

(19) 日本国特許庁 (JP)

(12) 公表特許公報 (A)

(11) 特許出願公表番号

特表2021-526375

(P2021-526375A)

(43) 公表日 令和3年10月7日 (2021. 10. 7)

(51) Int. Cl.			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	4 B 0 2 9		
G O 1 N	33/53	(2006.01)	G O 1 N	33/53	M	4 B 0 6 3		
G O 1 N	33/574	(2006.01)	G O 1 N	33/574	A			
C 1 2 N	15/09	(2006.01)	C 1 2 N	15/09	2 0 0			
C 1 2 M	1/00	(2006.01)	C 1 2 M	1/00	A			
			審査請求 未請求 予備審査請求 未請求			(全 58 頁)		
(21) 出願番号 特願2020-567097 (P2020-567097)			(71) 出願人 519156524					
(86) (22) 出願日 令和1年5月31日 (2019. 5. 31)			ジェネオスコピー インコーポレイテッド					
(85) 翻訳文提出日 令和3年1月29日 (2021. 1. 29)			アメリカ合衆国 6 3 1 2 4 ミズーリ州					
(86) 国際出願番号 PCT/US2019/035061			セント ルイス クレイトン ロード					
(87) 国際公開番号 W02019/232483			9 3 0 0					
(87) 国際公開日 令和1年12月5日 (2019. 12. 5)			(74) 代理人 100102978					
(31) 優先権主張番号 62/797, 763			弁理士 清水 初志					
(32) 優先日 平成31年1月28日 (2019. 1. 28)			(74) 代理人 100102118					
(33) 優先権主張国・地域又は機関 米国 (US)			弁理士 春名 雅夫					
(31) 優先権主張番号 62/679, 621			(74) 代理人 100160923					
(32) 優先日 平成30年6月1日 (2018. 6. 1)			弁理士 山口 裕孝					
(33) 優先権主張国・地域又は機関 米国 (US)			(74) 代理人 100119507					
			弁理士 刑部 俊					
			(74) 代理人 100142929					
			弁理士 井上 隆一					
			最終頁に続く					

(54) 【発明の名称】 検出方法

(57) 【要約】

本明細書において、対象、例えば患者由来の便試料中に存在する真核生物核酸内の便由来真核生物RNAバイオマーカーの発現レベルに基づいて結腸直腸新生物および結腸がんを検出するための材料および方法を提供する。該方法は、高リスク腺腫および結腸直腸新生物分子サブタイプの検出のために使用され得る。

【特許請求の範囲】

【請求項1】

a) 表1もしくは表2または表1と表2の組合せに列挙されたバイオマーカーから選択される2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28または29種の便由来真核生物RNAバイオマーカーの、対象由来の便試料から抽出した真核生物核酸における発現のレベルを測定する工程;

b) 該便試料における、該2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28または29種の便由来真核生物RNAバイオマーカーの測定された発現レベルを、対照における、該2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28または29種の便由来真核生物RNAバイオマーカーの測定された発現レベルと比較する工程であって、該対照における、該2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28または29種の便由来真核生物RNAバイオマーカーの測定された発現レベルに対する、該便試料における、該2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28または29種の便由来真核生物RNAバイオマーカーの測定された発現レベルの差は、該対象が結腸直腸新形成を有することを示す、該比較する工程

を含む、対象において結腸直腸新形成を検出する方法。

【請求項2】

対象がヒトである、請求項1記載の方法。

【請求項3】

結腸直腸新形成が、結腸直腸がん、高リスク腺腫、中リスク腺腫および低リスク腺腫からなる群より選択される、請求項1記載の方法。

【請求項4】

便由来真核生物RNAバイオマーカーが、ACY1、TNFRSF10B、DST、EGLN2、PER3、CTNBN1、ACHE、SMAD4、EDN1、ERBB2およびGAPDHからなる群より選択される、請求項1記載の方法。

【請求項5】

前記核酸が、RNA、全RNA、mRNA、seRNA、tRNA、rRNA、ncRNA、smRNAもしくはsnoRNA、またはRNA、全RNA、mRNA、seRNA、tRNA、rRNA、ncRNA、smRNAもしくはsnoRNAのうちの任意のものの組合せを含む、請求項1記載の方法。

【請求項6】

前記発現レベルが、核酸シーケンシング、マイクロアレイシーケンシング、分子バーコーディング、アンプリコンシーケンシング、プローブキャプチャー、ポリメラーゼ連鎖反応(PCR)、ddPCR、dPCR、RT-PCRまたはRT-qPCRによって測定される、請求項1記載の方法。

【請求項7】

対象の個体群統計情報を明らかにする工程をさらに含む、請求項1記載の方法。

【請求項8】

対象に対して免疫化学的便検査(FIT)を施行する工程をさらに含む、請求項1記載の方法。

【請求項9】

結腸直腸新形成を有するか、または結腸直腸新形成のリスクがある対象のための臨床計画を選択する方法であって、

a) 表1もしくは表2または表1と表2の組合せに列挙されたバイオマーカーから選択される2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28または29種の便由来真核生物RNAバイオマーカーの、該対象由来の便試料中に存在する真核生物核酸における発現のレベルを測定する工程;

b) 該便試料における、該2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28または29種の便由来真核生物RNAバイオマーカーの測定された発現レベルを、対照における、該2、3、4、5、6、7、8、9、10、11

10

20

30

40

50

、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28または29種の便由来真核生物RNAバイオマーカーの測定された発現レベルと比較する工程であって、該対照における、該2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28または29種の便由来真核生物RNAバイオマーカーの測定された発現レベルに対する、該便試料における、該2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28または29種の便由来真核生物RNAバイオマーカーの測定された発現レベルの差は、該対象が結腸直腸新形成を有するか、または結腸直腸新形成のリスクがあることを示す、該比較する工程；

c) 診断手順もしくは処置または診断手順と処置の組合せを施行する工程を含む、前記方法。

10

【請求項10】

対象がヒトである、請求項9記載の方法。

【請求項11】

結腸直腸新形成が、結腸直腸がん、高リスク腺腫、中リスク腺腫および低リスク腺腫からなる群より選択される、請求項9記載の方法。

【請求項12】

便由来真核生物RNAバイオマーカーが、ACY1、TNFRSF10B、DST、EGLN2、PER3、CTNNB1、ACHE、SMAD4、EDN1、ERBB2およびGAPDHからなる群より選択される、請求項9記載の方法。

【請求項13】

前記核酸が、RNA、全RNA、mRNA、seRNA、tRNA、rRNA、ncRNA、smRNAもしくはsnoRNA、またはRNA、全RNA、mRNA、seRNA、tRNA、rRNA、ncRNA、smRNAもしくはsnoRNAのうちの任意のものの組合せを含む、請求項9記載の方法。

20

【請求項14】

前記発現レベルが、核酸シーケンシング、マイクロアレイシーケンシング、分子バーコーディング、アンプリコンシーケンシング、プローブキャプチャー、ポリメラーゼ連鎖反応（PCR）、ddPCR、dPCR、RT-PCRまたはRT-qPCRによって測定される、請求項9記載の方法。

【請求項15】

対象の個体群統計情報を明らかにする工程をさらに含む、請求項9記載の方法。

【請求項16】

対象に対して免疫化学的便検査（FIT）を施行する工程をさらに含む、請求項9記載の方法。

30

【請求項17】

臨床計画が診断手順または処置を含む、請求項9記載の方法。

【請求項18】

診断手順が結腸内視鏡検査を含む、請求項9記載の方法。

【請求項19】

処置が、手術、化学療法、放射線療法、標的療法または免疫療法を含む、請求項9記載の方法。

【請求項20】

化学療法が、5-フルオロウラシル、ロイコボリン、カペシタビン、オキサリプラチン、イリノテカンまたはそれらの組合せの投与を含む、請求項19記載の方法。

40

【請求項21】

標的療法が、ベバシズマブ（抗VEGF）、ラムシラマブ（抗VEGFR2）、アフリベルセプト、レゴラフェニブ、セツキシマブ（抗EGFR）、パニツムマブ、トリプフルリジン-チピラシルまたはそれらの組合せの投与を含む、請求項19記載の方法。

【請求項22】

a) 表1もしくは表2または表1と表2の組合せに列挙されたバイオマーカーから選択される2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28または29種の便由来真核生物RNAバイオマーカーの発現レベルを含

50

むフィーチャーの所定のセットに基づいて結腸直腸新形成の存在を判定するように構成されるランダムフォレストモデルを作成する工程;

b) 表1もしくは表2または表1と表2の組合せに列挙されたバイオマーカーから選択される2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28または29種の便由来真核生物RNAバイオマーカーの、対象における発現レベルを測定する工程;および

c) 表1もしくは表2または表1と表2の組合せに列挙されたバイオマーカーから選択される該2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28または29種の便由来真核生物RNAバイオマーカーの、対象における発現レベルに基づいて、結腸直腸新形成の存在を、該ランダムフォレストモデルによって判定する工程

を含む、対象において結腸直腸新形成を検出する方法。

【請求項 2 3】

フィーチャーの所定のセットが個体群統計情報を含む、請求項22記載の方法。

【請求項 2 4】

ランダムフォレストモデルによって陽性免疫化学的便検査 (FIT) 強制モデルを適用する工程を含む、請求項22記載の方法。

【請求項 2 5】

前記2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28または29種の便由来真核生物RNAバイオマーカーが、ACY1、TNF、RSF10B、DST、EGLN2、PER3、CTNNB1、ACHE、SMAD4、EDN1、ERBB2およびGAPDHからなる群より選択される、請求項22記載の方法。

【請求項 2 6】

a) 患者由来の便試料が、対象由来の便試料中に存在する真核生物核酸において、表1もしくは表2または表1と表2の組合せに列挙されたバイオマーカーから選択される2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28または29種の便由来真核生物RNAバイオマーカーの増加したレベルを発現していると判定する工程;および

b) 該対象に対して結腸内視鏡検査、手術、化学療法、放射線療法、標的療法または免疫療法を施行する工程

を含む、対象の結腸直腸新形成を処置する方法。

【請求項 2 7】

a) 表3に列挙されたバイオマーカー遺伝子から選択される1つまたは複数のバリエントバイオマーカー遺伝子の、対象由来の便試料から抽出した真核生物核酸におけるバリエントアレル頻度を測定する工程;

b) 該便試料における、該1つまたは複数のバリエントバイオマーカー遺伝子の測定されたバリエントアレル頻度を、対照における、該1つまたは複数のバリエントバイオマーカー遺伝子の測定されたバリエントアレル頻度と比較する工程であって、該対照における、該1つまたは複数のバリエントバイオマーカー遺伝子のバリエントアレル頻度に対する、該1つまたは複数のバリエントバイオマーカー遺伝子のバリエントアレル頻度の差は、該対象が結腸直腸がんを有するか、または結腸直腸がんのリスクがあることを示す、該比較する工程

を含む、対象において結腸直腸新形成を検出する方法。

【請求項 2 8】

a) 表3に列挙されたバイオマーカー遺伝子から選択される1つまたは複数のバリエントバイオマーカー遺伝子の、対象由来の便試料から抽出した真核生物核酸におけるバリエントアレル頻度を測定する工程;

b) 該便試料における、該1つまたは複数のバリエントバイオマーカー遺伝子の測定されたバリエントアレル頻度を、対照における、該1つまたは複数のバリエントバイオマーカー遺伝子の測定されたバリエントアレル頻度と比較する工程であって、該対照における、

該1つまたは複数のバリエーションバイオマーカー遺伝子のバリエーションアレル頻度に対する、該1つまたは複数のバリエーションバイオマーカー遺伝子のバリエーションアレル頻度の差は、該対象が結腸直腸がんを有するか、または結腸直腸がんのリスクがあることを示す、該比較する工程

を含む、対象が結腸直腸がんを有するかどうか、または結腸直腸がんのリスクがあるかどうかを判定する方法。

【請求項 29】

結腸直腸がんを有するか、または結腸直腸がんのリスクがある対象のための臨床計画を選択する方法であって、

a) 該対象由来の生物学的試料中の真核生物核酸における、表3に列挙されたバイオマーカー遺伝子から選択されるバイオマーカー遺伝子の1つまたは複数のバリエーションアレルを検出する工程であって、該バリエーションが結腸直腸がんの腫瘍形成に関連している、該検出する工程;

b) 診断手順もしくは処置または診断手順と処置の組合せを施行する工程を含む、前記方法。

【請求項 30】

a) 対象由来の便試料中に存在する真核生物核酸における、表3に列挙されたバイオマーカー遺伝子から選択されるバイオマーカー遺伝子の1つまたは複数のバリエーションアレルを検出する工程であって、該バリエーションが結腸直腸がんの腫瘍形成に関連している、該検出する工程;および

b) 該対象に対して結腸内視鏡検査、手術、化学療法、放射線療法、標的療法または免疫療法を施行する工程を含む、対象の結腸直腸がんを処置する方法。

【請求項 31】

バリエーションアレルが、サイレント突然変異、ミスセンス突然変異、挿入、欠失、フレームシフト突然変異および/またはナンセンス突然変異を含む、請求項27~30のいずれか一項記載の方法。

【請求項 32】

生物学的試料が便試料である、請求項27~31のいずれか一項記載の方法。

【請求項 33】

便試料がヒトの便試料である、請求項27~32のいずれか一項記載の方法。

【請求項 34】

前記核酸が、cDNA、RNA、全RNA、mRNA、seRNA、tRNA、rRNA、ncRNA、smRNAもしくはsnoRNA、またはcDNA、RNA、全RNA、mRNA、seRNA、tRNA、rRNA、ncRNA、smRNAもしくはsnoRNAのうちの任意のものの組合せを含む、請求項27~33のいずれか一項記載の方法。

【請求項 35】

前記発現レベルが、核酸シーケンシング、マイクロアレイシーケンシング、分子バーコーディング、プローブキャプチャー、ポリメラーゼ連鎖反応(PCR)、ddPCR、RT-PCRまたはRT-qPCRによって測定される、請求項27~34のいずれか一項記載の方法。

【請求項 36】

前記臨床計画が、手術、化学療法、放射線療法、免疫療法および標的療法のうちの1つまたは複数を含む、請求項27~35のいずれか一項記載の方法。

【請求項 37】

a) 表4に列挙された結腸直腸新生物分子サブタイプバイオマーカー遺伝子のうちの任意のものから選択される2つまたはそれより多くのバイオマーカー遺伝子の、対象由来の便試料から抽出した真核生物核酸における発現のレベルを測定する工程;

b) 生物学的試料における、該2つまたはそれより多くの結腸直腸新生物分子サブタイプバイオマーカー遺伝子の測定された発現レベルを、対照における、該2つまたはそれより多くの結腸直腸新生物分子サブタイプバイオマーカー遺伝子の測定された発現レベルと比較する工程であって、該対照における、該2つまたはそれより多くの結腸直腸新生物分子

10

20

30

40

50

サブタイプバイオマーカー遺伝子に対する、該2つまたはそれより多くの結腸直腸新生物分子サブタイプバイオマーカー遺伝子の測定された発現レベルとの、該生物学的試料における、該2つまたはそれより多くの結腸直腸新生物分子サブタイプバイオマーカー遺伝子の測定された発現レベルの差は、結腸直腸がんの該分子サブタイプを示す、該比較する工程

を含む、対象において結腸直腸がんの分子サブタイプを検出する方法。

【請求項 38】

a) 表4に列挙された結腸直腸新生物分子サブタイプバイオマーカー遺伝子のうちの任意のものから選択される2つまたはそれより多くのバイオマーカー遺伝子の、対象由来の便試料から抽出した真核生物核酸における発現のレベルを測定する工程;

10

b) 生物学的試料における、該2つまたはそれより多くの結腸直腸新生物分子サブタイプバイオマーカー遺伝子の測定された発現レベルを、対照における、該2つまたはそれより多くの結腸直腸新生物分子サブタイプバイオマーカー遺伝子の測定された発現レベルと比較する工程であって、該対照における、該2つまたはそれより多くの結腸直腸新生物分子サブタイプバイオマーカー遺伝子に対する、該2つまたはそれより多くの結腸直腸新生物分子サブタイプバイオマーカー遺伝子の測定された発現レベルとの、該生物学的試料における、該2つまたはそれより多くの結腸直腸新生物分子サブタイプバイオマーカー遺伝子の測定された発現レベルの差は、該対象が結腸直腸がんの分子サブタイプを有するか、または結腸直腸がんの分子サブタイプのリスクがあることを示す、該比較する工程を含む、対象が結腸直腸がんの分子サブタイプを有するかどうか、または結腸直腸がんの分子サブタイプのリスクがあるかどうかを判定する方法。

20

【請求項 39】

a) ゲノム不安定性、マイクロサテライト不安定性または免疫浸潤に関係している表4に列挙された結腸直腸新生物分子サブタイプバイオマーカー遺伝子の任意のものから選択される2つまたはそれより多くのバイオマーカー遺伝子の、患者由来の便試料中の真核生物核酸における発現のレベルを測定する工程;

b) 生物学的試料における、該2つまたはそれより多くの結腸直腸新生物分子サブタイプバイオマーカー遺伝子の測定された発現レベルを、対照における、該2つまたはそれより多くの結腸直腸新生物分子サブタイプバイオマーカー遺伝子の測定された発現レベルと比較する工程であって、対照における、該2つまたはそれより多くの遺伝子の測定されたレベルに対する、マイクロサテライト不安定性または免疫浸潤に関係している該結腸直腸新生物分子サブタイプバイオマーカー遺伝子の変化したレベルは、該患者が免疫療法の候補であることを示す、該比較する工程;および

