,256		
(32)優先日	令和2年10月27日(2020.10.27)		
(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】 クラスターごとの強度補正及びベースコールのためのシステム及び方法

(57)【要約】

開示された技術は、改善されたベースコールのために、クラスター別に変動補正係数を生成して、クラスター間強度プロファイル変動を補正する。増幅係数は、スケール変動を補正する。チャンネル固有のオフセット係数は、それぞれの強度チャンネルに沿ったシフト変動を補正する。ターゲットクラスターについての変動補正係数は、配列決定ランの先行する配信決定サイクルでターゲットクラスターについて生成された履歴強度データの分析と、配列決定ランの現在の配列決定サイクルでターゲットクラスターについて生成された現在の強度データの分析との組み合わせに基づいて生成される。次いで、変動補正係数を使用して、配列決定ランの次の配列決定サイクルでターゲットクラスターについて生成された次の強度データが補正される。次いで、補正された次の強度データを使用して、次の配列決定サイクルでターゲットクラスターがベースコールされる。



10

20

【特許請求の範囲】

【請求項 1】

ターゲットクラスターをベースコールするコンピュータ実装方法であって、前記方法は

、
前記ターゲットクラスターについて、

配列決定ランの現在の配列決定サイクルに登録された現在のチャンネル固有の強度を、前記現在の配列決定サイクルで前記ターゲットクラスターがベースコールされる塩基固有の強度分布から読み取ることと、

前記塩基固有の強度分布の重心から現在のチャンネル固有の分布強度を読み取ることと、

前記現在のチャンネル固有の強度及び前記現在のチャンネル固有の分布強度に基づいて、前記現在の配列決定サイクルに対する現在の強度補正パラメータのセットを決定することと

10

、
前記配列決定ランの先行する配列決定サイクルについての前記現在の強度補正パラメータのセット及び先行する累積強度補正パラメータのセットを累積することによって、前記現在の配列決定サイクルについての現在の累積強度補正パラメータのセットを決定することと、

前記現在の累積強度補正パラメータのセットに基づいて、前記現在の配列決定サイクルに対する現在の増幅係数及び現在のチャンネル固有のオフセット係数を決定することと、

前記現在の増幅係数及び前記現在のチャンネル固有のオフセット係数を使用して、前記配列決定ランの次の配列決定サイクルに登録された次のチャンネル固有の強度を補正し、前記次の配列決定サイクルに対して補正された次のチャンネル固有の強度を生成することと、

20

前記補正された次のチャンネル固有の強度に基づいて、前記次の配列決定サイクルで前記ターゲットクラスターをベースコールすることと、を含む、コンピュータ実装方法。

【請求項 2】

前記現在の強度補正パラメータのセットは、前記現在のチャンネル固有の分布強度と、現在のチャンネル固有の強度誤差と、現在の分布重心から原点の距離と、現在の分布強度対強度誤差の類似性測定と、を含む、請求項 1 に記載のコンピュータ実装方法。

【請求項 3】

前記現在のチャンネル固有の強度誤差は、前記現在のチャンネル固有の強度と前記現在のチャンネル固有の分布強度との間のチャンネル単位の差である、請求項 1 ～ 2 のいずれか一項に記載のコンピュータ実装方法。

30

【請求項 4】

前記現在の分布重心から原点の距離は、前記重心と、前記塩基固有の強度分布を含む多次元空間の原点との間のユークリッド距離である、請求項 1 ～ 3 のいずれか一項に記載のコンピュータ実装方法。

【請求項 5】

前記多次元空間は、デカルト空間、極空間、円筒空間、及び球状空間のうちの少なくとも 1 つである、請求項 1 ～ 4 のいずれか一項に記載のコンピュータ実装方法。

【請求項 6】

前記現在の分布強度対強度誤差の類似性測定は、前記現在のチャンネル固有の分布強度と前記現在のチャンネル固有の強度誤差との間のチャンネル単位のドット積の総和である、請求項 1 ～ 5 のいずれか一項に記載のコンピュータ実装方法。

40

【請求項 7】

前記現在の累積強度補正パラメータのセットは、前記現在の強度補正パラメータのセット内の現在の強度補正パラメータと、前記先行する累積強度補正パラメータのセット内の先行する累積強度補正パラメータとの強度補正パラメータ単位の和である、請求項 1 ～ 6 のいずれか一項に記載のコンピュータ実装方法。

【請求項 8】

前記現在の累積強度補正パラメータのセットは、前記現在の強度補正パラメータ及び前記先行する累積強度補正パラメータの強度補正パラメータ単位の平均である、請求項 1 ～

50

7 のいずれか一項に記載のコンピュータ実装方法。

【請求項 9】

前記先行する累積強度補正パラメータのセット及び前記現在の累積強度補正パラメータのセットは、量子化された固定ビット幅フォーマットで記憶される、請求項 1 ～ 8 のいずれか一項に記載のコンピュータ実装方法。

【請求項 10】

前記現在のチャンネル固有のオフセット係数は、同一であるように構成されている、請求項 1 ～ 9 のいずれか一項に記載のコンピュータ実装方法。

【請求項 11】

前記現在の累積強度補正パラメータのセット内の現在の累積強度補正パラメータは、前記現在のチャンネル固有の分布強度についての第 1 の共通の現在の累積強度補正パラメータと、前記現在のチャンネル固有の強度誤差についての第 2 の共通の現在の累積強度補正パラメータと、を含む、請求項 1 ～ 10 のいずれか一項に記載のコンピュータ実装方法。

10

【請求項 12】

前記現在のチャンネル固有のオフセット係数は、次のチャンネル固有のシフト強度を生成するために前記次のチャンネル固有の強度からチャンネル単位で減算され、前記次のチャンネル固有のシフト強度は、前記補正された次のチャンネル固有の強度を生成するために前記現在の増幅係数で除算される、請求項 1 ～ 11 のいずれか一項に記載のコンピュータ実装方法。

【請求項 13】

重み付け関数を使用して、初期増幅係数を前記現在の増幅係数と組み合わせ、初期のチャンネル固有のオフセット係数を前記現在のチャンネル固有のオフセット係数と組み合わせ、前記現在の配列決定サイクルに対する重み付けされた現在の増幅係数及び重み付けされた現在のチャンネル固有のオフセット係数を生成することを更に含む、請求項 1 ～ 12 のいずれか一項に記載のコンピュータ実装方法。

20

【請求項 14】

前記重み付け関数は、最小重み (w_{min}) を初期増幅係数及び前記初期のチャンネル固有のオフセット係数に適用し、最大重み (w_{max}) を前記現在の増幅係数及び前記現在のチャンネル固有のオフセット係数に適用し、 $w_{min} = (1 - w_{max})$ である、請求項 1 ～ 13 のいずれか一項に記載のコンピュータ実装方法。

【請求項 15】

前記最大重み (w_{max}) は、 $(c - p) / c$ として定義され、 c は前記現在の配列決定サイクルのインデックスであり、 p は 2 ～ 7 の数値である、請求項 1 ～ 14 のいずれか一項に記載のコンピュータ実装方法。

30

【請求項 16】

前記重み付けされた現在の増幅係数及び前記重み付けされた現在のチャンネル固有のオフセット係数を使用して、前記次のチャンネル固有の強度を補正し、前記補正された次のチャンネル固有の強度を生成することを更に含む、請求項 1 ～ 15 のいずれか一項に記載のコンピュータ実装方法。

【請求項 17】

最尤法を使用して、前記現在の配列決定サイクルについて、前記現在の増幅係数及び前記現在のチャンネル固有のオフセット係数のそれぞれの現在の最尤重みを生成することと、

40

前記現在の最尤重みを前記現在の増幅係数及び前記現在のチャンネル固有のオフセット係数にそれぞれ適用して、前記現在の配列決定サイクルについて最尤重みが付けられた現在の増幅係数及び最尤重みが付けられた現在のチャンネル固有のオフセット係数を生成することと、

前記最尤重みが付けられた現在の増幅係数及び前記最尤重みが付けられた現在のチャンネル固有のオフセット係数を使用して、前記次のチャンネル固有の強度を補正し、前記補正された次のチャンネル固有の強度を生成することと、を更に含む、請求項 1 ～ 16 のいずれか一項に記載のコンピュータ実装方法。

【請求項 18】

50

減衰因子を前記現在の強度補正パラメータに適用して、前記現在の配列決定サイクルについて減衰された現在の強度補正パラメータを生成することと、

前記減衰された現在の強度補正パラメータと前記先行する累積強度補正パラメータとを強度補正パラメータ単位で累積することによって、前記現在の累積強度補正パラメータを決定することと、を更に含む、請求項 1～17 のいずれか一項に記載のコンピュータ実装方法。

【請求項 19】

前記減衰因子は、前記配列決定ランの特定の数の配列決定サイクルに対して固定されており、その後、減衰ロジックに基づいて指数関数的に減衰される、請求項 1～18 のいずれか一項に記載のコンピュータ実装方法。

10

【請求項 20】

前記減衰ロジックは $1 - 1 / t a u$ であり、 $t a u$ は事前定義された数値である、請求項 1～19 のいずれか一項に記載のコンピュータ実装方法。

【請求項 21】

前記配列決定ランの連続する配列決定サイクルで前記ターゲットクラスターについて前記読み取ること、前記読み取ること、前記決定すること、前記決定すること、前記決定すること、前記使用すること、及び前記ベースコールすることを繰り返すことを更に含む、請求項 1～20 のいずれか一項に記載のコンピュータ実装方法。

【請求項 22】

複数のクラスターについて前記読み取ること、前記読み取ること、前記決定すること、前記決定すること、前記決定すること、前記使用すること、及び前記ベースコールすることを並行して実行することを更に含む、請求項 1～21 のいずれか一項に記載のコンピュータ実装方法。

20

【請求項 23】

前記現在の強度補正パラメータのセット、前記現在の累積強度補正パラメータのセット、前記現在の増幅係数、及び前記現在のチャンネル固有のオフセット係数の閉形式の式は、最小二乗法を使用して決定される、請求項 1～22 のいずれか一項に記載のコンピュータ実装方法。

【請求項 24】

前記現在のチャンネル固有の強度は、強度チャンネルにそれぞれ対応する、請求項 1～23 のいずれか一項に記載のコンピュータ実装方法。

30

【請求項 25】

前記現在のチャンネル固有のオフセット係数は、前記現在のチャンネル固有の強度と前記現在のチャンネル固有の分布強度との間の前記チャンネル単位の差である、請求項 1～24 のいずれか一項に記載のコンピュータ実装方法。

【請求項 26】

ターゲットクラスターをベースコールするコンピュータ実装方法であって、前記方法は、

前記ターゲットクラスターについて、

現在の強度データ及び履歴強度データにアクセスすることであって、

前記現在の強度データは、配列決定ランの現在の配列決定サイクルに対するものであり

40

、前記履歴強度データは、前記配列決定ランの 1 つ以上の先行する配列決定サイクルに対するものである、ことと、

前記現在の強度データ及び前記履歴強度データに基づいて、スケール補正係数及びチャンネル固有のシフト補正係数を決定することと、

前記スケール補正係数及び前記チャンネル固有のシフト補正係数を使用して、次の強度データを補正し、補正された次の強度データを生成することであって、

前記次の強度データは、前記配列決定ランの次の配列決定サイクルに対するものである、ことと、

50

前記補正された次の強度データに基づいて、前記次の配列決定サイクルで前記ターゲットクラスターをベースコールすることと、を含む、コンピュータ実装方法。

【請求項 27】

ターゲットクラスターをベースコールするコンピュータ実装方法であって、前記方法は、

前記ターゲットクラスターについて、

現在の強度データ及び履歴強度データにアクセスすることであって、

前記現在の強度データは、配列決定ランの現在の配列決定サイクルに対するものであり

、
前記履歴強度データは、前記配列決定ランの1つ以上の先行する配列決定サイクルに対するものである、ことと、

前記現在の強度データ及び前記履歴強度データを使用して、次の強度データを補正し、補正された次の強度データを生成することであって、

前記次の強度データは、前記配列決定ランの次の配列決定サイクルに対するものである、ことと、

前記補正された次の強度データに基づいて、前記次の配列決定サイクルで前記ターゲットクラスターをベースコールすることと、を含む、コンピュータ実装方法。

【請求項 28】

メモリに結合された1つ以上のプロセッサを含むシステムであって、前記メモリは、ターゲットクラスターをベースコールするためのコンピュータ命令がロードされ、前記命令は、前記プロセッサ上で実行されると、請求項1に記載の命令を実施する、システム。

【請求項 29】

メモリに結合された1つ以上のプロセッサを含むシステムであって、前記メモリは、ターゲットクラスターをベースコールするためのコンピュータ命令がロードされ、前記命令は、前記プロセッサ上で実行されると、請求項26に記載の命令を実施する、システム。

【請求項 30】

メモリに結合された1つ以上のプロセッサを含むシステムであって、前記メモリは、ターゲットクラスターをベースコールするためのコンピュータ命令がロードされ、前記命令は、前記プロセッサ上で実行されると、請求項27に記載の命令を実施する、システム。

【発明の詳細な説明】

【技術分野】

【0001】

優先権出願

本出願は、2020年10月27日に出願された「Systems and Methods for Per-Cluster Intensity Correction and Base Calling」と題する米国特許仮出願第63/106,256号の優先権を主張する。

【0002】

開示された技術は、画像の自動分析又はパターンの認識のための装置及び対応する方法に関する。本明細書には、(a)認識前にその視覚的品質を向上させることと、(b)センサ又は記憶されたプロトタイプに対して画像を位置決めして位置合わせすること、又は無関係なデータを廃棄することによって画像データの量を減少させることと、(c)画像の有意な特性を測定することと、を目的として画像を変換するシステムが含まれる。特に、開示された技術は、画像データにおけるクラスター間強度プロファイルの変動を補正するための変動補正係数を生成することに関する。

【0003】

組み込み

2020年5月5日に出願された「EQUALIZATION-BASED IMAGE PROCESSING AND SPATIAL CROSSTALK ATTENUATOR」と題する米国特許仮出願第63/020,449号(代理人整理番号ILL

10

20

30

40

50

M1032-1/IP-1991-PRV)、

2018年3月26日に出願された「DETECTION APPARATUS HAVING A MICROFLUOROMETER, A FLUIDIC SYSTEM, AND A FLOW CELL LATCH CLAMP MODULE」と題する米国特許非仮出願第15/936,365号、

2019年9月11日に出願された「FLOW CELLS AND METHODS RELATED TO SAME」と題する米国特許非仮出願第16/567,224号、

2019年6月12日に出願された「DEVICE FOR LUMINESCENT IMAGING」と題する米国特許非仮出願第16/439,635号、

2017年5月12日に出願された「INTEGRATED OPTOELECTRONIC READ HEAD AND FLUIDIC CARTRIDGE USEFUL FOR NUCLEIC ACID SEQUENCING」と題する米国特許非仮出願第15/594,413号、

2019年3月12日に出願された「ILLUMINATION FOR FLUORESCENCE IMAGING USING OBJECTIVE LENS」と題する米国特許非仮出願第16/351,193号、

2009年12月15日に出願された「DYNAMIC AUTOFOCUS METHOD AND SYSTEM FOR ASSAY IMAGER」と題する米国特許非仮出願第12/638,770号、

2013年3月1日に出願された「KINETIC EXCLUSION AMPLIFICATION OF NUCLEIC ACID LIBRARIES」と題する米国特許非仮出願第13/783,043号、

2011年1月13日に出願された「DATA PROCESSING SYSTEM AND METHODS」と題する米国特許非仮出願第13/006,206号、

2014年10月31日に出願された「IMAGE ANALYSIS USEFUL FOR PATTERNED OBJECTS」と題する米国特許非仮出願第14/530,299号、

2014年12月3日に出願された「METHODS AND SYSTEMS FOR ANALYZING IMAGE DATA」と題する米国特許非仮出願第15/153,953号、

2013年9月6日に出願された「CENTROID MARKERS FOR IMAGE ANALYSIS OF HIGH DENSITY CLUSTERS IN COMPLEX POLYNUCLEOTIDE SEQUENCING」と題する米国特許非仮出願第14/020,570号、

2014年10月31日に出願された「IMAGE ANALYSIS USEFUL FOR PATTERNED OBJECTS」と題する米国特許非仮出願第14/530,299号、

2009年9月23日に出願された「METHOD AND SYSTEM FOR DETERMINING THE ACCURACY OF DNA BASE IDENTIFICATIONS」と題する米国特許非仮出願第12/565,341号、

2007年3月30日に出願された「SYSTEMS AND DEVICES FOR SEQUENCE BY SYNTHESIS ANALYSIS」と題する米国特許非仮出願第12/295,337号、

2008年1月28日に出願された「IMAGE DATA EFFICIENT GENETIC SEQUENCING METHOD AND SYSTEM」と題する米国特許非仮出願第12/020,739号、

2013年3月15日に出願された「BIOSENSORS FOR BIOLOGICAL OR CHEMICAL ANALYSIS AND SYSTEMS AND METHODS FOR SAME」と題する米国特許非仮出願第13/833,619号（代理人整理番号IP-0626-US）、

10

20

30

40

50

2016年6月7日に出願された「BIOSENSORS FOR BIOLOGICAL OR CHEMICAL ANALYSIS AND METHODS OF MANUFACTURING THE SAME」と題する米国特許仮出願第15/175,489号（代理人整理番号IP-0689-US）、

2013年4月26日に出願された「MICRODEVICES AND BIOSENSOR CARTRIDGES FOR BIOLOGICAL OR CHEMICAL ANALYSIS AND SYSTEMS AND METHODS FOR THE SAME」と題する米国特許仮出願第13/882,088号（代理人整理番号IP-0462-US）、

2012年9月21日に出願された「METHODS AND COMPOSITIONS FOR NUCLEIC ACID SEQUENCING」と題する米国特許仮出願第13/624,200号（代理人整理番号IP-0538-US）、

2019年3月21日に出願された「TRAINING DATA GENERATION FOR ARTIFICIAL INTELLIGENCE-BASED SEQUENCING」と題する米国特許仮出願第62/821,602号（代理人整理番号ILLM1008-1/IP-1693-PRV）、

2019年3月21日に出願された「ARTIFICIAL INTELLIGENCE-BASED GENERATION OF SEQUENCING METADATA」と題する米国特許仮出願第62/821,618号（代理人整理番号ILLM1008-3/IP-1741-PRV）、

2019年3月21日に出願された「ARTIFICIAL INTELLIGENCE-BASED BASE CALLING」と題する米国特許仮出願第62/821,681号（代理人整理番号ILLM1008-4/IP-1744-PRV）、

2019年3月21日に出願された「ARTIFICIAL INTELLIGENCE-BASED QUALITY SCORING」と題する米国特許仮出願第62/821,724号（代理人整理番号ILLM1008-7/IP-1747-PRV）、

2019年3月21日に出願された「ARTIFICIAL INTELLIGENCE-BASED SEQUENCING」と題する米国特許仮出願第62/821,766号（代理人整理番号ILLM1008-9/IP-1752-PRV）、

