

【特許請求の範囲】

【請求項 1】

結腸細胞増殖性障害の特徴を示すマイクロリボ核酸 (miRNA) シグネチャパネルであって、

表 1～表 11 に列挙される群から選択される 1 つもしくは複数、2 つ以上、3 つ以上、または 4 つ以上の miRNA からなる所定のセットであって、前記結腸細胞増殖性障害またはその亜型を患う対象から得た生体試料と、前記結腸細胞増殖性障害またはその亜型を患っていない対象から得た生体試料との間で差動的に発現されるセットを含む miRNA シグネチャパネル。

【請求項 2】

進行型腺腫の特徴を示すとともに、a) hsa-miR-1273a、hsa-miR-17-5p、hsa-miR-20a-3p、hsa-miR-20b-5p、b) hsa-miR-3065-5p、hsa-miR-4785、hsa-miR-5096、hsa-miR-5189-5p、または c) hsa-miR-545-3p、hsa-miR-570-3p、hsa-miR-624-3p、hsa-miR-1181、hsa-miR-6073 を含む miRNA からなる所定のセットを含んでおり、前記 miRNA が、進行型腺腫またはその亜型を患う対象から得た生体試料と、進行型腺腫またはその亜型を患っていない対象から得た生体試料との間で差動的に発現される、請求項 1 に記載の miRNA シグネチャパネル。

【請求項 3】

大腸癌の特徴を示すとともに、a) hsa-miR-1250-5p、hsa-miR-1255a、hsa-miR-223-3p、hsa-miR-338-3p、hsa-miR-338-5p、b) hsa-miR-424-5p、hsa-miR-424-3p、hsa-miR-450a-5p、hsa-miR-450b-5p、hsa-miR-4772-3p、c) hsa-miR-4772-5p、hsa-miR-625-5p、hsa-miR-7847-3p、hsa-miR-1181、hsa-miR-3651、hsa-miR-6073、d) hsa-miR-6125、hsa-miR-7704、hsa-miR-19b-3p、hsa-miR-19a-3p、hsa-miR-3157-5p、e) hsa-miR-142-3p、hsa-miR-30c-5p、hsa-miR-6741-5p、hsa-miR-590-3p、hsa-miR-4685-5p、f) hsa-miR-3648、hsa-miR-331-3p、hsa-miR-1303、hsa-miR-6790-3p、hsa-miR-6867-5p、hsa-miR-942-5p、g) hsa-miR-378a-3p、hsa-miR-1287-5p、hsa-miR-4785、hsa-miR-324-3p、hsa-miR-550b-2-5p、h) hsa-miR-200c-3p、hsa-miR-200b-3p、hsa-miR-3679-5p、hsa-miR-550a-3-5p、hsa-miR-3187-3p、i) hsa-miR-181b-5p、hsa-miR-3138、hsa-miR-146a-5p、hsa-miR-6721-5p、hsa-miR-23b-3p、hsa-miR-28-5p、j) hsa-miR-320d、hsa-miR-940、hsa-miR-320d-1、hsa-miR-10a-5p、hsa-miR-340-5p、k) hsa-miR-320b、hsa-miR-335-5p、hsa-miR-320c、hsa-miR-501-3p、hsa-miR-548n、または l) hsa-miR-27a-3p、hsa-miR-3065-3p、hsa-miR-548aa@、hsa-miR-584-3p、hsa-miR-22-3p を含む miRNA からなる所定のセットを含んでおり、前記 miRNA が、大腸癌またはその亜型を患う対象から得た生体試料と、大腸癌またはその亜型を患っていない対象から得た生体試料との間で差動的に発現される、請求項 1 に記載の miRNA シグネチャパネル。

【請求項 4】

miRNA からなる前記所定のセットが、表 1～表 11 に列挙される群から選択される

少なくとも1、少なくとも2、少なくとも3、少なくとも4、少なくとも5、少なくとも6、少なくとも7、少なくとも8、少なくとも9、少なくとも10、少なくとも11、少なくとも12、少なくとも13、少なくとも14、少なくとも15、少なくとも16、少なくとも17、少なくとも18、少なくとも19、少なくとも20、少なくとも21、少なくとも22、少なくとも23、少なくとも24、少なくとも25、少なくとも26、少なくとも27、少なくとも28、少なくとも29、少なくとも30、少なくとも40、少なくとも50、少なくとも60、少なくとも70、少なくとも80、少なくとも90、少なくとも100、少なくとも110、少なくとも120、少なくとも130、少なくとも140、少なくとも150、少なくとも160、少なくとも170、少なくとも180、少なくとも190、少なくとも200、または少なくとも250のmiRNAを含む、請求項1に記載のmiRNAシグネチャパネル。 10

【請求項5】

前記生体試料が、体液、便、結腸流出液、尿、血漿、血清、全血、単離した血液細胞、血液から単離した細胞、およびそれらの組合せからなる群から選択される、請求項1に記載のmiRNAシグネチャパネル。

【請求項6】

前記生体試料が、核酸、DNA、RNA、または無細胞核酸（cfDNAもしくはcfRNA）を含む、請求項1に記載のmiRNAシグネチャパネル。

【請求項7】

前記miRNAが、成熟miRNAおよびmiRNAヘアピンを含む、請求項1に記載のmiRNAシグネチャパネル。 20

【請求項8】

表1～表11に列挙される群から選択される6つ以上、または12以上のmiRNAにおける差動的発現を含む、請求項1に記載のmiRNAシグネチャパネル。

【請求項9】

前記結腸細胞増殖性障害が、腺腫（腺腫性ポリープ）、固着性鋸歯状腺腫（SSA）、進行型腺腫、大腸異形成、大腸腺腫、大腸癌、結腸癌、直腸癌、大腸癌腫、大腸腺癌、カルチノイド腫瘍、胃腸カルチノイド腫瘍、胃腸間質腫瘍（GIST）、リンパ腫、および肉腫からなる群から選択される、請求項1に記載のmiRNAシグネチャパネル。

【請求項10】

前記結腸細胞増殖性障害が、ステージ1大腸癌、ステージ2大腸癌、ステージ3大腸癌、およびステージ4大腸癌からなる群から選択される、請求項1に記載のmiRNAシグネチャパネル。 30

【請求項11】

健康な対象集団と、結腸細胞増殖性障害を患う対象とを判別可能な分類器であって、

a) 前記分類器は、表1～表11に列挙される群から選択される6つ以上または12以上の予め選択したmiRNAにおける差動的miRNA発現を表す測定値のセットを含み、該測定値は、健康な対象と結腸細胞増殖性障害を患う対象から得たmiRNA発現データから取得され、

b) 前記測定値は、前記差動的miRNA発現の特性に相当する一組の特徴を生成するために使用され、該一組の特徴は、機械学習モデルを使用してコンピュータにより処理され、 40

c) 前記機械学習モデルは、健康な対象集団と、結腸細胞増殖性障害を患う対象とを判別可能な分類器として有用な特徴ベクトルを提供する、分類器。

【請求項12】

前記予め選択したmiRNAが、表1～表11に列挙される群から選択される少なくとも1、少なくとも2、少なくとも3、少なくとも4、少なくとも5、少なくとも6、少なくとも7、少なくとも8、少なくとも9、少なくとも10、少なくとも11、少なくとも12、少なくとも13、少なくとも14、少なくとも15、少なくとも16、少なくとも17、少なくとも18、少なくとも19、少なくとも20、少なくとも21、少なくとも 50

22、少なくとも23、少なくとも24、少なくとも25、少なくとも26、少なくとも27、少なくとも28、少なくとも29、少なくとも30、少なくとも40、少なくとも50、少なくとも60、少なくとも70、少なくとも80、少なくとも90、少なくとも100、少なくとも110、少なくとも120、少なくとも130、少なくとも140、少なくとも150、少なくとも160、少なくとも170、少なくとも180、少なくとも190、少なくとも200、または少なくとも250のmiRNAを含む、請求項11に記載の分類器。

【請求項13】

前記結腸細胞増殖性障害が、腺腫（腺腫性ポリープ）、固着性鋸歯状腺腫（SSA）、進行型腺腫、大腸異形成、大腸腺腫、大腸癌、結腸癌、直腸癌、大腸癌腫、大腸腺癌、カルチノイド腫瘍、胃腸カルチノイド腫瘍、胃腸間質腫瘍（GIST）、リンパ腫、および肉腫からなる群から選択される、請求項11に記載の分類器。

10

【請求項14】

前記測定値のセットが、異なる数を有するフラグメントの観察の数または割合、生のmiRNA存在量、ハウスキーピング遺伝子へと正規化されるmiRNA存在量、合成配列へと正規化されるmiRNA存在量、対数正規化されたmiRNA存在量、フラグメント長、フラグメント中点、成熟miRNAまたはmiRNAヘアピンに沿ったリードマッピング位置およびリードパイルアップ、ならびにmiRNAクラスタの存在量からなる群から選択される差動的miRNA発現の特徴を記述する、請求項11に記載の分類器。

20

【請求項15】

前記機械学習モデルが、訓練生体試料、結腸細胞増殖性障害を患う対象に相当すると特定される前記訓練生体試料の第1のサブセット、および結腸細胞増殖性障害を患っていない対象に相当すると特定される前記訓練生体試料の第2のサブセットから取得した訓練データを用いて訓練される、請求項11に記載の分類器。

【請求項16】

a) miRNAシグネチャパネルに基づき前記対象を分類するのに操作可能な分類器を含むコンピュータ可読媒体と、

b) 前記コンピュータ可読記憶媒体に記憶される命令を実行するための1または複数のプロセッサと

を含む、結腸細胞増殖性障害を検出するためのシステムに設けられる、請求項11に記載の分類器。

30

【請求項17】

前記システムが、深層学習分類器、ニューラルネットワーク分類器、線形判別分析（LDA）分類器、二次判別分析（QDA）分類器、サポート・ベクター・マシン（SVM）分類器、ランダムフォレスト（RF）分類器、線形カーネル・サポート・ベクター・マシン分類器、一次または二次多項式カーネル・サポート・ベクター・マシン分類器、リッジ回帰分類器、弾性ネットアルゴリズム分類器、逐次最小問題最適化アルゴリズム分類器、単純ベイズアルゴリズム分類器、および主成分分析分類器からなる群から選択される機械学習分類器として構成される分類回路を含む、請求項16に記載の分類器。

【請求項18】

対象から得た生体試料のマイクロリボ核酸（miRNA）プロファイルを決定するための方法であって、

40

a) 前記生体試料からRNA分子を単離する工程と、

b) 前記RNA分子を相補的デオキシリボ核酸（cDNA）分子に逆転写する前または後にRNAアダプタを前記RNA分子にライゲートする工程と、

c) 前記cDNA分子を増幅する工程と、

d) 前記cDNA分子の核酸配列を決定する工程と、

e) 表1～表11に列挙される群から選択されるmiRNAのパネルにおいて前記核酸配列を参照核酸配列にアライメントする工程と、

f) アライメントされた前記核酸配列に少なくとも部分的に基づき前記対象のmiRNA

50

NA プロファイルを決する工程とを含む方法。

【請求項 19】

予め選択した miRNA が、表 1 ～ 表 11 に列挙される群から選択される少なくとも 1、少なくとも 2、少なくとも 3、少なくとも 4、少なくとも 5、少なくとも 6、少なくとも 7、少なくとも 8、少なくとも 9、少なくとも 10、少なくとも 11、少なくとも 12、少なくとも 13、少なくとも 14、少なくとも 15、少なくとも 16、少なくとも 17、少なくとも 18、少なくとも 19、少なくとも 20、少なくとも 21、少なくとも 22、少なくとも 23、少なくとも 24、少なくとも 25、少なくとも 26、少なくとも 27、少なくとも 28、少なくとも 29、少なくとも 30、少なくとも 40、少なくとも 50、少なくとも 60、少なくとも 70、少なくとも 80、少なくとも 90、少なくとも 100、少なくとも 110、少なくとも 120、少なくとも 130、少なくとも 140、少なくとも 150、少なくとも 160、少なくとも 170、少なくとも 180、少なくとも 190、少なくとも 200、または少なくとも 250 の miRNA を含む、請求項 18 に記載の方法。

10

【請求項 20】

前記結腸細胞増殖性障害が、腺腫（腺腫性ポリープ）、固着性鋸歯状腺腫（SSA）、進行型腺腫、大腸異形成、大腸腺腫、大腸癌、結腸癌、直腸癌、大腸癌腫、大腸腺癌、カルチノイド腫瘍、胃腸カルチノイド腫瘍、胃腸間質腫瘍（GIST）、リンパ腫、および肉腫からなる群から選択される、請求項 18 に記載の方法。

20

【請求項 21】

RNA 分子または cDNA 分子を富化または枯渇させる工程をさらに含む、請求項 18 に記載の方法。

【請求項 22】

前記参照核酸配列が、ゲノム、トランスクリプトーム、またはカスタムトランスクリプトームの参照核酸配列を含む、請求項 18 に記載の方法。

【請求項 23】

増幅前に miRNA ライブラリを調製する工程をさらに含む、請求項 18 に記載の方法。

【請求項 24】

前記 RNA アダプタをライゲートする工程が、アダプタブロッキング、アダプタ環状化、および二量体除去を実行することを含む、請求項 18 に記載の方法。

30

【請求項 25】

前記 RNA アダプタをライゲートする工程が、3' RNA アダプタライゲーション、5' RNA アダプタライゲーション、固有分子識別子（UMI）割付けによる逆転写、および cDNA クリーンアップを実行することを含む、請求項 18 に記載の方法。

【請求項 26】

対象から得た生体試料のマイクロリボ核酸（miRNA）プロファイルを決する方法であって、1）前記生体試料からの RNA 分子の抽出後、直接 RNA を計数すること、2）前記生体試料からの RNA 分子の抽出後、A テーリングを行い、次いで鋳型スイッチングにより cDNA へと逆転写（RT）すること、3）前記生体試料からの RNA 分子の抽出後、A テーリングを行い、次いで逆転写ポリメラーゼ連鎖反応（RT-PCR）と量的 PCR（qPCR）またはデジタルドロップレット PCR（ddPCR）を行うこと、4）前記生体試料からの RNA 分子の抽出後、配列特異的ライゲーションを行い、次いで RT-PCR と qPCR または ddPCR を行うこと、ならびに 5）RNA 単離を実行しない条件下で前記生体試料から得た RNA 分子の抽出不要 miRNA プロファイリングを行うことのうち、1 または複数を行う工程と、前記対象から得た前記生体試料の miRNA プロファイルを決する工程とを含む方法。

40

【請求項 27】

前記 miRNA プロファイルを決する工程が、ヒトゲノムまたはヒトトランスクリプ

50

トームのデータベースの一部である参照核酸配列の使用を含む、請求項 26 に記載の方法。

【請求項 28】

前記 m i R N A プロファイルを決定する工程が、発現された m i R N A のカウントテーブルを生成することを含む、請求項 26 に記載の方法。

【請求項 29】

前記 m i R N A プロファイルを決定する工程が、発現された m i R N A に基づき正規化されたカウントテーブルを生成することで、差動的に豊富な m i R N A を特定することを含む、請求項 26 に記載の方法。

【請求項 30】

前記 m i R N A プロファイルが、結腸細胞増殖性障害と関連付けられ、かつ、結腸細胞増殖性障害がある、または結腸細胞増殖性障害を患っていないとの対象の分類を提供する、請求項 26 に記載の方法。

【請求項 31】

前記対象から得た生体試料が、体液、便、結腸流出液、尿、血漿、血清、全血、単離した血液細胞、血液から単離した細胞、およびそれらの組合せからなる群から選択される、請求項 26 に記載の方法。

【請求項 32】

前記 m i R N A プロファイルを健康な対象から得た参照 m i R N A プロファイルのデータベースと比較する工程と、前記 m i R N A プロファイルの m i R N A 発現が前記参照 m i R N A プロファイルと比して少なくとも 15 % 変化することの測定に少なくとも部分的に基づき、前記対象が結腸細胞増殖性障害を患うリスクは増加すると決定する工程とをさらに含む、請求項 26 に記載の方法。

【請求項 33】

前記結腸細胞増殖性障害が、腺腫（腺腫性ポリープ）、固着性鋸歯状腺腫（S S A）、進行型腺腫、大腸異形成、大腸腺腫、大腸癌、結腸癌、直腸癌、大腸癌腫、大腸腺癌、カルチノイド腫瘍、胃腸カルチノイド腫瘍、胃腸間質腫瘍（G I S T）、リンパ腫、および肉腫からなる群から選択される、請求項 26 に記載の方法。

【請求項 34】

前記結腸細胞増殖性障害が、ステージ 1 大腸癌、ステージ 2 大腸癌、ステージ 3 大腸癌、およびステージ 4 大腸癌からなる群から選択される、請求項 26 に記載の方法。

【請求項 35】

前記進行型腺腫が、管状腺腫、尿細管腺腫、絨毛腺腫、腺癌、または過形成ポリープを含む、請求項 26 に記載の方法。

【請求項 36】

対象における結腸細胞増殖性障害の有無を検出するための方法であって、

a) 前記生体試料からリボ核酸（R N A）分子を単離する工程と、
b) 前記 R N A 分子を相補的デオキシリボ核酸（c D N A）分子に逆転写する前または後に R N A アダプタを前記 R N A 分子にライゲートする工程と、

c) 前記 c D N A 分子を増幅する工程と、

d) 前記 c D N A 分子の核酸配列を決定する工程と、

e) 表 1 ～ 表 11 に列挙される群から選択される m i R N A の予め特定されたパネルにおいて前記核酸配列を参照核酸配列にアライメントする工程と、

f) アライメントされた前記核酸配列に少なくとも部分的に基づき m i R N A プロファイルを決定する工程と、

g) 健康な対象と結腸細胞増殖性障害を患う対象とを判別可能となるよう訓練された機械学習モデルを用いて前記 m i R N A プロファイルをコンピュータにより処理することで、結腸細胞増殖性障害の有無に関連する出力値を提供し、それにより前記対象における結腸細胞増殖性障害の有無を示す工程とを含む方法。

10

20

30

40

50

【請求項 37】

b が、試料特異的バーコードおよび／または分子特異的固有分子識別子（UMI）を組み入れることを含む、請求項 36 に記載の方法。

【請求項 38】

予め選択した miRNA が、表 1 ～ 表 11 に列挙される群から選択される少なくとも 1、少なくとも 2、少なくとも 3、少なくとも 4、少なくとも 5、少なくとも 6、少なくとも 7、少なくとも 8、少なくとも 9、少なくとも 10、少なくとも 11、少なくとも 12、少なくとも 13、少なくとも 14、少なくとも 15、少なくとも 16、少なくとも 17、少なくとも 18、少なくとも 19、少なくとも 20、少なくとも 21、少なくとも 22、少なくとも 23、少なくとも 24、少なくとも 25、少なくとも 26、少なくとも 27、少なくとも 28、少なくとも 29、少なくとも 30、少なくとも 40、少なくとも 50、少なくとも 60、少なくとも 70、少なくとも 80、少なくとも 90、少なくとも 100、少なくとも 110、少なくとも 120、少なくとも 130、少なくとも 140、少なくとも 150、少なくとも 160、少なくとも 170、少なくとも 180、少なくとも 190、少なくとも 200、または少なくとも 250 の miRNA を含む、請求項 36 に記載の方法。

10

【請求項 39】

前記結腸細胞増殖性障害が、腺腫（腺腫性ポリープ）、固着性鋸歯状腺腫（SSA）、進行型腺腫、大腸異形成、大腸腺腫、大腸癌、結腸癌、直腸癌、大腸癌腫、大腸腺癌、カルチノイド腫瘍、胃腸カルチノイド腫瘍、胃腸間質腫瘍（GIST）、リンパ腫、および肉腫からなる群から選択される、請求項 36 に記載の方法。

20

【請求項 40】

前記参照核酸配列が、ヒトゲノムまたはヒトトランスクリプトームのデータベースの一部である、請求項 36 に記載の方法。

【請求項 41】

前記対象の miRNA プロファイルを決定する工程が、発現された miRNA のカウントテーブルを生成することを含む、請求項 36 に記載の方法。

【請求項 42】

前記対象の miRNA プロファイルを決定する工程が、発現された miRNA のカウントテーブルを生成することで、差動的に豊富な miRNA を特定することを含む、請求項 36 に記載の方法。

30

【請求項 43】

前記対象から得た前記生体試料が、体液、便、結腸流出液、尿、血漿、血清、全血、単離した血液細胞、血液から単離した細胞、およびそれらの組合せからなる群から選択される、請求項 36 に記載の方法。

【請求項 44】

前記 miRNA プロファイルを健康な対象から得た参照 miRNA プロファイルのデータベースと比較する工程と、前記 miRNA プロファイルの miRNA 発現が参照 miRNA プロファイルと比して少なくとも 15 % 変化することの測定に少なくとも部分的に基づき、前記対象が結腸細胞増殖性障害を患うリスクは増加すると決定する工程とをさらに含む、請求項 36 に記載の方法。

40

【請求項 45】

前記結腸細胞増殖性障害が、腺腫（腺腫性ポリープ）、固着性鋸歯状腺腫（SSA）、進行型腺腫、大腸異形成、大腸腺腫、大腸癌、結腸癌、直腸癌、大腸癌腫、大腸腺癌、カルチノイド腫瘍、胃腸カルチノイド腫瘍、胃腸間質腫瘍（GIST）、リンパ腫、および肉腫からなる群から選択される、請求項 36 に記載の方法。

【請求項 46】

前記結腸細胞増殖性障害が、ステージ 1 大腸癌、ステージ 2 大腸癌、ステージ 3 大腸癌、およびステージ 4 大腸癌からなる群から選択される、請求項 36 に記載の方法。

【請求項 47】

50

検出された前記結腸細胞増殖性障害に基づき前記対象に処置を施す工程をさらに含む、請求項 36 に記載の方法。

【請求項 48】

対象から得た生体試料の m i R N A プロファイルを決定するための方法であって、

a) 前記生体試料からリボ核酸 (R N A) 分子を単離する工程と、

b) 前記 R N A 分子を相補的デオキシリボ核酸 (c D N A) 分子に逆転写する工程と

、

c) R N A アダプタを前記 R N A 分子または前記 c D N A 分子にライゲートする工程と、

d) 前記 c D N A 分子を増幅する工程と、

e) 前記 c D N A 分子の核酸配列を決定する工程と、

f) 表 1 ～表 11 に列挙される群から選択される m i R N A のパネルにおいて前記核酸配列を参照核酸配列にアライメントする工程と、

g) アライメントされた前記核酸配列に少なくとも部分的に基づき前記 m i R N A プロファイルを決定する工程とを含む方法。

【請求項 49】

予め選択した前記 m i R N A が、表 1 ～表 11 に列挙される群から選択される少なくとも 1、少なくとも 2、少なくとも 3、少なくとも 4、少なくとも 5、少なくとも 6、少なくとも 7、少なくとも 8、少なくとも 9、少なくとも 10、少なくとも 11、少なくとも 12、少なくとも 13、少なくとも 14、少なくとも 15、少なくとも 16、少なくとも 17、少なくとも 18、少なくとも 19、少なくとも 20、少なくとも 21、少なくとも 22、少なくとも 23、少なくとも 24、少なくとも 25、少なくとも 26、少なくとも 27、少なくとも 28、少なくとも 29、少なくとも 30、少なくとも 40、少なくとも 50、少なくとも 60、少なくとも 70、少なくとも 80、少なくとも 90、少なくとも 100、少なくとも 110、少なくとも 120、少なくとも 130、少なくとも 140、少なくとも 150、少なくとも 160、少なくとも 170、少なくとも 180、少なくとも 190、少なくとも 200、または少なくとも 250 の m i R N A を含む、請求項 48 に記載の方法。

【請求項 50】

対象における結腸細胞増殖性障害の有無を検出するための方法であって、

a) 前記生体試料からリボ核酸 (R N A) 分子を単離する工程と、

b) 前記 R N A 分子を相補的デオキシリボ核酸 (c D N A) 分子に逆転写する工程と

、

c) R N A アダプタを前記 R N A 分子または前記 c D N A 分子にライゲートする工程と、

d) 前記 c D N A 分子を増幅する工程と、

e) 前記 c D N A 分子の核酸配列を決定する工程と、

f) 表 1 ～表 11 に列挙される群から選択される m i R N A のパネルにおいて前記核酸配列を参照核酸配列にアライメントする工程と、

g) アライメントされた前記核酸配列に少なくとも部分的に基づき m i R N A プロファイルを決定する工程と、

h) 前記結腸細胞増殖性障害を患っていない対象と前記結腸細胞増殖性障害を患う対象とを判別するよう訓練された機械学習モデルを用いて前記 m i R N A プロファイルをコンピュータにより処理する工程と、

i) 前記機械学習モデルにより、前記結腸細胞増殖性障害を患う対象または前記結腸細胞増殖性障害を患う対象に関連する値を出力することで、前記対象における結腸細胞増殖性障害の有無を検出する工程と

を含む方法。

【請求項 51】

10

20

30

40

50

予め選択した前記m i R N Aが、表1～表11に列挙される群から選択される少なくとも1、少なくとも2、少なくとも3、少なくとも4、少なくとも5、少なくとも6、少なくとも7、少なくとも8、少なくとも9、少なくとも10、少なくとも11、少なくとも12、少なくとも13、少なくとも14、少なくとも15、少なくとも16、少なくとも17、少なくとも18、少なくとも19、少なくとも20、少なくとも21、少なくとも22、少なくとも23、少なくとも24、少なくとも25、少なくとも26、少なくとも27、少なくとも28、少なくとも29、少なくとも30、少なくとも40、少なくとも50、少なくとも60、少なくとも70、少なくとも80、少なくとも90、少なくとも100、少なくとも110、少なくとも120、少なくとも130、少なくとも140、少なくとも150、少なくとも160、少なくとも170、少なくとも180、少なくとも190、少なくとも200、または少なくとも250のm i R N Aを含む、請求項50に記載の方法。

10

【請求項52】

前記結腸細胞増殖性障害が、腺腫（腺腫性ポリープ）、固着性鋸歯状腺腫（SSA）、進行型腺腫、大腸異形成、大腸腺腫、大腸癌、結腸癌、直腸癌、大腸癌腫、大腸腺癌、カルチノイド腫瘍、胃腸カルチノイド腫瘍、胃腸間質腫瘍（GIST）、リンパ腫、および肉腫からなる群から選択される、請求項50に記載の方法。

【請求項53】

検出された前記結腸細胞増殖性障害に基づき前記対象に処置を施す工程をさらに含む、請求項50に記載の方法。

20

【請求項54】

以前に疾患の処置を受けた対象における最小残存病変をモニタリングするための方法であって、表1～表11に列挙される群から選択されるマイクロリボ核酸（m i R N A）のパネルを用いて前記対象から得た生体試料のm i R N Aプロファイルを決定することで、ベースラインのm i R N A状態を生成する工程と、前記ベースラインのm i R N A状態の生成後に前記対象から得た生体試料のm i R N Aプロファイルを1または複数の時点で決定することで、現在のm i R N A状態を生成する工程と、前記ベースラインのm i R N A状態と前記現在のm i R N A状態との差異を決定することで、前記対象における前記最小残存病変の変化を検出する工程とを含む方法。

【請求項55】

前記最小残存病変が、処置への応答、腫瘍細胞量、術後残存腫瘍、再発、二次スクリーン、一次スクリーン、および癌の進展からなる群から選択される、請求項54に記載の方法。

30

【請求項56】

前記m i R N Aプロファイルが、少なくとも約40%の感度で前記対象における大腸癌の存在または感受性を示す、請求項54に記載の方法。

【請求項57】

前記対象における前記最小残存病変に対し検出された前記変化に基づき、前記対象に処置を施す工程をさらに含む、請求項54に記載の方法。

【発明の詳細な説明】

40

【技術分野】

【0001】

関連出願への相互参照

本出願は、2020年4月29日出願の米国仮特許出願第63／017,552号、および2020年5月14日出願の米国仮特許出願第63／024,875号の利益を主張するものであり、これら出願はそれぞれ、その全体を参照することで本明細書中に援用される。

【背景技術】

【0002】

処置が最も有効なのは症状発症前の早期段階で悪性腫瘍または前駆病変が検出されると

50

きであるため、癌のスクリーニングと早期検出は、癌に対し最も有効な戦略と認められている。例えば、大腸癌では、大腸内視鏡検査が早期診断の改善に役割を果たす。大腸内視鏡は早期検出に有用ではあるが、患者が遵守する割合は低く、手術の侵襲性が原因で推奨規則が満たされていない状態でスクリーニングが行われる。このため、早期癌検出におけるより有望な手法は、非侵襲的な方法により提供される。

【発明の概要】

【0003】

本開示は、結腸細胞増殖性障害（例えば、大腸癌）の検出および疾患進行に関連する遺伝子に対するマイクロリボ核酸（マイクロRNAまたはmiRNA）プロファイリングに関する。本開示のいくつかの実施形態は、結腸細胞増殖性障害を患っていない対象または結腸細胞増殖性障害を発症するリスクが低い対象の対応試料と比較して、結腸細胞増殖性障害を患う対象または結腸細胞増殖性障害を発症するリスクが高い対象の試料において差動的に発現されるmiRNAを提供する。いくつかの実施形態では、結腸細胞増殖性障害を発症するリスクが高い対象と、結腸細胞増殖性障害を発症するリスクが低い対象はそれぞれ、大腸粘膜内に非侵襲性前駆病変（以降、大腸病変とする）が生じる。結腸細胞増殖性障害を患う対象または結腸細胞増殖性障害を発症するリスクが高い対象の試料中に様々な存在量で存在する（多くの場合、「差動的に発現される」と称される）miRNAは、結腸細胞増殖性障害の診断、処置、および／または予防のバイオマーカーとして使用することができる。

10

【0004】

本明細書中で特定されるmiRNAは、結腸細胞増殖性障害を患う対象を特定して結腸細胞増殖性障害を患っていない対象と判別するために、もしくは結腸細胞増殖性障害を発症するリスクが高い対象を特定して結腸細胞増殖性障害を発症するリスクが低い対象と判別するために、または結腸細胞増殖性障害前駆体（膵管内乳頭粘液性腫瘍（IPMN）など）を有する対象とIPMNがない対象を特定するために、あるいは悪性IPMNを有する対象と良性IPMNを有する対象を特定するために使用することができる。このため、これらmiRNAは、結腸細胞増殖性障害のモニタリング、処置、および管理に伴う決断の指針となる補助ツールとして使用することができる。

