

(19)日本国特許庁(JP)

(12)公表特許公報(A)

(11)公表番号
特表2023-534094
(P2023-534094A)

(43)公表日 令和5年8月8日(2023.8.8)

(51)国際特許分類		F I		テーマコード (参考)	
C 1 2 Q	1/6886(2018.01)	C 1 2 Q	1/6886	Z Z N A	4 B 0 6 3
C 1 2 Q	1/6869(2018.01)	C 1 2 Q	1/6869	Z	4 C 0 8 4
C 1 2 Q	1/686(2018.01)	C 1 2 Q	1/686	Z	
C 1 2 Q	1/6813(2018.01)	C 1 2 Q	1/6813	Z	
A 6 1 P	35/00 (2006.01)	A 6 1 P	35/00		
		審査請求	未請求	予備審査請求	未請求 (全83頁) 最終頁に続く

(21)出願番号	特願2022-566136(P2022-566136)	(71)出願人	522105894 グレイル エルエルシー アメリカ合衆国 カリフォルニア 9 4 0 2 5 , メンロー パーク, オブライエン ドライブ 1 5 2 5
(86)(22)出願日	令和3年6月16日(2021.6.16)	(74)代理人	100147485 弁理士 杉村 憲司
(85)翻訳文提出日	令和4年12月1日(2022.12.1)	(74)代理人	230118913 弁理士 杉村 光嗣
(86)国際出願番号	PCT/US2021/037668	(74)代理人	100175477 弁理士 高橋 林太郎
(87)国際公開番号	WO2021/257729	(72)発明者	ルース イー マウンツ アメリカ合衆国 カリフォルニア州 9 4 0 2 5 メンロー パーク オブライエン ドライブ 1 5 2 5 グレイル インク内
(87)国際公開日	令和3年12月23日(2021.12.23)		
(31)優先権主張番号	63/039,886		
(32)優先日	令和2年6月16日(2020.6.16)		
(33)優先権主張国・地域又は機関	米国(US)		
(81)指定国・地域	AP(BW,GH,GM,KE,LR,LS,MW,MZ,NA ,RW,SD,SL,ST,SZ,TZ,UG,ZM,ZW),EA(AM,AZ,BY,KG,KZ,RU,TJ,TM),EP(AL,A T,BE,BG,CH,CY,CZ,DE,DK,EE,ES,FI,FR ,GB,GR,HR,HU,IE,IS,IT,LT,LU,LV,MC, 最終頁に続く		

(54)【発明の名称】 無細胞RNAの分析方法

(57)【要約】

【要約】無細胞リボ核酸（RNA）分子の亜集団を測定する方法が提供される。いくつかの実施形態では、対象から得られた試験サンプル中の複数のRNA分子からシーケンシングライブラリを生成する方法、および、シーケンシングライブラリを分析して、例えば、疾患の存在または非存在を検出する方法が提供される。

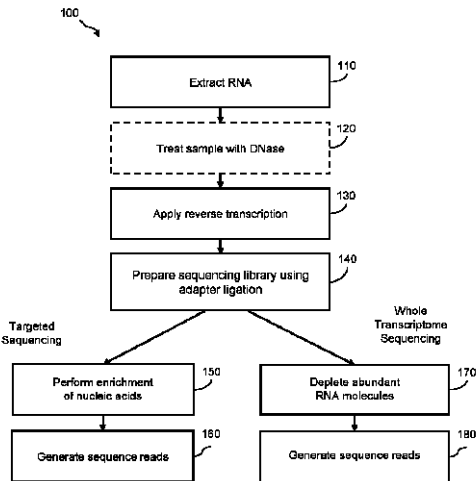


FIG. 1

【特許請求の範囲】

【請求項 1】

対象における癌を検出する方法であって、該方法は、以下を含む。

(a) 対象のサンプル中の複数の標的無細胞RNA (cfRNA) 分子を測定することであって、複数の標的cfRNA分子が表11の1つ以上の転写物から選択されること；及び

(b) 癌を検出することであって、癌を検出することが、閾値以上の標的cfRNA分子の1つ以上を検出することを含む、こと。

【請求項 2】

複数の標的cfRNA分子が、表8または表12～15の1つ以上から選択される、請求項1に記載の方法。

【請求項 3】

複数の標的cfRNA分子が、表8または11～14の1つ以上から少なくとも5、10、15、または20の遺伝子の転写産物から選択される、請求項1に記載の方法。

【請求項 4】

複数の標的cfRNA分子が、(i) 表11；(ii) 表2、5、および12の各々；(iii) 表3、4、6、および13の各々；(iv) 表14；または(v) 表8からの複数の転写物を含んでいる、請求項1に記載の方法。

【請求項 5】

複数の標的cfRNA分子が、表8または表11～15の1つ以上から少なくとも30個の遺伝子の転写物を含む、請求項1に記載の方法。

【請求項 6】

複数の標的cfRNA分子が、表14の転写物から選択される、請求項1に記載の方法。

【請求項 7】

閾値を超えて検出された複数の標的cfRNA分子が、以下からなる群から選択される複数の遺伝子に由来するcfRNA分子である、請求項1記載の方法。adipoq、agr3、ankrd30a、aqp4、bpifa1、ca12、ceacam5、cftr、cxcl17、cyp4f8、fabp7、foxy1、ggtlc1、gp2、IL20、Itih6、ldlr1、lemd1、lmx1b、mmp7、NKAIN1、NKX2-1、ROPN1、ROS1、SCGB1D2、SCGB2A2、SFTA2、SFTA3、SLC34A2、SOX9、STK32A、STMND1、TFAP2A、TFAP2B、TFF1、TRPV6、VGLL1および VTCN1。

【請求項 8】

複数の標的cfRNA分子が表8の転写物から選択される、請求項1に記載の方法。

【請求項 9】

閾値を超えて検出された複数の標的cfRNA分子が、以下からなる群より選択される複数の遺伝子に由来するcfRNA分子である、請求項1記載の方法。ceacam5、rhov、sfta2、scgb1d2、igf2bp1、sftpa1、ca12、sftpb、cdh3、muc6、src6a14、hoxc9、agr3、tmem125、tfap2b、irx2、potekp、agr3、tmem126、tfap2b、irx2、potekp、ARHGEF38、GPR87、LMX1B、ATP10B、NEL1、MUC21、SOX9、LINC00993、STMND1、ERVH48-1、SCTR、MAGEA3、MB、LEMD1、SIX4およびNXNL2。

【請求項 10】

複数の標的cfRNA分子が、(a) 表8または11～14の1つ以上の転写物、および(b) 表1～6の1つ以上の転写物からなる、請求項1に記載の方法。

【請求項 11】

複数の標的cfRNA分子が、(a) 表8または11～14の1つまたは複数の転写物、および(b) 表7の1つまたは複数の転写物を含む、請求項1に記載の方法。

【請求項 12】

i) 癌が肺癌であり、(ii) 閾値を超えて検出された複数の標的cfRNA分子が、表2、表5、または表12の1つ以上の転写産物から選択される、請求項1に記載の方法。

【請求項 13】

10

20

30

40

50

i) 癌が乳癌であり、(ii) 閾値を超えて検出された複数の標的cfRNA分子が、表3、4、6、又は13の1つ以上の転写産物から選択される、請求項1に記載の方法。

【請求項14】

測定が、配列決定、マイクロアレイ解析、逆転写PCR、リアルタイムPCR、定量リアルタイムPCR、デジタルPCR、デジタル液滴PCR、デジタルエマルジョンPCR、多重PCR、ハイブリッド捕捉、オリゴヌクレオチドライゲーションアッセイ、またはそれらの任意の組み合わせからなる、請求項1～13のいずれか一項に記載の方法。

【請求項15】

測定が、cfRNA配列リードを生成するためにcfRNA分子を配列決定することを含む、請求項1～13のいずれか一項に記載の方法。

【請求項16】

cfRNA分子の配列決定が、全トランスクリプトーム配列決定からなる、請求項15に記載の方法。

【請求項17】

cfRNA分子を配列決定することが、cDNA分子を生成するための逆転写、およびcDNA分子を配列決定してcfRNA配列リードを生成することを含む、請求項15に記載の方法。

【請求項18】

cfRNA分子を配列決定することが、標的cfRNA分子またはそのcDNA分子を濃縮することを含む、請求項15に記載の方法。

【請求項19】

試料が生体液からなる、請求項1～13のいずれか1項に記載の方法。

【請求項20】

生体液が、血液、血漿、血清、尿、唾液、胸水、心嚢液、脳脊髄液(CSF)、腹膜液、またはそれらの任意の組み合わせからなる、請求項19に記載の方法。

【請求項21】

生体物質が、被験者の血液、血液分画、血漿、または血清からなる、請求項19に記載の方法。

【請求項22】

閾値以上の標的cfRNA分子の1つ以上を検出することが、(i) 検出、(ii) バックグラウンド以上の検出、または(iii) 状態を有しない対象における標的cfRNA分子のレベルよりも大きいレベルでの検出を含む、請求項1～13のいずれか一項に記載の方法。

【請求項23】

閾値以上の1つ以上の標的cfRNA分子を検出することが、該状態を有しない被験者のレベルよりも少なくとも10倍大きいレベルで1つ以上の標的cfRNA分子を検出することを含む、請求項1～13のいずれか一項に記載の方法。

【請求項24】

標的cfRNA分子の1つ以上を閾値以上検出することが、0.5～5reads per million (RPM) の閾値以上検出することを含む、請求項15に記載の方法。

【請求項25】

閾値以上の標的cfRNA分子の1つ以上を検出することが、以下を含む、請求項1～13のいずれか一項に記載の方法である。

(a) 標的cfRNA分子の各々の発現レベルをRNA組織スコアマトリックスと比較することにより、各標的cfRNA分子について指標スコアを決定すること；及び

(b) 各ターゲットcfRNA分子の指標スコアを集計すること；及び

(c) 指標となるスコアが閾値を超えた場合に、がんを検出すること。

【請求項26】

閾値以上の標的cfRNA分子の1つ以上を検出することが、配列読み取りを機械学習又は深層学習モデルに入力することを含む、請求項15に記載の方法。

【請求項27】

機械学習または深層学習モデルが、ロジスティック回帰、ランダムフォレスト、勾配ブ

10

20

30

40

50

ースティング機械、ナイーブベイズ、ニューラルネットワーク、または多項回帰からなる、請求項26に記載の方法。

【請求項28】

前記機械学習又は深層学習モデルは、学習された重みを含む関数を通じて、前記1つ又は複数の特徴の値を前記被験者の疾患状態予測に変換する、請求項26に記載の方法。

【請求項29】

癌が以下のものを含む、請求項1～13のいずれか一項に記載の方法。

(i) 癌腫、肉腫、骨髄腫、白血病、リンパ腫、芽球腫、胚細胞腫瘍、またはそれらの任意の組合せ。

(ii) 腺癌、扁平上皮癌、小細胞肺癌、非小細胞肺癌、上咽頭癌、大腸癌、肛門癌、肝臓癌、膀胱癌、精巣癌、頸部癌、卵巣癌、胃癌、食道癌、頭頸部癌、すい臓癌、前立腺癌、腎癌、甲状腺癌、メラノーマおよび乳癌からなる群より選ばれた癌腫；

(iii) ホルモン受容体陰性乳癌またはトリプルネガティブ乳癌。

(iv) 骨肉腫、軟骨肉腫、平滑筋肉腫、横紋筋肉腫、中皮肉腫（中皮腫）、線維肉腫、血管肉腫、脂肪肉腫、グリオーマ、及び星細胞腫からなる群から選択される肉腫。

(v) 骨髄性、顆粒球性、リンパ性、リンパ球性、およびリンパ芽球性白血病からなる群から選択される白血病；または

(vi) からなる群より選択される リンパ腫。ホジキンリンパ腫および非ホジキンリンパ腫。

【請求項30】

癌を検出することが、癌の病期の決定、癌の進行の決定、癌のタイプの決定、癌の起源組織の決定、またはそれらの組み合わせを含む、請求項1～13のいずれか一項に記載の方法。

【請求項31】

検出された癌に基づいて治療法を選択することをさらに含む、請求項1～13のいずれか一項に記載の方法。

【請求項32】

治療が外科的切除、放射線療法、または抗癌剤の投与からなる、請求項31に記載の方法。

【請求項33】

本方法が、選択された治療で対象を治療することをさらに含む、請求項31に記載の方法。

【請求項34】

試料中の複数の標的無細胞RNA（cfRNA）分子を測定する方法であって、該方法は、以下を含む。

(a) 複数の標的cfRNA分子、またはそのcDNA分子を濃縮して、ポリヌクレオチドの濃縮試料を生成すること；および

(b) 該濃縮サンプルのポリヌクレオチド、またはその増幅産物を配列決定する工程。

ここで、複数の標的cfRNA分子は、表11の1つまたは複数の転写物から選択される。

【請求項35】

複数の標的cfRNA分子が、表8または表12～15の1つ以上から選択される、請求項34に記載の方法。

【請求項36】

複数の標的cfRNA分子が表8から選択される、請求項34に記載の方法。

【請求項37】

複数の標的cfRNA分子が表14から選択される、請求項34に記載の方法。

【請求項38】

複数の標的cfRNA分子が、表8または11～14の1つ以上から少なくとも5、10、15、または20の遺伝子の転写物から選択される、請求項34に記載の方法、。

【請求項 39】

請求項1～38のいずれか1項に記載の方法における1つまたは複数のステップを実施するためのコンピュータシステム。

【請求項 40】

請求項1～38のいずれか1項に記載の方法における1つ以上のステップを実施するためのコンピュータ可読命令をその上に格納した、非一時的な、コンピュータ可読媒体。

【発明の詳細な説明】

【関連出願へのクロスレファレンス】

【0001】

本出願は、2020年6月16日に提出された米国仮出願第63/039,886号の利益を主張し、この出願は、すべての目的のためにその全体が参照により本明細書に組み込まれるものとする。

【発明の背景】

【0002】

2017年現在、米国では毎年160万人以上の新規症例があり、がんは世界的に著名な公衆衛生問題を表している。参照：Siegel et al., 2017, "Cancer statistics," CA Cancer J Clin.67(1):7-30.スクリーニングプログラムと早期診断は、がん患者の無病生存率の向上と死亡率の減少に重要な影響を及ぼす。早期診断のための非侵襲的なアプローチは患者のコンプライアンスを促進するため、スクリーニングプログラムに含めることができる。

【0003】

無細胞核酸 (cfNA) は、血清、血漿、尿、その他の体液中に存在する (Chan et al., "Clinical Sciences Reviews Committee of the Association of Clinical Biochemists Cell-free nucleic acids in plasma, serum and urine: a new tool in molecular diagnosis," Ann Clin Biochem.2003;40(Pt 2):122-130) は、特定の疾患の循環像である「リキッドバイオプシー」を表している。参照：De Mattos-ArrudaおよびCaldas, 2016, "Cell-free circulating tumour DNA as a liquid biopsy in breast cancer," Mol Oncol.2016;10(3):464-474.同様に、無細胞RNAもがん検出のための分析物質として提案されている。Tzimagiorgis, et al., "Recovering circulating extracellular or cell-free RNA from bodily fluids," Cancer Epidemiology 2011; 35(6):580-589 を参照のこと。これらのアプローチは、癌などの様々な疾患のスクリーニングのための非侵襲的な方法の可能性を示している。

【0004】

しかしながら、がんは依然として世界中で頻繁に起こる死因の一つである。過去数十年の間に、治療の選択肢は改善されたが、生存率は依然として低いままである。外科的切除や薬物療法による治療の成功は、早期の腫瘍の同定に強く依存している。しかし、画像診断やバイオマーカーを用いたアプローチなどの現在の技術では、病期が進行するまで腫瘍を特定できないことが多い。

【0005】

そのため、治療が成功する可能性が高い初期の段階で疾患を特定することができる新しい非侵襲的な検出方法が依然として求められている。本発明は、このようなニーズやその他のニーズを満たすものである。

【発明の概要】

【0006】

様々な実施形態によれば、現在開示されている主題は、試料中の複数の標的無細胞RNA (cfRNA) 分子を測定する方法を提供する。実施形態では、方法は、(a) 複数の標的cfRNA分子、又はそのcDNA分子を濃縮して、ポリヌクレオチドの濃縮サンプルを生成すること；及び／又は (b) 濃縮サンプルのポリヌクレオチド、又はその増幅産物を配列決定することを含み、ここで複数の標的cfRNA分子は表11の1又は複数の転写物から

10

20

30

40

50

選択することが可能である。実施形態では、複数の標的cfRNA分子は、表8もしくは12～15の1つ以上、またはそれらの任意の組み合わせ、（例えば、表8もしくは11～14の1つ以上から5、10、15、または20の遺伝子の転写物）から選択することができる。

【0007】

いくつかの態様において、主題の開示は、対象における癌を検出する方法を提供する。実施形態では、方法は、（a）対象のサンプル中の複数の標的無細胞RNA（cfRNA）分子を測定することであって、複数の標的cfRNA分子が表11の転写物から選択されること；及び／又は（b）癌を検出することであって、癌を検出することが、標的cfRNA分子の一つ以上を閾値を超えて検出することから構成されることを含む。実施形態において、閾値を超える標的cfRNA分子の1つ以上を検出することは、（i）検出、（ii）バックグラウンド以上の検出、及び／又は（iii）状態を有さない対象における対応する配列読み取りのレベルよりも大きいレベルでの検出を含む。実施形態では、複数の標的cfRNA分子は、表8又は12～15の1つ以上から選択される（例えば、表8又は11～14の1つ以上から5、10、15又は20個の遺伝子の転写物）。

10

【0008】

様々な態様において、本開示は、被験者の疾患状態を検出するための方法及び組成物を提供する。実施形態では、方法は、無細胞リボ核酸（cfRNA）中の1つまたは複数のマーカーを検出することを含む。実施形態において、cfRNAを検出することは、対象からの生体試料からcfRNAを配列決定して、cfRNAリードを生成することを含む。実施形態において、本方法は、被験体の細胞からのRNAを配列決定して細胞リードを生成することと、cfRNAリードをフィルタリングして1つまたは複数の細胞リードに対応するcfRNAリードを除外することとをさらに含む。実施形態では、細胞は、血球である。実施形態において、方法は、1つまたは複数のリボソーム、ミトコンドリア、および／または血液関連転写物を除外するために、cfRNAリードをフィルタリングすることを含んでいる。実施形態では、エクソン-エクソン接合部に重なるcfRNAのリード（またはリードペア）のみが測定される。実施形態では、1つ以上のマーカーに対応するcfRNAが測定される（例えば、1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、20、25、またはそれ以上のマーカー）。1つ以上のマーカーは、本明細書に開示されるマーカーのいずれかであり、任意の組み合わせであり得る。実施形態において、1つ以上のマーカーは、疾患状態に関連する。実施形態において、方法は、被験者の疾患状態を治療することを含む。

20

30

【0009】

本開示の態様は、被験体における疾患状態を検出するための方法を含み、該方法は、以下を含む。対象から生体検査試料を分離することであって、生体検査試料が複数の無細胞リボ核酸（cfRNA）分子を含む、ことと、生体検査試料から複数のcfRNA分子を抽出することと、抽出されたcfRNA分子に配列決定手順を実行して複数の配列リードを生成することと、からなる方法である。フィルタリング手順を実行して、1つ又は複数の健康な細胞に由来する配列リードの除外集団、及び配列リードの非除外集団を生成すること；非除外配列リードに対して定量化手順を実行すること；及び定量化手順が閾値を超える値を生成したときに、対象における疾患状態を検出すること。実施形態において、閾値を超える1つ以上の非除去配列リードを検出することは、（i）検出、（ii）バックグラウンド以上の検出、又は（iii）状態を有しない対象における対応する配列リードのレベルよりも大きいレベルでの検出を含む。

40

【0010】

開示された主題の側面は、疾患状態を示す1つまたは複数のRNA配列を同定するためのコンピュータ実装方法をさらに含み、本方法は以下を含む。コンピュータシステムによって、疾患を有することが知られている対象からの第1の試験試料から複数のRNA分子から配列読み取り値の第1のセットを得ることであって、第1の試験試料が複数の無細胞RNA（cfRNA）分子を含む、ことと、コンピュータシステムによって、対照試料から複数のRNA分子から配列読み取り値の第2のセットを得ることである。コンピュータシステムに

50

よって、配列読み取り値の第1のセットに存在し、配列読み取り値の第2のセットに存在しない1つまたは複数のRNA配列を検出し、疾患状態を示す1つまたは複数のRNA配列を同定するステップ。

【0011】

他の態様において、主題は、対象における1つまたは複数の腫瘍由来RNA分子を検出するためのコンピュータ実装方法に向けられ、本方法は以下を包含する。コンピュータシステムによって、腫瘍を有することが知られている、または有することが疑われる対象からの第1の試験試料から複数のRNA分子から配列読み取り値の第1のセットを得ることであって、第1の試験試料が複数の無細胞RNA (cfRNA) 分子を含んでいる、ステップとコンピュータシステムにより、対象からの複数の血液細胞からの複数のRNA分子から第2の配列読み取りセットを取得し；コンピュータシステムにより、第1の配列読み取りセットに存在し、第2の配列読み取りセットに存在しない1つ以上のRNA配列を検出して、対象における1つ以上の腫瘍由来RNA分子を検出すること；。

10

【0012】

いくつかの態様において、開示された主題は、癌を有することが知られている又は癌を有することが疑われる対象において、癌の存在を検出し、癌の病期を決定し、癌の進行を監視し、及び／又は癌タイプ若しくは癌サブタイプを決定する方法に向けられており、該方法は、以下の工程を包含する。(a) 対象から生物学的試験サンプルを得ることであって、生物学的試験サンプルが複数の無細胞リボ核酸 (cfRNA) 分子を含むこと；(b) 生物学的試験サンプル中の1つ以上の標的RNA分子由来の1つ以上の核酸配列の存在を定量的に検出して腫瘍RNAスコアを決定し、ここで1つ以上の標的RNA分子は表8又は11～14のいずれか一つに記載された標的RNA分子から選択されること；(b) 対象から生物学的試験サンプルを得ること、ここで対象が、腫瘍RNA分子が表10又は表11～14に記載の分子であること。及び／又は(c) 腫瘍RNAスコアが閾値を超えた場合に、対象における癌の存在を検出すること、癌の病期を決定すること、癌の進行を監視すること、及び／又は癌のタイプ若しくはサブタイプを決定することを含む。

20

【0013】

様々な態様によれば、本開示は、対象における癌の存在を検出するためのコンピュータ実装方法であって、プロセッサ及びコンピュータ可読媒体を含むコンピュータにおいてデータセットを受け取ることであって、データセットが、対象からの生体検査サンプル中の複数のリボ核酸 (RNA) 分子から得られた複数の配列読み取りを含み、コンピュータ可読媒体が、プロセッサによって実行されるとコンピュータに以下を行わせる命令を含む方法に向けられている：コンピュータ可読媒体が、プロセッサに、以下を行わせる命令を含むことであって、コンピュータ可読媒体が、プロセッサに、以下を行わせる命令、及び、プロセッサが、コンピュータ可読媒体を含むことであって、コンピュータ可読媒体が、コンピュータ可読媒体が、プロセッサに、以下を行わせる命令、及び、コンピュータ可読媒体によって、コンピュータ可読媒体を含むことを特徴とする方法生体検査サンプル中の複数の標的RNA分子の発現レベルを決定すること；複数の標的RNA分子の各々の発現レベルをRNA組織スコア行列と比較して、複数の標的RNA分子の各々の癌指標スコアを決定すること；複数の標的RNA分子の各々の癌指標スコアを集約して、生体検査サンプルの癌指標スコアを生成すること；及び／又は生体検査サンプルの癌指標スコアが閾値を超えると、対象における癌の存在を検出すること；をコンピュータに行わせる命令を含む。

30

40

【0014】

様々な態様において、本開示の主題は、RNA組織スコアマトリックスを構築するための方法に向けられ、本方法は、複数の被験者から得られた複数のRNA配列読み取りをコンパイルしてRNA発現マトリックスを生成すること；及びRNA発現マトリックスを組織特異的RNA発現マトリックスで正規化してRNA組織スコアマトリックスを構築すること；を含む。いくつかの実施形態では、RNA配列リードは、既知の癌タイプを有する複数の被験者から取得され、癌RNA組織スコア行列を構築する。

【0015】

50

いくつかの態様において、現在開示されている主題は、被験体の無細胞RNA (cfRNA) 分子の垂集団を測定する方法を提供する。実施形態において、本方法は、(a) cfRNA 分子を配列決定してcfRNA配列読み取りを生成すること；(b) 対象の細胞から抽出した細胞RNAを配列決定して細胞配列読み取りを生成すること；(c) フィルタリング手順を実行してcfRNA配列読み取りの非除外集団を生成し、ここでフィルタリングは細胞配列読み取りの1つ以上に一致するcfRNA配列読み取りの除外を含む；及び／又は(d) 非除外配列読み取りの1つ以上を定量化すること、を備える。

【0016】

いくつかの態様において、本開示は、1人以上の被験者から収集された試料における癌バイオマーカー（本明細書では「マーカー」とも称される）を同定する方法を提供する。実施形態において、本方法は、(a) 癌でない被験者から収集した生体液のcfRNAを配列決定して、非癌配列決定リードを生成すること；(b) 癌を有する1人以上の被験者から収集した複数のマッチした試料について、以下の工程を包含する。(i)一致したサンプルの癌組織から収集したDNAおよびRNAを配列決定して、癌組織についての配列決定リードを生成し；(ii)一致したサンプルの一致した生体液から収集したcfDNAおよびcfRNAを配列決定して、一致した生体液についての配列決定リードを生成し；(iii)一致した生体液についてのcfDNA配列決定リードのカウントと癌組織についてのDNA配列決定リードの対応カウントとを関連付けることによって腫瘍率を測定するステップ。及び／又は(iv) 1つ以上の候補バイオマーカーのRNAシーケンシングリードのカウントを腫瘍分率に乘じることによって、1つ以上の候補バイオマーカーの腫瘍含量を測定し、ここで、1つ以上の候補バイオマーカーは、癌ではない対象から採取した生物液中よりもマッチした生物液中で高いレベルで発現される。(c) 腫瘍含有量を共変数として用いて、cfRNAにおける1つ又は複数の候補バイオマーカーの発現をモデル化すること、及び／又は(d) モデル化に基づいて、1つ又は複数の候補バイオマーカーの中から1つ又は複数のcfRNA癌バイオマーカーを同定すること。

【0017】

いくつかの態様において、本開示は、本明細書に開示される様々な態様のいずれかの方法における1つまたは複数のステップを実施するためのコンピュータシステムを提供する。

【0018】

いくつかの態様において、現在開示されている主題は、本明細書に開示されている様々な態様のいずれかの方法における1つまたは複数のステップを実施するためのコンピュータ可読命令をその上に格納した、非一過性のコンピュータ可読媒体を提供する。

【図面の簡単な説明】

【0019】

【図1】図1は、一実施形態による配列決定用核酸サンプルの調製方法のフローチャートである。

【図2】図2は、本発明の一実施形態に係る、疾患状態を示す1つ以上のRNA配列を同定する方法を示すフロー図である。

【図3】図3は、本発明の一実施形態に係る、1つ以上の腫瘍由来RNA配列を同定する方法を示すフロー図である。

【図4】図4は、本発明の一実施形態による、対象における癌の存在の検出、癌の状態の決定、癌の進行の監視、及び／又は癌の種類の決定のための方法を示す流れ図である。

【図5】図5は、本発明の一実施形態による、1つ以上の標的RNA分子に由来する1つ以上の配列リードから疾患状態を検出する方法を示すフロー図である。

【図6】図6は、本発明の一実施形態に係る、癌指標スコアに基づいて対象者の癌の存在を検出する方法を示すフロー図である。

【図7】図7は、実施形態による、サンプル分類スキームの感度及び特異度に関する結果の例を示す図である。

【図8】図8A～Cは、実施形態による、サンプル分類スキームの感度及び特異度に関す

10

20

30

40

50

る結果の例を示す図である。

【図 9】図 9 は、肺がんにおける 20 のダークチャンネル遺伝子の発現量を、がん性サンプルと非がん性サンプルとの発現量比が最も高いものとして描いたものである。Reads per million (RPM) は、ダークチャンネル遺伝子の関数としてプロットされている。各プロットにおいて、左から右への点の列は、それぞれ左から右への上の凡例で示されたグループに対応する（クラス、肛門、乳房、大腸、肺、および非がん）。

【図 10】図 10 は、ダークチャンネル遺伝子から集計された組織スコアを用いた決定木分類器の ROC 曲線である。

【図 11】図 11 は、いくつかの実施形態に従った方法を示すフローチャートである。

【図 12 A】図 12 A は、ステージ III の TCGA (The Cancer Genome Atlas) FFPE (ホルマリン固定パラフィン包埋) 組織 RNA-seq データの PCA (主成分分析) の一例の散布図である。遺伝子発現量は、read per million でプロットされている。

10

【図 12 B】図 12 B は、CCGA (Circulating Cell-free Genome Atlas) 腫瘍組織 RNA-seq データを TCGA PCA 軸に投影した結果の一例を示す散布図である。遺伝子発現レベルは、100 万回当たりの読み取り値でプロットされている。

【図 12 C】図 12 C は、CCGA 癌セルフリー RNA (cfRNA) RNA-seq データを TCGA PCA 軸に投影した結果の一例を示す散布図である。遺伝子発現レベルは、百万分の一の読み取り値でプロットされている。

【図 13】図 13 は、例示的なダークチャンネルバイオマーカー遺伝子のヒートマップである。各列は、1 つの cfRNA サンプルを描写し、各行は、1 つの遺伝子を描写する。行の色は、組織特異性をコードする（上から下へ、組織は、それぞれ、乳房、肺、および非特異的である）。列の色はサンプルグループを表している（左から右へ、それぞれ肛門癌、乳癌、大腸癌、肺癌、非癌である）。

20

【図 14 A】図 14 A は、異なるサンプルにおける cfRNA 発現レベル及び 2 つの例示的な乳房暗電流バイオマーカー (DCB) 遺伝子 (FABP7 及び SCGB2A2) の組織発現レベルを描写する箱ひげ図である。HER2+、HR+ / HER2-、トリプルネガティブ乳癌 (TNBC)、又は非癌のサンプルにおける cfRNA 発現量及び組織発現量を示す箱ひげ図である。

【図 14 B】図 14 B は、異なるサンプル：腺癌、小細胞肺癌、扁平上皮癌、又は非癌サンプルにおける cfRNA 発現レベル及び 4 つの例示的肺 DCB 遺伝子 (SLC34A2、ROS1、SFTPA2、及び CXCL17) の組織発現レベルを描いた箱ひげ図である。

30

【図 15 A】図 15 A は、マッチした腫瘍組織を有する乳癌サンプルについての 2 つの乳房 DCB 遺伝子 (FABP7 および SCGB2A2) の検出性を描写するフォレストプロットを示す。サンプル ID は、無細胞 DNA (cfDNA) 中の相対的な腫瘍割合 (95% CI) に基づいてプロットされている。FABP7 は、4653、4088、2037、3116、1202 の各検体で検出された。SCGB2A2 は、1656、2419、3911、2367、2037、1039、2139、3162 の各検体で検出された。cfDNA 中の腫瘍分画は、cfDNA 濃縮アッセイによる SNV allele 分画から測定した。

【図 15 B】図 15 B は、マッチした腫瘍組織を有する乳癌サンプルについての 2 つの乳房 DCB 遺伝子 (FABP7 および SCGB2A2) の検出性を描写するフォレストプロットを示す。サンプル ID は、腫瘍含有量 (腫瘍分率 * 腫瘍組織発現) の関数としてプロットされている。FABP7 は、サンプル 4088、1202、3116、2037 で検出された。SCGB2A2 は、1656、2419、2367、3911、1039、2139、3162、2037 の各検体で検出された。cfDNA 中の腫瘍分画は、cfDNA 濃縮アッセイで得られた SNV allele 分画から測定した。組織発現は、マッチした腫瘍組織の RNA-seq データから測定した。

40

【図 16】図 16 A ~ D は、乳癌、肺癌、または癌がない (正常) 被験者について、示された遺伝子についての cfRNA およびマッチした組織における DCB 遺伝子発現についての配列決定結果の例を示す図である。リードカウントの数は、Y 軸に表される。

【図 17】図 17 A - B は、分類器のワークフローの例を示す図である。

【図 18】図 18 A ~ C は、例示的な分類スキームの感度および特異性を示す ROC プロットである。

50

【図 19】図 19 は、本発明の一実施形態に係る試料処理方法及びパラメータ決定方法を示す図である。

【図 20】図 20 A-B は、実施形態に従った、選択された乳房及び肺特異的バイオマーカの分布を示し、乳癌及び肺癌由来（それぞれ）の cfRNA において、非癌由来の cfRNA に対してシグナルが増加していることを示す。乳癌、肺癌、及び非癌の CCGA 参加者の cfRNA から全トランスクリプトームサンプルを調製した。

【図 21】図 21 は、全トランスクリプトーム CCGA 乳がんサンプルの血漿と組織の遺伝子発現をマッチングさせたものである。この結果から、組織での高発現は必ずしも血漿への高流出率をもたらすとは限らないことがわかる。

【図 22】図 22 は、乳癌について、CCGA 血漿中のダークチャンネル発現が CCGA 腫瘍組織発現と相関していることを示す散布図である。平均血漿または組織発現がゼロである遺伝子は、ここでは視覚化のために $1e-4$ に変換されている。

【図 23】図 23 は、CCGA 血漿中のダークチャンネル発現が、肺癌の CCGA 腫瘍組織発現と相関していることを示す散布図である。平均血漿または組織発現がゼロである遺伝子は、ここでは視覚化のために $1e-4$ に変換されている。

【図 24】図 24 は、CCGA 血漿試料における腫瘍特異的マーカーを示すグラフである。血漿対数オッズ比は、全ての癌血漿から全ての非癌血漿への観察に基づいて、各遺伝子について計算された。示された遺伝子は、例示的なダークチャンネルバイオマーカーを示す。

【図 25】図 25 は、表 15 の cfRNA バイオマーカの分布を、ソース及び同定方法によってグループ分けして示すベン図である。図中の全てのグループ分けに存在する 38 のバイオマーカーは、表 14 に提供されている。遺伝子は、バイナリ検出のための最適化および tissue-of-origin (TOO) のための最適化のためにフィルタリングされる。バイナリ検出のための最適化のためにフィルタリングされた遺伝子は、CCGA 血漿においてログオッズ比 > 0.1 で観察され、乳癌および肺癌において TCGA 高発現 (> 5 RPM) 遺伝子であった。TOO の最適化のためにフィルタリングされた遺伝子は、TCGA の組織から multi class random forest 法で選択された遺伝子と、Human Protein Atlas で乳がん/肺がんまたは組織特異的とアノテーションされた遺伝子であった。

【図 26】図 26 A~D は、実施形態に従って、非癌被験者と比較して、乳癌及び／又は肺癌において検出される選択されたバイオマーカのレベルを示す。結果は、乳癌及び／又は肺癌由来（それぞれ）の cfRNA において、非癌由来の cfRNA と比較してシグナルが増加したことを示す。乳癌、肺癌、および非癌の CCGA 参加者の cfRNA から全トランスクリプトームサンプルを調製した。

【発明の詳細な説明】

【0020】

本発明をより詳細に説明する前に、本発明は、当然ながら異なり得るので、説明された特定の実施形態に限定されないことが理解されるであろう。また、本発明の範囲は添付の特許請求の範囲によってのみ限定されるので、本明細書で使用される用語は、特定の実施形態を説明する目的のためだけであり、限定することを意図していないことを理解されたい。

【0021】

値の範囲が提供される場合、文脈上明らかに他に指示されない限り、その範囲の上限と下限との間の、下限の単位の 10 分の 1 までの各介入値、およびその述べられた範囲内の他の述べられたまたは介在する値、ならびにその範囲の提供された各終点は、本発明の中に包含されることが理解される。これらの小さい範囲の上限および下限は、記載された範囲内の特に除外された限界値に従い、独立して本発明内に包含される小さい範囲に含まれることができる。

【0022】

他に定義されない限り、本明細書で使用される技術用語および科学用語は、この発明が属する技術分野における通常の当業者によって一般的に理解されるのと同じ意味を有する。Singleton ら、Dictionary of Microbiology and Molecular Biology 2nd ed

10

20

30

40

50

., J. Wiley & Sons (New York, NY 1994) は、当業者に本願で使用する多くの用語の一般的ガイドを提供し、以下のように、その各々は参照により全体が本明細書に組み込まれるものとする。Kornberg and Baker, DNA Replication, Second Edition (W.H. Freeman, New York, 1992); Lehninger, Biochemistry, Second Edition (Worth Publishers, New York, 1975); Strachan and Read, Human Molecular Genetics, Second Edition (Wiley-Liss, New York, 1999); Abbas et al, Cellular and Molecular Immunology, 6th edition (Saunders, 2007)である。

【0023】

本明細書で言及したすべての出版物は、その出版物が引用されている方法および／または材料に関連して開示および説明するために、参照により本明細書に明示的に組み込まれる。

【0024】

ポリヌクレオチド、核酸、オリゴヌクレオチドという用語は、互換的に使用される。これらは、任意の長さのヌクレオチド、デオキシリボヌクレオチドまたはリボヌクレオチドのいずれか、またはそれらのアナログの重合体を指す。ポリヌクレオチドは、任意の三次元構造を有していてもよく、既知または未知の任意の機能を果たすことができる。以下は、ポリヌクレオチドの非限定的な例である。遺伝子または遺伝子断片のコード領域または非コード領域、連鎖分析から定義される遺伝子座 (locus)、エクソン、イントロン、メッセンジャーRNA (mRNA)、転移RNA (tRNA)、リボソームRNA (rRNA)、ショートインターファーリングRNA (siRNA)、ショートヘアピンRNA (shRNA)、ショートヘアピンRNA (shRNA)、マイクロRNA (miRNA)、リボザイム、cDNA、組み換えポリヌクレオチド、分岐ポリヌクレオチド、プラスミド、ベクター、任意の配列の分離DNA、任意の配列の分離RNA、核酸プローブおよびプライマーを含む。ポリヌクレオチドは、メチル化ヌクレオチドおよびヌクレオチド類似体などの1つまたは複数の修飾ヌクレオチドを含んでいてもよい。存在する場合、ヌクレオチド構造への修飾は、ポリマーのアセンブリの前または後に付与されてもよい。ヌクレオチドの配列は、非ヌクレオチド成分によって中断されてもよい。ポリヌクレオチドは、標識成分とのコンジュゲーションなど、重合後にさらに修飾されてもよい。

【0025】

一般に、「標的ポリヌクレオチド」という用語は、標的配列を有する核酸分子の出発集団中の核酸分子またはポリヌクレオチドを指し、その存在、量、ヌクレオチド配列、またはこれらの1つ以上の変化を決定することが望まれる。一般に、「標的配列」という用語は、核酸の一本鎖上の核酸配列を指す。標的配列は、遺伝子の一部、制御配列、ゲノムDNA、cDNA、mRNA、miRNA、rRNAを含むRNA、またはその他であってもよい。標的配列は、試料からの標的配列であってもよいし、増幅反応の産物のような二次的な標的であってもよい。

