

(19)日本国特許庁(JP)

(12)特許公報(B2)

(11)特許番号
特許第7624929号
(P7624929)

(45)発行日 令和7年1月31日(2025.1.31)

(24)登録日 令和7年1月23日(2025.1.23)

(51)国際特許分類

F I

G 1 6 B 30/00 (2019.01)

G 1 6 B 30/00

G 1 6 B 40/20 (2019.01)

G 1 6 B 40/20

請求項の数 11 (全29頁)

(21)出願番号	特願2021-557678(P2021-557678)	(73)特許権者	500358711
(86)(22)出願日	令和2年12月4日(2020.12.4)		イルミナ インコーポレイテッド
(65)公表番号	特表2023-503739(P2023-503739 A)		アメリカ合衆国 カリフォルニア州 9 2
(43)公表日	令和5年2月1日(2023.2.1)		1 2 2 サンディエゴ イルミナ ウェイ
(86)国際出願番号	PCT/US2020/063496	(74)代理人	5 2 0 0
(87)国際公開番号	WO2021/113779		100108453
(87)国際公開日	令和3年6月10日(2021.6.10)		弁理士 村山 靖彦
審査請求日	令和5年12月1日(2023.12.1)	(74)代理人	100110364
(31)優先権主張番号	62/944,304		弁理士 実広 信哉
(32)優先日	令和1年12月5日(2019.12.5)	(74)代理人	100133400
(33)優先権主張国・地域又は機関	米国(US)		弁理士 阿部 達彦
		(72)発明者	ヴィラジ・デシュバンデ
			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

最終頁に続く

(54)【発明の名称】 遺伝子融合の迅速な検出

(57)【特許請求の範囲】

【請求項 1】

生体サンプル中の 1 つ以上の遺伝子融合を同定するためのコンピュータ実装方法であって、

1 つ以上のコンピュータによって、リード配列ユニットから複数の配列されたリードを表す第 1 のデータを取得することと、

1 つ以上のコンピュータによって、前記取得された第 1 のデータ内に含まれる複数の遺伝子融合候補を同定することと、

1 つ以上のコンピュータによって、前記複数の遺伝子融合候補をフィルタリングして、フィルタ処理された遺伝子融合候補のセットを決定することと、を含み、

前記フィルタ処理された遺伝子融合候補の特定の遺伝子融合候補のそれぞれについて、

1 つ以上のコンピュータによって、機械学習モデルへの入力のための入力データを生成することであって、前記入力データを生成することが、

(i) 前記特定の遺伝子融合候補が前記リード配列ユニットによって配列された参照配列の 1 つ以上のセグメントと、

(i i) 前記リード配列ユニットの出力に基づいて生成されたデータと、を含むデータから、前記特定の遺伝子融合候補を表す特徴データを抽出することを含み、前記リード配列ユニットの前記出力に基づいて生成された前記データが、変異体対立遺伝子頻度のカウント、特異リード配列のカウント、MAPQスコア、又は親遺伝子間の相同性を示すデータのうちの任意の 1 つ以上を含む、入力データを生成することと、

1つ以上のコンピュータによって、前記生成された入力データを前記機械学習モデルへの入力として提供することであって、前記機械学習モデルは、(i)前記特定の遺伝子融合候補が前記リード配列ユニットによって配列された参照配列の1つ以上のセグメントと、(ii)前記リード配列ユニットの出力に基づいて生成されたデータと、を表す機械学習モデル処理入力データに基づいて、遺伝子融合候補が有効な遺伝子融合である可能性を表す出力データを生成するように、標識された訓練ベクトルを使用して訓練されており、各標識された訓練ベクトルが、訓練融合候補を表すとともに、(i)前記訓練融合候補が配列された参照配列の1つ以上のセグメントと、(ii)変異体対立遺伝子頻度のカウント、特異リード配列のカウント、MAPQスコア、又は親遺伝子間の相同性を示すデータのうちの1つ以上とを含み、前記リード配列ユニットの出力に基づいて生成された前記データが、変異体対立遺伝子頻度のカウント、特異リード配列のカウント、MAPQスコア、又は親遺伝子間の相同性を示すデータのうちの1つ以上を含む、前記生成された入力データを前記機械学習モデルへの入力として提供することと、

10

1つ以上のコンピュータによって、前記生成された入力データを処理する前記機械学習モデルに基づいて、前記機械学習モデルによって生成された出力データを取得することと、

1つ以上のコンピュータによって、前記特定の融合候補が、前記出力データに基づいて、有効な遺伝子融合候補に対応するかどうかを決定することと、を含む、コンピュータ実装方法。

【請求項2】

20

前記入力データを生成することは、前記特定の遺伝子融合候補が前記リード配列ユニットによって配列された前記参照配列の前記セグメントのアノテーションを記述するアノテーションデータを含む特徴データを抽出することを更に含み、

前記機械学習モデルは、前記機械学習モデル処理入力データに基づいて、遺伝子融合候補が有効な遺伝子融合候補である可能性を表す出力データを生成するよう、標識された訓練ベクトルを使用して訓練されており、各標識された訓練ベクトルが、訓練融合候補を表すとともに、(i)前記訓練融合候補が配列された参照配列の1つ以上のセグメントと、(ii)前記訓練融合候補が配列された前記参照配列の前記セグメントのアノテーションを記述するアノテーションデータと、(iii)変異体対立遺伝子頻度のカウント、特異リード配列のカウント、MAPQスコア、又は親遺伝子間の相同性を示すデータのうちの1つ以上とを含み、前記機械学習モデル処理入力データは、

30

(i)前記特定の遺伝子融合候補が前記リード配列ユニットによって配列された参照配列の1つ以上のセグメントと、

(ii)前記特定の遺伝子融合候補が前記リード配列ユニットによって配列された前記参照配列の前記セグメントのアノテーションを記述するアノテーションデータと、

(iii)前記リード配列ユニットの出力に基づいて生成されたデータと、を表し、前記リード配列ユニットの出力に基づいて生成されたデータが、変異体対立遺伝子頻度のカウント、特異リード配列のカウント、MAPQスコア、又は親遺伝子間の相同性を示すデータのうちの1つ以上とを含む、請求項1に記載の方法。

【請求項3】

40

1つ以上のコンピュータによって、前記取得された第1のデータ内に含まれる複数の遺伝子融合候補を同定することが、1つ以上のコンピュータによって、複数の分割リード配列を同定することを含む、請求項1又は2に記載の方法。

【請求項4】

1つ以上のコンピュータによって、前記取得された第1のデータ内に含まれる複数の遺伝子融合候補を同定することが、1つ以上のコンピュータによって、複数の不一致リードペア配列を同定することを含む、請求項1～3のいずれか一項に記載の方法。

【請求項5】

前記リード配列ユニットが、操作を実行するように物理的に配置されたハードウェア論理回路を使用して構成されている1つ以上の処理エンジンのセットを使用して実装され、

50

前記ハードウェア論理回路を使用して、

(i) 第 1 のリードを表すデータを受信し、

(i i) 前記第 1 のリードを表す前記データを、参照配列の 1 つ以上の部分にマップして、整合参照配列の 1 つ以上の位置を同定し、

(i i i) 前記第 1 のリードの前記整合参照配列の位置のそれぞれに対応する 1 つ以上の配列スコアを生成し、

(i v) 前記 1 つ以上の配列スコアに基づいて、前記第 1 のリードのための 1 つ以上の候補配列を選択し、

(v) 前記第 1 のリードの候補配列を表すデータを出力する、請求項 1 ~ 4 のいずれか一項に記載の方法。

10

【請求項 6】

前記リード配列ユニットが、1 つ以上の中央演算処理装置 (C P U) 又は 1 つ以上のグラフィックス演算処理装置 (G P U) を使用することによって 1 つ以上の処理エンジンのセットを使用して実装され、前記 1 つ以上の C P U 又は 1 つ以上の G P U S に、

(i) 第 1 のリードを表すデータを受信させ、

(i i) 前記第 1 のリードを表す前記データを、参照配列の 1 つ以上の部分にマップして、前記第 1 のリードの整合参照配列の 1 つ以上の位置を同定させ、

(i i i) 前記第 1 のリードの前記整合参照配列の位置のそれぞれに対応する 1 つ以上の配列スコアを生成させ、

(i v) 前記 1 つ以上の配列スコアに基づいて、前記第 1 のリードのための 1 つ以上の候補配列を選択させ、

20

(v) 前記第 1 のリードの候補配列を表すデータを出力させる、ソフトウェア命令を実行する、請求項 1 ~ 4 のいずれか一項に記載の方法。

【請求項 7】

前記リード配列ユニットによって、まだ配列されていない複数のリードを受信することと、

前記リード配列ユニットによって、前記複数のリードの第 1 のサブセットを配列させることと、

前記リード配列ユニットによって、配列されたリードの前記第 1 のサブセットをメモリデバイス内に記憶することと、を更に含み、

30

1 つ以上のコンピュータによって、リード配列ユニットから複数の配列されたリードを表す第 1 のデータを取得することは、1 つ以上のコンピュータによって、前記メモリデバイスから配列されたリードの前記第 1 のサブセットを取得することと、前記リード配列ユニットが、まだ配列されていない前記複数のリードの第 2 のサブセットを配列させている間に、請求項 1 に記載の操作のうちの 1 つ以上を実行することと、を含む、請求項 1 ~ 6 のいずれか一項に記載の方法。

【請求項 8】

前記特定の融合候補が、前記出力データに基づいて有効な遺伝子融合候補に対応するかどうかを決定することが、

1 つ以上のコンピュータによって、前記出力データが所定の閾値を満たすかどうかを決定することと、

40

前記出力データが前記所定の閾値を満たすと決定することに基づいて、前記特定の融合候補が有効な遺伝子融合候補に対応すると決定することと、を含む、請求項 1 ~ 7 のいずれか一項に記載の方法。

【請求項 9】

前記特定の融合候補が、前記出力データに基づいて有効な遺伝子融合候補に対応するかどうかを決定することが、

1 つ以上のコンピュータによって、前記出力データが所定の閾値を満たすかどうかを決定することと、

前記出力データが前記所定の閾値を満たさないと決定することに基づいて、前記特定の

50

融合候補が有効な遺伝子融合候補に対応しないと決定することと、を含む、請求項 1 ~ 8 のいずれか一項に記載の方法。

【請求項 10】

生体サンプル中の 1 つ以上の遺伝子融合を同定するためのシステムであって、
操作可能である命令を記憶している 1 つ以上のコンピュータ及び 1 つ以上の記憶デバイスを含み、前記 1 つ以上のコンピュータによって実行された場合に、前記 1 つ以上のコンピュータに請求項 1 ~ 9 のいずれか一項に記載の方法を実行させる、システム。

【請求項 11】

1 つ以上のコンピュータによって実行可能な命令を含むソフトウェアを記憶している非一時的コンピュータ可読媒体であって、前記命令が、このような実行時に、前記 1 つ以上のコンピュータに請求項 1 ~ 9 のいずれか一項に記載の方法を実行させる、非一時的コンピュータ可読媒体。

10

【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、米国特許仮出願第 62 / 944 , 304 号 (2019 年 12 月 5 日出願) の利益を主張し、その全体が参照として本明細書に組み込まれる。

【背景技術】

【0002】

遺伝子融合は、癌などの疾患の治療において重要な診断及び治療標的である発癌動因として使用することができる。

20

【発明の概要】

【課題を解決するための手段】

【0003】

本開示の革新的な一態様によれば、生体サンプル中の 1 つ以上の遺伝子融合を同定するためのコンピュータ実装方法が開示される。一態様では、方法は、1 つ以上のコンピュータによって、リード配列ユニットから複数の配列されたリードを表す第 1 のデータを取得する操作と、1 つ以上のコンピュータによって、取得された第 1 のデータ内に含まれる複数の遺伝子融合候補を同定する操作と、1 つ以上のコンピュータによって、複数の遺伝子融合候補をフィルタリングして、フィルタ処理された遺伝子融合候補のセットを決定する操作と、を含み得、フィルタ処理された遺伝子融合候補の特定の遺伝子融合候補のそれぞれについて、1 つ以上のコンピュータによって、機械学習モデルへの入力のための入力データを生成することであって、入力データを生成することが、(i) 特定の遺伝子融合候補がリード配列ユニットによって配列された参照配列の 1 つ以上のセグメントと、(i i) リード配列ユニットの出力に基づいて生成されたデータと、を含むデータから、特徴データを抽出して特定の遺伝子融合候補を表すことを含む、入力データを生成することと、1 つ以上のコンピュータによって、生成された入力データを機械学習モデルへの入力として提供することであって、機械学習モデルは、(i) 特定の遺伝子融合候補がリード配列ユニットによって配列された参照配列の 1 つ以上のセグメントと、(i i) リード配列ユニットの出力に基づいて生成されたデータと、を表す機械学習モデル処理入力データに基づいて、遺伝子融合候補が有効な遺伝子融合である可能性を表す出力データを生成するように訓練されている、生成された入力データを機械学習モデルへの入力として提供することと、1 つ以上のコンピュータによって、生成された入力データを処理する機械学習モデルに基づいて、機械学習モデルによって生成された出力データを取得することと、1 つ以上のコンピュータによって、特定の遺伝子融合候補が、出力データに基づいて、有効な遺伝子融合候補に対応するかどうかを決定することと、を含み得る。