20

【0005】

本開示のいくつかの実施形態は、例えば対象が大腸病変を患っているとき、結腸細胞増殖性障害を患う対象または結腸細胞増殖性障害を発症するリスクが高い対象の試料において差動的に発現される本明細書に記載のmiRNAに対し訓練された、機械学習モデル分類器を提供する。一例では、組織学的重症度を精査するために大腸病変を患う対象に使用可能である、血液を用いた低侵襲性miRNAアッセイの方法が提供される。別の実施形態では、結腸細胞増殖性障害を示すmiRNAは、対象から得た無細胞試料、例えば、全血、血漿、または血清など対象から得た体液試料中に検出される。そのため、本発明は、結腸細胞増殖性障害、外科切除、免疫療法、放射線療法、または化学療法などの処置を正当化（warrant）する高リスクまたは低リスク大腸病変、およびモニタリング可能である低リスク大腸病変の有無を判別するために使用可能であるmiRNAを提供する。結腸細胞増殖性障害または病変の存在のモニタリングと確認は、例えば大腸内視鏡検査、超音波検査、MM、またはCTスキャンにより実行することができる。

30

40

【0006】

一態様では、本開示は、表1～表11に列挙される群から選択される1つもしくは複数、2つ以上、3つ以上、または4つ以上のmiRNAからなる所定のセットを含む、結腸細胞増殖性障害の特徴を示すマイクロリボ核酸（miRNA）シグネチャパネルを提供し、4つ以上のmiRNAが、結腸細胞増殖性障害またはその亜型を患う対象から得た生体試料と、結腸細胞増殖性障害またはその亜型を患っていない対象から得た生体試料との間で差動的に発現される。

【0007】

いくつかの実施形態では、miRNAシグネチャパネルは、進行型腺腫の特徴を示すと

50

ともに、a) hsa-miR-1273a、hsa-miR-17-5p、hsa-miR-20a-3p、hsa-miR-20b-5p、b) hsa-miR-3065-5p、hsa-miR-4785、hsa-miR-5096、hsa-miR-5189-5p、またはc) hsa-miR-545-3p、hsa-miR-570-3p、hsa-miR-624-3p、hsa-miR-1181、hsa-miR-6073を含むmiRNAからなる所定のセットを含んでおり、miRNAは、進行型腺腫またはその亜型を患う対象から得た生体試料と、進行型腺腫またはその亜型を患っていない対象から得た生体試料との間で差動的に発現される。

【0008】

いくつかの実施形態では、miRNAシグネチャパネルは、大腸癌の特徴を示すとともに、a) hsa-miR-1250-5p、hsa-miR-1255a、hsa-miR-223-3p、hsa-miR-338-3p、hsa-miR-338-5p、b) hsa-miR-424-5p、hsa-miR-424-3p、hsa-miR-450a-5p、hsa-miR-450b-5p、hsa-miR-4772-3p、c) hsa-miR-4772-5p、hsa-miR-625-5p、hsa-miR-7847-3p、hsa-miR-1181、hsa-miR-3651、hsa-miR-6073、d) hsa-miR-6125、hsa-miR-7704、hsa-miR-19b-3p、hsa-miR-19a-3p、hsa-miR-3157-5p、e) hsa-miR-142-3p、hsa-miR-30c-5p、hsa-miR-6741-5p、hsa-miR-590-3p、hsa-miR-4685-5p、f) hsa-miR-3648、hsa-miR-331-3p、hsa-miR-1303、hsa-miR-6790-3p、hsa-miR-6867-5p、hsa-miR-942-5p、g) hsa-miR-378a-3p、hsa-miR-1287-5p、hsa-miR-4785、hsa-miR-324-3p、hsa-miR-550b-2-5p、h) hsa-miR-200c-3p、hsa-miR-200b-3p、hsa-miR-3679-5p、hsa-miR-550a-3-5p、hsa-miR-3187-3p、i) hsa-miR-181b-5p、hsa-miR-3138、hsa-miR-146a-5p、hsa-miR-6721-5p、hsa-miR-23b-3p、hsa-miR-28-5p、j) hsa-miR-320d、hsa-miR-940、hsa-miR-320d-1、hsa-miR-10a-5p、hsa-miR-340-5p、k) hsa-miR-320b、hsa-miR-335-5p、hsa-miR-320c、hsa-miR-501-3p、hsa-miR-548n、またはl) hsa-miR-27a-3p、hsa-miR-3065-3p、hsa-miR-548aa@、hsa-miR-584-3p、hsa-miR-22-3pを含むmiRNAからなる所定のセットを含んでおり、miRNAは、大腸癌またはその亜型を患う対象から得た生体試料と、大腸癌またはその亜型を患っていない対象から得た生体試料との間で差動的に発現される。

【0009】

いくつかの実施形態では、miRNAからなる所定のセットは、表1～表11に列挙される群から選択される少なくとも1、少なくとも2、少なくとも3、少なくとも4、少なくとも5、少なくとも6、少なくとも7、少なくとも8、少なくとも9、少なくとも10、少なくとも11、少なくとも12、少なくとも13、少なくとも14、少なくとも15、少なくとも16、少なくとも17、少なくとも18、少なくとも19、少なくとも20、少なくとも21、少なくとも22、少なくとも23、少なくとも24、少なくとも25、少なくとも26、少なくとも27、少なくとも28、少なくとも29、少なくとも30、少なくとも40、少なくとも50、少なくとも60、少なくとも70、少なくとも80、少なくとも90、少なくとも100、少なくとも110、少なくとも120、少なくとも130、少なくとも140、少なくとも150、少なくとも160、少なくとも170、少なくとも180、少なくとも190、少なくとも200、または少なくとも250のmiRNAを含む。

【0010】

いくつかの実施形態では、生体試料は、体液、便、結腸流出液、尿、血漿、血清、全血、単離した血液細胞、血液から単離した細胞、およびそれらの組合せからなる群から選択される。

【0011】

いくつかの実施形態では、生体試料は、核酸、DNA、RNA、または無細胞核酸（cfDNAまたはcfRNA）を含む。

【0012】

いくつかの実施形態では、miRNAは、成熟miRNAおよびmiRNAヘアピンを含む。

10

【0013】

いくつかの実施形態では、シグネチャパネルは、表1～表11に列挙される群から選択される1つまたは複数、2つ以上、3つ以上、4つ以上、5つ以上、6つ以上、7つ以上、8つ以上、9つ以上、10以上、11以上、または12以上のmiRNAにおける差動的発現を含む。

【0014】

いくつかの実施形態では、結腸細胞増殖性障害は、腺腫（腺腫性ポリープ）、固着性鋸歯状腺腫（SSA）、進行型腺腫、大腸異形成、大腸腺腫、大腸癌、結腸癌、直腸癌、大腸癌腫、大腸腺癌、カルチノイド腫瘍、胃腸カルチノイド腫瘍、胃腸間質腫瘍（GIST）、リンパ腫、および肉腫からなる群から選択される。

20

【0015】

いくつかの実施形態では、結腸細胞増殖性障害は、ステージ1大腸癌、ステージ2大腸癌、ステージ3大腸癌、およびステージ4大腸癌からなる群から選択される。

【0016】

別の実施形態では、本開示は、健康な対象集団（例えば、結腸細胞増殖性障害がない対象）と、結腸細胞増殖性障害を患う対象とを判別可能な分類器であって、a）分類器は、表1～表11に列挙される群から選択される6つ以上または12以上の予め選択したmiRNAにおける差動的miRNA発現を表す測定値のセットを含み、この測定値は、健康な対象と結腸細胞増殖性障害を患う対象から得たmiRNA発現データから取得され、b）測定値は、差動的miRNA発現の特性に相当する一組の特徴を生成するために使用され、この一組の特徴は、機械学習モデル（例えば、統計モデル）を使用してコンピュータにより処理され、c）機械学習モデルは、健康な対象集団と、結腸細胞増殖性障害を患う対象とを判別可能な分類器として有用な特徴ベクトルを提供する、分類器を提供する。

30

【0017】

いくつかの実施形態では、予め選択したmiRNAは、表1～表11に列挙される群から選択される少なくとも1、少なくとも2、少なくとも3、少なくとも4、少なくとも5、少なくとも6、少なくとも7、少なくとも8、少なくとも9、少なくとも10、少なくとも11、少なくとも12、少なくとも13、少なくとも14、少なくとも15、少なくとも16、少なくとも17、少なくとも18、少なくとも19、少なくとも20、少なくとも21、少なくとも22、少なくとも23、少なくとも24、少なくとも25、少なくとも26、少なくとも27、少なくとも28、少なくとも29、少なくとも30、少なくとも40、少なくとも50、少なくとも60、少なくとも70、少なくとも80、少なくとも90、少なくとも100、少なくとも110、少なくとも120、少なくとも130、少なくとも140、少なくとも150、少なくとも160、少なくとも170、少なくとも180、少なくとも190、少なくとも200、または少なくとも250のmiRNAを含む。

40

【0018】

いくつかの実施形態では、結腸細胞増殖性障害は、腺腫（腺腫性ポリープ）、固着性鋸歯状腺腫（SSA）、進行型腺腫、大腸異形成、大腸腺腫、大腸癌、結腸癌、直腸癌、大腸癌腫、大腸腺癌、カルチノイド腫瘍、胃腸カルチノイド腫瘍、胃腸間質腫瘍（GIST

50

）、リンパ腫、および肉腫からなる群から選択される。

【0019】

いくつかの実施形態では、測定値のセットは、異なる数を有するフラグメントの観察の数または割合、生のmiRNA存在量、ハウスキーピング遺伝子へと正規化されるmiRNA存在量、合成配列へと正規化されるmiRNA存在量、対数正規化されたmiRNA存在量、フラグメント長、フラグメント中点、成熟miRNAまたはmiRNAヘアピンに沿ったリードマッピング位置およびリードパイルアップ、ならびにmiRNAクラスターの存在量からなる群から選択される差動的miRNA発現の特徴を記述する。

【0020】

いくつかの実施形態では、機械学習モデルは、訓練生体試料、結腸細胞増殖性障害を患う対象に相当すると特定される訓練生体試料の第1のサブセット、および結腸細胞増殖性障害を患っていない対象に相当すると特定される訓練生体試料の第2のサブセットから取得した訓練データを用いて訓練される。

【0021】

いくつかの実施形態では、分類器は、a) miRNAシグネチャパネルに基づき対象を分類するのに操作可能な分類器を含むコンピュータ可読媒体と、b) コンピュータ可読記憶媒体に記憶される命令を実行するための1または複数のプロセッサとを含む、結腸細胞増殖性障害を検出するためのシステムに設けられる。

【0022】

いくつかの実施形態では、上記システムは、深層学習分類器、ニューラルネットワーク分類器、線形判別分析(LDA)分類器、二次判別分析(QDA)分類器、サポート・ベクター・マシン(SVM)分類器、ランダムフォレスト(RF)分類器、線形カーネル・サポート・ベクター・マシン分類器、一次または二次多項式カーネル・サポート・ベクター・マシン分類器、リッジ回帰分類器、弾性ネットアルゴリズム分類器、逐次最小問題最適化アルゴリズム分類器、単純ベイズアルゴリズム分類器、および主成分分析分類器からなる群から選択される機械学習分類器として構成される分類回路を含む。

【0023】

別の実施形態では、本開示は、対象から得た生体試料のマイクロリボ核酸(miRNA)プロファイルを決断するための方法であって、a) 生体試料からRNA分子を単離する工程と、b) RNA分子を相補的デオキシリボ核酸(cDNA)分子に逆転写する前または後にRNAアダプタをRNA分子にライゲートする工程と、c) cDNA分子を増幅する工程と、d) cDNA分子の核酸配列を決断する工程と、e) 表1～表11に列挙される群から選択されるmiRNAのパネルにおいて核酸配列を参照核酸配列にアライメントする工程と、f) アライメントされた核酸配列に少なくとも部分的に基づき対象のmiRNAプロファイルを決断する工程とを含む方法を提供する。

【0024】

いくつかの実施形態では、予め選択したmiRNAは、表1～表11に列挙される群から選択される少なくとも1、少なくとも2、少なくとも3、少なくとも4、少なくとも5、少なくとも6、少なくとも7、少なくとも8、少なくとも9、少なくとも10、少なくとも11、少なくとも12、少なくとも13、少なくとも14、少なくとも15、少なくとも16、少なくとも17、少なくとも18、少なくとも19、少なくとも20、少なくとも21、少なくとも22、少なくとも23、少なくとも24、少なくとも25、少なくとも26、少なくとも27、少なくとも28、少なくとも29、少なくとも30、少なくとも40、少なくとも50、少なくとも60、少なくとも70、少なくとも80、少なくとも90、少なくとも100、少なくとも110、少なくとも120、少なくとも130、少なくとも140、少なくとも150、少なくとも160、少なくとも170、少なくとも180、少なくとも190、少なくとも200、または少なくとも250のmiRNAを含む。

【0025】

いくつかの実施形態では、結腸細胞増殖性障害は、腺腫(腺腫性ポリープ)、固着性鋸

10

20

30

40

50

歯状腺腫（SSA）、進行型腺腫、大腸異形成、大腸腺腫、大腸癌、結腸癌、直腸癌、大腸癌腫、大腸腺癌、カルチノイド腫瘍、胃腸カルチノイド腫瘍、胃腸間質腫瘍（GIST）、リンパ腫、および肉腫からなる群から選択される。

【0026】

いくつかの実施形態では、上記方法は、RNA分子またはcDNA分子を富化または枯渇させる工程をさらに含む。

【0027】

いくつかの実施形態では、参照核酸配列は、ゲノム、トランスクリプトーム、またはカスタムトランスクリプトームの参照核酸配列を含む。

【0028】

いくつかの実施形態では、上記方法は、増幅前にmiRNAライブラリを調製する工程をさらに含む。

【0029】

いくつかの実施形態では、アダプタのライゲーションは、上記cの前にRNAアダプタライゲーション、アダプタブロッキング、アダプタ環状化、および二量体除去を含む。

【0030】

いくつかの実施形態では、RNAアダプタをライゲートする工程は、アダプタブロッキング、アダプタ環状化、および／または二量体除去を実行することを含む。

【0031】

いくつかの実施形態では、RNAアダプタをライゲートする工程は、3' RNAアダプタライゲーション、5' RNAアダプタライゲーション、固有分子識別子（UMI）割付けによる逆転写、および／またはcDNAクリーンアップを実行することを含む。

【0032】

別の態様では、本開示は、対象から得た生体試料のマイクロリボ核酸（miRNA）プロファイルを決定するための方法であって、1）生体試料からのRNA分子の抽出後、直接RNAを計数すること、2）生体試料からのRNA分子の抽出後、Aテーリングを行い、次いで鋳型スイッチングによりcDNAへと逆転写（RT）すること、3）生体試料からのRNA分子の抽出後、Aテーリングを行い、次いで逆転写ポリメラーゼ連鎖反応（RT-PCR）と量的PCR（qPCR）またはデジタルドロップレットPCR（ddPCR）を行うこと、4）生体試料からのRNA分子の抽出後、配列特異的ライゲーションを行い、次いでRT-PCRとqPCRまたはddPCRを行うこと、ならびに5）RNA単離を実行しない条件下で生体試料から得たRNA分子の抽出不要miRNAプロファイリングを行うことのうち、1または複数を行う工程と、前記対象から得た生体試料のmiRNAプロファイルを決定する工程と含む方法を提供する。

【0033】

いくつかの実施形態では、miRNAプロファイルを決定する工程は、ヒトゲノムまたはヒトトランスクリプトームのデータベースの一部である参照核酸配列の使用を含む。

【0034】

いくつかの実施形態では、miRNAプロファイルを決定する工程は、発現されたmiRNAのカウントテーブルを生成することを含む。

【0035】

いくつかの実施形態では、miRNAプロファイルを決定する工程は、発現されたmiRNAに基づき正規化されたカウントテーブルを生成することで、差動的に豊富なmiRNAを特定することを含む。

【0036】

いくつかの実施形態では、miRNAプロファイルは、結腸細胞増殖性障害と関連付けられ、かつ、結腸細胞増殖性障害を患っている、または結腸細胞増殖性障害を患っていないとの対象の分類を提供する。

【0037】

いくつかの実施形態では、対象から得た生体試料は、体液、便、結腸流出液、尿、血漿

10

20

30

40

50

、血清、全血、単離した血液細胞、血液から単離した細胞、およびそれらの組合せからなる群から選択される。

【0038】

いくつかの実施形態では、上記方法は、miRNAプロファイルを健康な対象から得た参照miRNAプロファイルのデータベースと比較する工程と、miRNAプロファイルのmiRNA発現が参照miRNAプロファイルと比して少なくとも5%、少なくとも10%、少なくとも15%、少なくとも20%、少なくとも25%、少なくとも30%、少なくとも35%、少なくとも40%、少なくとも45%、または少なくとも50%変化することの測定に少なくとも部分的に基づき、対象が結腸細胞増殖性障害を患うリスクは増加すると決定する工程とをさらに含む

10

【0039】

いくつかの実施形態では、上記方法は、miRNAプロファイルを健康な対象から得た参照miRNAプロファイルのデータベースと比較する工程と、miRNAプロファイルのmiRNA発現が参照miRNAプロファイルと比して少なくとも15%変化することの測定に少なくとも部分的に基づき、対象が結腸細胞増殖性障害を患うリスクは増加すると決定する工程とをさらに含む。

【0040】

いくつかの実施形態では、結腸細胞増殖性障害は、腺腫（腺腫性ポリープ）、固着性鋸歯状腺腫（SSA）、進行型腺腫、大腸異形成、大腸腺腫、大腸癌、結腸癌、直腸癌、大腸癌腫、大腸腺癌、カルチノイド腫瘍、胃腸カルチノイド腫瘍、胃腸間質腫瘍（GIST）、リンパ腫、および肉腫からなる群から選択される。

20

【0041】

いくつかの実施形態では、結腸細胞増殖性障害は、ステージ1大腸癌、ステージ2大腸癌、ステージ3大腸癌、およびステージ4大腸癌からなる群から選択される。

【0042】

いくつかの実施形態では、進行型腺腫は、管状腺腫、尿細管腺腫、絨毛腺腫、腺癌、または過形成ポリープを含む。

【0043】

別の態様では、本開示は、対象における結腸細胞増殖性障害の有無を検出するための方法であって、a) 生体試料からリボ核酸（RNA）分子を単離する工程と、b) RNA分子を相補的デオキシリボ核酸（cDNA）分子に逆転写する前または後にRNAアダプタをRNA分子にライゲートする工程と、c) cDNA分子を増幅する工程と、d) cDNA分子の核酸配列を決定する工程と、e) 表1～表11に列挙される群から選択されるmiRNAの予め特定されたパネルにおいて核酸配列を参照核酸配列にアライメントする工程と、f) アライメントされた核酸配列に少なくとも部分的に基づきmiRNAプロファイルを決する工程と、g) 健康な対象と結腸細胞増殖性障害を患う対象とを判別可能となるよう訓練された機械学習モデルを用いてmiRNAプロファイルをコンピュータにより処理することで、結腸細胞増殖性障害の有無に関連する出力値を提供し、それにより対象における結腸細胞増殖性障害の有無を示す工程とを含む方法を提供する。

30

【0044】

いくつかの実施形態では、上記bは、試料特異的バーコードおよび／または分子特異的固有分子識別子（UMI）を組み入れることを含む。

40

【0045】

いくつかの実施形態では、予め選択したmiRNAは、表1～表11に列挙される群から選択される少なくとも1、少なくとも2、少なくとも3、少なくとも4、少なくとも5、少なくとも6、少なくとも7、少なくとも8、少なくとも9、少なくとも10、少なくとも11、少なくとも12、少なくとも13、少なくとも14、少なくとも15、少なくとも16、少なくとも17、少なくとも18、少なくとも19、少なくとも20、少なくとも21、少なくとも22、少なくとも23、少なくとも24、少なくとも25、少なくとも26、少なくとも27、少なくとも28、少なくとも29、少なくとも30、少なく

50

とも40、少なくとも50、少なくとも60、少なくとも70、少なくとも80、少なくとも90、少なくとも100、少なくとも110、少なくとも120、少なくとも130、少なくとも140、少なくとも150、少なくとも160、少なくとも170、少なくとも180、少なくとも190、少なくとも200、または少なくとも250のmiRNAを含む。

【0046】

いくつかの実施形態では、結腸細胞増殖性障害は、腺腫（腺腫性ポリープ）、固着性鋸歯状腺腫（SSA）、進行型腺腫、大腸異形成、大腸腺腫、大腸癌、結腸癌、直腸癌、大腸癌腫、大腸腺癌、カルチノイド腫瘍、胃腸カルチノイド腫瘍、胃腸間質腫瘍（GIST）、リンパ腫、および肉腫からなる群から選択される。

10

【0047】

いくつかの実施形態では、参照核酸配列は、ヒトゲノムまたはヒトトランスクリプトームのデータベースの一部である。

【0048】

いくつかの実施形態では、対象のmiRNAプロファイルを決定する工程は、発現されたmiRNAのカウントテーブルを生成することを含む。

【0049】

いくつかの実施形態では、対象のmiRNAプロファイルを決定する工程は、発現されたmiRNAのカウントテーブルを生成することで、差動的に豊富なmiRNAを特定することを含む。

20

【0050】

いくつかの実施形態では、対象から得た生体試料は、体液、便、結腸流出液、尿、血漿、血清、全血、単離した血液細胞、血液から単離した細胞、およびそれらの組合せからなる群から選択される。

【0051】

いくつかの実施形態では、上記方法は、miRNAプロファイルを健康な対象から得た参照miRNAプロファイルのデータベースと比較する工程と、miRNAプロファイルのmiRNA発現が参照miRNAプロファイルと比して少なくとも5%、少なくとも10%、少なくとも15%、少なくとも20%、少なくとも25%、少なくとも30%、少なくとも35%、少なくとも40%、少なくとも45%、または少なくとも50%変化することの測定に少なくとも部分的に基づき、対象が結腸細胞増殖性障害を患うリスクは増加すると決定する工程とをさらに含む

30

【0052】

いくつかの実施形態では、上記方法は、miRNAプロファイルを健康な対象から得た参照miRNAプロファイルのデータベースと比較する工程と、miRNAプロファイルのmiRNA発現が参照miRNAプロファイルと比して少なくとも15%変化することの測定に少なくとも部分的に基づき、対象が結腸細胞増殖性障害を患うリスクは増加すると決定する工程とをさらに含む。

【0053】

いくつかの実施形態では、結腸細胞増殖性障害は、腺腫（腺腫性ポリープ）、固着性鋸歯状腺腫（SSA）、進行型腺腫、大腸異形成、大腸腺腫、大腸癌、結腸癌、直腸癌、大腸癌腫、大腸腺癌、カルチノイド腫瘍、胃腸カルチノイド腫瘍、胃腸間質腫瘍（GIST）、リンパ腫、および肉腫からなる群から選択される。

40

【0054】

いくつかの実施形態では、結腸細胞増殖性障害は、ステージ1大腸癌、ステージ2大腸癌、ステージ3大腸癌、およびステージ4大腸癌からなる群から選択される。

【0055】

別の態様では、本開示は、対象から得た生体試料のmiRNAプロファイルを決定するための方法であって、a) 生体試料からリボ核酸（RNA）分子を単離する工程と、b) RNA分子を相補的デオキシリボ核酸（cDNA）分子に逆転写する工程と、c) RNA

50

アダプタをRNA分子またはcDNA分子にライゲートする工程と、d) cDNA分子を増幅する工程と、e) cDNA分子の核酸配列を決定する工程と、f) 表1～表11に列挙される群から選択されるmiRNAのパネルにおいて核酸配列を参照核酸配列にアライメントする工程と、g) アライメントされた核酸配列に少なくとも部分的に基づきmiRNAプロファイルを決

【0056】

いくつかの実施形態では、予め選択したmiRNAは、表1～表11に列挙される群から選択される少なくとも1、少なくとも2、少なくとも3、少なくとも4、少なくとも5、少なくとも6、少なくとも7、少なくとも8、少なくとも9、少なくとも10、少なくとも11、少なくとも12、少なくとも13、少なくとも14、少なくとも15、少なくとも16、少なくとも17、少なくとも18、少なくとも19、少なくとも20、少なくとも21、少なくとも22、少なくとも23、少なくとも24、少なくとも25、少なくとも26、少なくとも27、少なくとも28、少なくとも29、少なくとも30、少なくとも40、少なくとも50、少なくとも60、少なくとも70、少なくとも80、少なくとも90、少なくとも100、少なくとも110、少なくとも120、少なくとも130、少なくとも140、少なくとも150、少なくとも160、少なくとも170、少なくとも180、少なくとも190、少なくとも200、または少なくとも250のmiRNAを含む。

【0057】

いくつかの実施形態では、上記方法は、検出された結腸細胞増殖性障害に基づき対象に処置を施す工程をさらに含む。いくつかの実施形態では、処置は、化学療法、放射線療法、免疫療法、または手術を含む。

【0058】

別の態様では、本開示は、対象における結腸細胞増殖性障害の有無を検出するための方法であって、a) 生体試料からリボ核酸(RNA)分子を単離する工程と、b) RNA分子を相補的デオキシリボ核酸(cDNA)分子に逆転写する工程と、c) RNAアダプタをRNA分子またはcDNA分子にライゲートする工程と、d) cDNA分子を増幅する工程と、e) cDNA分子の核酸配列を決定する工程と、f) 表1～表11に列挙される群から選択されるmiRNAのパネルにおいて核酸配列を参照核酸配列にアライメントする工程と、g) アライメントされた核酸配列に少なくとも部分的に基づきmiRNAプロファイルを決

【0059】

いくつかの実施形態では、予め選択したmiRNAは、表1～表11に列挙される群から選択される少なくとも1、少なくとも2、少なくとも3、少なくとも4、少なくとも5、少なくとも6、少なくとも7、少なくとも8、少なくとも9、少なくとも10、少なくとも11、少なくとも12、少なくとも13、少なくとも14、少なくとも15、少なくとも16、少なくとも17、少なくとも18、少なくとも19、少なくとも20、少なくとも21、少なくとも22、少なくとも23、少なくとも24、少なくとも25、少なくとも26、少なくとも27、少なくとも28、少なくとも29、少なくとも30、少なくとも40、少なくとも50、少なくとも60、少なくとも70、少なくとも80、少なくとも90、少なくとも100、少なくとも110、少なくとも120、少なくとも130、少なくとも140、少なくとも150、少なくとも160、少なくとも170、少なくとも180、少なくとも190、少なくとも200、または少なくとも250のmiRNAを含む。

【0060】

いくつかの実施形態では、結腸細胞増殖性障害は、腺腫（腺腫性ポリープ）、固着性鋸歯状腺腫（SSA）、進行型腺腫、大腸異形成、大腸腺腫、大腸癌、結腸癌、直腸癌、大腸癌腫、大腸腺癌、カルチノイド腫瘍、胃腸カルチノイド腫瘍、胃腸間質腫瘍（GIST）、リンパ腫、および肉腫からなる群から選択される。

【0061】

いくつかの実施形態では、上記方法は、検出された結腸細胞増殖性障害に基づき対象に処置を施す工程をさらに含む。いくつかの実施形態では、処置は、化学療法、放射線療法、免疫療法、または手術を含む。

【0062】

別の態様では、本開示は、以前に疾患の処置を受けた対象における最小残存病変をモニタリングするための方法であって、表1～表11に列挙される群から選択されるマイクロリボ核酸（miRNA）のパネルを用いて対象から得た生体試料のmiRNAプロファイルを決定することで、ベースラインのmiRNA状態を生成する工程と、ベースラインのmiRNA状態の生成後に対象から得た生体試料のmiRNAプロファイルを1または複数の時点で決定することで、現在のmiRNA状態を生成する工程と、ベースラインのmiRNA状態と現在のmiRNA状態との差異を決定することで、対象における最小残存病変の変化を検出する工程とを含む方法を提供する。

【0063】

いくつかの実施形態では、最小残存病変は、処置への応答、腫瘍細胞量、術後残存腫瘍、再発、二次スクリーン（secondary screen）、一次スクリーン（primary screen）、および癌の進展からなる群から選択される。いくつかの実施形態では、上記方法は、対象における最小残存病変に対し検出された変化に基づき、対象に処置を施す工程をさらに含む。いくつかの実施形態では、処置は、化学療法、放射線療法、免疫療法、または手術を含む。

【0064】

別の態様では、処置に対する対象の応答を決定するための方法であって、表1～表11に列挙される群から選択されるマイクロリボ核酸（miRNA）のパネルを用いて対象から得た生体試料のmiRNAプロファイルを決定することで、ベースラインのmiRNA状態を生成する工程と、ベースラインのmiRNA状態の生成後に対象から得た生体試料のmiRNAプロファイルを1または複数の時点で決定することで、現在のmiRNA状態を生成する工程と、ベースラインのmiRNA状態と現在のmiRNA状態との差異を決定することで、処置に対する対象の応答を決定する工程とを含む方法が、提供される。いくつかの実施形態では、この方法は、対象において決定された処置に対する応答に基づき、対象に処置を施す工程をさらに含む。いくつかの実施形態では、処置は、化学療法、放射線療法、免疫療法、または手術を含む。

【0065】

別の態様では、対象の腫瘍細胞量をモニタリングするための方法であって、表1～表11に列挙される群から選択されるマイクロリボ核酸（miRNA）のパネルを用いて対象から得た生体試料のmiRNAプロファイルを決定することで、ベースラインのmiRNA状態を生成する工程と、ベースラインのmiRNA状態の生成後に対象から得た生体試料のmiRNAプロファイルを1または複数の時点で決定することで、現在のmiRNA状態を生成する工程と、ベースラインのmiRNA状態と現在のmiRNA状態との差異を決定することで、対象の腫瘍細胞量をモニタリングする工程とを含む方法が、提供される。いくつかの実施形態では、この方法は、対象の腫瘍細胞量に基づき、対象に処置を施す工程をさらに含む。いくつかの実施形態では、処置は、化学療法、放射線療法、免疫療法、または手術を含む。

【0066】

別の態様では、対象の術後残存腫瘍を検出するための方法であって、表1～表11に列挙される群から選択されるマイクロリボ核酸（miRNA）のパネルを用いて対象から得た生体試料のmiRNAプロファイルを決定することで、ベースラインのmiRNA状態

10

20

30

40

50

を生成する工程と、ベースラインのmiRNA状態の生成後に対象から得た生体試料のmiRNAプロファイルを1または複数の時点で決定することで、現在のmiRNA状態を生成する工程と、ベースラインのmiRNA状態と現在のmiRNA状態との差異を決定することで、対象の術後残存腫瘍を検出する工程とを含む方法が、提供される。いくつかの実施形態では、この方法は、対象の術後残存腫瘍に基づき、対象に処置を施す工程をさらに含む。いくつかの実施形態では、処置は、化学療法、放射線療法、免疫療法、または手術を含む。

【0067】

別の態様では、対象における再発を検出するための方法であって、表1～表11に列挙される群から選択されるマイクロリボ核酸(miRNA)のパネルを用いて対象から得た生体試料のmiRNAプロファイルを決

10

【0068】

別の態様では、対象のmiRNAプロファイルに少なくとも部分的に基づき、二次スクリー

20

【0069】

別の態様では、対象のmiRNAプロファイルに少なくとも部分的に基づき、一次スクリー

【0070】

別の態様では、対象の癌の進展をモニタリングするための方法であって、表1～表11に列挙される群から選択されるマイクロリボ核酸(miRNA)のパネルを用いて対象から得た生体試料のmiRNAプロファイルを決