【0026】

マーカー」及び「バイオマーカー」という用語は、特定の生物学的状態（例えば、一般的ながんの存在、又は特定のがんの種類及び／若しくは段階などの疾患状態）と関連するレベル又は濃度を有するポリヌクレオチド（例えば、遺伝子又はその識別可能な配列断片）を指すために本明細書において互換的に使用されている。実施形態において、マーカーは、特定の遺伝子のcfRNAであり、そのレベルの変化は、配列決定によって検出され得る。cfRNAバイオマーカーは、本明細書において、cfRNAが由来する遺伝子を参照して言及され得るが、遺伝子転写物全体の検出を必要とするものではない。実施形態では、特定の遺伝子転写物の断片のみが検出される。実施形態において、特定の遺伝子の存在及び／又はレベルを検出することは、同じ遺伝子の転写物に由来する異なる配列断片（重複又は非重複）を含む1つ又は複数のcfRNA断片を検出することを含み、これらは同じ「バイオマーカー」の一部としてまとめてスコア化されてもよい。配列情報（例えば、DNA、RNA、およびアミノ酸配列）、遺伝子記号の方法で一般に同定される遺伝子の完全名称な

どを含む、引用された遺伝子指定に関連する追加の情報は、当業者に周知の公にアクセス可能なデータベース、例えばGenBank (www.ncbi.nlm.nih.gov/genbank/) およびNCBI Proteinデータベース (www.ncbi.nlm.nih.gov/protein/)、およびUniProt (www.uniprot.org) などのNational Center for Biotechnology Informationから利用できるデータベースで入手可能である。

【0027】

本明細書で使用する「アンプリコン」という用語は、ポリヌクレオチド増幅反応の産物、すなわち、1本鎖または2本鎖であってよく、1つまたは複数の出発配列から複製されたポリヌクレオチドのクローン集団を意味する。1つ以上の出発配列は、同じ配列の1つ以上のコピーであってもよいし、異なる配列の混合物であってもよい。好ましくは、アンプリコンは、単一の出発配列の増幅によって形成される。アンプリコンは、その生成物が1つまたは複数の出発核酸、または標的核酸の複製を含んでいる様々な増幅反応によって生成されてもよい。1つの態様において、アンプリコンを生成する増幅反応は、ヌクレオチドまたはオリゴヌクレオチドのいずれかの反応物の塩基対が、反応生成物の生成に必要な鋳型ポリヌクレオチドにおける相補性を有するという点で「鋳型主導型」である。一態様では、鋳型駆動型反応は、核酸ポリメラーゼによるプライマー伸長、または核酸リガーゼによるオリゴヌクレオチドライゲーションである。このような反応には、ポリメラーゼ連鎖反応(PCR)、線状ポリメラーゼ反応、核酸配列ベース増幅(NASBA)、ローリングサークル増幅などが含まれるが、これらに限定されず、以下の文献に開示されており、その各々は参照によりその全体が本明細書に組み込まれるものとする。Mullisら、米国特許第4,683,195号;第4,965,188号;第4,683,202号;第4,800,159号(PCR);Gelfandら、米国特許第5,210,015号(PCR);Gelfandら、米国特許第5,210,015号(PCR);Gelfandら、米国特許第5,210,015号(PCR)。第5,210,015号(「taqman」プローブを用いたリアルタイムPCR);Wittwerら、米国特許第5,174,640号(PCR)。第6,174,670号;Kacianら、米国特許第5,399,493号。第5,399,491号(「NASBA」);Lizardi, U.S. Pat.第5,854,033号;青野ら、日本特許公開第4-262799号(ローリングサークル増幅);などが挙げられる。一態様において、本発明のアンプリコンは、PCRによって製造される。増幅反応の進行に伴って反応生成物を測定することを可能にする検出化学が利用可能であれば、増幅反応は「リアルタイム」増幅であってもよく、例えば、「リアルタイムPCR」、またはLeone et al, Nucleic Acids Research, 26: 2150-2155 (1998) などに記載の「リアルタイムNASBA」などが利用可能である。

【0028】

増幅する」とは、増幅反応を行うことを意味する。反応混合物」とは、反応を行うために必要な全ての反応物を含む溶液を意味し、反応中にpHを選択されたレベルに維持するための緩衝剤、塩類、補因子、スカベンジャーなどを含むことができるが、これらに限定されるものではない。

【0029】

本明細書で互換的に使用される「フラグメント」または「セグメント」という用語は、より大きなポリヌクレオチド分子の一部を指す。ポリヌクレオチドは、例えば、生物学的試料内に自然に存在し得るcfDNA断片の場合のように自然のプロセスによって、またはin vitro操作によって、複数のセグメントに分解され得る、または断片化され得る。核酸を断片化する様々な方法は、当技術分野でよく知られている。これらの方法は、例えば、化学的または物理的または酵素的な性質のいずれかであってよい。酵素的断片化には、DNaseによる部分分解;酸による部分脱プリンティング;制限酵素の使用;イントロンにコードされたエンドヌクレアーゼ;核酸セグメントの特定のハイブリダイゼーションに依存して切断剤を核酸分子の特定の位置に局在させるトリプレックスおよびハイブリッド形成法などのDNAベースの切断方法;または既知もしくは未知の位置でポリヌクレオチドを切断する他の酵素もしくは化合物などがありうる。物理的断片化方法は、ポリヌクレオチドを高い剪断速度にさらすことを含むことができる。高いせん断速度は、例えば、

ピットまたはスパイクを有するチャンバーまたはチャンネルを通してDNAを動かすこと、または制限されたサイズの流路、例えば、ミクロンまたはサブミクロンの範囲の断面寸法を有するアパーチャーを通してDNA試料を強制的に流すことによって生じることがある。その他の物理的方法としては、超音波処理、ネブライゼーションなどがある。物理的および化学的断片化方法の組み合わせも同様に、熱による断片化およびイオン媒介性加水分解のようなものを採用することができる。例えば、Sambrookら、「Molecular Cloning: A Laboratory Manual」、3rd Ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N. Y. (2001) ("Sambrook et al.") を参照し、これはすべての目的のために参照により本明細書に組み込まれる。これらの方法は、核酸を選択されたサイズ範囲の断片に消化するように最適化することができる。

10

【0030】

本明細書で互換的に使用される「ポリメラーゼ連鎖反応」または「PCR」という用語は、DNAの相補鎖の同時プライマー伸長による特定のDNA配列のインビトロ増幅のための反応を意味する。言い換えれば、PCRは、プライマー結合部位によって挟まれた標的核酸の複数のコピーまたは複製を作るための反応であり、そのような反応は、以下のステップの1回以上の繰り返しを含んでいる。(i) 標的核酸を変性させる、(ii) プライマー結合部位にプライマーをアニールする、(iii) ヌクレオシド三リン酸の存在下で核酸ポリメラーゼによりプライマーを伸長させる。通常、反応はサーマルサイ클ラーを用いて、各段階に最適化された異なる温度で循環させる。特定の温度、各ステップでの持続時間、およびステップ間の変化率は、当業者によく知られている多くの要因に依存し、例えば、以下の文献によって例示される。McPhersonら、編集者、PCR. A Practical ApproachおよびPCR2. A Practical Approach (IRL Press, Oxford, 1991 and 1995, respectively)に記載されている。例えば、Taq DNAポリメラーゼを用いる従来のPCRでは、二本鎖標的核酸を90℃を超える温度で変性させ、50～75℃の範囲の温度でプライマーをアニールさせ、72～78℃の範囲の温度でプライマーを伸長させることができる。採用されるPCRの特定の形式は、アプリケーションの文脈から当業者によって識別可能である。反応量は、数百ナノリットル、例えば、200nLから数百μL、例えば、200μLの範囲とすることができる。"逆転写PCR"、または"RT-PCR"は、標的RNAを相補的な一本鎖DNAに変換する逆転写反応を先行させ、それを増幅するPCRを意味し、その例は、Tecottら、米国特許第5,168,033号明細書に記載されている。第5,168,038号に記載されており、その開示は参照によりその全体が本明細書に組み込まれる。"リアルタイムPCR"とは、反応生成物、すなわちアンプリコンの量が反応の進行に伴って監視されるPCRを意味する。反応産物をモニターするために使用される検出化学物質が主に異なる多くの形態のリアルタイムPCRがあり、例えば、Gelfandら、U.S. Pat.第5,210,015号("taqman")；Wittwerら、U.S. Pat.第6,174,670号および第6,569,627号(インターカレート染料)；Tyagiら、米国特許第5,925,518号(インターカレート染料)；Tyagiら、米国特許第5,925,518号(インターカレート染料)。第5,925,517号(分子ビーコン)；これらの開示は、その全体が本明細書に参照として組み込まれる。リアルタイムPCRのための検出化学は、Mackayら、Nucleic Acids Research、30:1292-1305(2002)において概説されており、これも参照により本明細書に組み込まれるものとする。"Nested PCR"とは、最初のPCRのアンプリコンが、新しいプライマーセット(そのうちの少なくとも1つは最初のアンプリコンの内部の位置に結合する)を用いる2番目のPCRのためのサンプルとなる2段階PCRを意味する。本明細書で使用する場合、入れ子増幅反応に関する「最初のプライマー」は、最初のアンプリコンを生成するために使用されるプライマーを意味し、「二番目のプライマー」は、二番目の、または入れ子されたアンプリコンを生成するために使用される1つまたは複数のプライマーを意味する。"非対称PCR"とは、反応が主に標的核酸の2本の鎖の一方が優先的にコピーされる線形増幅であるように、採用される2つのプライマーの一方が大過剰濃度であるPCRを意味する。非対称PCRプライマーの過剰濃度は、濃度比として表すことができる。典型的な比率は10～100の範囲である。"Multiplexed PCR"とは、複数

20

30

40

50

の標的配列（または単一の標的配列と1つ以上の参照配列）が同一の反応混合物中で同時に実施されるPCRを意味し、例えば、Bernardら、Anal.Biochem., 273: 221-228 (1999)(2-colorリアルタイムPCR)に記載されている。通常、増幅される各配列に対して、異なるプライマーセットが採用される。通常、マルチプレックスPCRにおける標的配列の数は、2～50、または2～40、または2～30の範囲である。「定量的PCR」とは、サンプルまたは試料中の1つまたは複数の特定の標的配列の存在量を測定するように設計されたPCRを意味する。定量的PCRには、そのような標的配列の絶対的定量と相対的定量の両方が含まれる。定量的測定は、標的配列と別々にまたは一緒にアッセイされ得る1つ以上の参照配列または内部標準を使用して行われる。参照配列は、試料または検体に対して内因性であっても外因性であってもよく、後者の場合、1つまたは複数の競合テンプレートから構成されることがある。典型的な内因性参照配列としては、以下の遺伝子の転写産物のセグメントが挙げられる：β-アクチン、GAPDH、β₂-ミクログロブリン、リボソームRNAなどである。定量的PCRのための技術は、参照によりその全体が本明細書に組み込まれる以下の文献に例示されるように、当業者にはよく知られている。Freemanら、Biotechniques、26: 112-126 (1999) ; Becker-Andreら、Nucleic Acids Research、17: 9437-9447 (1989) ; Zimmermanら、Biotechniques、21: 268-279 (1996) ; Diviaccoら、Gene、122: 3013-3020 (1992) ; 及びBecker-Andreら、Nucleic Acids Research、17: 9437-9446 (1989)。

10

【0031】

本明細書で使用する「プライマー」という用語は、ポリヌクレオチド鋳型と二重鎖を形成する際に、核酸合成の開始点として働き、その3'末端から鋳型に沿って延長されて、延長二重鎖を形成することができる、天然または合成オリゴヌクレオチドを意味する。プライマーの伸長は、通常、DNAポリメラーゼやRNAポリメラーゼなどの核酸ポリメラーゼを用いて行われる。伸長過程で付加されるヌクレオチドの配列は、鋳型となるポリヌクレオチドの塩基配列によって決定される。通常、プライマーは、DNAポリメラーゼによって伸長される。プライマーの長さは、通常、14～40ヌクレオチド、または18～36ヌクレオチドの範囲にある。プライマーは、例えば、単一のプライマーを用いる線形増幅反応や、2つ以上のプライマーを用いるポリメラーゼ連鎖反応など、様々な核酸増幅反応に使用される。特定の用途のためのプライマーの長さおよび配列を選択するための指針は、その全体が参照により本明細書に組み込まれる以下の文献によって証明されるように、当業者にはよく知られている。Dieffenbach, editor, PCR Primer: A Laboratory Manual, 2nd Edition (Cold Spring Harbor Press, New York, 2003)。

20

30

【0032】

本明細書において、「被験者」及び「患者」という用語は互換的に用いられ、例えば癌などの医学的状態又は障害を有することが知られている、又は潜在的に有するヒト又は非ヒト動物を指す。

【0033】

本明細書で使用する「配列読み取り」という用語は、被験者から得られた試料から得られた核酸分子の一部または全部からのヌクレオチドの文字列を指す。配列読み取りは、核酸断片から配列決定された短いヌクレオチド列（例えば、20～150）、核酸断片の片端または両端の短いヌクレオチド列、または生体試料中に存在する核酸断片全体の配列決定であり得る。シーケンスリードは、当技術分野で知られている様々な方法によって得ることができる。例えば、配列読み取りは、様々な方法、例えば、配列決定技術を使用して、またはプローブ、例えば、ハイブリダイゼーションアレイまたは捕捉プローブにおいて、または増幅技術、例えば、ポリメラーゼ連鎖反応（PCR）または単一プライマーもしくは等温増幅を用いた線形増幅を用いて得ることができる。

40

【0034】

本明細書で使用する「リードセグメント」または「リード」という用語は、被験者から得られた配列リードおよび／またはサンプルからの最初の配列リードから得られたヌクレオチド配列を含む、任意のヌクレオチド配列を意味する。例えば、リードセグメントは、

50

アライメントされた配列リード、コラプスされた配列リード、またはスティッチされたリードを指すことができる。さらに、リードセグメントは、単一のヌクレオチドバリエーションなど、個々のヌクレオチド塩基を指すことができる。

【0035】

本明細書で使用される「濃縮」という用語は、サンプル中の1つ以上の標的核酸の割合を増加させることを意味する。したがって、「濃縮」試料または配列決定ライブラリーは、試料中の非標的核酸に対して1つ以上の標的核酸の割合が増加した試料または配列決定ライブラリーのことである。

【0036】

一般に、ポリヌクレオチド（例えば、「無細胞RNA」及び「無細胞DNA」）に適用される「無細胞」、「循環」及び「細胞外」という用語は、対象からの試料又はその一部中に存在するポリヌクレオチドを指すために互換的に用いられ、（例えば、細胞又はウイルスからの抽出のための溶解におけるように）当初採取した試料に溶解工程を適用せずに分離又はその他の操作が可能なポリヌクレオチドに用いられる。無細胞ポリヌクレオチドは、したがって、被験者の試料が採取される前であっても、それらが由来する細胞またはウイルスからカプセル化されていない、または「遊離」している。無細胞ポリヌクレオチドは、細胞死（例えば、アポトーシスまたはネクローシス）または細胞脱落の副産物として生成され、ポリヌクレオチドを周囲の体液または循環中に放出することがある。したがって、無細胞ポリヌクレオチドは、血液の非細胞画分（例えば、血清または血漿）、他の体液（例えば、尿）、または他の種類の試料の非細胞画分から単離されてもよい。無細胞RNAまたは「cfRNA」という用語は、被験者の体内（例えば、血流）を循環するリボ核酸断片を指し、1つまたは複数の健康な細胞および／または1つまたは複数の癌細胞に由来する可能性がある。同様に、「無細胞DNA」または「cfDNA」は、被験者の体内（例えば、血流）を循環するデオキシリボ核酸分子を指し、1つまたは複数の健康な細胞および／または1つまたは複数の癌細胞に由来する可能性がある。

【0037】

循環腫瘍RNAまたは「ctRNA」という用語は、腫瘍細胞または他の種類の癌細胞に由来するリボ核酸断片を指し、瀕死の細胞のアポトーシスまたは壊死などの生物学的プロセスの結果として被験者の身体（例えば、血流）に放出されるか、生存腫瘍細胞によって活発に放出される可能性がある。

【0038】

本明細書で使用する「ダークチャンネルRNA」または「ダークチャンネルcfRNA分子」または「ダークチャンネル遺伝子」という用語は、健康な細胞での発現が非常に低いまたは存在しないRNA分子または遺伝子を指す。したがって、ダークチャンネルRNA（cfRNA）分子の同定、検出、および／または定量化は、癌などの疾患状態の評価において、信号対雑音、および感度および特異性の改善をもたらす。

【0039】

本明細書で使用される「治療する」または「治療」は、臨床結果を含む、被験者の状態における有益なまたは所望の結果を得るための任意のアプローチを含む。有益なまたは所望の臨床結果には、1つ以上の症状または状態の緩和または改善、疾患の範囲の縮小、疾患の状態の安定化（すなわち、悪化させない）、疾患の感染または拡大の防止、疾患の進行の遅延または減速、疾患の状態の改善または緩和、疾患の再発の減少、および部分的または全体的であるかどうか、および検出可能または検出不可能であるかどうかにかかわらず、寛解が含まれるが、これらに限定されない。すなわち、本明細書で使用される「治療」は、疾患の任意の治療、改善、または予防を含む。治療は、疾患の発生を防止すること、疾患の広がりを抑制すること、疾患の症状を緩和すること、疾患の根本原因を完全にまたは部分的に取り除くこと、疾患の期間を短縮すること、またはこれらのことを組み合わせで行うことである。

【0040】

本明細書で使用される「治療する」及び「治療」は、予防的な治療を含む。治療方法は

、治療上有効な量の活性剤を対象に投与することを含む。投与するステップは、1回の投与で構成されてもよいし、一連の投与を含んでもよい。治療期間の長さは、状態の重症度、患者の年齢、活性剤の濃度、治療に使用される組成物の活性、またはそれらの組み合わせなどの様々な要因に依存する。また、治療または予防に使用される薬剤の有効量は、特定の治療または予防体制の経過とともに増加または減少する可能性があることが理解されよう。投与量の変化は、当該技術分野で知られている標準的な診断アッセイによって生じ、明らかになることがある。ある実施態様では、慢性投与が必要とされ得る。例えば、組成物は、患者を治療するのに十分な量および期間、対象に投与される。実施形態において、治療または処置は、予防的処置ではない。

【0041】

被験者の疾患または状態に関連する「予防」という用語は、被験者における1つ以上の対応する症状の発生を減少させることを意味する。上に示したように、予防は、完全（検出可能な症状がない）であってもよいし、治療を行わなかった場合に起こりそうな症状よりも少ない症状が観察される、及び／又は低い発生率であるような、部分的なものであってもよい。

【0042】

"抗癌剤"及び"抗癌剤"は、それらの明白な通常の意味に従って用いられ、抗腫瘍性特性又は細胞の成長若しくは増殖を阻害する能力を有する組成物（例えば、化合物、薬剤、アンタゴニスト、阻害剤、調節剤）を意味する。いくつかの実施形態では、抗癌剤は、化学療法剤である。いくつかの実施形態では、抗癌剤は、癌を治療する方法において有用性を有する本明細書で同定される薬剤である。いくつかの実施形態では、抗癌剤は、癌を治療するために、FDA又は米国以外の国の同様の規制機関によって承認された薬剤である。抗癌剤の例には、MEK（例えば、MEK1、MEK2、またはMEK1およびMEK2）阻害剤（例えば、以下が挙げられるが、これらに限定されない）が含まれる。XL518、CI-1040、PD035901、セルメチニブ／AZD6244、GSK1120212／トラメチニブ、GDC-0973、ARRY-162、ARRY-300、AZD8330、PD0325901、U0126、PD98059、TAK-733、PD318088、AS703026、BAY869766）、アルキル化剤（例えば、シクロホスファミド、イホスファミド、クロラムブシル、ブスルファン、メルファラン、メクロレタミン、ウラムスチン、チオテパ、ニトロソウレア、ナイトロジェンマスタード（例えば、メクロロエタミン、シクロホスファミド、クロラムブシル、マイファラン）、エチレンイミンおよびメチルメラミン（例えば、ヘキサメトリメラミン、チオテパ）、アルキルスルホン酸塩（例えば、ブスルファン）、ニトロソウレア（例えば、カルムスチン、ロムシチン、セムスチン、ストレプトゾシン）、トリアゼン（デカルバジン）、抗代謝物（例えば、(a,5-アザチオプリン、ロイコボリン、カペシタビン、フルダラビン、ゲムシタビン、ペメトレキセド、ラルテイトレキセド、葉酸アナログ（例えば、メトトレキサート）、またはピリミジンアナログ（例えば、フルオロウラシル、フロキシリジン、シタラビン）、プリンアナログ（例えば、メルカプトプリン、チオグアニン、ペントスタチン）など）、植物アルカロイド（例えば、ビンクリスチン、ビンブラスチン、ビノレルビン、ビンデシン、ポドフィロトキシン、パクリタキセル、ドセタキセルなど）、トポイソメラーゼ阻害剤（例えば、イリノテカン、トポテカン、アムサクリン、エトポシド（VP16）、リン酸エトポシド、テニポシドなど）、抗腫瘍性抗生物質（例、ドキソルビシン、アドリアマイシン、ダウノルビシン、エピルビシン、アクチノマイシン、ブレオマイシン、マイトマイシン、ミトキサントロン、プリカマイシンなど）、白金系化合物（例えば、シスプラチン、オキサロプラチン、カルボプラチン）、アントラセン系（例えば、ミトキサントロン）、置換尿素（例えば、ヒドロキシ尿素）、メチルヒドラジン誘導体（例えば、プロカルバジン）、副腎皮質抑制剤（例えば、ミトタン、アミノグルテチミド）、エポドフィロトキシン（例えば、エトポシド）、抗生物質（例えば、ダウノルビシン、ドキソルビシン、ブレオマイシン）、酵素（例えば、L-アスパラギナーゼ）、ミトジェン活性化タンパク質キナーゼシグナル伝達の阻害剤（例えば、U0126、PD98059、PD184352、PD0325901、ARRY-142886、SB239063、SP600125、BAY 43-

10

20

30

40

50

9006、wortmannin、またはLY294002、Syk阻害剤、mTOR阻害剤、抗体（例えば、リツキサン）、ゴシホル、ゲナーセンス、ポリフェノールE、クロロフシン、オール
 トランスレチノイン酸（ATRA）、ブリオスタチン、腫瘍壊死因子関連アポトーシス誘導
 リガンド（TRAIL）、5-アザ-2'-デオキシシチジン、オールトランスレチノイン酸、ド
 キソルピシン、ピンクリスチン、エトポシド、ゲムシタビン、イマチニブ（Gleevec.RT
 M.）、ゲルダナマシ、17-N-アリラミノ-17-デメトキシゲルダナマシ（17-AAG）
 、フラボピリドール、LY294002、ボルテゾミブ、トラスツズマブ、BAY11-7082、P
 KC412、PD184352、20-エピ-1、25ジヒドロキシビタミンD3;5-エチニルウラシル
 ;アピラテロン;アクラルピシン;アシルフルベンヌ;アデシペノール;アドゼレシン;
 アルデスルキンALL-TKアンタゴニスト;アルトレタミン;アンバムスチン;アミドック
 ス;アミフォスチン;アミノレブリン酸;アムルピシン;アムサクリン;アナグレリド;
 アナストロゾール;アンドログラフォリド;血管新生阻害剤;アンタゴニストD;アンタ
 ゴニストG;アンタレリックス;抗背側形態形成蛋白質1;抗アンドロゲン、前立腺がん
 。抗エストロゲン;アンチネオプラストン;アンチセンスオリゴヌクレオチド;アフィジ
 コリングリシネート;アポトーシス遺伝子モジュレーター;アポトーシスレギュレーター
 ;アプリン酸;ar-CDP-DL-ptba;アルギニンデアミナーゼ;アズラークリン;アタメ
 スタン;アトリムスチン;アキシナスタチン1;アキシナスタチン2;アキシナスタチン
 3;アザセトロン。アザトキシン;アザチロシン;バックチンIII誘導体;バラノール;
 バチマスタット;BCR/ABLアンタゴニスト;ベンゾクロリン;ベンゾイルスタウロス
 ポリン;βラクタム誘導体;β-アレチン;ベタクラマイシンB;ベツリン酸;bFGF阻害
 剤;ピカルタミド;ピサンドレン;ピサギリジニルスペルミン;ビスナファイドビストラ
 テンA;ビゼルシン;ブレフレート;ブロピリミン;ブドチタン;ブチオニンスルホキシ
 ミン;カリボトリオール;カルホスチンC;カンプトテシン誘導体;カナリーポックスIL
 -2;カペシタビン;カルボキサミド-アミノトリアゾール;カルボキシアミドトリアゾー
 ル;カレストM3;CARN700;軟骨由来阻害剤;カゼレジン。カゼインキナーゼ阻害剤
 (ICOS);カスタノスペルミン;セクロピンB;セトロレリックス;クロリン;クロロ
 キノキサリンサルフォンアミド;シカプロスト;シス-ボルフィリン;クラドリピン;ク
 ロミフェン類似体;クロトリマゾール;コリスマイシンA;コリスマイシンB;コンプレ
 タスチンA4;コンプレタスチンアナログ;コネゲニンがある。クランベシジン816;ク
 リスナトール;クリプトフィシン8;クリプトフィシンA誘導体;クラシンA;シクロペ
 ンタントラキノン;シクロプラタム;シペマイシン;シタラビンオクフォスフェート;細
 胞溶解因子;サイトスタチン;ダクリキシマブ;デシタビン;デヒドロデムニンB;デス
 ローレン;デキサメタゾン;デキシフォスファミド;デクスラゾキサンドクスベラパミル
 ;ジアジコン;ディデムニンB;ディドックス;ジエチルノルパミン;ジヒドロ-5-アザ
 シチジン;9-ジオキサマイシン;ジフェニルスピロムスチン;ドコサノール;ドラセト
 ロン;ドキシフルリジン;ドロキシフェン;ドロナビノール;デュカルマイシンSA;
 エブセレン;エコムスチン;エデルホシン;エトレコマブ;エフロニチン;エレメン;エ
 ミテファ。エピルピシン;エプリステリド;エストラムスチンアナログ;エストロゲンア
 ゴニスト;エストロゲンアンタゴニスト;エタニダゾール;リン酸エトポシド;エキセメ
 スタン;ファドロゾール;ファザラビン;フェンレチニド;フィルグラスチム;フィナス
 テリド;フラボピリドール;フレゼラスチン;フラステロン;フルダラビン;フルダノ
 ルニシン塩酸塩。フォーフェニメックス;フォルメスタン;フォストリエシン;フォテム
 スチン;ガドリニウムテキサフィリン;硝酸ガリウム;ガロシタビン;ガニレリクス;ゼ
 ラチナーゼ阻害剤;ゲムシタビン;グルタチオン阻害剤;ヘプスルファム;ヘレグリン;
 ヘキサメチレンビスアセタミド;ハイパーシン;イバンドロン酸;イダルピシン;イドキ
 シフェンイドラマントン;イルモフォシン;イロマスタット;イミダゾアクリドン;イミ
 キモド;免疫賦活ペプチド;インスリン様成長因子-1受容体阻害剤;インターフェロン
 アゴニスト;インターフェロン;インターロイキン;ヨーベングアン;ヨードドソルピシ
 ン;イボメノール、4-;イロプラクト;イルソグラジン;イソベンガゾール;イソホモ
 アリコンドリンB。イタセトロン;ジャスプラキノライド;カハライドF;ラメラリン-N

10

20

30

40

50

トリアセテート；ランレオチド；レイナマイシン；レノグラスチム；レンチナン硫酸；レ
 プトルスタチン；レトロゾール；白血病抑制因子；白血球αインターフェロン；リユーブ
 ロライド+エストロゲン+プロゲステロン；リユープロレリン；レバミゾール；リリアゾ
 レ。線状ポリアミンアナログ；親油性二糖ペプチド；親油性白金化合物；リソクリンアミ
 ド7；ロバプラチン；ロンブリシン；ロメトレキソール；ロニダミン；ロキサントロン；
 ロバスタチン；ロキソリビン；ルルトテカン；ルテチウムテキサフィリン；リゾフィリン
 ；溶解ペプチド；メイタンシン。マンノスタチンA；マリマスタット；マソプロコール；
 マスピン；マトリックスメタロプロテアーゼ阻害剤；メノガリル；メルバロン；メテレリ
 ン；メチオニナーゼ；メトクロプラミド；MIF阻害剤；ミフェプリストン；ミルテフォ
 シン；ミリモスティム；ミスマッチ2本鎖RNA；ミトグアゾン；ミトラクトール。マイト
 マイシン類似体；ミトナフィド；マイトトキシ線維芽細胞成長因子-サポリン；ミトキ
 サントロン；モファロテン；モルグラモスチム；モノクローナル抗体、ヒト絨毛性ゴナド
 トロフィン；モノフォスリピッドA+マイオバクテリア細胞壁sk；モピダモール；多重薬
 剤耐性遺伝子阻害剤。多重癌抑制因子1ベース療法；マスタード抗癌剤；マイカペロキサ
 イドB；マイコバクテリア細胞壁抽出物；ミリアポロン；N-アセチルジナリン；N-置換
 ベンズアミド；ナファレリン；ナグレシビ；ナロキソン+ペンタゾシン；ナパビン；ナフ
 ターピン；ナルトグラスチム；ネダプラチン；ネモルビシン；ネリドロロン酸。中性エンド
 ペプチダーゼ；ニルタミド；ニサマイシン；一酸化窒素調節剤；一酸化窒素抗酸化剤；ニ
 トルリン；O6-ベンジルグアニン；オクトレオチド；オキセノン；オリゴヌクレオチド；
 オナプリストン；オندانセトロン；オラシン；口腔サイトカイン誘導剤；オルマプラチ
 ン；オサテロン；オキサリプラチン；オキサウノミシンの順。パラオアミン；パルミトイ
 ルリゾキシシン；パミドロロン酸；パナキシトリオール；パノミフェン；パラバクチン；パゼ
 リプチン；ペガスパーガーゼ；ペルデシン；ペントサンポリ硫酸ナトリウム；ペントスタ
 チン；ペントロゾール；パーフルブロン；パーパスファミド；ペリルアルコール；フェナ
 ジノマイシン；フェニル酢酸；ホスファターゼ阻害剤ピシバニール；ピロカルピン塩酸塩
 ；ピラルビシン；ピリットレキシム；プラセチンA；プラセチンB；プラスミノーゲンア
 クチベーター阻害剤；白金錯体；白金トリアミン錯体；ポルフィマーナトリウム；ポルフ
 ィロマイシン；プレドニゾン；プロピルビスアクリドン；プロスタグランジンJ2；プロ
 テアソーム阻害剤。プロテインA系免疫調節剤；プロテインキナーゼC阻害剤；プロティ
 ンキナーゼC阻害剤、微細藻類；プロテインチロシンホスファターゼ阻害剤；プリンヌク
 レオシドホスホリラーゼ阻害剤；purpurins；ピラゾロアクリジン；ピリドキシ化ヘモ
 グロビンポリオキシエチル化コンジュゲート。rafアンタゴニスト；ラルティトレキセド
 ；ラモセトロン；rasファルネシル蛋白質転移酵素阻害剤；ras阻害剤；ras-GAP阻害剤
 ；レトリプチン脱メチル化；レニウムRe186エチドロネート；リゾキシシン；リボザイム
 ；RIIレチナミド；ログレティミド；ロヒツクイン；ロムルチド；ロキニメックス；ルビ
 ギノンB1；ruboxyl；safingol；saintopin；SarCNU；sarcophytol A；sargramos
 tim；Sdi 1 mimetics；semustine；senescence derived inhibitor 1；sense o
 ligonucleotides；signal transduction inhibitor；signal transduction mod
 ulators；single chain antigen-binding protein；sizofuran；sobuzoxane；ボ
 ロカプチン酸ナトリウム；フェニル酢酸ナトリウム；ソルベロール；ソマトメジン結合タ
 ンパク質；ソナーミン；スパーフォシン酸；スピカマイシンD；スピロムスチン；スプレ
 ノペンティン；スポンジスタチン1；スクアラミン；幹細胞阻害剤；幹細胞分裂阻害剤；
 ステイピアミド；ストロメリン阻害剤；サルフィノシンがある。超活性血管活性腸ペプチ
 ド拮抗薬；スラディスタ；スラミン；スワインソニン；合成グリコサミノグリカン；タリ
 ムスチン；タモキシフェンメチオダイド；タウラムスチン；タザロテン；テコガランナト
 リウム；テガフル；テルラピリリウム；テロメラーゼ阻害剤；テモポルフィン；テモゾ
 ロマイド；テニボシドテトラクロロデカオキシド；テトラゾミン；タリブラスチン；
 チオコラリン；トロンボポエチン；トロンボポエチン模倣品；チマルファシン；チモポエ
 チン受容体作動薬；チモトリナン；甲状腺刺激ホルモン；チン・エチル・エチオプルプリ
 ン；チラパザミン；チタノセンビクロリド；トップセンチン；トレミフェン totipotent

10

20

30

40

50

stem cell factor; translation inhibitors; tretinoin; triacetyluridine; triciribine; trimetrexate; triptorelin; tropisetron; turosteride; tyrosine kinase inhibitors; tyrphostins; UBC inhibitors; ubenimex; urogenital sinus-derived growth inhibitory factor; urokinase receptor antagonists; UTC; バブレオチド; バリオリンB; ベクターシステム、赤血球遺伝子治療; ベラレゾール; ベラミン; ベルジン; ベルテポルフィン; ビノレルビン; ピンクサルチン; ビタキシン; ボロゾール; ザノテロン; ゼニプラチン; ジラスコブ; ジノスタチンスチラマー、アドリアマイシン、ダクチノマイシン、プロマイシン、ビンブラスチン、シスプラチン、アシビジン。アクラルピシン; アコダゾール塩酸塩; アクロニン; アドゼルシン; アルデスロイキン; アルトレタミン; アンモマイシン; 酢酸アメタントロン; アミノグルテチミド; アムサクリン; アナストロゾール; アントラマイシン; アスパラキナーゼ; アスペリン; アザシチジン; アゼテパ; アゾトマイシン; バティマスタ; ベンゾデッパ; ビカルタミド塩酸ビスアントレン; ジメシル酸ビスナフィド; ビゼレシン; 硫酸ブレオマイシン; ブレキナーナトリウム; プロピリミン; プスルファン; カクチノマイシン; カルスステロン; カラセミド; カルベチマー; カルボプラチン; カルムスチン; 塩酸カルピシン; カゼレシン; セデフィングール; クロラムブチル; シロールマイシン; クラドリビンクリスナトールメシレート; シクロホスファミド; シタラビン; ダカルバジン; ダウノルビシン塩酸塩; デシタビン; デキソルマブラチン; デザグアニン; デザグアニンメシレート; ダイアジコン; ドキソルビシン; 塩酸ドキソルビシン; ドロキシフェン; クエン酸ドロキシフェン; ドロモスダノンプロピオン酸; ドュアゾマイシンの5種類である。エダトレキサート; エフルニチン塩酸塩; エルサミットルシン; エンロプラチン; エンプロメート; エピプロピジン; エピルビシン塩酸塩; エルプロゾール; エソルビシン塩酸塩; エストラムスチン; エストラムスチンリン酸ナトリウム; エタニダゾール; エトポシド; リン酸エトポシド; エトプライン; ファドロゾール塩酸塩ファザラビン; フェンレチニド; フロクスリジン; リン酸フルダラビン; フルオロウラシル; フルオロシタビン; フォスキドン; フォストリエシンナトリウム; ゲムシタビン; 塩酸ゲムシタビン; ヒドロキシウレア; 塩酸イダルビシン; イホスファミド; イモフォシン; インターロイキンI1 (組換えインターロイキンII、またはrILを含む、以下同じ。sub.2)、インターフェロンアルファ-2a; インターフェロンアルファ-2b; インターフェロンアルファ-n1; インターフェロンアルファ-n3; インターフェロンベータ-1a; インターフェロガンマ-1b; イプロプラチン; 塩酸イリノテカン; 酢酸ランレオチド; レトロゾール; 酢酸リウプロライド; 塩酸リウソゾル。ロメトレキソールナトリウム; ロムスチン; ロソキサントロン塩酸塩; マソプロコール; メイタンシン; メクロレタミン塩酸塩; メゲストロール酢酸塩; メレンゲストロール酢酸塩; メルファラン; メノガリル; メルカプトプリン; メトトレキサート; メトトレキサートナトリウム; メソプラインメツレデパ; ミチンドミド; ミトカルシン; ミトクロミン; ミトギリリン; ミトマルシン; マイトマイシン; ミトスパー; ミトタン; 塩酸ミトキサントン; ミコフェノール酸; ノコダゾイ; ノガマイシン; オルマブラチン; オキスラン; ペガスパガーゼ; ペリオマイシン; ペンタムスチンペプロマイシン硫酸塩; ペルホスファミド; ピボプロマン; ピボスルファン; ピロキサントロン塩酸塩; プリカマイシン; プロメスタン; ポルフィマーナトリウム; ポルフィロマイシン; プレドニムスチン; 塩酸プロカルバジン; プロマイシン; 塩酸プロマイシン; ピラゾフォーリンリボプリン; ログレチミド; サフィンゴル; サフィンゴル塩酸塩; セムスチン; シムトラゼン; スパーホサートナトリウム; スパーソマイシン; スピログルマニウム塩酸塩; スピロムスチン; スピロプラチン; ストレプトニグリン; ストレプトゾシン; スロフェノール; タリスマイシン; テコガランナトリウムテガフル; テロキサントロン塩酸塩; テモポルフィン; テニポシド; テロキシロン; テストラクトン; チアミプリン; チオグアニン; チオテパ; チアゾフリン; チラパザミン; クエン酸トレミフェン; 酢酸トレストロン; リン酸トリシリジン; トリメトレキサート。トリメトレキサートグルクロン酸塩; トリプトレリン; ツプロゾール塩酸塩; ウラシルマスタード; ウレデパ; バブレオチド; ベルテポルフィン; ビンブラスチン硫酸塩; ビンクリスチン硫酸塩; ビンデシン; ビンデシン硫酸塩; ビンピジン硫酸塩; ビニルシ

