

(19)日本国特許庁(JP)

(12)特許公報(B2)

(11)特許番号  
特許第7689494号  
(P7689494)

(45)発行日 令和7年6月6日(2025.6.6)

(24)登録日 令和7年5月29日(2025.5.29)

(51)国際特許分類 F I  
G 1 6 H 50/20 (2018.01) G 1 6 H 50/20  
G 1 6 H 10/60 (2018.01) G 1 6 H 10/60

請求項の数 22 (全66頁)

(21)出願番号	特願2021-538761(P2021-538761)	(73)特許権者	521288448 テンパス・エーアイ・インコーポレイテッド アメリカ合衆国・イリノイ・60654 ・シカゴ・ウェスト・シカゴ・アベニュー・600・スイート・510
(86)(22)出願日	令和1年12月31日(2019.12.31)	(74)代理人	100108453 弁理士 村山 靖彦
(65)公表番号	特表2022-516172(P2022-516172 A)	(74)代理人	100110364 弁理士 実広 信哉
(43)公表日	令和4年2月24日(2022.2.24)	(74)代理人	100133400 弁理士 阿部 達彦
(86)国際出願番号	PCT/US2019/069149	(72)発明者	ヘイリー・レフコフスキー アメリカ合衆国・イリノイ・60614 ・シカゴ・ウェスト・アーミテージ・ア 最終頁に続く
(87)国際公開番号	WO2020/142551		
(87)国際公開日	令和2年7月9日(2020.7.9)		
審査請求日	令和4年12月28日(2022.12.28)		
(31)優先権主張番号	62/786,739		
(32)優先日	平成30年12月31日(2018.12.31)		
(33)優先権主張国・地域又は機関	米国(US)		

(54)【発明の名称】 患者コホートの反応、増悪、および生存を予測し解析するための方法およびプロセス

(57)【特許請求の範囲】

【請求項1】

複数の対象について応答、増悪、または生存を予測するための予測モデルを実装するための、コンピュータシステムによって実施される方法であって、

複数の対象に対する複数のデータを受け取るステップであって前記複数のデータは一定の時間期間にわたり、前記複数の対象は、前記予測モデルを訓練するための前記複数のデータ及び、複数の予測のセットを生成するための複数の第1の特徴、複数の第2の特徴及びそれらの各々の状態を有する被験者である、ステップと、

前記複数の対象によって共有される共通のアンカーポイントに関する前記複数のデータに関連付けられたタイムラインをアライメントするステップと、

前記複数の対象の各々について、前記時間期間内の複数の対象時点を識別するステップであって、前記複数の対象時点が前記共通のアンカーポイントの後の少なくとも1つの時点を含み、各対象時点が前記複数のデータの各データに関連付けられる、ステップと、

前記複数の対象のうちの各対象について、および前記複数の対象時点のうちの各対象時点について、前記複数の対象に対する前記複数のデータに基づき、

転帰事象に対する転帰標的が、前記対象時点から測定された水平時間窓内に生じるか否かを決定するステップであって、前記転帰事象が、対象及び/または疾患の状態に関するものであり、前記転帰標的が、前記転帰事象が生じるか否かを示す、ステップと、

複数の第1の特徴を含む第1の特徴セットを生成するステップであって、前記複数の第1の特徴が、前記対象時点に関連して決定された時間的特徴を含む、ステップと、

10

20

前記対象時点における前記複数の第1の特徴の各々の状態を決定するステップと、  
有効な転帰標的を有する前記複数の対象時点のうちの各対象時点について、および水平時間窓および転帰事象の各組合せについて、複数の第2の特徴を識別するステップであって、前記複数の第2の特徴が、前記対象時点に関連して決定された時間的特徴を含む、ステップと、

有効な転帰標的を有する各対象時点における、水平時間窓及び転帰事象の各組合せについて、前記複数の第2の特徴のそれぞれの状態を決定するステップと、

前記複数の第1の特徴、前記複数の第2の特徴及びそれらの各々の状態を含む前記複数のデータに基づき機械学習を使用して生成された前記予測モデルを用いて前記複数の対象に対する予測の複数のセットを生成するステップとを含む方法。

10

【請求項2】

1)前記複数の対象のグループを含む対象のコホートを選択するステップと、  
2)対象の前記グループの各々に関連付けられているアンカーポイントのセットから共通のアンカー時点を識別するステップであって、前記共通のアンカーポイントは前記コホート内の対象の前記グループの各々によって共有される、ステップと、

3)対象の前記グループの各対象について、対象の前記グループの各対象に関連付けられているタイムラインを前記共通のアンカーポイントにアライメントするステップと、

4)転帰標的を識別するステップと、

5)対象の前記グループの各対象について、ならびに前記複数の第2の特徴および前記複数の第1の特徴の各々について、各々予測標的値を含む予測の前記生成された複数のセットを取り出すステップと、

20

6)各々予測標的値を含む予測の前記複数のセットに基づき分類モデルを生成するステップとをさらに含む請求項1に記載の方法。

【請求項3】

前記分類モデルは、複数の決定木を含み、前記方法は、

7)前記複数の決定木を生成するステップであって、前記複数の決定木の各決定木について、

a)前記複数の第2の特徴および前記複数の第1の特徴の特徴毎に、

i)前記予測された標的値と実際の標的値との間の差に基づき対象の前記グループを第1のサブグループと第2のサブグループとに分割し、

30

ii)前記第1のサブグループ内の対象数と前記第2のサブグループ内の対象数との間の差を決定し、

b)前記第1のサブグループ内の対象数と前記第2のサブグループ内の対象数との間の最大の差である差を結果としてもたらす特徴を選択するステップを含む、ステップをさらに含む請求項2に記載の方法。

【請求項4】

8)前記第1のサブグループ内の対象数と前記第2のサブグループ内の対象数との間の前記最大の差を結果としてもたらす前記特徴に基づき木構造の新しいノードを作成するステップと、

9)前記第1のサブグループに基づき前記新しいノードから第1の枝を作成するステップと

40

と、  
10)前記第2のサブグループに基づき前記新しいノードから第2の枝を作成するステップと、

11)前記第1の枝および前記第2の枝の各々について、7) a) i-ii)または7) b)及び8) - 10)のステップを、前記第1のサブグループおよび前記第2のサブグループ内の対象に基づき、それぞれ、

最大数のノードまたは枝がすでに作成されているか、または

ノードが最小数より少ない数の対象を含む、のいずれかまで、繰り返すステップとをさらに含む請求項3に記載の方法。

【請求項5】

50

前記分類モデルは、各々複数の終端ノードを含む複数のモデルを含み、  
前記方法は、

対象の前記グループの各対象について、

前記複数の終端ノードの各々の中に出現する前記対象と対象の前記グループの他方の各々との同時出現を識別し、

モデルの数で除算された同時出現の総和に基づき前記対象に対する類似度メトリックを決定するステップと、

前記複数の対象の各々について前記類似度メトリックを決定することに基づき対象-対象間の類似度メトリックのデータベースを生成するステップとをさらに含む請求項2に記載の方法。

10

【請求項6】

各対象に関する前記類似度メトリックをクラスタプロットとして表示するステップをさらに含む請求項5に記載の方法。

【請求項7】

前記複数の第1の特徴及び前記複数の第2の特徴のうち少なくとも1つを識別するためにステップ7) a) i-ii)または7) b)のうち1つまたは複数に関連付けられているデータを表示するステップをさらに含む請求項3に記載の方法。

【請求項8】

対象の前記グループと異なる新しい対象に対する類似度メトリックを決定するステップと、

20

決定された前記類似度メトリックに基づき複数の終端ノードのうち特定の終端ノードに対応する対象のサブグループと新しい対象をマッチングさせるステップと、

前記新しい対象を対象の前記サブグループとマッチングさせることに基づき前記新しい対象に対する治療を識別するステップとをさらに含む請求項5に記載の方法。

【請求項9】

共有されている特徴を有する対象の特定の cohorts を識別するために次元削減アルゴリズムを使用して対象-対象類似度メトリックのデータベースを処理するステップをさらに含む請求項5に記載の方法。

【請求項10】

前記共有されている特徴は、共有されている第1の特徴または共有されている第2の特徴のうち少なくとも一方を含む請求項9に記載の方法。

30

【請求項11】

対象の cohorts を選択するステップは、

特定の状態を有する対象の cohorts を選択するステップをさらに含む請求項2に記載の方法。

【請求項12】

前記特定の状態は、病気を含む請求項11に記載の方法。

【請求項13】

前記共通のアンカーポイントは、診断、治療、または転移事象のうち少なくとも1つに関連付けられている時間を含む請求項2に記載の方法。

40

【請求項14】

コンピュータシステムによって実施される、複数の対象について応答、増悪、または生存を予測するための予測モデルを実装するための方法であって、

一定の時間期間にわたって複数の対象に対する複数のデータを受け取るステップであって、前記対象は、前記予測モデルを訓練するための前記複数のデータ及び、複数の予測のセットを生成するための複数の第1の特徴、複数の第2の特徴及びそれらの各々の状態を有する被験者である、ステップと、

前記複数の対象によって共有される共通のアンカーポイントに対する前記複数のデータに関連付けられたタイムラインをアライメントするステップと、

前記複数の対象の各々について、前記時間期間内の複数の対象時点を識別するステップ

50

であって、前記複数の対象時点が前記共通のアンカーポイントの後の少なくとも1つの時点を含み、各対象時点が前記複数のデータの各データに関連付けられる、ステップと、

前記複数の対象のうちの各対象について、および前記複数の対象時点のうちの各対象時点について、前記複数の対象に対する前記複数のデータに基づき、

転帰事象に対する転帰標的が、前記対象時点から測定された水平線時間窓内に生じるか否かを決定するステップであって、前記転帰事象が対象及び/または疾患の状態に関するものであり、前記転帰標的が、前記転帰事象が生じるか否かを示す、ステップと、

複数の第1の特徴を識別するステップであって、前記第1の特徴が、前記対象時点に関連して決定された時間的特徴を含む、ステップと、

前記対象時点における前記複数の第1の特徴の各々の状態を決定するステップと、

有効な転帰標的を有する前記複数の対象時点のうちの各対象時点について、および水平時間窓および転帰事象の各組合せについて、複数の第2の特徴を識別するステップであって、前記複数の第2の特徴が、前記対象時点に関連して決定された時間的特徴を含む、ステップと、

有効な転帰標的を有する各対象時点における、水平時間窓及び転帰事象の各組合せについて、前記複数の第2の特徴のそれぞれの状態を決定するステップと、

前記複数の第1の特徴、前記複数の第2の特徴及びそれらの各々の状態を含む前記複数のデータに基づき機械学習を使用して生成された前記予測モデルを用いて前記複数の対象に対する予測の複数のセットを生成するステップとを含む方法。

#### 【請求項15】

前記複数のデータに基づき機械学習を使用して前記予測モデルを生成するステップは、前記複数のデータに基づき勾配ブースティングを使用して前記予測モデルを生成するステップをさらに含む請求項1または14に記載の方法。

#### 【請求項16】

前記複数の対象に対する予測の前記複数のセットは、複数のフォールドに分割され、各フォールドは、前記複数の対象のサブセットに対応するデータを含み、

前記複数のデータに基づき勾配ブースティングを使用して前記予測モデルを生成するステップは、

前記複数のフォールドのサブセットに基づき勾配ブースティングを使用して前記予測モデルを生成するステップをさらに含む請求項15に記載の方法。

#### 【請求項17】

前記予測モデルから前記複数の対象時点の各々に対応する複数の予測を受け取るステップと、

前記複数の予測の各々を対応する転帰標的に関連付けるステップとをさらに含む請求項1または14に記載の方法。

#### 【請求項18】

前記複数の予測と、転帰標的と、前記転帰標的に対応する前記複数の第2の特徴のサブセットと、前記複数の対象のサブセットを含む対象のコホートとを受け取るステップと、

アンカーポイントを受け取るステップと、

前記アンカーポイントを有する前記コホート内の各対象について、前記複数の第2の特徴の選択されたサブセットと、前記複数の予測の各々と前記転帰標的との間の差とを有する前記予測モデルを提供するステップと、

前記複数の第2の特徴の選択されたサブセットの各特徴について、前記複数の予測の各々と前記転帰標的との間の最大の差を決定することに基づき分類モデルを生成するステップとをさらに含む請求項17に記載の方法。

#### 【請求項19】

前記分類モデルは、決定木を含み、

前記方法は、前記複数の予測の各々と前記転帰標的との間の前記最大の差を決定することに基づき前記決定木を生成するステップをさらに含む、

前記決定木は、複数の葉ノードおよび1つまたは複数の枝ノードを含み、

10

20

30

40

50

前記1つまたは複数の枝ノードの各々は、各々葉ノードまたは枝ノードを含む枝の対を含み、

前記決定木の前記複数の葉ノードの各々は、対象の前記コホートからの多数の対象を含む請求項18に記載の方法。

【請求項20】

前記複数の予測の各々と前記転帰標的との間の前記最大の差を決定することに基づき前記決定木を生成するステップは、

前記決定木を、前記複数の葉ノードのうちの1つの葉ノード内の対象の数が対象の最小数よりも小さくなるまで生成するステップをさらに含む請求項19に記載の方法。

【請求項21】

前記複数の予測の各々と前記転帰標的との間の前記最大の差を決定することに基づき前記決定木を生成するステップは、

前記決定木を、前記決定木のレベルの数がレベルの最大数に等しくなるまで生成するステップをさらに含む請求項19に記載の方法。

【請求項22】

前記1つまたは複数の枝ノードの各々は、各々葉ノードまたは枝ノードを含む枝の対を含み、

前記枝は、前記複数の第2の特徴の前記サブセットから選択された特徴に基づき形成される請求項19に記載の方法。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、患者コホートの反応、増悪、および生存を予測し解析するための方法およびプロセスに関する。

【背景技術】

【0002】

いくつかの医療分野、たとえば、癌の研究および治療の分野では、各患者に対する膨大な量のデータが生成され、および収集され得る。このデータは、患者の年齢、性別、身長、体重、喫煙歴、地理的な位置などの人口統計学的情報、および他の非医療情報を含み得る。データは、腫瘍の腫瘍型、配置、サイズ、および病期などの臨床的構成要素、さらには投薬、投与量、治療法、死亡率、および他の転帰/反応データを含む治療データも含み得る。さらに、より高度な解析では、遺伝子マーカー、突然変異を含む、患者および/または腫瘍に関するゲノム情報、さらには、プロテオーム、トランスクリプトーム、エピゲノム、メタボローム、マイクロバイオーム、および他の多層オミクスの分野を含む分野からの他の情報も含み得る。

【0003】

このようにデータが豊富にあるにもかかわらず、データを迅速に、効率的に、および包括的に編集し、解析するための意味のある方法が不足している。

【0004】

したがって必要とされているのは、これらの課題の1つまたは複数克服するユーザインターフェース、システム、および方法である。

【先行技術文献】

【特許文献】

【0005】

【文献】米国仮特許出願第62/746,997号

【文献】米国特許出願第16/289,027号

【文献】米国特許第10,395,772号

【文献】PCT国際出願第PCT/US19/56713号

【文献】米国特許出願第16/679,054号

【発明の概要】

10

20

30

40

50

## 【課題を解決するための手段】

## 【0006】

一態様において、特定の治療を施されたときの特定の患者母集団またはコホートの予想される反応を予測するためのシステムおよびユーザインターフェースが提供される。それらの予測を遂行するために、システムは、既存のデータセットを使用して、サンプル患者母集団、すなわち「コホート」を定義し、一般的な患者母集団の分布に対して、コホート内の注目する各属性を示す患者の分布における1つまたは複数の重要な変曲点を識別し、それによって、特定の患者母集団に対する予想生存率および/または反応の予測を標的にする。

## 【0007】

本明細書において説明されているシステムは、患者の臨床、分子、表現型、および反応データ内に出現するパターンの自動解析を通じて、また完全に一体化された反応性ユーザインターフェースを介してさらなる調査を可能にすることにより、治療上の有意性の洞察の発見を円滑にする。

## 【0008】

一実施形態において、本発明は、患者の外れ値グループを識別するための方法を提供し、これは、1)複数の患者を含む患者のコホートを選択することと、2)患者のコホートに対する平均生存率を計算することと、3)患者のコホートに関連付けられている複数の臨床的または分子的特性を選択することと、4)複数の特性のうちの各特性について、a)特性に関連付けられている複数のデータ値を識別し、b)特性に関連付けられている複数のデータ値のうちの各データ値について、i)患者のコホートを、複数の患者の第1のサブグループと第2のサブグループとに分割することを、複数の患者のうちの各患者が外れ値時間期間において生存したかどうかに基づき行い、ii)第1のサブグループ内の患者数と第2のサブグループ内の患者数との間の差を決定し、iii)第1のサブグループ内の患者数と第2のサブグループ内の患者数との間の最大の差である差を結果としてもたらずデータ値を選択することと、5)第1のサブグループ内の患者数と第2のサブグループ内の患者数との間の最大の差を結果としてもたらずデータ値に基づき木構造の新しいノードを作成することと、6)第1のサブグループに基づき新しいノードから第1の枝を作成することと、7)第2のサブグループに基づき新しいノードから第2の枝を作成することと、8)第1の枝および第2の枝の各々について、4)b)i~iii)および5)のステップを、第1のサブグループおよび第2のサブグループ内の患者に基づき、それぞれ、最大数のノードまたは枝がすでに作成されているか、またはノードが最小数より少ない数の患者を含むまで、繰り返すことと、9)患者の外れ値グループを含む少なくとも1つのノードを識別することとを含む。

## 【0009】

さらに別の実施形態では、本発明は、予測モデルを実装するための方法を提供し、これは、一定期間の複数の患者に対する複数のデータを受信することと、複数の患者の各々について、一定期間内の複数の患者時点(patient timepoint)を識別することと、複数の患者のうちの各患者について、および複数の患者時点のうちの各患者時点について、複数の患者に対する複数のデータに基づき水平時間窓内の転帰事象(outcome event)に対する転帰標的(outcome target)を計算し、複数の事前特徴(prior feature)を識別し、患者時点における複数の事前特徴のうちの各々の状態を決定することと、有効な転帰標的を有する複数の時点のうちの各患者時点について、および水平時間窓および転帰事象の各組合せについて、複数の前方特徴(forward feature)を識別することと、複数の事前特徴および複数の前方特徴に基づき複数の患者に対する予測の複数のセットを生成することとを含む。

## 【0010】

なおも別の実施形態において、本発明は、方法を提供し、これは、1人または複数の患者に対する患者情報を受信することと、受信された患者情報に少なくとも一部は基づき1人または複数の患者の各々に対して1つまたは複数の相互作用を識別することと、1つまたは複数の相互作用の各々における1つまたは複数の標的について、1つまたは複数の標的の各々が相互作用の発生の時間期間内に出現するかどうかを識別する1つまたは複数のタイ

10

20

30

40

50

ムラインメトリクスを生成することと、1つまたは複数のタイムラインメトリクスのうちの各タイムラインメトリクスについて、患者がその時間期間内に1つまたは複数のステータス特性を受け得るかどうかを識別することと、1つまたは複数のステータス特性に少なくとも一部は基づき1つまたは複数の標的の各々について標的予測モデルを訓練することと、1つまたは複数の標的の各々について標的予測モデルからの各患者に対する予測を、1つまたは複数のタイムラインメトリクスのそれぞれの1つまたは複数のタイムラインメトリクスと関連付けることとを含む。

【0011】

いくつかの実施形態において、方法は、1)複数の患者の患者のグループを含む患者のコホートを選択することと、2)患者のグループの各々に関連付けられているアンカーポイントのセットから共通のアンカー時点を識別することと、共通のアンカーポイントはコホート内の患者のグループの各々によって共有される、識別することと、3)患者のグループの各患者について、患者のグループの各患者に関連付けられているタイムラインを共通のアンカーポイントにアライメントすることと、4)転帰標的を識別することと、5)患者のグループの各患者について、ならびに複数の前方特徴および複数の事前特徴の各々について、各々予測標的値を含む予測の生成された複数のセットを取り出すことと、6)複数の決定木を生成することと、複数の決定木の各木の決定について、a)複数の前方特徴および複数の事前特徴の特徴毎に、i)予測された標的値と実際の標的値との間の差に基づき患者のグループを第1のサブグループと第2のサブグループとに分割し、ii)第1のサブグループ内の患者数と第2のサブグループ内の患者数との間の差を決定し、iii)第1のサブグループ内の患者数と第2のサブグループ内の患者数との間の最大の差である差を結果としてもたらす特徴を選択することとを含む、生成することと、7)第1のサブグループ内の患者数と第2のサブグループ内の患者数との間の最大の差を結果としてもたらす特徴に基づき木構造の新しいノードを作成することと、8)第1のサブグループに基づき新しいノードから第1の枝を作成することと、9)第2のサブグループに基づき新しいノードから第2の枝を作成することと、10)第1の枝および第2の枝の各々について、6)a)i~iii)および7)のステップを、第1のサブグループおよび第2のサブグループ内の患者に基づき、それぞれ、最大数のノードまたは枝がすでに作成されているか、またはノードが最小数より少ない数の患者を含むまで、繰り返すこととをさらに含み得る。

【0012】

他の実施形態において、方法は、複数の予測と、転帰標的と、転帰標的に対応する複数の前方特徴のサブセットと、複数の患者のサブセットを含む患者のコホートとを受信することと、アンカーポイントを受信することと、アンカーポイントを有するコホート内の各患者について、複数の前方特徴の選択されたサブセットと、複数の予測の各々と転帰標的との間の差とを有する予測モデルを提供することと、複数の前方特徴の選択されたサブセットの各特徴について、複数の予測の各々と転帰標的との間の最大の差を決定することに基づき決定木を生成することと、決定木は、複数の葉ノードおよび1つまたは複数の枝ノードを含み、1つまたは複数の枝ノードの各々は、その各々が葉ノードまたは枝ノードを含む枝の対を含み、決定木の複数の葉ノードの各々は、患者のコホートからの患者の数を含み、生成することとをさらに含み得る。

【0013】

本発明の前述の、および他の態様および利点は、次の説明から明らかになるであろう。本明細書では、本明細書の一部を形成する添付図面を参照し、その中に本発明の好ましい実施形態が示されている。しかしながら、そのような実施形態は、必ずしも、本発明の全範囲を代表せず、したがって、本発明の範囲を解釈するために本明細書の請求項を参照する。

【0014】

本開示のさらなる目的、特徴、および利点は、本開示の例示的な実施形態を示す添付図と併せて次の詳細な説明を読むと明確になるであろう。

【図面の簡単な説明】

10

20

30

40

50

【 0 0 1 5 】

【 図 1 】 患者コホートの反応、増悪、および生存を予測し、解析するためのバックエンドおよびフロントエンドコンポーネントの例示的なシステム図である。

【 図 2 】 患者コホート選択フィルタ処理インターフェースの一例を示す図である。

【 図 3 】 コホートファネル(cohort funnel)および母集団解析ユーザインターフェースの一例を示す図である。

【 図 4 】 コホートファネルおよび母集団解析ユーザインターフェースの別の例を示す図である。

【 図 5 】 コホートファネルおよび母集団解析ユーザインターフェースの別の例を示す図である。

10

【 図 6 】 コホートファネルおよび母集団解析ユーザインターフェースの別の例を示す図である。

【 図 7 】 コホートファネルおよび母集団解析ユーザインターフェースの別の例を示す図である。

【 図 8 】 コホートファネルおよび母集団解析ユーザインターフェースの別の例を示す図である。

【 図 9 】 コホートファネルおよび母集団解析ユーザインターフェースの別の例を示す図である。

【 図 1 0 】 患者タイムライン解析ユーザインターフェース内のデータ要約ウィンドウの一例を示す図である。

20

【 図 1 1 】 患者タイムライン解析ユーザインターフェース内のデータ要約ウィンドウの別の例を示す図である。

【 図 1 2 】 患者タイムライン解析ユーザインターフェース内のデータ要約ウィンドウの別の例を示す図である。

【 図 1 3 】 患者タイムライン解析ユーザインターフェース内のデータ要約ウィンドウの別の例を示す図である。

【 図 1 4 】 患者タイムライン解析ユーザインターフェース内のデータ要約ウィンドウの別の例を示す図である。

【 図 1 5 】 患者生存率解析ユーザインターフェースの一例を示す図である。

【 図 1 6 】 患者生存率解析ユーザインターフェースの別の例を示す図である。

30

【 図 1 7 】 患者生存率解析ユーザインターフェースの別の例を示す図である。

【 図 1 8 】 患者生存率解析ユーザインターフェースの別の例を示す図である。

【 図 1 9 】 患者生存率解析ユーザインターフェースの別の例を示す図である。

【 図 2 0 】 患者生存率解析ユーザインターフェースの別の例を示す図である。

【 図 2 1 】 患者事象尤度解析ユーザインターフェースの一例を示す図である。

【 図 2 2 】 患者事象尤度解析ユーザインターフェースの別の例を示す図である。

【 図 2 3 】 患者事象尤度解析ユーザインターフェースの別の例を示す図である。

【 図 2 4 】 患者事象尤度解析ユーザインターフェースの別の例を示す図である。

【 図 2 5 A 】 患者事象尤度解析ユーザインターフェースに関して使用可能な外れ値を決定するための二分決定木の一例を示す図である。

40

【 図 2 5 B 】 患者事象尤度解析ユーザインターフェースに関して使用可能な外れ値を決定するための二分決定木の一例を示す図である。

【 図 2 6 】 関連付けられている増悪ウィンドウを有するアンカー事象のサンプルタイムラインの例を示す図である。

【 図 2 7 A 】 SAFEアルゴリズムの実施形態による適応型特徴ランキング(adaptive feature ranking)の例を示す図である。

【 図 2 7 B 】 SAFEアルゴリズムの実施形態による適応型特徴ランキングの例を示す図である。

【 図 2 7 C 】 SAFEアルゴリズムの実施形態による相関性のある特徴の取り扱いの一例を示す図である。

50

【図27D】SAFEアルゴリズムの実施形態によるサンプルレベル重要度割り当ての一例を示す図である。

【図27E】SAFEアルゴリズムの実施形態によるサンプルレベル重要度割り当ての一例を示す図である。

【図28】交差検証に患者フォールドを使用する一例を示す図である。

【図29】いくつかの実施形態による1つまたは複数のノートブックを介してアナリティクスを生成するためのインタラクティブ解析ポータルユーザインターフェースの一例を示す図である。

【図30】一実施形態による新しいワークブックを作成するためのインタラクティブ解析ポータルのワークブック生成インターフェースを例示する図である。

10

【図31】ノートブックユーザインターフェースのカスタムワークブックウィジェットからの事前構成されているテンプレートを開く動作を例示する図である。

【図32】ユーザがワークブックを表示ウィンドウ内にドラッグしたときのノートブックユーザインターフェースからの応答を例示する図である。

【図33】ユーザがワークブックをワークブックエディタにロードし、セルUIEから編集を選択した後の、カスタムワークブックのセル編集ビューを例示する図である。

【図34】本開示のいくつかの実装形態が動作し得るコンピュータシステムの一実装形態のブロック図の説明図である。

【発明を実施するための形態】

【0016】

20

添付図を参照し、特に図1を参照すると、患者コホートの応答、増悪、および生存を予測および解析するためのシステム10は、患者コホートタイムラインデータストレージ18と通信する患者コホートセレクトாமジュール16によってアクセス可能な患者データストア14を含むバックエンド層12を含み得る。患者コホートセレクトாமジュール16は、データストア14のオンデマンドフィルタ処理および解析を可能にするために、一例としてウェブブラウザを介して実装され得るインタラクティブ解析ポータル22を含むフロントエンド層20とインタラクティブにやり取りする。

【0017】

インタラクティブ解析ポータル22は、以下でより詳細に説明されるように、ユーザがデータストア14にクエリを実行し、データストア14の要素をフィルタ処理することを可能にするインタラクティブコホート選択フィルタ処理インターフェース24を備える複数のユーザインターフェースを具備し得る。以下でより詳細に説明されるように、ポータル22は、また、コホートファネルおよび母集団解析インターフェース26と、患者タイムライン解析ユーザインターフェース28と、患者生存率解析ユーザインターフェース30と、患者事象尤度解析ユーザインターフェース32とを含み得る。ポータル22は、さらに、患者次解析ユーザインターフェース34と、1つまたは複数の患者未来解析ユーザインターフェース36とを備え得る。

30

【0018】

再び図1を参照すると、バックエンド層12は、また、患者コホートタイムラインデータストレージ18からデータを受信して、複数のモジュールに入力を行う分散コンピューティングおよびモデリング層38を含むものとしてもよく、これは患者生存率解析ユーザインターフェース30を駆動する事象までの時間モデリングモジュール40と、患者事象尤度解析ユーザインターフェース32で受信された1つまたは複数の事象の尤度を、そのユーザインターフェースにその後表示するために計算する事象尤度モジュール42と、患者次事象解析ユーザインターフェース34にその後表示するために1つまたは複数の次事象のモデルを生成する次事象モデリングモジュール44と、1つまたは複数の患者未来解析ユーザインターフェース36にその後表示するために1つまたは複数の未来モデルを生成する1つまたは複数の未来モデリングモジュール46とを含む。

