

(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2018-514187

(P2018-514187A)

(43) 公表日 平成30年6月7日(2018.6.7)

(51) Int.Cl.

C 12 Q 1/68
G 06 F 19/22

F 1

C 12 Q 1/68
C 12 Q 1/68
G 06 F 19/22

A

Z

テーマコード(参考)

4 B 0 6 3

審査請求 未請求 予備審査請求 未請求 (全 54 頁)

(21) 出願番号	特願2017-546066 (P2017-546066)
(86) (22) 出願日	平成28年3月3日(2016.3.3)
(85) 翻訳文提出日	平成29年10月24日(2017.10.24)
(86) 國際出願番号	PCT/US2016/020583
(87) 國際公開番号	W02016/141127
(87) 國際公開日	平成28年9月9日(2016.9.9)
(31) 優先権主張番号	62/128, 463
(32) 優先日	平成27年3月4日(2015.3.4)
(33) 優先権主張国	米国(US)
(31) 優先権主張番号	62/238, 893
(32) 優先日	平成27年10月8日(2015.10.8)
(33) 優先権主張国	米国(US)
(31) 優先権主張番号	62/128, 469
(32) 優先日	平成27年3月4日(2015.3.4)
(33) 優先権主張国	米国(US)

(71) 出願人	511118469 ペラサイト 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 94 080, 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3 OO
(74) 代理人	100102978 弁理士 清水 初志
(74) 代理人	100102118 弁理士 春名 雅夫
(74) 代理人	100160923 弁理士 山口 裕孝
(74) 代理人	100119507 弁理士 刑部 俊

最終頁に続く

(54) 【発明の名称】発現レベルおよび配列変種情報を用いて疾患の発症または再発のリスクを評価するための方法

(57) 【要約】

対象管理の個別化を改善するために2つまたはそれ以上の特徴空間を組み合わせることによる対象から取得されたサンプルの疾患発症のリスクの層別化のための方法、システム、およびキットが、本明細書において提供される。

Performance Across Risk Groups

	INTERMEDIATE/HIGH RISK (N=51)	LOW RISK (N=28)
Classified as "Med-High Risk"	41	6
Classified as "Low Risk"	9	23
Sensitivity	82%	
Specificity	79%	

PATHWAY OR ONTOLOGY	NO. GENES EXPECTED	NO. GENES OBSERVED	FDR	P VALUE
Extracellular matrix	2	20		1.28×10^{-13}
ECM-receptor interaction	1	11		8.12×10^{-08}
Focal adhesion	2	12		3.71×10^{-05}
Tyrosine Kinase Activity	1	6		4.77×10^{-03}
Regulation of immune System Process	3	10		6.83×10^{-02}
Blood Vessel Development	1	7		1.0×10^{-02}

LIST OF GENES USED IN CLASSIFICATION				
COX6C	FANCA	KCTD17	MPRIP	TUBA1B
DCAKD	ICE2	MCM2AP	TINFRSF14	WSB2

FIG. 5

【特許請求の範囲】**【請求項 1】**

以下の工程を含む、対象における疾患の発症のリスクを判定するために該対象の組織サンプルを評価するための方法：

(a) 該対象から取得した針吸引サンプルにおける核酸サンプル中の第1の遺伝子セットの1つまたは複数の遺伝子の各々に対応する発現レベルを取得する工程であって、該第1の遺伝子セットが該対象における該疾患の発症のリスクに関連する、工程；

(b) 該核酸サンプル中の第2の遺伝子セットの1つまたは複数の遺伝子の各々に対応する核酸配列の存在を判定する工程であって、該第2の遺伝子セットが該対象における該疾患の発症のリスクに関連する、工程；

(c) 対照と、(i) (a)において取得した発現レベルおよび(ii) (b)において取得した核酸配列を別々に比較して、該対照に対する該発現レベルおよび該核酸配列の比較を提供する工程であって、該対照中の参照配列に対する該核酸配列の比較により、該第2の遺伝子セットの所定の遺伝子に関する1つまたは複数の配列変種の存在が示される、工程；ならびに

(d) 訓練されたアルゴリズムを用いてプログラムされたコンピュータプロセッサを使用して、(i) 該比較を分析し、かつ(ii) 該比較に基づき該疾患の発症のリスクを判定する工程。

【請求項 2】

前記疾患ががんである、請求項1に記載の方法。

【請求項 3】

(a) の前に、前記対象から前記針吸引サンプルを取得する工程をさらに含む、請求項1に記載の方法。

【請求項 4】

(a) の前に、前記針吸引サンプルにおける前記核酸サンプル由来の前記発現レベルを決定する工程をさらに含む、請求項1に記載の方法。

【請求項 5】

(b) の前に、前記針吸引サンプルにおける前記核酸サンプル由来の前記核酸配列を決定する工程をさらに含む、請求項1に記載の方法。

【請求項 6】

前記核酸配列を前記参照配列と比較して、前記1つまたは複数の配列変種を同定する工程をさらに含む、請求項5に記載の方法。

【請求項 7】

前記参照配列が、前記対象由来のハウスキーピング遺伝子である、請求項6に記載の方法。

【請求項 8】

前記第1の遺伝子セット中または前記第2の遺伝子セット中の前記1つまたは複数の遺伝子が、複数の遺伝子を含む、請求項1に記載の方法。

【請求項 9】

前記針吸引サンプルが、細胞学的に不明瞭であるとまたは疑わしいと判明している、請求項1に記載の方法。

【請求項 10】

前記針吸引サンプルが、約1マイクロリットルまたはそれ未満の容積を有する、請求項1に記載の方法。

【請求項 11】

前記針吸引サンプルが、約9.0またはそれ未満のRNA Integrity Number (RIN) 値を有する、請求項1に記載の方法。

【請求項 12】

前記針吸引サンプルが、約6.0またはそれ未満のRIN値を有する、請求項10に記載の方法。

。

10

20

30

40

50

【請求項 1 3】

前記疾患の発症のリスクが、前記対象における該疾患の再発のリスクを含む、請求項1に記載の方法。

【請求項 1 4】

前記がんの発症のリスクが、前記対象における転移のリスクを含む、請求項2に記載の方法。

【請求項 1 5】

前記訓練されたアルゴリズムが、前記疾患を有すると診断された少なくとも25体の対象由來の組織サンプルを用いて訓練される、請求項1に記載の方法。

【請求項 1 6】

前記訓練されたアルゴリズムが、前記疾患を有すると診断された少なくとも200体の対象由來の組織サンプルを用いて訓練される、請求項15に記載の方法。

【請求項 1 7】

(d) が手術前に行われる、請求項1に記載の方法。

【請求項 1 8】

前記対象が陽性の疾患診断を受ける前に (d) が行われる、請求項1に記載の方法。

【請求項 1 9】

(d) が、前記発症のリスクを低発症リスクまたは中～高発症リスクに層別化する工程をさらに含み、該低発症リスクが、約50%～約80%の発症率を有し、かつ該中～高発症リスクが、約80%～100%の発症率を有する、請求項1に記載の方法。

【請求項 2 0】

前記層別化する工程が、少なくとも80%の精度を有する、請求項19に記載の方法。

【請求項 2 1】

前記層別化する工程が、少なくとも80%の特異度を有する、請求項19に記載の方法。

【請求項 2 2】

1つもしくは複数のフィルター、1つもしくは複数のラッパー、1つもしくは複数の組み込みプロトコル、またはそれらの任意の組み合わせを前記比較に適用する工程をさらに含む、請求項1に記載の方法。

【請求項 2 3】

1つまたは複数の前記フィルターを前記比較に適用する工程をさらに含む、請求項22に記載の方法。

【請求項 2 4】

1つまたは複数の前記フィルターが、t検定、分散分析(ANOVA)分析、ベイズフレームワーク、ガンマ分布、ウィルコクソン順位和検定、二乗検定の級間・級内和、ランクプロダクト法(rank product method)、ランダム置換法、誤分類の閾値(TNoM)、二変数法、相関に基づく特徴選択(CFS)法、最小冗長性最大関連性(MRMR)法、マルコフプランケットフィルター法、非相関収縮重心法、またはそれらの任意の組み合わせを含む、請求項23に記載の方法。

【請求項 2 5】

前記1つまたは複数の配列変種が、点変異、融合遺伝子、置換、欠失、挿入、逆位、変換、転座、またはそれらの任意の組み合わせの1つまたは複数を含む、請求項23に記載の方法。

【請求項 2 6】

1つまたは複数の前記点変異が、約5個～約4000個の点変異である、請求項25に記載の方法。

【請求項 2 7】

1つまたは複数の前記融合遺伝子が、少なくとも2つの融合遺伝子である、請求項25に記載の方法。

【請求項 2 8】

前記第1のセットまたは前記第2のセットの前記1つまたは複数の遺伝子が、約15個未満

10

20

30

40

50

の遺伝子である、請求項1に記載の方法。

【請求項 2 9】

前記第1のセットまたは前記第2のセットの前記1つまたは複数の遺伝子が、約75個未満の遺伝子である、請求項1に記載の方法。

【請求項 3 0】

前記第1のセットまたは前記第2のセットの前記1つまたは複数の遺伝子が、約50個～約400個の遺伝子である、請求項1に記載の方法。

【請求項 3 1】

(b)における取得する工程が、前記核酸配列を取得するために前記FNAサンプルにおける核酸サンプルを配列決定する工程を含む、請求項1に記載の方法。

10

【請求項 3 2】

前記配列決定する工程が、前記第2の遺伝子セットの1つもしくは複数の前記遺伝子またはその変種を濃縮する工程を含む、請求項31に記載の方法。

【請求項 3 3】

(a)が、前記第1の遺伝子セットの前記1つまたは複数の遺伝子に選択的なプローブを用いるマイクロアレイを使用する工程を含む、請求項1に記載の方法。

【請求項 3 4】

前記組織サンプルが甲状腺組織サンプルである、請求項1に記載の方法。

【請求項 3 5】

前記第1の遺伝子セットおよび前記第2の遺伝子セットが、COL1A1、THBS2、またはそれらの任意の組み合わせを含む、請求項34に記載の方法。

20

【請求項 3 6】

前記第2の遺伝子セットが、EPHA3、COL1A1、EHF、RAPGEF5、PRICKLE1、TMEM92、ROBO1、C6orf136、SPAG4、GALNT15、LUM、NCAM2、NUP210L、NR2F1、THBS2、PSORS1C1、またはそれらの任意の組み合わせを含む、請求項34に記載の方法。

【請求項 3 7】

前記第1の遺伝子セットが、COL1A1、TMEM92、C1orf87、SPAG4、EHF、COL3A1、GALNT15、NUP210L、PDZRN3、C6orf136、NA、NRXN3、COL6A3、RAPGEF5、PRICKLE1、LUM、ROBO1、BGN、AC019117.2、PRSS3P1、またはそれらの任意の組み合わせを含む、請求項34に記載の方法。

30

【請求項 3 8】

前記第2の遺伝子セットが、EPHA3、COL1A1、EHF、RAPGEF5、PRICKLE1、TMEM92、ROBO1、C6orf136、SPAG4、GALNT15、LUM、NCAM2、SYNPO2、NUP210L、AMZ1、NR2F1、THBS2、PSORS1C1、FTH1P24、またはそれらの任意の組み合わせを含む、請求項34に記載の方法。

【請求項 3 9】

前記第2の遺伝子セットが、AKAP9、SPRY3、SPRY3、CAMKK2、COL1A1、FITM2、COX6C、VSG10L、CYC1、KDM1B、MAPK15、ARSG、PAXIP1、DAAM1、AVL9、DMGDH、HLA-DQA1、HLA-DQB1、HLA-DRA、HLA-DRB5、HLA-H、IRF1、MGAT1、P2RX1、PLEK、CCDC93、PPP1R12C、SLC41A3、METTL3、CCAR2、PTPRE、SRL、SLC30A5、BMP4、ZNF133、ICE2、DCAKD、TMX1、TNFSF12、PER2、MCM3AP、またはそれらの任意の組み合わせを含む、請求項34に記載の方法。

40

【請求項 4 0】

前記第1の遺伝子セットおよび前記第2の遺伝子セットが異なる、請求項1に記載の方法。

【請求項 4 1】

前記疾患の新規遺伝子バイオマーカーを同定する工程をさらに含む、請求項1に記載の方法。

【請求項 4 2】

(a)における取得する工程が、前記1つまたは複数の遺伝子の各々に対応する前記発現レベルについてアッセイする工程を含む、請求項1に記載の方法。

【請求項 4 3】

50

前記アッセイする工程が、前記1つまたは複数の遺伝子の各々について選択されたマークーを用いるアレイハイブリダイゼーション、核酸配列決定、または核酸増幅を含む、請求項42に記載の方法。

【請求項44】

前記マークーが、前記1つまたは複数の遺伝子の各々について選択されたプライマーである、請求項43に記載の方法。

【請求項45】

前記アッセイする工程が逆転写ポリメラーゼ連鎖反応（PCR）を含む、請求項43に記載の方法。

【請求項46】

前記判定する工程が、前記核酸サンプル中の前記第2の遺伝子セットの前記1つまたは複数の遺伝子の各々についてアッセイする工程を含む、請求項1に記載の方法。

10

【請求項47】

前記アッセイする工程が、前記1つまたは複数の遺伝子の各々について選択されたマークーを用いるアレイハイブリダイゼーション、核酸配列決定、または核酸増幅を含む、請求項46に記載の方法。

【請求項48】

前記マークーが、前記1つまたは複数の遺伝子の各々について選択されたプライマーである、請求項47に記載の方法。

【請求項49】

前記アッセイする工程が逆転写ポリメラーゼ連鎖反応（PCR）を含む、請求項47に記載の方法。

20

【請求項50】

前記針吸引サンプルが微細針吸引サンプルである、請求項1に記載の方法。

【請求項51】

以下を備える、対象における疾患の発症のリスクを判定するために該対象の組織サンプルを評価するためのシステム：

(a) 第1の遺伝子セットが該対象における該疾患の発症のリスクに関連する、該対象から取得した針吸引サンプルにおける核酸サンプル中の該第1の遺伝子セットの1つまたは複数の遺伝子の各々に対応する発現と、(b) 第2の遺伝子セットが該対象における該疾患の発症のリスクに関連する、該核酸サンプル中の該第2の遺伝子セットの1つまたは複数の遺伝子の各々に対応する核酸配列の存在の表示とを保存する、1つまたは複数のコンピュータメモリ；ならびに

30

該1つまたは複数のコンピュータメモリに接続され、かつ、

(i) 対照と、(1) 該コンピュータメモリ中の発現レベルおよび(2) 該核酸配列を別々に比較して、該対照に対する該発現レベルおよび該核酸配列の比較を提供し、該対照中の参照配列に対する該核酸配列の比較により、該第2の遺伝子セットの所定の遺伝子に関する1つまたは複数の配列変種の存在が示されるよう、かつ

(ii) 訓練されたアルゴリズムを使用して、(1) 該比較を分析し、かつ(2) 該比較に基づき該疾患の発症のリスクを判定するよう

40

プログラムされた、コンピュータプロセッサ。

【請求項52】

1つまたは複数のコンピュータプロセッサによって実行されると対象における疾患の発症のリスクを判定するために該対象の組織サンプルを評価するための方法を実施する機械実行可能なコードを備える非一時的コンピュータ読み取り可能媒体であって、該方法が以下の工程を含む、非一時的コンピュータ読み取り可能媒体：

(a) 該対象から取得した針吸引サンプルにおける核酸サンプル中の第1の遺伝子セットの1つまたは複数の遺伝子の各々に対応する発現レベルを取得する工程であって、該第1の遺伝子セットが該対象における該疾患の発症のリスクに関連する、工程；

(b) 該核酸サンプル中の第2の遺伝子セットの1つまたは複数の遺伝子の各々に対応す

50

る核酸配列の存在を判定する工程であって、該第2の遺伝子セットが該対象における該疾患の発症のリスクに関連する、工程；

(c) 対照と、(i) (a)において取得した発現レベルおよび(ii) (b)において取得した核酸配列を別々に比較して、該対照に対する該発現レベルおよび該核酸配列の比較を提供する工程であって、該対照中の参照配列に対する該核酸配列の比較により、該第2の遺伝子セットの所定の遺伝子に関する1つまたは複数の配列変種の存在が示される、工程；ならびに

(d) 訓練されたアルゴリズムを用いてプログラムされたコンピュータプロセッサを使用して、(i) 該比較を分析し、かつ(ii) 該比較に基づき該疾患の発症のリスクを判定する工程。

10

【発明の詳細な説明】

【技術分野】

【0001】

相互参照

本願は、各々全体が参照により本明細書に組み入れられる2015年3月4日に出願された米国特許出願第62/128,463号、2015年3月4日に出願された米国特許出願第62/128,469号および2015年10月8日に出願された米国特許出願第62/238,893号からの優先権を主張する。

【背景技術】

【0002】

20

背景

疾患治療、例えば甲状腺がん治療に対するリスク適応型アプローチは、疾患特異的に生存性を改善するのに加えて、疾患発症のリスクを最小限に抑制し得る。現在、初期の対象管理に対するこのリスク適応型アプローチは、大部分が、2009 米国甲状腺学会 (American Thyroid Association) (ATA) 病期体系を用いた高、中、または低疾患再発リスクのいずれかへの対象の手術後分類に基づいている。この解剖学的病期体系は、臨床的に有用であることが証明されているが、それは侵襲的な甲状腺摘出術の前に正確に評価されることができず、かつそれは疾患結果のいかなる分子的予測因子も含まない。

【発明の概要】

【0003】

30

概要

疾患の発症および／または再発のリスクを評価または層別化するための様々な方法が、本明細書において提供されている。診断前評価または診断評価、例えば微細針吸引 (FNA) の間に取得された転写データは、疾患、例えば甲状腺がんの発症のリスクの手術前予測を改善し得、かつ対象の治療および処置のさらなる個別化を提供し得る。本開示の方法は、比較的非侵襲的な様式でかつ少ないサンプル量を用いて疾患の発症および／または再発のリスクに関する評価を提供し得る。

【0004】

本開示の1つの局面は、対象における疾患の発症のリスクを判定するために対象の組織サンプルを評価するための方法を提供する。該方法は、(a) 対象から取得した針吸引サンプルにおける核酸サンプル中の第1の遺伝子セットの1つまたは複数の遺伝子の各々に対応する発現レベルを取得する工程であって、第1の遺伝子セットが対象における疾患の発症のリスクに関連する、工程；(b) 該核酸サンプル中の第2の遺伝子セットの1つまたは複数の遺伝子の各々に対応する核酸配列の存在を判定する工程であって、第2の遺伝子セットが対象における疾患の発症のリスクに関連する、工程；(c) 対照と、(i) (a)において取得した発現レベルおよび(ii) (b)において取得した核酸配列を別々に比較して、対照に対する発現レベルおよび核酸配列の比較を提供する工程であって、対照中の参照配列に対する核酸配列の比較により、第2の遺伝子セットの所定の遺伝子に関する1つまたは複数の配列変種の存在が示される、工程；ならびに(d) 訓練されたアルゴリズムを用いてプログラムされたコンピュータプロセッサを使用して、(i) 該比較を分析し、か

40

50

つ(ii)該比較に基づき疾患の発症のリスクを判定する工程を含む。

【0005】

いくつかの態様において、針吸引サンプルは、微細針吸引サンプルである。いくつかの態様において、疾患はがんである。いくつかの態様において、この方法は、(a)の前に、対象から針吸引サンプルを取得する工程をさらに含む。いくつかの態様において、この方法は、(a)の前に、針吸引サンプルにおける核酸サンプル由来の発現レベルを決定する工程をさらに含む。いくつかの態様において、この方法は、(b)の前に、針吸引サンプルにおける核酸サンプル由来の核酸配列を決定する工程をさらに含む。いくつかの態様において、この方法は、前記核酸配列を参照配列と比較して、1つまたは複数の配列変種を同定する工程をさらに含む。いくつかの態様において、参照配列は、対象由来のハウスキーピング遺伝子である。いくつかの態様において、第1の遺伝子セットまたは第2の遺伝子セットの1つまたは複数の遺伝子は、複数の遺伝子を含む。

10

【0006】

いくつかの態様において、針吸引サンプルは、細胞学的に不明瞭であるとまたは疑わしいと判明している。いくつかの態様において、針吸引サンプルは、約1マイクロリットルまたはそれ未満の容積を有する。いくつかの態様において、針吸引サンプルは、約9.0またはそれ未満のRNA Integrity Number (RIN) 値を有する。いくつかの態様において、針吸引サンプルから精製されたRNAは、約9.0またはそれ未満のRNA RIN値を有する。いくつかの態様において、針吸引サンプルは、約6.0またはそれ未満のRIN値を有する。いくつかの態様において、RNAサンプルは、約6.0またはそれ未満のRIN値を有する。

20

【0007】

いくつかの態様において、疾患の発症のリスクは、対象における疾患の再発のリスクを含む。いくつかの態様において、がんの発症のリスクは、対象における転移のリスクを含む。いくつかの態様において、がんの発症のリスクは、疾患の進行の加速のリスクを含む。いくつかの態様において、がんの発症のリスクは、治療の失敗のリスクを含む。

【0008】

いくつかの態様において、訓練されたアルゴリズムは、前記疾患を有すると診断された少なくとも25体または少なくとも100体の対象由来の組織サンプルを用いて訓練される。いくつかの態様において、訓練されたアルゴリズムは、該疾患を有すると診断された少なくとも200体の対象由来の組織サンプルを用いて訓練される。

30

【0009】

いくつかの態様において、(d)は手術前に行われる。いくつかの態様において、(d)は、対象が陽性の疾患診断を受ける前に行われる。いくつかの態様において、(d)は、発症のリスクを低発症リスクまたは中～高発症リスクに層別化する工程をさらに含み、低発症リスクは、約50%～約80%の発症率を有し、かつ中～高発症リスクは、約80%～100%の発症率を有する。

【0010】

いくつかの態様において、この方法は、1つもしくは複数のフィルター、1つもしくは複数のラッパー、1つもしくは複数の組み込みプロトコル、またはそれらの任意の組み合わせを上記比較に適用する工程をさらに含む。いくつかの態様において、1つもしくは複数のフィルターが上記比較に適用される。いくつかの態様において、1つまたは複数のフィルターは、t検定、分散分析(ANOVA)分析、ペイズフレームワーク、ガンマ分布、ウィルコクソン順位和検定、二乗検定の級間・級内和、ランクプロダクト法(rank product method)、ランダム置換法、誤分類の閾値(TNoM)、二変数法、相関に基づく特徴選択(CFS)法、最小冗長性最大関連性(MRMR)法、マルコフプランケットフィルター法、非相関収縮重心法、またはそれらの任意の組み合わせを含む。いくつかの態様において、1つまたは複数の配列変種は、点変異、融合遺伝子、置換、欠失、挿入、逆位、変換、転座の1つもしくは複数、またはそれらの任意の組み合わせを含む。いくつかの態様において、1つまたは複数の点変異は、約5個～約4000個の点変異である。いくつかの態様において、1つまたは複数の融合遺伝子は、少なくとも2つの融合遺伝子である。

40

50

【0011】

いくつかの態様において、層別化する工程は、約80%の精度を有する。いくつかの態様において、層別化する工程は、約80%の特異度を有する。いくつかの態様において、第1のセットまたは第2のセットの1つまたは複数の遺伝子は、約15個未満または約10個未満の遺伝子である。いくつかの態様において、第1のセットまたは第2のセットの1つまたは複数の遺伝子は、約75個未満の遺伝子である。いくつかの態様において、第1のセットまたは第2のセットの1つまたは複数の遺伝子は、約50個～約400個の遺伝子である。

【0012】

いくつかの態様において、(b)における取得する工程は、核酸配列を取得するために針吸引サンプルにおける核酸サンプルを配列決定する工程を含む。いくつかの態様において、配列決定する工程は、第2の遺伝子セットの1つもしくは複数の遺伝子またはその変種を濃縮する工程を含む。いくつかの態様において、(a)は、第1の遺伝子セットの1つまたは複数の遺伝子に選択的なプローブを用いるマイクロアレイを使用する工程を含む。いくつかの態様において、(a)は、標的化配列決定プラットホーム(例えば、Ion Torrent AmpliconまたはIllumina TruSeq Custom Amplicon)の使用を含む。

10

【0013】

いくつかの態様において、組織サンプルは、甲状腺組織サンプルである。いくつかの態様において、第1および第2の遺伝子セットは、COL1A1、THBS2、またはそれらの任意の組み合わせを含む。いくつかの態様において、第2の遺伝子セットは、EPHA3、COL1A1、EHF、RAPGEF5、PRICKLE1、TMEM92、ROBO1、C6orf136、SPAG4、GALNT15、LUM、NCAM2、NUP210L、NR2F1、THBS2、PSORS1C1、またはそれらの任意の組み合わせを含む。いくつかの態様において、第1の遺伝子セットは、COL1A1、TMEM92、C1orf87、SPAG4、EHF、COL3A1、GALNT15、NUP210L、PDZRN3、C6orf136、NA、NRXN3、COL6A3、RAPGEF5、PRICKLE1、LUM、ROBO1、BGN、AC019117.2、PRSS3P1、またはそれらの任意の組み合わせを含む。いくつかの態様において、第2の遺伝子セットは、EPHA3、COL1A1、EHF、RAPGEF5、PRICKLE1、TMEM92、ROBO1、C6orf136、SPAG4、GALNT15、LUM、NCAM2、SYNP02、NUP210L、AMZ1、NR2F1、THBS2、PSORS1C1、FTH1P24、またはそれらの任意の組み合わせを含む。いくつかの態様において、第2の遺伝子セットは、AKAP9、SPRY3、SPRY3、CAMKK2、COL1A1、FITM2、COX6C、VSIG10L、CYC1、KDM1B、MAPK15、ARSG、PAXIP1、DAAM1、AVL9、DMGDH、HLA-DQA1、HLA-DQB1、HLA-DRA、HLA-DRB5、HLA-H、IRF1、MGAT1、P2RX1、PLEK、CCDC93、PPP1R12C、SLC41A3、METTL3、CCAR2、PTPRE、SRL、SLC30A5、BMP4、ZNF133、ICE2、DCAKD、TMX1、TNFSF12、PER2、MCM3AP、またはそれらの任意の組み合わせを含む。

20

【0014】

いくつかの態様において、第1の遺伝子セットおよび第2の遺伝子セットは異なる。いくつかの態様において、この方法は、疾患の新規遺伝子バイオマーカーを同定する工程をさらに含む。

30

【0015】

いくつかの態様において、(a)における取得する工程は、1つまたは複数の遺伝子の各々に対応する発現レベルについてアッセイする工程を含む。いくつかの態様において、アッセイする工程は、1つまたは複数の遺伝子の各々について選択されたマーカーを用いるアレイハイブリダイゼーション、核酸配列決定、または核酸增幅を含む。いくつかの態様において、マーカーは、1つまたは複数の遺伝子の各々について選択されたプライマーである。

40

【0016】

いくつかの態様において、アッセイする工程は、逆転写ポリメラーゼ連鎖反応(PCR)を含む。いくつかの態様において、判定する工程は、核酸サンプル中の第2の遺伝子セットの1つまたは複数の遺伝子の各々についてアッセイする工程を含む。いくつかの態様において、アッセイする工程は、1つまたは複数の遺伝子の各々について選択されたマーカーを用いるアレイハイブリダイゼーション、核酸配列決定、または核酸增幅を含む。いくつかの態様において、マーカーは、1つまたは複数の遺伝子の各々について選択されたプ

50

ライマーである。いくつかの態様において、アッセイする工程は、逆転写ポリメラーゼ連鎖反応（PCR）を含む。

【0017】

本開示の別の局面は、1つまたは複数のコンピュータプロセッサによって実行されると本明細書の上記または他箇所の方法のいずれかを実施する機械実行可能なコードを備えるコンピュータ読み取り可能媒体（例えば、メモリ）を提供する。

【0018】

本開示の別の局面は、1つまたは複数のコンピュータプロセッサおよびそれに接続されたコンピュータ読み取り可能媒体を備えるコンピュータシステムを提供する。コンピュータ読み取り可能媒体は、1つまたは複数のコンピュータプロセッサによって実行されると本明細書の上記または他箇所の方法のいずれかを実施する機械実行可能なコードを備え得る。

10

【0019】

本開示のさらなる局面および利点は、本開示の例示にすぎない態様が示され説明されている以下の詳細な説明から当業者に直ちに明らかとなるであろう。明らかなように、本開示は、他のおよび異なる態様が可能であり、そのそれぞれの細部は様々な明白な局面において改変することが可能であり、それらはすべて本開示から逸脱することなく行われる。したがって、図面および説明は、本質的に例示的なものであり、限定的なものではないとみなされるべきである。

20

【0020】

参照による組み入れ

本明細書で言及されているすべての刊行物、特許、および特許出願は、各々個々の刊行物、特許、または特許出願が具体的かつ個別に参照により組み入れられることが示されているものとして、参照により本明細書に組み入れられる。参照により組み入れられる刊行物および特許または特許出願が本明細書に含まれる開示と相反する範囲では、本明細書があらゆるそのような相反するものに取って代わるかまたはそれよりも優先される。

【図面の簡単な説明】

【0021】

本発明の新規の特徴は、特に添付の特許請求の範囲に示されている。本発明の特徴および利点のさらなる理解は、本発明の原理が用いられている例示的な態様を示す以下の詳細な説明および添付図面（本明細書で「図（figure）」および「図（FIG）」とも称される）を参照することによって得られるであろう。

30

【図1】低がん発症リスクおよび中～高がん発症リスクに層別化された細胞学データおよび専門家による組織病理学データのサンプルコホートを示している。

【図2】組織病理学的なリスク特徴および各特徴に関するサンプルの数および比率を示している。

【図3】偽陽性率に対してプロットされた真陽性率の交差検証を示している。

【図4】中～高リスクの有病率に対して的中率をプロットした分類パフォーマンスデータを示している。

【図5】低リスクおよび中～高リスクグループにおける分類パフォーマンスデータを示している。

40

【図6-1】遺伝子発現レベルデータに基づく甲状腺がんの発症のリスクに関連する遺伝子の具体例リストを示している。

【図6-2】遺伝子発現レベルデータに基づく甲状腺がんの発症のリスクに関連する遺伝子の具体例リストを示している。

【図6-3】遺伝子発現レベルデータに基づく甲状腺がんの発症のリスクに関連する遺伝子の具体例リストを示している。

【図7-1】リボ核酸（RNA）配列決定から取得された遺伝子発現レベルデータに基づく甲状腺がんの発症のリスクに関連する遺伝子の具体例リストを示している。

【図7-2】リボ核酸（RNA）配列決定から取得された遺伝子発現レベルデータに基づく

50

体例リストを示している。

【図 8 - 1 1】配列変種データに基づく甲状腺がんの発症のリスクに関連する遺伝子の具体例リストを示している。

【図 8 - 1 2】配列変種データに基づく甲状腺がんの発症のリスクに関連する遺伝子の具体例リストを示している。

【図 8 - 1 3】配列変種データに基づく甲状腺がんの発症のリスクに関連する遺伝子の具体例リストを示している。

【図 8 - 1 4】配列変種データに基づく甲状腺がんの発症のリスクに関連する遺伝子の具体例リストを示している。

【図 8 - 1 5】配列変種データに基づく甲状腺がんの発症のリスクに関連する遺伝子の具体例リストを示している。 10

【図 9】本明細書において提供される方法を実行するようプログラムされているかまたはそうでなければ該方法を実行するよう構成されている、コンピュータ制御システムを示している。