30

40

【0004】

その他のバージョンは、コンピュータ可読記憶デバイス上に符号化された命令によって定義される方法のアクションを実行するための、対応するシステム、装置、及びコンピュ

50

ータプログラムを含む。

【 0 0 0 5 】

これら及びその他のバージョンは、任意で、以下の特徴のうちの1つ以上を含んでもよい。例えば、いくつかの実装形態では、入力データを生成することは、特定の遺伝子融合候補がリード配列ユニットによって配列された参照配列のセグメントのアノテーションを記述するアノテーションデータを含む特徴データを抽出することを更に含む。このような実装形態では、機械学習モデルは、(i) 特定の遺伝子融合候補がリード配列ユニットによって配列された参照配列の1つ以上のセグメントと、(i i) 特定の遺伝子融合候補がリード配列ユニットによって配列された参照配列のセグメントのアノテーションを記述するアノテーションデータと、(i i i) リード配列ユニットの出力に基づいて生成されたデータと、表す、機械学習モデル処理入力データに基づいて、遺伝子融合候補が有効な遺伝子融合候補である可能性を表す出力データを生成するよう訓練されている。

10

【 0 0 0 6 】

いくつかの実装形態では、1つ以上のコンピュータによって、取得された第1のデータ内に含まれる複数の遺伝子融合候補を同定することは、1つ以上のコンピュータによって、複数の分割リード配列を同定することを含み得る。

【 0 0 0 7 】

いくつかの実装形態では、1つ以上のコンピュータによって、取得された第1のデータ内に含まれる複数の遺伝子融合候補を同定することは、1つ以上のコンピュータによって、複数の不一致リードペア配列を同定することを含む。

20

【 0 0 0 8 】

いくつかの実装形態では、リード配列ユニットは、操作を実行するように物理的に配置されたハードウェア論理回路を使用して構成されている1つ以上の処理エンジンセットを使用して実装され、ハードウェア論理回路を使用して、(i) 第1のリードを表すデータを受信し、(i i) 第1のリードを表すデータを、参照配列の1つ以上の部分にマップして、整合参照配列の1つ以上の位置を同定し、(i i i) 第1のリードの整合参照配列位置のそれぞれに対応する1つ以上の配列スコアを生成し、(i v) 1つ以上の配列スコアに基づいて、第1のリードのための1つ以上の候補配列を選択し、(v) 第1のリードの候補配列を表すデータを出力する。

【 0 0 0 9 】

30

いくつかの実装形態では、リード配列ユニットは、1つ以上の中央演算処理装置(C P U)又は1つ以上のグラフィックス演算処理装置(G P U)を使用することによって1つ以上の処理エンジンのセットを使用して実装され、1つ以上のC P U又は1つ以上のG P U Sに、(i) 第1のリードを表すデータを受信させ、(i i) 第1のリードを表すデータを、参照配列の1つ以上の部分にマップして、第1のリードの整合参照配列の1つ以上の位置を同定させ、(i i i) 第1のリードの整合参照配列位置のそれぞれに対応する1つ以上の配列スコアを生成させ、(i v) 1つ以上の配列スコアに基づいて、第1のリードのための1つ以上の候補配列を選択させ、(v) 第1のリードの候補配列を表すデータを出力させる、ソフトウェア命令を実行する。

【 0 0 1 0 】

40

いくつかの実装形態では、方法は、リード配列ユニットによって、まだ配列されていない複数のリードを受信することと、リード配列ユニットによって、複数のリードの第1のサブセットを配列させることと、リード配列ユニットによって、配列されたリードの第1のサブセットをメモリデバイス内に記憶することと、を更に含み得る。このような実装形態では、1つ以上のコンピュータによって、リード配列ユニットから複数の配列されたリードを表す第1のデータを取得することは、1つ以上のコンピュータによって、メモリデバイスから配列されたリードの第1のサブセットを取得することと、リード配列ユニットが、まだ配列されていない複数のリードの第2のサブセットを配列させている間に、請求項1に記載の1つ以上の操作を実行することと、を含み得る。

【 0 0 1 1 】

50

いくつかの実装形態では、リード配列ユニットの出力に基づいて生成されたデータは、変異体対立遺伝子頻度のカウント、特異リード配列のカウント、転写物にわたるリード確率、MAPQスコア、又は親遺伝子間の相同性を示すデータのうちの任意の1つ以上を含み得る。

【0012】

いくつかの実装形態では、特定の融合候補が、出力データに基づいて有効な遺伝子融合候補に対応するかどうかを決定することは、1つ以上のコンピュータによって、出力データが所定の閾値を満たすかどうかを決定することと、出力データが所定の閾値を満たすと決定することに基づいて、特定の融合候補が有効な遺伝子融合候補に対応すると決定することと、を含み得る。

10

【0013】

いくつかの実装形態では、特定の融合候補が、出力データに基づいて有効な遺伝子融合候補に対応するかどうかを決定することは、1つ以上のコンピュータによって、出力データが所定の閾値を満たすかどうかを決定することと、出力データが所定の閾値を満たさないと決定することに基づいて、特定の融合候補が有効な遺伝子融合候補に対応しないと決定することと、を含み得る。

【0014】

本開示のこれら及びその他の革新的な態様は、発明を実施するための形態、添付図面、及び特許請求の範囲を考慮することで容易に明らかである。

【図面の簡単な説明】

20

【0015】

【図1】有効な遺伝子融合を迅速に検出するためのシステムの一例のブロック図である。

【0016】

【図2】有効な遺伝子融合の迅速な検出を行うためのプロセスの一例のフローチャートである。

【0017】

【図3】有効な遺伝子融合を迅速に検出するためのシステムの別の例のブロック図である。

【0018】

【図4】有効な遺伝子融合の迅速な検出のためのシステムを実装するために使用され得るシステム構成要素のブロック図である。

30

【発明を実施するための形態】

【0019】

本開示は、遺伝子融合を迅速に検出するためのシステム、方法、装置、コンピュータプログラム、又はこれらの任意の組み合わせに関する。特定の遺伝子融合の存在は、特定の疾患の重要な指標、特定の疾患の特定の治療法の使用を示唆する指標、又はこれらの組み合わせであり得る。例えば、特定の遺伝子融合は、特定の種類の癌、例えば、急性骨髄性白血病及び慢性骨髄性白血病、脊髄形成異常性症候群(MDS)、軟組織肉腫、又はその治療の指標であり得る。本開示は、各融合候補が有効な遺伝子融合であるかどうかを決定するために処理された遺伝子融合候補(本明細書では「融合候補」とも呼ばれる)の数を低減するために、フィルタリングエンジンを使用することによって、正確な遺伝子融合を迅速に検出することができる。このフィルタリングエンジンは、候補遺伝子融合のフィルタ処理されたサブセットのみを、本明細書に記載された更なる下流処理のために促進することができるため、後続分析のために融合候補をより正確に選択することを可能にする一方で、有効な遺伝子融合を同定するために費やされる必要性がある演算リソースの低減を達成することもできる。

40

【0020】

候補遺伝子融合セットの低減もまた、他の技術的利点を提供する。例えば、本明細書にて開示する方法及びシステムは、全ての遺伝子融合候補を処理及び採点する従来の方法と比較して、低減された実行時間を提供する。その操作を実行するための低減された実行時間はまた、処理リソース(例えば、CPU又はGPUリソース)の消費量、メモリ使用量

50

、及び電力消費の低減をもたらす。フィルタリングエンジンは、従来の方法と比較して低減された実行時間を提供するが、本明細書にて開示する方法及びシステムはまた、実行時間を低減するための他の方法も提供し得る。例えば、いくつかの実装形態では、実行時間における更なる低減は、候補遺伝子融合を処理するために使用されるメタデータのマッピング、配列、及び生成を実行するために、ハードウェア・アクセラレートリード配列ユニット (hardware-accelerated read alignment unit) を使用することによって達成され得る。

【 0 0 2 1 】

図 1 は、有効な遺伝子融合を迅速に検出するためのシステム 1 0 0 の一例のブロック図である。システム 1 0 0 は、核酸配列決定デバイス 1 1 0 と、メモリ 1 2 0 と、二次解析ユニット 1 3 0 と、融合候補同定モジュール 1 4 0 と、融合候補フィルタリングモジュール 1 5 0 と、特徴セット生成モジュール 1 6 0 と、機械学習モデル 1 7 0 と、遺伝子融合決定モジュール 1 8 0 と、出力アプリケーションプログラムインターフェース (A P I) モジュール 1 9 0 と、出力ディスプレイ 1 9 5 と、を含み得る。図 1 の実施例では、これらの構成要素のそれぞれは、核酸配列決定デバイス 1 1 0 内に実装されるものとして記載される。しかしながら、本開示は、このような実施形態に限定されない。

10

【 0 0 2 2 】

その代わりに、いくつかの実装形態では、図 1 に記載された構成要素のうちの 1 つ以上は、核酸配列決定デバイス 1 1 0 の外側のコンピュータ上で実行され得る。例えば、いくつかの実装形態では、二次解析モジュールは、核酸配列決定デバイス 1 1 0 内に実装されてもよく、融合候補同定モジュール 1 4 0、融合候補フィルタリングモジュール 1 5 0、特徴セット生成モジュール 1 6 0、機械学習モデル 1 7 0、遺伝子融合決定モジュール 1 8 0、出力アプリケーションプログラムインターフェース (A P I) モジュール 1 9 0 は、1 つ以上の異なるコンピュータ内に実装され得る。このような実装形態では、1 つ以上の異なるコンピュータ及び核酸配列決定デバイスは、1 つ以上の有線ネットワーク、1 つ以上の無線ネットワーク、又はこれらの組み合わせを使用して通信可能に連結され得る。

20

【 0 0 2 3 】

本明細書の目的のために、用語「モジュール」は、1 つ以上のソフトウェア構成要素、1 つ以上のハードウェア構成要素、又はこれらの任意の組み合わせを含み、これらは、本明細書によるそれぞれのモジュールに起因する機能性を実現するために使用され得る。一般に、本明細書に記載された「モジュール」は、本明細書に記載されたモジュールの機能性を実現するために、ソフトウェア命令を実行するための 1 つ以上のプロセッサを使用する。プロセッサは、中央演算処理装置 (C P U)、グラフィック演算処理装置 (G P U) 等を含み得る。

30

【 0 0 2 4 】

同様に、本明細書で使用される用語「ユニット」は、1 つ以上のソフトウェア構成要素、1 つ以上のハードウェア構成要素、又はこれらの任意の組み合わせを含み、これらは、本明細書によるそれぞれのユニットに起因する機能性を実現するために使用され得る。一般に、本明細書に記載される「ユニット」は、本明細書に記載されたユニットの機能を実現する操作を実行するように処理エンジンとして配置された、ハードワイヤードデジタル論理ゲート又はハードワイヤードデジタル論理ブロックなどの 1 つ以上のハードウェア構成要素を使用する。このようなハードワイヤードデジタル論理ゲート又はハードワイヤードデジタル論理回路は、フィールド・プログラマブル・ゲート・アレイ (F P G A)、特定用途向け集積回路 (A S I C) などを含み得る。

40

【 0 0 2 5 】

核酸配列決定デバイス 1 1 0 (本明細書では配列決定デバイス 1 1 0 とも呼ばれる) は、一次核酸配列解析を行うように構成されている。一次解析を実施することは、配列決定デバイス 1 1 0 によって、血液サンプル、組織サンプル、痰、又は核酸サンプルなどの生体サンプル 1 0 5 を受領することと、配列決定デバイス 1 1 0 によって、受領した生体サンプルの核酸配列のヌクレオチドの順序をそれぞれ表す 1 つ以上のリード 1 1 2 などの出

50

力データを生成することと、を含み得る。いくつかの実装形態では、核酸シーケンサ 110 による配列決定は、複数のリードサイクルで実行され得、第 1 のリードサイクル「リード 1」は、核酸配列フラグメントの第 1 の端部からのヌクレオチドの順序を表す 1 つ以上の第 1 のリードを生成し、第 2 のリードサイクル「リード 2」は 1 つ以上の第 2 のリードを生成し、それぞれ、核酸配列フラグメントのうちの 1 つのその他の端部からのヌクレオチドの順序を表す。いくつかの実装形態では、リードは、長さが約 80 ~ 120 のヌクレオチドの短いリードであり得る。しかしながら、本開示は、任意の特定のヌクレオチド長のリードに限定されない。その代わりに、本開示は、任意のヌクレオチド長のリードに使用され得る。

【0026】

いくつかの実装形態では、生体サンプル 105 は、DNA サンプルを含み得、核酸シーケンサ 110 は、DNA シーケンサを含み得る。このような実装形態では、核酸シーケンサによって生成される、リード中の配列決定されたヌクレオチドの順序は、グアニン (G)、シトシン (C)、アデニン (A)、及びチミン (T) のうちの 1 つ以上を任意の組み合わせで含み得る。いくつかの実装形態では、核酸シーケンサ 110 を使用することで、生体サンプル 105 の RNA リードを生成し得る。このような実装形態では、このことは、RNA-seq プロトコルを使用して行うことができる。例として、逆転写酵素を使用して相補的 DNA (cDNA) を形成するために、逆転写を使用して、生体サンプル 105 を前処理し得る。その他の実装形態では、核酸シーケンサ 110 は、RNA シーケンサを含み得、生体サンプルは、RNA サンプルを含み得る。cDNA を使用して、又は RNA シーケンサを介して生成された RNA リードは、C、G、A、及び U r a c i l (U) から構成され得る。本明細書に記載された図 1 の実施例は、RNA リードの生成及び解析に関して説明される。しかしながら、本開示は、DNA 又は RNA リードを含む任意の種類の核酸配列リードを生成及び解析するために使用され得る。