40

【0019】

患者データストア14は、人口統計、併存疾患、診断および再発、投薬、手術、ならびに

50

他の治療法などの、患者病歴を、その反応および副作用の詳細とともに含む既存のデータセットであってもよい。患者データストアは、患者に関係する患者遺伝子/分子シーケンシングおよび遺伝子突然変異の詳細、さらにはオルガノイドモデリングの結果も収め得る。一態様において、これらのデータセットは、1つまたは複数のソースから生成され得る。たとえば、システムを実装する機関は、それらの機関の記録のすべてから引き出すことができるものとしてよく、それらの機関に関わるすべての医師および/または患者からのすべての記録は、それらの機関の代理人、医師、研究者、または他の許可されたメンバーから利用可能であり得る。同様に、医師は、自分の記録のすべて、たとえば、自分の患者全員の記録から引き出すこともあり得る。代替的に、何人かのシステムユーザは、データセットの態様を購入するか、またはライセンスすることが、それらのユーザが十分に口バスタなデータセットに即時アクセスできないとき、それらのユーザがなおいっそう多くの記録を探しているとき、および/またはそれらのユーザが、いくつかの原発癌、発生部位および/もしくは診断部位毎の転移、発生部位、転移、または診断部位毎の再発などを有する患者を反映するデータなどの、特定のデータタイプを探しているときなどに、できる場合がある。

10

#### 【0020】

##### 特徴および特徴モジュール

患者データストアは、システム10内のすべての患者に対して利用可能な特徴のコレクションを備え得る1つまたは複数の特徴モジュールを含み得る。これらの特徴は、システム10内の人工知能分類器を生成し、モデル化するために使用され得る。すべての患者にわたる特徴範囲は情報的密度が高いが、患者の特徴セットは、すべての患者にわたるすべての特徴の集合的特徴範囲の全体にわたって要素がまばらに配置されているものとしてよい。たとえば、すべての患者にわたる特徴範囲は数万の特徴に拡大し得るが、患者の固有特徴セットは、その患者について利用可能な記録に基づき集合的特徴範囲の数百または数千個からなるサブセットのみを含み得る。

20

#### 【0021】

特徴コレクションは、患者の健康記録内で利用可能なフィールドの多様なセットを含み得る。臨床情報は、医師、看護師、または他の医療専門家もしくは代表者によって電子医療記録(EMR)または電子健康記録(EHR)に入力されたフィールドに基づくものとしてよい。遺伝子シーケンシングレポートからの分子フィールドなどの他の臨床情報は、他の情報源からキュレーションされ得る。シーケンシングは、次世代シーケンシング(NGS)を含むものとしてよく、ロングリード、ショートリード、または患者の体細胞および/または正常なゲノムをシーケンシングする他の形態であってもよい。追加の特徴モジュール内の機能の包括的コレクションは、診断、治療計画への反応、遺伝子プロファイル、臨床および表現型の特徴、および/または他の医学的、地理的、人口統計学的、臨床的、分子的、もしくは遺伝的な特徴を含み得る様々な医学分野にわたって様々な特徴を一緒に組み合わせ得る。たとえば、特徴のサブセットは、RNA特徴モジュールまたはDNA特徴モジュールシーケンシングに由来する特徴などの分子データ特徴を含み得る。

30

#### 【0022】

特徴の別のサブセット、イメージング特徴モジュールからのイメージング特徴は、染色されたH&EまたはIHCスライドのレビューなどの、病理学者のレビューによる標本のレビューを通じて識別される特徴を含み得る。別の例として、特徴のサブセットは、そのような特徴セットの個別の結果および組み合わせられた結果の解析から取得された派生的特徴を含み得る。DNAおよびRNAシーケンシングに由来する特徴は、シーケンシングされた組織中に存在するバリエーション科学モジュールからの遺伝子バリエーションを含み得る。遺伝的バリエーションのさらなる解析は、一塩基または複数塩基多型を識別するステップ、変異が挿入または欠失事象であるかどうかを識別するステップ、機能喪失または機能獲得を識別するステップ、融合を識別するステップ、コピー数変化を計算するステップ、マイクロサテライト不安定性を計算するステップ、腫瘍遺伝子変異量を計算するステップ、またはDNAおよびRNA内の他の構造的変化などの追加のステップを含み得る。H&E染色またはIHC染色の

40

50

ためのスライドの解析は、腫瘍浸潤、プログラム細胞死リガンド1(PD-L1)の状態、ヒト白血球抗原(HLA)の状態、または他の免疫学的特徴などの特徴を明らかにし得る。

#### 【0023】

構造化された、キュレーションされた、または電子的な医療もしくは健康記録から導出される特徴は、診断、症状、治療法、転帰などの臨床的特徴、患者名、生年月日、性別、民族、死亡日、住所、喫煙状態、癌、病気、疾患、糖尿病、鬱病、他の身体的または精神的な疾病の診断日などの患者人口統計、既往歴、家族病歴、初診日、転移診断日などの臨床診断、癌病期、腫瘍キャラクタリゼーション、原発組織、一連の療法、治療グループ、臨床試験、処方されるまたは服用される薬剤、外科手術、放射線療法、画像診断、副作用、関連する転帰などの治療および転帰、遂行度スコア、実験室検査、病理結果、予後指標などの遺伝子検査および実験室情報、遺伝子検査日、使用される検査提供者、遺伝子シーケンシング方法または遺伝子パネルなどの使用される検査方法、含まれる遺伝子、バリエーション、発現レベル/状態などの遺伝子結果、または上記のいずれかに対応する日付を含み得る。

10

#### 【0024】

特徴は、プロテオーム、トランスクリプトーム、エピゲノム、メタボローム、マイクロバイオーム、および他の多層オミクスの分野を含む、追加の医療または研究に基づくオミクス分野からの情報から導出され得る。オルガノイドモデリングラボから導出される特徴は、各オルガノイドと密接に関係するDNAおよびRNAシーケンシング情報、およびそれらのオルガノイドに適用される治療からの結果を含み得る。イメージングデータから導出される特徴は、染色されたスライドに関連付けられているレポート、腫瘍のサイズ、経時的な腫瘍サイズ差異をさらに含むものとしてよく、これは変化期間中の治療、さらにはイメージングデータからのPDL1ステータス、HLAステータス、または他の特性を分類するための機械学習アプローチを含む。他の特徴は、任意の新しい特徴および/または上記の特徴の組合せに少なくとも一部は基づき他の機械学習アプローチからの追加の派生特徴セットを含み得る。たとえば、イメージング結果は、追加のさらなるイメージング特徴を決定するためにRNA発現に導出されるMSI計算と組み合わせられる必要があり得る。別の例では、機械学習モデルは、患者の癌が特定の臓器に転移する可能性、または体内のさらに別の臓器への患者の将来の転移確率を生成し得る。医療情報から抽出され得る他の特徴も使用されてよい。何千もの特徴があり、また特徴のタイプの上記のリスティングは単に代表的なものにすぎず、特徴の完全なリスティングとして解釈されるべきでない。

20

30

#### 【0025】

変化モジュールは、特徴コレクションから非識別化患者特徴に関連付けられている変化特徴を生成する1つまたは複数のマイクロサービス、サーバ、スクリプト、または他の実行可能なアルゴリズムであってよい。変化モジュールは、特徴コレクションから入力を取り出し、保存する変化を提供し得る。例示的な変化モジュールは、変化モジュールのコレクションとして、次の変化のうちの1つまたは複数を含み得る。SNP(一塩基多型)モジュールは、ゲノム内の特定の位置に出現する一塩基の置換を識別するものとしてよく、各変化は、母集団内においてある程度目立って(たとえば、1%)存在する。たとえば、ヒトゲノムの特定の塩基位置、すなわち遺伝子座において、ほとんどの人々ではC-ヌクレオチドが出現し得るが、少数の人々ではA-ヌクレオチドが出現し得る。このことはこの特定の位置にSNPがあることを意味し、2つの可能なヌクレオチド変化、すなわちCまたはAはこの位置に対して対立遺伝子であると言われる。SNPは、広範な疾病に対するわれわれの脆弱性の違いを明らかにする(たとえば、鎌状赤血球貧血、 $\beta$ -サラセミア、嚢胞性線維症はSNPに起因する)。病気の重症度および身体が治療に反応する仕方は、遺伝的変異の現れでもある。たとえば、APOE(アポリポタンパク質E)遺伝子内単一塩基突然変異は、アルツハイマー病のより低いリスクに関連する。一塩基バリエーション(SNV)は、頻度の制限なしの一塩基の変異であり、体細胞で生じる可能性がある。体細胞一塩基変異(たとえば、癌によって引き起こされる)は、一塩基変化と呼ばれることもある。MNP(複数塩基多型)モジュールは、ゲノム内の特定の位置における連続するヌクレオチドの置換を識別し得る。InDelsモジュ

40

50

ールは、小さな遺伝的変異のうち分類される生命体のゲノム内の塩基の挿入または欠失を識別し得る。通常、長さが1から10,000個の塩基対を測定するが、マイクロインデルは、1から50個のヌクレオチドの正味の変化を結果としてもたらすインデルと定義される。インデルは、SNPまたは点突然変異と対比され得る。インデルは、配列からヌクレオチドを挿入し、削除するが、点変異は、DNA内の総数を変化させることなくヌクレオチドの1つを置き換える置換の一形態である。インデルは、挿入または欠失のいずれかであり、自然個体群の遺伝子マーカーとして、特に系統発生的研究において使用され得る。インデル頻度は、ホモポリマーおよびマイクロサテライトを含む、近い反復性の高い領域を除く、一塩基多型(SNP)の頻度よりも著しく低くなる傾向がある。MSI(マイクロサテライト不安定性)モジュールは、DNAミスマッチ修復遺伝子(MMR)欠陥の結果生じる遺伝的超変異性(突然変異を起こしやすい性質)を識別し得る。MSIの存在は、MMRが正常に機能していないことを示す表現型の証拠となる。MMRは、一塩基ミスマッチまたは短い挿入および欠失などの、DNA複製中に自然発生する誤りを修正する。MMRに参与するタンパク質は、ポリメラーゼ誤りを修正することを、DNAのミスマッチセクションに結合する複合体を形成し、誤りを削除し、その場所に正しい配列を挿入することによって行う。機能異常を示すMMRを有する細胞は、DNA複製中に発生する誤りを修正することができず、その結果、誤りを蓄積する。これは、新しいマイクロサテライト断片の作成を引き起こす。ポリメラーゼ連鎖反応に基づくアッセイで、これらの新しいマイクロサテライトを明らかにし、MSIの存在の証拠をもたらすことができる。マイクロサテライトは、DNAの反復配列である。これらの配列は、長さが1から6塩基対の繰り返し単位から作られ得る。これらのマイクロサテライトの長さは人によって変異が大きく、個人のDNA「指紋鑑定」に寄与するけれども、各個人は設定された長さのマイクロサテライトを有する。ヒトで最も一般的なマイクロサテライトは、C-ヌクレオチドとA-ヌクレオチドからなるジヌクレオチド反復であり、これはゲノム上に数万回出現する。マイクロサテライトは、単純配列反復(SSR)とも呼ばれる。TMB(腫瘍遺伝子変異量)モジュールは、腫瘍細胞によって伝えられる突然変異の測定を識別するものとしてよく、癌免疫(I-O)療法への反応との関連性を評価するために研究されている予測バイオマーカーである。TMBが高い腫瘍細胞は、より多くの新抗原を有し、それに伴って腫瘍内微小環境および周辺部内の抗癌T細胞が増加し得る。これらの新抗原は、T細胞によって認識され、抗腫瘍反応を引き起こし得る。近年、TMBは、メラノーマ、肺癌、膀胱癌を含む、様々な癌の間の免疫療法の潜在的反応を予測するのに助けることができる定量的マーカーとして浮上してきた。TMBは、腫瘍ゲノムのコーディング領域あたりの突然変異の総数として定義される。重要なのは、TMBは一貫して再現可能であるということである。これは、標的療法または免疫療法の選択、臨床試験における登録など、治療法決定を適切に伝えるために使用できる定量的な尺度を提供する。CNV(コピー数変化)モジュールは、遺伝子、バリエーション、対立遺伝子、またはヌクレオチド配列を解析することから、正常なゲノムからの逸脱、およびその後の影響を識別し得る。CNVは、反復、欠失、または逆位を含むヌクレオチドまたは塩基対のセクションに構造的変異が生じ得る現象である。融合モジュールは、2つの以前は別々だった遺伝子から形成されるハイブリッド遺伝子を識別し得る。これは、転座、中間部欠失、または染色体逆位の結果として生じ得る。遺伝子融合は、腫瘍形成において重要な役割を果たす。融合遺伝子は、非融合遺伝子に比べてかなり活性の高い異常タンパク質を産生することができるので腫瘍形成に寄与することができる。しばしば、融合遺伝子は癌を引き起こす癌遺伝子であり、BCR-ABL、TEL-AML1(すべてt(12;21)を有する)、AML1-ETO(t(8;21)を有するM2 AML)、およびしばしば前立腺癌内に出現する、21番染色体上の中間部欠失を伴うTMPRSS2-ERGを含む。TMPRSS2-ERGの場合、アンドロゲン受容体(AR)シグナリングを阻害し、発癌性ETS転写因子によるARの発現を抑制することによって、融合産物は前立腺癌を調節する。大半の融合遺伝子は、血液癌、肉腫、および前立腺癌から見つかる。BCAM-AKT2は、高悪性度漿液性卵巣癌に特異的で固有の融合遺伝子である。発癌性融合遺伝子は、2つの融合パートナーから新たな、または異なる機能を有する遺伝子産物を生み出し得る。代替的に、癌原遺伝子が強力なプロモーターに融合され、それによって、上流の融合パートナーの強

10

20

30

40

50

力なプロモーターによって引き起こされる上方調節によって発癌機能が機能するように設定される。後者はリンパ腫によく見られ、癌遺伝子が免疫グロブリン遺伝子のプロモーターに並列される。発癌性融合転写物は、トランススプライシングまたは読み過ごし事象によっても引き起こされ得る。染色体転座は新生組織形成にそのような重要な役割を果たしていることから、癌における染色体異常および遺伝子融合の専門データベースが作成されている。このデータベースは、Mitelman Database of Chromosome Aberrations and Gene Fusions in Cancerと呼ばれている。IHC(免疫組織化学)モジュールは、生体組織中の抗原に特異的に結合する抗体の原理を利用することによって組織切片の細胞中の抗原(タンパク質)を識別し得る。IHC染色は、癌性腫瘍に見られるような異常細胞の診断に広く使用されている。特定の分子マーカーは、増殖や細胞死(アポトーシス)などの特定の細胞的事象に特徴的である。また、IHCは、バイオマーカーの分布および局在ならびに生体組織の異なる部分で差別的に発現するタンパク質を理解するための基礎研究において広く使用されている。抗体-抗原相互作用を可視化することは、多数の方法で遂行され得る。最も一般的な事例では、抗体はペルオキシダーゼなどの酵素に結合され、免疫ペルオキシダーゼ染色の際に色生産反応を触媒することができる。代替的に、抗体は、また、免疫蛍光におけるフルオレセインやローダミンなどの発蛍光団にタグ付けされ得る。RNA発現データ、H&Eスライドイメージングデータ、または他のデータからの近似が生成され得る。療法モジュールは、成長および増殖を助け得る、癌細胞(または癌細胞の近くにある他の細胞)の差異と、これらの差異を「標的」とする薬物とを識別し得る。これらの薬物による治療は、標的療法と呼ばれる。たとえば、多くの標的薬物は、ほとんどの健康細胞をそのままにしながら、癌細胞を正常な健康細胞とは異なるものにしていく癌細胞の内部「プログラミング」を狙う。標的薬物は、癌細胞に成長と分裂を指令する化学信号を遮断するか、またはオフにするか、癌細胞が死滅するように癌細胞内のタンパク質を変化させるか、癌細胞に栄養を供給する新しい血管を作るのを停止するか、癌細胞を殺すように免疫システムをトリガーするか、または癌細胞を死滅させるために毒素を癌細胞に運ぶが、正常細胞には運ぶことがないものとしてよい。いくつかの標的薬物は、他の薬物に比べてより「標的」である。あるものは癌細胞の単一の変化のみを標的とし、他のものはいくつかの異なる変化に影響を及ぼし得る。他のものは、身体が癌細胞と戦う仕方をブーストする。これは、これらの薬物がどこに作用し、どのような副作用を引き起こすかに影響を及ぼし得る。標的療法のマッチングは、患者の療法標的を識別することと、他の包含または除外基準を満たすこととを含み得る。VUS(臨床的意義不明のバリエーション)モジュールは、コールされるが、コールの時点では病原性が良性かの分類ができないバリエーションを識別し得る。VUSは、良性または病原性として分類され得るかどうかを識別するためにVUSに関する出版物からカタログ化されてよい。試験モジュールは、患者の特徴と臨床試験とのマッチングによって特定の特性を有する癌を治療することに対する仮説を識別し、検証し得る。これらの試験は、出版物、試験レポート、または他の文書から取り込まれ、構造化され得る登録内容とマッチしなければならぬ包含および除外基準を有する。増幅モジュールは、他の遺伝子に関して不釣り合いにカウントが増える遺伝子を識別し得る。増幅は、カウントの大きい遺伝子が休眠状態になること、過活動になること、または別の予想外の様式で動作することを引き起こし得る。増幅は、遺伝子レベル、バリエーションレベル、RNA転写もしくは発現レベル、またはタンパク質レベルですら検出され得る。検出は、すべての異なる検出メカニズムまたはレベルで実行され、互いに突き合わせて検証され得る。アイソフォームモジュールは、複数のmRNA(アイソフォーム)がエクソンとイントロンの異なる組合せを通じて同じ遺伝子の転写産物から生成される生物学的プロセスである、選択的スプライシング(AS)を識別し得る。大規模ゲノミクス研究によって、哺乳類の遺伝子の30~60%が選択的スプライシングされていると推定される。1つの遺伝子に対する選択的スプライシングの可能なパターンは、非常に複雑であり得、その複雑さは遺伝子内のイントロンの数が増えるにつれて急激に増大する。コンピュータ内での選択的スプライシング予測は、ゲノム配列に対するmRNA配列の検索を通じてゲノム遺伝子座を識別し、ゲノム遺伝子座に対する配列を抽出して両端の配列を20kbまで伸長し、ゲノム配列を検索し(反復配列はマ

10

20

30

40

50

スクされている)、スプライシング対を抽出し(GT-AGコンセンサスを有するか、またはギャップの両端でアライメントされた2つよりも多い発現配列タグを有するアライメントギャップの2つの境界)、スプライシングペアをその座標に従ってアSEMBルし、遺伝子境界を決定し(スプライシングペア予測はこの点に合わせて生成される)、mRNA配列をゲノムテンプレートにアライメントすることによって予測遺伝子構造を生成し、スプライシングペア予測と遺伝子構造予測とを比較して選択的スプライスアイソフォームを見つけることによって、大部分のアライメントされた配列を共有するmRNAのセット内の大きな挿入または欠失を見つけ得る。パスウェイモジュールは、癌細胞がその悪性表現型に寄与するゲノム変化を蓄積することを可能にするDNA修復パスウェイ内の欠陥を識別し得る。癌性腫瘍は、遺伝毒性ストレスによって誘発された損傷を生き延びるために残存するDNA修復能力に依存しており、これは癌細胞内で孤立DNA修復パスウェイが不活性化されることを引き起こす。DNA修復パスウェイは、一般的に、異なる細胞周期の段階で異なるタイプの損傷を取り扱う相互排他的な仕組みのユニットと考えられている。しかしながら、最近の前臨床試験では、従来の複数のDNA修復パスウェイに関わる、多機能DNA修復ハブが、癌の中で頻繁に変化していることを示す強い証拠を提供している。影響を受け得るパスウェイを識別することは、患者の治療に関する重要な考慮事項をもたらす。ローカウントモジュールは、シーケンシングデータから検出されるバリエーションの数を識別し得る。DNAについては、これは、遺伝子内の特定のバリエーションに対応するシーケンシングからのリードの数となり得る。RNAについては、これは、シーケンシングからの遺伝子発現カウントまたはトランスクリプトームカウントであってもよい。

10

20

#### 【0026】

構造バリエーション分類は、特徴コレクションからの特徴、変化モジュールからの変化、および1つまたは複数の分類モジュールからのそれ自体の中からの他の分類を評価することを含み得る。構造バリエーション分類では、分類を記憶済み分類ストレージに提供し得る。例示的な分類モジュールは、CNVの分類を含むものとしてよく、この分類において、「報告可能」は、CNVが腫瘍癌キャラクターゼーション、病状、または薬理ゲノミクスに影響を及ぼすものとして1つまたは複数の参照データベースにおいて識別されていることを意味し、「報告不可能」は、CNVがそのようなものとして識別されていないことを意味し、「相矛盾する証拠」は、CNVが「報告可能」および「報告不可能」を示唆する両方の証拠を有することを意味し得る。さらに、治療関連度の分類は、CNVの検出(または非検出)の影響を受け得る療法の参照データセット言及から同様に確認される。他の分類は、機械学習アルゴリズム、ニューラルネットワーク、回帰法、グラフ技法、帰納的推論アプローチ、またはモジュール内の他の人工知能評価のアプリケーションを含み得る。臨床試験のための分類器は、有意であるかまたは報告可能であると識別されている変化モジュールから識別されるバリエーションの評価、包含および除外基準を識別するために利用可能なすべての臨床試験の評価、患者のバリエーションおよび他の情報を包含および除外基準にマッピングすること、および臨床試験を患者に適用可能である、または患者に適用不可能であるとして分類することを含み得る。同様の分類は、療法、機能喪失、機能獲得、診断、マイクロサテライト不安定性、腫瘍遺伝子変異量、インデル、SNP、MNP、融合、および変化モジュールの結果に基づき分類され得る他の変化に対して実行され得る。

30

40

#### 【0027】

特徴コレクション、変化モジュール、構造バリエーション、および特徴ストアの各々は、データベースに通信可能に結合され、処理および/または記憶のために各モジュール間でデータを転送し得る。別の実施形態では、特徴コレクション、変化モジュール、構造バリエーション、および特徴ストアの各々は、データベースを共有することなく独立した通信を行うように互いに通信可能に結合され得る。

#### 【0028】

上記の特徴および列挙されたモジュールに加えて、特徴モジュールは、サブモジュールとして、またはスタンドアロンモジュールとして、それぞれのモジュール内に次のモジュールの1つまたは複数を含み得る。

50

## 【 0 0 2 9 】

生殖細胞/体細胞DNA特徴モジュールは、患者または患者の腫瘍のDNA由来の情報に関連付けられている特徴コレクションを含み得る。これらの特徴は、FASTQ、BAM、VCF、または当技術分野で知られている他のシーケンシングファイルタイプに記憶されるような生のシーケンシング結果、遺伝子、突然変異、バリエーションコール、およびバリエーションキャラクター化を含み得る。患者の正常なサンプルからのゲノム情報は生殖細胞系として記憶され、患者の腫瘍サンプルからのゲノム情報は体細胞系として記憶され得る。

## 【 0 0 3 0 】

RNA特徴モジュールは、トランスクリプトーム情報などの、患者のDNA由来の情報に関連付けられている特徴コレクションを含み得る。これらの特徴は、生のシーケンシング結果、トランスクリプトーム発現、遺伝子、突然変異、バリエーションコール、およびバリエーションキャラクター化を含み得る。

10

## 【 0 0 3 1 】

メタデータモジュールは、ヒトゲノム、タンパク質構造、およびタンパク質構造に基づくエネルギー安定性の変化など、その効果に関連する特徴コレクションを含み得る。

## 【 0 0 3 2 】

臨床モジュールは、患者の臨床記録および患者の家族からの記録から導出される情報に関連付けられている特徴コレクションを含み得る。これらは、非構造化臨床文書、EMR、EHR、または患者病歴の他の情報源から抽出されてもよい。情報は、患者の症状、診断、治療、投薬、療法、ホスピス、治療への反応、実験室での検査結果、病歴、各々の地理的な位置、人口統計、または患者の医療記録に見られ得る患者の他の特徴を含み得る。治療、投薬、療法、および同様のものに関する情報は、推奨もしくは処方箋として、ならびに/またはそのような治療、投薬、療法、および同様のものなどが投与されたか、または服用されたことの確認として取り込まれ得る。

20

## 【 0 0 3 3 】

イメージングモジュールは、患者のイメージング記録から導出される情報に関連付けられている特徴コレクションを含み得る。イメージング記録は、H&Eスライド、IHCスライド、放射線画像、および様々な病気および疾患の診断および治療の過程で医師によってオーダーされ得る他の医療画像を含み得る。これらの特徴は、TMB、倍数性、純度、核-細胞質比、大核、細胞状態変化、生物学的パスウェイ活性化、ホルモン受容体変化、免疫細胞の浸潤、MMR、MSI、PDL1、CD3、FOXP3、HRD、PTEN、PIK3CAの免疫バイオマーカー、コラーゲンもしくは間質組成、外観、密度、または特性、腫瘍簇出、サイズ、悪性度、転移、免疫状態、クロマチン形態、および予後予測のための細胞、組織、もしくは腫瘍の他の特性を含み得る。

30

## 【 0 0 3 4 】

オミクスからのエピゲノムモジュールなどのエピゲノムモジュールは、DNA配列の変化ではなく、遺伝子の発現を調節するDNAの修飾に由来する情報に関連付けられている特徴コレクションを含み得る。これらの修飾は、患者が何を呼吸するか、食べるか、または飲むかに基づく環境要因の結果であることが多い。これらの特徴は、DNAメチル化、ヒストン修飾、または遺伝子内のヌクレオチドの配列を変更することなく遺伝子を不活性化するか、または遺伝子機能への変化を引き起こす他の要因を含み得る。

40

## 【 0 0 3 5 】

オミクスからのマイクロバイームモジュールなどのマイクロバイームモジュールは、患者のウイルスおよび細菌から導出された情報に関連付けられている特徴コレクションを含み得る。これらの特徴には、いくつかの病気の治療および診断に影響を及ぼし得るウイルス感染、さらには患者によって摂取された医薬品の有効性に影響を及ぼし得る患者の胃腸管内に存在する細菌を含み得る。

## 【 0 0 3 6 】

オミクスからのプロテオームモジュールなどのプロテオームモジュールは、患者体内で産生されるタンパク質から導出された情報に関連付けられている特徴コレクションを含み

50

得る。これらの特徴は、タンパク質の組成、構造、および活性、タンパク質がいつ、どこで発現されるか、タンパク質の産生率、分解率、および定常状態の存在量、タンパク質がどのように修飾されるか、たとえば、リン酸化などの翻訳後修飾、細胞内コンパートメント間のタンパク質の移動、代謝経路へのタンパク質の関与、タンパク質同士の相互作用、またはリン酸化、ユビキチン化、メチル化、アセチル化、グリコシル化、酸化、またはニトロシル化などの、RNAから翻訳された後のタンパク質の修飾を含み得る。

【0037】

追加のオミクスモジュールも、オミクスのすべての異なる分野に関連付けられている特徴コレクションなどの、オミクスに含まれるものとしてよく。これは、遺伝子プロファイルに関連付けられている認知プロセスの変化の研究を含む特徴のコレクションである、認知ゲノミクス、異なる生物学的種または株間のゲノム構造および機能の関係の研究を含む特徴のコレクションである、比較ゲノミクス、トランスクリプトームを含む遺伝子およびタンパク質の機能および相互作用の研究を含む特徴のコレクションである、機能ゲノミクス、遺伝子間、タンパク質間、またはタンパク質-リガンド間の相互作用を大規模に解析することに関する研究を含む特徴のコレクションである、インタラクトミクス、環境サンプルから直接回収された遺伝物質などのメタゲノムの研究を含む特徴のコレクションである、メタゲノミクス、神経系の発達および機能への遺伝的影響の研究を含む特徴のコレクションである、ニューロゲノミクス、所与の種内に見られる遺伝子ファミリーのコレクション全体の研究を含む特徴のコレクションである、パンゲノミクス、遺伝子型が判明した後に、個人の遺伝子型が公開されている文献と比較されて、形質発現の可能性および疾病リスクを決定し、オーダーメイド医療の提案を強化することができる、個人のゲノムのシーケンシングおよび解析に関するゲノミクスの研究を含む特徴のコレクションである、パーソナルゲノミクス、タンパク質およびRNAバインダー、代替DNA構造、ならびにDNA上の化学修飾を含む、ゲノムの構造をサポートすることの研究を含む特徴のコレクションである、エピゲノミクス、複雑な動的生物系として細胞核を形成する一揃いのゲノム成分の研究を含む特徴のコレクションである、ヌクレオミクス、患者によって産生された特定の脂質群に加えらるる修飾を含む、細胞脂質の研究を含む特徴のコレクションである、リピドミクス、患者によって産生された特定のタンパク質群に加えらるる修飾を含む、タンパク質の研究を含む特徴のコレクションである、プロテオミクス、免疫反応に関わる大規模なタンパク質群の研究を含む特徴のコレクションである、イムノプロテオミクス、タンパク質発現研究のためのプロテオミクス質量解析データの使用を含む食事の栄養成分および非栄養成分の分子標的を識別することに関する研究を含む特徴のコレクションである、ニュートリプロテオミクス、遺伝子アノテーションを識別するデータを含むプロテオミクスとゲノミクスとの交わりにおける生物学的研究の研究を含む、特徴のコレクションである、プロテオゲノミクス、モデリングアプローチの組合せを使用する所与のゲノムによってエンコードされるすべてのタンパク質の3次元構造の研究を含む特徴のコレクションである、構造ゲノミクス、糖質および炭水化物ならびに患者におけるその効果の研究を含む特徴のコレクションである、グリコミクス、消費者の幸福、健康、および知識を改善するための技術の応用および統合を通じての食品と栄養領域との間の交わりの研究を含む特徴のコレクションである、フードミクス、細胞内で産生されるmRNA、rRNA、tRNA、および他のノンコーディングRNAを含む、RNA分子の研究を含む特徴のコレクションである、トランスクリプトミクス、代謝産物を伴う化学プロセス、または特定の細胞プロセスが後に残す固有の化学指紋、およびそれらの小分子代謝物プロファイルの研究を含む特徴のコレクションである、メタボロミクス、病態生理学的刺激または遺伝的修飾に対する細胞の動的マルチパラメトリック代謝反応の定量的測定の研究を含む特徴のコレクションである、メタボノミクス、影響を受けやすいサブグループとの関連を有する食事と健康との間の相互作用における遺伝的変異の研究を含む特徴のコレクションである、ニュートリゲネティクス、遺伝子プロファイルに関連付けられている認知プロセスの変化の研究を含む特徴のコレクションである、認知ゲノミクス、ヒトゲノム内の変異の総和が薬物に及ぼす効果の研究を含む特徴のコレクションである、ファーマコゲノミクス、ヒトマイクロバイーム