【図 1 0】正確な訓練ラベルを決定する流れ図を示している。

【図 1 1 A】偽陽性率に対してプロットされた真陽性率の交差検証を示している。

【図 1 1 B】中／高リスクおよび低リスクグループにおける分類パフォーマンスデータを示している。

【図 1 2】各分割で分類器によって選択された変種の遺伝子の具体例リストを示している。 20

【図 1 3】10分割で分類器によって8～10回選択されたカウントの遺伝子の具体例リストを示している。

【図 1 4】5つの点変異パネルおよび融合対の表を示している。

【図 1 5】5つの変異および融合対のパネルにおける試験パフォーマンスの特異度および感度のグラフを示している。

【図 1 6】細胞学による図14および15のパネル3の変異パフォーマンスの表を示している。 30

【図 1 7】5つの変異および融合対のパネルにおける試験パフォーマンスの特異度および感度のグラフを示している。

【図 1 8】図18Aは、臨床検査室改善修正法 (Clinical Laboratory Improvement Amendment) (CLIA) 微細針吸引 (FNA) サンプルの変異頻度のグラフによる表示を示しており、図18Bは、その変異頻度の表による表示を示している。

【図 1 9】図19Aは、FNAサンプルの変異頻度のグラフによる表示を示しており、図19Bは、その変異頻度の表による表示を示している。

【図 2 0】図20Aは、組織サンプルの変異頻度のグラフによる表示を示しており、図20Bは、その変異頻度の表による表示を示している。

【発明を実施するための形態】

【0 0 2 2】

詳細な説明

本明細書において本発明の様々な態様が示され説明されているが、そのような態様は例として提供されるにすぎないことが当業者に明らかであろう。当業者は、本発明から逸脱することなく、多数のバリエーション、変更および置換を思い浮かべるであろう。本明細書に記載される本発明の態様に対する様々な代替物が用いられ得ることが理解されるべきである。 40

【0 0 2 3】

「対象」という用語は、本明細書で使用される場合、概ね、任意の動物または生きた有機体を表す。動物は、哺乳動物、例えばヒト、非ヒト霊長類、げっ歯類、例えばマウスおよびラット、イヌ、ネコ、ブタ、ヒツジ、ウサギ等であり得る。動物は、魚類、爬虫類等であり得る。動物は、生まれたばかりの、幼い、若い、または成体の動物であり得る。ヒトは、約1、2、5、10、20、30、40、50、60、65、70、75、または約80歳超であり得る。 50

対象は、疾患、例えばがんを有し得るまたはそれを有する疑いがあり得る。対象は、患者、例えば疾患の処置を受けている患者、例えばがん患者であり得る。対象は、疾患、例えばがんを発症するリスクにさらされている者であり得る。対象は、疾患からの回復期にあり得る、例えばがん患者であり得る。対象は、健常であり得る。

【0024】

「疾患」という用語は、本明細書で使用される場合、概ね、対象に影響を与える任意の異常または病理学的な状態を表す。疾患の例は、がん、例えば甲状腺がん、副甲状腺がん、肺がん、皮膚がん等を含む。疾患は、処置可能または処置不可能であり得る。疾患は、末期的または非末期的であり得る。疾患は、受け継いだ遺伝子、環境的曝露、またはそれらの任意の組み合わせの結果であり得る。疾患は、がん、遺伝病、増殖性障害、または本明細書に記載される他の疾患であり得る。

10

【0025】

「疾患の発症のリスク」という用語は、本明細書で定義される場合、概ね、対象における疾患の発症に関連するリスクまたは確率を表す。発症のリスクは、対象における疾患の最初の発症を含み得、または後続の発症、例えば第2、第3、第4またはそれ以降の発症を含み得る。疾患の発症のリスクは、(a) 疾患を最初に発症するリスク、(b) 疾患を再発するもしくは疾患を再び発症するリスク、(c) 疾患を将来発症するリスク、(d) 対象の生涯の中で疾患を発症する可能性があるというリスク、または(e) 幼いとき、若いとき、もしくは大人になってから疾患を発症する可能性があるというリスクを含み得る。疾患、例えばがんの発症のリスクは、がんが転移性になるリスクを含み得る。疾患、例えばがんの発症のリスクは、Ⅰ期がん、Ⅱ期がん、Ⅲ期がん、またはⅣ期がんの発症のリスクを含み得る。がんの発症のリスクは、血液がん、組織がん(例えば、腫瘍)または他の部位から1つもしくは複数の臓器部位に転移するようになるがんのリスクを含み得る。

20

【0026】

「配列変種」、「配列バリエーション」、「配列変更」、または「対立遺伝子変種」という用語は、本明細書で使用される場合、概ね、参照配列、例えば、ゲノムデオキシリボ核酸(DNA)参照配列、コーディングDNA参照配列もしくはタンパク質参照配列等との関連での特定の変化またはバリエーションを表す。参照DNA配列は、参照データベースから取得され得る。配列変種は、機能に影響する場合がある。配列変種は、機能に影響しない場合がある。配列変種は、1つもしくは複数のヌクレオチドにおいてDNAレベルで、1つもしくは複数のヌクレオチドにおいてリボ核酸(RNA)レベルで、1つまたは複数のアミノ酸においてタンパク質レベルで、またはそれらの任意の組み合わせで生じ得る。参照配列は、データベース、例えばNCBI レファレンス配列データベース(Reference Sequence Database)(RefSeq)データベースから取得され得る。配列バリエーションを構成し得る個々の変化は、1つもしくは複数のヌクレオチドまたは1つもしくは複数のアミノ酸における置換、欠失、挿入、逆位または変換を含み得る。配列変種は、点変異であり得る。配列変種は、融合遺伝子であり得る。融合対または融合遺伝子は、配列変種、例えば転座、中間部欠失、染色体逆位、またはそれらの任意の組み合わせから生じ得る。配列バリエーションは、反復配列の数の多様性、例えば三重配列、四重配列等を構成し得る。例えば、配列バリエーションは、所定の配列に関するコピー数の増加または減少であり得る(すなわち、コピー数バリエーションまたはCNV)。配列バリエーションは、異なる対立遺伝子における2つもしくはそれ以上の配列変化または1つの対立遺伝子における2つもしくはそれ以上の配列変化を含み得る。配列バリエーションは、1つの対立遺伝子内の1つの位置における2つの異なるヌクレオチド、例えばモザイクを含み得る。配列バリエーションは、1つの対立遺伝子内の1つの位置における2つの異なるヌクレオチド、例えばキメラを含み得る。配列変種は、悪性組織に存在し得る。配列変種は、良性組織に存在し得る。変種の非存在は、組織またはサンプルが良性であることを示し得る。あるいは、変種の非存在は、組織またはサンプルが良性であることを示さない場合がある。

30

【0027】

「変異パネル」という用語は、本明細書で使用される場合、概ね、リスク分類器を用い

40

50

て検出する（または調査される）規定数のゲノム部位および融合対を指定するパネルを表す。例えば、変異パネルは、調査する9つのゲノム部位および3つの融合対を含み得る。検出する点変異および融合対の数を増やすことによりリスク分類器の感度を高めることは、リスク分類器の感度を低下させることになり得る。

【0028】

変異パネルは、1つまたは複数のゲノム部位および1つまたは複数の融合対を含み得る。変異パネルは、約1、2、3、4、または5個超のゲノム部位を含み得る。変異パネルは、約15個超のゲノム部位を含み得る。変異パネルは、約100個超のゲノム部位を含み得る。変異パネルは、約200個超のゲノム部位を含み得る。変異パネルは、約500個超のゲノム部位を含み得る。変異パネルは、約1000個超のゲノム部位を含み得る。変異パネルは、約2000個超のゲノム部位を含み得る。変異パネルは、約3000個超のゲノム部位を含み得る。変異パネルは、約1または2個超の融合対を含み得る。変異パネルは、約5個超の融合対を含み得る。変異パネルは、約10個超の融合対を含み得る。変異パネルは、約15個超の融合対を含み得る。変異パネルは、約20個超の融合対を含み得る。変異パネルは、約25個超の融合対を含み得る。

10

【0029】

「疾患診断」という用語は、本明細書で使用される場合、概ね、疾患を診断もしくはスクリーニングすること、疾患の発症のリスクを層別化すること、疾患の進行もしくは寛解をモニタリングすること、疾患に対する処置計画を編成すること、またはそれらの任意の組み合わせを表す。疾患診断は、(a) 対象由来の1つもしくは複数の組織サンプルから情報を取得すること、(b) 取得した情報もしくは組織サンプルに基づき対象が特定の疾患有するかどうかについての判定を行うこと、(c) 対象における疾患の発症のリスクを層別化すること、(d) 対象が疾患有するかどうか、疾患を発症するかどうか、もしくは疾患寛解状態にあるかどうかを確認すること、またはそれらの任意の組み合わせを含み得る。疾患診断は、その疾患に対する特定の処置または治療的介入を通知し得る。疾患診断はまた、例えば、疾患、例えばがんの重篤度もしくはグレードを示すスコアまたは、例えばp値、補正されたp値もしくは統計的信頼性指標を通じた、正確な診断の尤度を提供し得る。疾患診断はまた、特定のタイプの疾患を示し得る。例えば、甲状腺がんに対する疾患診断は、サブタイプ、例えば濾胞性腺腫(FA)、結節性過形成(NHP)、リンパ球性甲状腺炎(LCT)、ハースル細胞腺がん(HA)、濾胞がん(FC)、甲状腺乳頭がん(PTC)、濾胞型甲状腺乳頭がん(FVPTC)、甲状腺髓様がん(MTC)、ハースル細胞がん(HC)、未分化甲状腺がん(ATC)、腎臓がん(RCC)、乳がん(BCA)、黒色腫(MMN)、B細胞リンパ腫(BCL)、副甲状腺(PTA)、または過形成乳頭がん(HPC)を示し得る。

20

【0030】

疾患の発症または再発のリスクを評価するための方法

本開示は、対象における疾患の発症または再発のリスクを判定するために、いくつかの例では、疾患の新規遺伝子バイオマーカーを決定するために対象の組織サンプルを評価するための方法を提供する。そのような方法は、対象から取得された核酸サンプル中の第1の遺伝子セットの1つまたは複数の遺伝子の各々に対応する発現レベルを取得する工程を含み得る。いくつかの例において、発現レベルは、第1の遺伝子セットの1つまたは複数の遺伝子に選択的なプローブを用いるマイクロアレイを用いて取得される。核酸サンプルは、対象によってまたは別の個人によって、例えば医療専門家によって取得され得る。第1の遺伝子セットは、対象における疾患の発症のリスクに関連し得る。いくつかの例において、核酸サンプルは、FNA、手術（例えば、外科生検）または対象からサンプルを取得する他のアプローチによって取得される。核酸サンプルは、対象から取得された組織サンプル（例えば、甲状腺組織サンプル）、血液サンプル中または体液サンプル中に含まれ得る。1つの例において、核酸サンプルは、対象から取得されたFNAサンプルに含まれ得る。

30

【0031】

次に、核酸サンプル中の第2の遺伝子セットの1つまたは複数の遺伝子の各々に対応する核酸配列の存在が判定される。第2の遺伝子セットは、対象における疾患の発症のリスク

40

50

に関連し得る。いくつかの例において、その配列の存在は、FNAサンプル中の核酸を配列決定し、その核酸配列を取得することによって判定される。配列決定はまた、第2の遺伝子セットの1つもしくは複数の遺伝子またはそれらの変種を濃縮し得る。

【0032】

次に、取得された発現レベルおよび取得された核酸配列が対照と比較され、対照に対する発現レベルおよび核酸配列の比較が提供される。対照内の参照配列に対する核酸配列の比較により、第2の遺伝子セットの所定の遺伝子に関する1つまたは複数の配列変種の存在が示され得る。参照配列は、例えば、対象から取得されたハウスキーピング遺伝子であり得る。

【0033】

次に、比較が分析され、その比較に基づき疾患の発症または再発のリスクが判定される。いくつかの例において、その比較を分析し疾患の発症または再発のリスクを判定するために、1つまたは複数のプログラムされたコンピュータプロセッサによって実行されるアルゴリズムが使用される。アルゴリズムは、訓練されたアルゴリズム（例えば、少なくとも10、200、100、または500個の参照サンプルに対して訓練されたアルゴリズム）であり得る。参照サンプルは、該疾患を有すると診断された対象からまたは健常対象から取得され得る。

10

【0034】

いくつかの例において、第1の遺伝子セットの1つまたは複数の遺伝子の各々の発現レベルは、その発現レベルについてアッセイすることによって取得され得る。いくつかの例において、第2の遺伝子セットの1つまたは複数の遺伝子の各々に対応する核酸配列の存在は、1つまたは複数の遺伝子の各々についてアッセイすることによって判定され得る。そのような例において、アッセイすることは、アレイハイブリダイゼーション、核酸配列決定、核酸増幅等を含み得る。アッセイすることは、配列決定、例えばDNAまたはRNA配列決定を含み得る。そのような配列決定は、次世代（NextGen）配列決定により得る。アッセイすることは、逆転写ポリメラーゼ連鎖反応（PCR）を含み得る。アッセイすることは、第1遺伝子セットまたは第2の遺伝子セットの1つまたは複数の遺伝子の各々について選択されたマーカー、例えばプライマーを用い得る。

20

【0035】

第1の遺伝子セットの1つまたは複数の遺伝子に対応する発現レベルを取得する前に、サンプルが対象から取得され得る。核酸サンプルの複数の遺伝子の発現レベルがまた、第1の遺伝子セットの1つまたは複数の遺伝子に対応する発現レベルを取得する前に決定され得る。いくつかの例において、第2の遺伝子セットの核酸配列の存在を判定する前に、サンプル中の複数の遺伝子の核酸配列が決定され得る。

30

【0036】

いくつかの例において、疾患は、がん、例えば甲状腺がん、乳がん等である。発症または再発のリスクの判定は、非がん性疾患、例えば遺伝病、過剰増殖性障害等においても判定され得る。

【0037】

対象から取得されたサンプルは、細胞学的に不明瞭であってもよいかまたは疑わしくてもよい（または不確定であってもよい）。いくつかの例において、サンプルは、疾患の存在を示唆していてもよい。対象から取得されるサンプルの容積は、少量、例えば約100マイクロリットル、50マイクロリットル、10マイクロリットル、5マイクロリットル、1マイクロリットルまたはそれ未満であり得る。サンプルは、少量または低品質のポリヌクレオチド、例えば、分解または部分分解したRNAを含む組織サンプルを含み得る。例えば、FNAサンプルは、少量または低品質のポリヌクレオチドを提供し得る。そのような例において、サンプルのRNA Integrity Number（RIN）値は、約9.0またはそれ未満であり得る。いくつかの例において、RIN値は、約6.0またはそれ未満であり得る。

40

【0038】

疾患の発症のリスクは、後続の発症、例えば第2、第3、第4、またはそれより後続の発

50

症のリスクを含み得る。疾患の発症のリスクは、(a) 疾患を最初に発症するリスク、(b) 疾患を再発するもしくは疾患を再び発症するリスク、(c) 疾患を将来発症するリスク、(d) 対象の生涯の中で疾患を発症する可能性があるというリスク、または(e) 幼いとき、若いときもしくは大人になってから疾患を発症する可能性があるというリスク、の1つまたは複数を含み得る。疾患ががんの場合、発症のリスクは、がんが転移性になるリスクを含み得る。

【0039】

リスクの判定は、手術前に、例えば患者の手術前に完了され得る。臨床医は、患者が例えば低リスクの甲状腺乳頭がんを有すると判定された場合に、手術を推奨するよりも患者が継続して観察されることを推奨し得る。いくつかの例において、臨床医は、患者が高リスクの甲状腺乳頭がんを有すると判定された場合、患者が手術を受けることを推奨する可能性が高い。判定は、患者が陽性の疾患診断を受ける前、例えば対象が疾患を有することが疑われている段階で、または通常の臨床手順の間に行われ得る。

10

【0040】

リスクの判定はさらに、リスクを低発症リスクまたは中～高発症リスクに層別化する工程を含み得る。いくつかの例において、低リスクは、約50%～約80%の発症率であり得、かつ中～高リスクは、約80%～100%の発症率であり得る。

【0041】

リスクを低および中～高リスク群に正確に層別化することは、分析されるサンプルの約80%で行われ得る。リスクを層別化することは、細胞学的に不明瞭であるとまたは疑わしいと同定されたサンプルを含む、分析されるサンプルの約50%、60%、70%、75%、80%、85%、90%、95%、96%、97%、98%、または約99%において正確に決定され得る。リスクを低および中～高リスク群に層別化することは、少なくとも約80%の特異度を有し得る。いくつかの例において、リスクの層別化の特異度は、細胞学的に不明瞭であるとまたは疑わしいと同定されたサンプルを含む、約50%、60%、70%、75%、80%、85%、90%、95%、96%、97%、98%、99%、またはそれ以上であり得る。

20

【0042】

第1の遺伝子セット中または第2の遺伝子セット中の1つまたは複数の遺伝子は、複数の遺伝子、例えば約2、10、20、40個の遺伝子、またはそれ以上を含み得る。第1の遺伝子セットまたは第2の遺伝子セットの1つまたは複数の遺伝子は、約10個の遺伝子、20個の遺伝子、50個の遺伝子、60個の遺伝子、または約75個の遺伝子未満であり得る。第1のセットまたは第2のセットの1つまたは複数の遺伝子は、約50個～約400個の遺伝子であり得る。第1の遺伝子セットは、図6または図7の遺伝子を含み得る。第2の遺伝子セットは、図8の遺伝子を含み得る。

30

【0043】

第1の遺伝子セットおよび第2の遺伝子セットは、同じセットであり得る。例えば、第1および第2の遺伝子セットは、COL1A1、THBS2、またはそれらの任意の組み合わせを含み得る。

【0044】

第1の遺伝子セットおよび第2の遺伝子セットは、異なるセットであり得る。第2の遺伝子セットは、EPHA3、COL1A1、EHF、RAPGEF5、PRICKLE1、TMEM92、ROBO1、C6orf136、SPAG4、GALNT15、LUM、NCAM2、NUP210L、NR2F1、THBS2、PSORS1C1、またはそれらの任意の組み合わせを含み得る。第1の遺伝子セットは、COL1A1、TMEM92、C1orf87、SPAG4、EHF、COL3A1、GALNT15、NUP210L、PDZRN3、C6orf136、NA、NRXN3、COL6A3、RAPGEF5、PRICKLE1、LUM、ROBO1、BGN、AC019117.2、PRSS3P1、またはそれらの任意の組み合わせを含み得る。第2の遺伝子セットは、EPHA3、COL1A1、EHF、RAPGEF5、PRICKLE1、TMEM92、ROBO1、C6orf136、SPAG4、GALNT15、LUM、NCAM2、SYNP02、NUP210L、AMZ1、NR2F1、THBS2、PSORS1C1、FTH1P24、またはそれらの任意の組み合わせを含み得る。第2の遺伝子セットは、AKAP9、SPRY3、SPRY3、CAMKK2、COL1A1、FITM2、COX6C、VSIG10L、CYC1、KDM1B、MAPK15、ARSG、PAXIP1、DAAM1、AVL9、DMGDH、HLA-DQA1、HLA-DQB1、HLA-DRA、HLA-DRB5、HLA-H、IRF1、M

40

50

GAT1、P2RX1、PLEK、CCDC93、PPP1R12C、SLC41A3、METTL3、CCAR2、PTPRE、SRL、SLC30A5、BMP4、ZNF133、ICE2、DCAKD、TMX1、TNFSF12、PER2、MCM3AP、またはそれらの任意の組み合わせを含み得る。

【0045】

サンプル

対象から取得されるサンプルは、組織、細胞、細胞フラグメント、細胞オルガネラ、核酸、遺伝子、遺伝子フラグメント、発現産物、遺伝子発現産物、遺伝子発現産物フラグメント、またはそれらの任意の組み合わせを含み得る。サンプルは、不均質または均質であり得る。サンプルは、血液、尿、脳脊髄液、精液、唾液、痰、排泄物、リンパ液、組織、またはそれらの任意の組み合わせを含み得る。サンプルは、組織特異的サンプル、例えば甲状腺組織、皮膚、心臓、肺、腎臓、乳房、脾臓、肝臓、筋肉、平滑筋、膀胱、胆嚢、結腸、腸、脳、食道または前立腺から取得されるサンプルであり得る。

10

【0046】

本開示のサンプルは、様々な方法、例えば微細針吸引（FNA）、コア針生検、真空補助下生検、切開生検、切除生検、パンチ生検、薄片生検、皮膚生検、またはそれらの任意の組み合わせによって取得され得る。

【0047】

FNAは、微細針吸引生検（FNAB）または針吸引生検（NAB）とも称される、対象から少量の組織を取得する方法である。FNAは、取得のために対象の手術および入院を必要とし得る組織生検よりも低侵襲性であり得る。FNA法の針は、さらなる分析のためのサンプル量を取得するために対象の組織塊に挿入され得る。いくつかの例において、2つの針が、組織塊に挿入され得る。組織塊から取得されるFNAサンプルは、組織塊を貫く針の1回または複数回の通過によって獲得され得る。いくつかの例において、FNAサンプルは、約 6×10^6 、 5×10^6 、 4×10^6 、 3×10^6 、 2×10^6 、 1×10^6 個未満の細胞またはそれ未満を含み得る。針は、超音波または他の画像化装置によって組織塊へと案内され得る。針は、吸引もしくは真空または他の吸い込み技術による針を通じたFNAサンプルの回収が可能なように中空であり得る。

20

【0048】

本明細書において開示される方法を用いて取得されるサンプル、例えばFNAサンプルは、少ないサンプル容積を含み得る。サンプル容積は、約500マイクロリットル（uL）、400 uL、300 uL、200 uL、100 uL、75 uL、50 uL、25 uL、20 uL、15 uL、10 uL、5 uL、1 uL、0.5 uL、0.1 uL、0.01 uL未満またはそれ未満であり得る。サンプル容積は、約1 uL未満であり得る。サンプル容積は、約5 uL未満であり得る。サンプル容積は、約10 uL未満であり得る。サンプル容積は、約20 uL未満であり得る。サンプル容積は、約1 uL～約10 uLであり得る。サンプル容積は、約10 uL～約25 uLであり得る。

30

【0049】

本明細書において開示される方法を用いて取得されるサンプル、例えばFNAサンプルは、少ないサンプル重量を含み得る。サンプル重量、例えば組織重量は、約100ミリグラム（mg）、75 mg、50 mg、25 mg、20 mg、15 mg、10 mg、9 mg、8 mg、7 mg、6 mg、5 mg、4 mg、3 mg、2 mg、1 mg、0.5 mg、0.1 mg未満またはそれ未満であり得る。サンプル重量は、約20 mg未満であり得る。サンプル重量は、約10 mg未満であり得る。サンプル重量は、約5 mg未満であり得る。サンプル重量は、約5 mg～約20 mgであり得る。サンプル重量は、約1 mg～約5 ngであり得る。

40

【0050】

本明細書において開示される方法を用いて取得されるサンプル、例えばFNAは、少数の細胞を含み得る。単一サンプルの細胞数は、約 10×10^6 、 5.5×10^6 、 5×10^6 、 4.5×10^6 、 4×10^6 、 3.5×10^6 、 3×10^6 、 2.5×10^6 、 2×10^6 、 1.5×10^6 、 1×10^6 、 0.5×10^6 、 0.2×10^6 、 0.1×10^6 細胞未満またはそれ未満であり得る。単一サンプルの細胞数は、約 5×10^6 細胞未満であり得る。単一サンプルの細胞数は、約 4×10^6 細胞未満であり得る。単一サンプルの細胞数は、約 3×10^6 細胞未満であり得る。単一サンプルの細胞数は、約 2×10^6 細胞未満

50

であり得る。単一サンプルの細胞数は、約 1×10^6 ～約 5×10^6 細胞であり得る。単一サンプルの細胞数は、約 1×10^6 ～約 10×10^6 細胞であり得る。

【0051】

本明細書において開示される方法を用いて取得されるサンプル、例えばFNAは、少量のデオキシリボ核酸(DNA)またはリボ核酸(RNA)を含み得る。個々のサンプル中のDNAまたはRNAの量は、約500ナノグラム/ng)、400 ng、300 ng、200 ng、100 ng、75 ng、50 ng、45 ng、40 ng、35 ng、30 ng、25 ng、20 ng、15 ng、10 ng、5 ng、1 ng、0.5 ng、0.1 ng未満、またはそれ未満であり得る。DNAまたはRNAの量は、約40 ng未満であり得る。DNAまたはRNAの量は、約25 ng未満であり得る。DNAまたはRNAの量は、約15 ng未満であり得る。DNAまたはRNAの量は、約1 ng～約25 ngであり得る。DNAまたはRNAの量は、約5 ng～約50 ngであり得る。10

【0052】

サンプルのRNA収量またはRNA量は、ナノグラムからマイクログラムの量で測定され得る。実験室で核酸収量を測定するのに使用され得る装置の例は、NANODROP(登録商標)分光光度計、QUBIT(登録商標)蛍光光度計またはQUANTUS(商標)蛍光光度計である。NANODROP(登録商標)測定の精度は、非常に低いRNA濃度で大きく低下し得る。本明細書に記載される方法により取得されるデータの質は、RNA量に依存的であり得る。有意義な遺伝子発現または配列変種のデータ等は、NANODROP(登録商標)によって測定される場合に低いまたは測定不能なRNA濃度を有するサンプルから生成され得る。いくつかの例において、遺伝子発現または配列変種のデータ等は、測定不能なRNA濃度を有するサンプルから生成され得る。20

【0053】

本明細書に記載される方法は、少量または低品質のポリヌクレオチド、例えばDNAまたはRNAを含むサンプルを用いて実施され得る。少量または低品質のRNAを含むサンプルは、例えば、分解または部分分解した組織サンプルであり得る。少量または低品質のRNAを含むサンプルは、微細針吸引(FNA)サンプルであり得る。サンプルのRNAの質は、算出されたRNA Integrity Number(RIN)値によって測定され得る。RIN値は、RNA測定に完全性の値を割り当てるアルゴリズムである。このアルゴリズムは、1～10のRIN値を割り当てることができ、10のRIN値は完全にインタクトなRNAであり得る。RNAを含む本明細書に記載されるサンプルは、約9.0、8.0、7.0、6.0、5.0、4.0、3.0、2.0、1.0、またはそれ未満のRIN値を有し得る。いくつかの例において、RNAを含むサンプルは、約8.0またはそれ未満のRIN値を有し得る。いくつかの例において、RNAを含むサンプルは、約6.0またはそれ未満のRIN値を有し得る。いくつかの例において、RNAを含むサンプルは、約4.0またはそれ未満のRIN値を有し得る。いくつかの例において、サンプルは、約2.0またはそれ未満のRIN値を有し得る。30

【0054】

サンプル、例えばFNAサンプルは、別の個人または主体、例えば保健(もしくは医療)専門家またはロボットによって対象から取得され得る。医療専門家は、医師、看護師、医療技術者等を含み得る。いくつかの例において、医師は、がん専門医、外科医または内分泌科医等の専門家であり得る。医療技術者は、細胞学者、鴉血専門医、放射線科医、呼吸器科医等の専門家であり得る。医療専門家は、試験のために対象からサンプルを取得し得るまたはサンプルの提出のために対象を試験センターまたは研究所に照会し得る。医療専門家は、試験センターまたは研究所に、サンプルに対して適切な試験またはアッセイ法、例えば、遺伝子配列データ、遺伝子発現レベル、配列変種データ、またはそれらの任意の組み合わせの決定を含む本開示の方法を実施することを指示し得る。40

【0055】

いくつかの例において、医療専門家は、疾患の初期診断または初期サンプル取得に関与している必要はない。その代わり、個体、例えば対象が、市販のキットの使用を通じてサンプルを取得し得る。キットは、本明細書に記載されるようにサンプルを取得するための回収ユニットまたはデバイス、サンプル分析の前にサンプルを保管するための保管ユニッ50

トおよびそのキットの使用説明書を含み得る。

【0056】

サンプルは、(a)手術前に、(b)手術後に、(c)がんの診断後に、(d)疾患の寛解もしくは治癒後の通常のスクリーニング後に、(e)対象が疾患を有することが疑われる場合に、(f)通常の通院または臨床スクリーニング中に、(g)臨床専門家の要望により、またはそれらの任意の組み合わせで取得され得る。別の時点の複数のサンプルが、例えば疾患処置開始前および処置終了後に、例えば経時的な対象のモニタリングのために、同じ対象から取得され得る。複数のサンプルは、対象における疾患の進行、退行または寛解の非存在または存在をモニタリングするために、別の時点で対象から取得され得る。

【0057】

細胞学的分析

疾患の発症のリスクの評価を含む、本明細書に記載される方法は、サンプルの細胞学的分析を含み得る。細胞学的分析の例は、エオシン・アズール(EA)染色、ヘマトキシリン染色、CYTO-STAIN(商標)、パパニコロウ染色、エオシン、ニッスル染色、トルイジンブルー、銀染色、アゾカルミン染色、ニュートラルレッドまたはヤヌスグリーンを含むがこれらに限定されない任意の多くの方法および適切な試薬によって実施される細胞染色技術および/または顕微鏡試験を含む。2つ以上の染色が、他の染色と組み合わせて使用され得る。いくつかの例において、細胞は、全く染色されない。細胞は、染色手順の前にまたはその中で、例えばメタノール、エタノール、グルタルアルデヒドまたはホルムアルデヒドを用いて固定および/または透過処理され得る。いくつかの例において、細胞は、固定されない場合がある。染色手順はまた、例えば臭化エチジウム、ヘマトキシリン、ニッスル染色または任意の他の核酸染色を用いて、サンプルの核酸量を測定するために使用され得る。

【0058】

サンプル中の細胞の顕微鏡試験は、細胞学的試験の標準的方法によって細胞をスライド上に塗りつけることを含み得る。液体細胞学(LBC)法が用いられ得る。いくつかの例において、LBC法は、細胞学的スライドの調製、より均質なサンプル、向上した感度および特異度、もしくは改善されたサンプル操作性、またはそれらの任意の組み合わせの改善されたアプローチを提供する。LBC法において、サンプルは、対象から、LBC調製溶液、例えばCYTYC THINPREP(登録商標)、SUREPATH(商標)もしくはMONOPREP(登録商標)または任意の他のLBC調製溶液を含む容器またはバイアルに移され得る。加えて、サンプルは、サンプルの実質量移動を確実にするために、LBC調製溶液を用いて回収デバイスから容器またはバイアルに洗い流され得る。LBC調製溶液中にサンプルを含む溶液は、その後、保管され得、かつ/またはガラススライド上に細胞の層を形成するよう機械もしくは当業者によって処理され得る。サンプルはさらに、従来の細胞学的調製物と同じ様式で染色され、顕微鏡下で試験され得る。