2019年6月14日に出願された「TRAINING DATA GENERATION FOR ARTIFICIAL INTELLIGENCE-BASED SEQUENCING」と題する蘭国特許出願第2023310号（代理人整理番号ILLM1008-11/IP-1693-NL）、

2019年6月14日に出願された「ARTIFICIAL INTELLIGENCE-BASED GENERATION OF SEQUENCING METADATA」と題する蘭国特許出願第2023311号（代理人整理番号ILLM1008-12/IP-1741-NL）、

2019年6月14日に出願された「ARTIFICIAL INTELLIGENCE-BASED BASE CALLING」と題する蘭国特許出願第2023312号（代理人整理番号ILLM1008-13/IP-1744-NL）、

2019年6月14日に出願された「ARTIFICIAL INTELLIGENCE-BASED QUALITY SCORING」と題する蘭国特許出願第2023314号（代理人整理番号ILLM1008-14/IP-1747-NL）、及び

2019年6月14日に出願された「ARTIFICIAL INTELLIGENCE-BASED SEQUENCING」と題する蘭国特許出願第2023316号（代理人整理番号ILLM1008-15/IP-1752-NL）。

2020年3月20日に出願された「TRAINING DATA GENERATION FOR ARTIFICIAL INTELLIGENCE-BASED SEQUENCING」と題する米国特許仮出願第16/825,987号（代理人整理番号ILLM1008-16/IP-1693-US）、

10

20

30

40

50

2020年3月20日に出願された「TRAINING DATA GENERATION FOR ARTIFICIAL INTELLIGENCE-BASED SEQUENCING」と題する米国特許仮出願第16/825,991号（代理人整理番号ILLM1008-17/IP-1741-US）、

2020年3月20日に出願された「ARTIFICIAL INTELLIGENCE-BASED BASE CALLING」と題する米国特許仮出願第16/826,126号（代理人整理番号ILLM1008-18/IP-1744-US）、

2020年3月20日に出願された「ARTIFICIAL INTELLIGENCE-BASED QUALITY SCORING」と題する米国特許仮出願第16/826,134号（代理人整理番号ILLM1008-19/IP-1747-US）、

2020年3月21日に出願された「ARTIFICIAL INTELLIGENCE-BASED SEQUENCING」と題する米国特許仮出願第16/826,168号（代理人整理番号ILLM1008-20/IP-1752-PRV）、

2019年5月16日に出願された「SYSTEMS AND DEVICES FOR CHARACTERIZATION AND PERFORMANCE ANALYSIS OF PIXEL-BASED SEQUENCING」と題する米国特許仮出願第62/849,091号（代理人整理番号ILLM1011-1/IP-1750-PRV）、

2019年5月16日に出願された「BASE CALLING USING CONVOLUTIONS」と題する米国特許仮出願第62/849,132号（代理人整理番号ILLM1011-2/IP-1750-PR2）、

2019年5月16日に出願された「BASE CALLING USING COMPACT CONVOLUTIONS」と題する米国特許仮出願第62/849,133号（代理人整理番号ILLM1011-3/IP-1750-PR3）、

2020年2月20日に出願された「ARTIFICIAL INTELLIGENCE-BASED BASE CALLING OF INDEX SEQUENCES」と題する米国特許仮出願第62/979,384号（代理人整理番号ILLM1015-1/IP-1857-PRV）、

2020年2月20日に出願された「ARTIFICIAL INTELLIGENCE-BASED MANY-TO-MANY BASE CALLING」と題する米国特許仮出願第62/979,414号（代理人整理番号ILLM1016-1/IP-1858-PRV）、

2020年2月20日に出願された「KNOWLEDGE DISTILLATION-BASED COMPRESSION OF ARTIFICIAL INTELLIGENCE-BASED BASE CALLER」と題する米国特許仮出願第62/979,385号（代理人整理番号ILLM1017-1/IP-1859-PRV）、

2020年2月20日に出願された「MULTI-CYCLE CLUSTER BASED REAL TIME ANALYSIS SYSTEM」と題する米国特許仮出願第62/979,412号（代理人整理番号ILLM1020-1/IP-1866-PRV）、

2020年2月20日に出願された「DATA COMPRESSION FOR ARTIFICIAL INTELLIGENCE-BASED BASE CALLING」と題する米国特許仮出願第62/979,411号（代理人整理番号ILLM1029-1/IP-1964-PRV）、及び

2020年2月20日に出願された「SQUEEZING LAYER FOR ARTIFICIAL INTELLIGENCE-BASED BASE CALLING」と題する米国特許仮出願第62/979,399号（代理人整理番号ILLM1030-1/IP-1782-PRV）。

【背景技術】

【0004】

10

20

30

40

50

このセクションで考察される主題は、単にこのセクションにおける言及の結果として、先行技術であると想定されるべきではない。同様に、このセクションで言及した問題、又は背景として提供された主題と関連付けられた問題は、先行技術において以前に認識されていると想定されるべきではない。このセクションの主題は、単に、異なるアプローチを表し、それ自体はまた、特許請求される技術の実施態様に対応し得る。

【0005】

本開示は、配列決定ラン中に、画像データを分析してクラスターをベースコールすることに関する。画像データの分析による1つの課題は、ベースコールされるクラスター集団におけるクラスターの強度プロファイルの変動である。これは、配列決定ラン中のデータスループットの低下及び誤差率の増加を引き起こす。

10

【0006】

クラスター間強度プロファイルの変動には多くの潜在的な理由がある。該変動は、クラスター集団の断片長分布によって引き起こされるクラスター輝度の違いに起因し得る。該変動は、クラスター内の分子がいくつかの配列決定サイクルでヌクレオチドを取り込まず、他の分子より遅れているとき、又は分子が単一の配列決定サイクルで2つ以上のヌクレオチドを取り込むときに生じる位相誤差に起因し得る。該変動は、退色、すなわち、配列決定ランが進行する際の、過剰な洗浄及びレーザ露光による配列決定サイクル数の関数としてのクラスターの信号強度の指数関数的減衰に起因し得る。該変動は、発育不良クラスターコロニー、すなわち、パターン化されたフローセル上に空又は部分的にしか充填されないウェルを生成する小さなクラスターサイズに起因し得る。該変動は、非排他的な増幅によって引き起こされるクラスターコロニーの重複に起因し得る。該変動は、例えば、クラスターがフローセルの縁部に位置することによる、照明不足又は不均一な照明に起因し得る。該変動は、放出された信号を不明確化するフローセル上の不純物に起因し得る。該変動は、多クローン性クラスター、すなわち、複数のクラスターが同一のウェルに堆積される場合に起因し得る。

20

【0007】

クラスター間強度プロファイルの変動を補正する機会が生じる。結果として、配列決定ラン中のベースコールスループットの改善及びベースコール誤差率の低減が得られ得る。

【図面の簡単な説明】

【0008】

30

特許又は出願ファイルは、カラーで創作された少なくとも1つの図面を含む。カラー図面（単数又は複数）を有するこの特許又は特許出願公開のコピーは、必要な料金の要求及び支払いの際に、庁によって提供される。カラー図面はまた、補足コンテンツタブを介してPAIR（patent application information retrieval：特許出願情報検索）で利用可能であり得る。

図面では、同様の参照文字は、概して、異なる図全体を通して同様の部分を指す。また、図面は必ずしも縮尺どおりではなく、その代わりに、開示された技術の原理を例示することを強調している。以下の説明において、開示された技術の様々な実施態様は、以下の図面を参照して説明される。

【図1】開示された技術によって発見され補正されたクラスター間強度プロファイルの変動の一例を示す。

40

【図2】本明細書に開示された変動補正ロジックを実装するベースコールパイプラインの一例を示す。

【図3】本明細書に開示された最小二乗法を実装する最小二乗法決定器を示す。

【図4】現在の配列決定サイクルでターゲットクラスターについてチャンネル固有の分布強度がどのように測定されるかを表す例を示す。

【図5】現在の配列決定サイクルでターゲットクラスターについてチャンネル固有の強度誤差がどのように計算されるかを表す例を示す。

【図6】現在の配列決定サイクルでターゲットクラスターについて分布重心から原点の距離がどのように計算されるかを表す例を示す。

50

【図 7】変動補正ロジックを実装するベースコールパイプラインの別の例を示す。

【図 8】本明細書に記載の重み付け関数の一実施態様を示す。

【図 9】最尤重みを変動補正係数に直接適用する一実施態様を示す。

【図 10】指数関数的減衰因子を変動補正係数に適用する一実施態様を示す。

【図 11】チャンネル固有のオフセット係数を決定する別の実施態様を示す。

【図 12】3つのアプローチ、すなわち、スケーリングのみの解法、オフセットのみの解法（図 11 で論じられる）、及び最小二乗法（図 3 で論じられる）の実行を比較する。

【図 13】3つのアプローチ、すなわち、スケーリングのみの解法、オフセットのみの解法（図 11 で論じられる）、及び最小二乗法（図 3 で論じられる）の実行を比較する。

【図 14】3つのアプローチ、すなわち、スケーリングのみの解法、オフセットのみの解法（図 11 で論じられる）、及び最小二乗法（図 3 で論じられる）の実行を比較する。

【図 15】開示された技術を実装するために使用され得るコンピュータシステムである。

【発明を実施するための形態】

【0009】

以下の考察は、開示される技術を当業者が作製及び使用することを可能にするために提示され、特定の用途及びその要件に関連して提供される。開示される実施態様に対する様々な修正は、当業者には容易に明らかとなり、本明細書で定義される一般原理は、開示される技術の趣旨及び範囲から逸脱することなく、他の実施態様及び用途に適用され得る。したがって、開示される技術は、示される実施態様に限定されることを意図するものではなく、本明細書に開示される原理及び特徴と一致する最も広い範囲を与えられるものである。

【0010】

序論

開示された技術の開発は、配列決定ラン中にベースコールされるクラスター集団内のクラスターの強度プロファイルの分析を開始した。分析により、クラスター集団内のクラスターの強度プロファイルは同様の形態（例えば、台形）を取るが規模は異なり、多次元空間 100 の原点 132 からシフトすることが明らかとなった。これを「クラスター間強度プロファイル変動」と称する。多次元空間 100 は、デカルト空間、極空間、円筒空間、又は球状空間であり得る。

【0011】

図 1 は、開示された技術によって発見され補正されたクラスター間強度プロファイルの変動の一例を示す。図 1 は、クラスター集団におけるクラスター 1、2、及び 3 の強度プロファイル 112、122、及び 132 をそれぞれ示す。ターゲットクラスターの強度プロファイルは、配列決定ランの複数の配列決定サイクル（例えば、150）でのターゲットクラスターへのヌクレオチドの取り込みに起因して生成された化学発光信号を捕捉する強度値を含む。

【0012】

図 1 に示される実施態様では、強度値は、複数の配列決定サイクルの中のそれぞれの配列決定サイクルでシーケンサによって生成された 2 つの異なる色／強度チャンネル配列決定画像から抽出される。シーケンサの例としては、Illumina の iSeq、HiSeq X、HiSeq 3000、HiSeq 4000、HiSeq 2500、NovaSeq 6000、NextSeq 550、NextSeq 1000、NextSeq 2000、NextSeq Dx、MiSeq、及び MiSeq Dx が挙げられる。

【0013】

一実施態様では、シーケンサは、配列決定画像を生成するための合成による配列決定（sequencing by synthesis、SBS）を使用する。SBS は、新たに添加されたそれぞれのヌクレオチドの放出信号を追跡しながら、蛍光標識されたヌクレオチドを有するクラスター鎖に相補的な新生鎖を伸長させることに依存する。蛍光標識されたヌクレオチドは、ヌクレオチド型のフルオロフォアシグナルをアンカーする、3' の取り外し可能なブロックを有する。SBS は、反復配列決定サイクルで行われ、それぞれは、（a）蛍光標

識されたヌクレオチドを添加することによって出現鎖を伸長させることと、(b)シーケンスの光学システムの1つ以上のレーザを使用してフルオロフォアを励起させ、光学システムの異なるフィルタを通した撮像によって配列決定画像を生成することと、(c)次の配列決定サイクルに備えてフルオロフォアを切断し、3'ブロックを除去することと、の3つの工程を含む。取り込み及び撮像サイクルは、指定された数の配列決定サイクルまで繰り返され、リード長を定義する。このアプローチを使用して、それぞれの配列決定サイクルはクラスター鎖に沿って新しい位置を照合する。

【0014】

図1では、

【0015】

【数1】

「X」

【0016】

記号は、クラスター1の強度値を表し、

【0017】

【数2】

「十」

【0018】

記号は、クラスター2の強度値を表し、

【0019】

【数3】

「●」

【0020】

記号は、クラスター3の強度値を表す。4つの異なるヌクレオチドタイプ／塩基A、C、T、及びGの同一性は、2つのカラー画像、すなわち、第1及び第2の強度チャンネルにおける強度値の組み合わせとして符号化される。例えば、核酸は、第1の強度チャンネル（多次元空間100のx軸）で検出される第1のヌクレオチドタイプ（例えば、塩基T）、第2の強度チャンネル（多次元空間100のy軸）で検出される第2のヌクレオチドタイプ（例えば、塩基C）、第1及び第2の強度チャンネルの両方で検出される第3のヌクレオチドタイプ（例えば、塩基A）、並びにいずれの強度チャンネルでも検出されないか、又は最小限しか検出されない、標識を欠く第4のヌクレオチドタイプ（例えば、塩基G）を提供することによって配列決定され得る。

【0021】

いくつかの実施態様では、強度プロファイルは、4つの強度分布（例えば、ガウス分布）が、第1及び第2の強度チャンネルでの強度値に反復的に適合されることによって生成される。4つの強度分布は、4つの塩基A、C、T、及びGに対応する。強度プロファイルでは、第1の強度チャンネルでの強度値は、第2の強度チャンネルでの強度値に対して（例えば、散布図として）プロットされ、強度値は4つの強度分布に分離される。

【0022】

強度プロファイルは、任意の形状をとることができる（例えば、台形、正方形、矩形、菱形など）。4つの強度分布がベースコールの強度値にどのように適合されるかについての追加の詳細は、米国特許出願公開第2018／0274023（A1）号に見出すことができ、そ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。

【0023】

10

20

30

40

50

一実施態様では、それぞれの強度チャンネルは、光学システムによって使用される複数のフィルタ波長帯域のうちの1つに対応する。別の実施態様では、それぞれの強度チャンネルは、配列決定サイクルにおける複数の撮像事象のうちの1つに対応する。更に別の実施態様では、それぞれの強度チャンネルは、特定のレーザによる照射と光学システムの特定の光学フィルタを通した撮像との組み合わせに対応する。

【0024】

開示された技術は、1チャンネルの実施態様、4チャンネルの実施態様などを使用して生成された配列決定画像に同様に適用され得ることは、当業者には明らかであろう。例えば、4チャンネルの実施態様の場合、4チャンネル固有のオフセット係数が、それぞれ4つの強度チャンネルのシフト変動を補正するために決定される。

【0025】

変動補正ロジック

クラスター集団内の多数（例えば、数千個、数百万個、数億個など）のクラスターの強度プロファイルにおけるクラスター間強度プロファイル変動は、ベースコールスループットの低下及びベースコール誤差率の増加を引き起こす。クラスター間強度プロファイル変動を補正するために、クラスター別に変動補正係数を生成する変動補正ロジックを開示する。

【0026】

2チャンネルの実施態様では、変動補正係数は、クラスター間強度プロファイル変動におけるスケール変動を考慮する増幅係数と、クラスター間強度プロファイル変動における第1及び第2の強度チャンネルに沿ったシフト変動をそれぞれ考慮する2チャンネル固有のオフセット係数と、を含む。別の実装態様では、シフト変動は、異なる強度チャンネル（例えば、第1及び第2の強度チャンネル）に共通のオフセット係数を使用することによって考慮される。

【0027】

ターゲットクラスターの変動補正係数は、配列決定ランの先行する配列決定サイクルでターゲットクラスターについて決定された履歴強度統計量の分析と、現在の配列決定サイクルでターゲットクラスターについて決定された現在の強度統計の分析との組み合わせに基づいて、配列決定ランの現在の配列決定サイクルで生成される。変動補正係数は、配列決定ランの次の配列決定サイクルでターゲットクラスターに登録された次の強度読み取り値を補正するために使用される。補正された次の強度読み取り値は、次の配列決定サイクルでターゲットクラスターをベースコールするために使用される。配列決定ランの連続する配列決定サイクルでそれぞれのクラスターのそれぞれの強度プロファイルにそれぞれの変動補正係数を繰り返し適用した結果、強度プロファイルは一致するようになり、原点132に（例えば、台形の下側隅にある）固定される。

【0028】

図2は、変動補正ロジックを実装するベースコールパイプライン200の例を示す。

【0029】

現在の配列決定サイクル

現在の配列決定サイクル*i*において、シーケンサは配列決定画像202を生成する。配列決定画像202は、クラスター集団内の複数のクラスターに登録された現在の強度データ202を含むと共に、現在の配列決定サイクル*i*でターゲットクラスターに登録された現在の強度データ202*t*を含む。現在の強度データ202*t*の「*t*」は、ターゲットクラスターを指す。

【0030】

現在の強度データ202*t*は、ベースコーラー212に提供される。ベースコーラー212は、現在の強度データ202*t*を処理し、現在の配列決定サイクル*i*でターゲットクラスターについての現在のベースコール222を生成する。ベースコーラー212の例としては、IlluminaのReal-Time Analysis (RTA) ソフトウェア、Illuminaのニューラルネットワークベースのベースコーラー（例えば、米

10

20

30

40

50

国特許公開第2020/0302297(A1)号に記載されている)、及び Illumina のイコライザーベースのベースコーラー (例えば、米国仮特許出願第63/020,449号に記載されている) が挙げられる。

【0031】

現在の配列決定サイクル i では、ターゲットクラスターの強度プロファイルは、現在の強度データ 202t と、現在の配列決定サイクル i に先行する配列決定ランの配列決定サイクル、すなわち、先行する配列決定サイクル $1 \sim i-1$ でターゲットクラスターに登録された現在の履歴強度データと、を含む。現在の強度データ 202t 及び現在の履歴強度データを総称して、現在の利用可能な強度データとする。

【0032】

強度プロファイルでは、4つの強度分布は、4つの塩基 A、C、T、及び G に対応する。一実施態様では、現在のベースコール 222 は、4つの強度分布のうち、現在の強度データ 202t が属する強度分布を決定することによって行われる。いくつかの実施態様では、これは、期待値最大化アルゴリズムを使用することによって達成される。期待値最大化アルゴリズムは、現在利用可能な強度データに最も適合する平均 (重心) 及び分布 (共分散) を観測する尤度を繰り返し最大化する。

【0033】

期待値最大化アルゴリズムを使用することによって、現在の配列決定サイクル i で4つの強度分布が決定されると、4つの強度分布のそれぞれに属する現在の強度データ 202t の尤度が計算される。最も大きい尤度が現在のベースコール 222 を与える。一例として、「m, n」が、それぞれ、第1及び第2の強度チャンネルにおける現在の強度データ 202t の強度値であるとする。期待値最大化アルゴリズムは、4つの強度分布のそれぞれに属する「m, n」の強度値の尤度を表す4つの値を生成する。4つの値の最大値は、呼び出された塩基を識別する。