30

c) 免疫療法を施行する工程

を含む、患者の結腸直腸がんを処置する方法。

【請求項 40】

免疫療法が阻害性チェックポイント分子または免疫チェックポイント阻害薬である、請求項39記載の方法。

【請求項 41】

結腸直腸がんを有するか、または結腸直腸がんのリスクがある対象のための臨床計画を選択する方法であって、

40

a) 表4に列挙された結腸直腸新生物分子サブタイプバイオマーカー遺伝子の任意のものから選択される2つまたはそれより多くのバイオマーカー遺伝子の、該対象由来の生物学的試料中の真核生物核酸における発現のレベルを測定する工程;

b) 該生物学的試料における、該2つまたはそれより多くの結腸直腸新生物分子サブタイプバイオマーカー遺伝子の測定された発現レベルを、対照における、該2つまたはそれより多くの結腸直腸新生物分子サブタイプバイオマーカー遺伝子の測定された発現レベルと比較し、それにより結腸直腸がんの該分子サブタイプを同定する工程;

c) 診断手順もしくは処置または診断手順と処置の組合せを施行する工程を含む、前記方法。

50

【請求項 4 2】

a) 患者由来の便試料が、対象由来の便試料中に存在する真核生物核酸において、表4に列挙された結腸直腸新生物分子サブタイプバイオマーカー遺伝子のうちの任意のものの増加したレベルを発現していると判定する工程;および

b) 該患者に対して結腸内視鏡検査、手術、化学療法、放射線療法、標的療法または免疫療法を施行する工程

を含む、結腸直腸新形成を処置する方法。

【請求項 4 3】

結腸直腸新生物分子サブタイプが、CMS1、CMS2、CMS3およびCMS4からなる群より選択される、請求項37~42のいずれか一項記載の方法。

10

【請求項 4 4】

生物学的試料が便試料である、請求項37~43のいずれか一項記載の方法。

【請求項 4 5】

便試料がヒトの便試料である、請求項37~44のいずれか一項記載の方法。

【請求項 4 6】

前記核酸が、cDNA、RNA、全RNA、mRNA、seRNA、tRNA、rRNA、ncRNA、smRNAもしくはsnoRNA、またはcDNA、RNA、全RNA、mRNA、seRNA、tRNA、rRNA、ncRNA、smRNAもしくはsnoRNAのうちの任意のものの組合せを含む、請求項37~45のいずれか一項記載の方法。

【請求項 4 7】

前記発現レベルが、核酸シーケンシング、マイクロアレイシーケンシング、分子バーコーディング、プローブキャプチャー、ポリメラーゼ連鎖反応（PCR）、ddPCR、RT-PCRまたはRT-qPCRによって測定される、請求項37~46のいずれか一項記載の方法。

20

【請求項 4 8】

前記臨床計画が、手術、化学療法、放射線療法、免疫療法および標的療法のうちの1つまたは複数を含む、請求項37~47のいずれか一項記載の方法。

【請求項 4 9】

a) 対照に対して、患者由来の便試料が、対象由来の便試料中に存在する真核生物核酸において、表1もしくは表2または表1と表2の組合せまたは表3または表4に列挙されたバイオマーカーから選択される2つまたはそれより多くの便由来真核生物バイオマーカーの増加したレベルを含むと判定する工程;

30

b) 該判定に基づいて、該患者は結腸直腸新形成のリスクがあると判定する工程;および

c) 該患者に対して結腸内視鏡検査を施行する工程を含む方法。

【請求項 5 0】

便試料を準備する工程をさらに含む、請求項1~49のいずれか一項記載の方法。

【発明の詳細な説明】

【技術分野】

【0 0 0 1】

発明の分野

本発明は、便試料からの真核生物核酸の抽出ならびに腸疾患の診断および処置のための該核酸の使用に関する。

40

【背景技術】

【0 0 0 2】

背景

胃腸障害、例えば胃腸がんならびに他の消化器疾患、例えば潰瘍性大腸炎、過敏性腸症候群およびクローン病が蔓延している。米国では、胃腸障害は、毎年6000万~7000万人が罹患していると推計されている。一部の障害では、早期のスクリーニングおよび診断により死亡率の減少および患者の生活の質の改善がもたらされている。しかしながら、標準的な診断方法、例えば結腸内視鏡検査は侵襲的で時間がかかり、比較的高額な費用を伴う。

ヒトおよび動物の両方の胃腸障害を診断する非侵襲的な方法が継続的に必要とされている

50

。

【発明の概要】

【0003】

概要

本明細書において、表1もしくは表2または表1と表2の組合せに列挙されたバイオマーカーから選択される2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28または29種の便由来真核生物RNAバイオマーカーの、対象由来の便試料から抽出した真核生物核酸における発現のレベルを測定する工程；該便試料における、該2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28または29種の便由来真核生物RNAバイオマーカーの測定された発現レベルを、対照における、該2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28または29種の便由来真核生物RNAバイオマーカーの測定された発現レベルと比較する工程であって、該対照における、該2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28または29種の便由来真核生物RNAバイオマーカーの測定された発現レベルの差は、該対象が結腸直腸新形成を有することを示す、該比較する工程を含む、対象において結腸直腸新形成を検出する方法を提供する。表3に列挙されたバイオマーカー遺伝子から選択される1つまたは複数のバリエーションバイオマーカー遺伝子の、対象の由来便試料から抽出した真核生物核酸におけるバリエーションアレル頻度を測定する工程；該便試料における、該1つまたは複数のバリエーションバイオマーカー遺伝子の測定されたバリエーションアレル頻度を、対照における、該1つまたは複数のバリエーションバイオマーカー遺伝子の測定されたバリエーションアレル頻度と比較する工程であって、該対照における、該1つまたは複数のバリエーションバイオマーカー遺伝子のバリエーションアレル頻度に対する、該1つまたは複数のバリエーションバイオマーカー遺伝子のバリエーションアレル頻度の差は、該対象が結腸直腸がんを有するか、または結腸直腸がんのリスクがあることを示す、該比較する工程を含む、対象において結腸直腸新形成を検出する方法も提供する。表4に列挙された結腸直腸新生物分子サブタイプバイオマーカー遺伝子のうちの任意のものから選択される2つまたはそれより多くのバイオマーカー遺伝子の、対象由来の便試料から抽出した真核生物核酸における発現のレベルを測定する工程；生物学的試料における、該2つまたはそれより多くの結腸直腸新生物分子サブタイプバイオマーカー遺伝子の測定された発現レベルを、対照における、該2つまたはそれより多くの結腸直腸新生物分子サブタイプバイオマーカー遺伝子の測定された発現レベルと比較する工程であって、該対照における、該2つまたはそれより多くの結腸直腸新生物分子サブタイプバイオマーカー遺伝子に対する、該2つまたはそれより多くの結腸直腸新生物分子サブタイプバイオマーカー遺伝子の測定された発現レベルとの、該生物学的試料における、該2つまたはそれより多くの結腸直腸新生物分子サブタイプバイオマーカー遺伝子の測定された発現レベルの差は、結腸直腸がんの該分子サブタイプを示す、該比較する工程を含む、対象において結腸直腸がんの分子サブタイプを検出する方法も提供する。

【図面の簡単な説明】

【0004】

本発明のこれらおよび他の特長および利点は、本発明の好ましい態様の以下の詳細説明においてより充分に開示され、あるいは該詳細説明によって自明となろう。本発明の好ましい態様の以下の詳細説明は添付の図面と一体であるとみなされるべきであり、添付の図面において同様の数値は同様の部分を示し、さらに、以下の通りである。

【0005】

【図1】図1Aは電気泳動ファイルの分離図である。電気泳動解析は、文献に記載された方法に基づいて抽出したRNAの品質をチェックするために使用した。図1Bは電気泳動ファイルの分離図である。電気泳動解析は、本明細書に記載の方法に基づいて抽出したRNAの品

質をチェックするために使用した。

【図 2】図2Aは電気泳動ファイルの分離図である。電気泳動解析は、安定化バッファー中でのインキュベーションなしで即座に抽出した試料でのseRNAの品質をチェックするために使用した。図2Bは電気泳動ファイルの分離図である。電気泳動解析は、安定化バッファー中でインキュベートし、抽出前に室温で24時間保存した試料でのseRNAの品質をチェックするために使用した。図2Cは電気泳動ファイルの分離図である。電気泳動解析は、安定化バッファー中でインキュベートし、抽出前に室温で48時間保存した試料でのseRNAの品質をチェックするために使用した。

【図 3】図3Aは、SVMの内部検証の際に得られた種々の患者集団でのROC解析を示す。図3Bは、独立したテストセットにおいて使用したSVMでの予測の感度を示す。

【図 4】図4Aは、大腸癌サブタイピングコンソーシアム (Colorectal Cancer Subtyping Consortium) 分類器において使用した274個の結腸直腸新生物分子サブタイプバイオマーカー遺伝子を列挙した表である。図4Bは、結腸直腸がんサブタイプCMS1の同定に有用な25個の例示的な結腸直腸新生物分子サブタイプバイオマーカー遺伝子を列挙した表である。

【図 5】図5は、大腸癌サブタイピングコンソーシアム分類器を用いた結腸直腸がんのCMS (コンセンサス分子サブタイプ) ごとの患者の階層化をまとめたヒートマップである。

【図 6】図6は、Affymetrix Human Transcriptome Array 2.0およびIllumina Targeted RNA Custom Panelによって測定された、398個の遺伝子の転写物発現量を比較した場合の4つ生物学的反復実施ペアの相関を示す。

【図 7】図7は、結腸直腸がん、腺腫および新生物所見なしを有する13人の患者の階層的クラスタリングを示す主成分分析グラフである。

【図 8】図8は、腺腫および結腸直腸がんを診断されたヒト対象由来の便試料において同定された6つの推定体細胞バリエーションを示す。

【図 9】図9は、推定体細胞バリエーションが同定され得るであろう、がん、結腸直腸新生物および/または胃腸の健康に関係しているバイオマーカーを列挙した表である。

【図 10】図10は、前向きトレーニングセット、前向きホールドアウトテストセット、後向きホールドアウトテストセットおよび全試験コホートと関連している患者の個体群統計および加工処理メトリクスをまとめた表である。

【図 11 A】図11Aは、テストセットのブートストラップ法を用いた候補フィーチャー選択のフローチャートである。

【図 11 B】図11Bは、選択された候補フィーチャーのグラフである。

【図 12】図12は、結腸内視鏡検査で所見なし、良性ポリープ、低リスク腺腫、中リスク腺腫、高リスク腺腫および結腸直腸がんを有する患者での未加工GAPDH値のグラフである。

【図 13】図13は、内部交差検証に基づいた高リスク腺腫 (HRA) の検出のためのモデルのパフォーマンスを示すグラフである (n=154人の患者)。

【図 14】図14は、Gini重要度によるランク付けを伴うフィーチャーを示す表である。

【図 15】図15Aは、免疫化学的便検査 (FIT) フィーチャーなしでの、独立したホールドアウトテストセット (n=110人の患者) に基づいたHRAの検出のためのモデルのパフォーマンスを示すグラフである。図15Bは、免疫化学的便検査 (FIT) フィーチャーありでの、独立したホールドアウトテストセット (n=110人の患者) に基づいたHRAの検出のためのモデルのパフォーマンスを示すグラフである。

【図 16】図16Aは、免疫化学的便検査 (FIT) フィーチャーなしでの、疾患の重症度でソートしたモデル予測値を示すグラフである。図16Bは、免疫化学的便検査 (FIT) フィーチャーありでの、疾患の重症度でソートしたモデル予測値を示すグラフである。

【図 17】図17Aは、免疫化学的便検査 (FIT) フィーチャーなしでの、増分ダウンサンプリング解析の結果を示すグラフである。図17Bは、免疫化学的便検査 (FIT) フィーチャーありでの、増分ダウンサンプリング解析の結果を示すグラフである。

【図 18】図18は、11例のさらなる結腸直腸がん (CRC) 試料を含むホールドアウトテストセットのすべての試料におけるモデルのパフォーマンスを示すグラフである。

【図19】図19は、一般化スクリーニング集団に外挿した、11例のさらなる結腸直腸がん（CRC）試料を含むホールドアウトテストセットのすべての試料におけるモデルのパフォーマンスを示すグラフである。

【発明を実施するための形態】

【0006】

詳細な説明

この好ましい態様の説明は、本発明の書面による説明全体の一部とみなされるべき添付の図面との関連において読まれることを意図している。図面の図は必ずしも均一の縮尺ではなく、本発明のある特定の長を、明瞭性および簡潔性のために縮尺を誇張して、あるいはいくぶん模式的な形態で示している場合があり得る。本説明において、相対語、例えば「水平な」、「垂直な」、「上」、「下」、「上部」および「下部」ならびにその派生形（例えば、「水平に」、「下方に」、「上方に」など）は、論考している当該図面の図を説明しているときのそのとおりの向きまたは論考している当該図面の図に示されているとおりの向きを示していると解釈されたい。このような相対語は説明の都合上のものであり、通常、特定の向きを必要とすることを意図するものではない。例えば、「内側に」に対して「外側に」、「縦方向」に対して「横方向」などの用語は、適宜、互いに対するもの、あるいは長軸または回転軸もしくは回転中心に対するものと解釈されたい。結合に関するカップリングなどの用語、例えば「連結される」および「相互連結される」は、そうでないことを明白に記載していない限り、構造体が互いに直接または介在構造体を介して間接的にいずれかで固定または結合されている関係、ならびに可動性の結合もしくは関係または固定された結合もしくは関係の両方を示す。用語「機能的に連結される」は、直接関係のある構造体が、該関係のおかげで意図されたとおりに機能することが可能となるような結合、カップリングまたは連結である。1つの機械しか図示されていない場合、用語「機械」は、本明細書で論考している方法論の任意の1つまたは複数を実施するための一組（または複数の組）の指示を単独または共同で実行する任意の機械集合体も包含していると解釈されたい。特許請求の範囲において、ミーンズ・プラス・ファンクション節は、使用されている場合は、書面による説明または図面に記載された、提案された、あるいは自明となる記載の機能を果たすための構造、例えば、構造的均等物だけでなく等価構造も包含していることを意図する。

【0007】

本発明は一部において、便試料、例えば哺乳動物から採取した便試料中の細菌細胞から真核生物細胞を分離するための方法の本発明者らの開発に基づいている。結腸内には、腸内容物1グラムあたりおよそ約 1×10^{13} 個の細菌細胞が存在している。この結腸内微生物叢には300~1000種が含まれ得る。便または糞便試料は、胃腸管の腸管腔から剥がれ落ちた真核生物細胞だけでなく、微生物、例えば細菌および任意の胃腸内寄生生物、消化の悪い未吸収の食物残渣、腸細胞からの分泌物ならびに排出物、例えば粘液および色素も含む複雑な巨大分子混合物である。正常便は約75%の水分と25%の固形物で構成されている。細菌は糞便の全乾燥質量の約60%を構成している。高い細菌量は、便試料からの真核生物バイオマーカーの検出に都合の悪い信号対ノイズ比の一因となり得る。さらに、真核生物シグナルがひどく劣化し得る。かかる真核生物核酸の抽出および加工処理によって劣化が促進または加速され得、これにより、さらなる解析が大幅に制限される。

【0008】

抽出方法は、便試料からの高品質の真核生物RNAの単離を可能にする。該方法は、参照によりその全体が本明細書に組み入れられる国際出願W02018/081580に記載されている。本発明者らは、大便生成過程において保存される真核生物RNAを特定するための便由来真核生物RNA（seRNA）、および続いてどれが便試料から国際出願W02018/081580に開示されている方法によって抽出されるかに言及し得る。

【0009】

かくして、本発明者らは、ヒトの結腸直腸がんおよび結腸直腸新形成のトランスクリプトームを非侵襲的に評価するための材料および方法を開発した。本明細書に開示の材料お

10

20

30

40

50

よび方法により、ヒトの便試料における真核生物核酸の効率的で感度のよい検出がもたらされる。本発明者らは、結腸直腸新生物が、対象由来の便試料中に存在する真核生物核酸内の便由来真核生物RNAバイオマーカの発現レベルとバリエーションに基づいて検出され得ることを見出した。該検出方法は、結腸直腸がんまたは結腸直腸新形成の種々の形態およびサブタイプを検出するのに有用な様式で構成され得る。