【0059】

サンプルは、免疫組織化学染色によって分析され得る。免疫組織化学染色は、サンプル(例えば細胞または組織)中での抗体の使用により特定の分子または抗原の存在、位置、および分布の分析を提供し得る。抗原は、抗体によって特異的に認識されることができる低分子、タンパク質、ペプチド、核酸または任意の他の分子であり得る。サンプルは、事前の固定および/または透過処理工程を用いるまたは用いない免疫組織化学法によって分析され得る。いくつかの例において、関心対象の抗原は、サンプルを抗原に特異的な抗体と接触させることによって検出され得、次いで非特異的な結合が、1回または複数回の洗浄によって除去され得る。特異的に結合した抗体は、次いで、抗体検出試薬、例えば標識された二次抗体または標識されたアビジン/ストレプトアビジンによって検出され得る。抗原特異的な抗体は、直接的に標識され得る。免疫組織化学に適した標識は、フルオロフオア、例えばフルオレセインおよびローダミン、酵素、例えばアルカリホスファターゼおよび西洋ワサビペルオキシダーゼ、または放射性核種、例えば³²Pおよび¹²⁵Iを含むがこれらに限定されない。免疫組織化学染色によって検出され得る遺伝子産物マークターは、He

10

20

30

40

50

r2/Neu、Ras、Rho、EGFR、VEGFR、UbcH10、RET/PTC1、サイトケラチン20、カルシトニン、GAL-3、甲状腺ペルオキシダーゼ、またはサイログロブリンを含むがこれらに限定されない。

【0060】

本明細書において開示される疾患発症のリスクに関するメトリクス、例えば、第1の遺伝子セットの遺伝子発現レベル、または第2の遺伝子セットの配列変種データは、疾患発症のリスクを有することが判明しているサンプルのあらゆる細胞の特徴である必要はない。したがって、本明細書において開示される方法は、遺伝子発現レベルもしくは配列変種データに関するまたは疾患の発症のリスクを示す他のデータに関する完全なパターンをサンプル中のすべてとは言えない細胞が示す組織における疾患、例えばがんの発症のリスクを評価するのに有用であり得る。遺伝子発現レベル、配列変種データ等は、サンプルの影響を受ける細胞および影響を受けない細胞内に完全に存在するか、部分的に存在するか、または存在しないかのいずれかであり得る。遺伝子発現レベル、配列変種データ等は、影響を受ける細胞内に様々な量で存在し得る。遺伝子発現レベル、配列変種データ等は、影響を受けない細胞内に様々な量で存在し得る。いくつかの例において、疾患発症のリスクに相関する第1の遺伝子セットの遺伝子発現レベルまたは第2の遺伝子セットにおける1つもしくは複数の配列変種の存在は、陽性検出され得る。いくつかの例において、陽性検出は、サンプルから抽出される細胞の少なくとも70%、75%、80%、85%、90%、95%、または100%において見られ得る。いくつかの例において、第1の遺伝子セットの遺伝子発現レベル、または第2の遺伝子セットにおける1つもしくは複数の配列変種の存在は、非存在であり得る。いくつかの例において、検出の非存在は、対応する正常、非疾患サンプルの細胞の少なくとも70%、75%、80%、85%、90%、95%、または100%において見られ得る。

10

20

30

40

50

【0061】

日常的な細胞学的アッセイ法または他のアッセイ法は、サンプルが、陰性（疾患なし）、診断（疾患、例えばがんについて陽性の診断）、不明瞭であるかもしくは疑わしい（疾患、例えばがんの存在が示唆される）または非診断（疾患の存在または非存在に関する情報を十分に提供しない）であることを示し得る。本明細書において記載される方法は、日常的な細胞学的評価からの結果を確認し得るかまたは日常的な細胞学的評価の非存在下でそれと同等の独自の評価を提供し得る。本明細書に記載される方法は、不明瞭であるまたは疑わしいと判明しているサンプルを含むサンプルを、悪性または良性に分類し得る。この方法はさらに、不明瞭であるまたは疑わしいと判明しているサンプルを含むサンプル、例えば悪性であることが分かっているサンプルを、低疾患発症リスク群および中～高疾患発症リスク群に層別化し得る。

【0062】

疾患

疾患は、本明細書で開示される場合、甲状腺がんを含み得る。甲状腺がんは、甲状腺の任意の悪性腫瘍、例えば、甲状腺乳頭がん（PTC）、濾胞性甲状腺がん（FTC）、濾胞型甲状腺乳頭がん（FVPTC）、甲状腺髓様がん（MTC）、濾胞がん（FC）、ハースル細胞がん（HC）、および／または未分化甲状腺がん（ATC）を含むがこれらに限定されない甲状腺がんの任意のサブタイプを含み得る。いくつかの例において、甲状腺がんは、分化型であり得る。いくつかの例において、甲状腺がんは、未分化型であり得る。

【0063】

甲状腺組織サンプルは、本開示の方法を用いて、濾胞性腺腫（FA）、結節性過形成（NHP）、リンパ球性甲状腺炎（LCT）およびハースル細胞腺腫（HA）、濾胞がん（FC）、甲状腺乳頭がん（PTC）、濾胞型乳頭がん（FVPTC）、甲状腺髓様がん（MTC）、ハースル細胞がん（HC）および未分化甲状腺がん（ATC）、腎がん（RCC）、乳がん（BCA）、黒色腫（MEN）、B細胞リンパ腫（BCL）または副甲状腺（PTA）を含むがこれらに限定されない1つまたは複数の良性または悪性組織タイプ（例えば、がんのサブタイプ）を含むものと分類され得る。

【0064】

本開示のがんの他のタイプは、副腎皮質がん、肛門がん、再生不良性貧血、胆管がん、膀胱がん、骨がん、骨転移、中枢神経系（CNS）がん、末梢神経系（PNS）がん、乳がん、キャッスルマン病、子宮頸がん、小児非ホジキンリンパ腫、リンパ腫、結腸直腸がん、子宮内膜がん、食道がん、ユーイング肉腫ファミリー腫瘍（Ewing's family of tumor）（例えば、ユーイング肉腫）、眼がん、胆嚢がん、胃腸カルチノイド腫瘍、胃腸間質腫瘍、妊娠性トロホblast疾患、ヘアリーセル白血病、ホジキン病、カポジ肉腫、腎臓がん、喉頭および下咽頭がん、急性リンパ性白血病、急性骨髓性白血病、小児白血病、慢性リンパ性白血病、慢性骨髓性白血病、肝臓がん、肺がん、肺カルチノイド腫瘍、非ホジキンリンパ腫、男性乳がん、悪性中皮腫、多発性骨髓腫、骨髓異形成症候群、骨髓増殖性疾患、鼻腔および副鼻腔がん、鼻咽腔がん、神経芽腫、口腔および中咽頭がん、骨肉腫、卵巣がん、膵臓がん、陰茎がん、下垂体腫瘍、前立腺がん、網膜芽細胞腫、横紋筋肉腫、唾液腺がん、肉腫（成人軟部組織がん）、黒色腫皮膚がん、非黒色腫皮膚がん、胃がん、精巣がん、胸腺がん、子宮がん（例えば、子宮肉腫）、腫がん、外陰がん、またはワルデンシュトレーームマクログロブリン血症を含み得るがこれらに限定されない。

10

【0065】

疾患は、本明細書で開示される場合、過剰増殖性障害を含み得る。悪性過剰増殖性障害は、リスク群、例えば、低リスク群および中～高リスク群に層別化され得る。過剰増殖性障害は、がん、過形成または新生物を含み得るが、これらに限定されない。いくつかの例において、過剰増殖性がんは、乳がん、例えば、乳腺の管組織における腺管がん、髓様がん、膠様がん、管状がんおよび炎症性乳がん；卵巣上皮腫瘍、例えば、卵巣における腺がんおよび卵巣から腹腔へ移動した腺がんを含む卵巣がん；子宮がん；子宮頸がん、例えば扁平上皮がんおよび腺がんを含む腹部頸部における腺がん；前立腺がん、例えば、腺がんまたは骨へ移動した腺がんから選択される前立腺がん；膵臓がん、例えば、膵管組織における類上皮細胞がんおよび膵管における腺がん；膀胱がん、例えば、膀胱における移行上皮がん、尿路上皮がん（移行上皮がん）、膀胱を裏打ちする尿路上皮細胞における腫瘍、扁平上皮がん、腺がん、および小細胞がんのような膀胱がん；白血病、例えば、急性骨髓性白血病（AML）、急性リンパ性白血病、慢性リンパ性白血病、慢性骨髓性白血病、ヘアリーセル白血病、脊髄形成異常症、骨髓増殖性障害、急性骨髓性白血病（AML）、慢性骨髓性白血病（CML）、肥満細胞症、慢性リンパ性白血病（CLL）、多発性骨髓腫（MM）および骨髓異形成症候群（MDS）；骨がん；肺がん、例えば、扁平上皮がん、腺がんおよび大細胞未分化がんに細分される非小細胞肺がん（NSCLC）ならびに小細胞肺がん；皮膚がん、例えば、基底細胞がん、黒色腫、扁平上皮がん、および扁平上皮がんへと進展する場合がある皮膚状態である日光角化症；眼網膜芽細胞腫；皮膚または眼球内（眼）黒色腫；原発性肝臓がん（肝臓において発生するがん）；腎臓がん；自己免疫不全症候群（AIDS）関連リンパ腫、例えば、びまん性大細胞型B細胞リンパ腫、B細胞免疫芽球性リンパ腫および小型非切れ込み核細胞性リンパ腫；カポジ肉腫；B型肝炎ウイルス（HBV）、C型肝炎ウイルス（HCV）および肝細胞がんを含むウイルスにより誘導されるがん；ヒトリンパ球向性ウイルス1型（HTLV-1）および成人T細胞白血病／リンパ腫；ならびにヒトパピローマウイルス（HPV）および子宮頸がん；中枢神経系（CNS）がん、例えば、神経膠腫（星状細胞腫、未分化星状細胞腫または多形神経膠芽腫）、乏突起細胞腫、上衣細胞腫、髓膜腫、リンパ腫、シュワン腫および髓芽腫を含む原発性脳腫瘍；末梢神経系（PNS）がん、例えば、聴神経腫瘍、ならびに神経纖維腫およびシュワン腫を含む悪性末梢神経鞘腫瘍（MPNST）、悪性線維性細胞腫、悪性線維性組織球腫、悪性髓膜腫、悪性中皮腫、ならびに悪性ミュラー管混合腫瘍；口腔および中咽頭がん、例えば、下咽頭がん、喉頭がん、上咽頭がん、および中咽頭がん；胃がん、例えば、リンパ腫、胃間質腫瘍およびカルチノイド腫瘍；精巣がん、例えば、精上皮腫および非精巣上皮腫を含む胚細胞性腫瘍（GCT）ならびにライディッヒ細胞腫瘍およびセルトリ細胞腫瘍を含む性腺間質腫瘍；胸腺がん、例えば、胸腺腫、胸腺がん、ホジキン病、非ホジキンリンパ腫カルチノイドまたはカルチノイド腫瘍；直腸がん；ならびに結腸がんであり得る。いくつかの例において、本開示の方法によって

20

30

40

50

層別化、分類、特徴決定、または診断される疾患は、濾胞性腺腫、ハースル細胞腺腫、リンパ性甲状腺炎および甲状腺過形成を含むがこれらに限定されない、甲状腺障害、例えば良性甲状腺障害を含むがこれらに限定されない。いくつかの例において、本開示の方法によって層別化、分類、特徴決定、または診断される疾患は、悪性甲状腺障害、例えば濾胞がん、濾胞型甲状腺乳頭がん、髓様がん、および乳頭がんを含むがこれらに限定されない。

【0066】

本開示の疾患は、遺伝性障害を含み得る。遺伝性障害は、遺伝子または染色体の異常にによって引き起こされる疾病である。遺伝性障害は、単一遺伝子障害および多因子性多遺伝子性（複合）障害という2つのカテゴリーに分類され得る。単一遺伝子障害は、単一の変異型遺伝子の結果であり得る。単一遺伝子障害の遺伝は、常染色体優性、常染色体劣性、X連鎖優性、X連鎖劣性、Y連鎖遺伝およびミトコンドリア遺伝を含み得るがこれらに限定されない。人が常染色体優性障害による影響を受けるには、その遺伝子の一つの変異型コピーのみで十分であり得る。常染色体優性型の障害の例には、ハンチントン病、神経纖維腫症1型、マルファン症候群、遺伝性非ポリポーシス結腸直腸がんまたは遺伝性多発性外骨腫を含み得るがこれらに限定されない。常染色体劣性障害においては、人が常染色体劣性障害による影響を受けるには、遺伝子の2つのコピーが変異していかなければならない。このタイプの障害の例は、囊胞性纖維症、鎌型赤血球症（部分鎌型赤血球症とも言う）、ティ・サックス病、ニーマン・ピック病または脊髄筋萎縮症を含み得るがこれらに限定されない。X連鎖優性障害、例えばX連鎖低リン血症性くる病は、X染色体上の遺伝子の変異により引き起こされる。いくつかのX連鎖優性状態、例えばレット症候群、色素失調症2型およびアイカルディ症候群は致死的であり得る。X連鎖劣性障害も、X染色体上の遺伝子の変異により引き起こされる。このタイプの障害の例は、血友病A、デュシェンヌ型筋ジストロフィー、赤緑色盲、筋ジストロフィーおよび男性型脱毛症を含み得るがこれらに限定されない。Y連鎖障害は、Y染色体上の変異により引き起こされる。例は、男性不妊および耳介多毛症（hypertrichosis pinnae）を含み得るがこれらに限定されない。母系遺伝としても公知のミトコンドリア遺伝の遺伝性障害、例えばレーバー遺伝性視神経萎縮症におけるそれは、ミトコンドリアDNA内の遺伝子において起こり得る。

【0067】

遺伝性障害は、複合型、多因子性、または多遺伝子性もあり得る。多遺伝子性遺伝性障害は、生活様式および環境因子と共に複数の遺伝子の影響に関連し得る。複合型遺伝性障害は家族内に集中するが、明確な遺伝パターンを有していない。多因子性または多遺伝子性の障害は、心疾患、糖尿病、喘息、自閉症、自己免疫疾患、例えば多発性硬化症、がん、纖毛病、口蓋裂、高血圧、炎症性腸疾患、精神遅滞または肥満を含み得る。

【0068】

他の遺伝性障害は、1p36欠失症候群、21水酸化酵素欠損症、22q11.2欠失症候群、無セルロプラスミン血症、軟骨無発生症II型、軟骨無形成症、急性間欠性ポルフィリン症、アデニロコハク酸リアーゼ欠損症、副腎白質ジストロフィー、アレキサンダー病、アルカブトン尿症、-1アンチトリプシン欠損症、アルストレーム症候群、アルツハイマー病（1型、2型、3型、および4型）、エナメル質形成不全、筋萎縮性側索硬化症、筋萎縮性側索硬化症2型、筋萎縮性側索硬化症4型、筋萎縮性側索硬化症4型、アンドロゲン不応症、貧血、アンジェルマン症候群、アペール症候群、毛細血管拡張性運動失調、ベアレ・ステイブンソン脳回状頭皮（Beare-Stevenson cutis gyrata）症候群、ベンジャミン症候群、

サラセミア、ビオチミダーゼ（biotimidase）欠損症、パート・ホッグ・デューべ症候群、膀胱がん、ブルーム症候群、骨疾患、乳がん、屈曲肢異形成症、カナバン病、がん、セリック病、慢性肉芽腫性障害（CGD）、シャルコー・マリー・トゥース病、シャルコー・マリー・トゥース病1型、シャルコー・マリー・トゥース病4型、シャルコー・マリー・トゥース病2型、シャルコー・マリー・トゥース病4型、コケーン症候群、コフィン・ローリー症候群、コラゲノパシー（collagenopathy）II型およびXI型、結腸直腸がん、先天性精管欠損症、先天性両側精管欠損症、先天性糖尿病、先天性赤血球生成性ポルフィリン症

10

20

30

40

50

、先天性心疾患、先天性甲状腺機能低下症、結合組織病、カウデン症候群、猫泣き、クローン病、線維性狭窄 (fibrostenosing) 、クルーゾン症候群、クルーゾン皮膚骨格 (Crouzonodermoskeletal) 症候群、囊胞性纖維症、ド・グルーシー (De Grouchy) 症候群、神経変性疾患、デント病、発達障害、ディジョージ症候群、遠位型脊髄性筋萎縮症V型、ダウン症候群、低身長症、エーラース・ダンロス症候群、エーラース・ダンロス症候群関節弛緩型、エーラース・ダンロス症候群古典型、エーラース・ダンロス症候群皮膚弛緩型、エーラース・ダンロス症候群後側彎型、血管型、骨髄性プロトポルフィリン症、ファブリ病、顔面損傷および顔面障害、第V因子ライデン栓友病、家族性大腸ポリポーシス、家族性自律神経異常症、ファンコニー貧血、FG症候群、脆弱X症候群、フリードライヒ運動失調症 (Friedreich ataxia) 、フリードライヒ運動失調症 (Friedreich's ataxia) 、G6PD欠損症、ガラクトース血症、ゴーシェ病 (1型、2型、および3型) 、遺伝性脳障害、グリシン脳症、ヘモクロマトーシス2型、ヘモクロマトーシス4型、道化師様魚鱗癬、頭部および脳の奇形、聴覚障害および聴覚消失、小児聴覚障害、ヘモクロマトーシス (新生児、2型、および3型) 、血友病、骨髄肝性ポルフィリン症、遺伝性コプロポルフィリン症、遺伝性多発性外骨腫症、遺伝性圧脆弱性ニューロパチー、遺伝性非ポリポーシス結腸直腸がん、ホモシスチン尿症、ハンチントン病、ハッチンソン・ギルフォード・プロジェクトニア症候群、原発性高シュウ酸尿症、高フェニルアラニン血症、軟骨低発生症、軟骨低形成症、idic15、色素失調症、乳児ゴーシェ病、乳児発症上行性遺伝性痙性麻痺、不妊症、ジャクソン・ワイス症候群、ジュベール症候群、若年型原発性側索硬化症、ケネディ病、クラインフェルター症候群、クニースト骨異形成症、クラッペ病、学習障害、レッシュ・ナイハン症候群、白質ジストロフィー、リー・フラウメニ症候群、家族性リポタンパク質リバーゼ欠損症、男性生殖器障害、マルファン症候群、マクキューン・オールブライト症候群、マクロード症候群、家族性地中海熱、メンケス病、メンケス症候群、代謝障害、グロビン型メトヘモグロビン血症、メトヘモグロビン血症、先天性メトヘモグロビン血症、メチルマロン酸血症、ミクロ (Micro) 症候群、小頭症、運動障害、モワット・ウィルソン症候群、ムコ多糖症 (MPS I) 、ムンケ (Muenke) 症候群、筋ジストロフィー、デュシェンヌ型筋ジストロフィーおよびベッカー型筋ジストロフィー、デュシェンヌ型筋ジストロフィーおよびベッカー型筋ジストロフィー、筋緊張性ジストロフィー、筋緊張性ジストロフィー1型および2型、新生児ヘモクロマトーシス、神経線維腫症、神経線維腫症1型、神経線維腫症2型、神経線維腫症I型、神経線維腫症II型、神経疾患、神経筋障害、ニーマン・ピック病、非ケトーシス型高グリシン血症、非症候性難聴、常染色体劣性非症候性難聴、ヌーナン症候群、骨形成不全症 (I型およびIII型) 、耳脊椎巨大骨端異形成症、パントテン酸キナーゼ関連神経変性症、パトー症候群 (13トリソミー) 、ペンドレッド症候群、ポイツ・ジェガース症候群、ブファイファー症候群、フェニルケトン尿症、ポルフィリン症、晚発性皮膚ポルフィリン症、プラダー・ウィリ症候群、原発性肺高血圧症、ブリオン病、早老症、プロピオン酸血症、プロテインC欠損症、プロテインS欠損症、偽性ゴーシェ (pseudo-Gaucher) 病、弾力線維性仮性黄色腫、網膜障害、網膜芽腫、網膜芽腫、FA - フリートライヒ運動失調症、レット症候群、ルビンシュタイン・テイビ症候群、サンドホフ病、感覚性自律神経性ニューロパチーIII型、鎌状赤血球貧血、骨格筋再生、皮膚色素異常症、スミス・レムリ・オピツ症候群、発話障害およびコミュニケーション障害、脊髄性筋萎縮症、球脊髄性筋萎縮症、脊髄小脳変性症、ストラドヴィック型脊椎骨端骨幹端異形成、先天性脊椎骨端骨幹端異形成、スティックラー症候群、スティックラー症候群COL2A1、ティ・サックス病、テトラヒドロビオブテリン欠損症、致死性骨異形成、糖尿病および感音難聴を伴うチアミン反応性巨赤芽球性貧血、甲状腺疾患、トウレット症候群、トリーチャー・コリンズ症候群、トリプルX症候群、結節性硬化症、ターナー症候群、アッシャー症候群、異型ポルフィリン症、フォンヒッペル・リンダウ病、ワールデンブルグ症候群、ワイセンバッハ・ツウェイミュラー (Weissenbacher-Zweymuller) 症候群、ウイルソン病、ウォルフ・ヒルシュホーン症候群、色素性乾皮症、X連鎖重症複合免疫不全、X連鎖鉄芽球性貧血、ならびにX連鎖球脊髄性筋萎縮症を含み得るがこれらに限定されない。

発症または再発のリスクの層別化

疾患の発症のリスクは、サンプルを、下位リスク群に層別化し得る。下位群は、低疾患発症性リスクを有するサンプルおよび中～高疾患発症性リスクを有するサンプルを含み得る。下位群は、低リスク、中リスクおよび高リスク群を含み得る。低リスクは、約1%、5%、10%、15%、20%、25%、30%、35%、40%、または約45%の疾患発症性リスクを有するサンプルを含み得る。低リスクは、約1%～約25%の疾患発症性リスクを有するサンプルを含み得る。低リスクは、約1%～約30%の疾患発症性リスクを有するサンプルを含み得る。低リスクは、約1%～約40%の疾患発症性リスクを有するサンプルを含み得る。中～高リスクは、約55%、60%、65%、70%、75%、80%、85%、90%、95%、または100%の疾患発症性リスクを有するサンプルを含み得る。中～高リスクは、約50%～約100%の疾患発症性リスクを有するサンプルを含み得る。中～高リスクは、約55%～約100%の疾患発症性リスクを有するサンプルを含み得る。中～高リスクは、約60%～約100%の疾患発症性リスクを有するサンプルを含み得る。

10

【0070】

サンプルは、細胞学的に不明瞭であるとまたは疑わしいとまたは不確定であると同定されたサンプルを含めて、少なくとも50%、60%、70%、75%、80%、85%、90%、95%、96%、97%、98%、99%、またはそれ以上の精度で低リスクまたは中～高リスク群に層別化され得る。サンプルは、少なくとも70%の精度で層別化され得る。サンプルは、少なくとも80%の精度で層別化され得る。サンプルは、少なくとも90%の精度で層別化され得る。サンプルは、細胞学的に不明瞭であるとまたは疑わしいとまたは不確定であると同定されたサンプルを含めて、50%、60%、70%、75%、80%、85%、90%、95%、96%、97%、98%、99%超、またはそれ以上の精度で、良性、悪性または非診断と同定され得る。精度は、分類器を用いて算出され得る。

20

【0071】

サンプルは、細胞学的に不明瞭であるとまたは疑わしいとまたは不確定であると同定されたサンプルを含めて、少なくとも50%、60%、70%、75%、80%、85%、90%、95%、96%、97%、98%、99%、またはそれ以上の特異度で低リスクまたは中～高リスク群に層別化され得る。サンプルは、少なくとも70%の精度で層別化され得る。サンプルは、少なくとも80%の精度で層別化され得る。サンプルは、少なくとも90%の精度で層別化され得る。サンプルは、細胞学的に不明瞭であるとまたは疑わしいとまたは不確定であると同定されたサンプルを含めて、50%、60%、70%、75%、80%、85%、90%、95%、96%、97%、98%、99%超、またはそれ以上の特異度で、良性、悪性、または非診断と同定され得る。特異度は、分類器を用いて算出され得る。

30

【0072】

本明細書に記載される疾患発症のリスクを層別化し、サンプルを良性、悪性、または非診断に分類するための方法は、少なくとも95%、95.5%、96%、96.5%、97%、97.5%、98%、98.5%、99%、99.5%、もしくはそれ以上の陽性的中率および/または少なくとも95%、95.5%、96%、96.5%、97%、97.5%、98%、98.5%、99%、99.5%、もしくはそれ以上の陰性的中率を有し得る。疾患の陽性的中率(PPV)または適合率または検査後確率は、正しく診断されるまたは正しくリスク群に層別化される陽性試験結果を有する対象の比率であり得る。それは、試験された根底にある疾患を陽性試験が反映している蓋然性を反映し得るものなので、重要な尺度であり得る。その値は、疾患の有病率に依存し得、これは様々であり得る。陰性的中率(NPV)は、正確に診断された陰性試験結果を有する対象の比率であり得る。PPVおよびNPVの測定は、適切な疾患サブタイプの有病率の見積もりを用いて行われ得る。サブタイプ特異的な見積もりにおいて、使用可能なサンプルが存在しない可能性があるという理由のため、疾患有病率は場合によっては算出不可能であり得る。

40

【0073】

サンプルは、以下の1つまたは複数に分類され得る：良性（疾患なし）、悪性（疾患の陽性診断）、または非診断（疾患の存在または非存在に関して十分な情報を提供しない）

50

。悪性であることが判明しているサンプルは、低疾患発症リスクまたは中～高疾患発症リスク等の疾患発症リスクに層別化され得る。サンプルは、良性か疑わしい（疾患に関して陽性であることが疑われる）のカテゴリーに分類され得る。サンプルはさらに、例えば1つまたは複数のがんサブタイプの存在または非存在を同定することによって、疾患サブタイプに分類され得る。ある特定の分子経路が疾患に関与することが示され得、または特定の疾患のある特定のグレードもしくは病期（例えばI、II、III、もしくはIV期がん）もまた示され得る。いくつかの例において、層別化された発症リスクは、適切な治療的介入、例えば特定の薬物計画、または甲状腺摘出術もしくは準甲状腺摘出術等の外科的介入を通知し得る。

【0074】

10

本開示の分類器または訓練されたアルゴリズムは、サンプルを低もしくは中～高リスク群に層別化するためおよび／またはサンプルを良性、悪性、疑わしい、もしくは非診断に分類するため等に使用され得る。1つまたは複数の選択された特徴空間、例えば遺伝子発現レベルおよび配列変種のデータは、単独でまたは組み合わせて分類器もしくは訓練されたアルゴリズムに提供され得る。例示的なアルゴリズムは、変数の数を減らす方法、例えば主成分分析アルゴリズム、部分最小二乗法または独立成分分析アルゴリズムを含み得るがこれらに限定されない。例示的なアルゴリズムは、多数の変数を直接扱う方法、例えば統計学的方法または機械学習技術に基づく方法を含み得る。統計学的方法は、罰則付き（penalized）ロジスティック回帰、マイクロアレイ予測分析（prediction analysis of microarray）（PAM）、収縮重心に基づく方法、サポートベクターマシン分析、または正則化線形判別分析（regularized linear discriminant analysis）を含み得る。機械学習技術は、バギング（bagging）法、ブースティング（boosting）法、ランダムフォレストアルゴリズム、またはそれらの組み合わせを含み得る。

20

【0075】

本開示の分類器または訓練されたアルゴリズムは、2つまたはそれ以上の特徴空間を含み得る。2つまたはそれ以上の特徴空間は、特有であり得るかまたは互いと異なり得る。個々の特徴空間は、サンプルについての情報タイプ、例えば遺伝子発現レベルデータまたは配列変種データを含み得る。分類器において2つまたはそれ以上の特徴空間を組み合わせることで、単一の特徴空間を用いてリスク層別化を行うよりも高レベルの精度でリスク層別化または分類を行うことができる。個々の特徴空間のダイナミックレンジは、異なり得る、例えば少なくとも1または2オーダー異なり得る。例えば、遺伝子発現レベルの特徴空間のダイナミックレンジは、0～約300であり得、配列変種の特徴空間のダイナミックレンジは、0～約20であり得る。

30

【0076】

個々の特徴空間は、遺伝子セット、例えば第1の特徴空間の第1の遺伝子セットおよび第2の特徴空間の第2の遺伝子セットを含み得る。個々の特徴空間の遺伝子セットは、リスクの発症のリスクに関連し得る。第1の遺伝子セットおよび第2の遺伝子セットは、同一セットであり得る。第1の遺伝子セットおよび第2の遺伝子セットは、異なるセットであり得る。第1の遺伝子セットまたは第2の遺伝子セットは、約1000、500、400、300、200、100、75、70、65、60、55、50、45、40、35、30、25、20、15、10、5個未満、またはそれ未満の遺伝子を含み得る。第1の遺伝子セットまたは第2の遺伝子セットは、約10個未満の遺伝子を含み得る。第1の遺伝子セットまたは第2の遺伝子セットは、約50個未満の遺伝子を含み得る。第1の遺伝子セットまたは第2の遺伝子セットは、約75個未満の遺伝子を含み得る。第1の遺伝子セットまたは第2の遺伝子セットは、約50個～約400個の遺伝子を含み得る。第1の遺伝子セットまたは第2の遺伝子セットは、約50個～約200個の遺伝子を含み得る。第1の遺伝子セットまたは第2の遺伝子セットは、約10個～約600個の遺伝子を含み得る。

40

【0077】

第1の遺伝子セットは、図6に列挙されている遺伝子を含み得る。第1の遺伝子セットは、図7に列挙されている遺伝子を含み得る。第1の遺伝子セットは、COL1A1、THBS2、またはそれらの任意の組み合わせを含み得る。第1の遺伝子セットは、COL1A1、TMEM92、C

50

10rf87、SPAG4、EHF、COL3A1、GALNT15、NUP210L、PDZRN3、C6orf136、NA、NRXN3、COL6A3、RAPGEF5、PRICKLE1、LUM、ROBO1、BGN、AC019117.2、PRSS3P1、またはそれらの任意の組み合わせを含み得る。