10

20

30

40

50

内の変異が薬物に及ぼす効果の研究を含む特徴のコレクションである、ファーマコマイクロバイオミクス、毒性物質に反応する生命体の特定の細胞または組織内の遺伝子およびタンパク質の活動の研究を含む特徴のコレクションである、トキシコゲノミクス、ミトコンドリアタンパク質が相互作用する際に使用されるプロセスの研究を含む特徴のコレクションである、ミトインタラクトーム、ゲノミクスおよびプロテオミクスという強力なツールを適用して正常な挙動の生体基質および行動障害として現れる脳の疾患についての理解を深めるプロセスの研究を含む、特徴のコレクションである、サイコゲノミクスであって、サイコゲノミクスを薬物依存症の研究に適用して、これらの疾患に対するより効果的な治療さらには客観的診断ツール、予防対策、および治療法を開発することを含む、サイコゲノミクス、ヒト生物学および病状を理解するためのモデルシステムとして幹細胞を確立する幹細胞生物学の研究を含む特徴のコレクションである、幹細胞ゲノミクス、脳内の神経連絡の研究を含む特徴のコレクションである、コネクトミクス、消化管内に生息する微生物群のゲノムの研究を含む特徴のコレクションである、マイクロバイオミクス、定量細胞解析の研究および生体イメージング法およびバイオインフォマティクスを使用する研究を含む特徴のコレクションである、セロミクス、イメージング質量解析データから高空間分解能で組織または細胞生化学を理解するためのトモグラフィーおよびオミクス法の研究を含む特徴のコレクションである、トモミクス、患者の行動のハイスループット機械測定の研究を含む特徴のコレクションである、エソミクス、ならびにゲノミクスの原理に着想を得たビデオ解析パラダイムの研究を含む特徴のコレクションであり、連続画像シーケンスすなわちビデオは、患者の洞察を明らかにする突然変異の時間経過にわたって展開する単一画像のキャプチャとして解釈され得る、ビデオミクスを含む。

10

20

## 【 0 0 3 8 】

DNA関連(分子)特徴に対する特徴セットは、遺伝子のシーケンシング結果から遺伝子が有し得る最大の効果の独自の計算を含むものとしてよく、これらの遺伝子は、ABCB1-somatic、ACTA2-germline、ACTC1-germline、ALK-fluorescence\_in\_situ\_hybridization\_(fish)、ALK-immunohistochemistry\_(ihc)、ALK-md\_dictated、ALK-somatic、AMER1-somatic、APC-gene\_mutation\_analysis、APC-germline、APC-somatic、APOB-germline、APOB-somatic、AR-somatic、ARHGAP35-somatic、ARID1A-somatic、ARID1B-somatic、ARID2-somatic、ASXL1-somatic、ATM-gene\_mutation\_analysis、ATM-germline、ATM-somatic、ATP7B-germline、ATR-somatic、ATR-somatic、AXIN2-germline、BACH1-germline、BCL11B-somatic、BCLAF1-somatic、BCOR-somatic、BCORL1-somatic、BCR-somatic、BMPR1A-germline、BRAF-gene\_mutation\_analysis、BRAF-md\_dictated、BRAF-somatic、BRCA1-germline、BRCA1-somatic、BRCA2-germline、BRCA2-somatic、BRD4-somatic、BRIP1-germline、CACNA1S-germline、CARD11-somatic、CASR-somatic、CD274-immunohistochemistry\_(ihc)、CD274-md\_dictated、CDH1-germline、CDH1-somatic、CDK12-germline、CDKN2A-immunohistochemistry\_(ihc)、CDKN2A-germline、CDKN2A-somatic、CEBPA-germline、CEBPA-somatic、CFTR-somatic、CHD2-somatic、CHD4-somatic、CHEK2-germline、CIC-somatic、COL3A1-germline、CREBBP-somatic、CTNNB1-somatic、CUX1-somatic、DICER1-somatic、DOT1L-somatic、DPYD-somatic、DSC2-germline、DSG2-germline、DSP-germline、DYNC2H1-somatic、EGFR-gene\_mutation\_analysis、EGFR-immunohistochemistry\_(ihc)、EGFR-md\_dictated、EGFR-germline、EGFR-somatic、EP300-somatic、EPCAM-germline、EPHA2-somatic、EPHA7-somatic、EPHB1-somatic、ERBB2-fluorescence\_in\_situ\_hybridization\_(fish)、ERBB2-immunohistochemistry\_(ihc)、ERBB2-md\_dictated、ERBB2-somatic、ERBB3-somatic、ERBB4-somatic、ESR1-immunohistochemistry\_(ihc)、ESR1-somatic、ETV6-germline、FANCA-germline、FANCA-somatic、FANCD2-germline、FANCI-germline、FANCL-germline、FANCM-somatic、FAT1-somatic、FBN1-germline、FBXW7-somatic、FGFR3-somatic、FH-germline、FLCN-germline、FLG-somatic、FLT1-somatic、FLT4-soma

30

40

50

tic、GATA2-germline、GATA3-somatic、GATA4-somatic、GATA6-somatic、GLA-germline、GNAS-somatic、GRIN2A-somatic、GRM3-somatic、HDAC4-somatic、HGF-somatic、IDH1-somatic、IKZF1-somatic、IRS2-somatic、JAK3-somatic、KCNH2-germline、KCNQ1-germline、KDM5A-somatic、KDM5C-somatic、KDM6A-somatic、KDR-somatic、KEAP1-somatic、KEL-somatic、KIF1B-somatic、KMT2A-fluorescence\_in\_situ\_hybridization\_(fish)、KMT2A-somatic、KMT2B-somatic、KMT2C-somatic、KMT2D-somatic、KRAS-gene\_mutation\_analysis、KRAS-md\_dictated、KRAS-somatic、LDLR-germline、LMNA-germline、LRP1B-somatic、MAP3K1-somatic、MED12-somatic、MEN1-germline、MET-fluorescence\_in\_situ\_hybridization\_(fish)、MET-somatic、MKI67-immunohistochemistry\_(ihc)、MKI67-somatic、MLH1-germline、MSH2-germline、MSH3-germline、MSH6-germline、MSH6-somatic、MTOR-somatic、MUTYH-germline、MYBPC3-germline、MYCN-somatic、MYH11-germline、MYH11-somatic、MYH7-germline、MYL2-germline、MYL3-germline、NBN-germline、NCOR1-somatic、NCOR2-somatic、NF1-somatic、NF2-germline、NOTCH1-somatic、NOTCH2-somatic、NOTCH3-somatic、NRG1-somatic、NSD1-somatic、NTRK1-somatic、NTRK3-somatic、NUP98-somatic、OTC-germline、PALB2-germline、PALLD-somatic、PBRM1-somatic、PCSK9-germline、PDGFRA-somatic、PDGFRB-somatic、PGR-immunohistochemistry\_(ihc)、PIK3C2B-somatic、PIK3CA-somatic、PIK3CG-somatic、PIK3R1-somatic、PIK3R2-somatic、PKP2-germline、PLCG2-somatic、PML-somatic、PMS2-germline、POLD1-germline、POLD1-somatic、POLE-germline、POLE-somatic、PREX2-somatic、PRKAG2-germline、PTCH1-somatic、PTEN-fluorescence\_in\_situ\_hybridization\_(fish)、PTEN-gene\_mutation\_analysis、PTEN-germline、PTEN-somatic、PTPN13-somatic、TPRD-somatic、RAD51B-germline、RAD51C-germline、RAD51D-germline、RAD52-germline、RAD54L-germline、RANBP2-somatic、RB1-germline、RB1-somatic、RBM10-somatic、RECQL4-somatic、RET-fluorescence\_in\_situ\_hybridization\_(fish)、RET-germline、RET-somatic、RICTOR-somatic、RNF43-somatic、ROS1-fluorescence\_in\_situ\_hybridization\_(fish)、ROS1-md\_dictated、ROS1-somatic、RPTOR-somatic、RUNX1-germline、RUNX1T1-somatic、RYY1-germline、RYY2-germline、SCN5A-germline、SDHAF2-germline、SDHB-germline、SDHC-germline、SDHD-germline、SETBP1-somatic、SETD2-somatic、SH2B3-somatic、SLIT2-somatic、SLX4-somatic、SMAD3-germline、SMAD4-germline、SMAD4-somatic、SMARCA4-somatic、SOX9-somatic、SPEN-somatic、STAG2-somatic、STK11-gene\_mutation\_analysis、STK11-germline、STK11-somatic、TAF1-somatic、TBX3-somatic、TCF7L2-somatic、TERT-somatic、TET2-somatic、TGFBR1-germline、TGFBR2-germline、TGFBR2-somatic、TMEM43-germline、TNNI3-germline、TNNT2-germline、TP53-gene\_mutation\_analysis、TP53-immunohistochemistry\_(ihc)、TP53-md\_dictated、TP53-germline、TP53-somatic、TPM1-germline、TSC1-germline、TSC1-somatic、TSC2-germline、TSC2-somatic、VHL-germline、WT1-germline、WT1-somatic、XRCC3-germline、およびZFHX3-somaticである。

【 0 0 3 9 】

特徴の十分に口バスタなコレクションは、上に開示されている特徴のすべてを含むものとしてよく、しかしながら、利用可能な特徴に基づくモデルおよび予測は、網羅的な特徴セットに比べてかなり制限されている特徴の選択から最適化および訓練されたモデルを含み得る。そのような制約された特徴セットは、数十から数百もの特徴を含み得る。たとえば、モデルの制約された特徴セットは、患者の腫瘍のシーケンシングのゲノム結果、ゲノム結果に基づく派生特徴、患者の腫瘍発生源、診断時の患者の年齢、患者の性別および人種、ならびに患者が定期検診中に医師に示した症状を含み得る。

【 0 0 4 0 】

特徴ストアは、任意の特徴、変化、または患者の特徴もしくはそれらの特徴の変化から導出される計算された出力から選択することによって、機械学習およびアナリティクスを適用することで患者の特徴セットを強化し得る。このような特徴ストアは、特徴モジュールで見つかった元の特徴から新しい特徴を生成するか、または特徴に基づき重要な洞察もしくは解析を識別して記憶し得る。特徴の選択は、生成されるべき変化または計算に基づくものとしてよく、ゲノムの一塩基もしくは複数塩基多型の挿入もしくは欠失、腫瘍遺伝子変異量、マイクロサテライト不安定性、コピー数変化、融合、または他のそのような計算を含み得る。将来の変化または計算を知らせ得る生成される変化または計算の例示的な出力には、肥大型心筋症(HCM)およびMYH7におけるバリエーションの所見を含む。以前に分類されたバリエーションは、患者のゲノム内で識別され得るが、これは新規バリエーションの分類を知らせるか、または疾患のさらなるリスクを示し得る。例示的なアプローチは、HCMに関連付けられているMYH7における領域を識別するためにバリエーションとそれぞれの分類を強化することを含み得る。この領域に局在化される患者のシーケンシングから検出された新規のバリエーションがあれば、HCMに対する患者のリスクを高めることになるであろう。このような変化検出に利用され得る特徴は、MYH7の構造およびその中のバリエーションの分類を含む。強化に焦点を当てたモデルは、そのようなバリエーションを分離し得る。

#### 【0041】

#### 人工知能モデル

本明細書で参照される人工知能モデルは、勾配ブースティングモデル、ランダムフォレストモデル、ニューラルネットワーク(NN)、回帰モデル、ナイーブベイズモデル、または機械学習アルゴリズム(MLA)であってもよい。MLAまたはNNは、訓練データセットから訓練され得る。例示的な予測プロファイルでは、訓練データセットは、EHRからキュレーションされたもの、または遺伝子シーケンシングレポートなど、患者のイメージング、病理、臨床、および/または分子レポートならびに詳細を含み得る。MLAは、線形回帰、ロジスティック回帰、決定木、分類および回帰木、ナイーブベイズ、最近傍クラスタリングを使用する教師ありアルゴリズム(データセット内の特徴/分類がアノテーションされているアルゴリズムなど)、アプリアリ、meansクラスタリング、主成分解析、ランダムフォレスト、適応型ブースティングを使用する教師なしアルゴリズム(データセット内の特徴/分類がアノテーションされていないアルゴリズムなど)、生成アプローチ(ガウス分布の混合、多項分布の混合、隠れマルコフモデルなど)、低密度分離、グラフベースのアプローチ(mincut、調和関数、多様体正則化など)、発見的アプローチ、またはサポートベクターマシンを使用する半教師ありアルゴリズム(データセット内の不完全な数の特徴/分類がアノテーションされているアルゴリズムなど)を含む。NNは、条件付き確率場、畳み込みニューラルネットワーク、注意ベースニューラルネットワーク、ディープラーニング、長短期記憶ネットワーク、または他のニューラルモデルを含み、訓練データセットは、複数の腫瘍サンプル、各サンプルに対するRNA発現データ、および各サンプルに対するイメージングデータをカバーする病理レポートを含む。MLAおよびニューラルネットワークは、機械学習の異なるアプローチを識別するが、これらの用語は、本明細書では交換可能に使用され得る。したがって、特に断りのない限り、MLAの言及は対応するNNを含むか、またはNNの言及は対応するMLAを含み得る。訓練は、最適化されたデータセットを提供すること、患者の記録に出現するときにこれらの特徴にラベルを付けすること、および新しい入力に基づき予測するかまたは分類するようMLAを訓練することを含み得る。人工NNは、効率的な計算モデルであり、人工知能の難しい問題を解くことに強みのあることを示している。これらが普遍的近似器であることも示されている(適切なパラメータを与えたときに広範な関数を表現することができる)。MLAの中には、重要な特徴を識別し、それらに対する係数、すなわち重みを識別するものもある。係数は、特徴の出現頻度と掛け合わされてスコアを生成し、1つまたは複数の特徴のスコアが閾値を超えたときに、MLAによっていくつかの分類が予測され得る。係数スキーマは、ルールベーススキーマと組み合わせられ、複数の特徴に基づく予測などの、より複雑な予測を生成するものとしてよい。たとえば、異なる分類において10個の重要な特徴が識別され得る。係数のリストは、重要な特徴に対

10

20

30

40

50

して存在し、ルールセットは、分類に対して存在し得る。ルールセットは、特徴の出現数、特徴のスケーリングされた重み、または当業者に知られているロジックで符号化された特徴の他の定性的および定量的な評価に基づくものとしてよい。他のMLAでは、特徴は、二分木構造で編成されてもよい。たとえば、大部分の分類を区別する主要な特徴は、二分木のルートとして、また木における後続の各枝として、木の終端ノードに到達することに基づき分類が与えられ得るまで、存在し得る。たとえば、二分木は、第1の特徴を検定するルートノードを有し得る。この特徴の出現または非出現は、存在していなければならず(二分決定)、ロジックは、分類される項目に対して真である枝をトラバースするものとしてよい。追加のルールは、閾値、範囲、または他の定性的および定量的検定に基づくものとしてよい。訓練データセットが多数の知られている値またはアノテーションを有するとき

10

に教師ありの方法は有用であるが、EMR/EHRドキュメントの性質上、多数のアノテーションが与えられていない場合もある。大量のラベル付けされていないデータを探索するときに、教師なしの方法はデータセット内のインスタンスのビン分割/バケット化に対して有益である。上記のモデルの単一のインスタンス、または2つもしくはそれ以上のそのようなインスタンスは組み合わせることで、本明細書において、モデル、人工知能、ニューラルネットワーク、または機械学習アルゴリズムの目的のためにモデルを構成し得る。

#### 【0042】

一連の変換ステップは、患者データストアからのデータを解析に適したフォーマットに変換するために実行され得る。様々な最新の機械学習アルゴリズムが、特定の患者母集団に対する予想生存率および/または反応の予測を対象とするモデルを訓練するために利用され得る。例示的なデータストア14は、2018年10月17日に出願された米国仮特許出願第62/746,997号、名称「Data Based Cancer Research and Treatment Systems and Methods」、2019年2月28日に出願され、米国特許第10,395,772号として2019年8月27日に発行された、米国特許出願第16/289,027号、名称「Mobile Supplementation, Extraction, and Analysis of Health Records」、および2019年10月17日に出願されたPCT国際出願第PCT/US19/56713号、名称「Data Based Cancer Research and Treatment Systems and Methods」においてさらに詳しく説明されており、各々全体が参照により本明細書に組み込まれている。

20

#### 【0043】

システムは、臨床および分子の非識別化記録を一括して伝送するためのデータ配信パイプラインを備え得る。また、システムは、データプライバシー、および医療保険の相互運用性と説明責任に関する法律などの、適用法またはガイドラインの遵守を維持するための非識別データおよび識別データ用に別々のストレージを備え得る。

30

#### 【0044】

生の入力データおよび/または変換された、正規化された、および/または予測データは、本明細書においてより詳細に説明されているように、1つもしくは複数の比較または解析機能を実行するためにシステムによりさらにアクセスできるように1つまたは複数のリレーショナルデータベースに記憶され得る。リレーショナルデータベースを構築するために使用されるデータモデルは、著しい量の、様々なデータ、たとえば、何百もの異なる列を含む何十ものテーブルを記憶し、編成し、表示し、および/または解釈するために使用され得る。OMOPまたはQDMなどの標準データモデルとは異なり、このデータモデルは、様々な臨床属性を直接関係付けるためにテーブル内に、またはテーブル間に固有のリンクを生成し、それによって、複雑な臨床属性を取り込み、解釈し、解析することを容易にし得る。

40

#### 【0045】

上で説明されているように、関連データが受信され、変換され、操作された後、図1のシステム図に関して上で説明されているように、所望の動的ユーザーインターフェースを生成できるようにシステムは複数のモジュールを備え得る。

#### 【0046】

患者コホートフィルタ処理ユーザーインターフェース

50

図2を参照すると、患者コホート選択フィルタ処理インターフェース24の第1の実施形態は、ディスプレイ画面の高さ(または、代替的に長さ)に沿って提供されるサイドペイン200として提供されてよく、このサイドペインを通して、属性基準202(臨床、分子、人口統計など)がユーザによって指定され、さらなる解析のために注目する患者母集団を定義することができる。サイドペイン200は、それを選択する、ドラッグする、ダブルクリックするなどのことを行うことによって、非表示にされるか、または拡大され得る。

#### 【0047】

それに加えて、または代替的に、システムは、システムによって記憶されている腫瘍データに対して定義されている1つまたは複数の属性を認識するものとしてよく、それらの属性は、たとえば、遺伝子型、表現型、家系、または人口統計的なものであってよい。様々な選択可能な属性基準は、患者データストア14内に記憶されている患者関係メタデータを反映するものとしてよく、例示的なメタデータは、たとえば、プロジェクト名(患者のリストを記憶するデータベースを反映し得る)204、性別206、人種208、癌、癌部位210、癌名212、転移、癌名214、腫瘍部位216(腫瘍が特定された場所を反映し得る)、病期218(I、II、III、IV、および不明など)、M病期220(m0、m1、m2、m3、および不明など)、薬剤(名前222または成分224など)、シーケンシング226(遺伝子名またはバリエーションなど)、MSI(マイクロサテライト不安定性)ステータス228、TMB(腫瘍遺伝子変異量)ステータス(図示せず)、手順230(名前別など)。または死亡(イベント名232または死因234別など)を含み得る。

#### 【0048】

システムは、また、「特徴および特徴モジュール」の見出しの下にリストされているものを含む、本明細書にリストされている基準のいずれかに従って患者データをフィルタ処理することをユーザに許すものとしてよく、追加基準、すなわち、施設、人口統計、分子データ、評価、診断部位、腫瘍キャラクタリゼーション、治療、または1つもしくは複数の内部基準のうちの1つまたは複数を含み得る。施設オプションは、ユーザが特定の設備に基づきフィルタ処理することを許し得る。人口統計オプションは、たとえば、性別、死亡状況、初診時年齢、または人種のうちの1つもしくは複数によってユーザがソートすることを許し得る。分子データオプションは、ユーザがバリエーションコール(たとえば、いつ患者に対して利用可能な分子データがあるか、特定の遺伝子名、突然変異、変異効果、および/またはサンプルタイプがどのようなものか)、抽象化されたバリエーション(たとえば、遺伝子名および/またはシーケンシング方法を含む)、MSIステータス(たとえば、安定、低、もしくは高)、またはTMBステータス(たとえば、ユーザ定義の範囲内または範囲外で選択可能)に従ってフィルタ処理することを許し得る。評価は、喫煙状態および/または閉経状態などの、様々なシステム定義基準に従ってユーザがフィルタ処理することを許し得る。診断部位は、原発および/または転移部位に応じてユーザがフィルタ処理することを許し得る。腫瘍キャラクタリゼーションは、1つまたは複数の腫瘍関係基準、たとえば、異型度、組織、病期、TNM悪性腫瘍分類(TNM)、および/または各それぞれのT値、N値、および/またはM値に従ってユーザがフィルタ処理することを許し得る。治療は、たとえば、成分、処方計画、治療タイプなどを含む様々な治療関係オプションのうちからユーザが選択することを許し得る。

#### 【0049】

いくつかの基準は、初期基準が選択された後に指示され得る複数の下位基準からユーザが選択することを許し得る。他の基準は、たとえば、死亡しているか否かの二者択一オプションをユーザに提示し得る。さらに他の基準は、スライダまたは範囲タイプオプションをユーザに提示するものとしてよく、たとえば、初期診断時の年齢は、ユーザが選択可能な下限値および上限値を有するスライダとして提示されてよい。なおもさらに、これらのオプションのいずれかについて、システムは、選択された基準に基づきシステムが患者を含めるか除外するべきかを交互に決めるラジオボタンまたはスライダをユーザに提示し得る。本明細書において説明されている例は、基準として使用され得る情報のタイプの範囲を制限することはないことを理解されたい。構造化形式で記憶することができる任意のタ

イブの医療情報が基準として使用されてよい。

【0050】

別の実施形態では、ユーザインターフェースは、たとえば、ユーザインターフェースの「Ask Gene」タブ236において、またはフィルタ処理インターフェースのテキスト入力を介して、コホートに対するフィルタ基準定義を円滑にするための自然言語検索スタイルバーを備え得る。一態様において、キーボードタイプの入力を介して、または機械解釈された口述を介して、クエリを指定することができることで、コホートファネル(次の節でより詳細に説明される)の後続の層の1つまたは複数が定義され得る。したがって、たとえば、従来の自然言語処理ソフトウェアまたは技術を採用したときに、「乳癌患者」という入力があると、システムは「"cancer\_site==乳癌"」というフィルタを認識し、これをフィルタ処理の次の層として追加する。同様に、システムは、「ゲムシタピンに対して拒絶反応のある膵臓患者」という入力を認識し、それを、フィルタ処理の複数の連続する層、たとえば、「"cancer\_site==膵臓癌" AND "薬剤==ゲムシタピン" AND "拒絶反応==not null"」に変換する。

10

【0051】

第2の態様では、自然言語処理は、ユーザがシステムを使用して一般的洞察について直接クエリを実行することを許し、それによって、1つまたは複数のファネルレベルを介して患者のコホートを絞り込み、またユーザインターフェースに適切な要約パネルを表示することをシステムに行わせるものとしてよい。したがって、システムが「病期IIIの直腸結腸癌患者の、放射線治療後の5年間の無増悪生存率は?」というクエリを受け取る状況にあれば、「"cancer\_site==直腸結腸" AND "stage==III" AND "treatment==放射線療法"」などの一連のフィルタに変換し、次いで、たとえば患者生存率解析ユーザインターフェース30を使用して5年間の無増悪生存率を表示することになる。同様に、「女性の肺癌患者のうち何パーセントが、診断時に閉経後であるか」というクエリは、これを"gender==女性"、"cancer\_site==肺"、"temporal==診断時"などの一連の患者に翻訳し、その結果の導出される患者のうち何人が閉経後の状況を反映するデータを有していたかを決定し、関連するパーセンテージを決定し、たとえば、1つまたは複数の統計的要約チャートを通じて結果を表示する。

20

【0052】

コホートファネルおよび母集団解析ユーザインターフェース

30

次に図3～図9を参照すると、コホートファネルおよび母集団解析ユーザインターフェース26は、一般患者母集団またはデータが患者データストア14に記憶されている患者母集団における分布に関して、注目する各属性を示す患者の分布における重要な変曲点を識別することを目的として、ユーザがコホートの解析を実行することを可能にするように構成され得る。一態様において、図2に関して上で説明されている追加の患者関係基準のフィルタ処理および選択は、コホートファネルおよび母集団解析ユーザインターフェース26に関連して使用され得る。

【0053】

別の実施形態では、システムは、複数の選択可能なタブを示すダイアログボックス238を開く選択可能なボタンまたはアイコンを備えるものとしてよく、各タブは、上で説明されている同じまたは類似するフィルタ処理基準(人口統計、分子データ、評価、診断部位、腫瘍キャラクターゼーション、および治療)を表す。各タブを選択すると、上で説明されているような各それぞれのフィルタに対する同じまたは類似するオプションがユーザに提示され得る(たとえば、「人口統計」を選択すると、性別、死亡状況、初期診断時の年齢、または人種に関係するさらなるオプションがユーザに対して提示される)。次いで、ユーザは、1つまたは複数のオプションを選択し、「次へ」を選択し、次いで、それが包含フィルタであるかまたは除外フィルタであるかを選択するものとしてよく、対応する選択は、ファネル(以下でより詳細に説明される)に追加され、アイコンはファネルの次の連続して狭くなる部分の下に移動する。

40

【0054】

50

それに加えて、または代替的に、データベース内のコホート、または患者のセットを見ることで、システムは、メニュー240を介して複数の臨床および分子因子によるフィルタ処理を可能にする。たとえば、臨床的要因に関して、システムは、患者人口統計242、癌部位244、腫瘍キャラクタリゼーション246、または分子データ248に基づくフィルタを含むことができ、これらはさらに、腫瘍キャラクタリゼーションについて組織250、病期252、および/または異型度ベースのオプション254(図4参照)などの、フィルタ処理可能なオプション242の独自のサブセットを含み得る。分子的因子に関して、システムは、バリエーションコール256、抽象化されたバリエーション258、MSI260、および/またはTMB262に従ってフィルタ処理を可能にし得る。

#### 【0055】

本明細書において説明されている例は、様々な癌型に関する解析を提供しているけれども、他の実施形態では、システムは、他の病状のフィルタ処理された表示を示すために使用されてよく、それらのような状況では、他の疾患に対する関連する状態に特に焦点を当てるために、選択項目が異なることが理解されるべきである。

#### 【0056】

コホートファネルおよび母集団解析ユーザインターフェース26は、データセット内の患者の数を、一度にすべて表示するか、または複数のフィルタ処理基準のユーザの選択を受けた後に徐々に表示するかのいずれかを視覚的に実行し得る。一態様において、フィルタ属性による患者頻度の表示は、インタラクティブファネルチャート264を使用して提供され得る。図3～図9を見るとわかるように、各選択で、ユーザインターフェース26は、フィルタ基準にマッチする結果の削減を例示するように更新され、たとえば、より多くのフィルタ基準が追加されると、ユーザのフィルタ処理因子の各々を受け取った後に、選択された基準のすべてにマッチするより少ない患者が存在する。

#### 【0057】

上記のフィルタ処理は、フィルタ基準の各ユーザの選択を受け取った後に実行されるものとしてよく、ファネル264は各フィルタ選択の後にデータセットの絞り込み範囲を示すように更新する。その状況では、上で説明されているようなフィルタ処理メニュー240は、切り換えられたときも各タブ内で見えたままであり得るか、またはサイドに折り畳まれ得るか、または選択されたフィルタ処理オプションの要約266として表現され、ユーザに縮小データセット/サイズを知らせるようにし得る。

#### 【0058】

上で説明されている各フィルタ処理方法に関して、因子の組合せは、ブール形式の組合せに基づき得る。例示的なブール式の組合せは、フィルタリング因子AおよびBについて、「A AND B」、「A OR B」、「A AND NOT B」、「B AND NOT A」などを有する患者を検索するかどうかをユーザが選択することを可能にすることを含み得る。

#### 【0059】

注目する最終のフィルタ処理済みコホートは、以下で説明されているモジュールまたは他のユーザインターフェースにおけるさらなる詳細な解析の基礎をなし得る。注目している母集団は「コホート」と呼ばれる。ユーザインターフェースは、患者データストア内の利用可能なデータ属性に基づき適切に事前に入力された固定機能属性セレクタを提供することができる。

#### 【0060】

表示は、患者の地理的位置クラスタリングプロット、および/または公開されている統計および/または非公開のキュレーションされた統計との人口統計学的分布の比較をさらに示し得る。

#### 【0061】

##### 患者タイムライン解析モジュール

それに加えて、システムは、ユーザが各患者の病床での生活における一連の事象を検討することを可能にする患者タイムライン解析モジュール28を備え得る。このデータは、患者データの機密性を保護するために、上で説明されているように匿名化され得ることは理

