(19)日本国特許庁(JP)　(12)公表特許公報(A)　(11)特許出願公表番号
特表2020-522690
(P2020-522690A)
(43)公表日　令和2年7月30日(2020.7.30)

(51)Int.Cl.		F I			テーマコード（参考）
G01N 33/50	***(2006.01)***	G01N	33/50	Z	2G045
G01N 33/574	***(2006.01)***	G01N	33/574	Z	4B063
G01N 33/53	***(2006.01)***	G01N	33/53	M	
C12Q 1/6809	***(2018.01)***	C12Q	1/6809	Z	

審査請求 未請求　予備審査請求 未請求　（全 133 頁）

(21)出願番号　特願2019-565941 (P2019-565941)
(86)(22)出願日　平成30年6月1日(2018.6.1)
(85)翻訳文提出日　令和2年1月24日(2020.1.24)
(86)国際出願番号　PCT/US2018/035702
(87)国際公開番号　WO2018/223066
(87)国際公開日　平成30年12月6日(2018.12.6)
(31)優先権主張番号　62/514,595
(32)優先日　平成29年6月2日(2017.6.2)
(33)優先権主張国・地域又は機関
米国(US)
(31)優先権主張番号　62/546,936
(32)優先日　平成29年8月17日(2017.8.17)
(33)優先権主張国・地域又は機関
米国(US)

(71)出願人　511118469
ベラサイト　インコーポレイテッド
アメリカ合衆国，カリフォルニア州　９４０８０，サウス　サンフランシスコ，ショアライン　コート　６０００　ナンバー３００
(74)代理人　100079108
弁理士　稲葉　良幸
(74)代理人　100109346
弁理士　大貫　敏史
(74)代理人　100117189
弁理士　江口　昭彦
(74)代理人　100134120
弁理士　内藤　和彦

最終頁に続く

(54)【発明の名称】肺疾病の特定又はモニタリング方法及びシステム

(57)【要約】
【課題】　肺疾病を同定又はモニタリングするための方法及びシステムに関する。
【解決手段】　不要な侵襲的処置を最小化し、疾患及び再発の早期発見を促進し、疾患の予防又は回復のための介入療法の有効性をモニタリングするために、現行の臨床治療経路における様々な決定ポイントにおいてゲノム分類器を用いて、肺疾病に対する現行の臨床治療経路を改善するための、方法、システム、及びキットを提供する。
【選択図】　図３

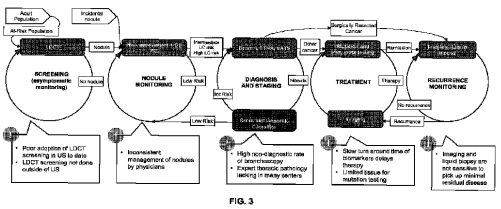


FIG. 3

【特許請求の範囲】
【請求項１】
肺疾病について被検体をスクリーニングする方法であって、
（ａ）（１）前記肺疾病を発症する１つ以上のリスク因子の存在についてコンピュータ分析を行い、（２）前記１つ以上のリスク因子の存在が特定された被検体から採取された第１の試料からの上皮組織をアッセイすることであって、前記第１の試料における前記肺疾病を発症するリスクに関連付けられた１つ以上のバイオマーカーの有無を特定する、こと、
（ｂ）前記１つ以上のバイオマーカーの有無を特定するにあたり、（ｉ）前記被検体において前記肺疾病に罹患する疑いのある肺領域の電子画像スキャンを取得するように指示すること、又は、（ｉｉ）前記被検体の第２の試料から他の上皮組織をアッセイすること、
を含む、方法。
【請求項２】
（ｂ）の前に、前記被検体の前記上皮組織を含む前記第１の試料のアッセイの要請を受信すること、を更に含む、請求項１に記載の方法。
【請求項３】
前記電子画像スキャンは、低線量断層撮影（ＬＤＣＴ）スキャン又は磁気共鳴画像法（ＭＲＩ）である、請求項１に記載の方法。
【請求項４】
前記ＬＤＣＴスキャンは、前記被検体に対する照射線量が約５ミリシーベルト（ｍＳｖ）未満である、請求項３に記載の方法。
【請求項５】
前記肺疾病は、肺癌、慢性閉塞性肺疾患（ＣＯＰＤ）、間質性肺疾患（ＩＬＤ）、又はこれらの任意の組み合わせである、請求項１に記載の方法。
【請求項６】
前記肺疾病は肺癌であり、前記肺癌は、小細胞肺癌、腺癌、扁平上皮癌、大細胞癌、小細胞肺癌、又はこれらの任意の組み合わせを含む、請求項１に記載の方法。
【請求項７】
前記第１の試料又は前記第２の試料は、気管支鏡検査により採取される、請求項１に記載の方法。
【請求項８】
前記第１の試料又は前記第２の試料は、穿刺吸引法により採取される、請求項１に記載の方法。
【請求項９】
前記第１の試料又は前記第２の試料は、粘膜上皮組織、鼻上皮組織、肺上皮組織、又はこれらの任意の組み合わせを含む、請求項１に記載の方法。
【請求項１０】
前記第１の試料又は前記第２の試料は、前記被検体の気道から採取された上皮組織を含む、請求項１に記載の方法。
【請求項１１】
前記第１の試料又は前記第２の試料の一部分に対して、前記第１の試料又は前記第２の試料を不明確なもの又は疑わしいものとして特定する細胞検査を行う、請求項１に記載の方法。
【請求項１２】
前記第１の試料又は前記第２の試料を不明確なもの又は疑わしいものとして特定するにあたり、前記試料の第２の部分について（ｂ）を行い、前記第２の部分は前記上皮組織を含む、請求項１１に記載の方法。
【請求項１３】
前記第２の試料は、前記第１の試料とは異なる、請求項１に記載の方法。

【請求項１４】
前記第２の試料は、前記第１の試料とは異なる種類の試料である、請求項１に記載の方法。
【請求項１５】
前記第１の試料は第１の時点にて前記被検体から採取されたものであり、前記第２の試料は、前記第１の時点より後である第２の時点にて前記被検体から採取されたものである、請求項１に記載の方法。
【請求項１６】
前記第２の時点は前記第１の時点から約１～２年以内である、請求項１５に記載の方法。
【請求項１７】
（ａ）は、前記１つ以上のバイオマーカーの有無を、１つ以上のバイオマーカーの参照セットと比較することを含む、請求項１に記載の方法。
【請求項１８】
前記被検体は、前記肺疾病の治療を必要とする、請求項１に記載の方法。
【請求項１９】
前記被検体は、肺疾病を発症するリスクが増加している疑いがある、請求項１に記載の方法。
【請求項２０】
前記被検体は、前記肺疾病に対して無症状である、請求項１に記載の方法。
【請求項２１】
前記被検体は、以前に前記電子画像スキャンを受けたことがない、請求項１に記載の方法。
【請求項２２】
前記被検体は、以前に確定診断を受けたことがない、請求項１に記載の方法。
【請求項２３】
前記１つ以上のリスク因子は、喫煙、受動喫煙、ラドン曝露、大気汚染曝露、放射線被曝、工業物質曝露、遺伝又は環境による遺伝子突然変異、被検体の年齢、被検体の二次疾患、又はこれらの任意の組み合わせを含む、請求項１に記載の方法。
【請求項２４】
前記被検体は、２つ以上のリスク因子を有する、請求項１に記載の方法。
【請求項２５】
前記１つ以上のバイオマーカーは、少なくとも５つのバイオマーカーを含む、請求項１に記載の方法。
【請求項２６】
前記１つ以上のバイオマーカーは、遺伝子又は遺伝子フラグメント、配列変異体、融合体、ミトコンドリア転写物、エピジェネティック修飾、コピー数多型、ヘテロ接合性の消失（ＬＯＨ）、又はこれらの任意の組み合わせのうち１つ以上を含む、請求項１に記載の方法。
【請求項２７】
前記１つ以上のバイオマーカーの有無は、発現レベルにより表される、請求項１に記載の方法。
【請求項２８】
前記方法は、前記被検体が前記肺疾病を発症するリスクが増加しているか否かを特定する、請求項１に記載の方法。
【請求項２９】
（ｂ）における前記特定は、訓練アルゴリズムの採用を含む、請求項１に記載の方法。
【請求項３０】
前記訓練アルゴリズムは、個人の気道から採取された上皮細胞を含む訓練セットにより訓練されている、請求項２９に記載の方法。

【請求項３１】
前記訓練アルゴリズムは、前記肺疾病に対して良性の試料と、前記肺疾病に対して悪性の試料とを含む訓練セットにより訓練されている、請求項２９に記載の方法。
【請求項３２】
前記訓練アルゴリズムは、１つ以上のリスク因子を有する複数の被検体から採取された試料を含む訓練セットにより訓練されている、請求項２９に記載の方法。
【請求項３３】
前記方法は、（ａ）の前に、前記被検体をコンピュータ分析することであって、前記被検体における前記肺疾病を発症する前記１つ以上のリスク因子の存在を特定する、ことを更に含む、請求項１に記載の方法。
【請求項３４】
肺疾病に罹患している又は肺疾病に罹患する疑いのある被検体をモニタリングする方法であって、
（ａ）前記肺疾病に罹患する疑いがあり１つ以上の肺結節について以前に陽性の徴候を示した被検体から採取された上皮組織を含む第１の試料をアッセイすることであって、前記肺疾病と関連付けられた１つ以上のバイオマーカーの有無を特定する、こと、
（ｂ）前記１つ以上のバイオマーカーの有無を特定するにあたり、（ｉ）前記被検体から第２の試料を採取すること、又は、（ｉｉ）（ａ）の結果に基づき前記被検体の肺領域の電子画像スキャンを取得するように前記被検体に指示すること、
を含む、方法。
【請求項３５】
前記陽性の徴候は、以前に電子画像スキャンによって特定されたものである、請求項３４に記載の方法。
【請求項３６】
前記１つ以上の肺結節は、少なくとも２つの結節である、請求項３４に記載の方法。
【請求項３７】
前記被検体から前記第２の試料を採取することは、前記被検体に対して気管支鏡検査、胸腔穿刺法（ＴＴＮＡ）、又は胸腔鏡補助下手術（ＶＡＴＳ）を行うことを含む、請求項３４に記載の方法。
【請求項３８】
前記被検体から前記第２の試料を採取することは、組織生検を行うことを含む、請求項３４に記載の方法。
【請求項３９】
前記１つ以上のバイオマーカーの有無が、前記被検体が前記肺疾病に罹患するリスクが高いか低いかを特定する、請求項３４に記載の方法。
【請求項４０】
（ｂ）は、評価されたリスクに応じて（ｉ）又は（ｉｉ）を推奨することを更に含む、請求項３９に記載の方法。
【請求項４１】
前記被検体の血液試料を分析すること、前記被検体に対する電子画像スキャンを行うこと、又はこれらの組み合わせを更に含む、請求項３４に記載の方法。
【請求項４２】
前記第２の試料は上皮の試料であり、（ｂ）に続いて、上皮組織の前記試料が１つ以上の追加バイオマーカーの有無についてアッセイされる、請求項３４に記載の方法。
【請求項４３】
前記１つ以上の追加バイオマーカーは、前記１つ以上のバイオマーカーである、請求項４２に記載の方法。
【請求項４４】
以前に前記肺疾病の予防又は治療のために介入療法を受けることを推奨されている、肺疾病に罹患している又は肺疾病に罹患する疑いのある被検体をモニタリングする方法であ

って、
（ａ）前記被検体が前記肺疾病に対して前記介入療法の少なくとも一部を完了した後、前記被検体から採取された上皮組織を含む第１の試料をアッセイすることであって、遺伝データを生成する、こと、
（ｂ）前記遺伝データを処理することであって、前記肺疾病と関連付けられた１つ以上のバイオマーカーの有無を特定する、こと、
（ｃ）前記被検体から第２の試料を採取することを推奨することを含むレポートをコンピュータで生成すること、
を含む、方法。
【請求項４５】
前記方法は、前記介入療法に対する被検体のコンプライアンスを特定する、請求項４４に記載の方法。
【請求項４６】
前記方法は、前記肺疾病の予防又は治療に対する前記介入療法の有効性を特定する、請求項４４に記載の方法。
【請求項４７】
前記介入療法は、前記被検体に対して医薬組成物を投与することを含む、請求項４４に記載の方法。
【請求項４８】
前記医薬組成物は化学療法薬を含む、請求項４７に記載の方法。
【請求項４９】
前記介入療法は、運動療法、食事療法、減煙又は禁煙、又はこれらの任意の組み合わせを含む、請求項４４に記載の方法。
【請求項５０】
（ｂ）は、前記遺伝データを処理することであって、前記１つ以上のバイオマーカーのそれぞれに対応する発現レベルを特定する、ことを含む、請求項４４に記載の方法。
【請求項５１】
肺疾病について前記被検体をモニタリングする方法であって、
（ａ）肺疾病に対する治療を以前に開始している被検体から採取された上皮組織を含む第１の試料をアッセイすること及び１つ以上のバイオマーカーの有無を特定すること、
（ｂ）寛解の確定を受けるにあたり、前記被検体から第２の試料を採取すること及び前記第２の試料で（ａ）を繰り返すこと、を含む、方法。
【請求項５２】
（ｂ）は、前記遺伝データを処理することであって、前記１つ以上のバイオマーカーにおいて少なくとも１つの遺伝子異常を特定する、ことを含む、請求項４４又は５１に記載の方法。
【請求項５３】
前記方法は、非侵襲的モニタリングにより肺疾病の再発を早期段階で特定する、請求項５１に記載の方法。
【請求項５４】
前記被検体は、前記肺疾病に対して再発のリスクが増加している疑いがある、請求項５１に記載の方法。
【請求項５５】
肺疾病に罹患している又は肺疾病に罹患する疑いのある被検体をモニタリングする方法であって、
（ａ）前記肺疾病に罹患する疑いがあり、１つ以上の肺結節について以前に陰性の徴候を示した被検体から採取された上皮組織を含む第１の試料をアッセイすることであって、前記肺疾病と関連付けられた１つ以上のバイオマーカーの有無を特定する、こと、
（ｂ）前記１つ以上のバイオマーカーの有無を特定するにあたり、（ｉ）前記被検体から第２の試料を採取すること、又は、（ｉｉ）（ａ）の結果に基づき前記被検体の肺領域

の電子画像スキャンを取得するように前記被検体に指示すること、
を含む、方法。
【請求項５６】
（ａ）の前に、前記被検体に対して前記肺疾病を発症する１つ以上のリスク因子の存在についてコンピュータ分析すること、及び前記被検体から前記１つ以上のリスク因子の存在を特定すること、を更に含む、請求項５５に記載の方法。
【請求項５７】
肺疾病について被検体をスクリーニングするシステムであって、
被検体のヘルスデータ又は生理データを含む１つ以上のコンピュータデータベースと、
（ｉ）前記被検体に対して前記肺疾病を発症する１つ以上のリスク因子の存在について前記ヘルスデータ又は生理データを分析し、（２）前記１つ以上のリスク因子を特定するにあたり、前記被検体の試料からの上皮組織に対して前記肺疾病を発症するリスクと関連付けられた１つ以上のバイオマーカーに関するアッセイの推奨を生成するように、個別又は集合的にプログラムされた、１つ以上のコンピュータプロセッサと
を含む、システム。
【請求項５８】
肺疾病について被検体をスクリーニングするシステムであって、
（ｉ）被検体において前記肺疾病を発症する１つ以上のリスク因子の存在を示すデータを含む第１のデータセットと、（ｉｉ）前記被検体の試料における上皮組織における１つ以上のリスク因子の有無を示すデータを含む第２のデータセットとを含み、１つ以上のバイオマーカーは前記肺疾病を発症するリスクと関連付けられる、１つ以上のコンピュータデータベースと、
（ｉ）前記第１のデータセットを分析して前記１つ以上のリスク因子の存在を特定し、（ｉｉ）前記第２のデータセットを分析して前記１つ以上のバイオマーカーの存在を特定し、（ｉｉｉ）前記１つ以上のバイオマーカーの有無を特定するにあたり、（１）前記被検体において前記肺疾病に罹患する疑いのある肺領域の電子画像スキャンを取得するように指示する、又は（２）前記被検体の第２の試料からの他の上皮組織をアッセイするように指示する、レポートを生成するように、個別又は集合的にプログラムされた、１つ以上のコンピュータプロセッサと
を含む、システム。
【請求項５９】
肺疾病に罹患している又は肺疾病に罹患する疑いのある被検体をモニタリングするシステムであって、
前記被検体の第１の試料における上皮組織における１つ以上のバイオマーカーの有無を示すデータを含み、１つ以上のバイオマーカーは前記肺疾病と関連付けられる、１つ以上のコンピュータデータベースと、
（ｉ）前記被検体が以前に１つ以上の肺結節の存在について陽性の徴候を示していたと判定し、（ｉｉ）（ｉ）に続いて、前記データセットを処理して前記１つ以上のバイオマーカーの有無を特定し、（ｉｉｉ）前記１つ以上のバイオマーカーの有無を特定するにあたり、（１）前記被検体から第２の試料を採取するように指示する、又は（２）前記被検体の肺領域の電子画像スキャンを更に取得するように指示する、レポートを生成するように、個別又は集合的にプログラムされた、１つ以上のコンピュータプロセッサと、
を含む、システム。
【請求項６０】
以前に前記肺疾病の予防又は治療のために介入療法を完了することを推奨されている、肺疾病に罹患している又は肺疾病に罹患する疑いのある被検体をモニタリングするシステムであって、
遺伝データを含むデータセットを含む１つ以上のコンピュータデータベースと、
（ｉ）前記被検体が前記肺疾病に対して前記介入療法の少なくとも一部を完了した後、前記遺伝データを処理して前記肺疾病と関連付けられる１つ以上のバイオマーカーの有無

を特定し、（ｉｉｉ）前記被検体から第２の試料を採取することを推奨することを含むレポートを生成するように、個別又は集合的にプログラムされた、１つ以上のコンピュータプロセッサと
を含む、システム。
【発明の詳細な説明】
【技術分野】
【０００１】
（関連出願）
[1] 本願は、２０１７年６月２日付で出願された米国仮出願第６２／５１４５９５号、及び、２０１７年８月１７日付けで出願された米国仮出願第６２／５４６９３６号の優先権を主張し、全体が参照により援用される。
【背景技術】
【０００２】
[2] 現在、肺癌などの肺疾病を検出する方法が複数存在する。そのような肺疾病に対する現在の臨床治療経路(clinical pathway of care)は、不要な侵襲的処置の割合が高く、肺疾病の早期発見ができないか、肺疾病を発症する被検体のリスクをアッセイできない。
【発明の概要】
【発明が解決しようとする課題】
【０００３】
[3] 本開示は、被検体が肺疾病、例えば肺癌に罹患するか、又は罹患するリスクを有するかを判定する方法及びシステムを提供する。本開示の方法によると、被検体における肺疾病の進行や退行を、例えば、被検体から非侵襲的に採取された試料（例えば、鼻組織試料）を用いてスクリーニングやモニタリングすることができる。これは、肺疾病に対して無症状を示すが、肺疾病を発症するリスク（例えば、受動喫煙や大気汚染曝露）がある被検体に対するスクリーニングや、肺疾病に罹患しているか、又は罹患している疑いのある被検体に対してモニタリングにも活用し得る。
【課題を解決するための手段】
【０００４】
[4] 本開示の一態様は、肺疾病について被検体をスクリーニングする方法であって、（ａ）（１）肺疾病を発症する１つ以上のリスク因子の存在についてコンピュータ分析を行い、（２）１つ以上のリスク因子の存在が特定された被検体から採取された第１の試料からの上皮組織をアッセイし、第１の試料における肺疾病を発症するリスクに関連付けられた１つ以上のバイオマーカーの有無を特定すること、（ｂ）１つ以上のバイオマーカーの有無を特定するにあたり、（ｉ）被検体において肺疾病に罹患する疑いのある肺領域の電子画像スキャンを取得するように指示すること、又は、（ｉｉ）被検体の第２の試料から他の上皮組織をアッセイすることを含む、方法を提供する。一部の実施形態において、方法は、（ｂ）の前に、被検体の上皮組織を含む第１の試料のアッセイの要請を受信することを更に含む。
【０００５】
[5] 一部の実施形態において、電子画像スキャンは、低線量断層撮影（ＬＤＣＴ）スキャン又は磁気共鳴画像法（ＭＲＩ）である。一部の実施形態において、ＬＤＣＴスキャンは被検体に対する照射線量が約５ミリシーベルト（ｍＳｖ）未満である。
【０００６】
[6] 一部の実施形態において、肺疾病は、肺癌、慢性閉塞性肺疾患（ＣＯＰＤ）、間質性肺疾病（ＩＬＤ）、又はこれらの任意の組み合わせである。一部の実施形態において、肺疾病は肺癌であり、肺癌は、小細胞肺癌、腺癌、扁平上皮癌、大細胞癌、小細胞肺癌、又はこれらの任意の組み合わせを含む。
【０００７】
[7] 一部の実施形態において、第１の試料又は第２の試料は気管支鏡検査により採取される。一部の実施形態において、第１の試料又は第２の試料は穿刺吸引法により採取され

る。一部の実施形態において、第１の試料又は第２の試料は、粘膜上皮組織、鼻上皮組織、肺上皮組織、又はこれらの任意の組み合わせを含む。一部の実施形態において、第１の試料又は第２の試料は、被検体の気道から採取された上皮組織を含む。

【０００８】

[8] 一部の実施形態において、第１の試料又は第２の試料の一部分に対して、第１の試料又は第２の試料を不明確なもの又は疑わしいものとして特定する細胞検査を行う。一部の実施形態において、第１の試料又は第２の試料を不明確なもの又は疑わしいものとして特定するにあたり、試料の第２の部分について（ｂ）を行い、第２の部分は上皮組織を含む。

【０００９】

[9] 一部の実施形態において、第２の試料は、第１の試料とは異なる。一部の実施形態において、第２の試料は、第１の試料とは異なる種類の試料である。一部の実施形態において、第１の試料は第１の時点にて被検体から採取されたものであり、第２の試料は第１の時点より後である第２の時点にて被検体から採取されたものである。一部の実施形態において、第２の時点は第１の時点から約１～２年以内である。

【００１０】

[10] 一部の実施形態において、（ａ）は、１つ以上のバイオマーカーの有無を特定するにあたり、１つ以上のバイオマーカーの参照セットと比較することを含む。一部の実施形態において、被検体は、肺疾病の治療を必要とする。一部の実施形態において、被検体は、肺疾病を発症するリスクが増加している疑いがある。一部の実施形態において、被検体は、肺疾病に対して無症状である。一部の実施形態において、被検体は、以前に電子画像スキャンを受けたことがない。一部の実施形態において、被検体は、以前に確定診断を受けたことがない。

【００１１】

[11] 一部の実施形態において、１つ以上のリスク因子は、喫煙、受動喫煙、ラドン曝露、大気汚染曝露、放射線被曝、工業物質曝露、遺伝又は環境による遺伝子突然変異、被検体の年齢、被検体の二次疾患、又はこれらの任意の組み合わせを含む。一部の実施形態において、被検体は、２つ以上のリスク因子を有する。

【００１２】

[12] 一部の実施形態において、１つ以上のバイオマーカーは、少なくとも５つである。一部の実施形態において、１つ以上のバイオマーカーは、遺伝子又は遺伝子フラグメント、配列変異体、融合体、ミトコンドリア転写物、エピジェネティック修飾、コピー数多型、ヘテロ接合性の消失（ＬＯＨ）、又はこれらの任意の組み合わせのうち１つ以上を含む。一部の実施形態において、１つ以上のバイオマーカーの有無は、発現レベルにより表される。

【００１３】

[13] 一部の実施形態において、方法は、被検体において肺疾病を発症するリスクが増加しているか否かを特定する。一部の実施形態において、（ｂ）における特定は、訓練アルゴリズムの採用を含む。一部の実施形態において、訓練アルゴリズムは、個人の気道から採取された上皮細胞を含む訓練セットにより訓練されている。一部の実施形態において、訓練アルゴリズムは、肺疾病に対して良性の試料と、肺疾病に対して悪性の試料とを含む訓練セットにより訓練されている。一部の実施形態において、訓練アルゴリズムは、１つ以上のリスク因子を有する複数の被検体から採取された試料を含む訓練セットにより訓練されている。

【００１４】

[14] 一部の実施形態において、方法は、（ａ）の前に、被検体に対してコンピュータ分析を行い、被検体から肺疾病を発症する１つ以上のリスク因子の存在を特定することを更に含む。

【００１５】

[15] 本開示の別の態様は、肺疾病に罹患している又は肺疾病に罹患する疑いのある被検

体をモニタリングする方法であって、（ａ）肺疾病に罹患する疑いがあり１つ以上の肺結節について陽性の徴候を示す被検体から採取された上皮組織を含む第１の試料をアッセイして肺疾病と関連付けられた１つ以上のバイオマーカーの有無を特定すること、（ｂ）１つ以上のバイオマーカーの有無を特定するにあたり、（ｉ）被検体から第２の試料を採取し、又は、（ｉｉ）（ａ）の結果に基づき被検体の肺領域の電子画像スキャンを取得するように被検体に指示することを含む、方法を提供する。

【００１６】

[16] 一部の実施形態において、陽性の徴候は、以前に電子画像スキャンによって特定されたものである。一部の実施形態において、電子画像スキャンは、低線量断層撮影（ＬＤＣＴ）スキャン又は磁気共鳴画像法（ＭＲＩ）である。一部の実施形態において、ＬＤＣＴスキャンは、被検体に対する照射線量が約５ｍＳｖ未満である。

【００１７】

[17] 一部の実施形態において、１つ以上の肺結節は、少なくとも２つの結節である。一部の実施形態において、被検体における第２の試料の採取は、被検体に対して気管支鏡検査、胸腔穿刺法（ＴＴＮＡ）、又は胸腔鏡補助下手術（ＶＡＴＳ）を行うことを含む。一部の実施形態において、被検体における第２の試料の採取は、組織生検を含む。

【００１８】

[18] 一部の実施形態において、１つ以上のバイオマーカーの有無により、被検体が肺疾病に罹患するリスクが高いか低いかを特定する。一部の実施形態において、（ｂ）は、評価されたリスクに応じて（ｉ）又は（ｉｉ）を推奨することを更に含む。

【００１９】

[19] 一部の実施形態において、肺疾病は、肺癌、慢性閉塞性肺疾病（ＣＯＰＤ）、間質性肺疾患（ＩＬＤ）、又はこれらの任意の組み合わせである。一部の実施形態において、肺疾病は肺癌であり、肺癌は、小細胞肺癌、腺癌、扁平上皮癌、大細胞癌、小細胞肺癌、又はこれらの任意の組み合わせを含む。

【００２０】

[20] 一部の実施形態において、第１の試料又は第２の試料は気管支鏡検査により採取される。一部の実施形態において、第１の試料又は第２の試料は穿刺吸引法により採取される。一部の実施形態において、第１の試料又は第２の試料は、粘膜上皮組織、鼻上皮組織、肺上皮組織、又はこれらの任意の組み合わせを含む。一部の実施形態において、第１の試料又は第２の試料は、被検体の気道から採取された上皮組織を含む。

【００２１】

[21] 一部の実施形態において、第２の試料は、第１の試料とは異なる。一部の実施形態において、第２の試料は、第１の試料とは異なる種類の試料である。一部の実施形態において、第２の試料は第１の試料の採取から時間を置いて被検体から採取されている。一部の実施形態において、時間は、約１～２年である。

【００２２】

[22] 一部の実施形態において、（ｂ）は、１つ以上のバイオマーカーの有無を、１つ以上のバイオマーカーの参照セットと比較することを含む。一部の実施形態において、被検体は、肺疾病の治療を必要とする。一部の実施形態において、被検体は、肺疾病を発症するリスクが増加している疑いがある。一部の実施形態において、被検体は、肺疾病に対して無症状である。一部の実施形態において、被検体は、以前に確定診断を受けたことがない。

【００２３】

[23] 一部の実施形態において、１つ以上のバイオマーカーは、少なくとも５つである。一部の実施形態において、１つ以上のバイオマーカーは、遺伝子又は遺伝子フラグメント、配列変異体、融合体、ミトコンドリア転写物、エピジェネティック修飾、コピー数多型、ヘテロ接合性の消失（ＬＯＨ）、又はこれらの任意の組み合わせのうち１つ以上を含む。一部の実施形態において、１つ以上のバイオマーカーの有無は、発現レベルにより表される。

【００２４】
[24] 一部の実施形態において、方法は、被検体において肺疾病を発症するリスクが増加しているか否かを特定する。一部の実施形態において、（ａ）における特定は、訓練アルゴリズムの採用を含む。一部の実施形態において、訓練アルゴリズムは、個人の気道から採取された上皮細胞を含む訓練セットにより訓練されている。一部の実施形態において、訓練アルゴリズムは、肺疾病に対して良性の試料と、肺疾病に対して悪性の試料とを含む訓練セットにより訓練されている。一部の実施形態において、訓練アルゴリズムは、１つ以上のリスク因子を有する複数の被検体から採取された試料を含む訓練セットにより訓練されている。一部の実施形態において、方法は、被検体の血液分析、被検体に対する電子画像スキャン、又はこれらの任意の組み合わせを更に含む。
【００２５】
[25] 一部の実施形態において、第２の試料は上皮の試料であり、（ｂ）に続いて、上皮組織の試料をアッセイし、１つ以上の追加バイオマーカーの有無を確認する。一部の実施形態において、１つ以上の追加バイオマーカーは、１つ以上のバイオマーカーである。
【００２６】
[26] 本開示の別の態様は、以前に肺疾病の予防又は治療のために介入療法を受けることを推奨されている、肺疾病に罹患している又は肺疾病に罹患する疑いのある被検体をモニタリングする方法を提供する。方法は、（ａ）被検体が肺疾病に対して介入療法の少なくとも一部を受けた後、被検体から採取された上皮組織を含む第１の試料をアッセイして遺伝データを生成すること、（ｂ）遺伝データを処理して肺疾病と関連付けられた１つ以上のバイオマーカーの有無を特定すること、（ｃ）被検体から第２の試料を採取することを推奨することを含むレポートをコンピュータで生成すること、を含む方法を提供する。
【００２７】
[27] 本開示の別の態様は、方法を提供する。方法は、（ａ）以前に肺疾病の予防又は治療のために介入療法を受けることを推奨されている被検体から採取された上皮組織を含む第１の試料をアッセイして１つ以上のバイオマーカーの有無を特定すること、（ｂ）肺疾病に対する介入療法の少なくとも一部を受けるにあたり、被検体が肺疾病に対して介入療法の少なくとも一部を受けた後、被検体から採取された上皮組織を含む第１の試料をアッセイして遺伝データを生成し、被検体から第２の試料を採取して第２の試料に対して（ａ）を繰り返すこと、を含む。
【００２８】
[28] 一部の実施形態において、方法は、介入療法に対する被検体のコンプライアンスを特定する。一部の実施形態において、方法は、肺疾病の予防又は治療に対する介入療法の有効性を特定する。一部の実施形態において、介入療法は、被検体に対する医薬組成物の投与を含む。一部の実施形態において、医薬組成物は化学療法薬を含む。一部の実施形態において、介入療法は、運動療法、食事療法、減煙又は禁煙、又はこれらの任意の組み合わせを含む。
【００２９】
[29] 一部の実施形態において、肺疾病は、肺癌、慢性閉塞性肺疾病（ＣＯＰＤ）、間質性肺疾患（ＩＬＤ）、又はこれらの任意の組み合わせである。一部の実施形態において、肺疾病は肺癌であり、肺癌は、小細胞肺癌、腺癌、扁平上皮癌、大細胞癌、小細胞肺癌、又はこれらの任意の組み合わせを含む。
【００３０】
[30] 一部の実施形態において、第１の試料又は第２の試料は気管支鏡検査により採取される。一部の実施形態において、第１の試料又は第２の試料は穿刺吸引法により採取される。一部の実施形態において、第１の試料又は第２の試料は、粘膜上皮組織、鼻上皮組織、肺上皮組織、又はこれらの任意の組み合わせを含む。一部の実施形態において、第１の試料又は第２の試料は、被検体の気道から採取された上皮組織を含む。
【００３１】
[31] 一部の実施形態において、第２の試料は、第１の試料とは異なる。一部の実施形態

において、第２の試料は、第１の試料とは異なる種類の試料である。一部の実施形態において、第２の試料は第１の試料の採取から時間を置いて被検体から採取されている。一部の実施形態において、時間は、約１～２年である。

【００３２】

[32] 一部の実施形態において、（ａ）は、１つ以上のバイオマーカーの有無を、１つ以上のバイオマーカーの参照セットと比較することを含む。一部の実施形態において、被検体は、肺疾病の治療を必要とする。一部の実施形態において、被検体は、肺疾病を発症するリスクが増加している疑いがある。一部の実施形態において、被検体は、肺疾病に対して無症状である。一部の実施形態において、被検体は、以前に確定診断を受けたことがない。

【００３３】

[33] 一部の実施形態において、１つ以上のバイオマーカーは、少なくとも５つである。一部の実施形態において、１つ以上のバイオマーカーは、遺伝子又は遺伝子フラグメント、配列変異体、融合体、ミトコンドリア転写物、エピジェネティック修飾、コピー数多型、ヘテロ接合性の消失（ＬＯＨ）、又はこれらの任意の組み合わせのうち１つ以上を含む。一部の実施形態において、１つ以上のバイオマーカーの有無は、発現レベルにより表される。

【００３４】

[34] 一部の実施形態において、（ａ）における特定は、訓練アルゴリズムの採用を含む。一部の実施形態において、訓練アルゴリズムは、個人の気道から採取された上皮細胞を含む訓練セットにより訓練されている。一部の実施形態において、訓練アルゴリズムは、肺疾病に対して良性の試料と、肺疾病に対して悪性の試料とを含む訓練セットにより訓練されている。一部の実施形態において、訓練アルゴリズムは、１つ以上のリスク因子を有する複数の被検体から採取された試料を含む訓練セットにより訓練されている。一部の実施形態において、方法は、被検体の血液分析、被検体に対する電子画像スキャン、又はこれらの任意の組み合わせを更に含む。

【００３５】

[35] 一部の実施形態において、（ｂ）は、遺伝データを処理して１つ以上のバイオマーカーのそれぞれに対応する発現レベルを特定することを含む。一部の実施形態において、（ｂ）は、遺伝データを処理し、１つ以上のバイオマーカーにおいて少なくとも１つの遺伝子異常を特定することを含む。

【００３６】

[36] 本開示の別の態様は、肺疾病について被検体をモニタリングする方法であって、（ａ）肺疾病に対する治療を開始している被検体から採取された上皮組織を含む第１の試料をアッセイして１つ以上のバイオマーカーの有無を特定すること、（ｂ）寛解を確定されるにあたり、被検体から第２の試料を採取し、第２の試料で（ａ）を繰り返すこと、を含む方法を提供する。

【００３７】

[37] 一部の実施形態において、方法は、非侵襲的モニタリングにより肺疾病の再発を早期段階で特定する。一部の実施形態において、肺疾病は、肺癌、慢性閉塞性肺疾病（ＣＯＰＤ）、間質性肺疾患（ＩＬＤ）、又はこれらの任意の組み合わせである。一部の実施形態において、肺疾病は肺癌であり、肺癌は、小細胞肺癌、腺癌、扁平上皮癌、大細胞癌、小細胞肺癌、又はこれらの任意の組み合わせを含む。

【００３８】

[38] 一部の実施形態において、第１の試料又は第２の試料は気管支鏡検査により採取される。一部の実施形態において、第１の試料又は第２の試料は穿刺吸引法により採取される。一部の実施形態において、第１の試料又は第２の試料は、粘膜上皮組織、鼻上皮組織、肺上皮組織、又はこれらの任意の組み合わせを含む。一部の実施形態において、第１の試料又は第２の試料は、被検体の気道から採取された上皮組織を含む。

【００３９】

[39] 一部の実施形態において、第２の試料は、第１の試料とは異なる。一部の実施形態において、第２の試料は、第１の試料とは異なる種類の試料である。一部の実施形態において、第２の試料は第１の試料の採取から時間を置いて被検体から採取されている。一部の実施形態において、時間は、約１～２年である。

【００４０】

[40] 一部の実施形態において、（ａ）は、１つ以上のバイオマーカーの有無を、１つ以上のバイオマーカーの参照セットと比較することを含む。一部の実施形態において、被検体は、肺疾病の治療を必要とする。一部の実施形態において、被検体は、肺疾病を発症するリスクが増加している疑いがある。一部の実施形態において、被検体は、肺疾病に対して無症状である。

【００４１】

[41] 一部の実施形態において、１つ以上のバイオマーカーは、少なくとも５つである。一部の実施形態において、１つ以上のバイオマーカーは、遺伝子又は遺伝子フラグメント、配列変異体、融合体、ミトコンドリア転写物、エピジェネティック修飾、コピー数多型、ヘテロ接合性の消失（ＬＯＨ）、又はこれらの任意の組み合わせのうち１つ以上を含む。一部の実施形態において、１つ以上のバイオマーカーの有無は、発現レベルにより表される。

【００４２】

[42] 一部の実施形態において、（ａ）における特定は、訓練アルゴリズムの採用を含む。一部の実施形態において、訓練アルゴリズムは、個人の気道から採取された上皮細胞を含む訓練セットにより訓練されている。一部の実施形態において、訓練アルゴリズムは、肺疾病に対して良性の試料と、肺疾病に対して悪性の試料とを含む訓練セットにより訓練されている。一部の実施形態において、訓練アルゴリズムは、１つ以上のリスク因子を有する複数の被検体から採取された試料を含む訓練セットにより訓練されている。一部の実施形態において、方法は、被検体の血液分析、被検体に対する電子画像スキャン、又はこれらの任意の組み合わせを更に含む。本開示の別の態様は、肺疾病に罹患している又は肺疾病に罹患する疑いのある被検体をモニタリングする方法であって、（ａ）肺疾病に罹患する疑いがあり、１つ以上の肺結節について陰性の徴候を示す被検体から採取された上皮組織を含む第１の試料をアッセイして、肺疾病と関連付けられた１つ以上のバイオマーカーの有無を特定すること、（ｂ）１つ以上のバイオマーカーの有無を特定するにあたり、（ｉ）被検体から第２の試料を採取し、又は、（ｉｉ）（ａ）の結果に基づき被検体の肺領域の電子画像スキャンを取得するように被検体に指示すること、を含む方法を提供する。一部の実施形態において、方法は、（ａ）の前に、被検体に対してコンピュータ分析を行い、被検体から肺疾病を発症する１つ以上のリスク因子の存在を特定することを更に含む。

【００４３】

[43] 本開示の別の態様は、肺疾病について被検体をスクリーニングするシステムを提供する。システムは、（ｉ）被検体に対して肺疾病を発症する１つ以上のリスク因子の存在についてヘルスデータ又は生理データを分析し、（２）１つ以上のリスク因子を特定するにあたり、被検体の試料からの上皮組織に対して肺疾病を発症するリスクと関連付けられた１つ以上のバイオマーカーに関するアッセイの推奨を生成するように、個別又は集合的にプログラムされた、１つ以上のコンピュータプロセッサとを含む。

【００４４】

[44] 本開示の別の態様は、肺疾病について被検体をスクリーニングするシステムを提供する。システムは、（ｉ）被検体において肺疾病を発症する１つ以上のリスク因子の存在を示すデータを含む第１のデータセットと、（ｉｉ）被検体の試料における上皮組織に対して１つ以上のリスク因子の有無を示すデータを含む第２のデータセットとを含み、１つ以上のバイオマーカーは肺疾病を発症するリスクと関連付けられる、１つ以上のコンピュータデータベースと、（ｉ）第１のデータセットを分析して１つ以上のリスク因子の存在を特定し、（ｉｉ）第２のデータセットを分析して１つ以上のバイオマーカーの存在を特

定し、（ｉｉｉ）１つ以上のバイオマーカーの有無を特定するにあたり、（１）被検体において肺疾病に罹患する疑いのある肺領域の電子画像スキャンを取得するように指示する、又は（２）被検体の第２の試料から得た他の上皮組織をアッセイするように指示するレポートを生成するように、個別又は集合的にプログラムされた、１つ以上のコンピュータプロセッサとを含む。

【００４５】

[45] 本開示の別の態様は、肺疾病に罹患している又は肺疾病に罹患する疑いのある被検体をモニタリングするシステムを提供する。システムは、被検体の第１の試料における上皮組織において１つ以上のバイオマーカーの有無を示すデータを含み、１つ以上のバイオマーカーは肺疾病と関連付けられる、１つ以上のコンピュータデータベースと、（ｉ）被検体が以前に１つ以上の肺結節の存在について陽性の徴候を示していたと判定し、（ｉｉ）（ｉ）に続いて、データセットを処理して１つ以上のバイオマーカーの有無を特定し、（ｉｉｉ）１つ以上のバイオマーカーの有無を特定するにあたり、（１）被検体から第２の試料を採取するように指示する、又は（２）被検体の肺領域の電子画像スキャンを更に取得するように指示するレポートを生成するように、個別又は集合的にプログラムされた、１つ以上のコンピュータプロセッサとを含む。

【００４６】

[46] 本開示の別の態様は、以前に肺疾病の予防又は治療のために介入療法を受けることを推奨されている、肺疾病に罹患している又は肺疾病に罹患する疑いのある被検体をモニタリングするシステムを提供する。システムは、遺伝データを含むデータセットを含む１つ以上のコンピュータデータベースと、（ｉ）被検体が肺疾病に対して介入療法の少なくとも一部を受けた後、遺伝データを処理して肺疾病と関連付けられる１つ以上のバイオマーカーの有無を特定し、（ｉｉｉ）被検体から第２の試料を採取することを推奨することを含むレポートを生成するように、個別又は集合的にプログラムされた、１つ以上のコンピュータプロセッサとを含む。

【００４７】

[47] 本開示の別の態様は、１つ以上のコンピュータプロセッサにより実行されたとき、本明細書に記載の方法のいずれかを実現する機械実行可能型コードを含む持続性コンピュータ可読媒体を提供する。

【００４８】

[48] 本開示の別の態様は、１つ以上のコンピュータプロセッサとそれに結合されたメモリとを含むコンピュータシステムを提供する。メモリは、１つ以上のコンピュータプロセッサにより実行されたとき、本明細書に記載の方法のいずれかを実現する機械実行可能型コードを含む持続性コンピュータ可読媒体を含む。

【００４９】

[49] 当業者であれば、本開示におけるその他の態様や利点について、以下の詳細な説明から容易に理解するはずである。ただし、以下に記載された本開示の実施形態は説明のために提示したものであり、他の実施形態を以って本開示を実現することもできる。本開示の趣旨から逸脱しない限り、修正や変更は全て許容範囲である。従って、図面及び説明は、本質的に例示に過ぎず、本開示を限定するものでない。

【００５０】

（参照による援用）

[50] 本明細書に言及される刊行物、特許、出願は全て、それぞれ全文が本明細書に援用されていると見なされる。援用されている刊行物、特許、出願の内容が本明細書の開示内容と矛盾する場合は、本明細書の内容を最優先とする。

【図面の簡単な説明】

【００５１】

[51] 本発明の新たな特徴は、下記の請求の範囲に詳細に記載されている。本発明の特徴及び利点をより理解できるように、本発明の原理を適用した実施形態を例示する詳細な説明、及び添付の図面を参照して説明する。本明細書の図面は、次の通りである。

【００５２】
【図１】[52] 肺癌診断における臨床的課題を示す。
【図２】[53] ゲノム分類分析を含む方法を肺癌の臨床治療経路に組み込んだ場合の利点を示す。
【図３】[54] ゲノム分類分析を含むことで改善された臨床決定経路を示す。
【図４】[55] ゲノム分類分析を含む方法を臨床治療経路に組み込むことによる、処置推奨の４７％低減を示す。
【図５】[56] 特発性肺線維症（ＩＰＦ）の臨床治療経路にゲノム分類分析を含む方法を組み込むことから得られる利益を示す。
【図６】[57] 通常の間質性肺炎（ＵＩＰ）を他の間質性肺疾患（ＩＬＤ）と区別するために臨床治療経路にゲノム分類分析を組み込むことにより治療決定に生じた肯定的変化を示す。
【図７】[58] 共通の経路を共有する損傷範囲の病因を示す。
【図８】[59] 被検体における発癌範囲と損傷範囲の差の例を示す。
【図９】[60] 発癌範囲と損傷範囲の分子的概念を示す。
【図１０】[61] ゲノム分類分析（気管支ゲノム分類器）を含むことで改善された、肺癌の臨床標準治療経路を示す。
【図１１Ａ－Ｂ】[62] 複数のゲノム分類器分析（気管支ゲノム分類器：Nasa-Detect、Nasa-RiskStratifier、Nasa-ProtectMonitoring、Nasa-Recurrence）を含めることにより、肺癌の改善された臨床治療経路を示す。
【図１２】[63] Nasa-Detect分類器のテスト特徴を示す。
【図１３】[64] Nasa-Risk階層化分類器のテスト特徴を示す。
【図１４】[65] Nasa-PROTECT分類器のテスト特徴を示す。
【図１５】[66] Nasa-Recurrence分類器のテスト特徴を示す。
【図１６】[67] 実演と予防におけるゲノミクスの評価を示す。
【図１７】[68] 本明細書に記載の方法で使用される試料特性と試料種の例を示す。
【図１８】[69] 鼻／気管支の擦過試料を含む、様々な被検体コホートを示す。
【図１９】[70] ゲノム分類器（Nasa-Detect分類器）の訓練に用いられる訓練用試料を示す。
【図２０】[71] ゲノム分類器（Nasa-Risk階層化分類器）の訓練に用いられる訓練用試料を示す。
【図２１】[72] バイオマーカーの種類、及び、様々な種類のバイオマーカーの検出に使用されるテクノロジープラットフォームを示す。
【図２２】[73] ゲノム分類器におけるＲＮＡシーケンシングの例を示す。
【図２３】[74] ＲＮＡシーケンシングの例を示す。
【図２４】[75] 訓練アルゴリズムを含むゲノム分類器の訓練と検証のフロー図を示す。
【図２５】[76] ゲノム分類器の訓練に使用される訓練セットで用いられる多様な細胞サブタイプ・組織学的サブタイプの例を示す。
【図２６】[77] 本明細書で提供される方法を実装するようにプログラム又は構成されたコンピュータ制御システムを示す。
【図２７】[78] 機械学習アプリケーションの課題と解決策を示す。
【図２８】[79] ＩＬＤ患者における一般の間質性肺炎（ＵＩＰ）パターンを予測する分子ゲノム分類器の開発・評価における分析パイプラインを示す。
【図２９】[80] DESeq2とvolcanoプロットを用いた分類器を使用した遺伝子選択による、DESeq2により選択された１５１個の遺伝子（調整済みｐ値＜０．０５、倍率変化＞２）と、分類器の１９０の予測遺伝子、そして両セットにおける３２個の共通遺伝子を示す。
【図３０】[81] DESeq2選択遺伝子のみを用いた全ての経気管支生検（ＴＢＢ）試料の主成分分析（ＰＣＡ）プロットを適用した、DESeq2と分類器による遺伝子選択（ＵＩＰ以外の試料（十字）とＵＩＰ試料（丸）を区別できていない）を示す。
【図３１】[82] 分類器による遺伝子を使用した全てのＴＢＢ試料のＰＣＡプロットを適