30

【0071】

いくつかの実施形態では、miRNAプロファイルは、少なくとも約40%の感度で対象における大腸癌の存在または感受性を示す。いくつかの実施形態では、miRNAプロファイルは、少なくとも約50%の感度で対象における大腸癌の存在または感受性を示す。いくつかの実施形態では、miRNAプロファイルは、少なくとも約60%の感度で対象における大腸癌の存在または感受性を示す。いくつかの実施形態では、miRNAプロファイルは、少なくとも約70%の感度で対象における大腸癌の存在または感受性を示す。いくつかの実施形態では、miRNAプロファイルは、少なくとも約80%の感度で対象における大腸癌の存在または感受性を示す。いくつかの実施形態では、miRNAプロファイルは、少なくとも約90%の感度で対象における大腸癌の存在または感受性を示す。いくつかの実施形態では、miRNAプロファイルは、少なくとも約95%の感度で対象における大腸癌の存在または感受性を示す。

40

【0072】

いくつかの実施形態では、miRNAプロファイルは、少なくとも約30%の陽性適中率(PPV)で対象における大腸癌の存在または感受性を示す。いくつかの実施形態では

50

、m i R N A プロファイルは、少なくとも約 4 0 % の陽性適中率 (P P V) で大腸癌の存在または感受性を示す。いくつかの実施形態では、m i R N A プロファイルは、少なくとも約 5 0 % の陽性適中率 (P P V) で大腸癌の存在または感受性を示す。いくつかの実施形態では、m i R N A プロファイルは、少なくとも約 6 0 % の陽性適中率 (P P V) で大腸癌の存在または感受性を示す。いくつかの実施形態では、m i R N A プロファイルは、少なくとも約 7 0 % の陽性適中率 (P P V) で大腸癌の存在または感受性を示す。いくつかの実施形態では、m i R N A プロファイルは、少なくとも約 8 0 % の陽性適中率 (P P V) で大腸癌の存在または感受性を示す。いくつかの実施形態では、m i R N A プロファイルは、少なくとも約 9 0 % の陽性適中率 (P P V) で大腸癌の存在または感受性を示す。いくつかの実施形態では、m i R N A プロファイルは、少なくとも約 9 5 % の陽性適中率 (P P V) で大腸癌の存在または感受性を示す。いくつかの実施形態では、m i R N A プロファイルは、少なくとも約 9 9 % の陽性適中率 (P P V) で大腸癌の存在または感受性を示す。

10

【 0 0 7 3 】

いくつかの実施形態では、m i R N A プロファイルは、少なくとも約 4 0 % の陰性適中率 (N P V) で大腸癌の存在または感受性を示す。いくつかの実施形態では、m i R N A プロファイルは、少なくとも約 5 0 % の陰性適中率 (N P V) で大腸癌の存在または感受性を示す。いくつかの実施形態では、m i R N A プロファイルは、少なくとも約 6 0 % の陰性適中率 (N P V) で大腸癌の存在または感受性を示す。いくつかの実施形態では、m i R N A プロファイルは、少なくとも約 7 0 % の陰性適中率 (N P V) で大腸癌の存在または感受性を示す。いくつかの実施形態では、m i R N A プロファイルは、少なくとも約 8 0 % の陰性適中率 (N P V) で大腸癌の存在または感受性を示す。いくつかの実施形態では、m i R N A プロファイルは、少なくとも約 9 0 % の陰性適中率 (N P V) で大腸癌の存在または感受性を示す。いくつかの実施形態では、m i R N A プロファイルは、少なくとも約 9 5 % の陰性適中率 (N P V) で大腸癌の存在または感受性を示す。いくつかの実施形態では、m i R N A プロファイルは、少なくとも約 9 9 % の陰性適中率 (N P V) で大腸癌の存在または感受性を示す。

20

【 0 0 7 4 】

いくつかの実施形態では、訓練されたアルゴリズムは、少なくとも約 0 . 5 0 の曲線下面積 (A U C) で対象における大腸癌の存在または感受性を決定する。いくつかの実施形態では、訓練されたアルゴリズムは、少なくとも約 0 . 6 0 の曲線下面積 (A U C) で対象における大腸癌の存在または感受性を決定する。いくつかの実施形態では、訓練されたアルゴリズムは、少なくとも約 0 . 7 0 の曲線下面積 (A U C) で対象における大腸癌の存在または感受性を決定する。いくつかの実施形態では、訓練されたアルゴリズムは、少なくとも約 0 . 8 0 の曲線下面積 (A U C) で対象における大腸癌の存在または感受性を決定する。いくつかの実施形態では、訓練されたアルゴリズムは、少なくとも約 0 . 9 0 の曲線下面積 (A U C) で対象における大腸癌の存在または感受性を決定する。いくつかの実施形態では、訓練されたアルゴリズムは、少なくとも約 0 . 9 5 の曲線下面積 (A U C) で対象における大腸癌の存在または感受性を決定する。いくつかの実施形態では、訓練されたアルゴリズムは、少なくとも約 0 . 9 9 の曲線下面積 (A U C) で対象における大腸癌の存在または感受性を決定する。

30

40

【 0 0 7 5 】

いくつかの実施形態では、上記方法は、ユーザの電子デバイスの報告またはグラフィカル・ユーザ・インターフェースを提示する工程をさらに含む。いくつかの実施形態では、ユーザは、対象、個体、または患者である。

【 0 0 7 6 】

いくつかの実施形態では、上記方法は、対象、個体、または患者における大腸癌の存在または感受性の決定の可能性を決定する工程をさらに含む。

【 0 0 7 7 】

いくつかの実施形態では、訓練されたアルゴリズム (例えば、機械学習モデルまたは分

50

類器)は、教師あり機械学習アルゴリズムを含む。いくつかの実施形態では、教師あり機械学習アルゴリズムは、深層学習アルゴリズム、サポート・ベクター・マシン (SVM)、ニューラルネットワーク、またはランダムフォレストを含む。

【0078】

いくつかの実施形態では、上記方法は、対象に、大腸癌患者を処置するための治療介入(例えば、化学療法、放射線療法、免疫療法、または手術)など、少なくとも部分的にmiRNAのプロファイルまたは解析に基づく治療介入を提供する工程をさらに含む。

【0079】

いくつかの実施形態では、上記方法は、大腸癌の存在または感受性をモニタリングする工程をさらに含み、モニタリングする工程は、上記対象における大腸癌の存在または感受性を複数の時点で精査することを含み、精査は、複数の時点それぞれで求められた大腸癌の存在または感受性に基づいて行われる。

10

【0080】

いくつかの実施形態では、複数の時点の間の、対象の大腸癌の存在または感受性の評価における差は、(i)対象の大腸癌の存在または感受性の診断、(ii)対象の大腸癌の存在または感受性の予後、および(iii)対象の大腸癌の存在または感受性を処置するための治療経過の有効性もしくは非有効性からなる群から選択される、1もしくは複数の臨床的指標を示す。

【0081】

いくつかの実施形態では、上記方法は、訓練されたアルゴリズムを用いて複数の別個のサブタイプまたは段階の大腸癌の中から対象の大腸癌のサブタイプを判定することにより、対象の大腸癌を階層化する工程をさらに含む。

20

【0082】

本開示の別の態様は、結腸細胞増殖性障害を患う対象の集団を、結腸細胞増殖性障害を患っていない対象と判別するための分類器を提供し、この分類器は、表1～表11に列挙される群から選択される6以上のmiRNAにおける差動的なmiRNA存在量を表す複数組の測定値を含み、測定値は、結腸細胞増殖性障害を患っていない対象および結腸細胞増殖性障害を患う対象から得たmiRNA発現データから取得され、測定値は、差動的miRNA存在量の特性に相当する一連の特徴を生成するのに使用され、この特徴は、機械学習モデルまたは統計モデルに組み入れられ、機械学習モデルまたは統計モデルは、結腸細胞増殖性障害を患っていない対象の集団を、結腸細胞増殖性障害を患う対象と判別することが可能な分類器として有用である特徴ベクトルを提供する。

30

【0083】

本開示の別の態様は、1または複数のコンピュータプロセッサによる実行後、上記または本明細書の他の場所のいずれかに記載の方法を実施する機械実行可能コードを備える、非一時的コンピュータ可読媒体を提供する。

【0084】

本開示の別の態様は、1または複数のコンピュータプロセッサおよびそれに繋がれたコンピュータメモリを備えたシステムを提供する。コンピュータメモリは、1または複数のコンピュータプロセッサによる実行後、上記または本明細書の他の場所のいずれかに記載の方法を実施する機械可読コードを備える。

40

【0085】

本開示の別の態様は、a)機械学習モデルを用いたmiRNAシグネチャパネルに基づき、結腸細胞増殖性障害を患う対象の集団と結腸細胞増殖性障害を患っていない対象とを判別するための分類器と、b)コンピュータ可読媒体に記憶された命令を実行するための1または複数のプロセッサとを備えるシステムを提供する。

【0086】

参照による引用

本明細書中で言及される刊行物、特許、および特許出願はすべて、あたかも個々の刊行物、特許、または特許出願がそれぞれ参照により引用されると具体的かつ個別に示される

50

のと同じ程度にまで、参照により本明細書に引用される。参照により引用される刊行物と特許または特許出願が本明細書に含まれる開示と矛盾する限り、本明細書は、このような矛盾が生じるあらゆる題材にした材料に取って代わる、および／または優先するように意図される。

【図面の簡単な説明】

【0087】

ここに、本開示の例を、ほんの一例として添付の図面を参照しつつ記述する。本発明の新規な特徴は、添付の特許請求の範囲において具体的に明記される。本発明の原理が利用される例示的实施形態を明記する以下の詳細な説明と、添付の図面（本明細書では「図面（Figure）」や「図（FIG.）」ともいう）を参照することで、本発明の特徴と利点がより良く理解されるであろう。

10

【0088】

【図1】本明細書に提供される方法を実施するために、機械学習モデルと分類器によりプログラムまたはその他の方法で構成されるコンピュータシステムの概略図である。

【図2】特徴選択中に選択されるmiRNAを示すヒストグラムを提供する図である。

【図3】頻繁に選択された上位10個のmiRNAのロジスティック回帰計数を示すグラフを提供する図である。

【発明を実施するための形態】

【0089】

本発明の様々な実施形態が本明細書中で示され、記述されているが、かかる実施形態はほんの一例として提供されることは当業者に明白であろう。多数の変形、変更、および置換えが、本発明から逸脱することなく当業者によりなされてよい。本明細書に記載の本発明の実施形態に対する様々な代替物が利用されてよいことを理解されたい。

20

【0090】

処置が最も有効なのは症状発症前の早期段階で悪性腫瘍または前駆病変が検出されるときであるため、癌のスクリーニングと早期検出は、癌に対し最も有効な戦略と認められている。例えば、大腸癌では、大腸内視鏡検査が早期診断の改善に役割を果たす。大腸内視鏡は早期検出に有用ではあるが、患者が遵守する割合は低く、手術の侵襲性が原因で推奨規則が満たされていない状態でスクリーニングが行われる。このため、早期癌検出におけるより有望な手法は、非侵襲的な方法により提供される。

30

【0091】

本開示は、結腸細胞増殖性障害（例えば、大腸癌）の検出および疾患進行に関連する遺伝子に対するマイクロリボ核酸（マイクロRNAまたはmiRNA）プロファイリングに関する。本開示のいくつかの実施形態は、結腸細胞増殖性障害を患っていない対象または結腸細胞増殖性障害を発症するリスクが低い対象の対応試料と比較して、結腸細胞増殖性障害を患う対象または結腸細胞増殖性障害を発症するリスクが高い対象の試料において差動的に発現されるmiRNAを提供する。いくつかの実施形態では、結腸細胞増殖性障害を発症するリスクが高い対象と、結腸細胞増殖性障害を発症するリスクが低い対象はそれぞれ、大腸粘膜内に生じる非侵襲性前駆病変（以降、大腸病変とする）を患っている。結腸細胞増殖性障害を患う対象または結腸細胞増殖性障害を発症するリスクが高い対象の試料中に様々な存在量で存在する（多くの場合、「差動的に発現される」と称される）miRNAは、結腸細胞増殖性障害の診断、処置、および／または予防のバイオマーカーとして使用することができる。

40

【0092】

本明細書中で特定されるmiRNAは、結腸細胞増殖性障害を患っている対象を特定して結腸細胞増殖性障害を患っていない対象と判別するために、もしくは結腸細胞増殖性障害を発症するリスクが高い対象を特定して結腸細胞増殖性障害を発症するリスクが低い対象と判別するために、または結腸細胞増殖性障害前駆体（膵管内乳頭粘液性腫瘍（IPMN）など）を有する対象とIPMNがない対象を特定するために、あるいは悪性IPMNを有する対象と良性IPMNを有する対象を特定するために使用することができる。この

50

ため、これら *miRNA* は、結腸細胞増殖性障害のモニタリング、処置、および管理に伴う決断の指針となる補助ツールとして使用することができる。

【0093】

本開示のいくつかの実施形態は、例えば対象が大腸病変を患っているとき、結腸細胞増殖性障害を患う対象または結腸細胞増殖性障害を発症するリスクが高い対象の試料において差動的に発現される本明細書に記載の *miRNA* に対し訓練された、機械学習モデル分類器を提供する。一例では、組織学的重症度を精査するために大腸病変を患う対象に使用可能である、血液を用いた低侵襲性 *miRNA* アッセイの方法が提供される。別の実施形態では、結腸細胞増殖性障害を示す *miRNA* は、対象から得た無細胞試料、例えば、全血、血漿、または血清など対象から得た体液試料中に検出される。そのため、本発明は、結腸細胞増殖性障害、外科切除、免疫療法、放射線療法、または化学療法などの処置を正当化 (warrant) する高リスクまたは低リスク大腸病変、およびモニタリング可能である低リスク大腸病変の有無を判別するために使用可能である *miRNA* を提供する。結腸細胞増殖性障害または病変の存在のモニタリングと確認は、例えば大腸内視鏡検査、超音波検査、MM、またはCTスキャンにより実行することができる。

10

【0094】

本開示は、概して癌検出と疾患モニタリングに関する。より具体的には、この分野は、早期大腸癌に対する癌関連マイクロRNA (*miRNA*) 検出と疾患モニタリングに関する。特異的には、大腸癌 (CRC) および／または大腸腺腫 (CA)、例えば進行型大腸腺腫 (AA) などの結腸細胞増殖性障害を患うか、それを発症するリスクがあるヒト対象を特定するために、循環 *miRNA* シグネチャパネルとその使用が提供される。

20

【0095】

本開示は、概して癌検出と疾患モニタリングに関する。本開示は、例えば対象が大腸病変を患うときに結腸細胞増殖性障害の存在を示すか、結腸細胞増殖性障害を発症する高いリスクを示す、対象の *miRNA* を特定する。早期検出により、増殖し拡散する前に癌を排除することができるため、癌のスクリーニングとモニタリングにより生存アウトカムが改善する。例えば、大腸癌では、大腸内視鏡検査が早期診断の改善に役割を果たす。しかし生憎、患者が遵守する割合は低く、手術の侵襲性が原因で推奨規則が満たされない状態でスクリーニングが行われる。

【0096】

本明細書には、疾患を患う対象において上方調節もしくは過剰発現される *miRNA* の発現プロファイルまたは存在量に基づき、結腸細胞増殖性障害を患うリスクのある対象をスクリーニングまたは特定するための方法が記載される。本明細書にはさらに、対象、例えばヒト対象に対する結腸細胞増殖性障害の診断に有用なデータを取得するための方法が記載される。結腸細胞増殖性障害の非限定的な例として、大腸癌、大腸腺腫、または進行型大腸腺腫が挙げられる。

30

【0097】

本明細書で使用するとき、「結腸細胞増殖性障害を発症する高リスク (high risk of developing a colon cell proliferative disorder)」という用語は、結腸細胞増殖性障害を患っていない、または近い将来結腸細胞増殖性障害を発症するリスクが低い対象に比べて、近い将来結腸細胞増殖性障害を発症するリスクが高い対象を表す。本明細書で使用するとき、「近い将来 (near future)」という用語は、約1か月～約2年、約6か月～約18か月、または約1年の期間を表す。

40

【0098】

結腸細胞増殖性障害は、あらゆる腫瘍段階 (例えば、TX、T0、Tis、T1、T2、T3、T4)、あらゆる局所リンパ節または遠隔転移段階 (例えば、NX、N0、N1、M0、M1)、あらゆる段階 (例えば、ステージ0 (Tis、N0、M0)、ステージIA (T1、N0、M0)、ステージIIA (T3、N0、M0)、ステージIIB (T1～3、N1、M0)、ステージIII (T4、あらゆるN、M0)、またはステージI

50

V（あらゆるT、あらゆるN、M1））、切除可能、局所進行型（切除不可能）、または転移性の障害であってよい。

【0099】

現行のスクリーニングツールは、偽陽性および偽陰性の結果が生じること、および理想的な特異性および感度を満たしていないことを理由に、課題に直面する場合がある。理想的な癌スクリーニングツールは、不必要な調査を最小限に抑える（偽陽性が低い）が癌の大部分を検出する（偽陰性が低い）、高い陽性予測値（PPV）を有する場合がある。別の重要な妥協案（compromise）は、一般的にサイズに基づき腫瘍を検出する下限を表す、検査感度とは異なる「検出感度」である。検出可能なレベルで循環腫瘍マーカを放出するのに十分なサイズへの腫瘍成長が可能になると、癌進行の早期検出と予防の目的が頓挫してしまう。したがって本開示は、大腸癌の早期診断のために、血液を用いた高感度で有効なスクリーニングの必要性に対処する。

10

【0100】

「液体生検」とも称される循環腫瘍DNA検出により、非侵襲的な形で腫瘍の検出と有益な調査が可能になる場合がある。これら液体生検における腫瘍特異的変異の特定は、結腸癌、乳癌、および前立腺癌を診断するために使用されてよい。しかし、循環中に存在する正常な（例えば、非腫瘍由来）DNAのバックグラウンドが高いことが原因で、これらの技法は感度が制限される場合がある。このため、リスク集団の再発スクリーニングと一次スクリーニングにおいて早期または低腫瘍量の大腸癌腫瘍マーカを検出するために、より高感度で特異的なスクリーニングツールが依然として必要とされる。

20

【0101】

本開示は、結腸細胞増殖性障害、例えば大腸癌とその進行に関連する循環miRNAをプロファイリングすることを目的とする方法とシステムを提供する。結腸細胞増殖性障害の存在、または結腸細胞増殖性障害を発症する高リスクを示すそれらのmiRNAは、例えば対象に大腸病変しかないときに、結腸細胞増殖性障害を可能な限り早期に診断、処置、またはその進行を予防するために使用されてよい。本明細書にはさらに、具体的に対象に大腸病変があるときに、結腸細胞増殖性障害を診断、または対象における結腸細胞増殖性障害を発症するリスクを精査するためのキットと方法が提供される。

【0102】

miRNAは、概して植物や動物に見出される約18~22個のヌクレオチドからなる小さなノンコーディングRNAを表す。miRNAは、3'非翻訳領域（3'-UTR）中の特定部位に結合することによりmRNA標的を転写後調節し、それによりこれらmRNA標的の分解を促進するか、その翻訳を阻害する場合がある。miRNAは、増殖、分化、およびアポトーシスなどの複数の生理学的細胞機能に寄与する場合がある。miRNAは癌遺伝子と腫瘍抑制遺伝子の発現を調節するため、miRNAの調節不全は癌に重要な役割を果たす場合がある。無細胞miRNA（cfmiRNA）、循環腫瘍細胞（CTC）、循環腫瘍DNA（ctDNA）、腫瘍教育血小板（TEP）、および細胞外小胞（EV）は、疾患状態の検出を補助するとともに、疾患状態についての関連する予後情報と予測情報を提供する場合がある。

30

【0103】

miRNAは、真核生物の核DNAによりコードされると、mRNA分子内の相補的配列との塩基対合を介して機能し、通常は翻訳抑制または標的分解を介して遺伝子サイレンシングをもたらす場合がある。miRNAは、pri-miRNAと呼ばれる大型RNA前駆物質としてRNAポリメラーゼにより転写される。pri-miRNAは、核内でさらに処理されて、pre-miRNAを産出する場合がある。pri-miRNAは、長さ約70のヌクレオチドであってよく、不完全なステムループ、または「ヘアピン」構造へと折り重ねられる。次いで、pri-miRNAは細胞質へと輸出され、さらなる処理を受けることで成熟miRNAを生成する場合がある。試料のmiRNAプロファイルは、試料中の様々なmiRNAの発現レベルを示す場合がある。

40

【0104】

50

差動的に発現された *miRNA* は、参照試料における発現レベルと比して、過剰発現、上方調節、過小発現、または下方調節される *miRNA* であってよい（例えば、組織試料の検査細胞が、対象細胞、または細胞もしくは無細胞流体試料、あるいは参照発現レベル（参照値）と比較される）。参照発現レベルは、例えば、関連集団（例えば、疫学的に関連のある集団）における「正常」状態（例えば、疾患がない）または対応する目的の疾患状態を反映し得る。

【0105】

いくつかの実施形態では、*miRNA* は、対象試料または参照発現レベルにおける対応 *miRNA* より少なくとも約 1.8 倍多く、またはそれよりも少なく試料中で発現される場合、「差動的に発現される」または「差動的に豊富」と特定され、あるいは、試料と対象試料または参照発現レベルとの間の発現レベル差の統計的有意差（*p* 値）は 0.05 未満である。

10

【0106】

いくつかの実施形態では、*miRNA* は、対照試料または参照発現試料における対応 *miRNA* より約 2 倍、約 3 倍、約 4 倍、約 5 倍、もしくは 5 倍より多く、またはそれよりも少なく試料中で発現される場合、「差動的に発現される」または「差動的に豊富」と特定される。いくつかの実施形態では、発現レベルは、限定されないが $\log_2$ 、CPM（counts per million）、合成スパイクインへの正規化、またはハウスキーピング遺伝子への正規化など、参照標準に基づき正規化される。

【0107】

20

差動的に発現された *miRNA* は、試料中には存在するが参照試料中ではめったに観察されない、または試料中に存在しないが参照試料中によく見られる *miRNA* であってよい（例えば、組織試料の検査細胞が、対象細胞、または細胞もしくは無細胞流体試料、あるいは参照発現レベル（参照値）と比較される）。

【0108】

本明細書中の一態様では、疾患状態に基づき試料を対象と判別するのに有用な *miRNA* パネルを使用する方法が提供される。本明細書中の他の態様では、*miRNA* パネルを使用して結腸細胞増殖性障害を検出、分化、および判別するための方法、アッセイ、ならびにキットが提供される。結腸細胞増殖性障害の非限定的な例として、腺癌、腺腫、ポリープ、扁平上皮癌、カルチノイド腫瘍、肉腫、およびリンパ腫が挙げられる。

30

【0109】

いくつかの実施形態では、上記方法は、結腸細胞増殖性障害の分化、検出、および判別のためのマーカーとして選択される 1 または複数の *miRNA* の使用を含む。

【0110】

I. 定義

本明細書と特許請求の範囲で使用する時、単数形「a」、「an」、および「the」は、前後関係から明らかでない限り、複数形を含む。例えば、「核酸」という用語は複数の核酸を含み、核酸の混合物も含まれる。

【0111】

本明細書で使用する時、「対象」という用語は、概して検査可能または検出可能な遺伝情報を有する実体または媒体を表す。対象は、個人、個体、または患者とすることができる。対照は、例えば哺乳動物などの脊椎動物とすることができる。哺乳動物の非限定的な例として、ヒト、サル、家畜、競技用動物、げっ歯類、およびペットが挙げられる。対象は、癌を患うか、または癌を患う疑いのある個人とすることができる。対象は、対象の癌または他の疾患、障害、もしくは疾病など、対象の健康上もしくは生理学的な状態または疾病を示す症状を呈している場合がある。代替的に、対象は、このような健康上もしくは生理学的な状態または疾病に対して無症候性の可能性がある。

40

【0112】

本明細書で使用する時、「試料」という用語は、概して 1 または複数の対象から取得されるか得られる生体試料を表す。生体試料は、無細胞生体試料または実質的に細胞を含

50

まない生体試料であるか、あるいは処理または分画されることで無細胞生体試料を産出する場合がある。例えば、無細胞生体試料として、無細胞リボ核酸 (c f RNA)、無細胞デオキシリボ核酸 (c f DNA)、無細胞胎児 DNA (c f F DNA)、血漿、血清、尿、唾液、羊水、およびそれらの誘導体が挙げられる場合がある。無細胞生体試料は、エチレンジアミンテトラ酢酸 (EDTA) 回収管、無細胞 RNA 回収管 (例えば、Streck RNA Complete BCT)、または無細胞 DNA 回収管 (例えば、Streck Cell-Free DNA BCT) を用いて取得されるか、得られる場合がある。無細胞生体試料は、分画化 (例えば、差動遠心分離) により全血試料から得られる場合がある。生体試料またはその誘導体は、細胞を含有してよい。例えば、生体試料は、血液試料またはその誘導体 (例えば、回収管または血液滴により採取した血液) であってよい。

10

【0113】

本明細書で使用する時、「核酸」という用語は、概して多量体型のあらゆる長さのヌクレオチド、すなわちデオキシリボヌクレオチド (d NTP) もしくはリボヌクレオチド (r NTP) のいずれか、またはそれらのアナログを表す。核酸は、あらゆる三次元構造を有し、既知または未知のあらゆる機能を実行する場合がある。核酸の非限定的な例として、デオキシリボ核酸 (DNA)、リボ核酸 (RNA)、遺伝子もしくは遺伝子フラグメントのコーディングまたはノンコーディング領域 連結解析から定められる座位 (loc i (locus))、エクソン、イントロン、メッセンジャ RNA (mRNA)、転移 RNA (tRNA)、リボソーム RNA (rRNA)、短鎖干渉 RNA (siRNA)、シ

20

【0114】

本明細書で使用する時、「標的核酸」という用語は、概して塩基配列を有する核酸分子集団中にある核酸分子を表し、その存在、量、もしくは配列、またはそれらの変更を判定することが所望される。標的核酸は、DNA、RNA、およびそのアナログを含むあらゆる種類の核酸であってよい。本明細書で使用する時、「標的リボ核酸 (RNA)」は、概して RNA である標的核酸を表す。本明細書で使用する時、「標的デオキシリボ核酸 (DNA)」は、概して DNA である標的核酸を表す。

30

【0115】

本明細書で使用する時、「増幅すること (amplifying)」および「増幅 (amplification)」という用語は、概して核酸分子のサイズまたは量を増加させることを表す。核酸分子は、一本鎖または二本鎖であってよい。増幅は、核酸分子の 1 または複数のコピーまたは「増幅産物」を生成することを含む場合がある。増幅は、例えば伸長 (例えば、プライマー伸長) またはライゲーションにより実行されてよい。増幅は、一本鎖核酸分子に相補的な鎖を生成するためにプライマー伸長反応を実行することを含む場合があり、場合によっては、この鎖および／または一本鎖核酸分子の 1 または複数のコピーを生成する。「DNA 増幅」という用語は、概して DNA 分子または「増幅 DNA 産物」の 1 または複数のコピーを生成することを表す。「逆転写増幅」という用語は、概して逆転写酵素の作用を介してリボ核酸 (RNA) 鋳型からデオキシリボ核酸 (DNA) の生成を表す。

40

【0116】

「無細胞核酸」または「c f NA」という用語は、本明細書で使用する時、概して細胞に含有されない生物学的試料中の核酸を表す。c f NA の非限定的な例として、無細胞

50

RNA (cfRNA) と無細胞DNA (cfDNA) が挙げられる。cfRNA は、血流中などにある体液中で自由に循環する場合がある。

【0117】

本明細書で使用する時、「無細胞試料」という用語は、概して無傷の細胞を実質的に欠いた生体試料を表す。無細胞試料は、それ自体が実質的に細胞を欠いた生体試料から得られるか、または細胞が除去されている試料から得られる場合がある。無細胞試料の非限定的な例として、血液、血清、血漿、尿、精液、痰、糞便、乳管渗出液 (ductal exudate)、リンパ液、および回収済洗浄液 (recovered lavage) から得たものが挙げられる。

【0118】

本明細書で使用する時、「循環腫瘍DNA」または「ctDNA」という用語は、概して腫瘍由来のcfDNAを表す。

【0119】

本明細書で使用する時、「結腸細胞増殖性障害」という用語は、概して結腸もしくは直腸における細胞の乱雑または異常な増殖を含む障害あるいは疾患を表す。結腸細胞増殖性障害の非限定的な例として、腺腫 (腺腫性ポリープ)、固着性鋸歯状腺腫 (SSA)、進行型腺腫、大腸異形成、大腸腺腫、大腸癌、結腸癌、直腸癌、大腸癌腫、大腸腺癌、カルチノイド腫瘍、胃腸カルチノイド腫瘍、胃腸間質腫瘍 (GIST)、リンパ腫、および肉腫が挙げられる。

【0120】

本明細書で使用する時、「健康な (healthy)」という用語は、概して結腸直腸細胞増殖障害を患っていない対象を表す。健康は動的な状態ではあるが、本明細書で使用する時、この用語は、特定の記述において言及がなされている疾患状態がない、対象の病理学的状態を表す。一例では、大腸癌を患う対象を分類可能なシグネチャパネルについて言及するとき、健康な個体、健康な試料、または健康な個体から得た試料は、大腸癌 (CRC)、進行性腺腫 (AA)、または良性腺腫 (NAA) がない個体を表す。他の疾患または健康状態が対象に存在する場合があるが、本明細書で使用する時、「健康な」という用語は、論じられている疾患を患う対象とその疾患がない対象との比較または分類を行う目的で、明記された疾患がない状態を示す。

【0121】

「最小残存疾患」または「MRD」という用語は、概して癌処置後の対象の身体に存在する癌細胞が少数であることを表す。癌処置の有効性を判定するとともにさらなる処置計画の指針を出すために、MRD検査が実行されてよい。

【0122】

本明細書で使用する時、「スクリーニング」という用語は、概して大腸癌または大腸腺腫を患うリスクのある対象の集団に対し審査または検査を行うことを表し、健康な対象を、未診断の大腸癌もしくは大腸腺腫を患う対象、またはこのような適応症を患うリスクが高い対象と区別することを目的とする。

【0123】

本明細書で使用する時、「大腸癌」という用語は、概して小腸より下にある腸管 (例えば、大腸 (結腸)、例えば、盲腸、上行結腸、横行結腸、下行結腸、s 状結腸、および直腸) の細胞の癌を特徴とする医学的状态を表す。

【0124】

本明細書で使用する時、「大腸腺腫」という用語は、概して大腸癌の良性および前期である、腺腫性ポリープとも呼ばれる結腸の腺腫を表す。大腸腺腫は、大腸癌に進行するリスクが高いことを示す場合がある。

【0125】

本明細書で使用する時、「進行型大腸腺腫」は、概してサイズが少なくとも 10 mm であるか、または組織学的に高悪性度の異形成または絨毛成分が 20 % を超える腺腫を表す。