【0034】

他の実施態様では、k-means クラスタリングアルゴリズム、k-means 様クラスタリングアルゴリズム、ヒストグラムベースの方法などがベースコールに使用され得る。

【0035】

次の配列決定サイクル

次の配列決定サイクル $i+1$ で、強度補正パラメータ決定器 232 は、現在のベースコール 222 に基づいて、ターゲットクラスターについての強度補正パラメータ 242 を決定する。2チャンネルの実施態様では、強度補正パラメータ 242 は、第1の強度チャンネルにおける分布強度、第2の強度チャンネルにおける分布強度、第1の強度チャンネルにおける強度誤差、第2の強度チャンネルにおける強度誤差、分布重心から原点の距離、及び分布強度対強度誤差の類似性測定を含む。

【0036】

強度補正パラメータ 242 のそれぞれを以下のように定義する。

【0037】

1) 第1の強度チャンネルにおける分布強度は、現在の配列決定サイクル i でターゲットクラスターが属する塩基固有の強度分布の重心での第1の強度チャンネルにおける強度値である。塩基固有の強度分布は、現在のベースコール 222 を呼び出すための基礎であることに留意されたい。

【0038】

2) 第2の強度チャンネルにおける分布強度は、塩基固有の強度分布の重心での第2の強度チャンネルにおける強度値である。

【0039】

3) 第1の強度チャンネルにおける強度誤差は、第1の強度チャンネルにおける現在の強度データ 202t の測定強度値と第1の強度チャンネルにおける分布強度との間の差である。

【0040】

4) 第2の強度チャンネルにおける強度誤差は、第2の強度チャンネルにおける現在の強度データ202tの測定強度値と第2の強度チャンネルにおける分布強度との間の差である。

【0041】

5) 分布重心から原点の距離は、塩基固有の強度分布の重心と、塩基固有の強度分布が（例えば、期待値最大化アルゴリズムを使用することによって）適合された多次元空間100の原点132との間のユークリッド距離である。他の実施態様では、マハラノビス距離及び最小共分散行列式（minimum covariance determinant、MCD）距離などの距離メトリック、並びにそれらの関連する重心推定値が使用され得る。

6) 分布強度対強度誤差の類似性測定は、第1及び第2の強度チャンネルにおける分布強度と強度誤差との間のチャンネル単位のドット積の総和である。

【0042】

累積強度補正パラメータ決定器252は、先行する配列決定サイクル*i* - 1からの履歴累積強度補正パラメータ254と共に強度補正パラメータ242を累積して、累積強度補正パラメータ262を決定する。累積の例としては、合計及び平均化が挙げられる。

【0043】

変動補正係数決定器272は、決定された累積強度補正パラメータ262に基づいて、変動補正係数282を決定する。

【0044】

次の配列決定サイクル*i* + 1で、シーケンサは配列決定画像294を生成する。配列決定画像294は、クラスター集団内の複数のクラスターに登録された次の強度データ294を含むと共に、次の配列決定サイクル*i* + 1でターゲットクラスターに登録された次の強度データ294tを含む。次の強度データ294tの「t」は、ターゲットクラスターを指す。

【0045】

強度補正器292は、変動補正係数282を次の強度データ294tに適用して、補正された次の強度データ296tを生成する。補正された次の強度データ296tの「t」は、ターゲットクラスターを指す。

【0046】

次の配列決定サイクル*i* + 1では、ターゲットクラスターの強度プロファイルは、補正された次の強度データ296tと、次の配列決定サイクル*i* + 1に先行する配列決定ランの配列決定サイクル、すなわち、先行する配列決定サイクル1～*i*でターゲットクラスターに登録された次の履歴強度データと、を含む。補正された次の強度データ296t及び次の履歴強度データを総称して、次の利用可能な強度データとする。

【0047】

補正された次の強度データ296tは、ベースコーラー212に提供される。ベースコーラー212は、補正された次の強度データ296tを処理し、次の配列決定サイクル*i* + 1でターゲットクラスターについて次のベースコール298を生成する。次のベースコール298を生成するために、期待値最大化アルゴリズムは、次の利用可能な強度データに最も適合するように、補正された次の強度データ296tに基づいて、平均（重心）及び分布（共変量）を観察する。

【0048】

期待値最大化アルゴリズムを使用することによって、次の配列決定サイクル*i* + 1で4つの強度分布が決定されると、4つの強度分布のそれぞれに属する補正された次の強度データ296tの尤度が計算される。最も大きい尤度が次のベースコール298を与える。

【0049】

ベースコールパイプライン200は、クラスター別に行われ、クラスター集団内の複数のクラスターについて並行して実行されることに留意されたい。また、ベースコールパイプライン200は、配列決定ランの連続する配列決定サイクルに対して（例えば、ペアエンドの配列決定ランにおいてリード1の連続する150の配列決定サイクル及びリード2の別の連続する150の配列決定サイクルに対して）繰り返し実行される。

10

20

30

40

50

【 0 0 5 0 】

最小二乗法

図 3 は、本明細書に開示された最小二乗法 3 0 0 を実装する最小二乗法決定器を示す。最小二乗法 3 0 0 は、累積強度補正パラメータ 2 6 2 及び変動補正係数 2 8 2 の閉形式の式を決定する。最小二乗法決定器 3 0 2 は、強度モデラー 3 1 2 及び最小化器 3 2 2 を含む。

【 0 0 5 1 】

強度モデル 3 1 2 は、以下の式に従って、ターゲットクラスターの測定強度と変動補正係数 2 8 2 との間の関係をモデル化する。

【 0 0 5 2 】

$y_{c,i} = a x_{c,i} + d_i + n_{c,i}$ 等式 (1)

【 0 0 5 3 】

式中、

a は、ターゲットクラスターの増幅係数である。

d_i は、強度チャンネル i のチャンネル固有のオフセット係数である。

$x_{c,i}$ は、現在の配列決定サイクル C でのターゲットクラスターについての強度チャンネル i における分布強度である。

$y_{c,i}$ は、現在の配列決定サイクル C でのターゲットクラスターについての強度チャンネル i における測定強度である。

$n_{c,i}$ は、現在の配列決定サイクル C でのターゲットクラスターについての強度チャンネル i の加法性ノイズである。

【 0 0 5 4 】

最小化器 3 2 2 は、最小二乗法 3 0 0 を使用して、以下の式を最小化する。

【 0 0 5 5 】

【数 4】

$$errorf(\hat{a}, \hat{d}_i) = \sum_{c=1}^C \sum_{i=1}^2 (\hat{a} x_{c,i} + \hat{d}_i - y_{c,i})^2 \quad \text{等式 (2)}$$

【 0 0 5 6 】

式中、

$errorf$ はエラー関数である。

【 0 0 5 7 】

【数 5】

$\hat{a}$

【 0 0 5 8 】

は、ターゲットクラスターの増幅係数である。

【 0 0 5 9 】

【数 6】

$\hat{d}_i$

【 0 0 6 0 】

は、強度チャンネル i のチャンネル固有のオフセット係数である。

C は、現在の配列決定サイクルである。

【 0 0 6 1 】

チェーンルールを使用して、最小化器 3 2 2 は、増幅係数

【 0 0 6 2 】

【 数 7 】

 $\hat{a}$

【 0 0 6 3 】

及びチャネル固有のオフセット係数

【 0 0 6 4 】

【 数 8 】

 $\hat{d}_i$

10

【 0 0 6 5 】

に対する誤差関数の2つの偏微分を計算する。誤差関数を最小限に抑えるために、偏微分は等式2をゼロに設定する。

【 0 0 6 6 】

【 数 9 】

$$\frac{\partial errorf}{\partial \hat{a}} = \frac{\partial}{\partial \hat{a}} \left[\sum_{c=1}^c \sum_{l=1}^2 (\hat{a} x_{c,l} + \hat{d}_i - y_{c,l})^2 \right] = 0 \quad \text{等式 (3)}$$

20

$$\frac{\partial errorf}{\partial \hat{d}_i} = \frac{\partial}{\partial \hat{d}_i} \left[\sum_{c=1}^c \sum_{l=1}^2 (\hat{a} x_{c,l} + \hat{d}_i - y_{c,l})^2 \right] = 0 \quad \text{等式 (4)}$$

【 0 0 6 7 】

チャネル固有の強度誤差 $e_{c,i}$ は、以下のように定義される。

【 0 0 6 8 】

$e_{c,i} = y_{c,i} - x_{c,i}$ 等式 (5)

30

【 0 0 6 9 】

閉形式の式

第1の偏微分は、以下のように増幅係数

【 0 0 7 0 】

【 数 1 0 】

 $\hat{a}$

【 0 0 7 1 】

の閉形式の式を決定する。

40

【 0 0 7 2 】

【 数 1 1 】

$$\frac{\partial errorf}{\partial \hat{a}} = \sum_{c=1}^C \sum_{i=1}^2 2x_{c,i} (\hat{a}x_{c,i} + \hat{d}_i - y_{c,i}) = 0 \quad \text{等式 (6)}$$

$$= \sum_{c=1}^C \sum_{i=1}^2 2x_{c,i} (\hat{a}x_{c,i} - x_{c,i} + \hat{d}_i - y_{c,i} + x_{c,i}) = 0 \quad \text{等式 (7)}$$

$$= \sum_{c=1}^C \sum_{i=1}^2 x_{c,i}^2 (\hat{a} - 1) + \hat{d}_1 x_{c,1} - e_{c,1} x_{c,1} = 0 \quad \text{等式 (8)}$$

10

$$= (\hat{a} - 1) \sum_{c=1}^C (x_{c,1}^2 + x_{c,2}^2) + \hat{d}_1 \sum_{c=1}^C x_{c,1} + \hat{d}_2 \sum_{c=1}^C x_{c,2} - \sum_{c=1}^C e_{c,1} x_{c,1} - \sum_{c=1}^C e_{c,2} x_{c,2} = 0 \quad \text{等式 (9)}$$

20

【 0 0 7 3 】

累積強度補正パラメータ 2 6 2 の閉形式の式

【 0 0 7 4 】

【数 1 2 】

 $\bar{x}_1, \bar{x}_2, \bar{e}_1, \bar{e}_2, \overline{xx},$ 及び $\overline{xe}$

【 0 0 7 5 】

は、以下のように等式 9 を再特性化する。

【 0 0 7 6 】

30

【数 1 3 】

$$= C(\hat{a} - 1)\overline{xx} + C\bar{x}_1\hat{d}_1 + C\bar{x}_2\hat{d}_2 - C\overline{xe} = 0 \quad \text{等式 (10)}$$

$$= (\hat{a} - 1)\overline{xx} + \bar{x}_1\hat{d}_1 + \bar{x}_2\hat{d}_2 - \overline{xe} = 0 \quad \text{等式 (11)}$$

【 0 0 7 7 】

式中、

【 0 0 7 8 】

【数 1 4 】

40

$$\bar{x}_1 = \sum_{c=1}^C x_{c,1} \quad \text{中間項 (1)}$$

$$\bar{x}_2 = \sum_{c=1}^C x_{c,2} \quad \text{中間項 (2)}$$

$$\bar{e}_1 = \sum_{c=1}^C e_{c,1} \quad \text{中間項 (3)} \quad 10$$

$$\bar{e}_2 = \sum_{c=1}^C e_{c,2} \quad \text{中間項 (4)}$$

$$\overline{xx} = \sum_{c=1}^C (x_{c,1}^2 + x_{c,2}^2) \quad \text{中間項 (5)} \quad 20$$

$$\overline{xe} = \sum_{c=1}^C (x_{c,1}e_{c,1} + x_{c,2}e_{c,2}) \quad \text{中間項 (6)}$$

【 0 0 7 9 】

累積強度補正パラメータ 2 6 2 のそれぞれを以下のように定義する。

1) 第 1 の累積強度補正パラメータ

【 0 0 8 0 】

【 数 1 5 】 30

$$\bar{x}_1$$

【 0 0 8 1 】

は、先行する配列サイクル 1 ～ i - 1 のそれぞれ、及び現在の配列決定サイクル i でターゲットクラスターについて測定された第 1 の強度チャンネルにおける分布強度の和である。

2) 第 2 の累積強度補正パラメータ

【 0 0 8 2 】

【 数 1 6 】 40

$$\bar{x}_2$$

【 0 0 8 3 】

は、先行する配列決定サイクル 1 ～ i - 1 のそれぞれ、及び現在の配列決定サイクル i でターゲットクラスターについて測定された第 2 の強度チャンネルにおける分布強度の和である。

3) 第 3 の累積強度補正パラメータ

【 0 0 8 4 】

【 数 1 7 】 50

$$\bar{e}_1$$

【 0 0 8 5 】

は、先行する配列決定サイクル 1 ～ i - 1 のそれぞれ、及び現在の配列決定サイクル i でターゲットクラスターについて計算された第 1 の強度チャンネルにおける強度誤差の和である。

4) 第 4 の累積強度補正パラメータ

【 0 0 8 6 】

【 数 1 8 】

10

$$\bar{e}_2$$

【 0 0 8 7 】

は、先行する配列決定サイクル 1 ～ i - 1 のそれぞれ、及び現在の配列決定サイクル i でターゲットクラスターについて計算された第 2 の強度チャンネルにおける強度誤差の和である。

5) 第 5 の累積強度補正パラメータ

【 0 0 8 8 】

【 数 1 9 】

20

$$\overline{xx}$$

【 0 0 8 9 】

は、先行する配列決定サイクル 1 ～ i - 1 のそれぞれ、及び現在の配列決定サイクル i でターゲットクラスターについて計算された分布重心から原点の距離の和である。

6) 第 6 の累積強度補正パラメータ

【 0 0 9 0 】

【 数 2 0 】

30

$$\overline{xe}$$

【 0 0 9 1 】

は、先行する配列決定サイクル 1 ～ i - 1 のそれぞれ、及び現在の配列決定サイクル i でターゲットクラスターについて計算された分布強度対強度誤差の類似性測定の和である。

【 0 0 9 2 】

第 2 の偏微分は、オフセット係数

【 0 0 9 3 】

【 数 2 1 】

40

$$\hat{d}_i$$

【 0 0 9 4 】

の閉形式の式を以下のように決定する。

【 0 0 9 5 】

【 数 2 2 】

50

$$\frac{\partial errorf}{\partial \hat{d}_i} = \sum_{c=1}^C \sum_{i=1}^2 2 (\hat{a}x_{c,i} + \hat{d}_i - y_{c,i}) = 0 \quad \text{等式 (12)}$$

$$= \sum_{c=1}^C \sum_{i=1}^2 (\hat{a}x_{c,i} - x_{c,i} + \hat{d}_i - y_{c,i} + x_{c,i}) = 0 \quad \text{等式 (13)}$$

$$= \sum_{c=1}^C \sum_{i=1}^2 x_{c,i}^2 \left((\hat{a} - 1) + \sum_{i=1}^2 \hat{d}_i + \sum_{c=1}^C x_{c,i} \right) = 0 \quad \text{等式 (14)} \quad 10$$

【 0 0 9 6 】

次いで、それぞれの強度チャネルについて：

【 0 0 9 7 】

【数 2 3】

$$= C(\hat{a} - 1)\bar{x}_i + C\hat{d}_i - C\bar{e}_i = 0 \quad \text{等式 (15)}$$

$$= (\hat{a} - 1)\bar{x}_i + \hat{d}_i - \bar{e}_i = 0 \quad \text{等式 (16)} \quad 20$$

【 0 0 9 8 】

第 1 の強度チャネル、すなわち、 $i = 1$ について：

【 0 0 9 9 】

【数 2 4】

$$\hat{d}_1 = \bar{e}_1 + (1 - \hat{a})\bar{x}_1 \quad \text{等式 (17)}$$

【 0 1 0 0 】

式中、

【 0 1 0 1 】

【数 2 5】

$$\hat{d}_1$$

【 0 1 0 2 】

は、第 1 の強度チャネルのオフセット係数である。

【 0 1 0 3 】

第 2 の強度チャネル、すなわち、 $i = 2$ について：

【 0 1 0 4 】

【数 2 6】

$$\hat{d}_2 = \bar{e}_2 + (1 - \hat{a})\bar{x}_2 \quad \text{等式 (18)}$$

【 0 1 0 5 】

式中、

【 0 1 0 6 】

【数 2 7】

$$\hat{d}_2$$

【 0 1 0 7 】

は、第 2 の強度チャネルに対するオフセット係数である。

【 0 1 0 8 】

等式 1 7 及び 1 8 の等式 1 1 への代入：

【 0 1 0 9 】

【 数 2 8 】

$$\frac{\partial errorf}{\partial \hat{a}} = (\hat{a} - 1)\overline{xx} + \bar{x}_1[\bar{e}_1 + (1 - \hat{a})\bar{x}_1] + \bar{x}_2[\bar{e}_2 + (1 - \hat{a})\bar{x}_2] - \overline{xe} = 0 \quad \text{等式 (19)}$$

10

$$= (\hat{a} - 1)\overline{xx} + (1 - \hat{a})\bar{x}_1^2 + (1 - \hat{a})\bar{x}_2^2 + \bar{x}_1\bar{e}_1 + \bar{x}_2\bar{e}_2 - \overline{xe} = 0 \quad \text{等式 (20)}$$

$$= (\hat{a} - 1)(\overline{xx} - \bar{x}_1^2 - \bar{x}_2^2) + \bar{x}_1\bar{e}_1 + \bar{x}_2\bar{e}_2 - \overline{xe} = 0 \quad \text{等式 (21)}$$

$$\hat{a} - 1 = \frac{\bar{x}_1\bar{e}_1 + \bar{x}_2\bar{e}_2 - \overline{xe}}{\overline{xx} - \bar{x}_1^2 - \bar{x}_2^2} \quad \text{等式 (22)}$$

20

$$\hat{a} = 1 + \frac{\bar{x}_1\bar{e}_1 + \bar{x}_2\bar{e}_2 - \overline{xe}}{\overline{xx} - \bar{x}_1^2 - \bar{x}_2^2} \quad \text{等式 (23)}$$