【 0 0 7 8 】

第1の遺伝子セットは、図13に列挙されている遺伝子を含み得る。第1の遺伝子セットは、COL1A1、NUP210L、TMEM92、C6orf136、SPAG4、EHF、RAPGEF5、COL3A1、GALNT15、PRICKLE1、LUM、COL6A3、ROBO1、SSC5D、PSORS1C1、またはそれらの任意の組み合わせを含み得る。第1の遺伝子セットは、COL1A1、NUP210L、TMEM92、C6orf136、SPAG4、EHF、RAPGEF5、COL3A1、GALNT15、PRICKLE1、LUM、COL6A3、ROBO1、SSC5D、PSORS1C1、およびそれらの任意の組み合わせからなる群より選択され得る。第1の遺伝子セットは、COL1A1を含み得る。第1の遺伝子セットは、NUP210Lを含み得る。第1の遺伝子セットは、TMEM92を含み得る。第1の遺伝子セットは、C6orf136を含み得る。第1の遺伝子セットは、SPAG4を含み得る。第1の遺伝子セットは、EHFを含み得る。第1の遺伝子セットは、RAPGEF5を含み得る。第1の遺伝子セットは、COL3A1を含み得る。第1の遺伝子セットは、GALNT15を含み得る。第1の遺伝子セットは、PRICKLE1を含み得る。第1の遺伝子セットは、LUMを含み得る。第1の遺伝子セットは、COL6A3を含み得る。第1の遺伝子セットは、ROBO1を含み得る。第1の遺伝子セットは、SSC5Dを含み得る。第1の遺伝子セットは、PSORS1C1を含み得る。

10

【 0 0 7 9 】

第2の遺伝子セットは、図8に列挙されている遺伝子を含み得る。第2の遺伝子セットは、COL1A1、THBS2、またはそれらの任意の組み合わせを含み得る。第2の遺伝子セットは、EPHA3、COL1A1、EHF、RAPGEF5、PRICKLE1、TMEM92、ROBO1、C6orf136、SPAG4、GALNT15、LUM、NCAM2、NUP210L、NR2F1、THBS2、PSORS1C1、またはそれらの任意の組み合わせを含み得る。第2の遺伝子セットは、EPHA3、COL1A1、EHF、RAPGEF5、PRICKLE1、TMEM92、ROBO1、C6orf136、SPAG4、GALNT15、LUM、NCAM2、SYNP02、NUP210L、AMZ1、NR2F1、THBS2、PSORS1C1、FTH1P24、またはそれらの任意の組み合わせを含み得る。第2の遺伝子セットは、AKAP9、SPRY3、SPRY3、CAMKK2、COL1A1、FITM2、COX6C、VSIG10L、CYC1、KDM1B、MAPK15、ARSG、PAXIP1、DAAM1、AVL9、DMGDH、HLA-DQA1、HLA-DQB1、HLA-DRA、HLA-DRB5、HLA-H、IRF1、MGAT1、P2RX1、PLEK、CCDC93、PPP1R12C、SLC41A3、METTL3、CCAR2、PTPRE、SRL、SLC30A5、BMP4、ZNF133、ICE2、DCAKD、TMX1、TNFSF12、PER2、MCM3AP、またはそれらの任意の組み合わせを含み得る。

20

【 0 0 8 0 】

第2の遺伝子セットは、図12に列挙されている遺伝子を含み得る。第2の遺伝子セットは、COL1A1、FITM2、AASDH、COX6C、COX10、VSIG10L、MAPK15、PAXIP1、AVL9、GIGYF2、HLA-DQA1、HLA-DQB1、HLA-DRA、HLA-H、MGAT1、SLC41A3、PTPRE、SRL、SLC30A5、BMP4、ICE2、DCAKD、TMX1、HAVCR2、TNFSF12、PER2、MCM3AP、またはそれらの任意の組み合わせを含み得る。第2の遺伝子セットは、COL1A1、FITM2、AASDH、COX6C、COX10、VSIG10L、MAPK15、PAXIP1、AVL9、GIGYF2、HLA-DQA1、HLA-DQB1、HLA-DRA、HLA-H、MGAT1、SLC41A3、PTPRE、SRL、SLC30A5、BMP4、ICE2、DCAKD、TMX1、HAVCR2、TNFSF12、PER2、MCM3AP、およびそれらの任意の組み合わせからなる群より選択され得る。第2の遺伝子セットは、COL1A1を含み得る。第2の遺伝子セットは、FITM2を含み得る。第2の遺伝子セットは、AASDHを含み得る。第2の遺伝子セットは、COX6Cを含み得る。第2の遺伝子セットは、COX10を含み得る。第2の遺伝子セットは、VSIG10Lを含み得る。第2の遺伝子セットは、MAPK15を含み得る。第2の遺伝子セットは、PAXIP1を含み得る。第2の遺伝子セットは、AVL9を含み得る。第2の遺伝子セットは、GIGYF2を含み得る。第2の遺伝子セットは、HLA-DQA1を含み得る。第2の遺伝子セットは、HLA-DQB1を含み得る。第2の遺伝子セットは、HLA-DRAを含み得る。第2の遺伝子セットは、HLA-Hを含み得る。第2の遺伝子セットは、MGAT1を含み得る。第2の遺伝子セットは、SLC41A3を含み得る。第2の遺伝子セットは、PTPREを含み得る。第2の遺伝子セットは、SRLを含み得る。第2の遺伝子セットは、SLC30A5を含み得る。第2の遺伝子セットは、BMP4を含み得る。第2の遺伝子セットは、ICE2を含み得る。第2の遺伝子セットは、DCAKDを含み得る。第2の遺伝子セットは、TMX1を含み得る。第2の遺伝子セット

30

40

50

は、HAVCR2を含み得る。第2の遺伝子セットは、TNFSF12を含み得る。第2の遺伝子セットは、PER2を含み得る。第2の遺伝子セットは、MCM3APを含み得る。

【0081】

本開示の分類器または訓練されたアルゴリズムは、サンプルセット、例えばサンプルコポートを用いて訓練され得る。サンプルコポートは、約5、10、20、30、40、50、60、70、80、90、100、150、200、250、300、350、400、450、500、600、700、800、900、1000、2000、3000、4000、5000個、またはそれ以上の独立したサンプルを含み得る。サンプルコポートは、約100個の独立したサンプルを含み得る。サンプルコポートは、約200個の独立したサンプルを含み得る。サンプルコポートは、約100個～約500個の独立したサンプルを含み得る。独立したサンプルは、疾患、例えばがんを有すると診断された対象由来、健常対象由来、またはそれらの任意の組み合わせであり得る。

10

【0082】

サンプルコポートは、約5、10、20、30、40、50、60、70、80、90、100、150、200、250、300、350、400、450、500、600、700、800、900、1000体、またはそれ以上の異なる個体由来のサンプルを含み得る。サンプルコポートは、約100体の異なる個体由来のサンプルを含み得る。サンプルコポートは、約200体の異なる個体由来のサンプルを含み得る。異なる個体は、疾患、例えばがんを有すると診断された個体、健常個体、またはそれらの任意の組み合わせであり得る。

【0083】

サンプルコポートは、少なくとも1、2、3、4、5、6、67、8、9、10、15、20、25、30、35、40、45、50、55、60、65、70、75、または80箇所の異なる地理的場所（例えば、国家、例えば米国全体、大陸全体または全世界に広がる場所）で生活している個体から取得されたサンプルを含み得る。地理的場所は、試験センター、診療施設、診療所、受信用住所、市、群、州、国または大陸を含むがこれらに限定されない。いくつかの例において、米国由来のサンプルコポートを用いて訓練され分類器は、他の地理的地域（例えば、インド、アジア、ヨーロッパ、アフリカ等）由来のサンプルコポートに対して使用される際には再訓練される必要があり得る。

20

【0084】

分類器または訓練されたアルゴリズムは、実行されるごとに特有の出力を生成し得る。例えば、同じ分類器で異なるサンプルを用いることで、その分類器が実行されるごとに特有の出力が生成され得る。同じ分類器で同じサンプルを用いた場合も、その分類器が実行されるごとに特有の出力が生成され得る。同じサンプルを用いて分類器を複数回訓練した場合も、その分類器が実行されるごとに特有の出力がもたらされ得る。

30

【0085】

サンプルの特徴は、参照セットの特徴と比較され得る。比較は、分類器によって実施され得る。疾患発症のリスクを算出するために、サンプルの2つ以上の特徴が組み合わされ得る。組み合わせは、分類器によって実施され得る。例えば、サンプルから取得された配列が、サンプル内の1つまたは複数の配列変種の存在を判定するために、参照セットと比較され得る。いくつかの例において、サンプル由来の1つまたは複数の遺伝子の遺伝子発現レベルが、1つまたは複数の遺伝子の異なる遺伝子発現の存在を判定するために、参照遺伝子セットの発現レベルと比較され得る。参照セットは、1つまたは複数のハウスキーピング遺伝子を含み得る。参照セットは、特定の疾患に関連することが知られているまたは非疾患状態に関連することが知られている遺伝子の既知の配列変種または発現レベルを含み得る。分類器または訓練されたアルゴリズムは、比較、組み合わせ、統計学的評価もしくは結果のさらなる分析またはそれらの任意の組み合わせを実施し得る。異なる特徴空間に対して別個の参照セットが提供され得る。例えば、配列変種データは、配列変種データの参照セットと比較され得る。遺伝子発現レベルデータは、遺伝子発現レベルの参照セットと比較され得る。いくつかの例において、複数の特徴空間が、同じ参照セットと比較され得る。

40

【0086】

50

いくつかの例において、特定の遺伝子の配列変種は、同じ遺伝子の遺伝子発現レベルに影響する場合もあり影響しない場合もある。特定の遺伝子の配列変種は、その配列変種を有する特定の遺伝子に隣接しておよびその特定の遺伝子から離れて位置し得る1つまたは複数の異なる遺伝子の遺伝子発現レベルに影響し得る。1つまたは複数の配列変種の存在は、1つまたは複数の遺伝子に対して下流効果を有し得る。特定の遺伝子の配列変種は、1つまたは複数のシグナル伝達経路を混乱させ得、リボ核酸（RNA）転写調節を変化させ得、デオキシリボ核酸（DNA）を増幅させ得、複数の転写コピーを生成させ得、過剰なタンパク質を生成させ得、単一の塩基対、複数の塩基対、部分遺伝子または1つもしくは複数の遺伝子をその配列から除去させ得る。

【0087】

10

記載される方法からのデータ、例えば遺伝子発現レベルまたは配列変種のデータは、特徴選択技術、例えばそのデータ固有の特性を見ることによって個々の特徴の関連性を評価することができるフィルター、特徴サブセット検索の中にモデル仮説を埋め込むラッパーまたは最適な特徴セットの検索を分類アルゴリズムに組み込む組み込みプロトコルを用いてさらに分析され得る。

【0088】

20

本開示の方法において有用なフィルターは、(1) パラメトリック法、例えば2サンプルt検定、分散分析（ANOVA）分析、ベイズフレームワークもしくはガンマ分布モデルの使用、(2) モデルフリー法、例えばウィルコクソン順位和検定、二乗検定の級間・級内和、ランクプロダクト法、ランダム置換法、2つのデータセット間の発現の変化倍率差の閾点を設定し、誤分類数を最小化する各遺伝子内の閾点を検出する誤分類の閾値（TNOM）、または(3) 多変数法、例えば二変数法、相関に基づく特徴選択法（CFS）、最小冗長性最大関連性（MRMR）法、マルコフプランケットフィルター法、および非相関収縮重心法を含み得る。本開示の方法において有用なラッパーは、配列検索法、遺伝的アルゴリズムまたは分散アルゴリズムの概算を含み得る。組み込みプロトコルは、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ーンアルゴリズムの重みベクトルまたはロジスティック回帰アルゴリズムの重みを含み得る。

【0089】

30

本明細書に記載される方法から取得された結果の統計学的評価は、量的な値または以下の1つもしくは複数を示す値を提供し得る：リスク評価が正確である確率；診断が正確である確率；疾患、例えばがんの確率；特定の疾患、例えば組織特異的がん、例えば甲状腺がんの確率；および特定の治療的介入の成功的確率。したがって、遺伝学または分子生物学に関して教育を受けていない場合がある医療専門家は、遺伝子発現レベルまたは配列変種のデータ結果を理解する必要はない。むしろ、データは、対象の治療または処置の手引きとなる最も有用な形式で医療専門家に直接提示され得る。統計学的評価、別のデータ結果の統合および有用な結果の報告は、分類器または訓練されたアルゴリズムによって実施され得る。結果の統計学的評価は、スチューデントT検定、両側T検定、ピアソンの順位和分析、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、一元配置分散分析（ANOVA）、二元配置ANOVA等を含むがこれらに限定されない多くの方法によって実施され得る。統計学的評価は、分類器または訓練されたアルゴリズムによって実施され得る。

【0090】

40

本明細書において開示される方法は、対象由来の1つまたは複数のサンプルからタンパク質または核酸（RNAもしくはDNA）を抽出および分析する工程を含み得る。核酸は、取得された総サンプルから抽出され得るまたは一部から抽出され得る。いくつかの例において、核酸抽出に供されないサンプルの一部は、細胞学的試験または免疫組織化学により分析され得る。生物学的サンプルからのRNAまたはDNAの抽出方法は、例えば、フェノール・クロロホルム抽出（例えば、チオシアニ酸グアニジンフェノール・クロロホルム抽出）、エタノール沈降、スピンドラムベースの精製等を含み得る。

【0091】

50

遺伝子発現レベルを決定する一般的方法は、以下の1つまたは複数を含み得るがこれら

に限定されない：追加の細胞学的アッセイ法、特定のタンパク質または酵素活性のアッセイ法、タンパク質またはRNAもしくは特定のRNAスプライス変種を含む特定の発現産物のアッセイ法、インサイチューハイブリダイゼーション、全または部分ゲノム発現分析、マイクロアレイハイブリダイゼーションアッセイ法、連続遺伝子発現分析（SAGE）、酵素連結免疫吸着アッセイ法、質量分析、免疫組織化学、プロッティング、配列決定、RNA配列決定、DNA配列決定（例えば、RNAから取得した相補的デオキシリボ核酸（cDNA）の配列決定）、次世代（Next-Gen）配列決定、ナノ細孔配列決定、ピロシーケンス、またはナノストリング配列決定。遺伝子発現産物レベルは、内部標準、例えば総メッセンジャーRNA（mRNA）または特定遺伝子の発現レベルに対して標準化され得る。相互に比較されるサンプル間、例えば対象由来のサンプルと参照サンプルとの間の遺伝子発現には、一定の差または一定範囲の差が存在し得る。遺伝子発現レベルの差は、少なくとも5%、10%、15%、20%、25%、30%、35%、40%、45%、もしくは50%またはそれ以上であり得る。いくつかの例において、遺伝子発現レベルの差は、少なくとも2、3、4、5、6、7、8、9、10倍、またはそれ以上であり得る。

10

【0092】

RNA配列決定は、2つまたはそれ以上の特徴空間、例えば特定のサンプルの遺伝子発現および配列変種の存在のカウントを生成し得る。例えば、RNA配列決定は、特定の組織または特定のサンプル、例えば甲状腺組織または甲状腺結節において発現された遺伝子内の変種を測定する。次世代配列決定は、特定のサンプルの遺伝子発現レベルデータを提供し得る。配列決定結果、例えばRNA配列決定および次世代配列決定結果は、単一の特徴空間を用いるよりも高い精度で疾患の発症のリスクを決定するために特有の特徴空間を組み合わせることができる分類器に入力され得る。分類器または訓練されたアルゴリズムは、既知の悪性、良性、および正常サンプルの参照セットを用いて構築されたアルゴリズムを含み得る。分類器または訓練されたアルゴリズムは、既知の低リスク、中リスク、および高リスクサンプルの参照セットを用いて構築されたアルゴリズムを含み得る。

20

【0093】

アレイハイブリダイゼーション、配列決定、増幅のためのマーカー

アレイハイブリダイゼーション、核酸配列決定、核酸増幅、または他の増幅反応を実施するのに適した試薬は、DNAポリメラーゼ、マーカー、例えばフォワードおよびリバースプライマー、デオキシヌクレオチド三リン酸（dNTP）および1つまたは複数の緩衝液を含むがこれらに限定されない。そのような試薬は、関心対象の所定配列、例えば第1の遺伝子セットおよび／または第2の遺伝子セットの1つまたは複数の遺伝子について選択されたプライマーを含み得る。

30

【0094】

そのような増幅反応において、プライマー対の一方のプライマーは、標的ポリヌクレオチド分子（例えば、第1のセットまたは第2のセットの1つまたは複数の遺伝子）の配列に相補的なフォワードプライマーであり得、プライマー対の一方のプライマーは、標的ポリヌクレオチド分子の第2の配列に相補的なリバースプライマーであり得、標的部位は、第1の配列と第2の配列の間に位置し得る。

40

【0095】

フォワードプライマーおよびリバースプライマーの長さは、標的ポリヌクレオチド（例えば、第1のセットまたは第2のセットの1つまたは複数の遺伝子）の配列および標的部位に依存し得る。いくつかの例において、プライマーは、約5、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50、51、52、53、54、55、56、57、58、59、60、65、70、75、80、85、90、95、または約100ヌクレオチド以上の長さであり得る。あるいは、プライマーは、約100、95、90、85、80、75、70、65、60、59、58、57、56、55、54、53、52、51、50、49、48、47、46、45、44、43、42、41、40、39、38、37、36、35、34、33、32、31、30、29、28、27、26、25、24、23、22、21、20、19、18、17、16、15、14、13、12、11、10、9、8、7、6ヌクレオチド未満の長さであり得る。い

50

くつかの例において、プライマーは、約15～約20、約15～約25、約15～約30、約15～約40、約15～約45、約15～約50、約15～約55、約15～約60、約20～約25、約20～約30、約20～約35、約20～約40、約20～約45、約20～約50、約20～約55、約20～約60、約20～約80、または約20～約100ヌクレオチドの長さであり得る。

【0096】

プライマーは、二次構造および自己ハイブリダイゼーション、例えばプライマーニ量体対を回避するための既知のパラメータにしたがい設計され得る。異なるプライマー対は、別のプライマー対とほぼ同じ温度で、例えば1、2、3、4、5、6、7、8、9、または10以内で、アニールおよび融解し得る。

【0097】

標的部位は、複数の鋳型ポリヌクレオチドの3'末端または5'末端から約10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50、100、150、200、220、230、240、250、260、270、280、290、300、310、320、330、340、350、360、370、380、390、400、410、420、430、440、450、460、470、480、490、500、510、520、530、540、550、560、570、580、590、600、650、700、750、800、850、900、または1000ヌクレオチドであり得る。

【0098】

記載される方法のためのマーカー（すなわち、プライマー）は、1つまたは複数の同一プライマーであり得る。いくつかの例において、マーカーは、1つまたは複数の異なるプライマー、例えば約2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、30、40、50、60、70、80、90、100、200、300、400、500、600、700、800、900、1000個、またはそれ以上の異なるプライマーであり得る。そのような例において、1つまたは複数のプライマーの各プライマーは、異なる標的または鋳型特異的領域または配列、例えば第1のセットまたは第2のセットの1つまたは複数の遺伝子を含み得る。

【0099】

1つまたは複数のプライマーは、プライマーの固定パネルを含み得る。1つまたは複数のプライマーは、少なくとも1つまたは複数のカスタムプライマーを含み得る。1つまたは複数のプライマーは、少なくとも1つまたは複数の対照プライマーを含み得る。1つまたは複数のプライマーは、少なくとも1つまたは複数のハウスキーピング遺伝子プライマーを含み得る。いくつかの例において、1つまたは複数のカスタムプライマーは、標的特異的領域またはその相補領域にアニールする。1つまたは複数のプライマーは、1つもしくは複数の標的もしくは鋳型ポリヌクレオチドを増幅するようまたはそのプライマー伸長、逆転写、直線的伸長、非指數関数的増幅、指數関数的増幅、PCR、もしくは任意の他の増幅法を行うよう設計され得る。

【0100】

プライマーは、プライマーの検出または固定化を可能にするがそのプライマーの基本的特性（例えば、DNA合成の開始点としての作用）を変化させない追加の特徴を含み得る。例えば、プライマーは、標的核酸にハイブリダイズしないが、クローニングもしくはさらなる増幅または増幅産物の配列決定を容易にする核酸配列を5'末端に含み得る。例えば、配列は、プライマー結合部位、例えばPCRプライミング配列、サンプルバーコード配列またはユニバーサルプライマー結合部位等を含み得る。

【0101】

ユニバーサルプライマー結合部位または配列は、ユニバーサルプライマーをポリヌクレオチドおよび/またはアンプリコンに付加させることができる。ユニバーサルプライマーは、-47F (M13F)、MF、AOX3'、AOX5'、BGHr、CMV-30、CMV-50、CVMf、LACrmt、gt10F、gt10R、gt11F、gt11R、M13 rev、M13フォワード(-20)、M13リバース、male、p10SEQPpQE、pA-120、pet4、pGAPフォワード、pGLRVpr3、pGLpr2R、pKLAC14、pQEFS、pQERs、pucU1、pucU2、reversA、seqIREStam、seqIRESzpet、seqori、seqPCR、seqpIRES-、seqpIRES+、seqpSecTag、seqpSecTag+、seqretro+PSI、SP6、T3-prom、T7-promおよびT7-t

10

20

30

40

50

ermInvを含み得る。本明細書で使用される場合、付加は、共有結合的相互作用および非共有結合的相互作用の両方またはいずれかを表し得る。ユニバーサルプライマー結合部位へのユニバーサルプライマーの付加は、ポリヌクレオチドおよび／またはアンブリコンの増幅、検出および／または配列決定のために使用され得る。

【0102】

リスク判定の使用

分類器の結果、例えば疾患発症のリスク、または本明細書において開示される方法からのデータ、例えば遺伝子発現レベルもしくは配列変種データは、分子プロファイリング事業、個人、医療専門家または保険事業の代表者または代理人によってアクセス用データベースに入力され得る。データのコンピュータまたはアルゴリズムによる分析は、自動的に提供され得る。結果は、コンピュータスクリーン上のレポートとしてまたは紙面による記録として提示され得る。結果は、いくつかの例において、データベースまたはリモートサーバに自動的にアップロードされ得る。レポートは、以下の1つまたは複数のような情報を含み得るがこれらに限定されない：原サンプルの適性、示差的に発現される遺伝子の名前および／もしくは数、配列変種を含む遺伝子の名前および／もしくは数、配列変種のタイプ、示差的に発現される遺伝子の発現レベル、数的な分類器のスコア、対象の診断、診断に対する統計学的信頼性、疾患の発症のリスク、指示された治療法またはそれらの任意の組み合わせ。

【0103】

対象は、本明細書に記載される方法を用いて、单一の時点でまたは複数の時点にわたってモニタリングされ得る。例えば、対象は、本明細書に記載される方法を用いて、疾患、例えばがんまたは遺伝性障害を有すると診断され得る。いくつかの例において、この初期診断は、本明細書に記載される方法の使用を伴わない場合がある。陽性の疾患診断、例えば甲状腺がんを有する対象は、その後、治療的介入、例えば甲状腺摘出術を指示され得るまたは薬物計画、例えば化学療法を開始され得る。治療的介入の結果は、本明細書に記載される方法を使用してその治療的介入の効果を検出することによって継続的にモニタリングされ得る。別の例において、がんを有さない対象が、がんの発症のリスクを有すると診断され、本明細書に記載される方法を使用して彼ら／彼女らの健康状態の状況の変化を検出し、後の時点でがんが存在し得るかどうかを判定するまたはスクリーニング法を実施する頻度に影響を与えることによって継続的にモニタリングされ得る。

【0104】

本明細書に記載される方法はまた、特定の治療的介入の潜在的効果を対象への実施前に確認するために使用され得る。例えば、対象は、がんを有すると診断された者であり得る。本明細書に記載される方法は、がんの悪性化に関与することが既知の遺伝子産物、例えばRASオンコジーンにおいて高レベルの遺伝子発現を示し得る。その高レベルを有する対象からサンプルが取得され、インビトロで培養され得る。その後、異常に活性化されたもしくは調節異常の経路の様々な阻害剤またはその経路の活動を阻害することが既知の薬物の適用が、サンプルの腫瘍細胞において成長阻害に関して試験され得る。分子プロファイリングがまた、例えば関与する経路の下流標的に対するこれらの阻害剤の効果をモニタリングするために使用され得る。分子プロファイリングはまた、これらの阻害剤の効果を予測するために使用され得る。

【0105】

本明細書に記載される方法は、疾患、例えばがんの診断のため、サンプル、例えば腫瘍細胞、細胞株、組織もしくは生物に対する薬物もしくは候補薬物の効果をモニタリングするため、または疾患の進行もしくは抑制、例えばがんの発生および／もしくは腫瘍の抑制のための新規経路を明らかにするための新規マーカーを同定するための調査ツールとして使用され得る。

【0106】

本明細書に記載される方法は、(1) 少量および／または低品質の核酸を含むサンプルの遺伝子発現分析、(2) 偽陽性および偽陰性の大きな減少、(3) 結果として生じる病態

10

20

30

40

50

を担う根底にある遺伝的経路、代謝的経路、またはシグナル伝達経路の決定、(4)疾患、例えば遺伝性障害の診断の正確さに対して統計学的確率を割り当てる能力、(5)不明瞭な結果を解消する能力、(6)疾患、例えばがんのサブタイプ間を区別する能力、ならびに(7)低疾患発症リスクと中～高疾患発症リスクの間を区別する能力を提供し得る。

【0107】

予測は、正確な訓練ラベルに依存し得る。例えば、図10に示されるように、Afirma Gene Expression Classifier (GEC) バージョン1において組織学的に悪性とラベル付けまたは分類されたサンプルは、米国甲状腺学会 (ATA) 病期体系を用いて低発症リスクまたは中／高発症リスクのいずれかとしてさらにラベル付けまたは分類される。低発症リスクとラベル付けされたサンプルにおいて、組織病理学レポートは、1つまたは複数のリスク特徴の非存在を記述し得る。中／高発症リスクとラベル付けされたサンプルにおいて、組織病理学レポートは、1つまたは複数のリスク特徴が正に存在することを記述し得る。リスク特徴は、リンパ節転移、血管侵襲、甲状腺外拡張、またはそれらの任意の組み合わせであり得る。

10

【0108】

リスク分類器は、がんの特定のサブタイプを含む単一の組織サンプル、例えば、甲状腺乳頭がん (PTC) を含む組織サンプルを用いて訓練され得る。いくつかの例において、リスク分類器は、がんの2つ、3つ、4つ、またはそれ以上のサブタイプ、例えばPTC、LCT、HA、およびFCを含む単一の組織サンプルを用いて訓練される。いくつかの例において、リスク分類器は、2つ以上の組織サンプル、例えば、がんの2つ、3つ、4つ、またはそれ以上のサブタイプ、例えばPTC、LCT、HA、およびFCを含む2つの組織サンプルを用いて訓練され得る。

20

【0109】

キット

疾患診断事業、分子プロファイリング事業、製薬事業、または患者の保健に関連する他の事業は、疾患の発症のリスクの判定を実施するためのキットを提供し得る。キットは、分類器、アルゴリズムを訓練するためのサンプルコホートおよび各特徴空間のための遺伝子リスト、例えば第1の遺伝子セットおよび第2の遺伝子セットを含み得る。いくつかの例において、キットは、分類器および各特徴空間のための遺伝子リストを含み得る。キットは、すべての疾患タイプのための汎用キットであり得る。キットは、特定の疾患、例えばがんのための特別キット、または疾患サブタイプ、例えば甲状腺がんに対する特別キットであり得る。キットは、キットには提供されないサンプルコホートを用いてすでに訓練されている分類器を提供し得る。キットは、分類器とともに使用するサンプルコホートまたは特徴空間のための遺伝子リストの定期的な更新を提供し得る。キットは、医療専門家によって報告され得もしくは表示され得もしくはダウンロードされ得る結果概要、および／またはデータベースに入力され得る結果概要を自動化するソフトウェアを提供し得る。結果概要是、患者に対する処置オプションの推奨および疾患の発症リスクを含む、本明細書の開示される結果のいずれかを含み得る。キットはまた、対象からサンプルを取得するためのユニットまたはデバイス（例えば、アスピレーターに接続された針を含むデバイス）を提供し得る。キットはまた、本明細書において開示される方法を実施するための説明書を提供し得、RNA配列決定および次世代 (NextGen) 配列決定のためのすべての必要な緩衝液および試薬を含み得る。キットはまた、結果を分析するための説明書を含み得る。そのような説明書は、使用者を結果分析のためのソフトウェア（例えば、訓練されたアルゴリズムを含むソフトウェア）およびデータベースに誘導することを含む。

30

【0110】

コンピュータ制御システム

本開示は、本開示の方法を実行するよう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9は、本明細書において提供される方法を実行するようプログラムされたまたはそれ以外の方法でそのように構成されたコンピュータシステム9001を示している。コンピュータシステム9001は、本開示の疾患発症のリスクを層別化する、例えば分類器を稼

40

50

働させ、アルゴリズムを訓練し、層別化された発症リスクを報告する様々な局面を制御し得る。コンピュータシステム9001は、使用者の電子デバイスまたはその電子デバイスから遠隔に設置されたコンピュータシステムであり得る。電子デバイスは、携帯型電子デバイスであり得る。

【0111】

コンピュータシステム9001は、シングルコアもしくはマルチコアプロセッサまたは並列処理のための複数のプロセッサであり得る中央処理装置（CPU、本明細書で「プロセッサ」および「コンピュータプロセッサ」とも称される）9005を備える。コンピュータシステム9001はまた、メモリまたはメモリロケーション9010（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）、電子記憶装置9015（例えば、ハードディスク）、1つまたは複数の他のシステムと連絡するためのコミュニケーションインターフェース9020（例えば、ネットワークアダプタ）ならびに端末デバイス9025、例えばキャッシュ、他のメモリ、データ記憶、および／または電子表示アダプタを備える。メモリ9010、記憶装置9015、インターフェース9020および端末デバイス9025は、コミュニケーションバス（実線）、例えばマザーボードを通じてCPU 9005に接続される。記憶装置9015は、データを記憶するためのデータ記憶装置（またはデータレポジトリ）であり得る。コンピュータシステム9001は、コミュニケーションインターフェース9020の支援により、コンピュータネットワーク（「ネットワーク」）9030に機能的に接続され得る。ネットワーク9030は、インターネット、インターネットおよび／もしくはエクストラネット、またはインターネットに接続されたインターネットおよび／もしくはエクストラネットであり得る。ネットワーク9030は、いくつかの例において、テレコミュニケーションおよび／またはデータネットワークである。ネットワーク9030は、分散コンピューティング、例えばクラウドコンピューティングを実現し得る1つまたは複数のコンピュータサーバを含み得る。ネットワーク9030は、いくつかの例において、コンピュータシステム9001の支援により、コンピュータシステム9001に接続されたデバイスがクライアントまたはサーバとしてふるまうことを可能にし得るピア・ツー・ピアネットワークを実現し得る。
10 20