10

20

30

40

50

ート硫酸塩；ビンレウロシン硫酸塩。酒石酸ビノレルピン；硫酸ビノロシジン；硫酸ビン
 ゴリジン；ボロゾール；ゼニプラチン；ジノスタチン；塩酸ゾルピシン、細胞をG2-M相
 で停止させる薬剤、および／または微小管の形成もしくは安定性を調節する薬剤、（e.
 例えば、タキソール.TM（すなわち、パクリタキセル）、タキソテール.TM、タキサン骨
 格を含む化合物、エルプロゾール（すなわち、R-55104）、ドラスタチン10（すなわ
 ち、DLS-10およびNSC-376128）、ミボブリンイソチオネイト（I.すなわちCI-980
 として）、ピンクリスチン、NSC-639829、ディスコデルモライド（すなわちNVP-XX
 -A-296として）、ABT-751（Abbott、すなわちE-7010）、アルトルヒルチン（すな
 わちアルトルヒルチンA、アルトルヒルチンC）、スポンジスタチン（すなわちスポンジ
 スタチン1、スポンジスタチン2、スポンジスタチン3、スポンジスタチン4、スポンジス
 タチン5、スポンジスタチン6、スポンジスタチン7、スポンジスタチン8、スポンジスタ
 チン9など）、セマドチン塩酸塩（I.E. LU-103793など）、セマドチン塩酸塩（I.E. L
 U-103793など）。LU-103793、NSC-D-669356など）、エポチロン（エポチロン
 A、エポチロンB、エポチロンC（デソキシエポチロンAまたはdEpoAなど）、エポチロ
 ンD（すなわち、KOS-862、dEpoB、デソキシエポシロンB）、エポシロンE、エポシ
 ロンF、エポシロンB N-オキシド、エポシロンA N-オキシド、16-アザ-エポシロンB、2
 1-アミノエポシロンB（i.BMS-310705等）、21-ヒドロキシエポシロンD（デソキシ
 エポシロンF、dEpoF等）、26-フルオロエポシロン、オーリスタチンPE（NSC-6546
 63等）、ソブリドチン（TZT-1027等）、LSR-1、LSR-2、LSR-3、LSR-4、LSR-5
 、LSR-6、LSR-7等。TZT-1027）、LS-4559-P（ファルマシア、すなわちLS-4577
 ）、LS-4578（ファルマシア、すなわちLS-477-P）、LS-4477（ファルマシア）、L
 S-4559（ファルマシア）、RPR-112378（Aventis）、Vincristine sulfate、DZ-
 3358（第一化学工業）、FR-182877（藤沢、すなわちWS-9885B。e. WS-9885B
 ）、GS-164（武田）、GS-198（武田）、KAR-2（ハンガリー科学アカデミー）、BSF
 -223651（BASF、in.すなわちILX-651およびLU-223651）、SAH-49960（Lilly
 /Novartis）、SDZ-268970（Lilly/Novartis）、AM-97（Armad/Kyowa Ha
 kko）、AM-132（Armad）、AM-138（Armad/Kyowa Hakko）、IDN-5005（I
 ndena）、Cryptophycin 52（I.LY-355703など）、AC-7739（味の素社製、すな
 わちAVE-8063AおよびCS-39・HCl）、AC-7700（味の素社製、すなわちAVE-806
 2、AVE-8062A、CS-39・L・Ser. HCl、およびRPR-258062A）、Vitilevuamid
 e、Tubulysin A、Canadensol、Centaureidin（すなわちNSC-106969）、T-13
 8067（Tularik、すなわちI.T-67、TL-138067およびTI-138067）、COBRA-1（
 パーカーヒューズ研究所、すなわちDDE-261およびWHI-261）、H10（カンザス州立
 大学）、H16（カンザス州立大学）、Oncocidin A1（I.e.BTO-956およびDIME）、
 DDE-313（Parker Hughes Institute）、Fijianolide B、Laulimalide、SPA-2
 （Parker Hughes Institute）、SPA-1（Parker Hughes Institute、すなわちSP
 IKET-P）、3-IABU（Cytoskeleton/Mt.Sinai School of Medicine、すなわちM
 F-569）、Narcosine（NSC-5366としても知られている）、Nascapine、D-2485
 1（アスタメディカ）、A-105972（アボット）、ヘミアスターリン、3-BAABU（Cyt
 oskeleton/Mt.Sinai School of Medicine、すなわちMF-191）、TMPN（Arizon
 a State University）、Vanadocene acetylacetonate、T-138026（Tularik）
 、Monsatrol、Inanocine（すなわちNSC-698666）、3-IABE（Cytoskeleton/
 Mt. Sinai School of Medicine）、A-204197（Abbott）、T-607（Tuiarik, i.e.
 T-900607）、RPR-115781（Aventis）、エレウテロビン（Desmethyleleuthero
 bin、Desaetyeleutherobin、Isoeleutherobin A、Z-Eleutherobinなど）、カリ
 バオサイド、カリバオリン、Halichondrin B、D-64131（アスタメディカ）、D-681
 44（アスタメディカ）、ジアゾナミドA、A-293620（アボット）、NPI-2350（ネレ
 ウス）、タカロノライドA、TUB-245（アベンティス）、A-259754（アボット）、Di
 ozostatin、(-) Phenylahistin（i.e.NSCL-96F037）、D-68838（アスタメディ
 カ）、D-68836（アスタメディカ）、ミオセベリンB、D-43411（ゼンタリス、すな

10

20

30

40

50

わち、D-81862)、A-289099(アボット)、A-318315(アボット)、HTI-286(すなわち、SPA-110、トリフルオロ酢酸塩)(Wyeth)、D-82317(Zentaris)、D-82318(Zentaris)、SC-12983(NCI)、レスベラスチンリン酸ナトリウム、BPR-OI-007(National Health Research Institutes)、SSR-250411(Sanofi))、ステロイド(例えば、デキサメタゾン)、フィナステリド、アロマターゼ阻害剤、ゴセレリンまたはリユープロリドなどのゴナドトロピン放出ホルモンアゴニスト(GnRH)、副腎皮質ホルモン(例えば、プレドニゾン)、プロゲステロン(例えば、ヒドロキシプロゲステロンカプロエート、メゲストロールアセテート、メドロキシプロゲステロンアセテート)、エストロゲン(例えば、ジエチルスチルベストロール、エチニルエストラジオール)、抗エストロゲン(例えば、タモキシフェン)、アンドロゲン(例えば、プロピオン酸テストステロン、フルオキシメステロン)、抗アンドロゲン(例えば、フルタミド)、免疫賦活剤(例えば、カルメット菌(BCG)、レバミソール、インターロイキン-2、 α -インターフェロンなど)、モノクローナル抗体(例えば、抗CD20、抗HER2、抗CD52、抗HLA-DR、抗VEGFモノクローナル抗体)、免疫毒素(抗CD33モノクローナル抗体-カリケアマイシン結合体、抗CD22モノクローナル抗体-シュードモナス外毒素結合体など)、放射免疫療法(例えば、抗CD20モノクローナル抗体と ^{111}In 、 ^{90}Y 、 ^{131}I との結合体など)、放射性免疫療法)、トリプトライド、ホモハリントニン、ダクチノマイシン、ドキシソルピシン、エピルピシン、トポテカン、イトラコナゾール、ビンデシン、セリバスタチン、ピンクリスチン、デオキシアデノシン、サートラリン、ピタバスタチン。イリノテカン、クロファジミン、5-ノニルオキシトリプタミン、ベムラフェニブ、ダブラフェニブ、エルロチニブ、ゲフィチニブ、EGFR阻害剤、上皮成長因子受容体(EGFR)標的治療または治療(e.g.ゲフィチニブ(イレッサ(登録商標))、エルロチニブ(タルセバ(登録商標))、セツキシマブ(アービタックス(登録商標))、ラパチニブ(タイカーブ(登録商標))、パニツムマブ(ベクティビックス(登録商標))、バンデタニブ(カプレッサ(登録商標))、アフアチニブ/BIBW2992、CI-1033/カナチニブ、ネラチニブ/HKI-272、CP-724714、TAK-285、AST-1306、ARRY334543、ARRY-380、AG-1478、ダコミチニブ/PF299804、OSI-420/デスメチルエルロチニブ、AZD8931、AEE788、ペリティニブ/EKB-569、CUDC-101、WZ8040、WZ4002、WZ3146、AG-490、XL647、PD153035、BMS-599626)、ソラフェニブ、イマチニブ、スニティニブ、ダサチニブ等があげられる。

10

20

30

【0043】

本明細書で使用される「エピジェネティック阻害剤」は、DNAメチル化(DNAメチル化阻害剤)またはヒストンの修飾(ヒストン修飾阻害剤)などのエピジェネティックプロセスの阻害剤を意味する。エピジェネティック阻害剤は、ヒストンデアセチラーゼ(HDAC)阻害剤、DNAメチルトランスフェラーゼ(DNMT)阻害剤、ヒストンメチルトランスフェラーゼ(HMT)阻害剤、ヒストンデメチラーゼ(HDM)阻害剤、またはヒストンアセチルトランスフェラーゼ(HAT)であってもよい。HDAC阻害剤の例には、ボリノスタット、ロミデプシン、CI-994、ベリノスタット、パノビノスタット、ジビノスタット、エンティノスタット、モセチノスタット、SRT501、CUDC-101、JNJ-26481585、またはPCI24781が含まれる。DNMT阻害剤の例には、アザシチジンおよびデシタピンが含まれる。HMT阻害剤の例には、EPZ-5676が含まれる。HDM阻害剤の例には、パルギリンおよびトラニルシプロミンが含まれる。HAT阻害剤の例には、CCT077791およびガルシニアが含まれる。

40

【0044】

「マルチキナーゼ阻害剤」とは、チロシンプロテインキナーゼおよびセリン/スレオニンキナーゼを含む少なくとも1つのタンパク質キナーゼの低分子阻害剤である。マルチキナーゼ阻害剤は、単一のキナーゼ阻害剤を含んでもよい。マルチキナーゼ阻害剤は、リン酸化を阻害してもよい。マルチキナーゼ阻害剤は、タンパク質キナーゼの共有結合修飾因子として作用してもよい。マルチキナーゼ阻害剤は、キナーゼ活性部位に結合してもよいし、プロテインキナーゼ活性を阻害する二次又は三次部位に結合してもよい。マルチキナ

50

ーゼ阻害剤は、抗がん性マルチキナーゼ阻害剤であってもよい。例示的な抗癌剤マルチキナーゼ阻害剤には、ダサチニブ、スニチニブ、エルロチニブ、ペバシズマブ、バタラニブ、ベムラフェニブ、バンデタニブ、カボザンチニブ、ポアチニブ、アキシチニブ、ルキシリチニブ、レゴラフェニブ、クリゾチニブ、ボスチニブ、セツキシマブ、ゲフィチニブ、イマチニブ、ラパチニブ、レンパチニブ、ムブリチニブ、ニロチニブ、パニツムマブ、パゾパニブ、トラスツズマブ、またはソラフェニブ。

【0045】

本明細書で使用される場合、用語「約」は、当業者が指定された値と合理的に類似すると考える、指定された値を含む数値の範囲を意味する。実施形態において、約とは、当技術分野で一般に許容される測定値を用いて標準偏差の範囲内であることを意味する。実施形態では、約とは、規定値の±10%に及ぶ範囲を意味する。実施形態において、約とは、指定された値を含む。

【0046】

開示される主題の側面は、対象からの試料中の1つまたは複数のRNA分子の分析に基づいて、対象における疾患状態、（例えば、癌の存在または不在）、および／または疾患の起源組織を検出するための方法を含む。いくつかの実施形態では、対象における疾患状態を検出する方法は、対象から生体検査試料を単離し、生体検査試料が複数の無細胞リボ核酸（cfRNA）分子を含み、生体検査試料からcfRNA分子を抽出すること、及び抽出されたcfRNA分子に対して配列決定手順を実行して、複数の配列リードを生成すること、フィルタリング手順を実行して、1つ以上の健康な細胞に由来する配列リードの除外集団、及び配列リードの非除外集団を生成すること、及び／又は非除外配列リードに対して定量化手順を実行することを含む。実施形態において、本方法は、定量化手順が閾値を超える値を生成する場合に、対象における疾患状態を検出することを含む。実施形態において、閾値を超える1つ以上の非除去配列リードを検出することは、（i）検出、（ii）バックグラウンド以上の検出、及び／又は（iii）状態を有しない対象における対応する配列リードのレベルよりも大きいレベルでの検出を含む。様々な実施形態において、閾値は、約2、3、4、5、6、7、8、または約またはちょうど9など、約1またはちょうど1から約またはちょうど10までの範囲の整数である。いくつかの実施形態では、閾値は、約または正確に0.1～約または正確に0.9、例えば約または正確に0.2、0.3、0.4、0.5、0.6、0.7または約または正確に0.8までの範囲の、非整数の値である。

【0047】

いくつかの実施形態では、方法は、生物学的試験サンプルから抽出されるcfRNA分子を検出及び定量化するための配列決定手順の使用を含む。例えば、様々な実施形態において、配列決定手順は、cfRNA分子に逆転写手順を実行して複数のcDNA/RNAハイブリッド分子を生成すること、ハイブリッド分子のRNAを分解して複数の一本鎖cDNA分子テンプレートを生成すること、及び、複数の一本鎖cDNA分子テンプレートから複数の二本鎖DNA分子を合成することを含む。一本鎖cDNA分子テンプレートから複数の二本鎖DNA分子を合成し、複数の二本鎖DNAアダプターを複数の二本鎖DNA分子に結紮して配列決定ライブラリーを生成し、配列決定ライブラリーの少なくとも一部に対して配列決定手順を実行して複数の配列リードを取得するステップと、を含む。様々な実施形態において、二本鎖DNA分子を合成することは、鎖置換逆転写酵素手順を実行することを含む。

【0048】

いくつかの実施形態では、本方法は、全トランスクリプトームシーケンス手順を利用する。他の実施形態では、配列決定手順は、標的配列決定手順を含み、配列決定ライブラリーを調製する前に、1つ以上のcfRNA分子が生体検査試料から濃縮される。この実施形態によれば、疾患状態を示す1つ以上のcfRNA分子が、濃縮の対象とされる。例えば、いくつかの実施形態では、1つ以上の標的化cfRNA分子は、以下からなる群から選択される1つ以上の遺伝子に由来する。agr2、bpifa1、casp14、csn1s1、disp2、eif2d、fabp7、gabrg1、gnat3、grhl2、hoxc10、idi2-AS1、krt16p2、lalba、linc00163、nkx2-1、opn1sw、padi3、ptprz1、ros1、S100A7、SCGB2A2、SERP

INB5、SFTA3、SFTPA2、SLC34A2、TFF1、VTCN1、WFDC2、MUC5B、SMIM22、CXCL17、RNU1-1、及びKLK5であり、これらの任意の組み合わせを含むことが可能である。いくつかの実施形態では、1つ以上の標的RNA分子は、ROS1、NKX2-1、GGTLC1、SLC34A2、SFTPA2、BPIFA1、SFTA3、GABRG1、AGR2、GNAT3、MUC5B、SMIM22、CXCL17及びWFDC2からなる群より選択される1つ以上の遺伝子に由来し、これらの任意の組み合わせを含むことが可能である。いくつかの実施形態では、1つ以上の標的RNA分子は、SCGB2A2、CSN1S1、VTCN、FABP7、LALBA、RNU1-1、OPN1SW、CASP14、KLK5、およびWFDC2からなる群から選択される1つ以上の遺伝子に由来し、その任意の組合せからなることが可能である。いくつかの実施形態では、1つ以上の標的RNA分子は、CASP14、CRABP2、FABP7、SCGB2A2、SERPINB5、TRGV10、VGLL1、TFF1、およびAC007563.5からなる群より選択される1つ以上の遺伝子に由来し、それらの任意の組合せからなることが可能である。さらに他の実施形態では、標的化RNA分子は、AKR1B10、C3、および／またはPIEXO2遺伝子（複数可）に由来するものである。

10

【0049】

開示される主題の側面は、健常者の血漿における発現が非常に低いか又は存在しない、1つ以上の暗黒チャネルRNA分子の分析を含む。健常者の血漿におけるその低い発現レベルのために、ダークチャネルRNA分子は、本発明の方法と組み合わせて使用することができる高い信号対雑音比を提供する。

【0050】

20

開示される主題のいくつかの態様は、1つ以上の健康な細胞に由来する配列リードの除外集団、及びその後の分析で使用される配列リードの非除去集団を生成するために使用されるフィルタリング手順を含む。様々な実施形態において、フィルタリング手順は、生体試験サンプルから抽出されたcfRNA分子からの各配列リードをRNA配列の対照データセットと比較し、RNA配列の対照データセット中の1つ又は複数の配列リードに一致する1つ又は複数の配列リードを識別し、RNA配列の対照データセット中の1つ又は複数の配列リードに一致する各配列リードを配列リードの除外集団に配置することを含む。

【0051】

いくつかの実施形態では、RNA配列の対照データセットは、1人または複数の健常被験者から得られた複数の配列読み取り値を含む。いくつかの実施形態では、RNA配列の対照データセットは、対象からの複数の血液細胞から得られた複数の配列リードを含む。例えば、いくつかの実施形態では、複数の配列リードは、被験者の白血球（WBC）から取得される。

30

【0052】

生物試料

【0053】

様々な実施形態において、本開示は、その中の複数の核酸（例えば、複数のcfRNA分子）を分析する目的で、対象から試験試料、例えば、組織及び／又は体液試料などの生物学的試験試料を得ることを含む。本発明の実施形態に従った試料は、臨床的に許容される任意の方法で収集することができる。複数の核酸を含むと疑われる任意の試料を、本発明の方法と組み合わせて使用することができる。いくつかの実施形態では、試料は、組織、体液、又はそれらの組合せを含むことができる。いくつかの実施形態では、生物学的試料は、健康な対象から採取される。いくつかの実施形態では、生体試料は、特定の疾患又は障害（例えば、特定の癌又は腫瘍）を有することが知られている対象から採取される。いくつかの実施形態では、生物学的試料は、特定の疾患又は障害を有することが疑われる対象から収集される。

40

【0054】

本明細書で使用される場合、「組織」という用語は、連結した細胞及び／又は細胞外マトリックス材料（複数可）の塊を指す。本発明の方法と併せて一般的に使用される組織の非限定的な例としては、例えば、ヒトまたは非ヒト哺乳動物に由来する、皮膚、毛髪、指

50

爪、子宮内膜組織、鼻腔組織、中枢神経系（CNS）組織、神経組織、眼組織、肝臓組織、腎組織、胎盤組織、乳腺組織、胃腸組織、筋骨格組織、生殖器組織、骨髄、および同様のものが挙げられる。本発明の実施形態に従う組織サンプルは、例えば、限定されないが、ホルマリン固定パラフィン包埋（FFPE）、新鮮、及び新鮮凍結（FF）組織サンプルのような、当該技術分野で公知の任意の組織サンプルタイプの形態で調製及び提供され得る。

【0055】

本明細書で使用する場合、「体液」及び「生体液」という用語は、対象、例えば、ヒト又はヒト以外の哺乳動物に由来する液体物質を指す。本発明の方法と関連して一般的に使用される体液の非限定的な例としては、粘液、血液、血漿、血清、血清誘導体、滑液、リンパ液、胆汁、痰、唾液、汗、涙、痰、羊水、月経液、膣液、精液、尿、脳脊髄液（CSF）、腰髄液や脳室髄液などの脳脊髄液、胃液、鼻腔、咽頭、頬の綿棒から得られた1つ以上の材料からなる液体試料、腹膜、胃、胸郭、管状の洗浄処置などの洗浄処置から得られた1つ以上の材料からなる液体試料、などなどである。

【0056】

いくつかの実施形態では、試料は、微細針吸引物または生検された組織を含むことができる。いくつかの実施形態では、試料は、細胞又は生物学的材料を含む培地からなることができる。いくつかの実施形態では、試料は、血栓、例えば、血清を除去した後の全血から得られた血栓から構成することができる。いくつかの実施形態では、試料は、便を含むことができる。好ましい一実施形態では、サンプルは全血を採取したものである。一態様では、血漿、赤血球、白血球、及び血小板など、全血サンプルの一部のみが使用される。いくつかの実施形態では、試料は、本方法と関連して、2つ以上の構成部分に分離される。例えば、いくつかの実施形態では、全血試料は、血漿、赤血球、白血球、及び血小板の各成分に分離される。

【0057】

いくつかの実施形態では、サンプルは、サンプルが採取された対象のみならず、サンプリング時に対象内に存在するウイルスDNA／RNAなど、1つ以上の他の生物からの複数の核酸を含む。

【0058】

核酸は、当技術分野で知られている任意の適切な方法に従って試料から抽出することができ、抽出された核酸は、本明細書に記載される方法と組み合わせて利用することができる。例えば、Maniatis, et al., Molecular Cloning.を参照されたい。A Laboratory Manual, Cold Spring Harbor, N.Y., pp.280-281, 1982を参照し、その内容は、参照によりその全体が本明細書に組み込まれる。好ましい一実施形態では、無細胞リボ核酸（例えば、cfRNA）が試料から抽出される。

【0059】

実施形態では、試料は「マッチド」または「ペア」試料である。一般に、「マッチドサンプル」及び「ペアサンプル」という用語は、同じ対象から、好ましくはほぼ同時に（例えば、単一の処置又は診察の一部として、又は同じ日に）採取された異なるタイプのサンプルのペアを指す。実施形態では、異なる種類の試料は、組織試料（例えば、切除試料又は生検試料のような癌組織）及び生体液試料（例えば、血液又は血液画分）である。用語はまた、マッチした試料（例えば、マッチした生体液試料からの無細胞ポリヌクレオチドと対になった、癌組織から抽出されたポリヌクレオチド）、またはその配列決定リードを指すために使用されてもよい。実施形態では、癌バイオマーカーの同定などにおいて、複数のペア試料が分析される。複数のペアサンプルは、異なる時期に収集された同一個体から（例えば、がんの初期段階からのペアサンプル、及びがんの後期段階からのペアサンプルのように）、同じ時期又は異なる時期に異なる個体から、又はこれらの組合せであってもよい。実施形態において、マッチしたサンプルは、異なる被験者からのものである。実施形態において、複数の一致した試料は、同じ癌の種類、及び任意に同じ癌の病期を有する対象からである。

10

20

30

40

50

【0060】

アッセイプロトコル例

【0061】

図1は、一実施形態による配列決定のための核酸試料を調製するための方法100のフローチャートである。方法100は、以下のステップを含むが、これに限定されるものではない。例えば、方法100の任意のステップは、品質管理又は当業者に知られている他の実験室アッセイ手順のための定量サブステップを構成することができる。

【0062】

ステップ110では、被験者からリボ核酸（RNA）試料が抽出される。RNAサンプルは、ヒトトランスクリプトーム全体、またはヒトトランスクリプトームの任意のサブセットから構成されてもよい。試料は、疾患（例えば、癌）を有することが知られている、又は有することが疑われる対象から抽出されてもよい。サンプルは、血液、血漿、血清、尿、糞便、唾液、他の種類の体液、又はそれらの任意の組み合わせを含んでもよい。いくつかの実施形態では、血液サンプルを採取するための方法（例えば、注射器又は指刺し）は、外科手術を必要とし得る組織生検を得るための手順よりも侵襲性が低い場合がある。抽出された試料は、cfDNAをさらに含んでもよい。被験者が疾患（例えば、癌）を有する場合、抽出サンプル中のcfRNAは、診断のために検出可能なレベルで存在する可能性がある。

10

【0063】

ステップ120において、RNA分子を含む核酸試料は、任意にDNase酵素で処理される。DNaseは、核酸サンプルからDNA分子を除去し、RNA分子のDNA汚染を低減することができる。RNA分子がDNAに変換された後、RNA変換されたDNAと核酸サンプルに元々含まれていたゲノムDNAとを区別することが困難な場合がある。DNaseを適用することで、cfRNAに由来する分子を標的として増幅することができる。DNase工程は、DNaseバッファーを加えるステップ、遠心分離機を使用してDNaseを適用した試料を混合するステップ、及びインキュベーションを含んでもよい。いくつかの実施形態では、ステップ120は、Qiagen QIAamp Circulating Nucleic Acid Handbookに記載されたDNase処理プロトコルに基づく1つ又は複数の工程を含む。

20

【0064】

ステップ130では、逆転写酵素を用いて、核酸サンプル中のRNA分子を相補的DNA（cDNA）に変換する。逆転写酵素プロセスは、第一鎖合成ステップ（逆転写によるcDNA鎖の生成）、RNA鎖の分解による一本鎖cDNA分子の生成、およびポリメラーゼを用いた一本鎖cDNA分子からの二本鎖DNA分子の合成を含んでもよい。一本鎖合成では、プライマーがRNA分子の3'末端にアニーリングする。第二鎖合成では、異なるプライマーがcDNA分子の3'末端にアニーリングする。

30

【0065】

ステップ140において、配列決定ライブラリーが調製される。例えば、当該技術分野において周知であるように、アダプターをdsDNA分子の片端または両端にライゲーションして、配列決定のためのライブラリーを調製することができる。一実施形態では、利用されるアダプターは、その後のクラスター生成及び／又は配列決定において使用するための1つ又は複数の配列決定オリゴヌクレオチド（例えば、合成による配列決定（SBS）（イルミナ、サンディエゴ、カリフォルニア）において使用するための既知のP5及びP7配列）を含んでもよい。別の実施形態では、アダプターは、ライブラリー調製後、ライブラリーを個々のサンプルから調製された1つまたは複数の他のライブラリーと組み合わせることができ、それによって多重配列決定を可能にするような、サンプル特異的インデックス配列を含む。サンプル特異的インデックス配列は、約2nt～約または正確に20nt、約2nt～約または正確に10nt、約2nt～約または正確に8nt、または約2nt～約または正確に6ntの長さを有する短いオリゴヌクレオチド配列からなることが可能である。別の実施形態では、サンプル特異的インデックス配列は、長さが約2、3、4、5、6、7、または8ヌクレオチド（nt）よりも大きい短いオリゴヌクレオチド配列から構成され得る。

40

50

【0066】

オプションとして、ライブラリー調製中に、アダプターライゲーションを通じてサンプル中の核酸分子にユニーク分子識別子（UMI）を付加することができる。UMIは、アダプターライゲーション中に核酸断片の一方または両方の末端に付加される短い核酸配列（例えば、4～10塩基対）である。いくつかの実施形態では、UMIは、特定の核酸断片に由来する配列読み取りを識別するために使用することができる固有のタグとして機能する縮退塩基対である。アダプターライゲーションに続くPCR増幅の間、UMIは、付着した核酸断片とともに複製され、これは、下流の分析において同じ元の核酸分子に由来する配列リードを識別する方法を提供する。

【0067】

RNAの標的配列決定を含む実施形態では、ステップ150において、標的核酸配列がライブラリーから濃縮される。濃縮の間、ハイブリダイゼーションプローブ（本明細書では「プローブ」とも呼ばれる）は、疾患（例えば、癌）、疾患状態（例えば、癌状態）、または疾患分類（例えば、癌タイプまたは起源組織）の有無について情報提供する核酸断片を標的とし、引き落とすために使用される。所定のワークフローにおいて、プローブは、標的（相補的）核酸鎖（例えば、RNAから変換されたDNA鎖）にアニール（またはハイブリダイズ）するように設計されてもよい。プローブは、10s、100s、または1000sの塩基対の長さの範囲であってよい。一実施形態では、プローブは、特定の癌又は他の種類の疾患に対応することが疑われるゲノム（例えば、ヒト又は他の生物の）の特定の標的領域を分析するための遺伝子パネルに基づいて設計される。さらに、プローブは、標的領域の重複する部分をカバーしてもよい。他の実施形態では、標的化RNA分子は、逆転写酵素（図示せず）を用いてRNA分子をcDNA鎖に変換する前に、ハイブリダイゼーションプローブを用いて濃縮することができる。一般に、当技術分野における任意の既知の方法を使用して、プローブハイブリダイゼーション標的核酸を単離し、そして濃縮することができる。例えば、当技術分野でよく知られているように、ストレプトアビジンでコーティングされた表面（例えば、ストレプトアビジンでコーティングされたビーズ）を用いてプローブとハイブリダイズした標的核酸の単離を容易にするために、ビオチン部分をプローブの5'末に加える（すなわち、ビオチン化）ことが可能である。

【0068】

さらに、標的配列決定のために、ステップ160において、配列リードが濃縮核酸試料から生成される。配列決定データは、当技術分野において既知の手段により、濃縮されたDNA配列（すなわち、RNA配列から誘導された、または変換されたDNA配列）から取得されてもよい。例えば、方法100は、合成技術（Illumina）、パイロシーケンス（454 Life Sciences）、イオン半導体技術（Ion Torrent sequencing）、1分子リアルタイムシーケンス（Pacific Biosciences）、ライゲーションによるシーケンス（SOLiD sequencing）、ナノポアシーケンス（Oxford Nanopore Technologies）、又はペアエンドシーケンスを含む次世代シーケンス（NGS）技術を含んでもよい。いくつかの実施形態では、可逆的色素ターミネーターを有するsequencing-by-synthesisを用いて、超並列シーケンシングを行う。

【0069】

他の実施形態では、例えば、全トランスクリプトーム配列決定アプローチにおいて（例えば、標的配列決定の代わりに）、ステップ170で、豊富なRNA種が核酸サンプルから枯渇される。例えば、いくつかの実施形態では、リボソームRNA（rRNA）及び／又はトランスファーRNA（tRNA）種が枯渇され得る。RiboMinus（登録商標）（ThermoFisher Scientific）又はAnyDeplete（NuGen）などの入手可能な市販キットを、豊富なRNA種の枯渇のために使用することができる。実施形態では、豊富なRNA分子に由来する核酸（例えば、変換されたDNA）の枯渇後、ステップ180で配列リードが生成される。

【0070】

いくつかの実施形態では、配列リードは、アライメント位置情報を決定するために、当

10

20

30

40

50

該技術分野において既知の方法を用いて参照ゲノムにアライメントされてもよい。アライメント位置情報は、所定の配列リードの開始ヌクレオチド塩基及び終了ヌクレオチド塩基に対応する参照ゲノム内の領域の開始位置及び終了位置を示してもよい。アライメント位置情報は、開始位置及び終了位置から決定され得る配列リードの長さを含んでもよい。参照ゲノム中の領域は、遺伝子又は遺伝子のセグメントに関連してもよい。参照ゲノムは、トランスクリプトーム全体、またはその任意の部分（例えば、複数の標的転写産物）から構成されてもよい。別の実施形態では、参照ゲノムは、試験される生物からの全ゲノムであり得、抽出されたRNA分子由来の（またはそこから逆転写された）配列リードは、位置、断片長、および／または開始位置と終了位置を決定するために参照ゲノムに整列される。例えば、一実施形態では、配列リードは、ヒト参照ゲノムhg19にアライメントされる。ヒト参照ゲノムhg19の配列は、参照番号GRCh37／hg19でGenome Reference Consortiumから入手可能であり、Santa Cruz Genomics Instituteが提供するGenome Browserからも入手可能である。アライメント位置情報は、所定の配列リードの開始塩基及び終了塩基に対応する参照ゲノム中の領域の開始位置及び終了位置を示していてもよい。アライメント位置情報は、開始位置及び終了位置から決定され得る配列リードの長さを含んでもよい。参照ゲノム中の領域は、遺伝子または遺伝子のセグメントに関連していてもよい。

10

【0071】

ダークチャネルRNA分子の同定

【0072】

20

本開示の態様は、被験体における疾患状態を示す1つ以上のRNA配列（又は「ダークチャネルRNA分子」）を同定するためのコンピュータ実装方法を含む。いくつかの実施形態では、方法は、コンピュータシステムによって、疾患を有することが知られている対象から得られた第1の試験試料から複数のRNA分子から配列読み取り値の第1のセットを得ることを含み、第1の試験試料は、複数の無細胞RNA（cfRNA）分子を含んでいる。及び対照試料からの複数のRNA分子からの第2の配列読み取り組を検出し、第1の配列読み取り組に存在し、第2の配列読み取り組に存在しない1つ又は複数のRNA配列を検出し、疾患状態を示す1つ又は複数のRNA配列を同定する。いくつかの実施形態では、患者から得られた第1の試験サンプルは、体液（例えば、血液、血漿、血清、尿、唾液、胸水、心臓液、脳脊髄液（CSF）、腹膜液、またはそれらの任意の組み合わせ）からなる。好ましい一実施形態では、患者から得られた試験サンプルは、血漿サンプルである。いくつかの実施形態では、対照試料は、対象からの健康な細胞（例えば、白血球）から得られた複数のRNA分子を含んでいる。

30

【0073】

図2は、本開示の一実施形態による、疾患状態を示す1つ又は複数のRNA配列を特定する方法を示すフロー図である。図2に示すように、ステップ210において、複数の無細胞RNA（cfRNA）分子を含む生物学的試験サンプルから配列リードの第1のセットが取得される。無細胞を含む生体検査サンプルは、任意の体液、例えば、血液、血漿、血清、尿、胸水、脳脊髄液、涙、唾液、又は腹水であり得る。本実施形態によれば、cfRNA生物学的試験サンプルは、疾患を有することが知られている、または疾患を有することが疑われる試験対象から得られ、サンプルからcfRNA分子が抽出され、配列読み取り値が決定される（本明細書の他の箇所に記載されるように）。例えば、一実施形態では、cDNA／RNAハイブリッド分子を生成する逆転写ステップを用いて相補的DNA鎖を合成し、RNA分子を分解し、ポリメラーゼを用いてcDNA鎖から二本鎖DNA分子を合成し、配列決定ライブラリを作成し、配列決定プラットフォームを用いて配列リードを測定する。配列決定ステップは、当該技術分野における任意の既知の配列決定プラットフォーム、例えば、合成による配列決定プラットフォーム（例えば、イルミナのHiSeq X）またはライゲーションによる配列決定プラットフォーム（例えば、ライフテクノロジーズのSOLiDプラットフォーム）、Ion Torrent／Ion Proton、半導体配列決定、ロシュ454、単一分子配列決定プラットフォーム（例えば、ヘリコス、パシフィック・バイオサイエンスおよび

40

50

ナノコア)、以前に記載したようなものなど、任意の超並列配列決定プラットフォームを使用して実施されうる。あるいは、配列リードを検出し定量化するための他の手段を用いることができ、例えば、アレイベースハイブリダイゼーション、プローブベース溶液内ハイブリダイゼーション、ライゲーションベースアッセイ、プライマー伸長反応アッセイは、当業者が容易に理解するように、DNA分子(例えば、RNA分子から変換される)から配列リードを決定するために用いることができる。

【0074】

ステップ220において、配列読み取りの第2のセットは、健康な対照試料から得られる。一実施形態では、健常対照サンプルは、同じ対象からであり、複数の細胞RNA分子を含む。例えば、対照試料は、白血球などの血液細胞であり得、複数の配列読み取りは、血液細胞から抽出されたRNA分子に由来する。この実施形態に従って、RNA分子は、健康な対照試料(例えば、血球)から抽出され、DNAに変換され、配列決定ライブラリーが調製され、配列読み取りの第2のセットが決定される(本明細書の他の箇所に記載されるように)。他の実施形態では、健康な対照サンプルは、健康な対象から得られたRNA配列について決定された配列データのデータベース、又は健康な細胞から得られた配列データであり得る。

【0075】

ステップ230において、配列読みの第1のセット及び配列読みの第2のセットからの配列読みを比較して、疾患状態を示す1つ又は複数のRNA分子を同定する。さらに、配列読み取りの第1のセットに存在し、配列読み取りの第2のセットに存在しない1つ以上の配列読み取り(RNA分子由来)が、疾患状態を示すRNA分子由来であると同定される。例えば、配列リードの第1のセットは、疾患(例えば、癌)を有することが知られている、又は有することが疑われる対象から得られた血漿試料からのcfRNA分子由来の配列リードから構成され得る。そして、第2のシーケンスリードのセットは、健康な細胞(例えば、白血球)からのRNA分子由来のシーケンスリードから構成され得る。無細胞RNAサンプルに由来する配列リードの第1のセットから、健康な細胞に由来する配列リードの第2のセットを比較し、除去することによって、疾患状態(例えば、癌)に由来する配列リードを特定することができる。

【0076】

いくつかの実施形態では、RNA配列の対照データセットは、1人以上の健康な被験者から得られた複数の配列リードを含む。様々な実施形態において、配列リードの第2のセットは、公共データベースから取得されたRNA配列情報を含んでいる。本発明の実施形態に従って使用することができる公共データベースには、組織RNA-seqデータベースGTEX(gtexportal.org/homeで入手可能)が含まれる。いくつかの実施形態では、RNA配列の対照データセットは、対象からの複数の血液細胞から得られた複数の配列リードを含む。例えば、いくつかの実施形態では、複数の配列リードは、被験体の白血球(WBC)から得られる。

【0077】

腫瘍由来RNA分子の検出

【0078】

本開示の態様は、対象における1つ又は複数の腫瘍由来RNA分子を検出するためのコンピュータ実装方法を含む。いくつかの実施形態において、本方法は以下を含む。コンピュータシステムによって、腫瘍を有することが知られている対象からの第1の試験試料から複数のRNA分子から配列読み取り値の第1のセットを取得することであって、第1の試験試料が複数の無細胞RNA(cfRNA)分子を含む、ことと、コンピュータシステムによって、対象からの複数の血球から複数のRNA分子から配列読み取り値の第2のセットを取得することである。及び／又は、コンピュータシステムによって、配列読み取り値の第1のセットに存在し、配列読み取り値の第2のセットに存在しない1つ又は複数のRNA配列を検出し、対象における1つ又は複数の腫瘍由来RNA分子を検出することである。

【0079】

いくつかの実施形態では、患者から得られた第1の試験サンプルは、血液、血漿、血清、尿、唾液、胸水、心嚢液、脳脊髄液（CSF）、腹膜液、又はそれらの任意の組み合わせからなる。好ましい一実施形態では、患者から得られた試験サンプルは、血漿サンプルである。いくつかの実施形態では、被験者から得られた複数の血球は、白血球（WBC）である。

【0080】

図3は、本発明の一実施形態による、1つ以上の腫瘍由来RNA配列を同定する方法を示すフロー図である。ステップ310において、複数の無細胞RNA（cfRNA）分子を含む生物学的試験サンプルから配列リードの第1のセットが取得される。本実施形態によれば、cfRNA生物学的試験サンプルは、疾患を有することが知られている、又は疾患を有することが疑われる試験対象から得られ、サンプルからcfRNA分子が抽出され、配列リードが（本明細書の他の箇所に記載されるように）決定される。例えば、一実施形態では、cDNA/RNAハイブリッド分子を生成する逆転写ステップを用いて相補的DNA鎖を合成し、RNA分子を分解し、ポリメラーゼを用いてcDNA鎖から二本鎖DNA分子を合成し、配列決定ライブラリを作成し、配列決定プラットフォームを用いて配列リードを測定する。配列決定ステップは、先に記載したように、当該技術分野において既知の配列決定プラットフォームを使用して、何れでも実施することができる。あるいは、配列リードを決定するための他の手段を用いることができ、例えば、当業者であれば容易に理解できるように、アレイベースハイブリダイゼーション、プローブベース溶液中ハイブリダイゼーション、ライゲーションベースアッセイ、プライマー伸長反応アッセイを用いて、DNA分子（例えば、RNA分子から変換された）から得られる配列リードを検出および／または定量することができる。

10

20

【0081】

ステップ315において、配列リードの第2のセットは、血液細胞（例えば、白血球又はバフィーコート）から得られる。一実施形態では、血球は、同じ対象から得られ、そこから抽出されたRNA分子が用いられる。この実施形態に従って、RNA分子は、血球から抽出され、DNAに変換され、配列決定ライブラリーが調製され、配列リードの第2のセットが決定される（本明細書の他の箇所に記載されるように）。一般に、試験試料から無細胞核酸を抽出及び精製するために、当技術分野における任意の既知の方法を使用することができる。例えば、無細胞核酸は、QIAamp循環核酸キット（Qiagen）などの1つ以上の既知の市販のプロトコルまたはキットを使用して抽出および精製することができる。