10

20

30

40

50

【0126】

本明細書で使用する時、「最小侵襲性生体試料」または「非侵襲性試料」という用語は、概して、対象から血液を取得するのに使用される細い針を除く器具を必要とすることなく、患者の身体から採取されるあらゆる試料を表す。いくつかの実施形態では、低侵襲性生体試料として、血液、血清、または血漿の試料が挙げられる。

【0127】

本明細書で使用する時、「上方調節された (up-regulated)」または「過剰発現された (over-expressed)」という用語は、概して、所与の「閾値」または「カットオフ値」に対して発現レベルが少なくとも5%、少なくとも10%、少なくとも15%、少なくとも20%、少なくとも25%、少なくとも30%、少なくとも35%、少なくとも40%、少なくとも45%、少なくとも50%、少なくとも55%、少なくとも60%、少なくとも65%、少なくとも70%、少なくとも75%、少なくとも80%、少なくとも85%、少なくとも90%、少なくとも95%、少なくとも100%、少なくとも110%、少なくとも120%、少なくとも130%、少なくとも140%、少なくとも150%、または150%より多く増加することを表す。

【0128】

本明細書で使用する時、「閾値」または「カットオフ値」という用語は、概して、対象の発現レベルが閾値、カットオフ値、または参照レベルを超える場合に、対象が所与の感度と特異性で大腸癌または大腸腺腫を患う可能性が高いことを示す参照発現レベルを表す。

【0129】

本明細書で使用する時、「キット」という用語は、あらゆる特定のデバイスに限定されず、限定されないがマイクロアレイ、バイオアレイ、バイオチップ、またはバイオチップアレイなど、本開示のシステムと方法を実施するのに適したあらゆるデバイスを含む場合がある。

【0130】

本明細書で使用する時、癌の「種類」と「サブタイプ」という用語は、概して、乳癌などの癌の1つの「種類」が、例えば、段階、形態学、組織学、遺伝子発現、受容体プロファイル、変異プロファイル、攻撃性、予後、および悪性特徴に基づき「サブタイプ」であり得るように、比較的明細書中で使用される。同様に、「種類」と「サブタイプ」は、より精密なレベルで、例えば、1つの組織学的な「種類」を、例えば変異プロファイルまたは遺伝子発現に従い定義される「サブタイプ」へと分化するために適用される場合がある。疾患進行に関する組織学的かつ病理学的特徴に基づき癌の種類の分類を表すために、癌の「ステージ」も使用される。

【0131】

本明細書で使用する時、「miRNA」もしくは「miR」または「マイクロRNA」という用語は、概してコーディングRNAにハイブリダイズするとともにその発現を調節する、長さ17~25の間である核酸塩基のノンコーディングRNAを表す。17~25個のヌクレオチドmiRNA分子は、天然処理経路を通じて（例えば、無傷の細胞または細胞溶解物を用いて）、または合成処理経路により（例えば、単離されたDicer、Argonaute、またはRNAase IIIなどの単離された処理酵素を用いて）miR前駆物質から取得することができる。17~25個のヌクレオチドRNA分子は、miR前駆物質から処理されることなく、生物学的合成または化学的合成により直接産出することもできる。

【0132】

本明細書で使用する時、「miRNA分子」という用語は、概してmiRNAを表現するあらゆる核酸分子を表す。非限定的な例として、天然miRNA分子、pre-miRNA、pri-miRNA、およびmiRNA分子が挙げられ、これらは、核酸配列において天然形態と同一であるほか、これら天然形態の核酸配列と同一であり、天然形態の核酸配列では、1または複数の核酸が、1または複数のDNAヌクレオチドおよび／また

10

20

30

40

50

は核酸アナログにより置き換えられるか、あるいは表現されている。場合により、miRNA分子は、miRNAをコードする核酸分子、または単に核酸分子と称される。

【0133】

本明細書で使用する時、「miRNAプロファイル」という用語は、概して複数のmiRNAの発現レベルまたは存在量の集まりを表す。miRNAプロファイルは、個々のmiRNAの発現レベルまたは存在量の量的尺度である。ここで、各miRNAは数値で表される。個々のmiRNAの値が高いほど、このmiRNAの発現レベルが高くなる。miRNAプロファイルは、生体試料のRNAから取得される。miRNAプロファイルを判定するのに使用されてよい技法の非限定的な例として、マイクロアレイ、RT-PCR、および次世代シーケンシングが挙げられる。RNA、全RNA、またはそのあらゆる画分は、解析の出発材料として使用することができる。miRNAプロファイルにより判定される複数のmiRNAは、1つの既知のmiRNAの選択から、すべての既知のmiRNAの選択にまで及ぶ可能性がある。

10

【0134】

本明細書で使用する時、「miRNAからなる所定のセット (pre-determined set of miRNAs)」または「miRNAシグネチャ」という用語は、概して条件1と別の条件2とを判別可能な、miRNAからなる固定のセットを表す。例えば、条件1は大腸癌であり、条件2は正常対照である。この場合、対応するmiRNAからなる所定のセットは、大腸癌患者または正常対照患者から得た試料を区別可能である。代替的に、条件1が大腸癌であり、条件2が進行性腺腫である場合、対応するmiRNAからなる所定のセットは、大腸癌患者と進行性腺腫患者とを区別可能である。サンプル分析を、例えばmiRNAプロファイルの判定に使用されてよいマトリクス上で実行するには、これらのmiRNAからなる固定のセットは、プローブ、またはmiRNAからなる所定のセットにより定義される他の方法により表される。標的方法を用いたシーケンシング、例えば、トランスクリプトームワイドmiRNAシーケンシングおよびdd/q-PCR法のための方法を、選択することができる。例えば、健康な対照から得た大腸癌診断用のmiRNAからなる所定のセットが25個のmiRNAからなる時、これら25個のmiRNAを検出可能なプローブまたは方法は、診断解析の実行において実施される必要がある。

20

【0135】

本明細書で使用する時、「共通miRNAシグネチャプロファイル」という用語は、概して条件1と別の条件2とを区別可能である、miRNAまたはノンコーディングRNAからなる固定のセットを表す。共通miRNAまたはノンコーディングRNAのシグネチャプロファイルは、例えばデータベースに記憶された複数のmiRNAプロファイルからその場で算出される。ある状態と別の状態2とを区別可能な共通のmiRNAシグネチャプロファイルは、健康状態1または別の状態2のいずれかに関連するデータベースに新たなプロファイルが追加されるとすぐに変化する。この点で、共通miRNAシグネチャプロファイルは、miRNAからなる所定のセットとは異なる。さらに、共通miRNAシグネチャプロファイル、例えば、データベースに記憶されたmiRNAプロファイルを生成するための基礎は、例えば、可能な限り多くのmiRNAを検出するために、可能な限り多くの異なる捕捉プローブを表しているマトリクス上で、捕捉プローブから生成される。

30

40

【0136】

本明細書で使用する時、「ノンコーディングRNA」または「ncRNA」という用語は、概してタンパク質へと翻訳されない機能性RNA分子を表す。場合によりncRNAは、非タンパク質コーディングRNA (npcRNA)、非メッセンジャRNA (nmRNA)、小型非メッセンジャRNA (snmRNA)、または機能性RNA (fRNA)を表す。スモールRNA (sRNA)という用語は、多くの場合、細菌ncRNAに使用される。ノンコーディングRNAが最終生成物として転写されるDNA配列は、多くの場合RNA遺伝子またはノンコーディングRNA遺伝子と呼ばれる。ノンコーディングRNA

50

A 遺伝子として、転写 RNA (tRNA) やリボソーム RNA (rRNA) などの高度に豊富で機能的に重要な RNA のほか、snoRNA、マイクロ RNA、siRNA、および piRNA などの RNA、Xist や HOTAIR などの長鎖 ncRNA が挙げられる。ヒトゲノム内にコードされる ncRNA の数は、不明である。しかし、近年のトランスクリプトミクス研究とバイオインフォマティクス研究で、数千もの ncRNA の存在が認められている。新たに特定された ncRNA の大半は機能が検証されていないため、多くの ncRNA は非機能性の場合がある。

【0137】

II. 試料のアッセイ

無細胞生体試料は、ヒト対象から取得されるか、または得られる場合がある。無細胞生体試料は、処理前に、様々な温度（例えば、室温、冷蔵または冷凍条件下、例えば 25℃、4℃、-18℃、-20℃、または -80℃）や様々な懸濁器具 (suspensions)（例えば、EDTA 回収管、無細胞 RNA 回収管、または無細胞 DNA 回収管）などの様々な保管条件で保管される場合がある。

10

【0138】

無細胞生体試料は、癌を患う対象、癌を患っている疑いのある対象、または癌を患っていないかその疑いがない対象から取得されてよい。

【0139】

無細胞生体試料は、癌を患う対象を処置する前および／または後に取得されてよい。無細胞生体試料は、処置または処置計画の間に対象から取得されてよい。処置の効果を経時的にモニタリングするために、複数の無細胞生体試料が対象から取得されてよい。無細胞生体試料は、臨床検査を介して確定的な陽性または陰性の診断が入手不能な癌を患っていることが明瞭であるかその疑いのある対象から入手される場合がある。試料は、癌を患う疑いのある対象から入手される場合がある。無細胞生体試料は、疲労、悪心、体重減少、痛みと疼痛、衰弱、または出血などの原因不明の症状を発症する対象から入手される場合がある。無細胞生体試料は、原因が明らかな症状を患う対象から入手される場合がある。無細胞生体試料は、家族歴、年齢、高血圧もしくは高血圧前段階、糖尿病もしくは糖尿病前段階、体重過多もしくは肥満、環境曝露、生活様式のリスク因子（例えば、喫煙、飲酒、または薬物使用）、または他のリスク因子の存在などの因子が原因で、癌を発症するリスクのある対象から入手される場合がある。

20

30

【0140】

無細胞生体試料は、アッセイを行いトランスクリプトームデータを生成するのに適した無細胞リボ核酸 (cfRNA) 分子、アッセイを行いゲノムデータを生成するのに適した無細胞デオキシリボ核酸 (cfDNA) 分子、またはそれらの混合物もしくは組合せなど、アッセイの対象とすることができる 1 または複数の分析物を含有する場合がある。このような 1 または複数の分析物（例えば、cfRNA 分子および／または cfDNA 分子）は、1 または複数の適切なアッセイを用いた下流アッセイ (downstream assay) を行うために、対象の 1 または複数の無細胞生体試料から単離されるか、または抽出されてよい。

【0141】

対象から無細胞生体試料を取得した後、無細胞生体試料は、対象の癌を示すデータセットを生成するために処理される場合がある。例えば、癌関連ゲノム座位のパネルにおける無細胞生体試料の核酸分子の存在、不在、または定量的評価（例えば、癌関連ゲノム座位における RNA 転写物または DNA の定量的測定）である。対象から取得した無細胞生体試料の処理は、(i) 無細胞生体試料を、複数の核酸分子を単離、富化、または抽出するのに十分な条件に晒すこと、および (ii) 複数の核酸分子を解析してデータセットを生成することを含む場合がある。

40

【0142】

いくつかの実施形態では、複数の核酸分子は、無細胞生体試料から抽出され、シーケンシングにかけられることで、複数のシーケンシングリードが生成される。核酸分子は、リ

50

ボ核酸（RNA）またはデオキシリボ核酸（DNA）を含んでよい。核酸分子（例えば、RNAまたはDNA）は、MagMAX mirVana Total RNA Isolation Kit、QIAamp ccfDNA/RNA Kit、Zymo Quick-ccfRNA Serum & Plasma Kit、MP BiomedicalsのFastDNA Kitプロトコル、QiagenのQIAamp DNA無細胞生体ミニキット、またはNorgen Biotekの無細胞生体DNA単離キットなど、様々な方法により無細胞生体試料から抽出される場合がある。抽出方法は、試料からすべてのRNAまたはDNA分子を抽出する場合がある。代替的に、抽出方法は、試料からRNAまたはDNA分子の一部を選択的に抽出する場合がある。試料から抽出したRNA分子は、逆転写（RT）によりDNA分子に変換される場合がある。

10

【0143】

シーケンシングは、大規模並列シーケンシング（MPS）、ペアエンドシーケンシング、抗スループットシーケンシング、次世代シーケンシング（NGS）、ショットガンシーケンシング、一分子シーケンシング、ナノポアシーケンシング、半導体シーケンシング、パイロシーケンシング、合成によるシーケンシング、ライゲーションによるシーケンシング、ハイブリダイゼーションによるシーケンシング、およびRNA-Seq（Illumina）などのあらゆる適切なシーケンシング法により実行されてよい。

【0144】

シーケンシングは、例えばRNA分子またはDNA分子の核酸増幅を含む場合がある。いくつかの実施形態では、核酸増幅は、ポリメラーゼ連鎖反応（PCR）である。適切な回数のPCR（例えば、PCR、qPCR、逆転写酵素PCR、デジタルPCRなど）が、核酸（例えば、RNAまたはDNA）の初期量を、その後の配列決定のために所望の入力量にまで十分に増幅するために実施される場合がある。場合により、PCRは、標的核酸の大規模増幅に使用されてよい。この大規模増幅は、最初に異なる分子へとライゲートされ得るアダプタ配列を使用すること、続いて普遍的プライマーを使用するPCR増幅を含む場合がある。PCRは、例えば、Life Technologies、Affymetrix、Promega、Qiagenなどが提供する多数の市販キットのうちいずれかを用いて実施される場合がある。他の場合、核酸集団内の特定の標的核酸のみが増幅されてよい。特定のプライマーは、場合によりアダプタライゲーションと組み合わせて、下流シーケンシングのための特定の標的を選択的に増幅するために使用される場合がある。PCRは、癌に関連するゲノム座位などの1または複数のゲノム座位の標的化増幅を含む場合がある。シーケンシングは、Qiagen、NEB、Thermo Fisher Scientific、またはBio-RadによるOneStep RT-PCRキットプロトコルなどの同時逆転写（RT）およびポリメラーゼ連鎖反応（PCR）の使用を含む場合がある。

20

30

【0145】

無細胞生体試料から単離または抽出したRNAまたはDNA分子は、複数の試料のマルチプレキシング（multiplexing）を可能にするために、例えば識別可能なタグでタグ付けされてよい。あらゆる数のRNAまたはDNA試料がマルチプレキシングされてよい。例えば、マルチプレキシングされた反応は、少なくとも2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、25、30、35、40、45、50、55、60、65、70、75、80、85、90、95、100、または100より多くの初期無細胞生体試料から得たRNAまたはDNAを含有する場合がある。例えば、複数の無細胞生体試料は、各DNA分子が、その由来となる試料（および対象）に戻って追跡され得るように、試料バーコードによりタグ付けされてよい。このようなタグは、ライゲーションにより、またはプライマーを用いたPCR増幅によりRNAまたはDNA分子に結合されてよい。

40

【0146】

核酸分子をシーケンシングにかけた後、癌の存在、不在、または相対的精査を示すデータを生成するために、適切なバイオインフォマティクスプロセスがシーケンスリードに対

50

し実行されてよい。例えば、シーケンスリードは、1または複数の参照ゲノム（例えば、ヒトゲノムなどの1または複数の種のゲノム）へとアライメントされる場合がある。アライメントされたシーケンスリードは、癌を示すデータセットを生成するために1または複数のゲノム座位において定量化されてよい。例えば、癌に関連付けられた複数のゲノム座位に相当する配列に対し定量化を行うことで、癌を示すデータセットを生成してよい。

【0147】

アッセイの読み出しは、癌を示すデータを生成するために1または複数のゲノム座位（例えば、癌関連ゲノム座位）にて定量化されてよい。例えば、複数のゲノム座位（例えば、癌関連ゲノム座位）に相当するアレイハイブリダイゼーションまたはポリメラーゼ連鎖反応（PCR）に対し定量化を行うことで、癌を示すデータを生成してよい。アッセイの読み出しは、量的PCR（qPCR）値、デジタルPCR（dPCR）値、デジタルドロップレットPCR（ddPCR）値、蛍光値など、またはそれらの正規化された値を含む場合がある。このアッセイは、家庭設定で実行されるように構成された家庭用検査であってよい。

【0148】

いくつかの実施形態では、対象の無細胞生体試料を同時に処理するために、複数のアッセイが使用されてよい。例えば、対象から取得されるか得られる第1の無細胞生体試料を処理して、癌を示す第1のデータセットを生成するために第1のアッセイが使用され、かつ、対象から取得されるか得られる第2の無細胞生体試料を処理して、癌を示す第2のデータセットを生成するために第2のアッセイが使用される場合がある。次いで、第1のデータセットおよび第2のデータセットのいずれかまたは両方を解析し、対象の癌を評価してよい。例えば、第1のデータセットと第2のデータセットの組み合わせに基づき、1つの診断指標または診断スコアを生成することができる。別の例として、第1のデータセットと第2のデータセットに基づき、別個の診断指標または診断スコアを生成することができる。

【0149】

III. シグネチャパネル

本開示は、結腸細胞増殖性障害の発症に関連付けられる試料中で特定されるmiRNA分子の組合せから測定可能な特徴を得るために生体試料を解析するための方法とシステムを提供する。特定されたmiRNA分子を回収することで、シグネチャが結腸細胞増殖性障害またはその段階の特徴を示すシグネチャパネルを形成してよい。シグネチャパネルパネルから得た特徴は、対象の集団を、結腸細胞増殖性障害を患っているか、または結腸細胞増殖性障害を患っていないと層別化するように構成された分類器を作製するために、訓練されたアルゴリズム（例えば、機械学習モデル）を用いて処理されてよい。この方法は、シグネチャパネルに記載されるmiRNAを有する1または複数の核酸を用いることを特徴とする場合がある。

【0150】

本明細書に記載のmiRNAシグネチャパネルは、結腸細胞増殖性障害に関連付けられた特定のmiRNAに対する迅速かつ特異的な分析を可能にし得る。本明細書に記載されるとともに本明細書中の方法に使用されるシグネチャパネルは、結腸細胞増殖性障害の診断、予後、処置選択、およびモニタリング（例えば、処置のモニタリング）の改善のために使用されてよい。

【0151】

シグネチャパネルと方法は、全血、血漿、または血清などの体液試料から早期結腸細胞増殖性障害を検出するために使用されるマーカまたはシグネチャパネルが必要であるという点で、現行の手法よりも顕著な改善を提供する。結腸細胞増殖性障害を検出かつ診断するのに使用される現行の方法として、大腸内視鏡検査、s状結腸鏡検査、および便潜血結腸癌検査（fecal occult blood colon cancer）が挙げられる場合がある。これらの方法と比較して、本明細書に提供される方法は、大腸内視鏡検査よりも侵襲性ははるかに低く、同様に、s状結腸鏡検査、糞便免疫化学検査（FIT

10

20

30

40

50

）、および糞便潜血検査（FOBT）より感度が高いと言わないまでも侵襲性が低い場合がある。これらマーカの現在の使用と比較して、本明細書で提供される方法は、遺伝子パネルと感度の高いアッセイ技法を組み合わせることで利点により、感度と特異性の観点から有意な利点を提供し得る。

【0152】

有益なmiRNAを含むシグネチャパネルは、意図したアッセイの目的に応じて選択され得る。標的とされた方法では、意図した標的miRNAのセットに基づきプライマー対が設計されてよい。いくつかの実施形態では、miRNAのセットは、表1に列挙される群から選択される少なくとも1つ、少なくとも2つ、少なくとも3つ、または3つより多くのmiRNAを含む。いくつかの実施形態では、miRNAのセットは、表1に列挙される領域すべてを含む。いくつかの実施形態では、大腸癌に関連するmiRNAのセットは、表1に列挙される群から選択される。

【0153】

10

20

30

40

50

【表 1 - 1】

表 1

hsa-mir-889	hsa-mir-6511a-3	hsa-mir-548l	hsa-mir-4484	hsa-mir-320c-1
hsa-mir-543	hsa-mir-376a-2	hsa-mir-548v	hsa-mir-4521	hsa-mir-320d-1
hsa-mir-376b	hsa-mir-155	hsa-mir-106b	hsa-mir-4706	hsa-mir-320d-2
hsa-mir-335	hsa-mir-3140	hsa-mir-133b	hsa-mir-487a	hsa-mir-324
hsa-mir-1185-1	hsa-mir-1277	hsa-mir-548h-2	hsa-mir-548ac	hsa-mir-331
hsa-mir-548k	hsa-mir-340	hsa-mir-136	hsa-mir-5588	hsa-mir-3529
hsa-mir-12135	hsa-mir-548n	hsa-mir-26b	hsa-mir-579	hsa-mir-374c
hsa-mir-369	hsa-mir-518b	hsa-mir-433	hsa-mir-6734	hsa-mir-431
hsa-mir-190a	hsa-mir-654	hsa-mir-4719	hsa-mir-6882	hsa-mir-450b
hsa-mir-6770-1	hsa-mir-5581	hsa-mir-3610	hsa-mir-9-1	hsa-mir-4651
hsa-mir-382	hsa-mir-409	hsa-mir-374b	hsa-mir-93	hsa-mir-4716
hsa-mir-1843	hsa-mir-628	hsa-mir-376c	hsa-let-7i	hsa-mir-4763
hsa-mir-142	hsa-mir-10399	hsa-mir-4779	hsa-mir-103b-1	hsa-mir-4772
hsa-mir-485	hsa-mir-3184	hsa-mir-496	hsa-mir-103b-2	hsa-mir-548a-1
hsa-mir-548ax	hsa-mir-423	hsa-mir-622	hsa-mir-12136	hsa-mir-548ah
hsa-mir-548e	hsa-mir-548z	hsa-mir-671	hsa-mir-1283-2	hsa-mir-548aj-1
hsa-mir-548al	hsa-mir-374a	hsa-mir-6876	hsa-mir-1287	hsa-mir-5703

10

20

30

40

【 0 1 5 4】

50

【表 1 - 2】

hsa-mir-548am	hsa-mir-548a-3	hsa-let-7b	hsa-mir-130a	hsa-mir-574
hsa-mir-590	hsa-mir-6770-2	hsa-mir-103a-1	hsa-mir-146a	hsa-mir-6729
hsa-mir-135a-2	hsa-mir-1185-2	hsa-mir-103a-2	hsa-mir-146b	hsa-mir-6781
hsa-mir-6770-3	hsa-mir-6077	hsa-mir-10400	hsa-mir-151b	hsa-mir-6802
hsa-mir-410	hsa-mir-3202-1	hsa-mir-135a-1	hsa-mir-154	hsa-mir-7-1
hsa-mir-376a-1	hsa-mir-548o-2	hsa-mir-139	hsa-mir-18a	hsa-mir-7111
hsa-mir-377	hsa-mir-3143	hsa-mir-191	hsa-mir-2278	hsa-mir-9902-1
hsa-mir-570	hsa-mir-5009	hsa-mir-2392	hsa-mir-2355	
hsa-mir-381	hsa-mir-548g	hsa-mir-26a-2	hsa-mir-3138	
hsa-mir-665	hsa-mir-656	hsa-mir-320c-2	hsa-mir-3168	
hsa-mir-758	hsa-mir-6818	hsa-mir-4468	hsa-mir-3202-2	

10

20

30

【0155】

いくつかの実施形態では、大腸癌に関連するmiRNAのセットは、表2に列挙される群から選択される。

【0156】

40

50

【表 2】

表 2

hsa-mir-889
hsa-mir-543
hsa-mir-376b
hsa-mir-335
hsa-mir-1185-1

10

【0157】

いくつかの実施形態では、大腸癌に関連するmiRNAのセットは、表3に列挙される群から選択される。

20

【0158】

【表 3】

表 3

hsa-mir-889	hsa-mir-548k
hsa-mir-543	hsa-mir-12135
hsa-mir-376b	hsa-mir-369
hsa-mir-335	hsa-mir-190a
hsa-mir-1185-1	hsa-mir-6770-1

30

【0159】

いくつかの実施形態では、大腸癌に関連するmiRNAのセットは、表4に列挙される群から選択される。

40

【0160】

50

【表 4】

表 4

hsa-mir-889	hsa-mir-548k	hsa-mir-382
hsa-mir-543	hsa-mir-12135	hsa-mir-1843
hsa-mir-376b	hsa-mir-369	hsa-mir-142
hsa-mir-335	hsa-mir-190a	hsa-mir-485
hsa-mir-1185-1	hsa-mir-6770-1	hsa-mir-548ax

10

【0161】

いくつかの実施形態では、大腸癌に関連するmiRNAのセットは、表5に列挙される群から選択される。

20

【0162】

【表 5】

表 5

hsa-mir-889	hsa-mir-548k	hsa-mir-382	hsa-mir-548e
hsa-mir-543	hsa-mir-12135	hsa-mir-1843	hsa-mir-548al
hsa-mir-376b	hsa-mir-369	hsa-mir-142	hsa-mir-548am
hsa-mir-335	hsa-mir-190a	hsa-mir-485	hsa-mir-590
hsa-mir-1185-1	hsa-mir-6770-1	hsa-mir-548ax	hsa-mir-135a-2

30

40

【0163】

差動的に発現されたmiRNAのプロファイルは、対照レベルまたは参照レベルと比較して流体または組織試料中に差動的に発現される一組のmiRNAを表す。特異的に発現されたmiRNAのプロファイルは、下方調節または過小発現されたmiRNAのプロファイル、および上方調節／過剰発現されたmiRNAのプロファイルを含む。

【0164】

いくつかの実施形態では、miRNAは、結腸細胞増殖性障害を発症するリスクが低い対象の対応試料と比較して、結腸細胞増殖性障害を発症するリスクが高い対象の試料中で差動的に発現される。いくつかの実施形態では、結腸細胞増殖性障害を発症するリスクが高い対象と、結腸細胞増殖性障害を発症するリスクが低い対象はそれぞれ、大腸病変を患

50

っている。結腸細胞増殖性障害を発症するリスクが高いまたはそれを患う対象の試料中に差動的に発現されるmiRNAは、結腸細胞増殖性障害の診断または予防のためのバイオマーカーとして使用することができる。例えば、結腸細胞増殖性障害を発症するリスクが低い対象の対応細胞と比較して、結腸細胞増殖性障害を発症するリスクが高い対象の試料中に差動的に発現されるmiRNAは、hsa-mir-889、hsa-mir-543、hsa-mir-376b、hsa-mir-335、hsa-mir-1185-1、hsa-mir-548k、hsa-mir-12135、hsa-mir-369、hsa-mir-190a、hsa-mir-6770-1、hsa-mir-382、hsa-mir-1843、hsa-mir-142、およびhsa-mir-485、hsa-mir-548ax、hsa-mir-548e、hsa-mir-548al、hsa-mir-548am、hsa-mir-590、hsa-mir-135a-2、hsa-mir-6770-3、hsa-mir-410、hsa-mir-376a-1、hsa-mir-377、hsa-mir-570、hsa-mir-381、hsa-mir-665、hsa-mir-758、hsa-mir-6511a-3、hsa-mir-376a-2、hsa-mir-155、hsa-mir-3140、hsa-mir-1277、hsa-mir-340、およびhsa-mir-548n、hsa-mir-518b、hsa-mir-654、hsa-mir-5581、hsa-mir-409、hsa-mir-628、hsa-mir-10399、hsa-mir-3184、hsa-mir-423、hsa-mir-548z、hsa-mir-374a、hsa-mir-548a-3、hsa-mir-6770-2、hsa-mir-1185-2、hsa-mir-6077、hsa-mir-3202-1、hsa-mir-548o-2、hsa-mir-3143、hsa-mir-5009、hsa-mir-548g、およびhsa-mir-656のうち、1または複数を含む。

10

20

30

40

【0165】

いくつかの実施形態では、結腸細胞増殖性障害を発症するリスクが低い対象の対応試料と比較して、結腸細胞増殖性障害を発症するリスクが高い対象の試料中に高度に発現されるmiRNAは、hsa-mir-889、hsa-mir-543、hsa-mir-376b、hsa-mir-335、hsa-mir-1185-1、hsa-mir-548k、hsa-mir-12135、hsa-mir-369、hsa-mir-190a、hsa-mir-6770-1、hsa-mir-382、hsa-mir-1843、hsa-mir-142、およびhsa-mir-485、hsa-mir-548ax、hsa-mir-548e、hsa-mir-548al、hsa-mir-548am、hsa-mir-590、hsa-mir-135a-2、hsa-mir-6770-3、hsa-mir-410、hsa-mir-376a-1、hsa-mir-377、hsa-mir-570、hsa-mir-381、hsa-mir-665、hsa-mir-758、hsa-mir-6511a-3、hsa-mir-376a-2、hsa-mir-155、hsa-mir-3140、hsa-mir-1277、hsa-mir-340、およびhsa-mir-548n、hsa-mir-518b、hsa-mir-654、hsa-mir-5581、hsa-mir-409、hsa-mir-628、hsa-mir-10399、hsa-mir-3184、hsa-mir-423、hsa-mir-548z、hsa-mir-374a、hsa-mir-548a-3、hsa-mir-6770-2、hsa-mir-1185-2、hsa-mir-6077、hsa-mir-3202-1、hsa-mir-548o-2、hsa-mir-3143、hsa-mir-5009、hsa-mir-548g、およびhsa-mir-656のうち、1または複数を含む。

【0166】

いくつかの実施形態では、結腸細胞増殖性障害は進行性腺腫である。

【0167】

いくつかの実施形態では、結腸細胞増殖性障害を発症するリスクが低い対象の対応試料

50

と比較して、結腸細胞増殖性障害を発症するリスクが高い対象の試料中に低度に発現されるmiRNAは、hsa-mir-889、hsa-mir-543、hsa-mir-376b、hsa-mir-335、hsa-mir-1185-1、hsa-mir-548k、hsa-mir-12135、hsa-mir-369、hsa-mir-190a、hsa-mir-6770-1、hsa-mir-382、hsa-mir-1843、hsa-mir-142、およびhsa-mir-485、hsa-mir-548ax、hsa-mir-548e、hsa-mir-548al、hsa-mir-548am、hsa-mir-590、hsa-mir-135a-2、hsa-mir-6770-3、hsa-mir-410、hsa-mir-376a-1、hsa-mir-377、hsa-mir-570、hsa-mir-381、hsa-mir-665、hsa-mir-758、hsa-mir-6511a-3、hsa-mir-376a-2、hsa-mir-155、hsa-mir-3140、hsa-mir-1277、hsa-mir-340、およびhsa-mir-548n、hsa-mir-518b、hsa-mir-654、hsa-mir-5581、hsa-mir-409、hsa-mir-628、hsa-mir-10399、hsa-mir-3184、hsa-mir-423、hsa-mir-548z、hsa-mir-374a、hsa-mir-548a-3、hsa-mir-6770-2、hsa-mir-1185-2、hsa-mir-6077、hsa-mir-3202-1、hsa-mir-548o-2、hsa-mir-3143、hsa-mir-5009、hsa-mir-548g、およびhsa-mir-656のうち、1または複数を含む。