用した、DESeq2と分類器による遺伝子選択（ＵＩＰ以外の試料（十字）とＵＩＰ試料（丸）に分類する）を示す。
【図３２】[83] 患者内でのin silico及びin vitro混合の比較を示す。Ｒ平方値が０．９９のEnsemble分類器による、in silicoとin vitroの混合比較の散布図である。
【図３３】[84] 患者内でのin silico及びin vitro混合の比較を示す。Ｒ平方値が０．９８の罰則付きロジスティック回帰分類器により得られたin silicoとin vitroの混合比較の散布図である。
【図３４】[85] Ensembleモデルの分類スコアを示す。諧調の異なる灰色は、組織病理ＵＩＰ、非ＵＩＰ、非診断の試料を区別する。丸、上向きの三角形、正方形、下向きの三角形はそれぞれ、in silico混合試料、上葉、中葉、下葉の試料を示す。
【図３５】[86] １人患者抜き交差検証を行った場合の罰則付きロジスティック回帰モデルの分類スコアを示す。諧調の異なる灰色は、組織病理ＵＩＰ、非ＵＩＰ、非診断の試料を区別する。丸、上向きの三角形、正方形、下向きの三角形はそれぞれ、in silico混合試料、上葉、中葉、下葉の試料を示す。
【図３６Ａ－Ｂ】[87] １人患者抜き交差検証（LOPO CV）及び独立テストセットの検証（テスト）からの受信者動作特性（ＲＯＣ）曲線を示す。各ＲＯＣ曲線上の小惑星は、各提案モデルの前向きに定義された決定境界に対応している。
【図３７】[88] １人患者抜き交差検証と、独立テストセットでの検証による分類性能を示す。
【図３８】[89] 複数の試料を含む代表患者６人のデータにおける患者内・患者間の異質性を示す相関行列のヒートマップを示す。
【図３９】[90] 非ＵＩＰ亜型とＵＩＰ試料を比較して選択した遺伝子を使用したＰＣＡプロットを示す。ＵＩＰ試料（丸）と呼吸細気管支炎（ＲＢ）を比較して有意な差次的発現遺伝子を使用した、全ての訓練用試料のＰＣＡにおける最初の２つの主成分である。
【図４０】[91] 非ＵＩＰ亜型とＵＩＰ試料を比較して選択した遺伝子を使用したＰＣＡプロットを示す。ＵＩＰ試料（丸）と細気管支炎を比較して有意な差次的発現遺伝子を使用した、全ての訓練用試料のＰＣＡにおける最初の２つの主成分である。
【図４１】[92] 非ＵＩＰ亜型とＵＩＰ試料を比較して選択した遺伝子を使用したＰＣＡプロットを示す。ＵＩＰ試料（丸）と過敏性肺炎（ＨＰ）を比較して有意な差次的発現遺伝子を使用した、全ての訓練用試料のＰＣＡにおける最初の２つの主成分である。
【図４２】[93] 非ＵＩＰ亜型とＵＩＰ試料を比較して選択した遺伝子を使用したＰＣＡプロットを示す。ＵＩＰ試料（丸）と非特異性間質性肺炎（ＮＳＩＰ）を比較して有意な差次的発現遺伝子を使用した、全ての訓練用試料のＰＣＡにおける最初の２つの主成分である。
【図４３】[94] 非ＵＩＰ亜型とＵＩＰ試料を比較して選択した遺伝子を使用したＰＣＡプロットを示す。ＵＩＰ試料（丸）と組織化肺炎（ＯＰ）を比較して有意な差次的発現遺伝子を使用した、全ての訓練用試料のＰＣＡにおける最初の２つの主成分である。
【図４４】[95] 非ＵＩＰ亜型とＵＩＰ試料を比較して選択した遺伝子を使用したＰＣＡプロットを示す。ＵＩＰ試料（丸）とサルコイドーシスを比較して有意な差次的発現遺伝子を使用した、全ての訓練用試料のＰＣＡにおける最初の２つの主成分である。
【図４５】[96] 遺伝子発現の変動を示す。濃い灰色の点は、訓練分類から削除された遺伝子である。
【図４６Ａ－Ｂ】[97] Ensembleモデル（図４６Ａ）及び罰則付きロジスティック回帰モデル（図４６Ｂ）の訓練セットを使用したin silico混合サンプルの閾値対感度／特異度を示す。
【図４７Ａ－Ｃ】[98] Ensembleモデルのスコア変動シミュレーションを示す。スコアの変動性の最終閾値である０．９０は、図４７Ａの特異度（縦の点線）で表している。感度（１．８０）及び反転率（１．１５）のスコア変動の個々の閾値は、図４７Ｂ及び図４７Ｃの縦の点線で示される。
【図４８Ａ－Ｃ】[99] 罰則付きロジスティック回帰モデルのスコア変動シミュレーショ

ンを示す。スコアの変動性の最終閾値である０．４８は、図４８Ａの特異度（垂直線）で表している。感度（０．７８）及び反転率（０．６８）のスコア変動性の個々の閾値は、図４８Ｂ及び４８Ｃの灰色の縦線で示される。

【発明を実施するための形態】

【００５３】

[100] 本発明の様々な実施形態が本明細書に示され説明されたが、そのような実施形態が例示に過ぎないことは当業者には明らかであろう。本発明から逸脱することなく、多数の変形、変更、及び置換が当業者には思いつくであろう。本明細書に記載された本発明の実施形態の様々な代替物を使用できることを理解されたい。

【００５４】

[101] 本明細書における「癌」とは、一般に、異常な細胞増殖の状態を指す。癌には、固形腫瘍又は循環癌細胞が含まれてもよい。癌は転移してもよい。癌は組織特異的な癌であってもよい。癌は肺癌であってもよい。癌は悪性又は良性であってもよい。

【００５５】

[102] 本明細書における「肺癌」とは、一般に、肺又は肺関連組織における癌又は腫瘍を指す。例えば、肺癌は、非小細胞肺癌、小細胞肺癌、肺カルチノイド腫瘍、又はこれらの任意の組み合わせを含んでもよい。非小細胞肺癌は、腺癌、扁平上皮癌、大細胞癌、又はこれらの任意の組み合わせを含んでもよい。肺カルチノイド腫瘍は気管支カルチノイドを含んでもよい。肺癌は、気管支、上皮細胞、平滑筋細胞、肺胞、又はこれらの任意の組み合わせなどの肺組織の癌を含んでもよい。肺癌は、気管、気管支、終末細気管支、又はこれらの任意の組み合わせの癌を含んでもよい。肺癌は、基底細胞、杯細胞、繊毛細胞、神経内分泌細胞、線維芽細胞、マクロファージ細胞、クララ細胞、又はこれらの任意の組み合わせを含んでもよい。

【００５６】

[103] 本明細書における「疾患（disease）又は疾病（condition）」とは、一般に、異常状態又は病的状態を指す。疾患又は疾病は、肺疾患又は肺疾病であってもよい。肺疾患又は肺疾病は、肺癌、間質性肺疾患（ＩＬＤ）、慢性閉塞性肺疾患（ＣＯＰＤ）、慢性気管支炎、嚢胞性線維症、喘息、肺気腫、肺炎、結核、肺水腫、急性呼吸窮迫症候群、又は塵肺症を含んでもよい。ＩＬＤの種類には、特発性肺線維症、非特異的間質性肺炎、落屑性間質性肺炎、呼吸細気管支炎、急性間質性肺炎、リンパ性間質性肺炎、又は原因不明の組織化肺炎が含まれる。

【００５７】

[104] 本明細書における「間質性肺疾患（ＩＬＤ）」とは、一般に、間質性肺組織の疾患を指す。ＩＬＤは、間質性肺炎、特発性肺線維症、非特異的間質性肺炎、過敏性肺炎、極低温組織化肺炎（ＣＯＰ）、急性間質性肺炎、落屑性間質性肺炎、サルコイドーシス、石綿肺、又はこれらの任意の組み合わせを含んでもよい。

【００５８】

[105] 低線量断層撮影（ＣＴ）スキャン（ＬＤＣＴ）は、一般に、被検体への照射線量を減らす画像診断法を指す。例えば、ＬＤＣＴの照射線量は、約１．５ミリシーベルト（ｍＳｖ）未満であってもよい。ＬＤＣＴの照射線量は、約５ｍＳｖ未満、約４ｍＳｖ未満、約３ｍＳｖ未満、約２ｍＳｖ未満、約１ｍＳｖ未満、約０．５ｍＳｖ未満、約０．１ｍＳｖ未満、又はそれ以下であってもよい。ＬＤＣＴの照射線量は、約１．０ｍＳｖ～約２．０ｍＳｖであってもよい。ＬＤＣＴの照射線量は、約０．５ｍＳｖ～約１．５ｍＳｖであってもよい。ＬＤＣＴの照射線量は、約１．０ｍＳｖ～約４．０ｍＳｖであってもよい。ＬＤＣＴの照射線量は、約１．０ｍＳｖ～約３．０ｍＳｖであってもよい。ＬＤＣＴの管電流設定は、約４０ｍＡｓ未満、約３５ｍＡｓ未満、約３０ｍＡｓ未満、約２５ｍＡｓ未満、約２０ｍＡｓ未満、約１５ｍＡｓ未満、約１０ｍＡｓ未満、約５ｍＡｓ未満、約１ｍＡｓ未満、又はそれ以下であってもよく、その場合でも十分な画像品質を得ることができる。ＬＤＣＴの管電流設定は、約２０ｍＡｓ～約４０ｍＡｓであってもよい。ＬＤＣＴの管電流設定は、約２０ｍＡｓ～約５０ｍＡｓであってもよい。ＬＤＣＴの管電流設定は

、約２０ｍＡｓ～約８０ｍＡｓであってもよい。ＬＤＣＴの管電流設定は、約２０ｍＡｓ～約１００ｍＡｓであってもよい。

【００５９】

[106] 中線量ＣＴスキャンの照射線量は、約１ｍＳｖ以上、約５ｍＳｖ以上、約６ｍＳｖ以上、約７ｍＳｖ以上、約８ｍＳｖ以上、約９ｍＳｖ以上、約１０ｍＳｖ以上、約１５ｍＳｖ以上、又はそれ以上であってもよい。中線量ＣＴスキャンの照射線量は、約８ｍＳｖであってもよい。中線量ＣＴスキャンの照射線量は、約７ｍＳｖ～約１０ｍＳｖであってもよい。中線量ＣＴスキャンの照射線量は、約１ｍＳｖ～約１０ｍＳｖであってもよい。中線量ＣＴスキャンの照射線量は、約５ｍＳｖ～約１０ｍＳｖであってもよい。中線量ＣＴスキャンの照射線量は、約１ｍＳｖ～約５ｍＳｖであってもよい。中線量ＣＴの管電流設定は、約１００ｍＡ以上、約１２５ｍＡｓ以上、約１５０ｍＡｓ以上、約１７５ｍＡｓ以上、約２００ｍＡｓ以上、約２２５ｍＡｓ以上、約２５０ｍＡｓ以上、約３００ｍＡｓ以上、約３５０ｍＡｓ以上、約４００ｍＡｓ以上、約５００ｍＡｓ、又はそれ以上であってもよい。中線量ＣＴの管電流設定は、約２００ｍＡｓ～約２５０ｍＡｓであってもよい。中線量ＣＴスキャンの管電流設定は、約１５０ｍＡｓ～約２５０ｍＡｓであってもよい。中線量ＣＴスキャンの管電流設定は、約１００ｍＡｓ～約３００ｍＡｓであってもよい。中線量ＣＴスキャンの管電流設定は、約１００ｍＡｓ～約２００ｍＡｓであってもよい。中線量ＣＴスキャンの管電流設定は、約１５０ｍＡｓ～約３００ｍＡｓであってもよい。中線量ＣＴスキャンの管電流設定は、約１５０ｍＡｓ～約４００ｍＡｓであってもよい。

【００６０】

[107] 本明細書における「相同性」とは、一般に、最適な比較を目的として、配列をアラインすることで判定できる、２つ以上のヌクレオチド配列又はアミノ酸配列の間の「相同性」又は「相同率」を算出することを指す（例えば、第１の配列において配列間ギャップを導入することができる）。次に、対応する位置のヌクレオチドを比較し、２つの配列間の同一性％は、配列が共有する同一位置の数に対して求める（つまり、相同率＝同一位置の数／位置の総数×１００）。例えば、第１の配列の位置と、第２の配列の対応位置が同じヌクレオチドを有していると、この位置における分子も同一になる。両配列を最適にアラインさせるために導入したギャップの数や各ギャップの長さを考慮して、両配列が共有する同一位置の数を以って両配列間の相同率を求める。一部の実施形態において、比較のためにアラインされる配列の長さは、参照配列の長さの少なくとも約３０％、少なくとも約４０％、少なくとも約５０％、少なくとも約６０％、少なくとも約６５％、少なくとも約７０％、少なくとも約７５％、少なくとも約８０％、少なくとも約８５％、少なくとも約９０％、少なくとも約９１％、少なくとも約９２％、少なくとも約９３％、少なくとも約９４％、少なくとも約９５％、少なくとも約９６％、少なくとも約９７％、少なくとも約９８％、又は少なくとも約９５％となる。一部において、配列相同性は約７０％～１００％であってもよい。一部において、配列相同性は約８０％～１００％であってもよい。一部において、配列相同性は約９０％～１００％であってもよい。一部において、配列相同性は約９５％～１００％であってもよい。一部において、配列相同性は約７０％～９９％であってもよい。一部において、配列相同性は約８０％～９９％であってもよい。一部において、配列相同性は約９０％～９９％であってもよい。一部において、配列相同性は約９５％～９９％であってもよい。ＢＬＡＳＴ（登録商標）サーチを用いて両配列間の相同性を判定してもよい。両配列は、遺伝子、ヌクレオチド配列、タンパク質配列、ペプチド配列、アミノ酸配列、又はそれらのフラグメントである。実際に両配列を比較する場合は、例えば、数学アルゴリズムを使用するなど、既知の方法を用いることができる。数学アルゴリズムの例としてKarlin, S. and Altschul, S., Proc. Natl. Acad. Sci. USA, 90- 5873-5877 (1993)が挙げられるが、これに限定されない。アルゴリズムは、Altschul, S. et al., Nucleic Acids Res., 25:3389-3402 (1997)に記載のNBLASTやXBLASTプログラム（バージョン２．０）に組み込まれている。BLASTやGappedBLASTプログラムを用いるときは、各プログラムの関連パラメータ（例えば、ＮＢＬＡＳＴ）を使用することがで

きる。例えば、配列比較のパラメータは、スコア＝１００、ワード長＝１２に設定してもよく、変更してもよい（Ｗ＝５又はＷ＝２０など）。他の例は、Myers and Miller, CABIOS (1989), ADVANCE, ADAM, BLAT, and FASTAのアルゴリズムを含む。他の実施例体において、両アミノ配列間の同一性％は、ＧＣＧソフトウェアパッケージのＧＡＰプログラム（イギリス国ケンブリッジ、アクセルリス）を用いて得ることができる。

【００６１】

[108] 本明細書における「フラグメント」とは、一般に、配列の一部であって、完全長配列よりも短いサブセットを指す。フラグメントは遺伝子の一部であってもよい。フラグメントはペプチドやタンパク質の一部であってもよい。フラグメントはアミノ酸配列の一部であってもよい。フラグメントはオリゴヌクレオチド配列の一部であってもよい。フラグメントの長さは、約２０個未満、約３０個未満、約４０個未満、又は約５０個未満のアミノ酸であってもよい。フラグメントの長さは、約２０個未満、約３０個未満、約４０個未満、又は約５０個未満のヌクレオチドあってもよい。フラグメントの長さは、約１０個～約５０個未満のアミノ酸であってもよい。フラグメントの長さは、約１０個～約４０個のアミノ酸であってもよい。フラグメントの長さは、約１０個～約３０個のアミノ酸であってもよい。フラグメントの長さは、約１０個～約２０個のアミノ酸であってもよい。フラグメントの長さは、約２０個～約５０個のアミノ酸であってもよい。フラグメントの長さは、約３０個～約５０個のアミノ酸であってもよい。フラグメントの長さは、約４０個～約５０個のアミノ酸であってもよい。フラグメントの長さは、約１０個～約５０個のヌクレオチドであってもよい。フラグメントの長さは、約１０個～約４０個のヌクレオチドであってもよい。フラグメントの長さは、約１０個～約３０個のヌクレオチドであってもよい。フラグメントの長さは、約１０個～約２０個のヌクレオチドであってもよい。フラグメントの長さは、約２０個～約５０個のヌクレオチドであってもよい。フラグメントの長さは、約３０個～約５０個のヌクレオチドであってもよい。フラグメントの長さは、約４０個～約５０個のヌクレオチドであってもよい。

【００６２】

[109] 本明細書における「被検体」とは、一般に、疾病（例えば、肺疾患）に罹患しているか、罹患している疑いがある個人を指す。被検体は動物であってもよい。動物は、ヒト、非ヒト霊長類、マウス又はラットなどのげっ歯類、イヌ、ネコ、ブタ、ヒツジ、又はウサギなどの哺乳動物であってもよい。動物は、魚類、爬虫類、又はその他の動物であってもよい。動物は、新生児、幼児、青年、成体であってもよい。被検体は微生物(living organism)であってもよい。被検体はヒトであってもよい。ヒトは、１歳以上、２歳以上、５歳以上、１０歳以上、２０歳以上、３０歳以上、４０歳以上、５０歳以上、６０歳以上、６５歳以上、７０歳以上、７５歳以上、８０歳以上、又はそれ以上であってもよい。ヒトは、約１８歳～約９０歳であってもよい。ヒトは、約１８歳～約３０歳であってもよい。ヒトは、約３０歳～約５０歳であってもよい。ヒトは、約５０歳～約９０歳であってもよい。被検体は、疾病に対するリスク因子を１つ以上有し、無症状であってもよい。被検体は、疾病に対して無症状であってもよい。被検体は、疾病に対するリスク因子を１つ以上有していてもよい。被検体は、疾病の症候を示していてもよい。被検体は、疾病の症候を示し、疾病に対するリスク因子を１つ以上有していてもよい。被検体は、癌や腫瘍などの疾患に罹患しているか、罹患している疑いがあってもよい。被検体は、疾患の治療を受けている患者、例えば癌患者、腫瘍患者、癌・腫瘍患者であってもよい。被検体は、癌や腫瘍などの疾患を発症するリスクを有しやすい傾向があってもよい。被検体は、癌や腫瘍などの疾患に対して寛解に至っていてもよい。被検体には癌に罹患していないか、腫瘍に罹患していないか、癌と腫瘍の両方に罹患していなくてもよい。被検体は健康であってもよい。

【００６３】

[110] 本明細書における「組織試料」とは、一般に、被検体から得たいずれかの組織試料を指す。組織試料は、気道の一部から得られた上皮細胞など、気道の一部から得られた細胞を含んでもよい。組織試料は、鼻組織、気管支組織、肺組織、食道組織、喉頭組織、

口腔組織、又はこれらの任意の組み合わせであってもよい。組織試料は、癌や腫瘍などの疾患又は疾病が疑われる又は確定されている試料であってもよい。組織試料は、組織擦過、拭き取り、組織生検、組織切除、穿刺吸引法、組織洗浄、細胞診標本、気管支鏡検査、又はこれらの任意の組み合わせなど、被検体から取り除いた試料であってもよい。組織試料は、穿刺吸引法、気管支鏡検査、又は少量の試料を採取するその他の方法によって採取された試料など、不明確な又は疑わしい試料であってもよい。組織試料は、放射線などの抗癌治療を受けている患者の体において無損傷の部位から採取されたものであってもよい。組織試料は、患者の体における腫瘍であってもよい。組織試料は、癌性細胞、腫瘍細胞、非癌性細胞、又はこれらの任意の組み合わせを含んでもよい。組織は、侵襲性細胞、非侵襲性細胞、又はこれらの任意の組み合わせを含んでもよい。組織試料は、鼻組織、気管組織、肺組織、咽頭組織、喉頭組織、気管支組織、胸膜組織、肺細胞組織、乳房組織、膀胱組織、腎臓組織、肝臓組織、結腸組織、甲状腺組織、子宮頸部組織、前立腺組織、心臓組織、筋肉組織、膵臓組織、肛門組織、胆管組織、骨組織、子宮組織、卵巣組織、子宮内膜組織、膣組織、外陰部組織、胃組織、眼組織、副鼻腔組織、陰茎組織、唾液腺組織、腸組織、胆嚢組織、胃腸組織、膀胱組織、脳組織、脊髄組織、血液試料、又はこれらの任意の組み合わせであってもよい。

【００６４】

[111] 本明細書において、肺疾病の発症又は罹患という文脈における「リスクの増加」とは、一般に、被検体における肺疾病の発生に関連付けられたリスク又は確率の増加を指す。肺疾病を発症するリスクの増加には、被検体における疾病の最初の発生、又は、２回目、３回目、４回目、又はそれ以降の発生など、後続の発生を含む。肺疾病を発症するリスクの増加には、ａ）最初に疾病を発症するリスク、ｂ）再発、又は再度疾病を発症するリスク、ｃ）将来疾病を発症するリスク、ｄ）被検体の生涯を通して疾病を発症する傾向があるリスク、又は、ｅ）幼児、青年、又は成人のときに疾病を発症する傾向があるリスクが含まれる。肺疾病が発生又は再発し得るリスクの増加には、疾病（癌など）が転移し得るリスクが含まれる。腫瘍又は癌が発生又は再発し得るリスクの増加には、Ｉ期癌、ＩＩ期癌、III期癌、又はＩＶ期癌が発生し得るリスクが含まれる。腫瘍又は癌が発生又は再発し得るリスクには、血液癌、組織癌（腫瘍など）、又は他の部位から１つ以上の臓器部位に転移するリスクが含まれる。

【００６５】

[112] 本明細書における「介入療法又は治療計画の有効性」とは、一般に、介入療法又は治療計画が意図していた結果を達成しているか否かに関する評価、又は判定を指す。例えば、抗癌剤の投与などを含む治療計画の有効性は、腫瘍又は癌細胞の侵襲性を低下させる、被検体の癌又は腫瘍細胞を殺すか排除する、疾患の進行を止める、疾患の発症を予防する抗癌剤の効能を評価する。治療計画は、手術（つまり、外科切除）、栄養療法、運動、放射線、化学療法、細胞移植、輸血などを含んでもよい。介入療法には、医薬組成物、運動療法、食事療法、１つ以上のリスク因子（喫煙や受動喫煙など）の減少又は除去、又はこれらの任意の組み合わせを被検体に適用することを含んでもよい。

【００６６】

[113] 図１に示すように、年間約２２５，０００人を超える患者が新たに肺癌の確定診断を受けている。肺癌と新たに診断された被検体の約９０％は、過去に喫煙歴がある被検体である。肺癌は年間約１６０，０００の死者を出す。本明細書に記載されているような新たな方法、システム、キットを開発すると、肺癌の早期発見や肺癌を発症するリスクの増加を改善することができる。早期発見は、死亡率を低減するための重要な鍵である。更に、現在の臨床基準では、手術や肺生検など、高リスク高費用の侵襲的処置を採用せず肺癌を正確に診断することは困難である。現行の臨床標準治療の一環として侵襲的肺生検を受ける被検体の約４０％に癌はない。従って、本明細書に記載されているような新たな方法、システム、キットは、肺癌の早期発見と診断の精度を改善する一方、（リスクと追加費用を伴う）不要な侵襲的処置の数を減らすこともできる。

【００６７】

[114] 図２に示すように、現在の臨床診療基準における様々な決定ポイントにゲノム分類器を組み込むことで、不要な侵襲性手法の数を減らし、肺癌のリスクが低い被験体を特定することができる。例えば、米国の場合、画像スキャンで偶然に肺結節を検出するケースが毎年約１８０万～２００万件に達する。現在の臨床基準では、画像スキャンで結節が検出された被検体は、結節が肺癌の存在を意味するものであるか否かを判定するために、侵襲性気管支鏡検査を受けなければならない。約１４０，０００人の被検体（又は、気管支鏡検査を受けた被検体３５０，０００人のうち約６０～７０％）が、不明確な又は疑わしい結果を受ける。現在の臨床基準によると、気管支鏡検査で不明確な又は疑わしい結果が得られた場合、診断手術を行って組織病理的に判断するしかない。しかしながら、不明確な又は疑わしい結果を受けた被検体の約７０～８０％は、結局、肺組織が良性であるとの判定を受ける。従って、本明細書に記載の新たな方法、システム、キットは、不明確な又は疑わしい結果の後に１つ以上のゲノム分類器で更に分析を行い、肺癌を発症するリスクが低い被検体を特定し、現在の臨床治療基準を改善することができる。肺癌に対して高いリスクを有する被検体、又はリスクが増加している被検体にのみ診断手術を行うと、低リスク集団に不要の侵襲的処置を適用せずともよい。

【００６８】

[115] 図３は、本明細書に記載の気管支ゲノム分類器に追加／改善を伴う現在の臨床標準治療を示す。一般的な成人集団から、肺癌のリスクがあると特定された個人は、低線量ＣＴスキャンなどの画像スキャンを受けることになる。結節が見つからない場合、間を置いて画像スキャンを再度行ってもよい。結節が見つかった場合、被検体は、リスクアセスメント、ＣＴスキャン、ＰＥＴスキャン、磁気共鳴画像（ＭＲＩ）スキャン、Ｘ線、又はこれらの任意の組み合わせを受ける。現在、米国では低線量ＣＴスキャンはあまり採用されていない。リスクアセスメント、ＣＴスキャン、ＰＥＴスキャン、磁気共鳴画像（ＭＲＩ）スキャン、Ｘ線、又はこれらの任意の組み合わせにより、被検体に肺癌のリスクが低いと特定された場合、間を置いてリスクアセスメント、ＣＴスキャン、ＰＥＴスキャン、磁気共鳴画像（ＭＲＩ）スキャン、Ｘ線、又はこれらの任意の組み合わせを再度行ってもよい。リスクアセスメント、ＣＴスキャン、ＰＥＴスキャン、磁気共鳴画像（ＭＲＩ）スキャン、Ｘ線、又はこれらの任意の組み合わせにより、被検体に肺癌のリスクが中又は高と特定された場合、被検体は、気管支鏡検査、胸腔穿刺法（ＴＴＮＡ）、胸腔鏡補助下手術（ＶＡＴＳ）、気道組織試料の採取法、又はこれらの任意の組み合わせの施術を受けることになる。採取された気道試料が不明確又は疑わしいと特定された場合、気管支ゲノム分類器を実行して肺癌のリスクを特定する。気管支ゲノム分類器により被検体に肺癌のリスクが低いと特定された場合、リスクアセスメント、ＣＴスキャン、ＰＥＴスキャン、磁気共鳴画像（ＭＲＩ）スキャン、Ｘ線、又はこれらの任意の組み合わせを再度行ってもよい。気管支ゲノム分類器により被検体に肺癌のリスクが中間程度であると特定された場合、気管支鏡検査、胸腔穿刺法（ＴＴＮＡ）、胸腔鏡補助下手術（ＶＡＴＳ）、気道組織試料の採取法、又はこれらの任意の組み合わせを再度行ってもよい。気管支鏡検査の試料は不明確又は疑わしいものであり得る。気管支鏡検査で採取された試料は、高い割合で不明確又は疑わしいものであり得る。従って、現行の臨床標準治療に気管支ゲノム分類器を追加すると、不明確な又は疑わしい結果の数を大幅に減らすことができる。被検体が肺癌に罹患していると特定された場合、被検体は肺癌の治療を受けながら、画像診断法、液体生検、又はこれらの任意の組み合わせにより肺癌の再発をモニタリングされてもよい。しかしながら、現在の画像診断法と液体生検による再発の特定は感度も低く、残存疾患の特定能も最小であるという問題を抱えている。

【００６９】

[116] 図４に示すように、肺癌の臨床標準治療に気管支ゲノム分類器を追加すれば、被検体管理を大幅に改善して良い結果をもたらすことができる。例えば、気管支ゲノム分類器を追加する前は、中～低リスクの被検体の約３７％以上が侵襲的処置を受けることになる。それに対して、気管支ゲノム分類器を臨床標準治療に追加すれば、中～低リスクの被検体に対する侵襲的処置の数が約４７％以上減少する。

【００７０】
[117] 図５に示すように、特発性肺線維症（ＩＰＦ）の臨床標準治療に気管支ゲノム分類器を追加すれば、不要な侵襲的処置の数を大幅に減らすことができる。例えば、米国及び欧州では、約２００，０００人がＩＰＦの疑いありと評価され、診断のために高解像度コンピュータ断層撮影（ＨＲＣＴ）を受け得る。これらの被検体のうち、約１５０，０００人（又は７０～７５％）はＨＲＣＴから不明確な又は疑わしい結果を受け得る。不明確な又は疑わしい結果を示す被験体は、組織病理学的真実（ＩＰＦの有無）を特定するために診断手術を受け得る。しかしながら、本明細書に記載のゲノム分類器を実施すれば、クラシック間質性肺炎パターン（ＵＩＰ）（ＩＰＦのパターン）の有無を特定し得る。クラシックなＵＩＰの存在が特定されると、被検体は診断手術又は治療を受け得る。クラシックＵＩＰが存在しないことが確認された場合、被検体は、侵襲的処置を受けなくてもよい。

【００７１】
[118] 図６は、生検回数における減少率のグラフであり、ＵＩＰを他のＩＬＤ病理と区別する際にゲノム分類器を用いることの臨床的有用性を強調している。例えば、ゲノム分類器を導入すると、ＩＬＤの管理アプローチの改善に大きな臨床的影響を与えることができる。ＵＩＰと他のＩＬＤ病理と区別する際にゲノム分類器を用いることで、侵襲的生検の数を大幅に減らせるはずである。

【００７２】
[119] 図７に示すように、損傷範囲（field of injury）の病因は共通の経路を有する疑いが高い。例えば、病因曝露と慢性気道損傷は、気道上皮環境などの組織の微小環境を変化させ得る。微小環境の変化は、１つ以上の分子異常と１つ以上の修復経路の活性化をもたらし得る。表現型は、損傷に対する内因性の宿主反応により決定され得る。ＣＯＰＤ、ＩＬＤ、喘息又はこれらの任意の組み合わせは、肺癌のリスクを高める宿主の反応を反映し得る。気道上皮に対するバイオマーカー分析は、変化の連続性を特定する重要な機会となり得る。

【００７３】
[120] 図８に示すように、発癌範囲と損傷範囲など、複数の範囲があり得る。損傷範囲は、肺癌の存在に関連付けられ、気道全体の細胞から見られるゲノム変化を含み得る。発癌範囲は、腫瘍源の近位など、周囲の気道に存在し得る腫瘍特異的ゲノム変化を含み得る。損傷範囲と発癌範囲の間には相互作用が生じ得る。例えば、気道上部で見られる分子の変化は、損傷範囲、発癌範囲、又はこれらの任意の組み合わせと関連している場合と関連していない場合があり得る。リスクのある分子指標（at-risk molecular signature）は、肺癌、ＩＬＤ、ＣＯＰＤ、喘息などの肺疾病として表され得る。

【００７４】
[121] 図９は、損傷範囲と発癌範囲における分子の概念を示したものである。損傷には、喫煙又は環境曝露が含まれ得る。損傷指標（ＲＮＡ発現の変化など）及び疾患指標（追加的な変異、転写調節不全など）は、癌、線維症、肺気腫など肺疾病の概略となり得る。

【００７５】
[122] 図１０は、図３の同様の経路であり、単一の気管支ゲノム分類器を追加して改善された臨床判断の現状を示す。ただし、臨床治療経路における他の決定ポイントに他のゲノム分類器を追加すれば、現状の臨床治療を更に改善することができる。

【００７６】
[123] 図１１ａ及び図１１ｂは、早期発見を改善し、不要な侵襲的処置を最小限に抑える、現行の臨床標準治療における特定の決定ポイントへの様々なゲノム分類器の追加を示す。例えば、リスク曝露集団は、包括的な集団内で特定し得る。リスク曝露集団には、肺疾病（肺癌など）を発症する又は罹患するリスクが高い被検体が含まれ得る。肺疾病に関連する１つ以上のリスク因子の存在を特定することによりリスク曝露集団を特定することができる。被検体には、１つ以上のリスク因子の存在を評価するために問診を行ってもよい。医療従事者は、質問に対する被検体の答えから１つ以上のリスク因子の存在を評価し

得る。肺疾病のリスクがあると特定された被検体から試料（鼻の擦過などにより採取された非侵襲的試料など）を採取してもよい。試料から得たデータ（発現レベルや配列変異体データなど）を、ゲノム分類器（Nasa-Detect分類器など）に入力し得る。ゲノム分類器により、試料が陽性又は陰性のいずれであるかを特定し得る。陽性の結果が出た被検体は、肺結節をスキャンするために画像スキャン（低線量ＣＴスキャンなど）を受け得る。陰性の結果が出た被検体については、後の時点で試料をもう一度採取し、そこから得たデータをゲノム分類器に入力してもよい。

【００７７】

[124] 画像スキャン（低線量ＣＴスキャンなど）により肺結節の存在が確認された被検体から、試料を採取してもよい。試料から得たデータ（発現レベルや配列変異体データなど）を、ゲノム分類器（Nasa-RISK分類器など）に入力する。このゲノム分類器により、試料の肺疾病（肺癌など）に対するリスクが高いか低いかを特定する。分類器により高リスクの結果が出た被検体は、肺疾病の有無を確認するために侵襲的処置（気管支鏡検査、ＴＴＮＡ、ＶＡＴＳなど）を受け得る。分類器により低リスクの結果が出た被検体については、後の時点でもう一度画像スキャンを行って肺結節の存在を確認し、他の試料から得たデータをゲノム分類器に入力してもよい。

【００７８】

[125] ゲノム分類器（Nasa-Risk階層化分類器又は気管支ゲノム分類器など）により肺疾病のリスクが低いと特定された被検体は、疾患の進行を遅らせるか食い止めるために、又は肺疾病の発生を予防するために介入療法を受けてもよい。介入療法の少なくとも一部が完了した後、被検体から試料を採取してもよい。試料から得たデータ（発現レベルや配列変異体データなど）を、ゲノム分類器（Nasa-PROTECT監視分類器など）に入力し得る。ゲノム分類器により、介入療法の有効性、被検体のコンプライアンス、疾患の退行又は肺疾病の予防、又はこれらの任意の組み合わせについて特定してもよい。

【００７９】

[126] 外科的に切除された癌などの治療計画又は根治的治療（医薬組成物の投与など）を受けている被検体から、治療処置の後、試料を採取してもよい。試料から得たデータ（発現レベルや配列変異体データなど）を、ゲノム分類器（Nasa-Recurrence分類器など）に入力し得る。ゲノム分類器により、肺疾病の再発を早期発見することができる。

【００８０】

[127] 図１２は、Nasa-Detect分類器の特徴を示す。分類器により、リスク曝露集団の肺損傷を検出し得る。この分類器は、（ｉ）画像スクリーニングの適用対象を選別するふるいの最適化、（ｉｉ）より確実な初期スクリーニングツールの使用による画像スキャンの効果増大、（ｉｉｉ）介入治療の効能を期待できる被検体における早期発見の強化、又は（ｉｖ）これらの任意の組み合わせを提供し得る。この分類器の評価対象である被検体は、既に肺癌のリスクありと判定を受けていてもよい。分類器による陽性の結果は、画像スキャン（ＬＤＣＴなど）による追跡検査の推奨を含み得る。ＬＤＣＴによる結節なしとの判定は、被検体を介入療法の適用候補と見なしてもよい。結果が陰性であれば、後の時点でこの分類器を用いて再度モニタリングを行ってもよい。

【００８１】

[128] 図１３はNasa-Risk階層化分類器の特徴を示す。分類器は結節のリスクを階層化し得る。分類器は、確定できない肺結節の数を最小に抑え得る。生検を必要とし得る被検体には生検を受けさせ、生検を必要としない被検体に侵襲性生検を行うことを回避することができる。この分類器の評価対象である被検体は、既に肺病変が特定されている被検体であってもよい。分類器による「リスク低」の結果は、介入療法の適用候補として被検体に対して監視を行い得る。分類器による「リスク中」の結果は、臨床上の判断を含み得る。分類器による「リスク高」の結果は、被検体に生検を受けさせ得る。分類器は、次世代シーケンシング（ＮＧＳ）プラットフォームで展開してもよく、シーケンシング情報、放射線学的特徴、又はこれらの任意の組み合わせを含んでもよい。

【００８２】

[129] 図１４はNasa-PROTECT分類器の特徴を示す。分類器は、肺損傷の回復をモニタリングするコンパニオン診断として用いられてもよい。分類器は、特定の処置や治療に対して被検体のコンプライアンスを特定してもよい。分類器は、推奨の処置や治療から効能を期待できる被検体を特定してもよい。分類器の適用対象である被検体は、Nasa-DETECT陽性・陰性の被検体集団を含んでもよい。分類器の適用対象である被検体は、肺結節陽性・Nasa-Risk階層化分類器による低リスクを含んでもよい。

【００８３】

[130] 図１５はNasa-Recurrence分類器の特徴を示す。分類器は、根治的外科切除又は根治的治療計画を受けている被検体の再発をテストするための非侵襲的モニタリング方法として用いられてもよい。分類器は、疾患の発病又は再発病を初期段階で特定することができる。分類器は再発特定に対する感度が高いものであり得る。この分類器の適用対象である被検体は、治療を目的とする肺癌の外科切除、又は根治的治療計画を受けている被検体を含んでもよい。

【００８４】

[131] 図１６は、遺伝子検査用のＡＣＣＥ評価プロセスを示す。遺伝子検査の評価における主要な基準の４つは、Ａ：分析的妥当性、Ｃ：臨床的妥当性、Ｃ：臨床的有用性、Ｅ：倫理的意味を含む。

【００８５】

[132] 図１７は、（ｉ）ゲノム分類器の訓練及び検証に用いられる試料の種類、並びに（ｉｉ）特定のためにゲノム分類器に入力される試料の種類の例を示す。試料には、既に良性の肺疾患を有する被検体、慢性肺感染症に罹患している被検体、免疫抑制療法を受けた被検体、肺疾病を発症する遺伝的リスクが高い被検体、受動喫煙にさらされる非喫煙者、又はこれらの任意の組み合わせから採取された試料が含まれてもよい。試料は、複数の異なる国から採取されたものであってもよい。コホートから部分集団を抽出し、特定の分類器の訓練や検証に用いてもよい。特定の集団、曝露の種類、又はこれらの任意の組み合わせに対する分類器を開発してもよい。例えば、分類器は、中国の環境汚染や肺疾病の遺伝的素因のために開発されてもよい。ゲノム分類器の開発において、肺疾病のスクリーニング、診断、肺疾病治療の評価、被検体の症状のモニタリング、又はこれらの任意の組み合わせを目的としてもよい。試料を被検体から毎年採取してもよい。毎年採取される試料には、鼻擦過、血液試料、画像スキャン、又はこれらの任意の組み合わせが含まれ得る。

【００８６】

[133] 図１８は、鼻又は気管支の擦過から試料を採取しているコホートを示す。各コホートは、識別し得る（ＡＥＧＩＳ、ＤＥＣＡＭＰ１、ＬＴＰ２，ＤＥＣＡＭＰ２、Ｌａｈｅｙ）。各試料コホートに対して、登録された被検体の数と、現行の標準治療における位置（気管支鏡検査中、事後画像スキャン中、又はスクリーニング中）を同定することができ、表記することができる。被検体の年齢と喫煙歴などの包含基準を提示し得る。それぞれの試料コホートに対して、試料の種類（鼻擦過、気管支擦過、血液、画像スキャン）と、経過観察期間（１２ヵ月、２４ヵ月、４８ヵ月）も記入し得る。

【００８７】

[134] 図１９は、分類器（Ｎａｓａ－ＤＥＴＥＣＴ分類器など）を訓練及び検証するために使用される訓練用試料の例を示す。コホートＤＥＣＡＭＰ２及びＬａｈｅｙは、この分類器の訓練のために使用され得る。試料は、鼻擦過、血液試料、又はそれらの組み合わせを含み得る。試料を提供する各被検体から追加データを収集することができる。これには：被検体が以前の喫煙者であるか現在の喫煙者であるか；禁煙後の期間；共存症の存在；肺疾病の家族歴；気管支前リスク；又はそれらの任意の組み合わせが含まれる。分類器を訓練し、検証するために使用される訓練用試料は、約：１００試料、２００試料、３００試料、４００試料、５００試料、６００試料、７００試料、８００試料、９００試料、１０００試料、１１００試料、１２００試料、１３００試料、１４００試料、１５００試料、１６００試料、１７００試料、１８００試料、１９００試料、２０００試料またはそれ以上（例えば異なる被検体から得られた１９５０の試料）であってもよい。ある場合に

は、訓練用試料は、約１００個から約２００個の試料を含むことができる。ある場合には、訓練用試料は、約１００個から約３００個の試料を含むことができる。ある場合には、訓練用試料は、約１００個から約４００個の試料を含むことができる。ある場合には、訓練用試料は、約１００個から約５００個の試料を含むことができる。ある場合には、訓練用試料は、約１００個から約６００個の試料を含むことができる。ある場合には、訓練用試料は、約１００個から約７００個の試料を含むことができる。ある場合には、訓練用試料は、約１００個から約８００個の試料を含むことができる。ある場合には、訓練用試料は、約１００個から約９００個の試料を含むことができる。ある場合には、訓練用試料は、約１００個から約１０００個の試料を含むことができる。ある場合には、訓練用試料は、約１００個から約１５００個の試料を含むことができる。ある場合には、訓練用試料は、約１００個から約２０００個の試料を含むことができる。ある場合には、訓練用試料は、約１００個から約３０００個の試料を含むことができる。ある場合には、訓練用試料は、約１００個から約４０００個の試料を含むことができる。ある場合には、訓練用試料は、約１００個から約５０００個の試料を含むことができる。試料を提供する被検体は、喫煙者、曝露リスクのある非喫煙者、又は喫煙歴も曝露リスクもない健康被検体であり得る。

【００８８】

[135] 図２０は、分類器（例えば、Nasa-RISK Stratifier分類器）を訓練及び検証するために使用される訓練用試料の例を示す。コホートＡＥＧＩＳ及びＤＥＣＡＭＰ１が、この分類器の訓練のために使用され得る。試料は、鼻擦過、気管支擦過、血液試料、又はそれらの任意の組合せを含み得る。試料を提供する各被検体から追加データを収集することができる。これには：被検体が以前の喫煙者であるか現在の喫煙者であるか；禁煙後の期間；共存症の存在；気管支前リスク；又はその組み合わせ；が含まれ得る。分類器を訓練し、かつ、検証するために使用される訓練用試料は、約１００試料、２００試料、３００試料、４００試料、５００試料、６００試料、７００試料、８００試料、９００試料、１０００試料、１１００試料、１２００試料、試料、試料、１３００試料、１４００試料、１５００試料、１６００試料、１７００試料、１８００試料、１９００試料、２０００試料、２１００試料、２２００試料、２３００試料、２４００試料、２５００試料、２６００試料、２７００試料、２８００試料、２９００試料、３０００試料、又はそれ以上（例えば異なる被検体から得られた２３５０の試料）とすることができる。ある場合には、訓練用試料は、約１００個から約２００個の試料を含むことができる。ある場合には、訓練用試料は、約１００個から約３００個の試料を含むことができる。ある場合には、訓練用試料は、約１００個から約４００個の試料を含むことができる。ある場合には、訓練用試料は、約１００個から約５００個の試料を含むことができる。ある場合には、訓練用試料は、約１００個から約６００個の試料を含むことができる。ある場合には、訓練用試料は、約１００個から約７００個の試料を含むことができる。ある場合には、訓練用試料は、約１００個から約８００個の試料を含むことができる。ある場合には、訓練用試料は、約１００個から約９００個の試料を含むことができる。ある場合には、訓練用試料は、約１００個から約１０００個の試料を含むことができる。ある場合には、訓練用試料は、約１００個から約１５００個の試料を含むことができる。ある場合には、訓練用試料は、約１００個から約２０００個の試料を含むことができる。ある場合には、訓練用試料は、約１００個から約３０００個の試料を含むことができる。ある場合には、訓練用試料は、約１００個から約４０００個の試料を含むことができる。ある場合には、訓練用試料は、約１００個から約５０００個の試料を含むことができる。試料を提供する対象は、喫煙者又は非喫煙者であり得る。

【００８９】

[136] 図２１は、バイオマーカーと、その存在の有無を検出するために用いられる技術を示す。例えば、ゲノムバイオマーカー（突然変異と不均衡を含む）は、次世代シーケンシング（ＮＧＳ）、マイクロアレイ、蛍光in situハイブリダイゼーション（ＦＩＳＨ）、ポリメラーゼ連鎖反応（ＰＣＲ）、又はこれらの任意の組み合わせにより検出すること

ができる。エピジェネティックバイオマーカー（５－ヒドロキシメチル化シトシン、５－メチル化シトシン、５－カルボキシメチル化シトシン、５－ホルミル化シトシンなどのＤＮＡメチル化など）は、ＮＧＳ、マイクロアレイ、ＰＣＲ、質量分析（ＭＳ）、又はこれらの任意の組み合わせにより検出することができる。トランスクリプトームバイオマーカー（ＲＮＡ発現レベルなど）は、ＮＧＳ、マイクロアレイ、ＰＣＲ、又はこれらの任意の組み合わせにより検出することができる。プロテオームバイオマーカー（タンパク質の存在など）は、タンパク質アレイ、免疫組織化学染色（ＩＨＣ）、又はこれらの任意の組み合わせにより検出することができる。