30

【0082】

ステップ320において、1つ以上のRNA配列が配列読みの第1のセットに存在し、配列読みの第2のセットに存在しない場合に、1つ以上の腫瘍由来のRNA分子が検出される。さらに、配列リードの第1のセットに存在し、配列リードの第2のセットに存在しない1つ以上の配列リード（RNA分子由来）は、疾患状態を示すRNA分子由来であると同定される。例えば、配列リードの第1のセットは、疾患（例えば、癌）を有することが知られている、又は有することが疑われる対象から得られた血漿試料からのcfRNA分子由来の配列リードから構成され得る。そして、第2のシーケンスリードのセットは、血球（例えば、白血球）由来のRNA分子由来のシーケンスリードから構成され得る。無細胞RNAサンプルに由来する配列リードの第1のセットから、血球に由来する配列リードの第2のセットを比較し、除去することにより、腫瘍に由来する配列リードを同定することができる。

40

【0083】

ダークチャネルRNA分子を用いた疾患状態の検出

【0084】

図4は、本発明の一実施形態による、対象における癌の存在の検出、癌の状態の決定、癌の進行の監視、及び／又は癌の種類の決定のための方法を示す流れ図である。ステップ410において、生物学的検査試料が対象から抽出される。前述のように、一実施形態では、試験サンプルは、複数の無細胞RNA分子を含む体液（例えば、血液、血漿、血清、尿

50

、唾液、胸水、心嚢液、脳脊髄液（CSF）、腹膜液、またはそれらの任意の組み合わせ）であり得る。

【0085】

ステップ415において、複数の無細胞RNA分子が試験サンプルから抽出され、配列決定ライブラリーが調製される。一般に、当技術分野における任意の既知の方法を使用して、試験試料から無細胞核酸を抽出及び精製することができる。例えば、無細胞核酸（cfRNA分子）は、QIAamp循環核酸キット（Qiagen）などの1つ以上の既知の市販のプロトコルまたはキットを使用して抽出および精製することができる。抽出後、cfRNA分子は、配列決定ライブラリーを調製するために使用される。一実施形態では、逆転写ステップを使用して、複数のcDNA／RNAハイブリッド分子を生成し、RNA鎖を分解して一本鎖cDNA分子を生成し、第2の鎖を合成して一本鎖cDNA分子テンプレートから複数の二本鎖DNA分子を生成し、DNAアダプターを複数の二本鎖DNA分子にライゲートして配列決定ライブラリーを作製する。先に記載したように、DNAアダプターは、その後のクラスター生成及び／又は配列決定において使用するための1つ又は複数の配列決定オリゴヌクレオチド（例えば、合成による配列決定（SBS）（イルミナ、サンディエゴ、カリフォルニア）において使用するための既知のP5及びP7配列）を含んでもよい。別の実施形態では、アダプターは、ライブラリー調製後、ライブラリーを個々のサンプルから調製された1つまたは複数の他のライブラリーと組み合わせることができ、それによって多重配列決定を可能にするような、サンプル特異的インデックス配列を含む。別の実施形態では、ユニークな分子識別子（UMI）が、アダプターライゲーションを通じて追加される。

10

20

【0086】

ステップ420において、複数の配列リードを生成するために配列決定反応が行われる。一般に、配列決定ライブラリーから配列データ又は配列リードを取得するために、当技術分野において既知の任意の方法を使用することができる。例えば、一実施形態では、配列決定ライブラリーからの配列データまたは配列リードは、次世代配列決定（NGS）を用いて取得することができる。次世代シーケンシング方法としては、例えば、合成技術によるシーケンシング（Illumina）、パイロシーケンス（454）、イオン半導体技術（Ion Torrent sequencing）、1分子リアルタイムシーケンス（Pacific Biosciences）、ライゲーションによるシーケンシング（SOLiD sequencing）、及びナノポアシーケンス（Oxford Nanopore Technologies）等が挙げられる。いくつかの実施形態では、配列決定は、可逆的色素ターミネーターを有するsequencing-by-synthesisを使用する超並列配列決定である。他の実施形態では、配列決定は、sequencing-by-ligationである。さらに他の実施形態では、シーケンシングは、単一分子シーケンシングである。さらに他の実施形態では、シーケンシングはペアエンドシーケンシングである。任意選択で、増幅ステップを配列決定の前に実施することができる。

30

【0087】

ステップ425では、cfRNAサンプルから得られた配列リードをフィルタリングして、除外されない配列リードのリストを生成し、ステップ430では、除外されない配列リードを定量化する。例えば、本明細書の他の箇所に記載されているように、cfRNAサンプルから得られた配列リードは、健康な細胞に存在することが知られている配列を除外するようにフィルタリングされ得る。一実施形態では、健康な細胞（例えば、白血球）から抽出されたRNA分子は、非除去配列リードを得るためにcfRNA由来の配列リードから除外された配列リードを導出する配列決定が行われる。別の実施形態では、データベース（例えば、公開データベース）からのRNA配列決定データを使用して、健康な細胞のリードに存在することが知られている配列をフィルタリングまたは除外して、除外されていない配列リードを得るために構成することが可能である。

40

【0088】

ステップ435において、定量化された非除去配列リードが閾値を超える場合に、疾患状態が検出される。様々な実施形態において、閾値は、約2、3、4、5、6、7、8、または約または約または約9など、約1または正確に1から約または正確に10までの範囲の整

50

数である。いくつかの実施形態では、閾値は、約または正確に0.1～約または正確に0.9、例えば約または正確に0.2、0.3、0.4、0.5、0.6、0.7または約または正確に0.8までの範囲の、非整数の値である。

【0089】

本開示の態様は、癌を有することが知られている、または癌を有することが疑われる対象において、癌の存在を検出し、癌の病期を決定し、癌の進行を監視し、および／または癌の種類を決定するための方法に関するものである。いくつかの実施形態において、本方法は以下を含む。(a) 対象からの生体検査試料中の複数のcfRNA分子から複数の配列読み取り値を得ること；(b) 生体検査試料中の1つ以上のRNAマーカー由来の1つ以上の配列の存在を定量的に検出して腫瘍RNAスコアを決定し、ここで1つ以上のRNAマーカーは1つ以上の標的RNA分子からなる群から選択されること。及び(c) 腫瘍RNAスコアが閾値を超える場合に、対象における癌の存在を検出すること、癌の病期を決定すること、癌の進行を監視すること、及び／又は癌の種類を決定することである。様々な実施形態において、閾値は、約2、3、4、5、6、7、8、または約または約または約9など、約1またはちょうど1から約またはちょうど10の範囲にある整数である。いくつかの実施形態では、閾値は、約または正確に0.1～約または正確に0.9、例えば約または正確に0.2、0.3、0.4、0.5、0.6、0.7または約または正確に0.8までの範囲の、非整数の値である。

10

【0090】

本開示の実施形態に従った定量的検出方法は、次世代シーケンシングなどの核酸シーケンシング手順を含むことができる。様々な実施形態において、配列決定には、全トランスクリプトーム配列決定が含まれ得る。様々な実施形態において、配列決定は、配列決定手順を実施する前に、関心のある1つまたは複数の標的RNA配列についてサンプルを濃縮することを含むことができる。あるいは、配列リードを検出および定量化するための他の手段を使用することができ、例えば、当業者であれば容易に理解するように、アレイベースハイブリダイゼーション、プローブベース溶液内ハイブリダイゼーション、ライゲーションベースアッセイ、プライマー延長反応アッセイは、DNA分子（例えば、RNA分子から変換される）から配列リードを決定するために使用することができる。

20

【0091】

図5は、本開示の別の実施形態による、1つ以上の標的RNA分子由来の1つ以上の配列読み取り値から疾患状態を検出する方法を示す流れ図である。ステップ510において、複数の無細胞RNA分子を含む生物学的試験サンプルが取得される。一実施形態では、生体検査サンプルは、体液（例えば、血液、血漿、血清、尿、唾液、胸水、心嚢液、脳脊髄液（CSF）、腹膜液サンプル、又はそれらの任意の組合せ）である。

30

【0092】

ステップ515において、生物学的試験サンプル中の1つ以上の標的RNA分子に由来する1つ以上の核酸配列の存在を検出し、定量化して、腫瘍RNAスコアを決定する。本明細書の他の箇所に記載されているように、RNA分子に由来する核酸は、当該技術分野において任意の既知の手段を用いて検出及び定量化することができる。例えば、一実施形態に従って、RNA分子由来の核酸は、次世代シーケンシングプラットフォーム（例えば、HiSeq又はNovaSeq、イルミナ、サンディエゴ、カリフォルニア州）等のシーケンシング手順を使用して検出及び定量化される。他の実施形態では、RNA分子に由来する核酸は、マイクロアレイ、逆転写PCR、リアルタイムPCR、定量的リアルタイムPCR、デジタルPCR、デジタル液滴PCR、デジタルエマルジョンPCR、多重PCR、ハイブリッド捕捉、オリゴヌクレオチドライゲーションアッセイ、又はそれらの任意の組み合わせを使用して検出及び定量化される。他の場所で説明したように、一実施形態では、無細胞核酸（cfRNA分子）は、QIAamp循環核酸キット（Qiagen）などの1つ以上の既知の市販のプロトコル又はキットを使用して抽出及び精製することができる。抽出後、cfRNA分子は、配列決定ライブラリーを調製するために使用される。一実施形態では、逆転写ステップを使用して、複数のcDNA／RNAハイブリッド分子を生成し、RNA鎖を分解して一本鎖cDN

40

50

A分子を生成し、第2の鎖を合成して一本鎖cDNA分子テンプレートから複数の二本鎖DNA分子を生成する。任意に、1つの実施形態において、1つ以上の標的RNA分子（またはそれに由来するDNA分子）は、本明細書の他の箇所に記載されるように、検出および定量の前に濃縮される。

【0093】

一実施形態では、腫瘍RNAスコアは、検出された標的RNA分子（または標的RNA分子に由来するDNA分子から得られた配列読み取り値）の量または数である。別の実施形態では、腫瘍RNAスコアは、検出された標的化RNA分子（又は標的化RNA分子由来のDNA分子から得られた配列読み取り値）の総数をRNA分子が標的化される遺伝子の総数で除した平均、最頻値、又は平均からなる。さらに他の実施形態では、腫瘍RNAスコアは、配列読み取りを予測モデルに入力することによって決定され、腫瘍RNAスコアは、本明細書の他の箇所に記載されるように、尤度又は確率として出力される。

10

【0094】

ステップ520において、腫瘍RNAスコアが閾値を超える場合、対象において、癌の存在が検出され、癌の状態が決定され、癌の進行が監視され、及び／又は癌の種類が決定される。閾値は、約2、3、4、5、6、7、8、又は約またはちょうど9など、約1又はちょうど1から約またはちょうど10までの範囲の整数であり得る。いくつかの実施形態では、閾値は、約または正確に0.1～約または正確に0.9、例えば約または正確に0.2、0.3、0.4、0.5、0.6、0.7または約または正確に0.8まで及ぶ非整数の値である。あるいは、腫瘍RNAスコアが予測モデルから出力される場合、その出力は単に、対象が癌、または癌タイプを有する可能性または確率を示す尤度または確率であり得る。

20

【0095】

がん指標スコア

【0096】

本開示の側面は、患者における癌の存在を検出するためのコンピュータ実装方法に向けられている。いくつかの実施形態では、方法は：プロセッサ及びコンピュータ可読媒体を含むコンピュータでデータセットを受信することであって、データセットが、生物学的標的における複数の核酸分子（例えば、患者からの生体検査サンプル中の複数の標的リボ核酸（RNA）分子に由来するDNA分子）であり、コンピュータ可読媒体は、プロセッサによって実行されると、コンピュータに以下を行わせる命令を含んでいる。生物学的検査試料から複数の標的化RNA分子の発現レベルを決定すること；標的化RNA分子の各々の発現レベルをRNA組織スコア行列と比較して、各標的化RNA分子の癌指標スコアを決定すること；各標的化RNA分子の癌指標スコアを集約して生物学的検査試料の癌指標スコアを生成すること；及び生物学的検査試料の癌指標スコアが閾値を超える場合に患者における癌の存在を検出すること；をコンピュータに行わせる命令を含む、請求項1に記載の方法。

30

【0097】

いくつかの実施形態では、標的RNA分子は、既知の癌の状態を有する患者において、健康な患者におけるその発現レベルを超える発現レベルを有する。様々な実施形態において、既知の癌の状態を有する患者における標的RNA分子の発現レベルは、健康な患者における標的RNA分子の発現レベルよりも、約2倍～約または正確に10倍、例えば、約3倍、4倍、5倍、6倍、7倍、8倍、または約または正確に9倍大きい範囲である。様々な実施形態において、標的RNA分子は、健康な患者からの生体検査サンプルにおいて検出されない、すなわち、標的RNA分子は検出不能な発現レベルを有する。

40

【0098】

いくつかの実施形態では、生物学的試験サンプル中の標的RNA分子の数は、約1又は正確に1～約又は正確に2000、約又は正確に10～約又は正確に1000、約又は正確に10～約又は正確に500、又は約又は正確に10～約又は正確に500の範囲である。他の実施形態では、標的RNA分子の数は、約2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、または約または正確に20など、約または正確に1～約また

50

は正確に50、約または正確に1～約または正確に40、約または正確に1～約または正確に30、または約または正確に1～約または正確に20の範囲である。

【0099】

いくつかの実施形態では、がん指標スコアは、生物学的試験サンプルから検出された標的RNA分子（または標的RNA分子に由来するDNA分子から得られた配列読み取り値）の総数の総計からなる。別の実施形態では、がん指標スコアは、検出された標的化RNA分子（または標的化RNA分子由来のDNA分子から得られた配列読み取り値）の総数をRNA分子が標的とされる遺伝子の総数で割った平均、最頻値、または平均値からなる。さらに他の実施形態では、本明細書の他の箇所に記載されるように、配列読み取りを予測モデルに入力し、癌指標スコアを尤度又は確率として出力することによって、癌指標スコアが決定される。

10

【0100】

いくつかの実施形態では、閾値は、約1～約10の範囲の整数であり、例えば、約2、3、4、5、6、7、8、または約またはちょうど9のような整数である。いくつかの実施形態では、閾値は、約または丁度0.2、0.3、0.4、0.5、0.6、0.7または約または丁度0.8など、約または丁度0.1～約または丁度0.9まで及ぶ非整数の値である。他の実施形態では、閾値は、約1、1.5、2、2.5、3、3.5、4、または約または正確に4.5RPMなどの、約または正確に0.5から約または正確に5RPM（reads per million）の範囲である。癌口ケータースコア閾値は、例えば健康な対象または既知の疾患状態を有する対象などの対照試料において検出された標的RNA分子（またはそれに由来する配列読み取り値）の量に基づいて決定することができる。また、予測モデルから癌口ケータースコアが出力される場合、その出力は、被験者が癌を有する可能性又は癌の種類を示す尤度又は確率とすることもできる。

20

【0101】

図6は、本開示の一実施形態による、癌指標スコアに基づいて対象における癌の存在を検出する方法を示す流れ図である。ステップ610において、生物学的試験サンプル中の複数のcfRNA分子に由来する複数の配列読み取りを含むデータセットが受信される。例えば、複数の配列読み取りは、本明細書に記載されるように、生物学的試験サンプルから抽出された複数のcfRNA分子について決定することができる。さらに、cfRNA分子を逆転写してDNA分子を作成し、DNA分子を配列決定して配列リードを生成する。

30

【0102】

ステップ615において、生物学的試験サンプル中の複数の標的RNA分子について発現レベルが決定される。例えば、一実施形態では、標的RNA分子の発現レベルは、関心のある1つまたは複数の標的RNA分子に由来する検出配列リードの定量化に基づいて決定することができる。

【0103】

ステップ620では、各標的RNA分子の発現レベルをRNA組織スコアマトリックスと比較して、各標的RNA分子の癌指標スコアを決定する。RNA組織スコアマトリックスは、癌の状態が既知の複数の癌トレーニングサンプルに由来する配列リードを含むトレーニングセットから決定することができる。

40

【0104】

ステップ625において、各標的RNA分子の癌指標スコアを集計して、癌指標スコアを生成する。いくつかの実施形態では、癌指標スコアは、生体検査サンプルから検出された標的RNA分子（または標的RNA分子に由来するDNA分子から得られた配列読み取り値）の総数の集計値からなる。別の実施形態では、がん指標スコアは、検出された標的化RNA分子（または標的化RNA分子由来のDNA分子から得られた配列読み取り値）の総数をRNA分子が標的化された遺伝子の総数で割った平均、最頻値、または平均からなる。

【0105】

ステップ630において、試験サンプルの癌指標スコアが閾値を超える場合に、対象における癌の存在を検出する。上述したように、一実施形態では、閾値は、約2、3、4、5

50

、6、7、8、又は約または正確に9など、約または正確に1～約または正確に10の範囲にある整数である。いくつかの実施形態では、閾値は、約または正確に0.1～約または正確に0.9、例えば、約または正確に0.2、0.3、0.4、0.5、0.6、0.7または約または正確に0.8まで及ぶ非整数の値である。他の実施形態では、閾値は、約1、1.5、2、2.5、3、3.5、4、または約または正確に4.5RPMなどの約1または正確に0.5から約または正確に5RPM (Reads per million) の範囲である。

【0106】

本開示の態様は、1つ以上の標的RNA分子の発現レベル、1つ以上の標的RNA分子に対する癌指標スコア、生物学的検査試料に対する癌指標スコア、またはそれらの任意の組み合わせに基づいて、患者における癌細胞型または癌の起源組織を決定するための方法を含む。様々な実施形態において、方法は、1つ以上の標的RNA分子の発現レベル、1つ以上の標的RNA分子に対する癌指標スコア、生物学的検査試料に対する癌指標スコア、又はそれらの任意の組み合わせに基づいて、患者を複数の治療カテゴリーのうちの1つ以上に治療的に分類することを更に含む。

10

【0107】

様々な実施形態において、コンピュータは、1つ以上の標的RNA分子の発現レベル、1つ以上の標的RNA分子の癌指標スコア、生体検査サンプルの癌指標スコア、患者における癌の有無の表示、患者における癌の起源組織の癌細胞タイプの表示、患者の治療分類、又はそれらの任意の組み合わせを含むレポートを生成するように構成されている。

【0108】

20

RNA組織マトリックススコア

【0109】

本開示の態様は、RNA組織スコアマトリックスを構築するための方法を含む。いくつかの実施形態では、方法は、複数の患者から得られた複数のRNA配列読み取りをコンパイルしてRNA発現行列を生成すること、及び／又はRNA発現行列を組織特異的RNA発現行列で正規化し、RNA組織スコア行列を構築することを含む。いくつかの実施形態では、組織特異的RNA発現マトリックスは、複数の参照ヒト組織を含んでいる。様々な実施形態において、RNA配列読み取り値は、健康なRNA組織スコア行列を構築するために、複数の健康な患者から得られる。様々な実施形態において、RNA配列リードは、癌RNA組織スコアマトリックスを構築するために、既知の癌タイプを有する複数の患者から取得される。

30

【0110】

RNAマーカーと解析技術

【0111】

本開示のいくつかの実施形態に従った方法は、cfRNA分子及び／又はctRNA分子に対して実施することができる。いくつかの実施形態では、主題の方法で使用されるRNA分子は、がん性細胞及び非がん性細胞からのRNA分子を含む。

【0112】

実施形態では、方法は、(a) 複数の標的cfRNA分子、またはそのcDNA分子を濃縮して、ポリヌクレオチドの濃縮サンプルを生成すること；および／または (b) 濃縮サンプルのポリヌクレオチド、またはその増幅産物を配列決定することを含み、複数の標的cfRNA分子が、表11の1つまたは複数の転写産物から選択される。実施形態では、複数の標的cfRNA分子は、表8または12～15の1つ以上から選択される（例えば、表8または11～14の1つ以上からの5、10、15、または20の遺伝子の転写物である）。実施形態では、複数のcfRNA分子を測定することは、配列決定などによる検出又は測定の前に、複数のcfRNA分子（又はそのcDNA分子）を濃縮することを含む。実施形態において、測定される標的cfRNA分子は、500以下の遺伝子（例えば、400以下、300以下、200以下、100以下、又は50以下の遺伝子）からのものである。

40

【0113】

実施形態では、方法は以下を含む。(a) 対象のサンプル中の複数の標的無細胞RNA (c

50

fRNA) 分子を測定し、複数の標的cfRNA分子が表8又は11～14の転写物から選択される；及び(b) 癌を検出し、癌を検出することが、閾値以上の標的cfRNA分子の1以上を検出することから構成される。実施形態では、複数の標的cfRNA分子は、表8又は11～14の1つ以上に記載された少なくとも2、3、4、5、6、7、8、9、10、11、12、13、15、16、17、18、19、20、25、30、35、40、50又はそれ以上の遺伝子から選択された転写産物である。標的cfRNA分子は、これらの表のいずれか1つから選択される遺伝子、またはそれらの任意の組み合わせからであることができる。実施形態において、表8又は11～14の中から選択される表の数は、2、3、4、又は全ての表、又はそれらの任意の組み合わせである。実施形態において、複数のcfRNA分子を測定することは、全トランスクリプトームシーケンスを構成しない。実施形態において、複数のcfRNA分子を測定することは、配列決定などによる検出又は測定の前に、複数のcfRNA分子（又はそのcDNA分子）を濃縮することを含んでいる。実施形態において、測定される標的cfRNA分子は、500未満の遺伝子（例えば、400、300、200、100、又は50未満の遺伝子）からのものである。

10

【0114】

いくつかの実施形態では、1つ以上の標的cfRNA分子は、表1に記載される遺伝子から選択される1つ以上の遺伝子に由来する。実施形態では、1つ以上の標的cfRNA分子は、表1からの少なくとも2、3、4、5、6、7、8、9、10、15、20、25、又は30の遺伝子を含む。実施形態において、1つ以上の標的cfRNA分子は、表1からの少なくとも5つの遺伝子を含む。実施形態では、1つ以上の標的cfRNA分子は、表1からの少なくとも100個の遺伝子を含む。実施形態では、1つ以上の標的cfRNA分子は、表1からの遺伝子を全て含む。実施形態では、1つ以上の標的cfRNA分子は、表1の最初の5つの遺伝子（AGR2、HOXC10、S100A7、BPIFA1、及び／又はIDI2-AS1）、並びに任意に表1の1つ以上の追加の遺伝子のうちの少なくとも1つを含む。実施形態では、1つまたは複数の標的cfRNA分子は、AGR2遺伝子の転写物を含む。実施形態では、1つ以上の標的cfRNA分子は、AGR2、HOXC10、S100A7、BPIFA1、およびIDI2-AS1の転写産物を含む。実施形態において、測定される標的cfRNA分子は、500未満の遺伝子（例えば、400、300、200、100、又は50未満の遺伝子）からのものである。以下の表1は、癌暗黒チャンネルバイオマーカーの例を提供する。

20

【0115】

30

[表1]

【0116】

いくつかの実施形態では、1つ以上の標的cfRNA分子は、表2に記載される遺伝子から選択される1つ以上の遺伝子に由来する。実施形態では、1つ以上の標的cfRNA分子は、表2からの少なくとも2、3、4、5、6、7、8、9、又は10個の遺伝子を含む。実施形態において、1つ以上の標的cfRNA分子は、表2からの少なくとも5つの遺伝子を含む。実施形態では、1つ以上の標的cfRNA分子は、表2からの少なくとも10個の遺伝子を含む。実施形態では、1つ以上の標的cfRNA分子は、表2からの遺伝子を全て含む。実施形態では、1つ以上の標的cfRNA分子は、表2の最初の5つの遺伝子（ROS1、NKX2-1、GGTLC1、SLC34A2、およびSFTPA2）の少なくとも1つ、ならびに任意選択で表2からの1つ以上の追加の遺伝子を含む。実施形態において、1つまたは複数の標的cfRNA分子は、ROS1遺伝子の転写物を含む。実施形態では、1つ以上の標的cfRNA分子は、ROS1、NKX2-1、GGTLC1、SLC34A2、およびSFTPA2の転写産物を含む。実施形態において、測定される標的cfRNA分子は、500未満の遺伝子（例えば、400、300、200、100、又は50未満の遺伝子）からのものである。以下の表2は、ダークチャンネル肺がんバイオマーカーの例を提供する。

40

【0117】

[表2]

【0118】

いくつかの実施形態では、1つ又は複数の標的cfRNA分子は、表3に記載された遺伝子

50

から選択される1つ又は複数の遺伝子に由来する。実施形態では、1つ以上の標的cfRNA分子は、表3からの少なくとも2、3、4、5、6、7、8、又は9個の遺伝子を含む。実施形態において、1つ以上の標的cfRNA分子は、表3からの少なくとも5つの遺伝子を含む。実施形態では、1つ以上の標的cfRNA分子は、表3からの遺伝子を全て含む。実施形態では、1つ以上の標的cfRNA分子は、表3の最初の5つの遺伝子（SCGB2A2、CSN1S1、VTCN1、FABP7、及びLALBA）の少なくとも1つ、及び任意で表3からの1つ以上の追加の遺伝子を含む。実施形態では、1つまたは複数の標的cfRNA分子は、SCGB2A2遺伝子の転写物を含む。実施形態では、1つ以上の標的cfRNA分子は、SCGB2A2、CSN1S1、VTCN1、FABP7、及びLALBAの転写産物を含む。実施形態において、測定される標的cfRNA分子は、500未満の遺伝子（例えば、400、300、200、100、又は50未満の遺伝子）からのものである。以下の表3は、乳癌暗黒チャネルバイオマーカーの例を提供する。

10

【0119】

[表3]

【0120】

いくつかの実施形態では、1つ以上の標的cfRNA分子は、表4に記載される遺伝子から選択される1つ以上の遺伝子に由来する。実施形態では、1つ以上の標的cfRNA分子は、表4から少なくとも2、3、4、又は5個の遺伝子を含む。実施形態では、1つ以上の標的cfRNA分子は、表4から少なくとも5つの遺伝子を含む。実施形態では、1つ以上の標的cfRNA分子は、表4からの遺伝子を全て含む。実施形態では、1つ以上の標的cfRNA分子は、表4の最初の5つの遺伝子（CASP14、CRABP2、FABP7、SCGB2A2、及びSERPINB5）の少なくとも1つ、及び任意で表4の1つ以上の追加的な遺伝子を含む。実施形態では、1つまたは複数の標的cfRNA分子は、CASP14遺伝子の転写物を含む。実施形態では、1つまたは複数の標的cfRNA分子は、CASP14、CRABP2、FABP7、SCGB2A2、およびSERPINB5の転写物を含む。実施形態において、測定される標的cfRNA分子は、500未満の遺伝子（例えば、400、300、200、100、又は50未満の遺伝子）からのものである。以下の表4は、本明細書に記載されるように、heteroDE法を使用して同定された乳癌バイオマーカーの例を提供する。

20

【0121】

[表4]

【0122】

いくつかの実施形態では、1つ以上の標的cfRNA分子は、表5に記載される遺伝子から選択される1つ以上の遺伝子に由来する。実施形態では、1つ以上の標的cfRNA分子は、表5からの少なくとも2、3、4、5、6、7、8、9、10、15、20、又は25個の遺伝子を含む。実施形態において、1つ以上の標的cfRNA分子は、表5からの少なくとも5つの遺伝子を含む。実施形態では、1つ以上の標的cfRNA分子は、表5からの少なくとも10個の遺伝子を含む。実施形態では、1つ以上の標的cfRNA分子は、表5からの遺伝子を全て含む。実施形態では、1つ以上の標的cfRNA分子は、表5の最初の5つの遺伝子（PTPRZ1、AGR2、SHANK1、PON1、及びMYO16__AS1）の少なくとも1つを含み、任意で表5の1つ以上の追加の遺伝子が含まれる。実施形態では、1つまたは複数の標的cfRNA分子は、PTPRZ1遺伝子の転写物を含む。実施形態では、1つ以上の標的cfRNA分子は、PTPRZ1、AGR2、SHANK1、PON1、及びMYO16__AS1の転写物を含む。実施形態において、測定される標的cfRNA分子は、500未満の遺伝子（例えば、400、300、200、100、又は50未満の遺伝子）からのものである。以下の表5は、本明細書に記載されるように、情報利得法を用いて同定された肺癌バイオマーカーの例を提供する。

30

【0123】

[表5]

【0124】

いくつかの実施形態では、1つ以上の標的cfRNA分子は、表6に記載される遺伝子から選択される1つ以上の遺伝子に由来する。実施形態では、1つ以上の標的cfRNA分子は、

40

50

表6からの少なくとも2、3、4、5、6、7、8、9、10、15、20、又は25個の遺伝子を含む。実施形態において、1つ以上の標的cfRNA分子は、表6からの少なくとも5つの遺伝子を含む。実施形態では、1つ以上の標的cfRNA分子は、表6からの少なくとも10個の遺伝子を含む。実施形態では、1つ以上の標的cfRNA分子は、表6からの遺伝子を全て含む。実施形態では、1つ以上の標的cfRNA分子は、表6の最初の5つの遺伝子（ADARB2、HORMAD2、SPDYE18、RPS19、及びCYP4F35P）の少なくとも1つ、及び任意で表6からの1つ以上の追加の遺伝子を含む。実施形態において、1つまたは複数の標的cfRNA分子は、ADARB2遺伝子の転写物を含む。実施形態では、1つ以上の標的cfRNA分子は、ADARB2、HORMAD2、SPDYE18、RPS19、およびCYP4F35Pの転写産物を含む。実施形態において、測定される標的cfRNA分子は、500未満の遺伝子（例えば、

10

【0125】

[表6]

【0126】

いくつかの実施形態では、1つ以上の標的cfRNA分子は、表7に記載される遺伝子から選択される1つ以上の遺伝子に由来する。実施形態では、1つ以上の標的cfRNA分子は、表7からの少なくとも2、3、4、5、6、7、8、9、又は10個の遺伝子を含む。実施形態では、1つ以上の標的cfRNA分子は、表7から少なくとも5つの遺伝子を含む。実施形態では、1つ以上の標的cfRNA分子は、表7からの少なくとも10個の遺伝子を含む。実施形態では、1つ以上の標的cfRNA分子は、表7からの遺伝子を全て含む。実施形態では、1つ以上の標的cfRNA分子は、表7の最初の5つの遺伝子（S100A7、FOXA1、BARX2、MMP7、及びPLEKHG4B）のうちの少なくとも1つ、及び任意で表7からの1つ以上の追加の遺伝子を含む。実施形態では、1つまたは複数の標的cfRNA分子は、S100A7遺伝子の転写物を含む。実施形態では、1つ以上の標的cfRNA分子は、S100A7、FOXA1、BARX2、MMP7、及びPLEKHG4Bの転写産物を含む。実施形態において、測定される標的cfRNA分子は、500未満の遺伝子（例えば、400、300、200、100、又は50未満の遺伝子）からのものである。以下の表7は、癌組織において比較的高いレベルで発現されるダークチャネル癌バイオマーカーの例を提供する。

20

30

【0127】

[表7]

【0128】

いくつかの実施形態では、1つ以上の標的cfRNA分子は、表11に記載される遺伝子から選択される1つ以上の遺伝子に由来する。実施形態において、1つ以上の標的cfRNA分子は、表11からの少なくとも2、3、4、5、10、25、50、100、150、200、300、又は400の遺伝子を含む。実施形態において、1つ以上の標的cfRNA分子は、表11からの少なくとも5つの遺伝子を含む。実施形態では、1つ以上の標的cfRNA分子は、表11から少なくとも25個の遺伝子を含む。実施形態では、1つ以上の標的cfRNA分子は、表11からの少なくとも100個の遺伝子を含む。実施形態では、1つ以上の標的cfRNA分子は、表11からの少なくとも200個の遺伝子を含む。実施形態では、1つ以上の標的cfRNA分子は、表11からの少なくとも300の遺伝子を含む。実施形態では、1つ以上の標的cfRNA分子は、表11からの遺伝子を全て含む。実施形態において、測定される標的cfRNA分子は、500未満の遺伝子（例えば、400、300、200、100、又は50未満の遺伝子）からのものである。以下の表11に、癌バイオマーカーの例を示す。

40

【0129】

[表8]

【0130】

いくつかの実施形態では、1つ以上の標的cfRNA分子は、表12に記載される遺伝子から選択される1つ以上の遺伝子に由来する。実施形態では、1つ以上の標的cfRNA分子は

50

、表12からの少なくとも2、3、4、5、10、20、30、40、50、又は60個の遺伝子を含む。実施形態では、1つ以上の標的cfRNA分子は、表12から少なくとも5つの遺伝子を含む。実施形態では、1つ以上の標的cfRNA分子は、表12からの少なくとも10個の遺伝子を含む。実施形態では、1つ以上の標的cfRNA分子は、表12から少なくとも25個の遺伝子を含む。実施形態では、1つ以上の標的cfRNA分子は、表12から少なくとも50個の遺伝子を含む。実施形態では、1つ以上の標的cfRNA分子は、表12からの遺伝子を全て含む。実施形態では、測定される標的cfRNA分子は、500未満の遺伝子（例えば、400、300、200、100、又は50未満の遺伝子）からのものである。以下の表12は、肺癌バイオマーカーの例を示す。

【0131】

10

[表9]

【0132】

様々な実施形態において、1つ以上の標的cfRNA分子は、表13に記載される遺伝子から選択される1つ以上の遺伝子に由来する。実施形態では、1つ以上の標的cfRNA分子は、表13からの少なくとも2、3、4、5、10、20、30、40、50、60、又は70個の遺伝子を含む。実施形態において、1つ以上の標的cfRNA分子は、表13からの少なくとも5つの遺伝子を含む。実施形態では、1つ以上の標的cfRNA分子は、表13からの少なくとも10個の遺伝子を含む。実施形態では、1つ以上の標的cfRNA分子は、表13からの少なくとも25個の遺伝子を含む。実施形態では、1つ以上の標的cfRNA分子は、表13からの少なくとも50個の遺伝子を含む。実施形態では、1つ以上の標的cfRNA分子は、表13からの遺伝子を全て含む。実施形態では、測定される標的cfRNA分子は、500未満の遺伝子（例えば、400、300、200、100、又は50未満の遺伝子）からのものである。以下の表13は、乳癌バイオマーカーの例を示す。

20

【0133】

[表10]

【0134】

いくつかの実施形態では、1つ以上の標的cfRNA分子は、表14に記載される遺伝子から選択される1つ以上の遺伝子に由来する。実施形態では、1つ以上の標的cfRNA分子は、表14からの少なくとも2、3、4、5、10、15、20、又は30個の遺伝子を含む。実施形態では、1つ以上の標的cfRNA分子は、表14から少なくとも5つの遺伝子を含む。実施形態では、1つ以上の標的cfRNA分子は、表14からの少なくとも10個の遺伝子を含む。実施形態では、1つ以上の標的cfRNA分子は、表14から少なくとも25個の遺伝子を含む。実施形態では、1つ以上の標的cfRNA分子は、表14からの遺伝子を全て含む。実施形態において、測定される標的cfRNA分子は、500未満の遺伝子（例えば、400、300、200、100、又は50未満の遺伝子）からのものである。実施形態において、閾値を超えて検出される複数の標的cfRNA分子は、以下からなる群から選択される複数の遺伝子に由来するcfRNA分子である。adipoq、agr3、ankrd30a、aqp4、bpifa1、ca12、ceacam5、cftr、cxcl17、cyp4f8、fabp7、foxy1、ggtlc1、gp2、IL20、Itih6、ldlradi1、lemd1、lmx1b、mmp7、next、NKAIN1、NKX2-1、ROPN1、ROS1、SCGB1D2、SCGB2A2、SFTA2、SFTA3、SLC34A2、SOX9、STK32A、STMND1、TFAP2A、TFAP2B、TFF1、TRPV6、VGLL1、及びVTCN1。以下の表14は、情報量の多い癌バイオマーカーの例である。

30

40

【0135】

[表11]

【0136】

いくつかの実施形態では、1つ以上の標的cfRNA分子は、表15に記載される遺伝子から選択される1つ以上の遺伝子に由来する。実施形態において、1つ以上の標的cfRNA分子は、表15からの少なくとも2、3、4、5、10、25、50、100、150、200、300、又は400の遺伝子を含む。実施形態において、1つ以上の標的cfRNA分子は、表15からの少なくとも5つの遺伝子を含む。実施形態では、1つ以上の標的cfRNA分子は、表15

50

からの少なくとも25個の遺伝子を含む。実施形態では、1つ以上の標的cfRNA分子は、表15からの少なくとも100個の遺伝子を含む。実施形態では、1つ以上の標的cfRNA分子は、表15からの少なくとも200個の遺伝子を含む。実施形態では、1つ以上の標的cfRNA分子は、表15から少なくとも300の遺伝子を含む。実施形態では、1つ以上の標的cfRNA分子は、表15からの遺伝子を全て含む。実施形態では、測定される標的cfRNA分子は、500未満の遺伝子（例えば、400、300、200、100、又は50未満の遺伝子）からのものである。

【0137】

[表12]

【0138】

10

実施形態において、1つ以上の標的cfRNA分子は、表1～6の1つ以上（例えば、2、3、5、またはそれ以上の遺伝子）から選択される1つ以上の遺伝子と組み合わせて表8または11～14の1つ以上（例えば、2、3、5、またはそれ以上の遺伝子）から選択される1つ以上の遺伝子から得られる。実施形態において、1つ以上の標的cfRNA分子は、表8または11～14の1つ以上から選択される1つ以上の遺伝子（例えば、2、3、5、またはそれ以上の遺伝子）と表7から選択される1つ以上の遺伝子（例えば、2、3、5、またはそれ以上の遺伝子）の組み合わせから誘導される。実施形態において、表8又は11～14から選択される表は、表11である。実施形態において、表8又は11～14から選択される表は、表12である。実施形態において、表8または11～14から選択される表は、表13である。実施形態において、表8又は11～14から選択される表は、表14である。実施形態において、表8又は11～14から選択される表は、表13である。実施形態において、表8または11～14から選択される表は、表8である。実施形態において、第1及び第2のテーブルからの遺伝子の選択は、第1及び第2のテーブルの両方において1つ以上の遺伝子を選択することを含む。実施形態において、第1及び第2の表からの遺伝子の選択は、第2の表にない第1の表から1つ以上の遺伝子を選択すること、及び第1の表にない第2の表から1つ以上の遺伝子を選択することを含む。実施形態において、測定される標的cfRNA分子は、500未満の遺伝子（例えば、400、300、200、100、又は50未満の遺伝子）からのものである。