【 0 1 1 0 】

式中、

【 0 1 1 1 】

【 数 2 9 】

$$\hat{a}$$

30

【 0 1 1 2 】

は、ターゲットクラスターについての増幅係数である。

【 0 1 1 3 】

別の実装態様では、クラスターごとのメモリ要件を低減するために、異なる強度チャネル（例えば、第 1 及び第 2 の強度チャネル）に対する共通オフセット係数は、制約

【 0 1 1 4 】

【 数 3 0 】

$$\hat{d}_1 = \hat{d}_2$$

40

【 0 1 1 5 】

を導入することによって以下のように決定される。

【 0 1 1 6 】

【 数 3 1 】

50

$$\bar{x} = \frac{1}{2C} \sum_{c=1}^C (x_{c,1} + x_{c,2}) \quad \text{中間項 (1.1)}$$

$$\overline{xx} = \frac{1}{C} \sum_{c=1}^C (x_{c,1}^2 + x_{c,2}^2) \quad \text{中間項 (2.1)}$$

$$\bar{e} = \frac{1}{2C} \sum_{c=1}^C (e_{c,1} + e_{c,2}) \quad \text{中間項 (3.1)}$$

$$\overline{xe} = \frac{1}{C} \sum_{c=1}^C (x_{c,1}e_{c,1} + x_{c,2}e_{c,2}) \quad \text{中間項 (4.1)}$$

$$\hat{a} = 1 + \frac{\overline{xe} - 2\bar{x}\bar{e}}{\overline{xx} - 2\bar{x}^2} \quad \text{等式 (24)}$$

$$\hat{d}_1 = \bar{e} + \bar{x}(1 - \hat{a}) \quad \text{等式 (25)}$$

$$\hat{d}_2 = \bar{e} + \bar{x}(1 - \hat{a}) \quad \text{等式 (26)}$$

10

20

30

40

50

【 0 1 1 7 】

最小二乗法 3 0 0 は、閉形式の式の決定するために配列決定ランの前に実行されることが当業者には明らかであろう。決定されると、閉形式の式は、配列決定ランのそれぞれの配列決定サイクルで反復的に、クラスターごとに配列決定ラン中に生成された強度値に適用される。

【 0 1 1 8 】

強度補正パラメータ

以下の考察では、6つの強度補正パラメータ、すなわち、第1の強度チャンネルにおける分布強度、第2の強度チャンネルにおける分布強度、第1の強度チャンネルにおける強度誤差、第2の強度チャンネルにおける強度誤差、分布重心から原点の距離、及び分布強度対強度誤差の類似性測定が、現在の配列決定サイクルでターゲットクラスターについてどのように決定されるかに焦点を当てる。

【 0 1 1 9 】

強度補正パラメータの数が強度チャンネルの数に応じて変化することは、当業者には明らかであろう。例えば、4チャンネルの実施態様の場合、4チャンネル固有の分布強度及び4チャンネル固有の強度誤差が、それぞれ4つの強度チャンネルについて計算される。

【 0 1 2 0 】

図4は、現在の配列決定サイクル i でターゲットクラスターについてチャンネル固有の分布強度がどのように測定されるかを表す例 4 0 0 を示す。図4では、

【 0 1 2 1 】

【 数 3 2 】

「X」

【 0 1 2 2 】

記号は、現在の配列決定サイクル i 及び先行する配列決定サイクル $1 \sim i-1$ でクラスター 1 に登録された第 1 及び第 2 の強度チャンネルの強度値を表す。

【0 1 2 3】

図 4 では、4 つの強度分布 C 4 0 2、A 4 0 6、G 4 6 2、及び T 4 6 6 が接続されて、クラスター 1 のコンステレーション 1 0 2 を形成している。図 4 では、「*」記号は、現在の配列決定サイクル i でクラスター 1 に登録された第 1 及び第 2 の強度チャンネルにおける測定強度「m, n」4 2 2 を表す。測定強度「m, n」4 2 2 は、強度分布 C 4 0 2 の重心 4 1 4 に最も近いいため、クラスター 1 は、強度分布 C 4 0 2 に属し、したがって、現在の配列決定サイクル i でベースコール C に割り当てられる。

【0 1 2 4】

更に、クラスター 1 は C 強度分布 4 0 2 に属するため、重心 4 1 4 の強度値「a, b」は、現在の配列決定サイクル i でのクラスター 1 についての分布強度である。また、「a」は、第 1 の強度チャンネルのチャンネル固有の分布強度であり、「b」は、第 2 の強度チャンネルのチャンネル固有の分布強度である。

【0 1 2 5】

図 5 は、現在の配列決定サイクル i でターゲットクラスターについてチャンネル固有の強度誤差がどのように計算されるかを表す例 5 0 0 を示す。第 1 の強度チャンネルにおける強度誤差

【0 1 2 6】

【数 3 3】

$$(\overline{er}_1)$$

【0 1 2 7】

5 3 2 は、第 1 の強度チャンネルにおけるチャンネル固有の測定強度 (m) と第 1 の強度チャンネルにおけるチャンネル固有の分布強度 (a) との差として、現在の配列決定サイクル i でクラスター 1 について計算される。

【0 1 2 8】

【数 3 4】

$$\overline{er}_1 = m - a$$

等式 (27)

【0 1 2 9】

第 2 の強度チャンネルにおける強度誤差

【0 1 3 0】

【数 3 5】

$$(\overline{er}_2)$$

【0 1 3 1】

5 0 2 は、第 2 の強度チャンネルにおけるチャンネル固有の測定強度 (n) と第 2 の強度チャンネルにおけるチャンネル固有の分布強度 (b) との差として、現在の配列決定サイクル i でクラスター 1 について計算される。

【0 1 3 2】

【数 3 6】

$$\overline{er}_2 = n - b$$

等式 (28)

【0 1 3 3】

図 6 は、現在の配列決定サイクル i でターゲットクラスターについて分布重心から原点

の距離がどのように計算されるかを表す例 600 を示す。クラスター 1 は、C 強度分布 402 に属し、重心 414 の強度値「a, b」は、現在の配列決定サイクル i でのクラスター 1 についての分布強度である。

【0134】

分布重心から原点の距離は、重心 414 と原点 132 「x, y」との間のユークリッド距離 (d) 652 として、現在の配列決定サイクル i でクラスター 1 について計算される。

【0135】

【数 37】

$$d = \sqrt{(a-x)^2 + (b-y)^2} \quad \text{等式 (29)}$$

10

【0136】

分布強度対強度誤差の類似性測定は、チャンネル固有の分布強度とチャンネル固有の強度誤差との間のチャンネル単位のドット積の総和として、現在の配列決定サイクル i でクラスター 1 について計算される。

【0137】

【数 38】

$$\text{類似性測定} = \overline{er_1} \bullet a + \overline{er_2} \bullet b \quad \text{等式 (30)}$$

20

【0138】

式中、

・ は、ドット積演算子である。

【0139】

ベースコールパイプライン

図 7 は、変動補正ロジックを実装するベースコールパイプライン 700 の別の例を示す。現在の配列決定サイクル i が、配列決定ランの第 25 の配列決定サイクル、すなわち、i = 25 であるとする。先行する配列決定サイクル i - 1 は、配列決定ランの第 24 の配列決定サイクル、すなわち、i - 1 = 24 である。次の配列決定サイクル i + 1 は、配列決定ランの第 26 の配列決定サイクル、すなわち、i + 1 = 26 である。後続の配列決定サイクル i + 2 は、配列決定ランの第 27 の配列決定サイクル、すなわち、i + 2 = 27 である。

30

【0140】

先行する配列決定サイクル

第 1 ~ 24 の配列決定サイクルのそれぞれにおいて、累積強度補正パラメータのそれぞれのセットは、強度補正パラメータのそれぞれのセットから決定される。ターゲットクラスターについての先行する累積強度補正パラメータ 702 は、24 セットの強度補正パラメータの強度補正パラメータ単位の累積である。2 チャンネルの実施態様では、24 セットの強度補正パラメータのうちのそれぞれは、6 つの強度補正パラメータ、すなわち、第 1 の強度チャンネルにおける分布強度、第 2 の強度チャンネルにおける分布強度、第 1 の強度チャンネルにおける強度誤差、第 2 の強度チャンネルにおける強度誤差、分布重心から原点の距離、及び分布強度対強度誤差の類似性を含む。24 セットの累積強度補正パラメータのうちのそれぞれは、6 つの累積強度補正パラメータ

40

【0141】

【数 39】

$$\bar{x}_1, \bar{x}_2, \bar{e}_1, \bar{e}_2, \overline{xx}, \text{ 及び } \overline{xe}$$

【0142】

50

を含む。

【0143】

先行する累積強度補正パラメータ702は、基礎となる先行する強度値及びそれらが計算される先行する強度補正パラメータに関するメタデータ（又は統計）である。結果として、基礎となる先行強度値及び先行強度補正パラメータと比較して、先行する累積強度補正パラメータ702は、はるかに小さいメモリフットプリントを有する。先行する累積強度補正パラメータ702は、配列決定ラン中にメモリにキャッシュされ、ターゲットクラスターについての現在の強度補正パラメータ732と共に累積されて、三角形734によって示されるように、ターゲットクラスターについての現在の累積強度補正パラメータ742を生成する。

10

【0144】

一実装態様では、先行する累積強度補正パラメータ702は、量子化された固定ビット幅フォーマットで記憶される。例えば、1つ又は2つのバイトを使用して、先行する累積強度補正パラメータ702にそれぞれの先行する累積強度補正パラメータが記憶され得る。

【0145】

現在の配列決定サイクル

ターゲットクラスターの現在の測定強度712は、第25の配列決定サイクルでターゲットクラスターに登録された強度値を含む。現在の測定強度712に基づいて、現在のベースコール722は、第25の配列決定サイクルでターゲットクラスターについて呼び出される（例えば、期待値最大化アルゴリズムを使用することによる）。

20

【0146】

次の配列決定サイクル

現在のベースコール722に基づいて、現在の強度補正パラメータ732がターゲットクラスターについて決定される。現在の累積強度補正パラメータ742は、三角形734によって示されるように、現在の強度補正パラメータ732と共に先行する累積強度補正パラメータ702を累積することに基づいて、ターゲットクラスターについて計算される。累積の一例は合計である。合計の実施態様では、現在の累積強度補正パラメータ742は、（上記の中間項1～6に示されるように）強度補正パラメータ別に先行する累積強度補正パラメータ702と現在の強度補正パラメータ732との合計によって計算される。

30

【0147】

累積の別の例は平均化である。

【0148】

【数40】

40

50

$$\bar{x}_1 = \frac{1}{C} \sum_{c=1}^C x_{c,1}$$

中間項 (1.2)

$$\bar{x}_2 = \frac{1}{C} \sum_{c=1}^C x_{c,2}$$

中間項 (2.2)

$$\bar{e}_1 = \frac{1}{C} \sum_{c=1}^C e_{c,1}$$

中間項 (3.2)

10

$$\bar{e}_2 = \frac{1}{C} \sum_{c=1}^C e_{c,2}$$

中間項 (4.2)

$$\overline{xx} = \frac{1}{C} \sum_{c=1}^C (x_{c,1}^2 + x_{c,2}^2)$$

中間項 (5.2)

20

$$\overline{xe} = \frac{1}{C} \sum_{c=1}^C (x_{c,1}e_{c,1} + x_{c,2}e_{c,2})$$

中間項 (6.2)

【 0 1 4 9 】

式中、C は、現在の配列決定サイクル i のインデックスである。すなわち、本明細書で論じられる例では、C = 25 である。

【 0 1 5 0 】

中間項 1 . 2 ~ 6 . 2 に基づいて、累積強度補正パラメータのそれぞれを以下のように定義する。

30

1) 第 1 の累積強度補正パラメータ

【 0 1 5 1 】

【 数 4 1 】

$$\bar{x}_1$$

【 0 1 5 2 】

は、先行する配列決定サイクル 1 ~ i - 1 のそれぞれ、及び現在の配列決定サイクル i でターゲットクラスターについて測定された第 1 の強度チャンネルにおける分布強度の平均である。

40

2) 第 2 の累積強度補正パラメータ

【 0 1 5 3 】

【 数 4 2 】

$$\bar{x}_2$$

【 0 1 5 4 】

は、先行する配列決定サイクル 1 ~ i - 1 のそれぞれ、及び現在の配列決定サイクル i でターゲットクラスターについて測定された第 2 の強度チャンネルにおける分布強度の平均であ

50

る。

3) 第3の累積強度補正パラメータ

【0155】

【数43】

$$\bar{e}_1$$

【0156】

は、先行する配列サイクル1～i-1のそれぞれ、及び現在の配列決定サイクルiでターゲットクラスターについて計算された第1の強度チャンネルにおける強度誤差の平均である。

10

4) 第4の累積強度補正パラメータ

【0157】

【数44】

$$\bar{e}_2$$

【0158】

は、先行する配列サイクル1～i-1のそれぞれ、及び現在の配列決定サイクルiでターゲットクラスターについて計算された第2の強度チャンネルにおける強度誤差の平均である。

20

5) 第5の累積強度補正パラメータ

【0159】

【数45】

$$\overline{xx}$$

【0160】

は、先行する配列サイクル1～i-1のそれぞれ、及び現在の配列決定サイクルiでターゲットクラスターについて計算された分布重心から原点の距離の平均である。

30

6) 第6の累積強度補正パラメータ

【0161】

【数46】

$$\overline{xe}$$

【0162】

は、先行する配列サイクル1～i-1のそれぞれ、及び現在の配列決定サイクルiでターゲットクラスターについて計算された分布強度対強度誤差の類似性測定の平均である。

【0163】

40

コンパクト表現

一実施態様では、先行する累積強度補正パラメータ702は、コンパクト表現（例えば、合計された表現又は平均化された表現）で記憶される。平均化の実施態様では、先行する累積強度補正パラメータ702は、それらの平均化された表現に記憶され、最初に、それらが累積される配列決定サイクルの数が乗算されて、平均化前表現が取り出される。すなわち、本明細書で論じられる実施例においては、24が乗数である。

【0164】

次いで、乗算の結果、すなわち、平均化前表現が、強度補正パラメータ別に現在の強度補正パラメータ732と合計される。次いで、和の結果が、現在の配列決定サイクルiのインデックスC（C=25）で除算されて、現在の累積強度補正パラメータ742が決定

50

される。

【 0 1 6 5 】

【 数 4 7 】

$$\bar{x}_1^{24}$$

【 0 1 6 6 】

が、第 2 4 の配列決定サイクルについての第 1 の累積強度補正パラメータであるとする。 $x^{25}_{c,1}$ が、第 2 5 の配列決定サイクルについての第 1 の強度チャンネルにおける分布強度であるとする。

【 0 1 6 7 】

【 数 4 8 】

$$\bar{x}_1^{25}$$

【 0 1 6 8 】

が、第 2 5 の配列決定サイクルに対する第 1 の累積強度補正パラメータであり、第 2 6 の配列決定サイクルについて測定された強度を補正するために使用されたとする。この場合、以下ようになる。

【 0 1 6 9 】

【 数 4 9 】

$$\bar{x}_1^{25} = \frac{[(\bar{x}_1^{24} * 24) + x^{25}_{c,1}]}{25}$$

【 0 1 7 0 】

【 数 5 0 】

$$\bar{x}_2^{24}$$

【 0 1 7 1 】

が、第 2 4 の配列決定サイクルに対する第 2 の累積強度補正パラメータであるとする。 $x^{25}_{c,2}$ が、第 2 5 の配列決定サイクルについての第 2 の強度チャンネルにおける分布強度であるとする。

【 0 1 7 2 】

【 数 5 1 】

$$\bar{x}_2^{25}$$

【 0 1 7 3 】

が、第 2 5 の配列決定サイクルについての第 2 の累積強度補正パラメータであり、第 2 6 の配列決定サイクルについて測定された強度を補正するために使用されたとする。この場合、以下ようになる。

【 0 1 7 4 】

【 数 5 2 】

$$\bar{x}_2^{25} = \frac{[(\bar{x}_2^{24} * 24) + x^{25}_{c,2}]}{25}$$

【 0 1 7 5 】

【数 5 3】

$$\bar{e}_1^{24}$$

【0 1 7 6】

が、第 2 4 の配列決定サイクルについての第 3 の累積強度補正パラメータであるとする。 $e^{25}_{c,1}$ が、第 2 5 の配列決定サイクルについての第 1 の強度チャンネルにおける強度誤差であるとする。

【0 1 7 7】

【数 5 4】

$$\bar{e}_1^{25}$$

10

【0 1 7 8】

が、第 2 5 の配列決定サイクルについての第 3 の累積強度補正パラメータであり、第 2 6 の配列決定サイクルについて測定された強度を補正するために使用されたとする。この場合、以下ようになる。

【0 1 7 9】

【数 5 5】

$$\bar{e}_1^{25} = \frac{[(\bar{e}_1^{24} * 24) + e^{25}_{c,1}]}{25}$$

20

【0 1 8 0】

【数 5 6】

$$\bar{e}_2^{24}$$

【0 1 8 1】

が、第 2 4 の配列決定サイクルについての第 4 の累積強度補正パラメータであるとする。 $e^{25}_{c,2}$ が、第 2 5 の配列決定サイクルについての第 2 の強度チャンネルにおける強度誤差であるとする。

30

【0 1 8 2】

【数 5 7】

$$\bar{e}_1^{25}$$

【0 1 8 3】

が、第 2 5 の配列決定サイクルについての第 4 の累積強度補正パラメータであり、第 2 6 の配列決定サイクルについて測定された強度を補正するために使用されたとする。この場合、以下ようになる。

40

【0 1 8 4】

【数 5 8】

$$\bar{e}_2^{25} = \frac{[(\bar{e}_2^{24} * 24) + e^{25}_{c,2}]}{25}$$

【0 1 8 5】

【数 5 9】

50

$$\overline{xx}^{24}$$

【 0 1 8 6 】

が、第 2 4 の配列決定サイクルについての第 5 の累積強度補正パラメータであるとする。

【 0 1 8 7 】

【 数 6 0 】

$$[(x_{c,1}^{25})^2 + (x_{c,2}^{25})^2]$$

10

【 0 1 8 8 】

が、第 2 5 の配列決定サイクルについての分布重心から原点の距離であるとする。

【 0 1 8 9 】

【 数 6 1 】

$$\overline{xx}^{25}$$

【 0 1 9 0 】

が、第 2 5 の配列決定サイクルについての第 5 の累積強度補正パラメータであり、第 2 6 の配列決定サイクルについて測定された強度を補正するために使用されたとする。この場合、以下になる。

20

【 0 1 9 1 】

【 数 6 2 】

$$\overline{xx}^{25} = \frac{[(\overline{xx}^{24} * 24) + [(x_{c,1}^{25})^2 + (x_{c,2}^{25})^2]]}{25}$$

【 0 1 9 2 】

【 数 6 3 】

$$\overline{xe}^{24}$$

30

【 0 1 9 3 】

が、第 2 4 の配列決定サイクルについての第 6 の累積強度補正パラメータであるとする。

【 0 1 9 4 】

【 数 6 4 】

$$[(x_{c,1}^{25}e_{c,1}^{25}) + (x_{c,2}^{25}e_{c,2}^{25})]$$

40

【 0 1 9 5 】

が、第 2 5 の配列決定サイクルについての分布強度対強度誤差の類似性測定であるとする。

【 0 1 9 6 】

【 数 6 5 】

$$\overline{xe}^{25}$$

【 0 1 9 7 】

50

が、第 25 の配列決定サイクルについての第 6 の累積強度補正パラメータであり、第 26 の配列決定サイクルについて測定された強度を補正するために使用されたとする。この場合、以下になる。

【0198】

【数 66】

$$\overline{xe}^{25} = \frac{[(\overline{xe}^{24} * 24) + [(x_{c,1}^{25} e_{c,1}^{25}) + (x_{c,2}^{25} e_{c,2}^{25})]]}{25}$$