【００９０】

[137] 図２２は、ゲノム分類器におけるＲＮＡシーケンシングと、ゲノム分類器の甲状腺ＦＮＡ分析を示す。図２３は、遺伝子Ａ、遺伝子Ｂ、遺伝子ＣのＲＮＡシーケンシングの例を示す。ＲＮＡへの転写の後、（ｉ）１つ以上の発現レベル（各転写産物の計数など）の検出、（ｉｉ）１つ以上の変異体（各転写産物の配列など）の検出、（ｉｉｉ）染色体コピーの検出（ヘテロ接合性の消失（ＬＯＨ）など）、又は（ｉｖ）これらの任意の組み合わせを行ってもよい。

【００９１】

[138] 図２４は、本明細書における訓練アルゴリズムのフローチャートである。例えばアルゴリズムは、試料から１つ以上の種類のシーケンシングデータを受け取ることができる。アルゴリズムが受け取ったデータを正規化し得る。必要に応じて、特徴抽出又は特徴選択が教師付き機械学習とともに行われ得る。１つ以上の臨床的共変量をアルゴリズムに追加してもよい。１つ以上の訓練ラベルをアルゴリズムに追加してもよい。１つ以上のロックをアルゴリズムに追加してもよい。分析検証を確認し得る。臨床検証を確認し得る。ゲノム分類器を稼働し得る。

【００９２】

[139] 図２５は、ベセスダシステムにおける細胞サブタイプと組織サブタイプを多く含む訓練セットの例を示す。例えば、図２５は総計６３４個の試料のうち、ベセスダシステムにおける細胞サブタイプと組織サブタイプを両方とも含む５０７個の試料を図示している。訓練セットは、生物学の全カテゴリーに及んでいてもよい。

【００９３】

精度、特異度、感度

[140] 本明細書に記載の方法は、（ｉ）肺癌など疾病の有無の判定、（ｉｉ）良性又は悪性組織の分類を行ってもよく、方法における診断の特異度は、約７０％を超えてもよい。一部の実施形態において、特異度は、少なくとも約７０％、少なくとも約７５％、少なくとも約８０％、少なくとも約８５％、少なくとも約８６％、少なくとも約８７％、少なくとも約８８％、少なくとも約８９％、少なくとも約９０％、少なくとも約９１％、少なくとも約９２％、少なくとも約９３％、少なくとも約９４％、少なくとも約９５％、少なくとも約９６％、少なくとも約９７％、少なくとも約９８％、少なくとも約９９％、又はそれ以上であってもよい。一部において、特異度は約７０～約９９％であってもよい。一部において、特異度は約８０～約９９％であってもよい。一部において、特異度は約８５～約９９％であってもよい。一部において、特異度は約９０～約９９％であってもよい。一部において、特異度は約９５～約９９％であってもよい。一部において、特異度は約７０～約９５％であってもよい。一部において、特異度は約８０～約９５％であってもよい。一部において、特異度は約８５～約９５％であってもよい。一部において、特異度は約９０～約９５％であってもよい。一部において、特異度は約７０～１００％であってもよい。一部において、特異度は約８０～１００％であってもよい。一部において、特異度は約８５～１００％であってもよい。一部において、特異度は約９０～１００％であってもよい。

【００９４】

[141] 本明細書に記載の方法は、（ｉ）肺癌など疾病の有無の判定、（ｉｉ）良性又は悪性組織の分類を行ってもよく、方法における診断の感度は、約７０％を超えてもよい。

一部の実施形態において、感度は、少なくとも約７０％、少なくとも約７５％、少なくとも約８０％、少なくとも約８５％、少なくとも約８６％、少なくとも約８７％、少なくとも約８８％、少なくとも約８９％、少なくとも約９０％、少なくとも約９１％、少なくとも約９２％、少なくとも約９３％、少なくとも約９４％、少なくとも約９５％、少なくとも約９６％、少なくとも約９７％、少なくとも約９８％、少なくとも約９９％、又はそれ以上であってもよい。一部において、感度は約７０～約９９％であってもよい。一部において、感度は約８０～約９９％であってもよい。一部において、感度は約８５～約９９％であってもよい。一部において、感度は約９０～約９９％であってもよい。一部において、感度は約９５～約９９％であってもよい。一部において、感度は約７０～約９５％であってもよい。一部において、感度は約８０～約９５％であってもよい。一部において、感度は約８５～約９５％であってもよい。一部において、感度は約９０～約９５％であってもよい。一部において、感度は約７０～１００％であってもよい。一部において、感度は約８０～１００％であってもよい。一部において、感度は約８５～１００％であってもよい。一部において、感度は約９０～１００％であってもよい。

【００９５】

[142] 本明細書に記載の方法は、（ｉ）肺癌など疾病の有無の判定、（ｉｉ）良性又は悪性組織の分類を行ってもよく、方法における診断の感度は約７０％を超えてもよく、特異度は約７０％を超えてもよい。感度は約７０％を超えてもよく、特異度は約８０％を超えてもよい。感度は約７０％を超えてもよく、特異度は約９０％を超えてもよい。感度は約７０％を超えてもよく、特異度は約９５％を超えてもよい。感度は約８０％を超えてもよく、特異度は約７０％を超えてもよい。感度は約８０％を超えてもよく、特異度は約８０％を超えてもよい。感度は約８０％を超えてもよく、特異度は約９０％を超えてもよい。感度は約８０％を超えてもよく、特異度は約９５％を超えてもよい。

感度は約９０％を超えてもよく、特異度は約７０％を超えてもよい。感度は約９０％を超えてもよく、特異度は約８０％を超えてもよい。感度は約９０％を超えてもよく、特異度は約９０％を超えてもよい。感度は約９０％を超えてもよく、特異度は約９５％を超えてもよい。感度は約９５％を超えてもよく、特異度は約７０％を超えてもよい。感度は約９５％を超えてもよく、特異度は約８０％を超えてもよい。感度は約９５％を超えてもよく、特異度は約９０％を超えてもよい。感度は約９５％を超えてもよく、特異度は約７５％を超えてもよい。

【００９６】

[143] 本明細書に記載の方法は、（ｉ）肺癌など疾病の有無の判定、（ｉｉ）良性又は悪性組織の分類を行ってもよく、方法における陰性反応適中度（ＮＰＶ）は、約９５％以上であってもよい。ＮＰＶは、少なくとも約９５％、少なくとも約９５．５％、少なくとも約９６％、少なくとも約９６．５％、少なくとも約９７％、少なくとも約９７．５％、少なくとも約９８％、少なくとも約９８．５％、少なくとも約９９％、少なくとも約９９．５％、又はそれ以上であってもよい。一部において、ＮＰＶは約９５％～約９９％であってもよい。一部において、ＮＰＶは約９６％～約９９％であってもよい。一部において、ＮＰＶは約９７％～約９９％であってもよい。一部において、ＮＰＶは約９８％～約９９％であってもよい。一部において、ＮＰＶは約９５％～１００％であってもよい。一部において、ＮＰＶは約９６％～１００％であってもよい。一部において、ＮＰＶは約９７％～１００％であってもよい。一部において、ＮＰＶは約９８％～１００％であってもよい。

【００９７】

[144] 一部の実施形態において、公称(nominal)特異度は約５０％以上である。一部の実施形態において、公称特異度は約６０％以上である。一部の実施形態において、公称特異度は約７０％以上である。一部の実施形態において、公称陰性反応適中度（ＮＰＶ）は約９５％以上である。一部の実施形態において、ＮＰＶは少なくとも約９０％、少なくとも約９１％、少なくとも約９２％、少なくとも約９３％、少なくとも約９４％、少なくとも約９５％、少なくとも約９５．５％、少なくとも約９６％、少なくとも約９６．５％、少

なくとも約９７％、少なくとも約９７．５％、少なくとも約９８％、少なくとも約９８．５％、少なくとも約９９％、少なくとも約９９．５％（例えば、９０％、９１％、９２％、９３％、９４％、９５％、９５．５％、９６％、９６．５％、９７％、９７．５％、９８％、９８．５％、９９％、９９．５％、又は１００％）、であり、特異度（又は陽性反応適中度（ＰＰＶ））は少なくとも約３０％、少なくとも約３５％、少なくとも約４０％、少なくとも約５０％、少なくとも約６０％、少なくとも約７０％、少なくとも約８０％、少なくとも約９０％、少なくとも約９５％、少なくとも約９５．５％、少なくとも約９６％、少なくとも約９６．５％、少なくとも約９７％、少なくとも約９７．５％、少なくとも約９８％、少なくとも約９８．５％、少なくとも約９９％、少なくとも約９９．５％（例えば３０％、３５％、４０％、５０％、６０％、７０％、８０％、９０％、９５％、９５．５％、９６％、９６．５％、９７％、９７．５％、９８％、９８．５％、９９％、９９．５％、又１００％）である。一部において、ＮＰＶは少なくとも約９５％、特異度は少なくとも約５０％である。一部において、ＮＰＶは少なくとも約９５％、特異度は少なくとも約７０％である。一部において、ＮＰＶは少なくとも約９５％、特異度は少なくとも約７５％である。一部において、ＮＰＶは少なくとも約９５％、特異度は少なくとも約８０％である。

【００９８】

[145] 感度はＴＰ／（ＴＰ＋ＦＮ）で表され、ここで、ＴＰは真陽性率、ＦＮは偽陰性率を指す。確定した組織病理診断に基づき、継続判定不能数を悪性結果の総数で割ったものである。特異度は通常ＴＮ／（ＴＮ＋ＦＰ）で表されるが、ＴＮは真陰性率、ＦＰは偽陽性率を指す。確定した組織病理診断に基づき、良性結果の数を良性結果の総数で割ったものである。陽性反応適中度（ＰＰＶ）はＴＰ／（ＴＰ＋ＦＰ）で、陰性反応適中度（ＮＰＶ）はＴＮ／（ＴＮ＋ＦＮ）で表される。

【００９９】

[146] 本発明による方法及び組成物は、更に、生体試料の特定、分類、診断、又はその他の特徴付けを目的とするバイオマーカーパネルの使用に関する。パネルは、損傷範囲、発癌範囲、症状（ＩＬＤ、ＣＯＰＤ、肺癌）の有無、疾患を発症するリスクの増加の有無、疾患再発の有無、疾患の回復、疾患の予防、又はこれらの任意の組み合わせのうち１つ以上を特定し得る。上記の方法及び組成物は、バイオマーカーパネル群を使用してもよい。通常、パネルにおけるバイオマーカーの遺伝子発現レベルパターン（損傷指標や発癌指標などの指標としても知られる）を判定し、その後、生体試料における同じバイオマーカーパネルの指標を、例えば、試料指標と参照指標の類似度を用いて評価し得る。一部の実施形態において、方法は、バイオマーカーパネル、及び／又は分類パネルに含まれてもよい２つ以上の遺伝子発現産物のレベルを測定（又は取得）する。例えば、一部の実施形態において、バイオマーカーパネル又は分類パネルは、少なくとも約１個、２個、３個、４個、５個、６個、７個、８個、９個、１０個、１５個、２０個、２５個、３０個、３３個、３５個、３８個、４０個、４３個、４５個、４８個、５０個、５３個、５８個、６３個、６５個、６８個、１００個、１２０個、１４０個、１４２個、１４５個、１４７個、１５０個、１５２個、１５７個、１６０個、１６２個、１６７個、１７５個、１８０個、１８５個、１９０個、１９５個、２００個、又は３００個のバイオマーカーを含み得る。一部の実施形態において、バイオマーカーパネル又は分類パネルは、約１個、２個、３個、４個、５個、６個、７個、８個、９個、１０個、１５個、２０個、２５個、３０個、３３個、３５個、３８個、４０個、４３個、４５個、４８個、５０個、５３個、５８個、６３個、６５個、６８個、１００個、１２０個、１４０個、１４２個、１４５個、１４７個、１５０個、１５２個、１５７個、１６０個、１６２個、１６７個、１７５個、１８０個、１８５個、１９０個、１９５個、２００個、又は３００個以下のバイオマーカーを含み得る。一部の実施形態において、バイオマーカーパネル又は分類パネルは、約１個～約５００個のバイオマーカーを含み得る。一部の実施形態において、バイオマーカーパネル又は分類パネルは、約１個～約４００個のバイオマーカーを含み得る。一部の実施形態において、バイオマーカーパネル又は分類パネルは、約１個～約３００個のバイオマーカーを含

み得る。一部の実施形態において、バイオマーカーパネル又は分類パネルは、約１個～約２００個のバイオマーカーを含み得る。一部の実施形態において、バイオマーカーパネル又は分類パネルは、約１個～約１００個のバイオマーカーを含み得る。一部の実施形態において、バイオマーカーパネル又は分類パネルは、約１個～約５００個のバイオマーカーを含み得る。一部の実施形態において、バイオマーカーパネル又は分類パネルは、約１００個～約５００個のバイオマーカーを含み得る。一部の実施形態において、バイオマーカーパネル又は分類パネルは、約２００個～約５００個のバイオマーカーを含み得る。一部の実施形態において、バイオマーカーパネル又は分類パネルは、約３００個～約５００個のバイオマーカーを含み得る。一部の実施形態において、バイオマーカーパネル又は分類パネルは、約４００個～約５００個のバイオマーカーを含み得る。一部の実施形態において、分類パネルは、少なくとも約１個、２個、３個、４個、５個、６個、７個、８個、９個、１０個、１５個、２０個、又は２５個の異なるバイオマーカーパネルを含む。他の実施形態において、分類パネルは、約１個、２個、３個、４個、５個、６個、７個、８個、９個、１０個、１５個、２０個、又は２５個以下の異なるバイオマーカーパネルを含み得る。バイオマーカーパネルは、損傷指標の特定、間質性肺炎パターン（ＵＩＰ）の存在の確認、疾患を発症するリスクの特定、疾患再発のリスクの特定、疾患の進行のモニタリング、又はこれらの任意の組み合わせを行い得る遺伝子のパネルを含んでもよい。

【０１００】

[147] 肺癌を発症するリスク又は尤度を高め得る１つ以上のリスク因子は、喫煙、受動喫煙（二次喫煙など）、ラドン曝露、工業物質曝露（石綿、ヒ素、ディーゼル排気ガス、マスタードガス、ウラン、ベリリウム、塩化ビニル、ニッケルクロム酸塩、石炭製品、クロロメチルエーテル、ガソリンなど）、遺伝又は環境による遺伝子突然変異、結核、大気汚染曝露、放射線被曝（過去の放射線療法）、被検体の年齢、二次症状（慢性閉塞性肺疾病（ＣＯＰＤ）、間質性肺疾患（ＩＬＤ）、喘息）の有無、栄養補助食品（ベータカロチンなど）の摂取、又はこれらの任意の組み合わせを含んでもよい。肺癌を発症するリスク又は疑いを高め得るリスク因子は、タバコ、葉巻、パイプ、又はこれらの任意の組み合わせによる喫煙を含んでもよい。

【０１０１】

[148] １つのリスク因子を持つ被検体を、リスク状態であると特定してもよい。２つのリスク因子を持つ被検体を、リスク状態であると特定してもよい。３つのリスク因子を持つ被検体を、リスク状態であると特定してもよい。個々のリスク因子は等しく重み付けされなくてもよい。喫煙などの単一のリスク因子の存在は、被験体をリスク状態と特定することができる。特定の遺伝子変異などの単一のリスク因子の存在は、それのみでは十分でないことがあり、他のリスク因子と組み合わせて有している場合のみ、リスク状態と特定することができる。

【０１０２】

[149] 質問に回答できるよう、被検体にアンケート（書面又はウェブ上）を提供し、１つ以上のリスク因子の存在を評価してもよい。医療従事者は、被検体から１つ以上の質問に対する回答を直接聞き、１つ以上のリスク因子の存在を評価してもよい。被検体から採取された非侵襲性試料により１つ以上のリスク因子の存在を評価してもよい。１つ以上のリスク因子の存在を評価するために被検体の過去の病歴を用いてもよい。医療従事者は、例えば被検体の病歴を含む、被検体のヘルスデータ又は生理データを保持してもよい。

【０１０３】

[150] 未確定の診断は、不要な手術、診断の遅れ、治療の遅れ、又はこれらの任意の組み合わせにつながり得る。現行の臨床経路では、診断の１５～７０％が不安定又は未確定（uncertain or inconclusive）となり得る。未確定の診断の場合、診断手術が推奨され得る。診断が確定していないことにより、手術が推奨される被検体の一部は良性であり得る。高い感度及び特異度で試料を診断又は分類するゲノム分類器の開発が必要であり得る。

【０１０４】

[151] 現在、年間約２２５，０００人を超える患者が新たに肺癌の確定診断を受けている可能性がある。新たに肺癌の診断を受けた被検体の約９０％は、喫煙の経験がある被検体であり得る。侵襲的生検を受けた被検体の約４０％は癌ではない。更に、早期発見も死亡率の低減に非常に重要な要素であり得る。しかしながら、現在の標準治療は、診断に侵襲的処置を必要としている。

【０１０５】

[152] 末梢肺結節などの肺組織は生検を行い難く、気管支鏡検査から未確定又は不明瞭な結果しか得られないことがよくある。従って、肺癌を診断するための他のオプションが望まれる。

【０１０６】

[153] 喫煙は、鼻、口、口腔、鼻腔、咽頭、喉頭、気管、肺、気管支、肺胞の上皮細胞、又はこれらの任意の組み合わせを含む気道全体の上皮細胞における遺伝子発現を変化させ得る。

【０１０７】

[154] 気道の一部から上皮細胞を分離し、分離した上皮細胞における遺伝子指標又はバイオマーカーパネルをアッセイすることで、癌を発症するリスクの判定、癌の確定診断、又は肺組織の良性か悪性かの分類が可能となる。上記のアッセイでは、例えば、核酸増幅（例えばＰＣＲ）、アレイハイブリダイゼーション、又はシーケンシングを行ってもよい。そのようなシーケンシングは、超並列型シーケンシングであってもよい（例えば、イルミナ、パシフィックバイオサイエンス・オブ・カリフォルニア、又はオックスフォードナノポア）。シーケンシングは、遺伝的（又はゲノム）異常（例えばコピー数多型、単一のヌクレオチド多型、単一のヌクレオチド変異体、挿入、欠失など）と、遺伝子に対応する発現レベルを特定できる、シーケンシング読み取りを行ってもよい。これにより、被検体のゲノムにおける遺伝子異常に関する情報と、同試料における転写メッセンジャーリボ核酸（ｍＲＡＮ）の発現レベルに関する情報を、同時に提供することができる。

【０１０８】

[155] 癌又は腫瘍の部位から離れている気道の一部から上皮細胞を分離し得る。例えば、上皮細胞は鼻皮細胞又は口腔上皮細胞であってもよい。分離した鼻上皮細胞から得たバイオマーカーパネルの発現レベルの遺伝子指標により、気管支組織や末梢肺結節において癌を発症するリスクを予測し、又は癌の存在を確定することができる。癌の存在に関連付けられるゲノム変化は、気道全体の細胞から見られることがある。

【０１０９】

[156] 間質性肺疾患（ＩＬＤ）の亜型は、区別するのが難しいことがあり、臨床的に確定診断するのが難しいことがある。例えば約４２％の、ＩＬＤに罹患する多くの被検体が、最初の症状から確定診断を受けるまで少なくとも１年以上かかると報告されている。誤診が一般的であり得る。ＩＬＤに罹患する被検体の少なくとも５５％は、少なくとも１回は誤診されているとの報告もある。

【０１１０】

[157] 米国及び欧州では、毎年約２００，０００人がＩＬＤの疑いありと評価される。高解像度ＣＴスキャンの施術を受けた被検体の約２５～３０％がＵＩＰの存在を示している。約７０～７５％（約１５０，０００人）の被検体は、高解像度ＣＴスキャン後に不確定又は確定の診断を受ける。不確定の診断を受けているこれらの被験体は、診断手術を推奨され得る。

【０１１１】

[158] 遺伝子指標（ＩＰＦにおける標準ＵＩＰパターン）を用いるゲノム分類器を開発し、診断の精度を高め、診断手術を受ける被検体の数を低減する必要があり得る。

【０１１２】

[159] 本明細書に記載の方法は、ＩＬＤ罹患の疑いがある被検体から採取された試料におけるバイオマーカーパネル（例えば標準ＵＩＰパターン）に対するアッセイを行い、ＩＬＤ（ＩＰＦ）の存在を特定するゲノム分類器を提供する。この方法は、少なくとも約８

８％の特異度と少なくとも約６７％の感度を有し得る。ゲノム分類器により陽性のＵＩＰパターンを有すると特定された被検体の場合、その後診断手術を受ける被検体の割合は、ゲノム分類器なしでは約５９％であったが、ゲノム分類器を使用すると約２９％まで減少した。

【０１１３】

[160] 標準ＵＩＰパターンに対して高解像度コンピュータ断層撮影（ＨＲＣＴ）を行う基準として、少なくとも４つ、つまり、胸膜下基底優位、網状異常、蜂巣状構造（牽引気管支拡張症の有無に拘らず）、ある特徴の欠如によるＵＩＰパターンとの矛盾が挙げられ得る。ＵＩＰパターンには、胸膜下基底優位、網状異常、ある特徴の欠如によるＵＩＰパターンとの矛盾の３つが含まれ得る。標準ＵＩＰパターンと矛盾する症候には、上肺野又は中肺野優位、気管支周囲優位、広範囲のスリガラス状異常、大量の微小結節、離散性嚢胞、びまん性モザイク減弱又は空気捕捉、気管支肺区域又は肺葉の硬化のうちいずれかが含まれ得る。

【０１１４】

[161] 被検体（肺癌を発症するリスクが低い被検体など）から、気管支鏡検査、胸腔穿刺法（ＴＴＮＡ）、胸腔鏡補助下手術（ＶＡＴＳ）、又はその他の方法により、気道組織試料、例えば肺組織試料を採取し得る。気管支鏡検査から確定診断が得られなかった場合は、分類器（気管支ゲノム分類器など）を適用し、気道組織試料を特定・分類して侵襲的処置を不要に行わないようにしてもよい。

【０１１５】

[162] 被検体は、経気管支生検などの生検を受けることもある。生検から得た１つ以上の発現レベルに対して分類器（ゲノム分類器など）を適用し、遺伝子パネルの１つ以上の遺伝子や、遺伝子発現パターン（標準ＩＰＦ「ＵＩＰパターン」など）の有無を検出してもよい。分類器は、生検試料におけるＩＬＤ（ＩＰＦなど）の有無を特定することができる。

【０１１６】

[163] 一般集団に対して肺癌を発症するリスク増加の疑いがあり得る（１つ以上のリスク因子に基づいて判定）被検体に対して、分類器（Nasa-Detect分類器など）を採用し、被検体における「損傷」指標の有無を判定し得る。これは、肺癌診断における早期発見につながり得る。分類器（Nasa-Detect分類器）を被検体から採取された試料からアッセイされた１つ以上の発現レベルに適用し、遺伝子パネルの１つ以上の遺伝子又は遺伝子発現パターンの有無を検出してもよい。遺伝子パネルは、被検体において肺癌を発症しやすくする、又は、疾患の早期指標にあたる、「損傷」の指標を含み得る。分類器を用いて、介入療法又は損傷治療の適用対象となり得る被検体を特定することができる。分類器（Nasa-Detect分類器）が陰性の結果、つまり被検体は「損傷」パネルにおける１つ以上の遺伝子を有せず、発現変化もないとの結果を出した場合、後の時点で被検体から採取された第２の試料を以って分類器を再実行し、遺伝子発現における変化をモニタリングしてもよい。分類器（Nasa-Detect分類器）が陽性の結果、つまり被検体が「損傷」パネルにおける１つ以上の遺伝子又は発現変化を示しているとの結果を出した場合、被検体は、低線量ＣＴスキャン（ＬＤＣＴ）を受け得る。

【０１１７】

[164] 「リスク状態」である被検体集団から「損傷」を検出するように、分類器を訓練してもよい。陽性結果とともにＬＤＣＴによる経過観察を推奨してもよく、陰性結果とともに、一定期間ごとに第２の分類器（Nasa-Detect分類器など）によりモニタリングを行うことを推奨してもよい。上記の一定期間とは、例えば、約０．５年ごと、約１年ごと、約１．５年ごと、約２年ごと、約２．５年ごと、約３年ごと、約３．５年ごと、約４年ごと、約４．５年ごと、約５年ごと、又はそれ以上である。一部において、一定期間は約０．５年～約３年であってもよい。一部において、一定期間は約１年～約３年であってもよい。一部において、一定期間は約２年～約３年であってもよい。一部において、一定期間は約０．５年～約２年であってもよい。一部において、一定期間は約０．５年～約１．５

年であってもよい。「リスク状態」の集団から「損傷」を検出するように訓練を受けた分類器は、（ｉ）ＬＤＣＴによりスクリーニングされる被検体の下位集団の最適化、（ｉｉ）より確実なスクリーニングツールの使用によるＬＤＣＴスクリーニングの効果増大、（ｉｉｉ）介入治療の効能を期待できる被検体の検出、又はこれらの任意の組み合わせを提供し得る。

【０１１８】

[165] 被検体は１つ以上の肺結節の有無を判定するために、低線量ＣＴスキャンを受けてもよい。ＬＤＣＴにより肺結節がないことが確認された場合は、（ｉ）後の時点で被検体から採取された第２の試料を以って分類器（Nasa-Detect分類器など）を再実行し、「損傷」パネルにおける１つの遺伝子の遺伝子発現変化をモニタリングしてもよく、（ｉｉ）被検体に介入療法を受けるように勧めてもよい。ＬＤＣＴにより肺結節の存在が確認された場合は、分類器（Nasa-Risk階層化分類器）を被検体から採取された試料からアッセイされた１つ以上の発現レベルに適用してもよい。

【０１１９】

[166] 介入療法を勧められた被検体（ＬＤＣＴにより肺結節を有しないことが確認された被検体）は、１つ以上の薬物療法を受けてもよい。１つ以上の薬物療法を行った後、被検体から試料を採集し、１つ以上の発現レベルについてアッセイを行い、分類器（Nasa-PROTECT監視分類器）に適用してもよい。分類器（Nasa-PROTECT監視分類器など）を訓練し、特定のバイオマーカーセットにおける変化をモニタリングし、特定の薬物療法を続けるべきか否かに関してアドバイスができるようにしてもよい。分類器（Nasa-PROTECT監視分類器）の結果は、薬物治療の中止、他の薬物治療への切り替え、非薬物治療への切り替え、現行の維持、又はこれらの任意の組み合わせとなり得る。分類器（Nasa-PROTECT監視分類器など）をコンパニオン診断ツールとして利用し、肺癌などの癌の進行を繰り止める「損傷範囲」の治療をモニタリングしてもよい。

【０１２０】

[167] 分類器（例えば、Nasa-PROTECT分類器）をコンパニオン診断ツールとして訓練させ、肺損傷の治療をモニタリングしてもよい。また、分類器を訓練させ、特定の治療や薬物療法の効能を期待できる被検体の下位集団を特定してもよい。

【０１２１】

[168] ＬＤＣＴにより１つ以上の肺結節の存在が確認されると、被検体から試料を採取し得る。試料に対して１つ以上の発現レベルをアッセイし得、その発現レベルを分類器（Nasa-Risk階層化分類器など）に入力し得る。気管支鏡検査やその他の侵襲的処置より前に分類器（Nasa-Risk階層化分類器など）を実行してもよい。分類器（Nasa-Risk階層化分類器など）により、被検体が「肺癌を発症するリスク低」「肺癌を発症するリスク高」「肺癌に罹患するリスク低」「肺癌に罹患するリスク高」のいずれであるかを特定し得る。分類器（Nasa-Risk階層化分類器など）の結果が「リスク低」であれば、後の時点で被検体に対してＬＤＣＴをもう一度行う。分類器（Nasa-Risk階層化分類器など）の結果が「リスク高」であれば、被検体は、気管支鏡検査、胸腔穿刺法（ＴＴＮＡ）、胸腔鏡補助下手術（ＶＡＴＳ）、又はその他の方法を受け得る。分類器（Nasa-Risk階層化分類器など）により、異なるカテゴリーの被検体（例えば、高リスクの被検体と低リスクの被検体）に対して、それぞれに合う次のステップを提示し得る。これによって、偽陽性率を抑え、癌の早期発見を改善することができる。

【０１２２】

[169] 分類器（Nasa-Risk階層化分類器など）を訓練し、ＬＤＣＴにより検出された肺結節の存在によるリスクを階層化し、次にとるべきステップをより明確に提示できるようにしてもよい。分類器は、放射線選択特性が含まれてもよく、次世代シーケンシング（ＮＧＳ）プラットフォームで展開してもよい。分類器の結果が「リスク低」であれば、持続的に被検体を経過観察又はモニタリングすることを勧める、又は、介入療法の潜在的な適用対象として被検体に介入療法を勧めることができる。分類器の結果が「リスク高」であれば、摘出生検を勧めるができる。分類器は、更なる検査を必要する被検体には摘出生検を

促し、必要としない被検体に摘出生検を行うことを回避することができる。分類器は、確定診断のできない肺結節の数を最小化することができる。分類器にかかる被検体集団は、ＬＤＣＴなどにより肺病変の存在が確定されている被検体を含んでもよい。

【０１２３】

[170] 気管支鏡検査又はその他の侵襲的処置（ＴＴＮＡやＶＡＴＳなど）により癌を診断すると、陽性の結果が出ることがあれば、気管支鏡検査を用いても確定診断ができないこともある。このような場合、気管支鏡検査を用いても確定診断ができないときは、試料を被検体から採取し、１つ以上の発現レベルをアッセイした後、その発現レベルを分類器（気管支ゲノム分類器など）に入力し得るる。分類器（気管支ゲノム分類器など）による結果が「中度のリスク」であると、被検体は気管支鏡検査又はその他の侵襲的処置をもう一度受け得る。分類器（気管支ゲノム分類器など）による結果が「リスク低」であると、被検体はＬＤＣＴをもう一度受け得る。ある場合には、気管支鏡検査により、癌、又は悪性の結果が得られることがある。気管支鏡検査又はその他の侵襲的処置から癌、又は悪性の結果が出された被検体は、罹患組織を外科的に切除し得る。罹患組織に対して外科切除を行うことができれば、試料を被検体から採取し、１つ以上の発現レベルをアッセイした後、その発現レベルを分類器（Nasa-Recurrence分類器など）に入力し得る。癌、例えば早期癌を検出して外科切除を行った後、分類器（Nasa-Recurrence分類器など）は、モニタリングを通じて再発を早期に予測することができる。分類器（Nasa-Recurrence分類器など）の結果が「再発のリスクなし」であれば、後の時点で被検体から第２の試料を採取し、１つ以上の発現レベルをアッセイした後、その発現レベルを分類器（Nasa-Recurrence分類器など）にかけ得る。分類器（Nasa-Recurrence分類器など）の結果が「再発のリスクあり」であれば、試料を被検体から採取し、試料に対して突然変異試験、免疫毒性試験、又はこれらの任意の組み合わせを行い得る。突然変異試験や免疫毒性試験の結果に基づいて、被検体に治療法を勧め、治療法のモニタリングや第２の突然変異試験又免疫毒性試験を行い得る。

【０１２４】

[171] 分類器（Nasa-Recurrence分類器など）を訓練させ、癌の再発につき、被検体に対して非侵襲的モニタリングを行うようにしてもよい。又は、分類器を訓練させ、腫瘍の根治的外科切除を受けた被検体に対して腫瘍や癌の再発をモニタリングできるようにしてもよい。分類器は再発を感知するか、再発する疑いが低いと判定することになる。被検体集団は、肺癌の治療のために外科切除を受けている被検体を含んでもよい。分類器は、疾患の再発を早期で特定することができる。

【０１２５】

[172] 癌又は悪性として特定された罹患組織に対して外科切除が行えない場合、被検体から試料を採取し、突然変異試験又免疫毒性試験を行ってもよい。

【０１２６】

試料

[173] 被検体から１つ以上の試料を採取し得る。１つ以上の試料は同種の試料、例えば１つ以上の生検試料であってもよい。被検体から採取された１つ以上の試料は異なる種類のもの、例えば、それぞれ生検と穿刺吸引法により得た試料であってもよい。

【０１２７】

[174] 試料の種類には、血液試料、組織試料、画像試料が含まれてもよい。試料は、無細胞ＤＮＡを含んでもよい。血液試料は、無細胞ＤＮＡも含んでもよい。血液試料は、血球を含んでもよい。血液試料は、血清又は血漿を含んでもよい。組織試料は、摘出生検、外科切除、針吸引法、穿刺吸引法、組織拭き取り、組織擦過、又はこれらの任意の組み合わせにより採取されることができる。組織試料は、上皮細胞、血液細胞、又はこれらの任意の組み合わせを含んでもよい。組織試料は、癌性細胞、非癌性細胞、又はこれらの任意の組み合わせを含んでもよい。画像試料は、気管支鏡検査、ＣＴスキャン（低線量ＣＴスキャンなど）、ＶＡＴＳ、又はＴＴＮＡ、又はこれらの任意の組み合わせによって取得できる。

【０１２８】
[175] 試料は、分離して精製した試料であってもよい。試料は、新たに分離した試料であってもよい。新たに分離した試料から細胞を分離し、培養してもよい。試料は１つ以上の細胞を含んでもよい。分離した試料は、細胞の不均一な混合物を含んでもよい。細胞の均一な混合物を含むように試料を精製してもよい。試料には約１００個、約１，０００個、約５，０００個、約１０，０００個、約２０，０００個、約３０，０００個、約４０，０００個、約５０，０００個、約６０，０００個、約７０，０００個、約８０，０００個、約９０，０００個、約１００，０００個、約１５０，０００個、約２００，０００個、約２５０，０００個、約３００，０００個、約３５０，０００個、約４００，０００個、約４５０，０００個、約５００，０００個、約５５０，０００個、約６００，０００個、約６５０，０００個、約７００，０００個、約７５０，０００個、約８００，０００個、約８５０，０００個、約９００，０００個、約９５０，０００個、又はそれ以上の細胞を含んでもよい。試料は、約３０，０００個～約１，０００，０００個の細胞を含んでもよい。試料は、約２０，０００個～約５０，０００個の細胞を含んでもよい。試料は、約１００，０００個～約４００，０００個の細胞を含んでもよい。試料は、約４００，０００個～約８００，０００個の細胞を含んでもよい。

【０１２９】
[176] 試料は上皮細胞が含んでもよい。試料は血球を含んでもよい。試料は、鼻組織、口腔組織（歯肉組織、頬組織、舌組織など）、咽頭組織、喉頭組織、気管組織、気管支組織、肺組織、又はこれらの任意の組み合わせが含まれてもよい。

【０１３０】
[177] 分類器は１つ以上の訓練用試料を用いて訓練されてもよい。分類器は種類の異なる１つ以上の訓練用試料を用いて訓練されてもよい。訓練用試料の種類には、摘出生検、組織切除、針吸引、穿刺吸引法、血液試料、無細胞ＤＮＡ試料、画像又は撮影データ（ＣＴスキャンなど）、又はこれらの任意の組み合わせが含まれてもよい。分類器は、摘出生検や穿刺吸引法による試料など、少なくとも２種類の訓練用試料を用いて訓練されてもよい。分類器は、摘出生検と穿刺吸引法による試料、血液試料など、少なくとも３種類の訓練用試料を用いて訓練されてもよい。分類器は、摘出生検と穿刺吸引法による試料、ＣＴスキャンから得た画像など、少なくとも３種類の訓練用試料を用いて訓練されてもよい。分類器は、摘出生検と穿刺吸引法による試料、血液試料、ＣＴスキャンから得た画像など、少なくとも４種類の訓練用試料を用いて訓練されてもよい。

【０１３１】
[178] 訓練用試料は、一人以上の被検体から採取してもよい。被検体は出生国が異なってもよい。被検体は居住地が異なってもよい。訓練用試料における被検体の出生国は、それぞれ異なる、少なくとも約２ヶ国、３ヶ国、４ヶ国、５ヶ国、６ヶ国、７ヶ国、８ヶ国、９ヶ国、１０ヶ国、１１ヶ国、１２ヶ国、１３ヶ国、１４ヶ国、１５ヶ国、１６ヶ国、１７ヶ国、１８ヶ国、１９ヶ国、又は２０ヶ国であってもよい。訓練用試料における被検体の出生国は、それぞれ異なる、少なくとも約３ヶ国であってもよい。訓練用試料における被検体の出生国は、それぞれ異なる、少なくとも約５ヶ国であってもよい。訓練用試料における被検体の出生国は、それぞれ異なる、少なくとも約１０ヶ国であってもよい。訓練用試料における被検体の出生国は、それぞれ異なる、約２ヶ国～約１０ヶ国であってもよい。訓練用試料における被検体の出生国は、それぞれ異なる、約３ヶ国～約１５ヶ国であってもよい。訓練用試料における被検体の出生国は、それぞれ異なる、約２ヶ国～約２０ヶ国であってもよい。訓練用試料における被検体の居住地は、少なくとも約２ヶ所、３ヶ所、４ヶ所、５ヶ所、６ヶ所、７ヶ所、８ヶ所、９ヶ所、１０ヶ所、１１ヶ所、１２ヶ所、１３ヶ所、１４ヶ所、１５ヶ所、１６ヶ所、１７ヶ所、１８ヶ所、１９ヶ所、又は２０ヶ所のそれぞれ異なる場所であってもよい。訓練用試料における被検体の居住地は、少なくとも約３ヶ所のそれぞれ異なる場所であってもよい。訓練用試料における被検体の居住地は、少なくとも約５ヶ所のそれぞれ異なる場所であってもよい。訓練用試料における被検体の居住地は、少なくとも約１０ヶ所のそれぞれ異なる場所であってもよい。訓練用

試料における被検体の居住地は、約２ヶ所～約１０ヶ所のそれぞれ異なる場所であってもよい。訓練用試料における被検体の居住地は、約３ヶ所～約１５ヶ所のそれぞれ異なる場所であってもよい。訓練用試料における被検体の居住地は、約２ヶ所～約２０ヶ所のそれぞれ異なる場所であってもよい。

【０１３２】

[179] 訓練用試料は、肺癌などの疾病罹患の疑いのある被検体、疾病（肺癌など）の確定診断を受けている被検体、既に疾病（良性肺疾患など）を有している被検体、ＬＤＣＴにより肺結節が特定されている被検体、非喫煙者の被検体、非喫煙者であるが受動喫煙にさらされている被検体、喫煙者の被検体、禁煙している被検体、生涯に少なくとも約１本、約１０本、約２０本、約１００本、約２００本、約３００本、約４００本、約５００本、約６００本、約７００本、約８００本、約９００本、約１，０００本、約２，０００本、約３，０００本、約４，０００本、約５，０００本、約１０，０００本、約１１，０００本、約１２，０００本、約１３，０００本、約１４，０００本、約１５，０００本、約１６，０００本、約１７，０００本、約１８，０００本、約１９，０００本、約２０，０００本、約３０，０００本、約４０，０００本、約５０，０００本、約６０，０００本、約７０，０００本、約８０，０００本、約９０，０００本、約１００，０００本、約２００，０００本、約３００，０００本、約４００，０００本、約５００，０００本、又はそれ以上のタバコ、葉巻、又は電子タバコを吸っている被検体、疾病（肺癌など）を発症する遺伝的リスクが増加している被検体、免疫抑制療法を受けた被検体、慢性肺感染症に罹患している被検体、又はこれらの任意の組み合わせから採取された１つ以上の試料を含んでもよい。一部において、被検体は、生涯に約１本～約１０本のタバコ、葉巻、又は電子タバコを吸っていてもよい。一部において、被検体は、生涯に約１本～約１００本のタバコ、葉巻、又は電子タバコを吸っていてもよい。一部において、被検体は、生涯に約１本～約１０００本のタバコ、葉巻、又は電子タバコを吸っていてもよい。一部において、被検体は、生涯に約１０００本～約１０，０００本のタバコ、葉巻、又は電子タバコを吸っていてもよい。一部において、被検体は、生涯に約１０，０００本～約５０，０００本のタバコ、葉巻、又は電子タバコを吸っていてもよい。一部において、被検体は、生涯に約１０，０００本～約１００，０００本のタバコ、葉巻、又は電子タバコを吸っていてもよい。

【０１３３】

[180] 喫煙者とは、生涯に少なくとも約１本、約５本、約１０本、約２０本、約３０本、約４０本、約５０本、約６０本、約７０本、約８０本、約９０本、約１００本、約２００本、約３００本、約４００本、又は約５００本のタバコ、葉巻、又は電子タバコを吸っている者であってもよい。喫煙者とは、生涯に少なくとも約１００本のタバコ、葉巻、又は電子タバコを吸っている者であってもよい。喫煙者とは、生涯に少なくとも約５００本のタバコ、葉巻、又は電子タバコを吸っている者であってもよい。喫煙者とは、１年に約５箱、約１０箱、約２０箱、約３０箱、約４０箱、又は約５０箱を超えるタバコ、葉巻、又は電子タバコを吸っている者であってもよい。喫煙者とは、１年に約５箱を超えるタバコ、葉巻、又は電子タバコを吸っている者であってもよい。喫煙者とは、１年に約１０箱を超えるタバコ、葉巻、又は電子タバコを吸っている者であってもよい。喫煙者とは、１年に約２０箱を超えるタバコ、葉巻、又は電子タバコを吸っている者であってもよい。喫煙者とは、１年に約３０箱を超えるタバコ、葉巻、又は電子タバコを吸っている者であってもよい。喫煙者とは、１年に約１箱～約１２箱（又はそれ以上）のタバコ、葉巻、又は電子タバコを吸っている者であってもよい。喫煙者とは、１年に約１０箱～約２５箱のタバコ、葉巻、又は電子タバコを吸っている者であってもよい。喫煙者とは、１年に約２５箱～約５０箱のタバコ、葉巻、又は電子タバコを吸っている者であってもよい。

喫煙者とは、１年に約１箱～約５０箱のタバコ、葉巻、又は電子タバコを吸っている者であってもよい。喫煙者とは、１年に約１０箱～約５０箱のタバコ、葉巻、又は電子タバコを吸っている者であってもよい。

【０１３４】

[181] 訓練用試料は、疾病（肺癌など）の陽性診断を受けた喫煙者、疾病（肺癌など）の陰性診断を受けた喫煙者、診断を受けたことがない喫煙者、受動喫煙にさらされ疾病（肺癌など）の陽性診断を受けた非喫煙者、受動喫煙にさらされたが疾病（肺癌など）の陰性診断を受けた非喫煙者、受動喫煙にさらされたが診断を受けたことがない非喫煙者、疾病（肺癌など）の陽性診断を受けた非喫煙者、疾病（肺癌など）の陰性診断を受けた非喫煙者、診断を受けたことがない非喫煙者、又はこれらの任意の組み合わせから採取された１つ以上の試料を含んでもよい。

【０１３５】

[182] 訓練用試料又は検証用試料から１つ以上の種類のゲノム情報を得てもよい。例えば、１つ以上の遺伝子（バイオマーカーパネルの遺伝子）の発現レベルに対して試料のアッセイを行い得る。１つ以上の遺伝子の有無について試料のアッセイを行ってもよい。発現レベル、リードの計数、配列変異体、融合体、ヘテロ接合性の消失（ＬＯＨ）、ミトコンドリア転写物、これらのうち１つ以上、又はこれらの任意の組み合わせについて試料のアッセイを行ってもよい。

【０１３６】

[183] 試料は、同被検体から１回以上採取されてもよい。例えば、被検体からまず第１の試料を採取することができ、それから約１年後に第２の試料を採取することができる。毎日、週に数回、隔週、毎週、隔月、毎月、隔年、毎年、２年ごと、３年ごと、４年ごと、又は５年ごとに同被検体から試料を採集してもよい。一部の例示では、第1試料は特定の時点で採取し、少なくとも１日、２日、３日、４日、５日、６日、１週間、２週間、３週間、１ヵ月、２ヵ月、３ヶ月、４ヶ月、５ヵ月、６ヶ月、１年、２年、３年、４年、５年、又はそれ以上の間を置いて、第２の試料を採取することができる。第２の試料から得た結果を第１の試料から得た結果と比べ、被検体における疾患の進行度、処方された治療又は処置の有効性、疾病を発症するリスクの変化、又はこれらの任意の組み合わせをモニタリングしてもよい。

【０１３７】

[184] 分類器は１つ以上の特徴を探し出せるように訓練されてもよい。特徴は、疾病（肺癌など）、組織型（肺組織など）、集団（類似の遺伝的構造を有する被検体の集まり）、曝露リスク（環境汚染又は受動喫煙）、損傷プロファイル、又はこれらの任意の組み合わせであってもよい。分類器は、スクリーニングアッセイ、診断アッセイ、治療計画、モニタリング計画、又はこれらの任意の組み合わせの一部であってもよい。

【０１３８】

[185] 本開示は、試料を採集した後、本開示における１つ以上の方法により分析するまで、秒、分、時、日、週、月、年単位の期間、試料を保存する方法を提供する。一部において、被検体から採取された試料は、試料の各部分に対してその後異なる方法や工程（保存、細胞分析、妥当性試験、核酸抽出、分子プロファイリング、又はこれらの任意の組み合わせを含むが、これに限らない）を行えるよう、保存や更なる分析の前に細分されてもよい。

【０１３９】

[186] 一部において、試料の一部を保存し、別の部分を更に操作してもよい。操作には、分子プロファイリング、組織染色、核酸（ＲＮＡ又はＤＮＡ）の抽出・検出・定量化、遺伝子発現物質（ＲＮＡ又はタンパク質）の抽出・検出・定量化、固定、検査が含まれるが、これらに限らない。グルタルアルデヒド、ホルムアルデヒド、又はメタノールを使用するなど、本分野で既知の方法により、保存の前、又は保存中に試料を固定してもよい。他の例示において、試料を採取して保存し、その後、試料の各部分に対してその後異なる方法や工程（保存、細胞分析、妥当性試験、核酸抽出、分子プロファイリング、又はこれらの任意の組み合わせを含むが、これに限らない）を行えるように細分してもよい。一部において、試料を採取して分析（例えば細胞分析）を行い、その結果として得られた試料を本明細書に記載された１つ以上の分子プロファイリング方法により更に分析してもよい。その場合、細胞分析と分子プロファイリングの間に、試料を保存しておくことになる。