10

20

30

40

50

解される。

【0062】

ユーザが、たとえば、コホートファネル&母集団解析ユーザインターフェース26を介して、所望のフィルタ基準をすべて提供した後、システムは、ユーザが患者のフィルタ処理されたサブセットを解析することを許す。図に描かれているユーザインターフェースに関して、この手順は、インターフェース26の右上隅に提示されている「Analyze Cohort」オプション268を選択することによって達成され得る。

【0063】

次に図10を参照すると、患者のフィルタ処理されたサブセットの解析を要求した後、ユーザインターフェースは、患者タイムライン解析ユーザインターフェース28内にデータ要約ウィンドウを生成するものとしてよく、1つまたは複数の領域300に、選択された患者サブセットに関する情報、たとえば、臨床および分子特徴における他の分布の数を提供する。一態様において、第1の領域300aは、平均患者年齢302および/または患者年齢のプロット304などの人口統計学的情報を含み得る。第2の領域300bは、患者のサブセットに対して、性別情報306などの追加の人口統計学的情報を含み得る。第3の領域300cは、たとえば、サブセット内の患者の各々が服用した薬剤308の解析を含む、特定の臨床データの要約を含み得る。同様に、第4の領域300dは、患者の各々に関する分子データ、たとえば、サブセット内の患者によって保有される各ゲノムバリエーションまたは変化310の解析結果を含み得る。

【0064】

ユーザインターフェース28は、また、ユーザが、データ要約ウィンドウまたは領域300内に提示されているデータ要約情報に関するクエリを実行し、たとえば制御パネル312を使用してそのデータをさらにソートすることを可能にする。たとえば、図11~図14を見るとわかるように、システムは、たとえば、性別314、組織316、閉経状態318、反応320、喫煙状態322、病期324、および外科的手技326を含む1つまたは複数の要因に基づき患者データをソートするように構成され得る。これらのオプションのうちの1つまたは複数を選択しても、データ要約ウィンドウ内に要約されているフィルタ処理について説明したときの上述の場合のように、患者のサンプルサイズは減少し得ない。その代わりに、ソート機能は、要約された情報を1つまたは複数のサブカテゴリに細分化し得る。たとえば、図11および図12は、データ要約ウィンドウ300c内で追加の応答データ328が上に重ねられることによってソートされる薬剤情報308を、重ねられた応答データを説明する凡例330とともに示している。

【0065】

次に図13~図14を参照すると、ユーザによって選択された患者のサブセットは、また、たとえばドロップダウンメニュー332を介して、患者の第2のサブセット(または「コホート」と比較されてよく、それによってグループのサイドバイサイド解析を円滑にし得る。そうすることで、サブセット間の任意の類似点、さらには任意の顕著な差異をユーザが素早く簡単に見ることを可能にし得る。

【0066】

一実施形態において、高水準の概要を示す事象タイムラインガントスタイルチャートが、表形式の詳細パネルと結合されて提供される。この表示は、患者サブセット間の重複領域および潜在的な途切れ領域の両方を識別することを目的として、正規化されたタイムライン上で同時に複数の患者を視覚化し比較することも可能にし得る。

【0067】

患者「生存率」解析モジュール

システムは、さらに、図15~図20を見るとわかるように、患者生存率解析ユーザインターフェース30を使用することで患者のサブセットの生存率解析を行い得る。このモデリングおよび視覚化コンポーネントは、ユーザが、注目しているフィルタ処理されたコホートのサブグループに対して、事象(および時間における確率)曲線までの時間およびその信頼区間をインタラクティブに探索することを可能にし得る。時系列の開始および標的事象

10

20

30

40

50

は、選択された母集団内の患者グループをクラスタ化する属性とともに、ユーザによって選択され、動的に修正され得るが、すべて、曲線ビジュアライザは提供されたパラメータに反動的に適應する。

【0068】

ユーザにその解析の範囲を定義する柔軟性を提供するために、システムは、ユーザが、その解析の基礎となる開始事象および終了事象の一方または両方を選択することを可能にし得る。例示的な開始事象は、初期原発性疾患の診断、増悪、転移、退行、最初の原発性癌の識別、薬剤の初期処方などを含む。逆に、例示的な終了事象は、増悪、転移、再発、死亡、時間期間、および治療開始/終了日を含む。開始事象を選択すると、曲線が始まるすべての患者に対するアンカーポイントが設定され、終了事象を選択すると、曲線が予測される水平線が設定される。

10

【0069】

図15を見るとわかるように、解析は、終了事象302、たとえば、無増悪生存率または全生存率の、時間304に対するプロット300の形態でユーザに提示され得る。これらの目的に関する増悪は、1つまたは複数の増悪事象、たとえば、転移事象、再発、薬物に対するまたは薬剤に依存しない増悪の特定の尺度、特定の腫瘍サイズもしくは腫瘍サイズの変化、または強化された測定値(基礎となる臨床データセットから間接的に抽出される測定値など)の出現を反映し得る。例示的な強化された測定値は、病期変化の検出(病期2のカテゴリ分類が病期3に変更されたことを検出することなどによる)、回帰、または推論を介したものの(病期3および転移の両方が病期2および4の検出から推論されるが、病期3の検出はないことなど)を含み得る。

20

【0070】

それに加えて、システムは、図16を見るとわかるように、ユーザがプロット内の特定のタイムスパンに焦点を当てるか、またはズームインすることを可能にするように構成され得る。特に、ユーザは、x軸のみ、y軸のみ、またはx軸とy軸の両方を同時にズームインすることができるものとしてよい。この機能は、解析される疾患のタイプによっては特に有用であるが、それはいくつかの侵襲性疾患では、他の疾患に比べて小さい時間窓を解析する方が都合がよいからである。たとえば、膵臓癌の患者の生存率は、他のタイプの癌に比べて著しく低くなる傾向があり、したがって、膵臓癌を解析するとき、たとえば、約5年の窓から約1年の窓へと、より短い期間にズームインすることがユーザには有用であり得る。

30

【0071】

次に図17~図20を参照すると、ユーザインターフェース30は、また、追加のグループ化またはソート基準に対応するユーザ入力を受け取ることによってその表示を修正し、サブセット内のより小さなグループの生存率情報を提示するようにも構成され得る。それらの基準は、臨床的または分子的要因であってよく、ユーザインターフェース30は、たとえば、開始事象306または終了事象308のいずれか、さらには性別310、遺伝子312、組織314、処方計画316、喫煙状態318、病期320、外科的手技322などをユーザが選択することを可能にする1つまたは複数のドロップダウンメニューなどのセレクトアを含み得る。

【0072】

次いで、図18に示されているように、基準の1つを選択することで、その基準に関連する複数のオプションをユーザに提示し得る。たとえば、「処方計画」を選択することで、1つまたは複数の値セットを使用して、ユーザインターフェース内に生成された選択可能なフィールドに書き込み、サブセット内の患者のうちの1人または複数を受ける特定の薬剤処方計画324のうちの1つまたは複数を選択するようにユーザに促すことをシステムに行わせ得る。したがって、図19が示しているように、「Gemcitabine+Paclitaxel」オプション326を、続いて「FOLFIRINOX」オプション328を選択した結果、システムは患者サブセットデータを解析し、どの患者の記録が選択された処方計画のいずれかに対応するデータを含むかを決定し、患者のそれらの別々のグループに対する生存率統計を再計算し、各処方計画に対する別々の生存率プロット330、332を含むようにユーザインターフェ

40

50

ースを更新する。グループを追加する/2つまたはそれ以上の選択を追加する結果、システムはそれらを並べて表示するために同じチャートにプロットし、ユーザインターフェースは、各グループを区別するために名前、色、およびサンプルサイズを有する凡例334を生成し得る。

#### 【0073】

図20を見るとわかるように、システムは、生存率解析に関して統計的範囲を計算して重ね合わせることによってより高いレベルの解析を可能にし得る。特に、システムは、ユーザによって要求された各データセットに関して信頼区間を計算し、生存率プロット330、332に対するそれらの信頼区間336、338を表示し得る。一事例において、所望の信頼区間は、ユーザ設定され得る。別の事例では、信頼区間は、システムによって事前設定されてもよく、たとえば、68%(1標準偏差)区間、95%(2標準偏差)区間、または99.7%(3標準偏差)区間であってよい。信頼区間は、当業者によって理解されるように、Kaplan Meier信頼区間として、または別のタイプの統計解析を使用して計算され得る。

10

#### 【0074】

前の説明から理解されるように、システムの有用性の基盤となるのは、これらの予測を推進する高い重要度の特徴および相互作用パスウェイを強調する能力、および予想される規範から著しく逸脱する反応のレベルを示す患者のコホートをさらにピンポイントで特定する能力である。この文脈において、高い重要度は、予測の結果に対する特徴の重要度に基づくものと考えられ得る。特に、予測に最大の重みを与える特徴は、高い重要度の特徴として指定され得る。システムおよびユーザインターフェースは、特定の包含および/または除外基準を与えられた場合に患者の選択およびコホートの定義に対する直観的で効率的な方法を提供する。システムは、特定の臨床的および/または分子的属性、さらには薬物の投与量、組合せ、および/または他の治療プロトコルが、潜在的に大きく、他の方法では扱いにくい患者サンプルサイズに対する治療帰結および患者生存率に与える影響についての研究および解析を含む、内部研究および解析を円滑にするロバスタなユーザインターフェースも提供する。

20

#### 【0075】

本明細書において説明されているモデリングおよび可視化フレームワークは、ユーザが、フィルタ処理された患者コホートの臨床的およびゲノムデータ中の自動検出されたパターンをインタラクティブに探索し、それらのパターンと治療反応および/または生存の可能性との関係を解析し得る。この解析は、ユーザを、本システムおよびユーザインターフェースを使用しない場合と比較してサイクルの早い段階で、患者に対する情報がより豊富な治療決定に導き得る。また、この解析は、臨床試験の文脈においても有用であり、ロバスタな、データに裏付けられた臨床試験の包含および/または除外解析を提供する。本システムは、臨床的および分子的データの広範なライブラリによって裏付けられ、臨床解析および機械学習に関係する様々なアルゴリズムおよびコンセプトを統合し適用して、完全に統合されたインタラクティブなユーザインターフェースを生成する。

30

#### 【0076】

##### 外れ値解析モジュール

次に、図21~図24を参照すると、別の態様では、システムは、解析されている患者のグループ内の1つまたは複数の外れ値の存在を素早く、効果的に決定するために、患者事象尤度解析ユーザインターフェース32などの追加のユーザインターフェースを備え得る。たとえば、図21のインターフェースは、無増悪生存率に基づくデータにおいて患者の1つまたは複数のグループがどのように自然に分離するかをユーザが視覚的に決定することを可能にする。このユーザインターフェースは、複数の患者グループを表す複数のインジケータ402を含む第1の領域400を含み、所与のグループ内の各患者は、そのグループ内の他の患者との間に共通性を有している。たとえば、共通性は、上述の属性、フィルタ処理に使用される追加のシステム定義済み腫瘍関係基準、およびシステムによって識別され得る構造化形式で記憶できる他の医療情報のうちの1つまたは複数に基づくものとしてよい。それに加えて、任意の属性の非存在からグループが形成され得る。たとえば、薬を飲ん

40

50

だことがない、治療を受けたことがない、または他の何らかの形で1つまたは複数の属性の非存在を共有するグループによって、共通点が見出され得る。この領域は、インジケータが中央インジケータ408から半径方向に離れてプロットされ、さらにはそのインジケータの周りで周方向にプロットされるという点で、レーダープロット406に類似しているものとしてよく、中央インジケータ408からの半径方向距離は、中央の半径方向に相隔てて並ぶインジケータによって表される患者間の類似度を反映し、半径方向に相隔てて並ぶインジケータ間の周方向距離は、それらのインジケータによって表される患者間の類似度を反映している。この事例では、半径方向距離に関する類似度は、外れ値解析を支配する1つ/複数の基準にもつぱら、または単独で基づくものとしてよい。たとえば、無増悪生存率(「PFS」)に関して患者グループを解析するとき、中心点またはインジケータ408は、評価される時間期間にわたるコホート全体のPFSの特定の割合またはパーセンテージ(たとえば、10%、25%、50%、75%、または他のパーセンテージ)に基づくものとしてよく、中心点またはインジケータ408からの半径方向距離は、特定のパーセンテージPFSよりもよい患者のグループが中心点またはインジケータ408よりも上にプロットされ、特定のパーセンテージPFSよりも悪い患者のグループが中心点またはインジケータ408よりも下にプロットされるように、それぞれのインジケータ402によって反映される患者のグループの無増悪生存率を示すものとしてよく、X軸上の中心点からの距離は、母集団のサイズ、観察され予想されたPFSの間の差、または同様のメトリックに基づき導出され得る。

#### 【0077】

それに加えて、ユーザインターフェースは、患者のサブセットを外れ値としてフィルタ処理するか、選択するか、または他の何らかの形で第1の領域内で強調表示するための制御パネル412を含む第2の領域410を備え得る。制御パネル内で値または範囲を設定することで、レーダープロット上にオーバーレイ414を生成するものとしてよく(図22参照)、オーバーレイは、中央インジケータ408を中心とする円の形態であってよく、円の半径は、第2の領域410内でユーザから受け取った値または範囲に関係付けられ得る。この態様では、ユーザは、参照患者に関して両方向に等しく適用される値を選択し得る。たとえば、ユーザは、「25%」を選択してよく、これは、オーバーレイが中心点またはインジケータ408を囲む一様な円であり得るように-25%から+25%の範囲として反映され得る。代替的に、システムは、ユーザから複数の値を受け取ってもよく、たとえば、「-20%から+25%」などの、正の範囲を表すものと負の範囲を表すものを受け取るものとしてよい。値は、テキスト入力、ドロップダウンを介して受け取られるか、またはグラフ上のそれぞれの位置をクリックすることによって選択されてもよい。その場合、オーバーレイは、異なる半径を有する2つの別々の半球の形態をとるものとしてよく、その半径はユーザから受け取った値を反映している。図21および図22を見るとわかるように、値は、中心点またはインジケータ408に関係するどのような値であってもその値からの偏差のパーセントを示し得る。たとえば、図21および図22は、0%のPFS値を有する患者を中心とした患者の様々なクラスタの無増悪生存率(PFS)パーセンテージを表示している。図21は、 $\pm 10\%$ の範囲のオーバーレイ414を含み、図22は、範囲が $\pm 30\%$ に修正されたときにオーバーレイがどのように調整されるかを示している。中心点またはインジケータ408が、非ゼロ値、たとえば20%のPFSの患者に関連付けられ得ることは理解されるであろう。その場合、 $\pm 10\%$ の範囲は、10~30%のPFSの範囲にある患者のクラスタをカプセル化し、 $\pm 30\%$ の範囲は、-10~50%の範囲内の患者のクラスタをカプセル化する。いずれの場合も、システムがユーザ入力を受け取った後、オーバーレイによって覆われるインジケータは視覚的外観を変化させ、たとえば、図22に示されているように、薄いグレーで表示された、または他の何らかの形で目立たない形態に変化するものとしてよく、外れ値閾値414の外側にある値416(図22の右上隅にヒストグラム形式で示されている)は濃い色(たとえば、青色または陰影)で表示され、外れ値閾値414内の値418は薄い色(たとえば、薄灰色または陰影なし)で表示される。すなわち、オーバーレイの外側にあるインジケータは、強調表示されたままであるか、さもなければより容易に視覚的に区別可能であり、それによって、それらのインジケータを外れ値を表すものとして識別し得る。

10

20

30

40

50

## 【0078】

別の態様では、図23～図24を見るとわかるように、ユーザインターフェースの第1領域400は、説明したばかりのレーダー型プロットとは異なる、複数の患者グループのプロット420を含み得る。この態様では、x軸422がインジケータで表される所与のグループ内の患者の数を表し、y軸434が考察されている1つ/複数の基準からの逸脱の度合いを表し得る。これらの表示パラメータの結果として、このユーザインターフェース32は、最大の患者グループ436をy軸から最も遠くに提示し、最大の外れ値グループ438をx軸422から最も遠くに提示する。(このユーザインターフェースおよび前に説明したユーザインターフェースの両方について、原点は、y軸または半径方向の寸法のいずれかについて、それぞれ0の値を反映していなくてもよいことは理解されるべきである。その代わりに、原点は、解析される1つ/複数の基準の基礎レベルを反映し得る。たとえば、無増悪生存率の場合、基礎グループは15%の2年生存率を有し得る。その場合、外れ値の存在を評価するために、その15%の値に関して偏差が決定され得る。そのような偏差は加法的であり、±20%は0%から35%(負の生存率はあり得ないので、-5%の代わりに0%)、または乗法的に、±20%は12%から18%であり得る。)

10

## 【0079】

前に説明されているユーザインターフェースと同様に、図23～図24のインターフェースは、第1のパネル400内の識別子の提示を修正するための制御パネル412を含む第2の領域410を備え得る。ここでもまた、そのインターフェースと同様に、制御パネルは、ユーザがスケールの正側および負側に均一または独立した選択を行うことを可能にし得る。特に、図24を見るとわかるように、この事例の制御パネル412は、ユーザが外れ値の検索において正および負の範囲を独立して選択することを可能にする。各選択を行った後、ユーザインターフェース32は、ユーザによって選択されたゾーン内に入るインジケータを、そのゾーンの外側に入る外れインジケータから覆うか、見えにくくするか、強調表示をなくすか、取り除くか、またはその他の何らかの形で区別するように動的に調整し得る。上で説明されているように、x軸およびy軸の構成により、このユーザインターフェース32は、どの外れ値グループが代表的な患者/グループから最も遠く隔たっているかをユーザが素早く識別することを可能にするように構成され得るが、それは、その外れ値グループが、x軸から、正の方向、負の方向、または両方向に最も遠く隔てられているからである。同様に、ユーザインターフェース32は、どの患者グループが最大数の患者を有しているかをユーザが素早く視覚的に決定することを容易にするように構成され得るが、それは、そのグループが、正の方向、負の方向、または両方向にy軸から最も遠く隔てられているからである。さらに、軸の組合せは、どのインジケータがさらなる検査を正当化するかに関してユーザが素早く視覚的に決定することを、たとえば、どのインジケータが逸脱/外れ値の程度と患者サイズとを理想的な形で両立させるかをユーザが視覚的に決定することによって行うことを可能にし得る。

20

30

## 【0080】

上で説明されているいずれかの外れ値ユーザインターフェースに関して、インターフェースは、図24を見るとわかるように、たとえば、インターフェースの第1の領域内でそのインジケータ436をクリックすることによって、システムが所与のインジケータに対応するユーザ入力を受け取ったときに、選択されたノードに特有の情報を提供する第3の領域440をさらに備え得る。一態様において、その追加情報は、第1の領域のインターフェースを生成するために使用される母集団全体の値と比較したときに評価されている1つ/複数の基準の比較を含み得る。この領域内の情報は、レコードセット内の患者の総数、レコードセットが1つまたは複数の異なる基準に基づきフィルタ処理された患者の数、およびインラインプロットの一部としての選択されたノードの母集団サイズの識別も含み得るが、これらのサイズ比較は、外れ値グループの潜在的有意性についてユーザに知らせるのに役立つ。

40

## 【0081】

それに加えて、上で説明されているいずれかの外れ値ユーザインターフェースに関して

50

、外れ値の存在を決定するアルゴリズムは、図25Aおよび図25Bに示されているような二分木500に基づくものとしてよい。このような木を生成するために、システムは、各特徴をそれぞれのカテゴリに分離し得る。各カテゴリについて、システムは、コホートのどのサブセットが無増悪生存率対非生存の最大の広がりを持つかを決定し、最大の広がりを生成した分割された特徴をノード間のエッジとして、特徴それ自体をノードとして扱うものとしてよい。システムは、葉に遭遇するまでこの解析を続けるものとしてよい。たとえば、突然変異列は「突然変異あり」と「突然変異なし」とに分けられ、年齢オプションは「50歳超」と「50歳未満」とにユーザによって設定され得る。次いで、システムは、生存に対する最大のカットオフ年齢が何であることを決定し、それを二分決定点として使用し得る。各々2つのグループに分割する二分選択を有するこれらのカテゴリのすべてのうちで、システムは、どれがよりよい生存率を有し、どれがより悪い生存率を有するかを決定し、それらの決定をすべての列にわたって比較して最大の差異を有するグループを見つけるものとしてよい。最大の差異を有するカテゴリは、追加のノードで分割し続ける木中で分割された最初のノードであり、グループに対するカテゴリ基準が各ノード間の辺である複数の枝を形成する。枝の各々は葉で終端し、その葉は、それより上の分割に従ってコホート内で最も高いPFSを有する人々のグループを識別するために、前に来たすべての特徴の分割にすぎない。一態様において、システムは、各葉を外れ値として扱ってよい。代替的に、外れ値は、いくつかの、特に乖離している特徴であってよい。たとえば、外れ値の葉は、ユーザ入力した値または期待値から何らかの閾値だけ逸脱したものの、たとえば、期待閾値から1標準偏差またはそれ以上離れたものであってもよい。

10

20

**【0082】**

いくつかの事例において、システムが葉まで完全に外挿するときには枝のデータが失われ得る。そのような事例では、システムは、外れ値患者と共通の、現在の患者が有する特徴を走査し、より高い外れ値を有する患者の新しいバケット(葉/ノード)内にそれらを置き得る臨床プロセスの変更を提案し得る。たとえば、枝がノード内で高いPFSを有するが、枝が葉に解決するときまでに区別を失う場合、システムは、最も高いPFSを有するノードを葉として識別するものとしてよい。

**【0083】**

母集団に対する予想生存率を生成するために、システムは、データセット14内の患者の生存率に基づき構築された予測アルゴリズムに頼るものとしてよい。代替的に、システムは、いくつかの癌または治療に対するFDA公開PFSなどの、PFS予測に対する外部ソースを使用し得る。次いで、システムは、外れ値を決定するために、予想生存率を母集団について観察されたPFS率と比較し得る。

30

**【0084】**

特定の一実施形態において、患者の1つまたは複数の外れ値グループを識別するための方法が提供される。方法は、患者のコホートを選択するステップを含み、コホートは複数の患者を含む。コホートの選択は、特定の疾患などの特定の病状を有する患者のグループを識別することに基づき得る。特定の一実施形態において、コホートは、非小細胞肺癌または乳癌を有する患者のグループ(たとえば、数十人、数百人、数千人、またはそれ以上)を含み得る。他の基準に基づく他のグループ化も可能である。

40

**【0085】**

様々な実施形態において、方法の次のステップは、患者のコホートに対する平均生存率を計算することを含み得る。たとえば、利用可能なデータに基づき、これらの患者が平均して特定の時間(たとえば、63ヶ月などの月数)の間生存していることが決定され得る。

**【0086】**

いくつかの実施形態において、方法の別のステップは、患者のコホートに関連付けられている複数の臨床的または分子的特性を選択することを含み得る。患者のコホートに関連付けられている臨床的または分子的特性は、遺伝子マーカー、患者に実施された手技、患者に与えられた薬物治療、患者が診断を受けた年齢、患者が治療を受けた年齢、またはライフスタイルインジケータのうちの1つまたは複数を含み得る。特定の実施形態において

50

、患者の臨床的または分子的特性は、患者の喫煙状態(たとえば、はい、いいえ、不明)、患者に関連付けられているDNA突然変異(たとえば、KRAS、BRAF、EGFRなど)、診断もしくは治療時の患者の年齢(たとえば、18～115歳などの特定の年齢範囲内の1つまたは複数の整数)、または患者が受けた1つもしくは複数の治療法もしくは医薬品を含み得る。

【0087】

いくつかの実施形態において、患者のコホートに関する情報は、木構造を生成するために使用されてもよく、木構造のノードは、外れ値である1人または複数の患者、すなわち、条件の所与のセットに対して有意に異なる生存率(より短いまたはより長い)を示している患者を含み得る。したがって、木構造を生成するために、複数の特性のうちの各特性について、方法は、特性に関連付けられている複数のデータ値を識別することを含み得る。特性に関連付けられている複数のデータ値のうちの各データ値について、方法は、患者のコホートを、複数の患者の第1のサブグループと第2のサブグループとに分割することを、複数の患者のうちの各患者が外れ値時間期間において生存したかどうかなどの基準に基づき行うことと、第1のサブグループ内の患者数と第2のサブグループ内の患者数との間の差を決定することと、第1のサブグループ内の患者数と第2のサブグループ内の患者数との間の最大の差である差を結果としてもたらすデータ値を選択することを含み得る。

10

【0088】

この手順は、各特性の各データ値について繰り返され得る。たとえば、特性が年齢に係る実施形態については、データ値は、年齢18、19、20、21、...などの低い年齢範囲から始まり、年齢115(または別の好適な値)などの上限値に至る、年齢の範囲を含む。次いで、特定の一例では、年齢=20で、時間期間がx年(たとえば5年)である場合、患者の第1のコホートは、20歳の診断からx年後に死亡した者であり、患者の第2のコホートは、20歳の診断からx年以内に死亡しなかった者であるものとしてよい。

20

【0089】

差を決定するために、特定の時間内に生存しなかった患者の数は、患者の第1のサブグループと考えられ、特定の時間内に生存した患者の数は、患者の第2のサブグループと考えられる。次いで、各特性に関連付けられている各データ値について、第1のサブグループの患者の数と第2のサブグループの患者の数との差が決定される。この差は、第1のサブグループと第2のサブグループの患者の総数で除算され、0と1との間の10進値で表され得る(たとえば、20歳の診断からx年後に400人の患者が死亡し、20歳の診断からx年後に100人の患者が死亡しなかった場合、差 $400-100=300$ となり、これは2つのグループ内の総数である500で除算され、差0.6を得る)。最大のこのような差を有する特定のデータ値は、木構造のノードを決定するために手順が実行されている間保持され得る(たとえば、最大の差は、年齢=44で0.7の差であり得る)。

30

【0090】

方法は、第1のサブグループ内の患者の数と第2のサブグループ内の患者の数との間の最大の差を結果としてもたらすデータ値に基づき木構造の新しいノードを作成することをさらに含み得る(たとえば、年齢=44に対してノードが1つ作成され得る)。特定のデータ値が最大の差を有するものとして識別された後、方法は、次いで、第1のサブグループに基づき新しいノードから第1の枝を作成することと、第2のサブグループに基づき新しいノードから第2の枝を作成することを含む、ノードから枝を作成することを含み得る。潜在的なノードのいくつかの例は、喫煙=あり、差=0.8、DNA突然変異=KRAS、差=0.78、年齢=82歳、差=0.9、性別=男性、差=0.6を含み得る。この情報に基づいて、「年齢」特性は最大の差を有しており、選択され、82歳以上の年齢および82歳未満の年齢に基づく枝が作成され得る。

40

【0091】

木構造は、各特性および各特性の各データ値についてコホートをサブグループに分割するステップを含む、上記のステップを繰り返すことによって構築され続け得る。その後繰り返される各ステップにおける開始コホートは、開始点である特定のノード内の患者グループである。この手順は、第1のサブグループおよび第2のサブグループ内の患者に基づき

50

各ノードにおいてそれぞれ繰り返される。この手順は、次の条件、すなわち、(1)最大数のノードまたは枝が作成されているか、または(2)ノードが最小数より少ない数の患者を含む、の一方または両方が満たされるまで続く。手順が完了したときに、方法は、患者の外れ値グループを含む木構造から少なくとも1つのノードを識別することを含むものとしてよい。

【0092】

スマートコホート

様々な実施形態において、疾患の増悪および/または生存の可能性が予想とは実質的に異なる、たとえば、予想よりも著しく長いまたは短い、患者の1つまたは複数のコホートの識別を円滑にする予測モデルが開発され得る。次いで、これらのコホートからの情報が調べられ、コホートの生存率プロファイルに潜在的に寄与することもあり得る1つまたは複数の主要な要因を識別し得る。スマートコホートの識別は、特定の患者に対する精密医療結果を提供し、投薬研究を対象とする注目している潜在的な領域の識別、および/または投薬患者の対象範囲を拡大するための予想外の潜在的な可能性の識別を補助するために使用され得る。

10

【0093】

一組の患者タイムラインが与えられた場合、様々な実施形態において、スマートコホートモジュールには3つの目的があり、次の質問のうちの1つまたは複数に答えることを試みる。

【0094】

1.患者のタイムラインにおける各事象ポイントで測定された、各患者がY年を超えて生存する(または、少なくともY年間無増悪で生きる)尤度(すなわち、「生存率」)は何か。

20

【0095】

2.予想生存転帰に最も影響を及ぼす主な要因は何か。

【0096】

3.患者のどのサブセットが、ユーザ指定アンカータイムライン事象(たとえば、病期IVの診断時)において、予想に対して、生存率プロファイルに関して外れ値コホートとして目立つようなこれらの要因の組合せを示すか、またこれらの患者の特性は何か。

【0097】

この問題は、特徴状態の時点スナップショットのある、時系列モデリングの観点、および二項分類目標からアプローチされ得る。いくつかの実施形態において、注目する患者グループを識別するのを補助するために、木構造に基づく教師ありクラスタリングアプローチが使用され得るが、他の実施形態では、他の解析および可視化方法も含まれる。

30

【0098】

問題の固有の時間的性質は、アンカーポイントTでの標的生存率が、ポイントTより前に起きたことに依存するのと同様にポイントTの後に患者に起きたことに依存する可能性があるという事実によって複雑になる。そのようなものとして、予想される将来の生存率は、事象履歴だけを使用して単純にはモデル化できず、将来の事象は、レコメンダとしてモデルを無効化することまたは情報漏洩を特徴に誤って導入することなしには、モデルに含めることはできないが、これがあると結果として過剰適合をもたらす可能性がある。

