

(19) 日本国特許庁 (JP)

(12) 特 許 公 報 (B2)

(11) 特許番号

特許第6063447号
(P6063447)

(45) 発行日 平成29年1月18日 (2017. 1. 18)

(24) 登録日 平成28年12月22日 (2016. 12. 22)

(51) Int. Cl.

F I

GO 1 N 33/48 (2006. 01)

GO 1 N 33/48 M

GO 6 T 1/00 (2006. 01)

GO 1 N 33/48 Z

GO 6 T 1/00 2 9 5

請求項の数 15 (全 40 頁)

(21) 出願番号 特願2014-505671 (P2014-505671)
 (86) (22) 出願日 平成24年4月23日 (2012. 4. 23)
 (65) 公表番号 特表2014-516413 (P2014-516413A)
 (43) 公表日 平成26年7月10日 (2014. 7. 10)
 (86) 国際出願番号 PCT/EP2012/057395
 (87) 国際公開番号 W02012/143562
 (87) 国際公開日 平成24年10月26日 (2012. 10. 26)
 審査請求日 平成27年4月21日 (2015. 4. 21)
 (31) 優先権主張番号 61/478, 224
 (32) 優先日 平成23年4月22日 (2011. 4. 22)
 (33) 優先権主張国 米国 (US)
 (31) 優先権主張番号 13/252, 069
 (32) 優先日 平成23年10月3日 (2011. 10. 3)
 (33) 優先権主張国 米国 (US)

(73) 特許権者 390041542
 ゼネラル・エレクトリック・カンパニー
 アメリカ合衆国、ニューヨーク州 1 2 3
 4 5、スケネクタデイ、リバーロード、1
 番
 (74) 代理人 100137545
 弁理士 荒川 聡志
 (74) 代理人 100105588
 弁理士 小倉 博
 (74) 代理人 100129779
 弁理士 黒川 俊久
 (74) 代理人 100113974
 弁理士 田中 拓人

最終頁に続く

(54) 【発明の名称】 細胞におけるバイオマーカーの発現のクラスターによる解析

(57) 【特許請求の範囲】

【請求項 1】

マルチプレックス化バイOMETリック画像データに基づいて、組織特徴を解析する方法であって、

個別の細胞が描出されて区画へと分割される複数の視野に関して、複数のバイオマーカーの発現を捕捉するマルチプレックス化バイOMETリック画像を含む細胞プロファイルデータを含むデータセットを保存する段階であって、細胞プロファイルデータが、共通性を有する患者コホートから得られた複数の組織試料から作成され、データセットが、共通性と関連する視野レベルの評価又は患者レベルの評価を含むメタ情報の少なくとも一部分との細胞プロファイルデータの関連をさらに含んでいる、段階と、

データセットから類似する細胞の複数のクラスターセットを作成する段階であって、複数のクラスターセットの各々がユニークな数のクラスターを含み、各細胞が、複数のクラスターセットの各々の中の単一のクラスターに割り当てられ、複数のクラスターセットの各々の中の複数のクラスターの各々が、複数の選択された属性であって、そのセットの中の他のクラスターの中の細胞の複数の選択された属性より、そのクラスターの中の他の細胞の複数の選択された属性への類似性が大きい属性を有する細胞を含んでいる、段階と、

複数のクラスターセットの各々の中で、各クラスターに割り当てられた細胞の比率を観察する段階と、

共通性と関連する視野レベルの評価又は患者レベルの評価を含むメタ情報の少なくとも一部分との関連について、観察された比率を検証する段階と、

10

20

複数のクラスターセットについての少なくとも1つのモデルの性能の比較に基づいて、予測的なクラスターセットを含む複数のクラスターセットの1つを選択する段階とを含んでおり、細胞データがトレーニングデータ及びテストデータを含み、類似する細胞の複数のクラスターセットがトレーニングデータから作成され、比較のための少なくとも1つのモデルの性能がテストデータから決定される、方法。

【請求項2】

データセットが複数のバッチと関連し、

複数のバッチの1つの中の全ての細胞についての細胞全体の中央値強度を、そのバッチの中の各細胞についての細胞全体の中央値強度、核の中央値強度、膜の中央値強度及び細胞質の中央値強度の各々から減じることにより、複数のバッチに関して細胞プロファイルデータを正規化する段階をさらに含んでいて、

複数のクラスターセットを作成する段階が、正規化されたデータセットから、類似する細胞の複数のクラスターセットを作成することを含む、請求項1記載の方法。

【請求項3】

細胞の類似性が、

a) 複数のバイオマーカーの少なくとも1つの発現に基づく、細胞の少なくとも1つの属性の比較、

b) 細胞の少なくとも2つの属性の比較であって、少なくとも2つの属性の各々が複数のバイオマーカーの少なくとも1つの発現に基づく、細胞の少なくとも2つの属性の比較

c) 細胞の少なくとも3つの属性の比較であって、少なくとも3つの属性の各々が複数のバイオマーカーの少なくとも1つの発現に基づく、細胞の少なくとも3つの属性の比較、又は

d) 細胞の少なくとも3つの属性の比較であって、少なくとも4つの属性の各々が複数のバイオマーカーの少なくとも1つの発現に基づく、細胞の少なくとも4つの属性の比較に基づく、請求項1記載の方法。

【請求項4】

細胞の少なくとも1つの属性が、細胞全体の中央値強度、核の強度比、膜の強度比及び細胞質の強度比からなる細胞の4つの特徴から選択され、

核の強度比が、膜の中央値強度と細胞質の中央値強度との合計の半分を、核の中央値強度から減じることにより計算され、

膜の強度比が、核の中央値強度と細胞質の中央値強度との合計の半分を、膜の中央値強度から減じることにより計算され、

細胞質の強度比が、膜の中央値強度と核の中央値強度との合計の半分を、細胞質の中央値強度から減じることにより計算される、請求項1記載の方法。

【請求項5】

請求項1記載の方法であって、

a) 正常細胞の細胞プロファイルが、類似する細胞の複数のクラスターセットの作成に用いられるデータセットから除外されるか、或いは

b) 観察された細胞の比率が、各クラスターに割り当てられた各視野又は各患者の細胞の観察された比率である、方法。

【請求項6】

観察された比率を検証することが、共通性と関連する視野レベルの評価を含むメタ情報の少なくとも一部分との関連について、観察された比率を検証することを含み、

予測的なクラスターセット選択することが、複数のクラスターセットに基づく視野レベルの評価モデルの性能の比較に基づいて、予測的なクラスターセットを選択することを含む、請求項1記載の方法。

【請求項7】

観察された比率を検証することが、状態又は疾患についての予後診断[生存期間]との関連について、観察された比率を検証することを含み、

複数のクラスターセットの1つを選択することが、複数のクラスターセットに基づく患者レベルの評価モデルの性能の比較に基づいて、複数のクラスターセットの1つを選択することを含む、請求項1記載の方法。

【請求項8】

a) 複数のクラスターセットの各々の中のクラスターの数に関して少なくとも1つのモデルの性能を比較すること、

b) メタ情報の少なくとも一部分との一変量の関連について、選択されたクラスターセットにおいて観察された比率を検証すること、

c) メタ情報の少なくとも一部分との多変量の関連について、選択されたクラスターセットにおいて観察された比率を検証すること、

d) 閾値を超える一致率に対応するクラスターセットについての少なくとも1つのモデルの性能に基づいて、予測的なクラスターセットを選択すること、

e) 少なくとも1つの予測的なクラスターを、予測的なクラスターセットから同定すること、又は

f) 複数のバイオマーカーの少なくとも1つの発現に基づいて、中央値又は平均によるクラスター化アルゴリズムを細胞の少なくとも1つの属性に適用することにより細胞の類似性を決定すること

をさらに含む、請求項1記載の方法。

【請求項9】

予測的なクラスターセットを選択することが、

a) その数を上回ってクラスターセットの中のクラスターの数が大きくなると、統計学的に有意な性能の増大がもたらされない数のクラスター、又は

b) その数を下回ってクラスターセットの中のクラスターの数が小さくなると、性能の減少がもたらされる数のクラスター

を有する複数のクラスターセットの1つを選択することをさらに含む、請求項1記載の方法。

【請求項10】

マルチプレックス化バイオメトリック画像データに基づいて、組織特徴を解析するシステムであって、

個別の細胞が描出されて区画へと分割される複数の視野に関して、複数のバイオマーカーの発現を捕捉するマルチプレックス化バイオメトリック画像を含む細胞プロファイルデータを含むデータセットを保存するための保存デバイスであって、細胞プロファイルデータを、共通性を有する患者コホートから得られた複数の組織試料から作成し、データセットが、細胞プロファイルデータの、共通性と関連する視野レベルの評価又は患者レベルの評価を含むメタ情報の少なくとも一部分との関連をさらに含む保存デバイスと、

少なくとも1つのプロセッサであって、

類似する細胞の複数のクラスターセットをデータセットから作成する段階であって、複数のクラスターセットの各々がユニークな数のクラスターを含み、各細胞を複数のクラスターセットの各々の中の単一のクラスターに割り当て、複数のクラスターセットの各々の中の複数のクラスターの各々が、複数の選択された属性であり、そのセットの中の他のクラスターの中の細胞の複数の選択された属性より、そのクラスターの中の他の細胞の複数の選択された属性への類似性が大きい属性を有する細胞を含む段階、

複数のクラスターセットの各々の中で、各クラスターに割り当てられた細胞の比率を観察する段階、及び

共通性と関連する視野レベルの評価又は患者レベルの評価を含むメタ情報の少なくとも一部分との関連について、観察された比率を検証する段階

を少なくとも1つのプロセッサに実行させるコードを実行するための少なくとも1つのプロセッサと、

複数のクラスターセットについての少なくとも1つのモデルの性能の比較に基づいて、予測的なクラスターセットを含む複数のクラスターセットの1つを選択できるようにする

10

20

30

40

50

視覚表示デバイス

を含んでおり、少なくとも1つのプロセッサが、細胞データをトレーニングデータ及びテストデータへとさらに分割し、類似する細胞の複数のクラスターセットをトレーニングデータから作成し、比較のための少なくとも1つのモデルの性能をテストデータから決定する段階を少なくとも1つのプロセッサに実行させるコードをさらに実行する、システム。

【請求項11】

データセットが複数のバッチと関連し、少なくとも1つのプロセッサが、複数のバッチの1つの中の全ての細胞についての細胞全体の中央値強度を、そのバッチの中の各細胞についての細胞全体の中央値強度、核の中央値強度、膜の中央値強度及び細胞質の中央値強度の各々から減じることにより、複数のバッチに関して細胞プロファイルデータを正規化する段階

10

を少なくとも1つのプロセッサに実行させるコードをさらに実行し、

複数のクラスターセットを作成することが、正規化されたデータセットから、類似する細胞の複数のクラスターセットを作成することを含む、請求項10記載のシステム。

【請求項12】

請求項10記載のシステムであって、

a) 細胞の類似性が、複数のバイオマーカーの少なくとも1つの発現に基づく、細胞の少なくとも1つの属性の比較に基づくか、或いは
b) 観察された細胞の比率が、各クラスターに割り当てられた各視野又は各患者の細胞の観察された比率を含む、システム。

20

【請求項13】

少なくとも1つのプロセッサが、

a) 細胞の少なくとも2つの属性の比較であって各々が複数のバイオマーカーの少なくとも1つの発現に基づく細胞の少なくとも2つの属性の比較に基づいて、細胞の類似性を決定する段階、

b) 細胞の少なくとも3つの属性の比較であって各々が複数のバイオマーカーの少なくとも1つの発現に基づく細胞の少なくとも3つの属性の比較に基づいて、細胞の類似性を決定する段階、

30

c) 細胞の少なくとも4つの属性の比較であって各々が複数のバイオマーカーの少なくとも1つの発現に基づく細胞の少なくとも4つの属性の比較に基づいて、細胞の類似性を決定する段階、

d) 正常細胞の細胞プロファイルを、類似する細胞の複数のクラスターセットの作成に用いられるデータセットから除外する段階、

e) 複数のバイオマーカーの少なくとも1つの発現に基づいて、中央値によるクラスター化アルゴリズムを細胞の少なくとも1つの属性に適用することにより細胞の類似性を決定する段階、

f) 複数のバイオマーカーの少なくとも1つの発現に基づいて、平均によるクラスター化アルゴリズムを細胞の少なくとも1つの属性に適用することにより細胞の類似性を決定する段階、

40

g) 複数のクラスターセットの各々の中のクラスターの数に関して少なくとも1つのモデルの性能を比較する段階、或いは

h) 少なくとも1つの予測的なクラスターを、予測的なクラスターセットから同定する段階

を少なくとも1つのプロセッサに実行させるコードをさらに実行する、請求項10記載のシステム。

【請求項14】

視覚表示デバイスが、

a) その数を上回ってクラスターセットの中のクラスターの数が大きくなると、統計学的

50

に有意な性能の増大がもたらされない数のクラスター、又は
b) その数を下回ってクラスターセットの中のクラスターの数が小さくなると、性能の減少がもたらされる数のクラスター
を有する複数のクラスターセットの1つを選択できるようにする、請求項10記載のシステム。

【請求項15】

視覚表示デバイスが、

a) 閾値を超える一致率に対応するクラスターセットについての少なくとも1つのモデルの性能に基づいて、複数のクラスターセットの1つを選択できるようにするか、或いは
b) 使用者が第1のクラスターの特徴を理解できるようにするために、第1のクラスターの中の第1の細胞のモンタージュと、第1のクラスターの中の第2の細胞のモンタージュとをさらに同時に表示する、
請求項10記載のシステム。

10

【発明の詳細な説明】

【技術分野】

【0001】

本発明は一般に、それらの元の組織においてインサイチュで検証される個別の細胞におけるバイオマーカーの発現を解析及び視覚化し、状態もしくは疾患の診断、予後診断又は状態もしくは疾患の処置に対する応答と関連する発現パターンを同定及び理解することに関する。

20

【背景技術】

【0002】

バイオマーカーの発現を明らかにするように処理された組織検体の検証は、生物学的探索及び臨床的研究の公知のツールである。このような処理の1つは、バイオマーカーに特異的な対象の抗体又は抗体断片などの抗体サロゲート、一般にはタンパク質の使用を伴う。このような抗体又は抗体サロゲートは、適切な条件下でシグナルを発生させることが可能な部分で直接又は間接に標識することができる。例えば、蛍光部分を抗体に接合させて、処理された組織を蛍光について精査することができる。得られるシグナルは一般に、バイオマーカーの存在だけでなく、存在するバイオマーカーの量もまた示す。

30

【0003】

組織の処理及び検証の技術は、特定の細胞、或いはさらに核、細胞質、もしくは膜など、所与の細胞区画における所与のバイオマーカーの発現レベルを定量的に決定しうるように洗練されている。これらの区画又は細胞全体の境界は、公知の組織学的染色を用いて位置特定される。一般に、処理された組織をデジタルイメージングにより検証し、その結果として、異なるバイオマーカーから発せられる異なるシグナルのレベルを容易に定量化することができる。

【0004】

所与の組織検体を多数のバイオマーカーの発現について調べることを可能とする技術がさらに開発されている。一般に、この技術は、フルオロフォアで標識したプローブで検体を染色して、1つ以上のプローブが結合したバイオマーカーについてシグナルを発生させ、これらのシグナルを化学的に漂白し、検体を再染色して、いくつかのさらなるバイオマーカーについてシグナルを発生させることを伴う。互いから容易に差別化されうる限定された数のシグナルが存在するに過ぎず、特定の段階において限定された数のバイオマーカーだけを検証しうるため、化学的漂白の段階は好都合である。漂白するだけで、複数段階のために試料を再プローブし、再評価することができる。このサイクリング法は、ホルマリン固定化パラフィン包埋組織(FFP E)による試料及び細胞において用いることができる。検体のデジタル画像は、各染色段階の後に収集する。このような検体の継起画像は、そのシグナルが化学的漂白法により改変されていないDAPI染色された細胞核などの形態特徴を用いて登録しうると好都合である。