20

【0139】

実施形態において、癌は肺癌であり、閾値を超えて検出された複数の標的cfRNA分子は、表2、5、又は12のうちの1つ以上（例えば、2、3、5、又はそれ以上の遺伝子）の転写産物から選択される。実施形態において、1つ以上の標的cfRNA分子は、表2、5、及び12の各々から選択される1つ以上の遺伝子（例えば、2、3、5、又はそれ以上の遺伝子）に由来する。実施形態において、第1及び第2の表からの遺伝子の選択は、第1及び第2の表の両方において1つ又は複数の遺伝子を選択することを含む。実施形態において、第1及び第2の表からの遺伝子の選択は、第2の表にない第1の表から1つ又は複数の遺伝子、及び第1の表にない第2の表から1つ又は複数の遺伝子を選択することからなる。実施形態において、測定される標的cfRNA分子は、500未満の遺伝子（例えば、400、300、200、100、又は50未満の遺伝子）からのものである。

30

【0140】

実施形態において、癌は乳癌であり、閾値を超えて検出された複数の標的cfRNA分子は、表3、4、6、又は13のうちの1つ以上（例えば、2、3、5、又はそれ以上の遺伝子）の遺伝子の転写産物から選択される。実施形態において、1つ以上の標的cfRNA分子は、表3、4、6、及び13の各々から選択される1つ以上の遺伝子（例えば、2、3、5、又はそれ以上の遺伝子）に由来する。実施形態において、第1及び第2の表からの遺伝子の選択は、第1及び第2の表の両方において1つ又は複数の遺伝子を選択することを含む。実施形態において、第1及び第2の表からの遺伝子の選択は、第2の表にない第1の表から1つ又は複数の遺伝子、及び第1の表にない第2の表から1つ又は複数の遺伝子を選択することからなる。実施形態において、測定される標的cfRNA分子は、500未満の遺伝子（例えば、400、300、200、100、又は50未満の遺伝子）からのものである。

40

50

【0141】

実施形態において、1つ以上の標的cfRNA分子は、(a)表5または表6から選択される1つ以上の遺伝子（例えば、2、3、5、またはそれ以上の遺伝子）、および／または(b)表7から選択される1つ以上の遺伝子（例えば、2、3、5、またはそれ以上の遺伝子）と組み合わせて表11から選択される1つ以上の遺伝子から誘導される。実施形態において、第1及び第2の表からの遺伝子の選択は、第1及び第2の表の両方において1つ又は複数の遺伝子を選択することを含む。実施形態において、第1及び第2の表からの遺伝子の選択は、第2の表にない第1の表から1つ又は複数の遺伝子、及び第1の表にない第2の表から1つ又は複数の遺伝子を選択することからなる。実施形態において、測定される標的cfRNA分子は、500未満の遺伝子（例えば、400、300、200、100、又は50未満の遺伝子）からのものである。

10

【0142】

実施形態において、1つ以上の標的cfRNA分子は、(a)表5から選択される1つ以上の遺伝子（例えば、2、3、5、またはそれ以上の遺伝子）、及び／または(b)表7から選択される1つ以上の遺伝子（例えば、2、3、5、またはそれ以上の遺伝子）と組み合わせて表12から選択される1つ以上の遺伝子から誘導される。実施形態において、第1及び第2の表からの遺伝子の選択は、第1及び第2の表の両方において1つ又は複数の遺伝子を選択することを含む。実施形態において、第1及び第2の表からの遺伝子の選択は、第2の表にない第1の表から1つ又は複数の遺伝子、及び第1の表にない第2の表から1つ又は複数の遺伝子を選択することからなる。実施形態において、測定される標的cfRNA分子は、500未満の遺伝子（例えば、400、300、200、100、又は50未満の遺伝子）からのものである。

20

【0143】

実施形態において、1つ以上の標的cfRNA分子は、表13から選択される1つ以上の遺伝子（例えば、2、3、5、またはそれ以上の遺伝子）に由来し、(a)表4から選択される1つ以上の遺伝子（例えば、2、3、5またはそれ以上の遺伝子）、(b)表6から選択される1つまたは複数の遺伝子（例えば、2、3、5またはそれ以上の遺伝子）、および／または(c)表7から選択される1つまたは複数の遺伝子（例えば、2、3、5またはそれ以上の遺伝子）。実施形態において、第1及び第2の表からの遺伝子の選択は、第1及び第2の表の両方において1つ又は複数の遺伝子を選択することを含む。実施形態において、第1及び第2の表からの遺伝子の選択は、第2の表にない第1の表から1つ又は複数の遺伝子、及び第1の表にない第2の表から1つ又は複数の遺伝子を選択することからなる。実施形態において、測定される標的cfRNA分子は、500未満の遺伝子（例えば、400、300、200、100、又は50未満の遺伝子）からのものである。

30

【0144】

実施形態において、1つ以上の標的cfRNA分子は、表4から選択される1つ以上の遺伝子（例えば、2、3、5、またはそれ以上の遺伝子）に由来し、(a)表3から選択される1つ以上の遺伝子（例えば、2、3、5、又はそれ以上の遺伝子）、(b)表6から選択される1つ又は複数の遺伝子（例えば、2、3、5、又はそれ以上の遺伝子）、及び／又は(c)表7から選択される1つ又は複数の遺伝子（例えば、2、3、5、又はそれ以上の遺伝子）。実施形態において、第1及び第2の表からの遺伝子の選択は、第1及び第2の表の両方において1つ又は複数の遺伝子を選択することを含む。実施形態において、第1及び第2の表からの遺伝子の選択は、第2の表にない第1の表から1つ又は複数の遺伝子、及び第1の表にない第2の表から1つ又は複数の遺伝子を選択することからなる。実施形態において、測定される標的cfRNA分子は、500未満の遺伝子（例えば、400、300、200、100、又は50未満の遺伝子）からのものである。

40

【0145】

いくつかの実施形態では、1つ以上の標的cfRNA分子は、表8に記載される遺伝子から選択される1つ以上の遺伝子に由来する。実施形態では、1つ以上の標的cfRNA分子は、表8から少なくとも2、3、4、5、10、15、20、又は30の遺伝子を含む。実施形態で

50

は、1つ以上の標的cfRNA分子は、表8からの少なくとも5つの遺伝子（例えば、最初の5つの遺伝子、CEACAM5、RH0V、SFTA2、SCGB1D2、及びIGF2BP1）を含む。実施形態では、1つ以上の標的cfRNA分子は、表8からの少なくとも10個の遺伝子を含む。実施形態では、1つ以上の標的cfRNA分子は、表8から少なくとも25個の遺伝子を含む。実施形態では、1つ以上の標的cfRNA分子は、表8からの遺伝子を全て含む。実施形態において、測定される標的cfRNA分子は、500未満の遺伝子（例えば、400、300、200、100、又は50未満の遺伝子）からのものである。実施形態において、閾値を超えて検出される複数の標的cfRNA分子は、以下からなる群から選択される複数の遺伝子に由来するcfRNA分子である。ceacam5、rhov、sfta2、scgb1d2、igf2bp1、sftpa1、ca12、sftpb、cdh3、muc6、src6a14、hoxc9、agr3、tmem125、tfap2b、irx2、potekp、agr3、tmem126、tfap2b、irx2、potekp、ARHGEF38、GPR87、LMX1B、ATP10B、NEL1、MUC21、SOX9、LINC00993、STMND1、ERVH48-1、SCTR、MAGEA3、MB、LEMD1、SIX4、およびNXNL2があげられる。以下の表8は、情報量の多いがんバイオマーカーの例である。

10

【0146】

[表13]

【0147】

実施形態において、閾値を超えて1つ以上の標的cfRNA分子を検出することは、(i) 検出、(ii) バックグラウンド以上の検出、又は(iii) 状態を有しない対象における標的cfRNA分子のレベルよりも大きいレベルでの検出を含む。実施形態において、閾値を超えて検出することは、検出を構成する。実施形態において、閾値より上に検出することは、閾値より上に検出することを構成する。実施形態において、閾値を超えて検出することは、状態を有していない被験者における標的cfRNA分子のレベルよりも大きいレベルで検出することを含んでいる。

20

【0148】

実施形態において、閾値を超える1つ以上の標的cfRNA分子を検出することは、状態を有さない被験者におけるレベルの少なくとも約10倍又は正確に10倍（例えば、15、20、50、100、又はそれ以上）であるレベルで1つ以上の標的cfRNA分子を検出することを含んでいる。実施形態において、閾値以上の検出は、状態を有さない被験者のレベルよりも少なくとも約または正確に25倍大きいレベルで1つまたは複数の標的cfRNA分子を検出することを含む。実施形態では、1つ以上の標的cfRNA分子を、状態を有さない被験者のレベルの少なくとも約または正確に50倍であるレベルで検出することを含む、閾値を超える検出。

30

【0149】

実施形態において、閾値を超えて1つ以上の標的cfRNA分子を検出することは、約1、1.5、2、2.5、3、3.5、4、または約4もしくは正確に4.5RPMなどの0.5~5リード/ミリオン（RPM）閾値を超えての検出から構成されている。実施形態において、閾値を超えて検出することは、1RPMを超えて検出することを含んでいる。実施形態において、閾値より上の検出は、1RPMより上の検出からなる。実施形態において、閾値より上の検出は、2RPMより上の検出からなる。実施形態において、閾値より上の検出は、5RPMより上の検出からなる。

40

【0150】

病気と障害

【0151】

本開示の実施形態に従う方法は、心血管疾患、肝臓疾患、又は癌を含むがこれらに限定されない様々な疾患又は状態のいずれかの存在又は不在を検出するために使用され得る。いくつかの実施形態において、本方法は、癌の病期を決定することを含む。いくつかの実施形態では、癌の病期は、I期癌、II期癌、III期癌、又はIV期癌である。

【0152】

いくつかの実施形態では、方法は、癌腫、肉腫、骨髄腫、白血病、リンパ腫、芽球腫、

50

胚細胞腫瘍、またはそれらの任意の組み合わせの存在または不在の検出、病期の決定、進行の監視、および／または分類を含む。いくつかの実施形態では、癌腫は、腺癌であってもよい。他の実施形態では、癌腫は、扁平上皮癌であってもよい。さらに他の実施形態では、癌腫は、小細胞肺癌、非小細胞肺、鼻咽頭、大腸、肛門、肝臓、膀胱、子宮頸、精巣、卵巣、胃、食道、頭頸部、すい臓、前立腺、腎臓、甲状腺、メラノーマ及び乳癌からなる群から選択される。いくつかの実施形態において、乳癌は、ホルモン受容体陰性乳癌又はトリプルネガティブ乳癌である。

【0153】

いくつかの実施形態では、方法は、肉腫の存在の検出、病期の決定、進行の監視、及び／又は分類を含む。実施形態において、肉腫は、骨肉腫、軟骨肉腫、平滑筋肉腫、横紋筋肉腫、中皮肉腫（中皮腫）、線維肉腫、血管肉腫、脂肪肉腫、神経膠腫、及び星細胞腫からなる群から選択され得る。さらに他の実施形態では、方法は、白血病の存在又は不在の検出、病期の決定、進行の監視、及び／又は分類を含む。様々な実施形態において、白血病は、骨髄性、顆粒球性、リンパ性、およびリンパ芽球性白血病からなる群から選択され得る。さらに他の実施形態では、方法は、リンパ腫の存在の検出、病期の決定、進行の監視、及び／又は分類を含む。様々な実施形態において、リンパ腫は、以下のものからなる群から選択され得る。ホジキンリンパ腫及び非ホジキンリンパ腫。

【0154】

本開示の態様は、疾患の起源組織を決定する方法を含み、起源組織は、脾臓組織、肝胆脾組織、肝臓組織、肺組織、脳組織、神経内分泌組織、子宮組織、腎臓組織、尿路上皮組織、腎臓組織、子宮頸部組織、乳房組織、脂肪、結腸組織、直腸組織、心臓組織、骨格筋組織、前立腺組織及び甲状腺組織からなる群から選択される。

【0155】

本開示の態様は、癌細胞型を決定する方法を含み、癌細胞型は、膀胱癌、乳癌、子宮頸癌、結腸直腸癌、子宮内膜癌、食道癌からなる群から選択される。胃がん、頭頸部がん、肝胆脾がん、血液がん、肝臓がん、肺がん、リンパ腫、メラノーマ、多発性骨髄腫、卵巣がん、脾臓がん、前立腺がん、腎臓がん、甲状腺がん、尿道がんおよび子宮がんからなる群から選択される、がん細胞のタイプを決定する方法。

【0156】

症状別治療法

【0157】

本明細書に開示された方法は、治療法の決定、指導、モニタリング、ならびに癌治療法の開発および臨床試験に使用することができる。例えば、治療効果は、分子標的療法（モノクローナル薬）、化学療法薬、放射線プロトコル等又はこれらの組み合わせ等の特定の療法による治療の前、間、及び後のサンプルにおける患者のcfRNAを比較することによって監視することができる。いくつかの実施形態では、cfRNAは、特定の癌バイオマーカーが治療後に増加又は減少するかどうかを確認するために監視され、これにより、医師は、従来の患者の症状を追跡する監視方法によって与えられるよりもはるかに短い期間で治療を変更（例えば、治療の継続、停止又は変更）することができる。いくつかの実施形態では、方法は、検出されたcfRNAバイオマーカーに関連する特定のステージ又はタイプの癌で対象を診断する、又は患者がそのような癌を有するか又は発症する可能性を報告するなど、RNA由来の配列に基づいて対象を診断する段階をさらに含む。実施形態において、本明細書に開示される方法は、検出された状態に基づいて治療法を選択することをさらに含む。実施形態では、選択された治療法は、対象に投与される。状態が癌、又は特定の癌の種類及び／若しくは病期である場合、適切な抗癌治療が選択され得る。抗癌治療の非限定的な例としては、放射線療法、外科的切除、抗癌剤（例えば、免疫療法剤、化学療法剤等）の投与、又はこれらの1つ以上の組み合わせが挙げられる。

【0158】

分類モデル

【0159】

本開示の態様は、分類モデルに向けられている。例えば、機械学習又は深層学習モデル（例えば、疾患分類器）は、1つ以上のRNA分子又は配列リード（1つ以上のcfRNA分子由来）から決定される1つ以上の特徴の値に基づいて疾患状態を決定するために使用され得る。様々な実施形態において、機械学習または深層学習モデルの出力は、疾患状態の予測スコアまたは確率（例えば、予測がんスコア）である。したがって、機械学習又は深層学習モデルは、予測スコア又は確率に基づいて、疾患状態分類を生成する。

【0160】

いくつかの実施形態では、機械学習されたモデルは、ロジスティック回帰分類器を含む。他の実施形態では、機械学習又は深層学習モデルは、決定木、アンサンブル（例えば、バギング、ブースティング、ランダムフォレスト）、勾配ブースティングマシン、イオン、ナイーブベイズ、サポートベクターマシン、又はニューラルネットワークのうちの1つであり得る。疾患状態モデルは、トレーニング中に調整される特徴量に対する学習された重みを含む。重みという用語は、ここでは、どの特定の機械学習技術が使用されるかにかかわらず、モデルの任意の所与の特徴に関連する学習された量を表すために一般的に使用される。いくつかの実施形態では、がん指標スコアは、1つ以上のRNA配列（またはそのDNA配列読み取り値）に由来する特徴の値を機械学習モデルまたは深層学習モデルに入力することによって決定される。

【0161】

トレーニング中、トレーニングデータは、疾患状態モデルの重みをトレーニングするために使用される特徴量の値を生成するために処理される。一例として、トレーニングデータは、トレーニングサンプルから得られたcfRNAデータ及び／又はWBC RNAデータ、並びに出力ラベルを含むことができる。例えば、出力ラベルは、個体が特定の疾患を有することが知られているか（例えば、癌を有することが知られている）、又は健康であることが知られているか（すなわち、疾患を有しない）についての指示であり得る。他の実施形態では、モデルは、疾患の種類、または起源組織（例えば、起源癌組織）、または疾患の重症度の表示（例えば、癌ステージ）を決定し、それに対する出力ラベルを生成するために使用され得る。実施形態によっては、疾患状態モデルは、cfRNA分子又はそれに由来する配列の検出及び定量化に用いられるRNAアッセイから決定される1つ又は複数の特徴の値、並びに学習されるモデルに関連する計算分析を受け取る。一実施形態では、1つ以上の特徴は、1つ以上のcfRNA分子又はそれに由来する配列の読み取りの量からなる。トレーニング中のモデルによって出力されたスコアとトレーニングデータの出力ラベルとの間の差異に応じて、癌予測モデルの重みが、疾患状態モデルがより正確な予測を行うことを可能にするように最適化される。様々な実施形態において、疾患状態モデルは、ノンパラメトリックモデル（例えば、k-nearest neighbors）であってもよく、したがって、予測的な癌モデルは、パラメータを最適化する必要なく、より正確に予測を行うようにトレーニングされ得る。

【0162】

学習された疾患状態モデルは、コンピュータ可読媒体に保存することができ、その後、例えばモデルの展開時など、必要なときに取り出すことができる。

【0163】

いくつかの実施形態では、遺伝子発現行列（G）に組織特異性行列（TS）を乗じることにより、遺伝子発現行列（G）を組織スコア行列（S）に変換することを含む方法である。 $G_{m,n}$ は、サンプルmにおける遺伝子nの発現レベルである。 $TS_{n,j}$ は、組織jに対する遺伝子nの組織特異性である。遺伝子nが組織jに対して特異的でない場合、 $TS_{n,j} = 0$ である。いくつかの実施形態では、組織特異性行列は、組織RNA-seqデータベース（GTEx）を用いて計算される。組織スコアは、例えば、癌対非癌サンプルを分類するためのモデルを構築するための特徴として使用することができる。非限定的な一実施形態では、肺がんサンプルから同定された暗チャンネル遺伝子（SFTPA2、SLC39A4、NKX2_1、SFTPA1、BPIFA1、SLC34A2、CXCL17、SFTA3、MUC1、AGR2、WFDC2、ABC A12、VSIG10、CRABP2）を使用して決定木分類器を構築して肺がんcfRNAサンプル

とがん以外を識別するために使用された。この解析結果を図10に示す。

【0164】

シーケンシングとバイオインフォマティクス

【0165】

本開示の側面は、複数の配列リードを生成するための核酸分子の配列決定、及び主題の方法を実行するための配列リードのバイオインフォマティクス操作を含む。

【0166】

様々な実施形態において、試料は、対象から収集され、次いで、関心のある遺伝子領域または遺伝子断片について濃縮される。例えば、いくつかの実施形態では、試料は、癌関連遺伝子または関心のある遺伝子断片を含むヌクレオチドアレイへのハイブリダイゼーションによって濃縮され得る。いくつかの実施形態では、サンプルは、ハイブリッド捕捉など、当技術分野で公知の他の方法を用いて、目的の遺伝子（例えば、がん関連遺伝子）について濃縮することができる。例えば、Lapidus（米国特許番号7,666,593）（その内容は参照によりその全体が本明細書に組み込まれる）を参照されたい。つのハイブリッド捕捉法では、ビオチン化オリゴヌクレオチドおよびストレプトアビジン被覆磁気ビーズの使用を含む、溶液ベースのハイブリダイゼーション法が使用される。例えば、Duncavageら、J Mol Diagn.13(3):325-333（2011）；およびNewmanら、Nat Med.20(5):548-554（2014）。本開示の方法に従って試料から核酸を単離することは、当技術分野で既知の任意の方法に従って行うことができる。

【0167】

配列決定は、当該技術分野で既知の任意の方法または方法の組み合わせによることができる。例えば、既知の核酸配列決定技術には、標識されたターミネーターまたはプライマーおよびスラブまたはキャピラリーでのゲル分離を用いた古典的なジデオキシ配列決定反応（サンガー法）、可逆的に終結した標識ヌクレオチドを用いた合成による配列決定、パイロ配列決定、454配列決定、などがあるが、それらに限定されるわけではない。標識オリゴヌクレオチドプローブのライブラリへのアレル特異的ハイブリダイゼーション、標識クローンのライブラリへのアレル特異的ハイブリダイゼーションとそれに続くライゲーションを用いた合成による配列決定、重合ステップ中の標識ヌクレオチドの組み込みのリアルタイムモニタリング、Polony配列決定およびSOLiD配列決定などである。分離した分子の配列決定は、最近では、ポリメラーゼヤリガーゼを用いた連続的または単一の伸長反応や、プローブのライブラリを用いた単一または連続的な差分ハイブリダイゼーションによって示されている。

【0168】

シーケンシングを行う従来の方法としては、Sangerら、Proc Natl. Acad.Sci.U S A, 74(12):5463-67（1977）に記載されており、その内容は参照によりその全体が本明細書に組み入れられる。別の従来の配列決定方法は、核酸断片の化学的分解を含む。参照、Maxamら、Proc. Natl.Acad.Sci., 74: 560-564（1977）、その内容は、参照によりその全体が本明細書に組み込まれる。ハイブリダイゼーションによる配列決定に基づいて、方法も開発されている。例えば、Harrisら、（米国特許出願番号2009/0156412）、その内容は参照によりその全体が本明細書に組み込まれる、を参照されたい。

【0169】

提供される開示の方法において使用することができる配列決定技術は、例えば、Helicos True Single Molecule Sequencing (tSMS)（Harris T. D. et al. (2008) Science 320:106-109）、その内容は、参照によりその全体が本明細書に組み入れられる。tSMSのさらなる説明は、例えば、Lapidusら（米国特許番号7,169,560）（その内容全体が参照により本明細書に組み込まれる）、Lapidusら（米国特許出願公開番号2009/0191565、その内容全体が参照により本明細書に組み込まれる）、Quakeら（U.S. 特許番号6,818,395、その内容全体が参照により本明細書に組み込まれる）、Harris（米国特許番号7,282,337、その内容全体が参照により本明細書に組み込まれる）、Quakeら（米国特許出願公開番号2002/0164629、その内容全体が参

10

20

30

40

50

照により本明細書に組み込まれる)、およびBraslavsky, et al,PNAS (USA)、100:3960-3964 (2003)、これらの内容は、参照によりその全体が本明細書に組み入れられる。

【0170】

提供される開示の方法において使用することができる核酸配列決定技術の別の例は、454配列決定(Roche)である(Margulies, M et al. 2005, Nature, 437, 376-380、その内容は参照によりその全体が本明細書に組み入れられる)。提供される開示の方法において使用することができるDNA配列決定技術の別の例は、SOLiD技術(Applied Biosystems)である。提供される開示の方法で使用することができるDNA配列決定技術の別の例は、Ion Torrent配列決定(米国特許出願公開番号2009/0026082、2009/0127589、2010/0035252、2010/0137143、2010/0188073、2010/0197507、2010/0282617、2010/300559、2010/300895、2010/301398、及び2010/030482、これらの各々の内容はその全体として参考としてここに組み入れる)である。

10

【0171】

いくつかの実施形態では、配列決定技術は、Illumina配列決定である。イルミナシーケンスは、フォールドバックPCR及びアンカープライマーを用いた固体表面上のDNAの増幅に基づく。ゲノムDNAは断片化することができるが、cfDNAの場合、すでに短い断片であるため断片化は必要ない。断片の5'末端と3'末端にアダプターをリゲートする。フローセル流路の表面に付着したDNA断片は伸長し、橋渡し増幅される。断片は二本鎖となり、二本鎖分子は変性する。固相増幅と変性を複数回繰り返すことにより、フローセルの各チャンネルに同一鑄型の一本鎖DNA分子が約1,000コピー集まった数百万個のクラスターを形成することができる。プライマー、DNAポリメラーゼ、および4つの蛍光標識された可逆的終止ヌクレオチドを使用して、シーケンシャルシーケンスが行われる。ヌクレオチドを取り込んだ後、レーザーで蛍光体を励起し、画像を取り込み、最初の塩基の識別を記録する。組み込まれた各塩基の3'ターミネーターと蛍光色素を除去し、組み込み、検出、同定のステップを繰り返す。

20

【0172】

提供される開示の方法において使用することができる配列決定技術の別の例は、Pacific Biosciencesの単一分子、リアルタイム(SMRT)技術を含む。提供される開示の方法において使用することができる配列決定技術のさらに別の例は、ナノポアシーケンス(Soni G V and Meller A. (2007) Clin Chem 53: 1996-2001、その内容は参照によりその全体が本明細書に組み込まれる)である。提供される開示の方法において使用することができる配列決定技術の別の例は、DNAの配列決定に化学感受性電界効果トランジスタ(chemFET)アレイを使用することを含む(例えば、米国特許出願公開第20090026082号に記載されており、その内容全体が参照により本明細書に組み入れられる)。提供される開示の方法において使用され得る配列決定技術の別の例は、電子顕微鏡を使用することを含む(Moudrianakis E. N. and Beer M. Proc Natl Acad Sci US A.1965 March; 53:564-71、その内容は参照によりその全体が本明細書に組み込まれる)。

30

40

【0173】

試料からの核酸が分解されているか、または試料から最小量の核酸しか得られない場合、配列決定に十分な量の核酸を得るために、核酸に対してPCRを行うことができる(例えば、Mullisら、米国特許番号4,683,195、その内容全体が参照によりここに組み込まれる、を参照されたい)。

【0174】

コンピュータシステム・機器

【0175】

本明細書に記載される開示の態様は、プロセッサ、例えば中央処理装置を含むコンピュータなどの任意のタイプのコンピューティングデバイス、または各デバイスがプロセスま

50

たは方法の少なくとも一部を実行するコンピューティングデバイスの任意の組合せを用いて実行することができる。いくつかの実施形態では、本明細書に記載されるシステム及び方法は、ハンドヘルドデバイス、例えば、スマートタブレット、又はスマートフォン、又はシステムのために製造される専門デバイスを用いて実行され得る。

【0176】

本開示の方法は、ソフトウェア、ハードウェア、ファームウェア、ハードワイヤリング、又はこれらのいずれかの組合せを用いて実行することができる。機能を実装する機能は、機能の一部が異なる物理的位置（例えば、画像形成装置がある部屋に、ホストワークステーションが別の部屋に、又は別々の建物に、例えば無線又は有線接続で）で実装されるように分散されることを含む、様々な位置に物理的に配置することも可能である。

10

【0177】

コンピュータ・プログラムの実行に適したプロセッサには、一例として、汎用および特殊目的のマイクロプロセッサ、およびあらゆる種類のデジタル・コンピュータの任意の1つまたは複数のプロセッサが含まれる。一般に、プロセッサは、読み取り専用メモリまたはランダムアクセスメモリ、あるいはその両方から命令とデータを受け取る。コンピュータの本質的な要素は、命令を実行するためのプロセッサと、命令とデータを格納するための1つ以上のメモリ装置である。一般に、コンピュータは、データを格納するための1つ以上の大容量記憶装置、例えば、磁気ディスク、光磁気ディスク、または光ディスクからデータを受信するか、またはその両方にデータを転送するように動作可能に結合されるか、またはその両方を含むことになる。コンピュータプログラム命令およびデータを具現化するのに適した情報担体は、一例として、半導体メモリデバイス、（例えば、EPROM、EEPROM、ソリッドステートドライブ（SSD）、およびフラッシュメモリ装置）；磁気ディスク、（例えば、内蔵ハードディスクまたはリムーバブルディスク）；光磁気ディスク；および光ディスク（例えば、CDディスクおよびDVDディスク）などの不揮発性メモリのあらゆる形式を含んでいる。プロセッサとメモリは、特別な目的の論理回路によって補完されるか、またはその中に組み込まれることが可能である。

20

【0178】

ユーザとの相互作用を提供するために、本明細書に記載の主題は、ユーザに対して情報を表示するためのI/Oデバイス、例えばCRT、LCD、LED、または投影デバイスと、ユーザがコンピュータに入力を提供できるキーボードおよびポインティングデバイス（例えばマウスまたはトラックボール）などの入力または出力デバイスを有するコンピュータで実施することができる。他の種類のデバイスも、ユーザとの対話を提供するために使用することができる。例えば、ユーザに提供されるフィードバックは、任意の形式の感覚フィードバック（例えば、視覚フィードバック、聴覚フィードバック、または触覚フィードバック）であり、ユーザからの入力は、音響、音声、または触覚入力などの任意の形式で受け取ることができる。

30

【0179】

本明細書に記載される主題は、バックエンド構成要素（例えば、データサーバ）、ミドルウェア構成要素（例えば、アプリケーションサーバ）、またはフロントエンド構成要素（例えば、ユーザが本明細書に記載される主題の実装と対話できるグラフィカルユーザインターフェースまたはWebブラウザを有するクライアントコンピュータ）、またはそのようなバックエンド、ミドルウェアおよびフロントエンド構成要素の任意の組み合わせが含まれる計算システムで実装されることができる。システムの構成要素は、デジタルデータ通信の任意の形態または媒体、例えば通信ネットワークによって、ネットワークを介して相互接続することができる。例えば、データの基準セットは遠隔地に保存され、コンピュータはネットワークを介して通信し、比較目的のために基準データセットにアクセスすることができる。しかし、他の実施形態では、基準データセットは、コンピュータ内にローカルに格納され得、コンピュータは、比較目的のためにCPU内で基準データセットにアクセスする。通信ネットワークの例としては、セルネットワーク（例えば、3G又は4G）、ローカルエリアネットワーク（LAN）、及びワイドエリアネットワーク（WAN）、

40

50

例えば、インターネットが挙げられるが、これらに限定されるものではない。

【0180】

本明細書に記載された主題は、データ処理装置（例えば、プログラマブルプロセッサ、コンピュータ、または複数のコンピュータ）による実行のため、またはその動作を制御するために情報キャリア（例えば、非一過性のコンピュータ可読媒体）において接離された1つまたは複数のコンピュータプログラムなどの1つまたは複数のコンピュータプログラム製品として実装することが可能である。コンピュータプログラム（プログラム、ソフトウェア、ソフトウェアアプリケーション、アプリ、マクロ、またはコードとも呼ばれる）は、コンパイル言語または解釈言語（例えば、C、C++、Perl）を含む任意の形式のプログラミング言語で書くことができ、スタンドアロンプログラムとしてまたはモジュール、コンポーネント、サブルーチン、またはコンピュータ環境での使用に適した他のユニットとして含む任意の形式で配備することができる。本開示のシステムおよび方法は、限定されないが、C、C++、Perl、Java、ActiveX、HTML5、Visual Basic、またはJavaScriptを含む、当該技術分野で知られている任意の適切なプログラミング言語で書かれた命令を含むことが可能である。

10

【0181】

コンピュータ・プログラムは、必ずしもファイルに対応するものではありません。プログラムは、他のプログラムやデータを保持するファイルまたはファイルの一部、当該プログラム専用の単一ファイル、または複数の連携ファイル（例えば、1つまたは複数のモジュール、サブプログラム、またはコードの一部を格納するファイル）に格納することができる。コンピュータ・プログラムは、1台のコンピュータで実行されるように配置することも、1つのサイトの複数のコンピュータで実行されるように配置することも、複数のサイトに分散して通信ネットワークで相互接続されるように配置することも可能である。

20

【0182】

ファイルは、例えば、ハードドライブ、SSD、CD、または他の有形非一時的媒体に格納されるデジタルファイルであることができる。ファイルは、ネットワークを介して（例えば、ネットワークインターフェースカード、モデム、無線カードなどを介して、サーバからクライアントに送信されるパケットとして）、あるデバイスから別のデバイスに送信することができる。

【0183】

本開示によるファイルの書き込みは、例えば、粒子（例えば、正味の電荷または双極子モーメントを読み取り／書き込みヘッドによる磁化のパターンに有する）を追加、除去、または並べ替えることによって、有形の非一過性のコンピュータ可読媒体を変換することを含み、パターンは次に、ユーザによって望まれ、ユーザにとって有用な、目的の物理現象に関する情報の新しいコロケーションを表わす。いくつかの実施形態では、書き込みは、有形で非一時的なコンピュータ可読媒体（例えば、光学リード／ライトデバイスが次に情報の新しい有用なコロケーションを読み取るできるように特定の光学特性を有する、例えば、CD-ROMを焼く）の材料の物理的変換を含む。いくつかの実施形態では、ファイルの書き込みは、NANDフラッシュメモリ装置などの物理的フラッシュメモリ装置を変形させ、フローティングゲートトランジスタから作られたメモリセルのアレイ内の物理的要素を変形させることによって情報を格納することを含む。ファイルを書き込む方法は、当該技術分野において周知であり、例えば、プログラムによって、またはソフトウェアからの保存コマンドまたはプログラミング言語からの書き込みコマンドによって、手動または自動的に呼び出すことができる。

30

40

【0184】

適切なコンピューティングデバイスは、典型的には、マスメモリ、少なくとも1つのグラフィカルユーザインタフェース、少なくとも1つのディスプレイデバイスを含み、典型的には、デバイス間の通信を含む。マスメモリは、コンピュータ可読媒体の一種、すなわちコンピュータ記憶媒体を例示する。コンピュータ記憶媒体は、コンピュータ可読命令、データ構造、プログラムモジュール、又は他のデータなどの情報を記憶するための任意の

50

方法又は技術で実装された揮発性、不揮発性、取り外し可能、及び非取り出し可能な媒体を含んでもよい。コンピュータ記憶媒体の例としては、RAM、ROM、EEPROM、フラッシュメモリ、または他のメモリ技術、CD-ROM、デジタル多用途ディスク（DVD）または他の光学ストレージ、磁気カセット、磁気テープ、磁気ディスクストレージまたは他の磁気ストレージデバイス、無線自動識別（RFID）タグまたはチップ、または所望の情報を格納するのに使用でき、コンピュータデバイスによってアクセスできる他の任意の媒体がある。

【0185】

本明細書で説明する機能は、ソフトウェア、ハードウェア、ファームウェア、ハードウェア、またはこれらのいずれかの組合せを用いて実装することができる。ソフトウェアのいずれかは、機能の一部が異なる物理的な位置で実装されるように分散されることを含め、物理的に様々な位置に配置され得る。

10

【0186】

当業者であれば、本開示の方法の実行に必要なまたは最適であると認識するように、説明した発明方法の一部または全部を実施するためのコンピュータシステムは、1つまたは複数のプロセッサ（例えば、中央処理装置（CPU）グラフィックス処理装置（GPU）またはその両方）、メインメモリおよびスタティックメモリを含み得、これらはバスを介して互いに通信し合う。

【0187】

プロセッサは、一般に、中央処理装置（CPU）を提供するために、シングルコアまたはマルチコアチップなどのチップを含むことになります。プロセスは、インテルまたはAMDのチップによって提供される場合がある。

20

【0188】

メモリは、開示されたコンピュータのいずれか1つのプロセッサ（複数可）によって実行されると、本明細書に記載された方法論または機能の一部または全部を達成することができる1つまたは複数の命令セット（例えば、ソフトウェア）が格納されている1つまたは複数の機械可読デバイスを含むことができる。ソフトウェアはまた、コンピュータシステムによるその実行中に、完全に又は少なくとも部分的に、メインメモリ内及び／又はプロセッサ内に存在することができる。好ましくは、各コンピュータは、ソリッドステートドライブ、フラッシュドライブ、ディスクドライブ、ハードドライブなどの非一時的なメモリを含む。

30

【0189】

機械可読デバイスは、例示的な実施形態では単一の媒体であり得るが、用語「機械可読デバイス」は、1つ以上の命令セット及び／又はデータを格納する単一の媒体又は複数の媒体（例えば、集中又は分散データベース、並びに関連キャッシュ及びサーバー）を含むと解釈されるべきである。これらの用語はまた、マシンによる実行のための命令セットを格納、符号化、または保持することができ、マシンに本開示の方法論のいずれか1つ以上を実行させる任意の媒体またはメディアを含むと解釈されるものとする。これらの用語は、それに応じて、1つ以上の固体メモリ（例えば、加入者IDモジュール（SIM）カード、セキュアデジタルカード（SDカード）、マイクロSDカード、または固体ドライブ（SSD））、光学および磁気媒体、および／または他の任意の有形記憶媒体または媒体を含むが、これらに限定されないものと解釈される。

40

【0190】

本開示のコンピュータは、一般に、例えば、ビデオ表示装置（例えば、液晶ディスプレイ（LCD）または陰極線管（CRT））、英数字入力装置（例えば、キーボード）、カーソル制御装置（例えば、タッチスクリーン、加速度計、マイクロフォン、携帯電話用無線周波数アンテナ、およびネットワークインタフェースデバイス（例えば、ネットワークインタフェースカード（NIC）、Wi-Fiカード、または携帯電話モデムなど）。

【0191】

いずれのソフトウェアも、機能の一部が異なる物理的な位置で実装されるように分散配

50

置されるなど、物理的に様々な位置に配置することができる。

【0192】

さらに、本開示のシステムは、参照データを含むように提供され得る。任意の適切なゲノムデータが、システム内で使用するために保存され得る。例としては、以下が挙げられるが、これらに限定されない。The Cancer Genome Atlas (TCGA) からの、がんの主要なタイプおよびサブタイプにおける主要なゲノム変化の包括的な多次元マップ；The International Cancer Genome Consortium (ICGC) からのゲノム異常のカタログ；COSMICからのがんにおける体細胞突然変異のカタログ；。ヒトゲノムおよびその他の一般的なモデル生物の最新ビルド、dbSNPからの最新の参照SNP、1000ゲノムプロジェクトおよびBroad Instituteからのゴールドスタンダードインデル、Illumina、Agilent、NimblegenおよびIon Torrentからのエクソームキャプチャーキットアノテーション、転写産物アノテーション、パイプラインの実験用の小規模テストデータ（e...
.g.,新規ユーザー向け）。

10

【0193】

いくつかの実施形態では、データは、システムに含まれるデータベースのコンテキスト内で利用可能にされる。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、及びその他を含む任意の適切なデータベース構造が使用されてもよい。いくつかの実施形態では、参照データは、「ノットオンリーSQL」（NoSQL）データベースなどのリレーショナルデータベース内に格納される。様々な実施形態において、グラフデータベースは、本開示のシステム内に含まれる。また、本明細書で使用される「データベース」という用語は、1つの単一のデータベースに限定されず、むしろ、複数のデータベースがシステムに含まれ得ることを理解されたい。例えば、データベースは、本開示の実施形態に従って、その中のデータベースの任意の整数を含む、2、3、4、5、6、7、8、9、10、15、20、またはそれ以上の個々のデータベースを含むことができる。例えば、1つのデータベースは、公的参照データを含むことができ、第2のデータベースは、患者からの試験データを含むことができ、第3のデータベースは、健康な対象からのデータを含むことができ、第4のデータベースは、既知の状態又は障害を有する病気の対象からのデータを含むことができる。そこに含まれるデータに関するデータベースの任意の他の構成も、本明細書に記載される方法によって企図されることが理解される。

20

【0194】

例示的な実施形態(a)

【0195】

本開示は、以下の実施形態を提供する。

【0196】

実施形態P1. 対象における癌を検出する方法であって、該方法は、以下を含む。

(a) 対象のサンプル中の複数の標的無細胞RNA (cfRNA) 分子を測定し、複数の標的cfRNA分子が、表11の1つ以上の転写物、及び任意選択で表12～15の転写物から選択され；及び