40

【0099】

いくつかの実施形態において、ハイブリッド2モデルアプローチが採用され得る。このアプローチの1つの部分では、履歴のみモデルが各時点での「期待値」を導出するように訓練され、アプローチの別の部分では、前向きクラスタリングモデルが、関連する特徴とともに、予想生存率と観察生存率との間の乖離を分離するように開発される。

【0100】

したがって、いくつかの実施形態において、ハイブリッドアプローチは次を含み得る。

【0101】

1.タイムライン上の各事象ポイントで導出される、後ろ向き特徴のみを利用するデータセットを構築する。

50

## 【0102】

2.そのようなデータセット上でモデルを訓練して、各時点における予想される将来の生存率の予測を導出する。

## 【0103】

3.すべての履歴情報コンテンツを使用して最良推定事前確率として働くように各時点でのこれらの予想生存率予測にタグを付ける。

## 【0104】

4.各時点で「前向き」特徴セットを構築し、暗黙の生存持続時間情報が特徴に組み込まれないようにすることを確実にする(いくつかの場合において、同様に履歴的事前確率がこのセット内に特徴としても含まれ得る)。

## 【0105】

5.前向き特徴セットを使用して「要約/クラスタリング」モデルを訓練する。

## 【0106】

この時点で、「訓練」ステップに続いて、この部分に対する特徴がどれだけ前向きであり得るかを制限するかどうかに関する決定がなされ得る。たとえば、1年生存可能性を予測しようとしている場合に将来2年間にわたって観察される特徴を含めることは意味がないことがある。それに加えて、アンカー事象から遠く離れて生じる特徴に低い重要度を与えることも考えられる。最後に、注目している転帰事象の後に観察される事象ポイントを、たとえその事象がX年境界内で発生したとしても、除外することを考え得る。たとえば、最初の増悪事象が6ヶ月以内に観察され、2年PFSを予測している場合、その患者については、6ヶ月から2年の間のすべての事象を除外すべきである。

## 【0107】

6.前向きクラスタの各々について、前向きモデルに基づき予想生存率予測を実際の生存率と比較し、予想生存率予測からの乖離が大きいクラスタを、その構成前向き特徴セットとともに識別する。

## 【0108】

したがって、モデルは、特定のサブクラスタに対する予想生存率予測が異なるクラスタに対する予想生存率予測よりも高いかどうかにとらわれず、将来の事象が事前事象によって予測される予想生存率にどのような影響を及ぼし得るかを決定することに向けられている(ただし、予想生存率予測における乖離の根本原因も注目する)。すなわち、次の行動が患者の生存に影響を及ぼすかどうか、または患者生存がすでに経験済みの事象によってもっぱら決定されるのかどうかを知ることは注目すべきことである。

## 【0109】

予測モデルは、患者の生存に関する情報とともに患者の病歴および治療に関する情報を使用して、多数の患者からのデータに基づき実装され得る。多数の患者からのデータの時間的アライメントを行うために、1つまたは複数のアンカーポイント(「患者時点」とも称される)がデータ内で識別され得る(図26)。アンカーポイントは、患者のすべてまたは少なくとも多くに共通し、疾患増悪などの事象に関するデータの時間経過を標準化するのに役立つ時点を識別する。アンカーポイントは、最初の診断の時期、最初の転移の時期、または最初の治療の時期などの事象を含むものとしてよいが、他のアンカーポイントの事象も可能である。図26は、共通のアンカー事象に基づく患者 $P_1$ 、 $P_2$ 、 $P_3$ 、...、 $P_n$ のタイムラインのアライメントを示している。

## 【0110】

いくつかのアンカーポイント事象の時間に関しては何らかの不正確さがあり得、たとえば、最初の診断の日、正確さの欠如を考慮して、患者が最初に症状に気付くか、または診断を受けるために医者診察を受ける時間に起因して所与の患者に対して(たとえば、病気が始まったときに関して)数週間早くまたは遅く発生し得る。したがって、いくつかの実施形態において、アンカーポイントは、モデル化手順に柔軟性を加えることができるアンカーポイントの日付の前および/または後の許容範囲窓を含み得る。様々な実施形態において、許容範囲窓は、 $\pm 1$ 日、3日、1週間、2週間、1ヶ月、2ヶ月、3ヶ月、または他の好

10

20

30

40

50

適な期間であってもよい。図26は、アンカー事象(1月1日に設定)と、その後の12ヶ月の増悪窓を示している。アンカー事象は、それに関連付けられている±15日の許容範囲窓を有し得る。それに加えて、増悪窓は、3ヶ月の許容範囲窓を有していてもよく、したがって、増悪基準点窓は、1月1日の3ヶ月前から10月1日まで時間をさかのぼって延びるものとしてよい。

【0111】

予測モデルに関して、様々な実施形態において、複数のデータが、一定期間にわたって、複数の患者について取得されるか、または受け取られる(たとえば、患者の診断時から現在または死亡時までの患者の病歴の各々にわたるタイムスパン、病歴は診断前に始まってよい)。

10

【0112】

データは、各患者のデータの対象となる時間期間内に発生する複数の患者時点(アンカーポイント)を識別するように処理される。上で説明されているように、アンカーポイントまたは患者時点は、医療提供者、遺伝子シーケンシング機関、病院の外来または入院施設などの、医療を提供するか、または医療情報を取得する個人もしくは施設との任意の相互作用を含む、医療システムとの任意の患者の相互作用に関連付けられている時点を含んでもよい。患者時点は、患者データの受け取ったセット内の各データに付けられているか、または関連付けられている日付によって識別され得る。

【0113】

一般に、時間的特徴と静的特徴の両方が患者データから導出され得るが、この病期での解析は、将来の情報漏洩を回避するために純粹に後ろ向きである。特徴の異なるカテゴリまたはクラスは、「最後/最初のXXX以降の時間」、「XXXの数」、または「人口統計」を含む。特徴を抽出することは、複数のルックバック水平線を含んでもよく、たとえば、特徴は、過去12ヶ月を限界とするか、または連続的履歴解析に基づいてもよい。

20

【0114】

特定の一例において、仮説的患者Aについて、4つの時点、生検収集日、2018年7月1日(高いSNP効果が識別されているKRAS PL1S147GLU変異)、アナストロザルおよびロチニブの投与開始、2018年8月1日、放射線療法の実施、2018年11月1日、治療転帰の報告:病期1から病期2への疾患の増悪、2019年1月1日、実施されたイメージング、2018年7月1日および2018年11月1日が識別され得る。他の患者B、C、D...は、各々、同じ事象のいくつか(たとえば、診断、投薬開始、イメージングなど)に対応するか、または異なる事象に対応するか、またはいくつかの同じ事象およびいくつかの異なる事象の組合せに対応し得る時点のそれぞれのセットを有する。

30

【0115】

患者の各々に対するデータおよび各患者時点に対するデータに基づき、転帰事象に対する転帰標的は、水平線時間窓内で計算されてよく、複数の事前特徴が識別されてよく、患者時点における複数の事前特徴の各々の状態が決定され得る。転帰事象は、増悪または死亡などの、患者および/または疾患の状態を含むものとしてよく、転帰標的は、患者時点/アンカーポイントから特定の水平時間窓内に転帰が生じるかどうかを示す「はい」または「いいえ」などの標的ラベルにより、終点の日付とともに記述され得る。水平線時間窓は、3ヶ月、6ヶ月、9ヶ月、12ヶ月、24ヶ月、36ヶ月、48ヶ月、または60ヶ月などの任意の好適な期間を含み得る。

40

【0116】

仮説的患者Aの場合、時点の6ヶ月以内に生じる増悪事象の解析は次の通りである。

【0117】

患者A:2018年7月1日--12ヶ月以内の増悪--はい、2019年1月1日

【0118】

患者A:2018年8月1日--12ヶ月以内の増悪--はい、2019年1月1日

【0119】

患者A:2018年11月1日--12ヶ月以内の増悪--はい、2019年1月1日

50

## 【 0 1 2 0 】

患者A:2019年1月1日--12月以内の増悪--無効

## 【 0 1 2 1 】

患者Aに対するデータは、2019年1月1日に病期1から病期2への増悪のレポートの情報を含んでいたため、最初の3つの時点の各々について、「12か月以内の増悪」に対する有効な転帰標的「はい」がある。しかしながら、最終時点に対する解析は、この日付以降、モデルに情報を与える利用可能な患者情報がないので、「null」と表示される。この日付で増悪が報告されたけれども、この日付以降、患者Aに対して利用可能なさらなる情報は無い。

## 【 0 1 2 2 】

事前特徴は、患者の病状および/または治療に関係する様々な特徴を含み得る。様々な実施形態において、事前特徴は、他にもカテゴリがあるがとりわけ、時間/時刻ベースの事象もしくは特徴、構造的もしくは生物学的特徴、または分子的/遺伝的特徴を含み得る。特定の実施形態において、事前特徴は、特定の薬物療法を開始してから経過した時間、特定の薬剤を服用してから経過した時間、最後の進行的治療転帰(たとえば、薬剤に対する患者の反応)から経過した時間、転移から経過した時間、現在までの最大の腫瘍サイズ/最後に記録された腫瘍サイズ、識別されたSNPの最も深刻な効果(たとえば、低効果、高効果)、またはRNA特徴(たとえば、遺伝子/転写物毎の発現レベル)のうちの1つまたは複数を含み得る。いくつかの実施形態において、データは、特徴空間の次元を減らすために、オートエンコーダを使用するなど、追加処理を必要とし得る。

## 【 0 1 2 3 】

各事前特徴の状態は、患者時点の各々において決定され得る。仮説的患者Aについて、4つの患者時点の各々に対する3つの特徴(投薬Aを開始してから経過した時間、最後のイメージングから経過した時間、およびラボAによって識別された最も高いSNP効果)の状態が以下に示されている(最初の患者時点における「投薬Aを服用してから経過した時間」の値は、患者Aが次の時点まで投薬Aを服用していなかったため「null」であることに留意されたい)。

## 【 0 1 2 4 】

患者A:2018年7月1日

## 【 0 1 2 5 】

投薬Aを開始してから経過した時間:null

## 【 0 1 2 6 】

最後のイメージングから経過した時間:0日

## 【 0 1 2 7 】

ラボAによって識別されるような最も高いSNP効果:生殖細胞:KRAS:高(5)

## 【 0 1 2 8 】

患者A:2018年8月1日

## 【 0 1 2 9 】

投薬Aを開始してから経過した時間:0日

## 【 0 1 3 0 】

最後のイメージングから経過した時間:1ヶ月

## 【 0 1 3 1 】

ラボAによって識別されるような最も高いSNP効果:生殖細胞:KRAS:高(5)

## 【 0 1 3 2 】

患者A:2018年11月1日

## 【 0 1 3 3 】

投薬Aを開始してから経過した時間:3ヶ月

## 【 0 1 3 4 】

最後のイメージングから経過した時間:0日

## 【 0 1 3 5 】

10

20

30

40

50

ラボAによって識別されるような最も高いSNP効果:生殖細胞:KRAS:高(5)

【0136】

患者A:2019年1月1日

【0137】

投薬Aを開始してから経過した時間:5ヶ月

【0138】

最後のイメージングから経過した時間:2ヶ月

【0139】

ラボAによって識別されるような最も高いSNP効果:生殖細胞:KRAS:高(5)

【0140】

次に、有効な転帰標的を有する複数の時点のうちの各患者時点について、また水平線時間窓および転帰事象の各組合せについて、複数の前方特徴が識別され得る。水平線時間窓と転帰事象との組合せは、「6ヶ月以内の増悪」、「12ヶ月以内の増悪」、「24ヶ月以内の増悪」、「60ヶ月以内の増悪」、「6ヶ月以内の死亡」、「12ヶ月以内の死亡」、「24ヶ月以内の死亡」、「60ヶ月以内の死亡」などを含み得る。

【0141】

患者Aでは、「12ヶ月以内の増悪」という水平線時間窓/転帰事象の組合せを使用することで、前方特徴は以下を含み得る。

【0142】

患者A:2018年7月1日～

【0143】

患者は、時点後から終点の日付までに薬Aを服用するか(はい)

【0144】

患者は時点前に薬Aを服用したか(いいえ)

【0145】

ラボAによって識別されるような最も高いSNP効果:生殖細胞:KRAS:高(5)

【0146】

患者A:2018年8月1日～

【0147】

患者は、時点後から終点の日付までに薬Aを服用するか(いいえ)

【0148】

ラボAによって識別されるような最も高いSNP効果:生殖細胞:KRAS:高(5)

【0149】

患者は時点前に薬Aを服用したか(はい)

【0150】

患者A:2018年11月1日～

【0151】

患者は、時点後から終点の日付までに薬Aを服用するか(いいえ)

【0152】

ラボAによって識別されるような最も高いSNP効果:生殖細胞:KRAS:高(5)

【0153】

患者は時点前に薬Aを服用したか(はい)

【0154】

この時点で、複数の患者に対する予測の複数のセットが、複数の事前特徴および複数の前方特徴に基づき生成され、予測モデルが、機械学習を使用して予測のセットに基づき生成され得る。いくつかの実施形態において、予測モデルは、勾配ブースティングを使用して生成され得る。

【0155】

予測の複数のセットは、複数のフォールドに分割されるものとしてよく、各フォールドは、各患者に対するデータが同じフォールド内に保たれるように複数の患者のサブセット

10

20

30

40

50

またはサブグループに対応するデータを含む(図28)。したがって、勾配ブースティングなどの機械学習手順は、フォールドのサブセットを使用して訓練され得る。たとえば、8つのフォールドがある場合、8つのフォールドのうち7つのフォールドで勾配ブースティングアルゴリズムが実行されてよい。訓練に使用されていない残りのフォールドは、予測目的でモデルを通して実行され、予測結果と実際の結果との間の違いは、訓練のその後のラウンドが実行される前にモデルを調整するために使用され得る。これは、訓練ステップから省かれ、モデルの予測および/または調整に使用される、異なるフォールドで繰り返され得る。より一般的には、N個のフォールドがある場合、訓練はX N個のフォールド上で実行され、予測はN-X個のフォールドを使用して実行され得る。予測モデルを生成する際に、学習率、木の最大深度、最小葉サイズなどを含む、様々なパラメータが調整され得る(モデルのタイプに依存する)。目標は、標的結果につながるすべての患者の事前特徴の間の関係を学習するモデルである。予測は、モデルからの各患者時点から受け取られ、対応する転帰標的と結び付けられるか、関連付けられる。いくつかの実施形態において、8つのフォールドが交差検証され、追加の2つのフォールドが別の試験目的のために完全なホールドアウトとなり得る。フォールドは、標的、性別、癌、患者事象カウントなどの複数の特徴の組合せによって層状であってよい。

10

## 【0156】

複数の予測を生成した後、この情報は、1つまたは複数の「スマートコホート」、すなわち、疾患の増悪および/または生存の可能性が予想とは実質的に異なる、たとえば、予想よりも著しく長いまたは短い患者の1つまたは複数のコホートを識別するために使用され得る。一般に、決定木が、予測情報を使用して構築され、それにより、様々な潜在的なスマートコホートを識別し、最終的に決定木の様々な葉ノードにグループ化され得る。本明細書では、オフラインスマートコホートおよびオンラインスマートコホートと称される決定木を構築するための2つのアプローチが開示されている。

20

## 【0157】

## オフラインスマートコホート

いくつかの実施形態において、患者のコホートを識別するための方法が開発され得る。方法は、複数の患者を含む患者のコホート、たとえば、500人の乳癌患者のコホートを選択することを含み得る。一般に、コホートは、患者が共通の特定の病状、たとえば、特定の疾患を有することに基づき選択され得る。

30

## 【0158】

方法は、患者群の各々に関連付けられているアンカーポイントのセットから共通のアンカー時点を識別することであって、共通のアンカーポイントはコホート内の患者のグループの各々によって共有される、識別することも含み得る。すべての患者の間の共通点を選択することで、データの可視化を円滑にし、また、患者の利用可能なアンカーの各々において同じ患者がモデル内に複数回出現することを防ぐことを可能にする。可能なアンカーポイントには、診断の時間、治療の時間、転移の時間、および他の時間を含む。特定の実施形態において、診断の時間は、アンカーポイントとして選択され得る。

## 【0159】

患者のグループ内の各患者について、患者のグループの各々に関連付けられているタイムラインが、共通のアンカーポイントにアライメントされ得る。次に、12ヶ月以内の疾患増悪などの、転帰目標が識別され得る。その後、各々予測された標的値を含む、以前に生成された予測の複数のセットは、患者のグループの各患者について、また複数の前方特徴および複数の事前特徴の各々について、取り出され得る。予測は、Table 1(表1)に示されているような情報を含み得る。

40

## 【0160】

50

【表 1】

患者	標的予測	標的実際	特徴セット
A	0.95	1	A B C D
B	0.93	1	A C D F G
C	0.25	0	B D F
D	0.1	0	A C D G

Table 1

10

## 【0161】

より一般的に、「標的予測」は、「Xヶ月以内の生存確率(PFS)」、「Xヶ月以内の死亡」、「Xヶ月内に薬を服用する尤度」、「Xヶ月以内の他の標的の尤度」などの形式をとり、0と1との間の10進値の形式をとり得る。「標的実際」値は、本質的に1または0で示される2進数のはい/いいえの値であり、Xヶ月以内の事象の発生または非発生を表す。様々な実施形態において、特徴セットは、事前特徴および/または前方特徴、たとえば、「特徴および特徴モデル」の見出しの下に記載されているものを含む本明細書において開示されている特徴のいずれかを含んでもよい。事前特徴は、年齢、性別、治療(たとえば、薬剤、手技、療法など)、シーケンシング/ラボ/イメージングの結果のうちの1つまたは複数を含み得る。以下でさらに説明されている、前方特徴は、アンカーポイントと観察標的との間の将来生じる事象、治療などを含み得る。

20

## 【0162】

様々な実施形態において、数百もしくは数千(または他の、より大きな数)の決定木が、この情報を使用して、たとえば、外れ値手順について上で説明されている手順に類似する手順を使用して生成され得る。構築される決定木の各々について、複数の前方特徴および複数の事前特徴のうちの各特徴に対して、次の手順が実行され得る。

## 【0163】

- 患者のグループは、予測された標的値と実際の標的値との間の差に基づき第1のサブグループと第2のサブグループとに分割され得る。

30

## 【0164】

- 第1のサブグループ内の患者数と第2のサブグループ内の患者数との間の差が決定され得る。

## 【0165】

- 第1のサブグループ内の患者数と第2のサブグループ内の患者数との間の最大の差である差を結果としてもたらず特徴が、選択され得る。

## 【0166】

木構造の新しいノードが、第1のサブグループ内の患者数と第2のサブグループ内の患者数との間の最大の差を結果としてもたらず特徴に基づき作成され得る。第1の枝が、第1のサブグループに基づき新しいノードから作成され、第2の枝が、第2のサブグループに基づき新しいノードから作成され得る。次いで、決定木を構築するステップは、第1のサブグループおよび第2のサブグループ内の患者に基づき第1の枝および第2の枝の各々について繰り返され得る。これは、最大数のノードまたは枝が作成されたか、または特定のノードがすべてのノードおよび枝に対する最小数よりも少ない患者を含むかのいずれかによって定義されているように完了するので続き得る。

40

## 【0167】

決定木を構築する目標は、各患者について、特徴セットの特徴に基づき、どの特徴が予測と実際の転帰との間の差を最も正確に予測するかに基づき患者をクラスタリングすることによって標的に対する予測と実際の転帰との間の差を予測することである。

50

【 0 1 6 8 】

いくつかの実施形態において、この方法は、数百または数千の決定木全体にわたって所与の患者が他の患者と木の同じ葉ノードで終わる頻度を決定することによって類似度メトリックを決定することを含み得る。したがって、患者のグループの各患者について、方法は、数百または数千の決定木全体にわたって、複数の患者のうちの他の患者の各々との、複数の葉ノードの各々の中に出現する所与の患者の同時出現を識別することを含み得る。類似度メトリックは、構築され解析された数百または数千の決定木のすべてにわたって所与の患者が入るノードの総数で除算された同時出現の総和に基づき所与の患者について決定され得る。いくつかの実施形態において、患者-患者間の類似度メトリックのデータベースが、複数の患者の各々について類似度メトリックを決定することに基づき生成され得る。他の実施形態において、類似度メトリックは、たとえば、コホートレーダープロットとして表示されてもよい。さらに、データは、複数の特徴のうちの少なくとも1つを識別するために上で概要を説明されているステップのうちの1つまたは複数に関連して表示され得る。

10

【 0 1 6 9 】

方法は、新しい患者、すなわち、患者の初期グループとは異なる患者に対する類似度メトリックを決定することをさらに含み得る。新しい患者は、類似度メトリックを決定することに基づき複数の葉ノードのうち特定の葉ノードに対応する患者のサブグループとマッチングされ得る。次いで、新しい患者を患者のサブグループとマッチングさせることに基づき新しい患者について治療が識別され得る。さらに、患者-患者間の類似度メトリックのデータベースは、共有されている事前特徴または共有されている前方特徴などの共有特徴を有する患者の特定の cohorts を識別するように次元削減アルゴリズムを使用して処理され得る。一般に、次元削減は、サブグループ1-kの各々が患者 cohort 全体(標準母集団グルーピング)から識別されるグルーピングにわたって共通のいくつかの特性を有するいくつかのサブグルーピング(K個のサブグループなど)を識別する。

20

【 0 1 7 0 】

オンラインスマートコホート

複数の予測に加えて、システムは、転帰標的と、転帰標的に対応する複数の前方特徴のサブセットと、複数の患者のサブセットを含む患者の cohort とを受け取るものとしてよい。Cohort は、注目する状態または形質を共有するグループであってよく、たとえば、Cohort は、20,000人の乳癌患者のグループであってよい。次いで、このグループは、さらなる調査のために注目する1つまたは複数の特定のサブグループを見つけるために決定木を使用して細分化される。

30

【 0 1 7 1 】

Table 2(表2)は、受け取られ得る予測データのタイプの一例を示している。

【 0 1 7 2 】

【表 2】

患者	時点	予測	標的	特徴セット
A	T1	.95	1	CD
A	T2	.75	1	BC
A	T3	.66	0	ABCD
B	T4	.92	1	A EFG

40

Table 2

【 0 1 7 3 】

前方特徴は、患者に関係する様々な将来のアクションまたは状態を含んでもよく、いく

50

つかの実施形態において、特定の病状を有する患者に助言するために使用され得る。前方特徴のうちのいくつかは「アクション可能」であってもよく、すなわち、所与の患者が予後または結果を変えるために行えることを含んでいてもよい。たとえば、医師または他の臨床医は、患者の予後を改善するためにいくつかのステップまたはアクションをとることが可能である(たとえば、薬剤または薬剤の組合せを処方する、外科手術、化学療法、もしくは放射線などの特定の治療を処方する、シーケンシングのために腫瘍サンプルを送りDNAマーカーに対する検査などの分子的情報を受け取る)。特定の分子的特徴は、得られた分子的情報が後続のアクションまたはステップに関連するかどうかに基づきアクション可能であるとみなされる場合もあれば、そうでない場合もある。様々な実施形態において、「より低いN病期」、「ヘモグロビン濃度を高くする」など、患者が自分の制御の範囲内

10

## 【0174】

様々な実施形態において、この情報は、特定の患者グループにカウンセリングするために使用することも可能であり、たとえば、X突然変異を有するN病期患者については、治療AおよびBを一緒に行えば12ヶ月以内の生存確率(PFS)を改善する。たとえば、KRAS突然変異を有する病期4の乳癌患者は、コホート内での配置に基づき増悪することが予想され(90%の増悪予測)、最初の転移の時点の選択されたアンカーポイント以降の予測に基づき12ヶ月以内のPFSを改善するために介入療法としてアナストロザルとロチニブを併用すべき

20

## 【0175】

予測の例は、Table 3(表3)に示されているように、0と1との間の確率値として表される、患者AおよびBならびに時点T1(2018年1月1日)およびT2(2018年5月1日)に対する、12ヶ月以内の生存確率の予測を含む。

## 【0176】

## 【表3】

患者	時点	予測
A	2018年1月1日	.95
A	2018年5月1日	.75
B	2018年1月1日	.92

Table 3

30

## 【0177】

転帰標的は、Table 4(表4)に示されているように、0または1として与えられる、12ヶ月以内の生存確率であってよい。

40

## 【0178】

50

【表 4】

患者	時点	予測
A	2018年1月1日	1
A	2018年5月1日	1
B	2018年1月1日	1

Table 4

10

## 【0179】

以下は、12ヶ月以内の生存確率に対応する前方データを含む転帰標的に対応する複数の前方特徴(FD1、FD2、FD3、各々以下に示されている)のサブセットの一例である。

## 【0180】

2018年1月1日:

## 【0181】

FD1(患者はアナストロザルとロチニブを服用する):(はい)

## 【0182】

FD2(患者は放射線治療を受ける):....

20

## 【0183】

FD3(患者は手術を受ける):....

## 【0184】

2018年5月1日:

## 【0185】

FD1(患者はアナストロザルとロチニブを服用する):(はい)

## 【0186】

FD2(患者は放射線治療を受ける):....

## 【0187】

FD3(患者は手術を受ける):....

30

## 【0188】

システムは、アンカーポイントまたは患者時点、たとえば、最初の診断の時間、最初の転移の時間、最初の治療の時間なども受け取るものとしてよい。

## 【0189】

複数の前方特徴のサブセットが選択され得る。これらの特徴は、薬剤(将来および過去)、さらにはシーケンシング(体細胞シーケンシング(将来または過去)、生殖細胞シーケンシングなど)を含み得る。アンカーポイントを有するコホート内の各患者について、予測モデルは、複数の前方特徴の選択されたサブセットを備えるものとしてよく、複数の予測の各々と転帰標的との間の差が決定され得る。

## 【0190】

たとえば、モデルは、以下のようなデータを受け取るものとしてよい。

40

## 【0191】

患者A:[.95~1]、[薬剤およびシーケンシングデータセット]

## 【0192】

患者B:[.92~1]、[薬剤およびシーケンシングデータセット]

## 【0193】

患者C:[.63~0]、[薬剤およびシーケンシングデータセット]

## 【0194】

データは、患者およびそれぞれの特徴のN×Mテーブルを含み得る「アンカーポイントでの薬剤およびシーケンシングデータセット」などの情報を含み得る。それぞれの特徴は

50

、以下のような情報を含み得る。

【0195】

患者A:2018年7月1日(アンカーポイントの日付)-

【0196】

列1:患者は、時点後から終点の日付までに薬Aを服用するか(はい)

【0197】

列2:患者は時点前に薬Aを服用したか(いいえ)

【0198】

列3:ラポAによって識別されるような最も高いSNP効果:生殖細胞:KRAS:高(5)

【0199】

その後、複数の前方特徴の選択されたサブセットの各特徴について、複数の予測の各々と転帰標的との間の最大の差を決定することに基づき決定木が生成され得る。決定木は、複数の葉ノードおよび1つまたは複数の枝ノードを含むものとしてよく、1つまたは複数の枝ノードの各々は、葉ノードまたは枝ノードを含む枝の対を含むものとしてよく、枝は、複数の前方特徴のサブセットから選択された特徴に基づき形成される。

【0200】

決定木の複数の葉ノードの各々は、患者のコホートからの多数の患者を含み得る。いくつかの実施形態において、決定木は、複数の葉ノードのうち特定の葉ノード内の患者の数が患者の最小数より少なくなるまで複数の予測の各々と転帰標的との間の差に基づき分割を続けてもよい。他の実施形態では、決定木は、決定木のレベルの数が特定の数に達するまで、すなわちレベルの最大数に等しくなるまで、複数の予測の各々と転帰標的との差に基づき分割を続けてもよい。特定の一例において、特徴「KRAS体細胞:病歴 3」に関する各患者の状態は、このマーカーに対する各患者の病歴重要度値が3より大きい(高い重要度)かどうかに基づき枝ノードを2つの枝に分割するために使用され得る。

【0201】

決定木の葉ノードは、注目するコホートを識別するために使用され得る情報を提供する。いくつかの場合において、葉ノードは、予測値が平均して標的値よりもかなり高いので予測標的に対して高い値を有し得る。上記の例における患者Cについて、予測は患者Cの病状が増悪する可能性が高いことを示していたが、実際にはそうならなかった。他の場合には、葉ノードは、また、「予測値-標的値」の差に対する低い負の値を生成することもあり得、たとえば、予測値-標的値は、 $[.05-1]=-0.95$ であってよく、これは、患者の病状が増悪することはありそうもないが、場合によってはそれでも増悪し得ることを示している。しかしながら、いくつかの場合において、葉ノードは、おおよそゼロの値を有することもあり得、これはモデルが正確な予測を行ったことを示している。スマートコホート手順は、患者の実際の転帰が予想された結果から大きく逸脱した事例に焦点を当てているが、これは、患者のこれらのグループが、疾患増悪の進行軌道を変えるために何ができるかについての情報を提供することができるからであり、一方、予測値-標的値差がゼロに最も近いコホートは、どのような特徴が信頼性の高い予測にとって最も重要であるかをモデルに知らせる。

【0202】

いくつかの実施形態において、アナリティクスが決定木の葉ノードの1つまたは複数に対して実行されてもよく、アナリティクスは葉の枝を解析してそれらを意味あるものにする。モデルに送られた特徴のサブセットのみが、分割を作成するために考慮される。特徴のサブセットが「薬剤」および「分子」を含む一実施形態において、特定の葉が「KRAS(体細胞)タンパク質(ポストアンカー)に対するバリエーション効果: 1」(分子的特徴)および「薬剤を服用しない:ベムプロリズマブ」(医学的特徴)を示し得る。このように、全体的な品質を改善し、分割および結果として得られる葉ノードの精度を改善するために、アナリティクスがデータに対して実行され得る。特定の場において(薬剤および分子的特徴が分割に使用される場合には関連しないけれども)、アナリティクスは、分岐情報を解析し、他の何らかの形で曖昧な情報を意味のあるものにするために使用され得る、すなわち、「性別