40

50

【0005】

別の手法は、凍結させた組織検体を、それらを繰り返し染色し、前の染色段階から標識を光漂白してから、次の染色セットを適用することにより検証することであった。次いで、評価される各バイオマーカーと関連する蛍光シグナルの強度を、適切な画像から抽出する。

【0006】

このデータを使用して、バイオマーカーの発現パターンを同定する試みがなされている。1つの手法は、組織検体の全体においてこのようなパターンを探し出し、閾値を用いてフルオロフォアシグナルを二値化し、多様な発現プロファイルを作成し、次いで、これを対象の組織画像にオーバーレイすることであった。

10

【0007】

「Method and System for Analyzing the Expression of Biomarkers in Cells in Situ in Their Tissue of Origin」と題する米国特許出願公開第2011/0091081号及び「Process and System for Analyzing the Expression of Biomarkers in Cells」と題する米国特許出願公開第2011/0091091号はいずれも、本発明に先行するGeneral Electricによる研究及び開発の成果について記載している。

【0008】

20

米国特許出願公開第2011/0091081号は、それらの元の組織内の細胞における複数のバイオマーカーの発現パターンを解析するためのデータを取得する工程について開示した。個別の細胞又は個別の細胞の細胞内区画における複数のバイオマーカーの発現レベルが、細胞の元の組織においてインサイチュで測定された。測定は、組織検体を、対象のバイオマーカーに特異的な抗体又は抗体サロゲートで処理することによれば、簡便に行うことができるであろう。抗体又は抗体サロゲートは、適切な波長の光で精査すると光学シグナルを発する部分で直接又は間接に標識された。組織検体は、繰り返し処理され、各処理は、他の任意の処理に關与するバイオマーカーとは異なるバイオマーカーに特異的な抗体又は抗体サロゲートを伴い、直前の処理から発生したシグナルは、光学的手段又は化学的手段により中和された。抗体又は抗体サロゲートを介して対象のバイオマーカーへと結合させた各標識の量は、検体を適切な波長の光の下に置き、応答をデジタルイメージングすることにより測定した。細胞は、個別の細胞単位へと分割するのが好都合であり、それらの細胞内区画（膜、細胞質及び核を含む）が、データ取得の部分であった。データベースは、元の測定値及び各測定値が取り出される位置、細胞又は細胞区画を保存した。

30

【0009】

米国特許出願公開第2011/0091081号はまた、それらの元の組織の中の細胞における複数のバイオマーカーの発現パターンを表すデータを解析する工程についても開示した。データベースを精査するのに用いられる数値的方法は、バイオマーカーの発現レベルの測定に基づき、特定の属性を対象の各細胞に割り当て、類似するバイオマーカーの発現属性を有する細胞を併せて群分けすることを伴った。群分けすることは、属性空間内のそれらの間の距離が最小である細胞を併せて群分けするアルゴリズム、すなわち、各属性が1つの次元を割り当てられるn次元空間におけるそれらの互いからの距離に基づき、2つの細胞が同じ群に含まれるアルゴリズムを伴った。

40

【0010】

米国特許出願公開第2011/0091081号は、特定のバイオマーカーの発現パターンが類似する細胞群であれば、生物学的状態と所与の細胞属性との間の関連を探索するための好都合な基盤でありうることに付いてさらに開示した。各群分けを検証すれば、所与の状態もしくは疾患の診断もしくは予後診断と関連する任意の細胞属性又は所与の状態もしくは疾患のための所与の治療に対する応答と関連する任意の細胞属性を同定しうるであろう。

50

【0011】

米国特許出願公開第2011/0091081号は、特定のバイオマーカーの発現パターンが類似する1つ以上の細胞群を表示する工程について開示した。バイオマーカーの発現レベルを測定するのに使用される視野についてのデジタル画像の1つ以上にわたるオーバーレイによれば、群分けを視覚化しうるであろう。オーバーレイによれば、元の画像のどこに、所与の群のプロファイルを保有する細胞が存在するのかが示しうるであろう。異なる組織検体に由来する、このようなオーバーレイを伴う画像を比較すれば、1つ以上のプロファイルを伴う細胞のパターン、すなわち、1つ以上の群に属する細胞のパターンが、任意の生物学的状態又は生物学的過程を示すかどうかを決定しうるであろう。

【0012】

米国特許出願公開第2011/0091091号は、細胞試料の個別の細胞における複数のバイオマーカーの発現レベルを測定し、各細胞のバイオマーカー発現の測定値を、データベースの中のデータ点として保存し、データベースを、バイオマーカーの発現パターンが類似するデータ点について、各バイオマーカーの発現レベルを少なくとも半連続変数として用いる数値解析によりこのような類似性を決定するコンピュータアルゴリズムを用いて、精査することを含む工程について開示した。分散が最小であるデータ点を同定し、併せて群分けした。群には、アルゴリズムにより考慮される各属性についての中央値に基づく新たなデータ点により表される新たなバイオマーカーの発現プロファイルを割り当て、こうして、新たなデータセットを形成した。所定数の群が作成されるまで、この段階を、新たなデータセットにより繰り返した。

【0013】

米国特許出願公開第2011/0091091号はまた、特定のバイオマーカーの発現パターンが類似する細胞群を表示するのに群分けデータを用いる方法についても開示した。方法は、細胞試料の視野の中の1つ以上の群についての画像であって、群の中の各細胞に、それらが同じ群に属することの目視可能な表示を与える画像を創出することを伴った。新たな画像は、試料の元の画像へと登録されて、視野の中の群についての画像が継的にオーバーレイされ、解析され、表示されることを可能とする。

【発明の概要】

【課題を解決するための手段】

【0014】

本発明は、先行技術の1つ以上の限界に取り組む。例えば、米国特許出願公開第2011/0091081号及び米国特許出願公開第2011/0091091号のいずれも、可能な関連を探索するための特定のデータセットに適する数の群をどのようにして選択するのかについて開示できていない。米国特許出願公開第2011/0091091号は、特定のデータセットの中で所定数の群を作成することについて開示しているが、作成する群の数をどのようにして選択するのかについては開示していない。特定のデータセットに適する数の群を選択する手法がなければ、適切な数の群を選択することはできない。群が少なすぎれば、重要な示差的特徴を伴う細胞が併せて群分けされる結果がもたらされうる。群分けされた細胞のサブセットの関連は、同定するのがより困難な場合もあり、不可能な場合もある。群が多すぎれば、不必要に複雑な計算及び解析が結果として必要となる。群が多すぎれば、重要な示差的特徴を伴わない細胞が個別に群分けされるように、データセットをオーバーフィッティングさせる結果がもたらされうる。重要な示差的特徴を有さない2つの細胞群による関連は、同定するのがより困難な場合もあり、不可能な場合もある。

【0015】

別の例として述べると、米国特許出願公開第2011/0091081号及び米国特許出願公開第2011/0091091号のいずれも、群に関する情報を表示する技術についての開示が限定されている。いずれの公報も、群に割り当てられる細胞の位置を、はるかに大きな視野の中で印づけうることに開示している。いずれの公報も、細胞を、はるかに大きな視野の中で印づけて、それらの同じ視野の中の複数の群の1つへの割当て

を示しうることについてさらに開示している。しかし、はるかに大きな視野の中のそれらの相対位置以外に、このような表示が任意の特定の群の中の細胞の特徴に対してもたらず洞察は限定されている。さらに、多次元的類似性による細胞の群分けによって得られる群は、医療従事者が理解するのに本質的に困難である。したがって、本明細書で教示される実施形態は、データセットを解析する異なる工程を伴う。

【0016】

以下の説明を、添付図を参照しながら読めば、本発明の特徴、態様及び利点がよりよく理解されるであろう。

【図面の簡単な説明】

【0017】

【図1】本明細書で教示される例示的な実施形態を実施するのに適する例示的な演算環境を示す図である。

【図2】本明細書で教示される実施形態に従い、類似する細胞の予測的なクラスターセットをデータセットから同定するためのモデルを開発する例示的な方法を示す図である。

【図3】本明細書で教示される実施形態に従い、細胞クラスターの特徴を表示する例示的な方法を示す図である。

【図4】本明細書で教示される実施形態に従い、モデルのクラスターセットを新たな細胞プロファイルデータに適用する例示的な方法を示す図である。

【図5】本明細書で教示される実施形態に従い、細胞特徴の積率の予測的なセットをデータセットから同定するためのモデルを開発する例示的な方法を示す図である。

【図6】本明細書で教示される実施形態に従い、積率のモデルセットを新たな細胞プロファイルデータに適用する例示的な方法を示す図である。

【図7】マーカーデータ及び形態特徴の第1の2つの積率を含む、がん/正常分類子についてのROC (receiver operating characteristic) 曲線である。

【図8】マーカーデータの第1の2つの積率を含む、がんだけの分類子についてのROC 曲線である。

【図9】マーカーデータ及び形態特徴の第1の2つの積率を含む、がん/正常分類子についての変数重要度プロットである。

【図10】マーカーデータの第1の2つの積率を含む、がんだけの分類子についての変数重要度プロットである。

【図11】がん/正常分類子における上位4つの特徴についての偏変数依存性プロットである。

【図12】高悪性度/低悪性度分類子における上位4つの特徴についての偏変数依存性プロットである。

【図13】全コホートの生存モデルについての変数重要度を示すグラフである。

【図14】全コホートの生存モデルについての偏変数依存性プロットのグラフである。

【図15】グリソンスコア>0のコホートに対する生存モデルについての変数重要度を示すグラフである。

【図16】グリソンスコア>0のコホートの生存モデルについての偏変数依存性プロットである。

【図17】バッチごとの侵襲性視野(FOV)において観察される膜P13K p110aの平均を示す図である。

【図18】数を変化させる細胞クラスターの特徴に基づき、がん/正常分類子についてのROC 曲線下面積(AUC)を示す図である。

【図19】数を変化させる細胞クラスターの特徴に基づき、高悪性度/低悪性度がん分類子についてのROC 曲線下面積を示す図である。

【図20】がん/正常FOVの20の細胞クラスターモデルについてのROC 曲線である。

【図21】高悪性度/低悪性度FOVの20の細胞クラスターモデルについてのROC 曲

10

20

30

40

50

線である。

【図 2 2】がん / 正常 F O V の 2 0 のクラスター分類子モデルについての変数重要度を示す図である。

【図 2 3】高悪性度 / 低悪性度ががん F O V の 2 0 のクラスター分類子モデルの変数重要度を示す図である。

【図 2 4】がん / 正常分類子における上位 4 つの特徴についての偏変数依存性プロットである。

【図 2 5】高悪性度 / 低悪性度ががん分類子における上位 4 つの特徴についての偏変数依存性プロットである。

【図 2 6】バッチごとに (各パネルにおいて) 、かつ、がん対正常 (真 / 偽で示した) の対比により、クラスター 7 の細胞について観察された F O V レベルの比率を示す図である。x 軸は、F O V 中のクラスター 7 の比率の平方根である。

【図 2 7】クラスター 7 / 2 0 のシグネチャーを示す図である。各水平線の黒丸端は、クラスター 7 中の平均であり、他端は、2 0 クラスター全ての平均である。

【図 2 8】全コホートに対する生存モデルについての性能メトリックを示す図である。R S F による一致率、並びに 3、5 及び 1 0 年以内の前立腺がんによる死亡を分類するための A U C である。年齢及びグリソンスコアだけを含むヌルモデルの性能を、水平線として示す。

【図 2 9】グリソンスコア > 0 のコホートに対する生存モデルについての性能メトリックを示す図である。R S F による一致率、並びに 3、5 及び 1 0 年以内の前立腺がんによる死亡を分類するための A U C である。年齢及びグリソンスコアだけを含むヌルモデルの性能を、水平の黒色線として示す。

【図 3 0】全コホートの生存モデルについての変数重要度を示す図である。

【図 3 1】全コホートの生存解析における上位 4 つの特徴についての偏変数依存性プロットである。

【図 3 2】グリソンスコア > 0 のコホートについての生存モデルの変数重要度を示す図である。

【図 3 3】グリソンスコア > 0 のコホートの 2 0 クラスターのモデルにおける上位 4 つの特徴の偏依存性を示す図である。

【図 3 4】いずれも生存期間の短縮の表示である、クラスター 6 / 8 及びクラスター 1 / 2 0 のシグネチャーを示す図である。

【図 3 5】本明細書で教示される実施形態に従う、クラスターの中の 2 つの細胞の例示的なモンタージュを示す図である。

【発明を実施するための形態】

【0018】

本明細書で教示される実施形態では、マルチプレックス化染色 - 脱染法を介するなど、公知の技術を介して作成されるマルチプレックス化バイオメトリック画像を活用する。画像は、個別の細胞を互いと比較することを可能とする、個別の細胞内におけるバイオマーカーの発現を例示する。個別の細胞は、より大きな細胞試料の部分である。例えば、細胞試料は、細胞培養物、組織試料、臓器、腫瘍又は病変に由来する細胞群でありうる。個別の細胞はまた、異なる対象に由来する類似する組織による検体群の部分でもありうる。これらの細胞群は、1 つ以上の疾患モデル又は状態モデルを表す場合もあり、疾患モデル又は状態モデルの中の異なる段階を表す場合もあり、疾患又は状態の処置に対する 1 つ以上の応答を表す場合もある。

【0019】

染色された各視野の画像は、適切な顕微鏡及び適切な品質管理ルーチンとカップリングされたデジタルカメラによるなど、公知の技術を介して作成される。自動式画像の登録及び解析はまた、描出される個別の細胞、或いはさらに核、細胞質及び膜などの細胞内区画について、バイオマーカーの濃度レベルを定量化するのにも用いることができる。細胞のマルチプレックス化解析及び画像解析から得られるデータ値は、単独で保存することでも

10

20

30

40

50

き、さらなる解析の結果であるデータと共に保存することもできる。データベースは、組織及びそれが得られた組織中の位置を含めて、バイオマーカーの発現強度の測定内容を保存する。位置は、ある測定値が得られた細胞を含んでいるべきであり、測定と関連する区画、核、細胞質又は膜を含んでいてもよい。データベースには、保存デバイス 116 又はネットワークデバイス 126 において維持されうる情報が保存される。

【0020】

図1は、本明細書で教示される例示的な実施形態を実施するのに適する例示的な演算環境を示す。環境は、関連する周辺デバイスを伴う演算デバイス100を含む。演算デバイス100は、本明細書で教示される多様な方法のための実行用コード150を実装するようにプログラム可能である。演算デバイス100は、ハードディスクドライブ、CD-ROM又は他の非揮発性コンピュータ読み取り可能媒体など、保存デバイス116を含む。保存デバイス116は、オペレーティングシステム118及び他の関連するソフトウェアを保存する。演算デバイス100は、メモリ106をさらに含んでいてもよい。メモリ106は、DRAM、SRAM、EDO RAMなど、コンピュータのシステムメモリ又はランダムアクセスメモリを含みうる。メモリ106は、他の種類のメモリ又はその組合せも含みうる。演算デバイス100は、保存デバイス116及び/又はメモリ106内に、実行用コード150のモジュールを実装及び処理するための命令を保存しうる。

【0021】

演算デバイス100はまた、メモリ106に保存されるソフトウェアを実行するためのプロセッサ102及び1つ以上のプロセッサ102'、並びにシステムのハードウェアを制御するための他のプログラムも含む。プロセッサ102及びプロセッサ102'の各々は、単一のコアプロセッサの場合もあり、複数のコアプロセッサの場合もある(104及び104')。演算デバイス100では、演算デバイスの中のインフラストラクチャー及びリソースを動的に共有しうるように、仮想化を使用することができる。仮想化されたプロセッサはまた、保存デバイス116内の実行用解析コード150及び他のソフトウェアと共に用いることもできる。仮想マシン114を用意して、工程が複数ではなくて唯一の演算リソースを用いていると考えられるように、複数のプロセッサ上で試行される工程を操作することができる。また、複数の仮想マシンも1つのプロセッサと共に用いることができる。