採取の際に試料を保存して輸送を促進するか、他の分析結果を待ってもよい。他の実施例体において、医師や他の医療従事者からの指示が来るまで試料を保存しておいてもよい。

【０１４０】

[187] 細胞アッセイは、例えば甲状腺腫瘍又は結節を含む様々な腫瘍に対する現行の標準診断法である。本開示の一部の実施形態において、陰性、不明、診断確定、診断未確定などとアッセイを受けた試料に対して次のアッセイを行い、更なる情報を収集してもよい。本開示において、後続のアッセイには、ゲノムＤＮＡ、ＲＮＡ、ｍＲＮＡ発現産物レベル、ｍｉＲＮＡレベル、遺伝子発現産物レベル、又は遺伝子発現産物の選択的スプライシングの分子プロファイリングが含まれ得る。本開示の一部の実施形態において、分子プロファイリングとは、生体試料におけるゲノムＤＮＡの数（コピー数など）及び／又は種類の判定を指す。一部において、数及び／又は種類を、対照試料や正常と見なされる試料と更に比較してもよい。一部の実施形態において、ゲノムＤＮＡを、コピー数多型（コピー数の増加（増幅）又は減少）や変異体（挿入、欠失、切断など）について分析することもできる。分子プロファイリングは、同じ試料、同じ試料の一部、又は本明細書に記載のうちいずれかの方法を用いて新しく得た試料に対して行ってもよい。分子プロファイリングを行う企業は、個人に直接、又は仲介者（医師、第三者の試験センタや研究所、医療従事者など）を介して、追加の試料を要請する場合もある。一部において、試料アッセイの際、分子プロファイリング分野で用いられる方法や組成物を、組織染色法やその他の診断法と組み合わせて使用してもよい。他の例示において、組織染色法やその他の診断法を事前に採用せず、分子プロファイリング分野で用いられる方法や組成物のみを用いて試料を直接アッセイしてもよい。一部において、分子プロファイリングのみを用いた場合でも、又は組織アッセイや他のアッセイと組み合わせて用いた場合でも、当業者であれば、被検体を診断し、治療を提案することができる。一部において、分子プロファイリングのみを用いた場合でも、又は組織アッセイや他のアッセイと組み合わせて用いた場合でも、腫瘍又は腫瘍と疑わしきものの悪性変改を経時的にモニタリングすることができる。

【０１４１】

[188] 本開示の分子プロファイリングは、被検体から採取された１つ以上の試料からタンパク質又は核酸（ＲＮＡ又はＤＮＡ）を抽出し分析する方法を提供する。一部において、採取された試料全体から核酸を抽出する。他の例示において、採取された試料の一部から核酸を抽出する。一部において、核酸を抽出していない試料の部分は、細胞検査又は免疫組織化学法により分析される。一部において、被検体において互いに近接する部位から複数の試料を採取し得る。例えば、最大約５００ミリメートル（ｍｍ）、４００ｍｍ、３００ｍｍ、２００ｍｍ、１００ｍｍ、９０ｍｍ、８０ｍｍ、７０ｍｍ、６０ｍｍ、５０ｍｍ、４０ｍｍ、３０ｍｍ、２０ｍｍ、１０ｍｍ、９ｍｍ、８ｍｍ、７ｍｍ、６ｍｍ、５ｍｍ、４ｍｍ、３ｍｍ、２ｍｍ、１ｍｍ、又はその以下の距離を置いて離れている異なる２ヵ所から２つの異なる試料を採取し得る。一部において、複数の試料（例えば、近接部位から採取されたもの）を異なる方法により分析し得る。例えば、細胞検査又は免疫組織化学法により第１の試料を分析してもよく、分子プロファイリングにより第２の試料を分析してもよい。

【０１４２】

[189] 一部の実施形態において、本開示の方法は、核酸分子（例えば、ＤＮＡ、ＲＮＡ）を、被検体の組織試料から抽出し、核酸シーケンシングライブラリーを生成することを含む。例えば、核酸ライブラリーは、逆転写（ＲＴ－ＰＣＲ）により分離されたＲＮＡから生成されたｃＤＮＡを増幅することにより生成することができる。一部において、ｃＤＮＡは、ポリメラーゼ連鎖反応（ＰＣＲ）により増幅され得る。

【０１４３】

分類器

[190] データの固有特性を調べることにより特徴の関連性を評価するFilter型手法、特徴サブセット検索にモデル仮設を埋め込むWrapper型手法、最適の特徴セットの検索を分類器アルゴリズムに組み立てるEmbedded型手法を含む、特徴選択法を用いて、試料の強度

値を分析することができる。

【０１４４】

[191] 本開示の方法に有用に用いられるFilter型手法には、（１）二標本ｔ検定、ＡＮＯＶＡ分析、ベイズフレームワーク、ガンマ分布モデルなどを使用するパラメトリック法、（２）Wilcoxon順位和検定、級間・級内平方和検定、順位積法、ランダム順列法、又は２つのデータ間の発現の円貨倍率ついて閾値を設定し、誤分類の数を最小化する各遺伝子の閾値を検出することを含むＴＮｏＭを使用するモデルフリー法、（３）二変量法、相関性による特徴選択法（ＣＦＳ）、最小重複性・最大関連性法（ＭＲＭＲ）、マルコフブランケットフィルター法、非層間収縮重心法のような多変量法が含まれる。本開示の方法に有用に用いられるWrapper型手法には、逐次検索法、遺伝的アルゴリズム、分布推定アルゴリズムが含まれる。本開示の方法に有用に用いられるEmbedded型手法には、ランダムフォレストアルゴリズム、サポートベクターマシン重みベクターアルゴリズム、ロジスティクス回帰重みアルゴリズムが含まれる。Bioinformatics. 2007 Oct 1; 23(19):2507-17では、強度データ分析に用いられる上記のFilter型手法の利点を概説している。

【０１４５】

[192] 次いで、分類器アルゴリズムを用いて選択された特徴を分類し得るる。アルゴリズムの例示には、主成分分析アルゴリズム、部分最小二乗法、独立成分分析アルゴリズムなど、変数の数を低減させる方法が含まれるが、これに限らない。また、アルゴリズムの例示には、統計的方法や機械学習に基づく方法など、大量の変数を直接取り扱う方法が含まれるが、これに限らない。統計的方法には、罰則付きロジスティック回帰、マイクロアレイ予測分析（ＰＡＭ）、収縮重心(shrunken centroids)に基づく方法、サポートベクターマシン分析、正則化線形判別分析が含まれる。機械学習には、バギング法、ブースティング法、ランダムフォレストアルゴリズム、及びこれらの任意の組み合わせが含まれる。Cancer Inform. 2008; 6: 77-97では、マイクロアレイ強度データの分析に用いられる上記の分類手法を概説している。

【０１４６】

[193] 本方法及びアルゴリズムは、１）低量及び／又は低品質の核酸を含有する試料の遺伝子発現分析、２）偽陽性率及び偽陰性率の顕著な低下、３）結果として生じる病理の原因となる遺伝的経路、代謝経路、又はシグナル伝達経路の確定、４）遺伝性障害の診断精度に対する統計的確率、疾病を発症するリスク、疾病における変化のモニタリング、介入療法の有効性、又はこれらの任意の組み合わせの割り当て、５）不明確な結果の解消、６）肺疾病又は肺疾病の亜型との区別、を可能とする。

【０１４７】

[194] 一部の実施形態において、本開示の方法は、特定の生体試料の細胞構成を判定し、結果として得た分子プロファイリング指標を他の細胞型及び／又は組織型の存在による希釈効果に対して較正させる、アップフロント法を提供する。一態様において、アップフロント法は、試料の各成分に対するアップフロントのミニ分類器として既知の細胞及び／又は組織に特異的な遺伝子発現パターンの組み合わせる用いるアルゴリズムである。アルゴリズムは、この分子指紋を用いることで、組成に応じて試料を予備分類し、補正／正規化因子を適用する。このデータは、場合によって、最終診断を補助する情報を取り組む最終分類アルゴリズムに送り込まれ得る。

【０１４８】

[195] 生遺伝子の発現レベルと選択的スプライシングデータは、一部において、データの正規化を目的として、及び／又はデータの信頼性を高めることを目的として設計されたアルゴリズムを適用することで改善することができる。本開示の一部の実施形態において、データ分析では処理すべき個々のデータポイントの数を非常に多いので、本明細書に記載の様々なアルゴリズムを適用するためには、コンピュータや他の装置、マシン、又は機器を必要とする。「機械学習アルゴリズム」とは、遺伝子発現プロファイルの特徴付けに採用されるコンピュータ実行型予測の方法論であり、当業者には「分類器」として公知である。通常、例えばマイクロアレイに基づくハイブリダイゼーション分析から得得る、特

定の発現レベルに対応するシグナルを、発現プロファイル分類のために、アルゴリズムにかけ得る。教師付きの学習では、一般的に、分類器を「訓練」してクラス間の差異を認識させ、独立したテストセットを以って分類器の精度を「テスト」する。新たに採取された未知の試料については、分類器を用いて試料が属するクラスを予測することができる。

【０１４９】

[196] 一部において、ロバストマルチアレイ平均（ＲＭＡ）法を用いて生データを正規化してもよい。ＲＭＡ法では、まず、多数のマイクロアレイ上のマッチする各細胞についてバックグラウンド補正強度を算出する。そしてIrizarry et al. Biostatistics 2003 Apr. 4 (2): 249-64に記載されているように、バックグラウンド補正値を正の値に制限し得る。バックグラウンド補正の後、マッチングとバックグラウンド補正を行った各細胞の強度の底を２とする対数が得られ得る。各マイクロアレイ上の、バックグラウンド補正、対数変換、マッチングをかけた強度を、各入力アレイ及び各プローブ発現値に対してアレイパーセンタイルプローブ値を全アレイパーセンタイル点の平均値と取り替えられる分立正規化法を用いて正規化し得る。この方法につき、より詳しくはBolstad et al. Bioinformatics 2003に記載され得る。各マイクロアレイ上の各プローブに対する発現測定値を得るために、分立正規化の後、正規化データを線形モデルに適合させることができる。TukeyのMedian Polishアルゴリズム(Tukey, J. W., Exploratory Data Analysis. 1977)を使用し、正規化プローブセットデータについて対数スケールの発現レベルを測定してもよい。

【０１５０】

[197] データを更にフィルタリングし、疑わしいデータを取り除いてもよい。一部の実施形態において、約１個、２個、３個、４個、５個、６個、７個、又は８個より少ないグアノシン＋シストシンヌクレオチドを有するマイクロアレイプローブから由来するデータは、異常ハイブリダイゼーションが生じる傾向や二次構造上の問題のため、信頼性が低いと考えられ得る。約４個以上のグアノシン＋シストシンヌクレオチドを有するマイクロアレイプローブは、信頼性が低いと考えられ得る。約６個以上のグアノシン＋シストシンヌクレオチドを有するマイクロアレイプローブは、信頼性が低いと考えられ得る。約８個以上のグアノシン＋シストシンヌクレオチドを有するマイクロアレイプローブは、信頼性が低いと考えられ得る。約４個～約８個のグアノシン＋シストシンヌクレオチドを有するマイクロアレイプローブは、信頼性が低いと考えられ得る。同様に、約８個、９個、１０個、１１個、１２個、13個、１４個、１５個、１６個、１７個、１８個、１９個、２０個、２１個、２２個、２３個、２４個、２５個のグアノシン＋シストシンヌクレオチドを有するマイクロアレイプローブから由来するデータは、異常ハイブリダイゼーションが生じる傾向や二次構造上の問題のため、信頼性が低いと考えられ得る。約１０個以上のグアノシン＋シストシンヌクレオチドを有するマイクロアレイプローブは、信頼性が低いと考えられ得る。約１５個以上のグアノシン＋シストシンヌクレオチドを有するマイクロアレイプローブは、信頼性が低いと考えられ得る。約２０個以上のグアノシン＋シストシンヌクレオチドを有するマイクロアレイプローブは、信頼性が低いと考えられ得る。約２５個以上のグアノシン＋シストシンヌクレオチドを有するマイクロアレイプローブは、信頼性が低いと考えられ得る。約８個～約３０個のグアノシン＋シストシンヌクレオチドを有するマイクロアレイプローブは、信頼性が低いと考えられ得る。約１０個～約３０個のグアノシン＋シストシンヌクレオチドを有するマイクロアレイプローブは、信頼性が低いと考えられ得る。約１２個～約３０個のグアノシン＋シストシンヌクレオチドを有するマイクロアレイプローブは、信頼性が低いと考えられ得る。約１５個～約３０個のグアノシン＋シストシンヌクレオチドを有するマイクロアレイプローブは、信頼性が低いと考えられ得る。

【０１５１】

[198] 一部において、信頼性の低いプローブセットを選択し、一連の参照データセットに対してプローブセットの信頼性にランクを付けることにより、データ分析から除外してもよい。例えば、RefSeqやEnsembl (EMBL)は、極めて高品質の参照データセットと考えられ得る。RefSeqやEnsembl配列とマッチするプローブセットから得たデータは、一部にお

いて、その高い信頼性により、特にマイクロアレイ分析実験に使用され得る。同様に、信頼性の低い参照データセットとマッチするプローブセットから得たデータは、更なる分析から取り除かれてもよく、含めるか否かを個別的に検討されてもよい。一部において、EnsemblのハイスループットｃＤＮＡ及び／又はｍＲＮＡ参照データセットは別々に、又は共に、プローブセットの信頼性判定に用いられ得る。他の例示において、プローブセットの信頼性はランク付けをされ得る。例えば、全ての参照データセットに完璧にマッチするプローブ及び／又はプローブセットは、最も信頼性が高いものとしてランク付けをされ得る（１）。更に、３つの参照データセットのうち２つとマッチするプローブ及び／又はプローブセットは、二番目に信頼性が高いものとしてランク付され得る（２）。３つの参照データセットのうち１つとマッチするプローブ及び／又はプローブセットは、三番目に信頼性が高いものとしてランク付けされ得る（２）。そして、マッチする参照データがないものは、信頼性が最も低いものとしてランク付けされ得る（４）。ランクに基づいて、プローブ又はプローブセットを分析に含んでもよく、取り除いてもよい。例えば、更なる分析に用いるために、カテゴリー１、２、３、４のプローブセットからのデータを選んでもよく、カテゴリー１、２、３のプローブセットからのデータを選んでもよく、カテゴリー１及び２のプローブセットからのデータを選んでもよく、又はカテゴリー１のプローブセットからのデータのみを選んでもよい。別の例において、プローブセットは、参照データセットエントリーに対する塩基対ミスマッチの数に基づいてランク付けされてもよい。言い換えれば、分子プロファイリングに用いられる所定のプローブ及び／プローブセットの信頼性を評価する従来の方法が多数あり得、本開示の方法は、これらの方法やその組み合わせを網羅するものである。

【０１５２】

[199] 遺伝子発現レベル、又は選択的スプライシングのデータ分析法には、本明細書に記載の特徴選択アルゴリズムの使用が更に含まれ得る。本開示の一部の実施形態において、ＬＩＭＭＡソフトウェアパッケージ（Smyth, G. K. (2005). Limma: linear models for microarray data. In: Bioinformatics and Computational Biology Solutions using R and Bioconductor, R. Gentleman, V. Carey, S. Dudoit, R. Irizarry, W. Huber (eds.), Springer, New York, pages 397-420）を用いて特徴選択を行う。

【０１５３】

[200] 遺伝子発現レベル及び／又は選択的スプライシングのデータ分析法には、予備分類器アルゴリズムの使用が更に含まれ得る。例えば、アルゴリズムは、細胞特異的な分子指紋を使用することで、組成に応じて試料を予備分類して、補正／正規化因子を適用し得る。次いで、このデータ／情報を、最終診断、予後、又はモニタリング評価を補助する情報を取り組む最終分類アルゴリズムに送り込み得る。

【０１５４】

[201] 遺伝子発現レベル及び／又は選択的スプライシングのデータ分析法には、本明細書に記載の分類器アルゴリズムの使用が更に含まれ得る。本開示の一部の実施形態において、サポートベクターマシン（ＳＶＭ）アルゴリズム、ランダムフォレストアルゴリズム、又はこれらの任意の組み合わせが、マイクロアレイデータ分類に用いられる。一部の実施形態において、試料の区別（例えば、良性か悪性か、正常か悪性か、低リスクか高リスクか）、又は種類の区別（例えば、ＩＬＤか肺癌か）を特定するマーカーを、統計的優位性に基づいて選択し得る。一部において、偽検出率（ＦＤＲ）に対するBenjamini Hochberg補正を適用し、統計的優位性選択を行う。

【０１５５】

[202] 一部において、Fishel and Kaufman et al. 2007 Bioinformatics 23(13): 1599-606に記載されているメタ分析アプローチにより分類器アルゴリズムを補足し得る。一部において、再現性分析などのメタ分析アプローチにより分類器アルゴリズムを補足し得る。一部において、再現性分析では、少なくとも１つの予測的発現産物マーカーセットに出現するマーカーを選択する。

【０１５６】

[203] 一部において、ヘイズ事後分析法を用いて、特徴選択・分類の結果のランク付けを行い得る。例えば、本明細書に記載のものなど、既知の方法を用いてマイクロアレイデータに対して抽出、正規化、要約を行い得る。次いで、本明細書に記載のアルゴリズム又は方法（ＬＩＭＭＡにより提供される特徴選択法を含むが、これに限らない）など、既知の特徴選択法を適用し、データに特徴分析をかけ得る。その後、本明細書に記載の方法（ＳＶＭやランダムフォレストアルゴリズムの使用を含むが、これに限らない）など、既知のアルゴリズムや方法を適用し、データに分類をかけ得る。次いで、事後確率関数に従って分類器アルゴリズムの結果をランク付けし得る。例えば、マーカーをカテゴリー（例えばＩＬＤ、ＣＯＰＤ、肺癌など）に割り当てる際の第１の種及び第２の種の誤差率から事後確率を導出するために、既知の分子プロファイリング結果、例えば出版公開された結果を検査して事後確率関数を導出し得る。誤差率は、推定変化倍率値（例えば、１．１倍、１．２倍、１．３倍、１．４倍、１．５倍、１．６倍、１．７倍、１．８倍、１．９倍、２倍、２．２倍、２．４倍、２．５倍、３倍、４倍、５倍、６倍、７倍、８倍、９倍、１０倍、又はそれ以上）を用いて、各研究で報告された試料サイズに基づき計算され得る。変化倍率値は、約０．５倍、約０．８倍、約１．０倍、約１．１倍、約１．２倍、約１．３倍、約１．４倍、約１．５倍、約１．６倍、約１．７倍、約１．８倍、約１．９倍、約２．０倍、約２．１倍、約２．２倍、約２．３倍、約２．４倍、約２．５倍、約２．６倍、約２．７倍、約２．８倍、約２．９倍、約３．０倍、約３．５倍、約４．０倍、約４．５倍、約５．０倍、約５．５倍、約６．０倍、約６．５倍、約７．０倍、約７．５倍、約８．０倍、約８．５倍、約９．０倍、約９．５倍、又は約１０．０倍であってもよい。変化倍率値は、約０．５倍～約１０．０倍であってもよい。変化倍率値は、約０．５倍～約１．０倍であってもよい。変化倍率値は、約０．５倍～約５．０倍であってもよい。変化倍率値は、約２．０倍～約８．０倍であってもよい。変化倍率値は、約２．０倍～約６．０倍であってもよい。変化倍率値は、約６．０倍～約１０．０倍であってもよい。変化倍率値は、約５．０倍～約１０．０倍であってもよい。変化倍率値は、約８．０倍～約１０．０倍であってもよい。次いで、事前確率を、本開示の分子プロファイリングデータセットと組み合わせ、差次的な遺伝子発現の事後確率を推定し得る。最後に、事後確率推定値を、本開示の第２のデータセットと組み合わせ、差次的な発現の最終事後確率を公式化し得る。事後確率を導出しマイクロアレイデータの分析に適用する付加的な方法は、本技術分野において既知のものであり得、例えば、Smyth, G.K. 2004 Stat. Appl. Genet. Mol. Biol. 3: Article 3に記載されている。一部において、事後確率を用いて分類器アルゴリズムにより提供されるマーカーをランク付けしてもよい。一部において、事後確率に従ってマーカーのランク付けを行い、所定の閾値を超えるものを、例えば、良性、悪性、正常、低リスク、高リスク、又は疾病の種類（ＩＬＤ、ＣＯＰＤ、肺癌）などの試料を示す、又はその試料に対する診断結果を示す差次的発現を有するマーカーとして選んでもよい。閾値の例には、少なくとも約０．７、約０．７５、約０．８、約０．８５、約０．９、約０．９２５、約０．９５、約０．９７５、約０．９８、約０．９８５、約０．９９、約０．９９５、又はそれ以上の事前確率が含まれる。確率は少なくとも約０．７であってもよく、少なくとも約０．７５であってもよく、少なくとも約０．８であってもよく、少なくとも約０．８５であってもよく、少なくとも約０．９であってもよく、少なくとも約０．９５であってもよく、少なくとも約０．９９であってもよい。確率は約０．７５～約０．９９５であってもよく、約０．８０～約０．９９５であってもよく、約０．８５～約０．９９５であってもよく、約０．９～約０．９９５であってもよく、約０．８５～約０．９５であってもよく、約０．８～約０．９５であってもよく、約０．７５～約０．９５であってもよい。

【０１５７】

[204] 分子プロファイリング結果の統計的評価は、診断精度の尤度、癌、疾患又は疾病の尤度、特定の癌、疾患又は疾病の尤度、特定の介入療法成功の尤度のうち１つ以上を示す定量値を提供してもよい。よって、遺伝学や分子生物学の素養を持たない疑いの高い医師が生データを理解する必要はなくなる。むしろデータは、患者の治療ガイドに最も有用

な形式で直接医師に提示され得る。分子プロファイリングの結果は、既知の方法（スチューデントのＴ検定、両側Ｔ検定、ピアソンの順位和検定、隠れマルコフモデル分析、ｑ－ｑプロットの分析、主成分分析、一元配置のＡＮＯＶＡ、二元配置のＡＮＯＶＡ、ＬＩＭＭＡなどを含むが、これに限らない）を用いて統計的に評価されることができる。

【０１５８】

[205] 本開示の一部の実施形態において、訓練アルゴリズムを用いて結果を分類し得る。本開示の訓練アルゴリズムには、既知の良性試料、悪性試料、正常試料の参照セットを用いて開発されたアルゴリズムが含まれる。訓練用試料は、ＦＮＡ試料、摘出生検試料、気管支鏡試料、又はこれらの任意の組み合わせを含んでもよい。試料のカテゴリー化に適合するアルゴリズムには、ｋ近傍アルゴリズム、概念ベクトルアルゴリズム、単純ベイズアルゴリズム、ニューラルネットワークアルゴリズム、隠れマルコフモデルアルゴリズム、遺伝的アルゴリズム、相互情報特徴選択アルゴリズム、又はこれらの任意の組み合わせが含まれ得るが、これに限らない。一部において、本開示の訓練アルゴリズムは、遺伝子発現又は選択的スプライシングのデータ以外のデータ（ＤＮＡ多型データ、シーケンシングデータ、本開示の細胞学者や病理学者による採点又は診断、本開示の予備分類器アルゴリズムにより提供される情報、又は本開示における被検体の病歴に関する情報を含むが、これに限らない）を取り組んでもよい。

【０１５９】

[206] 逐次分析の初期に使用され得る分類器により、良性又は疑わしい試料、低リスクや高リスクの試料、ＩＬＤなしの試料やＩＬＤありの試料を選択又は除外することができる。一部の実施形態において、連続分析は、前の分類器により除外されていない試料のデータに「主」分類器を適用することで終了する。「主」分類器は、多数の種類の組織における遺伝子発現レベルのデータ分析から得られ、また試料を「良性」又は「疑わしい」（又は悪性）として指定することができる。

【０１６０】

[207] 例示の分類プロセスにおける次のステップでは、試料の遺伝子発現レベルとバイオマーカーの第１のセット又は第１の分類器とを比べる第１の比較を行い得る。第１の比較の結果が「マッチ」であれば、「肺疾病を発症するリスク低」又は「高」の試料、又はＩＬＤ又は肺癌を有する試料である、などの結果を出力して、分類プロセスを終了する。結果が「マッチ」でなければ、試料の遺伝子発現レベルを、第２の比較としてバイオマーカーの第２のセット又は第２の分類器と比べ得る。第２の比較の結果が「マッチ」であれば、（ａ）被検体は肺疾病に罹患しているとの診断と報告、（ｂ）肺疾病を発症するリスクありとの報告、（ｃ）介入療法の有効性の報告、（ｄ）画像スキャン、他の試料採取法、気管支鏡検査、生検、外科切除、医薬組成物など、次にとるべき処置の推奨などの結果を出力して、分類プロセスを終了する。結果が「マッチ」でなければ、「マッチ」するまで、又は分類プロセスに含まれるバイオマーカーセットや分類器を全て使い切るまで、同様の段階的比較プロセスを繰り返す。一部の実施形態において、分類プロセスの最終比較は、本明細書に記載されているように、試料の遺伝子発現レベルと、主分類器との比較になる。

【０１６１】

[208] 一部において、１つ以上の機械学習アルゴリズムを方法に採用してもよい。例えば、約２個、約３個、約４個、約５個、約６個、約７個、約８個、約９個、約１０個、約１１個、約１２個、約１３個、約１４個、約１５個、約１６個、約１７個、約１８個、約１９個、約２０個、又はそれ以上の機械学習アルゴリズムを方法に採用してもよい。一部において、少なくとも約４個の機械学習アルゴリズムを方法に採用してもよい。一部において、少なくとも約５個の機械学習アルゴリズムを方法に採用してもよい。一部において、少なくとも約６個の機械学習アルゴリズムを方法に採用してもよい。一部において、少なくとも約７個の機械学習アルゴリズムを方法に採用してもよい。一部において、少なくとも約８個の機械学習アルゴリズムを方法に採用してもよい。一部において、少なくとも約９個の機械学習アルゴリズムを方法に採用してもよい。一部において、少なくとも約１

０個の機械学習アルゴリズムを方法に採用してもよい。一部において、少なくとも約４個～約１０個の機械学習アルゴリズムを方法に採用してもよい。一部において、少なくとも約６個～約１０個の機械学習アルゴリズムを方法に採用してもよい。一部において、少なくとも約４個～約８個の機械学習アルゴリズムを方法に採用してもよい。一部において、少なくとも約４個～約１５個の機械学習アルゴリズムを方法に採用してもよい。また、１つ以上の機械学習アルゴリズムを連続して方法に採用してもよい。一部において、機械学習アルゴリズムと融合呼び出しアルゴリズムの組み合わせを方法に採用してもよい。例えば、少なくとも１個の機械学習アルゴリズムと少なくとも１個の融合呼び出しアルゴリズムを方法に採用する。一部において、少なくとも５個の機械学習アルゴリズムと少なくとも１個の融合呼び出しアルゴリズムを方法に採用してもよい。一部において、少なくとも７個の機械学習アルゴリズムと少なくとも１つの融合呼び出しアルゴリズムを方法に採用してもよい。

【０１６２】

[209] 本方法及びシステムは、試料における一つ以上のバイオマーカーの存在又は非存在を同定し得る。例えば、バイオマーカーは、表１、２、３、４、５、６、７、８、９、１０、１１、１２、１３、又はそれらの任意の組合せからのバイオマーカーを含むことができる。いくつかの場合において、バイオマーカーは、表１、表２、又はそれらの組み合わせからのバイオマーカーを含み得る。いくつかの場合において、バイオマーカーは、表１、表２、表３、又はそれらの任意の組合せからのバイオマーカーを含み得る。いくつかの場合において、バイオマーカーは、表４、表５、表６、表７、又はそれらの任意の組合せからのバイオマーカーを含み得る。いくつかの場合において、バイオマーカーは、表８、表９、表１０又はそれらの任意の組合せからのバイオマーカーを含み得る。いくつかの場合において、バイオマーカーは、表１１、表１２、表１３、又はそれらの任意の組合せからのバイオマーカーを含み得る。いくつかの場合において、バイオマーカーは、表１からのバイオマーカー又はそれらの任意の組合せを含むことができる。いくつかの場合において、バイオマーカーは、表２からのバイオマーカー又はそれらの任意の組合せを含み得る。いくつかの場合において、バイオマーカーは、表３からのバイオマーカー又はそれらの任意の組合せを含むことができる。いくつかの場合において、バイオマーカーは、表４からのバイオマーカー又はそれらの任意の組合せを含み得る。いくつかの場合において、バイオマーカーは、表５からのバイオマーカー又はそれらの任意の組合せを含むことができる。いくつかの場合において、バイオマーカーは、表６からのバイオマーカー又はそれらの任意の組合せを含み得る。いくつかの場合において、バイオマーカーは、表７からのバイオマーカー又はそれらの任意の組合せを含むことができる。いくつかの場合において、バイオマーカーは、表８からのバイオマーカー又はそれらの任意の組合せを含み得る。いくつかの場合において、バイオマーカーは、表９からのバイオマーカー又はそれらの任意の組合せを含むことができる。いくつかの場合において、バイオマーカーは、表１０からのバイオマーカー又はそれらの任意の組合せを含み得る。いくつかの場合において、バイオマーカーは、表１１からのバイオマーカー又はそれらの任意の組合せを含み得る。いくつかの場合において、バイオマーカーは、表１２からのバイオマーカー又はそれらの任意の組合せを含み得る。いくつかの場合において、バイオマーカーは、表１３からのバイオマーカー又はそれらの任意の組合せを含み得る。

【０１６３】

[210] １つ以上のバイオマーカーの有無、又は差次的発現は、肺癌、ＩＰＦ、ＩＬＤ、ＣＯＰＤ、又はこれらの任意の組み合わせなどの疾病を発症する１つ以上のリスク因子の存在を示すものであってもよい。１つ以上のバイオマーカーの有無、又は差次的発現は、疾病（肺癌、ＩＰＦ、ＩＬＤ、ＣＯＰＤなど）の予防又は治療を目的とする介入療法の有効性を特定するものであってもよい。１つ以上のバイオマーカーの有無、又は差次的発現は、被検体における疾病（肺癌、ＩＰＦ、ＩＬＤ、ＣＯＰＤなど）のリスク又は寛解を特定するものであってもよい。１つ以上のバイオマーカーの有無、又は差次的発現は、疾病（肺癌、ＩＰＦ、ＩＬＤ、ＣＯＰＤなど）に罹患する喫煙者と罹患しない喫煙者を区別す

るものであってもよい。１つ以上のバイオマーカーの有無、又は差次的発現は、疾病（肺癌、ＩＰＦ、ＩＬＤ、ＣＯＰＤなど）の診断、疾病（肺癌、ＩＰＦ、ＩＬＤ、ＣＯＰＤなど）の予後、又はこれらの任意の組み合わせを特定するものであってもよい。１つ以上のバイオマーカーの有無、又は差次的発現は、損傷範囲を特定するものであってもよい。１つ以上のバイオマーカーの有無、又は差次的発現は、第１の部位から採取された第１の細胞型又は第１の細胞の発現プロファイルと、第２の部位から採取された第２の細胞型又は第２の細胞の発現プロファイルとの関係を特定するものであってもよい。例えば、鼻組織における１つ以上のバイオマーカーの有無、又は差次的発現は、気管支組織における疾病（肺癌、ＩＰＦ、ＩＬＤ、ＣＯＰＤなど）の存在を示すものであってもよい。

【０１６４】

[211]

【表１】

表１－IPFにおいて上方制御されるバイオマーカーの例

Mmp7/MMP7	マトリックスメタロペプチダーゼ7(マトリライシン、子宮)
Pla2g2a/PLA2G2A	ホスホリパーゼA2、グループIIA(血小板、滑液)
Lcn2/LCN2	リポカリン2
Cthrc1/CTHRC1	1含有コラーゲントリプルヘリックスリピート
C6/C6	補体成分6
Ctse/CTSE	カテプシンE
Dclk1/DCLK1	ダブルコルチン様キナーゼ1
Anln/ANLN	アニリン、アクチン結合タンパク質
Kcnn4/KCNN4	カリウム中間体／低コンダクタンスカルシウム活性化チャネル、サブファミリーN、メンバー4
Aspn/ASPN	アスポリン
Pkib/PKIB	プロテインキナーゼ(cAMP依存、触媒)阻害剤β
Fhl2/FHL2	4つ半のLIMドメイン2
Mnd1/MND1	減数分裂核分裂1ホモログ(*Saccharomyces cerevisiae*)
Mycn/MYCN	V-myc myelocytomatosisウイルス関連癌遺伝子、神経芽腫由来(鳥類)
Calca/CALCA	カルシトニン関連ポリペプチドα
Slc2a5/SLC2A5	溶質キャリアファミリー2(促進グルコース／フルクトーストランスポーター)、メンバー5
Fkbp11/FKBP11	FK506結合タンパク質11, 19 kD
Gdf15/GDF15	成長分化因子15
Gal/GAL	ガラニンプレプロペプチド
Top2a/TOP2A	トポイソメラーゼ(DNA)IIα、170 kD
Tmem213/TMEM213	膜貫通タンパク質213
Podnl1/PODNL1	ポドカン様1
Pln/PLN	ホスホランバン
Mia/MIA	メラノーマ阻害活性
Bik/BIK	BCL2-相互作用キラー(アポトーシス誘導)
Col1a2/COL1A2	コラーゲンI型、α2型
Ccnb2/CCNB2	サイクリンB2
MGC105649/C15orf48	染色体15オープンリーディングフレーム48
Ptges/PTGES	プロスタグランジンEシンターゼ
Ctsk/CTSK	カテプシンK
Nuf2/NUF2	NUF2、NDC80キネトコア複合体コンポーネント、ホモログ(*S. cerevisiae*)
Bub1b/BUB1B	ベンズイミダゾール1ホモログβ(酵母)により阻害されない出芽
Fap/FAP	線維芽細胞活性化タンパク質, α
Col5a1/COL5A1	コラーゲンV型、α1型
Fkbp10/FKBP10	FK506結合タンパク質10, 65 kD
Uchl1/UCHL1	ユビキチンカルボキシル末端エステラーゼL1(ユビキチンチオエステラーゼ)
Pla2g7/PLA2G7	ホスホリパーゼA2、グループVII(血小板活性化因子アセチルヒドロラーゼ、血漿)
Spc25/SPC25	SPC25、NDC80キネトコア複合体コンポーネント、ホモログ(*S. cerevisiae*)
Mlf1ip/MLF1IP	MLF1相互作用タンパク質
Sel1l3/SEL1L3	lin-12-like 3のsel-1サプレッサー(*Caenorhabditis elegans*)
Foxm1/FOXM1	フォークヘッドボックスM1

【０１６５】

[212]

【表２】

表２－IPF において下方制御されるバイオマーカーの例

Esm1/ESM1	内皮細胞特異的分子 1
Tmem100/TMEM100	膜貫通タンパク質 100
Stxbp6/STXBP6	シンタキシン結合タンパク質 6 (アミシン)
Gcom1/GCOM1	GRINL1A 複合遺伝子座 1
Hpgd/HPGD	ヒドロキシプロスタグランジンデヒドロゲナーゼ 15-(NAD)
Vegfa/VEGFA	血管内皮増殖因子 A
Mme/MME	膜メタロエンドペプチダーゼ
Emp2/EMP2	上皮膜タンパク質 2
Slc1a1/SLC1A1	溶質キャリアファミリー1 (ニューロン／上皮の高親和性グルタミン酸トランスポーター、システム Xag)、メンバー1
Clic5/CLIC5	塩化物細胞内チャネル 5
Ptprr/PTPRR	プロテインチロシンホスファターゼ、受容体型、R
Anxa3/ANXA3	アネキシン A3
Lrrn3/LRRN3	ロイシンリッチリピートニューロン 3
Rapgef5/RAPGEF5	ラップグアニンヌクレオチド交換因子(GEF) 5
Olfml2a/OLFML2A	オルファクトメジン様 2A
Sgef/ARHGEF26	ローグアニンヌクレオチド交換因子(GEF) 26
Sdpr/SDPR	血清欠乏対応
Adrb2/ADRB2	アドレナリン受容体 β2、表面
Ramp2/RAMP2	受容体(G タンパク質共役)活性修飾タンパク質 2
Ccdc68/CCDC68	68 を含むコイルドコイルドメイン
RGD1306437/C13orf1	NA
Cav2/CAV2	カベオリン 2
Npr3/NPR3	ナトリウム利尿ペプチド受容体 C／グアニル酸シクラーゼ C
Tal1/TAL1	T 細胞急性リンパ性白血病 1
Lifr/LIFR	白血病抑制因子受容体 α
Prkce/PRKCE	プロテインキナーゼ C、ε
Cav1/CAV1	カベオリン 1、カベオラタンパク質、22kD
RGD1311307/C6orf145	NA
Nebl/NEBL	ネブレット
Nedd9/NEDD9	神経前駆細胞の発現、発達上の下方制御 9
S1pr5/S1PR5	スフィンゴシン-1-リン酸受容体 5
Afap1l1/AFAP1L1	アクチンフィラメント関連タンパク質 1 様 1
Thbd/THBD	トロンボモジュリン
Pard6b/PARD6B	Par-6 分割欠陥 6 ホモログ β (*C. elegans*)
Radil/RADIL	Ras アソシエーションと DIL ドメイン
Dnase2b/DNASE2B	デオキシ RNase II β
LOC691221/C5orf4	染色体 5 オープンリーディングフレーム 4
Sh3bp5/SH3BP5	SH3 ドメイン結合タンパク質 5 (BTK 関連)
Fgg/FGG	フィブリノゲン γ 鎖
Epb4.1l5/EPB41L5	赤血球膜タンパク質バンド 4.1 様 5
Tspan12/TSPAN12	テトラスパニン 12
Slc4a1/SLC4A1	溶質キャリアファミリー4、陰イオン交換体、メンバー1 (赤血球膜タンパク質バンド 3、ディエゴ血液型)
Zfp365/ZNF365	ジンクフィンガープロテイン 365
Phactr1/PHACTR1	ホスファターゼ及びアクチン調節因子 1
Gpd1/GPD1	グリセロール-3-リン酸デヒドロゲナーゼ 1 (可溶性)
Veph1/VEPH1	PH ドメインホモログ 1 発現の心室帯 (zebrafish)
Selenbp1/SELENBP1	セレン結合タンパク質 1

【０１６６】

【表３－１】

表３－ＣＯＰＤにおいて差次的に発現され得るバイオマーカーの例

遺伝子
PCDH7
CCDC81
CEACAM5
PTPRH
C12orf36
B3GNT6
PLAG1
PDE7B
CACHD1
EPB41L2
FRMD4A
PRKCE
SULF1
TLE1
FAM114A1
ELF5
SGCE
SEC14L3
GPR155
ITGA9
PTGFR
ISLR
SLC5A7
ZNF483
DPYSL3
TNS3
FMNL2
GALE
CNTN3
HSD17B13
PTPRM
HLF

【表３－２】

遺伝子
PROS1
PLA2G4A
KAL1
TCN1
DPP4
GPR98
KCNA1
CABLES1
PEG10
PPP1R9A
POLA2
C17orf37
ABCC4
CA8
CYP2A13
SETBP1
ANKS1B
CHP
THSD4
MPDU1
CD109
STK32A
HHLA2
AMMECR1
NPAS3
GXYLT2
KLF12
CA12
C21orf121
SH3BP4
FABP6
GUCY1B3
FUT3
STX10
FTO

【表３－３】

遺伝子
CNTN4
ATP8A1
GMDS
ZNF671
WBP5
MYO5B
FLRT3
SCGB1A1
SCNN1G
CFTR
LOC339524
THSD7A
CACNB4
DQX1
GLI3
NFAT5
RUNX1T1
SNTB1
C16orf89
PRKD1
ANXA6
YIPF1
ATP10B
HK2
ABHD2
DNAH5
GGT7
FBN1
PRSS12
TMPRSS4
AMIGO2
TMEM54
CAPRIN2

【０１６７】

【表４－１】

表４－肺癌の喫煙者と肺癌でない喫煙者を区別し得るバイオマーカーの例

Affymetrix ID	GenBank ID	遺伝子名
1316_at	NM_003335	UBE1L
200654_at	NM_000918	P4HB
200877_at	NM_006430.1	CCT4
201530_x_at	NM_001416.1	EIF4A1
201537_s_at	NM_004090	DUSP3
201923_at	NM_006406.1	PRDX4
202004_x_at	NM_003001.2	SDHC
202573_at	NM_001319	CSNK1G2
203246_s_at	NM_006545.1	TUSC4
203301_s_at	NM_021145.1	DMTF1
203466_at	NM_002437.1	MPV17
203588_s_at	NM_006286	TFDP2
203704_s_at	NM_001003698 ///	RREB1
	NM_001003699 ///	
	NM_002955	
204119_s_at	NM_001123 ///	ADK
	NM_006721	
204216_s_at	NM_024824	FLJ11806
204247_s_at	NM_004935.1	CDK5
204461_x_at	NM_002853.1	RAD1
205010_at	NM_019067.1	FLJ10613
205238_at	NM_024917.1	CXorf34
205367_at	NM_020979.1	APS
206929_s_at	NM_005597.1	NFIC
207020_at	NM_007031.1	HSF2BP
207064_s_at	NM_009590.1	AOC2
207283_at	NM_020217.1	DKFZp547I014
207287_at	NM_025026.1	FLJ14107
207365_x_at	NM_014709.1	USP34
207436_x_at	NM_014896.1	KIAA0894
207953_at	AF010144	—
207984_s_at	NM_005374.1	MPP2

【表４－２】

Affymetrix ID	GenBank ID	遺伝子名
208678_at	NM_001696	ATP6V1E1
209015_s_at	NM_005494 ///	DNAJB6
	NM_058246	
209061_at	NM_006534 ///	NCOA3
	NM_181659	
209432_s_at	NM_006368	CREB3
209653_at	NM_002268 ///	KPNA4
	NM_032771	
209703_x_at	NM_014033	DKFZP586A0522
209746_s_at	NM_016138	COQ7
209770_at	NM_007048 ///	BTN3A1
	NM_194441	
210434_x_at	NM_006694	JTB
210858_x_at	NM_000051 ///	ATM
	NM_138292 ///	
	NM_138293	
211328_x_at	NM_000410 ///	HFE
	NM_139002 ///	
	NM_139003 ///	
	NM_139004 ///	
	NM_139005 ///	
	NM_139006 ///	
	NM_139007 ///	
	NM_139008 ///	
	NM_139009 ///	
	NM_139010 ///	
	NM_139011	
212041_at	NM_004691	ATP6V0D1
212517_at	NM_012070 ///	ATRN
	NM_139321 ///	
	NM_139322	
213106_at	NM_006095	ATP8A1
213212_x_at	AI632181	—
213919_at	AW024467	—
214153_at	NM_021814	ELOVL5

【表４－３】

Affymetrix ID	GenBank ID	遺伝子名
214599_at	NM_005547.1	IVL
214722_at	NM_203458	N2N
214763_at	NM_015547 ///	THEA
	NM_147161	
214833_at	AB007958.1	KIAA0792
214902_x_at	NM_207488	FLJ42393
215067_x_at	NM_005809 ///	PRDX2
	NM_181737 ///	
	NM_181738	
215336_at	NM_016248 ///	AKAP11
	NM_144490	
215373_x_at	AK022213.1	FLJ12151
215387_x_at	NM_005708	GPC6
215600_x_at	NM_207102	FBXW12
215609_at	AK023895	—
215645_at	NM_144606 ///	FLCN
	NM_144997	
215659_at	NM_018530	GSDML
215892_at	AK021474	—
216012_at	U43604.1	—
216110_x_at	AU147017	—
216187_x_at	AF222691.1	LNX1
216745_x_at	NM_015116	LRCH1
216922_x_at	NM_001005375 ///	DAZ2
	NM_001005785 ///	
	NM_001005786 ///	
	NM_004081 ///	
	NM_020363 ///	
	NM_020364 ///	
	NM_020420	
217313_at	AC004692	—
217336_at	NM_001014	RPS10
217371_s_at	NM_000585 ///	IL15
	NM_172174 ///	
	NM_172175	

【表４－４】

Affymetrix ID	GenBank ID	遺伝子名
217588_at	NM_054020 ///	CATSPER2
	NM_172095 ///	
	NM_172096 ///	
	NM_172097	
217671_at	BE466926	—
218067_s_at	NM_018011	FLJ10154
218265_at	NM_024077	SECISBP2
218336_at	NM_012394	PFDN2
218425_at	NM_019011 ///	TRIAD3
	NM_207111 ///	
	NM_207116	
218617_at	NM_017646	TRIT1
218976_at	NM_021800	DNAJC12
219203_at	NM_016049	C14orf122
219290_x_at	NM_014395	DAPP1
219977_at	NM_014336	AIPL1
220071_x_at	NM_018097	C15orf25
220113_x_at	NM_019014	POLR1B
220215_at	NM_024804	FLJ12606
220242_x_at	NM_018260	FLJ10891
220459_at	NM_018118	MCM3APAS
220856_x_at	NM_014128	
220934_s_at	NM_024084	MGC3196
221294_at	NM_005294	GPR21
221616_s_at	AF077053	PGK1
221759_at	NM_138387	G6PC3
222155_s_at	NM_024531	GPR172A
222168_at	NM_000693	ALDH1A3
222231_s_at	NM_018509	PRO1855
222272_x_at	NM_033128	SCIN
222310_at	NM_020706	SFRS15
222358_x_at	AI523613	—
64371_at	NM_014884	SFRS14

【０１６８】

【表５－１】

表５－癌の喫煙者と癌でない喫煙者を区別し得るバイオマーカーの例.