【0010】

より具体的には、本明細書に開示の材料および方法は、対象由来の便試料中に存在する真核生物核酸内の便由来真核生物RNAバイオマーカの発現レベルに基づいて高リスク腺腫（HRA）を検出するために使用され得る。本明細書において、結腸直腸新生物、具体的には高リスク腺腫の予測または同定のためのモデルベースのアプローチを開示する。一部の態様では、モデルは、対象由来の便試料中に存在する真核生物核酸内の表1および表2に
10
列挙された2種またはそれより多くの便由来真核生物RNAバイオマーカの発現レベルに基づくものであり得る。一部の態様では、モデルは、表1もしくは表2または表1と表2の組合せに列挙された便由来真核生物RNAバイオマーカから選択される2種またはそれより多くの便由来真核生物RNAバイオマーカ、例えば2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28または29種の発現レベルに基づくものであり得る。一部の態様では、モデルは、表1に列挙された便由来真核生物RNAバイオマーカのうちの2種またはそれより多く、例えば2、3、4、5、6、7、8、9、10または11種の便由来真核生物RNAバイオマーカの発現レベルに基づくものであり得る。
20
。モデルは、個体群統計フィーチャー、例えば対象の年齢および喫煙状態も含み得る。一部の態様では、モデルは、対象由来の便試料に対して施行された免疫化学的便検査（FIT）の結果も含み得る。一部の態様では、本明細書に開示の材料および方法が、中リスク腺腫（MRA）、低リスク腺腫（LRA）または良性ポリープを同定するために使用され得る。

【0011】

対象由来の便試料中の真核生物核酸内のバリエーションバイオマーカの検出に基づいて結腸直腸がんを検出するための材料および方法も提供する。一部の態様では、バリエーションバイオマーカが結腸直腸がんの腫瘍形成と関連し得る。バリエーションは、表3に列挙された任意のバイオマーカのバリエーションであり得る。バリエーションは、結腸直腸がんのドライバー遺伝子、例えばTP53、KRAS、PIK3CA、BRAF、APC、BMP3、NDRG4、SMAD4、MLH1、CTNNB1、EGFR、BRCA1、CDKN2A、CDH1、PTEN、VEGFA、MAPK3またはNRASのバリエーションであり得る。
30

【0012】

本発明者らは、大腸癌サブタイピングコンソーシアム（CRCSC）によって規定されるようなコンセンサス分子サブタイプ（CMS）と関連している遺伝子発現シグネチャーが便由来真核生物RNAにおいて有効に検出され得ることを見出した。より具体的には、本明細書に開示の材料および方法は、CRCSCによって規定されるような、結腸直腸がんの特定のサブタイプ（例えば、CMS1）の存在を示し得るseRNAを便試料から単離するために使用することができよう。結腸直腸がんと診断された個体のうち、およそ14%がCMS1クラス分類を有する。CMS1腫瘍は、マイクロサテライト不安定性（MSI-H）、超突然変異および免疫浸潤物の増大を特徴とする。このようなフィーチャーは、免疫系が腫瘍部位の検出および監視に積極的な役割を果たしている腫瘍と整合する。かかる腫瘍を有する患者は、標的免疫療法、例えば免疫チェックポイントブロック療法の恩恵を受け得る。例えば、Keytruda（
40
商標）（ペムブロリズマブ）およびOpdivo（商標）（ニボルマブ）はどちらも、MSI-Hである切除不能または転移性の充実性腫瘍を有し、ファーストライン化学療法の恩恵がない成人および小児の患者の処置に対してFDAに承認されている。

【0013】

したがって、本明細書において、結腸直腸がんを有するヒト対象がCMS1と関連している遺伝子発現シグネチャーを有するかどうかを判定するための材料および方法を提供する。疾患モニタリングの状況において、該方法は、この患者集団を非侵襲的に選択的に同定することができ、seRNAを用いた処置ガイダンスをもたらし得る。該方法は、血液試料また
50

は生検試料ではなく便試料を用いて効率的に非侵襲的に行なわれ得る。該方法は、結腸直腸がんを有するか、または結腸直腸がんのリスクがある対象のための臨床計画および処置方法の開発において有用である。一部の態様では、該2種またはそれより多くのバイオマーカーが、図4または表4のマーカーのうちの2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、30、35、40、45、50、60、70、80、90、100、120、140、160、180種またはそれより多くの組合せを含み得る。一部の態様では、マーカーが、結腸直腸がんに関連している差次的発現転写物クラスターおよび/または共通経路内に含まれ得る。例示的な経路としては、マイクロサテライト不安定性 (MSI)、染色体不安定性 (CIN) およびCpGアイランドメチル化表現型 (CIMP) が挙げられる。一部の態様では、該経路が、細胞構成要素経路、ストレスに対する細胞応答、ストレスおよびRNA結合経路であり得る。

10

【0014】

疾患モニタリングの状況において、該方法は患者集団を非侵襲的に選択的に同定することができ、処置ガイダンスをもたらし得る。該方法は、血液試料または生検試料ではなく便試料を用いて効率的に非侵襲的に行なわれ得る。該方法は、結腸直腸新生物もしくは結腸直腸がんを有するか、または結腸直腸新生物もしくは結腸直腸がんのリスクがある対象のための臨床計画および処置方法の開発において有用である。

【0015】

本明細書に開示の方法および材料は、真核生物核酸を便試料から単離するための方法を含む。かかる真核生物核酸は、真核生物、例えば哺乳動物における胃腸の障害または疾患、例えば結腸直腸新生物または結腸直腸がんを示し得る特異的バイオマーカーのレベルについて、評価され得る。哺乳動物は、ヒトまたは非ヒト動物、例えばヒト、イヌ、ネコ、非ヒト霊長類、反芻動物、クマ類、ウマ類、ブタ、ヒツジ、ヤギ、ラクダ、バッファロー、シカ、エルク、ムース、イタチ類、ウサギ、モルモット、ハムスター、ラット、マウス、厚皮動物、サイまたはチンチラであり得る。

20

【0016】

本発明者らは、真核生物の便試料において真核生物細胞が細菌細胞から有効に分離され得ることを見出した。また、本発明者らは、かかる真核生物細胞から単離されたRNAにおいて真核生物バイオマーカーが検出され得ることも見出した。かかるバイオマーカーは、胃腸障害、例えば結腸直腸がん、セリアック病、クローン病、潰瘍性大腸炎、胃炎、胃腸炎、胃がん、胃潰瘍、壊死性腸炎、消化管間質腫瘍、消化管リンパ腫、消化管内新形成、リンパ肉腫、腺腫、過形成性変化、腺癌、炎症性腸疾患、過敏性腸症候群、膵内新形成、肝内新形成、胆管癌、大腸炎の検出に有用であり得る。本明細書において、対象、例えばヒト、イヌまたはネコに胃腸疾患、例えば結腸直腸新生物、例えば高リスク腺腫または結腸直腸がんのリスクがあるかどうかを判定するための材料および方法を提供する。また、疾患の診断のための材料および方法ならびに対象の健康状態を特定する方法も提供する。

30

【0017】

本明細書に開示の方法および組成物は、胃腸障害、例えば結腸直腸新生物または結腸直腸がんの検出、診断、クラス分類および処置に広くかつに様々に有用である。検出方法は、胃腸障害を有するか、または胃腸障害を有することが疑われる対象、例えば患者由来の試料中の1種、2種またはそれより多くのバイオマーカーの便試料中における発現レベルを測定する工程、および測定された該発現レベルを、対照における、該1種、2種またはそれより多くのバイオマーカーの測定された発現レベルと比較する工程を含み得る。対照における、該1種、2種またはそれより多くのバイオマーカーの測定された発現レベルに対する、対象の試料における、1種、2種またはそれより多くのバイオマーカーの測定された発現レベルの差は、該対象が胃腸障害を有することの表示である。一部の態様では、対照における、該1種、2種またはそれより多くのバイオマーカーの測定された発現レベルに対する、対象の試料における、1種、2種またはそれより多くのバイオマーカーの測定された発現レベルの差は、該対象、例えば患者に胃腸障害のリスクがあることの表示である。

40

【0018】

50

一部の態様では、検出方法が、胃腸障害、例えば結腸直腸新生物を有するか、または胃腸障害、例えば結腸直腸新生物を有することが疑われる対象、例えば患者由来の試料中の2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30種またはそれより多くの便由来真核生物RNAバイオマーカの便試料中における発現レベルを測定する工程、および測定された該発現レベルを、対照における、該2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30種またはそれより多くの便由来真核生物RNAバイオマーカの測定された発現レベルと比較する工程を含み得る。対照における、該2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30種またはそれより多くの便由来真核生物RNAバイオマーカの測定された発現レベルに対する、対象の試料における、該2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30種またはそれより多くの便由来真核生物RNAバイオマーカの測定された発現レベルの差は、該対象が胃腸障害、例えば結腸直腸新生物を有することの表示である。一部の態様では、対照における、該2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30種またはそれより多くの便由来真核生物RNAバイオマーカの測定された発現レベルに対する、対象の試料における、該2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30種またはそれより多くの便由来真核生物RNAバイオマーカの測定された発現レベルの差は、該対象、例えば患者に胃腸障害、例えば結腸直腸新生物のリスクがあることの表示である。一部の態様では、対照における、該2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30種またはそれより多くの便由来真核生物RNAバイオマーカの測定された発現レベルに対する、対象の試料における、該2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30種またはそれより多くの便由来真核生物RNAバイオマーカの測定された発現レベルの差は、該対象、例えば患者に特定の型の結腸直腸新形成、例えば腺腫、より具体的には高リスク腺腫のリスクがあることの表示である。任意の前記態様において、便由来真核生物RNAバイオマーカは、表1もしくは表2または表1と表2の組合せに列挙された便由来真核生物RNAバイオマーカから選択され得る。また、検出方法に、特異的バイオマーカのバリエーションの解析も含めてもよい。

【0019】

別の態様では、疾患の検出方法が、対象の便試料における、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30種またはそれより多くの便由来真核生物RNAバイオマーカの発現レベルの相対的割合、例えば相対比を測定する工程、および、このような便由来真核生物RNAバイオマーカの相対的割合を、対照における、該2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30種またはそれより多くの便由来真核生物RNAバイオマーカの発現レベルの相対的割合と比較する工程を含み得る。対照に対する、対象の試料における、該2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30種またはそれより多くの便由来真核生物RNAバイオマーカの測定された発現レベルの相対的割合の差は、該対象が胃腸疾患、例えば結腸直腸新生物を有することの表示である。一部の態様では、対照における、該2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30種またはそれより多くの便由来真核生物RNAバイオマーカの測定された発現レベルの割合に対する、対象の試料における、該2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30種またはそれより多くの便由来真核生物RNAバイオマーカの測定された発現レベルの割合の差は、該対象に胃腸障害、例えば結腸直腸新生物のリスクがあることの表示である。一部の態様では、対照における、該2、3、4、5、6、7、8、9、10

、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30種またはそれより多くの便由来真核生物RNAバイオマーカーの測定された発現レベルの割合に対する、対象の試料における、該2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30種またはそれより多くの便由来真核生物RNAバイオマーカーの測定された発現レベルの割合の差は、該対象に特定の型の結腸直腸新形成、例えば腺腫、より具体的には高リスク腺腫のリスクがあることの表示である。任意の前記態様において、便由来真核生物RNAバイオマーカーは、表1もしくは表2または表1と表2の組合せに列挙された便由来真核生物RNAバイオマーカーから選択され得る。また、検出方法に、特異的バイオマーカーのバリエーションの解析も含めてもよい。

10

【0020】

方法は、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30種またはそれより多くの便由来真核生物RNAバイオマーカーのレベルが、対照における、同じ2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30種またはそれより多くの便由来真核生物RNAバイオマーカーのレベルに対して異なるかどうかを判定することによって、対象から採取した便試料から単離された真核生物RNAにおける、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30種またはそれより多くの便由来真核生物RNAバイオマーカーの発現のレベルを明らかにする工程を含み得る。例示的な便由来真核生物RNAバイオマーカーを表1および表2に列挙する。例示的な便由来真核生物RNAバイオマーカーとしては、ACY1、TNFRSF10B、DST、EGLN2、PER3、CTNNB1、ACHE、SMAD4、EDN1、ERBB2、GAPDH、ABCB1、MAPK3、VEZF1、KRAS、PTEN、CREBBP、SUZ12、CDHR5、CABLES1、AREG、SPATA2、PPARGC1A、DBP、CDH1、PDGFA、OGG1、CGNおよびTCF7L2が挙げられ得る。

20

【0021】

(表1) 便由来真核生物RNAバイオマーカー

バイオマーカー	例示的なGenbankエントリー
ACY1	NM_000666.3
TNFRSF10B	NM_003842.5
DST	XM_011514826.3
EGLN2	NM_080732.4
PER3	XM_024450585.1
CTNNB1	NM_001904.4
ACHE	KJ425573.1
SMAD4	NM_005359.5
EDN1	NM_001955.5
ERBB2	XM_024450643.1
GAPDH	NM_002046.7

30

【0022】

(表2) 便由来真核生物RNAバイオマーカー

40

バイオマーカー	例示的なGenbankエントリー
ABCB1	AF399931.1
MAPK3	BC013992.1
VEZF1	NM_007146.3
KRAS	M54968.1
PTEN	KX398936.1
CREBBP	U85962.3
SUZ12	NM_015355.4
CDHR5	NM_021924.4
CABLES1	EF028204.1
AREG	NM_001657.4
SPATA2	BC009481.2
PPARGC1A	HQ695733.1
DBP	NM_001352.4
CDH1	NM_004360.5
PDGFA	M22734.1
OGG1	AF003595.1
CGN	NM_020770.3
TCF7L2	CR536574.1

10

【0023】

20

一部の態様では、便由来真核生物RNAバイオマーカーが、表1および表2に列挙された便由来真核生物RNAバイオマーカーのサブセットもまた含み得る。表1もしくは表2または表1と表2の組合せに列挙された便由来真核生物RNAバイオマーカーのうちの一部または全部がパネルを構成し得る。例えば、表1の便由来真核生物RNAバイオマーカーのうちの一部または全部がパネル（パネルA）を構成し得る。例えば、パネルAには便由来真核生物RNAバイオマーカーACY1、TNFRSF10B、DST、EGLN2、PER3、CTNNB1、ACHE、SMAD4、EDN1、ERBB2、GAPDHのうちの一部または全部が含まれ得る。組成物には、本明細書に開示のマーカーパネルの特異的検出用に構成された遺伝子アレイとプローブセットが含まれ得る。また、組成物が、本明細書に開示のマーカーパネルの特異的検出用に構成された遺伝子アレイとプローブセットを備えたキットを含むものであってもよい。該方法は、便由来真核生物RNAバイオマーカー、例えば表1および表2に列挙された便由来真核生物RNAバイオマーカーの核酸配列におけるバリエーションの存在を同定する工程を含み得る。

30

【0024】

また、真核生物核酸（例えば、seRNA）におけるバイオマーカーの核酸配列におけるバリエーションの検出方法を提供する。バリエーションは、腫瘍の生存、腫瘍の進行または腫瘍の転移に寄与する任意の突然変異であり得る。本発明者らは、かかる突然変異を「ドライバー突然変異」または「プログレッサー突然変異」と称する場合があります。そのような突然変異としては、サイレント突然変異、ミスセンス突然変異、挿入、欠失、フレームシフト突然変異またはナンセンス突然変異が挙げられ得る。また、任意の特定のバリエーションの発現を「バリエーションアレル頻度」（VAF）と記載している場合があります。そのようなバリエーションとしては、図8または9に列挙されたバイオマーカーのうちのいずれかのバリエーションが挙げられ得る。バリエーションは、結腸直腸がんのドライバー遺伝子、例えばTP53、KRAS、PIK3CA、BRAF、APC、BMP3、NDRG4、SMAD4、MLH1、CTNNB1、EGFR、BRCA1、CDKN2A、CDH1、PTEN、VEGFA、MAKP3またはNRASのバリエーションであり得る。例示的な便由来真核生物RNAバリエーションバイオマーカーを表3に列挙する。

40

【0025】

（表3）便由来真核生物RNAバイオマーカーおよびバリエーション

バイオマーカー	例示的なGenbankエントリー	例示的なバリエーション
APC	M74088.1	chr5:112175639 C>T p.R1450* NM_000038 c.C4348T
KRAS	M54968.1	chr12:25398284 C>T p.G12D NM_033360 c.G35A
TP53	KX710182.1	chr17:7577538 C>T p.R248Q NM_000546 c.G743A
BMP3	NM_001201.4	chr4:81967371 C>T p.P266S NM_001201 c.C796T
NDRG4	BC011795.2	chr16:58538324 G>A p.G136R NM_020465 c.G406A
SMAD4	NM_005359.5	chr18:48591919 G>A p.R361H NM_005359 c.G1082A
MLH1	U07343.1	chr3:37067240 T>A p.V384D NM_000249 c.T1151A
CTNNB1	NM_001904.4	chr3: 41266137 C>T p.S45S NM_001904 c.C134T
EGFR	NM_005228.5	chr7: 55228007 p.S492R NM_005228 c.A1474C

10

20

BRCA1	U14680.1	chr17:41243770 A>C p.L1260V NM_007294 c.T3778G
CDKN2A	JQ694045.1	chr9:2197112 G>A p.R80* NM_000077 c.C238T
CDH1	NM_004360.5	chr16:68849598 G>A p.V501M NM_004360 c.G1501A
PIK3CA	NM_006218.4	chr3:178936091 G>A p.E545K NM_006218 c.G1633A
PTEN	KX398936.1	chr10:89692905 G>A p.R130Q NM_000314 c.G389A
VEGFA	NM_001171623.1	chr6:43745357 G>T p.E90D NM_001171623 c.G270T
BRAF	M95712.2	chr7:140453136 A>T p.V600E NM_004333 c.T1799A
MAPK3	BC013992.1	chr16:30128482 G>C p.D300E NM_002746 c.C900G
NRAS	AF493919.1	chr1:115256530 G>T p.Q61K NM_002524 c.C181A