【0022】

使用者は、使用者インターフェース124又は他の任意のインターフェースを表示しうるコンピュータモニターなどの視覚表示デバイス122を介して、演算デバイス100と相互作用しうる。また、視覚表示デバイス122は、例示的な実施形態の他の態様又は要素、例えば、保存デバイス116のアイコンも表示しうる。演算デバイス100は、キーボード又はマルチタッチインターフェース108及びポインティングデバイス110、例えば、使用者からの入力を受け取るマウスなど、他のI/Oデバイスを備えていてもよい。。キーボード108及びポインティングデバイス110は、視覚表示デバイス122へと接続することができる。演算デバイス100は、他の従来の適切なI/O周辺機器を備えていてもよい。。

【0023】

演算デバイス100は、標準的な電話線、LAN接続又はWAN接続(例えば、802.11、T1、T3、56kb、X.25)、ブロードバンド接続(例えば、ISDN、Frame Relay、ATM)、無線接続、コントローラエリアネットワーク(CAN)又は上記のいずれかもしくは全ての何らかの組合せが含まれるがこれらに限定されない多様な接続を介する、ローカルエリアネットワーク(LAN)、ワイドエリアネットワーク(WAN)又はインターネットを介してネットワークデバイス126とインターフェースされる、ネットワークインターフェース112を備えていてもよい。。ネットワークインターフェース112は、内蔵型ネットワークアダプター、ネットワークインターフェースカード、PCMCIAネットワークカード、カードバスネットワークアダプター、無線ネットワークアダプター、USBネットワークアダプター、演算デバイス100が、

10

20

30

40

50

通信及び本明細書で記載される操作を実施することが可能な任意の種類のネットワークとインターフェースされることを可能とするのに適するモデム又は他の任意のデバイスを含みうる。

【0024】

さらに、演算デバイス100は、ワークステーション、デスクトップコンピュータ、サーバー、ラップトップ、ハンドヘルドコンピュータ又は通信が可能であり、本明細書で記載される操作を実施するのに十分なプロセッサ電源及びメモリ容量を有する他の形態の演算デバイスもしくは遠隔通信デバイスなど、任意のコンピュータシステムでありうる。

【0025】

演算デバイス100は、任意のバージョンのMicrosoft（登録商標）Windows（登録商標）オペレーティングシステム、異なるリリースのUnixオペレーティングシステム及びLinuxオペレーティングシステム、Macintoshコンピュータ用の任意のバージョンのMac OS（登録商標）、任意の組込み型オペレーティングシステム、任意のリアルタイムのオペレーティングシステム、任意のオープンソースのオペレーティングシステム、任意の特許取得オペレーティングシステム、任意のモバイル演算デバイス用のオペレーティングシステム又は演算デバイス上で動作し、本明細書で記載される操作を実施することが可能な他の任意のオペレーティングシステムなど、任意のオペレーティングシステム118を実行しうる。オペレーティングシステムは、ネイティブモードで動作することも可能であり、エミュレートモードで動作することも可能である。

【0026】

図2は、本明細書で教示される実施形態に従い、類似する細胞の予測的なクラスターセットをデータセットから同定するためのモデルを開発する方法200を例示する。方法は、例えば、保存デバイス116又はネットワークデバイス126に保存しうるデータセットを活用する。データセットは、細胞プロファイルデータを含む。細胞プロファイルデータは、個別の細胞が描出され、区画へと分割される複数の視野に関して、複数のバイオマーカーの発現を捕捉するマルチプレックス化バイオメトリック画像を含む。細胞プロファイルデータは、共通性を有する患者コホートから得られた複数の組織試料から作成される。共通性は、例えば、患者が疾患又は状態を共有することでありうる。代替的に、共通性は、例えば、患者が同じ疾患又は状態についての予備診断を共有することでありうる。データセットは、細胞プロファイルデータの、共通性と関連する視野レベルの評価又は患者レベルの評価を含むメタ情報の少なくとも一部分との関連をさらに含む。患者レベルの評価は、例えば、手術後の生存期間でありうる。

【0027】

220では、類似する細胞の複数のクラスターセットが、データセットから作成される。一部の実施形態では、プロセッサ102、102'など、1つ以上のプロセッサが、複数のクラスターセットを作成する。作成された複数のクラスターセットの各々は、ユニークな数のクラスターを含む。各細胞を、複数のクラスターセットの各々の中の単一のクラスターに割り当てる。複数のクラスターセットの各々の中の複数のクラスターの各々は、複数の選択された属性であって、そのセットの中の他のクラスターの中の細胞の複数の選択された属性より、そのクラスターの中の他の細胞の複数の選択された属性への類似性が大きい属性を有する細胞を含む。

【0028】

細胞の類似性は、複数のバイオマーカーの少なくとも1つの発現に基づく、細胞の少なくとも1つの属性の比較から少なくとも部分的に決定される。方法200の一部の実施形態においてクラスターの作成に用いられる細胞属性は、膜の中央値強度と細胞質の中央値強度との合計の半分を、細胞核による複数のバイオマーカーの少なくとも1つの発現の中央値強度から減じることにより定義される核の強度比である。方法200の一部の実施形態においてクラスターの作成に用いられる細胞属性は、核の中央値強度と細胞質の中央値強度との合計の半分を、細胞膜による複数のバイオマーカーの少なくとも1つの発現の中央値強度から減じることにより定義される膜の強度比である。方法200の一部の実施形

態においてクラスターの作成に用いられる細胞属性は、膜の中央値強度と核の中央値強度との合計の半分を、細胞質による複数のバイオマーカーの少なくとも1つの発現の中央値強度から減じることにより定義される細胞質の強度比である。方法200の一部の実施形態においてクラスターの作成に用いられる細胞属性は、細胞全体の中央値強度である。例えば、複数のバイオマーカーの各々についての核の強度比は、クラスターセットを作成するためのベースでありうる。

【0029】

方法200の一部の実施形態では、細胞の類似性を、複数のバイオマーカーの少なくとも1つの発現に基づく、細胞の2つの属性の比較から少なくとも部分的に決定する。例えば、複数のバイオマーカーの少なくとも1つについての核の強度比及び膜の強度比は、クラスターセットを作成するためのベースでありうる。方法200の一部の実施形態では、細胞の類似性を、複数のバイオマーカーの少なくとも1つの発現に基づく、細胞の3つの属性の比較から少なくとも部分的に決定する。例えば、複数のバイオマーカーの少なくとも1つについての核の強度比、膜の強度比及び細胞質の強度比は、クラスターセットを作成するためのベースでありうる。方法200の一部の実施形態では、細胞の類似性を、複数のバイオマーカーの少なくとも1つの発現に基づく、細胞の4つの属性の比較から少なくとも部分的に決定する。例えば、複数のバイオマーカーの少なくとも1つについての核の強度比、膜の強度比及び細胞質の強度比は、クラスターセットを作成するためのベースでありうる。方法200の実施形態では、細胞の類似性を、他の属性の組合せから決定する。方法200の一部の実施形態では、細胞の類似性を、複数のバイオマーカーの少なくとも1つの発現に基づく、細胞の4つを超える属性の比較から決定する。

【0030】

方法200の一部の実施形態では、中央値によるクラスター化アルゴリズムを、細胞属性の関与性のセットに適用することにより、細胞の類似性のクラスターを作成する。方法200の他の実施形態では、平均によるクラスター化アルゴリズムを、細胞属性の関与性のセットに適用することにより、細胞の類似性のクラスターを作成する。一部の実施形態では、解析コード150が、クラスター化アルゴリズムを含む。

【0031】

一部の実施形態における複数のクラスターセットは、正規化されたデータセットから作成される。一部の実施形態では、測定値を正規化して、所与の研究における所与のバイオマーカーと関連する全ての測定値の平均及び標準偏差を決定し、この平均値を各測定値から減じ、次いで、得られる差を標準偏差で除することができる。一部の実施形態では、測定値を画像におけるバイオマーカーの発現強度の対数スケールで表す。これらの実施形態において対数スケールで表される測定値の減算は、元の生の測定スケールにおける除算に対応する。他の実施形態では、測定値を正規化して、測定バッチの中の全ての細胞について、細胞全体による発現の中央値強度を決定し、この中央値をバッチの中の各測定値から減じることができる。このような中央値強度は、特定のバイオマーカーの発現に適用することができる。この正規化又は標準化された値は、データベース内に保存することもでき、データベースにおけるデータセット処理の一部として作成することもできる。

【0032】

一部の実施形態における複数のクラスターセットは、フィルターをかけたデータセットから作成される。このようなフィルタリングは、品質管理措置として行うことができる。このようなフィルタリングにより、例えば、マルチプレックス化画像における閾値より少ない数のピクセルにより表される少なくとも1つの区画を含む細胞と関連する細胞プロファイルデータを除外することができる。フィルタリングはまた、品質管理を超えた理由でも行うことができる。このようなフィルタリングにより、例えば、正常細胞と関連する細胞プロファイルデータを、類似する細胞の複数のクラスターセットを作成するのに用いられるデータセットから除外することができる。

【0033】

230では、複数のクラスターセットの各々の中で各クラスターに割り当てられた細胞

10

20

30

40

50

の比率が観察される。240では、共通性と関連する視野レベルの評価又は患者レベルの評価を含むメタ情報の少なくとも一部分との関連について、観察された比率が検証される。観察された比率と視野レベルの評価又は患者レベルの評価との関連は、転帰としての評価を伴う分類モデルと、予測子としての観察されたクラスターの比率とをフィッティングさせることにより導出することができる。Random Forest、ニューラルネットワーク及びロジスティック回帰を含め、複数の分類解析フレームワークが存在する。例えば、一部の実施形態では、組織の悪性度と、所与の細胞クラスターから観察された細胞の存在及び数との関連が、組織の悪性度を転帰として伴うRandom Forestによる分類モデルと、予測子としての観察されたクラスターの比率とをフィッティングさせることにより導出される。他の実施形態では、組織の悪性度と、所与の細胞クラスターから観察された細胞の存在及び数との関連が、組織の悪性度を転帰として伴うニューラルネットワークによる分類モデルと、予測子としての観察されたクラスターの比率とをフィッティングさせることにより導出される。方法200の一部の実施形態は、選択されたクラスターセットにおいて観察された比率を、評価との一変量の関連について検証することをさらに含む。方法200の他の実施形態は、選択されたクラスターセットにおいて観察された比率を、評価との変量の関連について検証することをさらに含む。

10

【0034】

方法200の一部の実施形態では、観察された細胞の比率が、各クラスターに割り当てられた各視野の細胞の観察された比率である。これらの実施形態では、観察された比率が、共通性と関連する視野レベルの評価との関連について検証され、複数のクラスターセットに基づき、視野レベルの評価モデルの性能を比較することを介して、予測的なクラスターセットが選択される。

20

【0035】

方法200の一部の実施形態では、観察された細胞の比率が、各クラスターに割り当てられた各患者の細胞の観察された比率である。これらの実施形態では、観察された比率が、状態又は疾患の予後診断との関連について検証され、複数のクラスターセットに基づき、患者レベルの評価モデルの性能を比較することを介して、複数のクラスターセットが選択される。

【0036】

一部の実施形態では、評価が群分けされる。例えば、前立腺がん患者のコホートでは、2又は3のグリソンスコアを結果としてもたらす評価を併せて群分けすることができる。これらの実施形態では、複数のクラスターセットが、群分けされた、患者コホートの共通性と関連する評価との関連について検証される。例えば、属性の組合せを、グリソンスコアが2又は3である試料が併せて群分けされる低度のグリソンスコアとの関連について検証することができる。他の種類のがんのコホートについての視野レベルの評価は、それらに固有の関与性の腫瘍悪性度を有する他の種類の腫瘍についての評価を伴いうる。他のがんの評価システムには、例えば、乳がんについてのブルーム-リチャードソンシステム及び腎臓がんについてのフルマンシステムが含まれる。一部の実施形態では、がん又は他の疾患が、2つを超える悪性度又は類型の中に収まりうる評価を示す場合はいつでも、類似する悪性度又は類型を群分けすることができる。

30

40

【0037】

250では、複数のクラスターセットの1つが、複数のクラスターセットについての少なくとも1つのモデルの性能の比較に基づき選択される。一部の実施形態では、視覚表示デバイス122が、選択のなされることを可能とする。複数のクラスターセットの各々について、類似する分類モデルを創出することができる。一部の実施形態では、プロセッサ102、102'など、1つ以上のプロセッサが、分類モデルを創出する。各モデルは、対応するクラスターセットの中の細胞クラスターの比率に基づく評価を予測する。一部の実施形態では、例えば、各モデルは、対応するクラスターセットの中の細胞クラスターの比率に基づき、組織の悪性度を予測する。各クラスターセットについてのモデルの性能は、予測的な性能についての多様なメトリックにより、モデルを開発するのに用いられ

50

ないテストデータセットにおいて評価することができる。モデルに基づきクラスターセットを比較するのに用いる性能メトリックには、感度、特異度、ROC (receiver operating characteristic) 曲線下面積 (また、一致率とも称する) が含まれる。次いで、用いられるクラスターセットを、モデルの性能のメトリックの1つ以上に基づき選択することができる。例えば、一部の実施形態では、最高的一致率と関連するクラスターセットが選択される。他の実施形態では、データが見かけ上オーバーフィッティングであるために、最高的一致率と関連するクラスターセットが選択されない。選択されたセットは、予測的なクラスターセットを含む。方法200の一部の実施形態は、複数のクラスターセットの各々の中のクラスターの数に関して少なくとも1つのモデルの性能を比較することをさらに含む。

10

【0038】

方法200の一部の実施形態は、ある数のクラスターを有するクラスターセットであって、それを下回ってクラスターセットの中のクラスターの数が小さくなると、性能の減少をもたらされるクラスターセットを選択することをさらに含む。方法200の一部の実施形態は、ある数のクラスターを有するクラスターセットであって、それを上回ってクラスターセットの中のクラスターの数が大きくなると、統計学的に有意な性能の増大をもたらされないクラスターセットを選択することをさらに含む。方法200の一部の実施形態は、例えば、0.85以上的一致率でありうる、あらかじめ規定された閾値を超える性能メトリックに対応するクラスターセットについての少なくとも1つのモデルの性能に基づき、クラスターセットを選択することをさらに含む。方法200の一部の実施形態は、少なくとも1つの予測的なクラスターを、予測的なクラスターセットから同定することをさらに含む。

20

【0039】

方法200の一部の実施形態は、細胞データをトレーニングデータ及びテストデータへと分割し、類似する細胞の複数のクラスターセットをトレーニングデータから作成し、少なくとも1つのモデルの性能をテストデータから決定する。

【0040】

図3は、本明細書で教示される実施形態に従い、細胞クラスターの特徴を表示する例示的な方法300を示す。この方法は、例えば、保存デバイス116又はネットワークデバイス126に保存しうるデータセットを活用する。データセットは、細胞プロファイルデータを含む。細胞プロファイルデータは、個別の細胞が描出され、区画へと分割される複数の視野に関して、複数のバイオマーカーの発現を捕捉するマルチプレックス化バイオメトリック画像を含む。

30

【0041】

320では、類似する細胞の複数のクラスターの中の第1のクラスターが、データセットから同定される。各細胞は、複数のクラスターの1つに割り当てられる。複数のクラスターの中の各クラスターが、複数の選択された属性であって、そのセットの中の他のクラスターの中の細胞の複数の選択された属性より、そのクラスターの中の他の細胞の複数の選択された属性への類似性が大きい属性を有する細胞を含む。220に関して上記で論じた技術のいずれかにより、細胞の類似性を判定することができ、クラスター化を行うことができる。

40

【0042】

330では、第1のクラスターの中の第1の細胞のモンタージュを創出する。一部の実施形態では、プロセッサ102、102'など、1つ以上のプロセッサが、モンタージュを創出する。モンタージュは、第1の細胞による複数のバイオマーカーの各々の発現について描述する、少なくともいくつかのマルチプレックス化画像の部分を含む。少なくともいくつかの画像の各部分は、第1の細胞及び第1の細胞周囲の対象の小領域を含む。