【0112】

CPU 9005は、プログラムまたはソフトウェアに組み込まれ得る機械読み取り可能な指令のシーケンスを実行し得る。指令は、メモリロケーション、例えばメモリ9010に保存され得る。指令は、CPU9005に対するものであり得、それによってCPU 9005を、本開示の方法を実行するよう、プログラミ得るかまたはそれ以外の方法で構成し得る。CPU 9005によって実施される作業の例は、フェッチ、復号、実行、およびライトバックを含み得る。
30

【0113】

CPU 9005は、回路、例えば集積回路の一部であり得る。システム9001の1つまたは複数の他のコンポーネントが、回路に含まれ得る。いくつかの例において、回路は、特定用途向け集積回路（ASIC）である。

【0114】

記憶装置9015は、ファイル、例えばドライバ、ライブラリ、および保存されたプログラムを保存し得る。記憶装置9015は、ユーザデータ、例えばユーザプレファレンスおよびユーザプログラムを保存し得る。コンピュータシステム9001は、いくつかの例において、コンピュータ9001の外部に、例えばインターネットまたはインターネットを通じてコンピュータシステム9001に接続されているリモートサーバ上に位置する1つまたは複数の追加のデータ記憶装置を備え得る。
40

【0115】

コンピュータシステム9001は、ネットワーク9030を通じて1つまたは複数のリモートコンピュータシステムと連絡し得る。例えば、コンピュータシステム9001は、使用者のリモートコンピュータシステム（例えば、サービスプロバイダ）と連絡し得る。リモートコンピュータシステムの例は、パーソナルコンピュータ（例えば、携帯型PC）、スレートもしくはタブレットPC（例えば、Apple（登録商標）iPad、Samsung（登録商標）Galaxy Tab）、電話、スマートフォン（例えばApple（登録商標）iPhone、アンドロイド実装デバイス
50

、Blackberry（登録商標）、またはパーソナルデジタルアシスタントを含む。使用者は、ネットワーク9030を通じてコンピュータシステム9001にアクセスすることができる。

【0116】

本明細書に記載される方法は、コンピュータシステム9001の電子記憶ロケーションに、例えば、メモリ9010または電子記憶装置9015に保存された機械（例えば、コンピュータプロセッサ）実行可能なコードによって実行され得る。機械実行可能または機械読み取り可能なコードは、ソフトウェアの形式で提供され得る。使用時、コードはプロセッサ9005によって実行され得る。いくつかの例において、コードは、プロセッサ9005による容易なアクセスのために、記憶装置9015から読み出され、メモリ9010上に保存される。いくつかの状況において、電子記憶装置9015は排除され得、機械読み取り可能な指令はメモリ9010上に保存される。

10

【0117】

コードは、そのコードを実行するよう適合されたプロセッサを有する機械で使用するために事前コンパイルおよび構成され得、または実行中にコンパイルされ得る。コードは、そのコードを事前コンパイルまたは同時コンパイル様式で実行できるよう選択され得るプログラム言語で供給され得る。

【0118】

本明細書において提供されるシステムおよび方法の局面、例えばコンピュータシステム9001は、プログラミングで実現され得る。この技術の様々な局面は、典型的に、機械読み取り可能な媒体の形式で持ち運ばれるまたはそのような媒体で具現化される機械（もしくはプロセッサ）実行可能コードおよび／または関連データの形態の「製品」または「製造品」と考えられ得る。機械実行可能なコードは、電子記録装置、例えばメモリ（例えば、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）またはハードディスク上に保存され得る。「記憶」タイプの媒体は、ソフトウェアプログラミングのためにいつでも非一時的記憶を提供し得るコンピュータ、プロセッサ等の有形のメモリまたはその関連モジュール、例えば様々な半導体メモリ、テープドライブ、ディスクドライブ等のいずれかまたはすべてを含み得る。ソフトウェアのすべてまたは一部は、時に、インターネットまたは様々な他のテレコミュニケーションネットワークを通じて接続され得る。そのようなコミュニケーションは、例えば、1つのコンピュータまたはプロセッサから別のものへの、例えば管理サーバまたは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ホームへのソフトウェアのローディングを実現し得る。したがって、ソフトウェア要素を保持し得る別のタイプの媒体は、有線および光通信ネットワークを通じたおよび様々な無線リンクによる、例えばローカルデバイス間の物理的インターフェースとして使用される、光波、電波、および電磁波を含む。そのような波、例えば有線または無線リンク、光リンク等を有する物理的要素もまた、ソフトウェアを保持する媒体とみなされ得る。本明細書で使用される場合、非一時的な有形の「記憶」媒体に制限されない限り、コンピュータまたは機械「読み取り可能媒体」等の用語は、実行用プロセッサへの指令の提供に関する任意の媒体を表す。

20

【0119】

したがって、機械読み取り可能媒体、例えばコンピュータ実行可能なコードは、有形の記憶媒体、搬送波媒体、または物理的伝送媒体を含むがこれらに限定されない多くの形式をとり得る。不揮発性記憶媒体は、例えば、光学または磁気ディスク、例えば任意のコンピュータの任意の記憶デバイス等、例えば画面に示されるデータベース等を実行するために使用され得るもの、を含む。揮発性記憶媒体は、動的メモリ、例えばそのようなコンピュータプラットホームのメインメモリを含む。有形伝送媒体は、コンピュータシステム内のバスを含む線を含む、同軸ケーブル、銅線および光ファイバーを含む。搬送波伝送媒体は、電気もしくは電磁信号、または音響もしくは光波、例えばラジオ周波数（RF）および赤外線（IR）データコミュニケーションの間に生成されるものの形式をとり得る。したがって、コンピュータ読み取り可能媒体の共通形式は、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、任意の他の磁気媒体、CD-ROM、DVD

30

40

50

またはDVD-ROM、任意の他の光媒体、パンチカードペーパーテープ、穴のパターンを有する任意の他の物理的記憶媒体、RAM、ROM、PROMおよびEPROM、FLASH-EPROM、任意の他のメモリチップもしくはカートリッジ、搬送波移送データもしくは指令、ケーブルもしくはリンク移送、例えば搬送波、またはコンピュータがプログラミングコードおよび／もしくはデータを読み取り得る任意の他の媒体を含む。これらの形式のコンピュータ読み取り可能媒体の多くは、1つまたは複数の指令の1つまたは複数のシーケンスを実行用プロセッサに運ぶのに関与し得る。

【0120】

コンピュータシステム9001は、例えば分類器または訓練されたアルゴリズムの出力または読み出しを提供するためのユーザインターフェース（UI）9040を含む電子ディスプレイ9035を含み得るまたはそれに接続され得る。UIの例は、非限定的に、グラフィカルユーザインターフェース（GUI）およびウェブベースのユーザインターフェースを含む。

【0121】

本開示の方法およびシステムは、1つまたは複数のアルゴリズムによって実施され得る。アルゴリズムは、中央処理装置9005によって実行されると、ソフトウェアによって実施され得る。アルゴリズムは、例えば、疾患の発症のリスクを層別化し得るか、またはサンプルを良性、悪性、疑わしい、もしくは非診断と分類し得る。

【実施例】

【0122】

実施例1：リスク分類器を用いたサンプルのリスク層別化

甲状腺がんの初期管理に対する現在のリスク適応型アプローチは、2009米国甲状腺学会病期体系（ATA）を用いた高～中発症リスクまたは低発症リスクのいずれかへの対象の手術後分類に基づいている。この解剖学的病期体系は、臨床的に有用であり得るが、甲状腺摘出術前に正確に評価することができず、かつ対象の結果の分子予測因子を含み得ない。この研究は、悪性甲状腺結節の診断的微細針吸引（FNA）で取得された転写データを使用して、甲状腺手術前のリスク層別化を強化できるかどうかを判定する。

【0123】

サンプル由来のFNA材料を手術前に回収し（n = 79）、手術後に専門家のパネルによって、古典的な組織学的サブタイプを含む甲状腺乳頭がん（PTC）について診断させる（図1および図2）。各患者を、発症リスクの層別化のために確立されたガイドラインを用いて、「低リスク」または「中～高リスク」のいずれかに分類する。ゲノムワイドRNA配列（RNASeq）データ（サンプルあたり8000万の読み取り）を取得し、教師付き学習を使用して、サポートベクターマシン（SVM）、ランダムフォレスト（RF）、罰則付きロジスティック回帰（PLR）およびこの3つのアンサンブルを含む分類器を訓練する。分類器のパフォーマンスを、同じサンプルコホートにおいて10分割交差検証を用いて測定する。

【0124】

分類器を、320個の遺伝子およびBRAF遺伝子状態に関して制御されたオープンソースソフトウェアDESeqモデルを用いて構築する。最大の「低リスク」対「中～高リスク」分類パフォーマンスが、サポートベクターマシン（SVM）分類器において観察され、その最大受信者動作特性（ROC）曲線下面積（AUC）は0.86である（図3および図4）。すべての分類器が、類似のAUCを達成する：RF 0.82、PLR 0.82およびアンサンブル0.84。分類に有用であることが発見された遺伝子は、ECM受容体相互作用、接着斑、および細胞接着分子を含む様々な膜貫通シグナル伝達経路に属する（図5）。評価された分類器は、感度または特異度のいずれよりも全体精度について最適化された閾値を使用している。サンプルコホートに適用された場合、サポートベクターマシン（SVM）分類器は、米国甲状腺学会（ATA）の低リスク腫瘍の79.3%（23/29）およびATAの中～高リスク腫瘍の82.0%（41/50）を正確に判定する（図5）。

【0125】

実施例2：交差検証モデル

手術前リスク層別化が機械学習を使用することによって強化されるかどうかを判定する

10

20

30

40

50

ため、Gene Expression Classifier (GEC) を変異パネルに用いて不確定甲状腺結節を試験する。図10は、訓練ラベルの決定を示す流れ図である。組織学的に良性のサンプルと組織学的に悪性のサンプルの間を区別するためにAfirma GECバージョン1訓練ラベルを使用する。組織学的に悪性のサンプルをさらに、米国甲状腺学会 (ATA) リスク訓練ラベルを用いて低発症リスクと中 / 高発症リスクの間を区別する。中 / 高リスクの特徴は、リンパ節転移、血管侵襲、甲状腺外拡張、またはそれらの任意の組み合わせを含む。リスク訓練用サンプルコホートが、図1に示されている。中 / 高発症リスクの組織学的特徴を有するサンプルの比率が、図2に示されている。10分割交差検証を行い、線形サポートベクターマシン (SVM) 、ランダムフォレスト、GLMNetおよびアンサンブル分類器を含む異なる学習モデルで曲線下面積 (AUC) を評価する。この実施例において、最良のモデルは、AUCが0.871 (図11Aに示されている) 、感度が86% (図11Bに示されている) 、特異度が86% (図11Bに示されている) 、陽性的中率 (PPV) が91.3%、陰性的中率 (NPV) が78.3%であるアンサンブル分類である。初期特徴空間は、50カウントおよび800個の変種を含む850個の初期特徴空間である。最良のパフォーマンスは、240個の組み合わされた特徴を使用している。各分割で分類器によって選択された変種の上位の特徴が、図12に示されている。10分割で分類器によって8~10回選択されたカウントの上位の特徴が、図13に示されている。

10

20

30

40

50

【 0 1 2 6 】

実施例3：変異分析

微細針吸引 (FNA) サンプル (n=81) は、回収され、手術後に専門家のパネルによって悪性 (甲状腺乳頭がん (PTC) 、多発性甲状腺乳頭がん (mPTC) 、濾胞型甲状腺乳頭がん (FVPTC) 、甲状腺乳頭がん高細胞型 (PTC-TCV) 、甲状腺髓様がん (MTC) 、詳細不明高分化がん (well-differentiated carcinoma-not otherwise specified (WDC-NOS)) 、肝細胞がん (HCC) 、濾胞がん (FC)) または良性 (良性家族性好中球減少症 (BFN) 、線維腺腫 (FA) 、肝細胞腺腫 (HCA) 、硝子化索状腺腫 (HTA) 、ライディッヒ細胞腫 (LCT)) と診断される。組織病理学的に真である手術組織サンプル (n=57) も分析する。組織病理を示さない臨床検査室改善修正法 (CLIA) ラボからの連続する不確定FNAの系列 (n=101) も分析する。サンプルを次世代配列決定 (NGS) に供し、14個の遺伝子 (図14) を、5つの異なる変異パネルにおいて漸増数の調査ゲノム部位および融合対により評価する。図14に示されるように、上の表は、5つの変異パネルの各々についてのゲノム部位の数および融合対の数を示している。変異パネル1は、9箇所のゲノム部位および3対の融合対から構成される。変異パネル2は、19箇所のゲノム部位および25対の融合対から構成される。変異パネル3は、208箇所のゲノム部位および25対の融合対から構成される。変異パネル4は、929箇所のゲノム部位および25対の融合対から構成される。変異パネル5は、3670箇所のゲノム部位および25対の融合対から構成される。図14の下の表は、変異パネルの1つまたは複数において標的とされた14個の遺伝子を示している。

【 0 1 2 7 】

いくつかのフィルターを適用してデータを採点する。融合体または点変異が存在しない場合、サンプルを負に採点する。それらのマーカーが良性のマーカーであるとみなされるグアニンヌクレオチド結合タンパク質、 刺激 (GNAS) 変異を除いて少なくとも1つの融合体または点変異が検出された場合、サンプルを正に採点する。

【 0 1 2 8 】

悪性腫瘍を検出する感度は、すべてのサンプルコホートにおいて位置数の増加とともに改善する。特異度は、その反対の傾向を示し、すべてのサンプルコホートにおいて位置数の増加とともに低下する。図15のFNAサンプルにおいて、最小の9部位パネルは、53%の感度および93%の特異度を示す。図15における最大のパネル (3670箇所の部位) は、100%の感度および10%の特異度を示す。

【 0 1 2 9 】

図17の外科的組織 (n=38) において、同様の傾向が観察される。合計57個の組織を評価する。しかし、38個の組織のみが、試験パフォーマンス算出に使用される明確な組織学的

良性または組織学的悪性の病態を有する。図17の最小の9部位パネルにおいて、89%の特異度は、45%の感度を伴う。図17の最高密度パネル(3670部位)において、100%の感度は0%の特異度を伴う。

【0130】

全体として、図15および図17の2つの大きい方のパネルは、組織学的に良性のFNAの87~90%を誤って悪性とみなし、図15および図17の2つの小さい方のパネルは、既知のがんの48~58%を見逃す。5つのパネルを通して、CLIA FNAサンプルにおける変異および融合体の頻度は、それぞれ、13%、4%、21%、89%および92%である。点変異および融合の数を増やして検出することによって得られる感度は、特異度の犠牲の上に到達し、真に良性のサンプルにおいて悪性をオーバーコールするリスクを伴う。

10

【0131】

208部位を有するパネル3における細胞学による変異パフォーマンスが、図16に示されている。細胞学的に良性(Cyto B)、意義不明の異型/意義不明の濾胞性病変(AUS/FLUS)、濾胞性新生物/濾胞性新生物の疑い(FN/SFN)、悪性腫瘍の疑い(SFM)、細胞学的に悪性(Cyto M)、および全サンプルを含むベセスタ細胞学カテゴリーによってグループ分けする。総サンプル数、総数あたりの組織学的良性変異の数、総数あたりの組織学的悪性変異の数、感度、特異度を含む様々なパラメータが、図16において各グループに関して示されている。

【0132】

CLIA FNAサンプルで観察された変異頻度のグラフ表示が、図18Aに示されている。変異陽性サンプル(パネル3)は、ダークグレイ色で示されている。GNAS陽性結節は、ライトグレイ色で示されている。変異頻度の比率が、全グループ、AUS/FLUSグループおよびFN/SFNグループを含む異なるグループに细分されている。図18Bは、図18Aにも示されている様々なサブグループにおいてパネル3を用いて検出された遺伝子および変異の表を示している。

20

【0133】

FNAサンプルで観察された変異頻度のグラフ表示が、図19Aに示されている。変異陽性サンプル(パネル3)は、ダークグレイ色で示されている。結節は、最小結節=1センチメートル(cm)に比例するサイズで示されている。変異頻度の比率が、全グループ、組織学的悪性グループおよび組織学的良性グループを含む異なるグループに细分されている。図19Bは、図19Aにも示されている様々なサブグループにおいてパネル3を用いて検出された遺伝子および変異の表を示している。

30

【0134】

組織サンプルで観察された変異頻度のグラフ表示が、図20Aに示されている。変異陽性サンプル(パネル3)は、ダークグレイ色で示されている。GNAS陽性結節は、ライトグレイ色で示されている。変異頻度の比率が、全グループ、組織学的悪性グループ、組織学的良性グループおよび組織学的不十分または非診断グループを含む異なるグループに细分されている。図20Bは、図20Aにも示されている様々なサブグループにおいてパネル3を用いて検出された遺伝子および変異の表を示している。

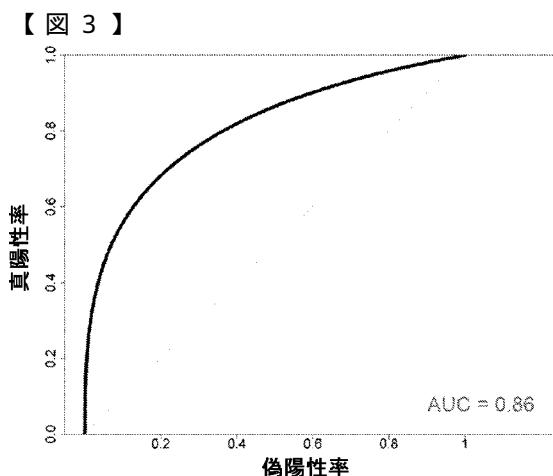
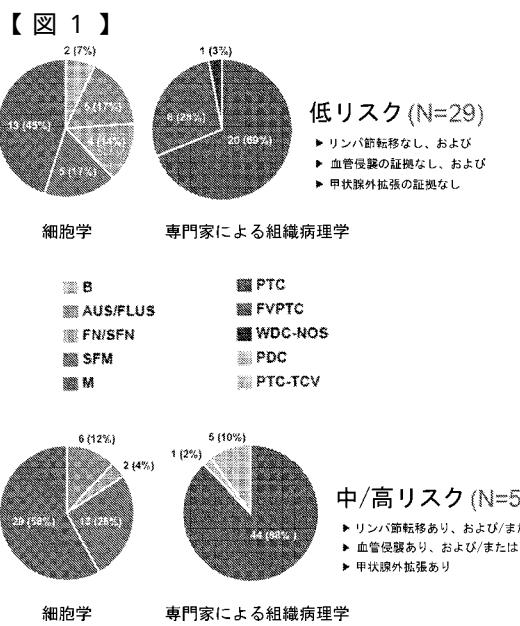
40

【0135】

本明細書において本発明の好ましい態様が示され説明されているが、そのような態様は実例として提供されるにすぎないことが当業者に明らかであろう。本発明が本明細書内に提供される特定の実施例によって限定されるこは意図されていない。本発明は、本明細書を参照して説明されているが、本明細書の態様の説明および実例は、限定の意味で捉えられることが意図されていない。今、本発明から逸脱することなく、多数の変更、変化および置換が当業者に想起されるであろう。さらに、本発明のすべての局面は、本明細書に示される特定の記述、構成または相対的な比率に限定されず、それらは様々な条件および変数に依存することが理解されるべきである。本発明を実施する際に、本明細書に記載される本発明の態様の様々な代替態様が用いられ得ることが理解されるべきである。したがって、本発明は、任意のそのような代替、改変、変更または等価物も網羅することが想定さ

50

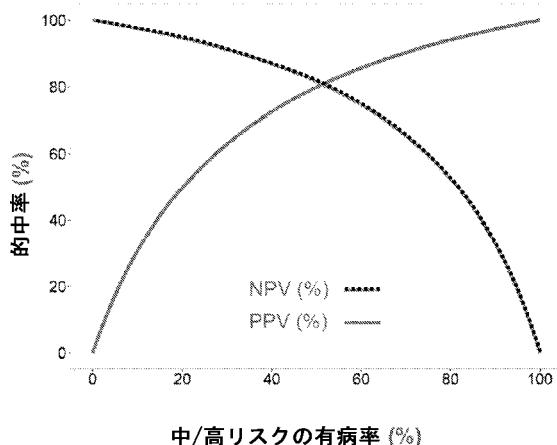
れている。添付の特許請求の範囲は、本発明の範囲を規定し、これらの特許請求の範囲内の方法および構造ならびにそれらの等価物がそれによって網羅されることが意図されている。



【図2】

組織病理学的特徴	サンプル数 (%)
リンパ節転移のみ	11 (22%)
血管侵襲のみ	1 (2%)
甲状腺外拡張のみ	15 (30%)
リンパ節転移+血管侵襲	1 (2%)
リンパ節転移+甲状腺外拡張	17 (34%)
血管侵襲+甲状腺外拡張	1 (2%)
リンパ節転移+血管侵襲+甲状腺外拡張	4 (8%)
合計	50 (100%)

【図4】



中/高リスクの有病率(%)

【図5】

複数のリスク群にわたるパフォーマンス

	中/高リスク (N=50)	低リスク (N=29)
--	------------------	----------------

「中/高リスク」に分類 41 6

「低リスク」に分類 9 23

感度 82%

特異度 79%

絆路または 遺伝子 オントロジー	予想された 遺伝子数	観察された 遺伝子数	FDR P値
------------------------	---------------	---------------	-----------

細胞外
マトリクス 2 20 1.28×10^{-13} ECM受容体
相互作用 1 11 6.12×10^{-08} 斑接着 2 12 3.71×10^{-05} チロシン
キナーゼ活性 1 6 4.77×10^{-03} 免疫系
プロセスの
調節 3 10 6.83×10^{-02} 血管発生 1 7 1.0×10^{-02}

分類に使用した遺伝子のサンプリング				
COX6C	FANCA	KCTD17	MPRIP	TUBA1B
DCAKD	ICE2	MCM3AP	TNFRSF14	WSB2

【図6-1】

Ensembl 遺伝子 ID	ENTREZID	遺伝子記号	概要	CV_分割_精度
ENSG00000044524	2042	EPHA3	EPH receptor A3 [Source:HGNC Symbol;Acc:HGNC:3387]	10
ENSG00000108821	1277	COL1A1	collagen, type I, alpha 1 [Source:HGNC Symbol;Acc:HGNC:2197]	10
ENSG00000135373	26298	EHF	ets homologous factor [Source:HGNC Symbol;Acc:HGNC:3246]	10
ENSG00000136237	9771	RAPGEF5	Rap guanine nucleotide exchange factor (GEF) 5 [Source:HGNC Symbol;Acc:HGNC:16862]	10
ENSG00000139174	144165	PICKLE1	pinickle homolog 1 (Drosophila) [Source:HGNC Symbol;Acc:HGNC:17019]	10
ENSG00000167105	162461	TMEM92	transmembrane protein 92 [Source:HGNC Symbol;Acc:HGNC:26579]	10
ENSG00000169855	6091	ROBO1	roundabout, axon guidance receptor, homolog 1 (Drosophila) [Source:HGNC Symbol;Acc:HGNC:10249]	10
ENSG00000204564	221545	C6orf136	chromosome 6 open reading frame 136 [Source:HGNC Symbol;Acc:HGNC:21301]	10
ENSG00000061656	6676	SPAG6	sperm associated antigen 4 [Source:HGNC Symbol;Acc:HGNC:11214]	9
ENSG00000131386	117248	GALNT15	polypeptide N-acetylgalactosaminyltransferase 15 [Source:HGNC Symbol;Acc:HGNC:21531]	9
ENSG00000139329	4060	LUM	lumican [Source:HGNC Symbol;Acc:HGNC:6724]	9
ENSG00000154654	4685	NLAM2	neural cell adhesion molecule 2 [Source:HGNC Symbol;Acc:HGNC:7657]	9
ENSG00000172403	171024	SYNP02	synaptopodin 2 [Source:HGNC Symbol;Acc:HGNC:17732]	9
ENSG00000143552	91181	NUP210L	nucleoporin 210kDa-like [Source:HGNC Symbol;Acc:HGNC:29915]	8
ENSG00000174945	155185	ANZ1	archaelysin family metallopeptidase 1 [Source:HGNC Symbol;Acc:HGNC:22231]	8
ENSG00000175745	7025	NR2F1	nuclear receptor subfamily 2, group F, member 1 [Source:HGNC Symbol;Acc:HGNC:7975]	8
ENSG00000186340	7058	THBS2	thrombospondin 2 [Source:HGNC Symbol;Acc:HGNC:11786]	8
ENSG00000204540	170679	PSORS1C1	psoriasis susceptibility 1 candidate 1 [Source:HGNC Symbol;Acc:HGNC:17202]	8
ENSG00000249302	NA	FTH1P24	ferritin, heavy polypeptide 1 pseudogene 24 [Source:HGNC Symbol;Acc:HGNC:37642]	8

【図6-2】

ENSG00000236039	NA	AC019117.2	7
ENSG00000158062	91544	UBXN11	UBX domain protein 11 [Source:HGNC Symbol;Acc:HGNC:30600]
ENSG00000178750	415117	STX19	syntaxin 19 [Source:HGNC Symbol;Acc:HGNC:19300]
ENSG00000196386	158055	C9orf153	chromosome 9 open reading frame 163 [Source:HGNC Symbol;Acc:HGNC:26718]
ENSG00000254352	NA	RP11-57R024.2	6
ENSG00000269845	641367	RP11-420K14.6	6
ENSG00000124740	3691	ITGB4	integrin, beta 4 [Source:HGNC Symbol;Acc:HGNC:6158]
ENSG00000185567	113146	AHNAK2	AHNAK nucleoprotein 2 [Source:HGNC Symbol;Acc:HGNC:20125]
ENSG00000232185	NA	CNOT7P2	CCR4 NOT transcription complex, subunit 7 pseudogene 2 [Source:HGNC Symbol;Acc:HGNC:44249]
ENSG00000232466	NA	RP11-337A23.5	5
ENSG00000000971	3075	CFH	complement factor H [Source:HGNC Symbol;Acc:HGNC:4883]
ENSG00000062428	25791	INGEF	neuronal guanine nucleotide exchange factor [Source:HGNC Symbol;Acc:HGNC:7807]
ENSG00000170454	9119	KRT75	keratin 75 [Source:HGNC Symbol;Acc:HGNC:24431]
ENSG00000226824	100996437	RP4-75H11.3	4
ENSG00000238961	677828	SNORD47	small nucleolar RNA, H/ACA box 47 [Source:HGNC Symbol;Acc:HGNC:32640]
ENSG00000243020	NA	RPL7P39	ribosomal protein L7 pseudogene 39 [Source:HGNC Symbol;Acc:HGNC:36214]
ENSG00000261649	728310	GOLGA6L7P	golgin A6 family-like 7, pseudogene [Source:HGNC Symbol;Acc:HGNC:37442]
ENSG00000141540	94015	TTVH2	Iwetely family member 2 [Source:HGNC Symbol;Acc:HGNC:13877]
ENSG00000141748	390790	ARL5C	ADP-ribosylation factor-like 5C [Source:HGNC Symbol;Acc:HGNC:31111]
ENSG00000149968	4314	MMP3	matrix metalloproteinase 3 (stromelysin, progelatinase) [Source:HGNC Symbol;Acc:HGNC:7173]

【図 6 - 3】

ENSG00000164100	9348	NDST3	N-deacetylase/N-sulfotransferase fibepean copine IV	30	1.19	1.47E-04	1.14E-01
ENSG00000164353	111034	CPRNE4		31	1.24	1.43E-04	1.14E-01
ENSG0000016498	NA	RPL5-564M11.2		32	1.27	1.50E-04	1.14E-01
ENSG00000292332	NA	ATP5ALP2	ATP synthase, H+ transporting, mitochondrial F1 complex, alpha subunit 1 pseudogene 3	33	1.24	1.48E-04	1.14E-01
ENSG0000016340	7058	THBS2	thrombospondin 2	34	1.04	1.66E-04	1.22E-01
ENSG0000017545	7025	NR3F1	nuclear receptor, family 5, subfamily B3, member 1 pseudogene	35	1.18	1.81E-04	1.30E-01
ENSG00000198261	NA	OR8BL1P	olfactory receptor, family 8, member 1	36	1.25	2.11E-04	1.47E-01
ENSG00000115414	2335	FN1	fibronectin 1	37	1.13	2.29E-04	1.56E-01
ENSG000000971	2015	C5H	complement factor H	38	1.02	2.45E-04	1.62E-01
ENSG000001337	11178	I2TS1	leucine zipper, putative tumor suppressor 1	39	1.12	2.97E-04	1.72E-01
ENSG0000013485	23475	QPR1	quinolinophosphoryltransferase	40	-1.01	4.18E-04	1.72E-01
ENSG00000105894	5764	PTN	pleiotrophin	41	1.09	3.09E-04	1.72E-01
ENSG00000109448	147166	TRIM16	imparine motif containing 16-like	42	-0.89	3.96E-04	1.72E-01
ENSG00000128675	1846	DUSP4	mitotic specificity phosphatase 4	43	0.63	4.14E-04	1.72E-01
ENSG00000128656	1123	CHN1	chitinase 1	44	0.89	3.98E-04	1.72E-01
ENSG00000133318	10313	RTNB	progesterin and adiponQ receptor family member V	45	0.31	3.80E-04	1.72E-01
ENSG00000158062	91544	UBXN11	UBX domain protein 11	46	-0.47	3.66E-04	1.72E-01
ENSG00000164078	4486	MSLR	macrophage stimulating 1 receptor (C-type lectin-related tyrosine kinase)	47	0.88	4.12E-04	1.72E-01
ENSG00000161666	115708	TRMT6A	tRNA methyltransferase 6A1 homolog A (S. cerevisiae)	48	-0.49	4.13E-04	1.72E-01
ENSG00000170370	2018	EWX2	empty spiracles homeobox 2	49	-1.12	4.14E-04	1.72E-01
ENSG00000174986	140807	Icat72	cathepsin D	50	-1.19	3.62E-04	1.72E-01
ENSG00000174203	171024	SMPD2	lysophosphatidylcholine acyltransferase 2	51	0.95	3.36E-04	1.72E-01
ENSG00000175197	1649	DOLT3	DNA cleavage-inducible transcript 3	52	-0.75	3.86E-04	1.72E-01
ENSG00000178756	415117	STX19	syntaxin 19	53	1.01	3.87E-04	1.72E-01
ENSG0000017134	1645	AK31C1	aldo-keto reductase family 1, member C1	54	-1.12	2.74E-04	1.72E-01
ENSG00000165655	3048	IHB32	hemoglobin, gamma G	55	-1.20	3.28E-04	1.72E-01
ENSG00000196639	3269	HRH1	histamine receptor H1	56	1.04	2.95E-04	1.72E-01
ENSG00000197997	54598	ELOV12	ELOV1 fatty acid long-chain 2	57	1.18	3.95E-04	1.72E-01
ENSG00000198805	4860	PNP	purine nucleoside phosphorylase	58	0.94	4.14E-04	1.72E-01
ENSG00000211797	NA	RP11-1036E70.9	purine nucleoside phosphorylase	59	1.10	4.17E-04	1.72E-01