10

20

【0026】

方法は、バイオマーカー、例えば表3に列挙されたバイオマーカーの核酸配列におけるバリエーションの存在を同定する工程を含み得る。表3に列挙された結腸直腸新生物バイオマーカー遺伝子のうちの一部または全部がパネル（パネルB）を構成し得る。一部の態様では、表3に列挙された結腸直腸新生物バイオマーカー遺伝子が、結腸直腸新生物サブタイプバイオマーカーのサブセットも含み得る。組成物には、本明細書に開示のマーカーパネルの特異的検出用に構成された遺伝子アレイとプローブセットが含まれ得る。また、組成物が、本明細書に開示のマーカーパネルの特異的検出用に構成された遺伝子アレイとプローブセットを備えたキットを含むものであってもよい。

30

【0027】

別の態様では、疾患の検出方法が、対象の便試料における、表3に列挙された任意のバイオマーカー遺伝子のうちの1種、2種またはそれより多くにおけるバリエーションの相対バリエーションアレル頻度、例えば相対比を測定する工程、およびこのようなバイオマーカーの該相対バリエーションアレル頻度を、対照における、該1種、2種またはそれより多くのバイオマーカーのバリエーションアレル頻度と比較する工程を含み得る。対照における測定されたバリエーションアレル頻度に対する、対象の試料における、1種、2種またはそれより多くのバイオマーカーの測定されたバリエーションアレル頻度の差は、該対象が胃腸疾患を有することの表示である。一部の態様では、対照における、該1種、2種またはそれより多くのバイオマーカーの測定されたバリエーションアレル頻度に対する、対象の試料における、該1種、2種またはそれより多くのバイオマーカーのバリエーションアレル頻度の差は、該対象に胃腸障害のリスクがあることの表示である。

40

【0028】

50

結腸直腸新生物分子サブタイプバイオマーカーの検出方法も提供する。結腸直腸がんは、特定のマーカーの発現に基づいて4つの異なる分子サブタイプにクラス分類され得る。この4つのコンセンサス分子サブタイプ（CMS1～4）は、図4Aに示す274個の遺伝子（固有のHUGO遺伝子名識別子に基づく）の発現に基づいて予測される。CRCSCによって示されたランダムフォレスト分類器では、この274個の遺伝子の発現が分子サブタイプクラス分類を正確に同定するためのフィーチャーとして使用される。4つのCMSサブタイプにはCMS1～4が含まれる。CMS1は超突然変異およびマイクロサテライト不安定性と関連している。CMS1腫瘍は典型的には免疫浸潤物を有する。CMS1腫瘍は、診断時に組織病理学的に高悪性度を有する傾向にあり、低生存率と関連している。CMS2は、「標準型」サブタイプとも称され、著しいWNTおよびMYCのシグナル伝達の活性化ならびにコピー数変化体の増加を特徴とする上皮性腫瘍であり、長期生存と関連している傾向にある。CMS3は、明白な代謝調節異常ならびにKRAS、受容体型チロシンキナーゼおよびMAPK経路における突然変異を特徴とする上皮性腫瘍である。CMS4腫瘍は、トランスフォーミング増殖因子- β の活性化、間質浸潤および血管新生を特徴とする間葉系腫瘍である。CMS4腫瘍は、進行期ステージ（ステージIIIおよびIV）と診断される傾向にあり、低い全生存率および低い無再発生存率と関連している。CMS1の予測において特に有力な25個の遺伝子（固有のHUGO遺伝子名識別子に基づく）を図4Bおよび表4に示す。

【0029】

（表4）CMS1結腸直腸がんサブタイプの便由来真核生物RNAバイオマーカー

バイオマーカー	例示的なGenbankエントリー
QPRT	BC010033.2
RNF43	BC109028.2
TFAP2A	BC017754.1
TSPAN6	BC012389.1
TRIM7	AF220032.1
GNLY	BC023576.2
AXIN2	AF205888.1
FITM2	NM_001080472.4
GNG4	AF493872.1
VAV3	AF067817.1
RETNLB	NM_032579.2
DUSP4	BC002671.2
TNFAIP6	NM_007115.4
HOXC6	CR456954.1
TRIB2	NM_021643.3
CEL	NM_001807.5
GPR143	NM_000273.3
ASCL2	NM_005170.2
SLC5A6	BC015631.2
GAS1	NM_002048.3
B3GNT6	NM_138706.5
CYP2B6	AF182277.1
BCAT1	NM_005504.7
FAP	NM_004460.5
BOC	AY358328.1

【0030】

方法は、対象由来の便試料における、2つまたはそれより多くの結腸直腸新生物サブタイプバイオマーカー遺伝子のレベルが、対照における、同じ2つまたはそれより多くの結腸直腸新生物サブタイプバイオマーカー遺伝子のレベルに対して異なるかどうかを判定することによって、対象から採取した便試料から単離されたヒトRNAにおける2種またはそれ

より多くの結腸直腸新生物サブタイプバイオマーカーの発現のレベルを明らかにする工程を含み得る。例示的な結腸直腸新生物サブタイプバイオマーカー遺伝子を表4に示す。表4に列挙された結腸直腸新生物バイオマーカー遺伝子のうちの一部または全部がパネル（パネルC）を構成し得る。一部の態様では、表4に列挙された結腸直腸新生物バイオマーカー遺伝子が、結腸直腸新生物サブタイプバイオマーカーのサブセットも含み得る。組成物には、本明細書に開示のマーカーパネルの特異的検出用に構成された遺伝子アレイとプローブセットが含まれ得る。また、組成物が、本明細書に開示のマーカーパネルの特異的検出用に構成された遺伝子アレイとプローブセットを備えたキットを含むものであってもよい。

【0031】

別の態様では、疾患の検出方法が、対象の便試料における1種、2種またはそれより多く2種またはそれより多くの結腸直腸新生物サブタイプバイオマーカーの発現レベルの相対的割合、例えば相対比を測定する工程、およびこのようなバイオマーカーの該相対的割合を、対照における、該1種、2種またはそれより多くのバイオマーカーの発現レベルの相対的割合と比較する工程を含み得る。対照に対する、対象の試料における、1種、2種またはそれより多くのバイオマーカーの測定された発現レベルの相対的割合の差は、結腸直腸がんの分子サブタイプを示し得る。一部の態様では、対照における、該1種、2種またはそれより多くのバイオマーカーの測定された発現レベルの割合に対する、対象の試料における、該1種、2種またはそれより多くのバイオマーカーの測定された発現レベルの割合の差は、該対象が結腸直腸がんの特定のサブタイプを発症するかもしれないことの表示である。

【0032】

MSI-H腫瘍とも称されるCMS1腫瘍を検出するための代替的方法を使用してもよい。DNAミスマッチ修復欠損に関与しているPOLE、MLH1、MSH2、MSH6およびPMS2のゲノムバリエーションは、免疫チェックポイントブロック療法の臨床試験において予測バイオマーカーとして使用されている。免疫阻害分子、例えばPD-1、PD-L1、CTLA-4、LAG-3およびIDOの発現に的を絞った遺伝子発現プロファイルが、MSI-H腫瘍の微小環境の免疫原性の増大を予測してチェックポイント免疫療法の恩恵を受ける患者の適格性をさらに予測するためにさらに使用され得る。

【0033】

本明細書において、結腸直腸新生物または特定のサブタイプの前がん性病変または結腸直腸がんの診断における使用のための、便由来真核生物RNAバイオマーカーと便由来真核生物RNAバイオマーカーパネルとを提供する。バイオマーカーは一般的に、生物学的プロセス、例えば結腸直腸新生物の発生、進行、寛解または再発を評価するために客観的に測定され、定量され、使用され得る特徴である。バイオマーカーは、核酸、ポリペプチド、代謝産物または物理的もしくは生理学的パラメータなどの多くの形態をとり得る。

【0034】

一般に、真核生物細胞由来のバイオマーカーとしては、以下が挙げられ得る：a) デオキシリボ核酸（DNA）の配列、b) リボ核酸（RNA）の配列、c) タンパク質の骨格を構成する予測アミノ酸配列、d) リボ核酸バイオマーカーの発現レベル、e) アミノ酸配列の予測発現レベル、またはf) 上記の任意の組合せ。一部の態様では、バイオマーカーが、大型配列の断片、例えば長鎖RNA配列、長鎖DNA配列または長鎖ポリペプチド配列の断片であり得る。一部の態様では、バイオマーカー、例えばGAPDH、ACTBなどが他のバイオマーカーの正規化のために使用され得る。他の態様では、例えば、全RNAカウント数、全RNA投入量などのフィーチャーがバイオマーカーとして、または他のバイオマーカーの正規化のために使用され得る。

【0035】

便由来真核生物RNAバイオマーカーはアンプリコンを用いて定量され得る。アンプリコンは、ゼロ、1つ、2つ、またはそれより多くのユニーク配列を含み得る。同じ便由来真核生物RNAバイオマーカーのアンプリコンは配列同一性パーセントが異なり得る。アンプリコンは、異なる遺伝子座が標的化されるように設計され得る。標的化対象の遺伝子座とし

ては、以下が挙げられ得る:a) 同じ遺伝子からの同じ転写物上の地理学的に類似している遺伝子座、b) 同じ遺伝子からの同じ転写物上の地理学的にユニークな遺伝子座、c) 同じ遺伝子からの異なる転写物上の地理学的にユニークな遺伝子座、またはd) 異なる遺伝子からの異なる転写物上の地理学的にユニークな遺伝子座。一部の態様では、便中で保護されるかもしれないかまたは優先的に分解されるかもしれない、特定のRNAの構造上の特色、例えば配列または二次構造を、異なる遺伝子座を標的とするように設計されるアンプリコンは反映し得る。一部の態様では、異なる遺伝子座を標的とするように設計されるアンプリコンが、例えば特定の選択的スプライシング転写物が増加または減少する疾患の特異的疾患パラメータを反映し得る。

【0036】

生物学的試料は、核酸または他の分析物が採取され得る細胞または他の細胞物質を含有している試料であり得る。生物学的試料は対照であっても実験試料であってもよい。生物学的試料は便試料であり得る。生物学的試料は、トイレ内、地面、リターボックス内または収集デバイス内への排便直後に採取され得る。一部の態様では、生物学的試料が、浣腸、糞便スワブまたは内視鏡検査などの手順後または手順中に採取され得る。生物学的試料は即座に試験され得る。あるいはまた、生物学的試料を、試験する前にバッファー中に、例えば水性バッファー、グリセロールベースのバッファー、極性溶媒ベースのバッファー、浸透圧バランスバッファー、または生物学的試料を保存するのに十分な他のバッファー中に保存してもよい。付加的または択一的に、生物学的試料を収集し、試験する前に例えば4℃の冷蔵状態で、または例えば0℃、-20℃、-80℃、-140℃もしくはそれより下の凍結状態で保存してもよい。生物学的試料は、試験する前に1ヶ月間、2ヶ月間、4ヶ月間、6ヶ月間、1年間、2年間またはそれより長く保存され得る。

【0037】

生物学的試料は、真核生物、例えば哺乳動物に由来し得る。哺乳動物は、ヒトまたは非ヒト動物、例えばヒト、イヌ、ネコ、非ヒト霊長類、反芻動物、クマ類、ウマ類、ブタ、ヒツジ、ヤギ、ラクダ、バッファロー、シカ、エルク、ムース、イタチ類、ウサギ、モルモット、ハムスター、ラット、マウス、厚皮動物、サイまたはチンチラであり得る。したがって、便試料は、ヒトまたは非ヒト動物、例えばヒト、イヌ、ネコ、非ヒト霊長類、反芻動物、クマ類、ウマ類、ブタ、ヒツジ、ヤギ、ラクダ、バッファロー、シカ、エルク、ムース、イタチ類、ウサギ、モルモット、ハムスター、ラット、マウス、厚皮動物、サイまたはチンチラから採取され得る。

【0038】

生物学的試料、例えば便試料由来の、真核生物核酸が濃縮された核酸の単離のための有用な方法を本明細書において提供する。該方法は、便試料をバッファーで崩壊させる工程を含み得る。試料は、ボルテックス、振盪、攪拌、回転または固形物および便中細菌を分散させるのに十分な他のアジテーション法に供され得る。アジテーションおよび遠心分離の工程が行なわれる温度はさまざまであることができ、例えば約4℃～約20℃、約4℃～約1℃、約4℃～約10℃、約4℃～約6℃であることができる。崩壊後、試料は1回または複数回の遠心分離に供され得る。一部の態様では、崩壊工程と遠心分離工程を1回、2回、3回またはさらに多くの回数繰り返してもよい。市販の試薬、例えばNuclisens（登録商標）EasyMag（登録商標）試薬は、便の崩壊、洗浄および細胞溶解のために使用され得る。また、溶解バッファーは真核生物細胞を溶解させるためにも使用され得る。ライセートを任意の温度、任意の期間、任意の回数でさらに遠心分離してもよい。遠心分離後、上清みは、自動RNA単離機、例えばEasyMag（登録商標）機器内への投入物として使用され得る。一部の態様では、抽出された核酸は、溶液中でDNAを分解するためにDNaseで処理され得る。他のRNA精製方法を使用してもよい;例えば、機械的または酵素的細胞破壊後、固相法、例えばカラムクロマトグラフィーまたは有機溶媒、例えばフェノール-クロロホルムでの抽出もしくはチオシアネート-フェノール-クロロホルム抽出が行なわれ得る。一部の態様では、核酸が官能基化ビーズ上に抽出され得る。一部の態様では、官能基化ビーズは磁気コアをさらに含み得る（「磁気ビーズ」）。一部の態様では、官能基化ビーズが荷電部分で官

10

20

30

40

50

能基化された表面を含み得る。荷電部分は、アミン、カルボン酸、カルボキシレート、第4級アミン、スルフェート、スルホネートまたはホスフェートから選択され得る。

【0039】

核酸の抽出のため、便試料を、バッファー、界面活性剤およびリボヌクレアーゼ阻害薬のうちの1種または複数種の存在下で崩壊させて、懸濁液を形成させることができる。バッファーは、生物学的に適合性のバッファー、例えばハンクス平衡塩溶液、アルセバー液、アール平衡塩溶液、ゲイ平衡塩溶液、リン酸緩衝生理食塩水、バック（Puck）平衡塩溶液、リングル平衡塩溶液、ジム（Simm）平衡塩溶液、TRIS緩衝生理食塩水またはタイロード平衡塩溶液であり得る。界面活性剤は、イオン系または非イオン系界面活性剤、例えばTween-20またはTriton-X-100であり得る。リボヌクレアーゼ阻害薬は、溶媒ベース、タンパク質ベース、またはRNA破壊を抑制するための別の型の方法、例えばProtector RNase Inhibitor（Roche）、RNasin（登録商標）（Promega）、SUPERase-In（商標）（Thermo Fisher Scientific）、RNaseOUT（商標）（Thermo Fisher Scientific）、ANTI-RNase、Recombinant RNase Inhibitorまたはクローン化RNase Inhibitorなどであり得る。便試料を、さまざまな様式で、例えばボルテックス、振盪、攪拌、回転または固形物および便中細菌を分散させるのに十分な他のアジテーション法によって崩壊させることができる。一部の態様では、便試料を、被覆ビーズ、磁気ビーズまたは攪拌器具、例えばガラス棒、金属棒、木製棒もしくは木製ブレードを用いて崩壊させることができる。

10

【0040】

懸濁液は次いで、液体部分と固形物部分に分離され得る。分離は、例えば遠心分離、濾過、真核生物細胞に特異的に結合するターゲティングプローブ、抗体、カラムベースの濾過、ビーズベースの濾過またはクロマトグラフィー法によって行なわれ得る。液体部分は細菌核酸が濃縮され、廃棄され得る。固形物部分は、界面活性剤の存在下または非存在下およびリボヌクレアーゼの存在下または非存在下のいずれかでバッファー中に再懸濁され得る。分離工程を1、2、3、4、5、6、7、8回またはさらに多くの回数繰り返してもよい。

20

【0041】

崩壊および分離の工程が行なわれる温度はさまざまであることができ、例えば約4℃～約20℃、約4℃～約15℃、約4℃～約10℃、約4℃～約6℃であることができる。

【0042】

分離工程で得られた結果物のペレットは溶解バッファー中に、例えば、カオトロピック剤および任意で界面活性剤を含むバッファー中に懸濁され、ライセートが形成され得る。一部の態様では、カオトロピック剤がグアニジウムチオシアネートであり得、界面活性剤がTriton-X-100であり得る。一部の態様では、溶解バッファーはTris-HCl、エチレンジアミン四酢酸（EDTA）、ドデシル硫酸ナトリウム（SDS）、Nonidet P-40、デオキシコール酸ナトリウムもしくはジチオトレイトールを含み得るか、またはこれらを含まないものであり得る。