(b) 癌を検出することであって、癌を検出することが、閾値を超える標的cfRNA分子の1つ以上を検出することを含む、工程。

40

【0197】

実施形態P2. 複数の標的cfRNA分子が、表11～14の少なくとも5、10、15、または20の転写物から選択される、実施形態P1に記載の方法。

【0198】

実施形態P3. 複数の標的cfRNA分子が、(i) 表11、(ii) 表2、5、及び12の各々、(iii) 表3、4、6、及び13の各々、又は(iv) 表14からの複数の転写物を含む、実施形態P1又はP2に記載の方法。

【0199】

実施形態P4. 複数の標的cfRNA分子が、表11～15のうちの1つ以上の転写物の全てを含む、実施形態P1～P3のいずれか1つに記載の方法。

50

【0200】

実施形態P5. 複数の標的cfRNA分子が、表14の転写物から選択される、実施形態P1～P4のいずれか1つに記載の方法。

【0201】

実施形態P6. 閾値を超えて検出された複数の標的cfRNA分子は、以下からなる群から選択される複数の遺伝子に由来するcfRNA分子である、実施形態P1～P5のいずれか一項に記載の方法。adipoq、agr3、ankrd30a、aqp4、bpifa1、ca12、ceacam5、cftr、cxcl17、cyp4f8、fabp7、foxy1、ggtlc1、gp2、IL20、Itih6、ldlr1、ldlr1、lmd1、lmx1b、mmp7、nkain1、nkx2-1、ropn1、ros1、scgb1d2、scgb2a2、sfta2、sfta3、slc34a2、sox9、stk32a、stmnd1、tfap2a、tfap2b、tff1、trpv6、vgll1、vtcn1。

10

【0202】

実施形態P7. 複数の標的cfRNA分子が、表11～14の1つ以上の転写物及び表1～6の1つ以上の転写物を含む、実施形態P1～P6のいずれか1つに記載の方法。

【0203】

実施形態P8. 複数の標的cfRNA分子が、表11～14の1つ以上の転写物及び表7の1つ以上の転写物を含む、実施形態P1～P7のいずれか1つに記載の方法。

【0204】

実施形態P9. i) 癌が肺癌であり、(ii) 閾値を超えて検出された複数の標的cfRNA分子が、表2、表5、又は表12の1つ以上の転写物から選択される、実施形態P1～P4のいずれか1つに記載の方法。

20

【0205】

実施形態P10. i) 癌が乳癌であり、(ii) 閾値を超えて検出された複数の標的cfRNA分子が、表3、4、6、又は13の1つ以上の転写物から選択される、実施形態P1～P4のいずれか1つに記載の方法。

【0206】

実施形態P11. 測定が、配列決定、マイクロアレイ解析、逆転写PCR、リアルタイムPCR、定量リアルタイムPCR、デジタルPCR、デジタル液滴PCR、デジタルエマルジョンPCR、マルチプレックスPCR、ハイブリッド捕捉、オリゴヌクレオチドライゲーションアッセイ、又はこれらの任意の組み合わせを含む、実施形態P1～P10のいずれか1つに記載の方法。

30

【0207】

実施形態P12. 測定が、cfRNA配列リードを生成するためにcfRNA分子を配列決定することを含む、実施形態P1～P11のいずれか1つに記載の方法。

【0208】

実施形態P13. cfRNA分子を配列決定することが、全トランスクリプトーム配列決定を含む、実施形態P12に記載の方法。

【0209】

実施形態P14. cfRNA分子を配列決定することが、cDNA分子を生成するための逆転写、及びcDNA分子を配列決定してcfRNA配列リードを生成することを含む、実施形態P12又はP13に記載の方法。

40

【0210】

実施形態P15. cfRNA分子を配列決定することが、標的cfRNA分子又はそのcDNA分子を濃縮することを含む、実施形態P12に記載の方法。

【0211】

実施形態P16. 試料が生体流体を含む、実施形態P1～P15のいずれか1つに記載の方法。

【0212】

実施形態P17. 生体液が、血液、血漿、血清、尿、唾液、胸水、心嚢液、脳脊髄液(CSF)、腹膜液、又はそれらの任意の組み合わせを含む、実施形態P16に記載の方法。

50

【0213】

実施形態P18. 生物学的が、被験者の血液、血液画分、血漿、又は血清からなる、実施形態P16に記載の方法。

【0214】

実施形態P19. 閾値以上の標的cfRNA分子の1つ以上を検出することが、(i) 検出、(ii) バックグラウンド以上の検出、又は(iii) 状態を有しない対象における標的cfRNA分子のレベルよりも大きいレベルでの検出を含む、実施形態P1～P18のいずれか1つに記載の方法。

【0215】

実施形態P20. 閾値以上の1つ以上の標的cfRNA分子を検出することが、状態を有しない被験者のレベルよりも少なくとも約10倍大きいレベルで1つ以上の標的cfRNA分子を検出することを含む、実施形態P1～P18のいずれか1つに記載の方法。

10

【0216】

実施形態P21. 閾値以上の標的cfRNA分子の1つ以上を検出することは、0.5～5リード／ミリオン(RPM)の閾値以上の検出を含む、実施形態P12～P18のいずれか1つに記載の方法。

【0217】

実施形態P22. 閾値以上の標的cfRNA分子の1つ以上を検出することは、以下のことを含む、実施形態P1～P18のいずれか1つに記載の方法。

(a) 標的cfRNA分子の各々の発現レベルをRNA組織スコアマトリックスと比較することによって、各標的cfRNA分子に対する指標スコアを決定するステップと

20

(b) 各ターゲットcfRNA分子の指標スコアを集計すること；及び

(c) 指標となるスコアが閾値を超えた場合に、がんを検出すること。

【0218】

実施形態P23. 閾値以上の標的cfRNA分子の1つ以上を検出することは、配列読み取りを機械学習又は深層学習モデルに入力することを含む、実施形態P12～P22のいずれか1つに記載の方法。

【0219】

実施形態P24. 機械学習又は深層学習モデルは、ロジスティック回帰、ランダムフォレスト、勾配ブースティング機械、ナイーブベイズ、ニューラルネットワーク、又は多項回帰からなる、実施形態P23に記載の方法。

30

【0220】

実施形態P25. 機械学習又は深層学習モデルは、学習された重みを含む関数を通じて、1つ以上の特徴の値を対象者の疾患状態予測に変換する、実施形態P23に記載の方法。

【0221】

実施形態P26. 癌が以下のものを含む、実施形態P1～P25のいずれか1つに記載の方法。

(i) 癌腫、肉腫、骨髄腫、白血病、リンパ腫、芽細胞腫、またはそれらの任意の組み合わせ。

(ii) 腺癌、扁平上皮癌、小細胞肺癌、非小細胞肺癌、鼻咽頭癌、大腸癌、肛門癌、肝臓癌、尿道癌、精巣癌、頸部癌、卵巣癌、胃癌、食道癌、頭頸部癌、すい臓癌、前立腺癌、腎癌、甲状腺癌、メラノーマおよび乳癌からなる群から選ばれる癌腫。

40

(iii) ホルモン受容体陰性乳癌またはトリプルネガティブ乳癌。

(iv) 骨肉腫、軟骨肉腫、平滑筋肉腫、横紋筋肉腫、中皮肉腫(中皮腫)、線維肉腫、血管肉腫、脂肪肉腫、グリオーマ、及び星細胞腫からなる群から選択される肉腫。

(v) 骨髄性、顆粒球性、リンパ性、リンパ球性、およびリンパ芽球性白血病からなる群から選択される白血病；または

(vi) からなる群より選択される リンパ腫。ホジキンリンパ腫および非ホジキンリンパ腫。

【0222】

50

実施形態P27. 癌を検出することが、癌の病期を決定すること、癌の進行度を決定すること、癌の種類を決定すること、起源癌組織を決定すること、又はそれらの組み合わせを含む、実施形態P1～P26のいずれか1つに記載の方法。

【0223】

実施形態P28. 検出された癌に基づいて治療法を選択することをさらに含む、実施形態P1～P27のいずれか1つに記載の方法。

【0224】

実施形態P29. 治療が、外科的切除、放射線療法、または抗癌剤の投与を含む、実施形態P28に記載の方法。

【0225】

実施形態P30. 選択された治療で対象を治療することをさらに含む、実施形態P28又はP29に記載の方法。

【0226】

実施形態P31. 実施形態P1～P30のいずれか1つに記載の方法における1つまたは複数のステップを実施するためのコンピュータシステム。

【0227】

実施形態P32. 非一過性の、コンピュータ可読媒体であって、実施形態P1～P30のいずれか1つの方法における1つ以上のステップを実施するためのコンピュータ可読命令をその上に格納した、非一過性のコンピュータ可読媒体。

【0228】

例示的な実施形態 (b)

【0229】

本開示は、以下の実施形態を提供する。

【0230】

実施形態1. 対象における癌を検出する方法であって、該方法は、以下を含む。

(a) 対象のサンプル中の複数の標的無細胞RNA (cfRNA) 分子を測定することであって、複数の標的cfRNA分子が表11の1つ以上の転写物から選択されること；及び

(b) 癌を検出することであって、癌を検出することが、閾値以上の標的cfRNA分子の1つ以上を検出することを含む、こと。

【0231】

実施形態2. 複数の標的cfRNA分子が、表8又は表12～15の1つ以上から選択される、実施形態1に記載の方法。

【0232】

実施形態3. 複数の標的cfRNA分子が、表8又は11～14の1つ以上から少なくとも5、10、15、又は20の遺伝子の転写物から選択される、実施形態1に記載の方法。

【0233】

実施形態4. 複数の標的cfRNA分子が、(i) 表11；(ii) 表2、5、及び12のそれぞれ；(iii) 表3、4、6、及び13のそれぞれ；(iv) 表14；又は(v) 表8からの複数の転写物を含む、実施形態1又は3記載の方法。

【0234】

実施形態5. 複数の標的cfRNA分子が、表8又は11～15の1つ以上から少なくとも30個の遺伝子の転写物を含む、実施形態1～4のいずれか1つに記載の方法。

【0235】

実施形態6. 複数の標的cfRNA分子が、表14の転写物から選択される、実施形態1～5のいずれか1つに記載の方法。

【0236】

実施形態7. 閾値を超えて検出された複数の標的cfRNA分子が、以下からなる群から選択される複数の遺伝子に由来するcfRNA分子である、実施形態1～6のいずれか一項に記載の方法。adipoq、agr3、ankrd30a、aqp4、bpifa1、ca12、ceacam5、cftr、cxcl17、cyp4f8、fabp7、foxy1、ggtlc1、gp2、IL20、Itih6、ldlr1、lem

10

20

30

40

50

d1、lmx1b、mmp7、NKAIN1、NKX2-1、ROPN1、ROS1、SCGB1D2、SCGB2A2、SFTA2、SFTA3、SLC34A2、SOX9、STK32A、STMND1、TFAP2A、TFAP2B、TFF1、TRPV6、VGLL1および VTCN1。

【0237】

実施形態8．複数の標的cfRNA分子が表8の転写物から選択される、実施形態1～5のいずれか1つに記載の方法。

【0238】

実施形態9．閾値を超えて検出された複数の標的cfRNA分子が、以下からなる群から選択される複数の遺伝子に由来するcfRNA分子である、実施形態1～5のいずれか一項に記載の方法。 ceacam5、rhov、sfta2、scgb1d2、igf2bp1、sftpa1、ca12、sftpb、cdh3、muc6、src6a14、hoxc9、agr3、tmem125、tfap2b、irx2、potekp、agr3、tmem126、tfap2b、irx2、potekp、ARHGEF38、GPR87、LMX1B、ATP10B、NEL1、MUC21、SOX9、LINC00993、STMND1、ERVH48-1、SCTR、MAGEA3、MB、LEMD1、SIX4、及びNXNL2。

【0239】

実施形態10．複数の標的cfRNA分子が、(a) 表8又は11～14の1つ以上の転写物、及び(b) 表1～6の1つ以上の転写物を含む、実施形態1～9のいずれか1つに記載の方法。

【0240】

実施形態11．複数の標的cfRNA分子が、(a) 表8又は11～14の1つ以上の転写物、及び(b) 表7の1つ以上の転写物を含む、実施形態1～10のいずれか1つに記載の方法。

【0241】

実施形態12． i) 癌が肺癌であり、(ii) 閾値を超えて検出された複数の標的cfRNA分子が、表2、表5、又は表12の1つ以上の転写物から選択される、実施形態1～5のいずれか一項に記載の方法。

【0242】

実施形態13． i) 癌が乳癌であり、(ii) 閾値を超えて検出された複数の標的cfRNA分子が、表3、4、6、又は13の1つ以上の転写物から選択される、実施形態1～5のいずれか1つに記載の方法。

【0243】

実施形態14．測定が、配列決定、マイクロアレイ解析、逆転写PCR、リアルタイムPCR、定量リアルタイムPCR、デジタルPCR、デジタル液滴PCR、デジタルエマルジョンPCR、マルチプレックスPCR、ハイブリッド捕捉、オリゴヌクレオチドライゲーションアッセイ、又はこれらの任意の組み合わせを含む、実施形態1～13のいずれか1つに記載の方法。

【0244】

実施形態15．測定が、cfRNA配列リードを生成するためにcfRNA分子を配列決定することを含む、実施形態1～14のいずれか1つに記載の方法。

【0245】

実施形態16．cfRNA分子を配列決定することが、全トランスクリプトーム配列決定を含む、実施形態15に記載の方法。

【0246】

実施形態17．cfRNA分子を配列決定することが、cDNA分子を生成するための逆転写、及びcDNA分子を配列決定してcfRNA配列リードを生成することを備える、実施形態15又は16に記載の方法。

【0247】

実施形態18．cfRNA分子を配列決定することが、標的cfRNA分子又はそのcDNA分子を濃縮することを含む、実施形態15に記載の方法。

【0248】

実施形態19．試料が生体流体を含む、実施形態1～18のいずれか1つに記載の方法。

10

20

30

40

50

【0249】

実施形態20．生体液が、血液、血漿、血清、尿、唾液、胸水、心嚢液、脳脊髄液（CSF）、腹膜液、又はそれらの任意の組み合わせを含む、実施形態19に記載の方法。

【0250】

実施形態21．生物学的が、被験者の血液、血液画分、血漿、又は血清からなる、実施形態19に記載の方法。

【0251】

実施形態22．閾値以上の標的cfRNA分子の1つ以上を検出することが、（i）検出、（ii）バックグラウンド以上の検出、又は（iii）状態を有しない対象における標的cfRNA分子のレベルよりも大きいレベルでの検出を含む、実施形態1～21のいずれか1つに記載の方法。

10

【0252】

実施形態23．閾値以上の1つ以上の標的cfRNA分子を検出することが、状態を有しない被験者のレベルよりも少なくとも10倍大きいレベルで1つ以上の標的cfRNA分子を検出することを含む、実施形態1～21のいずれか1つに記載の方法。

【0253】

実施形態24．標的cfRNA分子の1つ以上を閾値以上検出することが、0.5～5リード／ミリオン（RPM）の閾値以上の検出を含む、実施形態15～21のいずれか1つに記載の方法。

【0254】

実施形態25．閾値以上の標的cfRNA分子の1つ以上を検出することを含む、実施形態1～21のいずれか1つに記載の方法である。

20

（a）標的cfRNA分子の各々の発現レベルをRNA組織スコアマトリックスと比較することにより、各標的cfRNA分子に対する指標スコアを決定すること。

（b）各ターゲットcfRNA分子の指標スコアを集計すること；及び

（c）指標となるスコアが閾値を超えた場合に、がんを検出すること。

【0255】

実施形態26．閾値以上の標的cfRNA分子の1つ以上を検出することが、配列読み取りを機械学習又は深層学習モデルに入力することを含む、実施形態15～25のいずれか1つに記載の方法。

30

【0256】

実施形態27．機械学習又は深層学習モデルが、ロジスティック回帰、ランダムフォレスト、勾配ブースティングマシン、ナイーブベイズ、ニューラルネットワーク、又は多項回帰からなる、実施形態26に記載の方法。

【0257】

実施形態28．機械学習又は深層学習モデルは、学習された重みを含む関数を通じて、1つ以上の特徴の値を対象者の疾患状態予測に変換する、実施形態26に記載の方法。

【0258】

実施形態29．癌が以下のものを含む、実施形態1～28のいずれか1つに記載の方法。

（i）癌腫、肉腫、骨髄腫、白血病、リンパ腫、芽細胞腫、またはそれらの任意の組み合わせ。

40

（ii）腺癌、扁平上皮癌、小細胞肺癌、非小細胞肺癌、上咽頭癌、大腸癌、肛門癌、肝臓癌、膀胱癌、精巣癌、頸部癌、卵巣癌、胃癌、食道癌、頭頸部癌、すい臓癌、前立腺癌、腎癌、甲状腺癌、メラノーマおよび乳癌からなる群より選ばれた癌腫；

（iii）ホルモン受容体陰性乳癌またはトリプルネガティブ乳癌。

（iv）骨肉腫、軟骨肉腫、平滑筋肉腫、横紋筋肉腫、中皮肉腫（中皮腫）、線維肉腫、血管肉腫、脂肪肉腫、グリオーマ、及び星細胞腫からなる群から選択される肉腫。

（v）骨髄性、顆粒球性、リンパ性、リンパ球性、およびリンパ芽球性白血病からなる群から選択される白血病；または

（vi）からなる群より選択される リンパ腫。ホジキンリンパ腫および非ホジキンリ

50

ンパ腫。

【0259】

実施形態30. がんを検出することが、がんのステージを決定すること、がんの進行度を決定すること、がんのタイプを決定すること、起源のがん組織を決定すること、又はこれらの組み合わせを含む、実施形態1～29のいずれか1つに記載の方法。

【0260】

実施形態31. 検出された癌に基づいて治療法を選択することをさらに含む、実施形態1～30のいずれか1つに記載の方法。

【0261】

実施形態32. 治療が、外科的切除、放射線療法、または抗癌剤の投与を含む、実施形態31に記載の方法。

10

【0262】

実施形態33. 方法が、選択された治療で対象を治療することをさらに含む、実施形態31又は32に記載の方法。

【0263】

実施形態34. 試料中の複数の標的無細胞RNA (cfRNA) 分子を測定する方法であって、該方法は、以下を含む。

(a) 複数の標的cfRNA分子、又はそのcDNA分子を濃縮して、ポリヌクレオチドの濃縮試料を生成すること；及び

(b) 該濃縮サンプルのポリヌクレオチド、またはその増幅産物を配列決定する工程。

20

ここで、複数の標的cfRNA分子は、表11の1つまたは複数の転写物から選択される。

【0264】

実施形態35. 複数の標的cfRNA分子が、表8又は12～15の1つ以上から選択される、実施形態34に記載の方法。

【0265】

実施形態36. 複数の標的cfRNA分子が表8から選択される、実施形態34に記載の方法。

【0266】

実施形態37. 複数の標的cfRNA分子が、表14から選択される、実施形態34に記載の方法。

30

【0267】

実施形態38. 複数の標的cfRNA分子が、表8または11～14の1つ以上から少なくとも5、10、15、または20の遺伝子の転写物から選択される、実施形態34に記載の方法。

【0268】

実施形態39. 実施形態1～38のいずれか1つに記載の方法における1つまたは複数のステップを実施するためのコンピュータシステム。

【0269】

実施形態40. 非一過性のコンピュータ可読媒体であって、実施形態1～38のいずれか1つに記載の方法における1つ以上のステップを実施するためのコンピュータ可読命令をその上に記憶している、非一過性のコンピュータ可読媒体。

40

【0270】

例

【0271】

本明細書に記載された例および実施形態は、説明のためのものであり、それに照らして様々な修正または変更が当業者に示唆され、本願の精神および範囲、ならびに添付の請求項の範囲に含まれることが理解されるであろう。

【0272】

実施例1. がん患者の血漿中における組織特異的RNAの検出

【0273】

無細胞RNA (cfRNA) は、がん検出のための有望な分析物であるが、cfRNAの包括的

50

な評価は不足している。血漿中の腫瘍由来RNAの特徴を明らかにするため、Circulating Cell-free Genome Atlas (CCGA) のサブスタディから探索的解析を行い、がんの有無にかかわらず参加者のcfRNA発現を調査した。この解析では、一般集団およびCCGAでの発生率が高いことから、乳がん、肺がん、大腸がんに焦点を当てた。

【0274】

CCGAトレーニングセット (Kleinら、ASCO, 2018) から210名の参加者を選択した。合計98人の参加者は、採血時にステージIIIのがんと診断された (乳房 (47人)、肺 (32人)、大腸 (15人)、および肛門 (4人))。ステージIIIのサンプルは、血液中のシグナルを最大化し、潜在的な二次転移によるシグナルの交絡を避けるために選択された。癌グループと年齢を一致させた112人の非癌患者も含まれた。各患者について、バ

10

【0275】

参加者の血漿から核酸を抽出し、サンプルをDNase処理して無細胞DNA (cfDNA) とゲノムDNAを除去し、ランダムヘキサマープライマーを用いて逆転写を行い、各研究参加者の全転写物を捕捉した。得られたcDNAはDNAライブラリーに変換され、増幅され、リボソーム、ミトコンドリア、グロビンなどの血液関連転写産物から生じる豊富な配列が除去された。得られた全トランスクリプトーム RNA-seq ライブラリーは、サンプルあたり ~750M ペアエンドリードの深さで配列決定し、サンプルごとに各遺伝子の UM I-collapsed counts を生成するカスタムバイオインフォマティクスパイプラインを用いて解析した。この同じ手順で、マッチしたバフィーコートおよび組織 RNA から RNA-seq ライブラリーを作成し、解析した (入手可能な場合)。残留DNA汚染の存在のため、すべての下流分析は、この例では少なくとも1つのリードがエクソン-エクソン接合に重なるリードペアとして定義される、厳密なRNAリードの使用に依存した。図11は、エンドツーエンドワークフローの概要を示している。表9は、参加サンプルの概要を示している。

20

【0276】

[表14]

【0277】

TCGAのRNAサンプルと比較した (図12A)。CCGA腫瘍組織RNA-seqデータをTCGA腫瘍組織RNA-seqデータから得られた主成分に投影すると、CCGA腫瘍組織サンプルは、がんタイプによって分離可能であった (図12B)。これらの結果は、サンプル収集／ハンドリング／ライブラリー調製の違いにもかかわらず、CCGAとTCGA腫瘍の発現プロファイルが非常に類似していることを示唆しており、分析アプローチを検証している。CCGAコホートからのがんcfRNAサンプルのTCGA腫瘍組織RNA-seqデータから得られた主成分への投影は、がんタイプによるサンプルの分離を示さず (図12C)、がんタイプがcfRNAの分散の支配的な原因ではないことを示唆している。

30

【0278】

血漿中のcfRNAの大部分は、健康な免疫細胞に由来すると考えられている。そのため、これらの転写物をバックグラウンドノイズとして扱い、癌のシグナル源として腫瘍由来のcfRNAに着目した。解析の結果、cfRNAデータには2つの遺伝子クラスがあることがわかった。「ダークチャンネル」と「ダークチャンネルバイオマーカー」である。ダークチャンネルとは、がんではない被験者のcfRNAでは検出されなかった遺伝子のことである。57,783のアノテーションされた遺伝子のうち、39,564 (68%) がダークチャンネルとして同定された。ダークチャンネルバイオマーカー (DCB) 遺伝子は、以下の3つの基準を満たした。1) 非がん患者コホートにおける遺伝子の発現中央値がゼロ、2) がん患者コホートにおいて1人以上の参加者で遺伝子発現が検出された、3) がんグループで遺伝子発現がアップレギュレートされた。

40

【0279】

肺癌のDCB遺伝子は14個同定された。SLC34A2、GABRG1、ROS1、AGR2、GNA

50

T3、SFTPA2、MUC5B、SFTA3、SMIM22、CXCL17、BPIFA1、WFDC2、NKX2-1、GGTLC1（表2参照）の14のDCB遺伝子が同定されている。乳がんでは10個のDCB遺伝子が同定された。RNU1-1、CSN1S1、FABP7、OPN1SW、SCGB2A2、LALBA、CASP14、KLK5、WFDC2、VTCN1（表3参照）であった。大腸がんについては、DCB遺伝子は同定されなかった。

【0280】

DCB遺伝子は、いくつかの明確な特徴を示した。まず、DCB遺伝子は、組織特異的な遺伝子に富んでいた（図13）。57,783個の注釈付き遺伝子のうち、肺特異的なものは0.3%、乳腺特異的なものは0.2%であった。一方、肺DCB遺伝子の50%は肺特異的であり、乳DCB遺伝子の44%は乳腺特異的であった（タンパク質アトラスデータベース（Uhlenら, Science, 2015）により定義）。

10

【0281】

さらに、いくつかのDCB遺伝子は、特定のがんサブタイプでのみ検出されるサブタイプ特異的バイオマーカーであった（図14Aおよび図14B）。FABP7は、トリプルネガティブ乳がん（TNBC）サンプルにおいてのみ検出された。逆に、SCGB2A2は、TNBCでは検出されなかったが、HER2+およびHR+/HER-乳がんサンプルで検出された。SLC34A2、ROS1、SFTPA2、CXCL17遺伝子は肺腺癌患者サンプルのcfRNAで検出されたが、扁平上皮癌患者サンプルのcfRNAでは検出されなかった。また、これらのサブタイプ特異的な遺伝子は、同じ臓器に由来する他のサブタイプのがんと比較して、腫瘍組織で高い発現を示した。

20

【0282】

血中の腫瘍関連転写物のソースを決定するために、ダークチャネルバイオマーカー遺伝子のcfRNAと腫瘍組織RNAとの間の一致が評価された。cfRNAと腫瘍組織の発現の間には高い一致性が見られた（図15A）。腫瘍組織で検出されない遺伝子は、適合したcfRNAサンプルで検出される可能性は低く、腫瘍組織で検出された遺伝子は、適合したcfRNAサンプルで検出される可能性が高かった。さらに、所定の患者のcfDNA腫瘍分率と適合した腫瘍組織における遺伝子発現の積として測定される腫瘍含量は、乳癌患者のcfRNAにおけるDCB遺伝子の検出可能性の強い予測因子であった（図15B）。

【0283】

癌以外のcfRNAには見られない転写産物であるダークチャネルバイオマーカー（DCB）は、癌患者において高いS/N比を示す可能性があることがわかった。DCBシグナルは腫瘍量（血液中の腫瘍割合と組織中のRNA発現量の積として測定）と相関していた。cfRNA DCBは、がん患者において組織およびサブタイプに特異的な方法で同定された。腫瘍組織の高発現がDCBシグナルの増幅につながり、cfDNA腫瘍分率が低い患者でもがんの検出が可能になるケースが観察された。これらのデータを総合すると、組織特異的な転写産物は血液ベースの多発性がん検出に利用できる可能性があることが示唆される。

30

【0284】

例2：不均質なサンプルにおけるバイオマーカーの同定

【0285】

我々は、標準的な示差発現（DE）解析を用いた異種試料でのバイオマーカー探索において、2つの共通の偽陽性の原因を観察した。まず、対照群と癌群の両方において、遺伝子の不均一性や遺伝子増幅の脱落により、遺伝子発現が二峰性の分布になること。次に、1つの影響力のある異常値が一般化線形モデル（GLM）の傾きとp値を膨らませた。

40

【0286】

組織発現に基づくcfRNAのような不均一性の高いサンプルにおける差次的発現遺伝子を同定する方法（heteroDEと呼ばれる）を開発した。heteroDEモデルは、負の二項一般化線形モデル（NB-GLM）を使用している。（1）非がんグループの遺伝子発現が、遺伝的不均一性や遺伝子増幅の脱落による二峰性分布に従うかどうかをチェックする、（2）単一の外れ値サンプルだけがNB-GLMのp値に影響を及ぼしているかどうかをチェックする、の2つの機能を追加して、偽陽性を低減するためにheteroDEは構成され

50

ている。外れ値のサンプルはCookの距離を用いて同定される。NB-GLMは、Cookの距離が最も大きいサンプルを除いて、2回目に実行される。

【0287】

先行する差分発現 (DE) 法とは対照的に、heteroDEではNG-GLMの共変量として腫瘍含有率を用いている。非がんサンプルの腫瘍含有量はゼロに設定された。cfRNAの腫瘍バイオマーカー遺伝子の仮説は、その遺伝子の組織での発現が高く、cfDNA中の腫瘍割合が大きいほど、cfRNAでその遺伝子が検出される可能性が高いというものであった。この方法を乳がんサンプルに適用したところ、9つのcfRNAバイオマーカーが同定された。TRGV10、SCGB2A2、CASP14、FABP7、CRABP2、VGLL1、SERPINB5、TF F1、AC007563.5である (表4参照)。これらのバイオマーカーのうち3つ (FABP7、SCGB2A2、CASP14) は、DCB遺伝子として同定された遺伝子と重なる。

10

【0288】

heteroDEに従ったサンプル処理とパラメータ決定を示すワークフロー例を図19に示す。腫瘍含有量は、組織サンプルの不足のため、非癌被験者についてゼロに制約された。ワークフローの実施例は、以下のように与えられる。

【0289】

$K_{i,j}$: 患者jのcfRNAに含まれる遺伝子iのリード数。

【0290】

$\mu_{i,j}$: 患者jのcfRNAにおける遺伝子iの平均リードカウント。

【0291】

α_i : 遺伝子 i の分散 ;

20

【0292】

γ_i : 遺伝子iの血漿中に腫瘍が存在しない場合の平均読み取り数。

【0293】

$x_{i,j}$: 腫瘍内容、 $\log_{10}$ (マッチした cfDNA 中の腫瘍割合 * マッチした腫瘍組織中の遺伝子発現量)

【0294】

β_i : 腫瘍の内容物に対する係数。

【0295】

[数1]

30

【0296】

[数2]

【0297】

また、情報利得法を用いた特徴量選択も検証した。情報利得法とは、2値化されたcfRNA遺伝子発現と癌/非癌ラベルの相互情報量が高い遺伝子を選択する方法である。遺伝子発現のRPM行列を2値行列に変換した。遺伝子発現RPMが0以上の場合は1に変換し、RPM=0の場合は0とした。がん種 (例: 肺がん) と非がんラベルを与え、2値の発現値から各遺伝子の情報量を計算した。乳がん群の非がん群は、非がん群の女性被験者のみを選択して性別のバランスを取った。情報量が多い上位 100 個の遺伝子をモデル化のための特徴量として選択した。各遺伝子の値はバイナリ値に変換してモデリングした。これらの手順を乳癌対非癌、大腸癌対非癌について繰り返した。肺がんに対する情報利得の高い上位 30 遺伝子を表 5 に、乳がんに対する情報利得の高い上位 30 遺伝子を表 6 に示す。

40

【0298】

別の実施形態では、癌組織サンプルで発現し、非癌参加者で発現していない遺伝子を同定するために、癌組織サンプルから特徴選択を実施した。ライブラリーは、実施例1において上述したように調製され、配列決定された。各癌組織サンプルについて、Dark Channelsから、癌組織において比較的高いレベルで発現している遺伝子 (組織RPM > 10) を同定した。これらの遺伝子を "組織明チャンネル遺伝子" と分類した。同定された上位15個の組織ブライトチャンネル遺伝子を表7に示す。

【0299】

50

例3：別コホートでのDCBのバリデーション

【0300】

我々は、CCGAコホートで同定されたDCBを、商業ベンダー（Discovery Life Sciences）から入手した乳がん（38）および肺がん（18）の直交するサンプルで検証することに着手した。ステージI-IVの患者は、疾患の進行に伴うDCBの有病率を評価するために選択され、年齢をマッチさせた38の非がんサンプルが、がんのない患者におけるDCB発現の対照として含まれた。感度を上げ、シーケンスの必要性を減らすために、CCGAコホートで同定された23のDCBを選択するターゲットエンリッチメントアプローチを開発した。また、がん以外の血漿中に通常存在する33の陽性対照遺伝子についてもエンリッチした。非がんサンプルの大半はDCB転写物を含まないため、これらの転写物は濃縮ステップにおけるキャリア材料として機能する。得られたターゲットRNA-seqライブラリーは、サンプルあたり100Mペアエンドリードの深さで配列決定およびサブサンプリングされ、ターゲットおよびオフターゲットの両方の遺伝子について、ストリクトRNAリードの数が定量化された。全トランスクリプトームアッセイと比較すると、ターゲットアプローチは、ターゲットcfRNA転写物の変換効率を2-3倍向上させることが判明した。

【0301】

CCGAコホートで同定された23のDCBのうち、1つ（CRABP2）を除くすべてのDCBの発現中央値（RPM）は、非がんグループにおいて0であった。私たちのパネルに含まれる19のDCBは、検証コホートの少なくとも1つのがんサンプルで発現し（ユニークフラグメント2つ以上）、これらのDCBのうち16は、非がんサンプルと比較して少なくとも1つのがんタイプで差次的に発現していた。アッセイ効率とステージの向上に伴い、いくつかの組織特異的なマーカーが乳がんと肺がんの両方に存在することに気づいたが、それらは2つのグループ間で異なる発現が残っている。また、乳がんではSCGB2A2、肺がんではROS1、SFTA3、SFTPA2のように、あるがん種にのみ発現しているDCBも存在することがわかった。この検証コホートで観察されたすべてのDCBについて、がんサンプルにおけるDCBの発現レベルはステージとともに増加し、我々のコホートではステージIVのサンプルで最も高い発現が見られ、これらの特徴ががんの特異的マーカーとして有効であることを裏付けている。このような傾向がある一方で、早期がんにおいてもDCBの発現が確認されたことから、DCBを強調するアプローチを用いて早期がんを検出する機会があることが示唆された。図16A-Dに、読み取り回数がY軸に沿った例示的な結果を示す。

【0302】

例4：分類の結果

【0303】

ダークチャンネルバイオマーカー（DCB）、ヘテロDE、インフォメーションゲイン（IG）などの異なる特徴選択法を用いて、リーブワンアウト（LOO）および5重クロスバリデーションの分類を適用した。図解ワークフローを図17A～Bに示す。heteroDEはマッチした腫瘍組織を利用するため、肺組織サンプルの数が限られていることから、この特徴選択法は肺がん/非がん分類には適用されなかった。全体として、乳がん/非がん分類では、5重クロスバリデーションと比較して、LOOでは有意に優れた分類性能を示したが、これは、各トレーニングセットのサンプルサイズが小さいため、5重分類では乳がん分類器が十分に訓練されていないことを示唆している。肺がん/非がん分類器ではDCBが、乳がん/非がん分類器ではheteroDEが最も優れた性能（98%特異度での感度：0.303±0.046）を示した（表10）。

【0304】

[表15]

【0305】

例示的な結果は、図18A～Cにもプロットされており、これらはリーブワンアウトクロスバリデーションを用いて生成されたものである。図18Aは、heteroDE特徴選択法およびランダムフォレスト分類器を用いた乳房対非癌のリーブワンアウト（LOO）クロス

バリデーション分類からの受信者動作特性 (ROC) プロットと変数重要度プロットを示している。入力データは、サイズファクター正規化 (DESeq2 RパッケージのestimateSizeFactors) 関数を使用して正規化された遺伝子ごとのカウントであった。表10に示すように、95%での感度は0.394 +/- 0.049であった。

【0306】

図18Bは、ダークチャネル特徴選択法およびランダムフォレスト分類器を用いた肺対非癌ラベルのリーブワンアウト (LOO) クロスバリデーション分類からのROCプロットを示す。入力データは、百万分の一の読み取り数 (rpm) での遺伝子あたりの正規化されたカウントであった。表10に示すように、95%特異度での感度は0.3 +/- 0.042であった。

10

【0307】

図18Cは、ダークチャネル特徴選択法およびランダムフォレスト分類器を用いた乳房対非癌ラベルのリーブワンアウト (LOO) クロスバリデーション分類からのROCプロットと変数重要度プロットを示している。入力データは、遺伝子ごとの正規化カウント (reads per million (rpm)) である。表10に示すように、95%特異度での感度は0.212 +/- 0.041であった。

【0308】

例5：材料と方法

【0309】

シーケンシングデータ処理

20

【0310】

STAR version 2.5.3aを用い、全転写物を含むgencode v19 primary assemblyに生リードをアライメントした。重複配列リードは、ゲノムアライメント位置と非ランダムなUMI配列に基づいて検出・除去された。ペアエンドリードの大部分は、UMI配列が期待される配列と完全に一致した。一部のリードはUMI配列に誤りがあり、ヒューリスティックエラー補正が適用された。UMIが期待されるUMIとのハミング距離が1以内の場合、そのUMI配列に割り当てられた。ハミング距離が1を超える場合、またはハミング距離1以内に複数の既知配列がある場合は、UMIエラーが発生したリードは破棄された。アラインメント位置と修正されたUMIを共有するリードのセットは、メンバーリードの多重配列アラインメントによりエラー修正され、単一のコンセンサス配列/アラインメントが作成された。リードのアラインメントは、gencode v19の注釈付き転写産物と比較された。DNA汚染リードによる誤カウントを防ぐため、注釈付きエクソン-エクソン結合にまたがるリードのみをカウントした。

30

【0311】

サンプル採取

【0312】

全血をStreck Cell-free DNA BCTチューブに採取し、血漿分離前に常温で出荷・保存した。全血をスイングバケットローターで1600g、4℃、10分間スピンして血漿を分離した。血漿層を別のチューブに移し、1500g、4℃で12分間スピンして、さらに細胞性汚染物を除去した。二重紡糸した血漿は-80℃で保存し、クライオプレシピテートの形成を避けるため、抽出前に室温で融解した。

40

【0313】

サンプル抽出基準

【0314】

Circulating Cell-free Genome Atlas study (CCGA, NCT02889978) からステージ III の乳がん、肺がん、大腸がんサンプルのサブセットを選択した。我々は、選択された患者が、少なくとも2本のチューブの未処理グレード1-2血漿 (溶血なし) を持っており、患者あたり6-8mLの血漿を持っていることを条件とした。さらに、選択された患者は、以前の研究で得られたcfDNAシーケンスデータと一致することを条件とした。がん患者を選択した後、年齢、性別、民族ががん患者と一致する非がん患者を同数選択

50

した。この基準に基づいて、我々は210のサンプルを選択した。これらのサンプルは、各バッチ内でがんの種類（がんと非がんサンプル）がランダムに混在することを保証するRのランダム化関数を使用して、14のバッチに無作為化された。

【0315】

サンプル処理

【0316】

QIAamp Circulating Nucleic Acids kit (Qiagen, 55114) の循環miRNAプロトコルを用いて、最大8mLの凍結血漿から無細胞核酸を抽出した。抽出した材料は、RNase-free DNase Set (Qiagen, 79254) を用いてメーカーの指示に従いDNase処理し、High Sensitivity RNA Fragment Analyzer kit (Agilent, DNF-472) を用いて定量した。TruSeq RNA Exome kit (Illumina, 20020189)を用いて逆転写とアダプターライゲーションを行った。得られたライブラリは、AnyDeplete for Human rRNA and Mitochondrial Kit (Tecan, 9132) を用いて豊富な配列を除去し、除去対象のカスタムセットを追加した。

【0317】

配列決定されたサンプルはスクリーニングされ、品質管理指標が低いものはその後の解析から除外された。1つのアッセイメトリックと3つのパイプラインメトリックが「レッドフラッグ」として選ばれ、メトリックが低いサンプルを除外するために使用された。アッセイ指標は、サンプルがシーケンスのために十分な材料を有しているかどうかを測定し、パイプライン指標は、シーケンスの深さ、RNA純度、およびサンプル間の汚染であった。