10

20

30

40

50

が男性でない」を示す情報は「性別が女性である」に設定され得る。

【0203】

分割が薬剤および分子的特徴に基づくモデルに関係する、別の事例では、アナリティクスは、データを特定のカテゴリおよび/または範囲にマッピングしてデータを意味のあるものにするために使用され得る。たとえば、範囲を以下のように提示され得る。

【0204】

KRAS(体細胞)タンパク質(ポストアンカー)に対するパリアント効果:= 1

【0205】

これは以下にマッピングされ得る。

【0206】

KRAS(体細胞)タンパク質(ポストアンカー)に対するパリアント効果:=1(「陰性」)

【0207】

ここで、「陰性」という用語は、「検査され、突然変異していないことが確認された」ことを示す(不明な状態とは反対に)。

【0208】

いくつかの実施形態において、ノードから枝を生成することにつながる解析は、結果として得られる葉ノード内の患者のすべてが特定の要件を満たすことを必要とする、すなわち、この手順は、枝を形成するためにコホート100%参加を必要とし得る。しかしながら、いくつかの場合において、このコホート100%参加という要件があるので木から導出される特徴が、統計的に関連性のあるコホート特徴を見逃す可能性がある。したがって、いくつかの実施形態において、葉のコホートのすべてに満たない数の患者によって共有される(たとえば、95%によって共有される)が、コホート全体の中の患者のすべて(たとえば、95%)によって共有されない特徴が特定の葉に含まれることを可能にするようにサブセット認識特徴効果(SAFE)アルゴリズムが実行され得る。

【0209】

様々な実施形態において、スマートコホートアルゴリズムは、観察モード(予測を使用せず、標的のみを使用する、たとえば、0もしくは1)またはアルゴリズムモード(予測を使用する、たとえば、予測-標的[.95-1])で実行され得る。

【0210】

SAFEアルゴリズムは、基礎となるモデルの再訓練を必要とすることなく、患者の選択された部分集団に基づき実行可能な特徴重要度ランクを返すように開発されている。患者母集団について事前訓練済み大域的な多癌型モデルからの予測が与えられた場合、SAFEアルゴリズムは、近似的な高レベル重要度ランクをインタラクティブに、素早く導出し得る。それに加えて、特徴重要度ランクは、大域的モデルを再訓練しなくても、母集団の選択されたサブセットコホートが与えられた場合に関連するようにインテリジェントに、動的に調整され得る。解釈能力を最適化するために、いくつかの実施形態において、SAFE特徴重要度アルゴリズムは、使用された基礎的機械学習モデルを関知せず、関連する特徴に適切な重要度を割り当てることを明確に処理するように作られ得る。また、SAFEアルゴリズムは、標的が必ずしも定義されているとは限らない「特徴+予測」データセットにおける特徴重要度を探索する機能を提供し得る。最後に、より連続的な特徴に対して、SAFEアルゴリズムは、特徴値の変化に伴う特徴重要度の変化をより深く探索することを可能にし得る。

【0211】

一実施形態において、SAFEアルゴリズムは、母集団平均予測を計算することを含み得る。次いで、アルゴリズムは、カテゴリ的特徴レベルを、予測値と母集団平均予測との間の差分として符号化することを含むものとしてよく、めったに起きないレベルは一緒にグループ化されてもよい。アルゴリズムは、連続特徴のクラスタリングまたはバケット化、およびこれらの特徴を前のステップのように処理することをさらに含み得る。次に、アルゴリズムは、各特徴について、カテゴリレベル毎の平均値(p-E(p))を集計することを含み得る。最後に、アルゴリズムは、各特徴について、すべての値の絶対値の頻度加重和として全体的特徴重要度を割り当てることを含み得る。

10

20

30

40

50

## 【0212】

上で説明されているアプローチを使用することでわかるように、このアルゴリズムは、重要度ランキングを導出するために標的変数の存在に明示的に依存せず、その代わりに特徴および予測のみを必要とする。そのようなものとして、これはラベル付けされていないデータセットに対して行われる予測に効果的に適用され、さらには異なるタイプの機械学習(ML)アルゴリズムから取得される予測に一般化することができる。

## 【0213】

図27Aおよび図27Bは、SAFEアルゴリズムの実施形態による適応型特徴ランキングの例を示している。図27Aは、乳癌患者にもっぱら基づく全体的モデルからの上位10個の特徴のリストを示している。図27Bは、結腸直腸病期4患者を対象とするサブセットを作成した後の図27Aからのデータセットからの上位10個の特徴のリストを示している。図27Bを見るとわかるように、結腸直腸患者に関連付けられる可能性の高いいくつかの特徴(たとえば、「historical-took\_medication:irinotecan」および「historical-took\_medication:bevacizumab」)は、結腸直腸病期4患者を対象とするサブセットにおいてより高いランキングおよびより高い値を有する。他方では、結腸直腸病期4患者に関係しない特徴(たとえば、「cancer:lung\_cancer」や「cancer:pancreatic\_cancer」)は、図27Bのリストには現れない。図27Cでは、図27Aおよび図27Bの例を続けており、相関特徴の取り扱いの一例を示している。図27Bから結腸直腸の例を続けると、図27Cは、2つの特徴、すなわち、「historical-took\_medication:irinotecan」および「historical-took\_medication:capecitabine」に基づく重複ダミー列を追加した後、これらの重複列は、予想されるように、結腸直腸病期4に関連付けられている他の値で適切にソートされることを示している。

10

20

## 【0214】

図27Dおよび図27Eは、SAFEアルゴリズムの実施形態によるサンプルレベル重要度割り当ての一例を示している。SAFEアルゴリズムの導出が与えられた場合、1つの利点は、各特徴値の各インスタンスが、予測平均からの観察された逸脱との同時出現を表す「インパクト」値を割り当てられ、次いで、特徴値の変化毎のインパクトの変動を調べることができることである。図27Dは、「historical-took\_medication:irinotecan」という特徴に従ってグループ化されたボックスプロットを示している。図27Eは、最終病期に従ってグループ化されたボックスプロットを示している。図27Dは、結腸直腸病期4のサブセットについて予想されるように、値1の「historical-took\_medication:irinotecan」と同時出現する特徴は、値0に関連付けられている特徴よりも大きなインパクトを有することを示している。図27Eは、後の病期に関連付けられているより大きなインパクトを示している。

30

## 【0215】

SAFEアルゴリズムは、特徴の相互作用を直接的に考慮に入れていないけれども、これらの値は、手動で構築された複合特徴から導出され得る。それに加えて、SAFEアルゴリズムは、各特徴が基礎となるモデルからの予測値にどのようなインパクトを与えるかを伝えることを狙っており、これは、標的を予測することに対する特徴重要度の間接的代理として使用されるが、これはモデルの有効性を前提とする。

40

## 【0216】

## ノートブック

様々な実施形態において、1つまたは複数の統計モデルおよび解析は、特定の目的に対応できるように組み合わせられ、また、最初の解析のバリエーションを通じて、多くの問題を解決するために使用され得る。統計モデルおよび分析のそのような組合せは、インタラクティブ解析ポータル22のノートブックとして記憶され得る。ノートブックは、インタラクティブ解析ポータル22の一特徴であり、統計モデルおよび解析を構築するための容易にアクセス可能なフレームワークを提供する。統計モデルおよび解析が開発されてしまえば、その後、異なるユーザと共有され、それにより、最初に発せられた質問以外の科学的およびビジネス的質問を解析し、それへの答えを見つけるものとしてよい。

50

## 【0217】

1)インタラクティブ解析ポータル22は、単純で直感的なポイントアンドクリック/ドラッグアンドドロップのインターフェースによる入力のカスタマイズを可能にし、それにより解析のためにコホートを絞り込むことができる。インタラクティブ解析ポータル22、外れ値、スマートコホート、またはインタラクティブ解析ポータル22の他のポータルのいずれかを通じて選択されたコホートは、処理のためにノートブックに送られ得る。

## 【0218】

2)インタラクティブ解析ポータル22、基礎となる認可されたデータベース、ならびに任意のサポートされている統計モデル、可視化、算術モデル、および他の提供される操作とインターフェースする関数呼び出しのライブラリを有するカスタムアプリケーションインターフェース(API)が、ノートブックまたはワークブックをインタラクティブ解析ポータル22のデータ、関数呼び出し、および他のリソースと統合するためにユーザに提供され得る。例示的な関数呼び出しは、許可されたデータソースをリストすること、データソースを選択すること、データソースをフィルタ処理すること、現在のフィルタ処理済みコホート内の患者の臨床事象をリストすること、RNAまたはDNAからの融合の識別、RNAまたはDNAからの遺伝子の識別、マッチする臨床試験を識別すること、DNAバリエーション、免疫組織化学(IHC)を識別すること、RNA発現を識別すること、コホートにおける療法を識別すること、コホート内の患者を治療するのに適用可能な潜在的療法を識別すること、および他のコホートもしくはデータセット処理を含み得る。

## 【0219】

3)インタラクティブ解析ポータル22は、ノートブック自体の選択されたモデル、解析、可視化、またはレポートが、インタラクティブ解析ポータル22からコホートを受け入れ、コードレベルでのユーザの介入なしで、そのままコホートに対する解析を提供するように構成されているのでノートブック内でユーザに何かをコーディングさせることなく、ノートブック生成が、結果の1つまたは複数の統計モデル、解析、および可視化または報告を絞り込まれたコホートに対して実行することを可能にする。いくつかのモデルは、選択され得るハイパーパラメータまたはチューニングパラメータを有し得るか、またはモデルそれ自体が、実行時にコホートおよび/または他のモデル、解析、可視化、またはレポートに基づき適用されるべき最適なパラメータを識別し得る。

## 【0220】

4)インタラクティブ解析ポータル22は、選択されたノートブックに基づき準備された結果をユーザに表示する。

## 【0221】

5)次いで、関連付けられているユーザは、ノートブックそれ自体がインタラクティブ解析ポータル22からコホートを受け入れ、ユーザの介入なしでノートブック結果を提供するように構成されているのでユーザにノートブック内で何かをコーディングさせるか、または再コーディングさせることなく、選択された解析を絞り込まれたコホートに適用する前に生成されているノートブックを選択し得る。

## 【0222】

6)ユーザは、ネットワーク上のクラウドコンピューティングまたはハードウェアリソースのコストを理解するために自分のノートブックによって使用されている計算リソースを追跡し、ノートブックを通じて提供する統計分析の有効性を判断するために自分のノートブックの人気度を追跡するものとしてよい。

## 【0223】

いくつかの実施形態において、ノートブックは、インタラクティブ解析ポータル22がカスタムテンプレートをユーザの選択済みデータに与え、事前構築ヘルスケア統計モデルを活用してプログラミングに精通していないユーザに結果を提供することを可能にすることによってユーザに利益をもたらす。社内チームは、患者ケアを改善することと、ライフサイエンス研究を改善することの両方に役立つ新しいヘルスケアに関する洞察をサポートするためにキュレーションされたデータを解析し得る。同様に、外部ユーザは、分析のため

10

20

30

40

50

にこの専用現実世界データに容易にアクセスでき、また専用統計モデルにアクセスできる。

【0224】

ユーザに対する課金モデルは、サブスクリプションベースまたはオンデマンドベースで提供されてよい。たとえば、ユーザは、1つまたは複数のデータセットを、月払いまたは年払いサブスクリプションなど、一定期間、サブスクライブするか、または、ユーザは、特定の cohorts に対応するノートブックをロードし、消費のために瞬時結果を生成する料金を支払うなど、データおよびノートブックの使用に対してアクセス毎に支払うものとしてよい。ユーザは、ストレージおよびコンピューティングリソースの使用を確認し、最適化するために使用され得るベンチマークおよび最適化ポータルを望むこともあり得る。

【0225】

ノートブックを生成することは、ノートブック編集用のGUIで実行されてもよい。ユーザは、ノートブックに対するレポートページを構成してよい。報告ページは、ユーザによって選択され書き込まれているテキスト、画像、およびグラフを含み得る。事前構成された要素は、ドロップダウンリストまたはドラッグアンドドロップメニューなど、リストから選択され得る。事前構成された要素は、統計解析モジュールおよび機械学習モデルを含む。たとえば、ユーザは、特定の特徴に関するデータに対して線形回帰を実行したい場合がある。ユーザが線形回帰を選択すると、チェックボックスを備えたメニューが、線形回帰モデルに供給されるべきデータセットの特徴とともに表示され得る。記入後、選択された特徴に関する線形回帰の結果を報告するためのテンプレートが、アクティブカーソルまたはドラッグアンドドロップ要素のドロップ位置によって識別される場所で報告ページに追加され得る。ユーザが機械学習モデルを使用して問題を解決したい場合、モデルがシートに追加されるとよい。モデル、ハイパーチューニングパラメータ、および報告された結果を識別するヘッダが書き込まれ得る。次いで、いくつかの事例では、以前に訓練されたモデルは、現在の cohorts に適用され得る。他の事例では、モデルは、たとえば、モデルが訓練されるべきアノテーション付き特徴と関連付けられている転帰を選択することによってオンザフライで訓練され得る。教師なし機械学習モデルでは、特徴が訓練中に識別されるのでモデルはアノテーション付き特徴の選択を必要とし得ない。いくつかの実施形態において、選択された統計モデルが、テンプレート内で計算されていない訓練済みモデルからの結果を必要とする場合、テンプレートは、選択された統計モデルをノートブックに挿入する前に、必要な結果を生成するために訓練済みモデルを自動的に追加し得る。

【0226】

統計解析モデルは、選択された特徴に関する cohorts の算術平均、選択された特徴に対する cohorts の標準偏差/分布、選択された特徴に対する変数間の回帰関係、cohorts を解析のための最適な部分集団にサブセットするためのサンプルサイズ決定モデル、または cohorts における統計的に有意な特徴および相関を識別するための t 検定モジュールを計算するために事前設計され得る。他の事前計算された統計解析モジュールは、cohorts 内の有意な相関および/または特徴を識別するための cohorts 解析、意味のあるパターンを識別するためのデータマイニング、または統計モデルをデータにマッチさせ、どのモデルが適用可能であるかを報告し、それらのモデルをノートブックに追加するためのデータドレッシングを実行し得る。

【0227】

機械学習モデルは、線形回帰アルゴリズム、非線形回帰、ロジスティック回帰アルゴリズム、分類モデル、ブートストラップリサンプリングモデル、サブセット選択モデル、次元削減モデル、木ベースモデル(バギング、ブースティング、ランダムフォレストなど)、および他の教師ありまたは教師なしモデルを適用し得る。各モデルが選択されると、どの特徴をそのモデルが識別し、分類し、および/または報告すべきかを指定するユーザから標的出力が要求され得る。たとえば、ユーザは、どの特徴が cohorts 内の患者生存に最も密接に相関しているか、またはどの特徴が cohorts 内の陽性治療転帰に最も密接に相関しているかを識別するモデルを選択してよい。また、ユーザは、モデルの分類ラベルのうちどの分類ラベルをモデルに分類させたいかを選択し得る。モデルが5つのラベルに従ってコ

10

20

30

40

50

ホートを分類し得る例では、ユーザは、発生源不明の腫瘍を有する患者が乳房、肺、または脳に由来するのかなど、1つまたは複数のラベルを二項分類(患者がラベルを有する、患者がラベルを有しない)として指定し得る。ユーザは、乳房のみを選択して、腫瘍が乳房に由来するものとして分類され得るか、または乳房に由来しないものとして分類され得るかを、発生源不明の任意の腫瘍について識別し得る。

#### 【0228】

図29は、一実施形態による1つまたは複数のノートブックを介してアナリティクスを生成するためのインタラクティブ解析ポータル22のユーザインターフェースを例示している。

#### 【0229】

ノートブックユーザインターフェース2900は、インタラクティブコホート選択フィルタ処理24を介して患者のデータベースを患者の所望のコホートにフィルタ処理する前または後のいずれかにサイドバーメニュー2910などを介してインタラクティブ解析ポータル22からノートブックを選択することによってアクセスされ得る。

#### 【0230】

ノートブック、またはワークブックは、データサイエンス、機械学習、または患者データに対してアナリティクスを日常的に実行する他の分野に精通したチームメンバーによって会社ラベルで内部的にキュレーションされ、カスタムワークブックウィジェット2920を介してユーザに提示され得る。カスタムワークブックウィジェットは、検索可能なリスト、検索可能なアイコン、追加のワークブックを表示するために水平または垂直にスクロールし得るスクロールウィンドウ、またはユーザがアクセスを許可されているすべてのワークブックへのアクセスを提供するために拡張する拡張可能なウィンドウとして提示され得る。ワークブックは、ワークブック2960に例示されているような、アイコンおよび関連付けられているテキストによって表され得る。また、ユーザは、マイワークブックウィジェット2930を介してアクセスされ得る個人化されたワークブックを生成し得る。ウィジェット2920または2930から選択されたワークブックを表示するために、ワークブック表示ウィンドウ2950が提供され得る。空白のワークブック2940を選択することによって、新しいワークブックがユーザによって作成され得る。空白のワークブック2940を選択した後、ワークブック生成インターフェースが開くものとしてよい。

#### 【0231】

図30は、一実施形態による新しいワークブックを作成するためのインタラクティブ解析ポータル22のワークブック生成インターフェースを例示している。

#### 【0232】

ワークブック生成インターフェース3000は、ノートブックユーザインターフェースから空白のワークブックを選択した後、ユーザに提供され得る。テキスト入力ユーザインターフェース要素(UIE)3010は、生成後の識別、検索、および索引付けのためにワークブックに名前を付けるために提供され得る。一連のボタンおよびドロップダウンメニューUIE3020は、ユーザインターフェースのグループ化された要素を区分けするために提供され得る。UIE3020は、ワークブックのプレゼンテーションを構築し構造化するユーザを支援し得る。セルUIEは、現在選択されているセルを実行するためのコマンド、現在選択されているセルを終了するためのコマンド、セルを追加するためのコマンド、セルを削除するためのコマンド、すべてのセルを実行するためのコマンド、上にあるすべてのセルを実行するためのコマンド、下にあるすべてのセルを実行するためのコマンド、またはすべてのセルを終了するためのコマンドなどの、コードのブロックを有するウィンドウ3040の、現在選択されているセルに関連する選択を提供し得る。カーネルUIEは、Python、構造化照会言語(SQL)、R、Spark、Haskell、Ruby、Typescript、Javascript、Perl、Lua、C、C++、Matlab、Java、Emu86、他のカーネルなど、1つまたは複数のプログラミング言語および/またはユーザが利用可能な言語に関連する選択を提供し得る。カーネルUIEからカーネルを選択することで、セルがそれぞれの言語からコマンドを実行するようにワークブックをリロードする。ウィジェットUIEは、アクティブなカーネルに対してサポート

10

20

30

40

50

されている1つまたは複数のコードスニペットに関連する選択を提供し得る。コードスニペットは、グラフもしくはプロットなどの可視化物を作成するためのコード、平均もしくは標準偏差を計算することなどの単純な算術演算のためのコード、または分布を計算することおよびそれぞれの曲線を表示することなどのより複雑な演算のためのコードを含み得る。一連のアイコンUIE3030が提供されてもよく、各アイコンは、UIE3020から実行される人気のあるコマンドを表す。例示的な人気のあるコマンドは、ドキュメントを保存すること、新しいセルを追加すること、コードまたはセルを切り取るかまたは貼り付けること、任意の他のセルに関してページの上方もしくは下方に移動させることによってセルを再配置すること、またはアクティブなセル内のコードを実行/終了することを含み得る。

#### 【0233】

1つまたは複数のセルがウィンドウ3040内に存在し、これにユーザがアクティブなカーネルに対する1つまたは複数の行のコードを挿入するものとしてよい。ユーザは、アクティブなデータベースまたは患者のコホート上で動作し得るセルにコードまたはコマンドを入力してよい。セルを実行すると、入力されたコードまたはコマンドが実行される。標準出力、エラーメッセージ、印刷文などの出力が、実行後にセルの直下に表示される。それに加えて、1つまたは複数のセルからのコードに基づきフォーマットおよび関連するテキストを提供するテキストウィジェットが挿入され得る。そのようなテキストウィジェットは、コードを実行した結果を単純で読みやすいフォーマットを提供し得る。一実施形態において、テキストウィジェットは、HTML、インデントされたりリスト、テキストフォーマット、TeX/LaTeXの数式、およびインラインテーブルをサポートするマークダウンセルとして提示され得る。

#### 【0234】

一例において、コードブロックは、値の行列に対して算術演算を実行し得る。行列を印刷するなどの関連する出力の結果は、一連のブラケット、挿入句、およびカンマを理解するのが困難である。可視化ウィジェットは、行列を含む変数を受け取り、潜在的にわかりにくいテキスト出力の代わりに行列を表す表示可能テーブル形式で行列の値が見える画像を提供する。セルは、各サポートされているカーネルおよびプログラミング言語に関連付けられているすべてのコマンドを受け付ける。セルは、他のソースからモジュールまたはライブラリ(dask、fastparquet、pandas、または他のライブラリなど)をインポートし、データ構造をサポートし、条件文および論理ループをサポートし、さらには関数を確立し呼び出すものとしてよい。アクティブなコードからの瞬間的出力をユーザに見えるように、セル出力はコードの実行と非同期に生成される。出力が表示行数の事前構成されている限界を超えた場合、出力はスクロール可能なテキストになり、新しいエントリとともに自動スクロールするか、またはユーザ入力後にスクロールするものとしてよい。

#### 【0235】

1つまたは複数のテンプレートが、ユーザの便宜のためにテンプレートウィンドウ3050内に用意され得る。テンプレートは、フィルタ処理された患者コホートなどの入力データに働き、コードの1つまたは複数のセルを実行して論理的な結果を生成し、テキストまたは可視化の1つまたは複数のセルを実行して入力データ上の実行された論理の結果を使いやすい様式で報告するように事前構成されている1つまたは複数のセルを含み得る。チャート、グラフ、回帰、次元削減、分類、RNAまたはDNAの正規化、およびユーザから利用可能なテンプレート上で他の一般的に使用される特徴のためのテンプレートが存在し得る。テンプレートは、データセットとともに提供され得るか、または他のユーザと共有されるようにユーザによってカスタム作成され得る。

#### 【0236】

図31は、ノートブックユーザインターフェースのカスタムワークブックウィジェットからの事前構成されているテンプレートを開く動作を例示している。

#### 【0237】

ノートブックユーザインターフェース2900に戻ると、ユーザは、所望のワークブックをウィジェットから表示ウィンドウをクリックしてドラッグすることによって、カスタム

10

20

30

40

50

ワークブックウィジェット2920からのカスタムワークブックをワークブック表示ウィンドウ2950に入れてよい。一例において、ユーザは、3120のところで例示されているように、マウスカースルでワークブック2960を選択し、ワークブックを表示ウィンドウ2950にドラッグしてよい。他の直感的なマウス、キーボード、またはジェスチャーコマンドが、クリックおよびドラッグの代わりに、またはそれに加えて実装され得る。

【0238】

図32は、ユーザがワークブックを表示ウィンドウ内にドラッグしたときのノートブックユーザインターフェースからの応答を例示している。

【0239】

ノートブックエディタ3200は、ユーザが選択したワークブックに基づき、タイトル3210および1つまたは複数のセル3240A~Dを自動入力し得る。ユーザは、テキスト入力UIE3220をさらに使用しワークブックの編集を使用してワークブックの名前を変更し得る。ユーザは、一連のボタンを介してワークブックの構成を変更してもよく、ドロップダウンメニューUIE3220はユーザインターフェースのグループ化された要素を区分けするために提供され得る。UIE3220は、ワークブックのプレゼンテーションを構築し構造化するユーザを支援し得る。セルUIEは、現在選択されているセルを実行するためのコマンド、現在選択されているセルを終了するためのコマンド、セルを追加するためのコマンド、セルを削除するためのコマンド、すべてのセルを実行するためのコマンド、上にあるすべてのセルを実行するためのコマンド、下にあるすべてのセルを実行するためのコマンド、またはすべてのセルを終了するためのコマンドなどの、コードのブロックを有する現在選択されているセル3240A~Dに関連する選択を提供し得る。カーネルUIEは、Python、構造化照会言語(SQL)、R、Spark、Haskell、Ruby、Typescript、Javascript、Perl、Lua、C、C++、Matlab、Java、Emu86、他のカーネルなど、1つまたは複数のプログラミング言語および/またはユーザが利用可能な言語に関連する選択を提供し得る。カーネルUIEからカーネルを選択することで、セルがそれぞれの言語からコマンドを実行するようにワークブックをリロードする。ウィジェットUIEは、アクティブなカーネルに対してサポートされている1つまたは複数のコードスニペットに関連する選択を提供し得る。コードスニペットは、グラフもしくはプロットなどの可視化物を作成するためのコード、平均もしくは標準偏差を計算することなどの単純な算術演算のためのコード、または分布を計算することおよびそれぞれの曲線を表示することなどのより複雑な演算のためのコードを含み得る。ユーザは、提供され得る一連のアイコンUIE3230を介してワークブックの構成をさらに変更してもよく、各アイコンは、UIE3220から実行される人気のあるコマンドを表す。例示的な人気のあるコマンドは、ドキュメントを保存すること、新しいセルを追加すること、コードまたはセルを切り取るかまたは貼り付けること、任意の他のセルに関してページの上方もしくは下方に移動させることによってセルを再配置すること、またはアクティブなセル内のコードを実行/終了することを含み得る。

【0240】

ユーザは、セルを選択し、編集用のセルUIEオプションを選択するか、または関連付けられているキーボードショートカットを押すことによって、セル3240A~Dの各々のソースコードを編集してもよい。

【0241】

図33は、ユーザがワークブックをワークブックエディタ3300にロードし、セルUIEから編集を選択した後の、カスタムワークブックのセル編集ビューを例示している。

【0242】

セル3240A~Dを有するワークブックのセル編集ビューに入った後、セル3310Aおよび3310Bが見えるようになる(3310C~Dは図示せず)。セル3310Aは、患者の対照コホートと治療コホートとの間の傾向差に基づき生存曲線3240Aを生成するコードを表示する。セル3310Bは、患者のフィルタ処理されたコホートにおける2つの選択されたRNAトランスクリプトームについての正規化されたRNA発現に基づき散布図3240B(図示せず)を生成するコードを表示する。類似のセル3310C~D(図示せず)は、散布図およびボックスプロ

10

20

30

40

50

ット3240C~D(図示せず)に対してそれぞれ生成され得る。

【0243】

ユーザは、コードを編集して、自分の目的に合わせてワークブックを修正し、さらにはセルを追加するか、追加のセルを削除して新しいカスタマイズされたワークブックを作成し得る。

【0244】

セル編集ビューにおいて、ユーザは、1つまたは複数のテンプレートがユーザの便宜のためにテンプレートウィンドウ3050内に用意され得ることを確認し得る。テンプレートは、フィルタ処理された患者コホートなどの入力データに働き、コードの1つまたは複数のセルを実行して論理的な結果を生成し、テキストまたは可視化の1つまたは複数のセルを実行して入力データ上の実行された論理の結果を使いやすい様式で報告するように事前構成されている1つまたは複数のセルを含み得る。チャート、グラフ、回帰、次元削減、分類、RNAまたはDNAの正規化、およびユーザから利用可能なテンプレート上で他の一般的に使用される特徴のためのテンプレートが存在し得る。テンプレートは、データセットとともに提供され得るか、または他のユーザと共有されるようにユーザによってカスタム作成され得る。

10

【0245】

ユーザは、任意のテンプレートをセル内にドラッグして、テンプレートの関連付けられている可視化または算術演算を生成するためのコードをセルに入れるものとしてよい。

【0246】

ユーザは、各患者データベースへのサブスクリプションを有する機関または医療施設との関連によってユーザに提供されている患者のデータベースのユーザインターフェースにアクセスし得る。カスタムワークブックは、また、ワークブックが各データベース内の患者への適用性について選択されるデータベース毎に提供され得る。ユーザインターフェースにアクセスすることで、許可されたデータベースおよび/またはワークブックにアクセスできるクラウドコンピューティング環境内にリソースを生成し得る。クラウドコンピューティング環境内のユーザのリソース使用状況は、ユーザによって消費されるリソースに対する正確な課金を補完するために監視され、追跡され得る。ユーザは、患者の他のデータベースを要求し、購入し得る。患者のデータベースは、それらの中の患者の特性に基づき購入され得る。たとえば、ユーザは、乳癌と診断された患者のデータベースを望んでいる場合がある。ルックアップテーブル(LUT)または癌オンコロジーが参照され、それにより、乳管癌、乳癌、乳房癌、乳房上皮性悪性腫瘍、または他の関連する用語など、乳癌の代替マッチングが得られる。要求された診断、およびLUTまたは癌オンコロジーからの代替用語のうちのいずれかの条件を満たす患者は、データベースにまとめられ、ユーザに配信され得る。次いで、ユーザは、本明細書の開示に従ってデータに対する統計分析および研究を実施し得る。

20

30

【0247】

他のウェブインターフェースは、上記の外れ値ポータル、スマートコホートポータル、およびノートブックポータルに類似するインタラクティブ解析ポータル22に組み込まれ得る。そのような他のウェブインターフェースの1つは、傾向スコアリングを使用して患者の病状に対する療法、手技、臨床試験、または他の医療事象の効果を識別することを含み得る。傾向スコアリングおよび関連するウェブインターフェースは、全体が参照により本明細書に組み込まれている2019年11月8日に出願された米国特許出願第16/679,054号、名称「Evaluating Effect of Event on Condition Using Propensity Scoring」においてさらに詳細に説明されている。