30

【0043】

ライセートは、真核生物核酸が濃縮された部分に分画され得る。分画は、例えば遠心分離、濾過、真核生物核酸に特異的に結合するターゲティングプローブ、抗体、カラムベースの濾過、ビーズベースの濾過またはクロマトグラフィー法によって行なわれ得る。一部の態様では、遠心分離による分画により、細胞デブリを含む底層（ペレット）、真核生物核酸を含む親水性の中間層、および脂質と膜画分を含む疎水性の上層の形成がもたらされ得る。中間層は収集され得る。一部の態様では、中間層と上層が一緒に収集され得る。中間層は狭口径オリフィスを通して収集され得る。狭口径オリフィスは、ピペットチップ、または針を取り付けたシリンジであり得る。ピペットチップは、例えば1 uL、5 uL、10 uL、20 uLまたは100 uL容ピペットチップであり得る。針は、例えば18-ゲージまたは15-ゲージの針であり得る。

40

【0044】

収集された真核生物核酸を含む層はさらなる抽出に供され得る。さらなる抽出の方法はさまざまであり得る。例示的な方法としては、磁気粒子ベースの方法、カラムベースの方

50

法、フィルターベースの方法、ビーズベースの方法または有機溶媒ベースの方法が挙げられる。このような例示的な方法は、市販の試薬、例えばNuclisens（登録商標）EasyMag（登録商標）試薬（bioMerieux）を含み得る。

【0045】

抽出された核酸は、胃腸障害または胃腸細胞に関連性のある真核生物バイオマーカーについて解析され得る。バイオマーカーは、個体、すなわち対象の健康に関する情報をもたらし得る。真核生物細胞由来のこのようなバイオマーカーとしては、以下挙げられ得る：a) デオキシリボ核酸（DNA）の配列、b) リボ核酸（RNA）の配列、c) タンパク質の骨格を構成する予測アミノ酸配列、d) RNAバイオマーカーの発現レベルもしくは発現レベルの割合、e) アミノ酸配列の予測発現レベルもしくは予測発現レベルの割合、またはf) 上記の任意の組合せが。真核生物細胞由来のバイオマーカーの単離により実験試料と対照間の比較が可能となり得る。真核生物細胞由来のこのようなバイオマーカーの単離により、実験試料での腸疾患の検出のための方法がもたらされ得る。比較は、以下についての評価を含む：a) DNA配列の変動、b) RNA配列の変動、c) 予測アミノ酸配列の変動、d) RNAバイオマーカーの発現レベルの変動もしくは発現レベルの割合の変動、e) アミノ酸配列の予測発現レベルの変動もしくは予測発現レベルの割合の変動、またはf) 上記の任意の組合せを構成する変動。変動は、実験試料で測定されたバイオマーカーが対照で測定されたバイオマーカーと異なる場合に決定され得る。

【0046】

方法は、実験試料および対照、例えば便試料を採取する工程を含み得る。便試料には剥離した真核生物細胞が含まれており、これがバイオマーカーについて評価され得る。一部の態様では、真核生物細胞は、腸細胞、リンパ球、エンテロクロミフィン様細胞、腸内分泌細胞、神経内分泌細胞、膵細胞、肝細胞、胃の細胞または他の細胞であり得る。該方法により、便試料中の真核生物細胞が真核生物バイオマーカーについて評価され得る手段がもたらされる。バイオマーカーとしては、DNAの配列、RNAの配列、予測アミノ酸配列、RNAバイオマーカーの発現レベルもしくは発現レベルの割合、アミノ酸配列の予測発現レベルもしくは予測発現レベルの割合、または上記の任意の組合せが挙げられ得る。具体的な態様では、バイオマーカーは便由来真核生物RNAバイオマーカーである。一部の態様では、評価工程は、任意の型のマイクロアレイシーケンシング、ポリメラーゼ連鎖反応（PCR）、核酸シーケンシング、アンプリコンシーケンシング、分子バーコーディングまたはプローブ-キャプチャーを含む。

【0047】

また、該方法および組成物は、胃腸障害、例えば結腸直腸新生物または結腸直腸がんに苦しんでいる個体に対する臨床計画を選択するのに有用である。この方法を通して、臨床計画には、さらなる診断手順、例えば結腸内視鏡検査の施行が含まれ得る。一部の態様では、臨床計画に処置方法が含まれ得る。

【0048】

便由来真核生物RNAバイオマーカーのレベルは、さまざまな方法を用いて評価され得る。発現レベルは、核酸レベル、例えばRNAレベルまたはポリペプチドレベルのいずれかで測定され得る。RNA発現には、seRNA、全RNA、mRNA、tRNA、rRNA、ncRNA、smRNA、miRNAおよびsnoRNAの発現が包含され得る。RNAレベルでの発現は直接測定してもよく、関連性のあるRNAに対応するcDNAのレベルを測定することによって間接的に測定してもよい。択一的または付加的に、RNAにコードされているポリペプチド、関連性のある転写因子をコードしている遺伝子のRNA調節因子、および転写因子ポリペプチドのレベルをアッセイしてもよい。遺伝子発現をmRNAレベルで調べるための方法としては、例えばマイクロアレイ解析、遺伝子発現連鎖解析（SAGE）、RT-PCR、プロッティング、デジタルバーコード定量アッセイに基づいたハイブリダイゼーション、マルチプレックスRT-PCR、ドロップレットデジタルPCR（ddPCR）、デジタルPCR（dPCR）、NanoDrop分光測光器、RT-qPCR、qPCR、UV分光法、アンプリコンシーケンシング、RNAシーケンシング、次世代シーケンシング、分岐DNAシグナル増幅を用いるライセートベースのハイブリダイゼーションアッセイ、例えばQu

antiGene 2.0 Single Plex、および分岐DNA解析法が挙げられる。デジタルバーコード定量アッセイとしては、BeadArray (Illumina)、xMAPシステム (Luminex)、nCounter (NanoString)、HTG EdgeSe (High Throughput Genomics)、BioMark (Fluidigm) またはWaf ergenマイクロアレイが挙げられ得る。アッセイとしては、DASL (Illumina)、RNA-Seq (Illumina)、TruSeq (Illumina)、SureSelect (Agilent)、Bioanalyzer (Agilent)、T aqMan (ThermoFisher)、GeneReader (Qiagen) またはQIAseq (Qiagen) が挙げられ得る。

【0049】

本発明者らは、RNAおよびDNAの両方、例えばcDNA、ゲノムDNA、合成DNA、および核酸アナログを含むDNA (またはRNA) を示すために、用語「核酸」および「ポリヌクレオチド」を互換的に用いている場合があり得、これらはいずれも本発明のポリペプチドをコードしているものであり得、これらはすべて、本発明に包含される。ポリヌクレオチドは本質的に任意の3次元構造を有し得る。核酸は二本鎖であっても一本鎖 (すなわち、センス鎖またはアンチセンス鎖) であってもよい。ポリヌクレオチドの非限定的な例としては、遺伝子、遺伝子断片、エキソン、イントロン、メッセンジャーRNA (mRNA) およびその一部分、トランスファーRNA、マイクロRNA、リボソームRNA、siRNA、マイクロ-RNA、リボザイム、cDNA、組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、任意の配列の単離されたDNA、任意の配列の単離されたRNA、核酸プローブおよびプライマーならびに核酸アナログが挙げられる。本発明との関連において、核酸は、バイオマーカー、例えば、表1および表2に列挙された、もしくは表3に列挙された、もしくは表4に列挙されたバイオマーカーのうちの任意のものからの便由来真核生物RNAバイオマーカーまたはそのバリエーションの断片をコードし得る。

【0050】

「単離された」核酸は例えばDNA分子またはその断片であり得る、ただし、ゲノム内で該DNA分子に直接フランキングしている通常みられる核酸配列のうち少なくとも1つが除去されているか、または存在していないものとする。したがって、単離された核酸としては、非限定的に、他の配列とは独立した別個の分子として存在するDNA分子 (例えば、化学合成された核酸、またはポリメラーゼ連鎖反応 (PCR) もしくは制限エンドヌクレアーゼ処理によって作製されたcDNAもしくはゲノムDNA断片) が挙げられる。単離された核酸は、ベクター、自律複製プラスミド、ウイルス内に、または原核生物もしくは真核生物のゲノムDNA内に組み込まれるDNA分子もまた示す。また、単離された核酸としては、改変操作された核酸、例えばハイブリッドまたは融合核酸の一部であるDNA分子も挙げられ得る。例えばcDNAライブラリーもしくはゲノムライブラリー内またはゲノムDNA制限消化物を含むしているゲル切片中の多く (例えば、数ダース、または数百~数百万個) の他の核酸の中に存在している核酸は単離された核酸ではない。

【0051】

単離された核酸分子を、さまざまな様式で作製してもよい。例えば、ポリメラーゼ連鎖反応 (PCR) 手法が、本明細書に記載のヌクレオチド配列、例えば本明細書に記載のポリペプチドをコードしているヌクレオチド配列を含む単離された核酸を得るために使用され得る。PCRは、DNAならびにRNA由来の特定の配列、例えば全ゲノムDNAまたは全細胞RNA由来の配列を増幅させるために使用され得る。一般的に、関心対象の領域の両末端またはそれ以降からの配列情報を使用し、増幅対象の鋳型と反対の鎖と配列が同一または類似のオリゴヌクレオチドプライマーが設計される。また、部位特異的ヌクレオチド配列修飾が鋳型核酸に導入され得る種々のPCRストラテジーも利用可能である。

【0052】

また、単離された核酸は、単一の核酸分子 (例えば、ホスホロアミダイト技術を用いた3'から5'の方向への自動DNA合成を使用) として、または一連のオリゴヌクレオチドとしてのいずれかで化学合成され得る。例えば、所望の配列を含む1つまたは複数のペアの長鎖オリゴヌクレオチド (例えば、>50~100個のヌクレオチド) が合成され得、各ペアは、該オリゴヌクレオチドペアがアニーリングされる際に二本鎖が形成されるような相補性

の短鎖セグメント（例えば、約15個のヌクレオチド）を含む。DNAポリメラーゼは該オリゴヌクレオチドを伸長させるために使用され、オリゴヌクレオチドペア1つあたり1つの二本鎖核酸分子が生じ、次いでこれはベクター内にライゲートされ得る。

【0053】

2つの核酸またはこれらがコードしているポリペプチドが、互いにある特定度合の同一性を有すると記載している場合があり得る。例えば、表1もしくは表2または表1と表2の組合せから選択される、あるいは表3または表4の便由来真核生物RNAバイオマーカーと、その生物学的に活性なバリエーションは、ある特定度合の同一性を示すと記載され得る。アラインメントは、短い配列をProtein 情報 Research (PIR) 部位に配置し (<http://pir.georgetown.edu>)、続いてNCBIウェブサイト (<http://www.ncbi.nlm.nih.gov/blast>) の “short nearly identical sequences” Basic Local Alignment Search Tool (BLAST) アルゴリズムを用いて解析することによりアセンブリングされ得る。

10

【0054】

本明細書で用いる場合、用語「配列同一性パーセント」は、任意の所与のクエリー配列と対象配列間の同一性の度合を示す。例えば、表1もしくは表2または表1と表2の組合せあるいは表3または表4に列挙された便由来真核生物RNAバイオマーカー配列がクエリー配列であり得、表1もしくは表2または表1と表2の組合せあるいは表3または表4に列挙された便由来真核生物RNAバイオマーカー配列の断片が対象配列であり得る。同様に、表1もしくは表2または表1と表2の組合せあるいは表3または表4に列挙された便由来真核生物RNAバイオマーカー配列の断片がクエリー配列であり得、その生物学的に活性なバリエーションが対象配列であり得る。

20

【0055】

配列同一性を調べるため、クエリー核酸配列またはクエリーアミノ酸配列は、それぞれ1つまたは複数の対象核酸配列または対象アミノ酸配列と、コンピュータプログラム、例えばClustalW（バージョン1.83、デフォルトパラメータ）、HISAT、HISAT2またはSAMToolsを用いてアラインメントされ得、これらは、核酸配列またはタンパク質配列のアラインメントを全長にわたって行なうことが可能である（グローバルアラインメント）。

【0056】

本明細書に記載の核酸およびポリペプチドを「外来性」と称している場合があり得る。用語「外来性」とは、核酸またはポリペプチドが組換え核酸構築物の一部であるか、または組換え核酸構築物にコードされているか、あるいはその天然の環境において存在しないことを示す。例えば、外来性核酸は、ある種に由来する配列が別の種に導入されたもの、すなわち異種核酸であり得る。典型的には、かかる外来性核酸はその他の種に組換え核酸構築物によって導入される。また、生物体にとって生来のものであって、該生物体の細胞内に再導入された配列も外来性核酸であり得る。生来の配列を含む外来性核酸は、多くの場合、該生来の配列と、該外来性核酸に連結された非天然配列の存在、例えば組換え核酸構築物内の生来の配列にフランキングしている生来のものでない調節配列の存在によって区別され得る。また、安定的に形質転換された外来性核酸は典型的には、生来の配列がみられる位置以外の位置に組み込まれる。

30

【0057】

本発明の核酸は、表1もしくは表2または表1と表2の組合せあるいは表3または表4に列挙された便由来真核生物RNAバイオマーカーのうちのいずれか1種のヌクレオチド配列を有する核酸、または表1もしくは表2または表1と表2の組合せあるいは表3または表4に列挙された便由来真核生物RNAバイオマーカーのうちのいずれか1つの核酸配列と少なくとも約70%、少なくとも約75%、少なくとも約80%、少なくとも約85%、少なくとも約90%、少なくとも約95%、少なくとも約99%同一である核酸配列を有する核酸を含み得る。

40

【0058】

標的核酸に特異的な核酸、例えばオリゴヌクレオチド（例えば、プローブまたはプライマー）は該標的核酸に適当な条件下でハイブリダイズする。本発明者らは、単一のオリゴヌクレオチド鎖が相補鎖と既定のハイブリダイゼーション条件下での塩基対合によりアニ

50

ーリングするプロセスをハイブリダイゼーションまたはハイブリダイズすると称し得る。これは、2つの相補的なポリヌクレオチド間の特異的な、すなわちランダムでない相互作用である。ハイブリダイゼーションおよびハイブリダイゼーションの強度（すなわち、核酸間の会合の強度）は、核酸同士の相補度合、関与する条件のストリンジェンシーおよび形成されるハイブリッドの融解温度（ T_m ）などの要素に影響される。ハイブリダイゼーション産物は、溶液中または固相支持体上の標的と形成される二本鎖または三本鎖であり得る。

【0059】

一部の態様では、核酸に、表1もしくは表2または表1と表2の組合せあるいは表3または表4に列挙された便由来真核生物RNAバイオマーカーの解析および定量に有用な短い核酸配列が包含され得る。かかる単離された核酸はオリゴヌクレオチドプライマーであり得る。一般に、オリゴヌクレオチドプライマーは、標的ヌクレオチド配列、例えば表1もしくは表2または表1と表2の組合せあるいは表3または表4に列挙された任意の便由来真核生物RNAバイオマーカーのヌクレオチド配列に相補的なオリゴヌクレオチドであって、DNAポリメラーゼまたはRNAポリメラーゼの存在下での該プライマーの3'末端へのヌクレオチドの付加によるDNA合成の開始点としての機能を果たし得るオリゴヌクレオチドである。プライマーの3'ヌクレオチドは一般的に、最適な伸長および/または増幅のために、対応するヌクレオチド位置において標的配列と同一であるべきである。プライマーは、例えばペプチド核酸プライマー、ロックド核酸プライマー、アンロックド（unlocked）核酸プライマーおよび/またはホスホロチオエート修飾プライマーなどの多くの形態をとり得る。一部の態様では、フォワードプライマーがdsDNAのアンチセンス鎖に相補的なプライマーであり得、リバープライマーがdsDNAのセンス鎖に相補的なプライマーであり得る。また、本発明者らはプライマーペアにも言及し得る。一部の態様では、5'標的プライマーペアが、標的ヌクレオチド配列の5'領域を増幅する少なくとも1種のフォワードプライマーと少なくとも1種のリバープライマーを含むプライマーペアであり得る。一部の態様では、3'標的プライマーペアが、標的ヌクレオチド配列の3'領域を増幅する少なくとも1種のフォワードプライマーと少なくとも1種のリバープライマーのプライマーペアであり得る。一部の態様では、プライマーが、以下に論考するような検出可能な標識を含み得る。一部の態様では、検出可能な標識が定量可能な標識であり得る。