GenBank ID	遺伝子名	Affymetrix ID
NM_030757.1	MKRN4	208082_x_at
R83000	BTF3	214800_x_at
AK021571.1	MUC20	215208_x_at
NM_014182.1	ORMDL2	218556_at
NM_17932.1	FLJ20700	207730_x_at
U85430.1	NFATC3	210556_at
AI683552	—	217679_x_at
BC002642.1	CTSS	202901_x_at
AW024467	RIPX	213939_s_at
NM_030972.1	MGC5384	208137_x_at
BC021135.1	INADL	214705_at
AL161952.1	GLUL	215001_s_at
AK026565.1	FLJ10534	218155_x_at
AK023783.1	—	215604_x_at
BF218804	AFURS1	212297_at
NM_001281.1	CKAP1	201804_x_at
NM_024006.1	IMAGE3455200	217949_s_at
AK023843.1	PGF	215179_x_at
BC001602.1	CFLAR	211316_x_at
BC034707.1	—	217653_x_at
BC064619.1	CD24	266_s_at
AY280502.1	EPHB6	204718_at
BC059387.1	MYO1A	211916_s_at
	—	215032_at
AF135421.1	GMPPB	219920_s_at
BC061522.1	MGC70907	211996_s_at
L76200.1	GUK1	200075_s_at
U50532.1	CG005	214753_at
BC006547.2	EEF2	204102_s_at
BC008797.2	FVT1	202419_at
BC000807.1	ZNF160	214715_x_at
AL080112.1	—	216859_x_at

【表５－２】

GenBank ID	遺伝子名	Affymetrix ID
BC033718.1 ///	C21orf106	215529_x_at
BC046176.1 ///		
BC038443.1		
NM_000346.1	SOX9	202936_s_at
BC008710.1	SUI1	212130_x_at
Hs.288575	—	215204_at
(Unigene ID)		
AF020591.1	AF020591	218735_s_at
BC000423.2	ATP6V0B	200078_s_at
BC002503.2	SAT	203455_s_at
BC008710.1	SUI1	212227_x_at
	—	222282_at
BC009185.2	DCLRE1C	219678_x_at
Hs.528304	ADAM28	208268_at
(UNIGENE ID)		
U50532.1	CG005	221899_at
BC013923.2	SOX2	213721_at
BC031091	ODAG	214718_at
NM_007062	PWP1	201608_s_at
Hs.249591	FLJ20686	205684_s_at
(Unigene ID)		
BC075839.1 ///	KRT8	209008_x_at
BC073760.1		
BC072436.1 ///	HYOU1	200825_s_at
BC004560.2		
BC001016.2	NDUFA8	218160_at
Hs.286261	FLJ20195	57739_at
(Unigene ID)		
AF348514.1	—	211921_x_at
BC005023.1	CGI-128	218074_at
BC066337.1 ///	KTN1	200914_x_at
BC058736.1 ///		
BC050555.1		
	—	216384_x_at
Hs.216623	ATP8B1	214594_x_at

【表５－３】

GenBank ID	遺伝子名	Affymetrix ID
(Unigene ID)		
BC072400.1	THOC2	222122_s_at
BC041073.1	PRKX	204060_s_at
U43965.1	ANK3	215314_at
	—	208238_x_at
BC021258.2	TRIM5	210705_s_at
BC016057.1	USH1C	211184_s_at
BC016713.1 ///	PARVA	215418_at
BC014535.1 ///		
AF237771.1		
BC000360.2	EIF4EL3	209393_s_at
BC007455.2	SH3GLB1	210101_x_at
BC000701.2	KIAA0676	212052_s_at
BC010067.2	CHC1	215011_at
BC023528.2 ///	C14orf87	221932_s_at
BC047680.1		
BC064957.1	KIAA0102	201239_s_at
Hs.156701	—	215553_x_at
(Unigene ID)		
BC030619.2	KIAA0779	213351_s_at
BC008710.1	SUI1	202021_x_at
U43965.1	ANK3	209442_x_at
BC066329.1	SDHC	210131_x_at
Hs.438867	—	217713_x_at
(Unigene ID)		
BC035025.2 ///	ALMS1	214707_x_at
BC050330.1		
BC023976.2	PDAP2	203272_s_at
BC074852.2 ///	PRKY	206279_at
BC074851.2		
Hs.445885	KIAA1217	214912_at
(Unigene ID)		
BC008591.2 ///	KIAA0100	201729_s_at
BC050440.1 ///		
BC048096.1		

【表５－４】

GenBank ID	遺伝子名	Affymetrix ID
AF365931.1	ZNF264	205917_at
AF257099.1	PTMA	200772_x_at
BC028912.1	DNAJB9	202842_s_at

【０１６９】

【表６－１】

表６－肺がんの喫煙者と肺がんでない喫煙者を区別し得るバイオマーカーの例

GenBank ID	遺伝子名	Affymetrix ID
NM_007062.1	PWP1	201608_s_at
NM_001281.1	CKAP1	201804_x_at
BC000120.1		202355_s_at
NM_014255.1	TMEM4	202857_at
BC002642.1	CTSS	202901_x_at
NM_000346.1	SOX9	202936_s_at
NM_006545.1	NPR2L	203246_s_at
BG034328		203588_s_at
NM_021822.1	APOBEC3G	204205_at
NM_021069.1	ARGBP2	204288_s_at
NM_019067.1	FLJ10613	205010_at
NM_017925.1	FLJ20686	205684_s_at
NM_017932.1	FLJ20700	207730_x_at
NM_030757.1	MKRN4	208082_x_at
NM_030972.1	MGC5384	208137_x_at
AF126181.1	BCG1	208682_s_at
U93240.1		209653_at
U90552.1		209770_at
AF151056.1		210434_x_at
U85430.1	NFATC3	210556_at
U51007.1		211609_x_at
BC005969.1		211759_x_at
NM_002271.1		211954_s_at
AL566172		212041_at
AB014576.1	KIAA0676	212052_s_at
BF218804	AFURS1	212297_at
AK022494.1		212932_at
AA114843		213884_s_at
BE467941		214153_at
NM_003541.1	HIST1H4K	214463_x_at
R83000	BTF3	214800_x_at
AL161952.1	GLUL	215001_s_at

【表６－２】

GenBank ID	遺伝子名	Affymetrix ID
AK023843.1	PGF	215179_x_at
AK021571.1	MUC20	215208_x_at
AK023783.1	—	215604_x_at
AU147182		215620_at
AL080112.1	—	216859_x_at
AW971983		217588_at
AI683552	—	217679_x_at
NM_024006.1	IMAGE3455200	217949_s_at
AK026565.1	FLJ10534	218155_x_at
NM_014182.1	ORMDL2	218556_at
NM_021800.1	DNAJC12	218976_at
NM_016049.1	CGI-112	219203_at
NM_019023.1	PRMT7	219408_at
NM_021971.1	GMPPB	219920_s_at
NM_014128.1	—	220856_x_at
AK025651.1		221648_s_at
AA133341	C14orf87	221932_s_at
AF198444.1		222168_at

【０１７０】

【表７－１】

表７－肺がんの喫煙者と肺がんでない喫煙者を区別し得るバイオマーカーの例

GenBank ID	遺伝子名	Affymetrix ID
NM_007062.1	PWP1	201608_s_at
NM_001281.1	CKAP1	201804_x_at
BC002642.1	CTSS	202901_x_at
NM_000346.1	SOX9	202936_s_at
NM_006545.1	NPR2L	203246_s_at
BG034328		203588_s_at
NM_019067.1	FLJ10613	205010_at
NM_017925.1	FLJ20686	205684_s_at
NM_017932.1	FLJ20700	207730_x_at
NM_030757.1	MKRN4	208082_x_at
NM_030972.1	MGC5384	208137_x_at
NM_002268 /// NM_032771	KPNA4	209653_at
NM_007048 /// NM_194441	BTN3A1	209770_at
NM_006694	JBT	210434_x_at
U85430.1	NFATC3	210556_at
NM_004691	ATP6V0D1	212041_at
AB014576.1	KIAA0676	212052_s_at
BF218804	AFURS1	212297_at
BE467941		214153_at
R83000	BTF3	214800_x_at
AL161952.1	GLUL	215001_s_at
AK023843.1	PGF	215179_x_at
AK021571.1	MUC20	215208_x_at
AK023783.1	—	215604_x_at
AL080112.1	—	216859_x_at
AW971983		217588_at
AI683552	—	217679_x_at
NM_024006.1	IMAGE3455200	217949_s_at
AK026565.1	FLJ10534	218155_x_at
NM_014182.1	ORMDL2	218556_at

【表７－２】

GenBank ID	遺伝子名	Affymetrix ID
NM_021800.1	DNAJC12	218976_at
NM_016049.1	CGI-112	219203_at
NM_021971.1	GMPPB	219920_s_at
NM_014128.1	—	220856_x_at
AA133341	C14orf87	221932_s_at
AF198444.1		222168_at

【０１７１】

【表８－１】

表８－肺癌の診断又は予後を同定し得るバイオマーカーの例

Affymetrix ID	遺伝子シンボル (HUGO ID)
200729_s_at	ACTR2
200760_s_at	ARL6IP5
201399_s_at	TRAM1
201444_s_at	ATP6AP2
201635_s_at	FXR1
201689_s_at	TPD52
201925_s_at	DAF
201926_s_at	DAF
201946_s_at	CCT2
202118_s_at	CPNE3
202704_at	TOB1
202833_s_at	SERPINA1
202935_s_at	SOX9
203413_at	NELL2
203881_s_at	DMD
203908_at	SLC4A4
204006_s_at	FCGR3A /// FCGR3B
204403_x_at	KIAA0738
204427_s_at	RNP24
206056_x_at	SPN
206169_x_at	RoXaN
207730_x_at	HDGF2
207756_at	—
207791_s_at	RAB1A
207953_at	AD7C-NTP
208137_x_at	—
208246_x_at	TK2
208654_s_at	CD164
208892_s_at	DUSP6
209189_at	FOS
209204_at	LMO4
209267_s_at	SLC39A8

【表８－２】

Affymetrix ID	遺伝子シンボル (HUGO ID)
209369_at	ANXA3
209656_s_at	TMEM47
209774_x_at	CXCL2
210145_at	PLA2G4A
210168_at	C6
210317_s_at	YWHAE
210397_at	DEFB1
210679_x_at	—
211506_s_at	IL8
212006_at	UBXD2
213089_at	LOC153561
213736_at	COX5B
213813_x_at	—
214007_s_at	PTK9
214146_s_at	PPBP
214594_x_at	ATP8B1
214707_x_at	ALMS1
214715_x_at	ZNF160
215204_at	SENP6
215208_x_at	RPL35A
215385_at	FTO
215600_x_at	FBXW12
215604_x_at	UBE2D2
215609_at	STARD7
215628_x_at	PPP2CA
215800_at	DUOX1
215907_at	BACH2
215978_x_at	LOC152719
216834_at	—
216858_x_at	—
217446_x_at	—
217653_x_at	—
217679_x_at	—
217715_x_at	ZNF354A
217826_s_at	UBE2J1

【表８－３】

Affymetrix ID	遺伝子シンボル (HUGO ID)
218155_x_at	FLJ10534
218976_at	DNAJC12
219392_x_at	FLJ11029
219678_x_at	DCLRE1C
220199_s_at	FLJ12806
220389_at	FLJ23514
220720_x_at	FLJ14346
221191_at	DKFZP434A0131
221310_at	FGF14
221765_at	—
222027_at	NUCKS
222104_x_at	GTF2H3
222358_x_at	—

【０１７２】

【表９－１】

表９－肺癌の診断又は予後を同定し得るバイオマーカーの例

Affymetrix ID	(HUGO ID)
200729_s_at	ACTR2
200760_s_at	ARL6IP5
201399_s_at	TRAM1
201444_s_at	ATP6AP2
201635_s_at	FXR1
201689_s_at	TPD52
201925_s_at	DAF
201926_s_at	DAF
201946_s_at	CCT2
202118_s_at	CPNE3
202704_at	TOB1
202833_s_at	SERPINA1
202935_s_at	SOX9
203413_at	NELL2
203881_s_at	DMD
203908_at	SLC4A4
204006_s_at	FCGR3A /// FCGR3B
204403_x_at	KIAA0738
204427_s_at	RNP24
206056_x_at	SPN
206169_x_at	RoXaN
207730_x_at	HDGF2
207756_at	—
207791_s_at	RAB1A
207953_at	AD7C-NTP
208137_x_at	—
208246_x_at	TK2
208654_s_at	CD164
208892_s_at	DUSP6
209189_at	FOS
209204_at	LMO4
209267_s_at	SLC39A8

【表９－２】

Affymetrix ID	(HUGO ID)
209369_at	ANXA3
209656_s_at	TMEM47
209774_x_at	CXCL2
210145_at	PLA2G4A
210168_at	C6
210317_s_at	YWHAE
210397_at	DEFB1
210679_x_at	—
211506_s_at	IL8
212006_at	UBXD2
213089_at	LOC153561
213736_at	COX5B
213813_x_at	—
214007_s_at	PTK9
214146_s_at	PPBP
214594_x_at	ATP8B1
214707_x_at	ALMS1
214715_x_at	ZNF160
215204_at	SENP6
215208_x_at	RPL35A
215385_at	FTO
215600_x_at	FBXW12
215604_x_at	UBE2D2
215609_at	STARD7
215628_x_at	PPP2CA
215800_at	DUOX1
215907_at	BACH2
215978_x_at	LOC152719
216834_at	—
216858_x_at	—
217446_x_at	—
217653_x_at	—
217679_x_at	—
217715_x_at	ZNF354A
217826_s_at	UBE2J1

【表９－３】

Affymetrix ID	(HUGO ID)
218155_x_at	FLJ10534
218976_at	DNAJC12
219392_x_at	FLJ11029
219678_x_at	DCLRE1C
220199_s_at	FLJ12806
220389_at	FLJ23514
220720_x_at	FLJ14346
221191_at	DKFZP434A0131
221310_at	FGF14
221765_at	—
222027_at	NUCKS
222104_x_at	GTF2H3
222358_x_at	—
202113_s_at	SNX2
207133_x_at	ALPK1
218989_x_at	SLC30A5
200751_s_at	HNRPC
220796_x_at	SLC35E1
209362_at	SURB7
216248_s_at	NR4A2
203138_at	HAT1
221428_s_at	TBL1XR1
218172_s_at	DERL1
215861_at	FLJ14031
209288_s_at	CDC42EP3
214001_x_at	RPS10
209116_x_at	HBB
215595_x_at	GCNT2
208891_at	DUSP6
215067_x_at	PRDX2
202918_s_at	PREI3
211985_s_at	CALM1
212019_at	RSL1D1
216187_x_at	KNS2
215066_at	PTPRF

【表９－４】

Affymetrix ID	(HUGO ID)
212192_at	KCTD12
217586_x_at	—
203582_s_at	RAB4A
220113_x_at	POLR1B
217232_x_at	HBB
201041_s_at	DUSP1
211450_s_at	MSH6
202648_at	RPS19
202936_s_at	SOX9
204426_at	RNP24
206392_s_at	RARRES1
208750_s_at	ARF1
202089_s_at	SLC39A6
211297_s_at	CDK7
215373_x_at	FLJ12151
213679_at	FLJ13946
201694_s_at	EGR1
209142_s_at	UBE2G1
217706_at	LOC220074
212991_at	FBXO9
201289_at	CYR61
206548_at	FLJ23556
202593_s_at	MIR16
202932_at	YES1
220575_at	FLJ11800
217713_x_at	DKFZP566N034
211953_s_at	RANBP5
203827_at	WIPI49
221997_s_at	MRPL52
217662_x_at	BCAP29
218519_at	SLC35A5
214833_at	KIAA0792
201339_s_at	SCP2
203799_at	CD302
211090_s_at	PRPF4B

【表９－５】

Affymetrix ID	(HUGO ID)
220071_x_at	C15orf25
203946_s_at	ARG2
213544_at	ING1L
209908_s_at	—
201688_s_at	TPD52
215587_x_at	BTBD14B
201699_at	PSMC6
214902_x_at	FLJ42393
214041_x_at	RPL37A
203987_at	FZD6
211696_x_at	HBB
218025_s_at	PECI
215852_x_at	KIAA0889
209458_x_at	HBA1 /// HBA2
219410_at	TMEM45A
215375_x_at	—
206302_s_at	NUDT4
208783_s_at	MCP
211374_x_at	—
220352_x_at	MGC4278
216609_at	TXN
201942_s_at	CPD
202672_s_at	ATF3
204959_at	MNDA
211996_s_at	KIAA0220
222035_s_at	PAPOLA
208808_s_at	HMGB2
203711_s_at	HIBCH
215179_x_at	PGF
213562_s_at	SQLE
203765_at	GCA
214414_x_at	HBA2
217497_at	ECGF1
220924_s_at	SLC38A2
218139_s_at	C14orf108

【表９－６】

Affymetrix ID	(HUGO ID)
201096_s_at	ARF4
220361_at	FLJ12476
202169_s_at	AASDHPPT
202527_s_at	SMAD4
202166_s_at	PPP1R2
204634_at	NEK4
215504_x_at	—
202388_at	RGS2
215553_x_at	WDR45
200598_s_at	TRA1
202435_s_at	CYP1B1
216206_x_at	MAP2K7
212582_at	OSBPL8
216509_x_at	MLLT10
200908_s_at	RPLP2
215108_x_at	TNRC9
213872_at	C6orf62
214395_x_at	EEF1D
222156_x_at	CCPG1
201426_s_at	VIM
221972_s_at	Cab45
219957_at	—
215123_at	—
212515_s_at	DDX3X
203357_s_at	CAPN7
211711_s_at	PTEN
206165_s_at	CLCA2
213959_s_at	KIAA1005
215083_at	PSPC1
219630_at	PDZK1IP1
204018_x_at	HBA1 /// HBA2
208671_at	TDE2
203427_at	ASF1A
215281_x_at	POGZ
205749_at	CYP1A1

【表９－７】

Affymetrix ID	(HUGO ID)
212585_at	OSBPL8
211745_x_at	HBA1 /// HBA2
208078_s_at	SNF1LK
218041_x_at	SLC38A2
212588_at	PTPRC
212397_at	RDX
208268_at	ADAM28
207194_s_at	ICAM4
222252_x_at	—
217414_x_at	HBA2
207078_at	MED6
215268_at	KIAA0754
221387_at	GPR147
201337_s_at	VAMP3
220218_at	C9orf68
222356_at	TBL1Y
208579_x_at	H2BFS
219161_s_at	CKLF
202917_s_at	S100A8
204455_at	DST
211672_s_at	ARPC4
201132_at	HNRPH2
218313_s_at	GALNT7
218930_s_at	FLJ11273
219166_at	C14orf104
212805_at	KIAA0367
201551_s_at	LAMP1
202599_s_at	NRIP1
203403_s_at	RNF6
214261_s_at	ADH6
202033_s_at	RB1CC1
203896_s_at	PLCB4
209703_x_at	DKFZP586A0522
211699_x_at	HBA1 /// HBA2
210764_s_at	CYR61

【表９－８】

Affymetrix ID	(HUGO ID)
206391_at	RARRES1
201312_s_at	SH3BGRL
200798_x_at	MCL1
214912_at	—
204621_s_at	NR4A2
217761_at	MTCBP-1
205830_at	CLGN
218438_s_at	MED28
207475_at	FABP2
208621_s_at	VIL2
202436_s_at	CYP1B1
202539_s_at	HMGCR
210830_s_at	PON2
211906_s_at	SERPINB4
202241_at	TRIB1
203594_at	RTCD1
215863_at	TFR2
221992_at	LOC283970
221872_at	RARRES1
219564_at	KCNJ16
201329_s_at	ETS2
214188_at	HIS1
201667_at	GJA1
201464_x_at	JUN
215409_at	LOC254531
202583_s_at	RANBP9
215594_at	—
214326_x_at	JUND
217140_s_at	VDAC1
215599_at	SMA4
209896_s_at	PTPN11
204846_at	CP
222303_at	—
218218_at	DIP13B
211015_s_at	HSPA4

【表９－９】

Affymetrix ID	(HUGO ID)
208666_s_at	ST13
203191_at	ABCB6
202731_at	PDCD4
209027_s_at	ABI1
205979_at	SCGB2A1
216351_x_at	DAZ1 /// DAZ3 ///
	DAZ2 /// DAZ4
220240_s_at	C13orf11
204482_at	CLDN5
217234_s_at	VIL2
214350_at	SNTB2
201693_s_at	EGR1
212328_at	KIAA1102
220168_at	CASC1
203628_at	IGF1R
204622_x_at	NR4A2
213246_at	C14orf109
218728_s_at	HSPC163
214753_at	PFAAP5
206336_at	CXCL6
201445_at	CNN3
209886_s_at	SMAD6
213376_at	ZBTB1
213887_s_at	POLR2E
204783_at	MLF1
218824_at	FLJ10781
212417_at	SCAMP1
202437_s_at	CYP1B1
217528_at	CLCA2
218170_at	ISOC1
206278_at	PTAFR
201939_at	PLK2
200907_s_at	KIAA0992
207480_s_at	MEIS2
201417_at	SOX4

【表９－１０】

Affymetrix ID	(HUGO ID)
213826_s_at	—
214953_s_at	APP
204897_at	PTGER4
201711_x_at	RANBP2
202457_s_at	PPP3CA
206683_at	ZNF165
214581_x_at	TNFRSF21
203392_s_at	CTBP1
212720_at	PAPOLA
207758_at	PPM1F
220995_at	STXBP6
213831_at	HLA-DQA1
212044_s_at	—
202434_s_at	CYP1B1
206166_s_at	CLCA2
218343_s_at	GTF3C3
202557_at	STCH
201133_s_at	PJA2
213605_s_at	MGC22265
210947_s_at	MSH3
208310_s_at	C7orf28A /// C7orf28B
209307_at	—
215387_x_at	GPC6
213705_at	MAT2A
213979_s_at	—
212731_at	LOC157567
210117_at	SPAG1
200641_s_at	YWHAZ
210701_at	CFDP1
217152_at	NCOR1
204224_s_at	GCH1
202028_s_at	—
201735_s_at	CLCN3
208447_s_at	PRPS1
220926_s_at	C1orf22

【表９－１１】

Affymetrix ID	(HUGO ID)
211505_s_at	STAU
221684_s_at	NYX
206906_at	ICAM5
213228_at	PDE8B
217202_s_at	GLUL
211713_x_at	KIAA0101
215012_at	ZNF451
200806_s_at	HSPD1
201466_s_at	JUN
211564_s_at	PDLIM4
207850_at	CXCL3
221841_s_at	KLF4
200605_s_at	PRKAR1A
221198_at	SCT
201772_at	AZIN1
205009_at	TFF1
205542_at	STEAP1
218195_at	C6orf211
213642_at	—
212891_s_at	GADD45GIP1
202798_at	SEC24B
222207_x_at	—
202638_s_at	ICAM1
200730_s_at	PTP4A1
219355_at	FLJ10178
220266_s_at	KLF4
201259_s_at	SYPL
209649_at	STAM2
220094_s_at	C6orf79
221751_at	PANK3
200008_s_at	GDI2
205078_at	PIGF
218842_at	FLJ21908
202536_at	CHMP2B
220184_at	NANOG

【表９－１２】

Affymetrix ID	(HUGO ID)
201117_s_at	CPE
219787_s_at	ECT2
206628_at	SLC5A1
204007_at	FCGR3B
209446_s_at	—
211612_s_at	IL13RA1
220992_s_at	C1orf25
221899_at	PFAAP5
221719_s_at	LZTS1
201473_at	JUNB
221193_s_at	ZCCHC10
215659_at	GSDML
205157_s_at	KRT17
201001_s_at	UBE2V1 /// Kua-UEV
216789_at	—
205506_at	VIL1
204875_s_at	GMDS
207191_s_at	ISLR
202779_s_at	UBE2S
210370_s_at	LY9
202842_s_at	DNAJB9
201082_s_at	DCTN1
215588_x_at	RIOK3
211076_x_at	DRPLA
210230_at	—
206544_x_at	SMARCA2
208852_s_at	CANX
215405_at	MYO1E
208653_s_at	CD164
206355_at	GNAL
210793_s_at	NUP98
215070_x_at	RABGAP1
203007_x_at	LYPLA1
203841_x_at	MAPRE3
206759_at	FCER2

【表９－１３】

Affymetrix ID	(HUGO ID)
202232_s_at	GA17
215892_at	—
214359_s_at	HSPCB
215810_x_at	DST
208937_s_at	ID1
213664_at	SLC1A1
219338_s_at	FLJ20156
206595_at	CST6
207300_s_at	F7
213792_s_at	INSR
209674_at	CRY1
40665_at	FMO3
217975_at	WBP5
210296_s_at	PXMP3
215483_at	AKAP9
212633_at	KIAA0776
206164_at	CLCA2
216813_at	—
208925_at	C3orf4
219469_at	DNCH2
206016_at	CXorf37
216745_x_at	LRCH1
212999_x_at	HLA-DQB1
216859_x_at	—
201636_at	—
204272_at	LGALS4
215454_x_at	SFTPC
215972_at	—
220593_s_at	FLJ20753
222009_at	CGI-14
207115_x_at	MBTD1
216922_x_at	DAZ1 /// DAZ3 ///
	DAZ2 /// DAZ4
217626_at	AKR1C1 /// AKR1C2
211429_s_at	SERPINA1

【表９－１４】

Affymetrix ID	(HUGO ID)
209662_at	CETN3
201629_s_at	ACP1
201236_s_at	BTG2
217137_x_at	—
212476_at	CENTB2
218545_at	FLJ11088
208857_s_at	PCMT1
221931_s_at	SEH1L
215046_at	FLJ23861
220222_at	PRO1905
209737_at	AIP1
203949_at	MPO
219290_x_at	DAPP1
205116_at	LAMA2
222316_at	VDP
203574_at	NFIL3
207820_at	ADH1A
203751_x_at	JUND
202930_s_at	SUCLA2
215404_x_at	FGFR1
216266_s_at	ARFGEF1
212806_at	KIAA0367
219253_at	—
214605_x_at	GPR1
205403_at	IL1R2
222282_at	PAPD4
214129_at	PDE4DIP
209259_s_at	CSPG6
216900_s_at	CHRNA4
221943_x_at	RPL38
215386_at	AUTS2
201990_s_at	CREBL2
220145_at	FLJ21159
221173_at	USH1C
214900_at	ZKSCAN1

【表９－１５】

Affymetrix ID	(HUGO ID)
203290_at	HLA-DQA1
215382_x_at	TPSAB1
201631_s_at	IER3
212188_at	KCTD12
220428_at	CD207
215349_at	—
213928_s_at	HRB
221228_s_at	—
202069_s_at	IDH3A
208554_at	POU4F3
209504_s_at	PLEKHB1
212989_at	TMEM23
216197_at	ATF7IP
204748_at	PTGS2
205221_at	HGD
214705_at	INADL
213939_s_at	RIPX
203691_at	PI3
220532_s_at	LR8
209829_at	C6orf32
206515_at	CYP4F3
218541_s_at	C8orf4
210732_s_at	LGALS8
202643_s_at	TNFAIP3
218963_s_at	KRT23
213304_at	KIAA0423
202768_at	FOSB
205623_at	ALDH3A1
206488_s_at	CD36
204319_s_at	RGS10
217811_at	SELT
202746_at	ITM2A
221127_s_at	RIG
209821_at	C9orf26
220957_at	CTAGE1

【表９－１６】

Affymetrix ID	(HUGO ID)
215577_at	UBE2E1
214731_at	DKFZp547A023
210512_s_at	VEGF
205267_at	POU2AF1
216202_s_at	SPTLC2
220477_s_at	C20orf30
205863_at	S100A12
215780_s_at	SET /// LOC389168
218197_s_at	OXR1
203077_s_at	SMAD2
222339_x_at	—
200698_at	KDELR2
210540_s_at	B4GALT4
217725_x_at	PAI-RBP1
217082_at	—

【０１７３】

【表１０】

表１０－肺癌の診断又は予後を同定し得るバイオマーカーの例

Affymetrix ID	HUGO ID
207953_at	AD7C-NTP
215208_x_at	RPL35A
215604_x_at	UBE2D2
218155_x_at	FLJ10534
216858_x_at	—
208137_x_at	—
214715_x_at	ZNF160
217715_x_at	ZNF354A
220720_x_at	FLJ14346
215907_at	BACH2
217679_x_at	—
206169_x_at	RoXaN
208246_x_at	TK2
222104_x_at	GTF2H3
206056_x_at	SPN
217653_x_at	—
210679_x_at	—
207730_x_at	HDGF2
214594_x_at	ATP8B1

【０１７４】

【表１１－１】

表 11 – 喫煙に対する応答としての、気管支と上気道間の上皮細胞の発現プロファイルの関係を同定し得るバイオマーカーの例

AffyID	遺伝子名 (HUGO ID)
202437_s_at	CYP1B1
206561_s_at	AKR1B10
202436_s_at	CYP1B1
205749_at	CYP1A1
202435_s_at	CYP1B1
201884_at	CEACAM5
205623_at	ALDH3A1
217626_at	—
209921_at	SLC7A11
209699_x_at	AKR1C2
201467_s_at	NQO1
201468_s_at	NQO1
202831_at	GPX2
214303_x_at	MUC5AC
211653_x_at	AKR1C2
214385_s_at	MUC5AC
216594_x_at	AKR1C1
205328_at	CLDN10
209160_at	AKR1C3
210519_s_at	NQO1
217678_at	SLC7A11
205221_at	HGD /// LOC642252
204151_x_at	AKR1C1
207469_s_at	PIR
206153_at	CYP4F11
205513_at	TCN1
209386_at	TM4SF1
209351_at	KRT14
204059_s_at	ME1
209213_at	CBR1

【表１１－２】

AffyID	遺伝子名 (HUGO ID)
210505_at	ADH7
214404_x_at	SPDEF
204058_at	ME1
218002_s_at	CXCL14
205499_at	SRPX2
210065_s_at	UPK1B
204341_at	TRIM16 /// TRIM16L ///
	LOC653524
221841_s_at	KLF4
208864_s_at	TXN
208699_x_at	TKT
210397_at	DEFB1
204971_at	CSTA
211657_at	CEACAM6
201463_s_at	TALDO1
214164_x_at	CA12
203925_at	GCLM
201118_at	PGD
201266_at	TXNRD1
203757_s_at	CEACAM6
202923_s_at	GCLC
214858_at	GPC1
205009_at	TFF1
219928_s_at	CABYR
203963_at	CA12
210064_s_at	UPK1B
219956_at	GALNT6
208700_s_at	TKT
203824_at	TSPAN8
207126_x_at	UGT1A10 /// UGT1A8 /// UGT1A7 ///
	UGT1A6 /// UGT1A
213441_x_at	SPDEF
207430_s_at	MSMB
209369_at	ANXA3

【表１１－３】

AffyID	遺伝子名 (HUGO ID)
217187_at	MUC5AC
209101_at	CTGF
212221_x_at	IDS
215867_x_at	CA12
214211_at	FTH1
217755_at	HN1
201431_s_at	DPYSL3
204875_s_at	GMDS
215125_s_at	UGT1A10 /// UGT1A8 /// UGT1A7 ///
	UGT1A6 /// UGT1A
63825_at	ABHD2
202922_at	GCLC
218313_s_at	GALNT7
210297_s_at	MSMB
209448_at	HTATIP2
204532_x_at	UGT1A10 /// UGT1A8 /// UGT1A7 ///
	UGT1A6 /// UGT1A
200872_at	S100A10
216351_x_at	DAZ1 /// DAZ3 /// DAZ2 /// DAZ4
212223_at	IDS
208680_at	PRDX1
206515_at	CYP4F3
208596_s_at	UGT1A10 /// UGT1A8 /// UGT1A7 ///
	UGT1A6 /// UGT1A
209173_at	AGR2
204351_at	S100P
202785_at	NDUFA7
204970_s_at	MAFG
222016_s_at	ZNF323
200615_s_at	AP2B1
206094_x_at	UGT1A6

【表１１－４】

AffyID	遺伝子名 (HUGO ID)
209706_at	NKX3-1
217977_at	SEPX1
201487_at	CTSC
219508_at	GCNT3
204237_at	GULP1
213455_at	LOC283677
213624_at	SMPDL3A
206770_s_at	SLC35A3
217975_at	WBP5
201263_at	TARS
218696_at	EIF2AK3
212560_at	C11orf32
218885_s_at	GALNT12
212326_at	VPS13D
217955_at	BCL2L13
203126_at	IMPA2
214106_s_at	GMDS
209309_at	AZGP1
205112_at	PLCE1
215363_x_at	FOLH1
206302_s_at	NUDT4 /// NUDT4P1
200916_at	TAGLN2
205042_at	GNE
217979_at	TSPAN13
203397_s_at	GALNT3
209786_at	HMGN4
211733_x_at	SCP2
207222_at	PLA2G10
204235_s_at	GULP1
205726_at	DIAPH2
203911_at	RAP1GAP
200748_s_at	FTH1
212449_s_at	LYPLA1
213059_at	CREB3L1
201272_at	AKR1B1

【表１１－５】

AffyID	遺伝子名 (HUGO ID)
208731_at	RAB2
205979_at	SCGB2A1
212805_at	KIAA0367
202804_at	ABCC1
218095_s_at	TPARL
205566_at	ABHD2
209114_at	TSPAN1
202481_at	DHRS3
202805_s_at	ABCC1
219117_s_at	FKBP11
213172_at	TTC9
202554_s_at	GSTM3
218677_at	S100A14
203306_s_at	SLC35A1
204076_at	ENTPD4
200654_at	P4HB
204500_s_at	AGTPBP1
208918_s_at	NADK
221485_at	B4GALT5
221511_x_at	CCPG1
200733_s_at	PTP4A1
217901_at	DSG2
202769_at	CCNG2
202119_s_at	CPNE3
200945_s_at	SEC31L1
200924_s_at	SLC3A2
208736_at	ARPC3
221556_at	CDC14B
221041_s_at	SLC17A5
215071_s_at	HIST1H2AC
209682_at	CBLB
209806_at	HIST1H2BK
204485_s_at	TOM1L1
201666_at	TIMP1
203192_at	ABCB6

【表１１－６】

AffyID	遺伝子名 (HUGO ID)
202722_s_at	GFPT1
213135_at	TIAM1
203509_at	SORL1
214620_x_at	PAM
208919_s_at	NADK
212724_at	RND3
212160_at	XPOT
212812_at	SERINC5
200696_s_at	GSN
217845_x_at	HIGD1A
208612_at	PDIA3
219288_at	C3orf14
201923_at	PRDX4
211960_s_at	RAB7
64942_at	GPR153
201659_s_at	ARL1
202439_s_at	IDS
209249_s_at	GHITM
218723_s_at	RGC32
200087_s_at	TMED2
209694_at	PTS
202320_at	GTF3C1
201193_at	IDH1
212233_at	—
213891_s_at	—
203041_s_at	LAMP2
202666_s_at	ACTL6A
200863_s_at	RAB11A
203663_s_at	COX5A
211404_s_at	APLP2
201745_at	PTK9
217823_s_at	UBE2J1
202286_s_at	TACSTD2
212296_at	PSMD14
211048_s_at	PDIA4

【表１１－７】

AffyID	遺伝子名 (HUGO ID)
214429_at	MTMR6
219429_at	FA2H
212181_s_at	NUDT4
222116_s_at	TBC1D16
221689_s_at	PIGP
209479_at	CCDC28A
218434_s_at	AACS
214665_s_at	CHP
202085_at	TJP2
217992_s_at	EFHD2
203162_s_at	KATNB1
205406_s_at	SPA17
203476_at	TPBG
201724_s_at	GALNT1
200599_s_at	HSP90B1
200929_at	TMED10
200642_at	SOD1
208946_s_at	BECN1
202562_s_at	C14orf1
201098_at	COPB2
221253_s_at	TXNDC5
201004_at	SSR4
203221_at	TLE1
201588_at	TXNL1
218684_at	LRRC8D
208799_at	PSMB5
201471_s_at	SQSTM1
204034_at	ETHE1
208689_s_at	RPN2
212665_at	TIPARP
200625_s_at	CAP1
213220_at	LOC92482
200709_at	FKBP1A
203279_at	EDEM1
200068_s_at	CANX

【表１１－８】

AffyID	遺伝子名 (HUGO ID)
200620_at	TMEM59
200075_s_at	GUK1
209679_s_at	LOC57228
210715_s_at	SPINT2
209020_at	C20orf111
208091_s_at	ECOP
200048_s_at	JTB
218194_at	REXO2
209103_s_at	UFD1L
208718_at	DDX17
219241_x_at	SSH3
216210_x_at	TRIOBP
50277_at	GGA1
218023_s_at	FAM53C
32540_at	PPP3CC
43511_s_at	—
212001_at	SFRS14
208637_x_at	ACTN1
201997_s_at	SPEN
205073_at	CYP2J2
40837_at	TLE2
204447_at	ProSAPiP1
204604_at	PFTK1
210273_at	PCDH7
208614_s_at	FLNB
206510_at	SIX2
200675_at	CD81
219228_at	ZNF331
209426_s_at	AMACR
204000_at	GNB5
221742_at	CUGBP1
208883_at	EDD1
210166_at	TLR5
211026_s_at	MGLL
220446_s_at	CHST4

【表１１－９】

AffyID	遺伝子名 (HUGO ID)
207636_at	SERPINI2
212226_s_at	PPAP2B
210347_s_at	BCL11A
218424_s_at	STEAP3
204287_at	SYNGR1
205489_at	CRYM
36129_at	RUTBC1
215418_at	PARVA
213029_at	NFIB
221016_s_at	TCF7L1
209737_at	MAGI2
220389_at	CCDC81
213622_at	COL9A2
204740_at	CNKSR1
212126_at	—
207760_s_at	NCOR2
205258_at	INHBB
213169_at	—
33760_at	PEX14
220968_s_at	TSPAN9
221792_at	RAB6B
205752_s_at	GSTM5
218974_at	FLJ10159
221748_s_at	TNS1
212185_x_at	MT2A
209500_x_at	TNFSF13 /// TNFSF12-TNFSF13
215445_x_at	1-Mar
220625_s_at	ELF5
32137_at	JAG2
219747_at	FLJ23191
201397_at	PHGDH
207913_at	CYP2F1
217853_at	TNS3
1598_g_at	GAS6

【表１１－１０】

AffyID	遺伝子名 (HUGO ID)
203799_at	CD302
203329_at	PTPRM
208712_at	CCND1
210314_x_at	TNFSF13 /// TNFSF12-TNFSF13
213217_at	ADCY2
200953_s_at	CCND2
204326_x_at	MT1X
213488_at	SNED1
213505_s_at	SFRS14
200982_s_at	ANXA6
211732_x_at	HNMT
202587_s_at	AK1
396_f_at	EPOR
200878_at	EPAS1
213228_at	PDE8B
215785_s_at	CYFIP2
213601_at	SLIT1
37953_s_at	ACCN2
205206_at	KAL1
212859_x_at	MT1E
217165_x_at	MT1F
204754_at	HLF
218225_at	SITPEC
209784_s_at	JAG2
211538_s_at	HSPA2
211456_x_at	LOC650610
204734_at	KRT15
201563_at	SORD
202746_at	ITM2A
218025_s_at	PECI
203914_x_at	HPGD
200884_at	CKB
204753_s_at	HLF

【表１１－１１】

AffyID	遺伝子名 (HUGO ID)
207718_x_at	CYP2A6 /// CYP2A7 /// CYP2A7P1 /// CYP2A13
218820_at	C14orf132
204745_x_at	MT1G
204379_s_at	FGFR3
207808_s_at	PROS1
207547_s_at	FAM107A
208581_x_at	MT1X
205384_at	FXYD1
213629_x_at	MT1F
823_at	CX3CL1
203687_at	CX3CL1
211295_x_at	CYP2A6
204755_x_at	HLF
209897_s_at	SLIT2
40093_at	BCAM
211726_s_at	FMO2
206461_x_at	MT1H
219250_s_at	FLRT3
210524_x_at	—
220798_x_at	PRG2
219410_at	TMEM45A
205680_at	MMP10
217767_at	C3 /// LOC653879
220562_at	CYP2W1
210445_at	FABP6
205725_at	SCGB1A1
213432_at	MUC5B /// LOC649768
209074_s_at	FAM107A
216346_at	SEC14L3

【０１７５】

【表１２－１】

表 12 – 喫煙に対する応答において鼻上皮において大きく変化する遺伝子のうち、気管支上皮細胞において差次的に発現し得るバイオマーカーの例

AffxID	Hugo ID
203369_x_at	—
218434_s_at	AACS
205566_at	ABHD2
217687_at	ADCY2
210505_at	ADH7
205623_at	ALDH3A1
200615_s_at	AP2B1
214875_x_at	APLP2
212724_at	ARHE
201659_s_at	ARL1
208736_at	ARPC3
213624_at	ASM3A
209309_at	AZGP1
217188_s_at	C14orf1
200620_at	C1orf8
200068_s_at	CANX
213798_s_at	CAP1
200951_s_at	CCND2
202769_at	CCNG2
201884_at	CEACAM5
203757_s_at	CEACAM6
214665_s_at	CHP
205328_at	CLDN10
203663_s_at	COX5A
202119_s_at	CPNE3
221156_x_at	CPR8
201487_at	CTSC
205749_at	CYP1A1
207913_at	CYP2F1
206153_at	CYP4F11
206514_s_at	CYP4F3

【表１２－２】

AffxID	Hugo ID
216351_x_at	DAZ4
203799_at	DCL-1
212665_at	DKFZP434J214
201430_s_at	DPYSL3
211048_s_at	ERP70
219118_at	FKBP11
214119_s_at	FKBP1A
208918_s_at	FLJ13052
217487_x_at	FOLH1
200748_s_at	FTH1
201723_s_at	GALNT1
218885_s_at	GALNT12
203397_s_at	GALNT3
218313_s_at	GALNT7
203925_at	GCLM
219508_at	GCNT3
202722_s_at	GFPT1
204875_s_at	GMDS
205042_at	GNE
208612_at	GRP58
214040_s_at	GSN
214307_at	HGD
209806_at	HIST1H2BK
202579_x_at	HMGN4
207180_s_at	HTATIP2
206342_x_at	IDS
203126_at	IMPA2
210927_x_at	JTB
203163_at	KATNB1
204017_at	KDELR3
213174_at	KIAA0227
212806_at	KIAA0367
210616_s_at	KIAA0905
221841_s_at	KLF4
203041_s_at	LAMP2

【表１２－３】

AffxID	Hugo ID
213455_at	LOC92689
218684_at	LRRC5
204059_s_at	ME1
207430_s_at	MSMB
210472_at	MT1G
213432_at	MUC5B
211498_s_at	NKX3-1
201467_s_at	NQO1
206303_s_at	NUDT4
213498_at	OASIS
200656_s_at	P4HB
213441_x_at	PDEF
207469_s_at	PIR
207222_at	PLA2G10
209697_at	PPP3CC
201923_at	PRDX4
200863_s_at	RAB11A
208734_x_at	RAB2
203911_at	RAP1GA1
218723_s_at	RGC32
200087_s_at	RNP24
200872_at	S100A10
205979_at	SCGB2A1
202481_at	SDR1
217977_at	SEPX1
221041_s_at	SLC17A5
203306_s_at	SLC35A1
207528_s_at	SLC7A11
202287_s_at	TACSTD2
210978_s_at	TAGLN2
205513_at	TCN1
201666_at	TIMP1
208699_x_at	TKT
217979_at	TM4SF13
203824_at	TM4SF3

【表１２－４】

AffxID	Hugo ID
200929_at	TMP21
221253_s_at	TXNDC5
217825_s_at	UBE2J1
215125_s_at	UGT1A10
210064_s_at	UPK1B
202437_s_at	CYP1B1

【０１７６】

【表１３－１】

表 13 – バイオマーカーの例

AFFYID	遺伝子名 (HUGO ID)
213693_s_at	MUC1
211695_x_at	MUC1
207847_s_at	MUC1
208405_s_at	CD164
220196_at	MUC16
217109_at	MUC4
217110_s_at	MUC4
204895_x_at	MUC4
214385_s_at	MUC5AC
1494_f_at	CYP2A6
210272_at	CYP2B7P1
206754_s_at	CYP2B7P1
210096_at	CYP4B1
208928_at	POR
207913_at	CYP2F1
220636_at	DNAI2
201999_s_at	DYNLT1
205186_at	DNALI1
220125_at	DNAI1
210345_s_at	DNAH9
214222_at	DNAH7
211684_s_at	DYNC1I2
211928_at	DYNC1H1
200703_at	DYNLL1
217918_at	DYNLRB1
217917_s_at	DYNLRB1
209009_at	ESD
204418_x_at	GSTM2
215333_x_at	GSTM1
217751_at	GSTK1
203924_at	GSTA1
201106_at	GPX4

【表１３－２】

AFFYID	遺伝子名 (HUGO ID)
200736_s_at	GPX1
204168_at	MGST2
200824_at	GSTP1
211630_s_at	GSS
201470_at	GSTO1
201650_at	KRT19
209016_s_at	KRT7
209008_x_at	KRT8
201596_x_at	KRT18
210633_x_at	KRT10
207023_x_at	KRT10
212236_x_at	KRT17
201820_at	KRT5
204734_at	KRT15
203151_at	MAP1A
200713_s_at	MAPRE1
204398_s_at	EML2
40016_g_at	MAST4
208634_s_at	MACF1
205623_at	ALDH3A1
212224_at	ALDH1A1
205640_at	ALDH3B1
211004_s_at	ALDH3B1
202054_s_at	ALDH3A2
205208_at	ALDH1L1
201612_at	ALDH9A1
201425_at	ALDH2
201090_x_at	K-ALPHA-1
202154_x_at	TUBB3
202477_s_at	TUBGCP2
203667_at	TBCA
204141_at	TUBB2A
207490_at	TUBA4
208977_x_at	TUBB2C
209118_s_at	TUBA3

【表１３－３】

AFFYID	遺伝子名 (HUGO ID)
209251_x_at	TUBA6
211058_x_at	K-ALPHA-1
211072_x_at	K-ALPHA-1
211714_x_at	TUBB
211750_x_at	TUBA6
212242_at	TUBA1
212320_at	TUBB
212639_x_at	K-ALPHA-1
213266_at	76P
213476_x_at	TUBB3
213646_x_at	K-ALPHA-1
213726_x_at	TUBB2C

【０１７７】

【表１４】

表１４－試料の分布

代表的な組織病理型	訓練セット		テストセット
	試料数	患者数	患者数
一般の間質性肺炎(UIP)	136	34	11
診断し難いUIP	40	11	7
診断しやすいUIP	22	5	4
UIP(下葉)＋非特異性間質性肺炎(NSIP)(上葉)	5	1	
診断し難いUIP(下葉)＋NSIP(上葉)	4	1	
UIP(下葉)＋肺高血圧症(上葉)	5	1	
診断しやすいHP(下葉)＋診断し難いUIP(上葉)			1
UIP合計	212 (60%)	53 (59%)	23 (47%)
呼吸細気管支炎(RB)、喫煙による関連間質性線維症	26	7	7
過敏性肺炎、診断しやすいHP	19	4	4
サルコイドーシス	17	5	4
NSIP、セルラーNSIP、診断しやすいNSIP	18	5	3
びまん性肺胞損傷、ヘモジデローシスを伴うDAD	2	1	2
アミロイド沈着又は軽鎖沈着			1
細気管支炎	12	3	1
好酸球性肺炎(EP)	5	1	1
外因性リポイド肺炎			1
組織化肺胞出血			1
組織化肺炎(OP)	29	7	
ニューモシスチス肺炎(PP)	4	1	
肺気腫	10	3	
非UIP合計	142 (40%)	37 (41%)	26 (53%)
合計	354	90	49