【0027】

配列決定デバイス 110 は、所与のサンプルに関して、大規模並列配列決定技術 (massively parallel sequencing technology) の使用による超高スループット、超高スケラビリティ、及び超高速度を達成する方法で、リード 112-1、112-2、112-n (「n」は、0 を超える任意の正の整数である) などの配列リードを生成するように構成されている、次世代シーケンサ (NGS) を含み得る。NGS は、ゲノム全体の迅速な配列決定と、深く配列決定された標的領域にズームするか、RNA 配列決定 (RNA-seq) を利用して、新規 RNA 変異体及びスプライス部位を発見するか、又は遺伝子発現解析、ゲノムワイド DNA メチル化及び DNA-タンパク質相互作用などのエピジェネティック因子の解析、希少体細胞変異体及び腫瘍サブクローンを研究するための癌サンプルの配列決定、例えばヒト又は環境における微生物多様性の研究のための mRNA を定量化する能力と、を可能にする。

【0028】

配列決定デバイス 110 は、生体サンプル 105 を配列し、A、C、T、及び G を使用して表されるリードの対応するセットを生成し得る。配列決定デバイスは、次に、逆転写を実行して、対応する RNA 配列を表す cDNA 配列を生成し得る。これらの RNA 配列リード 112-1、112-2、112-n は、配列決定デバイス 110 によって出力され、メモリデバイス 120 内に記憶される。いくつかの実装形態では、RNA 配列リード 112-1、112-2、112-n は、メモリデバイス 120 内のリード 112-1、112-2、112-n の記憶に先立って、より小さいサイズのデータレコード内へと圧縮されてもよい。メモリデバイス 120 は、二次解析ユニット 130、融合候補同定モジュール 140、融合候補フィルタリングモジュール 150、特徴セット生成モジュール 160、機械学習モデル 170、遺伝子融合決定モジュール 180、及び出力 API モジュール 190 を含む、図 1 の構成要素のそれぞれによって、アクセス可能であり得る。それぞれのモジュールは、第 1 のモジュールの出力を第 2 のモジュールに提供するものとして示されてもよいが、このような機能の実際的な実装は、メモリ 120 などのメモリデバイ

10

20

30

40

50

ス内に出力を記憶する第1のモジュールと、メモリデバイスから記憶された出力にアクセスして、アクセスした出力を第2のモジュールへの入力として処理する第2のモジュールと、を含んでもよい。

【0029】

二次解析ユニット130は、メモリデバイス120内に記憶されたリード112-1、112-2、112-nにアクセスし、リード112-1、112-2、112-nに対して1つ以上の二次解析操作を実行し得る。いくつかの実装形態では、リード112-1、112-2、112-nは、圧縮データレコードでメモリデバイス120内に記憶されてもよい。このような実装形態では、二次解析ユニットは、リード記録に対して二次解析操作を実行するのに先立って、圧縮されたリード記録に対して展開操作を実行し得る。二次解析操作は、1つ以上のリードを参照ゲノムにマッピングすることと、1つ以上のリードを参照ゲノムに配列させること、又はその両方を含み得る。いくつかの実装形態では、二次解析操作はまた、変異体コーリング操作を含み得る。二次解析操作の実行に加えて、二次解析ユニット130はまた、分類操作を実行するように構成され得る。分類操作としては、例えば、配列されたリードがマップされた参照ゲノム内の位置に基づいて、二次解析ユニットによって配列されたリードを順序付けることを含み得る。

10

【0030】

図1の実施例などのいくつかの実装形態では、二次解析ユニット130は、メモリ132及びプログラマブル論理デバイス134を含み得る。プログラマブル論理デバイス134は、リード配列ユニット136などの1つ以上の二次解析操作ユニットを含むように動的に構成され得るハードウェア論理回路を有することができ、ハードウェア論理回路を使用して1つ以上の二次解析操作を実行するために使用され得る。プログラマブル論理デバイス134を、リード配列ユニット136などの二次解析操作ユニットを含むように動的に構成することは、例えば、リード配列ユニット136のハードウェア論理において、プログラマブル論理デバイス134に、機能を実現するように構成されているハードワイヤードデジタル論理構成内へとプログラマブル論理デバイス134のハードウェア論理ゲートを配列させる1つ以上の命令を、プログラマブル論理デバイス134に提供することを含み得る。

20

【0031】

プログラマブル論理デバイス134の動的構成を誘発する1つ以上の操作は、コンパイルされたハードウェア記述言語コードと、コンパイルされたハードウェア記述言語コードに基づいてそれ自体を構成するための、プログラマブル論理デバイス134のための1つ以上の命令等と、を含み得る。プログラマブル論理デバイス134の動的構成を誘発するこのような操作は、配列決定デバイス110、又は制御プログラムをホスティングするその他のコンピュータによって実行される制御プログラムによって、プログラマブル論理デバイス134に生成され得、かつプログラマブル論理デバイス134に展開され得る。いくつかの実装形態では、制御プログラムは、命令がメモリ120などのメモリデバイス内に存在するソフトウェアモジュールとすることができる。プログラマブル論理デバイス134を構成するための命令ハードウェア記述言語コード又はその他の命令を生成及び展開するための制御プログラムの機能性は、1つ以上のCPU又は1つ以上のGPUなどの1つ以上のプロセッサを使用して、制御プログラムソフトウェアモジュールを実行することによって実現し得る。

30

40

【0032】

リード配列ユニット136の機能性は、配列決定デバイス110によってメモリ120内に記憶されたRNAリード112-1、112-2、112-nなどの1つ以上の第1のリードを取得することと、取得した第1のリード112-1、112-2、112-nを参照配列の1つ以上の参照配列位置にマッピングすることと、次に、マップされた第1のリード112-1、112-2、112-nを参照配列に配列させることと、を含み得る。すなわち、マッピング段階は、特定のリードに一致する取得された第1のリードのそれぞれの特定のリードの候補参照配列位置のセットを同定し得る。次に、配列段階は、候

50

補参照配列位置のそれぞれを採点し、特定のリードのための正確な配列として最も高い配列スコアを有する特定の参照配列位置を選択し得る。参照配列は、周知のゲノムに対応する編成された一連のヌクレオチドを含み得る。

【 0 0 3 3 】

制御プログラムからの1つ以上の命令に応答してプログラマブル論理デバイス134のハードウェア論理ゲートを配置することは、ANDゲート、ORゲート、NORゲート、XORゲート、又はこれらの任意の組み合わせなどの論理ゲートを構成することで、リード配列ユニット136のデジタル論理機能を実行することを含み得る。あるいは又は加えて、ハードウェア論理ゲートを配列することは、加算、乗算、比較等を含む複雑な演算操作を実行するようにカスタマイズ可能なハードウェア論理ユニットを含む、動的に構成された論理ブロックを含み得る。ハードウェア論理ゲート、論理ブロック、又はこれらの組み合わせの正確な配列は、制御プログラムから受信した命令によって定義される。受信した命令は、エンティティによって書かれ、プログラマブル論理デバイス134へとプログラムされる二次解析操作ユニットの概略レイアウトを定義する、コンパイルされたハードウェア記述言語(HDL)プログラムコードを含み得るか、又はそれから導出される。HDLプログラムコードは、超高速集積回路ハードウェア記述言語(VHDL)、Verilog等の言語で書かれたプログラムコードを含み得る。このエンティティは、HDLプログラムコードを下書きする1人以上のヒトユーザー、HDLプログラムコードを生成する1つ以上の人工知能エージェント、又はこれらの組み合わせを含み得る。

【 0 0 3 4 】

プログラマブル論理デバイス134は、任意の種類のプログラマブル論理デバイスを含み得る。例えば、プログラマブル論理デバイス134は、制御プログラムによって特定のワークフローを実行するために、必要に応じて動的に構成可能であり、再構成可能である、1つ以上のフィールド・プログラマブル・ゲート・アレイ(FPGA)、1以上の複合プログラマブル論理デバイス(CPLD)、1つ以上のプログラマブル論理アレイ(PLA)、又はこれらの組み合わせを含み得る。例えば、いくつかの実装形態では、上述のように、プログラマブル論理デバイス134をリード配列ユニット136として使用することが望ましい場合がある。しかしながら、その他の実装形態では、プログラマブル論理デバイス134を使用して、変異体コーリング機能、又は隠れマルコフ・モデル(HMM)ユニットなどの変異体コーリングの支持における機能を実行することが望ましい場合がある。更にその他の実装形態では、プログラマブル論理デバイス134はまた、プログラマブル論理デバイス134のハードウェアロジックが、これらのタスクを実行することが可能であり、上記で同定されたその他のタスクが、1つ以上の演算処理ユニット150によって実行されるソフトウェア命令を使用した同じタスクの実行よりもはるかに速い故に、圧縮及び展開などの一般的な演算タスクを支持するように動的に構成され得る。いくつかの実装形態では、プログラマブル論理デバイス134は、異なる操作を実行するために、実行時間中に動的に再構成することができる。

【 0 0 3 5 】

例として、いくつかの実装形態では、プログラマブル論理デバイス134は、メモリデバイス120又は132内に記憶された第1のリード112-1、112-2、112-nの圧縮バージョンを表すデータにアクセスするために、展開ユニットとして動的に構成されるFPGAを使用して実装され得る。二次解析ユニット130は、(例えば、核酸シーケンサから受信したリードが圧縮された場合)展開ユニットを使用して、第1のリード112-1、112-2、122-nを表す圧縮データを展開し得る。展開ユニットは、メモリ120又は132内に展開されたリードを記憶し得る。このような実装形態では、FPGAは、次に、リード配列ユニット136として動的に再構成され、今度はメモリ132又は120内に記憶されている圧縮された第1のリード112-1、112-2、112-nのマッピング及び配列を実行するために使用され得る。リード配列ユニット136は、次に、メモリ132又は120内の、マップされ配列されたリードを表すデータを記憶し得る。一連の操作は、展開操作及びマッピング操作及び配列操作を含むものとして

記載されているが、本開示は、これらの操作を実行することに限定されるものではなく、又はこれらの操作のみを実行することに限定されるものではない。代わりに、プログラマブル論理デバイス 134 は、本明細書に記載された機能性を実現するために、必要に応じて、任意の順序で任意の操作ユニットの機能性を実行するように動的に構成され得る。

【0036】

図1の実施例は、リード配列ユニット136を実装するためにプログラマブル論理デバイス134の形態のハードウェア論理デバイスを使用する二次解析ユニット130を説明する。しかしながら、本開示は、リード配列ユニット136を実装するためにプログラマブル論理デバイスを使用することに限定されない。代わりに、他の種類の集積回路を使用して、二次解析ユニット130のハードワイヤードデジタル論理内にリード配列ユニット136を実装し得る。例えば、いくつかの実装形態では、二次解析ユニット143は、1つ以上の特定用途向け集積回路(ASIC)を使用して、1つ以上の二次解析操作ユニットの機能性を実現するように構成され得る。再プログラム可能ではないが、1つ以上のASICは、二次解析操作の実行を加速及び並行化するために、リード配列ユニット136、変異体コーリングユニット、変異体コーリング演算支持ユニット等などの1つ以上の二次解析操作ユニットのカスタムハードウェア論理を用いて設計され得る。いくつかの実装形態では、1つ以上の二次解析操作ユニットの機能性を実現する、二次解析ユニット130のハードワイヤード論理回路としての、1つ以上のASICの使用は、FPGAなどのプログラマブル論理デバイスを使用するよりも更に高速であり得る。したがって、当業者であれば、本明細書に記載された実施形態のいずれかにおけるFPGAなどのプログラマブル論理デバイスの代わりに、ASICを使用し得ることを理解するであろう。ASICが採用される実装形態では、ASICによって実行される二次解析操作ユニットごとに、専用のASIC又は単一のASICの専用論理グループを用いる必要がある。例として、リード配列のための1つ以上のASIC、展開のための1つ以上のASIC、圧縮のための1つ以上のASIC、又はこれらの組み合わせがある。あるいは、同じ機能性はまた、同じASIC内の専用論理グループで達成され得る。

【0037】

加えて、図1及び図3のシステム100及び300に関して説明される本開示の実施例は、それぞれ、プログラマブル論理デバイスにおけるリード配列ユニット136のハードウェア実装の使用に関して記載される。加えて、1つ以上のASICを使用して、リード配列エンジン又はその他の二次解析操作ユニットを実装し得ることが、上記に示されている。しかしながら、本開示は、このような二次解析操作を実施するためのハードウェアユニットの使用に限定されない。代わりに、いくつかの実装形態では、リード配列、圧縮、又は展開などのプログラマブル論理デバイスによって実行されるような本明細書に記載された操作のいずれかもまた、1つ以上のソフトウェアモジュールを使用して実施され得る。

【0038】

図1の実施例を参照すると、システム100の実行は、配列決定デバイス110が生体サンプル105を配列決定することで開始され得る。生体サンプルを配列決定することは、配列決定デバイス110によって、生体サンプル105中に存在するヌクレオチドの順序付けられた配列のデータ表現であるリード配列を生成することを含み得る。システム100がDNAリードを処理するように構成されている場合、次に、配列決定デバイス110によって生成されたリードは、メモリ120内に記憶され得る。

【0039】

あるいは、いくつかの実装形態では、システム100がRNAリードを処理するように構成されている場合、配列決定デバイス110は、逆転写酵素を使用して相補的DNA(cDNA)を形成する逆転写を使用して、生体サンプル110の前処理を実行するように構成され得る。図1の実施例での実装形態などの、このような実装形態では、配列決定デバイス110によって生成されたリードは、RNAリード112-1、112-2、112-nを含む。その他の実装形態では、核酸シーケンサ110はRNAシーケンサを含み得、生体サンプルはRNAサンプルを含み得る。RNAリードが、cDNAを使用するD

N A 配列決定デバイスによって、又は R N A シーケンサを介して生成されるかどうかにかかわらず、R N A リードはそれぞれ、C、G、A、及び U からなるヌクレオチドの配列を含む。リード 1 1 2 - 1、1 1 2 - 2、1 1 2 - n は、圧縮形式又は非圧縮形式でメモリ 1 2 0 内に記憶され得る。