【0060】

本明細書において提供するオリゴヌクレオチドプライマーは、表1および表2または表3もしくは表4に列挙された任意の便由来真核生物RNAバイオマーカーの増幅に有用である。一部の態様では、オリゴヌクレオチドプライマーは、本明細書に開示の便由来真核生物RNAバイオマーカー、例えば表1および表2または表3もしくは表4に列挙された便由来真核生物RNAバイオマーカーのうちの2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30種またはそれより多くに相補的であり得る。プライマーの長さは、プローブの具体的な核酸配列のヌクレオチド塩基の配列および組成ならびにプローブが使用される具体的な方法に応じてさまざまであり得る。一般に、有用なプライマーの長さは、約8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30個のヌクレオチド塩基であり得る。有用なプライマーの長さは、8個のヌクレオチド塩基～約60個のヌクレオチド塩基；約12個のヌクレオチド塩基～約50個のヌクレオチド塩基；約12個のヌクレオチド塩基～約45個のヌクレオチド塩基；約12個のヌクレオチド塩基～約40個のヌクレオチド塩基；約12個のヌクレオチド塩基～約35個のヌクレオチド塩基；約15個のヌクレオチド塩基～約40個のヌクレオチド塩基；約15個のヌクレオチド塩基～約35個のヌクレオチド塩基；約18個のヌクレオチド塩基～約50個のヌクレオチド塩基；約18個のヌクレオチド塩基～約40個のヌクレオチド塩基；約18個のヌクレオチド塩基～約35個のヌクレオチド塩基；約18個のヌクレオチド塩基～約30個のヌクレオチド塩基；約20個のヌクレオチド塩基～約30個のヌクレオチド塩基；約20個のヌクレオチド塩基～約25個のヌクレオチド塩基の範囲であり得る。

【0061】

表1および表2または表3もしくは表4に列挙された任意の便由来真核生物RNAバイオマーカに選択的に結合する、ならびに該バイオマーカに相補的な、プローブ、すなわち、単離された核酸断片も提供する。プローブは、オリゴヌクレオチドまたはポリヌクレオチド、DNAまたはRNA、一本鎖または二本鎖および天然型および修飾型であり得、ヌクレオチド塩基内または骨格内のいずれかに存在し得る。プローブは、さまざまな方法、例えば化学合成または酵素による合成によって作製され得る。

【0062】

プローブの長さは、プローブの具体的な核酸配列のヌクレオチド塩基の配列および組成ならびにプローブが使用される具体的な方法に応じてさまざまであり得る。一般に、有用なプローブの長さは、約8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、35、40、50、55、60、65、70、75、80、85、90、100、110、120、140、150、175または200個のヌクレオチド塩基であり得る。一般に、有用なプローブの長さは、約8～約200個のヌクレオチド塩基；約12～約175個のヌクレオチド塩基；約15～約150個のヌクレオチド塩基；約15～約100個のヌクレオチド塩基；約15～約75個のヌクレオチド塩基；約15～約60個のヌクレオチド塩基；約20～約100個のヌクレオチド塩基；約20～約75個のヌクレオチド塩基；約20～約60個のヌクレオチド塩基；約20～約50個のヌクレオチド塩基の長さの範囲である。一部の態様では、プローブセットは、表1もしくは表2または表1と表2の組合せあるいは表3または表4に列挙された便由来真核生物RNAバイオマーカのうちの2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30種またはそれより多くに指向されるプローブを含み得る。

【0063】

本明細書に開示のプライマーおよびプローブは検出可能に標識され得る。標識は、検出され得るか、あるいは検出可能な応答をもたらし得、核酸に直接または間接的に結合され得る分子性部分または化合物であり得る。直接標識には、標識とプローブを連結させるための結合または相互作用が使用され得、このようなものとしては共有結合、非共有結合性相互作用（水素結合、疎水性相互作用およびイオン相互作用）またはキレートもしくは配位錯体が挙げられる。間接標識には、直接または間接的に標識され、シグナルを増幅し得る架橋部分またはリンカー（例えば、抗体、オリゴマーまたは別の化合物）が使用され得る。標識としては任意の検出可能な部分、例えば放射性核種、リガンド、例えばビオチンもしくはアビジン、酵素、酵素基質、反応性基、クロモフォア（検出可能な色素、粒子もしくはビーズ）、フルオロフォアまたは発光化合物（生物発光標識、リン光標識もしくは化学発光標識）が挙げられる。標識は、混合物中の結合標識プローブが、未結合標識プローブのもの、例えば安定性または分解の差と比べて、結合形態の未結合形態からの物理的分離を必要とすることなく検出可能な変化を示す均一系アッセイにおいて検出可能であり得る。

【0064】

好適な検出可能な標識としては、それ自体が検出可能である分子（例えば、蛍光性部分、電気化学的標識、金属キレートなど）ならびに検出可能な反応生成物の生成によって間接的に検出され得る分子（例えば、酵素、例えばホースラディッシュペルオキシダーゼ、アルカリホスファターゼなど）またはそれ自体が検出可能であり得る特異的結合分子によって間接的に検出され得る分子（例えば、ビオチン、ジゴキシゲニン、マルトース、オリゴヒスチジン、2,4-ジントロベンゼン、フェニルアルセネート、ssDNA、dsDNAなど）が挙げられ得る。上記に論考したように、1つまたは複数のリガンドモチーフおよび/またはリガンドと検出可能な標識とのカップリングは直接であっても間接的であってもよい。検出は、インサイチュ、インビボ、組織切片でのインビトロまたは溶液中などであり得る。

【0065】

一部の態様では、方法が、アルカリホスファターゼコンジュゲートポリヌクレオチドプローブの使用を含む。アルカリホスファターゼ（AP）-コンジュゲートポリヌクレオチドプローブが使用される場合、適切な基質、例えばファーストブルーまたはファーストレッ

ド基質の逐次添加後、APは基質を分解し、特定の標的RNA分子のインサイチュ検出を可能にする沈殿物を形成する。アルカリホスファターゼは、いくつかの基質、例えばファーストブルー、ファーストレッドまたは5-ブromo-4-クロロ-3-インドリル-ホスフェート (BCIP) とともに使用され得る。

【0066】

一部の態様では、フルオロフォア-コンジュゲートプローブが蛍光色素コンジュゲート標識プローブであり得るか、あるいはアルカリホスファターゼの他に発色検出経路のための他の酵素的アプローチを使用するもの、例えば3,3'-ジアミノベンジジン (DAB) などの基質とのホースラディッシュペルオキシダーゼコンジュゲートプローブの使用であり得る。

10

【0067】

コンジュゲート型標識プローブに使用される蛍光色素は典型的には、例えばフルオレセインおよびその誘導体;ローダミンおよびその誘導体;シアニンおよびその誘導体;クマリールおよびその誘導体;Cascade Blue (商標) およびその誘導体;ルシファーイエローおよびその誘導体;BODIPYおよびその誘導体;などのファミリーに分けられ得る。例示的なフルオロフォアとしては、インドカルボシアニン (C3)、インドジカルボシアニン (C5)、Cy3、Cy3.5、Cy5、Cy5.5、Cy7、テキサスレッド、パシフィックブルー、オレゴングリーン488、Alexa Fluor (登録商標) -355、Alexa Fluor 488、Alexa Fluor 532、Alexa Fluor 546、Alexa Fluor-555、Alexa Fluor 568、Alexa Fluor 594、Alexa Fluor 647、Alexa Fluor 660、Alexa Fluor 680、JOE、リサミン、ローダミングリーン、BODIPY、フルオレセインイソチオシアネート (FITC)、カルボキシ-フルオレセイン (FAM)、フィコエリトリン、ローダミン、ジクロロローダミン (dRhodamine (商標))、カルボキシテトラメチルローダミン (TAMRA (商標))、カルボキシ-X-ローダミン (ROX (商標))、LIZ (商標)、VIC (商標)、NED (商標)、PET (商標)、SYBR、PicoGreen、RiboGreenなどが挙げられる。近赤外色素は、フルオロフォアおよび蛍光レポーター基という用語に意図される意味に明白に含まれる。

20

【0068】

一部の態様では、真核生物バイオマーカーのレベルが遺伝子アレイにおいて解析され得る。マイクロアレイ解析はカスタム遺伝子アレイにおいて行なわれ得る。択一的または付加的に、マイクロアレイ解析は、市販のシステムを製造業者の使用説明書およびプロトコルに従って用いて行なわれ得る。例示的な市販のシステムとしては、Affymetrix GENECHIP (登録商標) 技術 (ThermoFisher, Waltham, MA)、Agilentマイクロアレイ技術、NCOUNTER (登録商標) Analysis System (NanoString (登録商標) Technologies, Seattle, WA) およびBeadArray Microarray Technology (Illumina, San Diego, CA) が挙げられる。便試料から抽出される核酸は遺伝子アレイ上のプローブにハイブリダイズされ得る。プローブ-標的間のハイブリダイゼーションは化学発光によって検出され、特定の配列の相対存在度が調べられ得る。特定の配列の相対存在度は遺伝子アレイ間または遺伝子アレイ内で正規化され得る。

30

【0069】

一部の態様では、プローブおよびプローブセットが遺伝子アレイとして構成され得る。遺伝子アレイはマイクロアレイまたは遺伝子チップとしても知られ、複雑な生物学的試料の並行解析を可能にする、核酸が規則的に並んでいるアレイである。典型的には、遺伝子アレイは、固相基材、例えばマイクロチップ、ガラススライドまたはビーズに結合されたプローブを含む。該結合は一般的に、基材とプローブ間の共有結合をもたらし化学的カップリングを伴う。アレイ内のプローブの数はさまざまであり得るが、各プローブは、アレイまたはマイクロチップ上のアドレス指定可能な特定の位置に固定されている。一部の態様では、プローブが、約18個のヌクレオチド塩基、約20個のヌクレオチド塩基、約25個のヌクレオチド塩基、約30個のヌクレオチド塩基、約35個のヌクレオチド塩基または約40個のヌクレオチド塩基の長さであり得る。一部の態様では、プローブセットが、表1もしくは表2または表1と表2の組合せあるいは表3または表4に列挙された便由来真核生物RNAバイ

40

50

オマーカ-のうちの少なくとも2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30種またはそれより多くに指向されるプローブを含む。プローブセットは、5,000、10,000、20,000、50,000、100,000、200,000、300,000、400,000、500,000、600,000、700,000、800,000、900,000、1,000,000、2,000,000、3,000,000、4,000,000、5,000,000、6,000,000、7,000,000、8,000,000種またはそれより多くの異なるプローブを含む高密度アレイに組み込まれ得る。

【0070】

遺伝子アレイ合成の方法はさまざまであり得る。例示的な方法は、プローブの合成、続いて「スポットティング」によるアレイ表面上への堆積、例えばフォトリソグラフィまたは電気化学を用いた微小電極アレイ上でのインサイチュ合成を含む。

10

【0071】

一部の態様では、プローブおよびプローブセットが試薬として、すなわち複雑な生物学的試料の並行解析を可能にする核酸プールとして構成され得る。試薬は、例えば一組の増幅プローブ、ライブラリー調製物、アンプリコンパネルまたはキャプチャーパネルであり得る。典型的には、試薬は、溶液中に懸濁させたターゲティングプローブを含む。一部の態様では、該プローブが、特定の領域を標的化するように設計される。該プローブは、特定の核酸の捕捉が可能となるような様式で構成され得る。また、該プローブは、特定の核酸の増幅が可能となるように構成され得る。試薬中のプローブの数はさまざまであり得るが、各プローブは特異的配列に対して設計される。一部の態様では、プローブが、約10個のヌクレオチド塩基、約15個のヌクレオチド塩基、約20個のヌクレオチド塩基、約25個のヌクレオチド塩基、約30個のヌクレオチド塩基、約35個のヌクレオチド塩基または約40個のヌクレオチド塩基の長さであり得る。一部の態様では、プローブセットが、表1もしくは表2または表1と表2の組合せあるいは表3または表4に列挙された少なくとも2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30種またはそれより多くの便由来真核生物RNAバイオマーカ-に指向されるプローブを含む。プローブセットは、5,000、10,000、20,000、50,000、100,000、200,000、300,000、400,000、500,000、600,000、700,000、800,000、900,000、1,000,000、2,000,000、3,000,000、4,000,000、5,000,000、6,000,000、7,000,000、8,000,000種またはそれより多くの異なるプローブを含む高密度試薬に組み込まれ得る。

20

【0072】

試薬合成の方法はさまざまであり得る。例示的な方法は、核酸プローブの合成、続いて安定化溶液中への懸濁を含む。プローブ試薬は、分子識別子としての機能を果たすユニークな領域を含み得る。該試薬は、PCR、rtPCR ddPCR、dPCR、次世代シーケンシング、アンプリコンシーケンシング、RNA-seなどの方法および他の方法のために使用され得る。

30

【0073】

また、真核生物バイオマーカ-のレベルはDNAシーケンシングによっても解析され得る。DNAシーケンシングは、ターゲットシーケンシング、全ゲノムシーケンシング、アンプリコンシーケンシングまたはエクソームシーケンシングなどのシーケンシング法によって行なわれ得る。シーケンシング方法としては：サンガーシーケンシングまたはハイスループットシーケンシングが挙げられ得る。ハイスループットシーケンシングは、合成時解読、パイロシーケンシング、ライゲーションによるシーケンシング、リアルタイムシーケンシング、ナノポアシーケンシングまたはサンガーシーケンシングを伴い得る。一部の態様では、単離されたRNAが対応するcDNAを生成させるために使用され得、このcDNAがシーケンシングされ得る。

40

【0074】

本明細書に記載のシーケンシング法は、複数の異なる標的核酸が同時に操作されるようなマルチプレックス形式で行なわれ得る。一部の態様では、異なる標的核酸が共通の反応槽内または特定の基材の表面上で処理され、シーケンシング試薬の簡便な送達、未反応試薬の除去および組込み事象の検出がマルチプレックス様式で可能となり得る。表面結合標的核酸を伴う一部の態様では、標的核酸がアレイ形式であり得る。アレイ形式では、標的

50

核酸は典型的には表面に空間的に区別可能な様式でカップリングされ得る。例えば、標的核酸は、共有結合性直接結合、ビーズまたは他の粒子との結合によって結合され得るか、あるいはポリメラーゼまたは表面結合させる他の分子と会合され得る。アレイは、各部位（フィーチャーとも称される）に単一コピーの標的核酸を含んでもよく、または同じ配列を有する複数のコピーが各部位またはフィーチャーに存在してもよい。複数のコピーは、ブリッジ増幅、アンプリコン増幅、PCRまたはエマルジョンPCRなどの増幅方法によって作製される。

【0075】

一部の態様では、正規化工程が、試料間の核酸回収およびばらつきを制御するために使用され得る。一部の態様では、既定量の外来性対照核酸が、抽出された真核生物核酸に添加（「スパイクイン」）され得る。外来性対照核酸は、1つまたは複数の真核生物配列または非真核生物配列に対応する配列、例えばPhiXを有する核酸であり得る。択一的または付加的に、外来性対照核酸は、別の種にみられる配列に対応する配列、例えば細菌配列、例えば枯草菌（*Bacillus subtilis*）配列を有し得る。一部の態様では、該方法が、1つまたは複数のハウスキーピング遺伝子のレベルを調べる工程を含み得る。一部の態様では、該方法が、バイオマーカーの発現レベルをハウスキーピング遺伝子のレベルに対して正規化する工程を含み得る。

【0076】

方法は、実験試料における、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30種またはそれより多くの便由来真核生物RNAバイオマーカーの測定された発現レベルが、対照における、同じ2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30種またはそれより多くの便由来真核生物RNAバイオマーカーの測定された発現レベルと異なるかどうかを判定する工程を含む。別の態様では、方法が、実験試料における、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30種またはそれより多くの便由来真核生物RNAバイオマーカーの発現レベルの割合が、対照における、同じ2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30種またはそれより多くの便由来真核生物RNAバイオマーカーの測定された発現レベルの割合と異なるかどうかを判定する工程を含む。該発現レベルまたは発現レベルの割合の差は増加であっても減少であってもよい。

【0077】

本明細書に開示の組成物は結腸直腸新生物の検出、診断および処置に概してかつ様々な有用である。検出方法は、表1もしくは表2または表1と表2の組合せあるいは表3または表4に列挙されたバイオマーカーから選択される2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30種またはそれより多くの便由来真核生物RNAバイオマーカーの、便試料中における発現レベルを測定する工程、および、該試料における、表1もしくは表2または表1と表2の組合せあるいは表3または表4に列挙されたバイオマーカーから選択される該2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30種またはそれより多くの便由来真核生物RNAバイオマーカーの測定された発現レベルを、対照における、表1もしくは表2または表1と表2の組合せあるいは表3または表4に列挙されたバイオマーカーから選択される該2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30種またはそれより多くの便由来真核生物RNAバイオマーカーの測定された発現レベルに対する、患者の試料における、表1もしくは表2または表1と表2の組合せに列挙されたバイオマーカーから選択される該2、3、4、5、6、7、8、9

、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30種またはそれより多くの便由来真核生物RNAバイオマーカーの測定された発現レベルの差は、該患者が結腸直腸新生物、またはより具体的には高リスク腺腫を有することの表示である。一部の態様では、対照における、表1もしくは表2または表1と表2の組合せに列挙されたバイオマーカーから選択される該2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30種またはそれより多くの便由来真核生物RNAバイオマーカーの測定された発現レベルに対する、患者の試料における、表1もしくは表2または表1と表2の組合せに列挙されたバイオマーカーから選択される該2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30種またはそれより多くの便由来真核生物RNAバイオマーカーの測定された発現レベルの差は、該患者に結腸直腸新生物、またはより具体的には高リスク腺腫のリスクがあることの表示である。このような方法は、結腸直腸新生物、例えば結腸直腸がんもしくは前がん性病変を有するか、または結腸直腸新生物が発生するリスクがある対象（例えば、患者、より具体的にはヒト患者）を特定する工程をさらに含み得る。