10

20

【0168】

一例では、miRNAのパネルは、進行性腺腫がない健康な対象から得た試料中の発現と比して、進行性腺腫を患う対象から得た試料中の発現が増加している。一例では、パネルは、表6に列挙される群から選択される2つ以上のmiRNAを含む。他の例では、パネルは、列挙されたmiRNAのうち3つ以上、4つ以上、5つ以上、または6つ以上を含む。

【0169】

【表6】

表6

30

hsa-mir-12135	hsa-mir-376a-3p	hsa-mir-485-5p	hsa-mir-654-3p
hsa-mir-1277	hsa-mir-376b	hsa-mir-487b-3p	hsa-mir-889
hsa-mir-323b-3p	hsa-mir-376c-3p	hsa-mir-5096	
hsa-mir-335	hsa-mir-382-5p	hsa-mir-548	
hsa-mir-369-5p	hsa-mir-409-3p	hsa-mir-5585-3p	

40

【0170】

進行性腺腫がない健康な対象から得た試料中の発現と比して、進行性腺腫を患う対象から得た試料中の発現が増加しているmiRNAのパネルにおける別の例では、パネルは、表7に列挙される群から選択される2つ以上のmiRNAを含む。他の例では、パネルは、列挙されたmiRNAのうち3つ以上、4つ以上、5つ以上、または6つ以上を含む。

【0171】

50

いくつかの実施形態では、パネルは、a) hsa-miR-1273a、hsa-miR-17-5p、hsa-miR-20a-3p、hsa-miR-20b-5p、b) hsa-miR-3065-5p、hsa-miR-4785、hsa-miR-5096、hsa-miR-5189-5p、またはc) hsa-miR-545-3p、hsa-miR-570-3p、hsa-miR-624-3p、hsa-mir-1181、hsa-mir-6073を含む5つ以上のmiRNAから選択される。

【0172】

【表7】

表7

hsa-miR-1273a	hsa-miR-3065-5p	hsa-miR-545-3p	hsa-mir-6073
hsa-miR-17-5p	hsa-miR-4785	hsa-miR-570-3p	
hsa-miR-20a-3p	hsa-miR-5096	hsa-miR-624-3p	
hsa-miR-20b-5p	hsa-miR-5189-5p	hsa-mir-1181	

10

20

【0173】

一例では、miRNAのパネルは、進行性腺腫がない健康な対象から得た試料中の発現と比して、進行性腺腫を患う対象から得た試料中の発現が減少している。一例では、パネルは、表8に列挙される群から選択される2つ以上のmiRNAを含む。他の例では、パネルは、列挙されたmiRNAのうち3つ以上、4つ以上、5つ以上、または6つ以上を含む。

【0174】

30

40

50

【表 8】

表 8

hsa-mir-34a-5p
hsa-mir-3687
hsa-mir-4492
hsa-mir-455-3p
hsa-mir-4727-3p
hsa-mir-524-5p
hsa-mir-5701

10

20

【 0 1 7 5 】

一例では、m i R N Aのパネルは、大腸癌がない健康な対象から得た試料中の発現と比して、大腸癌を患う対象から得た試料中の発現が増加している。一例では、パネルは、表 9 に列挙される群から選択される 2 つ以上のm i R N Aを含む。他の例では、パネルは、列挙されたm i R N Aのうち 3 つ以上、4 つ以上、5 つ以上、または 6 つ以上を含む。

【 0 1 7 6 】

30

40

50

【表 9】

表 9

hsa-mir- 10a-3p	hsa-mir- 196a-5p	hsa-mir- 3183	hsa-mir- 6089-2
hsa-mir- 1207	hsa-mir- 200a-3p	hsa-mir- 3194-5p	hsa-mir- 6749
hsa-mir- 1233-5p	hsa-mir- 200a-5p	hsa-mir- 3683	hsa-mir- 6760-5p
hsa-mir- 1246	hsa-mir- 200b-3p	hsa-mir- 4436b-1	hsa-mir- 6846
hsa-mir- 1247-5p	hsa-mir- 200b-5p	hsa-mir- 4484	hsa-mir- 6867-5p
hsa-mir- 1273d	hsa-mir- 200c-3p	hsa-mir- 4670	hsa-mir- 6873
hsa-mir- 1273e	hsa-mir- 203a-3p	hsa-mir- 4683	hsa-mir- 7110
hsa-mir- 1273f	hsa-mir- 205-3p	hsa-mir- 4779	hsa-mir- 718
hsa-mir- 1275	hsa-mir- 3131	hsa-mir- 514b-5p	hsa-mir- 7704
hsa-mir- 1290	hsa-mir- 3132	hsa-mir- 552-3p	hsa-mir- 7847-3p
hsa-mir- 141-3p	hsa-mir- 3141	hsa-mir- 5692a-2	hsa-mir- 9902-1
hsa-mir- 1910-3p	hsa-mir- 3180-3p	hsa-mir- 5698	hsa-mir- 9902-2

10

20

30

【0177】

大腸癌がない健康な対象から得た試料中の発現と比して、大腸癌を患う対象から得た試料中の発現が増加しているmiRNAのパネルにおける別の例では、パネルは、表10に列挙される群から選択される2つ以上のmiRNAを含む。他の例では、パネルは、列挙されたmiRNAのうち3つ以上、4つ以上、5つ以上、または6つ以上を含む。

40

【0178】

別の例では、パネルは、a) hsa-miR-1250-5p、hsa-miR-1255a、hsa-miR-223-3p、hsa-miR-338-3p、hsa-miR-338-5p、b) hsa-miR-424-5p、hsa-miR-424-3p、hsa-miR-450a-5p、hsa-miR-450b-5p、hsa-miR-4772-3p、c) hsa-miR-4772-5p、hsa-miR-625-5p、hsa-miR-7847-3p、hsa-miR-1181、hsa-miR-3651、hsa-miR-6073、d) hsa-miR-6125、hsa-miR-7704、hsa-miR-19b-3p、hsa-miR-19a-3p、hsa-m

50

i R-3157-5p、e) hsa-miR-142-3p、hsa-miR-30c-5p、hsa-miR-6741-5p、hsa-miR-590-3p、hsa-miR-4685-5p、f) hsa-miR-3648、hsa-miR-331-3p、hsa-miR-1303、hsa-miR-6790-3p、hsa-miR-6867-5p、hsa-miR-942-5p、g) hsa-miR-378a-3p、hsa-miR-1287-5p、hsa-miR-4785、hsa-miR-324-3p、hsa-miR-550b-2-5p、h) hsa-miR-200c-3p、hsa-miR-200b-3p、hsa-miR-3679-5p、hsa-miR-550a-3-5p、hsa-miR-3187-3p、i) hsa-miR-181b-5p、hsa-miR-3138、hsa-miR-146a-5p、hsa-miR-6721-5p、hsa-miR-23b-3p、hsa-miR-28-5p、j) hsa-miR-320d、hsa-miR-940、hsa-miR-320d-1、hsa-miR-10a-5p、hsa-miR-340-5p、k) hsa-miR-320b、hsa-miR-335-5p、hsa-miR-320c、hsa-miR-501-3p、hsa-miR-548n、またはl) hsa-miR-27a-3p、hsa-miR-3065-3p、hsa-miR-548aa@、hsa-miR-584-3p、hsa-miR-22-3pを含む5つ以上のmiRNAから選択される。

10

【0179】

【表10】

表 10

20

hsa-miR-1250-5p	hsa-mir-6125	hsa-miR-378a-3p	hsa-miR-320d
hsa-miR-1255a	hsa-mir-7704	hsa-miR-1287-5p	hsa-miR-940
hsa-miR-223-3p	hsa-miR-19b-3p	hsa-mir-4785	hsa-miR-320d-1
hsa-miR-338-3p	hsa-miR-19a-3p	hsa-miR-324-3p	hsa-miR-10a-5p
hsa-miR-338-5p	hsa-miR-3157-5p	hsa-miR-550b-2-5p	hsa-miR-340-5p
hsa-miR-424-5p	hsa-miR-142-3p	hsa-miR-200c-3p	hsa-miR-320b
hsa-miR-424-3p	hsa-miR-30c-5p	hsa-miR-200b-3p	hsa-miR-335-5p
hsa-miR-450a-5p	hsa-miR-6741-5p	hsa-miR-3679-5p	hsa-miR-320c
hsa-miR-450b-5p	hsa-miR-590-3p	hsa-miR-550a-3-5p	hsa-miR-501-3p
hsa-miR-4772-3p	hsa-miR-4685-5p	hsa-miR-3187-3p	hsa-miR-548n
hsa-miR-4772-5p	hsa-miR-3648	hsa-miR-181b-5p	hsa-miR-27a-3p
hsa-miR-625-5p	hsa-miR-331-3p	hsa-miR-3138	hsa-miR-3065-3p
hsa-miR-7847-3p	hsa-miR-1303	hsa-miR-146a-5p	hsa-miR-548aa@
hsa-miR-1181	hsa-miR-6790-3p	hsa-miR-6721-5p	hsa-miR-584-3p
hsa-miR-3651	hsa-miR-6867-5p	hsa-miR-23b-3p	hsa-miR-22-3p
hsa-mir-6073	hsa-miR-942-5p	hsa-miR-28-5p	

30

40

【0180】

一例では、miRNAのパネルは、大腸癌がない健康な対象から得た試料中の発現と比べて、大腸癌を患う対象から得た試料中の発現が減少している。一例では、パネルは、表11に列挙される群から選択される2つ以上のmiRNAを含む。

【0181】

50

【表 11】

表 11

hsa-mir-3135b
hsa-mir-581

10

【0182】

いくつかの実施形態では、結腸細胞増殖性障害は進行性腺腫である。

【0183】

いくつかの実施形態では、結腸細胞増殖性障害は大腸癌である。

【0184】

いくつかの実施形態では、miR-889 単独から出発するバイオマーカの特定の具体的な組合せは、miR-889、miR-543、miR-376b、miR-335、およびmiR-1185-1 単独の使用と比較して、進行型大腸腺腫の存在を検出しかつ大腸癌の存在を検出することにおいてk-分割交差検証、AUC、感度、および特性における頻度の観点から良好な結果をもたらす。

20

【0185】

いくつかの実施形態では、このような例は、少なくとも(miR-889)もしくは少なくとも(miR-889およびmiR-543)もしくは少なくとも(miR-889およびmiR-376b)、または少なくとも(miR-889およびmiR-335)、または少なくとも(miR-889およびmiR-1185-1)、または少なくとも(miR-889、miR-543、およびmiR-376b)、または少なくとも(miR-889、miR-543、およびmiR-335)、または少なくとも(miR-889、miR-376b、およびmiR-335)、または少なくとも(miR-889、miR-543、およびmiR-1185-1)、または少なくとも(miR-889、miR-543、およびmiR-548k)、または少なくとも(miR-889、miR-543、およびmiR-12135)、または少なくとも(miR-889、miR-543、miR-376b、miR-335、およびmiR-1185-1)を含み、大腸癌および進行性腺腫を患う対象の血漿試料において有意に上方調節される。この意味では、表1に示すように、少なくとも(miR-889)もしくは少なくとも(miR-889およびmiR-543)もしくは少なくとも(miR-889およびmiR-376b)、または少なくとも(miR-889およびmiR-335)、または少なくとも(miR-889およびmiR-1185-1)、または少なくとも(miR-889、miR-543、およびmiR-376b)、または少なくとも(miR-889、miR-543、およびmiR-335)、または少なくとも(miR-889、miR-376b、およびmiR-335)、または少なくとも(miR-889、miR-543、およびmiR-1185-1)、または少なくとも(miR-889、miR-543、およびmiR-548k)、または少なくとも(miR-889、miR-543、およびmiR-12135)、または少なくとも(miR-889、miR-543、miR-376b、miR-335、およびmiR-1185-1)の組合せである。

30

40

【0186】

いくつかの実施形態では、本開示の方法は、(例えば、結腸増殖性障害を患うことに基づき)処置を必要とする対象に処置を施す工程を含む場合がある。大腸腺腫は大腸癌の前駆物質として見ることで見ることができ、確認された腺腫-癌腫の配列、および進行型大腸腺腫が癌に移行する可能性が高いという概念により、大腸腺腫(例えば、進行型大腸腺腫)は、例えば大腸内視鏡検査を通じて除去されることにより処置される場合がある(その後

50

、監視が行われる場合がある）。大腸癌の治療は、癌を発見した段階に左右される。早期大腸癌は、手術で処置される場合がある。ステージⅠ大腸癌の約95%、ステージⅡ大腸癌の65～80%が、手術により治癒することができる。しかし、直腸癌は、再発リスクを最小限に抑えるためにさらなる放射線療法を必要とする場合がある。進行期（ステージⅢとステージⅣ）での処置は、多くの場合、手術、化学療法、抗体による処置、抗VEGF／R、および放射線療法を含む治療法の組合せを含む。大腸癌を患う対象に対する処置は、例えば、その全体を参照することで本明細書に引用される、Wolpinらによる「Systemic Treatment of Colorectal Cancer」, Gastroenterology, Volume 134, Issue 5, 2008, Pages 1296–1310. e1, ISSN 0016–5085により記載され得る。

10

【0187】

処置は、対象のmiRNAプロファイルおよび／または対象の一組の生体形質に少なくとも部分的に基づき（例えば、複数の可能な処置選択肢の中から）選択され、施される場合がある。生体形質は、測定、診断、予後、または予測（例えば、訓練済み機械学習分類器を使用して判定される）であってよい。

【0188】

いくつかの実施形態では、生体形質は悪性腫瘍を含む。いくつかの実施形態では、生体形質は癌の種類を含む。いくつかの実施形態では、生体形質は癌のステージを含む。いくつかの実施形態では、生体形質は癌の分類を含む。いくつかの実施形態では、癌の分類は癌のグレードを含む。いくつかの実施形態では、癌の分類は組織学的分類を含む。いくつかの実施形態では、生体形質は代謝プロファイルを含む。いくつかの実施形態では、生体形質は突然変異を含む。いくつかの実施形態では、突然変異は疾患関連型突然変異である。いくつかの実施形態では、生体形質は臨床アウトカムを含む。いくつかの実施形態では、生体形質は薬物応答を含む。

20

【0189】

V. 分類器、機械学習モデル、およびシステム

いくつかの例では、miRNAシーケンシング機能は、配列組成と対象群（例えば、患者群）との相関性を見出すように訓練されたアルゴリズム（例えば、機械学習モデルまたは分類器）へ入力されるデータセットとして使用される。このような患者群の例は、疾患または疾病の有無、疾患または疾病の高リスクまたは低（non-elevated）リスク、疾患または疾病のステージ、疾患または疾病のサブタイプ、処置に対する応答と処置に対する不応答、および進行と非進行を含む。いくつかの例では、既知の条件または特徴を具備する対象から得た試料を比較するために、特徴マトリクスが生成される。いくつかの実施形態では、試料は、健康な対象、または明らかな適応症をいずれも患っていない対象、および癌を患っていることが明らかな患者から得られる。

30

【0190】

本明細書で使用する時、機械学習およびパターン認識に関する場合、「特徴（feature）」という用語は、一般に、ある減少における個々の測定可能な特性または特徴が観察されることを表す。「特徴」の概念は、例えば、限定されないが線形回帰およびロジスティック回帰などの統計的技法に使用される説明変数の概念に関連する。特徴は、数値または定言的であってよい（例えば、文字列やグラフなどの構造的特徴が構文パターン認識に使用される）。

40

【0191】

本明細書で使用する時、「入力特徴」（または「特徴」）という用語は、一般に、訓練済みアルゴリズム（例えば、機械学習モデルまたは分類器）が試料の出力分類（標識）を予測するのに使用される変数、例えば、状態、配列の内容（例えば、突然変異）、示唆されたデータ収集演算、または示唆された処置を表す。変数の値は、試料に対して判定されるとともに、分類を判定するために使用されてよい。

【0192】

50

複数のアッセイにおいて、上記システムは、訓練済みアルゴリズム（例えば、機械学習モデルまたは分類器）へと入力を行うように設定される。システムは、各生体試料に対しアッセイを行い、測定値から特徴ベクトルを形成する。システムは、特徴ベクトルを機械学習モデルに入力し、生体試料が特定の特性を有するかどうかの出力分類を取得する。

【0193】

いくつかの実施形態では、機械学習モデルは、対象集団またはこの集団の特徴における対象または特徴の2つ以上の群または分類を判別可能な分類器を出力する。いくつかの実施形態では、分類器は、訓練済み機械学習分類器である。

【0194】

いくつかの実施形態では、癌組織中のバイオマーカの有益な座位または特徴は、プロファイル形成するためにアッセイされる。2つの集団（例えば、治療剤に应答する対象と应答しない対象）を判別する際に特定の特性（例えば、本明細書に記載のバイオマーカのいずれか、および／または追加のバイオメディカル情報のいずれかの項目）のパフォーマンスをプロットすることにより、受信者操作特性（ROC）曲線が生成されてよい。いくつかの実施形態では、集団全体にわたる特徴データ（例えば、症例と対照）は、単一の特徴の値に基づき昇順に選別される。

【0195】

いくつかの例では、特定の特性は、健康 v s 癌、疾患の高リスク v s 低リスク、疾患のサブタイプ、疾患のステージ、進行 v s 非進行、および应答 v s 不应答から選択される。

【0196】

いくつかの実施形態では、結腸細胞増殖性障害は、腺腫（腺腫性ポリープ）、固着性鋸歯状腺腫（SSA）、進行型腺腫、大腸異形成、大腸腺腫、大腸癌、結腸癌、直腸癌、大腸癌腫、大腸腺癌、カルチノイド腫瘍、胃腸カルチノイド腫瘍、胃腸間質腫瘍（GIST）、リンパ腫、および肉腫からなる群から選択される。

【0197】

A. データ解析

いくつかの例では、本開示は、ソフトウェアアプリケーション、コンピューティングハードウェア、またはその両方においてデータ解析が実現されるシステム、方法、またはキットを提供する。いくつかの例では、解析アプリケーションまたはシステムは、少なくともデータ受信モジュール、データ前処理モジュール、データ解析モジュール（1または複数の種類のゲノムデータ上で動作可能なもの）、データ解釈モジュール、またはデータ可視化モジュールを含む。いくつかの実施形態では、データ受信モジュールは、室内実験ハードウェアまたは計測手段を、室内実験データを処理するコンピュータシステムと接続するコンピュータシステムを含むことができる。いくつかの実施形態では、データ前処理モジュールは、解析の準備のためにデータに対する演算を実行するハードウェアシステムまたはコンピュータソフトウェアを含むことができる。前処理モジュール中のデータに適用され得る演算の例として、アフィン変換、ノイズ除去演算、データクリーニング、再フォーマット、またはサブサンプリングが挙げられる。1または複数のゲノム材料からゲノムデータを解析するのに特化し得るデータ解析モジュールは、例えば、組み立てられたゲノム配列を取得し、確率および統計解析を実行して、疾患、病理学、状況、リスク、状態、または表現型に関連する異常パターンを特定することができる。データ解釈モジュールは、特定された異常パターンと健康状態、機能状況、予後、またはリスクとの関連性の理解を支援するのに、例えば統計、数学、または生物学から抽出される解析法を使用することができる。データ可視化モジュールは、結果の理解または解釈を容易にすることができるデータの視覚表現を作り出すために、数学的モデリング、コンピュータグラフィック、またはレンダリングの方法を使用することができる。

【0198】

いくつかの例では、機械学習法は、試料集団中の試料を判別するために適用される。いくつかの実施形態では、機械学習法は、健康な試料と疾患（例えば、腺腫）進行した試料との間で試料を判別するために適用される。

10

20

30

40

50

【0199】

いくつかの実施形態では、予測エンジンを訓練するのに使用される1または複数の機械学習演算として、一般化線形モデル、一般化加法モデル、ノンパラメトリック回帰演算、ランダムフォレスト 分類子、空間回帰演算、ベイズ回帰モデル、時系列解析、ベイズネットワーク、ガウスネットワーク、決定木学習演算、人工ニューラルネットワーク、リカレントニューラルネットワーク、強化学習演算、線形または非線形回帰演算、サポート・ベクター・マシン、クラスタリング演算、および遺伝的アルゴリズム演算のうち、1または複数が挙げられる。

【0200】

いくつかの例では、コンピュータ処理方法は、ロジスティック回帰、多重線形回帰（MLR）、次元削減、部分最小二乗（PLS）回帰、主成分回帰、オートエンコーダ、変分オートエンコーダ、特異値分解、フーリエ基底、ウェーブレット、判別解析、サポート・ベクター・マシン、決定木、分類および回帰木（CART）、ツリーベースの方法、ランダムフォレスト、勾配ブーストツリー、ロジスティック回帰、行列分解、多次元スケールリング（MDS）、次元性削減法、 t -分散確率的近傍埋め込み（ t -SNE）、多層パーセプトロン（MLP）、ネットワーククラスタリング、ニューロファジィ、および人工ニューラルネットワークから選択される。

【0201】

いくつかの例では、本明細書に開示される方法は、対象または複数の対象からの試料の核酸シーケンシングデータに関する計算解析を含むことができる。

【0202】

B. 分類子の生成

一態様では、開示されたシステムと方法は、cfRNAの生体試料からのmiRNA配列分析から導出した特徴情報に基づき生成された分類子を提供する。分類子は、cfDNAなどの生体試料中で特定された配列特徴に基づき、集団内の群を判別するための予測エンジンの一部を形成する。

【0203】

いくつかの実施形態では、分類子は、配列情報の類似部分を統一されたフォーマットおよび統一されたスケールへとフォーマット化することで配列情報を正規化し、正規化配列情報を柱状データベースに記憶し、記憶された正規化配列情報に1または複数の機械学習演算を適用することで、特定の集団に対して1または複数の特徴の組み合わせをマッピングする予測エンジンを訓練し、アクセスされたフィールド情報に予測エンジンを適用して群に関連付けられた対象を特定し、対象を群に分類することにより、作成される。

【0204】

特異性とは、本明細書で使用するとき、概して「疾患のない人々に対する検査が陰性の確率」を表す。これは、検査が陰性であった疾患のない人の数を、疾患のない対象の総数で割ることにより算出され得る。

【0205】

いくつかの例では、モデル、分類子、または予測検査の特異性は、少なくとも40%、少なくとも45%、少なくとも50%、少なくとも55%、少なくとも60%、少なくとも65%、少なくとも70%、少なくとも75%、少なくとも80%、少なくとも85%、少なくとも90%、少なくとも95%、または少なくとも99%である。

【0206】

感度とは、本明細書で使用するとき、概して「疾患を患う人々に対する検査が陽性の確率」を表す。これは、検査が陽性であった疾患を患う人の数を、疾患を患う対象の総数で割ることにより算出され得る。

【0207】

いくつかの例では、モデル、分類子、または予測検査の感度は、少なくとも40%、少なくとも45%、少なくとも50%、少なくとも55%、少なくとも60%、少なくとも65%、少なくとも70%、少なくとも75%、少なくとも80%、少なくとも85%、

少なくとも90%、少なくとも95%、または少なくとも99%である。

【0208】

C. デジタル処理デバイス

いくつかの例では、本明細書に記載される主題は、デジタル処理デバイスまたはその使用を含むことができる。いくつかの例では、デジタル処理デバイスは、デバイスの機能を実行する1または複数のハードウェア中央処理ユニット(CPU)、グラフィック処理ユニット(GPU)、あるいはテンソル・プロセッシング・ユニット(TPU)を含むことができる。いくつかの例では、デジタル処理デバイスは、実行可能命令を実行するように構成されたオペレーティングシステムを含むことができる。

【0209】

いくつかの例では、デジタル処理デバイスは、任意選択でコンピュータネットワークに接続することができる。いくつかの例では、デジタル処理デバイスは、任意選択でインターネットに接続されてもよい。いくつかの例では、デジタル処理デバイスは、任意選択でクラウド・コンピューティング・インフラストラクチャに接続されてもよい。いくつかの例では、デジタル処理デバイスは、任意選択でイントラネットに接続されてもよい。いくつかの例では、デジタル処理デバイスは、任意選択でデータ記憶デバイスに接続されてもよい。

【0210】

適切なデジタル処理デバイスの非限定的な例として、サーバコンピュータ、デスクトップコンピュータ、ラップトップコンピュータ、ノートブックコンピュータ、サブノートブックコンピュータ、ネットブックコンピュータ、ネットパッドコンピュータ、セットトップコンピュータ、ハンドヘルドコンピュータ、インターネットアプライアンス、モバイルスマートフォン、およびタブレットコンピュータが挙げられる。適切なタブレットコンピュータとして、例えば、ブックレット、スレート、および転換可能な構成を具備するものを挙げることができる。

【0211】

いくつかの例では、デジタル処理デバイスは、実行可能命令を実行するように構成されたオペレーティングシステムを含むことができる。例えば、オペレーティングシステムは、デバイスのハードウェアを管理し、アプリケーションの実行のためのサービスを提供するプログラムおよびデータを含む、ソフトウェアを含むことができる。オペレーティングシステムの非限定的な例として、Ubuntu、FreeBSD、OpenBSD、NetBSD(登録商標)、Linux(登録商標)、Apple(登録商標)Mac OS X Server(登録商標)、Oracle(登録商標)Solaris(登録商標)、Windows Server(登録商標)、およびNovell(登録商標)NetWare(登録商標)が挙げられる。適切なパーソナルコンピュータのオペレーティングシステムの非限定的な例として、Microsoft(登録商標)Windows(登録商標)、Apple(登録商標)Mac OS X(登録商標)、UNIX(登録商標)、および、GNU/Linux(登録商標)などのUNIX(登録商標)様オペレーティングシステムが挙げられる。いくつかの例では、オペレーティングシステムは、クラウドコンピューティングにより提供されてもよく、クラウド・コンピューティング・リソースは1または複数のサービスプロバイダにより提供されてもよい。

【0212】

いくつかの例では、デバイスは、記憶デバイスおよび／またはメモリデバイスを含むことができる。記憶デバイスおよび／またはメモリデバイスは、一時的または永続的にデータまたはプログラムを記憶するのに使用される1または複数の物理装置であってよい。いくつかの例では、デバイスは揮発性メモリであり、記憶した情報を維持するのに電力を要する場合がある。いくつかの例では、デバイスは不揮発性メモリであり、デジタル処理デバイスに電力が供給されていないときでも記憶情報を保持する場合がある。いくつかの例では、不揮発性メモリはフラッシュメモリを含むことができる。いくつかの例では、不揮発性メモリは、ダイナミック・ランダム・アクセス・メモリ(DRAM)を含むことがで

10

20

30

40

50

きる。いくつかの例では、不揮発性メモリは強誘電性ランダム・アクセス・メモリ（F R A M（登録商標））を含むことができる。いくつかの例では、不揮発性メモリは、相変化ランダム・アクセス・メモリ（P R A M）を含むことができる。

【0213】

いくつかの例では、デバイスは、例えば、C D - R O M、D V D、フラッシュメモリデバイス、磁気ディスクドライブ、磁気テープドライブ、光ディスクドライブ、およびクラウドコンピューティングベースのストレージを含む記憶デバイスであってよい。いくつかの例では、記憶デバイスおよび／またはメモリデバイスは、本明細書に開示されるものなどのデバイスの組合せであってよい。いくつかの例では、デジタル処理デバイスは、視覚情報をユーザに送信するためのディスプレイを含むことができる。いくつかの例では、ディスプレイは陰極線管（C R T）であってよい。いくつかの例では、ディスプレイは液晶ディスプレイ（L C D）であってよい。いくつかの例では、ディスプレイは、薄膜トランジスタ液晶ディスプレイ（T F T - L C D）であってよい。いくつかの例では、ディスプレイは、有機発光ダイオード（O L E D）ディスプレイであってよい。いくつかの例では、O L E Dディスプレイは、パッシブマトリクスO L E D（P M O L E D）またはアクティブマトリクスO L E D（A M O L E D）ディスプレイであってよい。いくつかの例では、ディスプレイはプラズマディスプレイであってよい。いくつかの例では、ディスプレイはビデオプロジェクトアである。いくつかの例では、ディスプレイは、本明細書に開示されるものなどのデバイスの組合せであってよい。

10

20

【0214】

いくつかの例では、デジタル処理デバイスは、ユーザから情報を受信するための入力デバイスを含むことができる。いくつかの例では、入力デバイスはキーボードであってよい。いくつかの例では、入力デバイスは、例えば、マウス、トラックボール、トラックパッド、ジョイスティック、ゲームコントローラ、またはスタイラスを含むポインティングデバイスであってよい。いくつかの例では、入力デバイスは、タッチスクリーンまたはマルチタッチスクリーンであってよい。いくつかの例では、入力デバイスは、音声または他の音声入力を捕えるためのマイクであってよい。いくつかの例では、入力デバイスは、動作または視覚入力を捕えるためのビデオカメラであってよい。いくつかの例では、入力デバイスは、本明細書に開示されるものなどのデバイスの組合せであってよい。

30

【0215】

D. 非一時的コンピュータ可読記憶媒体

いくつかの例では、本明細書に開示される主題は、任意選択でネットワーク化されたデジタル処理デバイスのオペレーティングシステムにより実行可能な命令を含むプログラムでコードされる、1または複数の非一時的コンピュータ可読記憶媒体を含むことができる。いくつかの例では、コンピュータ可読記憶媒体は、デジタル処理デバイスの有形成コンポーネントであってよい。いくつかの例では、コンピュータ可読記憶媒体は、任意選択でデジタル処理デバイスから取り外し可能であってよい。いくつかの例では、コンピュータ可読記憶媒体として、例えば、C D - R O M、D V D、フラッシュメモリデバイス、固体メモリ、磁気ディスクドライブ、磁気テープドライブ、光ディスクドライブ、クラウド・コンピューティング・システムおよびサービスなどを挙げることができる。いくつかの例では、プログラムと命令は、永久的に、実質的に永久的に、半永久的に、または非一時的に媒体上でコードされる場合がある。

40

【0216】

E. コンピュータシステム

本開示は、本明細書に記載の方法を実施するようにプログラムされたコンピュータシステムを提供する。図1は、患者データ、生体データ、生体配列、および参照配列を記憶、処理、特定、または解釈するようにプログラムあるいは他の方法で構成されるコンピュータシステム（101）を示す。コンピュータシステム（101）は、本開示の患者データ、生体データ、生体配列、または参照配列の様々な態様を処理することができる。コンピュータシステム（101）は、ユーザの電子デバイス、またはその電子デバイスに対して