【0040】

システム 1 0 0 の実行は、二次解析ユニット 1 3 0 がメモリ 1 2 0 内に記憶されたリード 1 1 2 - 1、1 1 2 - 2、1 1 2 - n を取得することで続行し得る。いくつかの実装形態では、二次解析ユニット 1 3 0 は、メモリデバイス 1 2 0 内のリード 1 1 2 - 1、1 1 2 - 2、1 1 2 - n にアクセスし、アクセスされたリード 1 1 2 - 1、1 1 2 - 2、1 2 2 - n を二次解析ユニット 1 3 0 のメモリ 1 3 2 内へと記憶し得る。その他の実装形態では、リード 1 1 2 - 1、1 1 2 - 2、1 1 2 - n の配列決定が完了し、二次解析ユニット 1 3 0 が二次解析操作を実行するために利用可能である制御プログラムによる決定の際に、制御プログラムは、リード 1 1 2 - 1、1 1 2 - 2、1 1 2 - n を二次解析ユニット 1 3 0 のメモリ 1 3 2 内へとロードし得る。

10

【0041】

リード 1 1 2 - 1、1 1 2 - 2、1 1 2 - n が圧縮された場合、二次解析ユニット 1 3 0 は、メモリ 1 3 2 又は 1 2 0 内のリード 1 1 2 - 1、1 1 2 - 2、1 1 2 - n にアクセスして、リード 1 1 2 - 1、1 1 2 - 2、1 1 2 - n を展開し、次に、展開されたリード 1 1 2 - 1、1 1 2 - 2、1 1 2 - n をメモリ 1 3 2 0 又は 1 2 0 内に記憶するために、展開ユニットとしてプログラマブル論理デバイス 1 3 4 を動的に構成し得る。いくつかの実装形態では、二次解析ユニットは、プログラマブル論理デバイスを動的に再構成し、制御プログラムからの命令に応答して展開を実行し得る。

20

【0042】

リード 1 1 2 - 1、1 1 2 - 2、1 2 2 - n が圧縮されない場合、二次解析ユニット 1 3 0 は、メモリ 1 3 2 又は 1 2 0 からのリードにアクセスし、リード配列操作を実行することができる。いくつかの実装形態では、二次解析ユニット 1 3 0 は、リード配列ユニット 1 3 6 を含むようにプログラマブル論理デバイス 1 3 4 を構成又は再構成するように二次解析ユニット 1 3 0 に命令する制御プログラムから、命令を受信し、次に、リード配列ユニット 1 3 6 を使用してリード 1 1 2 - 1、1 1 2 - 2、1 1 2 - n の配列を実行してもよい。あるいは、その他の実装形態では、プログラマブル論理デバイスは、リード配列ユニット 1 3 6 を含むように既に構成されていて、リード配列ユニット 1 3 6 を使用してリード 1 1 2 - 1、1 1 2 - 2、1 1 2 - n の配列を実行してもよい。更にその他の実装形態では、二次解析ユニット 1 3 0 は、リード配列を実行し、次に、A S I C を使用してリード 1 1 2 - 1、1 1 2 - 2、1 1 2 - n の配列を実行するように構成された A S I C を含んでもよい。

30

【0043】

二次解析ユニット 1 3 0 は、遺伝子融合解析と並行して、リード配列操作を実行するように構成され得る。例えば、二次解析ユニット 1 4 0 は、配列していない配列決定デバイス 1 1 0 によって生成されたリードの第 1 のバッチを取得し、リード配列ユニット 1 3 6 を使用してリードの第 1 のバッチを配列させることができ、プログラムされた論理デバイス 1 3 6 のハードウェア構成、又はプログラム命令を実行することによってソフトウェア内に実装され得るソーティングエンジンを使用して、配列されたリードを分類し、次に、メモリデバイス 1 3 2、1 3 0 内での記憶のために、配列され、分類されたリードの第 1 のバッチを出力する。いくつかの実装形態では、メモリ 1 3 2 は、リード配列ユニットによって処理されるデータをロードし、次に、リード配列ユニット 1 3 6 によって出力されたデータをオフロードする二次解析ユニット 1 3 2 のローカルキャッシュとして機能し得る。したがって、配列されたリードの第 1 のバッチがリード配列ユニット 1 3 6 によってメモリ 1 3 2 に出力されると、配列されたリードの第 1 のバッチを分類し、次に、メモリ 1 2 0 に出力することができる。次に、融合候補同定モジュール 1 4 0 は、メモリ 1 2 0 から配列され分類されたリードの第 1 のバッチにアクセスし、配列され分類されたリード

40

50

の第1のバッチを処理し得るが、一方で、二次解析ユニット130は、配列決定デバイス110によって生成され、従前に配列されていないリードの第2のバッチでの配列操作を実行する。本プロセスは、リードの各バッチがシステム100を通して処理されるまで反復的に実行され得る。本実施例は、配列され分類されたバッチを有するものとして記載されているが、配列されたリードのバッチも分類されることは、本開示の要件にはない。代わりに、後述のように、低減された実行時間などの性能向上を得るために、システム100又はシステム300において、配列及び分類されたリードの使用を用いることができる。

【0044】

融合候補同定モジュール140は、リード配列ユニット136によって配列された、配列され分類されたリードのバッチを取得して、配列され分類されたリードのバッチが1つ以上の遺伝子融合候補を含むかどうかを決定し得る。いくつかの実装形態では、受信されたバッチが配列され分類されたリードを含む場合、次に、融合候補同定モジュール140は、バッチに対応するゲノム間隔が少なくとも1つの融合候補の切断点と重複するバッチの分類されたリードを評価し得る。これにより、下流解析を必要とする融合候補の数を低減し得る。その他の実装形態では、受信されたバッチが分類されていない配列されたリードを含む場合、次に、融合候補同定モジュール140は、配列されたリードが融合候補であるかどうかを決定するために、バッチ内の配列されたリードのそれぞれを評価し得る。いくつかの実装形態では、リードのバッチが1つ以上の融合候補を含むかどうかを融合候補同定モジュール140によって決定する操作は、融合候補同定モジュール140によって、リードのバッチが、1つ以上の分割リード配列、1つ以上の不一致リードペア、1つ以上のソフトクリップ配列、又はこれらの組み合わせを含むかどうかを決定することを含む。

【0045】

いくつかの実装形態では、融合候補同定モジュール140は、融合候補として分割リード配列を同定するように構成され得る。融合候補同定モジュール140は、配列されたリードのバッチ内で各特定のリードが配列された参照配列の遺伝子を解析することによって、分割リード配列を同定し得る。融合候補同定モジュール140が、単一の遺伝子へのリードマップを決定する場合、融合候補同定モジュール140は、リードが分割リードではないと決定し得る。あるいは、融合候補同定モジュール140が、リードが2つの異なる遺伝子に配列していると決定した場合、次に、リードは分割リードであると決定され得る。このような実装形態では、分割リードは、融合候補であると決定され得る。例えば、リードのヌクレオチドの第1のサブセットが参照ゲノムの第1の親遺伝子に配列され、リードのヌクレオチドの第2のサブセットが参照ゲノムの第2の親遺伝子に配列される場合、リードは、2つの異なるリードに配列すると決定され得る。いくつかの実装形態では、ヌクレオチドの第1のサブセットはリードの接頭部であってもよく、ヌクレオチドの第2のサブセットはリードの接尾部であってもよい。融合候補同定モジュール140が分割リードを同定するように構成されている場合、分割リードを同定するデータは、もしあれば、メモリデバイス120内に記憶され得る。

【0046】

いくつかの実装形態では、融合候補同定モジュール140は、融合候補として不一致リードペアを同定するように構成され得る。融合候補同定モジュール140は、配列されたリードのバッチ内で各特定のリードペアが配列された参照配列の遺伝子を解析することによって、不一致リードペアを同定し得る。リードペアが参照配列に配列され、配列の配向及び範囲が予想される配向及び範囲である場合、次に、リードペアは、不一致リードではないと決定される。あるいは、リードペアが参照配列に配列され、配列の向き又は範囲が予想外である場合、次に、リードペアは、不一致リードペアであると決定される。このような実装形態では、1つのペアにおける1つのリードが1つの親遺伝子をマップし、他が別の親遺伝子をマップする場合、不一致リードは、融合候補であると決定され得る。融合候補同定モジュール140が不一致リードを同定するように構成されている場合、不一致リードを同定するデータは、もしあれば、メモリデバイス120内に記憶され得る。

10

20

30

40

50

【 0 0 4 7 】

いくつかの実装形態では、融合候補同定モジュール 1 4 0 は、ソフトクリップ配列を同定するように構成され得る。融合候補同定モジュール 1 4 0 は、配列されたリードのバッチ内で各特定の配列されたリードが配列された参照配列の遺伝子を解析することによって、ソフトクリップ配列を同定し得る。いくつかの実装形態では、融合候補同定モジュール 1 4 0 は、リードが参照ゲノム内の単一の位置にその全体が配列されているかどうかを決定し得る。リードが参照ゲノム内の単一の位置にその全体が配列されたら融合候補同定モジュール 1 4 0 が決定した場合、次に、融合候補同定モジュール 1 4 0 は、リードがソフトクリップリードではないと決定し得る。あるいは、融合候補同定モジュール 1 4 0 が、リードの一部分のみが参照ゲノムに配列されていると決定した場合、次に、融合候補同定モジュール 1 4 0 は、リードがソフトクリップリードであると決定し得る。リードマップの配列された部分が 1 つの親遺伝子に配列され、非配列部分が別の親遺伝子と類似の配列を有すると決定された場合、次に、ソフトクリップリードは、融合候補であると決定される。融合候補同定モジュール 1 4 0 がソフトクリップリードを同定するように構成されている場合、ソフトクリップリードを同定するデータは、もしあれば、メモリデバイス 1 2 0 内に遺伝子融合候補として記憶され得る。

10

【 0 0 4 8 】

融合候補フィルタリングモジュール 1 5 0 は、融合候補同定モジュール 1 4 0 によって同定された融合候補のセットを記述するデータを取得し得る。いくつかの実装形態では、融合候補フィルタリングモジュールは、メモリデバイス 1 2 0 にアクセスし、メモリデバイス 1 2 0 から融合候補を記述するデータを取得し得る。その他の実装形態では、融合候補フィルタリングモジュールは、融合候補同定モジュール 1 4 0 などの前述のモジュールの出力から融合候補を記述するデータを受信し得る。融合候補フィルタリングモジュール 1 5 0 は、遺伝子融合候補のセット全体よりも小さいフィルタ処理された遺伝子融合候補のセットを同定するために、融合候補のセットを記述するデータをフィルタ処理するための 1 つ以上のフィルタを使用し得る。いくつかの実装形態では、これらのフィルタは、単一の段階で適用される。例えば、1 つ以上のフィルタのそれぞれを適用することができ、融合候補のセット内の各融合候補を 1 つ以上のフィルタのそれぞれに対して評価し得る。しかしながら、その他の実装形態では、多段階のフィルタリングアプローチを採用し得る。このような実装形態では、融合候補同定モジュール 1 4 0 によって同定された融合候補の初期セットに、1 つ以上のフィルタの第 1 のセットが適用される。次に、第 1 のフィルタリング段階の適用後に残る、フィルタ処理された融合候補の第 1 のセットに、1 つ以上のフィルタの第 2 のセットが適用される。最適なフィルタ処理された融合候補のセットを達成するために、必要に応じて追加のフィルタリング段階を適用することもできる。

20

30

【 0 0 4 9 】

いくつかの実装形態では、融合候補フィルタリングモジュール 1 5 0 は、ショートリード配列決定の間に使用される高深度の適用範囲から生じる重複融合候補を考慮するために、融合候補のセットをフィルタ処理し得る。例えば、3 0 x 配列決定から生じるパイルアップは、融合候補同定モジュール 1 4 0 が、重複している最大 3 0 の融合候補を同定することにつながり得る。融合候補フィルタリングモジュール 1 5 0 は、融合候補の特性にフィルタを適用して重複を確認することによって、このような重複融合候補を除去し得る。例えば、融合候補フィルタリングモジュール 1 5 0 は、複数の融合候補が同じ親遺伝子に配列されているか、同じ又は類似の切断点に及ぶ参照ゲノムの一部に配列されるか、又はそれらの組み合わせに配列されるかどうかを決定し得る。融合候補フィルタリングモジュール 1 5 0 が、同じ親遺伝子に配列される、同じ又は類似の切断点に及ぶ参照ゲノムの一部に配列される、又はそれらの組み合わせに配列される、複数の融合候補を同定する場合、融合候補フィルタリングモジュール 1 5 0 は、融合候補が重複しており、融合候補のうちの 1 つのみを代表的な融合候補として選択することを決定し得る。このような場合、同じ又は類似の切断点又はこれらの組み合わせに及ぶ参照ゲノムの一部に配列された、同じ親遺伝子に配列された残りの融合候補を、更なる下流解析なしに廃棄し得る。次に、代表

40

50

的な融合候補を、メモリデバイス 120 などのメモリデバイス内のフィルタ処理された融合候補のセットに更に追加し得る。

【0050】

あるいは又は加えて、融合候補フィルタリングモジュール 150 は、1 つ以上の規定条件に基づいて、融合候補のセットをフィルタ処理し得る。例えば、融合候補フィルタリングモジュール 150 は、それぞれの融合候補を解析し、融合候補が 1 つ以上の属性を有するかどうかを決定することが、フィルタリングモジュール 150 によって採用される 1 つ以上の規定条件を満たすかどうかを決定し得る。いくつかの実装形態では、1 つ以上の規定条件は、融合候補の各部分の配列の位置、融合候補によって及んでいる切断点に対する配列の重複距離、融合候補の配列の向き、融合候補のリード配列品質、融合候補の更なるマッピング位置、又はそれらの任意の組み合わせを含み得る。