【0199】

現在の累積強度補正パラメータ 742 は、現在の測定強度 712 及び現在の強度補正パラメータ 732 に関するメタデータ（又は統計）である。結果として、現在の測定強度 712 及び現在の強度補正パラメータ 732 と比較して、現在の累積強度補正パラメータ 742 は、はるかに小さいメモリフットプリントを有する。現在の累積強度補正パラメータ 742 は、配列決定ラン中にメモリにキャッシュされ、三角形 784 によって示されるように、ターゲットクラスターについての次の累積強度補正パラメータ 796 を生成するために、ターゲットクラスターについての次の強度補正パラメータ 794 と共に累積される。

【0200】

一実施態様では、現在の累積強度補正パラメータ 742 は、量子化された固定ビット幅形式で記憶される。例えば、1 つ又は 2 つのバイトを使用して、現在の累積強度補正パラメータ 742 にそれぞれの先行する累積強度補正パラメータが記憶され得る。

【0201】

現在の累積強度補正パラメータ 742 は、ターゲットクラスターについての現在の増幅係数 752 を決定するために使用される。これは、現在の累積強度補正パラメータ 742 に応じて、等式 23 において閉形式の式を実行することを含む。

【0202】

現在の累積強度補正パラメータ 742 及び現在の増幅係数 752 は、ターゲットクラスターについての現在のチャンネル固有のオフセット係数 762 を決定するために使用される。これは、現在の累積強度補正パラメータ 742 及び増幅係数 752 に応じて、等式 17 及び 18 において閉形式の式を実行することを含む。

【0203】

現在の増幅係数 752 及び現在のチャンネル固有のオフセット係数 762 を使用して、第 26 の配列決定サイクルでターゲットクラスターについて測定された次の測定強度 772 が補正される。一実施態様では、補正することは、次の測定強度 772 から現在のチャンネル固有のオフセット係数 762 をチャンネル単位で減算して次のシフト強度を生成することと、次のシフト強度を現在の増幅係数 752 で除算してターゲットクラスターについての次の補正された測定強度 782 を生成することと、を含む。

【0204】

次いで、次のベースコール 792 は、次の補正された測定強度 782 を使用して、第 26 の配列決定サイクルでターゲットクラスターについて呼び出される。これは、次の補正された測定強度 782 を入力としてベースコーラー 212 に提供することによって達成される（例えば、期待値最大化アルゴリズムを使用することによる）。

【0205】

後続の配列決定サイクル

コントローラ（図示せず）は、動作 794、796、798、及び 799 によって例示されるように、配列決定ランの連続する配列決定サイクルについてベースコールパイプライン 700 を繰り返す。例えば、第 27 の配列決定サイクルについて、現在の累積強度補正パラメータ 742 は、三角形 784 によって示されるように、先行する累積強度補正パラメータ 702 として機能する。ベースコールパイプライン 700 は、クラスター別に実

10

20

30

40

50

行され、クラスター集団内の複数のクラスターについて並行して実行されることに留意されたい。

【0206】

重み付けされた最小二乗法

図8は、本明細書に記載の重み付け関数800の一実施態様を示す。最小二乗法300は、収束するまで数回の配列決定サイクルを必要とし得るため、重み付け関数800は、配列決定ランの初期配列決定サイクルで変動補正係数を減衰させ、配列決定ランの後の配列決定サイクルで変動補正係数を増幅させるために使用される。

【0207】

重み付け関数800は、以下のように作用する。最初に、初期増幅係数802並びに初期オフセット係数822及び832が初期化される。一実施態様では、初期増幅係数802は、所定の値（例えば、「1」）を用いて配列決定ランの第1の配列決定サイクルで初期化され、初期オフセット係数822及び832は、所定の値（例えば、「0」）を用いて第1の配列決定サイクルで初期化される。重み付け関数800は、初期増幅係数802を増幅係数806（最小二乗法300によって決定される）と組み合わせ（例えば、合計して）、初期の第1及び第2のオフセット係数822及び832を第1及び第2のオフセット係数826及び836（最小二乗法300によって決定される）と組み合わせ、その結果、増幅係数806並びに第1及び第2のオフセット係数826及び836は、初期配列決定サイクルで減衰し、後の配列決定サイクルで増幅する。

【0208】

一実施態様では、重み付け関数800は、初期の最小重み（inimin重み）804を初期増幅係数802並びに初期の第1及び第2のオフセット係数822及び832に、また、最小二乗法による最大重み（lsqmax重み）808を増幅係数806並びに第1及び第2のオフセット係数826及び836に適用（例えば、乗算）し、その結果、以下になる。

【0209】

【数67】

$$\text{lsqmax 重み} = \left(0, \frac{c-p}{c}\right) \quad \text{等式 (31)}$$

$$\text{inimin 重み} = \left(1, 1 - \left\lceil \frac{c-p}{c} \right\rceil\right) \quad \text{等式 (32)}$$

【0210】

式中、

cは、現在の配列決定サイクルのインデックスである。

pは2～7の数値である。

【0211】

第1の配列決定サイクル、すなわち、 $c = 1$ の場合、及び $p = 2$ の場合、式

【0212】

【数68】

$$\left(\frac{c-p}{c}\right)$$

【0213】

は「-1」に等しい。次いで、「0」と「-1」との間で、lsqmax重み808は、2つの値のうちの最大値、すなわち、0を選択する。式

【0214】

【数 6 9】

$$(1 - \left\lfloor \frac{c-p}{c} \right\rfloor)$$

【0 2 1 5】

は、「2」に等しい。次いで、「1」と「2」との間で、inimin 重み 8 0 4 は、2 つの値のうちの最小値、すなわち、1 を選択する。

【0 2 1 6】

引き続き、lsqmax 重み 8 0 8 からの 0 に増幅係数 8 0 6 並びに第 1 及び第 2 のオフセット係数 8 2 6 及び 8 3 6 が乗算され、inimin 重み 8 0 4 からの 1 に、初期増幅係数 8 0 2 並びに初期の第 1 及び第 2 のオフセット係数 8 2 2 及び 8 3 2 が乗算される。2 つの乗算の結果が合計されて、重み付けされた増幅係数 8 1 0 並びに重み付けされた第 1 及び第 2 のオフセット係数 8 2 0 及び 8 3 0 が生成される。

【0 2 1 7】

配列決定ランが進行し、インデックス「c」の値が増分すると、lsqmax 重み 8 0 8 及び inimin 重み 8 0 4 の値も変化し、同様に適用され、その結果、増幅係数 8 0 6 並びに第 1 及び第 2 のオフセット係数 8 2 6 及び 8 3 6（最小二乗法 3 0 0 から学習される）は、それぞれの連続する配列決定サイクルで漸進的に増幅する。

【0 2 1 8】

重み付け関数 8 0 0 は、重み付けされた増幅係数 8 1 0 並びに重み付けされた第 1 及び第 2 のオフセット係数 8 2 0 及び 8 3 0 を生成し、これらを使用して、次の配列決定サイクル i + 1 でターゲットクラスターについて測定強度が補正され、次の配列決定サイクル i + 1 でターゲットクラスターをベースコールするために補正された測定強度が生成される。

【0 2 1 9】

【数 7 0】

$$\hat{a} = 1 + \frac{W\overline{xe} - \bar{x}_1\bar{e}_1 - \bar{x}_2\bar{e}_2}{W\overline{xx} - \bar{x}_1^2 - \bar{x}_2^2} \quad \text{等式 (33)}$$

$$\hat{d}_1 = \frac{-\bar{x}_1W\overline{xe} + (W\overline{xx} - \bar{x}_2^2)\bar{e}_1 + \bar{x}_1\bar{x}_2\bar{e}_2}{W(W\overline{xx} - \bar{x}_1^2 - \bar{x}_2^2)} \quad \text{等式 (34)}$$

$$\hat{d}_2 = \frac{-\bar{x}_2W\overline{xe} + (W\overline{xx} - \bar{x}_1^2)\bar{e}_2 + \bar{x}_1\bar{x}_2\bar{e}_1}{W(W\overline{xx} - \bar{x}_1^2 - \bar{x}_2^2)} \quad \text{等式 (35)}$$

【0 2 2 0】

式中、

W は、重みである。

【0 2 2 1】

最尤法

図 9 は、最尤重み 9 0 6、9 0 8、及び 9 1 0 を変動補正係数に直接適用する一実施態様を示す。最尤重み 9 0 6、9 0 8、及び 9 1 0 は、最尤法 9 0 0 を、以前の配列決定ランで変動補正係数について観察された履歴値の確率分布 9 0 2 に適用することによって生成される。図 9 はまた、累積強度補正パラメータ 9 0 4 を示す。

【0 2 2 2】

最尤重み 9 0 6、9 0 8、及び 9 1 0 は、インデックス「C」で表されるように、現在の配列決定サイクルの関数である。最尤重み 9 0 6、9 0 8、及び 9 1 0 は、インデック

スCに依存して、配列決定サイクル別に変化する。最尤重み906、908、及び910はまた、文字「n」で表されるように、加法性ノイズの関数である。シグマ項「 σ 」は、それぞれの変動補正係数、すなわち、分散(σ^2)について観察された履歴値の変動範囲を表す。いくつかの実施態様では、加法性ノイズに対するシグマ項は、最尤法900を使用して推定され得るか、又はユーザ指定され得る。増幅係数、チャンネル固有のオフセット係数、及び加法性ノイズのシグマ項は、配列決定ラン別に決定され、配列決定ランの全ての配列決定サイクルに対して固定されたままである。シグマ項には、変動補正係数で観察された不確実性に関する事前の知識を組み込まれる。

【0223】

シグマ項のいくつかの例示的な値は、以下のとおりである。

【0224】

```
'm l _ _ c h a n e s t _ _ s i g m a _ _ a ' ,    0 . 1 5
'm l _ _ c h a n e s t _ _ s i g m a _ _ d 1 ' ,    0 . 1
'm l _ _ c h a n e s t _ _ s i g m a _ _ d 2 ' ,    0 . 0 2
'm l _ _ c h a n e s t _ _ s i g m a _ _ n ' ,    0 . 1 4
```

【0225】

一実施態様では、増幅係数の確率分布の中心／初期／平均値は、「1」に設定され、チャンネル固有のオフセット係数の確率分布の中心／初期／平均値は、「0」に設定される。

【0226】

最尤重み906、908、及び910における増幅係数及びチャンネル固有のオフセット係数のシグマ項のより小さい値は、それぞれの履歴値における低変動を示す。これにより、最尤重み906、908、及び910の値が高くなる。これは次いで、特に早期の配列決定サイクルにおいて、中心値1を優先して重み付けされる、重み付けされた増幅係数920チャンネルと、中心値0を優先して重み付けされる、重み付けされたチャンネル固有のオフセット係数930及び940と、をもたらす。

【0227】

逆に、最尤重み906、908、及び910における増幅係数及びチャンネル固有のオフセット係数のシグマ項のより大きい値は、それぞれの履歴値における高変動を示す。これにより、最尤重み906、908、及び910の値が低くなる。これは次いで、特に後期の配列サイクルにおいて、最小二乗法300（例えば、等式23）の出力を優先して重み付けされる、重み付けされた増幅係数920と、最小二乗法300（例えば、等式17及び18）の出力を優先して重み付けされる、重み付けされたチャンネル固有のオフセット係数930及び940と、をもたらす。

【0228】

最尤重み906、908、及び910は、それぞれ、重み付けされた増幅係数920並びに重み付けされたチャンネル固有のオフセット係数930及び940を計算するために直接組み込まれる。

【0229】

指数関数的減衰因子法

図10は、指数関数的減衰因子を変動補正係数に適用する一実施態様を示す。指数関数的減衰ロジック1000は、いわゆる「 τ 」及び「 stats.cycle 」に基づく。「 stats.cycle 」という用語は、現在の配列決定サイクルを指す。

【0230】

τ は、強度補正パラメータで観察された時間分散の程度に応じて所定の値に設定される。強度補正パラメータが時間不変である場合、 τ は無限に設定され得る。強度補正パラメータが急速に時間変化する場合、 τ は小さい値に設定され得る。一実施態様では、 τ は32に設定される。

【0231】

τ が32であるとする。次いで、ステートメント1002、1004、及び1006によれば、減衰因子は、配列サイクル1～31については「1」であり、これは、累積

10

20

30

40

50

強度補正パラメータの減衰をもたらさない。配列決定サイクル32以上については、ステートメント1008に基づき、減衰因子は32に対して31である。指数関数的減衰特性は、ステートメント1010、1012、1014、1016、1018、及び1020において、以下に示すように、それぞれの連続する配列決定サイクルで累積強度補正パラメータのそれぞれに減衰因子が乗算されるという事実由来する。

【0232】

配列決定サイクル32、減衰因子

【0233】

【数71】

$$= \frac{31}{32}$$

10

【0234】

の場合

【0235】

【数72】

$$\bar{x}_1^D = \frac{31}{32} * \bar{x}_1$$

減衰中間項 (1.3)

$$\bar{x}_2^D = \frac{31}{32} * \bar{x}_2$$

減衰中間項 (2.3)

20

$$\bar{e}_1^D = \frac{31}{32} * \bar{e}_1$$

減衰中間項 (3.3)

$$\bar{e}_2^D = \frac{31}{32} * \bar{e}_2$$

減衰中間項 (4.3)

$$\overline{xx^D} = \frac{31}{32} * \overline{xx}$$

減衰中間項 (5.3)

30

$$\overline{xe^D} = \frac{31}{32} * \overline{xe}$$

減衰中間項 (6.3)

【0236】

配列決定サイクル33、減衰因子

【0237】

【数73】

$$= \left[\frac{31}{32} \right]^2$$

40

【0238】

の場合

【0239】

【数74】

50

$$\bar{x}_1^D = \left[\frac{31}{32} \right]^2 * \bar{x}_1 \quad \text{減衰中間項 (1.4)}$$

$$\bar{x}_2^D = \left[\frac{31}{32} \right]^2 * \bar{x}_2 \quad \text{減衰中間項 (2.4)}$$

$$\bar{e}_1^D = \left[\frac{31}{32} \right]^2 * \bar{e}_1 \quad \text{減衰中間項 (3.4)}$$

$$\bar{e}_2^D = \left[\frac{31}{32} \right]^2 * \bar{e}_2 \quad \text{減衰中間項 (4.4)}$$

$$\overline{xx}^D = \left[\frac{31}{32} \right]^2 * \overline{xx} \quad \text{減衰中間項 (5.4)}$$

$$\overline{xe}^D = \left[\frac{31}{32} \right]^2 * \overline{xe} \quad \text{減衰中間項 (6.4)}$$

10

20

【 0 2 4 0 】

図 1 0 では、累積強度補正パラメータは、和演算を使用して累積される。指数関数的減衰因子の平均化の実施態様では、累積強度補正パラメータは、平均演算を使用して累積される。指数関数的減衰因子の平均化の実施態様では、中間項 1 . 2 ~ 6 . 2 における除数「C」は、配列決定ランの t a u 数の配列決定サイクルの後、すなわち、配列決定ランの第 3 2 の配列決定サイクルの後に固定されたままである。

【 0 2 4 1 】

いくつかの実施態様では、重み付けされた最小二乗法（図 8）、最尤法（図 9）、及び指数関数的減衰因子法（図 1 0）を組み合わせ、配列決定ランのそれぞれの配列決定サイクルで重み付けされた変動補正係数が生成される。

30

【 0 2 4 2 】

チャンネル固有のオフセット係数

図 1 1 は、チャンネル固有のオフセット係数を決定する別の実施態様を示す。2チャンネルの実施態様では、第 1 の強度チャンネルに対する、及び現在の配列決定サイクルでのターゲットクラスターについての第 1 のチャンネル固有のオフセット係数（「Δ x」）が、現在の配列決定サイクルでのターゲットクラスターについての第 1 の強度チャンネルにおける測定強度（「p」）と、（例えば、期待値最大化アルゴリズムによって決定されるように）現在の配列決定サイクルでターゲットクラスターが属する塩基固有の強度分布 A 1 1 0 2 の重心 1 1 0 4 での第 1 の強度チャンネルにおける強度値（「u」）との間の差として計算される。

40

【 0 2 4 3 】

2チャンネルの実施態様では、第 2 の強度チャンネルに対する、及び現在の配列決定サイクルでのターゲットクラスターについての第 2 のチャンネル固有のオフセット係数（「Δ y」）が、現在の配列決定サイクルでのターゲットクラスターについての第 2 の強度チャンネルにおける測定強度（「q」）と、現在の配列決定サイクルでターゲットクラスターが属する塩基固有の強度分布 A 1 1 0 2 の重心 1 1 0 4 での第 2 の強度チャンネルにおける強度値（「v」）との間の差として計算される。

【 0 2 4 4 】

一実施態様では、第 1 のチャンネル固有のオフセット係数（「Δ x」）及び第 2 のチャネ

50

ル固有のオフセット係数（「 Δy 」）は、配列決定ランのそれぞれの配列決定サイクルで決定される。いくつかの実施態様では、構成可能な数の配列決定サイクル（例えば、10又は20の配列決定サイクル）後に、第1のチャンネル固有のオフセット係数（「 Δx 」）及び第2のチャンネル固有のオフセット係数（「 Δy 」）が所定の値（例えば、「0」）で初期化される。

【0245】

ローリング平均の実施態様では、平均は、構成可能な数の配列決定サイクル後に、第1のオフセットチャンネル固有の係数（「 Δx 」）及び第2のチャンネル固有のオフセット係数（「 Δy 」）について計算される。次いで、平均は、次の平均が構成可能な数の配列決定サイクルの次のセットについて計算されるまで、第1のオフセットチャンネル固有の係数（「 Δx 」）及び第2のチャンネル固有のオフセット係数（「 Δy 」）の代用として使用される。

10

【0246】

いくつかの実装態様では、第1のチャンネル固有のオフセット係数（「 Δx 」）及び第2のチャンネル固有のオフセット係数（「 Δy 」）は、ターゲットクラスターがA、C、及びT塩基固有の強度分布に属するときのみ計算され、ターゲットクラスターがG（暗い）塩基固有の強度分布に属するときは計算されない。配列決定ランが、ペアエンドリードを含む場合の実施態様では、第1のチャンネル固有のオフセット係数（「 Δx 」）及び第2のチャンネル固有のオフセット係数（「 Δy 」）は、第2のリードに対して、第1のリードの最後に利用可能な値で初期化されるが、その後、構成可能な数の配列決定サイクルのそれぞれのセットで更新される。

20

【0247】

実行結果

図12、図13、及び図14は、3つのアプローチ、すなわち、スケーリングのみの解法、オフセットのみの解法（図11で論じられる）、及び最小二乗法300の実行を比較する。3つのアプローチは、IlluminaのシーケンサNextSeq2000からの20個のデータセットにわたってIlluminaのReal-Time Analysis（RTA）ソフトウェアを使用して生成された強度データに適用される。

【0248】

図12では、スケーリングのみの解法（青色）、オフセットのみの解法（オレンジ）、及び最小二乗法（灰色）の実行が、RTAの2チャンネルチャスティティフィルターを通過するクラスターのパーセンテージについて比較プロットされている。比較は、20個のデータセット（x軸として示される）にわたって行われる。3つ全てのアプローチが2チャンネルチャスティティフィルターを通過する>65%のクラスターを達成し、20件中の16件（80%）は、75%より高い通過率を記録し、8/20（又は20%）は>80%の通過率を記録している。最小二乗法300の実行が最良であることに留意されたい。