50

遠隔に位置するコンピュータシステムであってよい。電子デバイスは、モバイル電子デバイスであってもよい。

【0217】

コンピュータシステム(101)は、中央処理ユニット(CPU、本明細書では「プロセッサ」および「コンピュータプロセッサ」とも呼ばれる)(105)を含み、この中央処理ユニットは、シングルコアもしくはマルチコアプロセッサ、または並列のための複数のプロセッサであってもよい。コンピュータシステム(101)はまた、メモリもしくはメモリ位置(110)(例えば、ランダム・アクセス・メモリ、読取り専用メモリ、フラッシュメモリ)、電子記憶ユニット(115)(例えば、ハードディスク)、1または複数の他のシステムと通信するための通信インターフェース(120)(例えば、ネットワークアダプタ)、および、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などの周辺デバイス(125)も含む。メモリ(110)、記憶ユニット(115)、インターフェース(120)、および周辺デバイス(125)は、マザーボードなどの通信バス(実線)を介してCPU(105)と通信状態にある。記憶ユニット(115)は、データを記憶するためのデータ記憶ユニット(すなわちデータリポジトリ)であってもよい。コンピュータシステム(101)は、通信インターフェース(120)の補助により、コンピュータネットワーク(「ネットワーク」)(130)に動作可能に結合される場合がある。ネットワーク(130)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と通信状態にあるイントラネットおよび／もしくはエクストラネットであってもよい。いくつかの例では、ネットワーク(130)は、電気通信および／またはデータネットワークである。ネットワーク(130)は、クラウドコンピューティングなどの分散コンピューティングを可能にすることができる1または複数のコンピュータサーバを含むことができる。いくつかの例では、ネットワーク(130)は、コンピュータシステム(101)の補助により、ピア・ツー・ピアネットワークを実装することができ、これにより、コンピュータシステム(101)に結合されたデバイスがクライアントまたはサーバとして挙動することが可能になる場合がある。

10

20

【0218】

CPU(105)は、プログラムまたはソフトウェアに組み込まれ得る一連の機械可読命令を実行することができる。命令は、メモリ(110)などのメモリ位置に記憶されてよい。命令はCPU(105)に向けられてよく、その後、CPU(105)を、本開示の方法を実装するようにプログラムまたはその他の方法で構成することができる。CPU(105)により実行される演算の例として、フェッチ、デコード、実行、およびライトバックを挙げることができる。

30

【0219】

CPU(105)は、集積回路などの回路の一部であってよい。システム(101)の1または複数の他のコンポーネントが回路に含まれてもよい。いくつかの例では、回路は特定用途集積回路(ASIC)である。

【0220】

記憶ユニット(115)は、ドライバ、ライブラリ、および保存プログラムなどのファイルを記憶することができる。記憶ユニット(115)は、ユーザデータ、例えば、ユーザ嗜好およびユーザプログラムを記憶することができる。いくつかの例では、コンピュータシステム(101)は、イントラネットまたはインターネットを介してコンピュータシステム(101)と通信状態にあるリモートサーバ上に位置するなど、コンピュータシステム(101)の外部にある1または複数の追加のデータ記憶ユニットを含むことができる。

40

【0221】

コンピュータシステム(101)は、ネットワーク(130)を介して1または複数の遠隔コンピュータシステムと通信することができる。例えば、コンピュータシステム(101)は、ユーザの遠隔コンピュータシステムと通信することができる。遠隔コンピュー

50

タシステムの例として、パーソナルコンピュータ（例えば、携帯用 P C）、スレートもしくはタブレット P C（例えば、A p p l e（登録商標）i P a d（登録商標）、S a m u s n（登録商標）G a l a x y T a b）、電話、スマートフォン（例えば、A p p l e（登録商標）i P h o n e（登録商標）、A n d r o i d - e n a b l e d d e v i c e、B l a c k b e r y（登録商標））、またはパーソナル・デジタル・アシスタントが挙げられる。ユーザは、ネットワーク（130）を介してコンピュータシステム（101）にアクセスすることができる。

【0222】

本明細書に記載される方法は、例えば、メモリ（110）または電子記憶ユニット（115）上など、コンピュータシステム（101）の電子記憶位置に記憶される機械（例えば、コンピュータプロセッサ）実行可能コードにより実施される場合がある。機械実行可能コードまたは機械可読コードは、ソフトウェアの形で提供されてよい。使用中、コードはプロセッサ（105）により実行されてよい。いくつかの例では、コードは、記憶ユニット（115）から検索されて、プロセッサ（105）によるアクセス準備のためにメモリ（110）に記憶されてもよい。いくつかの例では、電子記憶ユニット（115）は排除されてもよく、機械実行可能命令はメモリ（110）に記憶される。

10

【0223】

コードは、予めコンパイルされ、コードを実行するように適合されたプロセッサを有するマシンとともに使用するために構成されるか、または実行中に解釈またはコンパイルされてもよい。コードは、予めコンパイルされるか、解釈されるか、またはコンパイルされたままの形でコードの実行を可能にするように選択され得るプログラミング言語で供給されてもよい。

20

【0224】

コンピュータシステム（101）など、本明細書で提供されるシステムと方法の態様は、プログラミングにおいて具体化される場合がある。この技術の様々な態様は、典型的には機械（またはプロセッサ）実行可能コード、および／または1種の機械可読媒体に担持されるか具体化される関連データの形にある、「製品」または「製造品」と考慮される場合がある。機械実行可能コードは、メモリ（例えば、読取り専用メモリ、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などの電子記憶ユニットに記憶されてよい。「記憶」型媒体として、コンピュータやプロセッサなどの有形メモリ、または、ソフトウェアプログラミングにおいてあらゆる時間に非一時的な記憶を提供し得るそれらの関連モジュール、例えば様々なセミコンダクタメモリやテープドライブのうちいずれかまたは両方を挙げることができる。ソフトウェアのすべてまたは一部は、インターネットまたはその他様々な電気通信ネットワークを介して通信されることがある。このような通信は、例えば、1つのコンピュータまたはプロセッサから別のコンピュータへの、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フォームへのソフトウェアのローディングを可能にする場合がある。このため、ソフトウェア要素を有し得る別の種類の媒体は、有線および光回線ネットワークを介して、および様々なエアリンク上で、ローカルデバイス間の物理インターフェース全体において使用されるなどの、光波、電波、および電磁波を含む。有線または無線リンク、光リンクなどの波を運ぶ物理要素はまた、ソフトウェアを有する媒体とみなされる場合がある。本明細書で使用するとき、非一時的で有形の「記憶」媒体に限定されない限り、コンピュータまたは機械「可読媒体」などの用語は、実行のためにプロセッサに命令を送るのに関与するあらゆる媒体を表す。

30

40

【0225】

したがって、コンピュータ実行可能コードなどの機械可読媒体は、有形記憶媒体、搬送波媒体、または物理伝送媒体を含むがこれらに限定されない、多くの形態を呈する場合がある。不揮発性記憶媒体として、例えば、図に示されるデータベースを実装するために使用され得るなどの、あらゆるコンピュータ等における記憶デバイスのいずれかなどの光ディスクまたは磁気ディスクが挙げられる。揮発性記憶媒体として、このようなコンピュー

50

タブレットフォームのメインメモリなどの動的メモリが挙げられる。有形伝送媒体として、コンピュータシステム内のバスを備えるワイヤを含む、同軸ケーブル、銅線および光ファイバが挙げられる。搬送波伝送媒体は、電気信号もしくは電磁信号、または無線周波数（RF）および赤外線（IR）データの通信中に生成されるものなどの音響波または光波の形態を呈する場合がある。そのため、コンピュータ可読媒体によく見られる形態として、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、あらゆる他の磁気媒体、CD-ROM、DVD、もしくはDVD-ROM、あらゆる他の光学媒体、パンチカード紙テープ、穴のパターンを有する他のあらゆる物理記憶媒体、RAM、ROM、PROM、およびEPROM、FLASH（登録商標）-EPROM、あらゆる他のメモリチップもしくはカートリッジ、データもしくは命令を運ぶ搬送波、このような搬送波を運ぶケーブルもしくはリンク、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他のあらゆる媒体が挙げられる。コンピュータ可読媒体におけるこれら形態の多くは、実行のために1または複数の命令の1または複数のシーケンスをプロセッサに運ぶことに関与する場合がある。

10

20

30

40

50

【0226】

コンピュータシステム（101）は、例えば、核酸配列、濃縮核酸試料、miRNAプロファイル、発現プロファイル、およびRNA発現プロファイルの解析を提供するためのユーザインターフェース（UI）（140）を含む電子ディスプレイ（135）を含むか、またはそれと通信状態とすることができる。UIの例として、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が挙げられるが、これらに限定されない。

【0227】

本開示の方法とシステムは、1または複数のアルゴリズムにより実装される場合がある。アルゴリズムは、中央処理ユニット（105）による実行に際してソフトウェアを介して実施されてもよい。このアルゴリズムは、例えば、患者データ、生体データ、生体配列、および参照配列を記憶、処理、特定、または解釈することができる。

【0228】

いくつかの例では、本明細書に開示される主題は、少なくとも1つのコンピュータプログラムまたはその使用を含むことができる。コンピュータプログラムは、指定されたタスクを実行するために書き込まれる、デジタル処理デバイスのCPU、GPU、またはTPU内で実行可能な一連の命令とすることができる。コンピュータ可読命令は、特定のタスクを実行するか、または特定の抽象データ型を実施する、関数、オブジェクト、アプリケーション・プログラミング・インターフェース（API）、データ構造などのプログラムモジュールとして実施されてもよい。本明細書で提供される開示に照らして、コンピュータプログラムは、様々なバージョンの様々な言語で書かれてよい。

【0229】

コンピュータ可読命令の機能性は、様々な環境で所望されるように組み合わせられるか、または分散される場合がある。いくつかの例では、コンピュータプログラムは、1つの命令シーケンスを含むことができる。いくつかの例では、コンピュータプログラムは、複数の命令シーケンスを含むことができる。いくつかの例では、コンピュータプログラムは、1つの位置から提供されてよい。いくつかの例では、コンピュータプログラムは、複数の位置から提供されてよい。いくつかの例では、コンピュータプログラムは、1または複数のモジュールを含むことができる。いくつかの例では、コンピュータプログラムは、部分的または全体的に、1または複数のウェブアプリケーション、1または複数のモバイルアプリケーション、1または複数のスタンドアロンアプリケーション、1または複数のウェブ・ブラウザ・プラグイン、エクステンション、アドイン、またはアドオン、あるいはそれらの組合せを含むことができる。

【0230】

いくつかの例では、コンピュータ処理は、統計、数学、生物学、またはそれらのあらゆる組合せの方法であってよい。いくつかの例では、コンピュータ処理方法は、例えば、口

ジスティック回帰、次元削減、主成分解析、オートエンコーダ、特異値分解、フーリエ基底、特異値分解、ウェーブレット、判別解析、サポート・ベクター・マシン、ツリーベース方法、ランダムフォレスト、勾配ブーストツリー、ロジスティック回帰、行列分解、ネットワーククラスタリング、およびニューラルネットワークを含む、次元削減法を含む。

【0231】

いくつかの例では、コンピュータ処理方法は、例えば、回帰、サポート・ベクター・マシン、ツリーベース方法、およびネットワークを含む教師あり機械学習方法である。

【0232】

いくつかの例では、コンピュータ処理方法は、例えば、クラスタリング、ネットワーク、主成分解析、および行列分解を含む教師なし機械学習方法である。

【0233】

F. データベース

いくつかの例では、本明細書に開示される主題は、患者データ、生体データ、生体配列、または参照配列を記憶するための1または複数のデータベース、またはその使用を含むことができる。参照配列は、データベースから導出される場合がある。本明細書で提供される開示に照らすと、多くのデータベースが、配列情報の記憶と検索に適し得る。いくつかの例では、適切なデータベースとして、例えば、リレーショナルデータベース、非リレーショナルデータベース、オブジェクト指向データベース、オブジェクトデータベース、実体関係モデルデータベース (entity-relationship model database)、連想データベース、およびXMLデータベースを挙げることができる。いくつかの例では、データベースは、インターネットベースであってよい。いくつかの例では、データベースは、ウェブベースであってよい。いくつかの例では、データベースは、クラウドコンピューティングベースであってよい。いくつかの例では、データベースは、1または複数のローカルコンピュータ記憶デバイスに基づく場合がある。

【0234】

一態様では、本開示は、本明細書に開示される方法を実行するようにプロセッサを指示する命令を含む非一時的コンピュータ可読媒体を提供する。

【0235】

一態様では、本開示は、コンピュータ可読媒体を備えるコンピュータデバイスを提供する。

【0236】

別の態様では、本開示は、生体試料の分類を実行するためのシステムであって、

a) 複数の訓練試料であって、それぞれが分子の複数の分類を有しており、1または複数の既知のラベルを含む、複数の訓練試料を受けるためのリザーバと、

b) 複数の訓練試料のそれぞれにおいて機械学習モデルに入力されるように操作可能なアッセイに相当する一組の特徴を特定するための特徴モジュールであって、一組の特徴は複数の訓練データ中の分子の特性に相当し、複数の訓練データのそれぞれにおいて、システムは複数組の測定値を得るために複数の異なるアッセイに訓練試料中の分子の複数の分類を晒すように操作可能であり、各組の測定値は、1つのアッセイから訓練試料中の分子の分類に適用され、複数組の測定値は複数の訓練試料において得られる、特徴モジュールと、

c) 訓練試料の訓練ベクトルを得るために複数組の測定値を解析するための解析モジュールであって、訓練ベクトルは対応するアッセイのN組の特徴の特徴値を含み、各特徴値は1つの特徴に相当するとともに1または複数の測定値を含み、訓練ベクトルは、複数の異なるアッセイの第1のサブセットに相当するN組の特徴のうち少なくとも2つから少なくとも1つの特徴を用いて形成される、解析モジュールと、

d) 複数の訓練試料の出力ラベルを得るために機械学習モデルのパラメータを用いて訓練ベクトル上でシステムに通知を行うラベリングモジュールと、

e) 出力ラベルと訓練試料の既知のラベルとを比較するコンパレータモジュールと、

f) 出力ラベルと訓練サンプルの既知のラベルとの比較に基づき、機械学習モデルの

10

20

30

40

50

訓練の一部としてパラメータの最適値を反復的に検索するための訓練モジュールと、

g) 機械学習モデルのパラメータおよび機械学習モデルの一組の特徴を提供する出力モジュールと

を含むシステムを提供する。

【0237】

VI. 集団中の対象を分類する方法

開示された方法は、対象に発現されたmiRNAの解析により、結腸細胞増殖性障害に関連するゲノムDNA発現のパラメータを確認することに関する。この方法は、結腸細胞増殖性障害の改善された診断、治療、およびモニタリングにおける、より具体的には、この障害のステージまたはサブクラスと障害に対する遺伝的素因との改善された特定およびそれらの判別を可能にすることによる使用のための方法である。

10

【0238】

いくつかの実施形態では、上記方法は、集団中の対象から得た生体試料中でのmiRNAの差動発現を解析する工程を含む。

【0239】

本開示は、例えば、結腸細胞増殖性障害を患う対象と患っていない対象との間で差動的に発現された無細胞miRNAを検出するために、無細胞試料に適用され得る結腸細胞増殖性障害を検出するための方法を提供する。この方法は、結腸細胞増殖性障害を患っていない健康な対象と比較して結腸細胞増殖性障害シグナルが基本的に「陽性」または「陰性」とする、miRNAの検出を利用する。

20

【0240】

いくつかの実施形態では、結腸細胞増殖性障害は、腺腫（腺腫性ポリープ）、固着性鋸歯状腺腫（SSA）、進行型腺腫、大腸異形成、大腸腺腫、大腸癌、結腸癌、直腸癌、大腸癌腫、大腸腺癌、カルチノイド腫瘍、胃腸カルチノイド腫瘍、胃腸間質腫瘍（GIST）、リンパ腫、および肉腫からなる群から選択される。

【0241】

ある態様では、本開示は、対象から得た生体試料のmiRNAプロファイルを判定するための方法であって、

a) 生体試料からRNAを単離する工程と、

b) RNAをcDNAに逆転写する前または後に、生体試料から得た核酸にRNAアダプタをライゲートする工程と、

30

c) 工程b)のcDNAを増幅する工程と、

d) cDNA分子の核酸配列を決定する工程と、

e) 表1～表11に列挙される群から選択されるmiRNAの予め特定されたパネルにおいて核酸分子の核酸配列を参照核酸配列にアライメントして、対象のmiRNAプロファイルを判定する工程と

を含む方法を提供する。

【0242】

いくつかの実施形態では、核酸シーケンシングライブラリは、増幅前に調製される。

【0243】

いくつかの実施形態では、アダプタのライゲーションは、上記cの前にRNAアダプタライゲーション、アダプタブロッキング、アダプタ環状化、および二量体除去を含む。

40

【0244】

いくつかの実施形態では、参照核酸配列は、ヒトゲノムまたはヒトトランスクリプトームのデータベースの一部である。

【0245】

いくつかの実施形態では、対象のmiRNAプロファイルを決定する工程は、発現されたmiRNAのカウントテーブルを生成することを含む。

【0246】

いくつかの実施形態では、対象のmiRNAプロファイルを決定する工程は、発現され

50

たm i R N Aのカウントテーブルを生成することで、差動的に豊富なm i R N Aを特定することを含む。

【0247】

いくつかの実施形態では、m i R N Aプロファイルは、結腸細胞増殖性障害と関連付けられ、かつ、結腸細胞増殖性障害があるとの対象の分類を提供する。

【0248】

いくつかの実施形態では、対象から得た生体試料は、体液、便、結腸流出液、尿、血漿、血清、全血、単離した血液細胞、血液から単離した細胞、およびそれらの組合せからなる群から選択される。

【0249】

いくつかの実施形態では、上記方法は、対象から測定したm i R N Aシグネチャパネルを、健康な対象から測定したm i R N Aシグネチャパネルのデータベースであって、コンピュータシステムに記憶されているデータベースに適用する工程、および、健康な対象から得たm i R N A状態と比較したときにm i R N Aシグネチャパネルに少なくとも5%、少なくとも10%、少なくとも15%、少なくとも20%、少なくとも25%、少なくとも30%、少なくとも35%、少なくとも40%、少なくとも45%、または少なくとも50%の変化を測定することにより、対象が結腸細胞増殖性障害を患うリスクが増加したと判定する工程を含む。

【0250】

いくつかの実施形態では、結腸細胞増殖性障害は、腺腫（腺腫性ポリープ）、固着性鋸歯状腺腫（SSA）、進行型腺腫、大腸異形成、大腸腺腫、大腸癌、結腸癌、直腸癌、大腸癌腫、大腸腺癌、カルチノイド腫瘍、胃腸カルチノイド腫瘍、胃腸間質腫瘍（GIST）、リンパ腫、および肉腫からなる群から選択される。

【0251】

いくつかの実施形態では、結腸細胞増殖性障害は、ステージ1大腸癌、ステージ2大腸癌、ステージ3大腸癌、およびステージ4大腸癌からなる群から選択される。

【0252】

別の態様では、本開示は、対象における結腸細胞増殖性障害を検出するための方法であって、

a) 生体試料からRNAを単離する工程と、

b) 生体試料から得たRNAにRNAアダプタをライゲートして、RNAをcDNAに逆転写する工程と、

c) 工程b)のcDNAを増幅する工程と、

d) cDNA分子の核酸配列を決定する工程と、

e) 表1～表11に列挙される群から選択されるm i R N Aの予め特定されたパネルの核酸配列をアライメントして、対象のm i R N Aプロファイルを判定する工程と、

f) 健康な対象と結腸細胞増殖性障害を患う対象とを判別可能となるよう訓練された機械学習モデルにm i R N Aプロファイルを入力することで、結腸細胞増殖性障害の存在に関連する出力値を提供し、それにより前記対象における結腸細胞増殖性障害の存在を示す工程と

を含む方法を提供する。

【0253】

いくつかの実施形態では、参照核酸配列は、ヒトゲノムまたはヒトトランスクリプトームのデータベースの一部である。

【0254】

いくつかの実施形態では、対象のm i R N Aプロファイルを決定する工程は、発現されたm i R N Aのカウントテーブルを生成することを含む。

【0255】

いくつかの実施形態では、対象のm i R N Aプロファイルを決定する工程は、発現されたm i R N Aのカウントテーブルを生成することで、差動的に豊富なm i R N Aを特定す

10

20

30

40

50

ることを含む。

【0256】

いくつかの実施形態では、miRNAプロファイルは、2つの対象群を区別可能な分類器（例えば、健康vs癌、疾患のステージ、進行腺腫vs癌）を取得するために機械学習モデルに入力される。

【0257】

いくつかの実施形態では、対象から得た生体試料は、体液、便、結腸流出液、尿、血漿、血清、全血、単離した血液細胞、血液から単離した細胞、およびそれらの組合せからなる群から選択される。

【0258】

いくつかの実施形態では、上記方法は、対象から測定したmiRNAシグネチャパネルを、健康な対象から測定したmiRNAシグネチャパネルのデータベースであって、コンピュータシステムに記憶されているデータベースに適用する工程、および、健康な対象から得たmiRNA状態と比較したときにmiRNAシグネチャパネルのmiRNA発現に少なくとも5%、少なくとも10%、少なくとも15%、少なくとも20%、少なくとも25%、少なくとも30%、少なくとも35%、少なくとも40%、少なくとも45%、または少なくとも50%の変化を測定することにより、対象が結腸細胞増殖性障害を患うリスクが増加したと判定する工程を含む。

【0259】

いくつかの実施形態では、結腸細胞増殖性障害は、腺腫（腺腫性ポリープ）、固着性鋸歯状腺腫（SSA）、進行型腺腫、大腸異形成、大腸腺腫、大腸癌、結腸癌、直腸癌、大腸癌腫、大腸腺癌、カルチノイド腫瘍、胃腸カルチノイド腫瘍、胃腸間質腫瘍（GIST）、リンパ腫、および肉腫からなる群から選択される。

【0260】

いくつかの実施形態では、結腸細胞増殖性障害は、ステージ1大腸癌、ステージ2大腸癌、ステージ3大腸癌、およびステージ4大腸癌からなる群から選択される。

【0261】

本明細書に記載される訓練された機械学習方法、モデル、および識別分類器は、癌の検出、診断、および処置応答性を含む様々な医療用途に適用される場合がある。モデルは対象メタデータおよび分析物由来の特徴を用いて訓練されてもよい。アプリケーションは、集団内の対象を層別化して、適宜処置の決定をガイドするように調整される場合がある。

【0262】

診断

本明細書で提供される方法とシステムは、人工知能を用いた手法を使用して予測解析を行い、対象（患者）から取得したデータを解析して、癌（例えば、大腸癌、CRC）を患う対象の診断の出力を生成する場合がある。例えば、アプリケーションは、取得したデータに予測アルゴリズムを適用して、癌を患う対象の診断を生成することができる。予測アルゴリズムは、取得したデータを処理して癌を患う対象の診断を生成するように構成された、機械学習ベース予測器などの人工知能ベース予測器を含んでよい。

【0263】

機械学習予測器は、データセット、例えば、対象の生体試料上で、本明細書に記載のシグネチャパネルを用いて、機械学習予測器への入力として癌を患った患者の1または複数の組のコホート、および機械学習予測器への出力として対象の明らかな診断（例えば、病期分類、および／または腫瘍画分）アウトカムからRNAアッセイを行うことにより生成されたデータセットを用いて訓練される場合がある。

【0264】

訓練データセット（例えば、対象の生体試料上で、本明細書に記載のシグネチャパネルを用いてメチル化アッセイを行うことにより生成されたデータセット）は、例えば、よく見られる特性（特徴）およびアウトカム（ラベル）を有する対象の1または複数の組から

10

20

30

40

50

生成される場合がある。訓練データセットは、診断に関連する特徴に対応する一組の特徴とラベルを含んでもよい。特徴は、例えば、c f R N A アッセイ測定の特定の範囲またはカテゴリなどの特性、例えば、基準ゲノムの一組のビン（ゲノムウィンドウ）のそれぞれの中で重なるか、またはその中にある、健康な試料および疾患試料から得た生体試料中の c f R N A またはそのフラグメントの数を含む場合がある。例えば、所与の時点で所与の対象から採取した一組の特徴は、総体的に所与の時点での特定された対象の癌を示し得る診断シグネチャとして機能する場合がある。特性はさらに、1または複数の癌などの対象の診断結果を示すラベルを含む場合もある。

【0265】

ラベルは、例えば、対象の明らかな診断（例えば、病期分類および／または腫瘍画分）アウトカムなどのアウトカムを含む場合がある。アウトカムは、対象の癌に関連付けられる特性を含む場合がある。例えば、特性は、1または複数の癌を患う対象を示す場合がある。

10

【0266】

訓練セット（例えば、訓練データセット）は、1または複数組の対象（例えば、1または複数の癌を患うか、または患っていない患者の後向きおよび／または前向きコホート）に相当する一組のデータに対するランダムなサンプリングにより選択されてもよい。代替的に、訓練セット（例えば、訓練データセット）は、1または複数組の対象（例えば、1または複数の癌を患うか、または患っていない患者の後向きおよび／または前向きコホート）に相当する一組のデータに対する比例サンプリングにより選択されてもよい。訓練セットは、1または複数組の対象（例えば、異なる臨床施設または検査の患者）に対応するデータセット全体で平衡化される場合がある。機械学習予測器は、診断精度測定値に相当する最小の所望の値を有するなど、精度または性能に関する特定の予め定めた条件が満たされるまで訓練される場合がある。例えば、診断精度測定は、対象における1または複数の癌の診断、病期分類、または腫瘍画分の予測に相当し得る。

20

【0267】

診断精度測定の例として、癌（例えば、大腸癌）を検出または予測する診断精度に相当する受信者操作特性（R O C）曲線の感度、特異性、陽性予測値（P P V）、陰性予測値（N P V）、精度、および曲線下面積（A U C）が挙げられる場合がある。

【0268】

30

一態様では、本開示は、対象の集団を判別可能な分類器を使用する方法であって、

a) 生体試料中の R N A をアッセイすることにより、生体試料中の R N A を表す一組の測定値を提供する工程と、

b) 機械学習または統計モデルに入力する生体試料中の R N A の特性に相当する一組の特徴を特定する工程と、

c) 複数組の測定値それぞれから、特徴値であって、それぞれが一組の特徴のうち1つの特徴に相当するとともに1または複数の測定値を含む、特徴値の特徴ベクトルを調製する工程であって、特徴ベクトルは複数組の測定値の各組を用いて得られる少なくとも1つの特徴値を含む、工程と、

d) コンピュータシステムのメモリに、分類器を含む機械学習モデルであって、訓練生体試料から得られる訓練ベクトルを用いて訓練される機械学習モデル、特定された特性を有すると特定された訓練生体試料の第1のサブセット、および特定された特性を有しないと特定された訓練生体試料の第2のサブセットをロードする工程と、

40

e) 機械学習モデルに特徴ベクトルを入力して、特定された特性を生体試料が有するかどうかの出力分類を取得することにより、特定された特性を有する対象の集団を判別する工程を含む。

【0269】

別の態様では、本開示は、対象における癌を特定するための方法であって、

a) 生体試料から R N A を単離する工程と、

50

b) 生体試料から得たRNAにRNAアダプタをライゲートして、RNAをcDNAに逆転写する工程と、

c) 工程b)のcDNAを増幅する工程と、

d) cDNA分子の核酸配列を決定する工程と、

e) 表1～表11に列挙される群から選択されるmiRNAの予め特定されたパネルにおいて核酸分子の核酸配列を参照核酸配列にアライメントして、対象のmiRNAプロファイルを判定する工程と、

f) 健康な対象と結腸細胞増殖性障害を患う対象とを判別可能となるよう訓練された機械学習モデルにmiRNAプロファイルを入力することで、結腸細胞増殖性障害の存在に関連する出力値を提供し、それにより対象における結腸細胞増殖性障害の存在を示して、対象が癌を患っている可能性を生成する工程とを含む方法を提供する。

10

【0270】

いくつかの実施形態では、少なくとも約10個の別個のmiRNAは、少なくとも約20個の別個のmiRNAを含み、少なくとも約20個の別個のmiRNAはそれぞれ、表1～表11に列挙されるmiRNAのうち少なくとも一部を含む。いくつかの例では、少なくとも約10個の別個のmiRNAは、少なくとも約30個の別個のmiRNAを含み、少なくとも約30個の別個のmiRNAはそれぞれ、表1～表11に列挙されるmiRNAのうち少なくとも一部を含む。

【0271】

いくつかの実施形態は、特に対象が膵臓病変を患うときに、結腸細胞増殖性障害を患うか、または結腸細胞増殖性障害を発症するリスクが高い対象の試料に差動的に発現された、miRNAのプロファイルを提供する。結腸細胞増殖性障害を患うか、または結腸細胞増殖性障害を発症するリスクが高い対象の試料に差動的に発現されたmiRNAのプロファイルは、上方調節／過剰発現されたmiRNAのプロファイル、および下方調節または過小発現されたmiRNAのプロファイルを含む。

20

【0272】

いくつかの実施形態では、対象において、結腸細胞増殖性障害の存在、または結腸細胞増殖性障害を発症する高リスクを検出するための方法は、

a) 対象中の1または複数のmiRNAの発現レベルを対象から検出する工程と、

b) 検出した発現レベルと参照発現レベルとを比較する工程であって、試料中の1または複数のmiRNAの差動的発現は、参照発現レベルと比較したとき、結腸細胞増殖性障害の不在または結腸細胞増殖性障害を発症する低リスクに対して、結腸細胞増殖性障害の存在または結腸細胞増殖性障害を発症する高リスクを示す、工程とを含む。

30

【0273】

試料中の1または複数のmiRNAの差動的発現は、参照発現レベルと比較して、結腸細胞増殖性障害前駆体を示す場合がある。

【0274】

いくつかの実施形態では、試料は組織試料であり、1または複数のmiRNAは、結腸細胞増殖性障害を発症するリスクが低い対象の対応する細胞と比較して、結腸細胞増殖性障害を発症するリスクが高い対象の細胞中で差動的に発現されるmiRNAのプロファイルに属する。

40

【0275】

いくつかの実施形態では、対象は大腸病変を患っており、1または複数のmiRNAは、膵臓病変を患うとともに、結腸細胞増殖性障害を発症するリスクが低い対象の対応する試料と比較して、大腸病変を患うとともに、結腸細胞増殖性障害を発症するリスクが高い対象の試料中で差動的に発現されたmiRNAのプロファイルに属する。

【0276】

いくつかの方法は、試料中の1または複数のmiRNAの発現レベルを検出するために

50

使用されてよい。例えば、*miRNA*の測定は、バーコードベースアッセイ、*miRNA*マイクロアレイ解析（例えば、チップ）、デジタルポリメラーゼ連鎖反応（PCR）、リアルタイムPCR、量的逆転写PCR（qRT-PCR）、半量的PCR、ノーザンブロット、または*in situ*ハイブリダイゼーションにより実施することができる。例えば、成熟*miRNA*は、例えば*in vitro*アッセイを用いて測定される。