【0043】

340では、使用者が第1のクラスターの特徴を理解することを可能とするために、第1のクラスターの中の第1の細胞のモンタージュを表示する。一部の実施形態では、モン

50

ターゲットを、視覚表示デバイス 122 上に表示する。方法 300 の他の実施形態で表示される第 1 の細胞のモニタージュは、第 1 の細胞による複数のバイオマーカーの各々の発現について描述する視野についての、少なくともいくつかの画像の一連の並置部分を含む。方法 300 の一部の実施形態で表示される第 1 の細胞のモニタージュは、第 1 の細胞による複数のバイオマーカーの各々の発現について描述する視野についての、少なくともいくつかの画像の一連のスーパーインポーズ部分を含む。

【0044】

方法 300 の一部の実施形態は、第 1 のクラスターの中の第 2 の細胞のモニタージュを創出し、表示することをさらに含む。第 2 の細胞のモニタージュは、第 2 の細胞による複数のバイオマーカーの各々の発現について描述する視野についての、少なくともいくつかの画像の部分を含む。少なくともいくつかの画像の各部分は、第 2 の細胞及び第 2 の細胞周囲の対象の小領域を含む。図 35 は、本明細書で教示される実施形態に従い、2 つの細胞の例示的なモニタージュを示す。とりわけ、図 35 は、いずれもが 20 クラスターのセットによるクラスター 15 の中の 2 つの細胞のモニタージュであって、左側の細胞が正常視野 (GLO) から採取されるのに対し、右側の細胞はグリソン悪性度 3 の視野 (GL3) から採取されるモニタージュを例示する。方法 300 の一部のこのような実施形態は、使用者が第 1 のクラスターの特徴を理解することを可能とするために、第 1 のクラスターの中の第 1 の細胞のモニタージュと、第 1 のクラスターの中の第 2 の細胞のモニタージュとを同時に表示することをさらに含む。同様に、第 1 のクラスターの中のさらなる細胞のモニタージュを創出し、表示することもできる。

【0045】

図 4 は、本明細書で教示される実施形態に従い、モデル化されたクラスターセットを、新たな細胞プロファイルデータに適用する方法 400 を例示する。モデル化されたクラスターセットは、例えば、保存デバイス 116 又はネットワークデバイス 126 に保存することができる。モデル化されたクラスターセットは、例えば、本明細書で教示される方法 200 の任意の実施形態を介して展開することができる。

【0046】

方法 400 は、患者に由来する少なくとも 1 つの組織試料についての少なくとも 1 つの視野に関する細胞プロファイルデータを伴う。細胞プロファイルデータは、複数のバイオマーカーの発現を捕捉するマルチプレックス化バイオメトリック画像を含む。視野の中の個別の細胞は、描出され、区画へと分割される。得られる情報はまた、細胞プロファイルデータ内にも包含される。この方法では、細胞プロファイルデータを、例えば、保存デバイス 116 又はネットワークデバイス 126 に保存することができる。

【0047】

方法 400 の一部の実施形態は、患者から少なくとも 1 つの組織試料を得ることをさらに含む。方法 400 の一部の実施形態は、患者に由来する少なくとも 1 つの組織試料を染色し、イメージングすることをさらに含む。方法 400 の一部の実施形態は、複数のバイオマーカーの各々の発現を捕捉するマルチプレックス化画像に基づき、患者に由来する少なくとも 1 つの組織試料の個別の細胞を描述することをさらに含む。方法 400 の一部の実施形態は、複数のバイオマーカーの各々の発現を捕捉するマルチプレックス化画像に基づき、患者に由来する少なくとも 1 つの組織試料の個別の細胞を区画へと分割することをさらに含む。

【0048】

420 では、少なくとも 1 つの組織試料の視野の中の細胞の各々が、選択されたクラスターセットの中の類似する細胞の複数のクラスターの中の単一のクラスターに割り当てられる。一部の実施形態では、プロセッサ 102、102' など、1 つ以上のプロセッサが、細胞を適切なクラスターに割り当てる。選択されたクラスターセットの中の各クラスターは、複数の選択された属性であって、そのセットの中の他のクラスターの中の細胞の複数の選択された属性より、そのクラスターの中の他の細胞の複数の選択された属性への類似性が大きい属性を有する細胞を含む。220 に関して上記で論じた技術のいずれか

により、細胞の類似性を判定することができ、クラスター化を行うことができる。一部の実施形態では、解析コード150が、クラスター化アルゴリズムを含む。方法200に関して上記で論じた技術のいずれかにより、クラスターセットを選択することができた。

【0049】

430では、選択されたクラスターセットにおいて各クラスターに割り当てられた細胞の比率が観察される。方法400の一部の実施形態では、観察された細胞の比率が、各クラスターに割り当てられた各視野の細胞の観察された比率である。方法400の一部の実施形態では、観察された細胞の比率が、各クラスターに割り当てられた各患者の細胞の観察された比率である。

【0050】

440では、観察された比率が、状態もしくは疾患の診断、予後診断又は状態もしくは疾患の処置に対する応答との関連について検証される。関連は、選択されたクラスターセットの、視野レベルの評価又は患者レベルの評価を含むメタ情報の少なくとも一部分との公知の関連に由来しうる。関連は、例えば、方法200の実施形態に従う解析を介して知ることができる。一部の実施形態では、関連が、グリソンの組織悪性度との関連である。一部の実施形態では、関連が、疾患又は状態による生存期間との関連である。

【0051】

方法400の一部の実施形態は、観察された比率を、選択されたクラスターセットにおいて、選択されたクラスターセットの公知の一変量の関連に由来しうる一変量の関連について検証することをさらに含む。方法400の他の実施形態は、観察された比率を、選択されたクラスターセットにおいて、選択されたクラスターセットの公知の多変量の関連に由来しうる多変量の関連について検証することをさらに含む。

【0052】

図5は、本明細書で教示される実施形態に従い、細胞特徴の積率の予測的なセットを、データセットから同定するモデルを開発する方法500を例示する。この方法は、例えば、保存デバイス116又はネットワークデバイス126に保存しうるデータセットを活用する。データセットは、細胞プロファイルデータを含む。細胞プロファイルデータは、個別の細胞が描出され、区画へと分割される複数の視野に関して、複数のバイオマーカーの発現を捕捉するマルチプレックス化バイオメトリック画像を含む。細胞プロファイルデータは、共通性を有する患者コホートから得られた複数の組織試料から作成される。共通性は、例えば、患者が疾患又は状態を共有することでありうる。代替的に、共通性は、例えば、患者が同じ疾患又は状態についての予備診断を共有することでありうる。データセットは、細胞プロファイルデータの、共通性と関連する視野レベルの評価又は患者レベルの評価を含むメタ情報の少なくとも一部分との関連をさらに含む。患者レベルの評価は、例えば、手術後の生存期間でありうる。

【0053】

520では、少なくとも1つの細胞特徴が、細胞による複数のバイオマーカーの各々の発現に基づき計算される。少なくとも1つの細胞特徴が計算される前に、細胞プロファイルデータを正規化することができる。一部の実施形態では、測定値を標準化して、所与の研究における所与のバイオマーカーと関連する全ての測定値の平均及び標準偏差を決定し、この平均値を各測定値から減じ、次いで、得られる差を標準偏差で除することができる。一部の実施形態では、測定値が、画像におけるバイオマーカーの発現強度の対数スケールで表される。これらの実施形態において対数スケールで表される測定値の減算は、元の生の測定スケールにおける除算に対応しうる。他の実施形態では、測定値を標準化して、測定バッチの中の全ての細胞について、細胞全体による発現の中央値強度を決定し、この中央値をバッチの中の各測定値から減じることができる。このような中央値強度は、特定のバイオマーカーの発現に適用することができる。この正規化又は標準化された値は、データベース内に保存することもでき、データベースにおけるデータセット処理の一部として作成することもできる。

【0054】

少なくとも1つの細胞特徴が計算される前に、一部の実施形態では、さらなる計算に由来する細胞プロファイルデータのサブセットにフィルターをかける。このようなフィルタリングは、品質管理措置として行うことができる。このようなフィルタリングにより、マルチプレックス化画像における閾値より少ない数のピクセルにより表される少なくとも1つの区画を含む細胞と関連する細胞プロファイルデータを除外することができる。フィルタリングはまた、品質管理を超えた理由でも行うことができる。このようなフィルタリングにより、複数の形態バイオマーカーの各々の発現を、さらなる計算から除外することができる。したがって、本明細書で教示される一部の実施形態では、少なくとも1つの細胞特徴を計算することが、細胞による複数の非形態バイオマーカーの各々の発現に基づき、少なくとも1つの細胞特徴を計算することを伴う。

10

【0055】

方法500の一部の実施形態は、細胞による複数の非形態バイオマーカーの各々の発現に基づき、2つ、3つ、4つ以上の細胞特徴を計算することを伴う。一部の実施形態では、プロセッサ102、102'など、1つ以上のプロセッサが、細胞特徴を計算する。一部の実施形態では、解析コード150が、各細胞特徴についての定義を含む。一部の実施形態における細胞特徴は、膜の中央値強度と細胞質の中央値強度との合計の半分を、細胞核による複数のバイオマーカーの少なくとも1つの発現の中央値強度から減じることにより定義される核の強度比を含む。一部の実施形態における細胞特徴は、核の中央値強度と細胞質の中央値強度との合計の半分を、細胞膜による複数のバイオマーカーの少なくとも1つの発現の中央値強度から減じることにより定義される膜の強度比を含む。一部の実施形態における細胞特徴は、膜の中央値強度と核の中央値強度との合計の半分を、細胞質による複数のバイオマーカーの少なくとも1つの発現の中央値強度から減じることにより定義される細胞質の強度比を含む。

20

【0056】

530では、第1の積率が、細胞特徴の各々に由来する複数の視野の各々について計算される。一部の実施形態では、プロセッサ102、102'など、1つ以上のプロセッサが、細胞特徴の第1の積率を計算する。本明細書で教示される実施形態は、第2の積率及び/又は第3の積率を、細胞特徴の各々に由来する複数の視野の各々について計算することをさらに伴いうる。

【0057】

30

540では、属性の複数の組合せが、共通性と関連する視野レベルの評価又は患者レベルの評価を含むメタ情報の少なくとも一部分との関連について検証される。属性の複数の組合せは、少なくとも計算された第1の積率を含む。視野において観察された全てのバイオマーカーの第1の積率と、視野レベルの評価又は患者レベルの評価との関連は、転帰としての評価を伴う分類モデルと、予測子としてのバイオマーカーの第1の積率とをフィッティングさせることにより導出することができる。Random Forest、ニューラルネットワーク及びロジスティック回帰を含め、複数の分類解析フレームワークが存在する。例えば、一部の実施形態では、組織の悪性度と、視野において観察された全てのバイオマーカーの第1の積率との関連が、組織の悪性度を転帰として伴うRandom Forestによる分類モデルと、予測子としてのバイオマーカーの第1の積率とをフィッティングさせることにより導出される。他の実施形態では、組織の悪性度と、視野において観察された全てのバイオマーカーの第1の積率との関連が、組織の悪性度を転帰として伴うニューラルネットワークによる分類モデルと、予測子としてのバイオマーカーの第1の積率とをフィッティングさせることにより導出される。一部の実施形態では、関連が、特定のグリソン悪性度など、試料の視野レベルの評価との関連である。他の実施形態では、関連が、疾患又は状態による生存期間など、患者レベルの評価との関連である。

40

【0058】

一部の実施形態では、プロセッサ102、102'など、1つ以上のプロセッサが、組合せを検証する。第2の積率を計算することを伴う実施形態では、540における検証することが、計算された第1の積率及び第2の積率を含む属性の複数の組合せを、共通

50

性と関連する視野レベルの評価又は患者レベルの評価を含むメタ情報の少なくとも一部分との関連について検証することを伴う。第3の積率を計算することを伴う実施形態では、540における検証することが、計算された第1の積率及び第3の積率を含む属性の複数の組合せを、共通性と関連する視野レベルの評価又は患者レベルの評価を含むメタ情報の少なくとも一部分との関連について検証することを伴う。一部の実施形態は、計算された第1の積率、第2の積率及び第3の積率を検証することをさらに伴う。

【0059】

一部の実施形態では、540における検証することが、共通性と関連する視野レベルの評価又は患者レベルの評価を含むメタ情報の少なくとも一部分との一変量の関連について計算された積率を検証することを伴う。一部の実施形態では、540における検証することが、共通性と関連する視野レベルの評価又は患者レベルの評価を含むメタ情報の少なくとも一部分との多変量の関連について計算された積率を検証することを伴う。第2の積率及び/又は第3の積率が計算される方法500の実施形態では、計算された積率により、共通性と関連する視野レベルの評価又は患者レベルの評価を含むメタ情報の少なくとも一部分との一変量又は多変量の関連について検証することができる。

【0060】

一部の実施形態では、視野レベルの評価が群分けされる。前立腺がん患者のコホートでは、例えば、2又は3のグリソンスコアを結果としてもたらず評価を併せて群分けすることができる。これらの実施形態では、属性の複数の組合せを、群分けされた、患者コホートの共通性と関連する視野レベルの評価との関連について検証する。例えば、属性の組合せを、グリソンスコアが2又は3である試料が併せて群分けされる低度のグリソンスコアとの関連について検証することができる。他の種類のがんのコホートについての視野レベルの評価は、それらの固有の関与性の腫瘍悪性度を有する他の種類の腫瘍についての評価を伴いうる。他のがんの評価システムには、例えば、乳がんについてのブルーム-リチャードソンシステム及び腎臓がんについてのフルマンシステムが含まれる。一部の実施形態では、がん又は他の疾患が、2つを超える悪性度又は類型の中に収まりうる評価を示す場合はいつでも、類似する悪性度又は類型を群分けすることができる。

【0061】

550では、属性の複数の組合せの1つが、属性の複数の組合せについての少なくとも1つのモデルの性能の比較に基づき選択される。一部の実施形態では、視覚表示デバイス122が、選択のなされることを可能とする。属性の複数の組合せの各々について、類似する分類モデルを創出することができる。一部の実施形態では、プロセッサ102、102'など、1つ以上のプロセッサが、分類モデルを創出する。各モデルは、対応する属性の組合せに基づく評価を予測する。一部の実施形態では、例えば、各モデルは、対応する属性のセットに基づき、組織の悪性度を予測する。各属性の組合せについてのモデルの性能は、予測的な性能についての多様なメトリックにより、モデルを開発するのに用いられないテストデータセットにおいて評価することができる。モデルに基づき属性の組合せを比較するのに用いる性能メトリックには、感度、特異度、ROC (receiver operating characteristic) 曲線下面積 (また、一致率とも称する) が含まれる。次いで、用いられる属性の組合せを、モデルの性能メトリックの1つ以上に基づき選択することができる。例えば、一部の実施形態では、最高的一致率と関連する属性の組合せが選択される。他の実施形態では、データが見かけ上オーバーフィッティングであるために、最高的一致率と関連する属性の組合せが選択されない。例えば、一部の実施形態は、例えば0.85以上的一致率でありうる、あらかじめ規定された閾値を超える性能メトリックに対応する属性の組合せについての少なくとも1つのモデルの性能に基づき、属性の組合せを選択することを伴う。他の実施形態は、他の組合せについてのモデルの性能と比較したその組合せについてのモデルの性能に基づき、組合せを選択することを伴いうる。選択された属性の組合せは、予測的な属性の組合せを含む。方法500の実施形態は、積率モデルに由来する少なくとも1つの予測的な非形態的マーカーを同定することをさらに含んでいてもよい。

10

20

30

40

50

【 0 0 6 2 】

図 6 は、本明細書で教示される実施形態に従い、積率のモデルセットを新たな細胞プロファイルデータに適用する方法 6 0 0 を例示する。積率のモデルセットは、例えば、保存デバイス 1 1 6 又はネットワークデバイス 1 2 6 に保存することができる。積率のモデルセットは、例えば、本明細書で教示される方法 5 0 0 の任意の実施形態を介して開発することができる。