10

【0051】

例として、融合候補フィルタリングモジュール 150 によって 1 つ以上の規定条件を使用して、配列位置に基づいて融合候補をフィルタ処理し得る。いくつかの実装形態では、例えば、融合候補フィルタリングモジュール 150 は、配列の全長が、所定の数を超えるヌクレオチドによって融合切断点と交差する様式で、参照配列に配列されたリードを有する融合候補をフィルタ処理する規定条件を使用するように構成され得る。いくつかの実装形態では、本規定条件の所定の数 of のヌクレオチドは、8 個のヌクレオチドであり得る。あるいは又は加えて、融合候補フィルタリングモジュール 150 は、参照配列上の配列の全長が、融合切断点の所定の閾値数のヌクレオチド内に到達しないような様式で、参照配列に配列されたリードを有する融合候補をフィルタ処理するように構成され得る。いくつかの実装形態では、本規定条件についての所定の閾値数のヌクレオチドは、50 個のヌクレオチドであり得る。あるいは又は加えて、融合候補フィルタリングモジュール 150 は、2 つの融合切断点におけるリードの配列された位置が少なくとも所定の数 of のヌクレオチドを共有するような様式で、参照配列に配列されたリードを有する融合候補をフィルタ処理する規定条件を使用するように構成され得る。いくつかの実装形態では、所定の数 of の共有ヌクレオチドは、少なくとも 8 個のヌクレオチドを含み得る。

20

【0052】

別の例として、融合候補フィルタリングモジュール 150 によって 1 つ以上の規定条件を使用して、配向に基づいて融合候補をフィルタ処理し得る。いくつかの実装形態では、例えば、融合候補フィルタリングモジュール 150 は、融合転写物内で親遺伝子のうちの少なくとも 1 つのヌクレオチド配列が転写したことを示す配列の配向を有する融合候補をフィルタ処理する規定条件を使用するように、構成され得る。

30

【0053】

別の例として、融合候補フィルタリングモジュール 150 によって 1 つ以上の規定条件を使用して、マッピング品質に基づいて融合候補をフィルタリングし得る。いくつかの実装形態では、例えば、融合候補フィルタリングモジュール 150 は、所定の閾値を満たさないマッピング品質スコアを有するリード配列を有する融合候補をフィルタ処理する規定条件を使用するように、構成され得る。

【0054】

別の例として、融合候補フィルタリングモジュール 150 によって 1 つ以上の規定条件を使用して、追加のマッピング位置に基づいて融合候補をフィルタリングし得る。いくつかの実装形態では、例えば、融合候補フィルタリングモジュール 150 は、融合候補のリード部分が参照配列の複数の位置にマップするとの決定に基づいて融合候補をフィルタ処理する規定条件を使用するように、構成され得る。いくつかの実装形態では、融合候補フィルタリングモジュール 150 は、相同遺伝子であることが注釈された位置を除外するように構成され得る。

40

【0055】

1 つ以上の規定条件のそれぞれを満たす融合候補は、メモリデバイス 120 などのメモリデバイス内のフィルタ処理された融合候補のセットに追加され得る。1 つ以上の規定条

50

件のそれぞれを満たさない融合候補は、更なる下流解析なしに廃棄され得る。いくつかの実装形態では、融合候補の規定条件に基づくフィルタリングは、第１段階の重複排除フィルタの適用後に、第２段階のフィルタとして適用され得る。その他の実装形態では、第１のフィルタリング段階として、融合候補の規定条件に基づくフィルタリングを適用し得、次に、重複排除フィルタを第２段階のフィルタとして適用し得る。その他の実装形態では、規定条件に基づくフィルタリングは、従来の重複排除フィルタリングなしに、単一段階のフィルタとして適用され得る。これらの規定条件のうちの１つ以上に基づいて融合候補をフィルタリングすることは、下流で更に処理する必要がある融合候補の数を大幅に低減し得る。

【００５６】

10

融合候補フィルタリングモジュール１５０によって、フィルタ処理された融合候補のセット内のそれぞれの融合候補に対して下流処理を実行し得る。下流処理は、特徴セット生成モジュール１６０、機械学習モデル１７０、遺伝子融合決定モジュール１８０、及び出力ＡＰＩモジュール１９０の実行を含む。このような下流処理は、候補融合候補が有効な遺伝子融合に対応するかどうかを決定するために使用され得る。

【００５７】

特徴セット生成モジュール１６０は、複数のデータソースからデータを引き出して、特徴抽出を実行するデータ属性のセットを同定し得る。これらのデータソースは、融合候補についてメモリ１２０内に記憶された属性データを含み、（ｉ）融合候補のリード（複数可）と、（ｉｉ）融合候補のリードが配列された参照配列位置の部分（複数可）と、（ｉｉｉ）特定の遺伝子融合候補が配列された参照ゲノムのセグメントのアノテーションと、を含む。いくつかの実装形態では、アノテーションは、遺伝子エクソンのアノテーション、相同遺伝子の存在を示すアノテーション、富化遺伝子のリストを示すアノテーション、又はこれらの組み合わせを含み得る。

20

【００５８】

データは、特徴セット生成モジュール１６０が配列プロセス中にリード配列ユニット１３６によって生成されたデータも含み得ることをソースする。いくつかの実装形態では、特徴セット生成モジュール１６０は、融合候補の配列中に、リード配列ユニット１３６によって生成されたデータから特徴データを導出し得る。例えば、特徴セット生成モジュール１６０は、リード配列ユニット１３６によって生成されたデータから、例えば、変異体対立遺伝子頻度数、特異リード配列数、転写物にわたるリード確率、ＭＡＰＱスコア、親遺伝子間の相同性を示すデータ、又はこれらの組み合わせなどの情報を導出し得る。

30

【００５９】

特徴セット生成モジュール１６０は、複数のデータソースから引き出された融合候補の上述の属性のうち１つ以上を表す特徴データを生成し、機械学習モデル１７０への入力のために、特徴データを１つ以上のデータ構造１６２へとコード化するために使用され得る。例えば、いくつかの実装形態では、融合候補の属性から抽出された特徴のセット全体は、機械学習モジュール１７０へと組み込まれた単一のベクトル１６２へとコード化されてもよい。例えば、分割リード配列又はソフトクリップ配列のシナリオでは、これらの種類の融合候補の属性から抽出された特徴のそれぞれは、単一のベクトル１６２へとコード化され得る。

40

【００６０】

その他の実装形態では、特徴データは、すなわち、融合候補の属性から抽出されてもよく、コード化された複数の入力ベクトルであってもよい。このようなシナリオでは、入力ベクトル１６２は、一対の入力ベクトル１６２a、１６２bから構成されてもよい。例えば、分割リード融合候補のシナリオでは、分割リードの接頭部に関連する属性から抽出された特徴のそれぞれは、分割リードの接頭部のヌクレオチドを表す特徴を含み、接頭部が配列する参照配列のセグメントを表す特徴、及び接頭部に関連する上述の属性から抽出された任意のその他の特徴、又はこれらの任意の組み合わせを、入力ベクトル１６２aへとコード化してもよい。同様に、このような実装形態では、各特徴は、分割リードの接尾部

50

のヌクレオチドを表す特徴を含む、分割リードの接尾部に関連する属性から抽出され、接尾部が配列する参照配列のセグメントを表す特徴、及び接尾部に関連する上述の属性から抽出された任意のその他の特徴、又はこれらの任意の組み合わせを、入力ベクトル 1 6 2 b へとコード化してもよい。別の例として、不一致リードペアが融合候補として同定される場合、次に、不一致リードペアの第 1 のリードを表す抽出された特徴、そこへ配列された参照配列の部分を表す抽出された特徴、不一致リードペアの第 1 のリードに関連する属性から抽出された特徴、又はこれらの任意の組み合わせは、入力ベクトル 1 6 2 a へとコード化されてもよい。同様に、このような実施例では、不一致リードペアの第 2 のリードを表す抽出された特徴、そこへ配列された参照配列の部分を表す抽出された特徴、不一致リードペアの第 2 のリードに関連する属性から抽出された特徴、又はこれらの任意の組み合わせは、入力ベクトル 1 6 2 b へとコード化されてもよい。

10

【 0 0 6 1 】

1 つ以上のベクトル 1 6 2 のそれぞれは、生成された特徴データを数値的に表し得、特徴データは、融合候補から抽出された特徴のうちのいずれか、又は融合候補に関連するリード配列ユニット 1 3 6 から受信され、メモリ 1 2 0 内に記憶されたデータから抽出された特徴のいずれかを含む。例えば、各ベクトル 1 6 2 又は 1 6 2 a、1 6 2 b は、それぞれが特定の融合候補の特定のリードの特定の特徴に対応する複数のフィールドを含み得る。特定の融合候補に依存して、これは、上述のように、1 つ以上の入力ベクトルをもたらし得る。特徴セット生成モジュール 1 6 0 は、融合候補の特定のリードの属性において特定の特徴が表現された程度を説明するフィールドのそれぞれについての数値を決定し得る。フィールドのそれぞれについて決定された数値を使用して、融合候補のリードの属性を表す生成された特徴データを、1 つ以上のそれぞれのベクトル 1 6 2 へとコード化し得る。融合候補の対応するリードを数値的に表す、生成された 1 つ以上のベクトル 1 6 2 a、1 6 2 b は、機械学習モデル 1 7 0 への入力として提供される。いくつかの実装形態では、複数の概念ベクトルが融合候補のために生成される場合であっても、複数の概念ベクトルは、機械学習モデル 1 7 0 へと入力され得る単一ベクトル 1 6 2 に接触し得る。このような実装形態では、(i) 接頭部の特徴が第 1 のベクトルに割り当てられ、接尾部の特徴が第 2 のベクトルに割り当てられる、特定の分割リード実装において、又は (i i) 単一ベクトルの第 1 の部分が第 1 の概念ベクトルに対応し得、単一ベクトルの第 2 の部分が第 2 の概念ベクトルに対応し得る、不一致ペア実装において、複数のベクトルが保証された。

20

30

【 0 0 6 2 】

機械学習モデル 1 7 0 は、融合候補の特徴を表す 1 つ以上の入力ベクトル 1 6 2 を入力する処理に基づいて、融合候補が有効な遺伝子融合に対応する可能性を発生させるように訓練された深層ニューラルネットワークを含み得る。有効な遺伝子融合は、1 つの親遺伝子の接頭部を別の親遺伝子の接尾部と接続するゲノム中の再配列故に、複数の遺伝子からの配列を含有するキメラ転写物である。本開示との関係においては、有効な遺伝子融合は、例えば、機械学習モデルによって生成された出力データ 1 7 8 が所定の閾値を満たす場合に、モデル 1 7 0 によって予測されていると決定される。機械学習モデル 1 7 0 は、入力データを受信するための入力層 1 7 2 と、入力層 1 7 2 を介して受信した入力データを処理するための 1 つ以上の隠れ層 1 7 4 a、1 7 4 b、1 7 4 c、及び出力データ 1 7 8 を提供するための出力層 1 7 6 と、を含み得る。各隠れ層 1 7 4 a、1 7 4 b、1 7 4 c は、1 つ以上の重量又はその他のパラメータを含む。訓練中、それぞれ対応する隠れ層 1 7 4 a、1 7 4 b、1 7 4 c の重量又は他のパラメータを調節することができるが、これにより、訓練された深層ニューラルネットワークが、1 つ以上の入力ベクトル 1 6 2 を処理する機械学習モデル 1 7 0 に基づいて 1 つ以上の入力ベクトル 1 6 2 が有効な遺伝子候補を示す可能性を示す、所望の標的出力 1 7 8 を生成する。

40

【 0 0 6 3 】

機械学習モデル 1 7 0 は、多数の異なる方法で訓練され得る。一実装形態では、機械学習モデル 1 7 0 は、(i) 有効な融合候補の属性から抽出された特徴を表す 1 つ以上の入力ベクトルと、(i i) 無効な融合候補の属性から抽出された特徴を表す 1 つ以上の入力

50

ベクトルと、を区別するように訓練され得る。いくつかの実装形態では、このような訓練は、標識された訓練ベクトル対を使用して達成され得る。各訓練ベクトルは、訓練融合候補を表すことができるが、上記の1つ以上の入力ベクトル162と同じ種類の特徴データから構成され得る。このような実装形態では、融合候補の属性から抽出された特徴を表す1つ以上の入力ベクトル162は、有効な遺伝子融合又は無効な遺伝子融合であると標識され得る。いくつかの実装形態では、有効な遺伝子融合標識又は無効な遺伝子融合標識は、数値として表すことができる。例えば、いくつかの実装形態では、有効な遺伝子融合標識は「1」であり得、無効な遺伝子融合標識は「0」であり得る。その他の実装形態では、例えば、有効な遺伝子融合標識は、所定の閾値を満たす「0」～「1」の数であり得、無効な遺伝子融合標識は、所定の閾値を満たさない「0」～「1」の数であり得る。このような実装形態では、数が所定の閾値を満たす又は満たさない大きさは、入力ベクトルの訓練対が有効な遺伝子融合又は無効な遺伝子融合を表すという信頼度の指標である。いくつかの実装形態では、所定の閾値を満たすことは、所定の閾値を超えることを含み得る。しかしながら、実装形態は、閾値を満たすことが所定の閾値を超えないことを意味するように構成することもできる。このような実装形態は、例えば、コンパレータ及びパラメータが両方とも否定された実装形態を含み得る。