【0277】

発現の閾値またはカットオフ値を確立するための様々な統計的かつ数学的な方法が、使用されてよい。特定のバイオマーカの閾値またはカットオフの発現レベルは、例えば、本開示の実施例と図面に記載されるものなどの、受信者操作特性（ROC）プロットからのデータに基づき選択されてよい。いくつかの実施形態では、これらの閾値またはカットオフの発現レベルは、例えば、特定のバイオマーカまたはそれらの組合せに対してROCプロットに沿って移動し、感度または特異性の様々な値を得ることで全体的なアッセイパフォーマンスに影響を及ぼすことにより、変動することができる。例えば、臨床的観点から強固な診断方法を有することを目的とする場合、高感度を優先する必要がある。しかし、費用対効果の高い方法を目指とする場合、高特異性を優先する必要がある。最良のカットオフは、最高の感度と特異性を生成する特定のバイオマーカについてのROCプロットから得られた値を表す。感度および特異度の値は、閾値（カットオフ）の範囲にわたって算出される。このため、閾値またはカットオフ値は、感度および／または特異度が少なくとも約70%であり、例えば、アッセイされた患者集団の少なくとも60%、またはアッセイされた患者集団の少なくとも65%、70%、75%、もしくは80%において、少なくとも75%、少なくとも80%、少なくとも85%、少なくとも90%、少なくとも95%、少なくとも96%、少なくとも97%、少なくとも98%、少なくとも99%、または少なくとも100%とすることができるように、選択することができる。

【0278】

結果的に、本開示の実施形態の一部は、診断またはスクリーニングされることとなる対象から単離された侵襲性が最小の試料において、前述された少なくともマイクロRNAの発現レベルを判定して、マイクロRNAの発現レベルを所定の閾値またはカットオフ値と比較することにより実行されてもよく、所定の閾値またはカットオフ値は、大腸癌または大腸腺腫を患うリスクにある患者集団において判定されたマイクロRNAの発現レベルに基づき算出されたROC曲線において、所望の感度で最高の特異度と相関するマイクロRNAの発現レベルに相当しており、所定のカットオフ値に対するマイクロRNAのうち少なくとも1つの過剰発現は、対象が大腸癌または大腸腺腫を患うことを所望の感度で示す。

【0279】

別の例として、このような所定の状態は、結腸細胞増殖性障害の予測感度が、例えば、少なくとも約50%、少なくとも約55%、少なくとも約60%、少なくとも約65%、少なくとも約70%、少なくとも約75%、少なくとも約80%、少なくとも約85%、少なくとも約90%、少なくとも約95%、少なくとも約96%、少なくとも約97%、少なくとも約98%、または少なくとも約99%の値を含む状態であってもよい。

【0280】

別の例として、このような所定の状態は、結腸細胞増殖性障害の予測の陽性予測値（PPV）が、例えば、少なくとも約50%、少なくとも約55%、少なくとも約60%、少なくとも約65%、少なくとも約70%、少なくとも約75%、少なくとも約80%、少なくとも約85%、少なくとも約90%、少なくとも約95%、少なくとも約96%、少なくとも約97%、少なくとも約98%、または少なくとも約99%の値を含む状態であってもよい。

【0281】

別の例として、このような所定の状態は、結腸細胞増殖性障害の予測の陰性予測値（NPV）が、例えば、少なくとも約50%、少なくとも約55%、少なくとも約60%、少なくとも約65%、少なくとも約70%、少なくとも約75%、少なくとも約80%、少

なくとも約 85%、少なくとも約 90%、少なくとも約 95%、少なくとも約 96%、少なくとも約 97%、少なくとも約 98%、または少なくとも約 99% の値を含む状態であってもよい。

【0282】

別の例として、このような所定の状態は、結腸細胞増殖性障害の予測における受信者操作特性 (ROC) 曲線の曲線下面積 (AUC) が、少なくとも約 0.50、少なくとも約 0.55、少なくとも約 0.60、少なくとも約 0.65、少なくとも約 0.70、少なくとも約 0.75、少なくとも約 0.80、少なくとも約 0.85、少なくとも約 0.90、少なくとも約 0.95、少なくとも約 0.96、少なくとも約 0.97、少なくとも約 0.98、または少なくとも約 0.99 の値を含む状態であってもよい。

10

【0283】

処置反応性

本明細書に記載の予測分類器、システム、および方法は、(例えば、個体の生体試料に対して本明細書に記載のシグネチャパネルを用いて RNA アッセイを行うことに基づき) 多数の臨床適用のために個体集団を分類するために適用される場合がある。このような臨床適用の例として、初期段階の癌の検出、癌の診断、特定の疾患段階への癌の分類、癌を処置するための治療剤に対する応答性または抵抗性の判定が挙げられる。

【0284】

本明細書に記載の方法とシステムは、等級や段階など、結腸細胞増殖性障害の特性に適用されてよい。そのため、異なる組織中の異なる癌の種類にわたる癌治療薬の応答性を予測して、処置応答性に基づき対象を分類するために、分析物とアッセイとの組合せが本システムと方法に使用される場合がある。いくつかの実施形態では、本明細書に記載の分類器は、対象の群を処置応答者および非応答者へと階層化することが可能である。

20

【0285】

別の態様では、本開示は、疾患を以前に処置された対象における最小残存病変をモニタリングするための方法であって、ベースラインの miRNA 状態として本明細書に記載される miRNA プロファイルを判定する工程と、解析を繰り返して 1 または複数の所定の時点で miRNA プロファイルを判定する工程であって、ベースラインからの変化は対象におけるベースラインでの最小残存病変の変化を示す、工程とを含む方法を、提供する。

【0286】

いくつかの実施形態では、最小残存病変は、処置への応答、腫瘍細胞量、術後残存腫瘍、再発、二次スクリーン、一次スクリーン、および癌の進展から選択される。

30

【0287】

別の態様では、処置に対する応答を判定するための方法が提供される。

【0288】

別の態様では、腫瘍細胞量をモニタリングするための方法が提供される。

【0289】

別の態様では、術後残存腫瘍を検出するための方法が提供される。

【0290】

別の態様では、再発を検出するための方法が提供される。

40

【0291】

別の態様では、二次スクリーンとして使用するための方法が提供される。

【0292】

別の態様では、一次スクリーンとして使用するための方法が提供される。

【0293】

別の態様では、癌の進展をモニタリングするための方法が提供される。

【0294】

本開示はまた、目的の疾病または疾患の薬物標的 (例えば、特定の分類に関連するかそれにおいて重要な遺伝子) を判定するための方法であって、少なくとも 1 つの遺伝子の遺伝子発現レベルについて対象から得られた試料を評価する工程と、近傍解析ルーチン (n

50

neighborhood analysis routine) を用いて試料の分類に関連する遺伝子を判定することで、その分類に関連する1または複数の薬物標的を確認する工程とを含む方法を、提供する。

【0295】

本開示はまた、疾患の分類を処置するように設計された薬物の有効性を判定するための方法であって、疾患の分類を有する個体から試料を得る工程と、試料を薬物に曝露する工程と、少なくとも1つの遺伝子の遺伝子発現のレベルについて薬物曝露を受けた試料を評価する工程と、加重投票スキーム (weighted voting scheme) で構築したコンピュータモデルを用いて、薬物曝露を受けた試料を、モデルの遺伝子発現レベルに対する試料の相対的遺伝子発現レベルに応じて疾患の分類へと分類する工程とを含む方法を、提供する。

10

【0296】

本開示はまた、疾患の分類を処置するように設計された薬物の有効性を判定するための方法であって、個体が薬物に晒されており、当該方法が、薬物に晒された個体から試料を得る工程と、少なくとも1つの遺伝子の遺伝子発現レベルについて試料を評価する工程と、加重投票スキームを用いて構築したモデルを用いて試料を疾患の分類へと分類する工程と、モデルの遺伝子発現レベルと比較して試料の遺伝子発現レベルを評価する工程とを含む方法を、提供する。

【0297】

本開示はまた、対象が表現型分類 (例えば、知能、処置への応答、寿命、ウイルス感染の可能性、または肥満症) に属するかどうかを判定する方法であって、対象から試料を得る工程と、少なくとも1つの遺伝子の遺伝子発現レベルについて試料を評価する工程と、加重投票スキームを用いて構築したモデルを用いて試料を疾患の分類へと分類する工程であって、モデルの遺伝子発現レベルと比較して試料の遺伝子発現レベルを評価することを含む、工程とを含む方法を、提供する。

20

【0298】

一態様では、処置応答性に基づき集団を分類することに関連する、本明細書に記載のシステムと方法は、DNA 損傷剤、DNA 修復標的治療薬、DNA 損傷シグナル伝達の阻害剤、DNA 損傷により誘導される細胞周期休止の阻害剤、およびDNA 損傷を間接的に生じさせるプロセスの阻害といった分類であるが、これらの分類に限定されない化学療法剤により処置される癌について言及している。これら化学療法剤はそれぞれ、用語が本明細書で使用されるとき、「DNA 損傷治療剤」とみなされる場合がある。

30

【0299】

患者の分析物データに基づき、患者は、臨床的再発のリスクが高いまたは低い患者など、高リスクおよび低リスクの患者群に分類される場合があり、その結果を使用して処置の期間が決定されてもよい。例えば、高リスク患者であると判定された患者は、手術後に補助化学療法により処置される場合がある。低リスク患者とみなされた患者では、補助化学療法は手術後に保留される場合がある。したがって、いくつかの態様では、本開示は再発のリスクを示す大腸癌腫瘍の遺伝子発現プロファイルを調製するための方法を提供する。

【0300】

いくつかの例では、本明細書に記載の分類器は、対象の集団を処置応答者と処置非応答者との間で階層化することが可能である。

40

【0301】

別の態様では、本明細書に開示される方法は、癌の検出またはモニタリングを含む臨床用途に適用される場合がある。

【0302】

いくつかの実施形態では、本明細書に開示される方法は、処置に対する応答を判定および／または予測するために適用される場合がある。

【0303】

大腸癌のモニタリング

50

データセットを処理するために訓練されたアルゴリズムを使用した後、大腸癌は対象において特定またはモニタリングされてよい。この特定は、大腸癌関連miRNAのパネルにおけるデータセットの配列リードの量的測定（例えば、RNA転写物の量的測定）に少なくとも部分的に基づく場合がある。例えば、モニタリングは、2つ以上の異なる時点それぞれにおいて対象の大腸癌を評価することを含む場合がある。

【0304】

いくつかの実施形態では、本明細書に開示される方法は、腫瘍細胞量をモニタリングおよび／または予測するために適用される場合がある。

【0305】

いくつかの実施形態では、本明細書に開示される方法は、術後の残存腫瘍を検出および／または予測するために適用される場合がある。

【0306】

いくつかの実施形態では、本明細書に開示される方法は、術後の最小残存病変を検出および／または予測するために適用される場合がある。

【0307】

いくつかの実施形態では、本明細書に開示される方法は、再発を検出および／または予測するために適用される場合がある。

【0308】

一態様では、本明細書に開示される方法は、二次スクリーンとして適用される場合がある。

【0309】

一態様では、本明細書に開示される方法は、一次スクリーンとして適用される場合がある。

【0310】

一態様では、本明細書に開示される方法は、癌の発症をモニタリングするために適用される場合がある。

【0311】

一態様では、本明細書に開示される方法は、癌のリスクをモニタリングおよび／または予測するために適用される場合がある。

【0312】

大腸癌は、少なくとも約50%、少なくとも約55%、少なくとも約60%、少なくとも約65%、少なくとも約70%、少なくとも約75%、少なくとも約80%、少なくとも約81%、少なくとも約82%、少なくとも約83%、少なくとも約84%、少なくとも約85%、少なくとも約86%、少なくとも約87%、少なくとも約88%、少なくとも約89%、少なくとも約90%、少なくとも約91%、少なくとも約92%、少なくとも約93%、少なくとも約94%、少なくとも約95%、少なくとも約96%、少なくとも約97%、少なくとも約98%、少なくとも約99%、またはそれよりも高い精度で、対象において特定される場合がある。訓練済みアルゴリズムにより大腸癌を特定する精度は、大腸癌を患うか患っていないと正確に特定または分類される独立した検査試料（例えば、大腸癌を患うことが明らかな対象、または大腸癌の臨床検査結果が陰性である対象）の割合として算出される場合がある。

【0313】

大腸癌は、少なくとも約5%、少なくとも約10%、少なくとも約15%、少なくとも約20%、少なくとも約25%、少なくとも約30%、少なくとも約35%、少なくとも約40%、少なくとも約45%、少なくとも約50%、少なくとも約55%、少なくとも約60%、少なくとも約65%、少なくとも約70%、少なくとも約75%、少なくとも約80%、少なくとも約81%、少なくとも約82%、少なくとも約83%、少なくとも約84%、少なくとも約85%、少なくとも約86%、少なくとも約87%、少なくとも約88%、少なくとも約89%、少なくとも約90%、少なくとも約91%、少なくとも約92%、少なくとも約93%、少なくとも約94%、少なくとも約95%、少なくとも

10

20

30

40

50

約 96%、少なくとも約 97%、少なくとも約 98%、少なくとも約 99%、またはそれよりも高い陽性予測値（PPV）で、対象において特定される場合がある。訓練済みアルゴリズムを用いて大腸癌を特定する PPV は、実際に大腸癌を患っている対象に相当する、大腸癌を患っていると特定されるか、または分類される無細胞生体試料の割合として算出される場合がある。

【0314】

大腸癌は、少なくとも約 5%、少なくとも約 10%、少なくとも約 15%、少なくとも約 20%、少なくとも約 25%、少なくとも約 30%、少なくとも約 35%、少なくとも約 40%、少なくとも約 45%、少なくとも約 50%、少なくとも約 55%、少なくとも約 60%、少なくとも約 65%、少なくとも約 70%、少なくとも約 75%、少なくとも約 80%、少なくとも約 81%、少なくとも約 82%、少なくとも約 83%、少なくとも約 84%、少なくとも約 85%、少なくとも約 86%、少なくとも約 87%、少なくとも約 88%、少なくとも約 89%、少なくとも約 90%、少なくとも約 91%、少なくとも約 92%、少なくとも約 93%、少なくとも約 94%、少なくとも約 95%、少なくとも約 96%、少なくとも約 97%、少なくとも約 98%、少なくとも約 99%、またはそれよりも高い陰性予測値（NPV）で、対象において特定される場合がある。訓練済みアルゴリズムを用いて大腸癌を特定する NPV は、実際に大腸癌がない対象に相当する、大腸癌を患っていないと特定されるか、または分類される無細胞生体試料の割合として算出される場合がある。

10

【0315】

大腸癌は、少なくとも約 5%、少なくとも約 10%、少なくとも約 15%、少なくとも約 20%、少なくとも約 25%、少なくとも約 30%、少なくとも約 35%、少なくとも約 40%、少なくとも約 45%、少なくとも約 50%、少なくとも約 55%、少なくとも約 60%、少なくとも約 65%、少なくとも約 70%、少なくとも約 75%、少なくとも約 80%、少なくとも約 81%、少なくとも約 82%、少なくとも約 83%、少なくとも約 84%、少なくとも約 85%、少なくとも約 86%、少なくとも約 87%、少なくとも約 88%、少なくとも約 89%、少なくとも約 90%、少なくとも約 91%、少なくとも約 92%、少なくとも約 93%、少なくとも約 94%、少なくとも約 95%、少なくとも約 96%、少なくとも約 97%、少なくとも約 98%、少なくとも約 99%、少なくとも約 99.1%、少なくとも約 99.2%、少なくとも約 99.3%、少なくとも約 99.4%、少なくとも約 99.5%、少なくとも約 99.6%、少なくとも約 99.7%、少なくとも約 99.8%、少なくとも約 99.9%、少なくとも約 99.99%、少なくとも約 99.999%、またはそれよりも高い臨床的感度で、対象において特定される場合がある。訓練済みアルゴリズムを用いて大腸癌を特定する臨床的感度は、大腸癌を患っていると正確に特定または分類される、大腸癌の存在に関連付けられる独立した検査試料（例えば、大腸癌を患っていることが明らかな対象）の割合として算出される場合がある。

20

30

【0316】

大腸癌は、少なくとも約 5%、少なくとも約 10%、少なくとも約 15%、少なくとも約 20%、少なくとも約 25%、少なくとも約 30%、少なくとも約 35%、少なくとも約 40%、少なくとも約 45%、少なくとも約 50%、少なくとも約 55%、少なくとも約 60%、少なくとも約 65%、少なくとも約 70%、少なくとも約 75%、少なくとも約 80%、少なくとも約 81%、少なくとも約 82%、少なくとも約 83%、少なくとも約 84%、少なくとも約 85%、少なくとも約 86%、少なくとも約 87%、少なくとも約 88%、少なくとも約 89%、少なくとも約 90%、少なくとも約 91%、少なくとも約 92%、少なくとも約 93%、少なくとも約 94%、少なくとも約 95%、少なくとも約 96%、少なくとも約 97%、少なくとも約 98%、少なくとも約 99%、少なくとも約 99.1%、少なくとも約 99.2%、少なくとも約 99.3%、少なくとも約 99.4%、少なくとも約 99.5%、少なくとも約 99.6%、少なくとも約 99.7%、少なくとも約 99.8%、少なくとも約 99.9%、少なくとも約 99.99%、少なくとも約 99.999%、またはそれよりも高い臨床的特異性で、対象において特定される場

40

50

合がある。訓練済みアルゴリズムを用いて大腸癌を特定する臨床的特異性は、大腸癌を患っていないと正確に特定または分類される、大腸癌の不在に関連付けられる独立した検査試料（例えば、大腸癌の臨床検査結果が陰性の対象）の割合として算出される場合がある。

【0317】

いくつかの実施形態では、訓練済みアルゴリズムは、対象が、少なくとも約5%、少なくとも約10%、少なくとも約15%、少なくとも約20%、少なくとも約25%、少なくとも約30%、少なくとも約35%、少なくとも約40%、少なくとも約45%、少なくとも約50%、少なくとも約55%、少なくとも約60%、少なくとも約65%、少なくとも約70%、少なくとも約75%、少なくとも約80%、少なくとも約81%、少なくとも約82%、少なくとも約83%、少なくとも約84%、少なくとも約85%、少なくとも約86%、少なくとも約87%、少なくとも約88%、少なくとも約89%、少なくとも約90%、少なくとも約91%、少なくとも約92%、少なくとも約93%、少なくとも約94%、少なくとも約95%、少なくとも約96%、少なくとも約97%、少なくとも約98%、少なくとも約99%、またはそれよりも高い大腸癌のリスクにあることを判定する場合がある。

10

【0318】

訓練済みアルゴリズムは、対象が、少なくとも約50%、少なくとも約55%、少なくとも約60%、少なくとも約65%、少なくとも約70%、少なくとも約75%、少なくとも約80%、少なくとも約81%、少なくとも約82%、少なくとも約83%、少なくとも約84%、少なくとも約85%、少なくとも約86%、少なくとも約87%、少なくとも約88%、少なくとも約89%、少なくとも約90%、少なくとも約91%、少なくとも約92%、少なくとも約93%、少なくとも約94%、少なくとも約95%、少なくとも約96%、少なくとも約97%、少なくとも約98%、少なくとも約99%、少なくとも約99.1%、少なくとも約99.2%、少なくとも約99.3%、少なくとも約99.4%、少なくとも約99.5%、少なくとも約99.6%、少なくとも約99.7%、少なくとも約99.8%、少なくとも約99.9%、少なくとも約99.99%、またはそれよりも高い精度で大腸癌のリスクにあることを判定する場合がある。

20

【0319】

大腸癌を患っていると対象を特定するに際して、対象は、任意選択で治療的介入（例えば、対象の大腸癌を処置するための適切な処置期間を定める）が提供されてよい。治療的介入は、薬物の有効量の処方、大腸癌に対するさらなる検査または評価、大腸癌のさらなるモニタリング、またはそれらの組合せを含む場合がある。対象が現在、処置期間により大腸癌の処置を受けている場合、治療的介入は、（例えば、現在の処置期間が有効でないことにより処置の有効性を増加させるために）その後の異なる処置の期間を含む場合がある。

30

【0320】

治療的介入は、大腸癌の診断を確認するために二次臨床検査に対象を推薦することを含み得る。二次臨床検査は、画像検査、血液検査、コンピュータ断層撮影（CT）スキャン、磁気共鳴画像法（MRI）スキャン、超音波スキャン、胸部X線、陽電子放出断層撮影（PET）スキャン、PET-CTスキャン、無細胞生体細胞学（cell-free biological cytology）、FIT検査、FOBT検査、またはそれらのあらゆる組合せを含む場合がある。

40

【0321】

大腸癌関連ゲノム座位のパネルにおけるデータセットの配列リードの量的測定（例えば、大腸癌関連ゲノム座位におけるRNA転写物の量的測定）は、患者（例えば、大腸癌を患う対象または大腸癌の処置を受けている対象）をモニタリングするために一定期間にわたり評価される場合がある。このような場合、患者のデータセットの量的測定は、処置期間の間に変化する場合がある。例えば、有効な処置により大腸癌リスクが低下した患者の

50

データセットの量的測定は、健康な対象（例えば、大腸癌がない対象）のプロファイルまたは分布に向けて推移し得る。反対に、例えば、有効でない処置により大腸癌のリスクが増加した患者のデータセットの量的測定は、大腸癌が高いか、または大腸癌がさらに進行した対象のプロファイルまたは分布に向けて推移し得る。

【0322】

対象の大腸癌は、対象の大腸癌を処置するための処置期間をモニタリングすることにより、モニタリングされてよい。モニタリングは、2つ以上の異なる時点において対象の大腸癌を評価することを含む場合がある。この評価は、2つ以上の時点それぞれにおいて判定される大腸癌関連ゲノム座位のパネルに対する量的測定を含む、大腸癌関連ゲノム座位のパネルにおけるデータセットの配列リードの量的測定（例えば、大腸癌関連ゲノム座位

10

【0323】

いくつかの実施形態では、2つ以上の時点の間に判定された大腸癌関連ゲノム座位のパネルに対する量的測定を含む、大腸癌関連ゲノム座位のパネルにおけるデータセットの配列リードの量的測定（例えば、大腸癌関連ゲノム座位におけるRNA転写物またはDNAの量的測定）における差異は、(i) 対象の大腸癌の診断、(ii) 対象の大腸癌の予後、(iii) 対象の大腸癌のリスク増加、(iv) 対象の大腸癌のリスク低下、(v) 対象の大腸癌を処置するための処置期間の有効性、および(vi) 対象の大腸癌を処置するための処置期間の非有効性など、1または複数の臨床的指標を示すことがある。

【0324】

いくつかの実施形態では、2つ以上の時点間で判定される大腸癌関連のゲノム座位のパネルの量的測定を含む、大腸癌関連のゲノム座位のパネルにおけるデータセットの配列リードの量的測定（例えば、大腸癌関連のゲノム座位におけるRNA転写物またはDNAの量的測定）の差は、対象の大腸癌の診断を示す場合がある。例えば、大腸癌が以前の時点で対象に検出されなかったが、後の時点で対象に検出された場合、その差は、対象の大腸癌の診断を示す。対象の大腸癌の診断に対するこのような指標に基づき、例えば、対象に新たな治療介入を実施するなどの臨床行為または決定を行ってもよい。臨床行為または決定は、大腸癌の診断を確認するために二次臨床検査に対象を推薦することを含み得る。二次臨床検査は、画像検査、血液検査、コンピュータ断層撮影(CT)スキャン、磁気共鳴画像法(MRI)スキャン、超音波スキャン、胸部X線、陽電子放出断層撮影(PET)スキャン、PET-CTスキャン、無細胞生体細胞学、FIT検査、FOBT検査、またはそれらのあらゆる組合せを含む場合がある。

20

30

【0325】

いくつかの実施形態では、2つ以上の時点間で判定される大腸癌関連のゲノム座位のパネルの量的測定を含む、大腸癌関連のゲノム座位のパネルにおけるデータセットの配列リードの量的測定（例えば、大腸癌関連のゲノム座位におけるRNA転写物またはDNAの量的測定）の差は、対象の大腸癌の予後を示す場合がある。

【0326】

いくつかの実施形態では、2つ以上の時点間で判定される大腸癌関連のゲノム座位のパネルの量的測定を含む、大腸癌関連のゲノム座位のパネルにおけるデータセットの配列リードの量的測定（例えば、大腸癌関連のゲノム座位におけるRNA転写物またはDNAの量的測定）の差は、大腸癌のリスクが増加した対象を示す場合がある。例えば、大腸癌が早い時点と遅い時点の両方で対象に検出され、その差が正の差である場合（例えば、大腸癌関連ゲノム座位のパネルにおけるデータセットの配列リードの量的測定（例えば、大腸癌関連ゲノム座位におけるRNA転写物またはDNAの量的測定）が、早い時点から遅い時点まで増加した場合）、その差は、大腸癌のリスクが増加したことを示す場合がある。大腸癌のリスク増加に対するこのような指標に基づき、例えば、対象に新たな治療介入を実施するか、治療介入を切り替える（例えば、現在の処置を終了し、新たな処置を実施する）などの臨床行為または決定を行ってもよい。臨床行為または決定は、大腸癌のリスク増加を確認するために二次臨床検査に対象を推薦することを含み得る。二次臨床検査は、

40

50

画像検査、血液検査、コンピュータ断層撮影（ＣＴ）スキャン、磁気共鳴画像法（ＭＲＩ）スキャン、超音波スキャン、胸部Ｘ線、陽電子放出断層撮影（ＰＥＴ）スキャン、ＰＥＴ－ＣＴスキャン、無細胞生体細胞学、ＦＩＴ検査、ＦＯＢＴ検査、またはそれらのあらゆる組合せを含む場合がある。

【０３２７】

いくつかの実施形態では、２つ以上の時点間で判定される大腸癌関連のゲノム座位のパネルの量的測定を含む、大腸癌関連のゲノム座位のパネルにおけるデータセットの配列リードの量的測定（例えば、大腸癌関連のゲノム座位におけるＲＮＡ転写物またはＤＮＡの量的測定）の差は、大腸癌のリスクが減少した対象を示す場合がある。例えば、大腸癌が早い時点と遅い時点の両方で対象に検出され、その差が負の差である場合（例えば、大腸癌関連ゲノム座位のパネルの量的測定を含む、大腸癌関連ゲノム座位のパネルにおけるデータセットの配列リードの量的測定（例えば、大腸癌関連ゲノム座位におけるＲＮＡ転写物またはＤＮＡの量的測定）が、早い時点から遅い時点まで減少した場合）、その差は、大腸癌のリスクが減少したことを示す場合がある。大腸癌のリスク減少に対するこのような指標に基づき、対象に対し臨床行為または決定（例えば、現在の治療介入を継続するか、または終了させる）を行ってもよい。臨床行為または決定は、大腸癌のリスク減少を確認するために二次臨床検査に対象を推薦することを含み得る。二次臨床検査は、画像検査、血液検査、コンピュータ断層撮影（ＣＴ）スキャン、磁気共鳴画像法（ＭＲＩ）スキャン、超音波スキャン、胸部Ｘ線、陽電子放出断層撮影（ＰＥＴ）スキャン、ＰＥＴ－ＣＴスキャン、無細胞生体細胞学、ＦＩＴ検査、ＦＯＢＴ検査、またはそれらのあらゆる組合せを含む場合がある。

10

20

【０３２８】

いくつかの実施形態では、２つ以上の時点間で判定される大腸癌関連のゲノム座位のパネルの量的測定を含む、大腸癌関連のゲノム座位のパネルにおけるデータセットの配列リードの量的測定（例えば、大腸癌関連のゲノム座位におけるＲＮＡ転写物またはＤＮＡの量的測定）の差は、対象の大腸癌を処置するための処置期間の有効性を示す場合がある。例えば、大腸癌が以前の時点で対象に検出されたが、後の時点で対象に検出されなかった場合、その差は、対象の大腸癌を処置するための処置期間の有効性を示す場合がある。対象の大腸癌を処置するための処置期間の有効性に対するこのような指標に基づき、例えば、対象において現在の治療介入を継続するか、または終了するなどの臨床行為または決定を行ってもよい。臨床行為または決定は、大腸癌を処置するための処置期間の有効性を確認するために二次臨床検査に対象を推薦することを含み得る。二次臨床検査は、画像検査、血液検査、コンピュータ断層撮影（ＣＴ）スキャン、磁気共鳴画像法（ＭＲＩ）スキャン、超音波スキャン、胸部Ｘ線、陽電子放出断層撮影（ＰＥＴ）スキャン、ＰＥＴ－ＣＴスキャン、無細胞生体細胞学、ＦＩＴ検査、ＦＯＢＴ検査、またはそれらのあらゆる組合せを含む場合がある。

30

【０３２９】

いくつかの実施形態では、２つ以上の時点間で判定される大腸癌関連のゲノム座位のパネルの量的測定を含む、大腸癌関連のゲノム座位のパネルにおけるデータセットの配列リードの量的測定（例えば、大腸癌関連のゲノム座位におけるＲＮＡ転写物またはＤＮＡの量的測定）の差は、対象の大腸癌を処置するための処置期間の非有効性を示す場合がある。例えば、大腸癌が早い時点と遅い時点の両方で対象に検出され、その差が正またはゼロの差である場合（例えば、大腸癌関連ゲノム座位のパネルの量的測定を含む、大腸癌関連ゲノム座位のパネルにおけるデータセットの配列リードの量的測定（例えば、大腸癌関連ゲノム座位におけるＲＮＡ転写物またはＤＮＡの量的測定）が、早い時点から遅い時点まで増加するか、または一定のレベルを維持した場合）、かつ、有効な処置が早い時点で認められた場合、その差は、対象の大腸癌を処置するための処置期間の非有効性を示す場合がある。対象の大腸癌を処置するための処置期間の非有効性に対するこのような指標に基づき、例えば、対象において現在の治療介入を終了する、および／または異なる新たな治療介入に切り替える（例えば、実施する）などの臨床行為または決定を行ってもよい。臨

40

50

床行為または決定は、大腸癌を処置するための処置期間の非有効性を確認するために二次臨床検査に対象を推薦することを含み得る。二次臨床検査は、画像検査、血液検査、コンピュータ断層撮影（ＣＴ）スキャン、磁気共鳴画像法（ＭＲＩ）スキャン、超音波スキャン、胸部Ｘ線、陽電子放出断層撮影（ＰＥＴ）スキャン、ＰＥＴ－ＣＴスキャン、無細胞生体細胞学、ＦＩＴ検査、ＦＯＢＴ検査、またはそれらのあらゆる組合せを含む場合がある。

【０３３０】

V I I . キット

本開示は、対象の癌を特定またはモニタリングするためのキットを提供する。キットは、対象の無細胞生体試料中の複数の癌関連ゲノム座位それぞれにおける配列の量的測定（例えば、存在、不在、または相対量を示すもの）を特定するためのプローブまたはプライマーを含む場合がある。無細胞生体試料中の複数の癌関連ゲノム座位それぞれにおける配列の量的測定（例えば、存在、不在、または相対量を示すもの）は、１または複数の癌を示す場合がある。プローブは、無細胞生体試料中の複数の癌関連ゲノム座位における配列に対して選択的であり得る。キットは、対象の無細胞生体試料中の複数の癌関連ゲノム座位それぞれにおける配列の量的測定（例えば、存在、不在、または相対量を示すもの）を示すデータセットを生成するために、プローブを用いて無細胞生体試料を処理するための指示書（instructions）を含んでもよい。