【 0 0 6 3 】

方法 6 0 0 は、患者に由来する少なくとも 1 つの組織試料についての少なくとも 1 つの視野に関する細胞プロファイルデータを伴う。細胞プロファイルデータは、複数のバイオマーカーの発現を捕捉するマルチプレックス化バイオメトリック画像を含む。視野の中の個別の細胞は、描出され、区画へと分割される。得られる情報はまた、細胞プロファイルデータ内にも包含される。細胞プロファイルデータは、例えば、保存デバイス 1 1 6 又はネットワークデバイス 1 2 6 に保存することができる。

10

【 0 0 6 4 】

方法 6 0 0 の一部の実施形態は、患者から少なくとも 1 つの組織試料を得ることをさらに含む。方法 6 0 0 の一部の実施形態は、患者に由来する少なくとも 1 つの組織試料を染色し、イメージングすることをさらに含む。方法 6 0 0 の一部の実施形態は、複数のバイオマーカーの各々の発現を捕捉するマルチプレックス化画像に基づき、患者に由来する少なくとも 1 つの組織試料の個別の細胞を描述することをさらに含む。方法 6 0 0 の一部の実施形態は、複数のバイオマーカーの各々の発現を捕捉するマルチプレックス化画像に基づき、患者に由来する少なくとも 1 つの組織試料の個別の細胞を区画へと分割することをさらに含む。

20

【 0 0 6 5 】

6 2 0 では、少なくとも 1 つの細胞特徴が、細胞による複数のバイオマーカーの各々の発現に基づき計算される。一部の実施形態では、プロセッサ 1 0 2、1 0 2' など、1 つ以上のプロセッサが、少なくとも 1 つの細胞特徴を計算する。一部の実施形態では、解析コード 1 5 0 が、各細胞特徴についての定義を含む。細胞特徴は、方法 5 0 0 に関して論じた任意の細胞特徴でありうる。方法 6 0 0 の一部の実施形態は、方法 5 0 0 に関して論じた細胞特徴の任意の組合せを備えていてもよい。複数の細胞特徴を計算することをさらに含む。細胞特徴は、細胞による非形態的バイオマーカーの発現から計算することができる。

30

【 0 0 6 6 】

6 3 0 では、第 1 の積率が、各視野の各細胞特徴について計算される。一部の実施形態では、プロセッサ 1 0 2、1 0 2' など、1 つ以上のプロセッサが、細胞特徴の第 1 の積率を計算する。方法 5 0 0 と同様に、方法 6 0 0 は、各細胞特徴について第 2 の及び/又は第 3 の積率を計算することをさらに含んでもよい。

【 0 0 6 7 】

6 4 0 では、計算された第 1 の積率が、状態もしくは疾患の診断、予後診断又は状態もしくは疾患の処置に対する応答との関連について検証される。関連は、例えば、方法 5 0 0 に関して記載されたデータセットなどの既存のデータセットに基づく積率のモデルセットから知ることができる。一部の実施形態では、関連が、特定のグリソン悪性度などの細胞悪性度との関連である。他の実施形態では、関連が、疾患又は状態による生存期間との関連である。

40

【 0 0 6 8 】

第 2 の積率を計算することを伴う方法 6 0 0 の実施形態では、6 4 0 における検証することが、計算された第 1 の積率及び第 2 の積率を、状態もしくは疾患の診断、予後診断又は状態もしくは疾患の処置に対する応答との関連について検証することを伴う。第 3 の積率を計算することを伴う実施形態では、6 4 0 における検証することが、計算された第 1 の積率及び第 3 の積率を、状態もしくは疾患の診断、予後診断又は状態もしくは疾患の処置に対する応答との関連について検証することを伴う。一部の実施形態は、計算された第

50

1の積率、第2の積率及び第3の積率を検証することをさらに伴う。

【0069】

一部の実施形態では、プロセッサ102、102'など、1つ以上のプロセッサが、計算された第1の積率を検証する。方法600の一部の実施形態では、640における検証することが、計算された第1の積率を、状態もしくは疾患の診断、予後診断又は状態もしくは疾患の処置に対する応答との一変量の関連について検証することを伴う。方法600の他の実施形態では、640における検証することが、計算された第1の積率を、状態もしくは疾患の診断、予後診断又は状態もしくは疾患の処置に対する応答との多変量の関連について検証することを伴う。第2の積率及び/又は第3の積率が計算される方法600の実施形態では、計算された積率を、状態もしくは疾患の診断、予後診断又は状態もしくは疾患の処置に対する応答との一変量の関連又は多変量の関連について検証することができる。

10

例示的な解析及び視覚化

データセット

がんのために前立腺の手術を受けた患者コホートに由来する組織試料に由来する情報を用いて、本明細書で教示される例示的な方法に従う解析を実施した。組織試料は、組織培養物として定義することができ、*in vivo*試料を備えていてもよい。。80例の患者に由来する前立腺の組織試料が、解析に利用可能であった。寄与集団のうち、62例が前立腺がんを有した。これら62例の前立腺がん患者のうち、11例が追跡時になおも存命であり、22例が前立腺がんにより死亡しており、残りの29例は他の原因で死亡していた。表1は、本発明者らのデータへの寄与集団の年齢、生存期間及び病理学者により導出されたグリソンスコアについての集団統計を示す。

20

【0070】

【表1】

表1：研究集団についての統計

| | 全例 (n=80) | 前立腺がん例 (n=62) | 前立腺がんによる死亡例 (n=22) |
|---------|------------|---------------|--------------------|
| 年齢 | 70.9(10.2) | 72.1(10.1) | 76.2(11.9) |
| 生存期間 | 8.76(6.49) | 7.64(6.35) | 3.73(3.44) |
| グリソンスコア | 26(32%) | 10(16%) | 1(5%) |
| 2-4 | 4(5%) | 4(6%) | 0 |
| 5-6 | 13(16%)(?) | 12(19%) | 1(5%) |
| 7 | 10(12%) | 10(16%) | 4(18%) |
| 8-10 | 20(25%) | 20(32%) | 13(59%) |
| 除外例 | 7(9%) | 6(10%) | 3(14%) |

30

本発明の他の実施形態は、異なる共通性を共有する患者コホートに由来する組織試料を伴う。例えば、一実施形態は、患者が乳がんなど、別の形態のがんを有するか否かを決定するために患者コホートから採取された組織試料を伴いうる。別の実施形態は、患者がパーキンソン病など、別の疾患を有するか否かを決定するために患者コホートから採取された組織試料を伴いうる。同様に、本発明の他の実施形態は、より大きい患者コホート又はより小さい患者コホートも伴う。

40

【0071】

組織試料は、蛍光ベースのマルチプレックス化免疫組織化学を用いて処理した。解析では14のバイオマーカーを用いた。14のうち5つのバイオマーカー：NaKATPアーゼ、PCAD、DAP I、S6及びケラチンを、個別の細胞を分割及び区画化するのに用いた。残りのマーカーは、AR、p m T O R、P I 3 K p 1 1 0 a、P I 3 K p 8 5 a、ベータカテニン、E G F R、切断型カスパーゼ3、p G S K 3 a及び切断型P A R Pであった。全てのバイオマーカーは、定性的な染色品質の点検に合格した。

50

【 0 0 7 2 】

本発明の他の実施形態は、異なるバイオマーカーを伴う。同様に、本発明の他の実施形態は、より多くのバイオマーカー又はより少ないバイオマーカーも伴う。

【 0 0 7 3 】

自己蛍光の除去、照度補正及び細胞分割の後、データは、全ての対象で各視野に分割された各細胞の3つの区画における各タンパク質画像の中央値強度を包含した。細胞は、以下のフィルターを適用することにより品質管理した。

- 1．細胞が、バックグラウンド（登録ミスのために不完全なマーカーデータを伴う画像の境界領域）とオーバーラップしないこと
- 2．細胞の核の分割が2つ以下であること
- 3．細胞核が少なくとも50ピクセルを含有すること
- 4．細胞質が少なくとも50ピクセルを含有すること
- 5．細胞膜が少なくとも50ピクセルを含有すること。

10

【 0 0 7 4 】

本発明の他の実施形態は、異なる品質管理特徴を伴う。同様に、本発明の他の実施形態は、より多くの品質管理特徴又はより少ない品質管理特徴も伴う。

【 0 0 7 5 】

イメージング、分割及び品質管理の後に、54例の患者対象が残った。患者1例当たりの視野数は、6～90の範囲であった。54例の対象においてイメージングされたのべ1757の視野のうち、1349の視野が、解析に十分な組織を含有した。これらの1349の視野の各々は、解析チームの病理学者（QL）により成功裏に評価された。

20

【 0 0 7 6 】

特に、グリソンスコアは、全ての視野について、解析チームの病理学者（QL）により、0～5のスケールで、手作業により記録された。グリソン悪性度2のデータがまれであるために、悪性度2の視野を、グリソン悪性度3の視野と組み合わせた。表2は、視野レベルのグリソン悪性度についての概要を示す。

【 0 0 7 7 】

【表2】

表2：FOVレベルのグリソン悪性度

30

| がんによる死亡 年齢（歳） 生存期間（年） | なし | | | | あり | | | |
|-----------------------------|-------|------|-------|------|-------|------|-------|------|
| | 48-72 | | 73-94 | | 48-72 | | 73-94 | |
| | 0-6 | 7-21 | 0-6 | 7-21 | 0-6 | 7-21 | 0-6 | 7-21 |
| グリソン悪性度の評価 | | | | | | | | |
| 0 | 64 | 304 | 99 | 29 | 7 | 18 | 63 | 36 |
| 2-3 | 32 | 54 | 36 | 10 | 9 | 3 | 13 | 9 |
| 4 | 34 | 73 | 24 | 1 | 8 | 11 | 125 | 38 |
| 5 | 11 | 3 | 3 | 0 | 0 | 6 | 120 | 20 |

本発明の他の実施形態は、異なる視野レベルの評価であって、関与性の患者コホートに影響を及ぼす疾患又は状態に適切でありうる評価を伴いうる。

【 0 0 7 8 】

40

対象試料は、5つのバッチに受容され、解析された。表3は、5つのバッチに対するグリソンスコアについての解析結果であって、表示項目が組織試料の総数である解析結果を示す。一部の対象が複数のバッチにおいて解析されたために、表3は、54例の固有の対象に由来するのべ63の組織試料を含む。9例の対象が複数の組織試料を有した：これらの対象のうちの4例が2つのバッチで解析され、2例が3つのバッチで解析され、2例は単一のバッチにおいて2回にわたり解析された。最後の対象は、4つの異なるバッチで解析された。

【 0 0 7 9 】

【表 3】

表 3：5つのバッチにおける対象レベルのグリソンスコア

| バッチ | | | | | | |
|---------|-------|-------|-------|-------|-------|----|
| グリソンスコア | バッチ 1 | バッチ 2 | バッチ 3 | バッチ 4 | バッチ 5 | 合計 |
| 0 | 1 | 0 | 1 | 4 | 4 | 10 |
| 2-4 | 3 | 0 | 0 | 0 | 1 | 4 |
| 5-6 | 4 | 4 | 3 | 2 | 1 | 14 |
| 7 | 3 | 1 | 3 | 2 | 0 | 9 |
| 8-10 | 4 | 7 | 9 | 4 | 2 | 26 |
| 合計 | 15 | 12 | 16 | 12 | 8 | 63 |

10

無病生存は、手術と死亡又は追跡との間の時間として定義された。この尺度は、対象が追跡時において存命であるか又は前立腺がん以外の原因で死んだ場合は右側打ち切りとして扱われた。患者対象のうちの18例は、追跡以前に、前立腺がん で死亡した。また、各患者対象に享受可能な術後の生存期間もデータセットに付加し、これにより、生のデータセットを補完した。

【0080】

本発明の他の実施形態は、異なる患者レベルの評価であって、関与性の患者コホートに影響を及ぼす疾患又は状態に適切でありうる評価を伴いうる。

【0081】

細胞全体及び区画の中央値強度は、各バッチの中で、バッチの中の全ての対象における全ての細胞についての全ての細胞全体の測定値の中央値を減じることにより正規化した。複数のバッチにおいて解析された8例の対象については、視野をバッチで正規化し、次いで、その後、単一のバッチにおいて解析された対象と同じ形で取り扱った。本発明の他の実施形態は、収集されたデータのより多くの正規化を伴う場合もあり、より少ない正規化を伴う場合もあり、異なる正規化を伴う場合もあり、正規化を伴わない場合もある。

20

さらなる細胞特徴

各タンパク質について個別に、4つの細胞特徴を細胞レベルのデータから計算した。各々がlog2スケールで定義される4つの特徴は、細胞全体の中央値強度、核の強度比、膜の強度比及び細胞質の強度比であった。3つの区画比は、核、膜又は細胞質の発現の中央値強度を、他の2つの区画の平均の中央値強度へと関連付ける。3つの区画比は、以下の通りに定義した。

30

【0082】

$$R_n = I_n - (I_m + I_c) / 2$$

$$R_m = I_m - (I_n + I_c) / 2$$

$$R_c = I_c - (I_m + I_n) / 2$$

[式中、 I_n 、 I_m 及び I_c は、核、膜及び細胞質のそれぞれにおけるlog2スケールによる中央値強度である]。区画の中のマーカーの発現レベル、例えば、膜におけるNaK ATPアーゼは、記載される通り、1つの区画の他の2つの区画の平均に対する比と解釈した。本発明の他の実施形態は、より多くの細胞特徴を伴う場合もあり、より少ない細胞特徴を伴う場合もあり、異なる細胞特徴を伴う場合もある。

40

【0083】

上記のデータセットは保存された。計算される任意のさらなる細胞特徴を、保存されたデータセットに付加することができる。

分類モデル及び生存モデル

2つの異なる種類の解析(積率解析及び細胞クラスター解析)を行った。次いで、各種類の解析の結果を、分類モデル及び生存モデルと個別に比較した。

【0084】

視野レベルの評価モデルのために、本発明の実施形態では、上記の特徴を伴う、L. Breiman's "Random Forests" in Machine Learning 45(1), 5-32 (2001)において記載されているRandom Forest分類子を適用した。その結果、グリソン悪性度の視野と

50

関連する2つの個別のモデルが得られた。第1のモデルは、グリソン悪性度（すなわち、2、3、4又は5）の視野を、グリソン悪性度0の視野から識別した。第2のモデルは、グリソン悪性度4又は5の視野を、グリソン悪性度2又は3の視野から識別した。第2のモデルでは、グリソン悪性度0の視野を解析から除去した。R (v. 2.11.0) 用の Random Forest パッケージ (v. 4.5-36) をデフォルトの設定で用いた。200ツリーを構築した後で、OOB (out-of-bag) の誤差率が収束したので、500ツリーを分類子に用いた。フィッティングでは、データをサンプリングし、対象により層別化して（層変数を Random Forest に用いて）、対象を多数の視野で過剰に重みづけすることを回避した。OOB (out-of-bag) 予測から予測されるクラス確率を閾値化することにより、ROC (receiver operating characteristic) 解析を行った。ROC 曲線下面積 (AUC) を推定した。変数重要度の結果は、所与の変数に由来するデータにスクランブルをかけるときの分類精度の低下に基づいた。変数依存性プロットは、予測されるクラス確率の対数値に基づいた。本発明の他の実施形態は、より多くの視野レベルの評価モデルを用いる場合もあり、より少ない視野レベルの評価モデルを用いる場合もあり、異なる視野レベルの評価モデルを用いる場合もある。

10

【0085】

生存（患者レベルの転帰）との関連のために、対象の侵襲性視野（グリソンスコア > 0）にわたるスポットレベルの特徴の平均及び対象の正常視野にわたる第2の平均を記録した。特定の種類の視野を伴わない対象のマーカーの特徴データは、集団の中央値に帰属させた。