【０１７８】

[225]

【表１５】

表１５

	上方制御される遺伝子	下方制御される遺伝子	差次的に発現した遺伝子の合計	#非 UIP 試料と重複する遺伝子
非 UIP 全部 (N = 147)	55	96	151	151 (100%)
細気管支炎(N = 10)	41	34	75	6 (8%)
HP (N = 13)	32	53	85	14 (16%)
NSIP (N =12)	37	49	86	13 (15%)
OP (N = 23)	1	15	16	31 (52%)
RB (N = 16)	549	152	701	64 (9%)
サルコイドーシス (N = 11)	448	726	1174	93 (8%)

【０１７９】

[226] 表１５は、非ＵＩＰ亜型とＵＩＰ試料の間で有意に発現した遺伝子の数（ｐ調整＜０．０５、倍率変化＞２）（ｎ＝２１２）を示す。差次的に発現する遺伝子のうち、ＵＩＰ試料と非ＵＩＰ試料間のそれと重複するものの数を３番目の列にまとめた。

【０１８０】

[227]

【表１６】

表１６

分類器	実行間	実行中（残余）	実行間（合計）
Ensemble	0.28 (4.0%)	0.37 (5.3%)	0.46 (6.5%)
罰則付きロジスティック回帰	0.10 (2.6%)	0.19 (4.9%)	0.22 (5.6%)

【０１８１】

[228] 表１６は、線形混合効果モデルを用いる２つの分類器から得たスコアの変動性の推定を示す。百分率（％）は、分類スコアの５％～９５％の変位値の範囲に対する推定変動性の比率である。

【０１８２】

[229] 本明細書に記載の分類器は、侵襲的処置を使用せず、ＩＰＦ又は肺癌などの疾病を診断してもよい。教師なしのクラスタリング分析には、（ｉ）小さい試料サイズ、（ｉｉ）疾患の異質性（例えば、非ＵＩＰパターン亜型疾患）、（ｉｉｉ）異なる試料のプール効果及びバッチ効果、又は（ｉｖ）これらの任意の組み合わせにより、（ａ）悪性組織と良性組織の区別、（ｂ）ＵＩＰパターンと非ＵＩＰパターンの区別、（ｃ）特定の発現パターンを有する試料とそうでない試料の区別、又は（ｄ）これらの任意の組み合わせを行うことができないという短所がある。訓練済みの機械学習アルゴリズムであれば、上記の短所を補える。本明細書に記載の方法は、侵襲的処置の必要性を排除し、小さい試料サイズ、疾患の不均一性、又は異なる試料のプール効果及びバッチ効果という限界にも拘らず、高い臨床精度を備えた非侵襲的予後ツール、診断ツール、又はこれらの任意の組み合わせを提供することができる。一部において、RNA-seqデータは、機械学習アルゴリズムに入力される。同被検体から採取された試料でも異質性が生じることがある。例えば、組織病理的特徴は、組織（肺組織など）全体で均一性を有しないことがあり、遺伝子発現プロファイルは、試料の採取部位により異なり得る。疾患により異質性が生じることもある。例えば、非ＵＩＰパターンの存在は、異なる種類の疾患など、１つ以上の亜型疾患を示す。

【０１８３】

[230] 一部において、被検体から１つ、２つ、３つ、４つ、５つ、６つ、７、８つ、９つ、１０個、又はそれ以上の試料を採取して個別的に分析してもよい。一部において、被検体から２つの試料を採取して個別的に分析してもよい。一部において、被検体から３つの試料を採取して個別的に分析してもよい。一部において、被検体から４つの試料を採取して個別的に分析してもよい。一部において、被検体から５つの試料を採取して個別的に

分析してもよい。一部において、被検体から６つの試料を採取して個別的に分析してもよい。一部において、被検体から７つの試料を採取して個別的に分析してもよい。部において、被検体から８つの試料を採取して個別的に分析してもよい。一部において、被検体から９つの試料を採取して個別的に分析してもよい。一部において、被検体から１０個の試料を採取して個別的に分析してもよい。一部において、被検体から１～１０個の試料を採取して個別的に分析してもよい。一部において、被検体から１～５の試料を採取して個別的に分析してもよい。一部において、被検体から１～２０個の試料を採取して個別的に分析してもよい。

【０１８４】

[231] 分類器、例えばロック分類器は、検証セット（分類器の検証に用いられるもの）と比べ、互いに独立したテストセットにおいて、実質的に同じ精度、ＮＰＶ、ＰＰＶ、感度、特異度、又はこれらの任意の組み合わせを出してもよい。分類器は、少なくとも約５つの独立テスト用試料において、実質的に同じ精度、ＮＰＶ、ＰＰＶ、感度、特異度、又はこれらの任意の組み合わせを維持してもよい。分類器は、少なくとも約１０個の独立テスト用試料において、実質的に同じ精度、ＮＰＶ、ＰＰＶ、感度、特異度、又はこれらの任意の組み合わせを維持してもよい。分類器は、少なくとも約５０個の独立テスト用試料において、実質的に同じ精度、ＮＰＶ、ＰＰＶ、感度、特異度、又はこれらの任意の組み合わせを維持してもよい。分類器は、少なくとも約１００個の独立テスト用試料において、実質的に同じ精度、ＮＰＶ、ＰＰＶ、感度、特異度、又はこれらの任意の組み合わせを維持してもよい。分類器は、少なくとも約５００個の独立テスト用試料において、実質的に同じ精度、ＮＰＶ、ＰＰＶ、感度、特異度、又はこれらの任意の組み合わせを維持してもよい。分類器は、少なくとも約１０００個の独立テスト用試料において、実質的に同じ精度、ＮＰＶ、ＰＰＶ、感度、特異度、又はこれらの任意の組み合わせを維持してもよい。分類器は、少なくとも約１つ～約１０個の独立テスト用試料において、実質的に同じ精度、ＮＰＶ、ＰＰＶ、感度、特異度、又はこれらの任意の組み合わせを維持してもよい。分類器は、少なくとも約１つ～約１００個の独立テスト用試料において、実質的に同じ精度、ＮＰＶ、ＰＰＶ、感度、特異度、又はこれらの任意の組み合わせを維持してもよい。分類器は、少なくとも約１つ～約５００個の独立テスト用試料において、実質的に同じ精度、ＮＰＶ、ＰＰＶ、感度、特異度、又はこれらの任意の組み合わせを維持してもよい。分類器は、少なくとも約１つ～約１０００個の独立テスト用試料において、実質的に同じ精度、ＮＰＶ、ＰＰＶ、感度、特異度、又はこれらの任意の組み合わせを維持してもよい。分類器は、少なくとも約１つ～約５０００個の独立テスト用試料において、実質的に同じ精度、ＮＰＶ、ＰＰＶ、感度、特異度、又はこれらの任意の組み合わせを維持してもよい。

【０１８５】

[232] 複数の独立テスト用試料において実質的に同じ精度、ＮＰＶ、ＰＰＶ、感度、特異度、又はこれらの任意の組み合わせを維持するために、バッチ効果を取り除くことができる。除去すると試料間で大きい変動をもたらすバイオマーカーを、分類器の選択特性やそれ以降の分析から取り除いてもよい。バッチ効果に非常に敏感なバイオマーカーを、以降の分析や選択特性から取り除いてもよい。それにより、複数の独立試料をかけても分類器の性能（精度、NPV、PPV、感度、特異度など）は、実質的に変わらないことになる。

【０１８６】

[233] 本方法は、被検体から採取された複数の試料のうち、異質性を持つ被検体を特定することを含み得る。例えば、この方法は、一方の試料には非ＵＩＰパターンが割り当てられ、他方にはＵＩＰパターンが割り当てられる被検体を特定することを含み得る。組織病理診断、遺伝子発現、又はこれらの組み合わせにより、同被検体から採取された試料から異質性を見付けることができる。例えば、ＵＩＰパターンと非ＵＩＰパターンの疾患は、異質である。非ＵＩＰパターンの疾患の区別又は診断に用いられるバイオマーカーは、それとは異なる非ＵＩＰパターンの疾患の区別又は診断には適用できなくてもよい。この場合、それぞれの疾患、亜型疾患、ＵＩＰパターン疾患、非ＵＩＰパターン疾患に対して

、新規のバイオマーカーセットを作成し得る。又は、非ＵＩＰパターンの疾患の区別又は診断に用いられ得るバイオマーカーを、それとは異なる非ＵＩＰパターンの疾患の区別又は診断に適用し得る。

【０１８７】

[234] 訓練セットの試料は、複数の症状（例えば、疾患や亜型疾患）を含んでもよい。独立テストセットの試料は、複数の症状（例えば、疾患や亜型疾患）を含んでもよい。独立テストセットの試料は、訓練セットの試料とは異なる少なくとも１種の疾患又は亜型疾患を含んでもよい。訓練セットの試料は、独立テストセットの試料とは異なる少なくとも１つの疾患又は亜型疾患を含んでもよい。独立テストセットの試料は、訓練セットの試料に加え、更に少なくとも２種の疾患又は亜型疾患を含んでもよい。例えば、更に含まれる少なくとも２種の疾患又は亜型疾患は、アミロイド沈着又は軽鎖沈着、外因性リポイド肺炎、組織化肺胞出血、又はこれらの任意の組み合わせである。独立テストセットに、訓練セットには含まれない１種以上の新規疾患又は亜型疾患が出現してもよい。訓練セットの試料は、独立テストセットの試料に加え、更に少なくとも２種の疾患又は亜型疾患を含んでもよい。

【０１８８】

[235] この方法では、in silico試料で分類器の性能を評価してもよい。in silico試料は、特に試料サイズが小さい場合、独立テストセットでin vitro試料の混合をシミュレートすることができる。in silico試料は分類器の決定境界、分類器の最適性能に必要な試料の最適数、又はこれらの任意の組み合わせを決定するとき、役に立つ。この方法は、例えば、試料サイズが小さい場合など、プール試料に適用できる。

【０１８９】

[236] 小さい試料サイズとは、１００体、９０体、８０体、７０体、６０体、５０体、４０体、３０体、２５体、２０体、１５体、１０体、又は５体未満の異なる被検体から採取された試料を指してもよい。小さい試料サイズとは、約５０体～約１００体の異なる被検体から採取された複数の試料を指してもよい。小さい試料サイズとは、約１体～約５０体の異なる被検体から採取された複数の試料を指してもよい。小さい試料サイズとは、約１体～約１００体の異なる被検体から採取された複数の試料を指してもよい。小さい試料サイズとは、約１体～約２００体の異なる被検体から採取された複数の試料を指してもよい。小さい試料サイズとは、約１体～約１０体の異なる被検体から採取された複数の試料を指してもよい。小さい試料サイズとは、約１体～約５体の異なる被検体から採取された複数の試料を指してもよい。小さい試料サイズとは、約１体又は約２体の異なる被検体から採取された複数の試料を指してもよい。小さい試料サイズとは、約１体～約１５体の異なる被検体から採取された複数の試料を指してもよい。小さい試料サイズとは、約１体～約８体の異なる被検体から採取された複数の試料を指してもよい。小さい試料サイズとは、約５体～約５０体の異なる被検体から採取された複数の試料を指してもよい。小さい試料サイズとは、約５体～約１００体の異なる被検体から採取された複数の試料を指してもよい。小さい試料サイズは、小さいサイズの独立テスト用試料と訓練用試料を含んでもよい。小さい試料サイズは、被検体が罹患する疾患が大変珍しい亜型であるなどして、被検体の数が限られたことを意味してもよい。単一試料の複製、例えば単一試料の１つ、２つ、３つ、４つ、５つ、又はそれ以上の複製を含むことで、小さい試料サイズを拡張してもよい。単一試料の約１つ又は約２つの複製を含むことで、小さい試料サイズを拡張してもよい。単一試料の約１つ～約３つの複製を含むことで、小さい試料サイズを拡張してもよい。単一試料の約１つ～約４つの複製を含むことで、小さい試料サイズを拡張してもよい。単一試料の約１つ～約５つの複製を含むことで、小さい試料サイズを拡張してもよい。単一試料の約１つ～約１０個の複製を含むことで、小さい試料サイズを拡張してもよい。単一試料の約１つ～約１５個の複製を含むことで、小さい試料サイズを拡張してもよい。単一試料の約１つ～約２０個の複製を含むことで、小さい試料サイズを拡張してもよい。

【実施例】

【０１９０】

[237] 実施例１
[238] 背景：侵襲的処置を使用せずとも特発性肺線維症（ＩＰＦ）を正確に診断するために、ＩＰＦの特徴である一般の間質性肺炎（ＵＩＰ）の組織病理的パターンを特定するRNA-seqデータを用いて、分類器を開発する。この場合、分類器の開発における様々な問題、例えば臨床設定のゲノムデータに機械学習を適用する際に生じる試料のサイズ、異質性、バッチ効果などに直面することになる。
【０１９１】
[239] 方法論：エクソーム濃縮ＲＮＡシーケンシングを、訓練アルゴリズムで用いられる、９０人の患者から得た３５４個の経気管支生検（ＴＢＢ）試料に対して実行する。別途の検証として、追加の患者４９人から得た３～５個の個別ＴＢＢ試料から構成されるプールＴＢＢ試料に対してシーケンシングを行う。教師なしのクラスタリングと差次的発現遺伝子分析を行い、疾患の異質性を特徴付け、ＵＩＰと非ＵＩＰを区別できるゲノム特徴を選択する。小さな試料サイズと、潜在的な疾患異質性による問題を克服できるように、患者ごとに複数の試料を用いて機械学習アルゴリズムを訓練する。テストセットのプール試料を模倣するようにシミュレートしたin silico混合試料に対して評価を行う。テストセットで機械学習アルゴリズムを検証し、複数のバッチに亘る技術的な複製を用いてアルゴリズムの堅牢性を更に評価する。
【０１９２】
[240] 結果：教師なしのクラスタリングと差次的遺伝子発現分析は、患者、特に非ＵＩＰ群の間で高い異質性を示す。罰則付きロジスティック回帰モデルとEnsembleモデルを用いて開発された分類器は、患者ごとに複数の試料をテストする場合、交差検証で約０．９の受信者動作特特性曲線下面積（ＡＵＣ）を以って組織病理的ＵＩＰを分類する。個々の訓練セットの試料からin silicoでシミュレートされるＴＢＢプールを用いて決定境界を定義し、８５％以上の特異度を最適化する。罰則付きロジスティック回帰モデルは、技術的な複製全体において高い再現性を示すので、最終モデルで選択してもよい。最終モデルは、分子テストの前に実験室でプールされた試料を使用すれば、独立テストセットにおいて感度７０％・特異度８８％を示す。
【０１９３】
[241] 結論：試料のサイズ、疾患とサンプリングの異質性、プール効果やバッチ効果による問題を克服し、本明細書に記載の方法は、機械学習とRNA-seqを活用してＵＩＰを特定する、非常に正確且つ堅牢な分類器を提供し得る。
【０１９４】
[242] 前書き：間質性肺疾患（ＩＬＤ）は、同様の臨床症状を伴い、肺間質に影響を与える様々な疾患で構成されている。特発性肺線維症（ＩＰＦ）は、予後が最も悪く、最も一般的なＩＬＤである。ＩＰＦの原因はほとんど不明のままなので、正確且つタイムリーな診断を下すのは難しい。多くの場合、ＩＰＦの正確な診断には、臨床的特徴、放射線特徴、組織病理的特徴の学際的評価が必要なので（Flaherty等2004、Travis等2013：両方とも本文に援用されている）、患者は、明確な結果も出ないまま、長々と続く診断プロセスに苦労させられる。特に、ＩＰＦの特徴である一般の間質性肺炎（ＵＩＰ）の存在を判定するためには、多くの場合、侵襲的手術による組織病理的判定が必要である。更に、組織病理的判定の質は、クリニックによって大きく変動し得るものである（Flaherty等2007：本文に援用されている）。従って、患者の苦労を軽減し、医師が自信を持って臨床診断を迅速に行い、より良い治療決定を下せるようにするためには、手術を必要とせずＵＩＰと非ＵＩＰを区別する、一貫性と正確性を兼ね備えた非侵襲的な診断ツールが非常に重要である。
【０１９５】
[243] このような新規の診断ツールを作り上げるためには、エクソーム濃縮ＲＮＡシーケンシングデータは、手術より非侵襲性である気管支鏡検査を介して採取された経気管支生検試料（ＴＢＢ）から利用できる。研究によると、トランスクリプトミクスデータのゲノム情報は、癌やその他の慢性疾患などの表現型の変化を示し（Tuch等2010、Twine等201

1：両方とも本文に援用されている）、その複雑な特性は疾患に対して明らかな関連性を持たないものを含む、ゲノム全体に広がる複数の遺伝子によるものである（Boyle 2017：本文に援用されている）。更に重要なことに、トランスクリプトミクスデータを使用してＵＩＰを特定することの実現可能性は、既に確立されている（Pankratz等2017：本文に援用されている）。本明細書に記載の方法及びシステムは、上記の問題に解消する分析的解決策を提供する。

【０１９６】

[244] 機械学習法は生物医学的問題の解決に幅広く適用され、それによって研究者は生物学的経路を構築し、臨床的に関連する疾患を特定し、疾患リスクをより正確に予測できるようになり、乳癌（Sorlie等：本文に援用されている）、膠芽腫（Brennan等：本文に援用されている）などの疾患への理解を深めてきた。ただし、最近の機械学習は、医療画像データやソーシャルメディアデータなどの大規模なデータセット向けに設計されている。しかしながら臨床研究では患者の数を増やせないので、多くの場合、試料のサイズが限られてしまう。この場合、多くの患者は生検を受けられる状態ではなくなるので、採取された試料のうち、相当の数が確定診断を得られず、教師付きの学習に適するものではなかった。その上、非ＵＩＰカテゴリーに属するのは１種の疾患ではなく、異なる種類の疾患の集まりである。これは小さな試料サイズと相まって、それぞれの非ＵＩＰ疾患カテゴリーから極一部の試料のみを利用できるようになり、分類が更に難しくなってしまう。この例の更なる特徴は、患者における異質性である。組織病理的特徴は肺全体に亘って均一ではなく、ゲノムの特徴は生検試料を採取された部位により異なる（Kim等：本文に援用されている）。そのような異質性をよく理解するために、患者ごとに複数の試料（最大５つ）を採取し、訓練セットの患者について個別にシーケンシングを行う。このデータセットは、解くべき課題であり、また新しい機会でもある。詳細については、以下に説明する。

【０１９７】

[245] 分類器は診断製品の基盤としても用いられ得るので、更に２つの要件があり得る。まず、費用対効果を高めるために、患者ごとに1回のシーケンシングのみが許容され得るので、独立テストセットにこの実情を反映する必要があり得る。訓練セットの各試料とテストセットのプール試料を分析的にブリッジすることが必要になり得る。また、最終ロック分類器が独立テストセットで良好に稼働することも重要であり得るが、今後の全試料において性能を維持することも重要であり得る。従って、予測可能なバッチ効果に対して高い堅牢性を有する分類器を将来開発することは非常に重要になり得る。

【０１９８】

[246] 下記セクションは、定量分析において解消すべき問題を説明し、この問題を克服するための実用的な解決策を説明する一方、改善の証拠と、このようなアプローチの限界について示し得る。

【０１９９】

[247] 材料及び方法

[248] 研究設計

[249] １８歳以上であり、ＩＬＤに対する組織病理診断を目的として臨床的に必要である肺生検処置計画の適用対象である患者は、試料採取を必要とする複数のラボ実験に志願してもよい（新規ゲノムテスト用の気管支試料、BRAVE）（Pankratz等）。気管支鏡検査の必要としない、勧められていない、又は検査そのものが困難である患者は、研究に参加することができない。病理診断のために行われた生検の種類に基づき、患者のグループ分けをする。BRAVE-1の患者は外科肺生検（ＳＬＢ）を受け、BRAVE-2の患者は病理診断用のＴＢＢを受け、BRAVE-3の患者は凍結生検を受ける。本研究は、各機関の治験審査委員会により承認を受け、患者全員から参加の前に同意を得る必要がある。

【０２００】

[250] 研究において、２０１人のBRAVE患者を、訓練で使用される１１３人のグループ（２０１２年１２月～２０１５年７月に登録）と、検証に用いられる８８人のグループ（２０１４年８月及び２０１６年５月に登録）に前向きに分け得る。訓練グループは最終的に

、使用できるＲＮＡシーケンシングデータと、モデルの訓練や交差検証に使用できる参照標準病理の真理値ラベルを持つ９０人の患者を選別し得る。検証グループは、試料の処理、試料の妥当性、参照標準の真理値ラベルの判定に関するテストセット選別基準を満足する４９人の患者を精選し得る。参照ラベルと関連病理を含むテストセットに関する全ての臨床情報は、分類器パラメータを最終的に確定してロックされたテストセットを前向きに採点するまで、アルゴリズム開発チームに知らせないようにし得る。

【０２０１】

[251] トータルＲＮＡを抽出し、TruSeq RNA Access Library Prep procedure（カリフォルニア州サンディエゴ、イルミナ）に入力して発現エクソン配列を濃縮し、NextSeq v2ケミストリー150サイクルキット（カリフォルニア州サンディエゴ、イルミナ）を用いてNextSeq 500機器でシーケンシングを行い得る。訓練セットとして、９０人の患者から採取された３５４個のＴＢＢ試料のそれぞれに対して個別的にＲＮＡシーケンシングデータを生成し、更に８つのＴＢＢ試料を品質管理のために選択して、センチネルと呼ばれ得る８つの異なるバッチにおいてシーケンシングを繰り返すことができる。独立テストセットとして、各患者の利用可能なＴＢＢ試料から抽出したトータルＲＮＡを等量混合することができ、訓練セットと同様の手順によりシーケンシングを行うが、実行時点をその後にして、異なるバッチを用いる。従って、訓練セットの場合、患者ごとに最大５つのシーケンシングデータがあり得、そのうち１つは個別のＴＢＢ試料に対応し得る。一方、テストセットの場合、全てのＴＢＢ試料と、同患者から得られた対応のＲＮＡ材料を市販試料の実行方法としてシーケンシング前に一緒にプールするため、各患者のシーケンシングデータは１つのみであり得る。

【０２０２】

[252] 病理検討、及びラベル割当

[253] 組織病理診断は、Pankratz等とKim等に記載されているプロセスに従い、病理判定を目的として特別に採取された生検試料とスライドを使用して、病理学者３人の総意により一元的に決定し得る。中心病理診断は、病理判定用の肺葉試料ごとに個別に定められ得る。以下の規則に従って、葉レベルの診断から各患者の参照標準レベルを定める。ＵＩＰ亜型、例えば標準ＵＩＰ（ＵＩＰの特徴を全て表す）、困難ＵＩＰ（標準ＵＩＰより特徴の数が少ない）、指示ＵＩＰ（差が生じるＵＩＰを用いた侵入プロセスの細分化）、又はこれらの任意の組み合わせに基づいて肺葉の診断を行い、その患者の参照ラベルとして「ＵＩＰ」を割り当て得る。肺葉のいずれかが「非ＵＩＰ」病理と診断され（Pankratz等）、他の肺に対して、診断不能、又は分類不能の繊維症ありとの診断を受けた場合、「非ＵＩＰ」を患者レベルの参照ラベルとして患者に割り当て得る。全ての肺葉が分類不能な線維症（例えば、他に分類できないか、「ＣＩＦ、ＮＯＣ」の慢性間質性線維症）との診断を受ける、又は未確定である場合、参照ラベルは割り当てられず、患者は除外され得る。この患者レベルの参照ラベルプロセスは、訓練セットとテストセットにおいて同一であり得るが、訓練セットの各ＴＢＢ試料は、患者レベルで定められた参照ラベルに加え、元の肺葉から試料レベルの参照ラベルを直接継承してもよい。

【０２０３】

[254] 分子テスト、シーケンシングパイプライン、データＱＣ

[255] 気管支鏡検査では、各患者から最大５つのＴＢＢ試料をサンプリングし得る。通常、臨床的に必要である診断手法の間に、２つの上肺葉と３つの下肺葉の試料を採取し得る。分子テスト用のＴＢＢ試料には、核酸防腐剤を入れ、開発研究所に送る出荷前に、又は出荷中に４℃で最大１８日間保存し、その後、凍結保存し得る。トータルＲＮＡを抽出し、定量して、必要に応じて患者ごとにプールし得る。次いで、複数ラウンドの増幅とハイブリダイゼーションを活用してエクソン配列に特定的なプローブする、TruSeq RNA Access Library Prep（カリフォルニア州サンディエゴ、イルミナ）に１５ｎｇを入力することにより、コーディングトランスクリプトームを濃縮させる。処理中の収量基準を満たすライブラリーは、High Outputキット（カリフォルニア州サンディエゴ、イルミナ）を用いて、NextSeq 500機器（2×75bpのペアエンドリード）でシーケンシングを行うことがで

きる。生のシーケンシング（ＦＡＳＴＱ）ファイルは、STAR RNA-seqアライナソフトウェア（Dobin等：本文に援用されている）を使用して、Human Referenceアセンブリ３７（Genome Reference Consortium）にアラインされ得る。63,677のEnsembl注釈付き遺伝子レベルの特徴に対する未加工リードカウントを、HTSeqを用いてまとめ得る（Anders等2015：本文に援用されている）。データ品質指標は、RNA-seqC（DeLuca等：本文に援用されている）を用いて生成し得る。トータルリード、マップユニークリード、塩基ごとの平均カバレッジ、塩基複製率、コーディング領域に整列された塩基の割合、塩基ミスマッチ率、遺伝子内カバレッジの均一性に対する最低基準を満たすライブラリー配列データをダウンストリーム分析で用いてもよい。

【０２０４】

[256] 正規化

[257] ライブラリーアッセイによる濃縮対象にならない機能を除外するためにシーケンシングデータをフィルタリングし、その結果として、２６，２６８個の遺伝子が残り得る。訓練セットの場合、２６，２６８個のEnsembl遺伝子の発現カウントデータを、比率中央値法で推定されたサイズ係数によって正規化し、閉形式であるパラメトリック法を使用して分散安定化変換（ＶＳＴ）により約ｌｏｇ２に変換し得る（DESeq2パッケージ）（Love等2014：本文に援用されている）。訓練セットからの幾何学的アプローチ及びＶＳＴのベクトルを凍結し、正規化用独立テストセットに個別に再度適用して、臨床パターンを模倣してもよい。

【０２０５】

[258] アルゴリズムの訓練及び開発につき、９０人の患者から採取された３５４個のＴＢＢ試料それぞれより、ＲＮＡ配列データを生成し得る。トータルＲＮＡから配列データに至る処理を８回行い、その度に追加ＴＢＢ試料（「センチネル」）８つを複製して、バッチ効果をモニタリングし得る。また、検証のために、ライブラリーの準備とシーケンシングの前に、患者ごとに最低３つから最高５つまで採取された等量混合ＴＢＢからトータルＲＮＡを抽出し得る。従って、訓練セットに属する患者は最大５つのシーケンスライブラリー訓練に寄与することができ、テストセットの患者は、臨床試料を用いる実験計画と同様に、単一のシーケンスライブラリーで表すものとする。

【０２０６】

[259] 差次的発現分析

[260] 標準パイプライン（Anders等2013：本明細書に援用されている)を用いて検出された差次的発現遺伝子を、非ＵＩＰ試料とＵＩＰを分類するために直接使用できるか否かについて検討し得る。差次的発現遺伝子は、Bioconductor RパッケージであるDESeq2を使用して特定し得る（Love等2014）。訓練セットにおける遺伝子レベルの未加工発現カウントを使用して、差次的分析を行い得る。テスト条件調整後のｐ値のカットオフ値を＜０．０５に、倍数変化を＞２にし、差次的発現遺伝子を選択する。訓練セットでは、非ＵＩＰ試料とＵＩＰ試料、又はＵＩＰ試料同士に対して一対差分分析を実行することができ、非ＵＩＰ疾患のそれぞれに対して１０個以上の試料、例えば、細気管支炎（Ｎ＝１０）、過敏性肺炎（ＨＰ）（Ｎ＝１３）、非特異的間質性肺炎（ＮＳＩＰ）（Ｎ＝１２）、組織化肺炎（ＯＰ）（Ｎ＝２３）、呼吸細気管支炎（ＲＢ）（Ｎ＝１６）、サルコイドーシス（Ｎ＝１１）を用意する。主成分分析（ＰＣＡ）プロットを、上記で特定した差次的発現遺伝子を用いて、全ての訓練用試料に対して作成してもよい。

【０２０７】

[261] 遺伝子発現相関ヒートマップ

[262] 代表患者６人からの採取された試料の相関値ｒ２を、ＶＳＴ遺伝子発現を使用して計算し、患者の順序を保存した相関行列のヒートマップをプロットして、遺伝子発現の患者内、及び患者間の異質性を視覚化し得る。患者６人は、患者内・患者間の異質性の全範囲で表せるように厳選した患者であることができ、それぞれ上肺葉対下肺葉のラベルが異なるＵＩＰ患者１人と非ＵＩＰ患者１人、上肺葉対下肺葉のラベルが同じ、又は類似するＵＩＰ患者２人と非ＵＩＰ患者２人を含む。ヒートマップは、gplots Rパッケージのhe

atmap.2機能を用いて作成できる。

【０２０８】

[263] 分類器の開発

[264] 分類器の開発と評価を、図２８にまとめた。目標は、正確且つ再現性のあるＵＩＰ／非ＵＩＰ予測を提供し、ＩＬＤ患者に対する侵襲的処置を減らして臨床的ニーズを満たす、ＴＢＢ試料に対して堅牢なバイナリ分類器を構築することである。正の予測値が高くなるように、特異度テスト（特異度＞８５％）を高く設計してもよい。テストでＵＩＰが予測される場合、この結果の信頼度は相当に高くなり得る。

【０２０９】

[265] 分類器開発における特徴フィルタリング

[266] まず、生物学的に無意味な特徴、又は情報量の少ないであろう特徴は、試料間のバラツキのない発現レベルのために除去され得る。Ensemblで偽遺伝子、リボソームＲＮＡ、Ｔ細胞受容体又は免疫グロブリン遺伝子の個々のエクソン、無情報及び低発現遺伝子として注釈付きの遺伝子は、訓練セット全体において未加工カウント発現レベル＜５として取り除かれるか、又は訓練セットにおける試料の５％未満に対してカウント＞０として発現され得る。

【０２１０】

[267] 複数のバッチで処理される同試料で可変的に発現する遺伝子は、生物学的要因ではなく技術的要因に対する感度を示唆するため、排除され得る。上記遺伝子を特定するために、複数のアッセイプレートで処理されたセンチネルＴＢＢ試料を線形混合効果モデルに適合させ得る。このモデルは、各遺伝子に個別に適合させることができる。ここで、g_{ij}は試料ｊ及びバッチｉの遺伝子発現、μは平均遺伝子発現であり、次のように表される。

【０２１１】

【数１】

$$g_{ij} = \mu + \beta sample_{ij} + batch_i + e_{ij} \qquad (1)$$

[268] セット全体において、$sample_{ij}$は生物学的に異なる試料の固定効果であり得、$batch_i$はバッチ固有のランダム効果であり得る。全変動を用いて、非常に可変性の高い遺伝子を特定し得る。この測定による上位５％の遺伝子は除外し得る（図３９から図４４）。その結果、１７，６０１個のEnsembl遺伝子がダウンストリーム解析の候補として残り得る。

【０２１２】

[269] 患者内のIn silico混合

[270] 個々のＴＢＢ試料により分類器の訓練と最適化を行い、サンプリングの多様性と機能選択、重み付けプロセス中に利用可能な情報コンテンツを最大化し得る。複数のＴＢＢ試料は、抽出後の段階でＲＮＡとしてプールされることができ、プールＲＮＡは、ライブラリーの準備、シーケンシング、分類を通じて１回の反応で処理され得る（Pankratz等）。個々の試料で開発された分類器が、プール試料ルで高い性能を達成できるかどうかを評価することができる。個々の試料データからプール試料を「in silico」でシミュレートする方法を開発できる。最初に、訓練セット全体における遺伝子全体の幾何学的アプローチを用いて計算されたサイズ係数により、未加工のリードカウントを正規化し得る。試料ｉ＝１、…、ｎ、遺伝子ｊ＝１、…、ｍの正規化カウントC_{ij}は、次の式で算出され得る。

【０２１３】

【数２】

$$C_{ij} = K_{ij} / S_j$$

【０２１４】

ここで、

【０２１５】
【数３】

$$s_j = \mathrm{median}_i \frac{K_{ij}}{\left(\prod_{v=1}^{m} K_{iv}\right)^{1/m}}$$

とＫ$_{ij}$は試料ｉと遺伝子ｊの未加工カウントを意味する。次に、各訓練患者ｐ＝１、…、Ｐについて、in silico混合カウントＫｐ$_{ij}$は次の式で算出される。
【０２１６】
【数４】

$$K_{ij}^{p} = \frac{1}{n_p} \sum_{i \in I(p)} C_{ij}$$

【０２１７】
ここで、Ｉ（ｐ）は、患者ｐに属する個々の試料ｉの指標セットである。訓練セットの凍結分散安定化変換（ＶＳＴ）をＫｐ$_{ij}$に適用できる。
【０２１８】
[271] 分類器の訓練
[272] テストは病理学により定義された参照ラベルを認識して呼び出すことを目的とし得るので、参照ラベルは分類器の訓練（Tuch等）の応答変数、及びエクソーム濃縮、フィルター処理、正規化を行った予測特徴としてのＲＮＡ配列データで定義され得る。ランダムフォレスト、サポートベクターマシン（ＳＶＭ）、勾配ブースティング、ニューラルネットワーク、罰則付きロジスティック回帰を含む複数の分類モデルを評価し得るる（Dobson等：本文に援用されている）。各分類器は、５分割交差検証と１人患者抜き交差検証（LOPO CV）に基づいて評価し得る（Friedman等：本文に援用されている）。Ensembleモデルは、個々のモデルのスコアの加重平均を介して個々の機械学習方法を組み合わせることで検査することもできる。
【０２１９】
[273] 過剰適合を最小限に抑えるために、訓練・評価中に、各交差検証分割を層化して、単一の患者からの全てのデータを特定の分割に含めるか保持し得る。ハイパーパラメータ調整は、ネスト交差検証法を用いて各交差検証分割内で実行し得る（Krstajic D等2014：本文に援用されている）。ランダム検索と１つの標準エラールール（Hastie、Tibshirani、Friedman 2009：本文に援用されている）を選択して、内部ＣＶから最適なパラメータを選択し、潜在的な過剰適合を更に最小限に抑え得る。最終的に、完全な訓練セットでハイパーパラメータ調整を繰り返し、最終ロック分類器のパラメータを定義する。様々な機械学習アルゴリズムのトレーニングのパイプラインは、Ｒパッケージ（DESeq2、hclust、cv.glmnet、caret、caretEnsemble）を使用して自動化と実行を行い得る。
【０２２０】
[274] 完全に独立した検証のベストプラクティスでは、テスト決定境界を含む全ての分類器パラメータを前向きに定義する必要があり得る。従って、訓練セットのデータのみを使用して行ってもよい。テストセットはプールＴＢＢを患者レベルで分類するため、提案されたin silico混合モデルを使用して、訓練セット内の患者レベルのスコアの分布をシミュレートし得る。患者内の混合物は、各LOPO CV分割で１００回シミュレートし、ＶＳＴ発現に遺伝子レベルの技術的変動が追加され得る。遺伝子レベルの技術的変動は、混合効果モデルを使用して推定し得る。ＴＢＢ試料の式（１）は、複数の処理バッチに亘って複製してもよい。最終決定境界は、感度を著しく損なわず（≧０．６５）特異度（＞０．８５）を最適化するように選択し得る。性能は、複製されたin silico混合シミュレーションからの患者レベルのLOPO CVスコアを使用して推定し得る。特異度を保守的にするために、９０％を超える平均特異度の基準を設定して最終決定境界を選択し得る。シミュレーションで同様の推定性能を備えた決定境界の場合、最も高い特異度を持つ決定境界を選択できる（図４６Ａから図４６Ｂ）。

【０２２１】
[275] 試料のバッチ効果と監視体系の評価
[276] 未確定の臨床患者集団に対する分類性能の拡張性を確保するため、グローブシフト、回転、圧縮、又はスコア分布の拡大を引き起こし得るバッチ効果と呼ばれる深刻な技術的要因がないことを保証することが重要であり得る。既存のデータのバッチ効果を定量化し、観察可能なバッチ効果に対する候補分類器の堅牢性を評価するために、９つの異なるＴＢＢ試料を採点し、各バッチ内で３回複製し、３つの異なる処理バッチで処理し、線形混合効果モデルを使用して各分類器のスコアの変動を評価した。線形混合モデルの低スコアの変動によって示されるように、バッチ効果に対してより堅牢なモデルを、独立した検証の最終モデルとして選択することができる。バッチ効果を監視するために、新しい処理バッチごとにＵＩＰ試料と非ＵＩＰ対照試料を処理し得る。潜在的なバッチ効果をキャプチャするには、これらの複製対照試料のスコアを比較し、推定スコアの変動性が事前に指定された閾値σ_{sv}よりも小さいか否かを、in silicoの患者レベルのLOPO CVスコアを使用して訓練により決定し得る。
【０２２２】
[277] 独立検証
[278] 最終候補分類器は、４９人の患者からのＴＢＢ試料の盲検化独立テストセットで前向きに検証され得る。ロックアルゴリズムを使用してテストセットの分類スコアを導出し、事前に設定された決定境界と比較して、ＵＩＰ対非ＵＩＰ呼び出しのバイナリ予測を行い得る。決定境界より上の分類スコアはＵＩＰ、決定境界以下は非ＵＩＰと見なし得る。ＲＯＣを構築し、ＡＵＣを計算するために、連続分類スコアを組織病理学ラベルと比較し得る。バイナリ分類予測を組織病理ラベルと比較して、感度や特異度などのバイナリ分類性能を算出し得る。
【０２２３】
[279] スコア変動シミュレーション
[280] 臨床現場では、臨床試料の分類スコアが安定しているか否か、潜在的な技術的要因の影響を受けていないか否かを監視することが重要であり得る。これを行うには、分類器が許容できるスコアの変動の限界を前向きに検討する必要があり得る。LOPO CVスコアが対象集団の分類スコアの分布を表すという仮定の下で、ＵＩＰコールと非ＵＩＰコール間の感度、特異度、フリップレートについてシミュレーションを実行し得る。最初のステップとして、in silicoの患者レベルのLOPO CVスコアにシミュレートされたノイズを追加し得る。この場合、ノイズはｅ～Ｎ（Ｏ、σ^2）としてシミュレートでき、σ^2は０、０．０１、…、１０になり得る。次に、シミュレートされたノイズのスコアを用いて、感度、特異度、フリップレートを算出し得る。シミュレーションを１，０００回繰り返し得る。１，０００セットのシミュレートスコアを使用して、個々の閾値、σ_{spec}、σ_{sens}、σ_{flip}は、ノイズの標準偏差σの最大値として定義されることができ、推定（平均）特異度＞０．９、感度＞０．６５、フリップレート＜０．１５になる。分類スコアの変動性の最終閾値は、$\sigma_{sv} = \min(\sigma_{spec}, \sigma_{sens}, \sigma_{flip})$のように定義できる。
【０２２４】
[281] Ensembleモデルの閾値は、特異度、感度、フリップレートに対してそれぞれ０．９、１．８、１．１５であることができ、最終閾値はσ^{E}_{sv}＝０．９であり得る（図４８Ａから図４８Ｃ）。罰則付き回帰モデルの閾値は、特異度、感度、フリップレートに対してそれぞれ０．４８、０．７８、０．６８であることができ、最終値はσ^{PL}_{sv}＝０．４８であり得る。
【０２２５】
[282] 結果
[283] ＩＬＤ疾患の分布
[284] 表14に、ＵＩＰ及び非ＵＩＰグループ内のＩＬＤ疾患患者の分布をまとめる。患者の中で、ＵＩＰパターンの患者の有病率は、ｐ値０．２７のテストセット（４７％）よりも訓練セット（５９％）の方が高くなり得る。訓練セットの患者３人とテストセットの

患者１人は、潜在的な異質性を有し得る。肺葉のうち一方には、非ＵＩＰ疾患（非特異的間質性肺炎、肺高血圧症、又は過敏性肺炎）の１つが割り当てられ、他方にはＵＩＰパターンが割り当てられ、最終的な患者レベルのラベルはＵＩＰとなり得る。

【０２２６】

[285] 非ＵＩＰグループには、臨床診療で一般的に遭遇する多様な異種疾患が含まれ得る。試料サイズが小さいため、一部の疾患は患者が１～２人のみであり得る。３種の新種疾患（アミロイド沈着又は軽鎖沈着、外因性リポイド肺炎、診断しやすい組織化肺胞出血）がテストセットには存在し得るが、訓練セットには存在しないことがある。

【０２２７】

[286] 患者内の異質性

[287] 同じ患者から採取された試料の異質性は、組織病理診断と遺伝子発現の両方で観察され得る。ＵＩＰ及び非ＵＩＰグループにまたがる疾患に罹患するこのような患者３人は、患者レベルの診断分類に算出上の課題をもたらし得る。６人の患者から採取された試料の相関行列はまた、発現プロファイルにおける顕著な患者内・患者間変動を明らかにし得る（図３８）。図３８は、異なる肺葉で同じラベルと類似の遺伝子発現パターンを持つ２人の非ＵＩＰ患者と（図３８の患者１と２）、同一又は類似のラベルと高度に相関する発現プロファイルを持つ２人のＵＩＰ患者（図３８の患者５と６）、更に、異なるラベル及び異種発現を有するＵＩＰ患者１人と非ＵＩＰ患者１人（図３８の患者３と４）とともに、患者内・患者全体で観察され得る異質性の全スペクトルの代表的な視覚化を図示する。

【０２２８】

[288] ＵＩＰと非ＵＩＰ間のＤＥ分析

[289] ＵＩＰと非ＵＩＰの間でDESeq2により検出された差次的発現遺伝子が２つの診断クラスを予測できるか否かをまず調べる。ＵＩＰと非ＵＩＰの間で１５１の有意に差次的発現遺伝子を特定することができ（調整済みｐ＜０．０５、倍率変化＞２）、ＵＩＰで５５個の上方制御遺伝子と９６個の下方制御遺伝子である（図２９、表１５）。しかし、これらの差次的発現遺伝子のみを使用すると、ＰＣＡプロット（図３０）に示されているように、２つのクラスを完全に分離できないことがある。それに対して、１９０個の分類器遺伝子にまたがるＰＣＡは、２つのクラスをよりよく分離し得る（図３１）。

【０２２９】

[290] 非ＵＩＰ疾患患者の異質性

[291] 異質性は、１０を超える臨床的に定義された疾患で構成される非ＵＩＰ試料の遺伝子発現で観察され得る。ＵＩＰ試料と、試料サイズが１０を超える各非ＵＩＰ疾患サブタイプとの間で有意に異なる（調整済みｐ＜０．０５、倍率変化＞２）遺伝子が特定され得る（表１５）。差次的発現遺伝子の数が多いほど、非ＵＩＰ疾患の亜型はＵＩＰとは異なり得る。各非ＵＩＰ亜型の差次的遺伝子のリストと全ての非ＵＩＰ試料の差次的遺伝子のリストを比較すると、重複する遺伝子の数は、個々の非ＵＩＰ亜型で特定された差次的遺伝子の数に大きく依存し得る。それによると、一部の非ＵＩＰ疾患は、全ての非ＵＩＰ試料とＵＩＰ試料間で見られる全体的な差次的遺伝子により支配的な影響を与え得る（表１５）。更に、個々の非ＵＩＰ疾患で特定された遺伝子の中で重複する差次的遺伝子はほとんどない。例えば、サルコイドーシスの１１７４個の差次的遺伝子とＲＢの７０１個の間で共通しているのは１７２個で、サルコイドーシス、ＲＢ、ＮＳＩＰの差次的遺伝子の中に共通するのは６個であり得る。細気管支炎、ＮＳＩＰ、ＨＰの差次的遺伝子には共通の遺伝子が存在しない。これは、非ＵＩＰ試料の疾患内の異なる分子発現パターンを示唆し得る。

【０２３０】

[292] 非ＵＩＰ亜型とＵＩＰ試料間で差次的に発現した遺伝子を使用したＰＣＡプロットは、特定の非ＵＩＰ疾患亜型がＲＢやＨＰなどの疾患のＵＩＰ試料から十分に分離される傾向があることを示している（図３９、図４１）。ただし、他の非ＵＩＰ試料にＵＩＰ試料が散在し得る（図４０、図４３）。これは、１つの非ＵＩＰ亜型に由来する差次的遺