30

【0249】

図13では、スケーリングのみの解法（青色）、オフセットのみの解法（オレンジ）、及び最小二乗法（灰色）の実行が、低多様性試料における誤差率について比較プロットされている。比較は、既知のファージゲノム（PhiX）をスパイクした20個の低多様性データセット（x軸として示される）にわたって行われる。20件のうちの17件（17/20又は85%）が、<35%の誤差率を達成し、大多数は25%未満の誤差率を享受している。最小二乗法300の実行が最良であることに留意されたい。

40

【0250】

図14では、スケーリングのみの解法（青色）、オフセットのみの解法（オレンジ）、及び最小二乗法（灰色）の実行が、Q30（すなわち、ベースコール誤差<10⁻³（-30/10）又は0.1%）を超える品質スコアを有する配列決定データのパーセンテージについて比較プロットされている。比較は、同じ20個のデータセット（x軸として示される）にわたって行われる。3つ全てのアプローチは、高Q30品質スコア（例えば、80%超）を獲得している一方、20件中の16件（80%）は75%より高い通過率を記

50

録し、8 / 20（又は20%）は>80%の通過率を記録している。最小二乗法300は、他の解法より>2%ポイント優れていることに留意されたい。

【0251】

コンピュータシステム

図15は、開示された技術を実施するために使用することができるコンピュータシステム1500である。コンピュータシステム1500は、バスサブシステム1555を介して多数の周辺デバイスと通信する少なくとも1つの中央処理装置（central processing unit、CPU）1572を含む。これらの周辺デバイスとしては、例えば、メモリデバイス及びファイル記憶サブシステム1536を含む記憶サブシステム1510、ユーザインターフェース入力デバイス1538、ユーザインターフェース出力デバイス1576、並びにネットワークインターフェースサブシステム1574が挙げられ得る。入力デバイス及び出力デバイスは、コンピュータシステム1500とのユーザ対話を可能にする。ネットワークインターフェースサブシステム1574は、他のコンピュータシステム内の対応するインターフェースデバイスへのインターフェースを含む外部ネットワークへのインターフェースを提供する。

10

【0252】

一実施態様では、最小二乗法決定器302は、記憶サブシステム1510及びユーザインターフェース入力デバイス1538に通信可能にリンクされる。

【0253】

ユーザインターフェース入力デバイス1538は、キーボード、マウス、トラックボール、タッチパッド、又はグラフィックスタブレットなどのポインティングデバイス、スキャナ、ディスプレイに組み込まれたタッチスクリーン、音声認識システム及びマイクロフォンなどのオーディオ入力デバイス、並びに他のタイプの入力デバイスを含んでもよい。一般に、用語「入力デバイス」の使用は、コンピュータシステム1500に情報を入力するための全ての可能なタイプのデバイス及び方式を含むことを意図している。

20

【0254】

ユーザインターフェース出力デバイス1576は、ディスプレイサブシステム、プリンタ、ファックス装置、又はオーディオ出力デバイスなどの非視覚ディスプレイを含むことができる。ディスプレイサブシステムは、LEDディスプレイ、陰極線管（Cathode Ray Tube、CRT）、液晶ディスプレイ（Liquid Crystal Display、LCD）などのフラットパネルデバイス、投影デバイス、又は可視画像を作成するための何らかの他の機構を含むことができる。ディスプレイサブシステムはまた、音声出力デバイスなどの非視覚ディスプレイを提供することができる。一般に、用語「出力デバイス」の使用は、コンピュータシステム1500からユーザ又は別のマシン若しくはコンピュータシステムに情報を出力するための、全ての可能なタイプのデバイス及び方式を含むことを意図している。

30

【0255】

記憶サブシステム1510は、本明細書に記載されるモジュール及び方法のうちのいくつか又は全ての機能を提供するプログラミング及びデータ構築物を記憶する。これらのソフトウェアモジュールは、一般にプロセッサ1578によって実行される。

40

【0256】

プロセッサ1578は、グラフィック処理ユニット（graphics processing unit、GPU）、フィールドプログラマブルゲートアレイ（field-programmable gate array、FPGA）、特定用途向け集積回路（application-specific integrated circuit、ASIC）、及び／又は粗粒化再構成可能構造（coarse-grained reconfigurable architecture、CGRA）であり得る。プロセッサ1578は、Google Cloud Platform（商標）、Xilinx（商標）及びCirrascule（商標）などの深層学習クラウドプラットフォームによってホスティングすることができる。プロセッサ1578の例は、GoogleのTensor Processing Unit（TPU）（商標）、GX4 Rackmount Series（商標）、GX1

50

5 Rackmount Series (商標) のようなラックマウントソリューション、NVIDIA DGX-1 (商標)、Microsoft' Stratix V FPGAs (商標)、GraphcoreのIntelligent Processor Unit (IPU) (商標)、Snapdragon processors (商標) を有するQualcommのZeroth Platform (商標)、NVIDIAのVolta (商標)、NVIDIAのDRIVE PX (商標)、NVIDIAのJETSON TX1/TX2 MODULE (商標)、IntelのNirvana (商標)、Movidius VPU (商標)、Fujitsu DPI (商標)、ARMのDynamicIQ (商標)、IBM TrueNorth (商標)、Testa V100s (商標) を有するLambda GPU Server、及び他のものを含む。

10

【0257】

記憶サブシステム1510で使用されるメモリサブシステム1522は、プログラム実行中に命令及びデータを記憶するためのメインランダムアクセスメモリ (random access memory、RAM) 1532と、固定命令が記憶された読み取り専用メモリ (read only memory、ROM) 1534とを含む多数のメモリを含むことができる。ファイル記憶サブシステム1536は、プログラム及びデータファイルのための永続的な記憶装置を提供することができ、これには、ハードディスクドライブ、関連する取り外し可能な媒体を伴うフロッピーディスクドライブ、CD-ROMドライブ、光学ドライブ、又は取り外し可能な媒体カートリッジが挙げられ得る。特定の実施態様の機能を実装するモジュールは、ファイル記憶サブシステム1536によって記憶サブシステム1510内に、又はプロセッサによってアクセス可能な他のマシン内に記憶され得る。

20

【0258】

バスサブシステム1555は、コンピュータシステム1500の様々な構成要素及びサブシステムを、意図されるように互いに通信させるための機構を提供する。バスサブシステム1555は、単一のバスとして概略的に示されているが、バスサブシステムの代替実施態様は、複数のバスを使用することができる。

【0259】

コンピュータシステム1500自体は、パーソナルコンピュータ、ポータブルコンピュータ、ワークステーション、コンピュータ端末、ネットワークコンピュータ、テレビ、メインフレーム、サーバーム、緩くネットワーク化されたコンピュータの広く分散されたセット、又は任意の他のデータ処理システム若しくはユーザデバイスを含む様々なタイプのものであり得る。コンピュータ及びネットワークは絶え間なく変化する性質のものであるため、図15に示されるコンピュータシステム1500の説明は、本発明の好ましい実施態様を例示する目的のための特定の実施例としてのみ意図される。コンピュータシステム1500の多くの他の構成は、図15に示されるコンピュータシステムよりも多くの又は少ない構成要素を有することができる。

30

【0260】

プロセッサ又はモジュールのそれぞれは、特定のプロセスを実行するためのアルゴリズム (例えば、有形及び／又は非一時的コンピュータ可読記憶媒体上に記憶された命令) 又はサブアルゴリズムを含んでもよい。変動補正器232は、モジュールの集合として概念的に例示されるが、専用ハードウェアボード、DSP、プロセッサなどの任意の組み合わせを利用して実装されてもよい。あるいは、変動補正器232は、単一のプロセッサ又は複数のプロセッサを備えた既製のPCを利用して実装されてもよく、機能動作はプロセッサ間に分散される。更なる選択肢として、以下に記載されるモジュールは、特定のモジュール式機能が専用ハードウェアを利用して実施されるハイブリッド構成を利用して実装されてもよく、残りのモジュール式機能は、既製のPCなどを利用して実施される。モジュールはまた、処理ユニット内のソフトウェアモジュールとして実装されてもよい。

40

【0261】

本明細書 (例えば、図9) に記載の方法の様々なプロセス及び工程は、コンピュータを使用して行われ得る。コンピュータは、検出デバイスの一部であるか、コンピュータによ

50

って処理されるデータを取得するために使用される検出デバイスとネットワーク化されているか、又は検出デバイスから分離しているプロセッサを含み得る。いくつかの実施態様では、情報（例えば、画像データ）は、本明細書に開示されるシステムの構成要素間で直接又はコンピュータネットワークを介して送信され得る。ローカルエリアネットワーク（LAN）又はワイドエリアネットワーク（WAN）は、システムを含むコンピュータ及びコンピューティングデバイスが接続されている、インターネットへのアクセスを含む企業コンピューティングネットワークであり得る。一実施態様では、LANは、伝送制御プロトコル／インターネットプロトコル（TCP／IP）業界標準に準拠する。場合によっては、情報（例えば、画像データ）は、入力デバイス（例えば、ディスクドライブ、コンパクトディスクプレーヤ、USBポートなど）を介して本明細書に開示されるシステムに入力される。場合によっては、情報は、例えば、ディスク又はフラッシュドライブなどの記憶デバイスから情報をロードすることによって受信される。

10

【0262】

本明細書に記載されるアルゴリズム又は他のプロセスを実行するために使用されるプロセッサは、マイクロプロセッサを含み得る。マイクロプロセッサは、Intel Corporation製のPentium（商標）プロセッサなどの任意の従来の汎用シングルチップ又はマルチチップマイクロプロセッサであり得る。特に有用なコンピュータは、128GBのRAM、及び2TBのソリッドステートディスクドライブを有する、Intel Ivybridge dual-12コアプロセッサ、LSI RAIDコントローラを利用し得る。更に、プロセッサは、デジタル信号プロセッサ又はグラフィックスプロセッサなどの任意の従来の専用プロセッサを含み得る。プロセッサは、典型的には、従来のアドレスライン、従来のデータライン、及び1つ以上の従来の制御ラインを有する。

20

【0263】

本明細書に開示される実施態様は、ソフトウェア、ファームウェア、ハードウェア、又はそれらの任意の組み合わせを生成するための標準的なプログラミング技術又は工学技術を使用して、方法（例えば、図2及び図7）、装置、システム、又は物品として実装されてもよい。本明細書で使用する時、用語「物品」は、光学記憶デバイスなどのハードウェア又はコンピュータ可読媒体、並びに揮発性又は不揮発性メモリデバイス内に実装されるコード又はロジックを指す。そのようなハードウェアとしては、フィールドプログラマブルゲートアレイ（FPGA）、特定用途向け集積回路（ASIC）、複合プログラマブル論理デバイス（Complex Programmable Logic Device、CPLD）、プログラマブルロジックアレイ（Programmable Logic Array、PLA）、マイクロプロセッサ、又は他の同様の処理装置が挙げられ得るが、これらに限定されない。特定の実施態様では、本明細書に記載される情報又はアルゴリズムは、非一過性記憶媒体中に存在する。

30

【0264】

特定の実施態様では、本明細書に記載される（例えば、（図11で論じられる））コンピュータ実装の方法は、物体の複数の画像が取得されている間に、リアルタイムで行われ得る。このようなリアルタイム分析は、核酸配列が流体及び検出工程の繰り返しサイクルに供される核酸配列決定用途に特に有用である。配列決定データの分析は、多くの場合、本明細書に記載される方法をリアルタイム又は背景で実行するのに有益であり得る一方で、他のデータ収集又は分析アルゴリズムがプロセス中である間に、本明細書に記載される方法を実行することが有益であり得る。本方法で使用する事ができるリアルタイム分析法の例は、Illumina, Inc (San Diego, Calif) から市販されており、及び／又は、参照により本明細書に組み込まれる米国特許出願公開第2012/0020537 (A1) 号に記載されているMiSeq及びHiSeq配列決定機器に使用されるものである。

40

【0265】

本出願において、用語「クラスター」、「ウェル」、「試料」、「検体」、及び「蛍光試料」は、ウェルが対応するクラスター／試料／検体／蛍光試料を含有するので、交換可能に使用される。本明細書中で定義されるように、「試料」及びその誘導体は、最も広い

50

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。いくつかの実施態様において、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つ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を含む試料中の汚染細菌DNAの存在に由来し得ることも想定される。いくつかの実施態様において、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

【0266】

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。別の実施態様では、低分子量物質は、酵素的又は機械的にフラグメント化されたDNAを含む。試料は、無細胞循環DNAを含むことができる。いくつかの実施態様において、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み得る。いくつかの実施態様において、試料は、疫学的、農業的、法医学又は病原性試料であり得る。いくつかの実施態様において、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。別の実施態様では、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。いくつかの実施態様において、核酸分子の供給源は、保存又は絶滅した試料又は種であり得る。

20

【0267】

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／又はフラグメント化されたゲノム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。一実施態様では、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以上の軍隊若しくはそのような要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。それ自体、いくつかの実施態様において、核酸試料は、ゲノムDNAのようなDNAの少量又はフラグメント化された部分を含み得る。いくつかの実施態様において、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。いくつかの実施態様において、標的配列は、毛髪、皮膚、組織試料、剖検又は犠牲者の遺体から得ることができる。いくつかの実施態様において、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。いくつかの実施態様において、標的配列は、微生物、植物細胞又は昆虫学的のような非ヒトから得られた核酸を含むことができる。いくつかの実施態様において、標的配列又は増幅された標的配列は、ヒト同定を対象とする。いくつかの実施態様において、本開示は、一般に、法医学試料の特徴を同定するための方法に関する。いくつかの実施態様において、本開示は、一般に、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。一実施態様において、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

【0268】

50

開示された技術は、画像データのクラスター間強度プロファイルの変動を補正するための変動補正係数を生成する。開示された技術は、システム、方法、又は製品として実施することができる。実施態様の1つ以上の特徴を、塩基実施態様と組み合わせることができる。相互に排他的でない実施態様は、組み合わせ可能であると教示されている。実施態様の1つ以上の特徴を他の実施態様と組み合わせることができる。本開示は、これらのオプションのユーザを定期的に通知する。これらの選択肢を繰り返す列挙のいくつかの実施態様からの省略は、前述のセクションで教示されている組み合わせを制限するものとして解釈されるべきではない。これらの記載は、以下の実施態様の各々に参照することにより本明細書に組み込まれる。

【0269】

このセクションで説明される方法の他の実施態様は、上述の方法のいずれかを実行するためにプロセッサによって実行可能な命令を記憶する非一時的コンピュータ可読記憶媒体を含むことができる。このセクションで説明される方法の更に別の実施態様は、メモリと、メモリ内に記憶された命令を実行して上記の方法のいずれかを実行するように動作可能な1つ以上のプロセッサとを含むシステムを含むことができる。

【0270】

別の実施態様では、変動補正は、非強度データ、例えば、分子伸長中の水素イオンの放出によって誘発されるpH変化に対して実行される。pH変化は検出され、組み込まれた塩基の数に比例する電圧変化に変換される（例えば、Ion Torrentの場合）。

【0271】

更に別の実施態様では、非強度データは、塩基の同一性を決定すると同時に、分析物がナノ細孔を通過する際、又はその開口部付近を通過する際に電流の破壊を測定するためにバイオセンサを使用するナノ細孔検知から構築される。例えば、Oxford Nanopore Technologies (ONT) 配列決定は、以下の概念に基づく：ナノ細孔を介して膜を介してDNA（又はRNA）の単一鎖を通過させ、膜にわたって電位差を印加する。細孔内に存在するヌクレオチドは、細孔の電気抵抗に影響を及ぼし、そのため、経時的な電流測定は、細孔を通過するDNA塩基の配列を示すことができる。この電流信号（プロットされたときにその外観に起因する「押しつぶし」）は、ONTシーケンサによって収集された生データである。これらの測定値は、4 kHz周波数（例えば）で取られた16ビットの整数データ取得（Data Acquisition、DAC）値として記憶される。1秒当たり～450塩基対のDNA鎖速度を用いて、これは、平均して、塩基当たり約9つの生観察を与える。次いで、この信号を処理して、個々の読み取りに対応する開孔信号の破断を識別する。これらの生信号の伸長は、ベースコールされ、DAC値をDNA塩基の配列に変換するプロセスである。いくつかの実施態様では、非強度データは、正規化又はスケーリングされたDAC値を含む。

【0272】

開示される技術、又はその要素の1つ以上の実施態様は、示された方法ステップを実行するためのコンピュータ使用可能なプログラムコードを備えた非一時的コンピュータ可読記憶媒体を含むコンピュータ製品の形態で実装することができる。更に、開示される技術、又はその要素の1つ以上の実施態様は、メモリと、メモリに結合され、例示的な方法ステップを実行するように動作する少なくとも1つのプロセッサと、を含む装置の形態で実装することができる。更に、別の態様では、開示される技術又はその要素の1つ以上の実施態様は、本明細書に記載の方法ステップのうちの1つ以上を実行するための手段の形態で実装ことができ、この手段は、(i) ハードウェアモジュール、(ii) 1つ以上のハードウェアプロセッサ上で実行されるソフトウェアモジュール、又は(iii) ハードウェア及びソフトウェアモジュールの組み合わせ、を含むことができ、(i)～(iii)のいずれかが、本明細書に記載の特定の技術を実施し、ソフトウェアモジュールは、コンピュータ可読記憶媒体（又は複数のそのような媒体）に記憶される。

【0273】

本出願は、「累積強度補正パラメータ（複数可）」及び「中間項（複数可）」という用

10

20

30

40

50

語を互換的に使用する。

【0274】

本出願は、「純粋な強度（複数可）」及び「分布強度（複数可）」という用語を互換的に使用する。

【0275】

本出願は、「強度プロファイル（複数可）」及び「コンステレーション（複数可）」という用語を互換的に使用する。

【0276】

本出願は、「変動補正係数（複数可）」及び「強度補正係数（複数可）」という用語を互換的に使用する。

【0277】

本出願は、「増幅係数」及び「スケール因子」という用語を互換的に使用する。

【0278】

本出願は、「オフセット係数（複数可）」及び「オフセット（複数可）」という用語を互換的に使用する。

【0279】

本出願は、「ターゲットクラスター」及び「特定のクラスター」という用語を互換的に使用する。

【0280】

本出願は、「次の」、「後続の」、及び「連続する」という用語を互換的に使用する。

【0281】

本出願は、「ヌクレオチド（複数可）」及び「塩基（複数可）」という用語を互換的に使用する。

【0282】

本出願は、「累積強度補正パラメータ決定器」及び「アキュムレータ」という用語を互換的に使用する。

【0283】

本出願は、「累積強度補正パラメータ決定器」及び「アキュムレータ」という用語を互換的に使用する。

【0284】

項目

1. ターゲットクラスターをベースコールするコンピュータ実装方法であって、方法は、

ターゲットクラスターについて、

配列決定ランの現在の配列決定サイクルに登録された現在のチャンネル固有の強度を、現在の配列決定サイクルでターゲットクラスターがベースコールされる塩基固有の強度分布から読み取ることと、