20

【0086】

患者レベルの評価モデルのために、本発明の実施形態では、H. Ishwaran et al.'s "Random Survival Forests" in the Ann. App. Statist. 2:841-860 (2008)において開示されている Random Survival Forest モデルなどの Random Survival Forest モデルを適用した。R (v. 2.11.0) 用の Random Survival Forest パッケージ (v. 3.6.3) をデフォルト変数で用いた。5000ツリーを用いてモデルを構築した。表中記載の誤差メトリックは、1 - [Harrell の一致率指数（ランダムに選択された対象の対において、最初に死亡する対象のモデルで予測される転帰が不良であった確率）] であった。Harrell, F.E. et al. in "Evaluating the Yield of Medical Tests," J. Amer. Med. Assoc. 247:2543-2546 (1982)によれば、50%誤差はランダムモデルであり、0%が完全モデルである。本発明の他の実施形態は、より多くの患者レベルの評価モデルを用いる場合もあり、より少ない患者レベルの評価モデルを用いる場合もあり、異なる患者レベルの評価モデルを用いる場合もある。

30

【0087】

この誤差メトリックは、OOB (out-of-bag) 試料において推定した。変数重要度の結果は、ランダムな派生割付けを特徴に関するツリーノードに用いるときの、所与の特徴についての一致率誤差の増大に基づいた。偏変数依存性プロットは、集団内の対象において常に観察される所与の特徴の関数として集団内で予測される死亡率である相対死亡率に基づいた。さらに、時間の閾値を3、5及び10年後に設定し、患者が閾値以前に前立腺がんで死亡したかどうかを分類することにより、3つの個別の二元分類モデルを生存データへとフィッティングさせた。

40

積率解析

FOVレベルのグリソン悪性度との関連のために、本発明の実施形態の積率ベースの解析では、4つの細胞レベルの特徴を、視野レベルの統計にまとめた。視野の中の細胞集団に基づき、14のマーカー全てについて、4つずつの発現レベルの特徴全ての平均、標準偏差及び歪度を記録した。悪性度FOVとの関連のために、構造マーカー及び標的マーカーを含めた14のマーカー全てを、予測子として考えた。これは、14のバイオマーカーの各々に4つずつの細胞特徴の各々に3つずつの積率（合計168のFOV属性）を結果

50

としてもたらした。本発明の他の実施形態は、より多くの視野レベルの属性を用いる場合もあり、より少ない視野レベルの属性を用いる場合もあり、異なる視野レベルの属性を用いる場合もある。

【0088】

例えば、単一の細胞分割に由来する以下の細胞の形態特徴は、多様な実施形態における積率ベースのモデル内に包含されうる：Eccentricity__Cell、Solidity__Cell、MajorAxisLength__Cell、MajorAxisAngle__Cell、Perimeter__Cell、Area__Cell、Area__Nuclei、Area__Mem、and Area__Cyto。

積率解析を用いる視野評価の予測

10

視野評価モデルを構築するに当たり、FOV属性に関して3つの選択肢を考慮した。

(1) 蛍光データに基づく特徴を含むか否か、

(2) 細胞の形態データを含むか否か及び

(3) 蛍光データの何次の積率を含むのか：平均(m1)、平均及び標準偏差(m12)又は平均、標準偏差及び歪度(m123)。

【0089】

本発明の他の実施形態は、視野属性に関してより多くの選択肢を考慮する場合もあり、より少ない選択肢を考慮する場合もあり、異なる選択肢を考慮する場合もある。

【0090】

表4は、異なる積率ベースの特徴セットに基づき、がん性悪性度(グリソン2、3、4又は5)視野を、正常悪性度(グリソン0)視野と対比して比較する分類子の性能を示す。積率の次数(m1、m12又はm123)の少なくとも1つを含めた全てのFOV属性の複数の組合せについて調べた。一部の組合せには、蛍光マーカーデータが含まれ、一部には、細胞の形態特徴が含まれた。ROC曲線下面積(AUC)は、蛍光マーカーデータの少なくとも第1の積率を含む全てのモデルについて少なくとも98%であった。形態特徴はAUCをわずかに増大させるに過ぎなかった。

20

【0091】

【表4】

表4: がん視野の正常視野との対比に対する積率ベースの分類子の性能

30

| 包含される積率 | 包含される蛍光特徴 | 包含される形態特徴 | AUC |
|---------|-----------|-----------|-------|
| m12 | あり | あり | 0.983 |
| m1 | あり | あり | 0.982 |
| m123 | あり | あり | 0.982 |
| m123 | あり | なし | 0.982 |
| m12 | あり | なし | 0.981 |
| m1 | あり | なし | 0.980 |
| m12 | なし | あり | 0.896 |
| m123 | なし | あり | 0.892 |
| m1 | なし | あり | 0.845 |

40

表5は、高悪性度(グリソン4又は5)のがん性視野を、低悪性度(グリソン2又は3

50

）のがん性視野と対比して比較する分類子の性能を示す。ここでもまた、蛍光マーカーデータの少なくとも第 1 の積率を包含しないモデルでは、AUC が減少した。

【 0 0 9 2 】

【表 5】

表 5：高悪性度のがん視野の低悪性度がん視野との対比に対する積率ベースの分類子の性能

| 包含される積率 | 包含される蛍光特徴 | 包含される形態特徴 | AUC |
|---------|-----------|-----------|-------|
| m12 | あり | なし | 0.929 |
| m12 | あり | あり | 0.928 |
| m1 | あり | なし | 0.928 |
| m123 | あり | あり | 0.928 |
| m1 | あり | あり | 0.926 |
| m123 | あり | なし | 0.926 |
| m12 | なし | あり | 0.834 |
| m123 | なし | あり | 0.817 |
| m1 | なし | あり | 0.781 |

上位のモデルについての ROC 曲線を、図 7 及び 8 に示す。

【 0 0 9 3 】

上位のモデルについての変数重要度プロットを、図 9 及び 10 に示す。いずれの場合にも、上位の特徴は、NaKATPアーゼと関連し、膜の外側で定量化されるか又は FOV レベルの標準偏差が大きい。がん / 正常分類子における第 1 の形態特徴は、一覧表の 2 4 位である核の面積である。

積率解析を用いる患者レベルの評価の予測

患者レベルの評価モデルを構築するに当たり、FOV 属性に関して 4 つの選択肢を考慮した。

- (1) 蛍光データに基づく特徴を含むか否か、
- (2) 細胞の形態データを含むか否か、
- (3) 蛍光データの何次の積率を含むのか：平均 (m 1) 、平均及び標準偏差 (m 1 2) 又は平均、標準偏差及び歪度 (m 1 2 3) 及び
- (4) 患者に由来するどの視野を含むのか：侵襲性組織だけの視野、正常組織だけの視野、全ての組織の視野又は侵襲性組織の平均から正常組織の平均を減じた視野。

【 0 0 9 4 】

本発明の他の実施形態は、視野属性に関してより多くの選択肢を考慮する場合もあり、より少ない選択肢を考慮する場合もあり、異なる選択肢を考慮する場合もある。

【 0 0 9 5 】

表 6 は、全患者のデータセットへとフィッティングさせた全ての積率ベースのモデルについての性能メトリックを示す。「包含される FOV」列における「inv - norm」というコードは、対象について用いられる特徴が、それらの侵襲性視野において認められる平均からそれらの正常視野において観察される平均を減じた差であることを意味する。特定の場合には、年齢及びグリソンスコアだけを伴うモデルを 11 回にわたりフィッティ

ングさせたが、これらの行は太字で強調する。太字の11行についての異なる結果は、Random Survival Forest手順及びRandom Forest手順に固有のランダムサンプリング誤差と関連する。

【0096】

侵襲性視野にマーカーの第1の積率を伴い、形態特徴を伴わないモデルが、好ましいモデルであった。RSFによる一致率のメトリックにおいてこのモデルを凌ぐモデルも存在するが、3年後及び10年後のAUCではこのモデルの方が優れており、第1の積率及び第2の積率を含むモデルを0.8%下回るに過ぎない。さらに、このモデルは、5年後のAUCが、ヌルモデルより73%~93%大きい。ヌルモデルのRSFによる一致率を大幅に凌ぐモデルは見られない。

10

【0097】

表7は、グリソンスコアが0を超える患者を除外した患者データセットに適用された、モデルについての同じ性能メトリックを示す。侵襲性視野におけるマーカー特徴の第1の積率を含む、表7における上位のモデルは、RSFによる一致率(69% 81%)、5年後のAUC(68% 89%)及び10年後のAUC(64% 87%)において、ヌルモデルを大幅に改善している。表6と同様に、表7の太字で強調された行は、年齢及びグリソンスコアだけを包含したモデルである。

【0098】

上位2つのモデルにおける上位4つの特徴についての偏変数依存性プロットを、図11及び12に示す。

20

【0099】

【表 6】

表 6：全ての対象を包含する生存データに適用された、全ての積率ベースのモデルについての性能メトリック

| 包含される積率 | 包含される蛍光特徴 | 包含される形態特徴 | 包含されるFOV | RSFによる一致率 | 3年後におけるAUC | 5年後におけるAUC | 10年後におけるAUC |
|---------|-----------|-----------|----------|-----------|------------|------------|-------------|
| m12 | あり | なし | 侵襲性 | 0.810 | 0.891 | 0.938 | 0.827 |
| m123 | あり | なし | 侵襲性+通常 | 0.808 | 0.878 | 0.901 | 0.791 |
| m12 | あり | なし | 侵襲性+通常 | 0.807 | 0.918 | 0.916 | 0.807 |
| m1 | あり | なし | 侵襲性+通常 | 0.805 | 0.920 | 0.934 | 0.856 |
| m123 | あり | あり | 侵襲性 | 0.804 | 0.904 | 0.901 | 0.843 |
| m1 | あり | なし | 侵襲性 | 0.802 | 0.931 | 0.932 | 0.852 |
| m12 | あり | あり | 侵襲性 | 0.801 | 0.876 | 0.914 | 0.830 |
| m12 | あり | あり | 侵襲性+通常 | 0.799 | 0.887 | 0.901 | 0.807 |
| m1 | あり | あり | 侵襲性+通常 | 0.799 | 0.900 | 0.934 | 0.836 |
| m123 | あり | なし | 侵襲性 | 0.799 | 0.889 | 0.870 | 0.813 |
| m123 | あり | あり | 侵襲性+通常 | 0.798 | 0.927 | 0.883 | 0.830 |
| m1 | あり | あり | 侵襲性 | 0.793 | 0.898 | 0.927 | 0.856 |
| m1 | なし | なし | 通常 | 0.776 | 0.893 | 0.744 | 0.706 |
| m123 | なし | なし | 侵襲性+通常 | 0.769 | 0.887 | 0.721 | 0.705 |
| m1 | なし | なし | 侵襲性+通常 | 0.769 | 0.871 | 0.741 | 0.711 |
| m12 | なし | なし | 侵襲性+通常 | 0.767 | 0.887 | 0.720 | 0.692 |
| m1 | なし | あり | 侵襲性+通常 | 0.766 | 0.869 | 0.786 | 0.751 |
| m123 | なし | なし | 侵襲性+通常 | 0.764 | 0.891 | 0.737 | 0.703 |
| m123 | なし | なし | 侵襲性 | 0.764 | 0.878 | 0.749 | 0.714 |
| m1 | なし | なし | 侵襲性 | 0.764 | 0.867 | 0.734 | 0.703 |
| m12 | なし | なし | 侵襲性+通常 | 0.763 | 0.882 | 0.722 | 0.710 |
| m123 | なし | なし | 通常 | 0.763 | 0.878 | 0.741 | 0.704 |
| m1 | なし | なし | 侵襲性+通常 | 0.760 | 0.869 | 0.751 | 0.699 |
| m12 | なし | なし | 侵襲性 | 0.760 | 0.878 | 0.744 | 0.690 |
| m12 | なし | なし | 通常 | 0.755 | 0.880 | 0.729 | 0.707 |
| m1 | なし | あり | 侵襲性 | 0.749 | 0.836 | 0.766 | 0.800 |
| m12 | なし | あり | 侵襲性+通常 | 0.735 | 0.847 | 0.697 | 0.740 |
| m123 | なし | あり | 侵襲性+通常 | 0.726 | 0.824 | 0.672 | 0.729 |
| m1 | なし | あり | 通常 | 0.726 | 0.856 | 0.652 | 0.675 |
| m1 | なし | あり | 侵襲性+通常 | 0.715 | 0.811 | 0.760 | 0.759 |
| m123 | なし | あり | 侵襲性 | 0.712 | 0.884 | 0.810 | 0.746 |
| m12 | なし | あり | 侵襲性 | 0.712 | 0.833 | 0.755 | 0.772 |
| m12 | なし | あり | 通常 | 0.705 | 0.760 | 0.648 | 0.616 |
| m1 | あり | なし | 通常 | 0.703 | 0.848 | 0.791 | 0.781 |
| m123 | なし | あり | 侵襲性+通常 | 0.700 | 0.847 | 0.782 | 0.751 |
| m12 | なし | あり | 侵襲性+通常 | 0.693 | 0.824 | 0.701 | 0.721 |
| m1 | あり | あり | 侵襲性+通常 | 0.686 | 0.887 | 0.793 | 0.731 |
| m1 | あり | あり | 通常 | 0.682 | 0.773 | 0.745 | 0.724 |
| m1 | あり | なし | 侵襲性+通常 | 0.680 | 0.831 | 0.784 | 0.688 |
| m12 | あり | あり | 侵襲性+通常 | 0.671 | 0.822 | 0.755 | 0.678 |
| m12 | あり | なし | 通常 | 0.670 | 0.698 | 0.777 | 0.656 |
| m123 | なし | あり | 通常 | 0.668 | 0.744 | 0.608 | 0.559 |
| m12 | あり | なし | 侵襲性+通常 | 0.653 | 0.840 | 0.824 | 0.655 |
| m123 | あり | あり | 侵襲性+通常 | 0.651 | 0.829 | 0.663 | 0.649 |
| m123 | あり | なし | 通常 | 0.651 | 0.744 | 0.701 | 0.631 |
| m12 | あり | あり | 通常 | 0.627 | 0.638 | 0.672 | 0.622 |
| m123 | あり | なし | 侵襲性+通常 | 0.624 | 0.804 | 0.767 | 0.616 |
| m123 | あり | あり | 通常 | 0.607 | 0.691 | 0.685 | 0.639 |