10

【0064】

訓練中、1つ以上の訓練ベクトルの各標識されたセットは、機械学習モデル170への入力として提供され、機械学習モデル170によって処理され、次に、機械学習モデル170によって生成された訓練出力を使用して、1つ以上の訓練ベクトルの標識されたセットそれぞれについて予測標識を決定する。訓練融合候補のための一対のリードに対応する標識された1つ以上の訓練ベクトルの機械学習モデルの処理に基づいて、機械学習モデル170によって生成された予測標識を、訓練融合候補の1つ以上のリード（又はリード部分）に対応する1つ以上の訓練ベクトルの訓練標識と比較することができる。次に、機械学習モデル170のパラメータは、予測標識と訓練標識との差に基づいて調整され得る。本プロセスは、訓練融合候補に対応する1つ以上の訓練ベクトルのセットの処理に基づいて、機械学習モデル170によって生成された予測融合候補標識が、所定のレベルの誤差内で、それぞれの訓練融合候補に対応する1つ以上の訓練ベクトルのセットの訓練標識と一致するまで、それぞれの訓練融合候補に対応する複数の標識訓練ベクトル（複数可）の各々について、反復継続し得る。

20

30

【0065】

いくつかの実装形態では、標識された訓練融合候補は、1人以上のヒトユーザーによってレビュー及び標識された訓練融合候補のライブラリから取得され得る。しかしながら、その他の実装形態では、標識された訓練融合候補は、シミュレータによって生成及び標識された訓練融合候補を含み得る。このような実装形態では、シミュレータを使用して、機械学習モデル170を訓練するために使用され得る訓練融合候補の異なるカテゴリの分布を作成し得る。一般に、実行時間機械学習モデル170が単一の入力ベクトル162を許容する場合、融合候補の抽出された特徴のそれぞれは、単一の入力ベクトル162をコード化しており、次に、機械学習モデル170は、上記の訓練プロセスを使用して入力ベクトル162と同じ特徴の単一入力ベクトルを使用して訓練される。同様に、実行時間機械学習モジュール170が、上述のように2つの訓練ベクトル162a、162bを許容する場合、次に、機械学習モデル170は、上記の入力ベクトル162a、162bの同じ対応する特徴をそれぞれ有する2つの入力ベクトルを使用して訓練される。すなわち、実行時間に処理される入力ベクトルの種類は、上述の訓練プロセスを使用してモデル170を訓練するために使用されるベクトルと同じ時間である。

40

【0066】

融合候補の属性から抽出された特徴に対応する入力データ162の処理中に、各隠れ層174a、174b、174cの出力は活性化ベクトルを含み得る。各それぞれの隠れ層によって出力される活性化ベクトルは、深層ニューラルネットワークの後続層を通して伝搬され、出力層によって使用されて、出力データ178を生成し得る。図1の実施例では

50

、機械学習モデル１７０は、それぞれが融合候補のリードのうちの１つに対応する別個の入力ベクトル１６２ a、１６２ bの機械学習モデル処理に基づいて、機械学習モデル１７０によって生成された組み合わせスコアを表す出力データ１７８を生成するように訓練される。この組み合わせスコア１７８は、最終隠れ層１７４ cから受信した活性化ベクトル上で、訓練された機械学習モデル１７０の出力層１７６によって実行される演算に基づいて、訓練された機械学習モデルの出力層１７６によって最終的に生成される。

【００６７】

訓練された機械学習モデル１７０によって生成された出力データ１７８は、１つ以上の入力ベクトル１６２に対応する融合候補が有効な融合候補であることを示すかどうかを決定するために、遺伝子融合決定モジュール１８０によって評価され得る。いくつかの実装形態では、出力データ１７８は、訓練された機械学習モデル１７０によって遺伝子融合決定モジュール１８０に提供され得る。その他の実装形態では、システム１００は、訓練された機械学習モデル１７０の出力１７８を、遺伝子融合決定モジュール１８０による後続のアクセスのためにメモリデバイス１２０などのメモリデバイスに記憶し得る。

遺伝子融合決定モジュール１８０は、機械学習モデル１７０によって生成された出力データ１７８を取得し、出力データ１７８を評価して、出力データ１７８に基づいて、入力ベクトル１６２ a、１６２ bの対１６２に対応する融合候補が有効な遺伝子融合であるかどうかを決定し得る。いくつかの実装形態では、遺伝子融合決定モジュール１８０は、１つ以上の入力ベクトル１６２に対応する融合候補が、機械学習モデルによって生成された出力データ１７８を所定の閾値と比較することによって、有効な遺伝子融合であるかどうかを決定し得る。遺伝子融合決定モジュール１８０が、出力データ１７８が所定の閾値を満たすと決定した場合、次に、遺伝子融合決定モジュール１８０は、１つ以上の入力ベクトル１６２に対応する融合候補が有効な遺伝子融合であると決定し得る。あるいは、遺伝子融合決定モジュール１８０が、出力データ１７８が所定の閾値を満たさないと決定した場合、次に、遺伝子融合決定モジュール１８０は、１つ以上の入力ベクトル１６２に対応する融合候補が有効な遺伝子融合ではないと決定し得る。

【００６８】

いくつかの実装形態では、遺伝子融合決定モジュール１８０は、機械学習モデル１７０によって生成された出力データ１７８の遺伝子融合決定モジュール１８０の評価に基づいて、遺伝子融合決定モジュール１８０によって行われた決定の結果を示す出力データ１８２を生成し得る。本出力データ１８２は、１つ以上の入力ベクトル１６２に対応する遺伝子融合候補を同定するデータと、遺伝子融合決定モジュール１８０の決定を同定するデータと、を含み得る。遺伝子融合決定モジュール１８０の決定を同定するデータは、１つ以上の入力ベクトル１６２に対応する遺伝子融合候補が有効な遺伝子融合又は無効な遺伝子融合であるかどうかを示すデータを含み得る。いくつかの実装形態では、出力データ１８２は、出力データ１７８に基づいて同定された有効な遺伝子融合のリスト、出力データ１７８に基づいて同定された無効な遺伝子融合のリスト、有効な遺伝子融合が同定されなかったことを示すデータ、又はこれらの任意の組み合わせのみを示してもよい。いくつかの実装形態では、本出力データ１８２は、別の演算モジュールによる後続の使用のために、ユーザーデバイスへの後続の出力等のために、メモリ１８２内に記憶され得る。

【００６９】

あるいは又は加えて、遺伝子融合決定モジュール１８０は、出力アプリケーション・プログラミング・インターフェース（ＡＰＩ）モジュール１９０の入力として提供され得る出力データ１８４を生成し得る。出力データ１８４は、１つ以上の入力ベクトル１６２に対応する遺伝子融合候補が有効な遺伝子融合又は無効な遺伝子融合であるかどうかを示す出力を、出力ディスプレイに発生させるように、出力ＡＰＩに指示し得る。いくつかの実装形態では、命令は、出力ＡＰＩモジュール１９０に、メモリデバイス１２０内に記憶された出力データ１８２にアクセスし、レンダリングデータを生成させ得るが、これは、出力ディスプレイ１９５に結合された演算デバイスによってレンダリングされた場合に、出力ディスプレイ１９５に、（ i ）１つ以上の入力ベクトル１６２に対応する融合候補を同

10

20

30

40

50

定するデータと、(i i) 同定された融合候補が有効な遺伝子融合又は無効な遺伝子融合であるかどうかを示すデータと、を生起させる。これは、出力ディスプレイ 195 に、メモリ 184 内に記憶された出力データ 182 のいずれかを表示させることを含み得る。いくつかの実装形態では、本出力はレポートの形態で表示され得る。

【0070】

いくつかの実装形態では、遺伝子融合決定モジュール 180 は、フィルタ処理された遺伝子融合候補のセットの各融合候補に対して実行される下流処理の性能に基づいて、メモリデバイス 120 内の各遺伝子融合候補に関する出力データ 182 を記憶する。このような実装形態では、遺伝子融合決定モジュール 180 は、各融合候補の下流処理が完了すると、フィルタ処理された遺伝子融合候補のセットのそれぞれの融合候補について、メモリ 120 内に記憶された遺伝子融合解析の結果を出力するように、出力 A P I モジュール 190 にのみ指示してもよい。このようなシナリオでは、出力ディスプレイ 195 上に表示するために提供される出力 192 は、有効な遺伝子融合のリスト、無効な遺伝子融合のリスト、又はその両方を含み得る。その他の実装形態では、遺伝子融合決定モジュール 180 は、その特定の融合候補に対する下流処理の完了時に、もしあれば、出力 A P I モジュール 190 に、同定された遺伝子融合のリストを示す結果データを出力させ得る。

【0071】

その他の種類の出力 192 は、出力 A P I モジュール 190 によって提供され得る。例えば、いくつかの実装形態では、出力 192 は、(i) 1 つ以上のベクトル 162 に対応する融合候補を同定するデータと、(i i) 同定された融合候補が有効な遺伝子であるかどうかを示すデータと、を含むレポートを出力させるためのプリンタなどの、別のデバイスを生起させるデータとすることができる。その他の実装形態では、本出力データ 192 は、(i) 1 つ以上のベクトル 162 に対応する融合候補を同定するデータと、(i i) 同定された融合候補が有効な遺伝子であるかどうかを示すデータと、を含むオーディオデータを、スピーカに出力させ得る。その他の種類の出力データはまた、出力 A P I R モジュール 190 によって誘発され得る。

【0072】

いくつかの実装形態では、出力ディスプレイ 195 は、配列決定デバイス 110 のディスプレイパネルであり得る。その他の実装形態では、出力ディスプレイ 195 は、1 つ以上のネットワークを使用して配列決定デバイス 110 に接続されたユーザーデバイスのディスプレイパネルであり得る。実際に、配列決定デバイス 110 は、出力データ 192 を、任意のディスプレイを有する任意のデバイスに通信するために使用され得る。

【0073】

図 2 は、有効な遺伝子融合の迅速な検出を行うためのプロセス 200 の一例のフローチャートである。システム 100 などのシステムは、1 つ以上のコンピュータを使用して、リード配列ユニット (210) からの複数の配列されたリードを表す第 1 のデータを取得することによって、プロセス 200 の実行を開始し得る。システムは、取得された第 1 のデータ (220) 内に含まれる複数の遺伝子融合候補を同定し得る。システムは、複数の遺伝子融合候補をフィルタ処理して、遺伝子融合候補 (230) のフィルタ処理されたセットを決定し得る。

【0074】

本システムは、フィルタ処理された遺伝子融合候補のセット (240) の特定の遺伝子融合候補を取得し得る。システムは、機械学習モデルへの入力のための入力データを生成し得るが、入力データを生成することは、(i) 特定の遺伝子融合候補がリード配列ユニットによって配列された参照配列の 1 つ以上のセグメントと、(i i) リード配列ユニット (250) の出力に基づいて生成されたデータと、を含むデータから特定の遺伝子融合候補を表す特徴データを抽出することを含む。

【0075】

システムは、生成された入力データを機械学習モデルへの入力として提供し得るが、機械学習モデルは、(i) 特定の遺伝子融合候補がリード配列ユニットによって配列された

10

20

30

40

50

参照ゲノムのセグメントと、(i i) リード配列ユニット (2 6 0) の出力に基づいて生成されたデータと、を表す機械学習モデル処理入力データに基づいて、遺伝子融合候補が有効な遺伝子融合である可能性を表す出力データを生成するように訓練されている。システムは、入力データ (2 7 0) を処理する機械学習モデルに基づいて、機械学習モデルによって生成された出力データを取得し得る。システムは、特定の融合候補が、出力データ (2 8 0) に基づいて、有効な遺伝子融合候補に対応するかどうかを決定し得る。

【 0 0 7 6 】

段階 2 8 0 の完了時に、システムは、フィルタ処理された融合候補のセットの別の融合候補が評価されるべきかどうかを決定し得る (2 9 0) 。システムが、評価されるべきフィルタ処理された融合候補のセットの別の融合候補が存在すると決定した場合、次に、システムは、段階 2 4 0 でプロセス 2 0 0 の実行を継続し得る。あるいは、システムが、評価されるべきフィルタ処理された融合候補のセットの別の融合候補が存在しないと決定した場合、次に、システムは、段階 2 9 5 でプロセスの実行を終了し得る。融合候補のセットのセットが使い尽されていない場合、別の融合候補が、フィルタ処理された融合候補のセット内に存在し得る。

【 0 0 7 7 】

図 3 は、有効な遺伝子融合を迅速に検出するためのシステム 3 0 0 の別の例のブロック図である。システム 3 0 0 が、配列決定デバイス 1 1 0 を使用して R N A (又は D N A) 配列リード 1 1 2 を生成する、R N A 配列リード 1 1 2 を参照配列に配列させるために二次解析ユニット 1 3 0 を使用する、融合候補同定モジュール 1 4 0 を使用して融合候補を同定する、融合候補フィルタリングモジュール 1 5 0 を使用して下流解析のためにフィルタ処理された融合候補のセットを決定する、次に、フィルタ処理された融合候補のセットの下流解析を実行して、特徴セット生成モジュール 1 6 0 と、機械学習モジュール 1 7 0 と、遺伝子融合決定モジュール 1 9 0 と、出力 A P I モジュール 1 9 0 と、を使用して、有効な遺伝子融合を同定する、という点で、システム 3 0 0 はシステム 1 0 0 と同じ機能を実行する。これらの機能ユニット、モジュール、又はモデルのそれぞれは、図 1 のシステム 1 0 0 の説明においてそれらに起因したものと同一機能を実行する。

【 0 0 7 8 】

システム 3 0 0 とシステム 1 0 0 との間の差異は、融合候補の同定、融合候補フィルタリング、及びフィルタ処理された融合候補のセットの下流解析が、異なるコンピュータ 3 2 0 上で実行され、配列決定デバイス 1 1 0 内ではない、という点である。したがって、システム 3 0 0 とシステム 1 0 0 との間の差異は、ネットワーク 3 1 0 を使用して、遺伝子融合解析のために配列されたリードがどのように実装されてコンピュータ 3 2 0 に連結されるか、コンピュータ 3 2 0 によってどのように取り出されるか、また遺伝子融合結果がどのように実装されて、出力のための対応するディスプレイを有する別のデバイスへと送信されるか、という点にある。