40

【0248】

図34は、本明細書で説明されている方法のうちの1つまたは複数を実行させるために、命令セットが実行され得るコンピュータシステム3400の例示的なマシンを例示している。代替の実装形態において、マシンは、LAN、イントラネット、エクストラネット、および/またはインターネットにおいて他のマシンに(たとえば、ネットワーク)接続さ

50

れ得る。

【0249】

マシンは、クライアントサーバネットワーク環境内のサーバまたはクライアントマシンの能力内で、またはピアツーピア(または分散)ネットワーク環境内のピアマシンとして、またはクラウドコンピューティングインフラストラクチャまたは環境内のサーバもしくはクライアントマシンとして動作し得る。マシンは、パーソナルコンピュータ(PC)、タブレットPC、セットトップボックス(STB)、携帯情報端末(PDA)、携帯電話、ウェブアプリケーション、サーバ、ネットワークルーター、スイッチもしくはブリッジ、またはそのマシンによって実行されるアクションを指定する命令セット(順次または他の方法で)実行することができる任意のマシンであってよい。さらに、単一のマシンが図示されているが、「マシン」という用語は、本明細書で説明されている方法のうちのいずれか1つまたは複数を実行するために命令セット(または複数の命令セット)を個別に、または連携して実行するマシンの任意の集合体を含むと解釈されるものとする。

10

【0250】

例示的なコンピュータシステム3400は、処理デバイス3402、メインメモリ3404(リードオンリーメモリ(ROM)、フラッシュメモリ、シンクロナスDRAM(SDRAM)またはDRAMなどのダイナミックランダムアクセスメモリ(DRAM)、など)、スタティックメモリ3406(フラッシュメモリ、スタティックランダムアクセスメモリ(SRAM)、など)、およびデータストレージデバイス3418を備え、これらはバス3430を介して互いに通信する。

【0251】

処理デバイス3402は、マイクロプロセッサ、中央演算処理装置、または同様のものなどの1つまたは複数の汎用処理デバイスを表す。より具体的には、処理デバイスは、複合命令セットコンピューティング(CISC)マイクロプロセッサ、縮小命令セットコンピューティング(RISC)マイクロプロセッサ、超長命令語(VLIW)マイクロプロセッサ、または他の命令セットを実装するプロセッサもしくは命令セットの組合せを実装するプロセッサであってよい。処理デバイス3402は、特定用途向け集積回路(ASIC)、フィールドプログラマブルゲートアレイ(FPGA)、デジタルシグナルプロセッサ(DSP)、ネットワークプロセッサ、または同様のものなどの1つまたは複数の専用処理デバイスであってもよい。処理デバイス3402は、本明細書で説明されているオペレーションおよびステップを実行する命令3422を実行するように構成される。

20

30

【0252】

コンピュータシステム3400は、LAN、イントラネット、インターネット、および/またはエクストラネットに接続するためのネットワークインターフェースデバイス3408をさらに含んでもよい。コンピュータシステム3400は、ビデオディスプレイユニット3410(液晶ディスプレイ(LCD)または陰極線管(CRT)など)、英数字入力デバイス3412(キーボードなど)、カーソル制御デバイス3414(マウスなど)、信号発生デバイス3416(スピーカーなど)、およびグラフィックプロセッシングユニット3424(グラフィックスカードなど)も備え得る。

【0253】

データストレージデバイス3418は、本明細書で説明されている方法または機能のうちのいずれか1つまたは複数を使用する1つまたは複数の命令セットまたはソフトウェア3422が記憶されるマシン可読記憶媒体(コンピュータ可読媒体としても既知である)であってよい。命令3422は、コンピュータシステム3400による実行中に完全にまたは少なくとも部分的にメインメモリ3404および/または処理デバイス3402内にも常駐するものとしてよく、メインメモリ3404および処理デバイス3402もマシン可読媒体を構成する。

40

【0254】

一実施形態において、命令3422は、インタラクティブ解析ポータル(図1のインタラクティブ解析ポータル22など)に対する命令および/またはインタラクティブ解析ポータルとして機能する方法を含むソフトウェアライブラリを含む。命令3422は、患者フィルタ処理モジュール3426(図1のインタラクティブコホート選択フィルタ処理インターフェース2

50

4など)および患者アナリティクスモジュール3428(図1のコホートファネルおよび母集団解析インターフェース26、患者タイムライン解析ユーザインターフェース28、患者生存率解析ユーザインターフェース30、および/または患者事象尤度解析ユーザインターフェース32など)に対する命令をさらに含み得る。データストレージデバイス3148/マシン可読記憶媒体は、単一の媒体であるものとして例示的な実装形態において示されているが、「マシン可読記憶媒体」という用語は、1つまたは複数の命令セットを記憶する単一の媒体または複数の媒体(集中もしくは分散データベース、および/または関連するキャッシュおよびサーバなど)を含むものとして解釈されるべきである。「マシン可読記憶媒体」という用語は、マシンによって実行される命令セットを記憶するか、または符号化することができ、本開示の方法のうちのいずれか1つまたは複数を実行させる任意の媒体を含むものとしても解釈されるべきである。「マシン可読記憶媒体」という用語は、限定はしないが、ソリッドステートメモリ、光媒体、および磁気媒体を含むものとしてしかるべく解釈されるものとする。「マシン可読記憶媒体」という用語は、マシン可読記憶媒体を一時的記憶媒体または一時的マシン可読記憶媒体として識別することによって別段の指定がなされていない限り信号などの一時的記憶媒体をしかるべく除外するものとする。

10

**【0255】**

別の実装形態において、仮想マシン3440は、患者フィルタ処理モジュール3426(図1のインタラクティブコホート選択フィルタ処理インターフェース24など)および患者アナリティクスモジュール3428(図1のコホートファネルおよび母集団解析インターフェース26、患者タイムライン解析ユーザインターフェース28、患者生存率解析ユーザインターフェース30、および/または患者事象尤度解析ユーザインターフェース32など)に対する命令を実行するためのモジュールを含み得る。計算では、仮想マシン(VM)がコンピュータシステムのエミュレーションである。仮想マシンは、コンピュータアーキテクチャに基づいており、物理的コンピュータの機能を提供する。それらの実装形態は、専用ハードウェア、ソフトウェア、またはハードウェアとソフトウェアとの組合せを伴い得る。

20

**【0256】**

先行する詳細な説明のいくつかの部分は、コンピュータメモリ内のデータビットに対するオペレーションのアルゴリズムおよび記号表現に関して提示されている。これらのアルゴリズムに関する説明および表現は、作業の内容を当業者に最も効果的に伝えるためにデータ処理の技術者によって使用される手段である。アルゴリズムは、本明細書では、また一般的に、望む結果をもたらす演算の自己矛盾のないシーケンスであることが企図される。これらの演算は、物理的量の物理的操作を必要とする演算である。通常、必ずというわけではないが、これらの量は、記憶され、転送され、組み合わせられ、比較され、および他の何らかの形で操作可能である電気的および磁気的信号の形態をとる。主に共通使用の理由から、これらの信号をビット、値、要素、記号、文字、語、数、または同様のものとして参照することがときには都合がよいことが実証されている。

30

**【0257】**

しかしながら、これらおよび類似の語のすべては、適切な物理的量と関連付けられるべきであり、これらの量に付けられた都合のよいラベルにすぎないことを覚えておくべきである。上記の説明から明らかなように特に断りのない限り、説明全体を通して、「識別すること」または「提供すること」または「計算すること」または「決定すること」または同様の語句などの語を利用する記述は、コンピュータシステムのレジスタおよびメモリ内で物理的(電子的)量として表されているデータを操作し、コンピュータシステムのメモリまたはレジスタまたは他のそのような情報記憶デバイス内で物理的量として同様に表現される他のデータに変換するコンピュータシステム、または類似の電子コンピューティングデバイスのアクションおよびプロセスを指すことは理解される。

40

**【0258】**

本開示は、本明細書のオペレーションを実行するための装置にも関係する。この装置は、意図された目的のために特別に制作され得るか、またはコンピュータに記憶されているコンピュータプログラムによって選択的にアクティブ化されるかもしくは再構成される汎

50

用コンピュータを備え得る。このようなコンピュータプログラムは、限定はしないが、フロッピーディスク、光ディスク、CD-ROM、および光磁気ディスクを含む任意の種類のディスク、リードオンリーメモリ(ROM)、ランダムアクセスメモリ(RAM)、EPROM、EEPROM、磁気もしくは光カード、または電子的命令を記憶するのに適した任意の種類の媒体などのコンピュータ可読記憶媒体に記憶されるものとしてよく、これらは各々コンピュータバスに結合されている。

【0259】

本明細書で提示されているアルゴリズムおよび表示は、特定のコンピュータまたは他の装置に本質的に関係しているわけではない。様々な汎用システムが、本明細書の教示に従ってプログラムとともに使用され得るか、方法を実行するためにより専門化された装置を製作することが都合よいことが実証され得る。様々なこれらのシステムに対する構造は、以下の説明で述べられている通りに見える。それに加えて、本開示は、任意の特定のプログラミング言語を参照して説明されてはいない。本明細書で説明されているように本開示の教示を実装するために様々なプログラミング言語が使用され得ることは理解されるであろう。

10

【0260】

本開示は、本開示に従ってプロセスを実行するようにコンピュータシステム(または他の電子デバイス)をプログラムするために使用され得る、命令が記憶されている機械可読媒体を含むものとしてよい、コンピュータプログラム製品またはソフトウェアとして提供され得る。機械可読記憶媒体は、機械(コンピュータなど)によって読み込むことが可能な形式で情報を記憶するためのメカニズムを備える。たとえば、機械可読(コンピュータ可読など)媒体は、リードオンリーメモリ(「ROM」)、ランダムアクセスメモリ(「RAM」)、磁気ディスク記憶媒体、光記憶媒体、フラッシュメモリデバイスなどを含む。

20

【0261】

前記指定において、本開示の実装形態は、特定の例示的な実装形態を参照しつつ説明されている。次に示す請求項に規定されているような本開示の実装形態のより広い精神および範囲から逸脱することなくそれに対して様々な修正が行われ得ることは明白であろう。したがって、本明細書および図面は、制限的な意味ではなく例示的な意味で考慮されるべきである。

【0262】

当業者にとっては、本発明の範囲から逸脱することなく上で説明されている発明の特定の実施形態において多数の変更および修正を行うことができることは明白であろう。したがって、前記の説明の全体は制限的な意味ではなく例示的な意味で解釈されるべきである。

30

【符号の説明】

【0263】

- 10 システム
- 12 バックエンド層
- 14 患者データストア
- 16 患者コホートセレクトアモジュール
- 18 患者コホートタイムラインデータストレージ
- 20 フロントエンド層
- 22 インタラクティブ解析ポータル
- 24 インタラクティブコホート選択フィルタ処理インターフェース
- 26 コホートファネルおよび母集団解析インターフェース、コホートファネルおよび母集団解析ユーザインターフェース
- 28 患者タイムライン解析ユーザインターフェース
- 30 患者生存率解析ユーザインターフェース
- 32 患者事象尤度解析ユーザインターフェース
- 34 患者次解析ユーザインターフェース
- 36 患者未来解析ユーザインターフェース

40

50

38	分散コンピューティングおよびモデリング層	
40	事象までの時間モデリングモジュール	
42	事象尤度モジュール	
204	プロジェクト名(患者のリストを記憶するデータベースを反映し得る)	
206	性別	
208	人種	
210	癌部位	
212、214	癌名	
216	腫瘍部位	
218	病期	10
220	M病期	
222	名前	
224	成分	
226	シーケンシング	
228	MSI(マイクロサテライト不安定性)ステータス	
230	手順	
232	イベント名	
234	死因	
236	「Ask Gene」タブ	
238	ダイアログボックス	20
242	患者人口統計、フィルタ処理可能なオプション	
244	癌部位	
246	腫瘍キャラクターゼーション	
248	分子データ	
250	組織	
252	病期	
254	異型度ベースのオプション	
256	バリエーションコール	
258	抽象化されたバリエーション	
260	MSI	30
262	TMB	
264	インタラクティブファネルチャート	
266	要約	
268	「Analyze Cohort」オプション	
300	領域	
300a	第1の領域	
300b	第2の領域	
300c	第3の領域、データ要約ウィンドウ	
300d	第4の領域	
302	平均患者年齢、終了事象	40
304	患者年齢のプロット、時間	
306	性別情報、開始事象	
308	薬剤、薬剤情報、終了事象	
310	ゲノムバリエーションまたは変化、性別	
312	制御パネル、遺伝子	
314	性別、組織	
316	組織、処方計画	
318	閉経状態、喫煙状態	
320	反応、病期	
322	喫煙状態、外科的手技	50

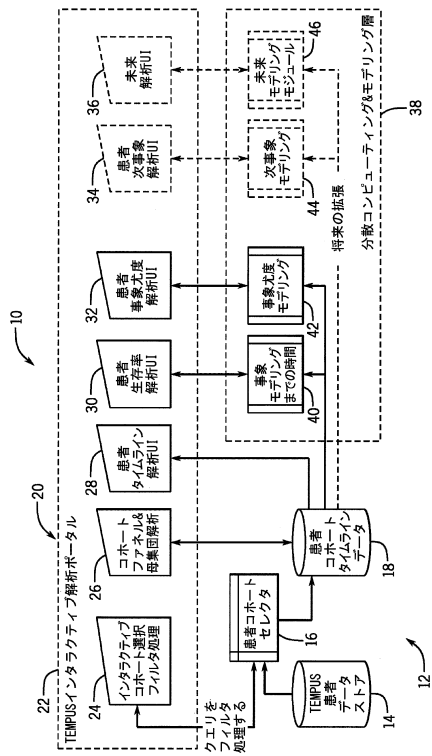
324	病期、特定の薬剤処方計画	
326	外科的手技	
328	応答データ	
330	凡例、生存率プロット	
332	ドロップダウンメニュー、生存率プロット	
334	凡例	
336、338	信頼区間	
400	第1の領域、第1のパネル	
402	インジケータ	
406	レーダープロット	10
408	中央インジケータ、中心点またはインジケータ	
410	第2の領域	
412	制御パネル	
414	オーバーレイ、外れ値閾値	
416、418	値	
420	プロット	
422	x軸	
434	y軸	
436	最大の患者グループ、インジケータ	
438	最大の外れ値グループ	20
440	第3の領域	
500	二分木	
2900	ノートブックユーザインターフェース	
2910	サイドバーメニュー	
2920	カスタムワークブックウィジェット	
2930	マイワークブックウィジェット	
2940	空白のワークブック	
2950	ワークブック表示ウィンドウ	
2960	ワークブック	
3000	ワークブック生成インターフェース	30
3010	テキスト入力ユーザインターフェース要素(UIE)	
3020	ボタンおよびドロップダウンメニューUIE	
3030	アイコンUIE	
3040	ウィンドウ	
3050	テンプレートウィンドウ	
3200	ノートブックエディタ	
3210	タイトル	
3220	テキスト入力UIE、ドロップダウンメニューUIE	
3240AD	セル、生存曲線	
3240B	セル、散布図	40
3240C	セル、散布図およびボックスプロット	
3240D	セル、散布図およびボックスプロット	
3300	ワークブックエディタ	
3310A、3310B、3310C、3310D	セル	
3400	コンピュータシステム	
3402	処理デバイス	
3404	メインメモリ	
3406	スタティックメモリ	
3408	ネットワークインターフェースデバイス	
3410	ビデオディスプレイユニット	50

- 3412 英数字入力デバイス
- 3414 カーソル制御デバイス
- 3416 信号発生デバイス
- 3418 データストレージデバイス
- 3422 命令、1つまたは複数の命令セットまたはソフトウェア
- 3424 グラフィックプロセッシングユニット
- 3426 患者フィルタ処理モジュール
- 3428 患者アナリティクスモジュール
- 3430 バス
- 3440 仮想マシン

10

【図面】

【図 1】



【図 2】

TEMPUS		コホートファネル	生存率曲線	生存木	患者タイムライン	診査データビジュアライザ	Ask	Gene
メタ	200							236
プロジェクト名								
SELECT	204							
性別								
SELECT	206							
人種								
SELECT	208							
癌								
癌部位名								
SELECT	210							
癌名								
SELECT	212							
転移								
癌名								
SELECT	214							
腫瘍発見								
腫瘍発見	202							
SELECT	216							
病期								
SELECT	218							
M病期								
SELECT	220							
薬剤								
薬剤名								
SELECT	222							
成分名								
SELECT	224							
シーケンシング								
遺伝子名+バリエーション								
SELECT	226							
免疫MSIステータス								
SELECT	228							
手順								
手順名								
SELECT	230							
死亡								
事象								
SELECT	232							
死因								
SELECT	234							

20

30

40

50

FIG. 3

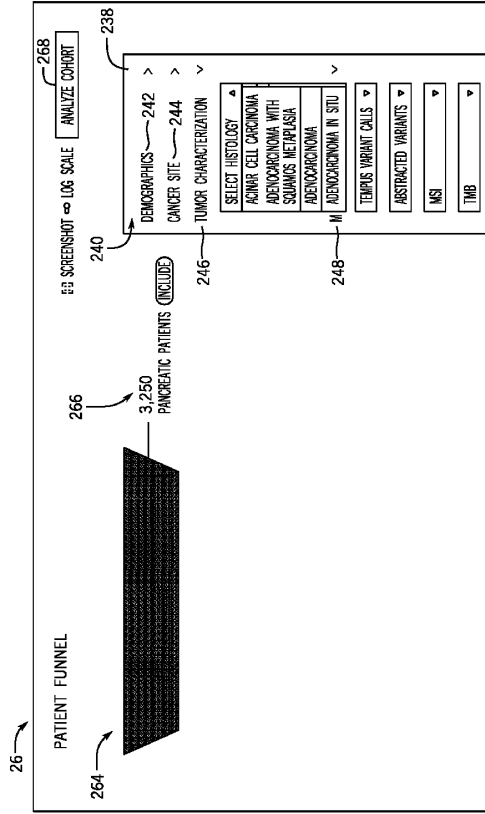


FIG. 3

FIG. 4

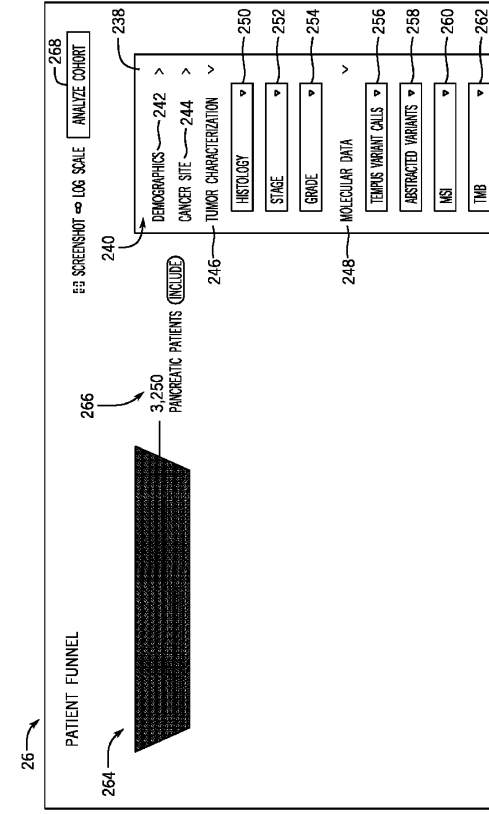


FIG. 4

FIG. 5

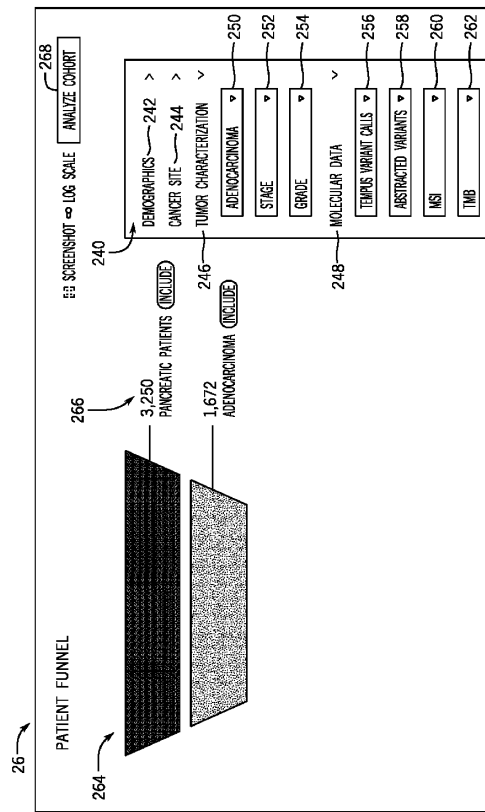


FIG. 5

FIG. 6

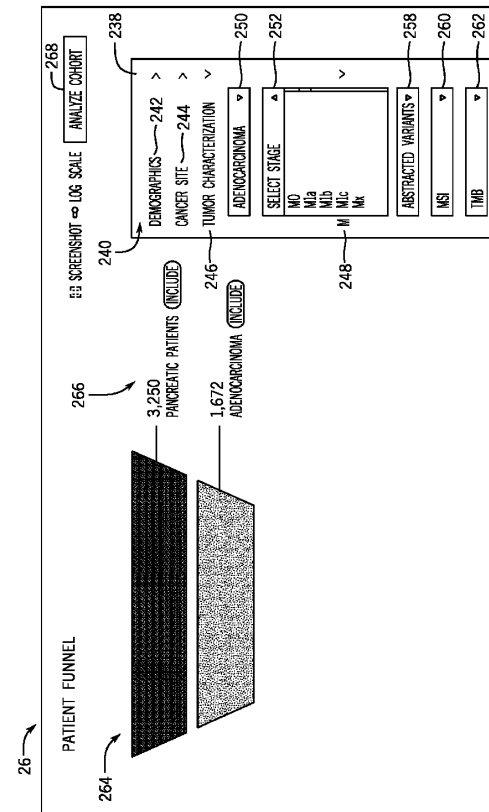


FIG. 6

10

20

30

40

50

[ 7 ]

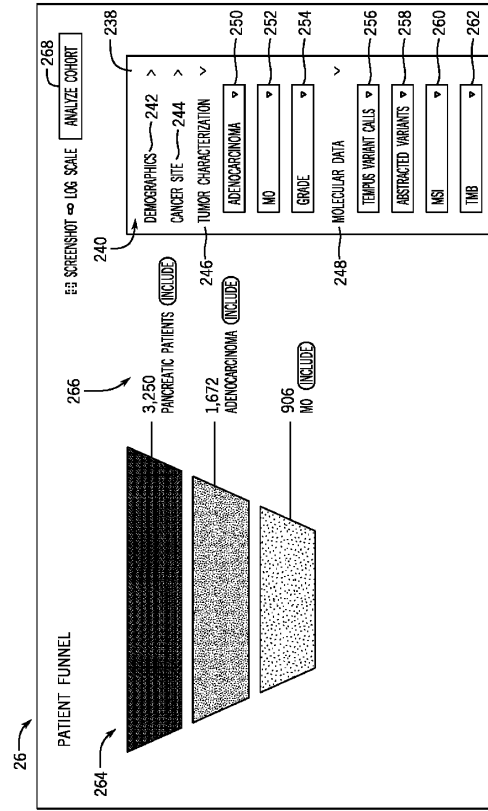


FIG. 7

[ 8 ]

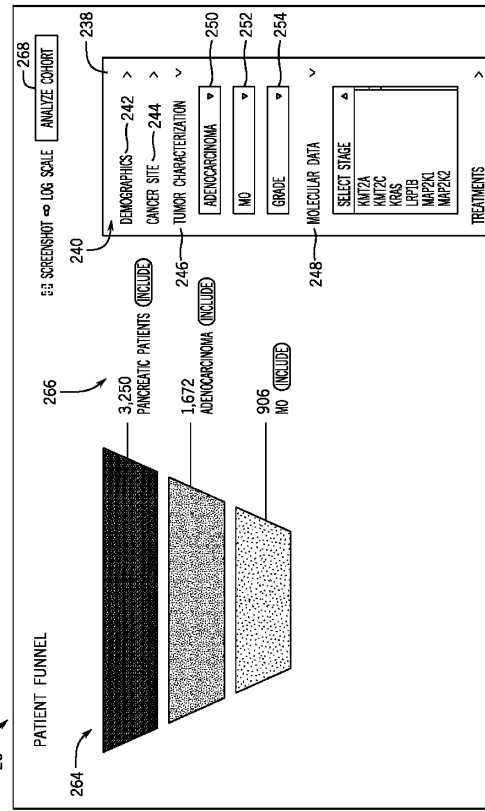


FIG. 8

[ 9 ]

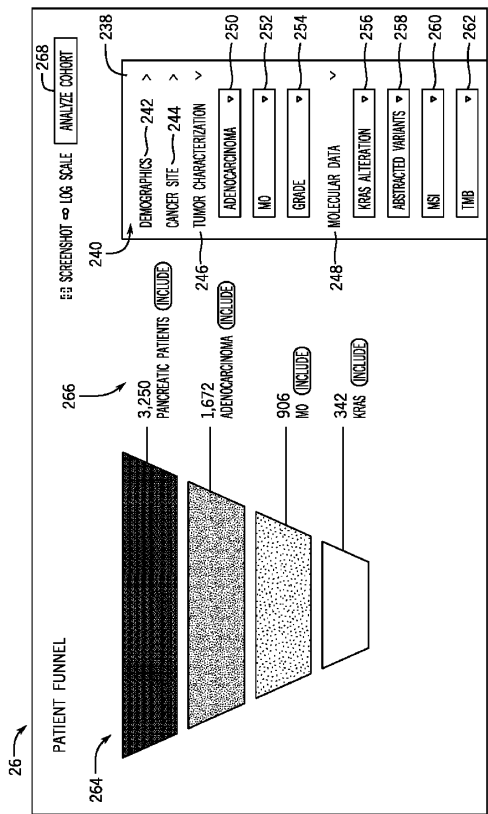


FIG. 9

[ 10 ]