【図 7 - 1】

ENSG00000117069	81849	ST6GALNAC5	glutathione S-transferase mu 4 [Source:HGN C Symbol:Acc:HGNC:2213]	3	3	3	3
ENSG000000267337	101927921	LINC01478	long intergenic non-protein coding RNA 1478 [Source:HGN C Symbol:Acc:HGNC:51121]	2	2	2	2
ENSG00000050344	9603	NFE2L3	nuclear factor, erythroid 2-like 3 [Source:HGN C Symbol:Acc:HGNC:7783]	1	1	1	1
ENSG00000100362	5816	PVALB	parvalbumin [Source:HGN C Symbol:Acc:HGNC:9704]	2	2	2	2
ENSG00000102780	160851	DGKH	diacylglycerol kinase, eta [Source:HGN C Symbol:Acc:HGNC:2854]	1	1	1	1
ENSG00000115414	2335	FN1	fibronectin 1	1	1	1	1
ENSG00000125897	100874241	TM4SF19	TM4SF19 antisense RNA 1 [Source:HGN C Symbol:Acc:HGNC:41085]	1	1	1	1
ENSG00000240518	NA	BPI1_515C16.1		1	1	1	1
ENSG00000241679	NA	RP11-80H8.4		1	1	1	1
ENSG00000250970	NA	RPS-1063M23.1		1	1	1	1
ENSG00000254429	NA	CTD-2562I17.7		1	1	1	1
ENSG00000267857	NA	RP5-1023B21.1		1	1	1	1

【図 7 - 2】

ENSG00000174594	10381	TUBB3	tubulin, beta 3 class III	60	1.05	3.59E-04	1.72E-01
ENSG00000266970	NA	NA	CD207 molecule, langrin	61	0.86	3.81E-04	1.72E-01
ENSG0000016031	50489	CD207	cyclochrome P450 family 21, subfamily A, endoplasmic reticulum 3	62	1.18	4.29E-04	1.72E-01
ENSG00000231852	1589	CYP21A2	cytochrome P450 family 21, subfamily A, member 1	63	1.18	4.33E-04	1.72E-01
ENSG00000124205	1908	EN1N3	aldo-keto reductase family 1, member C1	64	-1.18	4.66E-04	1.72E-01
ENSG00000174945	155185	ANX1	arachidonate metabolism	65	-0.99	4.63E-04	1.72E-01
ENSG00000204966	147920	IgHL2	IGI-like family member 2	66	1.16	4.69E-04	1.72E-01
ENSG00000137648	36649	TMRPS54	transmembrane protease, serine 4	67	1.06	4.88E-04	1.72E-01
ENSG00000049323	4052	I1BP1	latent transforming growth factor beta	68	1.02	5.15E-04	1.84E-01
ENSG0000011799	1303	CD12A1	collagen, type XII, alpha 1	69	1.01	5.20E-04	1.84E-01
ENSG00000188783	5549	PHE1P	proline-rich end leucine-rich repeat	70	1.10	5.10E-04	1.84E-01
ENSG0000023482	728190	NUTM2A-AS1	NUTM2A antisense RNA 1	72	-0.42	5.17E-04	1.84E-01
ENSG00000160355	3696	ITGB8	integrin, beta 3	73	1.01	5.41E-04	1.84E-01
ENSG00000169908	4071	TM4SF1	transmembrane 4, six family member 1	73	1.06	5.60E-04	1.84E-01
ENSG0000013361	1004	CDH6	cadherin 6, type 2, catenin (fetal kidney)	74	1.08	5.83E-04	1.86E-01
ENSG000002160774	NA	CID-2083E4.4	CID-2083E4.4	75	0.85	5.88E-04	1.86E-01
ENSG00000254708	NA	RP11-145M24.1	RP11-145M24.1	76	0.98	5.92E-04	1.86E-01
ENSG00000206195	503637	DUXAP3	double homeobox A pseudogene 8	77	-1.15	6.20E-04	2.04E-01
ENSG00000196139	8644	AKR1C8	alcohol dehydrogenase family 1, member C3	78	-1.10	6.34E-04	2.04E-01
ENSG00000160718	1301	COL1A1	collagen, type XI, alpha 1	79	1.12	6.98E-04	2.17E-01
ENSG00000169200	6590	S1P1	secretory leukocyte peptidase inhibitor	80	1.01	6.88E-04	2.17E-01
ENSG00000161295	NA	RPL11-524D16_A.3		81	1.10	6.94E-04	2.17E-01
ENSG00000103154	54550	NICAB2	N-terminal fl- band calcium binding protein 2	82	-1.12	7.15E-04	2.17E-01
ENSG00000160203	2043	EHH4	EPH receptor A4	83	0.76	7.29E-04	2.17E-01
ENSG00000128641	4430	MYO18	myosin 18	84	0.74	7.48E-04	2.17E-01
ENSG00000134873	9071	CLDN10	claudin 10	85	1.13	7.64E-04	2.17E-01
ENSG00000182110	347516	DGA1216	diacylglycerol O-acyltransferase 2-like 6	86	1.13	7.61E-04	2.17E-01
ENSG00000185567	113145	AIINAK2	AIINAK nuclear protein 2	87	0.94	7.68E-04	2.17E-01
ENSG00000236510	NA	A011284.3		88	1.13	7.64E-04	2.17E-01
ENSG00000249302	NA	I-H1P24	femin, heavy polypeptide 1 pseudogene 24	89	1.00	7.39E-04	2.17E-01

【図 7 - 3】

ENSG00000117069	1277	COL1A1	collagen, type I, alpha 1	1	1	1.47	1.66E-07
ENSG00000010821	162463	TMEV92	transmembrane protein 92	2	1.60	1.95E-07	2.44E-03
ENSG000000167105	162795	C1orf187	chromosome 1 open reading frame 87	3	1.58	9.19E-07	7.70E-03
ENSG00000016298	167055	SPAG4	spirocyclic associated antigen 4	4	1.17	6.74E-06	2.55E-02
ENSG000000155737	167298	EFH	EF-hand protein factor	5	1.24	6.80E-06	2.55E-02
ENSG000000131386	172148	GALM10	galactose 1-phosphate N-acetylpolymerase	6	5.88E-06	3.28E-02	
ENSG000000143552	91181	PBL210	pancreopancreatic domain-containing protein 210	7	1.41	9.33E-06	3.28E-02
ENSG000000121440	23024	PZDN3	panzicaurotide containing finger 3	8	-1.25	1.04E-05	3.28E-02
ENSG000000204564	225155	Gen136	chromosome 6 open reading frame 36	9	1.40	1.47E-05	4.11E-02
ENSG000000158906	NA	NA	neuregulin 3	11	1.43	1.84E-05	4.35E-02
ENSG000000216465	93639	NNKX3	neuregulin 3	12	-1.41	2.52E-05	5.28E-02
ENSG000000163352	1233	COL1A3	collagen, type VI, alpha 3	13	1.22	2.73E-05	5.33E-02
ENSG000000136337	9712	RGCEES	RGAE and RGNE nucleotide exchange factor	14	0.77	4.42E-05	5.33E-02
ENSG000000139174	141455	PICRE1	pancreole homolog 1 (pancreola)	15	0.87	4.22E-05	5.33E-02
ENSG000000339229	4060	LUM	lumican	16	1.25	4.64E-05	5.33E-02
ENSG000000161649	723310	GolG16L2P	golgin 16 like Golgi-associated protein	22	-1.34	6.38E-05	7.38E-02
ENSG000000169855	6091	ROBO1	1 (Drosophila)	23	1.13	6.38E-05	7.38E-02
ENSG000000132492	633	BN1	bigranin	24	1.19	3.52E-05	8.52E-02
ENSG000000236039	NA	A0119117.2		25	1.34	4.15E-05	8.52E-02
ENSG000000250591	170679	PRSS31	protease, serine, 3 pseudogene 1	26	1.39	3.30E-05	8.52E-02
ENSG000000179554	284297	SSCD	scavenger receptor cysteine rich family 5	27	1.14	1.07E-04	9.79E-02
ENSG000000261649	723310	ColG16L2P	golgin 16 like 7 pseudogene	28	1.13	1.07E-04	9.79E-02
ENSG000000254352	NA	RT1-578212.2		29	0.87	1.28E-04	1.11E-01
ENSG000000132470	11764	ITGB4	integrin, beta 4	30	1.00	8.14E-05	8.52E-02
ENSG000000204542	170679	PSORS1C1	psoriasis susceptibility 1 candidate 1	31	1.09	9.03E-05	9.05E-02
ENSG000000144543	2042	EPHA3	EPH receptor A3	32	1.10	1.03E-04	9.79E-02
ENSG000000154654	6555	NCAM2	neural cell adhesion molecule 2	33	1.13	1.03E-04	9.79E-02
ENSG000000231172	101927884	AC0017099.1		34	0.87	1.28E-04	1.11E-01
ENSG000000136197	79020	C7orf25	chromosome 7 open reading frame 25	35	0.87	1.28E-04	1.11E-01

ENSG00000156562	9173	IL1RL1	Interleukin 1 receptor-like 1	90	1.13	7.8E-04	2.18E-01
ENSG00000170699	81849	ST16GALNAC5	S16 alpha(N-acetylneuraminy)-2,3-beta-galactosyl-N-acetylglucosamine	91	1.11	8.4E-04	2.33E-01
ENSG00000116681	4053	L1BP2	transforming growth factor beta promoter/leucemia	92	0.73	9.1E-04	2.41E-01
ENSG00000140964	5371	NMT1	NMT serine/threonine protein kinase	93	0.36	9.0E-04	2.41E-01
ENSG00000174653	167359	NNTIK	Rho GTPase nucleotide exchange factor	94	-1.10	9.3E-04	2.41E-01
ENSG00000184110	389337	ARHGEP7	immunoglobulin lambda variable 3-9	95	0.94	9.7E-04	2.41E-01
ENSG00000134797	1104	RP1-3T3-9	ADAM metallopeptidase with thrombospondin type 1 motif 14	96	1.11	8.8E-04	2.41E-01
ENSG00000124948	11000465912	CDKN2A-AS1	CDKN2B antisense RNA 1	97	1.11	9.4E-04	2.41E-01
ENSG00000138316	140765	ADAM14	mannose-6-phosphate alpha, chit-14, member 2	99	1.10	9.6E-04	2.43E-01
ENSG00000267799	NA	MAN1A2P1	mannose-6-phosphate alpha, chit-14, member 2	100	-1.05	9.7E-04	2.43E-01
ENSG00000272077	NA	RP11-24P10.2	asporin	101	0.79	9.7E-04	2.43E-01
ENSG00000106819	54829	ASPN	tumor necrosis factor ligand superfamily, type 14 effector	102	0.99	1.0E-03	2.56E-01
ENSG00000120387	64995	TNSF18	APC-rhinosatatin factor-like 2	103	1.10	1.0E-03	2.56E-01
ENSG00000168273	644410	ATP14E	uncarboxylated ODC92942	104	1.09	1.0E-03	2.56E-01
ENSG00000268460	93426	DKK3-AS24D0226	neuin 4	105	1.09	1.0E-03	2.60E-01
ENSG00000147527	59277	NTN4	GUP1	106	0.91	1.1E-03	2.63E-01
ENSG00000144366	51454	NTN4	GUP1 signal peptide adaptor PIB domain	107	0.90	1.1E-03	2.63E-01
ENSG00000022695	NA	R2P1-11G13.3	ATP synthase, Ht transporting, mitochondrial F1 complex, subunit B1 pseudogene 7	108	0.90	1.1E-03	2.63E-01
ENSG00000162424	NA	ATPS1P17	neuronal calcium nucleotide exchange factor	109	0.74	1.1E-03	2.63E-01
ENSG00000062428	25731	NA	glutathione S-transferase mu 5	110	1.04	1.1E-03	2.63E-01
ENSG00000134201	23469	GSTM5	lrophilin 3	111	0.90	1.2E-03	2.66E-01
ENSG00000159471	23284	NA	prosite, serine, 23	112	-1.07	1.2E-03	2.66E-01
ENSG00000159367	11098	PRSS23	collagen, type I, alpha 2	113	1.06	1.2E-03	2.69E-01
ENSG0000016191	12728	COL1A2	prophoblastocyte 9A	115	0.90	1.2E-03	2.69E-01
ENSG00000141540	94015	PDE9A	twenty-four inulin member 2	116	0.65	1.2E-03	2.72E-01
ENSG00000131940	12428	TLYH2	dystonin (part 22)	117	-0.67	1.2E-03	2.75E-01
ENSG00000165183	34569	DCTN3	asparagine	118	-0.50	1.3E-03	2.78E-01
ENSG00000165183	34569	ASNG	asparagine	119	0.03	1.3E-03	2.78E-01

Gene ID	Gene Name	Description	Chromosome	Start Position	End Position	Strand	RefSeq	Uniprot	Protein ID	Length	EC Number	GO Terms	KEGG Pathway	UniProt
ENSG00000176722	ZNF404	zinc finger protein 404	22	349098	350400	-	ENSG00000176722	Q9HJG1	P48331	453				
ENSG00000155189	ENSG00000203707	zinc finger and B1B domain-containing 10	11	655866	657101	-	ENSG00000155189	AB741011.2	P48332	151	0.99	1.9E-03	3.28E-01	
ENSG00000203707	ENSG00000203707	RNA, U6 small nuclear-104S pseudogene	10	152	154	-	ENSG00000203707	SNRNU104SP	P48333	0.98	1.9E-03	3.28E-01	3.28E-01	
ENSG00000231320	NTANP3	N-terminal expansion endopeptidase gene 3	13	NA	NA	-	ENSG00000231320	RP11-21L2.3	P48334	354	1.00	2.0E-03	3.28E-01	
ENSG000002056255	RP11-21L2.2	doublecortin-like kinase 1	12	NA	NA	-	ENSG000002056255	DCLK1	P48335	155	0.98	2.0E-03	3.28E-01	
ENSG00000205383	ENSG00000135295	wingless-type MMTV integration site family, member 1	1	802326	802326	-	ENSG00000205383	WNT10A	P48336	157	0.90	2.1E-03	3.30E-01	
ENSG00000135295	ENSG00000135295	SMC3	4	NA	NA	-	ENSG00000135295	SMC3	P48337	158	0.74	2.0E-03	3.30E-01	
ENSG00000135295	ENSG00000135295	mitochondrial protease carrier 1	2	NA	NA	-	ENSG00000135295	TPC1	P48338	159	0.97	2.0E-03	3.30E-01	
ENSG00000135295	ENSG00000135295	ribosomal p40 ribosomal protein S3	3	NA	NA	-	ENSG00000135295	TRIB3	P48339	160	0.93	2.1E-03	3.30E-01	
ENSG00000135295	ENSG00000135295	secreted frizzled-related protein 1	1	NA	NA	-	ENSG00000135295	SFRP1	P48340	161	0.98	2.1E-03	3.30E-01	
ENSG00000135295	ENSG00000135295	secreted frizzled-related protein 2	2	NA	NA	-	ENSG00000135295	SFRP2	P48341	162	0.94	2.1E-03	3.31E-01	
ENSG00000135295	ENSG00000135295	pleckstrin homology domain containing, family G (with RhoGef domain) member 1	1	NA	NA	-	ENSG00000135295	PLEKHM1	P48342	163	0.66	2.1E-03	3.32E-01	
ENSG00000135295	ENSG00000135295	mitochondrial p40 ribosomal protein L1	1	NA	NA	-	ENSG00000135295	TPC1	P48343	164	-0.37	2.2E-03	3.34E-01	
ENSG00000135295	ENSG00000135295	ribosomal p40 ribosomal protein S3	3	NA	NA	-	ENSG00000135295	TRIB3	P48344	165	0.84	2.2E-03	3.34E-01	
ENSG00000135295	ENSG00000135295	secreted frizzled-related protein 1	1	NA	NA	-	ENSG00000135295	SFRP1	P48345	166	-1.03	2.2E-03	3.34E-01	
ENSG00000135295	ENSG00000135295	delta-like 4, E3 ubiquitin ligase	1	NA	NA	-	ENSG00000135295	DIX1A	P48346	167	0.67	2.2E-03	3.34E-01	
ENSG00000135295	ENSG00000135295	myostatin	6	NA	NA	-	ENSG00000135295	MYOG6	P48347	168	0.65	2.2E-03	3.34E-01	
ENSG00000135295	ENSG00000135295	hemoglobin, gamma A	1	NA	NA	-	ENSG00000135295	HGB1	P48348	169	-1.03	2.2E-03	3.34E-01	
ENSG00000135295	ENSG00000135295	chemokine (C-C motif) ligand 20	20	NA	NA	-	ENSG00000135295	CC2L20	P48349	170	1.01	2.2E-03	3.36E-01	
ENSG00000135295	ENSG00000135295	turner-associated calcium signal transducer 2	2	NA	NA	-	ENSG00000135295	TACSTD2	P48350	171	0.89	2.2E-03	3.36E-01	
ENSG00000135295	ENSG00000135295	arylsulfatase G	6	NA	NA	-	ENSG00000135295	ARSG	P48351	172	-0.37	2.3E-03	3.44E-01	
ENSG00000135295	ENSG00000135295	activating signal contegulator 1 complex	6	NA	NA	-	ENSG00000135295	ASC3	P48352	173	0.87	2.3E-03	3.44E-01	
ENSG00000135295	ENSG00000135295	leucine-rich repeat containing 16A	16A	NA	NA	-	ENSG00000135295	LRRC16A	P48353	174	0.85	2.4E-03	3.57E-01	
ENSG00000135295	ENSG00000135295	metallothionein 1G	1G	NA	NA	-	ENSG00000135295	M1L15	P48354	175	0.73	2.4E-03	3.57E-01	
ENSG00000135295	ENSG00000135295	chromosomal origin reading frame 163	163	NA	NA	-	ENSG00000135295	CORF163	P48355	176	0.87	2.5E-03	3.58E-01	
ENSG00000135295	ENSG00000135295	chromosomal origin reading frame 163	163	NA	NA	-	ENSG00000135295	CORF163	P48356	177	0.89	2.5E-03	3.58E-01	

ENSG000001017152	5999	R052	regulator of G-protein signalling 4	120	1.06	1.45E-03	2.99E-01
ENSG000001823185	5890	RAD51B	RAD51 paralog 8	121	0.32	1.45E-03	2.99E-01
ENSG0000017318	NA	AP000645.5		122	1.07	1.44E-03	2.99E-01
ENSG00000186254	220	AUDH1A3	aldehyde dehydrogenase 1 family, member	123	0.86	1.41E-03	3.00E-01
ENSG000001093255	10374	WNU	neurotrophin U	124	1.07	1.49E-03	3.02E-01
ENSG00000151914	667	DST	dystonin	125	0.73	1.50E-03	3.02E-01
ENSG00000173132	NA	R11P...350B10.12	ICL2-associated althiocigene 4	126	0.94	1.58E-03	3.08E-01
ENSG000001567305	9520	BAG4	KRIT177	127	-0.27	1.58E-03	3.08E-01
ENSG00000189182	374454	ST1440	keratin 77	128	-0.98	1.63E-03	3.10E-01
ENSG00000166182	83391	WAWS1	serine/threonine kinase 40	129	-0.44	1.65E-03	3.10E-01
ENSG0000025802	NA	RPS1	tryptophanyl tRNA synthetase pseudogene	130	1.01	1.57E-03	3.10E-01
ENSG000001306770	NA	RPS10B3M2.3	tryptophanyl tRNA synthetase	131	1.04	1.60E-03	3.10E-01
ENSG00000059715	27238	PCDH11Y	grionocadherin 11Y linked	132	0.94	1.65E-03	3.13E-01
ENSG00000059344	9603	NFE2L3	nuclear factor, erythroid 2-like 3	133	0.62	1.69E-03	3.19E-01
ENSG000001022780	160851	DNKR2	diacylglycerol kinase, eta	134	0.47	1.76E-03	3.22E-01
ENSG000001155816	56776	FNNK2	formin 2	135	-1.01	1.74E-03	3.22E-01
ENSG00000189253	2272	FHIT	Lynx6/neurocigin 1	136	0.69	1.76E-03	3.22E-01
ENSG00000189253	10148	ERB3	fragile histone 3rd	137	-0.40	1.75E-03	3.22E-01
ENSG00000105246	2272	AS12	Eps-tein-Born virus induced 3	138	0.83	1.77E-03	3.23E-01
ENSG00000180817	5464	PPA1	v-akt murine thymoma viral oncogene	139	-0.38	1.81E-03	3.24E-01
ENSG00000123901	10388	GPR83	prophorbol ester receptor 3	140	0.43	1.81E-03	3.24E-01
ENSG00000103585	1050747	SLC2A15.1	G protein-coupled receptor 83	141	-1.00	1.83E-03	3.26E-01
ENSG00000131725	21838	LINC0384	long intergenic nonprotein coding RNA 264	142	0.98	1.84E-03	3.26E-01
ENSG00000073584	6605	SAR1ACE1	SWI/SNF-related, matrix associated, actin dependent regulator of chromatin subfamily	143	1.05	1.86E-03	3.28E-01
ENSG00000109494	51561	I23A	interleukin 23, alpha subunit p19	144	0.22	1.92E-03	3.28E-01
ENSG000001013936	130	C7	complement component 7	145	0.75	1.96E-03	3.28E-01
ENSG00000117983	72897	WUCB	muco 5, disulfonated muco 5-binding protein	146	1.04	2.00E-03	3.28E-01
ENSG00000183933	8447.1	LORX4	lysyl oxidase-like 4	147	1.03	1.99E-03	3.28E-01
ENSG00000162444	11632	RSPB7	retinol binding protein 7, cellular	148	-0.90	1.94E-03	3.28E-01
				149	-0.61	2.03E-03	3.28E-01

ENSG00000244731	720	C4A	complement component 4A (Rodgers' blood group)	210	0.80	3.2E-03	3.81E-01
ENSG00000244734	10132939911	RP1-57D7.39		211	-0.85	3.12E-03	3.81E-01
ENSG00000244739	NA	RP1-32Q2.6		212	0.79	3.12E-03	3.81E-01
ENSG00000245059	NA	[C11]	enoyl CoA hydratase 1, peroxisomal	213	-0.53	3.2E-03	3.82E-01
ENSG00000245283	1891	E2AP2	epoxide-induced 2A esterase gene 2	214	-0.58	3.2E-03	3.82E-01
ENSG00000245287	NA	[NPV172]	MPV17 mitochondrial membrane protein-like	215	0.55	3.3E-03	3.88E-01
ENSG00000245486	84769	ELMO1	ELMO/CED-12 domain-containing 1	216	-0.56	3.3E-03	3.89E-01
ENSG00000245685	64367	ELMO2D1	Kappa-like factor 4 (klf4)	217	0.82	3.40E-03	3.94E-01
ENSG00000245707	55531	KL14	Abi family member 1 (Abi1) binding protein	218	-0.56	3.46E-03	3.94E-01
ENSG00000245826	9314	A2BP3	long intergenic non-protein coding RNA 541	219	0.93	3.46E-03	3.94E-01
ENSG00000245917	52830	INC0511	RP1-UD5D4.4	220	0.93	3.47E-03	3.94E-01
ENSG00000246036	60061	NA	kinesin family member 7	221	0.84	3.48E-03	3.94E-01
ENSG00000246058	374654	KIF7	tumor protein p53 inducible protein 11	222	0.76	3.5E-03	3.94E-01
ENSG00000246074	9537	TPM3	embigin-32	223	-0.44	3.48E-03	3.94E-01
ENSG000002460754	NA	SP1-1016G14.4	Uglier transposible element derived 6	224	0.46	3.5E-03	3.94E-01
ENSG00000246256	1948	EFNB2	Rho guanine nucleotide exchange factor (GEF) 13	225	0.70	3.5E-03	3.94E-01
ENSG00000246269	61789	ARHGFE18	amotom	226	0.41	3.6E-03	4.00E-01
ENSG000002462699	54848	AMOT	collagen, type V, alpha 2	227	0.80	3.61E-03	4.04E-01
ENSG000002462616	154796	COL5A2	TAFB RNA polymerase II TATA box binding protein (TBP)-associated factor, 31 kDa	228	0.63	3.67E-03	4.04E-01
ENSG00000246262	1290	NA	neurofibromatosis 1, beta 1	229	0.81	3.73E-03	4.04E-01
ENSG00000246274	NA	TAZ9921	RP1-3P3	230	0.67	3.77E-03	4.04E-01
ENSG000002462742	NA	NBF1-3P	neurofibromatosis 1, beta 1	231	0.78	3.77E-03	4.04E-01
ENSG000002462753	1417	CKR833	crystallin, beta B3	232	0.92	3.77E-03	4.04E-01
ENSG000002462769	26953	RANBP3	RAN binding protein 6	233	0.26	3.77E-03	4.04E-01
ENSG000002462894	5796	IPTRK	protein tyrosine phosphatase, receptor type, K	234	0.50	3.77E-03	4.04E-01
ENSG000002462922	9120	SLC16A6	solute carrier family 16, member 6	235	-0.95	3.82E-03	4.05E-01
ENSG000002462987	164312	LRNMM	Reactive oxygen reductant 4	236	0.92	3.84E-03	4.05E-01
ENSG000002467694	67079	SPTAN1	specific, alpha non-erythrocytic 1	237	0.37	3.87E-03	4.05E-01
ENSG000002479808	92283	ZINC1	zinc finger protein 1	238	0.34	3.88E-03	4.05E-01
ENSG00000233929	NA	M13X1P1	metallothionein 1X pseudogene 1	239	0.84	3.88E-03	4.05E-01

【☒ 7 - 1 2】

ENSG00000078607	2081	ERIN1	endoplasmic reticulum to nucleus signaling, 1	390	-0.35	7.64E-03	4.32E-01			RPI1-357N13.1	330	0.81	6.09E-03	4.64E-01
ENSG00000030164	3949	LDR	low density lipoprotein receptor	391	0.60	7.78E-03	4.33E-01			CB8-6IM7.2	331	-0.82	6.15E-03	4.65E-01
ENSG00000042207	9875	URB1	URB1 ribosome biogenesis 1 homolog (S-	392	0.28	7.76E-03	4.33E-01				332	0.87	6.25E-03	4.67E-01
ENSG00000050938	51232	CHRM1	cysteine-rich transmembrane BMP-regulator 1	393	0.45	7.69E-03	4.33E-01			C-FTR				
		SQPP2	sphingosine-1-phosphate phosphatase 2	394	0.63	7.71E-03	4.33E-01			KCNJ8				
ENSG00000063082	130367	CDDC159	coiled-coil domain containing 159	395	-0.32	7.76E-03	4.33E-01			ENSG00000121351	3764			
ENSG00000083401	126075	M11X	metallocine 11 X	396	0.66	7.76E-03	4.33E-01			ENSG00000180562	NA			
ENSG00000078193	4501	RP11-485G7.5		397	-0.85	7.85E-03	4.35E-01			ENSG00000023742	NA			
ENSG00000052050	NA	AOC2	amine oxidase, copper-containing 2 (retina-	398	-0.83	7.89E-03	4.38E-01			ENSG00000006210	5376			
ENSG00000131480	314	TACR1	tachykinin receptor 1	399	0.80	8.04E-03	4.39E-01			ENSG00000132329	NA			
ENSG0000015353	6869	EPC8	ER membrane protein complex subunit 8	400	-0.23	7.96E-03	4.39E-01			ENSG00000164898	23524			
ENSG00000031148	10928	EXTL1	exostosin-like/glycosyltransferase 1	401	0.80	8.04E-03	4.39E-01			SP10R				
ENSG00000058008	2134	CD-2020127.4		402	0.70	7.95E-03	4.39E-01			ENSG00000187045	156456			
ENSG00000039463	NA	FGR	CD-202112.1	403	0.82	7.97E-03	4.39E-01			TM-PRSS55				
ENSG00000074620	400058	MRNBP	natron lute ring finger protein 5, pseudogene	404	0.86	7.98E-03	4.39E-01			RNA2032				
ENSG00000020209	NA	LINCO01240	long intergenic non-protein coding RNA, 1240	405	0.62	8.01E-03	4.39E-01			TM-PRSS11E				
ENSG00000015709	NA	CB48	calcium and inositol binding family member 4	406	-0.89	8.13E-03	5.01E-01			ENSG0000013229	56256			
ENSG00000029555	NA	RP1-5054		407	0.72	8.14E-03	5.01E-01			SERTAD4				
ENSG00000065398	2268	microRNA-3.28		408	0.74	8.13E-03	5.01E-01			CD1C146				
ENSG00000003938	2268	FGR	FEK proto-oncogene, Src family tyrosine kinase	409	-0.77	8.22E-03	5.04E-01			ENSG0000018325	NA			
ENSG000000187116	353514	ULRAS	leukocyte immunoglobulin-like receptor, subfamily A (with TM domain), member 5	410	-0.72	8.22E-03	5.04E-01			WAC				
ENSG00000015058	284837	AR0653.11	lysine/arginine-rich protein coding RNA, 1240	411	-0.66	8.26E-03	5.04E-01			UBBAA				
ENSG00000028282	NA	AC02957.8		412	-0.83	8.24E-03	5.04E-01			immunoglobulin superfamily containing 2				
ENSG000000249235	6101927096	CTB-113P19.1	uncharacterized LOC_0101927096	413	0.57	8.22E-03	5.04E-01			SLR				
ENSG000000267756	6495	SIX1	SIX homeobox 1	414	0.88	8.31E-03	5.04E-01			SMN/M4				
ENSG00000013525	3567	I5	interleukin 5	415	0.85	8.35E-03	5.06E-01			SMN/D2				
ENSG00000013525	3567	WIF1C	WIF1C	416	0.36	8.45E-03	5.09E-01			XKR4				
ENSG000000187079	4641	RNA1	RNA, tSL, cytoplasmic 64d, pseudogene	417	-0.98	8.47E-03	5.10E-01			PR1-202937				
ENSG00000024426	NA	RN7SL16P		418	0.50	8.47E-03	5.10E-01			TC4				
ENSG000000261485	1002288730	PAN5.1	PAN5 antisense RNA 1	419	0.83	8.74E-03	5.13E-01			TRNP31				
ENSG00000073282	8626	TIP63	tumor protein 63	420	0.83	8.74E-03	5.13E-01			AGAP antisense RNA 1				
				421	0.57	8.22E-03	5.04E-01			RNA1				
ENSG000000267756	6495	CM1	cytokeratin 5	422	0.88	8.86E-03	5.13E-01			PR1-21-33159.3				
ENSG00000013525	3567	SOX4	SOX4 sex determining region Y-like box 4	423	0.51	8.84E-03	5.13E-01			PR1-1402937				
ENSG000000216267	1340	CD66B1	CD66B1	424	-0.23	8.92E-03	5.13E-01			PR1-202937.5				
ENSG00000011101	28987	NOB1	NOB1	425	-0.28	8.92E-03	5.13E-01			PR1-2124126				
ENSG000000143013	8543	LM04	LM domain only 4	426	0.33	8.74E-03	5.13E-01			PR1-3000B6.5				
ENSG000000249235	6101927096	CTP17A1	cytokeratin 5	427	-0.32	8.87E-03	5.13E-01			PR1-3114834				
ENSG000000143013	8543	C7orf151	chromosome 7 open reading frame 31	428	-0.32	8.92E-03	5.13E-01			PR1-3114835				
ENSG000000153790	136895	ARCA10	ATP-binding cassette, sub-family A (ARCA), member 10	429	0.62	8.68E-03	5.13E-01			PR1-3124933				
ENSG00000015263	10349	PRO51	protein S fidalgi	430	0.63	8.70E-03	5.13E-01			PR1-312568				
ENSG00000018450	5627	FCAR	FC fragment of IgA receptor for	431	-0.76	8.80E-03	5.13E-01			PR1-315127				
ENSG000000186431	2204	TRAP	transformation/transcription activator, tissue	432	0.92	8.69E-03	5.13E-01			PR1-315211				
ENSG000000196357	8295	CM1	immunoglobulin lambda variable 9-49	433	0.57	8.75E-03	5.13E-01			PR1-315211.5				
ENSG000000148547	1586	GSNR	GSNR antisense RNA 1 (head to head)	434	0.75	8.62E-03	5.13E-01			PR1-315211.7				
ENSG000000174197	21370	CM1	cytokeratin 5	435	0.88	8.91E-03	5.13E-01			PR1-315211.9				
ENSG00000015263	10349	RP1-146N11.3	ubiquitin protein ligase	436	0.88	8.69E-03	5.13E-01			PR1-31724				
ENSG000000272761	100506551	UN012179	long intergenic non-protein coding RNA 1279	437	0.77	8.75E-03	5.13E-01			PR1-318586				
ENSG000000135773	10753	CAPN9	calpain 9	438	-0.82	8.66E-03	5.14E-01			PR1-31851				
ENSG000000104368	5327	PLAT	plasminogen activator, tissue	439	0.81	9.02E-03	5.15E-01			PR1-17A1.3				
ENSG000000223350	NA	ICL9-49	inscutable homolog (Drosophila)	440	0.88	9.01E-03	5.15E-01			PR1-17A15.3				
ENSG000000249240	NA	CDP22	cotransporter complex, subunit eta 2	441	-0.54	9.07E-03	5.15E-01			PR1-1840D2				
ENSG000000254850	NA	VHL	von Hippel-Lindau tumor suppressor, E3	442	-0.30	9.08E-03	5.15E-01			PR1-181715.2				
ENSG000000271225	NA	GPKW	Guaiacum procumbens	443	-0.24	9.20E-03	5.15E-01			PR1-186.1				
ENSG0000002613005	26255	PTG3P	phosphatidylcholine transferase, 3, glyceride	444	-0.24	9.21E-03	5.15E-01			PR1-186.2				
ENSG000000239116	131294625	RP1-12015.3	lipoprotein lipase	445	0.86	9.21E-03	5.19E-01			PR1-187.1				
ENSG000000234079	NA	RP1-1304.3	lipoprotein lipase	446	-0.85	9.22E-03	5.19E-01			PR1-187.2				
ENSG000000140511	152864	HARLN3	hyaluronan and proteoglycan link protein 3	447	0.63	9.28E-03	5.19E-01			PR1-187.3				
ENSG000000178764	22882	ZH2	zinc fingers and homeoboxes 2	448	0.34	9.28E-03	5.21E-01			PR1-187.4				
ENSG000000151422	2241	FGR	tyrosine kinase	449	0.28	9.36E-03	5.23E-01			PR1-187.5				