10

【０３３１】

キット中のプローブは、無細胞生体試料中の複数の癌関連ゲノム座位における配列に対して選択的であり得る。キット中のプローブは、複数の癌関連ゲノム座位に相当する核酸（例えば、RNAまたはDNA）分子を選択的に富化するように構成され得る。キット中のプローブは、核酸プライマーであってよい。キット中のプローブは、複数の癌関連miRNAまたはそのフラグメントのうち１もしくは複数から得られる核酸配列に対し、部分的または完全な配列相補性を有する場合がある。複数の癌関連miRNAは、少なくとも２、少なくとも３、少なくとも４、少なくとも５、少なくとも６、少なくとも７、少なくとも８、少なくとも９、少なくとも１０、少なくとも１１、少なくとも１２、少なくとも１３、少なくとも１４、少なくとも１５、少なくとも１６、少なくとも１７、少なくとも１８、少なくとも１９、少なくとも２０、またはそれ以上の別個の癌関連miRNAを含む場合がある。複数の癌関連miRNAは、表１～表１１に列挙されるmiRNAからなる群から選択される１または複数の員を含む場合がある。

20

30

【０３３２】

キット中の指示書は、無細胞生体試料中の複数の癌関連ゲノム座位における配列に対して選択的であるプローブを用いて、無細胞生体試料をアッセイするための指示書を含み得る。これらのプローブは、複数の癌関連ゲノム座位のうち１または複数から得られる核酸配列（例えば、RNAまたはDNA）に対して配列相補性を有する核酸分子（例えば、RNAまたはDNA）であってよい。これらの核酸分子は、プライマーまたは富化配列であり得る。無細胞生体試料をアッセイするための指示書は、アレイハイブリダイゼーション、ポリメラーゼ連鎖反応（PCR）、または核酸シーケンシング（例えば、DNAシーケンシングまたはRNAシーケンシング）を行い、無細胞生体試料を処理して、無細胞生体試料中の複数の癌関連miRNAそれぞれの量的測定（例えば、存在、不在、または相対量を示すもの）を示すデータセットを生成するための指示書を含み得る。無細胞生体試料中の複数の癌関連miRNAそれぞれにおける配列の量的測定（例えば、存在、不在、または相対量を示すもの）は、１または複数の癌を示す場合がある。

40

【０３３３】

キット中の指示書は、アッセイの読み出しを測定かつ解釈するための指示書を含む場合があり、この読み出しは、複数の癌関連ゲノム座位のうち１または複数において定量されて、無細胞生体試料中の複数の癌関連ゲノム座位それぞれにおける配列の量的測定（例えば、存在、不在、または相対量を示すもの）を示すデータセットを生成する場合がある。例えば、複数の癌関連miRNAに相当する、アレイハイブリダイゼーションまたはポリ

50

メラーゼ連鎖反応（PCR）の定量は、無細胞生体試料中の量的測定（例えば、存在、不在、または相対量を示すもの）、およびmiRNAの生体特性が明らかな対象間の差動的発現を示すデータセットを生成する場合がある。アッセイの読み出しは、量的PCR（qPCR）値、デジタルPCR（dPCR）値、デジタルドロップレットPCR（ddPCR）値、蛍光値など、またはそれらの正規化された値を含む場合がある。

【実施例】

【0334】

実施例1：無細胞核酸に対するmiRNA解析

本検査では、計276名の被験者を前向きに対象とした。145名は新たに散発性大腸腫瘍と診断された患者（CRC39名、進行性腺腫（AA）49名、非進行性腺腫57名）、131名はいずれの癌の病歴もなく、最近の大腸内視鏡検査で大腸腫瘍性病変の欠如を認めた健康な個人である。AA患者は、大きさが少なくとも10mmの腺腫、もしくは組織学的に高悪性度の異形成、または20%を超える絨毛成分を認めた患者であった。全個体で血液試料を採取してから、内視鏡検査または手術を行った。

【0335】

検査コホートの説明は表12に示し、この表には、（病期、性別、年齢による）分類モデルに対するCRC実験に使用した健康な試料と癌試料の数を示す。

【0336】

【表12】

表 12

CRC		癌 (n=39)	AA (n=49)	NAA (n= 57)	対照 (n=131)
性別	女性 119 (43.1%)	17	17	20	65
	男性 157 (56.9%)	22	32	37	66
年齢	中央値, IQR	中央年齢: 61 IQR: 55.0-68.5	中央年齢: 63 IQR: 53.0-67.0	中央年齢: 63 IQR: 56.0-68.0	中央年齢: 59.0 IQR: 54.0-67.5

【0337】

276名の個体それぞれにおいて、MagMax mirVana Total RNA単離キットを用いて、血漿100マイクロリットル（uL）からRNAを抽出した。抽出の間、分量が明らかな52個の合成miRNAを、各血漿試料に対するスパイクインとして添加した。

【0338】

シーケンシングライブラリの構築

A. ライブラリ構築

無細胞核酸試料をライブラリ調製にかけ、次世代シーケンシングを行った。本明細書中で使用される次世代シーケンシング手法において、「ライブラリ調製」には、末端修復、A-ターミング、アダプタライゲーション、または、後のRNAのシーケンシングを可能にするために無細胞RNA上で行われるあらゆる他の調製が含まれる。いくつかの例では

、調製された無細胞核酸ライブラリ配列は、無細胞核酸試料分子上にライゲートされるアダプタ、配列タグ、インデックスバーコードを伴う。次世代シーケンシングのためのライブラリ調製を容易にするために、様々な市販のキットが利用可能である。次世代シーケンシングライブラリの構築は、高スループットシーケンシングのために、まとまった一連の (coordinated series of) 酵素反応を用いて核酸標的を調製して、特定のサイズのRNAフラグメントの寄せ集めを生成することを含む。新たなライブラリ調製技法の進歩と開発により、次世代シーケンシングの適用は、トランスクリプトミクスやエピジェネティクスなどの分野にまで拡大している。

【0339】

いくつかの実施形態では、ライブラリ調製キットは、Nextera Flex (Illumina)、Ion Ampliseq (Thermo Fisher Scientific)、および Genexus (Thermo Fisher Scientific)、Agilent ClearSeq (Agilent)、Agilent SureSelect Capture (Agilent)、RealSeq (RealSeq Biosciences)、Archer FusionPlex (Illumina)、BiooScientific NEXTFlex (Perkin Elmer)、IDT xGen (Illumina)、Illumina TruSight (Illumina)、SMARTer smRNA-Seq (Takara)、Nimblegen SeqCap (Illumina)、Qiaseq (Qiagen)、または Qiagen GeneRead (Qiagen) から選択される。

10

20

【0340】

RealSeq (登録商標) (RealSeq Biosciences、旧 SomaGenics) は、次世代シーケンシング (NSG) における組み込みバイアスを大幅に低減する、小型のRNAシーケンシングライブラリを調製するための方法である。この技法により、よく用いられるシーケンシングライブラリ調製において、一部では10,000倍にもなる多くのmiRNAの検出不足が生じるという課題が解決する。大半のバイアスは、配列決定ライブラリの調製中にmiRNA/小RNAの3'および5'末端に2個のアダプタを取り付ける酵素ライゲーション反応における配列依存的な変動が原因である。新規な単一のアダプタと分子内ライゲーション (circularization) を用いることで、RealSeq (登録商標) はライブラリ調製のバイアスを大幅に低減する。

30

【0341】

B. シーケンシング

シーケンシングライブラリを2ナノモル (nM) に希釈した。個体1名につき約10万回のリード深さにまでシーケンシングを行うために、試料をIllumina NextSeq機器の上に載せた。

【0342】

C. RealSeq解析の概要

Trimomatic、Skewer、SeqPurge、Atroposなどのトリミングソフトウェアを使用して配列リードをトリミングし、RealSeq特有のアダプタ、ならびに他のプライマーおよびアダプタを取り除いた。次いで、bowtie2アライメントソフトウェアを用いて、トリミングされたリードを、成熟miRNA配列、tRNA、rRNA、ミトコンドリアRNA、U snRNA、Y RNA、および合成スパイクイン配列で作製したカスタムトランスクリプトームへとアライメントした。samtools idxstatsを使用し、各特徴に対するマッピングのリード数を計数した。

40

【0343】

様々なデータ前処理工程を行い、解析のためにデータをさらに特色付けた。データの前処理は、欠損値の追加、情報の集計、カテゴリによるデータの標識化 (データビンニング)、および軌道の平滑化を目的とする。主成分分析や特徴選択のような、より高度な技法は

50

統計式を用いて実行し、複雑なデータセットに適用した。

【0344】

ヘアピン配列へのマッピングのリードが300,000未満の試料は、その後の解析から除外した。全試料にわたり数が25未満の特徴も除外した。比正規化平均 (mean of ratios normalization)、M値、RUVseq、およびlog2CPMのトリミング平均を含む、いくつかの正規化戦略を比較した。比正規化平均法を使用し、データを正規化した。

【0345】

カスタムスクリプトを使用して配列をトリミングし、miRNA配列からIlluminaアダプタを除外した。次いで、bowtieを用いてmiRNA配列を、(1) ヒトゲノム、(2) 把握された成熟ヒトmiRNAをすべて含むカスタムトランスクリプトーム、および(3) ヘアピンmiRNA配列の第2のトランスクリプトームのほか、非ヒト由来の合成miRNA52個のリストにもマッピングした。マッピング位置から、miRNAヘアピンとmiRNAヘアピン両方の計数表を各個体から作製した。

【0346】

これらの計数データセットを用いて、健康な個体と大腸癌または進行腺腫を患う個体との血漿間で差動的に豊富であるmiRNAを特定した。両方の計数マトリクスを、(1) M値のトリミング平均、(2) 比の平均、および(3) RUVseqを含む複数の方法を用いて正規化した。これら3つの方法はそれぞれ、正規化因子を生成するために全RNAライブラリ上に加え、第2の正規化因子のセットを生成するために合成RNAスパイクイン上でも実行した。各方法から正規化した計数を別々に使用して、有意に異なる成熟miRNA、および健康な状態でのmiRNAヘアピン存在量を表す配列を特定するモデルを生成した。2つ以上の方法においてpが0.05未満であるヘアピンとmiRNAは強力に支持されていることを認め、CRCおよび／またはAAのマーカとして特定した。

【0347】

分類モデルでは、第4の正規化法としてCombatを追加して、同じ正規化を両データセット(成熟miRNAとmiRNAヘアピン)に使用した。正規化した計数は、血漿の質、年齢、性別、および採取部位を含む把握された交絡因子に対しても調整した。得られた計数データセットから、 $k=8$ のk-bestモデルを構築した。それぞれデータを4つのフォールドに分割し、データの3つのフォールド上で訓練を行い、最後の4番目のフォールドで検査を行う500個の独立したシード上で、モデルを実行した。次いで、各モデルで特定した8つの特徴を、差動的に豊富なmiRNAとヘアピンのリストと比較した。

【0348】

実施例2：大腸癌および進行腺腫の分類シグネチャの生成

評価された機械学習法には、場合により次元削減、続いて教師あり分類アルゴリズムを含む、一連の変換が含まれていた。

【0349】

分類モデルの評価における交差検証(CV)手順の目的は、モデルの構築に使用しなかった、以前に確認されなかった新たなデータに対するモデルのパフォーマンスを推定することである。目標は、データの異なるサブセット上でモデルを繰り返し訓練して、訓練中にモデルにより確認されないデータの提供されたサブセット上で検査することにより、近似値を提供することである。K分割交差検証の手順では、データセット全体をk個の群に分割する必要がある。k個の群(すなわちフォールド)それぞれにおいて、機械学習モデルを他のk-1個のフォールドを用いて訓練し、提供されたフォールドを検査セットとして使用する。層別化したk分割交差検証により、試料のおおよその割合がフォールド全体でほぼ等価となるように、フォールドへの分割前に試料が分類ごとに層別化される。

【0350】

多くのモデルをk分割交差検証($k=5$)で訓練し、続いて最良の実行モデルを、追加の交差検証手順で評価した。上位10個の特徴を、交差検証中にANOVA F-統計に

基づき選択した。ロジスティック回帰を使用して試料を分類した。100個の異なる乱数シードを使用してデータをフォールドに分割したため、計500個のモデルを実行した（100個のシード、シード1個につき5個のフォールド）。

【0351】

特徴選択中にmiRNAを選択した。シードのカラム数は、実行された500個のモデルに対する特徴選択中にmiRNAが取り上げられた回数を示す。hsa-mir-889において数が多い、例えば499の場合、このmiRNAが500個のモデルのうち499において選択されたことが認められる。

【0352】

まず、外れ値（全訓練試料においてその特徴の99パーセンタイルを上回った所与の試料の特徴値として定めた値）を、99パーセンタイルの値に代入した。続いて、平均を減算し、標準偏差で除算することにより、各特徴を全訓練試料に対して標準化した。訓練セットの平均と標準偏差を用いての同じ外れ値の置換を使用して、検査セットを標準化した。次元削減変換法を選択した場合、これを訓練セット上で訓練して、訓練セットと検査セット両方の試料すべてに適用した。本試験で使用した次元低減変換は、打ち切り特異値分解（SVD）と主成分分析（PCA）であった。

【0353】

候補となる2つの分類アルゴリズム、すなわちロジスティック回帰とサポート・ベクター・マシン（SVM）を、変換入力上で訓練した。訓練データから無作為に選択した20%を含む検証セットを有するフォールドごとに100回の反復の無作為検索を使用して、各方法について複数のハイパーパラメータを検討した。最良の実行検証に相当するハイパーパラメータを選択して機械学習モデルを訓練し、検査フォールドを評価した。ロジスティック回帰には2つのハイパーパラメータとして、正則化強度の逆数と、 L_1 または L_2 ペナルティのいずれか選択があった。表13は、k分割交差検証において大腸癌を患う個体に関連付けられると特定された無細胞核酸試料中のmiRNAの集団を、ロジスティック回帰モデルに対する関連平均係数値（例えば、miRNAが選択されたシードの数の平均）とともに示す。

【0354】

10

20

30

40

50

【表 13 - 1】

表 13

miRNA	(500の中から) 選択されたシードの数	ロジスティック回帰の平均係数値 (miRNA が選択されたシードの数の平均)
hsa-mir-889	499	0.01227171412
hsa-mir-543	392	0.02242937901
hsa-mir-376b	366	0.001838778481
hsa-mir-335	342	0.00000399410613
hsa-mir-1185-1	262	-0.008112394768
hsa-mir-548k	233	0.03467194037
hsa-mir-12135	208	0.1467320116
hsa-mir-369	168	-0.001607343291
hsa-mir-190a	165	0.001031765066
hsa-mir-6770-1	143	0.09696291942
hsa-mir-382	135	0.003105079756
hsa-mir-1843	129	0.001999311556
hsa-mir-142	112	-0.00002348964374
hsa-mir-485	110	0.005757528395
hsa-mir-548ax	98	0.01063443146
hsa-mir-548e	98	0.06965708027
hsa-mir-548al	89	0.05388619835
hsa-mir-548am	89	0.001803661864
hsa-mir-590	82	-0.0005913171732
hsa-mir-135a-2	81	0.209409359
hsa-mir-6770-3	79	0.130261975
hsa-mir-410	77	-0.000443395015
hsa-mir-376a-1	72	0.00006790511459
hsa-mir-377	69	-0.003048876316

【 0 3 5 5 】

10

20

30

40

50

【表 1 3 - 2】

hsa-mir-570	62	0.01211643368
hsa-mir-381	55	0.007964852332
hsa-mir-665	48	0.01925932227
hsa-mir-758	46	0.06822055355
hsa-mir-6511a-3	43	-0.03469441469
hsa-mir-376a-2	42	-0.0004471845472
hsa-mir-155	38	0.004215096372
hsa-mir-3140	38	0.07299821889
hsa-mir-1277	34	0.006571441215
hsa-mir-340	32	-0.001637170648
hsa-mir-548n	28	0.03937433917
hsa-mir-518b	27	0.0866790012
hsa-mir-654	26	-0.003946375455
hsa-mir-5581	23	0.1074504023
hsa-mir-409	22	-0.002519196915
hsa-mir-628	21	0.003309890234
hsa-mir-10399	19	0.01400376339
hsa-mir-3184	19	0.0005741457151
hsa-mir-423	14	-0.0007777398231
hsa-mir-548z	12	0.002962701142
hsa-mir-374a	11	0.0009378886457
hsa-mir-548a-3	11	0.01152914924
hsa-mir-6770-2	10	0.006885855981
hsa-mir-1185-2	9	0.003843378367
hsa-mir-6077	9	0.1848254301
hsa-mir-3202-1	8	-0.1448356476
hsa-mir-548o-2	8	-0.01159264787

10

20

30

40

【 0 3 5 6 】

50

【表 1 3 - 3】

hsa-mir-3143	7	0.02527664699
hsa-mir-5009	7	0.1928650892
hsa-mir-548g	7	0.08575802863
hsa-mir-656	7	0.01393539188
hsa-mir-6818	7	0.08412729176
hsa-mir-548l	6	0.008159810018
hsa-mir-548v	6	0.06831132882
hsa-mir-106b	5	0.0001420887548
hsa-mir-133b	5	0.2074284074
hsa-mir-548h-2	5	0.08619580177
hsa-mir-136	4	-0.003615557403
hsa-mir-26b	4	0.00001385575386
hsa-mir-433	4	0.006869534333
hsa-mir-4719	4	0.3432401024
hsa-mir-3610	3	0
hsa-mir-374b	3	-0.004197640226
hsa-mir-376c	3	-0.0009238950727
hsa-mir-4779	3	0.1750502307
hsa-mir-496	3	0.005980723795
hsa-mir-622	3	0.1956975342
hsa-mir-671	3	-0.007865086858
hsa-mir-6876	3	0.1454046126
hsa-let-7b	2	-0.000008662808689
hsa-mir-103a-1	2	-0.001822456296
hsa-mir-103a-2	2	0.001520027525
hsa-mir-10400	2	0.1984508462
hsa-mir-135a-1	2	0.05744402254

10

20

30

40

【 0 3 5 7 】

50

【表 1 3 - 4】

hsa-mir-139	2	0.0005742609812
hsa-mir-191	2	-0.00001105403881
hsa-mir-2392	2	0.2984791466
hsa-mir-26a-2	2	-0.00002437884444
hsa-mir-320c-2	2	0.001248926099
hsa-mir-4468	2	0.06782728547
hsa-mir-4484	2	0.01846424611
hsa-mir-4521	2	0.1751726514
hsa-mir-4706	2	0.1722001142
hsa-mir-487a	2	0.001771573345
hsa-mir-548ac	2	0.06019265588
hsa-mir-5588	2	0.02659738315
hsa-mir-579	2	0.001881716483
hsa-mir-6734	2	-0.0009272692234
hsa-mir-6882	2	0.07055954999
hsa-mir-9-1	2	0.01230175494
hsa-mir-93	2	0.00001165527636
hsa-let-7i	1	-0.00006929515082
hsa-mir-103b-1	1	0.001366745737
hsa-mir-103b-2	1	-0.0004961485176
hsa-mir-12136	1	0.0006137906358
hsa-mir-1283-2	1	0.1251113007
hsa-mir-1287	1	0.0217592237
hsa-mir-130a	1	-0.00002556546956
hsa-mir-146a	1	-0.0000105260071
hsa-mir-146b	1	0.0003532849175
hsa-mir-151b	1	0.0002453616989

10

20

30

40

【 0 3 5 8 】

50

【表 1 3 - 5】

hsa-mir-154	1	0.001228707094
hsa-mir-18a	1	-0.0005226497861
hsa-mir-2278	1	0.006264706585
hsa-mir-2355	1	0.009943231933
hsa-mir-3138	1	0.102350404
hsa-mir-3168	1	-0.1457401201
hsa-mir-3202-2	1	-0.1529618687
hsa-mir-320c-1	1	-0.003100822914
hsa-mir-320d-1	1	0.003877366341
hsa-mir-320d-2	1	0.005518469682
hsa-mir-324	1	0.001810403358
hsa-mir-331	1	-0.001317093481
hsa-mir-3529	1	0.002602265853
hsa-mir-374c	1	0.008427415
hsa-mir-431	1	0
hsa-mir-450b	1	0.02620627163
hsa-mir-4651	1	0
hsa-mir-4716	1	-0.1316026223
hsa-mir-4763	1	-0.09201122096
hsa-mir-4772	1	0.03507199649
hsa-mir-548a-1	1	0.05655605303
hsa-mir-548ah	1	0.160519459
hsa-mir-548aj-1	1	0.006007596705
hsa-mir-5703	1	0.1132904105
hsa-mir-574	1	0.003831642737
hsa-mir-6729	1	-0.2068859174
hsa-mir-6781	1	0.1265329942

10

20

30

40

【 0 3 5 9 】

50

【表 1 3 - 6】

hsa-mir-6802	1	0.1171690204
hsa-mir-7-1	1	0
hsa-mir-7111	1	0.0425464288
hsa-mir-9902-1	1	0.01191197396

10

【0 3 6 0】

合計で、138個のmiRNAが大腸癌に関連することが分かった。健康な個体と大腸癌を患う個体とを判別するために、特定されたmiRNAをすべて分類モデルに含める必要はなかった。このため、一部の領域は、全体として評価された様々な種類の癌を示すと考えられる。他のmiRNAは、大腸癌のサブグループにおいてより頻繁である。本アッセイと、検査した癌の種類から、特定のmiRNAは、CVにおいてより高い頻度で「大腸癌に特異的に関連する」と説明され、試料配列を予測モデルで訓練したときにシグネチャにより大きな加重を運び得る。大腸癌に関連するこれらの高い頻度のmiRNAは、健康な人とCRCとの間で個体集団を区別するように訓練された特定のモデルに使用する。図2は、特徴選択中に選択されるmiRNAを示すヒストグラムを提供する図である。バーは、miRNAが選択されたモデルの数（最大＝500）を表す。図3は、頻繁に選択された上位10個のmiRNAのロジスティック回帰計数を示すグラフを提供する図である。

20

【0 3 6 1】

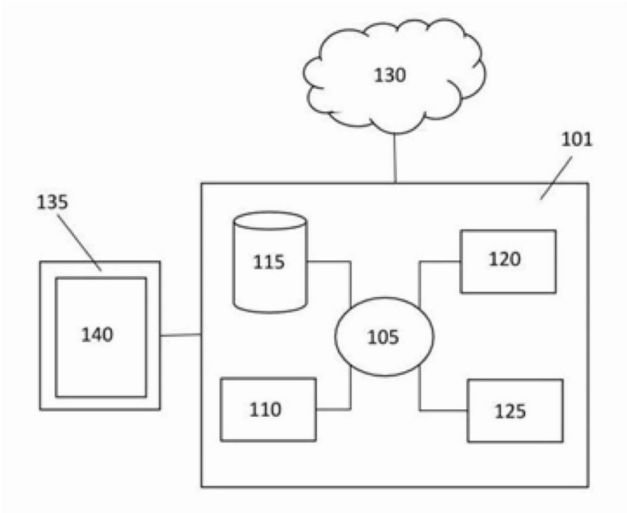
本発明の好ましい実施形態が本明細書中で示され、記述されているが、かかる実施形態はほんの一例として提供されることは当業者に明白であろう。本発明は本明細書内で提供される特定の例により限定されることは意図されていない。本発明は前述の明細書を参照して説明されているが、本明細書中の実施形態の記載と例示は、限定的な意味で解釈されるようには意図されていない。多数の変形、変更、および置換えが、本発明から逸脱することなく当業者によりなされるであろう。さらに、本発明の態様はすべて、様々な条件と変数に依存する、本明細書に明記する特定の描写、構成、または相対的比率に限定されないことを理解されたい。本明細書に記載の本発明の実施形態に対する様々な代替物が、本発明の実施に利用されてよいことを理解されたい。そのため、本発明は、あらゆるこのような代替物、修正、変形、または等価物も包含することが企図される。以下の特許請求の範囲は本発明の範囲を規定し、これら特許請求の範囲およびその等価物内にある方法と構造はそれにより包含されることが意図される。

30

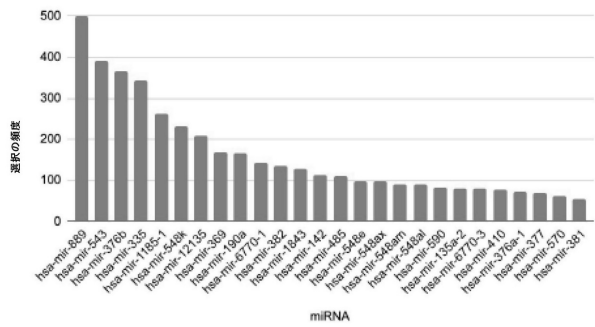
40

50

【図面】
【図 1】

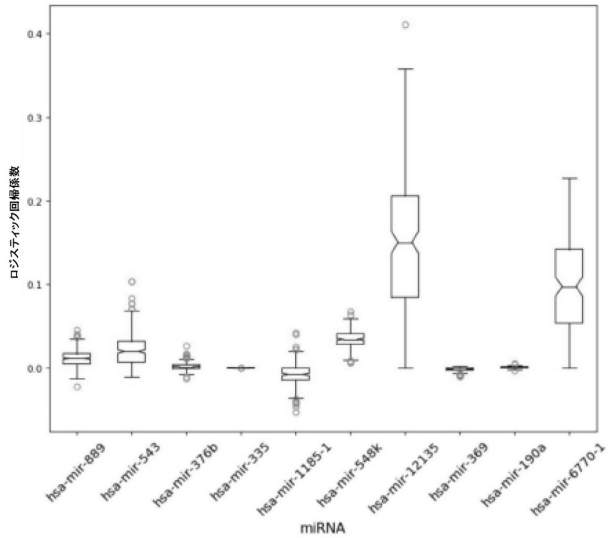


【図 2】



10

【図 3】



20

30

40

50

【国際調査報告】

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US2021/029361

A. CLASSIFICATION OF SUBJECT MATTER

IPC(8) - C12N 15/09; C12N 15/11; C12N 15/113; C12Q 1/68; C12Q 1/6883; C12Q 1/6886 (2021.01)
 CPC - C12N 15/1096; C12N 15/113; C12Q 1/6806; C12Q 1/6851; C12Q 1/686; C12Q 1/6883; C12Q 1/6886; C12Q 2600/178; G01N 33/57419 (2021.08)

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)

see Search History document

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

see Search History document

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

see Search History document

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X — Y	WANG et al. "Novel circulating microRNAs expression profile in colon cancer: a pilot study," Eur J Med Res, 29 November 2017 (29.11.2017), Vol. 22, 51, Pgs. 1-11, entire document	1, 4-6, 9 — 7, 10, 18-25, 30, 32-55, 57
X — Y	MARINER et al. "Improved Detection of Circulating miRNAs in Serum and Plasma Following Rapid Heat/Freeze Cycling," Microma, 16 April 2018 (18.04.2018), Vol. 7, Pgs. 138-147, entire document	28, 31 — 27-30, 32-35
Y	US 2013/0137593 A1 (NISSAN et al) 30 May 2013 (30.05.2013) entire document	7, 54, 55, 57
Y	US 2013/0102492 A1 (GENOMIC HEALTH, INC. et al) 25 April 2013 (25.04.2013) entire document	10, 34, 46
Y	US 2018/0258486 A1 (SRNALYTICS, LLC.) 13 September 2018 (13.09.2018) entire document	18-25, 27-29, 32, 36-53
Y —	BARBERAN-SOLER et al. "Decreasing miRNA sequencing bias using a single adapter and circularization approach," Genome Biol, 03 September 2018 (03.09.2018), Vol. 19, Article number. 105, Pgs. 1-9, entire document	24
Y	US 2011/0217779 A1 (HARRINGTON et al) 08 September 2011 (08.09.2011) entire document	25
Y	US 2005/0221372 A1 (BAUER et al) 06 October 2005 (06.10.2005) entire document	35

☒ Further documents are listed in the continuation of Box C.☐ See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"D" document cited by the applicant in the international application

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search

13 September 2021

Date of mailing of the international search report

OCT 22 2021

Name and mailing address of the ISA/US

Mail Stop PCT, Attn: ISA/US, Commissioner for Patents
 P.O. Box 1450, Alexandria, VA 22313-1450
 Facsimile No. 571-273-8300

Authorized officer

Harry Kim

Telephone No. PCT Helpdesk: 571-272-4300

Form PCT/ISA/210 (second sheet) (July 2019)

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US2021/029361

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	WO 2019/200410 A1 (FREENOME HOLDINGS, INC.) 17 October 2019 (17.10.2019) entire document	36-47, 50-53
Y	US 2018/0010176 A1 (PATEL) 11 January 2018 (11.01.2018) entire document	37

Form PCT/ISA/210 (continuation of second sheet) (July 2019)

フロントページの続き

(51)国際特許分類

F I

テーマコード (参考)

C 1 2 Q 1/6851(2018.01)

C 1 2 Q

1/6851

Z

C 1 2 Q 1/686(2018.01)

C 1 2 Q

1/686

Z

,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,D
K,EE,ES,FI,FR,GB,GR,HR,HU,IE,IS,IT,LT,LU,LV,MC,MK,MT,NL,NO,PL,PT,RO,RS,SE,SI,SK,SM,TR),O
A(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,B
B,BG,BH,BN,BR,BW,BY,BZ,CA,CH,CL,CN,CO,CR,CU,CZ,DE,DJ,DK,DM,DO,DZ,EC,EE,EG,ES,FI,GB,GD
,GE,GH,GM,GT,HN,HR,HU,ID,IL,IN,IR,IS,IT,JO,JP,KE,KG,KH,KN,KP,KR,KW,KZ,LA,LC,LK,LR,LS,LU,
LY,MA,MD,ME,MG,MK,MN,MW,MX,MY,MZ,NA,NG,NI,NO,NZ,OM,PA,PE,PG,PH,PL,PT,QA,RO,RS,
RU,RW,SA,SC,SD,SE,SG,SK,SL,ST,SV,SY,TH,TJ,TM,TN,TR,TT,TZ,UA,UG,US,UZ,VC,VN,WS,ZA,ZM,Z
W

(72)発明者 ハーヴェイ, アダム

アメリカ合衆国 9 4 0 8 0 カリフォルニア州 サウス・サンフランシスコ イースト・グランド
・アベニュー 2 7 9 5 階

(72)発明者 テワリ, アニーシャ

アメリカ合衆国 9 4 0 8 0 カリフォルニア州 サウス・サンフランシスコ イースト・グランド
・アベニュー 2 7 9 5 階

(72)発明者 ワインバーグ, デイビッド

アメリカ合衆国 9 4 0 8 0 カリフォルニア州 サウス・サンフランシスコ イースト・グランド
・アベニュー 2 7 9 5 階

(72)発明者 イートン, ジェシー

アメリカ合衆国 9 4 0 8 0 カリフォルニア州 サウス・サンフランシスコ イースト・グランド
・アベニュー 2 7 9 5 階

F ターム (参考) 4B029 AA07 BB11 BB20 CC01 FA15

4B063 QA19 QQ03 QQ42 QQ52 QS24 QX10