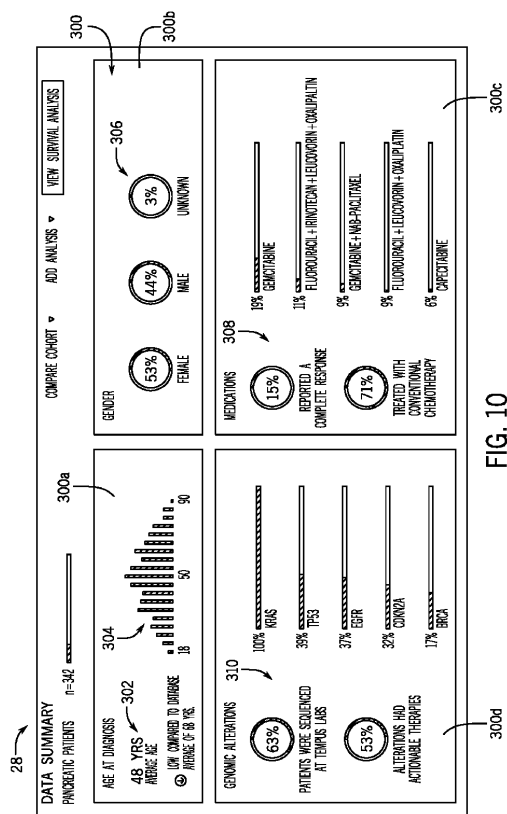


FIG. 10

10

20

30

40

50

1 1

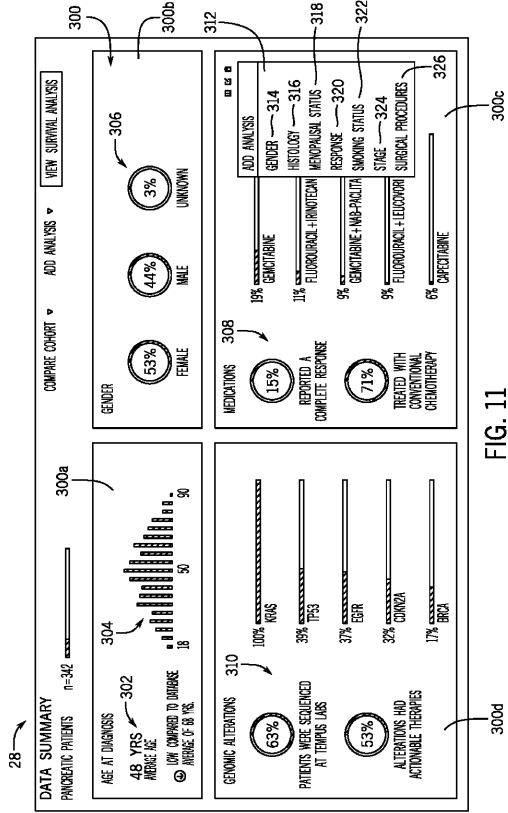


FIG. 11

1 2

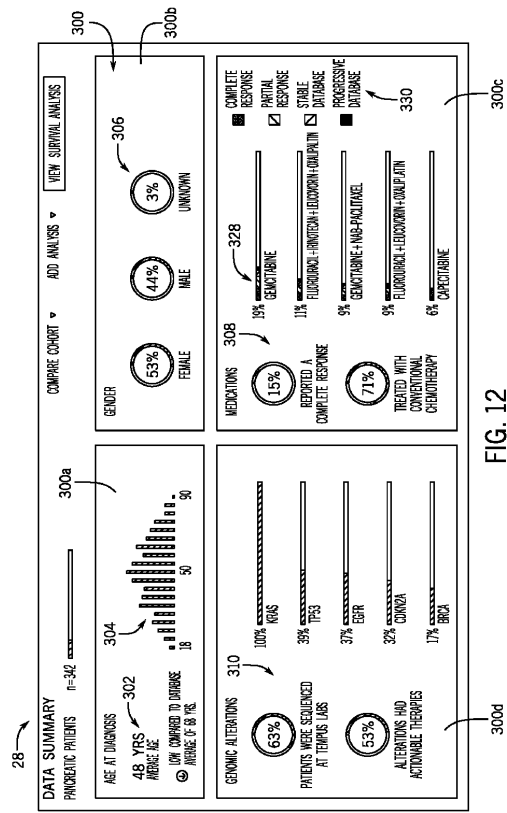


FIG. 12

1 3

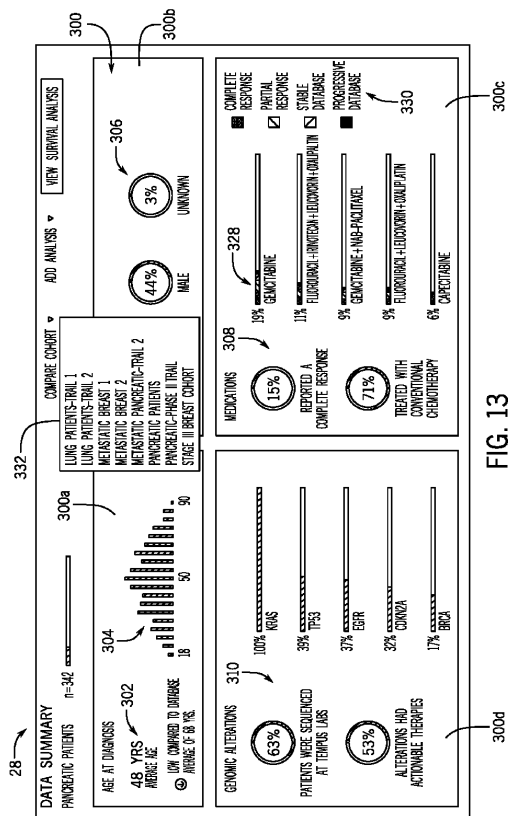


FIG. 13

1 4

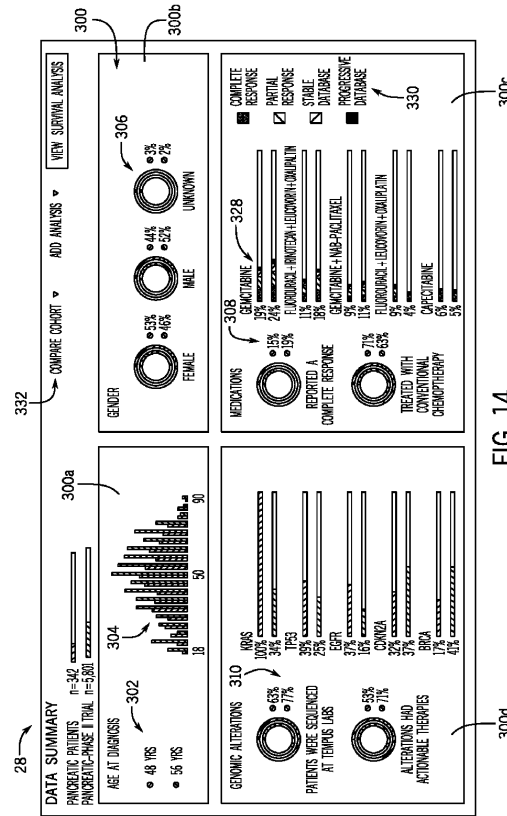


FIG. 14

10

20

30

40

50

【 1 5 】

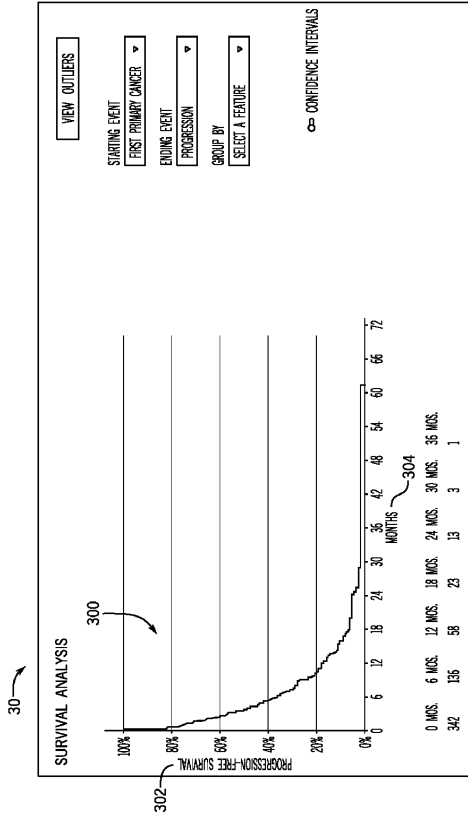


FIG. 15

【 1 6 】

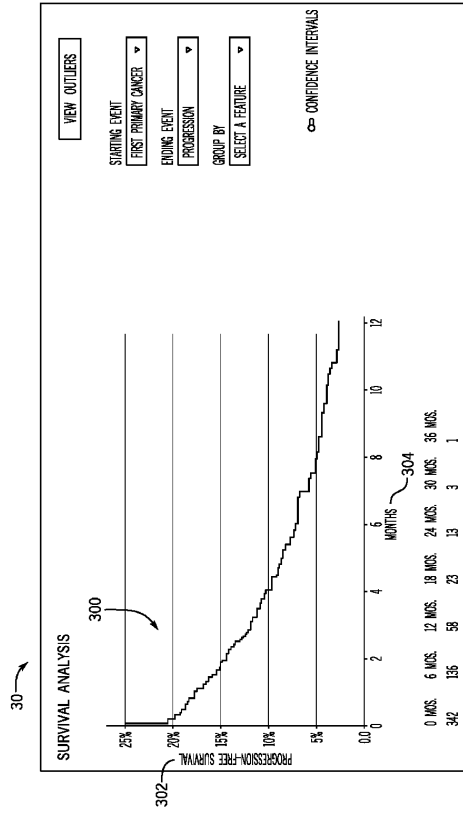


FIG. 16

【 1 7 】

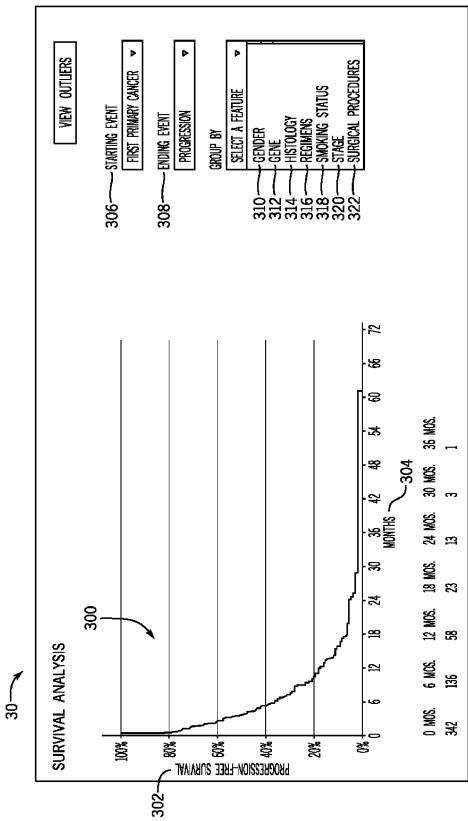


FIG. 17

【 1 8 】

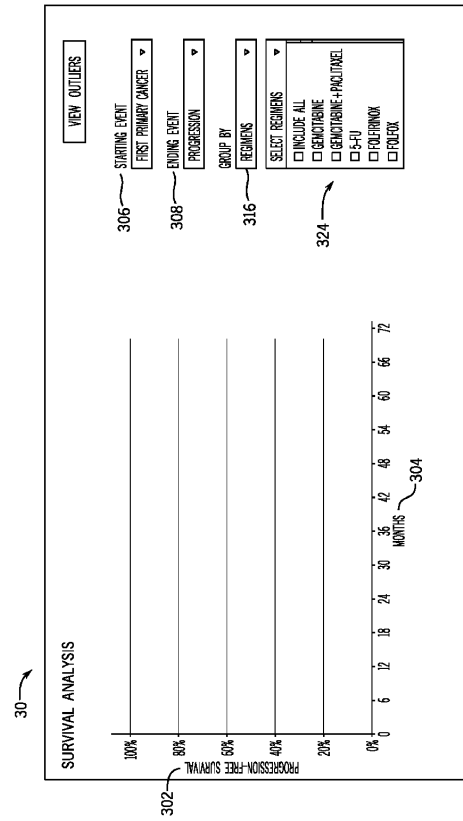


FIG. 18

10

20

30

40

50

【 19 】

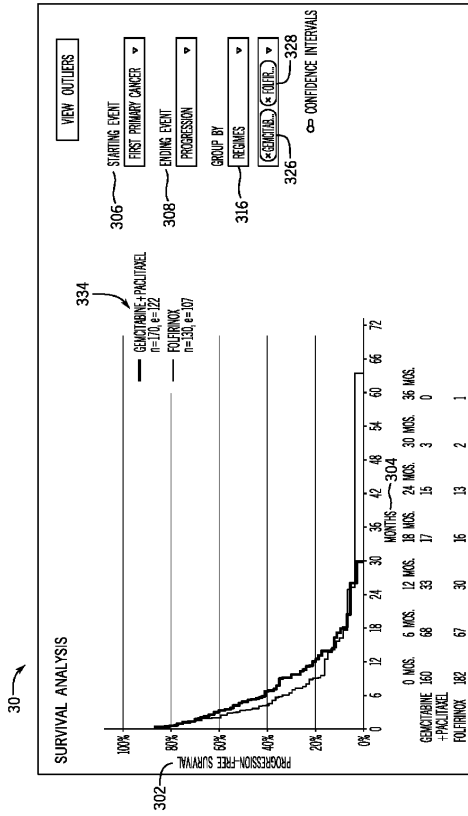


FIG. 19

【 20 】

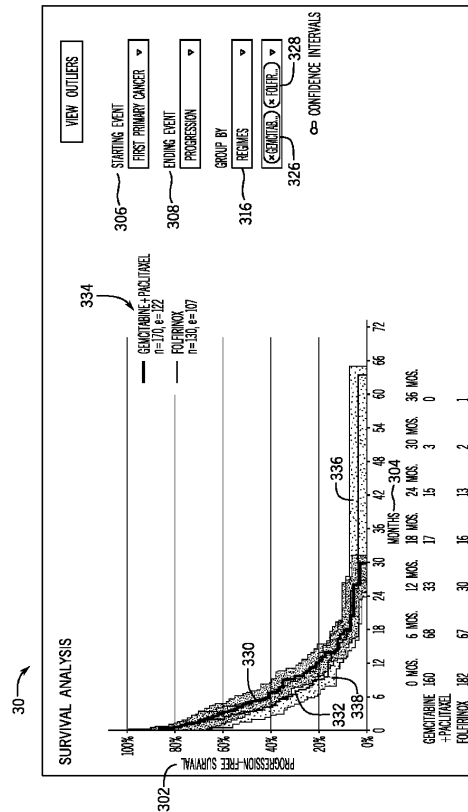


FIG. 20

【 21 】

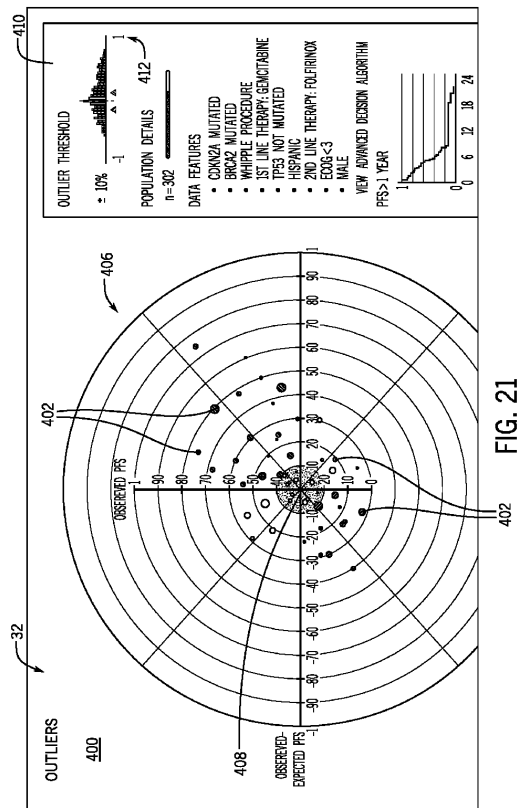


FIG. 21

【 22 】

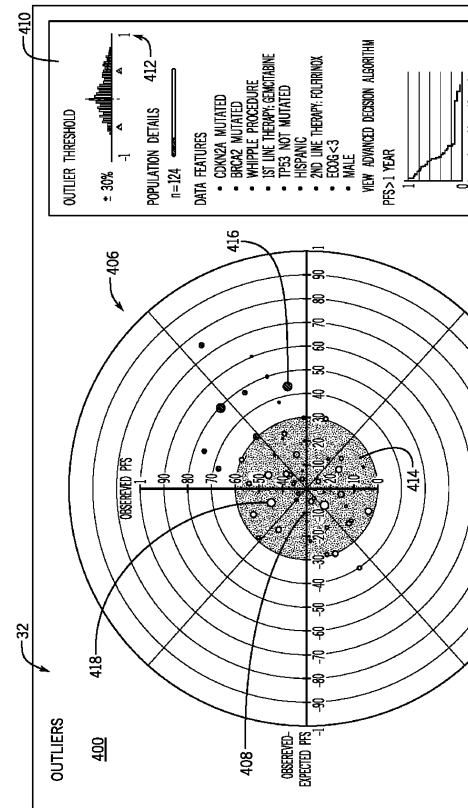


FIG. 22

10

20

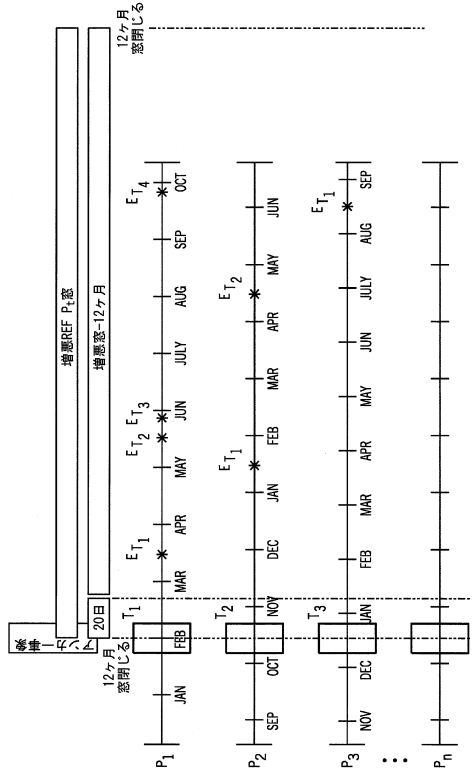
30

40

50



【 2 6 】



【 2 7 A 】

特徴	値
last stage	0.14485
cancer: breast_cancer	0.06266
historical-took_medication: gemcitabine	0.04721
historical-took_medication: paclitaxel	0.04686
cancer: lung_cancer	0.03740
sequencing: pgr:positive	0.03522
historical-took_medication: carboplatin	0.03478
historical-took_medication: bevacizumab	0.03350
cancer: pancreatic_cancer	0.03221
historical-got_radiotherapy: radiotherapy,_not_otherwise_specified	0.03046

10

20

【 2 7 B 】

特徴	値
historical-took_medication: irinotecan	0.06544
historical-took_medication: bevacizumab	0.05002
historical-took_medication: oxaliplatin	0.03810
historical-took_medication: capecitabine	0.03777
took_medication: oxaliplatin	0.03711
historical-took_medication: leucovorin	0.03680
historical-took_medication: fluorouracil	0.03317
historical-took_medication: cetuximab	0.03271
took_medication: fluorouracil	0.02646
sequencing: kras:mutation	0.02238

【 2 7 C 】

特徴	値
duplicate_column_1	0.06544
historical-took_medication: irinotecan	0.06544
historical-took_medication: bevacizumab	0.05002
historical-took_medication: oxaliplatin	0.03810
duplicate_column_2	0.03777
historical-took_medication: capecitabine	0.03777
took_medication: oxaliplatin	0.03711
historical-took_medication: leucovorin	0.03680
historical-took_medication: fluorouracil	0.03317
historical-took_medication: cetuximab	0.03271

30

40

50

【 27D 】

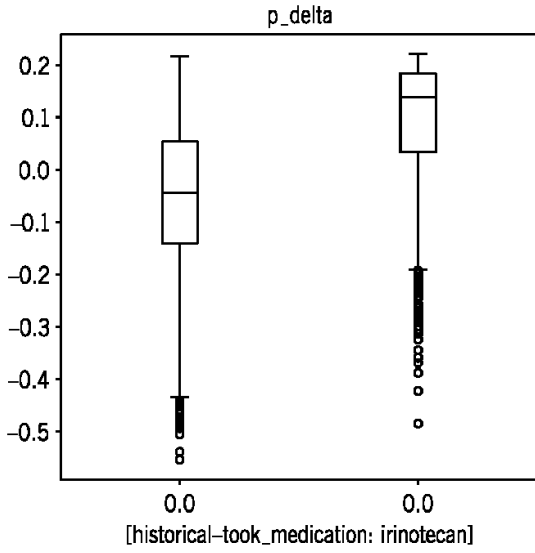


FIG. 27D

【 27E 】

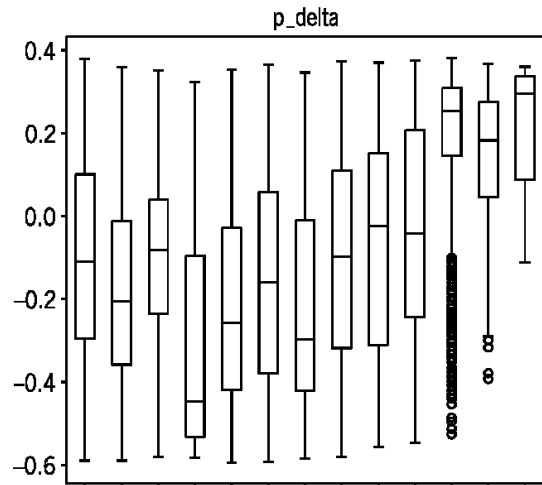
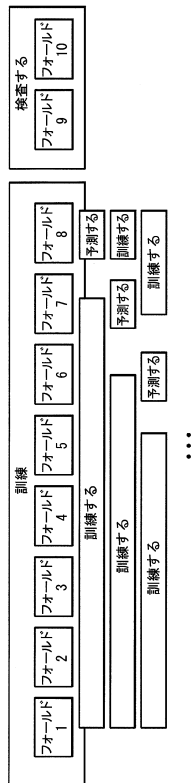


FIG. 27E

10

【 28 】



【 29 】

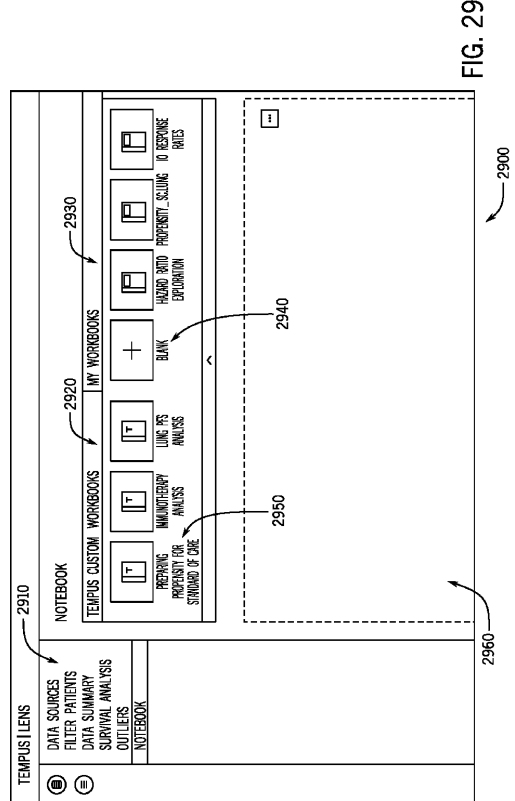


FIG. 29

20

30

40

50

【 3 0 】

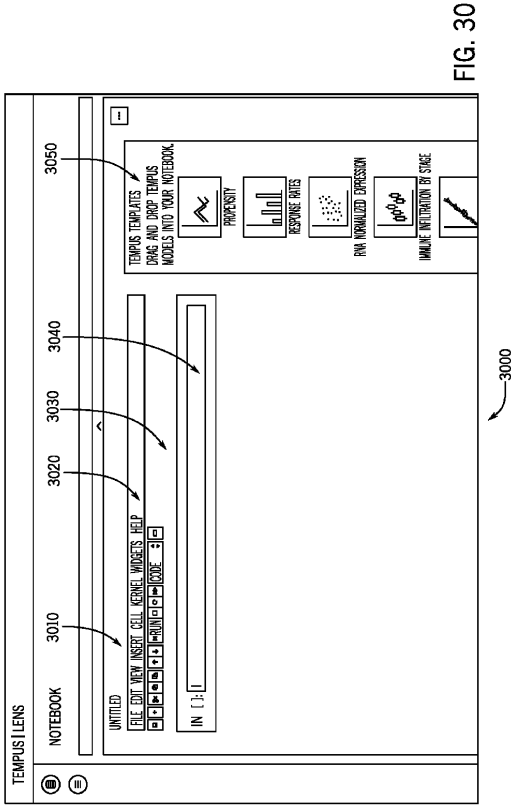


FIG. 30

【 3 1 】

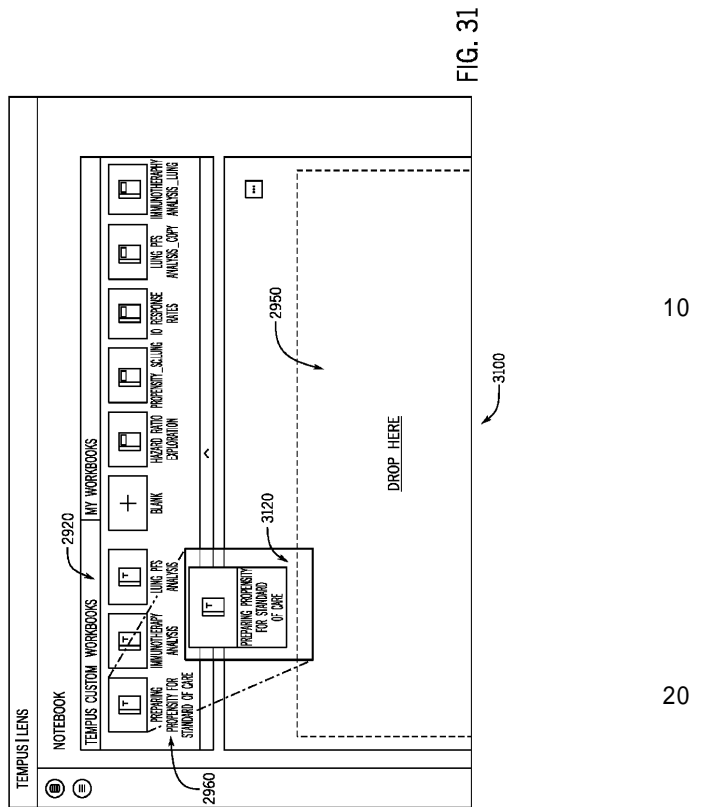


FIG. 31

【 3 2 】

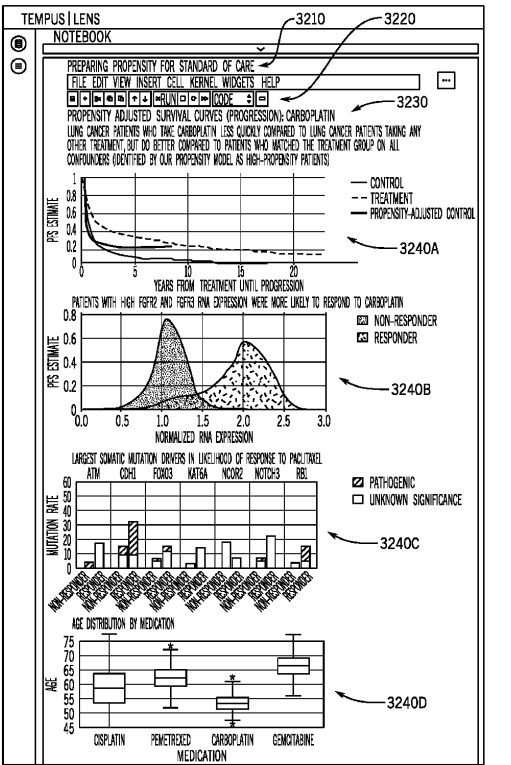


FIG. 32

【 3 3 】

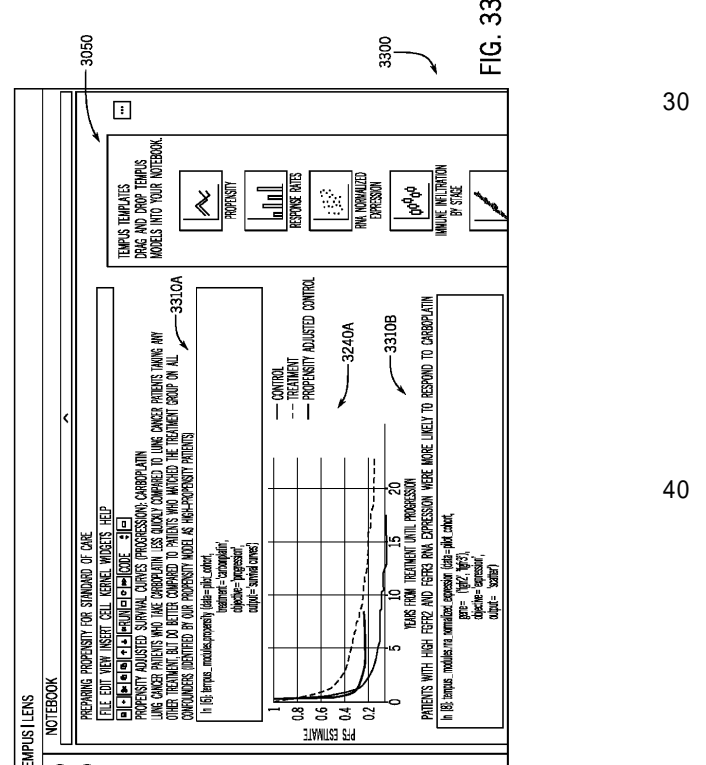


FIG. 33

10

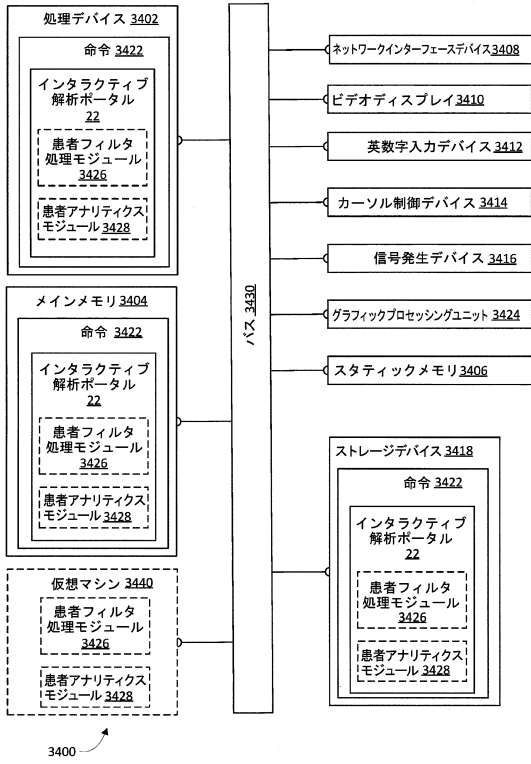
20

30

40

50

【 図 3 4 】



10

20

30

40

50

## フロントページの続き

- ベニュー・ 1 1 6 6  
(72)発明者 アシュラフ・ハーフェズ  
アメリカ合衆国・イリノイ・ 6 0 1 8 7 ・ウィートン・チャイルズ・ストリート・ 8 1 0
- (72)発明者 ジュリアン・ハビブ  
アメリカ合衆国・イリノイ・ 6 0 6 1 4 ・シカゴ・ノース・リンカーン・アベニュー・ 1 9 4 0
- (72)発明者 カリン・フィシェル  
アメリカ合衆国・イリノイ・ 6 0 6 2 2 ・シカゴ・ウェスト・ハードン・アベニュー・ 2 4 2 9 ・  
ユニット・ 2
- (72)発明者 キャロライン・エプスタイン  
アメリカ合衆国・イリノイ・ 6 0 6 1 0 ・シカゴ・ウェスト・チェスナット・ストリート・ 1 2 1  
・ユニット・ 2 8 0 5
- 審査官 森田 充功
- (56)参考文献 米国特許出願公開第 2 0 1 7 / 0 1 7 7 8 2 2 ( U S , A 1 )  
米国特許出願公開第 2 0 1 7 / 0 1 9 9 9 6 5 ( U S , A 1 )  
米国特許出願公開第 2 0 1 6 / 0 2 0 3 2 8 7 ( U S , A 1 )  
米国特許出願公開第 2 0 1 2 / 0 0 4 1 7 7 2 ( U S , A 1 )
- (58)調査した分野 (Int.Cl. , D B 名)  
G 1 6 H 1 0 / 0 0 - 8 0 / 0 0