伝子が他の非ＵＩＰ疾患に一般化できないことがあることを示している。
【０２３１】
[293] 患者におけるin silico混合とin vitroプーリングの比較
[294] 各患者におけるin silico混合試料を使用して、訓練セット内で評価するためのin vitroプール試料をモデル化し得る。in silico混合試料とin vitroプール試料が合理的に一致するように、１１人の患者のプール試料をシーケンシングし、in silico混合試料と比較し得る。in silico混合試料とin vitroプール試料のペアの２６，２６８遺伝子の発現レベルに基づく平均ｒ平方値は０．９９（ＳＤ＝０．００３）であり、平均ｒ平方値が、技術的複製では０．９８（ＳＤ＝０．００８）、生物学的複製では０．９４（０．０４）であることを考慮すると、in silico混合試料のシミュレートされた発現レベルはin vitroでプールされた試料のそれと良好に一致し得る。
【０２３２】
[295] 2つの候補分類器、つまりEnsembleと罰則付きロジスティック回帰モデル（以下で説明）によるin silico及びin vitro混合試料の分類スコアも散布図で比較できる（図３２、図３３）。各in vitroプール試料の複製数は３～５の範囲であるため、複数の複製の平均スコアを使用できる。in silico混合試料の分類スコアは、両方の分類器でピアソンの相度度が０．９９なので、in vitroでプールされた試料の分類スコアと非常に高い相関関係を示し得る（図３２、図３３）。ポイントは、明らかなシフトや回転なしで、Ｘ＝Ｙの線の周りに落ち得る。
【０２３３】
[296] 訓練セットに対する交差検証
[297] ９０人の患者から採取された３５４個のＴＢＢ試料の訓練セットでの特徴選択と機械学習アルゴリズムを評価し得る。最初の試みとして、個々の方法とEnsembleモデルは、各分割の経験的ＡＵＣの平均を使用して推定された５分割ＣＶ及び交差検証ＡＵＣ（ｃｖＡＵＣ）に基づいて個別に評価され得る。全体として、罰則付き回帰モデル（ｃｖＡＵＣ＝０．８９）などの線形モデルの方が、ランダムフォレスト（ｃｖＡＵＣ＝０．８３）や勾配ブースティング（ｃｖＡＵＣ＝０．８４）などの非線形ツリーベースのモデルよりも優れ得る。ニューラルネットワーク分類器のｃｖＡＵＣは０．８未満になり得る。最も優れた性能を示すのは、（１）線形カーネルと放射状カーネルを備えたＳＶＭのアンサンブルモデル、及び（２）罰則付きロジスティック回帰であり得る。両方ともｃｖＡＵＣ＝０．８９である。ただし、疾患間の異質性と小さいサイズの試料のため、全てのモデルにおけるＣＶ性能は、分割によって大きく異なることがある。
【０２３４】
[298] LOPO CVでは、LOPO CV分割内の患者ごとにインシリコ混合試料を１００回複製して、患者レベルの性能を評価できる。算出した各試料の分類スコアとin silico混合試料の平均スコアを、図３４と図３５に図示する。全体的に、患者レベルの性能が、試料レベルの性能に比べてわずかに高い場合がある。LOPO CV分割全体の合計スコアに基づいて、Ensembleモデルと罰則付きロジスティック回帰モデルは、それぞれ、試料レベルでＡＵＣの０．９［０．８７－０．９３］及び０．８７［０．８３－０．９１］、in silico混合患者レベルで０．９３［０．８８－０．９８］及び０．９１［０．８５－０．９７］を出し、最も優れた性能示す（図３６Ａ）。
【０２３５】
[299] 分類器の堅牢性
[300] 推定スコアの変動は、Ensembleモデルと罰則付きロジスティック回帰モデルでそれぞれ０．４６と０．２２であり得る（表１６）。両方とも、スコア変動許容範囲の事前に指定された閾値である０．９及び０．４８未満であり得る（図４７Ａから図４７Ｃ、図４８Ａから図４８Ｃ）。Ensemble分類器のスコア範囲が罰則付きロジスティック回帰分類器より広いことを考慮して、スコアの５％及び９５％分位の範囲に対する変動性の割合を比較し得る。全体的に、罰則付きロジスティック回帰分類器の方が、アンサンブルモデルよりもスコアの変動が少ない。これは、罰則付きロジスティック回帰が技術的（試薬／実

験室）バッチ効果に対してより堅牢であり、技術的複製に対してより一貫したスコアを提供することを示唆する（表１６）。高い交差検証性能と堅牢性により、独立検証の最終候補モデルとして、罰則付きロジスティック回帰モデルを選択し得る。

【０２３６】

[301] 独立検証性能

[302] 事前に指定された決定境界０．８７でロックされた罰則付きロジスティック分類器を使用し、in vitro混合試料の独立テストセットに基づいて検証性能を評価し得る。最終分類器は、ＡＵＣ０．８７［０．７６－０．９８］で特異度０．８８［０．７０－０．９８］、感度０．７０［０．４７－０．８７］であり得る（図３６Ｂ、図３７）。検証性能のポイント推定値は、in silicoの患者レベルの訓練ＣＶ性能よりは低いが、ＡＵＣ、感度、特異度のｐ値がそれぞれ０．６、０．７、１であり、誤差範囲である。

【０２３７】

[303] 論議

[304] 本研究では、重大な課題が存在する場合でも、正確性と堅牢性を兼ね備えた分類が可能である。適切な統計的手法、機械学習アプローチ、ＲＮＡシーケンシング技術を活用することにより、有意味な診断テストを提供し、間質性肺疾患に罹患する患者の治療を改善することができる。

【０２３８】

[305] 機械学習、特にディープラーニングは、ここ数年で革命的な進歩を成し遂げた。最近開発された高度に洗練されたツールにより、多くのアプリケーションで分類性能が劇的に改善され得る（Lecun等：本文に援用されている）。しかしながら、ツールのほとんどは、容易に入手でき、信頼性の高いラベルと大きな試料サイズを必要とし得る。性能改善は、高品質のラベルを持つ試料の数と、直接且つ明確に関係している（Gu等、Sun等）。本プロジェクトでは、患者試料に基づいた他の臨床研究と同様に、試料サイズが制限され得る。例えば、訓練セットに患者は９０人に過ぎない（表１４）。それに加えて、非ＵＩＰグループは生理学的に均質な疾患ではなく、それぞれ独自の生理を持つ様々な種類の疾患の集まりであり、一部の疾患に至っては、訓練セットに１～２人の患者しか含まれていない（Libbrecht等：本文に援用されている）（表１４）。当然のことながら、これらの非ＵＩＰ疾患は、生理学的に異なるだけでなく、分子レベルや遺伝子レベルでも異なり得る。訓練用試料を利用して、ＵＩＰグループとの差別化のために非ＵＩＰ疾患に共通する特徴の特定を試みたところ、何も現れなかった（表１５、図３８）。更に、３種以上の疾患（アアミロイド沈着又は軽鎖沈着、外因性リポイド肺炎、組織化肺胞出血）がテストセットには存在するが、訓練セットには存在しない（表１４）。ＵＩＰの割合の変化は、訓練（５９％）とテスト（４７％）の間にも現れる。最後の２つの要因により、訓練セットの交差検証性能と比較して、テストセットでの性能がわずかに低い理由を説明できる。この状況では、大きな試料サイズを活用する最近の機械学習は適用できない。場合によっては、より伝統的な線形モデルやツリーベースのモデルに焦点を合わせる。また、個々の非ＵＩＰ疾患グループの試料サイズが小さすぎて、ツリーモデルがキャプチャしようとする相互作用を強化できないため、候補の中で、線形モデルが非線形ツリーベースのモデルよりも優れている理由も理解できる。

【０２３９】

[306] 訓練用試料のサイズに直接対処するために、同じ患者から採取された最大５つの異なるＴＢＢ試料に対してＲＮＡ抽出からシーケンシングまで一通りこなして、９０人の患者セットを試料３５４個に成功的に拡張し得る（表１４）。概念的にはデータ増強のアイデアに似ているが、増強データに対してシミュレートや外挿を行う代わり、同じ患者の複数のＴＢＢ試料での実際の実験からシーケンシングデータを生成し得る。目標は、分類情報を強化するための追加性能を提供することである。交差検証分割を定義し、性能を評価する際に、患者を最小単位として使用する場合は特に注意が必要である。これにより、試料数の多い患者の重みが増すことや、同じ患者の試料がモデル作成とモデル評価の両側にまたがった過剰適合を引き起こすことを防止することができる。ネスト交差検証は、モ

デルの選択とパラメータの最適化に用いられる１つのＳＤ（標準偏差）ルールと同様に適用され、小さな試料サイズによる性能の高い変動性を正しく考慮し、モデルの複雑さを積極的に削減し、過剰適合を防止し得る。

【０２４０】

[307] 訓練セットで患者ごとに複数のＴＢＢ試料を実行すると、試料サイズの制限に役立つが、新しい問題が生じ得る。商業的環境では、費用の問題により患者ごとにシーケンシング実行は１回に制限され得る。そのため、１人の患者内の複数のＴＢＢ試料からのＲＮＡ材料に対するシーケンシングを行う前にプールする必要があり得る。しかしながら、個々のＴＢＢ試料で訓練された分類器をプールＴＢＢ試料に適用できるか否かは、検証実験を開始する前に対処する必要がある重要な問題になり得る。よって、一連のin silico混合シミュレーションを実行して、テストセットの患者レベルのin vitroプールを模倣する。このアプローチは、分類器の将来の決定境界と、最良の分類性能を達成するために必要な最適なＴＢＢの数を定義するための基本的な構成要素にもなる（Pankratz等）。シミュレートされたin silicoデータは、実験in vitroデータとよく一致し（図３２、図３３）、このアプローチを使用してプール試料に期待される性能を推定し、プール設定で検証実験を進めることができる。この例では、プール試料が同じタイプ（ＴＢＢ）で、同じ患者のものであり、よって重複リードの割合やリードの総カウントなど、類似の特性を持っているので、このin silicoアプローチはうまく機能するはずである。しかしながら、提案されているin silico混合モデルを拡張して、異なる特性や品質の試料、例えばＵＩＰと非ＵＩＰ試料、又は血液などの異なる種類の試料と混合したＴＢＢを混合するのは難しい。これらの場合、リードの総カウント大幅に多い試料は、ここで提案する混合モデルの基本的な仮定に違反する結合試料の表現を支配する傾向がある。このような複雑な手順と生物学的相互作用を正確にモデル化するためには、より高度な方法論が必要になる。

【０２４１】

[308] 成功的な検証により必要な臨床性能を成し遂げたことが確認されると（図３６Ａから図３６Ｂ、図３７）、患者の治療を改善することを目的とした有用な市販製品を完成させる最初のステップとなり得る。同様に重要であるが見落とされがちなのが、将来の患者の流れに一貫した信頼できる性能を提供することの重要性である。これには、分類スコアに体系的な変化をもたらし、誤った臨床予測を生じさせる疑いのある、患者から採取されたデータを配列する潜在的なバッチ効果に対処するための予防的な予測が必要となり得る。これは、上流側の機能選択（図３９から図４４）から取り組み、バッチ効果に非常に敏感な遺伝子を以降の分析から取り除くことで対処可能である。更に、３つの異なるバッチの１０個の異なるＴＢＢ試料について、追加の実験データが生成されることがあり、訓練用試料の生成にはいずれも役に立たない。この実験を活用して、各候補モデルの不可視バッチに対する堅牢性を直接評価し、最終モデルを選択する。ただし、実験データは、有限数のバッチを評価することになる。従って、予期しない変化を予測するために、市販のプレート／バッチのそれぞれで実行される制御試料に基づいて監視体系を開発し、予測できない潜在的な変化を検出するべきである。このような予期しない変更が発生する疑いがある場合は、新しいスコアを検証分類スコアの空間にマッピングするために、バッチ修正に直接対処する正規化方法が必要となる。

【０２４２】

[309] 結論

[310] 非ＵＩＰクラスの限られた試料サイズと高い異質性は、この例で直面する2つの主要な課題であり得、臨床研究では一般的な問題でもあり得る。更に、成功的な市販製品は、以来の着信試料全てに対して経済性と一貫性を兼ね備える必要があり得る。非常に正確性と堅牢性を兼ね備えた分類を実現することは難しいが、不可能ではない。本実施例により、この方法論が成功することが証明されている。同様の問題に直面している他の臨床シナリオにも適用できるはずである。

【０２４３】

[311] 実施例２：分子プロファイリング及び細胞診

[312] 被検体は肺癌の症状を見せている。被検体は、被検体を診察して内分泌専門医に紹介する主治医にまず相談をもちかける。内分泌専門医は気管支鏡検査で試料を採取し、細胞検査室に試料を送る。細胞検査室では、気管支鏡検査の一部に対して定期的な細胞検査を行うが、結果は不明確である（つまり、未確定診断）。細胞検査室は、内分泌専門医に残りの試料が分子プロファイリングに適している疑いがあると提案し、内分泌専門医はそれに同意する。

【０２４４】

[313] 残りの試料は、本明細書に記載の方法及び組成物を使用して分析される。分子プロファイリング分析の結果は、初期段階の肺癌の疑いが高いことを示唆している。更に、分子プロファイリング分析を患者データと組み合わせる。内分泌専門医は結果を確認し、推奨される治療法を処方する。

【０２４５】

[314] 細胞検査室は、内分泌専門医に定期的な細胞検査と分子プロファイリングに対する代金を請求する。内分泌専門医は細胞検査室に代金を支払い、提供された全ての製品とサービスに対して被検体の保険会社に請求する。細胞検査室は、分子プロファイリングの支払いを分子プロファイリングの専門企業に渡し、差額を残す。

【０２４６】

[315] 実施例３

[316] 被検体は、受動喫煙にさらされ、肺癌のリスク状態である。被検体は肺癌の無症状を示す。医療従事者は、被検体から鼻組織試料を採取する。本明細書に記載の分子分類器で、鼻組織試料を分析する。複数のバイオマーカーの有無に基づいて、医療従事者は、低線量ＣＴスキャンを受けること、又は、分子分類器を使用して１年後に別の鼻組織試料を分析することを勧める。

【０２４７】

[317] 実施例４

[318] 被検体は、肺結節が存在することが既に確認されている。医療従事者は、被検体から鼻組織試料を採取する。本明細書に記載の分子分類器で、鼻組織試料を分析する。複数のバイオマーカーの有無に基づいて、医療従事者は、気管支鏡検査を受けること、又は、分子分類器を使用して１年後に別の鼻組織試料を分析することを勧める。

【０２４８】

[319] 実施例５

[320] 被検体は現在、介入療法を受けている。医療従事者は、被検体から鼻組織試料を採取する。本明細書に記載の分子分類器で、鼻組織試料を分析する。複数のバイオマーカーの有無に基づいて、医療従事者は、介入療法を続けること、又は、介入療法を中止して別の介入療法を受けることを勧める。

【０２４９】

[321] 実施例６

[322] 被検体は悪性腫瘍の外科切除を前に受けている。本明細書に記載の分子分類器で、鼻組織試料を分析する。複数のバイオマーカーの有無に基づいて、医療従事者は、被検体に治療計画を勧める、又は、分子分類器を使用して１年後に別の鼻組織試料を分析することを勧める。

【０２５０】

コンピュータ制御システム

[323] 本開示は、本開示の方法を実装するようにプログラムされたコンピュータ制御システムを提供する。図２６は、ここで提供される方法を実装するようにプログラム又は構成されたコンピュータシステム２６０１を図示する。コンピュータシステム２６０１は、肺疾病の診断、被検体の肺疾病を発症するリスクの予測、肺疾病を有する被検体の治療の有効性の予測、又はこれらの任意の組み合わせなどを規制できるし、様々な態様、例えば、（ｉ）試料の１つ以上のバイオマーカーとバイオマーカーの参照セットとの比較、（ｉｉ）アルゴリズムの訓練による分類器の開発、（ｉｉｉ）分類器を適用して診断、予測、

又は試料の入力に基づく推奨、又は（ｉｖ）これらの任意の組み合わせを実現することができる。コンピュータシステム２６０１は、ユーザの電子デバイス、又は電子デバイスから離れた場所にあるコンピュータシステムである。電子デバイスは、モバイル電子デバイスである。

【０２５１】

[324] コンピュータシステム２６０１は、中央処理装置（ＣＰＵ、本明細書では「プロセッサ」及び「コンピュータプロセッサ」）２６０５を含む。中央処理装置は、シングルコア又はマルチコアプロセッサ、もしくは並列処理用の複数のプロセッサであってもよい。

コンピュータシステム２６０１は、メモリ又はメモリ領域２６１０（例えば、ランダムアクセスメモリ、読み取り専用メモリ、フラッシュメモリ）、電子記憶部２６１５（例えば、ハードディスク）、１つ以上の他のシステムと通信を行う通信インターフェース２６２０（例えば、ネットワークアダプタ）、周辺機器２６２５（キャッシュ、他のメモリ、データストレージ、及び／又は電子ディスプレイアダプタ）を備える。メモリ２６１０、記憶部２６１５、インターフェース２６２０、周辺機器２６２５は、マザーボードなどの通信バス（実線）を介してＣＰＵ２６０５と通信している。記憶部２６１５は、データを保存するためのデータストレージユニット（又はデータリポジトリ）であってもよい。コンピュータシステム２６０１は、通信インターフェース２６２０により、コンピュータネットワーク（「ネットワーク」）２６３０に動作可能に結合する。ネットワーク２６３０は、インターネット及び／又はエクストラネット、又は、インターネットと通信しているイントラネット及び／又はエクストラネットであってもよい。ネットワーク２６３０は、場合によっては、通信ネットワークやデータネットワークであってもよい。ネットワーク２６３０には、クラウドコンピューティングなどの分散コンピューティングを可能にする１つ以上のコンピュータサーバが含まれる。ネットワーク２６３０は、場合によってはコンピュータシステム２６０１の助けを借りて、ピアツーピアネットワークを実装できる。これにより、コンピュータシステム２６０１に接続されたデバイスは、クライアント又はサーバとして動作できるようになる。

【０２５２】

[325] ＣＰＵ２６０５は、プログラム又はソフトウェアで具現化できる一連の機械可読命令を実行する。命令は、メモリ２６１０などのメモリ領域に保存する。命令は、その後、本開示の方法を実装ためにプログラムされた、又はその他の方法で構成される、ＣＰＵ２６０５に送られる。ＣＰＵ２６０５により実行される動作の例には、フェッチ、デコード、実行、ライトバックが含まれる。

【０２５３】

[326] ＣＰＵ２６０５は、集積回路などの回路の一部にしてもよい。システム２６０１の１つ以上の他の構成要素が回路に含まれてもよい。一部において、回路は特定用途向け集積回路（ＡＳＩＣ）である。

【０２５４】

[327] 記憶部２６１５は、ドライバー、ライブラリー、保存されたプログラムなどのファイルを保存する。記憶部２６１５は、ユーザデータ（ユーザ設定やユーザプログラムなど）を保存する。一部において、コンピュータシステム２６０１には、イントラネット又はインターネットを介してコンピュータシステム２６０１と通信しているリモートサーバなど、コンピュータシステム２６０１の外部にある１つ以上の追加のデータストレージユニットが含まれてもよい。

【０２５５】

[328] コンピュータシステム２６０１は、ネットワーク２６３０を介して１つ以上のリモートコンピュータシステムと通信を行うことができる。例えば、コンピュータシステム２６０１は、ユーザのリモートコンピュータシステム（サービスプロバイダーなど）と通信を行う。リモートコンピュータシステムの例には、パーソナルコンピュータ（ポータブルＰＣなど）、スレート又はタブレットＰＣ（Apple（登録商標）iPad、Samsung（登録商

標）Galaxy Tab）、電話、スマートフォン（Apple（登録商標）iPhone、Android対応デバイス、Blackberry（登録商標）など）、携帯情報端が含まれる。ユーザは、ネットワーク２６３０を介してコンピュータシステム２６０１にアクセスできる。

【０２５６】

[329] 本明細書に記載の方法は、コンピュータシステム２６０１の電子記憶場所、例えば、メモリ２６１０や電子記憶部２６１５などに保存されている機械（コンピュータプロセッサなど）実行可能コードにより実装することができる。機械実行可能コード又は機械可読コードは、ソフトウェアの形で提供される。使用中、プロセッサ２６０５でコードを実行できる。一部では、プロセッサ２６０５がすぐにアクセスできるように電子記憶部２６１５からコードを取得してメモリ２６１０に保存する。場合によっては、電子記憶部２６１５を除外し、機械実行可能命令がメモリ２６１０に保存される。

【０２５７】

[330] コードは、コードの実行に適合されたプロセッサを備えたマシンで使用するためにプリコンパイル、又は構成できる。又は、実行時にコンパイルしてもよい。コードはプログラミング言語で提供される。プログラミング言語を選択すると、コードをプリコンパイル済み、又はコンパイル済みのまま実行できる。

【０２５８】

[331] コンピュータシステム２６０１など、本明細書で提供しているシステムと方法の態様は、プログラミングで具体化することができる。それぞれの態様は、一般的には、機械（又はプロセッサ）可読媒体に保存される機械実行可能コード及び／又は関連するデータによる「製品」又は「製造品」として具現される。機械実行可能コードは、メモリ（読み取り専用メモリ、ランダムアクセスメモリ、フラッシュメモリなど）又はハードディスクなどの電子ストレージユニットに格納できる。「ストレージ」タイプの媒体には、ソフトウェアプログラミングに用いられる非持続性ストレージとして機能する、コンピュータ、プロセッサなどの有形メモリ、又はその関連モジュール、例えば各種の半導体メモリ、テープドライブ、ディスクドライブが含まれる。全体又は一部のソフトウェアは、インターネット又は他の通信ネットワークを通じて通信を行う。通信により、例えば、あるコンピュータ又はプロセッサから別のコンピュータ、例えば管理サーバやホストコンピュータからアプリケーションサーバのコンピュータプラットフォームへのソフトウェアをロードすることができる。従って、ソフトウェア要素を保持できる別の媒体には、ローカルデバイスの物理インターフェースの間を飛び交い、有線及び光学固定電話ネットワーク、様々なエアリンク経由で使用される、光学、電気、電磁波が含まれる。有線又は無線リンク、光リンクなど、そのような波を運ぶ物理要素も、ソフトウェアを搭載した媒体として見なされる。本明細書で使用されるように、非持続的な有形「記憶」媒体に限定されない限り、コンピュータ又は機械の「読み取り可能な媒体」などの用語は、実行のためにプロセッサに命令を提供することに関与する任意の媒体を指す。

【０２５９】

[332] したがって、コンピュータ実行可能コードなどの機械可読媒体は、有形の記憶媒体、搬送波媒体又は物理的伝送媒体を含むが、これに限らない。不揮発性記憶媒体には、例えば、図面に示されているデータベースなどを実装するために使用される、任意のコンピュータなどの記憶装置などの光ディスクや磁気ディスクが含まれる。揮発性記憶媒体には、このようなコンピュータプラットフォームのメインメモリなどの動的メモリも含まれる。有形伝送媒体には、同軸ケーブル、例えば、コンピュータシステム内のバスを構成するワイヤを含む、銅線と光ファイバーが含まれる。搬送波伝送媒体は、電気信号又は電磁信号、もしくは無線周波数（ＲＦ）及び赤外線（ＩＲ）データ通信中に生成される音響波や光波の形をとることができる。従って、一般的な形式のコンピュータ可読可能媒体には、例えば、フロッピーディスク、フレキシブルディスク、ハードディスク、磁気テープ、その他の磁気メディア、ＣＤ－ＲＯＭ、ＤＶＤ又はＤＶＤ－ＲＯＭ、その他の光学メディア、パンチカード用紙、テープ、穴のパターンを持つ他の物理記憶媒体、ＲＡＭ、ＲＯＭ、ＰＲＯＭ、ＥＰＲＯＭ、ＦＬＡＳＨ－ＥＰＲＯＭ、その他のメモリチップ又はカートリ

ッジ、データや命令を搬送する搬送波、搬送波を搬送するケーブル又はリンク、コンピュータがプログラミングコードやデータを読み取ることができるその他の媒体が含まれる。このようなコンピュータ可読媒体の多くは、実行のためにプロセッサに複数の命令の複数のシーケンスを運ぶことに関する。

【０２６０】

[333] コンピュータシステム２６０１は、例えば、分類器や訓練アルゴリズムの出力又は読み出しを提供するユーザインターフェース（ＵＩ）２６４０を含む電子ディスプレイ２６３５を含むか、それと通信することができる。ＵＩの例としては、グラフィカルユーザインターフェース（ＧＵＩ）とＷｅｂベースのユーザインターフェースが挙げられるが、これに限らない。

【０２６１】

[334] 本開示の方法及びシステムは、１つ以上のアルゴリズムにより実装できる。アルゴリズムは、中央処理装置２６０５による実行されるフトウェアにより実装できる。アルゴリズムは、例えば、（ｉ）バイオマーカーの参照セットと比較して、１つ以上のバイオマーカーの試料内の存在を確定することができる。

【０２６２】

参考文献

[335] Flaherty KR, King TE, Jr., Raghu G, Lynch JP, 3rd, Colby TV, Travis WD, Gross BH, Kazerooni EA, Toews GB, Long Q, et al: Idiopathic interstitial pneumonia: what is the effect of a multidisciplinary approach to diagnosis? Am J Respir Crit Care Med 2004, 170:904-910.

[336] Travis WD, Costabel U, Hansell DM, King TE, Jr., Lynch DA, Nicholson AG, Ryerson CJ, Ryu JH, Selman M, Wells AU, et al: An official American Thoracic Society/European Respiratory Society statement: Update of the international multidisciplinary classification of the idiopathic interstitial pneumonias. Am J Respir Crit Care Med 2013, 188:733-748.

[337] Flaherty KR, Andrei AC, King TE, Jr., Raghu G, Colby TV, Wells A, Bassily N, Brown K, du Bois R, Flint A, et al: Idiopathic interstitial pneumonia: do community and academic physicians agree on diagnosis? Am J Respir Crit Care Med 2007, 175:1054-1060.

[338] Tuch BB, Laborde RR, Xu X, Gu J, Chung CB, Monighetti CK, Stanley SJ, Olsen KD, Kasperbauer JL, Moore EJ, et al: Tumor transcriptome sequencing reveals allelic expression imbalances associated with copy number alterations. PLoS One 2010, 5:e9317.

[339] Twine NA, Janitz K, Wilkins MR, Janitz M: Whole transcriptome sequencing reveals gene expression and splicing differences in brain regions affected by Alzheimer's disease. PLoS One 2011, 6:e16266.

[340] Boyle EA, Li YI, Pritchard JK: An Expanded View of Complex Traits: From Polygenic to Omnigenic. Cell 2017, 169:1177-1186.

[341] Pankratz DG, Choi Y, Imtiaz U, Fedorowicz GM, Anderson JD, Colby TV, Myers JL, Lynch DA, Brown KK, Flaherty KR, et al: Usual Interstitial Pneumonia Can Be Detected in Transbronchial Biopsies Using Machine Learning. Ann Am Thorac Soc 2017.

[342] Sorlie T, Tibshirani R, Parker J, Hastie T, Marron JS, Nobel A, Deng S, Johnsen H, Pesich R, Geisler S, et al: Repeated observation of breast tumor subtypes in independent gene expression data sets. Proc Natl Acad Sci U S A 2003, 100:8418-8423.

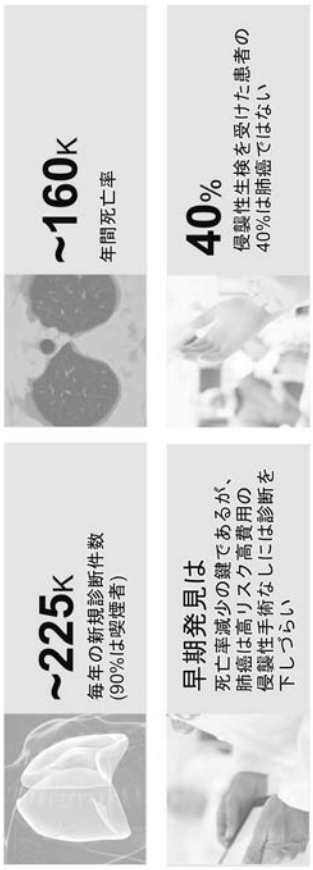
[343] Brennan CW, Verhaak RG, McKenna A, Campos B, Noushmehr H, Salama SR, Zheng S, Chakravarty D, Sanborn JZ, Berman SH, et al: The somatic genomic landscape of glioblastoma. Cell 2013, 155:462-477.

[344] Kim SY, Diggans J, Pankratz D, Huang J, Pagan M, Sindy N, Tom E, Anderson J, Choi Y, Lynch DA, et al: Classification of usual interstitial pneumonia in patients with interstitial lung disease: assessment of a machine learning approach using high-dimensional transcriptional data. Lancet Respir Med 2015, 3:473-482.
[345] Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR: STAR: ultrafast universal RNA-seq aligner. Bioinformatics 2013, 29:15-21.
[346] Anders S, Pyl PT, Huber W: HTSeq--a Python framework to work with high-throughput sequencing data. Bioinformatics 2015, 31:166-169.
[347] DeLuca DS, Levin JZ, Sivachenko A, Fennell T, Nazaire MD, Williams C, Reich M, Winckler W, Getz G: RNA-SeQC: RNA-seq metrics for quality control and process optimization. Bioinformatics 2012, 28:1530-1532.
[348] Love MI, Huber W, Anders S: Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 2014, 15:550.
[349] Anders S, McCarthy DJ, Chen Y, Okoniewski M, Smyth GK, Huber W, Robinson MD: Count-based differential expression analysis of RNA sequencing data using R and Bioconductor. Nat Protoc 2013, 8:1765-1786.
[350] Dobson AJ, Barnett A: An introduction to generalized linear models. CRC press; 2008.
[351] Krstajic D, Buturovic LJ, Leahy DE, Thomas S: Cross-validation pitfalls when selecting and assessing regression and classification models. J Cheminform 2014, 6:10.
[352] Friedman J, Hastie T, Tibshirani R: The elements of statistical learning. Springer series in statistics New York; 2001.
[353] LeCun Y, Bengio Y, Hinton G: Deep learning. Nature 2015, 521:436-444.
[354] Gu B, Hu F, Liu H: Modelling classification performance for large data sets. Advances in Web-Age Information Management 2001:317-328.
[355] Sun C, Shrivastava A, Singh S, Gupta A: Revisiting Unreasonable Effectiveness of Data in Deep Learning Era. arXiv preprint arXiv:170702968 2017.
[356] Libbrecht MW, Noble WS: Machine learning applications in genetics and genomics. Nat Rev Genet 2015, 16:321-332.
[357] Wong SC, Gatt A, Stamatescu V, McDonnell MD: Understanding data augmentation for classification: when to warp? In. IEEE; 2016: 1-6; arXiv:1609.08764.

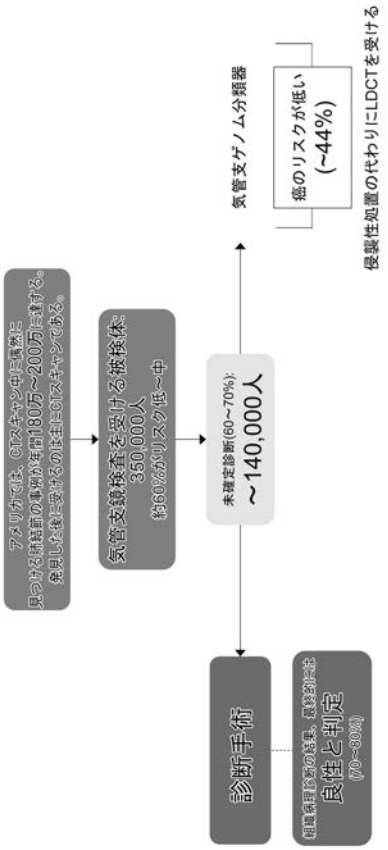
【０２６３】

[358] 本発明の好ましい実施形態が本明細書に示され、説明されているが、実施形態は例としてのみ提供されることを当業者は理解するであろう。本発明は、本明細書内で提供される特定の例により限定されることを意図するものではない。上記の明細書を参照して本発明を説明したが、本明細書の実施形態の説明や図面は、限定的な意味で解釈されることを意図していない。本発明から逸脱しない限り、多数の変形、変更、置換が可能である。更に、本発明の全ての態様は、様々な条件や変数に依存する本明細書に記載の特定の描写、構成、又は相対的割合に限らない。本明細書に記載された本発明の実施形態につき、本発明を実施する際に様々に取り替えることができる。従って、本発明は、代替、修正、変形、又は均等物を全て網羅する。以下の特許請求の範囲が本発明の範囲を定義し、請求項に記載の方法及び構造、それらの同等物が全て含まれる。

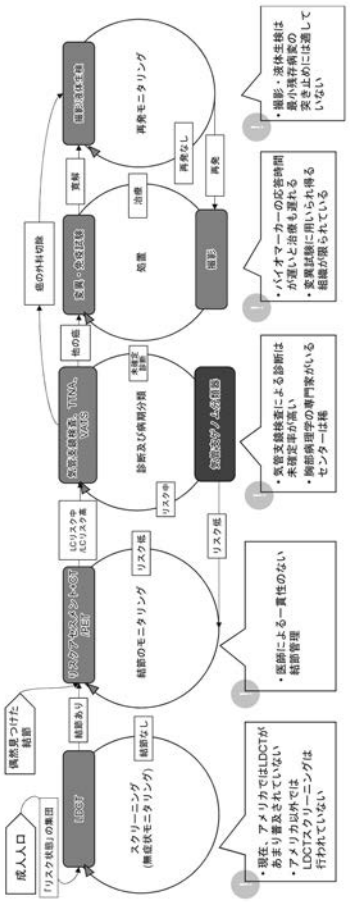
【図１】


~225K
毎年の新規診断件数
(90%は喫煙者)
~160K
年間死亡率
早期発見は
死亡率減少の鍵であるが、
肺癌は高リスク高費用の
侵襲性手術なしには診断を
下しづらい
40%
侵襲性生検を受けた患者の
40%は肺癌ではない

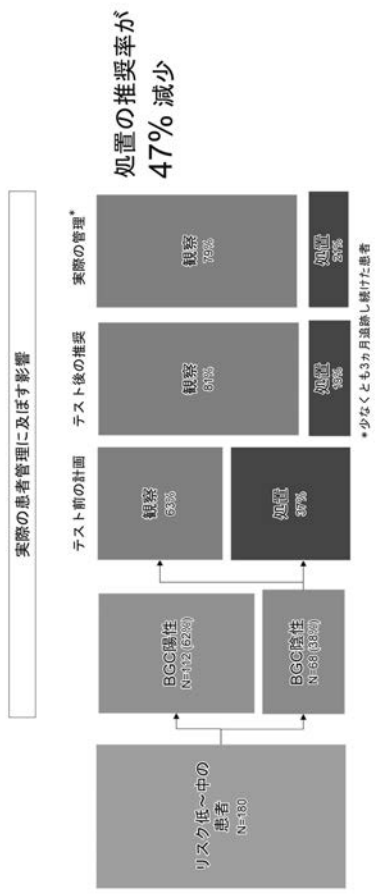
【図２】


アメリカでは、CTスキャン中に偶然に
見つかる結節の事例が年間100万~200万に達する。
発見した後に受けるのは主にCTスキャンである。
気管支鏡検査を受ける被検体:
350,000人
(約60%がリスク低~中
未確定診断(60~70%):
~140,000人
気管支ゲノム分類器
癌のリスクが低い
(~44%)
侵襲性処置の代わりにLDCTを受ける
診断手術
組織病理学の結果、確定診断は
良性と判定
(70~80%)

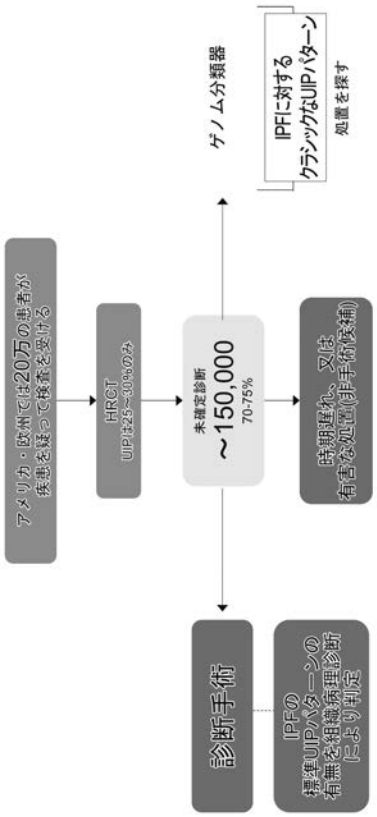
【図３】


成人人口
「リスク状態」の集団
偶然見つけた結節
結節なし
結節あり
スクリーニング
(無症状モニタリング)
結節のモニタリング
診断及び病期分類
処置
再発モニタリング
・現在、アメリカではLDCTが
あまり普及されていない
・アメリカ以外では
LDCTスクリーニングは
行われていない
・医師による一貫性のない
結節管理
・気管支鏡検査による診断は
未確定率が高い
・肺部病理学の専門家がいる
センターは稀
・バイオマーカーの応答時間
が遅いと治療も遅れる
・臨床試験に用いられ得る
組織が限られている
・撮影・侵襲生検は
最小臨存病変の
突き止めには適して
いない

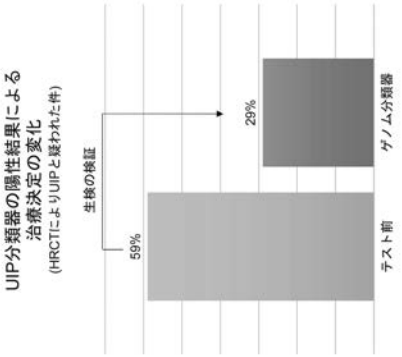
【図４】


処置の推奨率が
47% 減少
実際の患者管理に及ぼす影響
テスト前の計画
テスト後の推奨
実際の管理*
リスク低~中の
患者
N=180
BGC陽性
N=112 (62%)
BGC陰性
N=68 (38%)
観察
63%
処置
37%
観察
61%
処置
39%
観察
76%
処置
24%
*少なくとも3カ月追跡し続けた患者

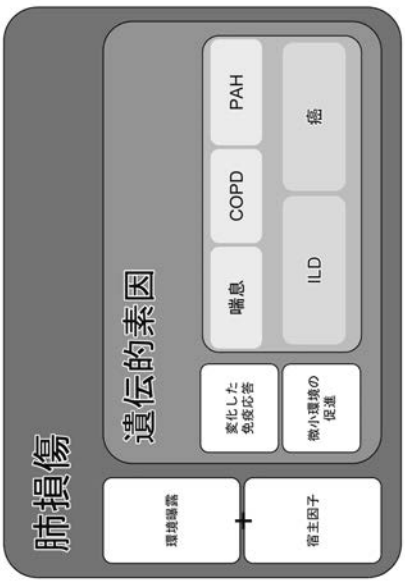
【図５】


アメリカ・欧州では20万の患者が
疾患を疑って検査を受ける
HRCT
UIPは25～30%のみ
未確定診断
～150,000
70-75%
ゲノム分類器
IPFに対する
クラシックなUIPパターン
処置を探す
時期遅れ、又は
有害な処置(非手術候補)
診断手術
IPFの
標準UIPパターンの
有無を組織病理診断
により判定

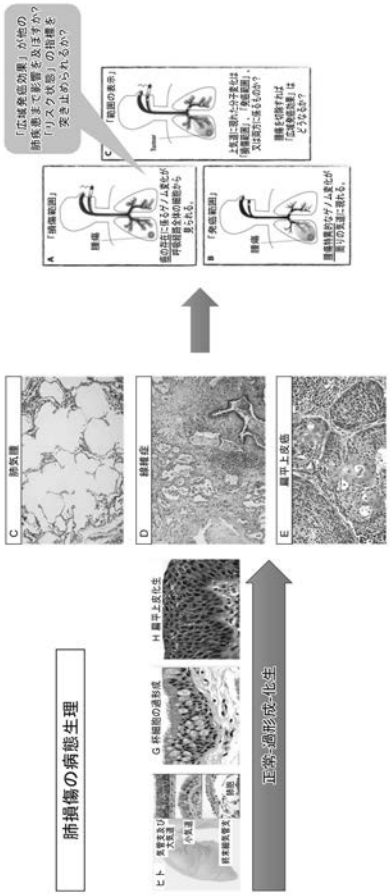
【図６】


UIP分類器の陽性結果による
治療決定の変化
(HRCTによりUIPと疑われた件)
生検の検証
59%
29%
テスト前
ゲノム分類器

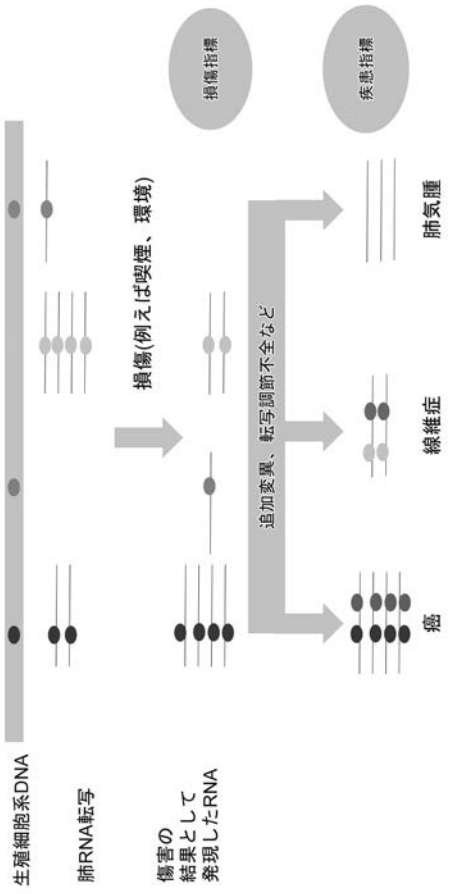
【図７】


肺損傷
遺伝的素因
環境曝露
宿主因子
変化した
免疫応答
微小環境の
促進
喘息
COPD
PAH
ILD
癌

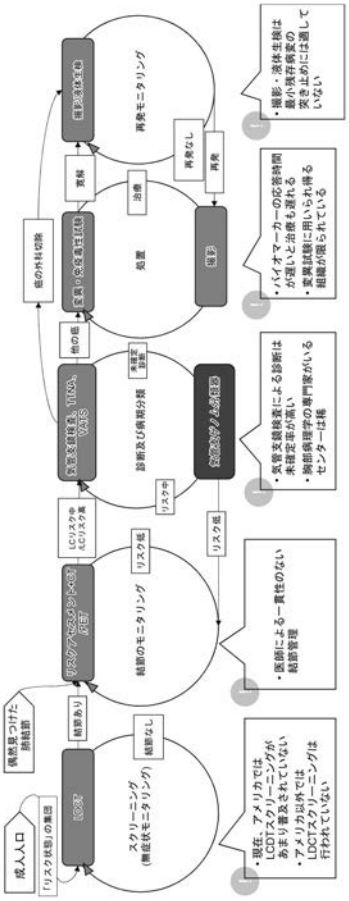
【図８】


肺損傷の病態生理
正常-過形成-化生

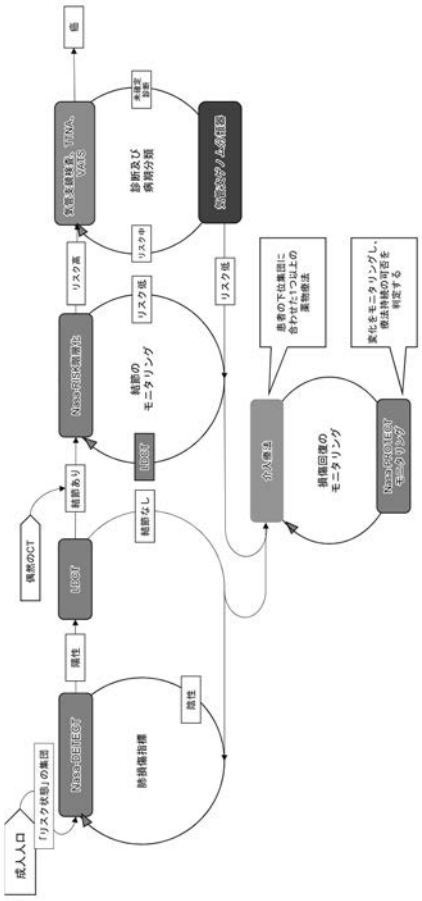
【図９】



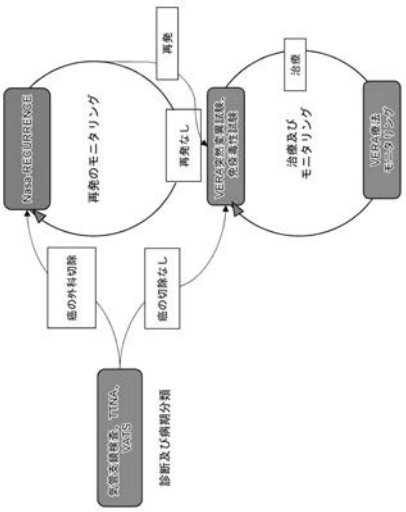
【図１０】



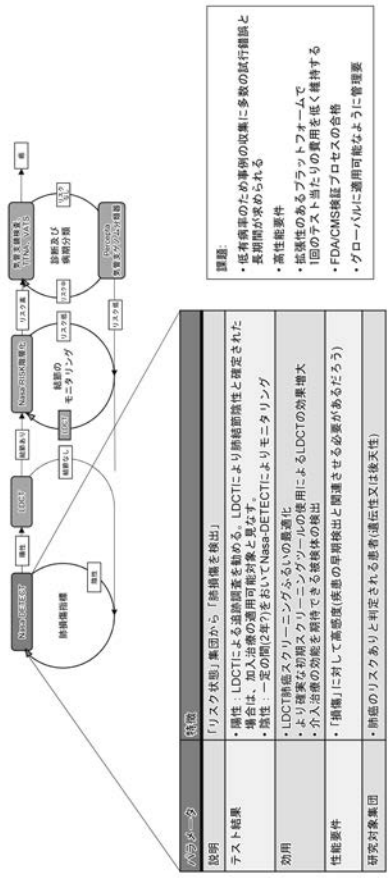
【図１１Ａ】



【図１１Ｂ】

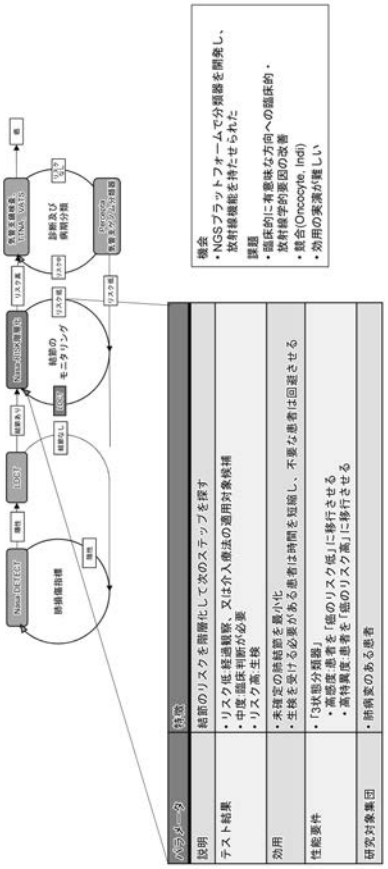


【図１２】



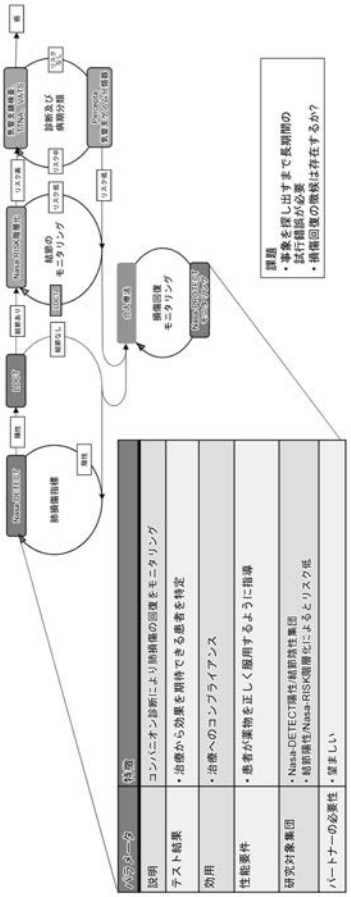
パラメータ	特徴
説明	「リスク状態」集団から「肺病変を検出」
テスト結果	・陽性：LDCTによる追跡調査を始める。LDCTにより肺結節陰性と確定された場合は、加入治療の適用可能対象と見なす。 ・陰性：一定の間(2年)をおいてNasa-DETECTによりモニタリング
効用	・LDCT肺癌スクリーニングふるいの最適化 ・より効果な初期スクリーニングツールの使用によるLDCTの効果増大 ・介入治療の効能を維持できる被検体の検出
性能要件	・「損傷」に対して高感度(疾患の早期検出と関連させる必要があるだろう)
研究対象集団	・肺癌のリスクありと判定される患者(遺伝性又は後天性)