【0318】

遺伝子発現の定量化

【0319】

データの初期検査では、ライブラリー調製時のDNase消化ステップにもかかわらず、cfRNAサンプルに様々なレベルのDNAが残存していることが明らかになった。汚染のレベルはごくわずかであり（サンプルあたり6ハプロイドゲノム相当未満）、消化前のcfDNAの量やバッチ固有の問題とは相関がなかった。むしろ、これまでの報告と同様に、確率的であるように思われる。

【0320】

QC指標として、発現量の95分位以下の遺伝子の鎖特異性として定義される「quantile 95 strand specificity」を用いて、各サンプルにおけるDNA汚染の程度を評価した。UHRポジティブコントロールサンプルは高いquantile 95 strand specificity (> 0.85)を示し、cfRNA quantile 95 strand specificity値は広い範囲(0.52 - 0.89)に広がっていた。参考までに、cfDNAサンプルのquantile 95 strand specificityは約0.5であり、一部のcfRNAサンプルは残存DNAからのシグナルが支配的であることが示唆される。NC67ではセンスリードとアンチセンスリードが均等に分布しているのに対し、NC3ではセンスリードのみであることがリード鎖の色で示されている。また、NC67では、DNAが存在する場合に予想されるように、イントロンとエクソンの両方に渡って豊富なカバレッジがある。DNAが多く含まれるサンプルの断片長分布は、cfDNAの長さ分布と類似しており（中央値~160）、未消化のcfDNAが主な汚染物質であることが強く示唆された。

【0321】

分位数 95 の鎖特異性が 0.84 未満のサンプルはフラグを立て、以降の解析から除外した。DNAコンタミネーションによるRNAカウントのインフレーションを防ぐため、ここで紹介する遺伝子カウントは、2つのリードのうち少なくとも1つがエクソン-エクソン接合部にマップされるリードペアと定義されるストリクトカウントを用いて作成されている。cfRNA サンプルに混入した cfDNA の量を変化させた実験を行ったところ、Strict Count を用いた RNA レベルの推定値に変化はなく、パイロット試験サンプルにおける遺伝子発現の定量化と比較に Strict Count を使用することが支持された。

【0322】

ダークチャンネル機能選

【0323】

暗黒チャンネル遺伝子は、以下の基準で同定した。1) 非がんグループにおける当該遺伝子の発現量 (RPM) の中央値が0、かつ、当該遺伝子の標準偏差が0.1RPM未満である。各がん種のダークチャンネルバイオマーカー (DCB) は、以下の基準で同定した。1) 指定されたがんグループで、その遺伝子が発現しているサンプルが少なくとも2つある、2) 2番目に高い発現サンプルのRPMが0.1より大きい、3) その遺伝子が非がんグループと比較して指定されたがんグループで差次的に発現する (肺がんは p 値 $< 2e-02$ 、乳がんは p 値 $< 2e-01$)。2群間差分発現の p 値はedgeRパッケージにより算出した。肺癌群と非癌群の間でFDR < 0.05 の遺伝子は816個ある。乳がんと非がんグループ間でFDR < 0.05 の遺伝子は28個ある。大腸がんとは非がんグループ間でFDR < 0.05 の遺伝子は4つある。boxplotとheatmapでは、最も有意な差分発現遺伝子のみを表示した (肺癌と乳癌はFDR $< 2e-06$ 、大腸癌はFDR $< 2e-02$)。

10

【0324】

組織特異的な遺伝子のアノテーションは、以下のように行った。肺癌、乳癌、大腸癌の組織特異的遺伝子ファイルは、Human Protein Atlas のウェブサイト (www.proteinatlas.org/) からダウンロードした。組織特異的遺伝子は、3つのカテゴリーに分類されている。1) Tissue Enriched: 特定の組織において、他のすべての組織と比較して少なくとも4倍以上のmRNAレベルがあるもの、2) Group Enriched: 特定の組織において、他のすべての組織と比較して少なくとも4倍以上のmRNAレベルがあるもの。2) Groupエンリッチド: 2~5組織のグループで少なくとも4倍以上のmRNAレベル、3) Tissueエンハンスト: 2~5組織のグループで少なくとも4倍以上のmRNAレベル。3) Tissue Enhanced: 特定の組織のmRNAレベルが、全組織の平均レベルに対して少なくとも4倍以上高い。この3つのカテゴリーはすべて組織特異的遺伝子の定義に含まれる。

20

【0325】

組織特異的な遺伝子の濃縮度を検定するために1) フィッシャーの正確検定を適用し、すべてのアノテーションされたヒト遺伝子について、肺DCBと肺特異的遺伝子の間の独立性を検定した。2) 乳腺DCBと乳腺特異的遺伝子の独立性を検定するために、フィッシャーの正確検定を適用し、すべての注釈付きヒト遺伝子について検定した。

30

【0326】

実施例6: cfRNAがんバイオマーカーパネル

【0327】

本研究は、肺がんおよび乳がんの特異的なcfRNAバイオマーカーを、がん以外の正常なコホートとは異なる全トランスクリプトームから同定すること、およびがんサンプルのcfRNAに特異的に表され、がんのバイナリ検出や血漿からの起源組織の特定に有用と考えられる生物学的シグナルを同定することを目的としている。私たちは、早期発見が困難ながんサブタイプ、すなわち肺腺がん、HR+およびトリプルネガティブ (TNBC) 乳がんに関連する遺伝子の特徴を特定することに重点を置いて研究を行った。

40

【0328】

この解析を行うために使用したデータには、1) CCGAおよび商業ベンダーから配列決定した全トランスクリプトーム血漿データ、2) TCGAからの全トランスクリプトーム組織データ、3) Human Protein Atlasからの遺伝子注釈 (Uhlen et al, Science 2015) などがある。ステージIIIの乳がんおよび肺がんサンプルのサブセットは、Circulating Cell-free Genome Atlas研究 (CCGA, NCT02889978) から選択し、配列決定した。ステージIIIのサンプルは、潜在的な二次転移からの交絡するシグナルを避けながら、血液中のシグナルを最大化するように選択された。合計で、CCGAの乳がん47検体、肺腺がん14検体、がん以外の血漿93検体を分析した。さらに、商業ベンダー (Conversant) から入手した全トランスクリプトームサンプルの追加セットも含まれている

50

。これには、ステージ IV の乳がん血漿サンプル 14 個が含まれており、血液中のバイオマーカーの後期シグナルを捕捉するために含まれている。これらの血漿データを用いて、健康な血漿で発現している遺伝子と、がん血漿で差次的に発現し、これらのサブタイプにおけるがんのバイナリ検出に有用と思われる遺伝子を定義した。各サンプルの遺伝子発現をRPM (reads per million) 正規化遺伝子特徴行列にまとめ、各サンプルを列、各行を遺伝子特徴としている。

【0329】

また、本研究では、GDC ポータルからダウンロードした TCGA コンソーシアムの乳がん (BRCA) および肺腺がん (LUAD) の組織全体のトランスクリプトームデータも含まれている。これには、ステージI~IVにわたる533の肺腺癌と1102の乳癌サンプルが含まれている。これらのデータは、バイナリ検出のための高発現腫瘍由来遺伝子特徴の同定に使用された。さらに、この高次元データは、TOOに使用できる組織特異的な遺伝子特徴を特定するのに有用であった。各サンプルの遺伝子発現をRPM (reads per million) 正規化遺伝子特徴行列にまとめ、各サンプルを列、各行を遺伝子特徴量とした。

【0330】

これは、がん腫瘍サンプルと健常組織から得られた様々なオミックス技術 (トランスクリプトームと抗体ベース) をオープンアクセスで編集し、組織区画と疾患のアノテーションを提供するものである。これらのアノテーションを用いて、診断時の腫瘍での発現量や患者の全生存率から、その遺伝子のがん種に富むか/富むか、疾患予後に有利か/不利かを捉え、がん種の診断に役立てた。

【0331】

バイナリー検出とTOO分類のためのターゲットセットを確立するために、まず、GDC データポータルからダウンロードしたTCGA組織発現データを使用して、バイオマーカーの可能性が高いかどうかを評価した。各遺伝子について、両コホートにおけるサンプル間の平均遺伝子発現を計算し、コホート間のピアソン相関を計算した。一般に、TCGA組織における高い平均遺伝子発現は、CCGA血漿における高い平均遺伝子発現とほぼ相関することがわかった (乳癌のスピアマンのRhoは0.568、肺癌は0.509であった)。このことから、TCGAの組織データは特徴選択において有益であると考えられる。TCGA組織の平均発現量が1RPMを超える遺伝子は、がん由来血漿で検出される可能性が高く、がんの二次検出または起源組織の検出に役立つ可能性があると考え、優先順位をつけた。アーティファクトの原因となる転写物 (HLA、IGH、IGL、リボソーム遺伝子にマッピングされる転写物) をフィルターにかけた結果、2898個の遺伝子フィーチャーの候補が得られた。

【0332】

しかし、これらの遺伝子の特徴がTCGAの組織で高度に発現していたとしても、血漿でこれらの遺伝子の特徴がどの程度普及しているかは不明であった。血漿と比較した組織における平均RPMのプロットを、図22 (乳癌) および図23 (肺癌) に示す。図21は、癌組織サンプルで高レベルで発現し、血漿ではほとんど検出可能な転写物がない遺伝子についての結果の一例を示している。また、CCGAによる血漿中の発現から得られた情報を活用して、遺伝子特徴量の選択を行った。CCGA血漿サンプルで遺伝子発現が検出されたものと検出されなかったものを2値化し、検出されたものは0.005 reads per million (RPM) 以上の発現であるとした。次に、すべての癌血漿とすべての非癌血漿の観察結果に基づいて、各遺伝子の血漿対数オッズ比 (LOR) を計算した。これは、ある遺伝子が癌サンプルに出現する可能性を、その遺伝子が非癌サンプルに出現する可能性と比較して定量化したものである。LOR > 0は、ある遺伝子のがん症例と非がん症例で検出される可能性が高いことを示し、LOR < 0は、ある遺伝子が非がん症例とがん症例で検出される可能性が高いことを示す。血漿中のLOR > 0.1の最も情報量の多い遺伝子を選択し、281の遺伝子フィーチャーを得た。cfRNAバイオマーカーのLORのプロットの一例を図24に示す。

【0333】

さらに、どの遺伝子特徴がT00の分類に特に有用であるかを評価することに着手した。cfRNAのCCGAデータセットは200サンプル未満に限られているため、TCGA腫瘍遺伝子マトリックスを使用し、再帰的特徴除去アルゴリズムを行って、肺腺癌、乳房HR+、乳房TNBC癌を区別するために重要な遺伝子特徴を特定することに決定した。ランダムフォレスト・マルチクラスモデルを使用し、すべての遺伝子特徴に対して10倍のクロスバリデーションでトップK遺伝子を再帰的に選択した。フォールド間の精度を最適化することにより、反復間で特徴が排除される。クロスバリデーションされたモデルは、750の遺伝子特徴を使用した場合、96.7%の精度でTCGAサンプルを分類するため、これらのトップ750バイオマーカーは、組織におけるサブタイプ分類に重要であると同定された。
【0334】

10

Human Protein Atlasは、がん腫瘍サンプルだけでなく、健康な組織サンプルからのTCGAトランスクリプトミクスと抗体ベースのタンパク質データをコンパイルし、バイナリ検出とT00のための遺伝子特徴の優先順位付けに使用した2つの特定のアトラスを提供している。Tissue Atlasには、正常組織におけるmRNAおよびタンパク質レベルに基づき、組織濃縮型（他の組織と比較して組織で上昇）および組織増強型（低い特異性で組織で発現）の遺伝子に対するアノテーションが含まれている。さらに、Pathology Atlasでは、診断時の腫瘍での発現量や患者の全生存率から、癌種に富む遺伝子（他の腫瘍と比較して腫瘍種で上昇）または増強される遺伝子（腫瘍種で特異性が低く発現）、および疾患予後に有利/不利な遺伝子のアノテーションが含まれている。乳がんと肺がんについて、これらのアノテーションを持つ遺伝子をバイオマーカー候補としてマークした（3028遺伝子の特徴）。
【0335】

20

血漿中に見出される転写産物の大部分は、健康な免疫細胞に由来すると考えられている。がん検出の妨げとなる健康な白血球には存在しないバイオマーカーを選択するために、CCGAコホートの健康な個体の血漿で発現が低い遺伝子フィーチャーをフィルタリングした（中央値RPM < 1、標準偏差RPM < 0.1）。この結果得られた4139個の遺伝子フィーチャーを「ダークチャンネル」と呼ぶことにする。さらに、バイナリ癌検出とT00バイオマーカーの同定における前述のアプローチを統合して、これらのダークチャンネルをフィルタリングした。暗いチャンネルは、がん関連遺伝子の特徴について、遺伝子の2値化LOR > 0.1であるか、またはその遺伝子がランダムフォレストモデルによって選択された750個の遺伝子に含まれるようにフィルタリングされた。これらの遺伝子はさらに、Human Protein Atlasでアノテーションされているか、TCGAコホートで平均発現が5RPM以上であるようにフィルタリングされた。実施例1~4からの追加の陽性対照及びDCB遺伝子が、この更新されたバイオマーカーセットに加えられ、cfRNAバイオマーカーの総数が467となり、これらは表15に記載されている（そのサブセットが表11に提供されている）。表14の遺伝子は、特に情報量の多いcfRNAバイオマーカーのサブセットを表す。乳癌及び肺癌の選択されたバイオマーカーの結果の例は、それぞれ図10A及び図20Bに示されている。
【0336】

30

実施例7：がん検体におけるcfRNAバイオマーカーの評価

40

【0337】

表15に記載された467のcfRNAバイオマーカーは、腫瘍割合の低い検出困難な乳がんや肺がんががんを識別し、非がんを区別する能力についてテストされた。すべてのサンプルは、サンプル中の任意の遺伝子で観察された最も高いエビデンスに基づいてスコア化された。高シグナル癌において何らかのシグナルを示す遺伝子をすべて選択した。各サンプルについて、そのサンプルで他のすべての非癌よりも多くのエビデンスを持つすべての遺伝子を同定し、各サンプルで最もエビデンスの高い遺伝子によってサンプルをランク付けした。以下の基準を用いて、順番に（1）すべての非癌で観察される最大カウント（低いほど良い）、（2）すべての高シグナル癌で観察される最大カウント（高いほど良い）、および（3）そのサンプルで観察されるカウントを行った。これらのバイオマーカー

50

をトレーニングセットとホールドアウトアンブルセットで使用して、リーブワンアウト分類器を評価した。結果は図7に示されている。アスタリスクで示すように、トレーニングコホートと比較して、バリデーションコホートの特異性は有意に低下している ($p=0.02$)。理論に縛られることなく、これはこの特定の実験におけるオーバーフィットの可能性を示していると思われる。

【0338】

DNAメチル化癌バイオマーカーのシグナルが低い、あるいは高い癌サンプルに対して、cfRNAバイオマーカーに基づくleave-one-out分類器を適用した。サンプルには、肺がんおよび乳がんサンプルが含まれる。図8A-8Cに示されるように、この分類器は高い特異性を示した。

【0339】

いくつかの遺伝子は、乳がんや肺がんの特異的であったり、乳がんと肺がんの両方で上昇するなど、特に有用なcfRNAがんバイオマーカーであることが証明された。これらの33の遺伝子は、上記の表8に記載されている。結果は、図26A~26Dにおいて、厳密な読み取り回数についてグラフで示される。これらの33の遺伝子の結果に関する追加の詳細は、以下の表16に提供される。

【0340】

[表16]

【0341】

参考文献

【0342】

Klein 他：複数の種類の腫瘍を早期に検出するための包括的な無細胞 DNA (cfDNA) アッセイの開発。循環型セルフリーゲノムアトラス (CCGA) 研究。ASCO (2018年)。

【0343】

Uhlen et al. Tissue-based map of the human proteome (www.proteinatlas.org).サイエンス doi:10.1126/science.1260419 (2015).

【0344】

A.M. Newman, 他。循環腫瘍DNAを定量化するための超高感度な方法であり、幅広い患者をカバーする。Nat.Med. 20, 548-554 (2014).

【0345】

E.Kirkizlar, 他。乳がん患者の無細胞DNAにおけるクローン性およびサブクローン性コピー数バリエーションの検出。Transl.Oncol. 8, 407-416 (2015).

【0346】

S.Y. Shen, et al.血漿細胞由来DNAメチロームを用いた高感度な腫瘍の検出と分類。ネイチャー 563, 579-583 (2018).

【0347】

C.ベットゴード 他。早期および後期のヒト悪性腫瘍における循環腫瘍DNAの検出。Sci.Transl.Med. 6, 224ra24 (2014).

【0348】

K.C. A. Chan, 他。血漿中 DNA バイサルファイトシーケンスによる、がん関連ゲノムワイドのヒポメチル化およびコピー数異常の非侵襲的検出。Proc.Natl. Acad.Sci. U. S. A. 110, 18761-18768 (2013).

【0349】

I.S. Haque, O. Elemento, Challenges in Using ctDNA to Achieve Early Detection of Cancer (癌の早期発見のためのctDNA利用の課題). bioRxiv, 237578 (2017).

【0350】

K.C. A. Chan, 他。血漿中のがんゲノムスキャン：超並列シーケンシングによる腫瘍関連コピー数異常、一塩基変異、腫瘍の不均一性の検出。C.A.ら、「Clin.Chem. 59,

10

20

30

40

50

211-224 (2013).

【0351】

C. Abbosh, 他。系統的な ctDNA 解析により、早期肺癌の進化が明らかになった。
ネイチャー 545, 446-451 (2017).

【0352】

K.-W. LoLo, 他。上咽頭癌患者の血漿中の無細胞エプスタイン・バー・ウイルス関連
RNAの分析(Analysis of Cell-free Epstein-Barr Virus-associated RNA in the
Plasma of Patients with Nasopharyngeal Carcinoma). Clin.Chem. 45, 12
92-1294 (1999).

【0353】

M.S. Kopreski, F. A. Benko, L. W. Kwak, C. D. Gocke, Detection of tum
or messenger RNA in serum of patients with malignant melanoma (悪性黒
色腫患者の血清中の腫瘍メッセンジャーRNAの検出). Clin.Cancer Res.Off.J. Am.
Cancer Res. 5, 1961-1965 (1999).

【0354】

J.D.アロヨ et al.アルゴノート2複合体は、ヒト血漿中の小胞とは無関係に、循環す
るマイクロRNAの集団を運ぶ。 Proc.Natl. Acad.Sci. U. S. A. 108, 5003-5008
(2011).

【0355】

P.M. Godoy, らヒトの生体内流体間の小分子RNA組成の大きな違い。 Cell Rep. 2
5, 1346-1358 (2018).

【0356】

M.F. de Souza, 他。前立腺癌の診断と予後のための候補マーカーとしての循環mRN
AとmiRNAs。 PLoS ONE 12(2017).

【0357】

G. Y. F. Ho, et al.大腸新生物の重症度に応じた循環型マイクロRNAの発現の違い。
Transl.Res. 166, 225-232 (2015).

【0358】

I.Lee, D. Baxter, M. Y. Lee, K. Scherler, K. Wang, The importance of st
andardization on analyzing circulating RNA. Mol.Diagn.Ther. 21, 259-2
68 (2017).

【0359】

X.Q. Chen, et al.乳がん患者の血清中の検出マーカーとしてのテロメラゼRNA。
Clin.Cancer Res.Off.J. Am.Assoc. Cancer Res. 6, 3823-3826 (2000).

【0360】

17.R. C. Kamm, A. G. Smith, Ribonuclease activity in human plasma. Cl
in.Biochem. 5, 198-200 (1972).

【0361】

T.エル-ヘフナウィ 他。血漿中の増幅可能な循環RNAの特性とがん診断のためのツ
ールとしての可能性。 Clin.Chem. 50, 564-573 (2004).

【0362】

N.B. Y. Tsui, E. K. O. Ng, Y. M. D. Lo, 血液検体、血清、血漿中の内因性およ
び付加RNAの安定性。 Clin.Chem. 48, 1647-1653 (2002).

【0363】

J.D.アロヨ et al.アルゴノート2複合体は、ヒト血漿中の小胞とは無関係に、循環す
るマイクロRNAの集団を運ぶ。 Proc.Natl. Acad.Sci. U. S. A. 108, 5003-5008
(2011).

【0364】

G. J. S. Talhouarne, J. G. Gall, 脊椎動物の赤血球における7SL RNA. RNA 24,
908-914 (2018).

10

20

30

40

50

【0365】

L.A. Hancock, ら Muc5b の過剰発現は、マウスの粘膜纖毛機能障害を引き起こし、肺線維化を促進させる。 *Nat. Commun.* 9, 1-10 (2018).

【0366】

T. Handa, et al. Caspase14 の発現は、乳がん患者のトリプルネガティブ表現型およびがん幹細胞マーカーの発現と関連している。 *J. Surg. Oncol.* 116, 706-715 (2017).

【0367】

R. Hrstka, 他 転移促進タンパク質である前方勾配-2 は、タモキシフェン治療を受けた乳癌の予後不良を予測する。 *オンコジーン* 29, 4838-4847 (2010).

【0368】

M. ピッツィ。 他 肺腺癌における前方勾配2の過剰発現。 *Appl. Immunohistochem. Mol. Morphol. AIMM* 20, 31-36 (2012).

【0369】

H. Cho, A. B. Mariotto, L. M. Schwartz, J. Luo, S. Woloshin, When do changes in cancer survival mean progress? 人口罹患率と死亡率からの洞察。 *J. Natl. Cancer Inst. Monogr.* 2014, 187-197 (2014).

【0370】

Y. M. Lo, ら 母体血漿からの胎児DNAの迅速なクリアランス。 *Am. J. Hum. Genet.* 64, 218-224 (1999).

【0371】

M. A. Watson, T. P. Fleming, Mamaglobin, a mammary-specific member of the uteroglobin gene family, is overexpressed in human breast cancer. *Cancer Res.* 56, 860-865 (1996).

【0372】

G・H・Lewis. 他。浸潤性乳癌の分子サブタイプと肉眼的嚢胞性疾患液蛋白15およびマンマグロビンの発現との関係。 *Am. J. Clin. Pathol.* 135, 587-591 (2011).

【0373】

R.-Z. Liu, 他 脂肪酸結合タンパク質7/RXR β 経路は、トリプルネガティブ乳がんの生存と増殖を促進する。 *J. Pathol.* 228, 310-321 (2012).

【0374】

A. コルデロ et al. FABP7 は HER2+ 乳癌の脳転移における重要な代謝調節因子である。 *オンコジーン* 38, 6445-6460 (2019).

【0375】

H. Zhang, et al. FABP7 と OATP2 というタンパク質は、ヒト乳癌の基底表現型と患者の転帰に関連している。 *Breast Cancer Res. Treat.* 121, 41-51 (2010).

【0376】

J. Xiao, et al. 肺腺癌と扁平上皮癌を区別するための8つの潜在的なバイオマーカー。 *オンコターゲット* 8, 71759-71771 (2017).

【0377】

M. Grageda, P. Silveyra, N. J. Thomas, S. L. DiAngelo, J. Floros, DNA methylation profile and expression of surfactant protein A2 gene in lung cancer (肺癌におけるサーファクタントタンパク質A2遺伝子のDNAメチル化プロファイルと発現). *Exp. Lung Res.* 41, 93-102 (2015).

【0378】

Z. Zhang, et al. SLC34A2 の高発現は肺腺がん患者における良好な予後マーカーである。 *Tumour Biol. J. Int. Soc. Oncodevelopmental Biol. Med.* 39, 1010428317720212 (2017).

【0379】

F. Diehl, et al. 腫瘍の動態を評価するための循環変異型DNA。 *Nat. Med.* 14, 985-990 (2008).

10

20

30

40

50

【0380】

Liu M.C.ら、無細胞DNAのメチル化シグネチャーを用いた高感度かつ特異的な多発性がんの検出と局在化。Ann Oncol.31(6), 745-59 (2020).

【0381】

本開示では、特許、特許出願、特許公報、雑誌、書籍、論文、ウェブコンテンツなど、他の文書への参照や引用が行われている。そのような文書はすべて、すべての目的のために、参照することにより本明細書に組み込まれる。

【0382】

本明細書に示され、説明されたものに加えて、本発明の様々な変更およびその多くのさらなる実施形態は、本明細書に引用された科学および特許文献への参照を含む本書の全内容から当業者には明らかになるであろう。本明細書の主題は、その様々な実施形態及びその等価物における本発明の実施に適応させることができる重要な情報、例示及び指針を含む。本明細書を通じて引用されたすべての文献は、参照により本明細書に明示的に組み込まれる。

10

【0383】

前述の実施形態の詳細な説明は、本開示の特定の実施形態を示す添付の図面を参照する。異なる構造及び動作を有する他の実施形態は、本開示の範囲から逸脱しない。本発明等の用語は、本明細書に記載された出願人の発明の多くの代替的な側面又は実施形態の特定の具体例を参照して使用され、その使用又はその不在は、出願人の発明の範囲又は請求項の範囲を制限することを意図していない。本明細書は、読者の便宜のためにのみ節に分けられている。見出しは、発明の範囲を限定するものとして解釈されるべきではない。定義は、発明の説明の一部として意図されている。本発明の様々な詳細は、本発明の範囲から逸脱することなく変更され得ることが理解されよう。さらに、前述の説明は、例示のみを目的とするものであり、限定を目的とするものではない。

20

【0384】

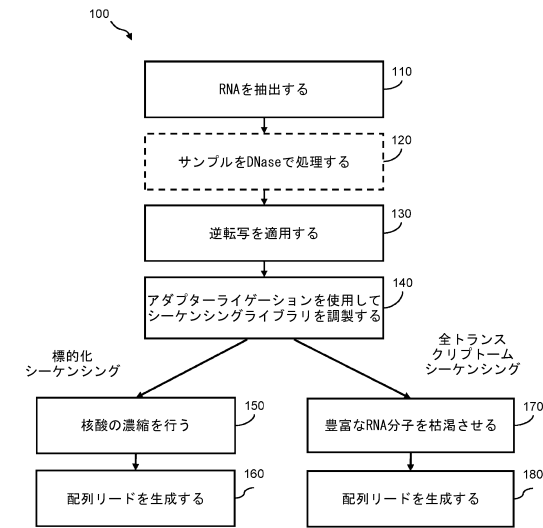
本発明をその特定の実施形態を参照して説明したが、本発明の真の精神および範囲から逸脱することなく、様々な変更がなされ、等価物が置換されてもよいことは、当業者によって理解されるべきである。さらに、特定の状況、材料、物質組成、プロセス、プロセスステップまたはステップに適合させるために、本発明の目的、精神および範囲に多くの変更を加えることができる。全てのそのような修正は、本明細書に添付された特許請求の範囲の範囲内にあることを意図している。

30

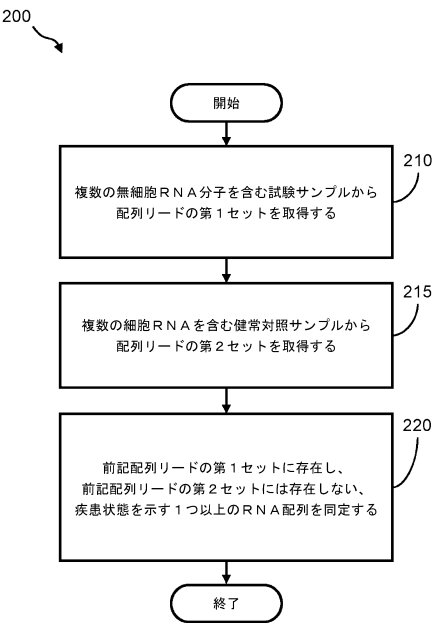
40

50

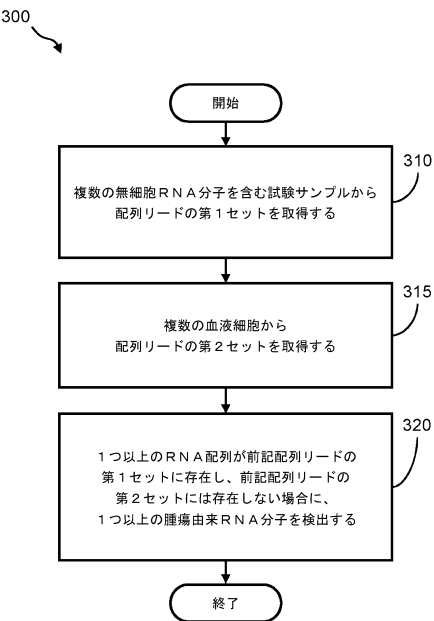
【図面】
【図 1】



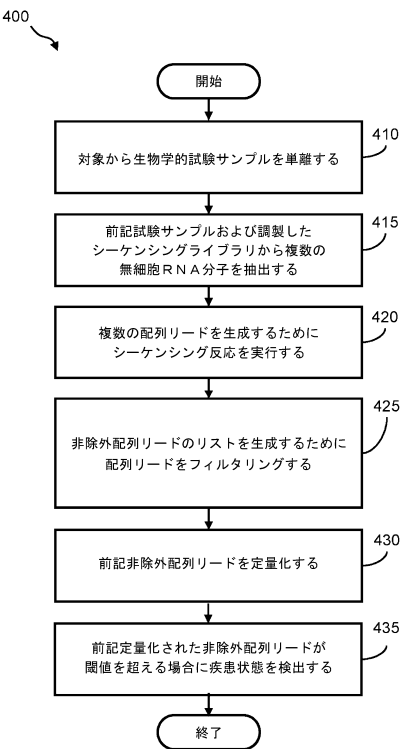
【図 2】



【図 3】



【図 4】



10

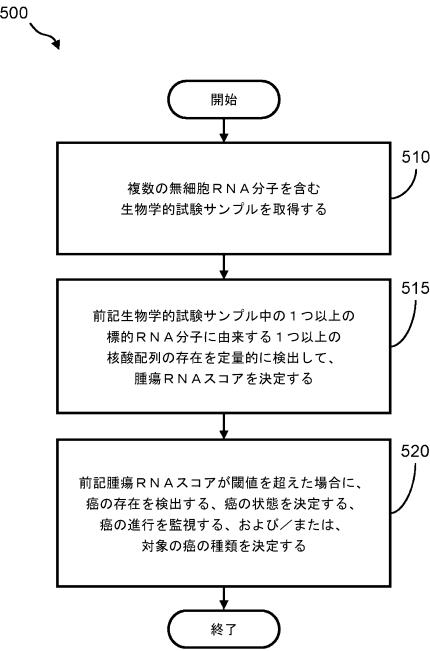
20

30

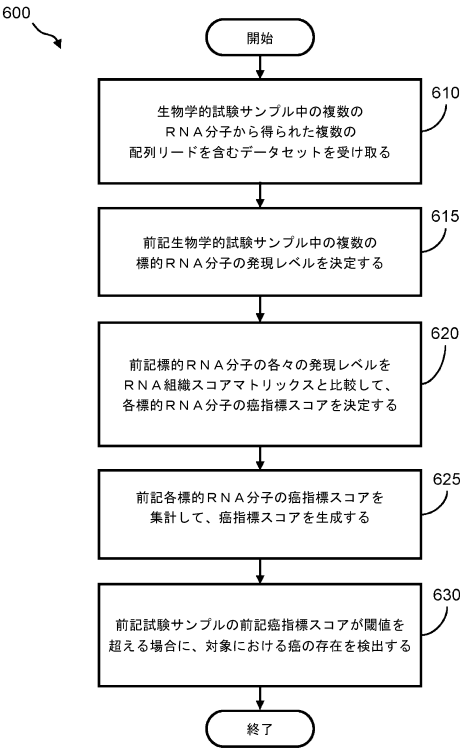
40

50

【図 5】



【図 6】



【図 7】

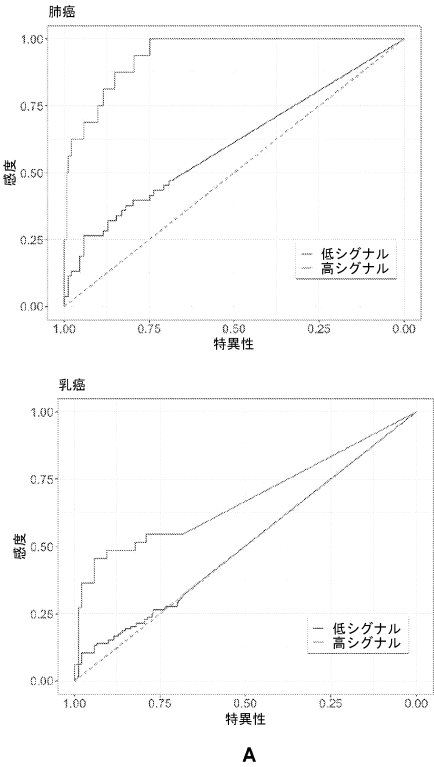
訓練 (119 乳房, 32 肺, 194 非癌):

特異性	感度. 肺	感度. 乳房
.98 [.95 - .99]	.16 [.05 - .33]	.14 [.09 - .22]

ホールドアウト (25 乳房, 21 肺, 49 非癌):

特異性. 標的	特異性. 観測された	感度. 肺	感度. 乳房
.98	.92 [.80 - .98]*	.19 [.05 - .42]	.12 [.03 - .31]

【図 8 A】



10

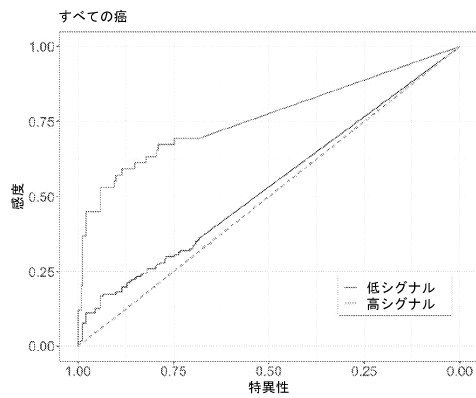
20

30

40

50

【図 8 B】



B

【図 8 C】

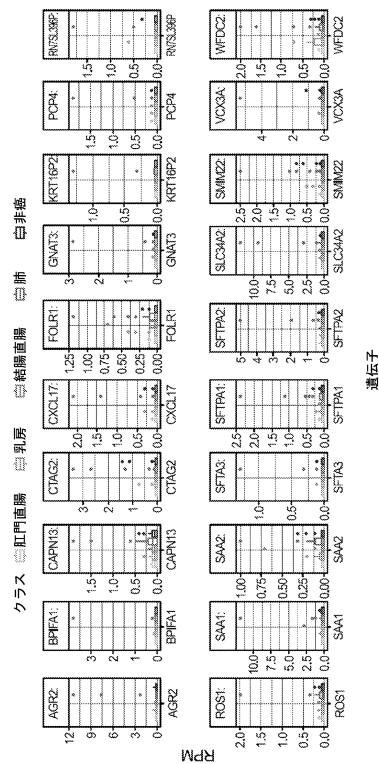
特異性 98 [95 ~ 99]	聴度、肺		聴度、乳房		聴度、すべて	
	低シグナル	高シグナル	低シグナル	高シグナル	低シグナル	高シグナル
	0.13 [05 ~ 25]	0.63 [35 ~ 85]	0.10 [05 ~ 16]	0.36 [20 ~ 55]	0.11 [07 ~ 16]	0.45 [31 ~ 60]

U

10

20

【圖 9】



【図 10】

$$S = G \cdot TS$$

$$\begin{bmatrix} S_{1,1} & S_{1,2} & \dots & S_{1,n} \\ S_{2,1} & S_{2,2} & \dots & S_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ S_{m,1} & S_{m,2} & \dots & S_{m,n} \end{bmatrix} = \begin{bmatrix} G_{1,1} & S_{1,2} & \dots & S_{1,n} \\ G_{2,1} & S_{2,2} & \dots & S_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ G_{m,1} & S_{m,2} & \dots & S_{m,n} \end{bmatrix} \begin{bmatrix} TS_{1,1} & TS_{1,2} & \dots & TS_{1,j} \\ TS_{2,1} & TS_{2,2} & \dots & TS_{2,j} \\ \vdots & \vdots & \ddots & \vdots \\ TS_{n,1} & TS_{n,2} & \dots & TS_{n,j} \end{bmatrix}$$

30

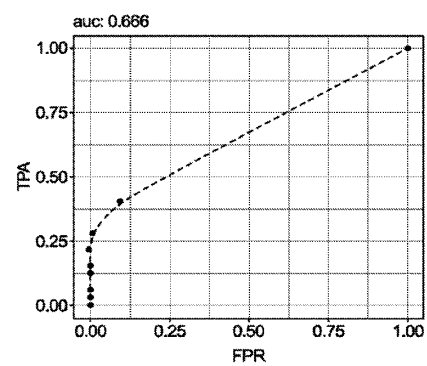
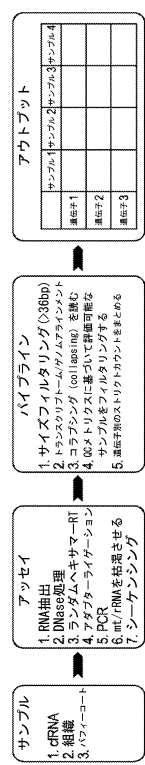