【 0 0 7 9 】

より詳細には、配列決定デバイス 1 1 0 は、生体サンプル 1 0 5 を配列し、R N A リード 1 1 2 - 1、1 1 2 - 2、1 1 2 - n を生成し得るが、ここで「 n 」は、システム 1 0 0 に関して記載されるように、0 を超える任意の正の整数である。R N A リードは一例として使用されるが、システムはまた、D N A リード上で同じプロセスを実施し得る。配列決定デバイス 1 1 0 は、メモリ 1 2 0 内にリード 1 1 2 - 1、1 1 2 - 2、1 1 2 - n を記憶し得る。いくつかの実装形態では、リード 1 1 2 - 1、1 1 2 - 2、1 1 2 - n は、圧縮形式であってもよい。

【 0 0 8 0 】

二次解析ユニット 1 3 0 は、リード 1 1 2 - 1、1 1 2 - 2、1 1 2 - n を取得し、二次解析ユニット 1 3 0 のメモリ 1 3 2 内にリード 1 1 2 - 1、1 1 2 - 2、1 1 2 - n を記憶し得る。いくつかの実装形態では、これは、配列決定デバイス 1 1 0 の制御プログラムを含んで、リード 1 1 2 - 1、1 1 2 - 2、1 1 2 - n を二次解析ユニット 1 3 0 のメモリ 1 3 2 内へとストリーミングし得る。その他の実装形態では、二次解析ユニット 1 3

10

20

30

40

50

0 は、リード 1 1 2 - 1、1 1 2 - 2、1 2 2 - n を要求し得る。リード 1 1 2 - 1、1 1 2 - 2、1 1 2 - n が圧縮される場合、二次解析ユニット 1 3 0 のプログラマブル論理デバイス 1 3 4 は、展開ユニット 1 3 8 として状態 B へと構成されて、リード 1 1 2 - 1、1 1 2 - 2、1 1 2 - n を展開するために使用され得る。次に、プログラマブル論理デバイス 1 3 4 は、リード配列ユニットとして状態 A に再構成されて、リード 1 1 2 - 1、1 1 2 - 2、1 1 2 - n を参照配列に配列させるために使用され得る。

【0081】

二次解析ユニット 1 3 0 は、圧縮ユニットとして状態 B に再構成されて、圧縮ユニットを使用して配列されたリードを圧縮して、コンピュータ 3 2 0 への送信のために配列されたリードを準備し得る。本実施例では、配列されたリードの第 1 のバッチの圧縮は、配列されたリードのみならず、遺伝子融合解析に使用される配列されたリードに関連するリード配列ユニット 1 3 6 によって生成されたデータもまた圧縮することを含む。本データは、図 1 のシステム 1 0 0 に関して説明されており、例えば、変異体対立遺伝子頻度数、特異リード配列数、転写物にわたるリード確率、MAPQ スコア、親遺伝子間の相同性を示すデータ、又はこれらの組み合わせを含み得る。加えて、配列されたリードの第 1 のバッチ内へと圧縮され得るその他のデータは、(i) 融合候補のリードと、(i i) 融合候補のリードが配列された参照配列位置の部分と、(i i i) 特定の遺伝子融合候補が配列された参照ゲノムのセグメントのアノテーションと、を含み得る。いくつかの実装形態では、アノテーションは、遺伝子エクソンのアノテーション、相同遺伝子の存在を示すアノテーション、富化遺伝子のリストを示すアノテーション、又はこれらの組み合わせを含み得る。

【0082】

配列されたリードを圧縮した後、二次解析ユニット 1 3 0 は、メモリ 1 2 0 内に圧縮されたリードの第 1 のバッチを記憶し得る。次に、配列決定デバイス 1 1 0 は、遺伝子融合解析のために、ネットワーク 3 1 0 にわたって、配列されたリードの第 1 のバッチ 1 2 5 をコンピュータ 3 2 0 に送信し得る。ネットワーク 3 1 0 は、1 つ以上の有線ネットワーク、1 つ以上の無線ネットワーク、又はこれらの組み合わせを含み得る。異なる実装形態では、ネットワーク 3 1 0 は、有線イーサネット、有線光ネットワーク、LAN、WAN、セルラーネットワーク、インターネット、又はこれらの組み合わせのうちの 1 つ以上であってもよい。いくつかの実装形態では、コンピュータ 3 2 0 は、リモートクラウドサーバであり得る。しかしながら、他の実装形態では、コンピュータ 3 2 0 は、直接イーサネット接続、USB - C 接続等の直接接続を介して、配列決定デバイス 1 1 0 に接続され得る。図 3 0 0 の本実施例では、リードの第 1 のバッチは通信前に圧縮されるが、圧縮が使用される必要はない。その代わりに、圧縮は、ネットワーク帯域幅の消費を低減し、記憶コストを最小化する方法として提供されるが、これは、ゲノムの大きなデータサイズに対処する場合に、著しい技術的利益を提供し、コストを削減し得る。

【0083】

いくつかの実装形態では、配列されたリードの第 1 のバッチは、サンプル 1 0 5 に対して生成されたリードのセット全体を含む。その他の実装形態では、配列されたリードの第 1 のバッチは、サンプル 1 0 5 に対して生成されたリードのセット全体の一部分のみであり、バッチ処理システムを使用して、平行処理を容易にし得る。例えば、いくつかの実装形態では、二次解析ユニットがメモリ 1 2 0 内で配列されたリードの第 1 のバッチを記憶した後、二次解析ユニット 1 3 0 は、メモリ 1 3 2 内に記憶するために、まだ配列されていないリードの第 2 のバッチを取得する。次に、リードの第 2 のバッチが圧縮された場合、二次解析ユニット 1 3 0 は展開を実行し得、コンピュータ 3 2 0 がリードの第 1 のバッチの遺伝子融合解析を実施している間に、リードの第 2 のバッチの配列を実行し得る。リードのバッチ処理によって容易になるこのような平行処理は、サンプル 1 0 5 のリードのための有効な遺伝子融合を決定するために必要とされるシステム 3 0 0 の実行時間を著しく低減し得る。

【0084】

コンピュータ 320 は、ネットワーク 310 を介してリードの第 1 のバッチ 125 を受信し、メモリ 320 内にリードの第 1 のバッチを記憶し得る。リードの第 1 のバッチ 125 が圧縮された場合、コンピュータ 320 は、圧縮 / 展開モジュール 325 を使用して、リードの第 1 のバッチを展開し、メモリ 320 内にリードの第 1 のバッチを記憶し得る。次に、コンピュータ 320 は、図 1 のシステム 100 に関して説明したのと同じ様式で、融合候補同定モジュール 140 と、融合候補フィルタリングモジュール 150 と、特徴セット生成モジュール 160 と、機械学習モデル 170 と、遺伝子融合決定モジュール 180 と、出力 API モジュール 190 と、の遺伝子融合解析パイプラインを実行し得る。

【0085】

出力 192 は、ネットワーク 310 を介して多数の異なるデバイスに提供され得る。例として、出力データは、シーケンサのディスプレイ 195 上で出力するために配列決定デバイスに送信され得る。あるいは又は加えて、出力 192 は、ネットワーク 310 を介してユーザーデバイス 330 のディスプレイ上に表示するために提供され得る。ユーザーデバイス 330 は、スマートフォン、タブレットコンピュータ、ラップトップコンピュータ、デスクトップコンピュータ、又はディスプレイを備える任意の他のコンピュータを含み得る。あるいは又は加えて、出力 192 はまた、ネットワーク 310 を介してプリンタ 340 を介して出力するために提供され得る。このような実装形態では、出力は、決定された有効な遺伝子融合のハードコピーレポートであってもよい。

【0086】

図 4 は、遺伝子融合の迅速な検出のためのシステムを実行するために使用され得るシステム構成要素のブロック図である。

【0087】

コンピューティングデバイス 400 は、ラップトップ、デスクトップ、ワークステーション、パーソナルデジタルアシスタント、サーバ、ブレードサーバ、メインフレーム、及び他の適切なコンピュータなどの種々の形態のデジタルコンピュータを表すことが意図されている。コンピューティングデバイス 450 は、パーソナルデジタルアシスタント、携帯電話、スマートフォン、及び他の同様のコンピューティングデバイスなどの種々の形態のモバイルデバイスを表すことが意図されている。これに加えて、コンピューティングデバイス 400 又は 450 は、ユニバーサルシリアルバス (USB) フラッシュドライブを含み得る。USB フラッシュドライブは、オペレーティングシステム及び他のアプリケーションを記憶し得る。USB フラッシュドライブは、別のコンピューティングデバイスの USB ポートに挿入できる無線送信機又は USB コネクタなどの入力 / 出力構成要素を含み得る。本明細書に示される構成要素、この構成要素の接続及び関係、並びにこの構成要素の機能は、単なる例であることを意味し、本文書に記載及び / 又は特許請求される発明の実装形態を限定することを意味するものではない。

【0088】

コンピューティングデバイス 400 は、プロセッサ 402 と、メモリ 404 と、記憶デバイス 406 と、メモリ 404 及び高速拡張ポート 410 に接続する高速インターフェース 408 と、低速バス 414 及び記憶デバイス 408 に接続する低速インターフェース 412 と、を含む。402、404、406、408、410、及び 412 の各構成要素は、種々のバスを使用して相互接続されており、共通のマザーボード上に、又は適切な他の様式で装着され得る。プロセッサ 402 は、メモリ 404 又は記憶デバイス 408 上に記憶された命令を含むコンピューティングデバイス 400 内での実行のための命令を処理して、高速インターフェース 408 に結合されたディスプレイ 416 などの外部入力 / 出力デバイス上の GUI に関するグラフィカル情報を表示し得る。他の実装形態では、複数のプロセッサ及び / 又は複数のバスを、適切な複数のメモリ及び複数の種類のメモリと共に使用し得る。また、複数のコンピューティングデバイス 400 を接続して、各デバイスが、例えば、サーババンク、ブレードサーバ群、又はマルチプロセッサシステムとして、必要な操作の部分を提供するようにすることができる。

【0089】

10

20

30

40

50

メモリ４０４は、コンピューティングデバイス４００内に情報を記憶する。一実装形態では、メモリ４０４は、揮発性メモリユニット又は複数の揮発性メモリユニットである。別の実装形態では、メモリ４０４は、不揮発性メモリユニット又は複数の不揮発性メモリユニットである。メモリ４０４を、磁気ディスク又は光ディスクなどの別の形態のコンピュータ可読媒体とすることもできる。

【００９０】

記憶デバイス４０８は、コンピューティングデバイス４００のための大規模な記憶機構を提供することができる。一実装形態では、記憶デバイス４０８は、フロッピーディスクデバイス、ハードディスクデバイス、光ディスクデバイス、若しくはテープデバイス、フラッシュメモリ若しくはその他の類似のソリッドステートメモリデバイス、又はストレージエリアネットワーク若しくはその他の構成内のデバイスを含むデバイスのアレイなどのコンピュータ可読媒体であり得るか、又はそれを内包し得る。コンピュータプログラム製品は、情報キャリア内で有形に具現化され得る。コンピュータプログラム製品はまた、実行された場合、上述したものなどの１つ以上の方法を実行する命令を内包し得る。情報キャリアは、メモリ４０４、記憶デバイス４０８、又はプロセッサ４０２上のメモリなどのコンピュータ可読媒体又は機械可読媒体である。

【００９１】

高速コントローラ４０８は、コンピューティングデバイス４００の帯域幅集約操作を管理する一方、低速コントローラ４１２は、低帯域幅集約操作を管理する。このような機能の割り当ては、一実施例に過ぎない。一実装形態では、高速コントローラ４０８は、例えば、グラフィックプロセッサ又はアクセラレータを介してメモリ４０４、ディスプレイ４１６に、及び種々の拡張カード（図示せず）を受容し得る高速拡張ポート４１０に連結されている。本実装形態では、低速コントローラ４１２は、記憶デバイス４０８及び低速拡張ポート４１４に連結されている。種々の通信ポート、例えば、ＵＳＢ、Bluetooth、イーサネット、無線イーサネットを含み得る低速拡張ポートは、例えば、ネットワークアダプタを介して、キーボード、ポインティングデバイス、マイクロフォン／スピーカペア、スキャナ、又はスイッチ若しくはルータなどのネットワーキングデバイスなどの１つ以上の入力／出力デバイスに連結され得る。コンピューティングデバイス４００は、図に示されるように、多数の異なる形態で実装され得る。例えば、コンピューティングデバイス４００は、標準サーバ４２０として、又はこのようなサーバのグループ内で複数回、実装され得る。コンピューティングデバイス４００はまた、ラックサーバシステム４２４の一部として実装され得る。加えて、コンピューティングデバイス４００は、ラップトップコンピュータ４２２などのパーソナルコンピュータに実装され得る。これに代えて、コンピューティングデバイス４００からの構成要素を、デバイス４５０などのモバイルデバイス（図示せず）内のその他の構成要素と組み合わせることができる。このようなデバイスのそれぞれは、コンピューティングデバイス４００、４５０のうちの１つ以上を内包し得、システム全体を、互いに通信する複数のコンピューティングデバイス４００、４５０から構成することができる。