【☒ 7 - 1 2 3】

ENSG0000002616447	NA	TM-PRSS55	transmembrane protease, serine/arginine-rich, 5	330	-0.81	6.09E-03	4.64E-01			PR1-187N13.1	331	-0.82	6.15E-03	4.65E-01
ENSG000000180562	NA	PR1-2119A8	ubiquosomal protein 21s endogenous 8	331	-0.90	6.21E-03	4.67E-01			PR1-187N13.2	332	0.87	6.25E-03	4.67E-01
ENSG000000187452	NA	PR1-2120A8	ubiquosomal protein 21s endogenous 8	332	0.66	6.18E-03	4.67E-01			PR1-187N13.3	333	-0.72	6.31E-03	4.67E-01
ENSG00000006210	NA	PR1-2124126	ubiquosomal protein 21s endogenous 8	333	-0.76	6.31E-03	4.67E-01			PR1-187N13.4	334	0.66	6.38E-03	4.67E-01
ENSG000000132329	NA	PR1-2124126	ubiquosomal protein 21s endogenous 8	334	-0.70	6.32E-03	4.67E-01			PR1-187N13.5	335	-0.88	6.38E-03	4.67E-01
ENSG000000164898	NA	PR1-2124126	ubiquosomal protein 21s endogenous 8	335	-0.88	6.38E-03	4.67E-01			PR1-187N13.6	336	-0.23	6.71E-03	4.81E-01
ENSG00000023742	NA	PR1-2124126	ubiquosomal protein 21s endogenous 8	336	-0.90	6.38E-03	4.67E-01			PR1-187N13.7	337	0.84	6.77E-03	4.81E-01
ENSG000000182909	NA	PR1-2124126	ubiquosomal protein 21s endogenous 8	337	-0.89	6.38E-03	4.67E-01			PR1-187N13.8	338	0.90	6.77E-03	4.81E-01
ENSG00000013229	NA	PR1-2124126	ubiquosomal protein 21s endogenous 8	338	-0.90	6.38E-03	4.67E-01			PR1-187N13.9	339	-0.52	6.91E-03	4.81E-01
ENSG000000177459	284119	PR1-2119A8	ubiquosomal protein 21s endogenous 8	339	0.52	6.91E-03	4.81E-01			PR1-187N13.10				

【☒ 7 - 1 3 】

ENSG000000224389	NA	complement component 4B (C4B)	complement component 4B (C4B)	340	0.70	6.92E-03	4.81E-01			PR1-187N13.11	340	0.70	6.92E-03	4.81E-01

</

【図 7 - 1 6】

ENSG0000000084106	340066	Irh-Mi_3p	triglyceride receptor expressed on myeloid cells like 3, pseudogene	450	-0.86	9.41E-03	5.23E-01
ENSG000000215840	NA	RP1-127G16.8.7	pregnancy-specific beta-1-glycoprotein 4	451	-0.75	9.46E-03	5.23E-01
ENSG000000241337	5672	P594	ZSCAN5D	452	0.54	9.49E-03	5.23E-01
ENSG000000267938	NA	RPI-155D18.1	linc finger and SCAN domain containing 4, 5D	453	0.85	9.44E-03	5.23E-01
ENSG000000270186	NA	C16orf70	luminescence 16, open reading frame 70	454	0.87	9.31E-03	5.23E-01
ENSG000000271549	82052	IKT	transmembrane 4 L-like ion channel, subfamily A, member 19	455	-0.50	9.46E-03	5.24E-01
ENSG000000163931	7285	TMAS5.19	TMAS5.19	456	-0.37	9.52E-03	5.24E-01
ENSG00000023945	10087-2207	S13GA16.4S1	lactoembryonic antigen-related cell adhesion molecule 6, non-specific cross-reactive antigen, type II, beta	457	-0.83	9.54E-03	5.24E-01
ENSG000000086548	4680	CEACAM6	CEACAM6	458	0.85	9.60E-03	5.26E-01
ENSG000000153347	158643	FAM81B	family with sequence similarity 81, member 8	459	-0.67	9.60E-03	5.26E-01
ENSG000000231985	343263	MYBP1	myosin binding protein H-like	460	0.85	9.65E-03	5.27E-01
ENSG000000262999	10192-7131	C10-308E15.6	transmembrane 4 L-like ion channel, subfamily A, member 19	461	-0.36	9.61E-03	5.27E-01
ENSG000000145107	116211	TMAS5.19	TMAS5.19	462	-0.87	9.77E-03	5.28E-01
ENSG000000240705	NA	TPARP	TPARP antisense RNA 1	463	0.64	9.77E-03	5.28E-01
ENSG000000262111	1025062211	MIR210HG	MIR210 host gene, non-protein coding	464	0.86	9.76E-03	5.29E-01
ENSG000000249575	6232	SNORD2B	small nucleolar ribonucleoprotein, type II, beta	465	0.76	9.82E-03	5.29E-01
ENSG00000007547499	NA	SLC75A4(P1)	vacuole carrier family 7, member 47	466	0.84	9.81E-03	5.29E-01
ENSG000000108236	NA	NA	vacuole carrier family 7, member 47	467	-0.22	9.86E-03	5.30E-01
ENSG000000196704	51321	AM22	parcelsin family metallopeptidase 2	468	-0.27	9.89E-03	5.31E-01
ENSG000000243926	NA	L1K	leukocyte receptor tyrosine kinase	469	-0.87	9.93E-03	5.32E-01
ENSG000000053409	NA	LR8V7.4	cell receptor beta variable 7.4	470	0.86	9.97E-03	5.34E-01
ENSG000000261222	101923343	C10-208E12.31	cell receptor beta variable 7.4	471	-0.86	1.00E-02	5.35E-01
ENSG000000173153	2101	FSRRA	testosterolelated receptor alpha	472	-0.39	1.00E-02	5.34E-01
ENSG000000256625	NA	RP1-1-S82E3.4	RP1-1-S82E3.4	473	0.65	1.01E-02	5.35E-01
ENSG000000169758	123591	C15orf27	luminescence 15, open reading frame 27	474	-0.74	1.01E-02	5.36E-01
ENSG000000213516	494115	RBX01	mRNA binding motif protein X-linked-like 1	475	0.22	1.02E-02	5.36E-01
ENSG000000215159	NA	RP1-174F11.3	parvalbumin	476	0.75	1.02E-02	5.36E-01
ENSG00000010362	5816	PVALB	receptor tyrosine kinase-like orphan receptor 1	477	-0.86	1.03E-02	5.38E-01
ENSG000000185483	4919	RROR1	ribosomal protein S26 pseudogene 3	478	0.75	1.03E-02	5.38E-01
ENSG000000212889	NA	RP52693	ribosomal protein S26 pseudogene 3	479	0.65	1.03E-02	5.38E-01

【図 7 - 1 7】

ENSG000000130876	56301	SLC7A10	solute carrier family 7, neutral amino acid transporter, light chain, asc system, member	480	0.70	1.04E-02	5.43E-01
ENSG000000261505	NA	LA16-35B27.3	calcium channel, voltage-dependent, alpha 2/delta subunit 3	481	0.60	1.05E-02	5.48E-01
ENSG000000157445	55799	CACNA2D3	potato leafroll virus 3	482	-0.72	1.06E-02	5.41E-01
ENSG000000184226	5101	20D19	proteasome 9	483	0.66	1.06E-02	5.41E-01
ENSG000000267375	NA	CB-18G5.2.4	tripartite motif containing 16	484	0.79	1.06E-02	5.42E-01
ENSG000000087308	94234	FOXP1	forkhead box Q1	485	0.70	1.07E-02	5.43E-01
ENSG000000245571	101927204	AN200125B8.4	NA	486	0.54	1.07E-02	5.43E-01
ENSG000000187534	NA	PRR13P5	proline rich 13 pseudogene 5	487	0.43	1.07E-02	5.43E-01
ENSG000000095059	1725	DHPS	deoxyhypusine synthase	488	-0.79	1.08E-02	5.45E-01
ENSG000000235695	NA	R15-22C21.1	serine/threonine kinase 11 interacting protein	489	-0.24	1.10E-02	5.61E-01
ENSG000000213799	91664	ZNH845	zinc finger protein 845	490	-0.45	1.09E-02	5.62E-01
ENSG000000066136	4802	NPYC	nuclear transcription factor Y gamma	491	0.27	1.10E-02	5.62E-01
ENSG000000109625	6532	CIZ2	carboxyphosphatase Z	492	-0.10	1.13E-02	5.63E-01
ENSG000000115221	3694	11S16	Intertin, beta 6	493	0.83	1.13E-02	5.63E-01
ENSG00000041738	2886	GRB7	growth factor receptor bound protein 7	494	0.71	1.13E-02	5.63E-01
ENSG000000144589	11470	STKL1P	serine/threonine kinase 11 interacting protein	495	0.60	1.13E-02	5.63E-01
ENSG000000151790	65999	U002	tryptophan 2,3-dioxygenase	497	-0.95	1.12E-02	5.63E-01
ENSG000000166116	25841	A1TB2	akaryin repeat and 15 (ODF) domain	498	0.49	1.11E-02	5.63E-01
ENSG000000170866	NA	NA	zinc finger protein 443	499	-0.74	1.12E-02	5.63E-01
ENSG000000180855	10224	ZNF443	zinc finger protein 443	500	0.48	1.13E-02	5.63E-01

【図 8 - 1】

ENTREZID	記号	遺伝子名		CV 分割 頻度
		遺伝子名	記号	
10142	AKAP9	A kinase (PRKA) anchor protein 9	10	
10251	SPRY3	sprouty homolog 3 (Drosophila)	10	
10251	SPRY3	sprouty homolog 3 (Drosophila)	10	
10645	CAMKK2	calcium/calmodulin-dependent protein kinase	10	
1277	COL1A1	collagen, type I, alpha 1	10	
128486	FITM2	fat storage-inducing transmembrane protein 2	10	
1345	COX6C	cytochrome c oxidase subunit Vtc	10	
147645	VSIG10L	V-set and immunoglobulin domain containing 10 like	10	
1537	CYCV1	cytochrome c-1	10	
221656	KDM1B	lysine (K)-specific demethylase 1B	10	
225689	MAPK15	mitogen-activated protein kinase 15	10	
22901	ARSG	arylsulfatase G	10	
22976	PAXIP1	PAX interacting (with transcription-activation domain) protein 1	10	
23002	DAAM1	dishevelled associated activator of morphogenesis 1	10	
23080	AVL9	AVL9 homolog (S. cerevisiae)	10	
29958	DMGDH	dimethylglycine dehydrogenase	10	
3117	HLA-DQA1	major histocompatibility complex, class II, DQ alpha 1	10	
3119	HLA-DQB1	major histocompatibility complex, class II, DQ beta 1	10	
3122	HLA-DRA	major histocompatibility complex, class II, DR alpha	10	
3127	HLA-DRB5	major histocompatibility complex, class II, DR beta 5	10	
3136	HLA-H	major histocompatibility complex, class I, H (pseudogene)	10	
3659	IRF1	interferon regulatory factor 1	10	
4245	MGAT1	N-acetylglycosaminyltransferase, mannosyl (alpha-1,3)-glycoprotein beta-1,N-acetylglucosaminyltransferase	10	
5023	P2RX1	purinergic receptor P2X, ligand-gated ion channel, 1	10	
5341	PLEK	pleckstrin	10	
54520	CCDC93	coiled-coil domain containing 93	10	
54776	PPP1R12C	protein phosphatase 1, regulatory subunit 12C	10	
54946	SLC41A3	solute carrier family 41, member 3	10	
56339	METTL3	methyltransferase like 3	10	

【図 8 - 2】

57805	CCAR2	cell cycle and apoptosis regulator 2	10
5791	PTPRE	protein tyrosine phosphatase, receptor type, E	10
6345	SRL	sarcalumenin	10
64924	SLC30A5	solute carrier family 30 (zinc transporter), member 5	10
652	BMP4	bone morphogenic protein 4	10
7692	ZNF133	zinc finger protein 133	10
79664	ICE2	interactor of little elongator complex ELL subunit 2	10
79877	DCAKD	dephospho-CoA kinase domain containing thioredoxin-related transmembrane protein 1	10
81542	TMX1	TMX1	10
8742	TNFSF12	tumor necrosis factor (ligand) superfamily, member 12	10
8864	PER2	period circadian clock 2	10
8888	IVCM3AP	minichromosome maintenance complex component 3 associated protein	10
100462981	MTRNR2L2	MT-RNR2-like 2	9
10568	SLC34A2	solute carrier family 34 (type II sodium/phosphate transporter), member 2	9
10626	TRIM16	tripartite motif containing 16	9
11276	SYNRG	synerg, gamma	9
126017	ZNF813	zinc finger protein 813	9
1352	COX10	cytochrome c oxidase assembly homolog 10 (yeast)	9
140685	ZBTB46	zinc finger and BTB domain containing 46	9
163081	ZNF567	zinc finger protein 567	9
2153	F5	coagulation factor V (proaccelerin, labile factor)	9
23065	EMC1	ER membrane protein complex subunit 1	9
23157	SEPT16	septin 6	9
23164	MPRIP	myosin phosphatase Rho interacting protein	9
23379	ICE1	interactor of little elongator complex ELL subunit 1	9
26058	GIGYF2	GRB10 interacting GYF protein 2	9
3728	JUP	junction plakophilin	9
4318	MMP9	matrix metalloproteinase 9 (gelatinase B, 92kDa gelatinase, 92kDa type IV collagenase)	9
5518	PPP2R1A	protein phosphatase 2, regulatory subunit A, alpha	9
55200	PLEKHG6	pleckstrin homology domain containing, family G (with RhoGef domain) member 6	9

【図 8 - 3】

58191	CXCL16	chemokine (C-X-C motif) ligand 16	9
629	CFB	complement factor B	9
633	BGN	biglycan	9
6990	DYNLT3	dynein, light chain, Tctex-type 3	9
7846	TUBA1A	tubulin, alpha 1a	9
79968	WDR76	WD repeat domain 76	9
8544	PIR	pirin (iron-binding nuclear protein)	9
871	SERPINH1	serpin peptidase inhibitor, clade H (heat shock protein 47), member 1, (collagen binding protein 1)	9
8773	SNAP23	synaptosomal-associated protein, 23kDa	9
914	CD2	CD2 molecule	9
91662	NLRP12	NLR family, pyrin domain containing 12	9
100130015	URAHP	urate (hydroxyl-) hydrolase, pseudogene	8
10053	AP1M2	adaptor-related protein complex 1, mu 2 subunit	8
10376	TUBA1B	tubulin, alpha 1b	8
10422	UBAC1	UBA domain containing 1	8
1062	CENPE	centromere protein E, 312kDa	8
11022	TDRKH	tudor and KH domain containing	8
11138	TBC1D8	TBC1 domain family, member 8 (with GRAM domain)	8
115361	GBP4	guanylate binding protein 4	8
1173	AP2M1	adaptor-related protein complex 2, mu 1 subunit	8
150468	CKAP2L	cytoskeleton associated protein 2-like	8
155061	ZNF746	zinc finger protein 746	8
1595	CYP51A1	cytochrome P450, family 51, subfamily A, polypeptide 1	8
1677	DFFB	DNA fragmentation factor, 40kDa, beta polypeptide (caspase-activated DNase)	8
199777	ZNF626	zinc finger protein 626	8
221476	PI16	peptidase inhibitor 16	8
222235	FBXL13	F-box and leucine-rich repeat protein 13	8
222696	ZSCAN23	zinc finger and SCAN domain containing 23	8
22822	PHLDA1	pleckstrin homology-like domain, family A, member 1	8
22900	CARD8	caspase recruitment domain family, member 8	8

【図 8 - 4】

22948	CCT5	chaperonin containing TCP1, subunit 5 (epsilon)	8
2313	FLI1	FLI-1 proto-oncogene, ETS transcription facto	8
23426	GRIP1	glutamate receptor interacting protein 1	8
25	ABL1	ABL proto-oncogene 1, non-receptor tyrosine kinase	8
2532	ACKR1	atypical chemokine receptor 1 (Duffy blood group)	8
25821	MTO1	mitochondrial tRNA translation optimization	8
27102	EIF2AK1	eukaryotic translation initiation factor 2-alpha kinase 1	8
27295	PDLIM3	PDZ and LIM domain 3	8
2842	GPR19	G protein-coupled receptor 19	8
29882	ANAPC2	anaphase promoting complex subunit 2	8
3337	DNAJB1	DnaJ (Hsp40) homolog, subfamily B, member 1	8
3486	IGFBP3	insulin-like growth factor binding protein 3	8
3712	IVD	isovaleryl-CoA dehydrogenase	8
3821	KLRC1	killer cell lectin-like receptor subfamily C, member 1	8
4882	NPR2	nutriuretic peptide receptor 2	8
51084	CRYLY1	crystallin, lambda 1	8
51295	ECSIT	ECSIT signalling integrator	8
55366	LGR4	leucine-rich repeat containing G protein-coupled receptor 4	8
56479	KCNQ5	potassium voltage-gated channel, KQT-like subfamily, member 5	8
56906	THAP10	THAP domain containing 10	8
56942	CMC2	C-x(9)-C motif containing 2	8
56999	ADAMTS9	ADAM metallopeptidase with thrombospondin type 1 motif, 9	8
57188	ADAMTS3	ADAMTS-like 3	8
57613	KIAA1467	KIAA1467	8
58487	CREB2F	CREB/ATF bZIP transcription factor	8
6338	SCNN1B	sodium channel, non-voltage-gated 1, beta subunit	8
63894	VIPAS39	VPS33B interacting protein, apical-basolateral polarity regulator, spe-39 homolog	8
65082	VPS33A	vacuolar protein sorting 33 homolog A (S. cerevisiae)	8
7053	TGM3	transglutaminase 3	8
7058	THBS2	thrombospondin 2	8

【図 8 - 5】

7881	KCNAB1	potassium voltage-gated channel, shaker-related subfamily, beta member 1	8
79143	MBOAT7	membrane bound O-acyltransferase domain containing 7	8
79152	FA2H	fatty acid 2-hydroxylase	8
80298	MTERF2	mitochondrial transcription termination factor 2	8
84034	EMILIN2	elastin microfibril interfacer 2	8
84951	TNS4	tensin 4	8
8578	SCARF1	scavenger receptor class F, member 1	8
8642	DCHS1	dachshous cadherin-related 1	8
8738	CRADD	CASP2 and RIPK1 domain containing adaptor with death domain	8
8874	ARHGEF7	Rho guanine nucleotide exchange factor (GEF) 7	8
89932	PAPLN	papilin, proteoglycan-like sulfated glycoprotein	8
9139	CBFA2T2	core-binding factor, runt domain, alpha subunit 2; translocated to, 2	8
9194	SLC16A7	solute carrier family 16 (monocarboxylate transporter), member 7	8
928	CD9	CD9 molecule	8
9957	HS3ST1	heparan sulfate (glucosamine) 3-O-sulfotransferase 1	8
100101467	ZSCAN30	zinc finger and SCAN domain containing 30	7
10101	NUBP2	nucleotide binding protein 2	7
10766	TOB2	transducer of ERBB2, 2	7
151194	METTL21A	methyltransferase like 21A	7
1975	EIF4B	eukaryotic translation initiation factor 4B	7
1994	ELAVL1	ELAV like RNA binding protein 1	7
221785	ZSCAN25	zinc finger and SCAN domain containing 25	7
3305	HSPA1L	heat shock 70kDa protein 1-like	7
359821	MRPL42P5	mitochondrial ribosomal protein L42 pseudogene 5	7
3880	KRT19	keratin 19	7
441155	LOC441155	zinc finger CCCH-type domain-containing-like	7
54472	TOLLIP	toll interacting protein	7
55005	RMND1	required for meiotic nuclear division 1 homolog (S. cerevisiae)	7
55827	DCAF6	DDB1 and CUL4 associated factor 6	7
56992	KIF15	kinesin family member 15	7

【図 8 - 6】

5918	RARRES1	retinoic acid receptor responder (tazarotene induced) 1	7
64236	PDLIM2	PDZ and LIM domain 2 (mystique)	7
79022	TMEM106C	transmembrane protein 106C	7
79168	LILRA6	leukocyte immunoglobulin-like receptor, subfamily A (with TM domain), member 6	7
80000	GREB1L	growth regulation by estrogen in breast cancer like	7
80219	COQ10B	coenzyme Q10 homolog B (S. cerevisiae)	7
80279	CDK5RAP3	CDK5 regulatory subunit associated protein 3	7
84131	CEP78	centrosomal protein 78kDa	7
84984	CEP19	centrosomal protein 19kDa	7
8925	HERC1	HECT and RLD domain containing E3 ubiquitin protein ligase family member 1	7
9559	VPS26A	vacuolar protein sorting 26 homolog A (S. pombe)	7
994	CDC25B	cell division cycle 25B	7
100506736	SLFN12L	schlafen family member 12-like	6
10260	DFNND4A	DFNN/MADD domain containing 4A	6
10985	GCN1L1	GCN1 general control of amino-acid synthesis 1-like 1 (yeast)	6
165140	OXR1	oxo-icosanoid (OXE) receptor 1	6
23294	ANKS1A	ankyrin repeat and sterile alpha motif domain containing 1A	6
3123	HLA-DRB1	major histocompatibility complex, class II, DR beta 1	6
338692	ANKRD13D	ankyrin repeat domain 13 family, member D	6
3516	RBPJ	recombination signal binding protein for immunoglobulin kappa J region	6
3673	ITGA2	integrin, alpha 2 (CD49B, alpha 2 subunit of VLA-2 receptor)	6
54434	SS11	slingshot protein phosphatase 1	6
546	ATRX	alpha thalassemia/mental retardation syndrome X-linked	6
57335	ZNF286A	zinc finger protein 286A	6
57511	COG6	component of oligomeric golgi complex 6	6
64848	YTHDC2	YTH domain containing 2	6
653361	NC1	neutrophil cytosolic factor 1	6
7572	ZNF24	zinc finger protein 24	6
81027	TUBB1	tubulin, beta 1 class VI	6
84656	GLYR1	glyoxylate reductase 1 homolog (Arabidopsis)	6

【図8-7】

84824	FCRLA	Fc receptor-like A	6
84868	HAVCR2	hepatitis A virus cellular receptor 2	6
9648	GCC2	GRIP and coiled-coil domain containing 2	6
10482	NXF1	nuclear RNA export factor 1	5
1106	CHD2	chromodomain helicase DNA binding protein 2	5
27237	ARHGEF16	Rho guanine nucleotide exchange factor (GEF) 16	5
3716	JAK1	Janus kinase 1	5
3831	KLC1	kinesin light chain 1	5
3965	LGALS9	lectin, galactoside-binding, soluble, 9	5
4052	LTBP1	latent transforming growth factor beta binding protein 1	5
440145	MZT1	mitotic spindle organizing protein 1	5
56107	PCDHGA9	protocadherin gamma subfamily A, 9	5
6490	PME1	premelanosome protein	5
80781	COL18A1	collagen, type XVIII, alpha 1	5
84334	APOPT1	apoptogenin 1, mitochondrial	5
100532731	COMM3-BM11	COMM3-BM11 readthrough	4
100532746	PPT2-EGFL8	PPT2-EGFL8 readthrough (NMD candidate)	4
10521	DDX17	DEAD (Asp-Glu-Ala-Asp) box helicase 17	4
10569	SLU7	SLU7 splicing factor homolog (<i>S. cerevisiae</i>)	4
1066	CES1	carboxylesterase 1	4
10767	HBS1L	HBS1-like translational GTPase	4
115677	NOSTRIN	nitric oxide synthase trafficking	4
149473	CCDC24	coiled-coil domain containing 24	4
1806	DGKA	diacylglycerol kinase, alpha 80kDa	4
		microtubule associated serine/threonine kinase 3	4
23031	MAST3	peroxiredoxin 5	4
25824	PRDX5	adenylate kinase 5	4
26289	AK5	torsin family 2, member A	4
27433	TOR2A	glycerol-3-phosphate dehydrogenase 2 (mitochondrial)	4
2820	GPD2	aquaporin 11	4
282679	AQP11		4

【図8-8】

3073	HEXA	hexosaminidase A (alpha polypeptide)	4
3159	HMGAA1	high mobility group AT-hook 1	4
3371	TNC	tenascin C	4
4671	NAIP	NLR family, apoptosis inhibitory protein	4
468	ATF4	activating transcription factor 4	4
51154	MRTO4	mRNA turnover 4 homolog (<i>S. cerevisiae</i>)	4
54439	RBM27	RNA binding motif protein 27	4
54471	MIEF1	mitochondrial elongation factor 1	4
		Rho guanine nucleotide exchange factor (GEF) 16	4
54848	ARHGEF38	ER membrane-associated RNA degradation	4
55780	ERMARD	pregnancy specific beta-1-glycoprotein 1	4
5669	PSG1	ribonucleotide reductase M2	4
6241	RRM2	secreted frizzled-related protein 2	4
6423	SFRP2	limb development membrane protein 1	4
64327	LMBR1	inverted formin, FH2 and WH2 domain containing	4
64423	INF2	tubulin folding cofactor D	4
6904	TBCD	family with sequence similarity 160, member A1	4
729830	FAM160A1	autophagy related 9A	4
79065	ATG9A	histone cluster 1, H2bo	4
8348	HIST1H2BO	phosphatidylinositol acid phosphatase type 2 domain containing 1B	4
84513	PPAPDC1B	carboxypeptidase Z	4
8532	CPZ	ubiquitin specific peptidase 2	4
9099	USP2	palmitoyl-protein thioesterase 2	4
9374	PPT2	RNAse K-C17orf49 readthrough	3
100529209	RNASEK-C17orf49	nine-amino acid residue-repeats	3
100861437	NARR	testis-specific kinase 2	3
10420	TESK2	myotubularin related protein 11	3
10903	MTMR11	V-set and immunoglobulin domain containing 4	3
11326	VSIG4	XK, Kell blood group complex subunit-related family, member 4	3
114786	XKR4	MAS-related GPR, member E	3
116534	MRGPRE		3

【図8-9】

122773	KLHDC1	kelch domain containing 1	3
1286	COL4A4	collagen, type IV, alpha 4	3
1491	CTH	cystathione gamma-lyase	3
151648	SGOL1	shugoshin-like 1 (<i>S. pombe</i>)	3
157769	FAM91A1	family with sequence similarity 91, member A1	3
162073	ITPR1PL2	inositol 1,4,5-trisphosphate receptor interacting protein-like 2	3
166378	SPATA5	spermatogenesis associated 5	3
1730	DIAPH2	diaphanous-related formin 2	3
197135	PATL2	protein associated with topoisomerase II homolog 2 (yeast)	3
203523	ZNF449	zinc finger protein 449	3
2051	EPHB6	EPH receptor B6	3
22862	FNDC3A	fibronectin type III domain containing 3A	3
23158	TBC1D9	TBC1 domain family, member 9 (with GRAM domain)	3
23264	ZC3H7B	zinc finger CCCH-type containing 7B	3
3303	HSPA1A	heat shock 70kDa protein 1A	3
4689	NCF4	neutrophil cytosolic factor 4, 40kDa	3
55331	ACER3	alkaline ceramidase 3	3
6929	TCF3	transcription factor 3	3
100131439	CD300LD	CD300 molecule-like family member d	2
100526740	ATP5J2-PTCD1	ATP5J2-PTCD1 readthrough	2
10380	BPN11	3'(2'), 5'-bisphosphate nucleotidase 1	2
10687	PNMA2	paraneoplastic Ma antigen 2	2
107	ADCY1	adenylyl cyclase 1 (brain)	2
113146	AHNAK2	AHNAK nucleoprotein 2	2
116443	GRIN3A	glutamate receptor, ionotropic, N-methyl-D-aspartate 3A	2
1211	CLTA	clathrin, light chain A	2
127731	VWA5B1	von Willebrand factor A domain containing 5B1	2
130162	CLHC1	clathrin heavy chain linker domain containing 1	2
131870	NUDT16	nudix (nucleoside diphosphate linked moiety X)-type motif 16	2
140691	TRIM69	tripartite motif containing 69	2