塩基固有の強度分布の重心から現在のチャンネル固有の分布強度を読み取ることと、

現在のチャンネル固有の強度及び現在のチャンネル固有の分布強度に基づいて、現在の配列決定サイクルに対する現在の強度補正パラメータのセットを決定することと、

配列決定ランの先行する配列決定サイクルについての現在の強度補正パラメータのセット及び先行する累積強度補正パラメータのセットを累積することによって、現在の配列決定サイクルについての現在の累積強度補正パラメータのセットを決定することと、

現在の累積強度補正パラメータのセットに基づいて、現在の配列決定サイクルに対する現在の増幅係数及び現在のチャンネル固有のオフセット係数を決定することと、

現在の増幅係数及び現在のチャンネル固有のオフセット係数を使用して、配列決定ランの次の配列決定サイクルに登録された次のチャンネル固有の強度を補正し、次の配列決定サイクルに対して補正された次のチャンネル固有の強度を生成することと、

補正された次のチャンネル固有の強度に基づいて、次の配列決定サイクルでターゲットクラスターをベースコールすることと、を含む、コンピュータ実装方法。

10

20

30

40

50

【0285】

2. 現在の強度補正パラメータのセットは、現在のチャンネル固有の分布強度と、現在のチャンネル固有の強度誤差と、現在の分布重心から原点の距離と、現在の分布強度対強度誤差の類似性測定と、を含む、項目1に記載のコンピュータ実装方法。

【0286】

3. 現在のチャンネル固有の強度誤差は、現在のチャンネル固有の強度と現在のチャンネル固有の分布強度との間のチャンネル単位の差である、項目2に記載のコンピュータ実装方法。

【0287】

4. 現在の分布重心から原点の距離は、重心と、塩基固有の強度分布を含む多次元空間の原点との間のユークリッド距離である、項目2に記載のコンピュータ実装方法。

10

【0288】

5. 多次元空間は、デカルト空間、極空間、円筒空間、及び球状空間のうちの少なくとも1つである、項目4に記載のコンピュータ実装方法。

【0289】

6. 現在の分布強度対強度誤差の類似性測定は、現在のチャンネル固有の分布強度と現在のチャンネル固有の強度誤差との間のチャンネル単位のドット積の総和である、項目2に記載のコンピュータ実装方法。

【0290】

7. 現在の累積強度補正パラメータのセットは、現在の強度補正パラメータのセット内の現在の強度補正パラメータと、先行する累積強度補正パラメータのセット内の先行する累積強度補正パラメータとの強度補正パラメータ単位の和である、項目1に記載のコンピュータ実装方法。

20

【0291】

8. 現在の累積強度補正パラメータのセットは、現在の強度補正パラメータ及び先行する累積強度補正パラメータの強度補正パラメータ単位の平均である、項目7に記載のコンピュータ実装方法。

【0292】

9. 先行する累積強度補正パラメータのセット及び現在の累積強度補正パラメータのセットは、量子化された固定ビット幅フォーマットで記憶される、項目1に記載のコンピュータ実装方法。

30

【0293】

10. 現在のチャンネル固有のオフセット係数は、同一であるように構成されている、項目1に記載のコンピュータ実装方法。

【0294】

11. 現在の累積強度補正パラメータのセット内の現在の累積強度補正パラメータは、現在のチャンネル固有の分布強度についての第1の共通の現在の累積強度補正パラメータと、現在のチャンネル固有の強度誤差についての第2の共通の現在の累積強度補正パラメータと、を含む、項目10に記載のコンピュータ実装方法。

【0295】

12. 現在のチャンネル固有のオフセット係数は、次のチャンネル固有のシフト強度を生成するために次のチャンネル固有の強度からチャンネル単位で減算され、次のチャンネル固有のシフト強度は、補正された次のチャンネル固有の強度を生成するために現在の増幅係数で除算される、項目1に記載のコンピュータ実装方法。

40

【0296】

13. 重み付け関数を使用して、初期増幅係数を現在の増幅係数と組み合わせ、初期のチャンネル固有のオフセット係数を現在のチャンネル固有のオフセット係数と組み合わせ、現在の配列決定サイクルに対する重み付けされた現在の増幅係数及び重み付けされた現在のチャンネル固有のオフセット係数を生成することを更に含む、項目1に記載のコンピュータ実装方法。

【0297】

50

14. 重み付け関数は、最小重み (w_{min}) を初期増幅係数及び初期のチャンネル固有のオフセット係数に適用し、最大重み (w_{max}) を現在の増幅係数及び現在のチャンネル固有のオフセット係数に適用し、 $w_{min} = (1 - w_{max})$ である、項目13に記載のコンピュータ実装方法。

【0298】

15. 最大重み (w_{max}) は、 $(c - p) / c$ として定義され、 c は現在の配列決定サイクルのインデックスであり、 p は2～7の数値である、項目14に記載のコンピュータ実装方法。

【0299】

16. 重み付けされた現在の増幅係数及び重み付けされた現在のチャンネル固有のオフセット係数を使用して、次のチャンネル固有の強度を補正し、補正された次のチャンネル固有の強度を生成することを更に含む、項目15に記載のコンピュータ実装方法。

【0300】

17.

最尤法を使用して、現在の配列決定サイクルについて、現在の増幅係数及び現在のチャンネル固有のオフセット係数のそれぞれの現在の最尤重みを生成することと、

現在の最尤重みを現在の増幅係数及び現在のチャンネル固有のオフセット係数にそれぞれ適用して、現在の配列決定サイクルについて最尤重みが付けられた現在の増幅係数及び最尤重みが付けられた現在のチャンネル固有のオフセット係数を生成することと、

最尤重みが付けられた現在の増幅係数及び最尤重みが付けられた現在のチャンネル固有のオフセット係数を使用して、次のチャンネル固有の強度を補正し、補正された次のチャンネル固有の強度を生成することと、を更に含む、項目1に記載のコンピュータ実装方法。

【0301】

18.

減衰因子を現在の強度補正パラメータに適用して、現在の配列決定サイクルについて減衰された現在の強度補正パラメータを生成することと、

減衰された現在の強度補正パラメータと先行する累積強度補正パラメータとを強度補正パラメータ単位で累積することによって、現在の累積強度補正パラメータを決定することと、を更に含む、項目1に記載のコンピュータ実装方法。

【0302】

19. 減衰因子は、配列決定ランの特定の数の配列決定サイクルに対して固定されており、その後、減衰ロジックに基づいて指数関数的に減衰される、項目18に記載のコンピュータ実装方法。

【0303】

20. 減衰ロジックは $1 - 1 / \tau$ であり、 τ は事前定義された数値である、項目19に記載のコンピュータ実装方法。

【0304】

21. 配列決定ランの連続する配列決定サイクルでターゲットクラスターについて読み取ること、読み取ること、決定すること、決定すること、決定すること、使用すること、及びベースコールすることを繰り返すことを更に含む、項目1に記載のコンピュータ実装方法。

【0305】

22. 複数のクラスターについて読み取ること、読み取ること、決定すること、決定すること、決定すること、使用すること、及びベースコールすることを並行して実行することを更に含む、項目1に記載のコンピュータ実装方法。

【0306】

23. 現在の強度補正パラメータのセット、現在の累積強度補正パラメータのセット、現在の増幅係数、及び現在のチャンネル固有のオフセット係数の閉形式の式は、最小二乗法を使用して決定される、項目1に記載のコンピュータ実装方法。

【0307】

10

20

30

40

50

24. 現在のチャンネル固有の強度は、強度チャンネルにそれぞれ対応する、項目1に記載のコンピュータ実装方法。

【0308】

25. 現在のチャンネル固有のオフセット係数は、現在のチャンネル固有の強度と現在のチャンネル固有の分布強度との間のチャンネル単位の差である、項目1に記載のコンピュータ実装方法。

【0309】

26. ターゲットクラスターをベースコールするためにコンピュータプログラム命令が記憶された非一時的コンピュータ可読記憶媒体であって、命令は、プロセッサ上で実行されると、

10

ターゲットクラスターについて、

配列決定ランの現在の配列決定サイクルに登録された現在のチャンネル固有の強度を、現在の配列決定サイクルでターゲットクラスターがベースコールされる塩基固有の強度分布から読み取ることと、

塩基固有の強度分布の重心から現在のチャンネル固有の分布強度を読み取ることと、

現在のチャンネル固有の強度及び現在のチャンネル固有の分布強度に基づいて、現在の配列決定サイクルに対する現在の強度補正パラメータのセットを決定することと、

配列決定ランの先行する配列決定サイクルについての現在の強度補正パラメータのセット及び先行する累積強度補正パラメータのセットを累積することによって、現在の配列決定サイクルについての現在の累積強度補正パラメータのセットを決定することと、

20

現在の累積強度補正パラメータのセットに基づいて、現在の配列決定サイクルに対する現在の増幅係数及び現在のチャンネル固有のオフセット係数を決定することと、

現在の増幅係数及び現在のチャンネル固有のオフセット係数を使用して、配列決定ランの次の配列決定サイクルに登録された次のチャンネル固有の強度を補正し、次の配列決定サイクルに対して補正された次のチャンネル固有の強度を生成することと、

補正された次のチャンネル固有の強度に基づいて、次の配列決定サイクルでターゲットクラスターをベースコールすることと、を含む方法を実装する、非一時的コンピュータ可読記憶媒体。

【0310】

27. 最終的に項目1に従属するそれぞれの項目を実装する、項目26に記載の非一時的コンピュータ可読記憶媒体。

30

【0311】

28. メモリに結合された1つ以上のプロセッサを含むシステムであって、メモリは、ターゲットクラスターをベースコールするためのコンピュータ命令がロードされ、命令は、プロセッサ上で実行されると、

ターゲットクラスターについて、

配列決定ランの現在の配列決定サイクルに登録された現在のチャンネル固有の強度を、現在の配列決定サイクルでターゲットクラスターがベースコールされる塩基固有の強度分布から読み取ることと、

塩基固有の強度分布の重心から現在のチャンネル固有の分布強度を読み取ることと、

40

現在のチャンネル固有の強度及び現在のチャンネル固有の分布強度に基づいて、現在の配列決定サイクルに対する現在の強度補正パラメータのセットを決定することと、

配列決定ランの先行する配列決定サイクルについての現在の強度補正パラメータのセット及び先行する累積強度補正パラメータのセットを累積することによって、現在の配列決定サイクルについての現在の累積強度補正パラメータのセットを決定することと、

現在の累積強度補正パラメータのセットに基づいて、現在の配列決定サイクルに対する現在の増幅係数及び現在のチャンネル固有のオフセット係数を決定することと、

現在の増幅係数及び現在のチャンネル固有のオフセット係数を使用して、配列決定ランの次の配列決定サイクルに登録された次のチャンネル固有の強度を補正し、次の配列決定サイクルに対して補正された次のチャンネル固有の強度を生成することと、

50

補正された次のチャネル固有の強度に基づいて、次の配列決定サイクルでターゲットクラスターをベースコールすることと、を含むアクションを実装する、システム。

【0312】

29. 最終的に項目1に従属するそれぞれの項目を実装する、項目28に記載のシステム。

【0313】

30. ターゲットクラスターをベースコールするコンピュータ実装方法であって、方法は、

ターゲットクラスターについて、

現在の強度データ及び履歴強度データにアクセスすることであって、

10

現在の強度データは、配列決定ランの現在の配列決定サイクルに対するものであり、

履歴強度データは、配列決定ランの1つ以上の先行する配列決定サイクルに対するものである、ことと、

現在の強度データ及び履歴強度データに基づいて、スケール補正係数及びチャネル固有のシフト補正係数を決定することと、

スケール補正係数及びチャネル固有のシフト補正係数を使用して、次の強度データを補正し、補正された次の強度データを生成することであって、

次の強度データは、配列決定ランの次の配列決定サイクルに対するものである、ことと

、
補正された次の強度データに基づいて、次の配列決定サイクルでターゲットクラスターをベースコールすることと、を含む、コンピュータ実装方法。

20

【0314】

31. 最終的に項目1に従属するそれぞれの項目を実装する、項目30に記載のコンピュータ実装方法。

【0315】

32. ターゲットクラスターをベースコールするためにコンピュータプログラム命令が記憶された非一時的コンピュータ可読記憶媒体であって、命令は、プロセッサ上で実行されると、

ターゲットクラスターについて、

現在の強度データ及び履歴強度データにアクセスすることであって、

30

現在の強度データは、配列決定ランの現在の配列決定サイクルに対するものであり、

履歴強度データは、配列決定ランの1つ以上の先行する配列決定サイクルに対するものである、ことと、

現在の強度データ及び履歴強度データに基づいて、スケール補正係数及びチャネル固有のシフト補正係数を決定することと、

スケール補正係数及びチャネル固有のシフト補正係数を使用して、次の強度データを補正し、補正された次の強度データを生成することであって、

次の強度データは、配列決定ランの次の配列決定サイクルに対するものである、ことと

、
補正された次の強度データに基づいて、次の配列決定サイクルでターゲットクラスターをベースコールすることと、を含む方法を実装する、非一時的コンピュータ可読記憶媒体。

40

【0316】

33. 最終的に項目1に従属するそれぞれの項目を実装する、項目32に記載の非一時的コンピュータ可読記憶媒体。

【0317】

34. メモリに結合された1つ以上のプロセッサを含むシステムであって、メモリは、ターゲットクラスターをベースコールするためのコンピュータ命令がロードされ、命令は、プロセッサ上で実行されると、

ターゲットクラスターについて、

50

現在の強度データ及び履歴強度データにアクセスすることであって、
現在の強度データは、配列決定ランの現在の配列決定サイクルに対するものであり、
履歴強度データは、配列決定ランの1つ以上の先行する配列決定サイクルに対するものである、ことと、

現在の強度データ及び履歴強度データに基づいて、スケール補正係数及びチャンネル固有のシフト補正係数を決定することと、

スケール補正係数及びチャンネル固有のシフト補正係数を使用して、次の強度データを補正し、補正された次の強度データを生成することであって、

次の強度データは、配列決定ランの次の配列決定サイクルに対するものである、ことと、

補正された次の強度データに基づいて、次の配列決定サイクルでターゲットクラスターをベースコールすることと、を含むアクションを実装する、システム。

【0318】

35. 最終的に項目1に従属するそれぞれの項目を実装する、項目34に記載のシステム。

【0319】

36. ターゲットクラスターをベースコールするコンピュータ実装方法であって、方法は、

ターゲットクラスターについて、

現在の強度データ及び履歴強度データにアクセスすることであって、

現在の強度データは、配列決定ランの現在の配列決定サイクルに対するものであり、

履歴強度データは、配列決定ランの1つ以上の先行する配列決定サイクルに対するものである、ことと、

現在の強度データ及び履歴強度データを使用して、次の強度データを補正し、補正された次の強度データを生成することであって、

次の強度データは、配列決定ランの次の配列決定サイクルに対するものである、ことと、

補正された次の強度データに基づいて、次の配列決定サイクルでターゲットクラスターをベースコールすることと、を含む、コンピュータ実装方法。

【0320】

37. 最終的に項目1に従属するそれぞれの項目を実装する、項目36に記載のコンピュータ実装方法。

【0321】

38. ターゲットクラスターをベースコールするためにコンピュータプログラム命令が記憶された非一時的コンピュータ可読記憶媒体であって、命令は、プロセッサ上で実行されると、

ターゲットクラスターについて、

現在の強度データ及び履歴強度データにアクセスすることであって、

現在の強度データは、配列決定ランの現在の配列決定サイクルに対するものであり、

履歴強度データは、配列決定ランの1つ以上の先行する配列決定サイクルに対するものである、ことと、

現在の強度データ及び履歴強度データを使用して、次の強度データを補正し、補正された次の強度データを生成することであって、

次の強度データは、配列決定ランの次の配列決定サイクルに対するものである、ことと、

補正された次の強度データに基づいて、次の配列決定サイクルでターゲットクラスターをベースコールすることと、を含む方法を実装する、非一時的コンピュータ可読記憶媒体。

【0322】

39. 最終的に項目1に従属するそれぞれの項目を実装する、項目38に記載の非一時

10

20

30

40

50

的コンピュータ可読記憶媒体。

【0323】

40. メモリに結合された1つ以上のプロセッサを含むシステムであって、メモリは、ターゲットクラスターをベースコールするためのコンピュータ命令がロードされ、命令は、プロセッサ上で実行されると、

ターゲットクラスターについて、

現在の強度データ及び履歴強度データにアクセスすることであって、

現在の強度データは、配列決定ランの現在の配列決定サイクルに対するものであり、

履歴強度データは、配列決定ランの1つ以上の先行する配列決定サイクルに対するものである、ことと、

現在の強度データ及び履歴強度データを使用して、次の強度データを補正し、補正された次の強度データを生成することであって、

次の強度データは、配列決定ランの次の配列決定サイクルに対するものである、ことと

、
補正された次の強度データに基づいて、次の配列決定サイクルでターゲットクラスターをベースコールすることと、を含む、アクションを実装する、システム。

【0324】

41. 最終的に項目1に従属するそれぞれの項目を実装する、項目40に記載のシステム。

【0325】

本発明は、上述の好ましい実施形態及び実施例を参照して開示されているが、これらの実施例は、限定的な意味ではなく例示的な意味で意図されていることが理解されるべきである。当業者であれば、変更及び組み合わせが容易に生じ、その変更及び組み合わせは、本発明の趣旨及び以下の特許請求の範囲の範囲内にあると考えられる。