10

【0078】

対照における、表3に列挙されたバイオマーカーから選択される1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17または18種の便由来真核生物RNAバリエーションバイオマーカーのバリエーションアレル頻度に対する、対象の試料における、表3に列挙されたバイオマーカーから選択される該1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17または18種の便由来真核生物RNAバリエーションバイオマーカーのバリエーションアレル頻度の差は、該患者が結腸直腸新生物を有することの表示である。一部の態様では、対照における、表3に列挙されたバイオマーカーから選択される2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17または18種の便由来真核生物RNAバイオマーカーの測定されたバリエーションアレル頻度に対する、患者の試料における、表3に列挙されたバイオマーカーから選択される該1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17または18種の便由来真核生物RNAバイオマーカーの測定されたバリエーションアレル頻度の差は、該患者に結腸直腸新形成のリスクがあることの表示である。このような方法は、結腸直腸新形成、例えば結腸直腸がんもしくは前がん性病変を有するか、または結腸直腸新生物が発生するリスクがある対象（例えば、患者、より具体的にはヒト患者）を特定する工程をさらに含み得る。

20

30

【0079】

対照における、図4に列挙された結腸直腸新生物分子サブタイプバイオマーカー遺伝子から選択される1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30種またはそれより多くの便由来真核生物RNAバイオマーカーの測定された発現レベルに対する、患者の試料における、図4に列挙された結腸直腸新生物分子サブタイプバイオマーカー遺伝子から選択される該1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30種またはそれより多くの便由来真核生物RNAバイオマーカーの測定された発現レベルの差は、該患者が結腸直腸がんの分子サブタイプ、例えばCMS1を有することの表示である。一部の態様では、対照における、図4に列挙された結腸直腸新生物分子サブタイプバイオマーカー遺伝子から選択される2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30種またはそれより多くの便由来真核生物RNAバイオマーカーの測定された発現レベルに対する、患者の試料における、図4に列挙された結腸直腸新生物分子サブタイプバイオマーカー遺伝子から選択される該1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30種またはそれより多くの便由来真核生物RNAバイオマーカーの測定された発現レベルの差は、該患者に結腸直腸がんの分子サブタイプ、例えばCMS1のリスクがあることの表示である。このような方法は、結腸直腸新形成、例えば結腸直腸がんもしくは前がん性病変を有するか、または結腸直腸新生物が発生す

40

50

るリスクがある対象（例えば、患者、より具体的にはヒト患者）を特定する工程をさらに含み得る。

【0080】

結腸直腸新生物には任意の形態の結腸直腸がんが包含され得る。また、結腸直腸新生物にはポリープ、例えば前がん性病変も包含され得る。結腸直腸がんは典型的には結腸または直腸の管腔内層内の、ポリープと称される増殖物として発症する。結腸直腸ポリープは一般的に2つのカテゴリー：腺腫性ポリープと良性ポリープに分けられる。腺腫性ポリープは腺腫とも称され得る。良性ポリープは、過形成性ポリープ、過誤腫性ポリープまたは炎症性ポリープとも称され得る。1つの腺腫性ポリープまたは複数の腺腫性ポリープを有する患者は、高リスク腺腫、中リスク腺腫または低リスク腺腫を有するとクラス分類され得る。高リスク腺腫には、上皮内癌もしくは高度異形成を有する任意のサイズの腺腫、25%であるか、もしくはそれより多い絨毛増殖パターンを有する任意のサイズの腺腫、1.0 cmであるか、もしくはそれより大きいサイズの任意の腺腫、または1.0 cmであるか、もしくはそれより大きいサイズの任意の鋸歯状病変が包含される。中リスク腺腫には、5.0 mm～1.0 cmの範囲のサイズの1つ、または2つの非高リスク腺腫あるいは1.0 cm未満のサイズの3つであるか、またはそれより多くの非高リスク腺腫が包含される。低リスク腺腫には、5.0 mmであるか、あるいはそれ未満のサイズの1つ、または2つの非高リスク腺腫が包含される。腺腫性ポリープは結腸直腸がんを生じ得る。結腸直腸がんの最も一般的な形態である腺癌は、結腸および/または直腸内部の内側に存在する腸腺細胞に発生する。腺癌には、有茎部上の腺癌である管状腺癌が包含され得る。また、腺癌には、結腸表面上に平坦に存在する腺癌である絨毛腺癌も包含され得る。他の結腸直腸がんは、その起源組織によって区別される。このようなものとしては、カハール介在細胞から生じる消化管間質腫瘍（GIST）；血液系細胞から生じる原発性結腸直腸リンパ腫；結合組織または平滑筋から生じる肉腫である平滑筋肉腫；メラニン細胞から生じる黒色腫；重層扁平上皮組織から生じ、直腸に限局性の扁平上皮癌；および一般的に予後不良が付随する上皮がんである粘液性癌が挙げられる。

【0081】

結腸直腸新形成または結腸直腸がんの症状としては、非限定的に、用便習慣の変化、例えば下痢もしくは便秘または4週間より長く持続する便の硬さの変化、直腸出血または便に血が混じる、持続性の腹部不快感、例えば急激な腹痛、ガスまたは痛み、腸が完全に空にならない感じ、脱力感または疲労感および説明のつかない体重減少が挙げられ得る。結腸直腸新形成または結腸直腸がんを有することが疑われる患者には末梢血検査、例えば全血算（CBC）、糞便潜血検査（FOBT）、肝臓機能解析、免疫化学的便検査（FIT）および/または特定の腫瘍マーカー、例えば癌胎児性抗原（CEA）およびCA19-9の他の解析が行なわれ得る。結腸直腸新形成または結腸直腸がんは多くの場合、結腸内視鏡検査に基づいて診断される。結腸内視鏡検査の際、確認されたポリープがあれば、これは取り出され、生検され、解析されて、このポリープに結腸直腸がん細胞または前がん性病変が起こっている細胞が含まれているかどうか判断される。上記に挙げた具体的ながんは各々、内視鏡で見ると異なる外観を有し得る。絨毛腺腫 黒色腫および扁平上皮癌は典型的には平坦または無茎性であるが、管状腺腫、リンパ腫、平滑筋肉腫およびGIST腫瘍は典型的には有茎性である。しかしながら、平坦で無茎性の腺腫は、結腸内視鏡検査の際に胃腸科専門医が見逃す場合があり得る。生検試料は、特定の遺伝子の遺伝子変化またはマイクロサテライト不安定性に基づいてさらなる解析に供され得る。

【0082】

他の診断方法としては、S状結腸鏡検査；画像診断検査、例えばコンピュータ断層撮影（CTもしくはCAT）スキャン；超音波、例えば腹部、直腸内もしくは術中超音波；または核磁気共鳴画像法（MRI）スキャン、例えば経直腸的MRIが挙げられ得る。他の検査、例えば血管造影および胸部x線検査は、結腸直腸がんが転移しているかどうかを調べるために行なわれ得る。

【0083】

10

20

30

40

50

結腸直腸がんの進行度分類のためのさまざまな方法が開発されている。最も一般的に使用されているシステムである、TNMシステムは、3つの要素:1) 原発性腫瘍が腸の壁内および隣接する領域内に増殖している距離 (T) ;2) 腫瘍が隣接する所属リンパ節まで広がっているかどうか (N) ;3) がんが他の器官に転移しているかどうか (M) に基づいている。他の進行度分類方法としては、デュークス進行度分類およびアストラ・カラー分類が挙げられる。

【0084】

TNMシステムは結腸直腸がんの4つのステージクラス分類を提供する。ステージ1 (T1) の結腸直腸がんでは、腫瘍が結腸壁の層内に増殖しているが結腸壁の外側またはリンパ節内には広がっていない。がんが管状腺腫ポリープの一部であれば、単純切除が行なわれ、患者は将来的ながんの発症について定期的な検査を受け続け得る。がんが高悪性度であるか、または平坦な/無茎性ポリープの一部である場合、さらなる手術が必要とされるかもしれない、広めのマージンが取られる;これには、結腸片が切除される部分結腸切除術が含まれ得よう。ステージ2 (T2) の結腸直腸がんでは、腫瘍が結腸の壁内に、もしかすると隣接する組織内にも増殖しているが隣接するリンパ節には広がっていない。腫瘍の外科的除去および部分結腸切除術が一般的に行なわれる。補助療法、例えば5-フルオロウラシル、ロイコボリンまたはカペシタビンなどの薬剤を用いる化学療法が施行される場合があり得る。かかる腫瘍は再発する可能性は低い、患者に対してスクリーニングを増やすことが一般的に必要とされる。ステージ3 (T3) の結腸直腸がんでは、腫瘍が隣接するリンパ節にまで広がっているが他の身体部分までには広がっていない。結腸片および冒されたすべてのリンパ節を除去するための手術が必要とされる。5-フルオロウラシル、ロイコボリン、オキサリプラチンまたはカペシタビンなどの薬剤をオキサリプラチンと併用して用いる化学療法が典型的に推奨される。また、患者の年齢および腫瘍の侵襲性の性質に応じて放射線療法も使用され得る。ステージ4 (T4) の結腸直腸がんでは、腫瘍が結腸から血液を介して遠隔器官に広がっている。結腸直腸がんは高頻度で肝臓、肺および/または腹膜に転移する。手術でこのようながんが治癒することは起こりにくく、生存率を改善するためには化学療法およびまたは放射線が一般的に必要とされる。

【0085】

本明細書に開示の方法は結腸直腸新形成の診断および処置に概して有用である。2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30種またはそれより多くの便由来真核生物RNAバイオマーカー、例えば表1もしくは表2または表1と表2の組合せから選択される便由来真核生物RNAバイオマーカーあるいは表3または表4の便由来真核生物RNAバイオマーカーの発現レベルが生物学的試料において、例えば対象由来の便試料において測定される。対象は、該患者に結腸直腸がんのリスクがあることを示すものであろう上記の症状のうちの1つまたは複数を有する患者であり得る。また、対象は、症状を有していないが、結腸直腸がんの高リスクと関連している、年齢 (例えば、50歳より上)、家族歴、肥満、食生活、アルコール消費量、タバコ使用、結腸直腸ポリープの診断歴、人種および民族的背景に基づいた結腸直腸新形成、炎症性腸疾患または遺伝的症候群、例えば家族性腺腫性ポリポーシス、ガードナー症候群、リンチ症候群、ターコット症候群、ポイツ・ジェガース症候群ならびにMUTYH関連ポリポーシスのリスクがあると思われる患者であってもよい。また、本明細書に開示の方法は、以前に結腸直腸新形成または結腸直腸がんとして診断され、処置された患者を、寛解をモニタリングするため、および病変の再発を検出するためにモニタリングするためにも有用である。

【0086】

一部の態様では、対象、すなわちヒト患者または非ヒト動物患者の疾患状態が病理学的評価によって判定される。例えば、ある型の疾患、例えば結腸直腸がんでは、疾患の程度がステージ1 (T1)、ステージ2 (T2)、ステージ3 (T3) およびステージ4 (T4) とクラス分類される。結腸直腸がんは、管状腺癌、絨毛腺癌、消化管間質腫瘍、原発性結腸直腸リンパ腫、平滑筋肉腫、黒色腫、扁平上皮癌または粘液性癌であり得る。別の型の疾患、例

例えば炎症性腸疾患では、疾患状態が、腸管に沿った疾患の場所ならびに組織学的フィーチャー、例えば肉芽腫、白血球浸潤部および/または陰窩膿瘍によって判定される。また、疾患状態を判定するための他の方法、例えば医師の判断、身体症状、糞便潜血検査、免疫化学的便検査、S状結腸鏡検査、FIT-DNA、CTコロノグラフィーまたは結腸内視鏡検査も本明細書に開示の方法と併せて使用され得る。

【0087】

対象に腸疾患のリスクがあるかどうかを判定する方法も提供する。腸疾患としては、腸のがん、結腸直腸がん、前がん性変化を示す腺腫性ポリープ、過敏性腸症候群、壊死性腸炎、潰瘍性大腸炎、クローン病 セリアック病または他の腸疾患が挙げられ得る。対象に腸疾患のリスクがあるかどうかを判定する方法は、a) デオキシリボ核酸 (DNA) の配列、
b) リボ核酸 (RNA) の配列、c) タンパク質の骨格を構成する予測アミノ酸配列、d) リボ核酸バイオマーカーの発現レベル、e) アミノ酸の配列変動の予測、またはf) 上記の任意の組合せを検出するために本発明を使用することにより判定され得、ここで、対照と実験試料との間の差は対象に腸疾患のリスクがあることを示し得る。

【0088】

また、該方法および組成物は、腸疾患を有する対象のための臨床計画を選択するのに有用である。この方法を通して、臨床計画には、さらなる診断手順の施行が含まれ得る。一部の態様では、臨床計画に処置方法が含まれ得る。

【0089】

例えば診断、状態、または処置に対する応答を調べるためのアルゴリズムが具体的な臨床条件に対して決定され得る。本明細書において提供する方法で使用されるアルゴリズムは、非限定的に医療デバイス、臨床評価スコアまたは生物学的試料の生物学的/化学的/物理的検査を用いて定量され得る複数のパラメータを組み込む数学関数であり得る。各数学関数は、選択された臨床条件に関連性があると判定されたパラメータレベル (例えば、測定レベル) の加重調整 (weight-adjusted) 表現であり得る。複数のマーカーパネルの加重および評価に関与する手法のため、データ解析するためには妥当な計算能力を有するコンピュータが使用され得る。

【0090】

したがって、診断方法は、結腸直腸新生物のリスクがあるか、あるいは結腸直腸新生物を有することが疑われる患者由来の便試料採取する工程;表1もしくは表2または表1と表2の組合せあるいは表3または表4に列挙された便由来真核生物RNAバイオマーカーから選択される2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30種またはそれより多くの便由来真核生物RNAバイオマーカーの発現を明らかにする工程、および複数種の便由来真核生物RNAバイオマーカーを所定の係数を用いて組み込む機械学習アルゴリズムによって試験値を得る工程を含み得る。例示的な機械学習アルゴリズムとしては、サポートベクターマシン、勾配ブースティング、適応的ブースティング、ランダムフォレスト、ナイーブベイズ、決定木またはk近傍法などが挙げられる。対照、例えば健常個体集団に対する複数種の結腸直腸新生物バイオマーカーの発現の有意な変化は患者が結腸直腸新形成を有する可能性が高いことを示す。一部の態様では、試料において測定された発現レベルを用いて確率または信頼度スコアが誘導または算出される。この値は発現レベルから誘導され得る。択一的または付加的に、該値は、発現レベルと他の要素、例えば患者の病歴、民族、性別、年齢、喫煙状態、以前のゲノム検査結果、以前の組織病理学検査結果および遺伝的背景との組合せから誘導され得る。択一的または付加的に、該値は発現レベルと免疫化学的便検査 (FIT) との組合せから誘導され得る。一部の態様では、該方法が、試験値を患者に伝達する工程をさらに含み得る。この方法としては、例えばマーカーの視覚的表示、マーカーの数値出量または他の伝達方法が挙げられ得よう。

【0091】

一部の態様では、1人または複数人の患者に対する予測がモデルベースのアプローチを用いて行なわれ得る。例えば、一部の態様では、ランダムフォレストモデルが、1つまた

は複数の群における疾患の非存在、疾患の存在および/または疾患の重症度、例えば結腸直腸がん、HRA、MRA、LRA、良性ポリープまたは所見なしが予測されるように構成され得る。一部の態様では、バリデーションデータセットおよび/またはテストデータセットを適用し、モデルの試験または絞り込みがなされ得る。モデルは、作成されたら、1人または複数人の特定患者の疾患の非存在、疾患の存在および/または疾患の重症度を、例えば複数のアンプリコンなどの提供した入力に基づいて予測するために使用される。具体的な態様を本明細書で論考しているが、任意の適当なモデルが任意の数の決定木、ノード、入力層、出力層、隠れ層または他の多様なパラメータを含み得ることは認識されよう。一部の態様では、より多くの数および/またはより少ない数の決定木、より多くの数および/またはより少ない数の候補フィーチャーなどを使用するランダムフォレストモデルが作成され得る。

10

【0092】

一部の態様では、該1つまたは複数のモデルが、疾患の検出用に構成されたシステムを用いて作成、試験および/または実行され得る。一部の態様では、該システムが、1つまたは複数のプロセッサを有するコンピュータシステムを含む。各プロセッサは、通信基盤（例えば、通信バス、クロスオーバーバー（cross-over bar）またはネットワーク）に接続される。プロセッサは、中央処理装置、組み込み型プロセッサもしくはマイクロコントローラ、特定用途向け集積回路（ASIC）および/または1つもしくは複数の工程を行なうためのコンピュータ実行可能命令が実行されるように構成された任意の他の回路として実装され得る。プロセッサは、上記で論考したプロセッサと同様であり、本明細書において同様の説明は繰り返さない。コンピュータシステムには、図、テキストおよび他のデータをディスプレイユニットのディスプレイ用通信基盤から（またはフレームバッファから）ユーザーに転送するディスプレイインターフェイスが含まれ得る。