FIG. 10

40

50

【図 1 1】



【図 1 2 A】

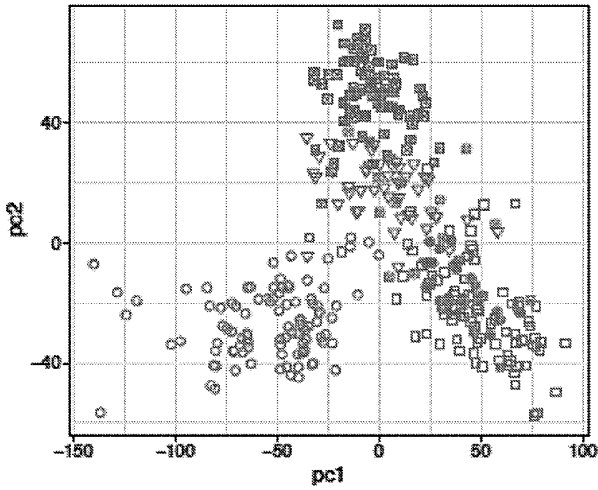


FIG. 12A

【図 1 2 B】

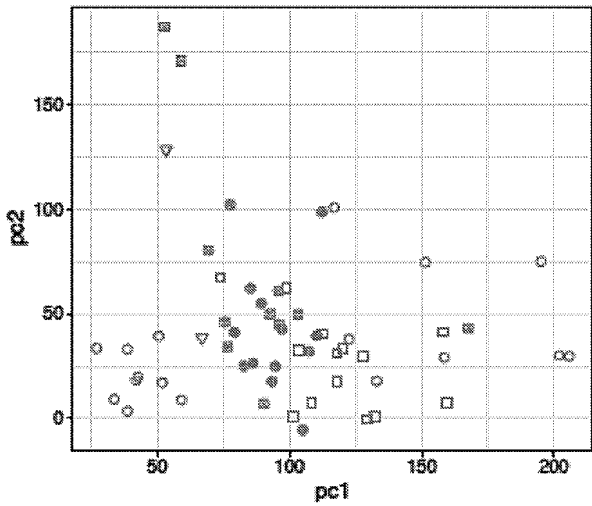


FIG. 12B

【図 1 2 C】

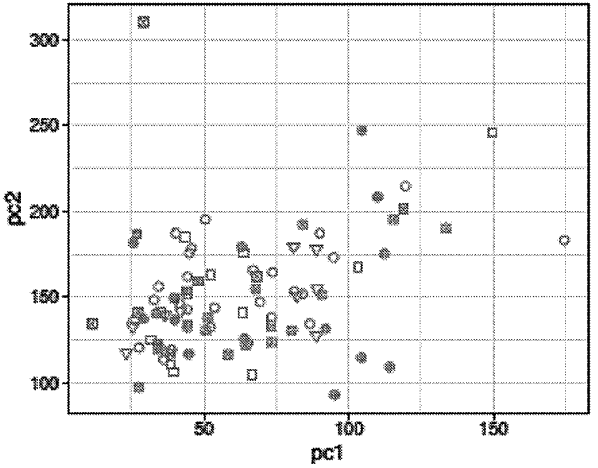


FIG. 12C

10

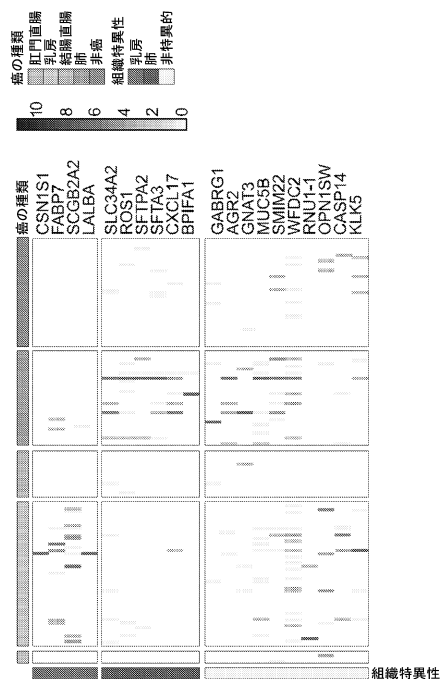
20

30

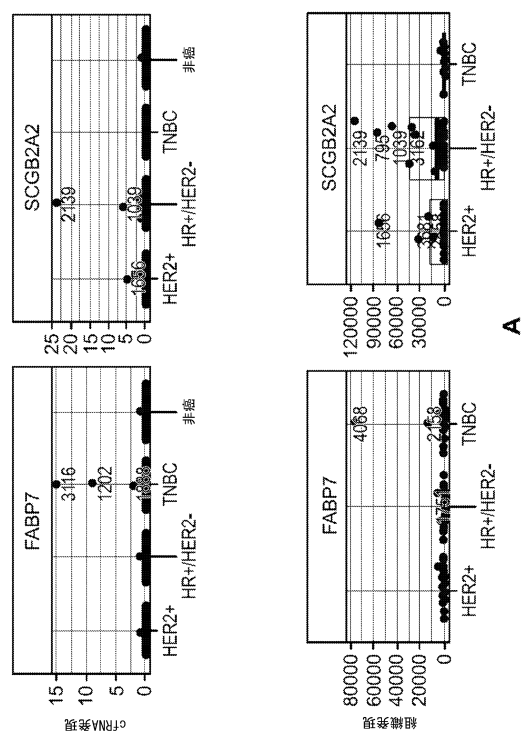
40

50

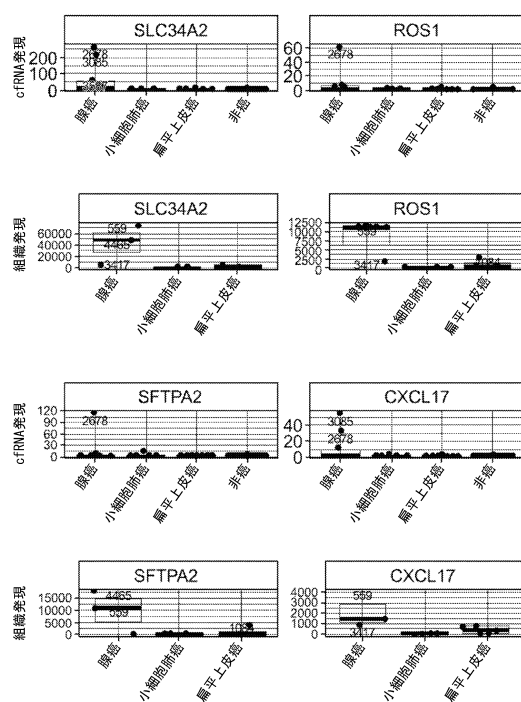
【圖 13】



【図 14 A】

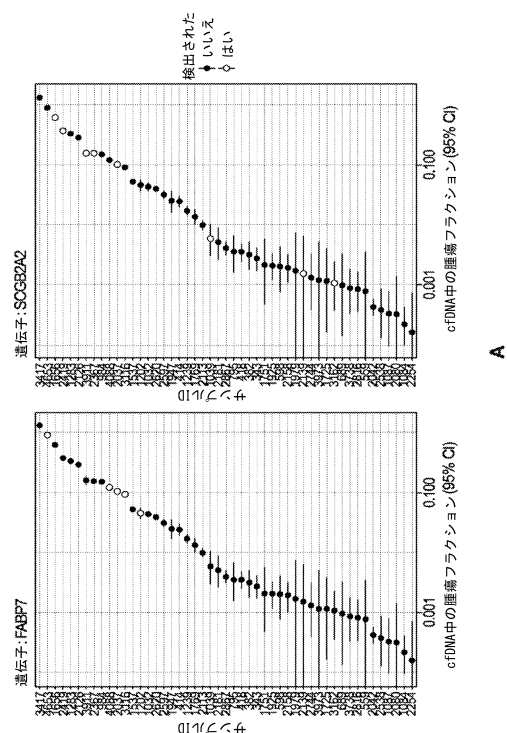


【図 14 B】



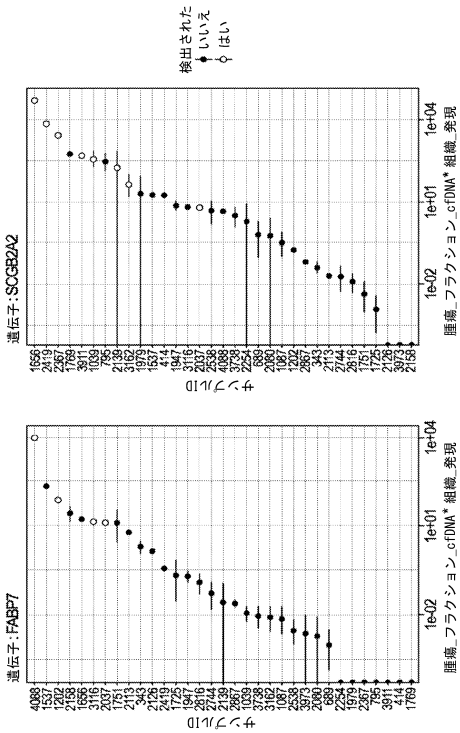
B

【図 15 A】



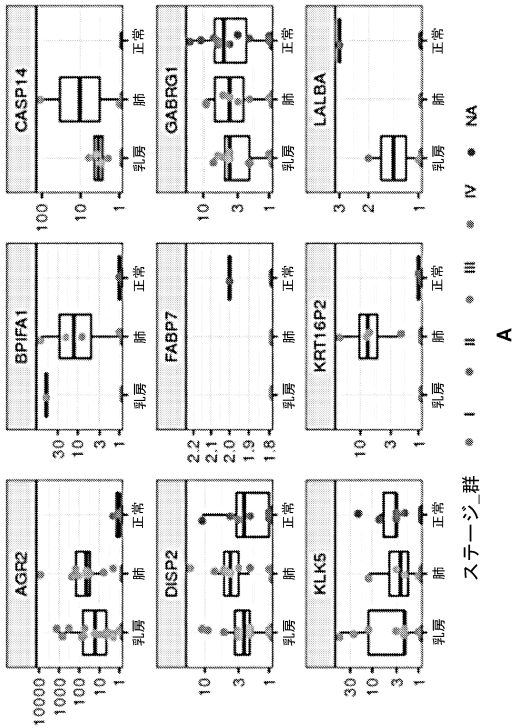
A

【図 1 5 B】



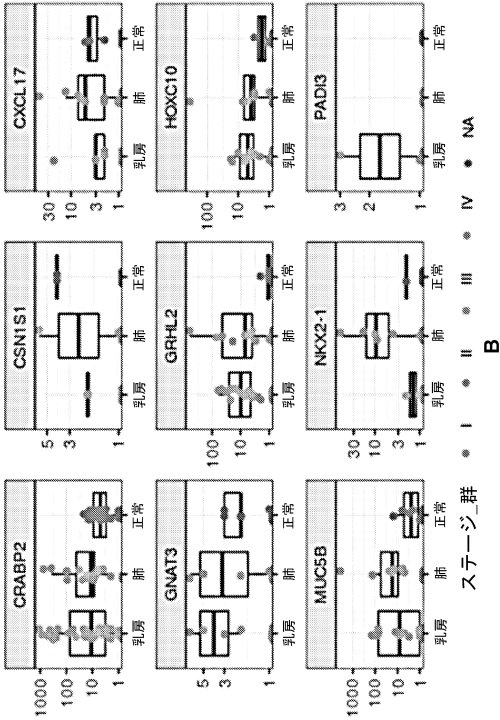
B

【図 1 6 A】



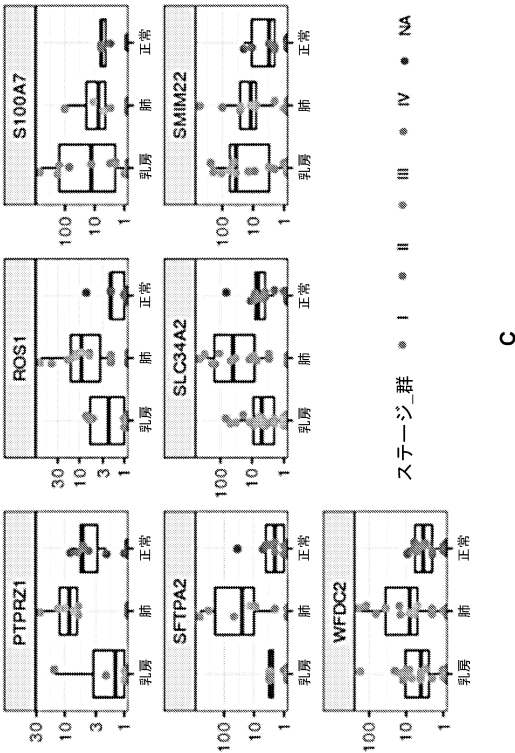
A

【図 1 6 B】



B

【図 1 6 C】



C

10

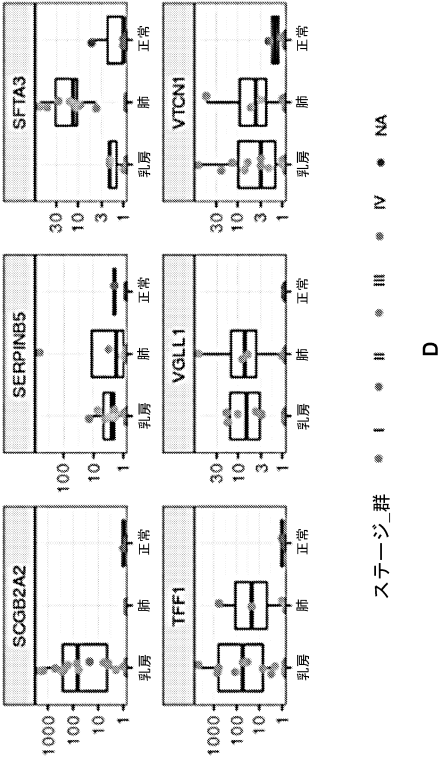
20

30

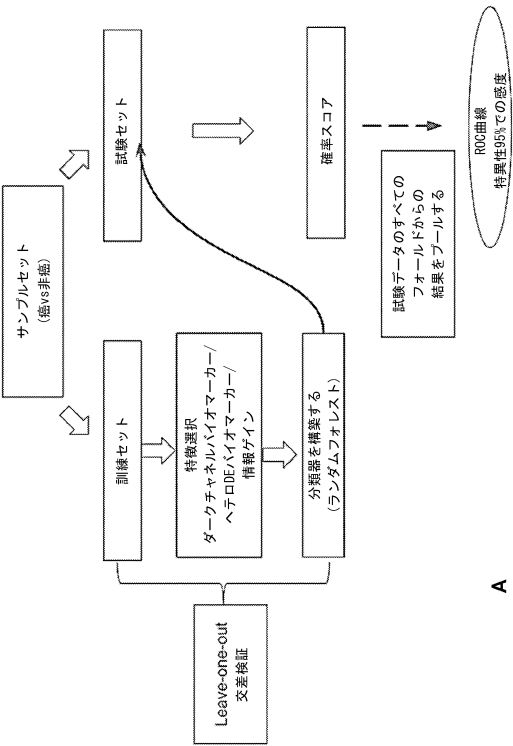
40

50

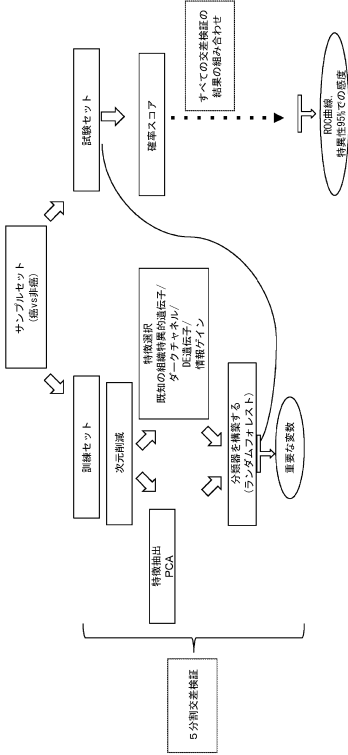
【図 1 6 D】



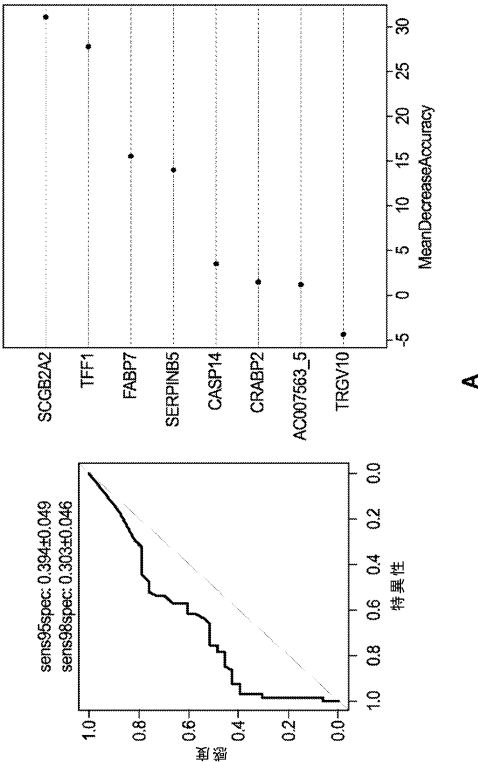
【図 1 7 A】



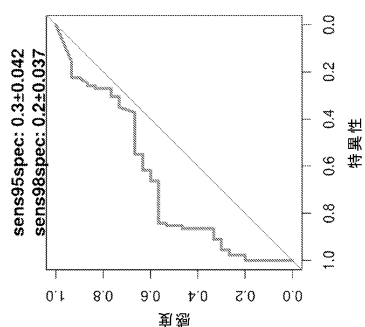
【図 1 7 B】



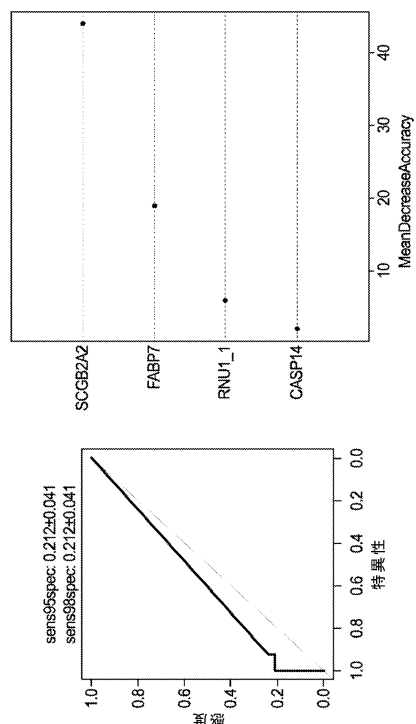
【図 1 8 A】



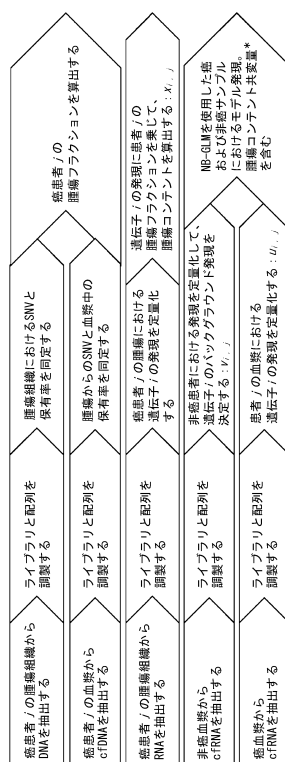
【図 18 B】



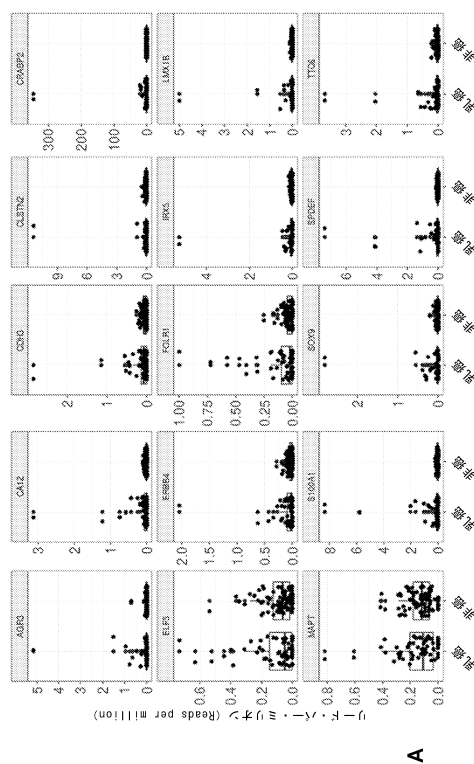
【図 18 C】



【圖 19】

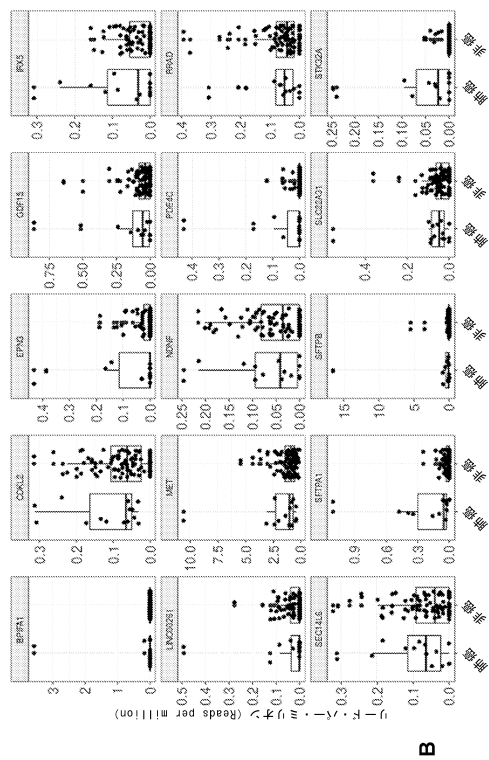


【図 20 A】

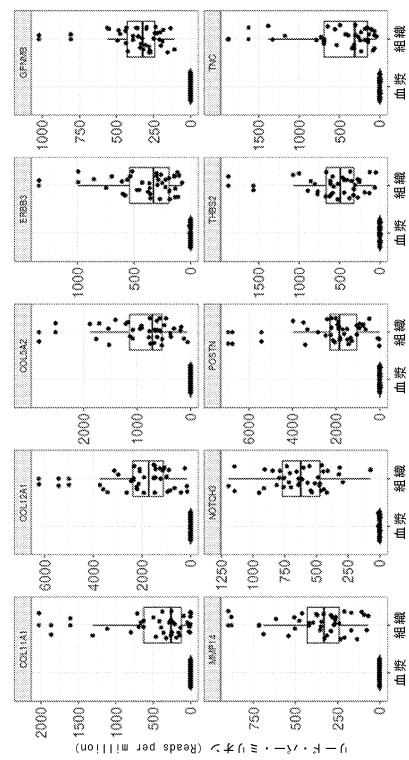


A

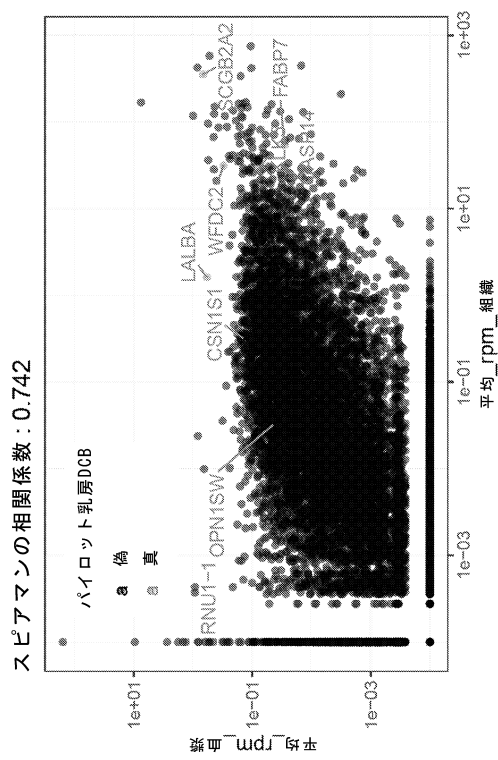
【図 2 0 B】



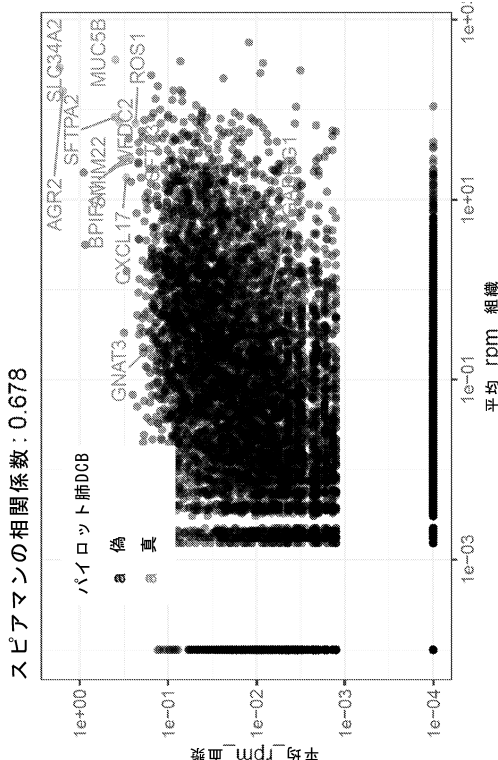
【図 2 1】



【図 2 2】



【図 2 3】



10

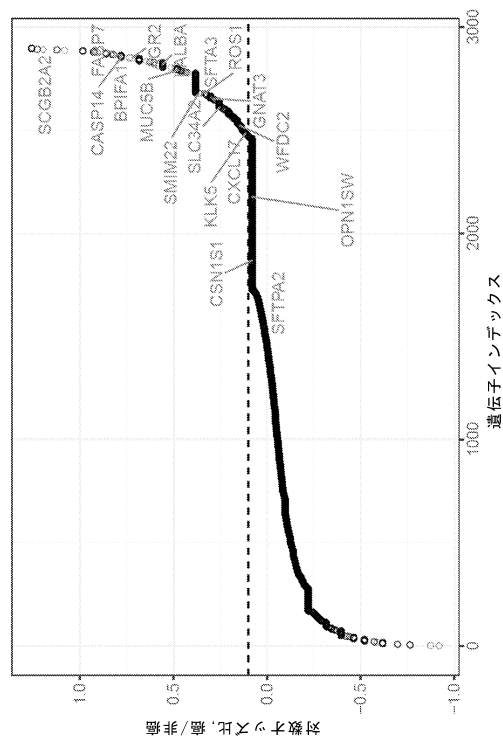
20

30

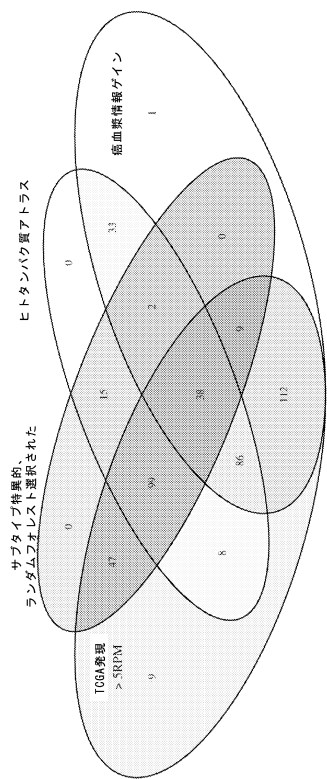
40

50

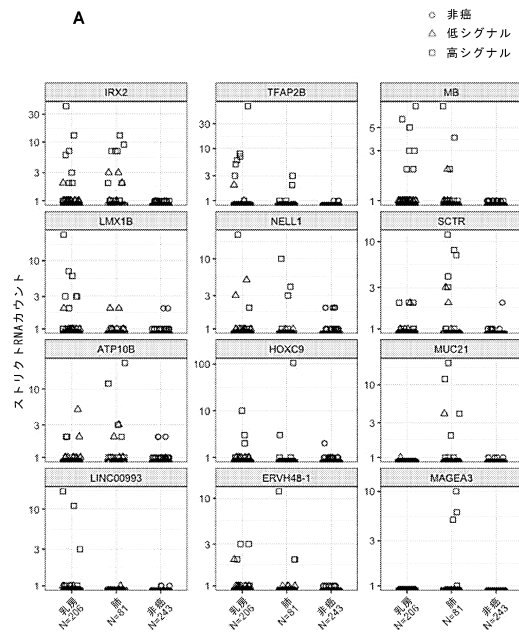
【図 2 4】



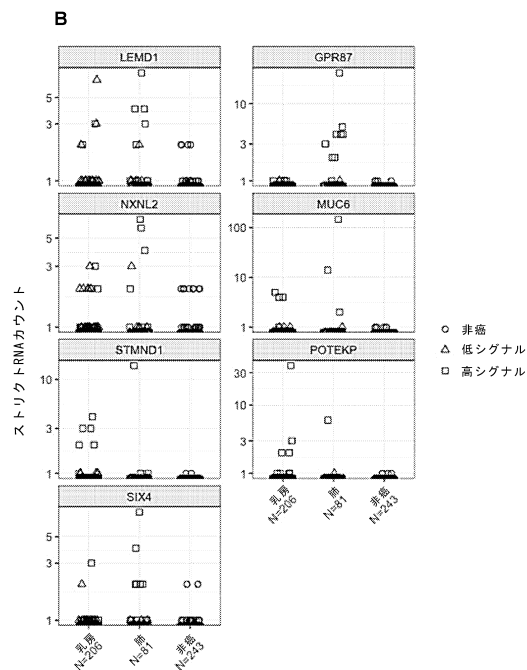
【図 2 5】



【図 2 6 A】



【図 2 6 B】



10

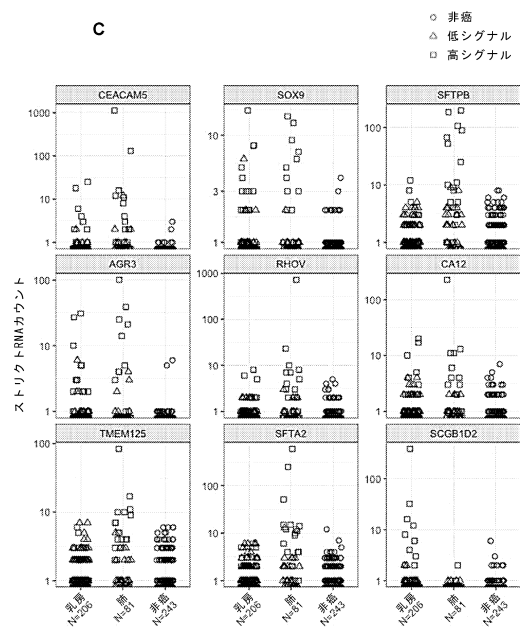
20

30

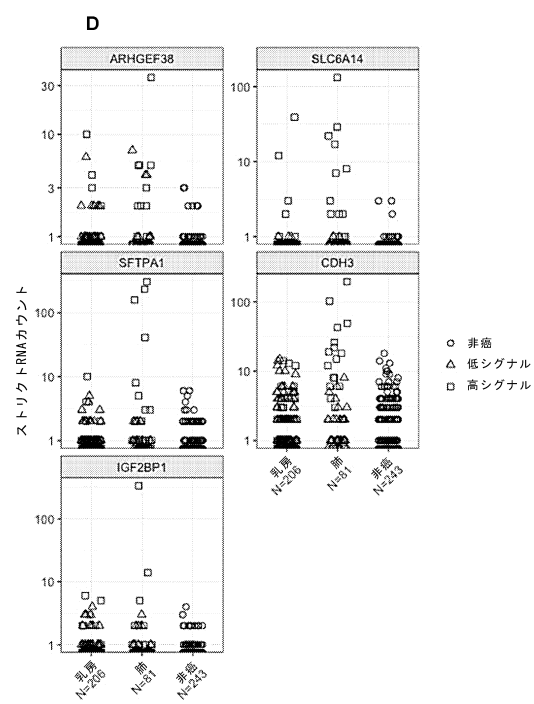
40

50

【図 2 6 C】



【図 2 6 D】



10

20

30

40

50

【国際調査報告】

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US2021/037668

A. CLASSIFICATION OF SUBJECT MATTER

IPC(8) - C12Q 1/6809; C12Q 1/6886 (2021.01)

CPC - C12Q 1/6886; C12Q 2600/112; C12Q 2600/158; G01N 2800/7028 (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 ⁴	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	US 2020/0081008 A1 (PRIME GENOMICS INC.) 12 March 2020 (12.03.2020) entire document	1, 14-34
Y	WO 2019/246625 A1 (BLUESTAR GENOMICS INC.) 26 December 2019 (26.12.2019) entire document	1, 14-34
Y	WO 2019/200410 A1 (FREENOME HOLDINGS INC.) 17 October 2019 (17.10.2019) entire document	25, 28
A	US 2020/0105375 A1 (GAIL INC.) 02 April 2020 (02.04.2020) entire document	1, 14-34
A	US 2019/0316184 A1 (NATERA INC.) 17 October 2019 (17.10.2019) entire document	1, 14-34
P, X	US 2020/0199671 A1 (GAIL INC.) 25 June 2020 (25.06.2020) entire document	1, 14-34

☐ 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 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

04 November 2021

Date of mailing of the international search report

DEC 07 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/US2021/037668

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

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

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:
2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:
3. ☒ Claims Nos.: 39, 40
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

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

This International Searching Authority found multiple inventions in this international application, as follows:
See extra sheet(s).

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying additional fees, this Authority did not invite payment of additional fees.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:
4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:
1, 14-34

Remark on Protest

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

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US2021/037668

Continued from Box No. III Observations where unity of invention is lacking

This application contains the following inventions or groups of inventions which are not so linked as to form a single general inventive concept under PCT Rule 13.1. In order for all inventions to be examined, the appropriate additional examination fees need to be paid.

Group I+: claims 1-38 are drawn to methods of detecting cancer.

The first invention of Group I+ is restricted to a plurality of target cell-free RNAs, selected to be a first target cfRNA AARD and a second target cfRNA ABCA12, and methods comprising the same. It is believed that claims 1 and 14-34 read on this first named invention and thus these claims will be searched without fee to the extent that they read on AARD and ABCA12.

Applicant is invited to elect additional target cfRNAs to be searched in a specific combination by paying additional fee for each set of election. An exemplary election would be a plurality of target cell-free RNAs, selected to be a first target cfRNA ABCC11 and a second target cfRNA ABCC8, and methods comprising the same. Additional target cfRNAs will be searched upon the payment of additional fees. Applicants must specify the claims that read on any additional elected inventions. Applicants must further indicate, if applicable, the claims which read on the first named invention if different than what was indicated above for this group. Failure to clearly identify how any paid additional invention fees are to be applied to the "+" group(s) will result in only the first claimed invention to be searched/examined.

The inventions listed in Groups I+ do not relate to a single general inventive concept under PCT Rule 13.1, because under PCT Rule 13.2 they lack the same or corresponding special technical features for the following reasons:

The Groups I+ formulas do not share a significant structural element responsible for detecting cancer requiring the selection of alternative target cfRNAs where "A method of detecting cancer in a subject, the method comprising: (a) measuring a plurality of target cell-free RNA (cfRNA) molecules in a sample of the subject, wherein the plurality of target cfRNA molecules are selected from one or more transcripts of Table 11; and (b) detecting the cancer, wherein detecting the cancer comprises detecting one or more of the target cfRNA molecules above a threshold level."

Additionally, even if Groups I+ were considered to share the technical features of a method of detecting cancer in a subject, the method comprising: (a) measuring a plurality of target cell-free RNA (cfRNA) molecules in a sample of the subject; and (b) detecting the cancer, wherein detecting the cancer comprises detecting one or more of the target cfRNA molecules above a threshold level; a method of measuring a plurality of target cell-free RNA (cfRNA) molecules in a sample, the method comprising: (a) enriching for the plurality of target cfRNA molecules, or cDNA molecules thereof, to produce an enriched sample of polynucleotides; and (b) sequencing the polynucleotides of the enriched sample, or amplification products thereof. However, these shared technical features do not represent a contribution over the prior art.

Specifically, US 2019/0316184 A1 to Natera, Inc. discloses a method of detecting cancer in a subject (a method for monitoring and detection of early relapse or metastasis of cancer ...a sample of blood or urine or a fraction thereof from a patient, Para. [0003]), the method comprising: (a) measuring a plurality of target cell-free RNA (cfRNA) molecules in a sample of the subject (levels of total or particular ...cfRNA, Para. [0311]; wherein a higher relative allele frequency compared to the other single nucleotide variants of the plurality of single nucleotide variant loci is indicative of a clonal single nucleotide variant in the breast cancer, Para. [0018]; nucleic acids isolated from a sample of blood or urine or a fraction thereof from a patient, Para. [0003]; and (b) detecting the cancer (detected and the presence of the at least 2 or at least 5 SNVs is indicative of early relapse or metastasis of breast cancer, Para. [0007]), wherein detecting the cancer comprises detecting one or more of the target cfRNA molecules above a threshold level (detecting early relapse or metastasis of MIBC when two or more patient-specific single nucleotide variants are detected above a predetermined confidence threshold, Para. [0078]; levels of total or particular ...cfRNA, Para. [0311]); a method of measuring a plurality of target cell-free RNA (cfRNA) molecules in a sample (levels of total or particular ...cfRNA, Para. [0311]; wherein a higher relative allele frequency compared to the other single nucleotide variants of the plurality of single nucleotide variant loci is indicative of a clonal single nucleotide variant in the breast cancer, Para. [0018]; nucleic acids isolated from a sample of blood or urine or a fraction thereof from a patient, Para. [0003]), the method comprising: (a) enriching for the plurality of target cfRNA molecules, or cDNA molecules thereof, to produce an enriched sample of polynucleotides (the sample is enriched for cancer cells or for DNA or RNA from cancer cells, Para. [0533]; levels of total or particular ...cfRNA, Para. [0311]; wherein a higher relative allele frequency compared to the other single nucleotide variants of the plurality of single nucleotide variant loci is indicative of a clonal single nucleotide variant in the breast cancer, Para. [0018]; nucleic acids isolated from a sample of blood or urine or a fraction thereof from a patient, Para. [0003]; and (b) sequencing the polynucleotides of the enriched sample, or amplification products thereof (a SNV amplification/sequencing reaction is performed on one or more tumor samples from the individual, Para. [0207]; the sample is enriched for cancer cells or for DNA or RNA from cancer cells, Para. [0533]).

The inventions listed in Groups I+ therefore lack unity under Rule 13 because they do not share a same or corresponding special technical features.

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)
A 6 1 K 45/00 (2006.01)	A 6 1 K 45/00	
A 6 1 P 35/02 (2006.01)	A 6 1 P 35/02	
G 0 1 N 33/53 (2006.01)	G 0 1 N 33/53	M
G 0 1 N 33/574 (2006.01)	G 0 1 N 33/574	Z
G 1 6 B 40/00 (2019.01)	G 1 6 B 40/00	
C 1 2 N 15/12 (2006.01)	C 1 2 N 15/12	

MK,MT,NL,NO,PL,PT,RO,RS,SE,SI,SK,SM,TR),OA(BF,BJ,CF,CG,CI,CM,GA,GN,GQ,GW,KM,ML,MR,N
E,SN,TD,TG),AE,AG,AL,AM,AO,AT,AU,AZ,BA,BB,BG,BH,BN,BR,BW,BY,BZ,CA,CH,CL,CN,CO,CR,CU,
CZ,DE,DJ,DK,DM,DO,DZ,EC,EE,EG,ES,FI,GB,GD,GE,GH,GM,GT,HN,HR,HU,ID,IL,IN,IR,IS,IT,JO,JP,K
E,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,N
G,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,ZW

(特許庁注：以下のものは登録商標)

1. J A V A

2. V I S U A L B A S I C

3. J A V A S C R I P T

(72)発明者 マシュー ラーソン

アメリカ合衆国 カリフォルニア州 9 4 0 2 5 メンロー パーク オブライエン ドライブ 1 5 2
5 グレイル インク内

(72)発明者 アーチャナ シェノイ

アメリカ合衆国 カリフォルニア州 9 4 0 2 5 メンロー パーク オブライエン ドライブ 1 5 2
5 グレイル インク内

(72)発明者 アラッシュ ジャムシディ

アメリカ合衆国 カリフォルニア州 9 4 0 2 5 メンロー パーク オブライエン ドライブ 1 5 2
5 グレイル インク内

(72)発明者 デイヴィッド ブルクハルト

アメリカ合衆国 カリフォルニア州 9 4 0 2 5 メンロー パーク オブライエン ドライブ 1 5 2
5 グレイル インク内

F ターム (参考) 4B063 QA12 QA19 QQ03 QQ42 QQ52 QR08 QR32 QR35 QR72 QS25
QS34

4C084 AA17 NA05 ZB261 ZB262 ZB271