【図１３】



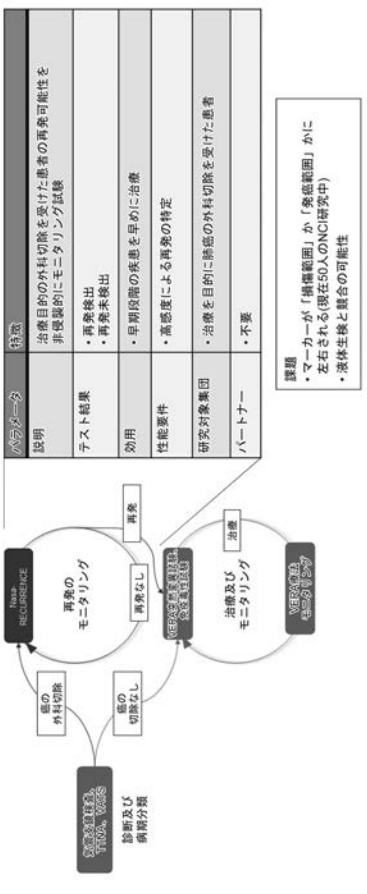
パラメータ	特徴
説明	結節のリスクを階層化して次のステップを探す
テスト結果	・リスク低:経過観察、又は介入療法の適用対象候補 ・中度:臨床判断が必要 ・リスク高:生検
効用	・未特定の肺結節を最小化 ・生検を受ける必要がある患者は時間を短縮し、不要な患者は回避させる
性能要件	・「3状態分類器」 ・「高感度:患者を「癌のリスク低」に移行させる ・「高特異度:患者を「癌のリスク高」に移行させる
研究対象集団	・肺病変のある患者

【図１４】



パラメータ	特徴
説明	コンパニオン診断により肺病変の回復をモニタリング
テスト結果	・治療から効果を期待できる患者を特定
効用	・治療へのコンプライアンス
性能要件	・患者が薬物を正しく服用するように指導
研究対象集団	・Nasa-DETECT陽性/結節陽性集団 ・結節陽性/Nasa-RISK階層化によるとリスク低
パートナーの必要性	・望ましい

【図１５】

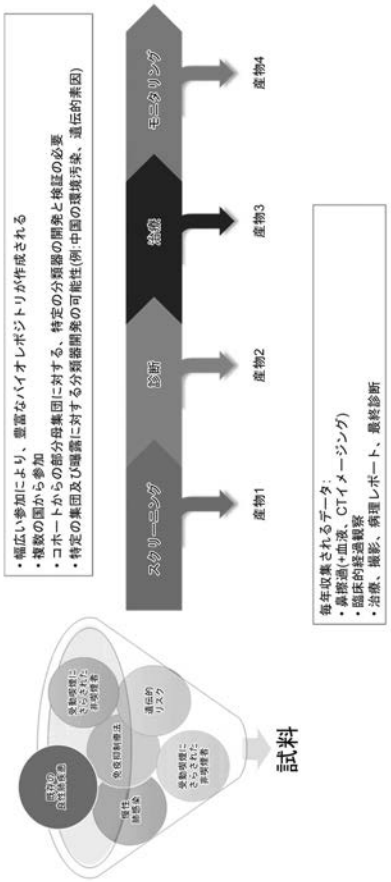


パラメータ	特徴
説明	治療目的の外科切除を受けた患者の再発可能性を非侵襲的にモニタリング試験
テスト結果	・再発検出 ・再発未検出
効用	・早期段階の疾患を早めに治療
性能要件	・高感度による再発の特定
研究対象集団	・治療を目的に肺癌の外科切除を受けた患者
パートナー	・不要

【図１６】



【図１７】

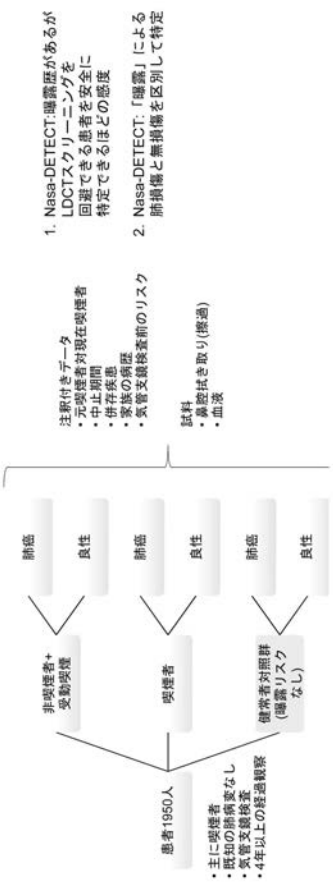


【図１８】

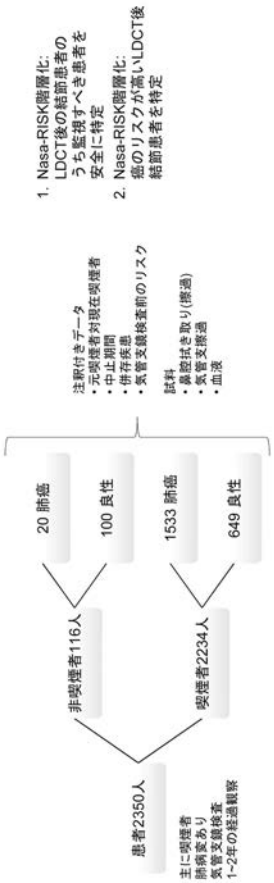
略語	予測される参加者数(陽性率)	現在のアプローチ、その問題	陽性率(陽性率)	参加者数	試験デザイン	肺癌	気管支鏡検査	血液	CT	追跡期間(月)
AEGIS	2000	気管支鏡検査中	75%	・28ヵ所 (ほぼ研究機関)	・気管支鏡検査による肺癌診断 ・>18歳 ・喫煙経験あり(現在まで>100本)	✓	✓	✓		12
DECAMP1	500 (350)	気管支鏡検査中	30-35%	・1ヵ所目論及び single ・研究機関3ヵ所	・気管支鏡検査による肺癌診断 肺結節7～30mm ・現・元喫煙者(>20箱/年) ・~45歳	✓	✓	✓	✓	24
LTP2 (NCI/EPRN)	200 (100)	CT後	20-25%	・4ヵ所 (BU UCLA Vanderbilt NYU)	・IPNの診断 ・肺結節>4-5mm ・現・元喫煙者(>30箱/年) ・>50歳	✓		✓	✓	24
DECAMP2	800 (450)	スクリーニング中	1.5% 年間発生率	・1ヵ所目論及び single ・研究機関3ヵ所	・50～79歳 ・現・元喫煙者(20箱/年) ・COPD、非喫煙血縁者のうちに肺癌病歴あり	✓ (0、12、24ヵ月)	✓ (0、24ヵ月)	✓ (0、12、24ヵ月)	✓ (年間)	48+
Lahey	3000 (1500)	スクリーニング中	1% 年間発生率	・Lahey+その他	・Laheyのスクリーニングプログラムに登録 ・スクリーニングNCDに基づいて	✓			✓	48+

■ 現在、試料に対するVCYTアクセスあり　■ 現在、試料に対するVCYTアクセスなし

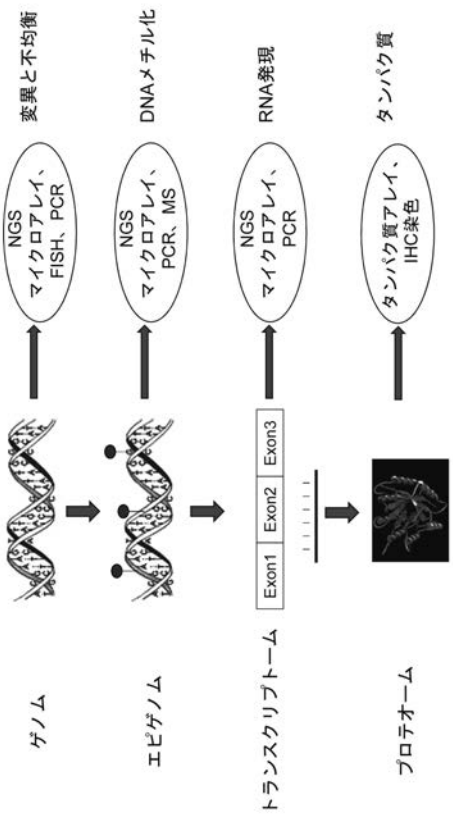
【図１９】



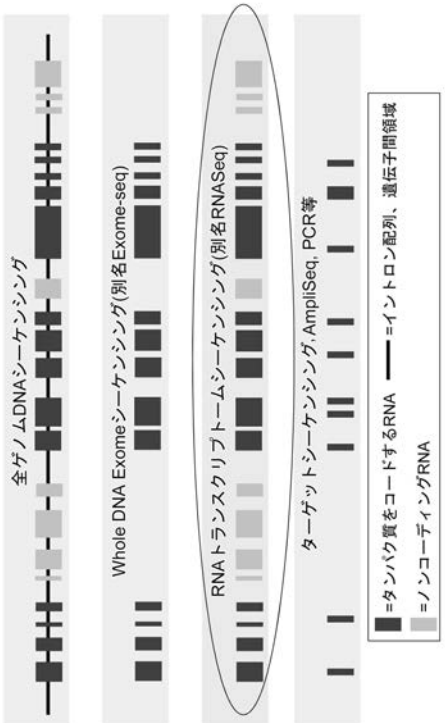
【図２０】



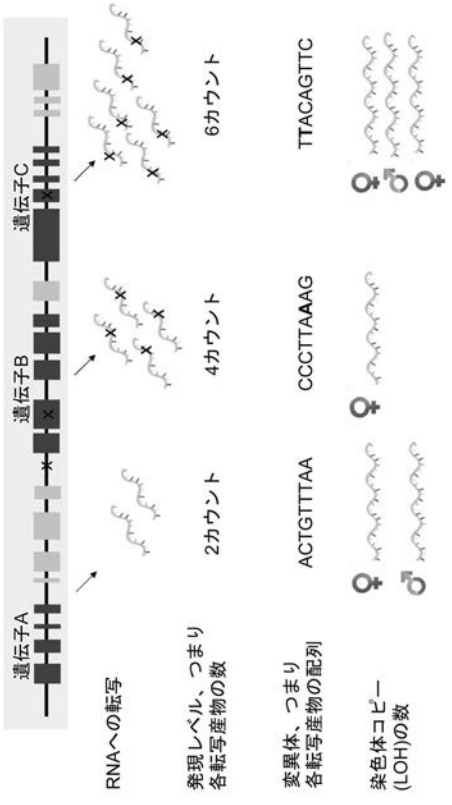
【図２１】



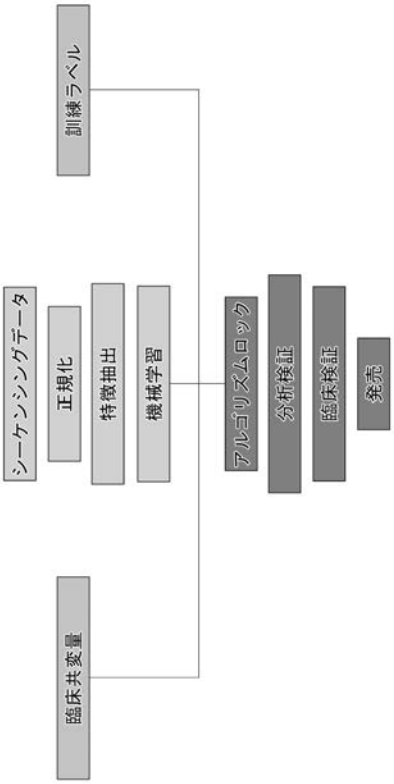
【図２２】



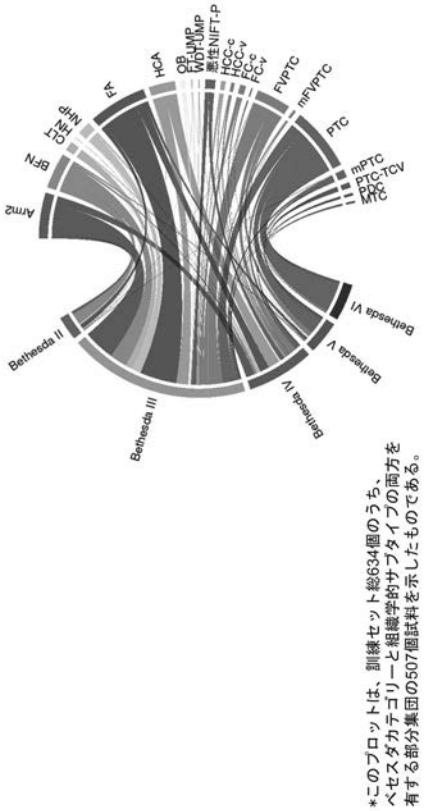
【図２３】



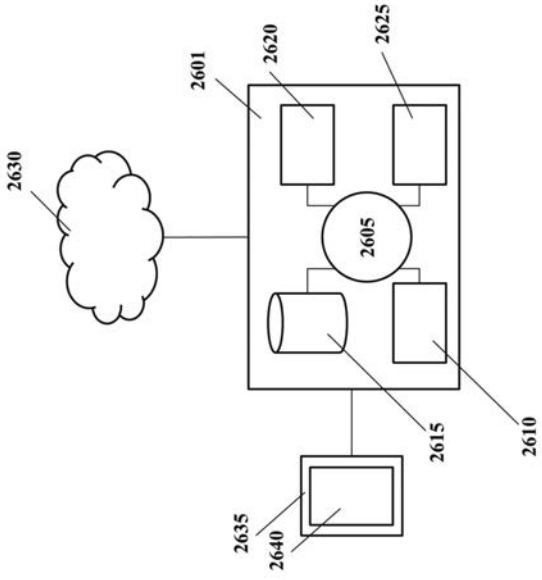
【図２４】



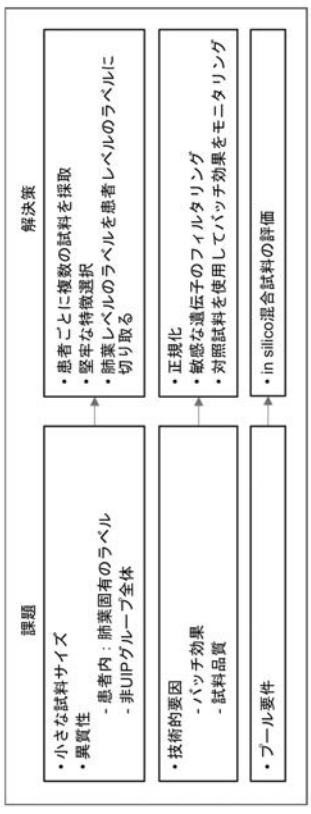
【図２５】



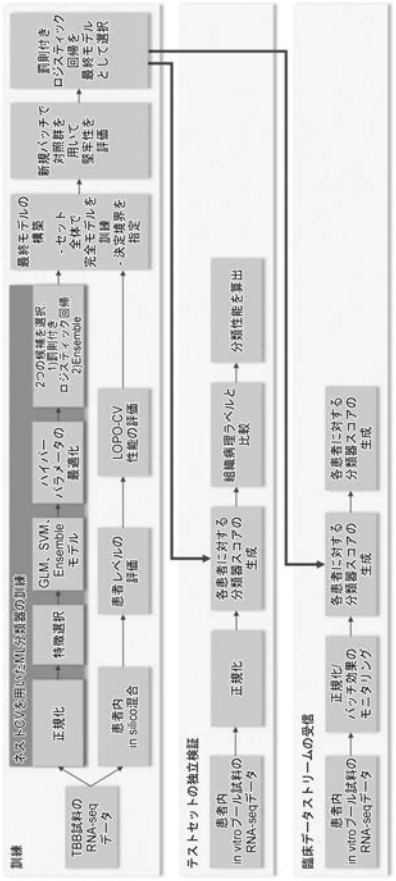
【図２６】



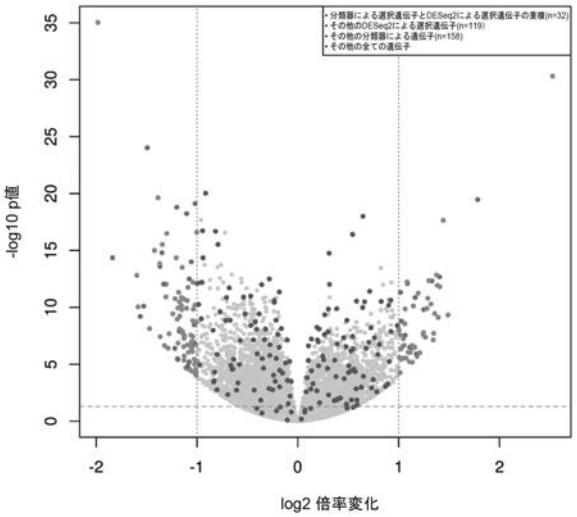
【図２７】



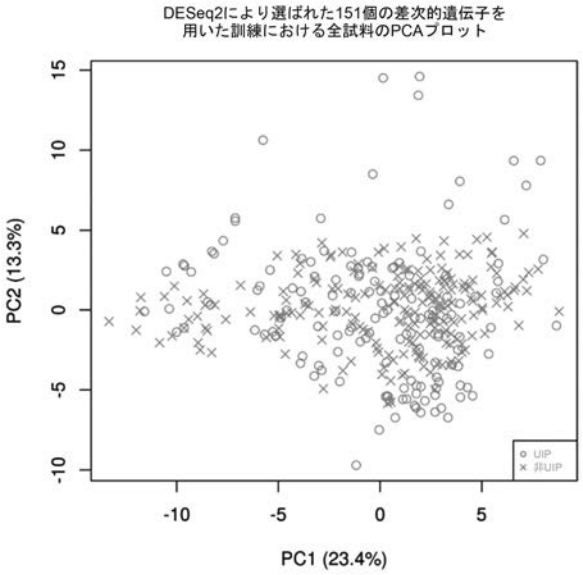
【図２８】



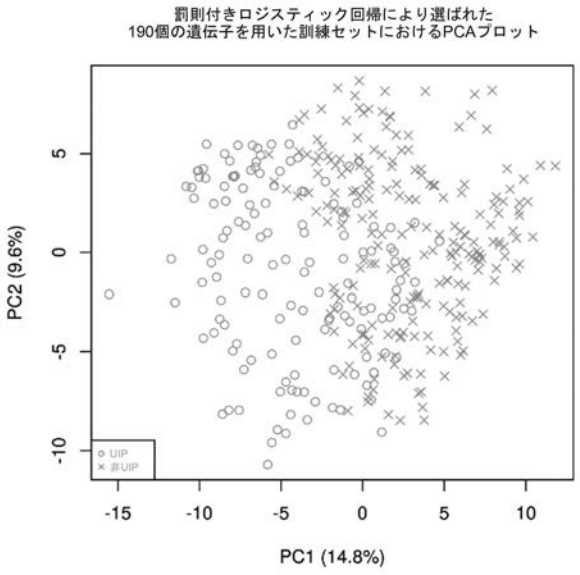
【図２９】



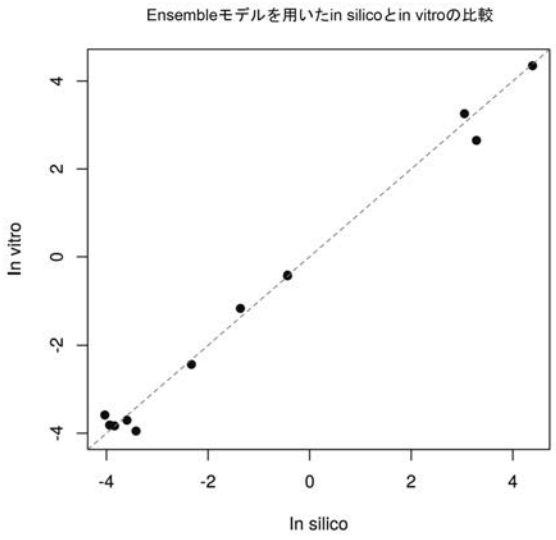
【図３０】



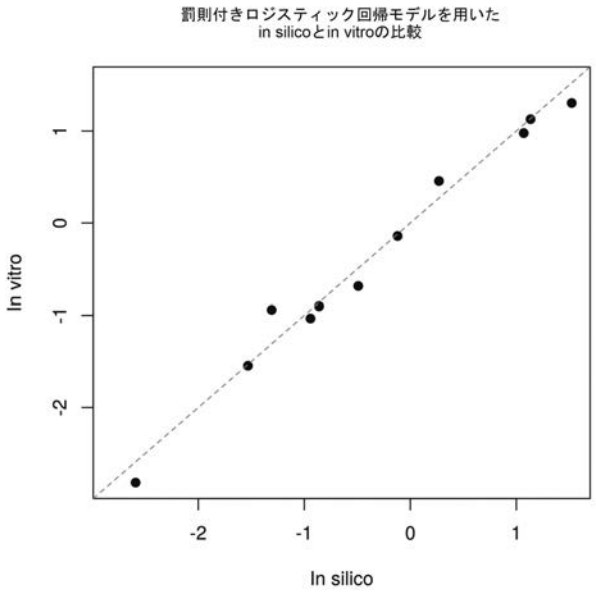
【図３１】



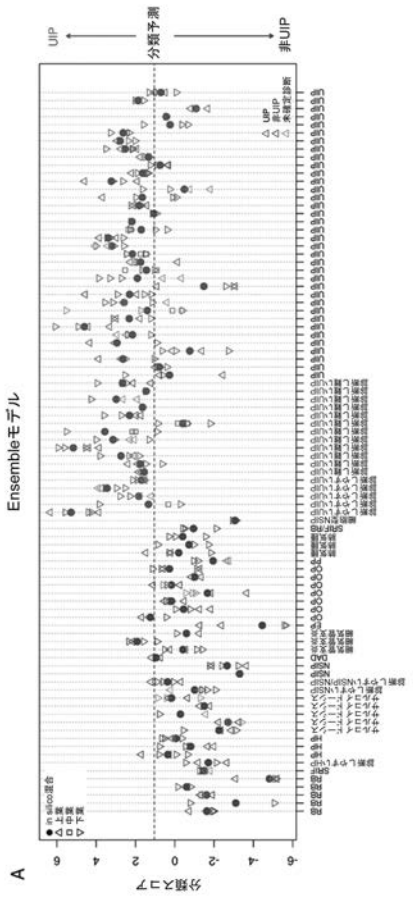
【図３２】



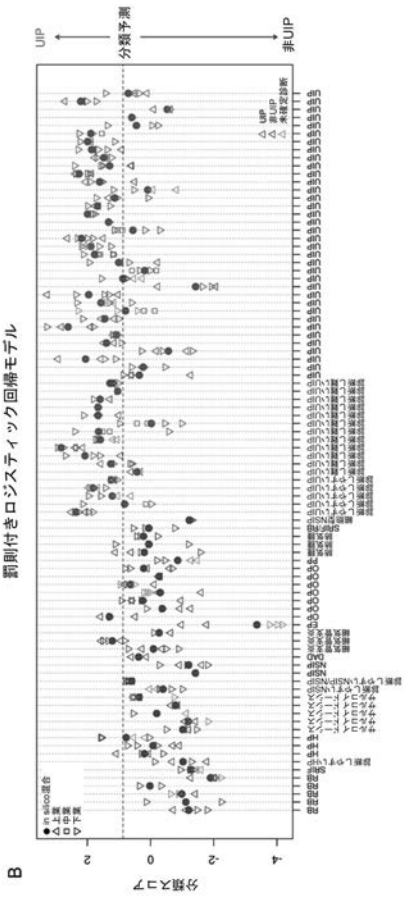
【図３３】



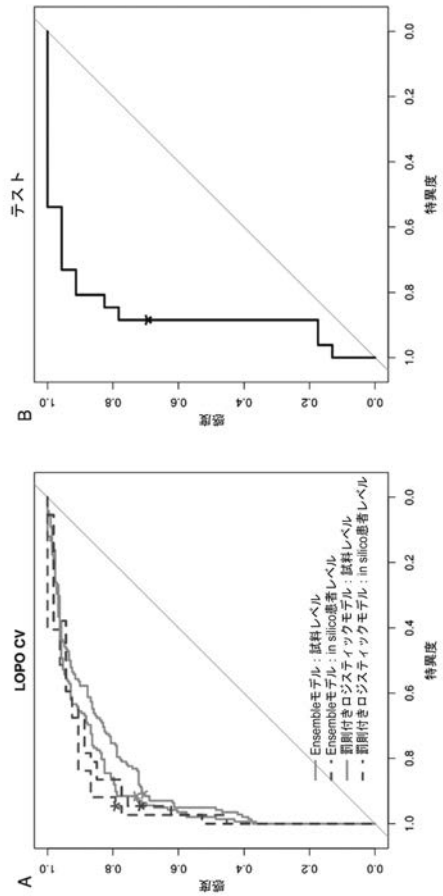
【図３４】



【図３５】



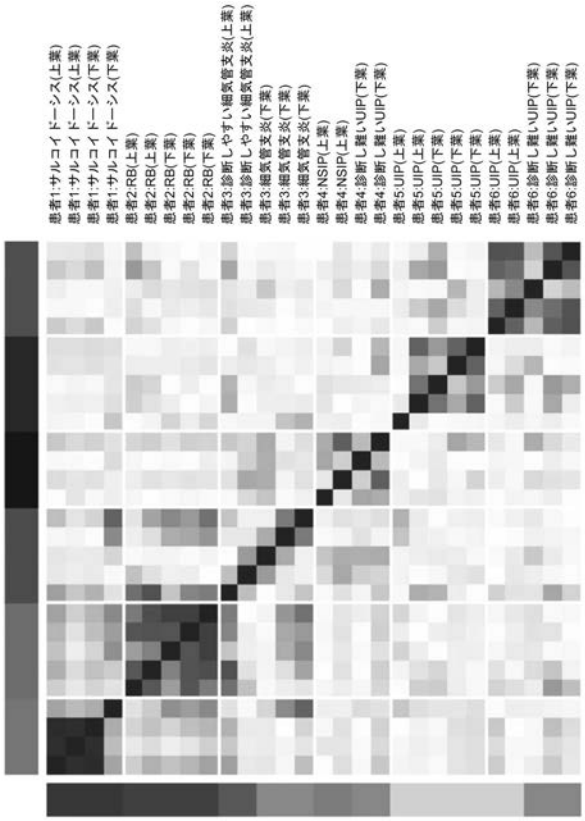
【図３６】



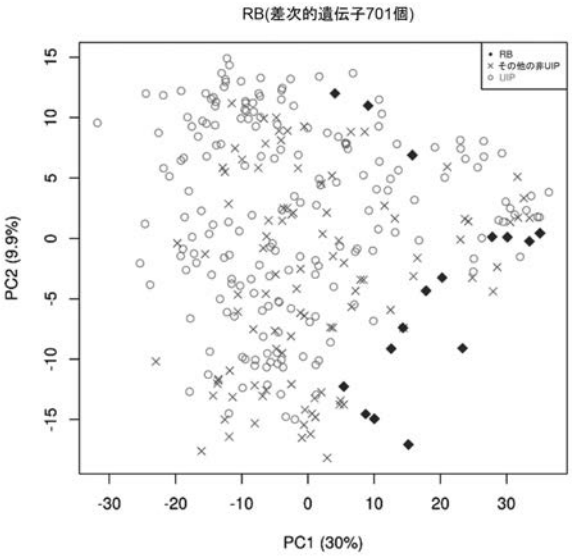
【図３７】

C		LOPO CV				独立テスト	
		Ensembleモデル		罰則付きロジスティックモデル		罰則付きロジスティックモデル	
		試料レベル	in silico混合	試料レベル	in silico混合	in vitro混合	
	AUC	0.90 [0.87-0.93]	0.93 [0.88-0.98]	0.87 [0.83-0.91]	0.91 [0.85-0.97]	0.87 [0.76-0.98]	
	特異度	0.92 [0.86-0.96]	0.95 [0.82-0.99]	0.91 [0.85-0.95]	0.95 [0.82-0.99]	0.88 [0.70-0.98]	
	感度	0.73 [0.67-0.79]	0.79 [0.66-0.89]	0.71 [0.64-0.77]	0.72 [0.58-0.83]	0.70 [0.47-0.87]	

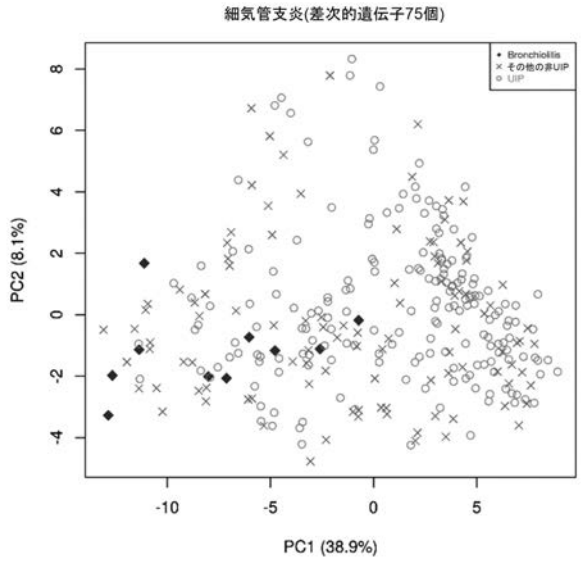
【図３８】



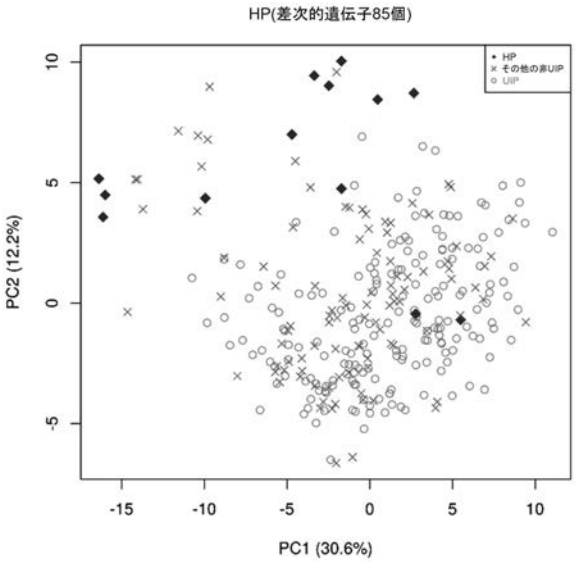
【図３９】



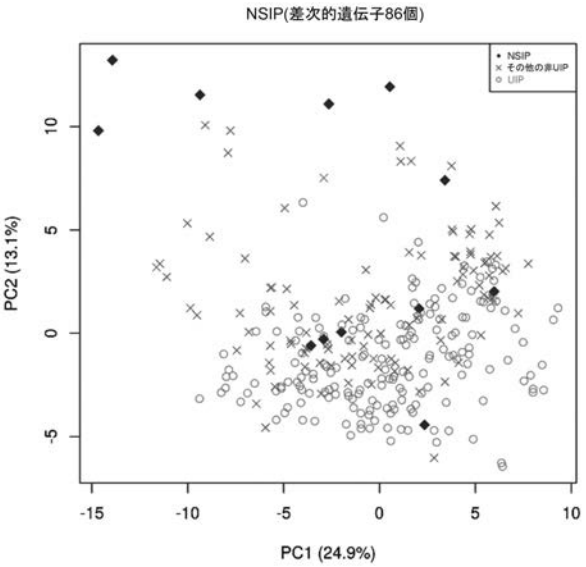
【図４０】



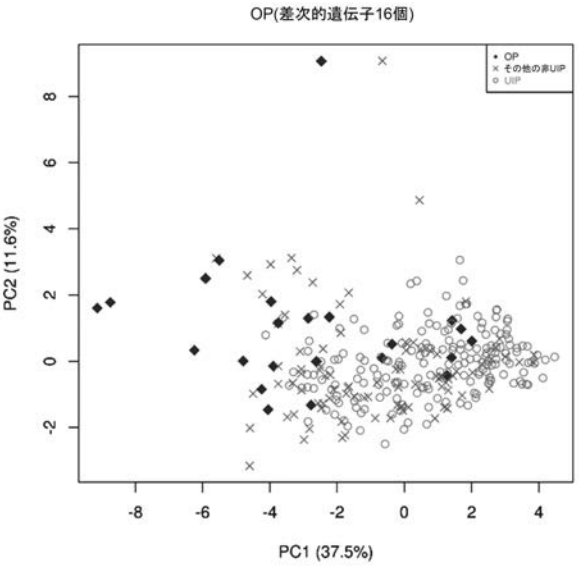
【図４１】



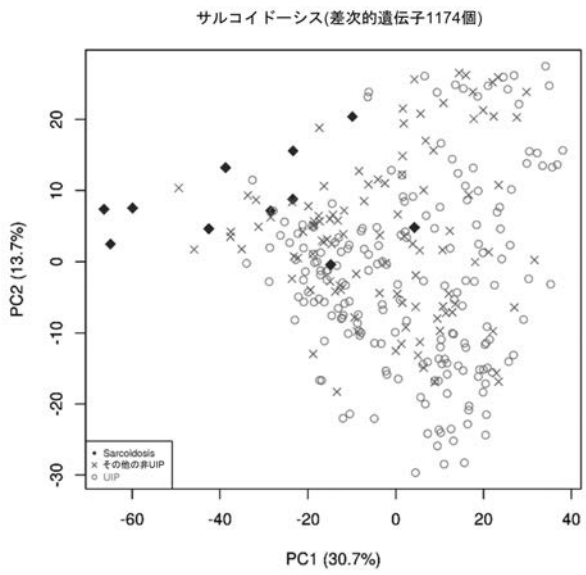
【図４２】



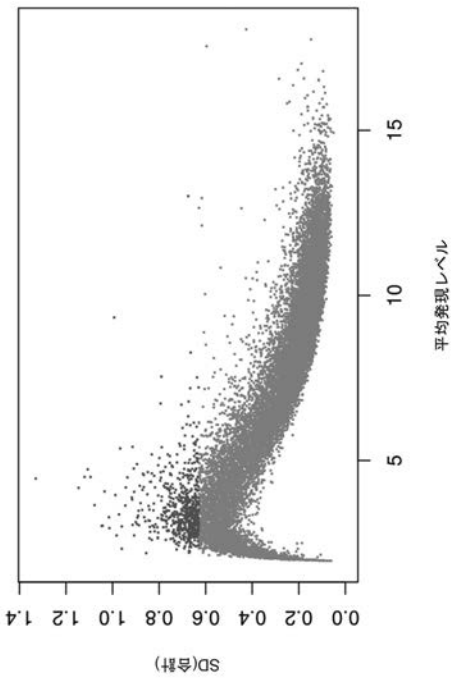
【図４３】



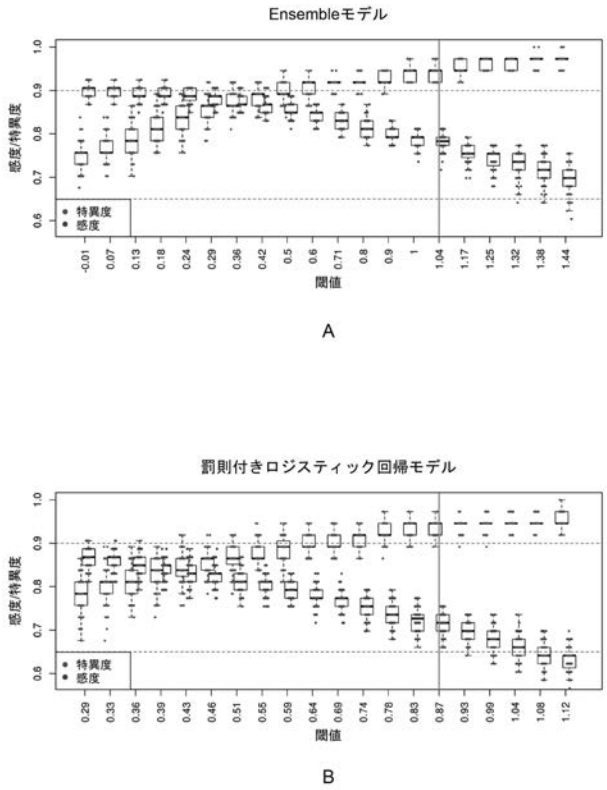
【図４４】



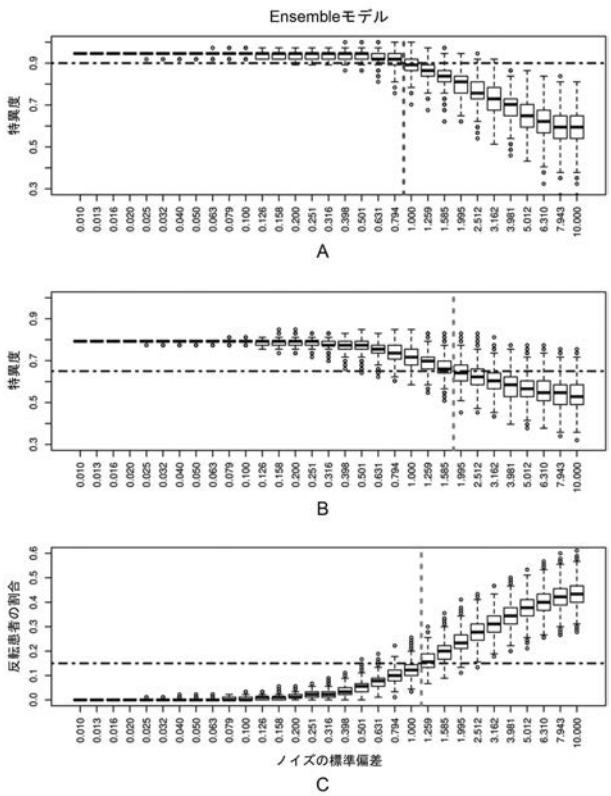
【図４５】



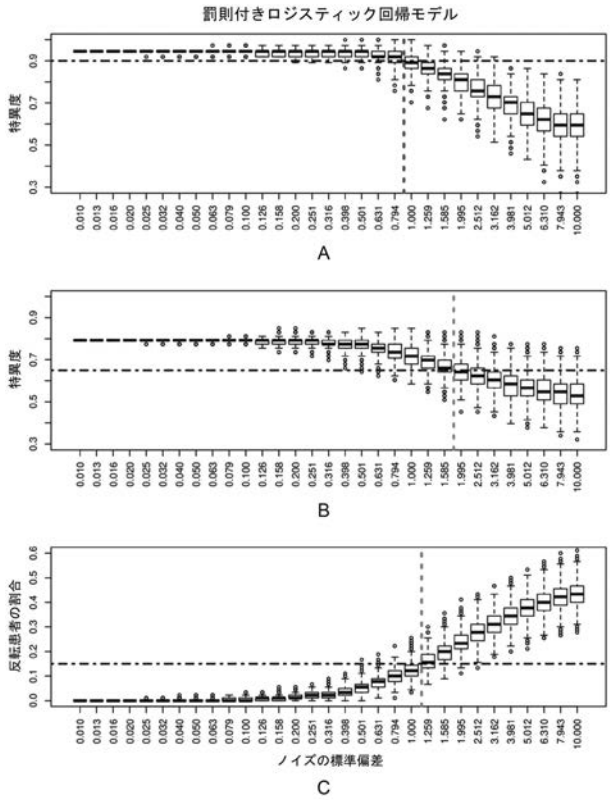
【図４６】



【図４７】



【図４８】



【国際調査報告】

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US2018/035702

A. CLASSIFICATION OF SUBJECT MATTER

A61B 5/00(2006.01)i, A61B 5/055(2006.01)i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
A61B 5/00; A61N 5/10; A61B 5/08; G06F 19/00; G06F 19/18; A61M 11/00; C12Q 1/68; A61B 5/055

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched
Korean utility models and applications for utility models
Japanese utility models and applications for utility models

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
eKOMPASS(KIPO internal) & Keywords: lung, biomarker, sample, genetic, processor

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 2015-0337385 A1 (THE UNITED STATES OF AMERICA, AS REPRESENTED BY THE SECRETARY, DEPT. OF HEALTH & HUMAN SERVICES) 26 November 2015 See paragraphs [10]-[178] and figures 1-19.	1-60
A	US 2017-0127976 A1 (MICHAEL PHILLIPS) 11 May 2017 See paragraphs [39]-[66] and figures 1-8A.	1-60
A	US 2015-0088430 A1 (ALLEGRO DIAGNOSTICS CORP.) 26 March 2015 See paragraphs [4]-[30] and figures 1-4.	1-60
A	KR 10-2016-0079086 A (MEDIAL RESEARCH LTD.) 05 July 2016 See claims 1-10 and figures 1-2.	1-60
A	JP 2015-519966 A (KONINKLIJKE PHILIPS N.V.) 16 July 2015 See claims 1-10 and figures 1-3.	1-60

☐ Further documents are listed in the continuation of Box C.

☒ See patent family annex.

* Special categories of cited documents:
"A" document defining the general state of the art which is not considered to be of particular relevance
"E" earlier application or patent but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed
"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents,such combination being obvious to a person skilled in the art
"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
11 September 2018 (11.09.2018)	**12 September 2018 (12.09.2018)**

Name and mailing address of the ISA/KR	Authorized officer
International Application Division Korean Intellectual Property Office 189 Cheongsa-ro, Seo-gu, Daejeon, 35208, Republic of Korea	Kim, Yeonkyung
Facsimile No. +82-42-481-8578	Telephone No. +82-42-481-3325

Form PCT/ISA/210 (second sheet) (January 2015)

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/US2018/035702

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 2015-0337385 A1	26/11/2015	EP 2885425 A1	24/06/2015
		EP 2885425 B1	26/07/2017
		JP 2015-533477 A	26/11/2015
		WO 2014-031609 A1	27/02/2014
US 2017-0127976 A1	11/05/2017	US 2016-0363581 A1	15/12/2016
US 2015-0088430 A1	26/03/2015	EP 2841603 A2	04/03/2015
		WO 2013-163568 A2	31/10/2013
		WO 2013-163568 A3	13/02/2014
KR 10-2016-0079086 A	05/07/2016	AU 2014-347669 A1	23/06/2016
		AU 2014-347669 A1	14/05/2015
		CA 2928271 A1	14/05/2015
		CN 105814571 A	27/07/2016
		EP 3065630 A1	14/09/2016
		IL 245515 A	30/06/2016
		IL 245515 D0	30/06/2016
		JP 2016-535262 A	10/11/2016
		MX 2016005825 A	09/12/2016
		RU 2016121682 A	11/12/2017
		RU 2016121682 A3	22/06/2018
		US 2016-0292379 A1	06/10/2016
		WO 2015-068157 A1	14/05/2015
JP 2015-519966 A	16/07/2015	BR 112014030569 A2	27/06/2017
		CN 104470427 A	25/03/2015
		CN 104470427 B	21/11/2017
		EP 2858560 A1	15/04/2015
		JP 6305994 B2	04/04/2018
		RU 2014153881 A	10/08/2016
		RU 2641975 C2	23/01/2018
		US 2015-0099994 A1	09/04/2015
		WO 2013-182951 A1	12/12/2013

フロントページの続き

(81)指定国・地域　AP(BW,GH,GM,KE,LR,LS,MW,MZ,NA,RW,SD,SL,ST,SZ,TZ,UG,ZM,ZW),EA(AM,AZ,BY,KG,KZ,RU,TJ,TM),EP(AL,AT,BE,BG,CH,CY,CZ,DE,DK,EE,ES,FI,FR,GB,GR,HR,HU,IE,IS,IT,LT,LU,LV,MC,MK,MT,NL,NO,PL,PT,RO,RS,SE,SI,SK,SM,TR),OA(BF,BJ,CF,CG,CI,CM,GA,GN,GQ,GW,KM,ML,MR,NE,SN,TD,TG),AE,AG,AL,AM,AO,AT,AU,AZ,BA,BB,BG,BH,BN,BR,BW,BY,BZ,CA,CH,CL,CN,CO,CR,CU,CZ,DE,DJ,DK,DM,DO,DZ,EC,EE,EG,ES,FI,GB,GD,GE,GH,GM,GT,HN,HR,HU,ID,IL,IN,IR,IS,JO,JP,KE,KG,KH,KN,KP,KR,KW,KZ,LA,LC,LK,LR,LS,LU,LY,MA,MD,ME,MG,MK,MN,MW,MX,MY,MZ,NA,NG,NI,NO,NZ,OM,PA,PE,PG,PH,PL,PT,QA,RO,RS,RU,RW,SA,SC,SD,SE,SG,SK,SL,SM,ST,SV,SY,TH,TJ,TM,TN,TR,TT

（特許庁注：以下のものは登録商標）
１．ＡＮＤＲＯＩＤ

(72)発明者　ケネディ，ジュリア　シー.
アメリカ合衆国，カリフォルニア州　９４１１６，サン　フランシスコ，カスタネダ　アベニュー　３６０
(72)発明者　アンダーソン，ボニー　エイチ.
アメリカ合衆国，カリフォルニア州　９４０１９，ハーフ　ムーン　ベイ，フェアウェイ　プレイス　４０

Ｆターム(参考) 2G045 AA24 AA25 AA26 CB01 CB02 DA13 JA01
4B063 QA01 QA13 QA19 QQ42 QQ52 QR55 QS24 QS32