【 0 1 0 0 】

【表 7】

表 7 : グリソンスコア>0の対象を包含する生存データに適用された、全ての積率ベースのモデルについての性能メトリック

| 包含される積率 | 包含される蛍光特徴 | 包含される形態特徴 | 包含される F O V | R S F に よる一致率 | 3 年後に おける A U C | 5 年後に おける A U C | 1 0 年後に おける A U C |
|---------|-----------|-----------|-------------|------------------|--------------------|--------------------|----------------------|
| m1 | あり | なし | 侵襲性 | 0.812 | 0.869 | 0.892 | 0.875 |
| m12 | あり | なし | 侵襲性+通常 | 0.802 | 0.883 | 0.901 | 0.843 |
| m12 | あり | あり | 侵襲性+通常 | 0.800 | 0.831 | 0.870 | 0.790 |
| m1 | あり | なし | 侵襲性+通常 | 0.800 | 0.886 | 0.897 | 0.875 |
| m1 | あり | あり | 侵襲性 | 0.798 | 0.860 | 0.889 | 0.879 |
| m123 | あり | なし | 侵襲性+通常 | 0.792 | 0.817 | 0.875 | 0.810 |
| m12 | あり | なし | 侵襲性 | 0.792 | 0.851 | 0.885 | 0.860 |
| m12 | あり | あり | 侵襲性 | 0.790 | 0.834 | 0.880 | 0.834 |
| m1 | あり | あり | 侵襲性+通常 | 0.788 | 0.880 | 0.897 | 0.869 |
| m123 | あり | あり | 侵襲性+通常 | 0.786 | 0.831 | 0.839 | 0.851 |
| m123 | あり | なし | 侵襲性 | 0.786 | 0.840 | 0.875 | 0.825 |
| m123 | あり | あり | 侵襲性 | 0.781 | 0.897 | 0.892 | 0.851 |
| m1 | なし | あり | 侵襲性+通常 | 0.767 | 0.886 | 0.702 | 0.782 |
| m12 | なし | あり | 侵襲性+通常 | 0.742 | 0.817 | 0.647 | 0.735 |
| m1 | なし | あり | 侵襲性 | 0.720 | 0.800 | 0.736 | 0.790 |
| m123 | なし | あり | 侵襲性+通常 | 0.713 | 0.814 | 0.615 | 0.724 |
| m123 | なし | なし | 侵襲性+通常 | 0.705 | 0.849 | 0.692 | 0.662 |
| m1 | なし | なし | 通常 | 0.703 | 0.851 | 0.690 | 0.647 |
| m12 | なし | なし | 侵襲性+通常 | 0.701 | 0.857 | 0.656 | 0.644 |
| m1 | なし | なし | 侵襲性+通常 | 0.697 | 0.854 | 0.695 | 0.649 |
| m123 | なし | なし | 侵襲性+通常 | 0.697 | 0.849 | 0.695 | 0.640 |
| m12 | なし | なし | 侵襲性+通常 | 0.695 | 0.834 | 0.675 | 0.640 |
| m1 | なし | なし | 侵襲性 | 0.695 | 0.837 | 0.675 | 0.644 |
| m123 | なし | なし | 通常 | 0.693 | 0.849 | 0.675 | 0.634 |
| m1 | なし | あり | 侵襲性+通常 | 0.691 | 0.749 | 0.620 | 0.763 |
| m12 | なし | なし | 侵襲性 | 0.691 | 0.857 | 0.678 | 0.642 |
| m12 | なし | なし | 通常 | 0.689 | 0.840 | 0.675 | 0.651 |
| m1 | なし | なし | 侵襲性+通常 | 0.687 | 0.843 | 0.673 | 0.640 |
| m123 | なし | なし | 侵襲性 | 0.678 | 0.849 | 0.691 | 0.627 |
| m123 | なし | あり | 侵襲性 | 0.676 | 0.847 | 0.764 | 0.763 |
| m12 | なし | あり | 侵襲性 | 0.676 | 0.771 | 0.690 | 0.738 |
| m1 | あり | あり | 侵襲性+通常 | 0.654 | 0.866 | 0.728 | 0.696 |
| m123 | なし | あり | 侵襲性+通常 | 0.650 | 0.786 | 0.757 | 0.744 |
| m12 | なし | あり | 侵襲性+通常 | 0.649 | 0.760 | 0.584 | 0.642 |
| m12 | あり | あり | 侵襲性+通常 | 0.639 | 0.840 | 0.695 | 0.670 |
| m1 | なし | あり | 通常 | 0.629 | 0.740 | 0.563 | 0.653 |
| m123 | あり | あり | 侵襲性+通常 | 0.621 | 0.800 | 0.650 | 0.610 |
| m1 | あり | なし | 侵襲性+通常 | 0.617 | 0.853 | 0.716 | 0.664 |
| m123 | なし | あり | 通常 | 0.610 | 0.637 | 0.464 | 0.509 |
| m12 | なし | あり | 通常 | 0.606 | 0.709 | 0.486 | 0.614 |
| m12 | あり | なし | 侵襲性+通常 | 0.594 | 0.820 | 0.728 | 0.677 |
| m1 | あり | なし | 通常 | 0.588 | 0.740 | 0.739 | 0.698 |
| m1 | あり | あり | 通常 | 0.579 | 0.691 | 0.690 | 0.657 |
| m123 | あり | なし | 通常 | 0.569 | 0.667 | 0.685 | 0.700 |
| m123 | あり | なし | 侵襲性+通常 | 0.561 | 0.777 | 0.677 | 0.565 |
| m123 | あり | あり | 通常 | 0.548 | 0.649 | 0.647 | 0.631 |
| m12 | あり | なし | 通常 | 0.532 | 0.660 | 0.678 | 0.608 |
| m12 | あり | あり | 通常 | 0.518 | 0.617 | 0.611 | 0.584 |

図 1 3 に示される通り、全コホート解析では、P I 3 K p 1 1 0 a、P C A D 及び p G S K 3 a が、マーカーのうちで最も予測的であった。図 1 4 は、膜における P I 3 K p 1 1 0 量及び p G S K 3 a 量の増大、並びに細胞全体における P C A D 量の減少が、生存の短縮と関連しうること示す。図 1 5 及び 1 6 で認められる通り、グリソンスコアが 0 を超える対象のコホートでは、重要な特徴の覧表が類似した。

【 0 1 0 1 】

全ての上位の特徴を明白なバッチ効果について点検したが、バッチ効果は見いだされなかった。例えば、バッチ 1 においてわずかな差分が認められるに過ぎない図 1 7 を参照されたい。

細胞クラスター解析

本発明の実施形態の細胞クラスター解析では、対象により層別化した全コホートからサンプリングされた 2 0 , 0 0 0 個の細胞についての 中央値によるクラスター化を用いて、1 4 のマーカー及び 4 つの細胞レベルの特徴、5 6 次元のマーカー空間に基づき、細胞を 群へとクラスター化した。 を 2 ~ 5 0 の範囲として、R (v . 2 . 1 1 . 0) 用の F l e x c l u s t ライブラリーの s t e p F l e x c l u s t 関数 (v . 1 . 3 - 1) を 2 0 連で実行した。次いで、クラスターのセントロイドからの距離を計算することによ

り、全コホート内の細胞を クラスターの1つと関連させた。これは、Flexclust 内の予測関数を用いて達成した。次いで、FOVレベルの細胞クラスターの特徴を、クラスターの各々に属するFOVの中の細胞の比率として定義した。作成される 群セットの各々について、個別の分類モデル及び生存モデルをフィッティングさせた。本発明の他の実施形態では、異なるクラスター化アルゴリズムを用いる場合もあり、アルゴリズムを細胞属性の異なるセットに適用する場合もあり、作成するクラスターセットの異なる範囲を指定する場合もあり、作成するクラスターセットの特定の数を同定する場合もある。

クラスター解析を用いる視野レベルの評価の予測

図18及び19に見られる通り、がん視野を正常視野と対比する分類子及び高悪性度ががん視野を低悪性度がん視野と対比する分類子のいずれの性能も、約20の細胞クラスターを包含した後で定常化した。細胞クラスターが20のとき、正常をがんと対比する分類子のAUCは、トレーニングセット及びテストセットのそれぞれにおいて96.1%及び95.7%であった。細胞クラスターが20のとき、高悪性度ががんを低悪性度がんと対比する分類子のAUCは減少した：トレーニングセットにおける88.0%及びテストセットにおける88.7%であった。形態特徴は、これらのモデル内に包含されなかった。

【0102】

20の細胞クラスターモデルについてのROC曲線を、図20及び21に示す。図22及び23において示される通り、がんを正常と対比する分類子及び 高悪性度ががんを低悪性度がんと対比する分類子のいずれの分類子においても、単一のクラスター7が、FOV悪性度について高度に予測的であるものとして突出している。クラスター7は、いずれのモデルにおいても、上位4つの特徴の残りの特徴と同様、正常組織の指標である(図24及び25を参照されたい)。高悪性度のがんにおいてクラスター7の細胞が少数であるパターンは、5つのバッチ全てにおいて明らかであった(図26を参照されたい)。

【0103】

クラスター7の細胞のFOV比率をバッチ効果について点検したが、バッチ効果は見いだされなかった。

【0104】

クラスター7のシグネチャーを、図27にプロットする。このクラスターの著明な特徴は、それら両者の細胞質における量の減少と関連する、NaKATPアーゼ及びベータカテニンの両方の核及び膜における量の増大である。

クラスター解析を用いる患者レベルの評価の予測

全コホート解析において、年齢及びグリソンスコアによるヌルモデルを幾分か上回って改善することができるのは、遠隔期生存の予測だけである。これを、Random Survival Forestによる一致率(RSF_CONC)、並びに3、5及び10年以内の前立腺がんによる死亡を分類するためのAUC(AUC_YR)を、モデル内に包含されるクラスター数と対比してプロットする図28に示す。正常FOVと対比した侵襲性FOVの組入れは、図中の色により差別化される。ヌルモデルより性能が大きい可能性があるモデルは、5年後以降における生存予測の改善を示したので、侵襲性の特徴を含むモデルである。一般に、6つのクラスターにより良好な性能がもたらされる。

【0105】

グリソンスコアが0を超えるコホート解析では、少なくとも5つの細胞クラスターを含むと、生存期間についての一致率のメトリック及び5年後及び10年後における死亡分類率が、ヌルモデルより良好である(図29を参照されたい)。約20クラスターを含むまでは、生存期間についての一致率が上昇するのに対し、5年後における死亡は、5つだけのクラスターによるときの分類が最も良好である。正常FOVに由来する特徴を含むことは一般に、モデルの性能を改善しない。

【0106】

図30における、全コホートに適用された、侵襲性組織の中の6つのクラスターを包含したモデルについての変数重要度プロットは、クラスター6が、モデルの中の他の5つのクラスターのいずれよりはるかに大きく予測的であることを示す。図31に示される通り

、クラスター 6 は、生存期間の短縮と関連する。

【 0 1 0 7 】

グリソンスコアが 0 を超えるコホートについての 2 0 クラスターによる解析では、2 つのクラスター：クラスター 7 及びクラスター 1 が、生存期間を予測するのに比較的重要である。図 3 2 は、グリソンスコアが 0 を超えるコホートについての生存モデルの変数重要度を示す図である。クラスター 7 が生存期間の延長と関連するのに対し、クラスター 1 は、生存期間の短縮と関連する。図 3 3 は、グリソンスコアが 0 を超えるコホートについての 2 0 クラスターのモデルにおける上位 4 つの特徴の偏変数依存性を示す図である。

【 0 1 0 8 】

上位のクラスター全てをバッチ効果について点検したが、バッチ効果は見いだされなかった。

10

【 0 1 0 9 】

クラスター 6 / 8 及びクラスター 1 / 2 0 のシグネチャーを、図 3 4 に示す。これらの 2 つのクラスターは、N a K A T P アーゼ、S 6、ベータカテニン、P C A D、P I 3 K p 1 1 0 a 及びケラチンの明確な局在化を顕徴とする、類似するシグネチャーを示す。これらの 2 つのクラスターはまた、細胞全体における N a K A T P アーゼ、ベータカテニン及びケラチンが幾分か低度であることも示す。

【 0 1 1 0 】

本明細書では、本発明の特定の特徴だけを例示及び記載してきたが、当業者は、多くの改変及び変更に想到するであろう。したがって、付属の特許請求の範囲は、このような改変及び変更の全てを、本発明の真の精神の範囲内に収まるものとして対象とすることを意図することが理解される。

20

【 0 1 1 1 】

特許請求の範囲では、特定の限定の組合せについて列挙するが、本発明は、各独立請求項を、それ自体で、また、明らかに不適合な限定の組合せを除き、関連する従属請求項において明記される任意の可能な限定の組合せと共に明示的に包摂する。例えば、本発明は、異なるクラスター化アルゴリズムの適用を要請する 2 つの従属請求項のうちの 1 つだけを除き、関連する従属請求項の各々において列挙される限定と組み合わせて、独立請求項 1 を明示的に包摂する。