【図8-10】

149175	MANEAL	mannosidase, endo-alpha-like	2
158358	KIAA2026	KIAA2026	2
1622	DBI	diazepam binding inhibitor (GABA receptor modulator, acyl-CoA binding protein)	2
1979	EIF4EBP2	eukaryotic translation initiation factor 4E binding protein 2	2
2033	EP300	E1A binding protein p300	2
206358	SLC36A1	solute carrier family 36 (protryptophanyl transporter), member 1	2
221150	SKA3	spindle and kinetochore associated complex subunit 3	2
23148	NACAD	NAC alpha domain containing	2
23163	GGA3	golgi-associated, gamma adaptin ear containing, ARF binding protein 3	2
23171	GPD1L	glycerol-3-phosphate dehydrogenase 1-like	2
23237	ARC	activity-regulated cytoskeleton-associated protein	2
23227	FMO2	flavin containing monooxygenase 2 (non-functional)	2
23299	BICD2	bicaudal D homolog 2 (<i>Drosophila</i>)	2
23365	ARHGEFF12	Rho guanine nucleotide exchange factor (GEF) 12	2
23412	COMM3	COMM domain containing 3	2
246176	GAS2L2	growth arrest-specific 2 like 2	2
25794	FSCN2	fascin actin-bundling protein 2, retinal	2
26024	PTCD1	pentatricopeptide repeat domain 1	2
26145	IRFBP1	interferon regulatory factor 2 binding protein 1	2
261734	NPHIP4	nephronophthisis 4	2
283431	GAS2L3	growth arrest-specific 2 like 3	2
284996	RNF149	ring finger protein 149	2
286207	C9orf117	chromosome 9 open reading frame 117	2
2907	GRINA	glutamate receptor, ionotropic, N-methyl-D-aspartate-associated protein 1 (glutamate binding)	2
2949	GSTM5	glutathione S-transferase mu 5	2
29929	ALG6	ALG6, alpha-1,3-glucosyltransferase	2
339804	C2orf74	chromosome 2 open reading frame 74	2
340543	TCEAL5	transcription elongation factor A (SII)-like 5	2
342897	NCCRP1	non-specific cytotoxic cell receptor protein 1 homolog (<i>zebrafish</i>)	2
348094	ANKDD1A	ankyrin repeat and death domain containing 1A	2

【図 8 - 1 1】

348180	CTU2	cytosolic thiouridylase subunit 2 homolog (S. pombe)	2
3611	ILK	integrin-linked kinase	2
387032	ZKSCAN4	zinc finger with KRAB and SCAN domains 4	2
387111	LINC00222	long intergenic non-protein coding RNA 222	2
400673	VMAC	vimentin-type intermediate filament associated coiled-coil protein	2
4171	MCM2	minichromosome maintenance complex component 2	2
4327	MMP19	matrix metallopeptidase 19	2
494470	RNF165	ring finger protein 165	2
5087	PBX1	pre-B-cell leukemia homeobox 1	2
51559	NT5DC3	5'-nucleotidase domain containing 3	2
54756	IL17RD	interleukin 17 receptor D	2
55197	RPRD1A	regulation of nuclear pre-mRNA domain containing 1A	2
55238	SLC38A7	solute carrier family 38, member 7	2
55326	AGPAT5	1-acylglycerol-3-phosphate O-acyltransferase 5	2
55686	MREG	melanoregulin	2
55759	WDR12	WD repeat domain 12	2
56898	BDH2	3-hydroxybutyrate dehydrogenase, type 2	2
5891	MOK	MOK proto-oncogene	2
64174	DPEP2	dipeptidase 2	2
645121	CCNI2	cyclin I family, member 2	2
692312	PPAN-P2RY11	PPAN-P2RY11 readthrough	2
7355	SLC35A2	solute carrier family 35 (UDP-galactose transporter), member A2	2
7465	WEE1	WEE1 G2 checkpoint kinase	2
		mitochondrial transcription termination factor 1	2
7978	MTERF1	serine palmitoyltransferase, long chain base subunit 2	2
9517	SPTLC2	piezo-type mechanosensitive ion channel component 1	2
9780	PIEZ01	ZNF816-ZNF321P	1
100529240	ZNF816-ZNF321P	ZNF816-ZNF321P readthrough	1
10385	BTN2A2	butyrophilin, subfamily 2, member A2	1
10523	CHERP	calcium homeostasis endoplasmic reticulum protein	1
10630	PDPN	podoplanin	1

【図 8 - 1 2】

11060	WWP2	WW domain containing E3 ubiquitin protein ligase 2	1
11162	NUDT6	nudix (nucleoside diphosphate linked moiety X)-type motif 6	1
114904	C1QTNF6	C1q and tumor necrosis factor related protein 6	1
116541	MRPL54	mitochondrial ribosomal protein L54	1
1281	COL3A1	collagen, type III, alpha 1	1
145173	B3GALT1	beta 1,3-galactosyltransferase-like	1
146722	CD300LF	CD300 molecule-like family member f	1
154796	AMOT	angiomotin	1
163259	DENN2C	DENN/MADD domain containing 2C	1
165055	CCDC138	coiled-coil domain containing 138	1
165215	FAM171B	family with sequence similarity 171, member B	1
2125	EVPL	envoplakin	1
222487	GPR97	G protein-coupled receptor 97	1
2235	FECH	ferrochelatase	1
		TBC1 (tre-2/USP6, BUB2, cdc16) domain family, member 1	1
23216	TBC1D1	KIAA0922	1
23240		solute carrier family 7 (anionic amino acid transporter light chain, xc- system), member 11	1
23683	PRKD3	protein kinase D3	1
25800	SLC39A6	solute carrier family 39 (zinc transporter), member 6	1
26278	SACS	sacsin molecular chaperone	1
283284	IGSF22	immunoglobulin superfamily, member 22	1
283578	TMED8	transmembrane emp24 protein transport domain containing 8	1
283848	CE54A	carboxylesterase 4A	1
284759	SIRPB2	signal-regulatory protein beta 2	1
2891	GRIA2	glutamate receptor, ionotropic, AMPA 2	1
2931	GSK3A	glycogen synthase kinase 3 alpha	1
29841	GRHL1	grainyhead-like 1 (Drosophila)	1
29964	PRICKLE4	prickle homolog 4 (Drosophila)	1
30836	DNTTIP2	deoxynucleotidyltransferase, terminal, interacting protein 2	1
3094	HINT1	histidine triad nucleotide binding protein 1	1

【図 8 - 1 3】

3310	HSPA6	heat shock 70kDa protein 6 (HSP70B)	1
340481	ZDHHC21	zinc finger, DHHC-type containing 21	1
3431	SP110	SP110 nuclear body protein	1
3613	IMPA2	inositol(myo)-1(or 4)-monophosphatase 2	1
376267	RAB15	RAB15, member RAS oncogene family	1
3816	KLK1	kallikrein 1	1
389677	RBM12B	RNA binding motif protein 12B	1
399669	ZNF321P	zinc finger protein 321, pseudogene	1
400569	MED11	mediator complex subunit 11	1
4058	LTK	leukocyte receptor tyrosine kinase	1
440482	ANKRD20A5P	ankyrin repeat domain 20 family, member A5, pseudogene	1
441108	C5orf56	chromosome 5 open reading frame 56	1
4668	NAGA	N-acetylgalactosaminidase, alpha-	1
4735	SEPT2	septin 2	1
4853	NOTCH2	notch 2	1
5032	P2RY11	purinergic receptor P2Y, G-protein coupled, 11	1
51252	FAM178B	family with sequence similarity 178, member B	1
5127	CDK16	cyclin-dependent kinase 16	1
51291	GMIP	GEM interacting protein	1
5433	POLR2D	polymerase (RNA) II (DNA directed) polypeptide D	1
54662	TBC1D13	TBC1 domain family, member 13	1
54977	SLC25A38	solute carrier family 25, member 38	1
55001	TTC22	tetratricopeptide repeat domain 22	1
55030	FBXO34	F-box protein 34	1
55039	TRMT12	tRNA methyltransferase 12 homolog (S. cerevisiae)	1
55314	TMEM144	transmembrane protein 144	1
55593	OTUD5	OTU deubiquitinase 5	1
55723	ASF1B	anti-silencing function 1B histone chaperone	1
55734	ZFP64	ZFP64 zinc finger protein	1
5670	PSG2	pregnancy specific beta-1-glycoprotein 2	1

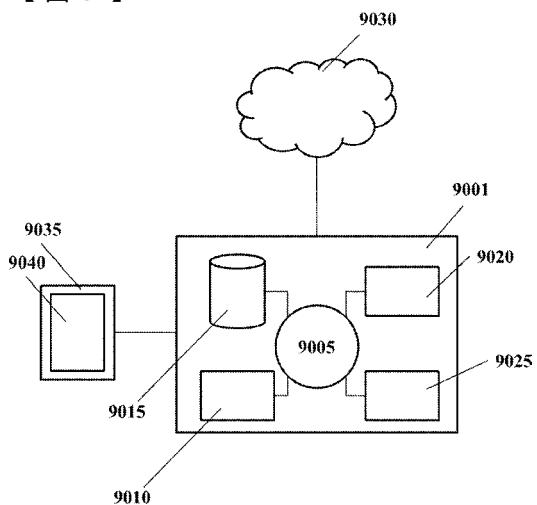
【図 8 - 1 4】

56832	IFNK	interferon, kappa	1
56987	BBX	bobby sox homolog (Drosophila)	1
57191	VN1R1	vomeronasal 1 receptor 1	1
57455	REXO1	REX1, RNA exonuclease 1 homolog (S. cerevisiae)	1
57458	TMCC3	transmembrane and coiled-coil domain family 3	1
57553	MICAL3	microtubule associated monooxygenase, calponin and LIM domain containing 3	1
58526	MID1IP1	MID1 interacting protein 1	1
5887	RAD23B	RAD23 homolog B (S. cerevisiae)	1
59067	IL21	interleukin 21	1
5970	RELA	v-rel avian retroviroblastosis viral oncogene homolog A	1
60490	PPCDC	phosphopantetheoylcysteine decarboxylase	1
6137	RPL13	ribosomal protein L13	1
6273	S100A2	S100 calcium binding protein A2	1
637	BID	BH3 interacting domain death agonist	1
63916	ELMO2	engulfment and cell motility 2	1
64397	ZNF105	zinc finger protein 106	1
64754	SMYD3	SET and MYND domain containing 3	1
64761	PARP12	poly (ADP-ribose) polymerase family, member 12	1
684	BST2	bone marrow stromal cell antigen 2	1
		transcription factor AP-2 gamma (activating enhancer binding protein 2 gamma)	1
7022	TFAP2C	trichorhinophalangeal syndrome I	1
7227	TRPS1	DnaJ (Hsp40) homolog, subfamily C, member 7	1
7266	DNAJC7	zinc finger protein 286B	1
729288	ZNF286B	SEC14-like 6 (S. cerevisiae)	1
730005	SEC14L6	ubiquitin-conjugating enzyme E2H	1
7328	UBE2H	zinc finger protein 16	1
7564	ZNF16	interferon-related developmental regulator 2	1
79066	IFRD2	methyltransferase like 16	1
79070	METTL16	KDEL (Lys-Asp-Glu-Leu) containing 1	1
79661	NEIL1	nei endonuclease VIII-like 1 (E. coli)	1

【図 8 - 15】

79665	DHX40	DEAH (Asp-Glu-Ala-His) box polypeptide 40	1
79800	CARF	calcium responsive transcription factor	1
79937	CNTNAP3	contactin associated protein-like 3	1
80210	ARMC9	armadillo repeat containing 9	1
84216	TMEM117	transmembrane protein 117	1
84255	SLC37A3	solute carrier family 37, member 3	1
84989	JMJD1C-AS1	JMJD1C antisense RNA 1	1
8570	KHSRP	KH-type splicing regulatory protein	1
8702	B4GALT4	UDP-Gal:betaGlcNAc beta 1,4-galactosyltransferase, polypeptide 4	1
8767	RIPK2	receptor-interacting serine-threonine kinase 2	1
8897	MTMR3	myotubularin related protein 3	1
9013	TAF1C	TATA box binding protein (TBP)-associated factor, RNA polymerase I, C, 110kDa	1
9033	PKD2L1	polycystic kidney disease 2-like 1	1
9130	FAM150A	family with sequence similarity 50, member A	1
91947	ARRDC4	arrestin domain containing 4	1
9249	DHRS3	dehydrogenase/reductase (SDR family) member 3	1
93594	TBC1D31	TBC1 domain family, member 31	1
954	ENTPD2	ectonucleoside triphosphate diphosphohydrolase 2	1
9744	ACAP1	ArfGAP with coiled-coil, ankyrin repeat and PH domains 1	1
9860	LRIG2	leucine-rich repeats and immunoglobulin-like domains 2	1
9936	CD302	CD302 molecule	1

【図 9】



【図 10】

Afirma GEcv1訓練ラベル

組織学的良性 細胞学的悪性

ATA 2009リスク訓練ラベル

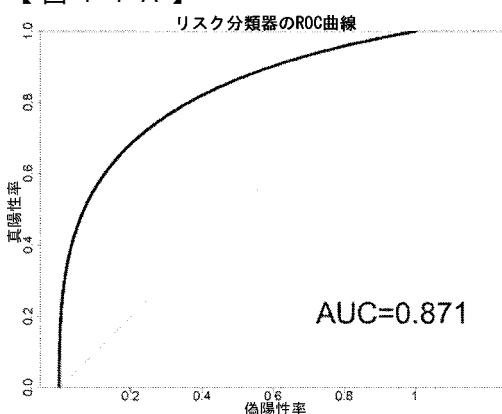
低リスク

中/高リスク

病理学的報告は
リスク特徴の
非存在を記述している
はずである

- 1.リンパ節転移
2.血管侵襲
3.甲状腺外拡張

【図 11 A】



【図 12】

[1分割]での頻度	遺伝子記号	説明
10	COL1A1	collagen, type I, alpha 1
10	FITM2	fat storage-inducing transmembrane protein 2
10	AASDH	aminoacidate-semialdehyde dehydrogenase
10	COX6C	cytochrome c oxidase subunit VIc
10	COX10	cytochrome c oxidase assembly homolog 10 (酵母)
10	VSIG10L	V-set and immunoglobulin domain containing 10 like
10	MAPK15	mitogen-activated protein kinase 15
10	PAXIP1	PAX interacting (with transcription-activation domain) protein 1
10	AVL9	AVL9 homolog (酵母 S. cerevisiae)
10	GIGYF2	GIGYF1 interacting GYF protein 2
10	HLA-DQA1	major histocompatibility complex, class II, DQ alpha 1
10	HLA-DQB1	major histocompatibility complex, class II, DQ beta 1
10	HLA-DRA	major histocompatibility complex, class II, DR alpha
10	HLA-H	major histocompatibility complex, class I, H (pseudogene)
10	MGAT1	mannosyl (alpha-1,3-) glycoprotein beta-1,2-N-acetylglucosaminyltransferase
10	SLC41A3	solute carrier family 41, member 3
10	PTPRE	protein tyrosine phosphatase, receptor type, E
10	SRL	sarcalumenin
10	SLC30A3	solute carrier family 30 (zinc transporter), member 5
10	BMP4	bone morphogenic protein 4
10	ICE2	interactor of little elongator complex ELL subunit 2
10	DCAKD	dephospho-CoA kinase domain containing
10	TMX1	thioredoxin-related transmembrane protein 1
10	HAVCR2	hepatitis A virus cellular receptor 2
10	TNFSF12	tumor necrosis factor (ligand) superfamily, member 12
10	PER2	period circadian clock 2
10	MCM3AP	minichromosome maintenance complex component 3 associated protein

【図 11 B】

	中/高リスク (n=50)	低リスク (n=25)
「中/高リスク」 に分類	43	7
「低リスク」 に分類	7	25

感度 = 86%

特異度 = 86%

(义 1 3)

10割合での 頻度	遺伝子記号	説明
10	COL1A1	collagen, type I, alpha 1
10	NUP210L	nucleoporin 210kDa-like
10	TMEM92	transmembrane protein 92
10	C6orf136	chromosome 6 open reading frame 136
9	SPAG4	sperm associated antigen 4
9	EHF	ets homologous factor
9	RAPGEFS	Rap guanine nucleotide exchange factor (GEF) 5
9	COL3A1	collagen, type III, alpha 1
8	GALNT15	polypeptide N-acetylgalactosaminyltransferase 15
8	PRICKLE1	prickle homolog 1(シワコウヤエ)(Drosophila)
8	LUM	lumican
8	COL6A3	collagen, type VI, alpha 3
8	ROBO1	roundabout, axon guidance receptor, homolog 1(シワコウヤエ)
8	SSC5D	scavenger receptor cysteine rich family, 5 domains
8	PSORS1C1	psoriasis susceptibility 1 candidate 1

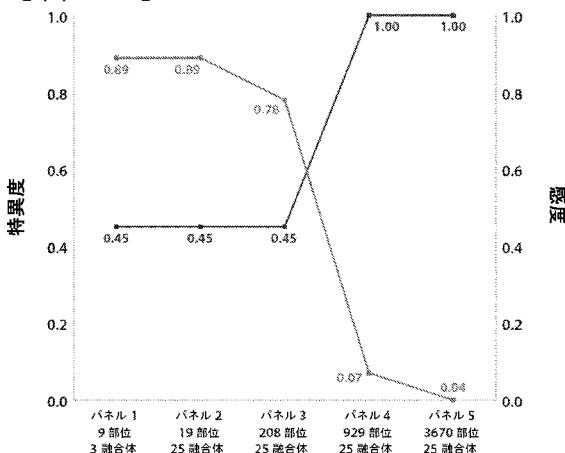
【 四 1 4 】

変異パネル	ゲノム部位	融合対
パネル 1	9	3
パネル 2	19	25
パネル 3	208	25
パネル 4	929	25
パネル 5	3670	25

変異バナルにおいて標的とされた遺伝子*						
AKT1	BRAF**	CTNNB1	EIF1AX	GNAS	HRAS**	KRAS**
NRAS**	PIK3CA	PTEN	RET	TERT***	TSHZ***	TP53

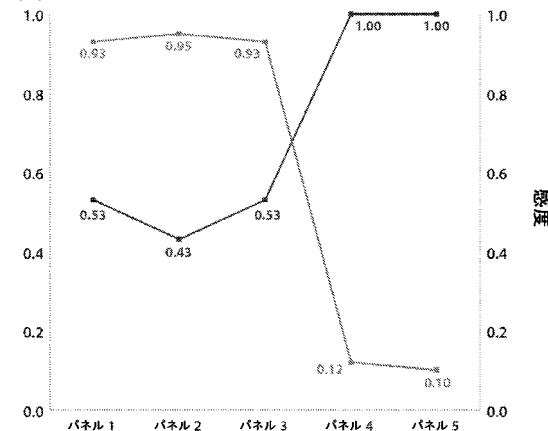
* パネル2~5は、Nikiforov et al.によって報告されたこれらの遺伝子および融合体を含む
 ** パネル1は、Beaudenon et al.によって報告された遺伝子および融合体のサブセットのみを評価する
 *** パネル5にのみ含まれる

(17)



・合計57個の組織を評価し、38個のみが試験パフォーマンス算出に使用される
明確な組織学的良性または組織学的悪性の病態を有していた。

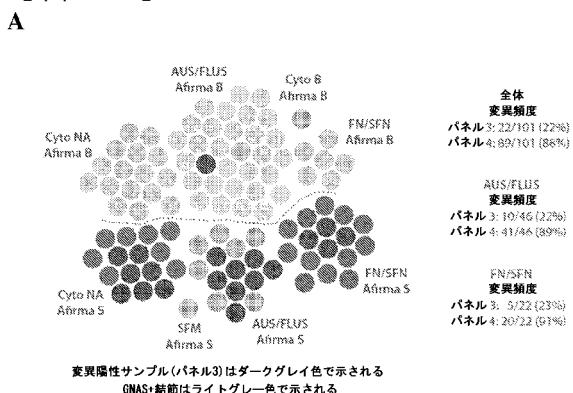
(15)



3 融合体

ペセダ 細胞学 カテゴリー	総サンプル 数	組織学的 良性変異 (+)/全体	組織学的 悪性変異 (+)/全体	感度 (95% CI)	特異度 (95% CI)
Cyto B	28	2/22	3/6	0.50(0.12-0.88)	0.91(0.71-0.99)
AUS/FLUS	12	1/11	0/1	0(0-0.98)	0.91(0.59-1.00)
FN/SFN	9	0/7	1/2	0.50(0.01-0.99)	1.00(0.59-1.00)
SFM	12	0/1	4/11	0.36(0.11-0.69)	1.00(0.03-1.00)
Cyto M	20	0/0	13/20	0.65(0.41-0.85)	NA
全サンプル	81	3/41	21/40	0.53(0.36-0.68)	0.93(0.80-0.98)

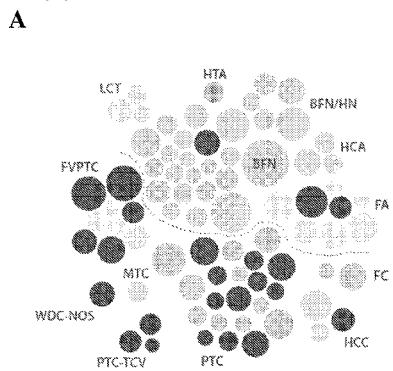
【 図 1 8 】



B

分子	変異	Affirma B (n=54)		Affirma S (n=47)		NA (n=16)
		AUS/FLUS (n=31)	AUS/FLUS (n=15)	FN/SFN (n=16)		
BRAF	K601E					1
GNAS	R201H	1				
HRAS	Q61K	1				
HRAS	Q61R				1	
KRAS	G12D		1			
KRAS	Q61R		1		2	1
NRAS	G12V		1			
NRAS	Q61R		1		1	4
TP53	P152L		1			
TP53	R213*					1
融合体	PAX8/PPARG			3		

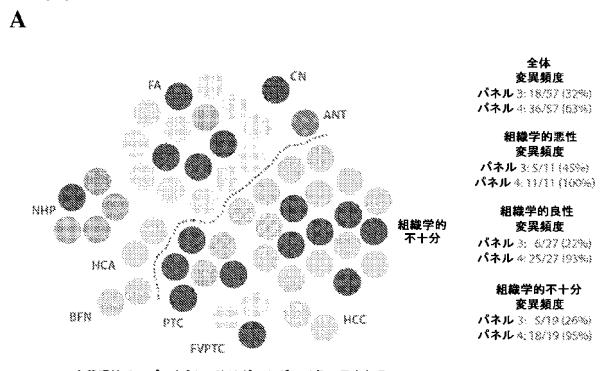
【 図 1 9 】



変異陽性結節(パネル3)はダークグレイ色で示される。結節はサイズに比例して示される。最小の結節=1cm。

B

【 図 2 0 】



変異陽性サンプル(パネル3)はダークグレイ色で示される
GNAS+結節はライトグレー色で示される

B

パネル3で検出	遺伝子名	組織学的良性(n=27)			組織学的悪性(n=33)		組織学的不十分(n=19)	
		Affirma B(n=8)	Affirma S(n=18)	Affirma NA(n=1)	Affirma A5(n=11)	Affirma S(n=16)		
BRAF	V600E	-	-	-	-	-	5	-
CNAAS	Q225H	-	-	1	-	-	-	-
NRAS	G12K	-	-	-	1	-	-	3
NRAS	G51R	-	-	-	1	-	-	2
TSHR	D93Y	-	2	-	-	-	-	-
TSHR	I486F	-	-	-	1	-	-	-
融合体	PAX8/PARG	-	-	-	-	-	-	5
融合体	CCDC6/RET	-	-	-	1	-	-	-

【国際調査報告】

INTERNATIONAL SEARCH REPORT		International application No. PCT/US2016/020583
A. CLASSIFICATION OF SUBJECT MATTER C12Q 1/68(2006.01)i, G06F 19/18(2011.01)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) C12Q 1/68; A61N 5/00; G06F 19/18		
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: recurrence, evaluation, biomarker, expression level, variant, needle aspirate sample, thyroid		
C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 2013-0303826 A1 (JURISICA et al.) 14 November 2013 See abstract; paragraphs [0089], [0101], [0110], [0127], [0206], [0320], [0332], [0333]; and claims 2-15.	1,2,4-52
A	WO 2013-063544 A1 (ASURAGEN, INC.) 02 May 2013 See the whole document.	1,2,4-52
A	US 2009-0280490 A1 (BAKER et al.) 12 November 2009 See the whole document.	1,2,4-52
A	MIYAMOTO et al., 'Potential marker of oral squamous cell carcinoma aggressiveness detected by fluorescence in situ hybridization in fine-needle aspiration biopsies' Cancer, Vol. 95, No. 10, pp. 2152-2159 (2002) See the whole document.	1,2,4-52
A	CHENG et al., 'A multi-cancer mesenchymal transition gene expression signature is associated with prolonged time to recurrence in glioblastoma' PLoS One, Vol. 7, Issue 4, Article No. e34705, pages 1-8 (2012) See the whole document.	1,2,4-52
<input type="checkbox"/> Further documents are listed in the continuation of Box C.		<input checked="" type="checkbox"/> See patent family annex.
<p>* 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</p> <p>"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</p>		
Date of the actual completion of the international search 02 June 2016 (02.06.2016)	Date of mailing of the international search report 02 June 2016 (02.06.2016)	
Name and mailing address of the ISA/KR International Application Division Korean Intellectual Property Office 189 Cheongsa-ro, Seo-gu, Daejeon, 35208, Republic of Korea Facsimile No. +82-42-481-8578	Authorized officer KIM, Seung Beom Telephone No. +82-42-481-3371	

INTERNATIONAL SEARCH REPORT

International application No. PCT/US2016/020583

Box No. II Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. Claims Nos.: 3
because they relate to subject matter not required to be searched by this Authority, namely:
Claim 3 pertains to a method for treatment of the human body by surgery, and thus relates to a subject matter which this International Searching Authority is not required, under PCT Article 17(2)(a)(i) and PCT Rule 39.1(iv), to search.
2. Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:
3. Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. As all searchable claims could be searched without effort justifying an additional fees, this Authority did not invite payment of any additional fees.
3. As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:
4. No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

- The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.
- The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.
- No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/US2016/020583

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 2013-0303826 A1	14/11/2013	EP 2663672 A1 WO 2012-094744 A1	20/11/2013 19/07/2012
WO 2013-063544 A1	02/05/2013	EP 2771487 A1 US 2013-142728 A1	03/09/2014 06/06/2013
US 2009-0280490 A1	12/11/2009	EP 1644858 A2 EP 1644858 A4 JP 2007-527220 A JP 4906505 B2 US 2005-0048542 A1 US 7526387 B2 US 7939261 B2 WO 2005-008213 A2 WO 2005-008213 A3	12/04/2006 13/05/2009 27/09/2007 28/03/2012 03/03/2005 28/04/2009 10/05/2011 27/01/2005 24/03/2005

フロントページの続き

(81)指定国 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,R0,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,DK,DM,D0,DZ,EC,EE,EG,ES,FI,GB,GD,GE,GH,GM,GT,HN,HR,HU,ID,IL,IN,IR,IS,JP,KE,KG,KN,KP,KR,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,TZ,UA,UG,US

(特許庁注：以下のものは登録商標)

1. アンドロイド

(74)代理人 100142929
弁理士 井上 隆一
(74)代理人 100148699
弁理士 佐藤 利光
(74)代理人 100128048
弁理士 新見 浩一
(74)代理人 100129506
弁理士 小林 智彦
(74)代理人 100205707
弁理士 小寺 秀紀
(74)代理人 100114340
弁理士 大関 雅人
(74)代理人 100114889
弁理士 五十嵐 義弘
(74)代理人 100121072
弁理士 川本 和弥
(72)発明者 ケネディー ジュリア シー。
アメリカ合衆国 94116 カリフォルニア州 サンフランシスコ カステナダ アベニュー
360
(72)発明者 パガン モライマ
アメリカ合衆国 94117 カリフォルニア州 サンフランシスコ スコット ストリート 6
25 アパートメント 104
(72)発明者 リン チュー - ファン
アメリカ合衆国 94080 カリフォルニア州 サウス サンフランシスコ ショアライン コ
ート 7000 スイート 250
(72)発明者 ファン チン
アメリカ合衆国 94080 カリフォルニア州 サウス サンフランシスコ ショアライン コ
ート 7000 スイート 250
(72)発明者 ウォルシュ ピー. ショーン
アメリカ合衆国 94080 カリフォルニア州 サウス サンフランシスコ ショアライン コ
ート 7000 スイート 250
(72)発明者 松崎 新
アメリカ合衆国 95014 カリフォルニア州 クバチーノ デービソン アベニュー 104
71
(72)発明者 トラバース ケビン
アメリカ合衆国 94080 カリフォルニア州 サウス サンフランシスコ ショアライン コ
ート 7000 スイート 250

(72)発明者 キム ス ヨン

アメリカ合衆国 94080 カリフォルニア州 サウス サンフランシスコ ショアライン コート 7000 スイート 250

F ターム(参考) 4B063 QA13 QA19 QQ02 QQ42 QQ53 QR08 QR55 QR62 QS25 QX02

【要約の続き】
複数のリスク群にわたるパフォーマンス

	中/高リスク (n=50)	低リスク (n=23)
「中/高リスク」に分類	41	6
「低リスク」に分類	9	23
感度	82%	
特異度	79%	

経路または 遺伝子 オントロジー	予想された 遺伝子数	観察された 遺伝子数	P値
細胞外 マトリクス	2	20	1.28×10^{-13}
ECM受容体 相互作用	1	11	6.12×10^{-08}
班接着	2	12	3.71×10^{-05}
チロシン キナーゼ活性	1	6	4.77×10^{-03}
免疫系 プロセスの 調節	3	10	6.83×10^{-02}
血管発生	1	7	1.0×10^{-02}

分類に使用した遺伝子のサンプリング				
COX6C	FANCA	KCTD17	MPRIP	TUBA1B
DCAKD	ICE2	MCM3AP	TNFRSF14	WSB2