【符号の説明】

【0326】

1 クラスター

2 クラスター

3 クラスター

302 最小二乗法決定器

312 強度モデラー

322 最小化器

1500 コンピュータシステム

1510 記憶サブシステム

1522 メモリサブシステム

1532 RAM

1534 ROM

1536 ファイル記憶サブシステム

1538 ユーザインターフェース入力デバイス

1555 バスサブシステム

1572 CPU

1574 ネットワークインターフェースサブシステム

1576 ユーザインターフェース出力デバイス

1578 プロセッサ

10

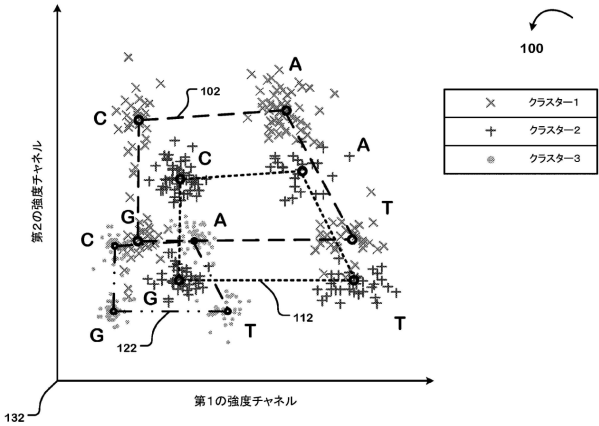
20

30

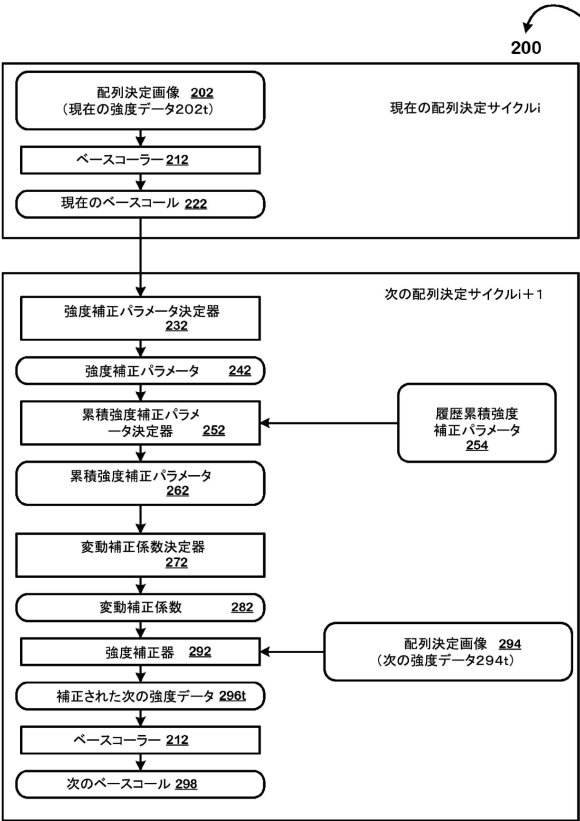
40

50

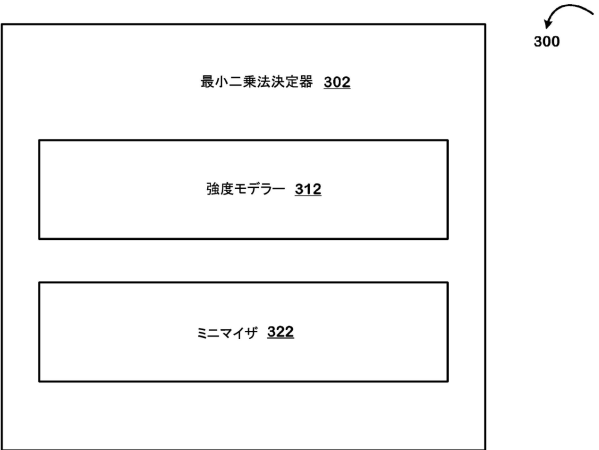
【図面】
【図 1】



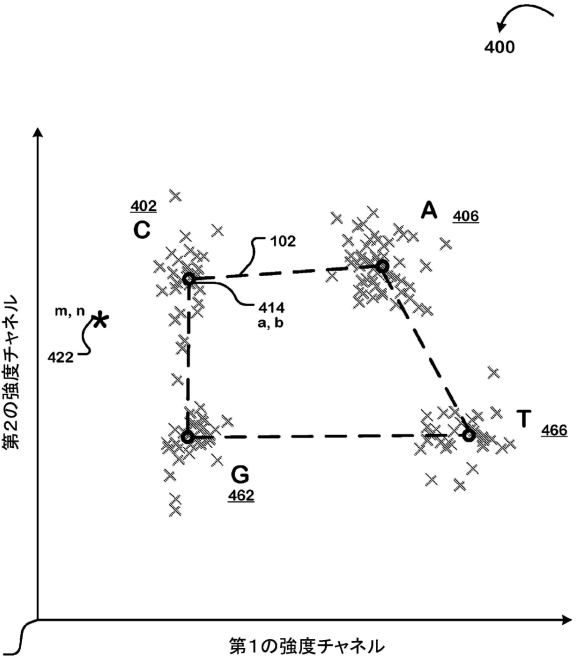
【図 2】



【図 3】



【図 4】



10

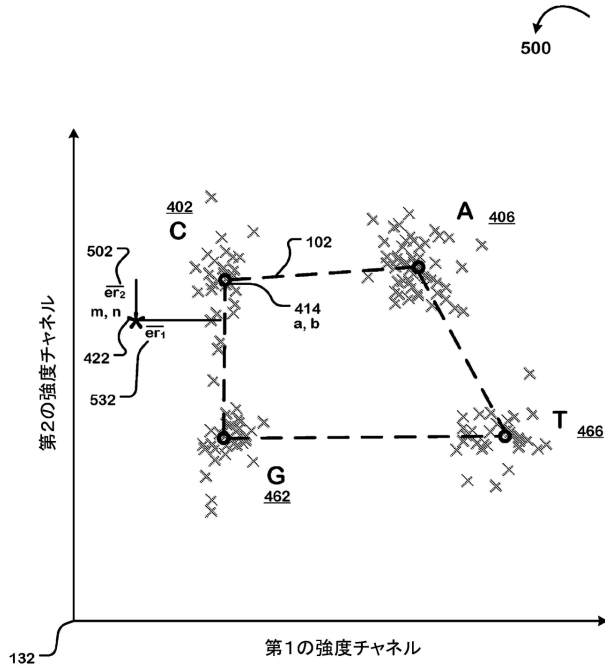
20

30

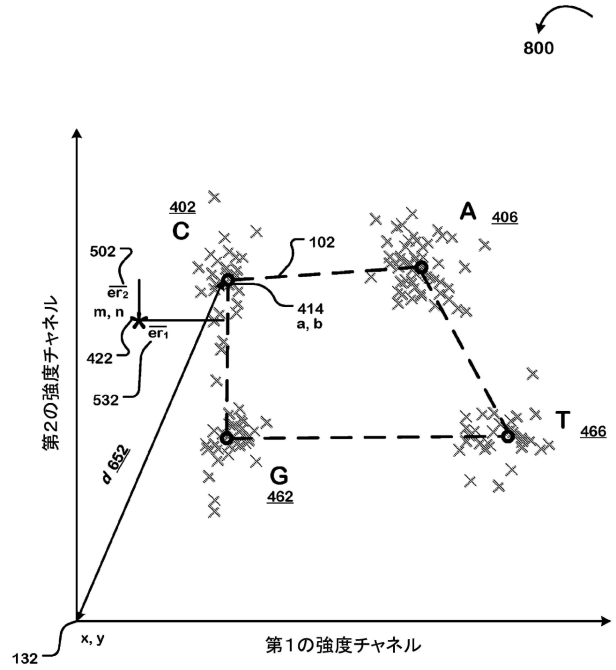
40

50

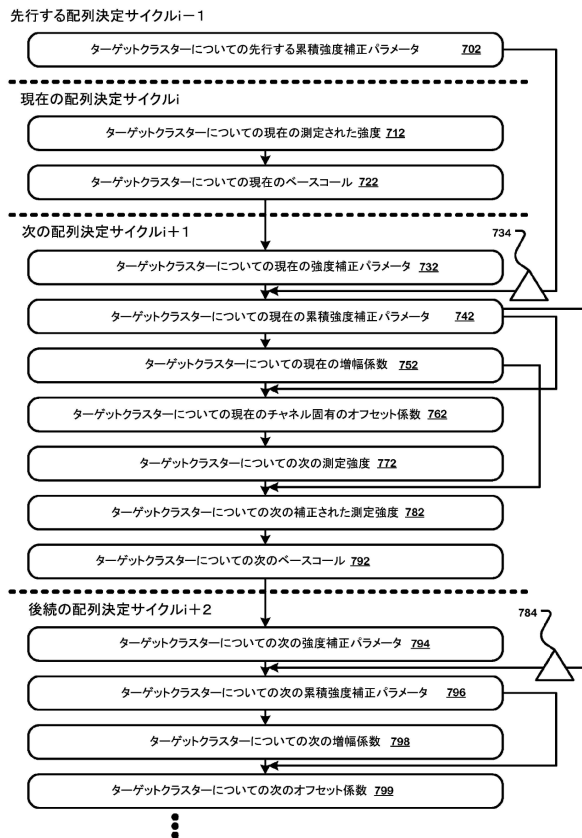
【図 5】



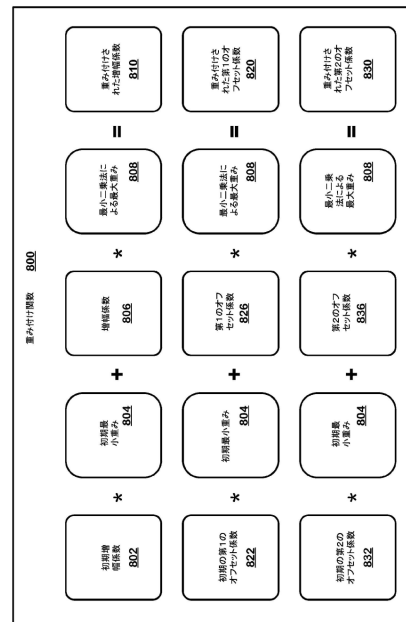
【図 6】



【図 7】



【図 8】



10

20

30

40

50

【図 9】

α, d_1, d_2 及び n_{e_i} が、最尤法を見つけることができる事前確率分布 $\alpha \sim N(1, \sigma_\alpha^2), d_1 \sim N(0, \sigma_{d_1}^2), d_2 \sim N(0, \sigma_{d_2}^2), n_{e_i} \sim N(0, \sigma_{n_{e_i}}^2)$ を有する場合、 $\theta_{e_i,1} = y_{e_i,1} - x_{e_i,1}$ とする。

$$\bar{x}_1 \triangleq \frac{1}{C} \sum_{c=1}^C x_{c,1}, \quad \bar{x}_2 \triangleq \frac{1}{C} \sum_{c=1}^C x_{c,2}, \quad \overline{xx} \triangleq \frac{1}{C} \sum_{c=1}^C (x_{c,1}^2 + x_{c,2}^2) \quad 902$$

$$\bar{\theta}_1 \triangleq \frac{1}{C} \sum_{c=1}^C \theta_{e1,1}, \quad \bar{\theta}_2 \triangleq \frac{1}{C} \sum_{c=1}^C \theta_{e2,2}, \quad \overline{x\theta} \triangleq \frac{1}{C} \sum_{c=1}^C (x_{c,1}\theta_{e1,1} + x_{c,2}\theta_{e2,2}) \quad 904$$

$$k_a(C) = \frac{\sigma_n^2}{C\sigma_a^2}, \quad k_{d_1}(C) = \frac{C\sigma_{d_1}^2}{C\sigma_{d_1}^2 + \sigma_n^2}, \quad k_{d_2}(C) = \frac{C\sigma_{d_2}^2}{C\sigma_{d_2}^2 + \sigma_n^2} \quad 906, 908, 910$$

$$\hat{a} = 1 + \frac{\overline{x\theta} - \bar{x}_1\bar{\theta}_1k_{d_1}(C) - \bar{x}_2\bar{\theta}_2k_{d_2}(C)}{\overline{xx} + k_a(C) - \bar{x}_1^2k_{d_1}(C) - \bar{x}_2^2k_{d_2}(C)} \quad 920$$

$$\hat{d}_1 = (\bar{\theta}_1 + \bar{x}_1(1 - \hat{a}))k_{d_1}(C) \quad 930$$

$$\hat{d}_2 = (\bar{\theta}_2 + \bar{x}_2(1 - \hat{a}))k_{d_2}(C) \quad 940$$

C → ∞ のとき、これは最小二乗法に近づく。

【図 10】

```

1000
1002
1004
1006
1008
1010
1012
1014
1016
1018
1020

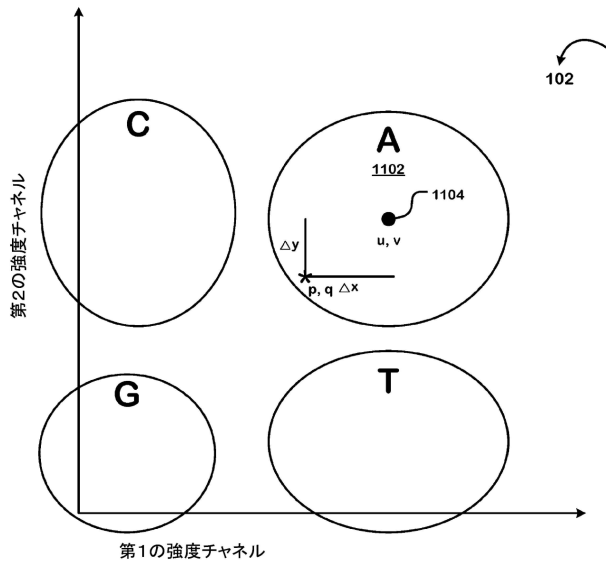
if stats.cycle < 1000;
    stats.cycle = stats.cycle + 1;
    decay_factor = 1.0;
else
    decay_factor = 1.0 / 1000;
end

stats.xx_sum = decay_factor * stats.xx_sum + channel1(x).^2 + channel2(x).^2;
stats.xy_sum = decay_factor * stats.xy_sum + channel1(x) * channel2(x);
stats.x2_sum = decay_factor * stats.x2_sum + channel2(x).^2;
stats.xe_sum = decay_factor * stats.xe_sum + channel1(x).^2 + channel2(x).^2;
stats.e1_sum = decay_factor * stats.e1_sum + channel1(e);
stats.e2_sum = decay_factor * stats.e2_sum + channel2(e);
    
```

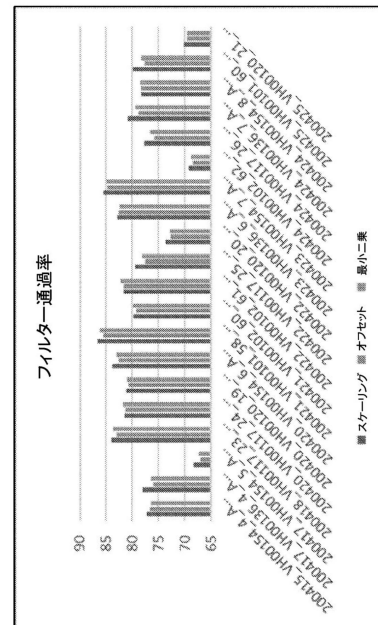
10

20

【図 11】



【図 12】

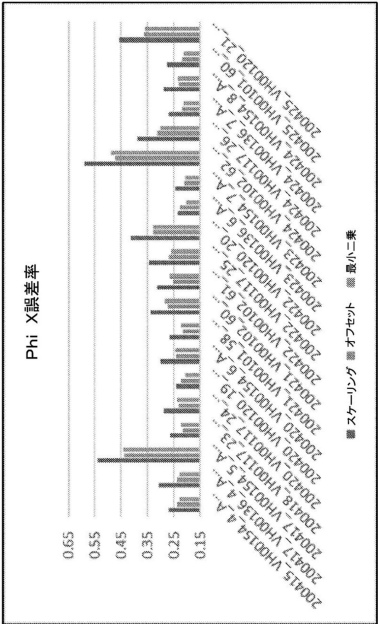


30

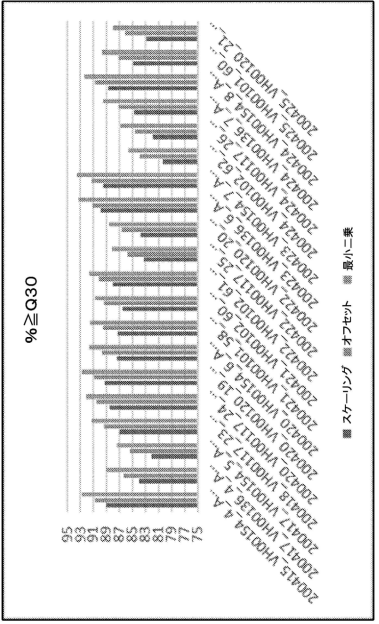
40

50

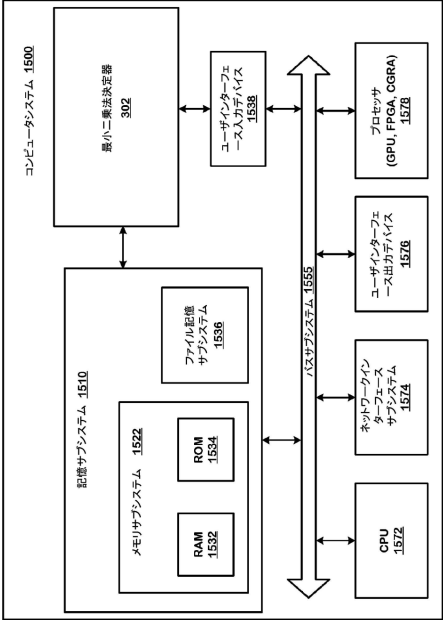
【図 1 3】



【図 1 4】



【図 1 5】



10

20

30

40

50

【国際調査報告】

INTERNATIONAL SEARCH REPORT

International application No
PCT/US2021/056711

A. CLASSIFICATION OF SUBJECT MATTER
INV. G16B40/10 G16B40/30
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.
X	WEI-CHUN KAO ET AL: "naiveBayesCall: An Efficient Model-Based Base-Calling Algorithm for High-Throughput Sequencing", 25 April 2010 (2010-04-25), RESEARCH IN COMPUTATIONAL MOLECULAR BIOLOGY, SPRINGER BERLIN HEIDELBERG, BERLIN, HEIDELBERG, PAGE(S) 233 - 247, XP019141683, ISBN: 978-3-642-12682-6 the whole document in particular sections 2 and 3 ----- -/--	1-30

☒ 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

Date of mailing of the international search report

4 February 2022

18/02/2022

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

Rákossy, Z

Form PCT/ISA/210 (second sheet) (April 2005)

INTERNATIONAL SEARCH REPORT

International application No
PCT/US2021/056711

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	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 the whole document in particular "methods" section -----	1-30
X	US 2018/195953 A1 (LANGLOIS ROBERT [US] ET AL) 12 July 2018 (2018-07-12) the whole document in particular paragraph [0070] - paragraph [0089]; figures 6, 9, 10 -----	1-30
X	TIM MASSINGHAM ET AL: "All your base: a fast and accurate probabilistic approach to base calling", GENOME BIOLOGY, BIOMED CENTRAL LTD, vol. 13, no. 2, 29 February 2012 (2012-02-29), page R13, XP021126088, ISSN: 1465-6906, DOI: 10.1186/GB-2012-13-2-R13 the whole document in particular "methods" section -----	1-30

1

Form PCT/ISA/210 (continuation of second sheet) (April 2005)

page 2 of 2

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No

PCT/US2021/056711

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 2018195953 A1	12-07-2018	AU 2018205218 A1	18-07-2019
		BR 112019013886 A2	03-03-2020
		CA 3049142 A1	12-07-2018
		CN 110383386 A	25-10-2019
		EP 3566158 A1	13-11-2019
		JP 2020506677 A	05-03-2020
		KR 20190104336 A	09-09-2019
		RU 2019122320 A	08-02-2021
		US 2018195953 A1	12-07-2018
		US 2021389236 A1	16-12-2021
		WO 2018129314 A1	12-07-2018

10

20

30

40

50

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,
CZ,DE,DJ,DK,DM,DO,DZ,EC,EE,EG,ES,FI,GB,GD,GE,GH,GM,GT,HN,HR,HU,ID,IL,IN,IR,IS,IT,JO,JP,K
E,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,N
G,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 アブデ・アリ・フナイド・カガルワラ
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
- (72)発明者 ニティン・ウドパ
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 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