【図 1】

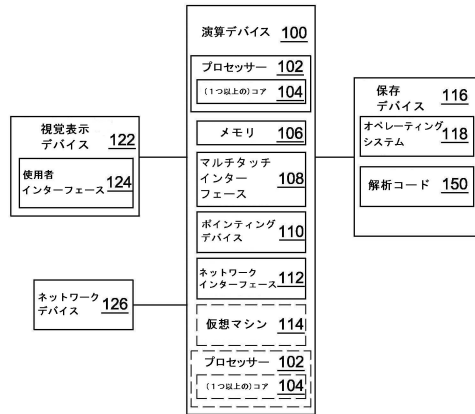


FIG. 1

【図 2】

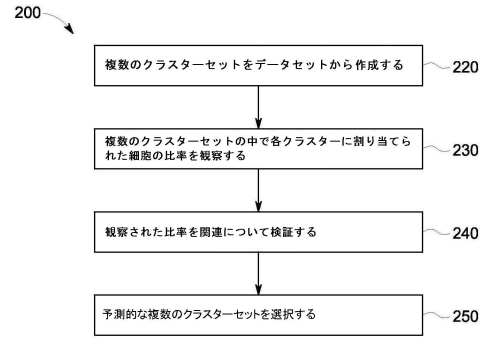


FIG. 2

【図 3】

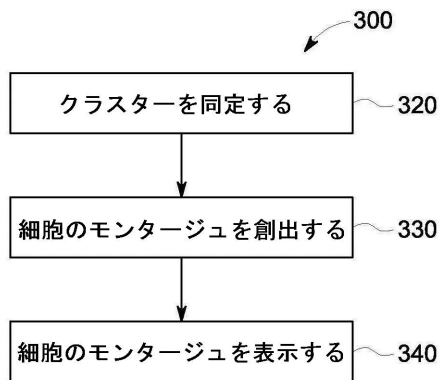


FIG. 3

【図 4】

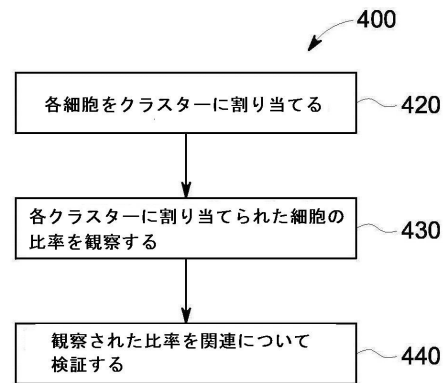


FIG. 4

【図 5】

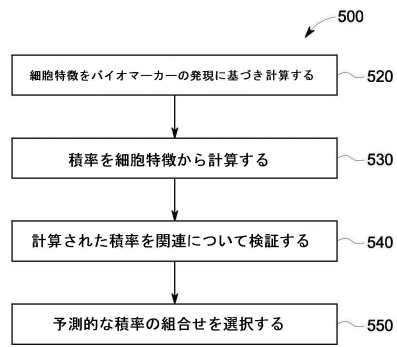


FIG. 5

【図 6】

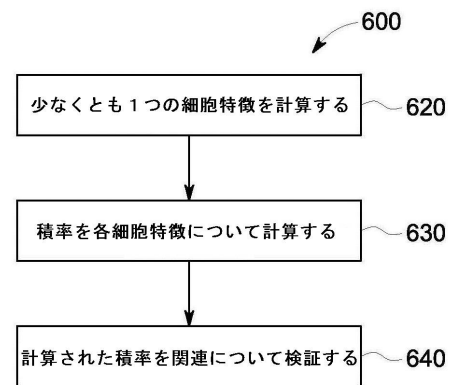


FIG. 6

【図 7】

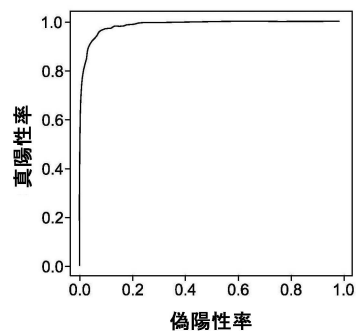


FIG. 7

【図 8】

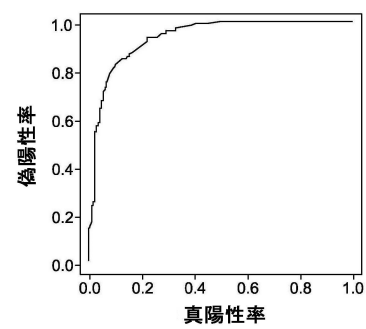


FIG. 8

【図 9】

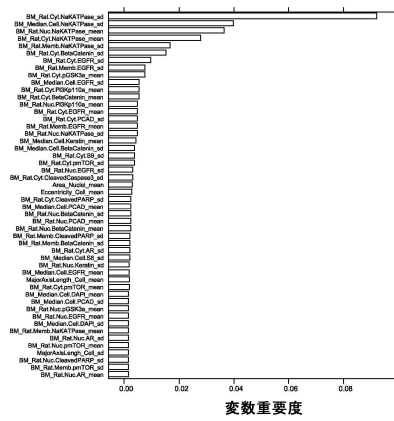


FIG. 9

【図 10】

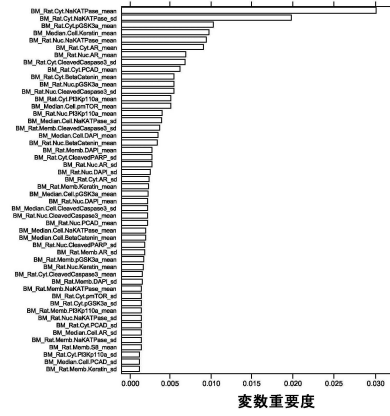


FIG. 10

【図 11】

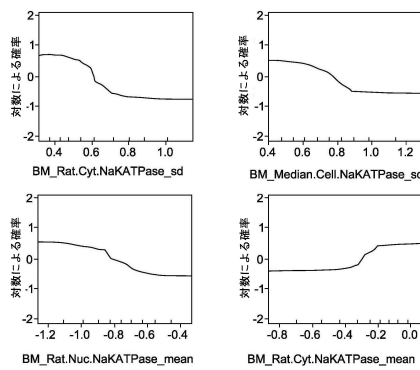


FIG. 11

【図 12】

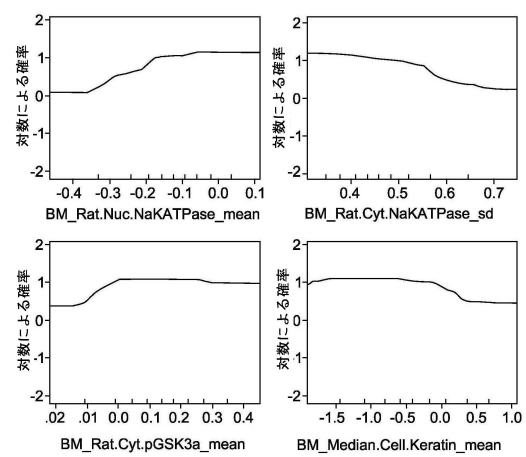


FIG. 12

【 図 1 4 】



【 図 1 6 】



【図 17】

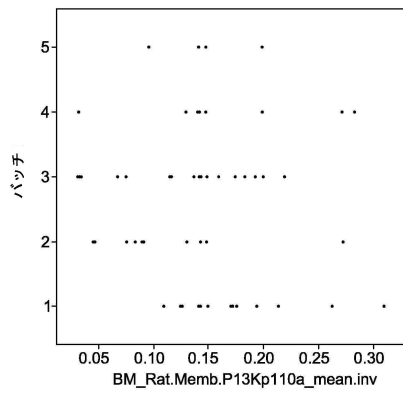


FIG. 17

【図 18】

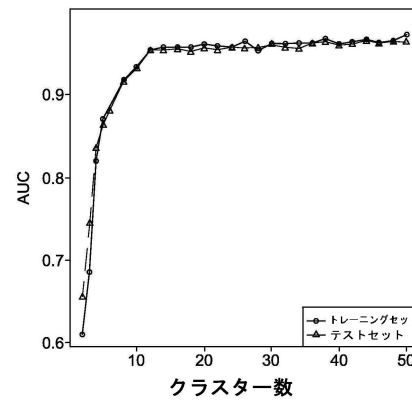


FIG. 18

【図 19】

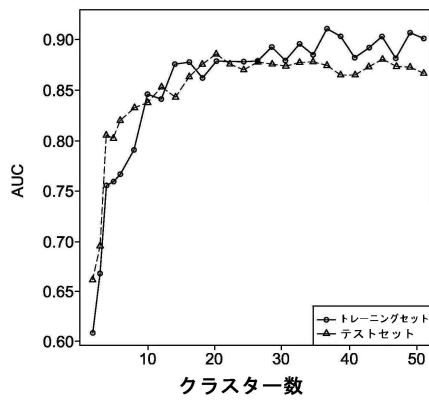


FIG. 19

【図 20】

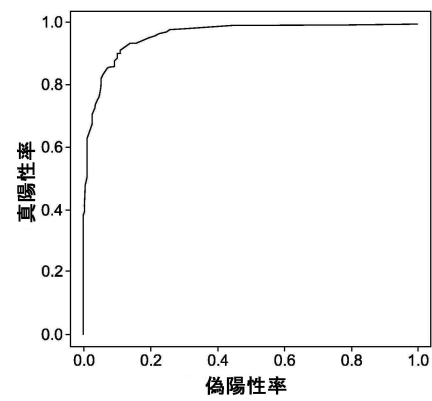


FIG. 20

【図 2 1】

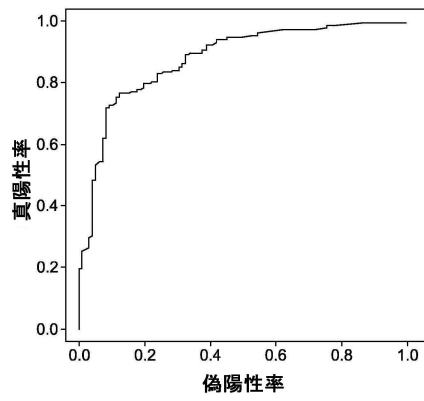


FIG. 21

【図 2 2】

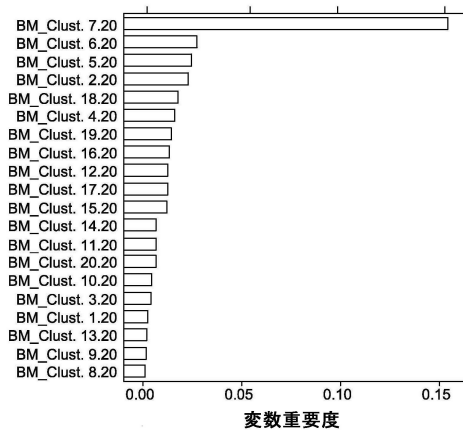


FIG. 22

【図 2 3】

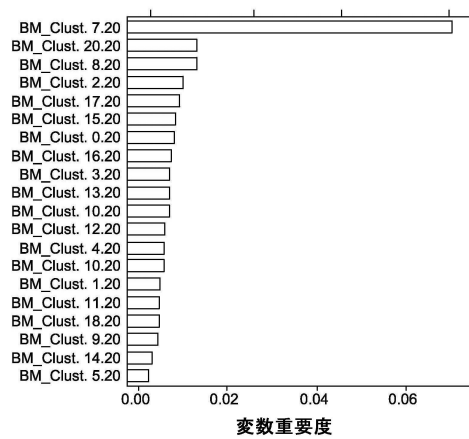


FIG. 23

【図 2 4】

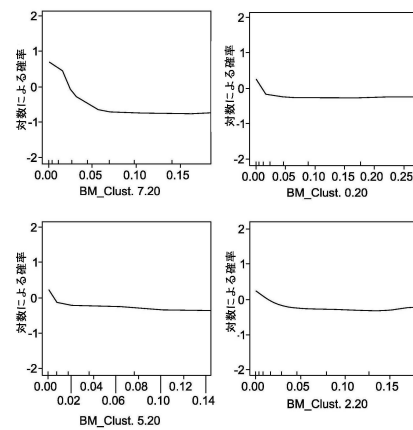


FIG. 24

【図 25】

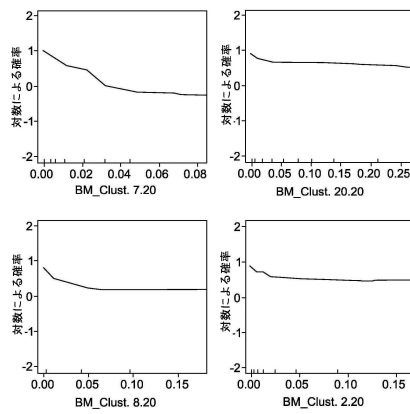


FIG. 25

【図 26】

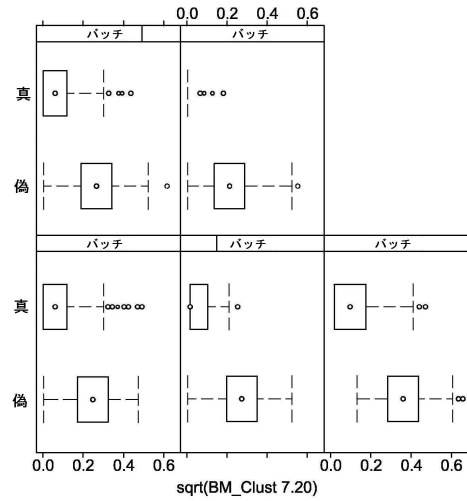


FIG. 26

【図 27】

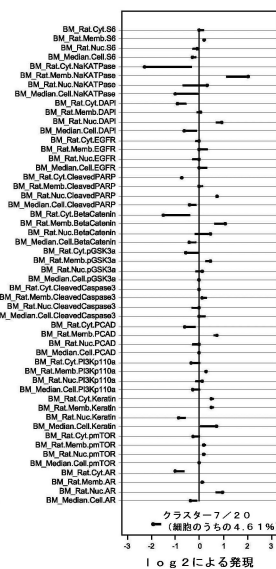


FIG. 27

【図 28】

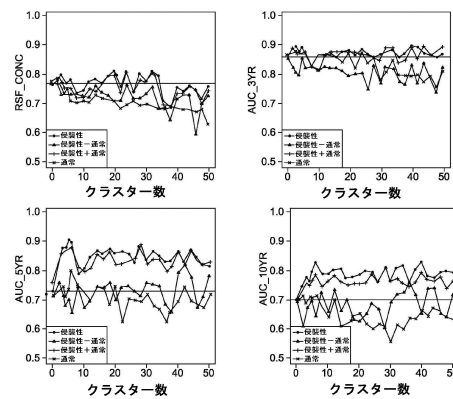


FIG. 28

【図 29】

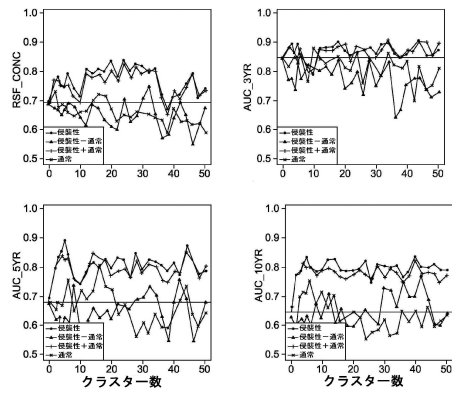


FIG. 29

【図 30】

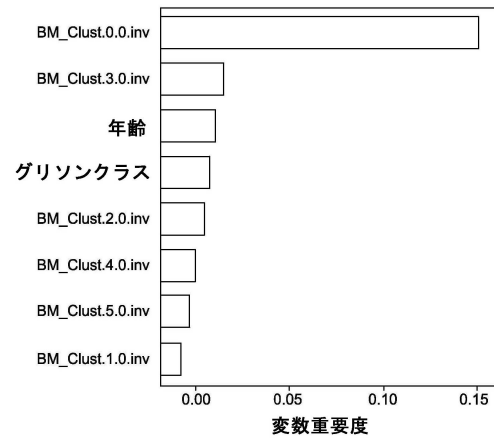


FIG. 30

【図 31】

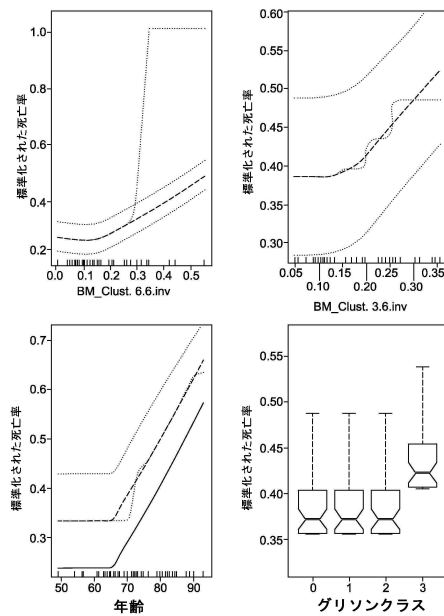


FIG. 31

【図 32】

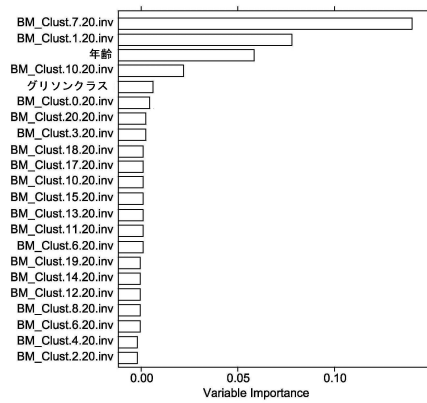


FIG. 32

【図 33】

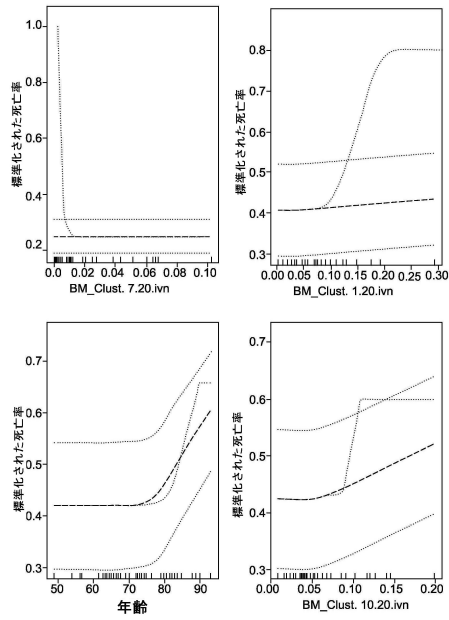


FIG. 33

【図 34】

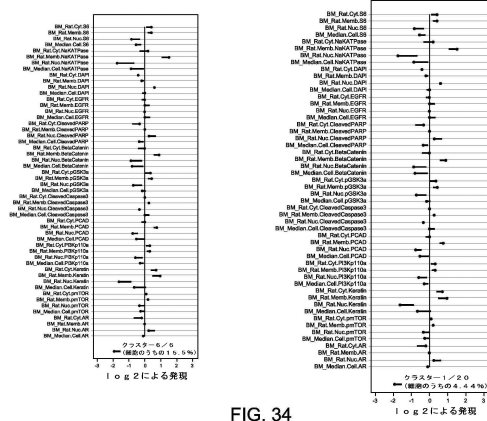


FIG. 34

【図 35】

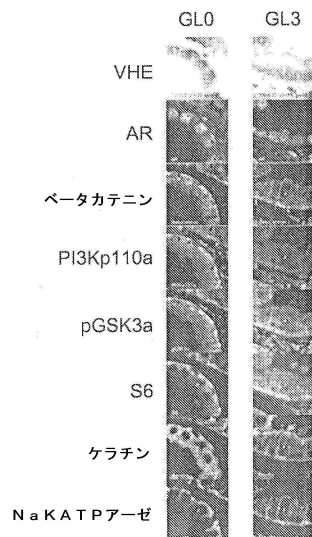


FIG. 35

フロントページの続き

- (31)優先権主張番号 13/252,072
(32)優先日 平成23年10月3日(2011.10.3)
(33)優先権主張国 米国(US)
(31)優先権主張番号 13/252,080
(32)優先日 平成23年10月3日(2011.10.3)
(33)優先権主張国 米国(US)
(31)優先権主張番号 13/252,078
(32)優先日 平成23年10月3日(2011.10.3)
(33)優先権主張国 米国(US)

- (72)発明者 マコラック, コリン・クレイグ
アメリカ合衆国、ニューヨーク州・12309、ニスカユナ、ワン・リサーチ・サークル、ジーイー・グローバル・リサーチ

審査官 大瀧 真理

- (56)参考文献 米国特許出願公開第2011/0091081(US, A1)
特表2010-527448(JP, A)
国際公開第2007/028944(WO, A1)

- (58)調査した分野(Int.Cl., DB名)
G01N 33/48 - 33/98
G06T 1/00