【００９２】

コンピューティングデバイス４００は、図に示されるように、多数の異なる形態で実装され得る。例えば、コンピューティングデバイス４００は、標準サーバ４２０として、又はこのようなサーバのグループ内で複数回、実装され得る。コンピューティングデバイス４００はまた、ラックサーバシステム４２４の一部として実装され得る。加えて、コンピューティングデバイス４００は、ラップトップコンピュータ４２２などのパーソナルコンピュータに実装され得る。これに代えて、コンピューティングデバイス４００からの構成要素を、デバイス４５０などのモバイルデバイス（図示せず）内のその他の構成要素と組み合わせることができる。このようなデバイスのそれぞれは、コンピューティングデバイス４００、４５０のうちの１つ以上を内包し得、システム全体を、互いに通信する複数のコンピューティングデバイス４００、４５０から構成することができる。

【００９３】

コンピューティングデバイス 450 は、構成要素の中でもとりわけ、プロセッサ 452 と、メモリ 464 と、ディスプレイ 454、通信インターフェース 466、及び送受信機 468 などの入出力デバイスと、を含む。デバイス 450 はまた、追加の記憶機構を提供するために、マイクロドライブ又はその他のデバイスなどの記憶デバイスを備え得る。構成要素 450、452、464、454、466、及び 468 のそれぞれは、種々のバスを使用して相互接続されており、構成要素のうちのいくつかは、共通のマザーボード上に、又は適切なその他の様式で装着され得る。

【0094】

プロセッサ 452 は、メモリ 464 に記憶された命令を含む、コンピューティングデバイス 450 内の命令を実行し得る。プロセッサは、別個及び複数のアナログ及びデジタルプロセッサを含むチップのチップセットとして実装できる。これに加えて、プロセッサは、いくつかのアーキテクチャのうちのいずれかを使用して実装できる。例えば、プロセッサ 410 は、CISC（複合命令セットコンピュータ）プロセッサ、RISC（縮小命令セットコンピュータ）プロセッサ、又はMISC（最小命令セットコンピュータ）プロセッサであり得る。プロセッサは、例えば、ユーザーインターフェースの制御、デバイス 450 によって実行されるアプリケーション、及びデバイス 450 による無線通信などの、デバイス 450 の他の構成要素の協調を提供することができる。

【0095】

プロセッサ 452 は、制御インターフェース 458 と、ディスプレイ 454 に連結されたディスプレイインターフェース 456 と、を介して、ユーザーと通信し得る。ディスプレイ 454 は、例えば、TFT（薄膜トランジスタ液晶ディスプレイ）ディスプレイ、OLED（有機発光ダイオード）ディスプレイ、又はその他の適切なディスプレイ技術であり得る。ディスプレイインターフェース 456 は、ディスプレイ 454 を駆動してグラフィカル情報及びその他の情報をユーザーに提示するための適切な回路を含み得る。制御インターフェース 458 は、ユーザーから指令を受信して、この指令をプロセッサ 452 への発信用に変換し得る。加えて、デバイス 450 とその他のデバイスとの近接領域通信を可能にするために、プロセッサ 452 と通信する外部インターフェース 462 を提供し得る。外部インターフェース 462 は、例えば、いくつかの実装形態では有線通信を、又は他の実装形態では無線通信を提供し得るが、複数のインターフェースもまた使用し得る。

【0096】

メモリ 464 は、コンピューティングデバイス 450 内に情報を記憶する。メモリ 464 は、コンピュータ可読媒体（単数）若しくはコンピュータ可読媒体（複数）、揮発性メモリユニット（単数）若しくは揮発性メモリユニット（複数）、又は不揮発性メモリユニット（単数）若しくは不揮発性メモリユニット（複数）のうちの 1 つ以上として、実装され得る。また、例えばSIMM（シングルインラインメモリモジュール）カードインターフェースを含み得る、拡張インターフェース 472 を介して、デバイス 450 に拡張メモリ 474 を提供及び接続し得る。このような拡張メモリ 474 は、デバイス 450 のための増設記憶空間を提供し得るか、又はデバイス 450 のためのアプリケーション若しくはその他の情報もまた記憶し得る。具体的には、拡張メモリ 474 は、上述したプロセスを実行又は補完する命令を含み得るが、機密情報もまた含み得る。したがって、例えば、拡張メモリ 474 は、デバイス 450 のためのセキュリティモジュールとして提供され得るが、デバイス 450 の安全な使用を可能にする命令を用いてプログラムされ得る。加えて、安全なアプリケーションは、ハッキング不能な様式でSIMMカード上に識別情報を配置するなど、追加情報と共に、SIMMカードを介して提供され得る。

【0097】

メモリは、例えば、後述するように、フラッシュメモリ及び/又はNVRAMメモリを含み得る。一実装形態では、コンピュータプログラム製品は、情報キャリア内で有形に具現化され得る。コンピュータプログラム製品は、実行された場合、上述したものなどの 1 つ以上の方法を実行する命令を内包する。情報キャリアは、例えば、送受信機 468 又は外部インターフェース 462 を介して受信し得るメモリ 464、拡張メモリ 474、若し

10

20

30

40

50

くはプロセッサ 4 5 2 上のメモリなどの、コンピュータ可読媒体又は機械可読媒体である。

【 0 0 9 8 】

デバイス 4 5 0 は、必要に応じてデジタル信号処理回路を含み得る通信インターフェース 4 6 6 を介して、無線通信し得る。通信インターフェース 4 6 6 は、とりわけ G S M 音声通話、S M S、E M S、若しくは M M S メッセージング、C D M A、T D M A、P D C、W C D M A (登録商標)、C D M A 2 0 0 0、若しくは G P R S などの種々のモード又はプロトコル下での通信を提供し得る。このような通信は、例えば、高周波送受信機 4 6 8 を介して行うことができる。加えて、B l u e t o o t h、W i - F i、又はその他のこのような送受信機 (図示せず) を使用するなど、短距離通信を行うことができる。加えて、G P S (全地球測位システム) 受信機モジュール 4 7 0 が、デバイス 4 5 0 に追加のナビゲーション関連及び位置関連の無線データを提供し得るが、これは、デバイス 4 5 0 上で作動するアプリケーションによって適切に使用され得る。

10

【 0 0 9 9 】

デバイス 4 5 0 はまた、音声符復号器 4 6 0 を使用して可聴的に通信し得るが、これは、ユーザーから発話情報を受信し、この発話情報を使用可能なデジタル情報へと変換し得る。音声符復号器 4 6 0 は同様に、例えば、デバイス 4 5 0 のハンドセット内のスピーカを介してなど、ユーザーのための可聴音を発生させ得る。このような音は、音声電話通話からの音を含み得るが、録音された音、例えば、音声メッセージ、音楽ファイル等などを含み得、また、デバイス 4 5 0 上で操作するアプリケーションによって生成される音もまた含み得る。

20

【 0 1 0 0 】

コンピューティングデバイス 4 5 0 は、図に示されるように、多数の異なる形態で実装され得る。例えば、コンピューティングデバイス 4 5 0 は、携帯電話 4 8 0 として実装され得る。また、コンピューティングデバイス 4 5 0 を、スマートフォン 4 8 2、携帯情報端末、又はその他の類似したモバイルデバイスの一部として実装し得る。

【 0 1 0 1 】

本明細書に記載されたシステム及び方法の種々の実装形態は、デジタル電子回路、集積回路、専用に設計された A S I C (特定用途向け集積回路)、コンピュータハードウェア、ファームウェア、ソフトウェア、及び / 又はこのような実装形態の組み合わせで実現され得る。これらの種々の実装形態は、少なくとも 1 つのプログラマブル・プロセッサを含むプログラマブルシステム上で実行可能及び / 解釈可能な 1 つ以上のコンピュータプログラムにおける実装形態を含み得るが、これは、専用又は汎用であり得、記憶システム、少なくとも 1 つの入力デバイス、及び少なくとも 1 つの出力デバイスからデータ及び命令を受信し、かつこれらにデータ及び命令を送信するように連結される。

30

【 0 1 0 2 】

これらのコンピュータプログラム (プログラム、ソフトウェア、ソフトウェアアプリケーション、又はコードとしても知られている) は、プログラマブル・プロセッサのための機械命令を含み、高水準言語 (high-level procedural language) 及び / 又はオブジェクト指向プログラミング言語で、及び / 又はアセンブリ言語 / 機械言語で実装され得る。本明細書で使用される場合、用語「機械可読媒体」「コンピュータ可読媒体」とは、任意のコンピュータプログラム製品、装置、及び / 又はデバイス、例えば、磁気ディスク、光ディスク、メモリ、機械命令及び / 又はデータをプログラマブル・プロセッサに提供するために使用されるプログラマブル論理デバイス (P L D) を意味し、機械可読信号として機械命令を受信する機械可読媒体を含む。用語「機械可読信号」とは、プログラマブル・プロセッサに機械命令及び / 又はデータを提供するために使用される任意の信号を意味する。

40

【 0 1 0 3 】

ユーザーとの相互作用を提供するために、本明細書に記載されたシステム及び技術は、ユーザーに情報を表示するためのディスプレイデバイス、例えば、C R T (陰極線管) 又は L C D (液晶ディスプレイ) モニタ、並びにユーザーがコンピュータに入力を提供し得

50

るポインティングデバイス、例えばマウス又はトラックボールを有するコンピュータ上に実装され得る。同様に他の種類のデバイスを使用して、ユーザーとの相互作用を提供し得るが、例えば、ユーザーに提供されるフィードバックは、任意の形態の感覚フィードバック、例えば、視覚フィードバック、聴覚フィードバック、又は触覚フィードバックであり得、ユーザーからの入力、音響入力、発話入力、又は触覚入力を含む任意の形態で受信され得る。

【0104】

本明細書に記載されるシステム並びに技術は、例えばデータサーバとしてのバックエンド構成要素を含むコンピューティングシステムで、若しくはミドルウェアコンポーネント、例えばアプリケーションサーバを含むコンピューティングシステムで、若しくはフロントエンド構成要素、例えばユーザーがそれらを介して本明細書に記載されたシステム及び技術の実装形態と相互作用し得るグラフィカル・ユーザー・インターフェース若しくはウェブブラウザを有するクライアントコンピュータを含むコンピューティングシステムで、又はこのようなバックエンド、ミドルウェア、若しくはフロントエンドの構成要素の任意の組み合わせで、実装できる。システムの構成要素は、デジタルデータ通信、例えば通信ネットワークの任意の形態又は媒体によって相互接続され得る。通信ネットワークの例としては、ローカルエリアネットワーク（「LAN」）、広域ネットワーク（「WAN」）、及びインターネットが挙げられる。

10

【0105】

コンピューティングシステムは、クライアント及びサーバを含み得る。クライアント及びサーバは、一般に、互いにリモートであり、典型的には、通信ネットワークを介して相互作用する。クライアントとサーバとの関係は、それぞれのコンピュータ上で作動し、かつ互いにクライアント - サーバ関係を有するコンピュータプログラムによって生じる。

20

他の実施形態

【0106】

いくつかの実施形態が記載されてきた。それにもかかわらず、本発明の趣旨及び範囲から逸脱することなく、種々の変更を行い得ることが理解されよう。加えて、図に描示される論理フローは、所望の結果を達成するために、示される特定の順序、又は連続的な順序を必要としない。加えて、記載されたフローからその他の工程を提供し得る、又は工程を排除し得る、及び記載されたシステムにその他の構成要素を追加し得る、又はそこから除去し得る。したがって、他の実施形態は、以下の特許請求の範囲内にある。

30

40

50

フロントページの続き

(72)発明者 ヨハン・フェリックス・ウィルヘルム・シュレジンジャー
 アメリカ合衆国・カリフォルニア・９２１２２・サン・ディエゴ・イルミナ・ウェイ・５２００
(72)発明者 ショーン・トゥルオン
 アメリカ合衆国・カリフォルニア・９２１２２・サン・ディエゴ・イルミナ・ウェイ・５２００
(72)発明者 ジョン・クーパー・ロディ
 アメリカ合衆国・カリフォルニア・９２１２２・サン・ディエゴ・イルミナ・ウェイ・５２００
(72)発明者 マイケル・ルーレ
 アメリカ合衆国・カリフォルニア・９２１２２・サン・ディエゴ・イルミナ・ウェイ・５２００
(72)発明者 セヴリーヌ・カトリュクス
 アメリカ合衆国・カリフォルニア・９２１２２・サン・ディエゴ・イルミナ・ウェイ・５２００
(72)発明者 ラミ・メヒオ
 アメリカ合衆国・カリフォルニア・９２１２２・サン・ディエゴ・イルミナ・ウェイ・５２００

 審査官 前田 侑香
(56)参考文献 米国特許出願公開第２０１６／００１９３４０（ＵＳ，Ａ１）
 米国特許第１０３５４７４７（ＵＳ，Ｂ１）
 Andrew McPherson, 外16名, “ deFuse: An Algorithm for Gene Fusion Discovery in Tumor
RNA-Seq Data ”, PLoS Computational Biology [online], 米国, PLoS, 2011年05月19日, V
olume 7, Issue 5, p. 1-16, [2024年08月20日検索], インターネット: URL: <https://journals.plos.org/ploscompbiol/article/file?id=10.1371/journal.pcbi.1001138&type=printable>
, DOI: 10.1371/journal.pcbi.1001138
 Francesco Abate, 外9名, “ Pegasus: a comprehensive annotation and prediction tool for det
ection of driver gene fusions in cancer ”, BMC Systems Biology [online], 英国, BioMed Ce
ntral, 2014年09月04日, Volume 8, article number 97, (2014), p. 1-14, [2024年08月20
日検索], インターネット: URL: <https://link.springer.com/article/10.1186/s12918-014-0097-z>
097-z, DOI: 10.1186/s12918-014-0097-z
(58)調査した分野 (Int.Cl., D B 名)
 G 0 6 Q 1 0 / 0 0 - 9 9 / 0 0
 G 1 6 B 5 / 0 0 - 9 9 / 0 0
 G 1 6 C 1 0 / 0 0 - 9 9 / 0 0