20

【0093】

また、コンピュータシステムには主記憶装置、例えばランダムアクセスメモリ（RAM）および二次記憶装置も含まれ得る。主記憶装置および/または二次記憶装置はダイナミックランダムアクセスメモリ（DRAM）を含む。二次記憶装置としては、例えば、ソリッドステートメモリであり得るハードディスクドライブ（HDD）および/またはリムーバブルストレージドライブ、光ディスクドライブ、フラッシュドライブ、磁気テープドライブなどが挙げられ得る。リムーバブルストレージドライブはリムーバブルストレージユニットからの読み出しおよび/またはリムーバブルストレージユニットへの書き込みを行なう。リムーバブルストレージユニットは、光ディスク、磁気ディスク、フロッピーディスク、磁気テープなどであり得る。リムーバブルストレージユニットには、データがタンジブルに記憶されている（もしくは具体化されている）コンピュータ可読記憶媒体および/または例えばプロセッサ（1つもしくは複数）に種々の操作および/または1つもしくは複数の工程を行なわせるためのコンピュータ実行可能ソフトウェアの命令が含まれ得る。

30

【0094】

択一的な態様では、二次記憶装置としては、コンピュータプログラムまたは他の命令がコンピュータシステムにロードされることを可能にするための他のデバイスが挙げられ得る。二次記憶装置としては、リムーバブルストレージドライブと同様であり得、独自のリムーバブルストレージユニットを有するリムーバブルストレージユニットおよび対応するリムーバブルストレージインターフェイスが挙げられ得る。かかるリムーバブルストレージユニットの例としては、非限定的に、ソフトウェアおよびデータがリムーバブルストレージユニットからコンピュータシステムに転送されることを可能にするユニバーサルシリアルバス（USB）またはフラッシュドライブが挙げられる。

40

【0095】

また、コンピュータシステムには通信インターフェイス（例えば、ネットワーキングインターフェイス）も含まれ得る。通信インターフェイスは、命令およびデータがコンピュータシステムと1つまたは複数のさらなるシステム間で転送されることを可能にする。また、通信インターフェイスは他の外部デバイスとの通信も提供する。通信インターフェイス

50

スの例としては、モデム、イーサネットインターフェイス、ワイヤレスネットワークインターフェイス（例えば、無線周波数、IEEE 802.11 インターフェイス、ブルートゥースインターフェイスなど）、Personal Computer Memory Card International Association (PCMCIA) スロットおよびカードなどが挙げられ得る。通信インターフェイスを介して転送される命令およびデータは、通信インターフェイスによって受信され得る、電子、電磁気、光などであり得るシグナルの形態であり得る。このようなシグナルは通信インターフェイスに、ワイヤ、ケーブル、光ファイバ、電話線、移動体リンク、無線周波数 (RF) リンクおよび他の通信チャネルを用いて実装され得る通信路（例えば、チャネル）を介して提供され得る。

【0096】

本明細書に記載の方法およびシステムは、コンピュータ実行プロセスを実施するための該プロセスおよび装置の形態で少なくとも部分的に具体化され得る。また、本開示の方法は、コンピュータ実行可能プログラムコードでコードされるタンジブルな非一時的機械可読記憶媒体の形態で少なくとも部分的に具体化され得る。該媒体としては、例えばRAM、ROM、CD-ROM、DVD-ROM、BD-ROM、ハードディスクドライブ、フラッシュメモリ、またはコンピュータプログラムコードがコンピュータにロードされて実行されると、該コンピュータが、その方法を実施するための装置となる任意の他の非一時的機械可読記憶媒体が挙げられ得る。また、該方法は、コンピュータが、その方法を実施するための専用コンピュータとなるようにコンピュータプログラムコードがロードおよび/または実行されるコンピュータの形態で少なくとも部分的に具体化され得る。コンピュータプログラムコードセグメントは、汎用プロセッサで実行されると、該プロセッサを、本明細書に開示の方法を実行するための特定の接続、回路およびアルゴリズムが創出されるように構成する。

【0097】

本明細書に記載の方法を行なうため、例えば本明細書に記載の値を求めるために必要とされる計算を行なうためには標準的なコンピューティングデバイスおよびシステムが使用、実行、例えば適切にプログラミングされ得る。コンピューティングデバイスとしては、種々の形態のデジタルコンピュータ、例えばラップトップ、デスクトップ、モバイルデバイス、ワークステーション、パーソナルデジタルアシスタント、サーバ、ブレードサーバ、メインフレームおよび他の適切なコンピュータが挙げられる。一部の態様では、コンピューティングデバイスがモバイルデバイス、例えばパーソナルデジタルアシスタント、携帯電話、スマートフォン、タブレットまたは他の同様のコンピューティングデバイスである。

【0098】

一部の態様では、コンピュータが、情報を例えば保健医療専門家に通信するために使用され得る。情報は専門家に、該情報を電子的に入手可能にすることにより通信され得る（例えば、安全な様式で）。例えば、情報は、保健医療専門家が該情報にアクセスできるようにコンピュータデータベース上に配置され得る。また、情報は、専門家のためのエージェントとしての機能を果たす病院、クリニックまたは研究施設に通信され得る。オープンネットワーク（例えば、インターネットまたはe-メール）で転送される情報は暗号化され得る。患者の遺伝子発現のデータおよび解析は、暗号化してクラウドに保存してもよい。改ざん防止を有する256ビットAESの方法がディスク暗号化のために使用され得る；SSLプロトコルでは、データ伝送中の保護が好ましく確実にされ得、SHA2-HMACキーマネージメント手法ではデータへの認証アクセスが可能であり得る。他の確実なデータ保存手段もまた使用され得る。

【0099】

上記のかかる解析の結果、例えば、発現レベルと他の要素、例えば患者の病歴、民族、性別、年齢、喫煙状態、以前のゲノム検査結果、以前の組織病理学検査結果、遺伝的背景または免疫化学的便検査 (FIT) との組合せから誘導された確率または信頼度スコアは、担当医師による追跡および処置の根拠であり得る。2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30種または

10

20

30

40

50

それより多くの便由来真核生物RNAバイオマーカー、例えば表1もしくは表2または表1と表2の組合せから選択される便由来真核生物RNAバイオマーカーあるいは表3または表4の便由来真核生物RNAバイオマーカーの発現レベルが対照における同じ便由来真核生物RNAバイオマーカーの発現レベルと有意に異なる場合、医師は、現時点では患者に結腸直腸新生物のリスクはないと判断し得る。かかる患者は、将来的に再スクリーニングのために再度来院するよう促され得る。2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30種またはそれより多くの便由来真核生物RNAバイオマーカー、例えば表1もしくは表2または表1と表2の組合せから選択される便由来真核生物RNAバイオマーカーあるいは表3または表4の便由来真核生物RNAバイオマーカーの発現レベルが対照における同じ便由来真核生物RNAバイオマーカーの発現レベルと有意に異なる程度は、必要とされる追跡までの期間を判断するために使用され得る。一部の態様では、医師は、患者に追跡のために1ヶ月後、2ヶ月後、3ヶ月後、6ヶ月後、1年後、2年後、3年後、5年後または10年後に再度来院することを推奨し得る。本明細書に開示の方法は、結腸直腸新生物マーカーのレベルのなんらかの変化を経時的にモニタリングするために使用され得る。対象は、最初のスクリーニングおよび/または診断後、任意の長さの期間、モニタリングされ得る。例えば、対象は、少なくとも2、4、6、8、10、12、14、16、18、20、25、30、35、40、45、50、55もしくは60ヶ月またはそれより長く、あるいは少なくとも1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20年間またはそれより長くモニタリングされ得る。

10

【0100】

20

本明細書に開示の方法および組成物は、結腸直腸新形成または結腸直腸がんのリスクがあるか、あるいは結腸直腸新形成または結腸直腸がんに苦しんでいる対象のための臨床計画を選択するのに有用である。臨床計画としては、さらなる診断手順、例えば糞便潜血検査、免疫化学的便検査、またはがん、ポリープもしくは前がん性病変を除去するための結腸内視鏡検査の施行が挙げられ得る。一部の態様では、臨床計画に処置方法が含まれ得る。一部の態様では、該方法が、結腸直腸新生物または結腸直腸がんを有する対象に対する処置を選択する工程を含む。表1もしくは表2または表1と表2の組合せあるいは表3または表4から選択される便由来真核生物RNAバイオマーカーのうちの2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30種またはそれより多くのバイオマーカーの発現レベルが、対照における同じ2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30種またはそれより多くの便由来真核生物RNAバイオマーカーの発現レベルと有意に異なる場合、患者は結腸直腸新生物または結腸直腸がんを有し得る。このような場合、さらなるスクリーニング、例えば本明細書に開示の方法を用いたスクリーニングの頻度の増加、ならびに胎児潜血検査、免疫化学的便検査および/または結腸内視鏡検査が推奨され得る。表1もしくは表2または表1と表2の組合せあるいは表3または表4から選択される便由来真核生物RNAバイオマーカーのうちの2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30種またはそれより多くのバイオマーカーの発現レベルが、対照における、表1もしくは表2または表1と表2の組合せあるいは表3または表4から選択される便由来真核生物RNAバイオマーカーのうちの同じ2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30種またはそれより多くのバイオマーカーの発現レベルと有意に異なる場合、患者は特定の型の結腸直腸新生物、例えば高リスク腺腫を有し得る。一部の態様では、例えば、ポリープの除去を伴う結腸内視鏡検査、化学療法、免疫療法または手術、例えば腸切除などの処置が推奨され得る。したがって、該方法は、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30種またはそれより多くの便由来真核生物RNAバイオマーカー、例えば表1もしくは表2または表1と表2の組合せあるいは表3または表4から選択される便由来真核生物RNAバイオマーカーのうちの2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30種またはそれよ

30

40

50

り多くのバイオマーカーあるいはそのバリエーションの発現のレベルを調べるため、次いで処置コースを決定するために使用され得る。対象、すなわち患者は、臨床的に有益な結果がもたらされる場合はいつでも有効に処置される。これは、例えば疾患の症状の完全な解消、疾患の症状の重症度の低減または疾患の進行の低速化を意味し得る。このような方法は、a) 結腸直腸新形成または結腸直腸がんを有する対象（例えば、患者、より具体的にはヒト患者）を特定する工程、およびb) 該対象に抗がん処置、例えば治療用薬剤および例えば免疫療法剤、手術または放射線療法を提供する工程をさらに含み得る。疾患の症状の完全な解消、疾患の症状の重症度の低減または疾患の進行の低速化をもたらす、対象に提供される治療用薬剤の量は治療有効量とみなされる。また、本発明の方法には、投薬およびスケジューリングの最適化ならびに転帰の予測を補助するためのモニタリング工程も含められ得る。モニタリングは、薬物耐性の発現を検出するため、奏効患者を非奏効患者と速やかに区別するため、またはがんの再発を評価するためにも使用され得る。耐性または非奏効の徴候がみられる場合、医師は、腫瘍にさらなる回避機構が生じる前に択一的または補助的な薬剤を選出することができる。

【0101】

また、本明細書に開示の方法は、結腸直腸がんの診断および処置のための慣用的な方法と併用して使用することもできる。したがって、該診断方法は、結腸直腸がんの標準的な診断方法とともに使用され得る。例えば、該方法は、糞便潜血検査、免疫化学的便検査または結腸内視鏡検査と併用して使用され得る。また、該方法を、他の結腸直腸がんマーカー、例えばKRAS、NRAS、BRAF、CEA、CA 19-9、p53、MSL、DCC、MSIおよびMMRを用いて使用してもよい。

【0102】

また、本明細書に開示の診断方法を、結腸直腸がんの処置と併用して使用してもよい。結腸直腸がんの処置方法は、いくつかの一般カテゴリー：手術、化学療法、放射線療法、標的療法および免疫療法に分類される。手術としては、結腸切除術、部分肝切除術を伴う結腸人工肛門造設術、またはプロテクトミーが挙げられ得る。化学療法は全身性化学療法であってもよく、化学療法剤が罹患器官のすぐ近くに配置される局所化学療法であってもよい。例示的な化学療法剤としては、5-フルオロウラシル、オキサリプラチンもしくはその誘導体、イリノテカンもしくはその誘導体、ロイコボリンまたはカペシタビン、マイトマイシンC、シスプラチンおよびドキソルビシンが挙げられ得る。放射線療法は、がん放射線を指向する機械を使用する放射線外部照射療法であってもよく、結腸直腸がん内に直接または結腸直腸がん付近に放射性物質を配置する放射線内部照射療法であってもよい。標的剤としては、抗血管新生剤、例えばベバシズマブ）もしくはEGFR阻害モノクローナル抗体（セツキシマブ、パニツムマブ）、ラムシラマブ（抗VEGFR2）、アフリベルセプト、レゴラフェニブ、トリプフルリジン-チピラシルまたはその組合せが挙げられ得る。また、標的剤を標準的な化学療法剤と組み合わせてもよい。免疫療法としては、特異的抗体、例えば抗PD-1抗体、抗PD-L1抗体、およびtime-CTLA-4抗体、抗CD27抗体；がんワクチン、養子細胞療法薬、腫瘍溶解性ウイルス療法薬、アジュバント免疫療法薬およびサイトカインベース療法薬の投与が挙げられ得る。例示的な免疫療法薬としては、Keytruda、Opdivaおよびイブリムマブが挙げられ得る。他の処置方法としては、幹細胞移植、温熱療法、光線力学的療法、血液製剤献血および輸血またはレーザー処置が挙げられる。

【0103】

本発明者らは、バイオマーカーのレベルの統計学的に有意な量の増加/増大を一般的に意図するために「増加/増大した」、「増加/増大する」または「上方調節された」という用語を使用し得る。一部の態様では、増加/増大が対照と比べて少なくとも10%の増加/増大、例えば対照と比べて少なくとも約20%、もしくは少なくとも約30%、もしくは少なくとも約40%、もしくは少なくとも約50%、もしくは少なくとも約60%、もしくは少なくとも約70%、もしくは少なくとも約80%、もしくは少なくとも約90%の増加/増大もしくは100%を含む100%までの増加/増大または10~100%の任意の増加/増大あるいは対照と比べて少なくとも約0.5倍、もしくは少なくとも約1.0倍、もしくは少なくとも約1.2倍、もし

くは少なくとも約1.5倍、もしくは少なくとも約2倍、もしくは少なくとも約3倍、もしくは少なくとも約4倍、もしくは少なくとも約5倍もしくは少なくとも約10倍の増加/増大または1.0倍～10倍以上の任意の増加/増大であり得る。

【0104】

本発明者らは、真核生物バイオマーカーのレベルの統計学的に有意な量の減少を示すために「減少する」、「減少した」、「低減された」、「低減」または「下方調節された」という用語を使用し得る。一部の態様では、減少が対照と比べて少なくとも10%の減少、例えば対照と比べて少なくとも約20%、もしくは少なくとも約30%、もしくは少なくとも約40%、もしくは少なくとも約50%、もしくは少なくとも約60%、もしくは少なくとも約70%、もしくは少なくとも約80%、もしくは少なくとも約90%の減少もしくは100%を含む100%までの減少（すなわち対照と比べてレベルの非存在）または10～100%の任意の減少あるいは対照と比べて少なくとも約0.5倍、もしくは少なくとも約1.0倍、もしくは少なくとも約1.2倍、もしくは少なくとも約1.5倍、もしくは少なくとも約2倍、もしくは少なくとも約3倍、もしくは少なくとも約4倍、もしくは少なくとも約5倍もしくは少なくとも約10倍の減少または1.0倍～10倍以上の任意の減少であり得る。

10

【0105】

真核生物バイオマーカーにおける増加/増大または真核生物バイオマーカーにおける減少の統計学的有意性はp値またはq値として表示され得る。具体的な真核生物バイオマーカーによるが、p値またはq値は、0.05未満、0.01未満、0.005未満、0.002未満、0.001未満または0.0005未満であり得る。q値はp値から誘導される値であり得る。一部の態様では、q値が、偽発見率について調整されたp値であり得る。

20

【0106】

対照は、患者または患者群から採取された生物学的試料であってもよい。一部の態様では、対照が参照値であり得る。対照は、健常と診断されている個体または個体集団から得たものであり得る。健常個体としては、例えば、直近の1年以内の糞便寄生虫検査、糞便細菌検査、結腸内視鏡検査または内視鏡検査において検査で陰性であった個体が挙げられ得る。対照は、罹病状態と診断されている個体または個体集団から得たものであってもよい。罹病状態の個体としては、例えば、直近の1年以内の糞便寄生虫検査、糞便細菌検査、結腸内視鏡検査または内視鏡検査において検査で陽性であった個体が挙げられ得る。対照は、疾患を有すると以前に診断されたが現在は寛解状態であるか、活動性疾患を有していないか、または現在は該疾患に苦しんでいない個体または個体集団から得たものであってもよい。対照は、個体から1回、2回またはそれより多くの時点で得たものであってもよい。例えば、対照は、対象から早期の時点で採取された生物学的試料であり得る。対照は、具体的なバイオマーカーの基準参照値であり得る。基準参照値は、同様の年齢、性別、ジェンダー、体格、血統、民族的背景または一般健康状態の個体の評価に基づいて誘導してもよい。一部の態様では、対照がアルゴリズムから誘導された値（1つまたは複数）であり得る。

30

【0107】

実験試料は、対象から採取された生物学的試料であり得る。実験試料は、健康状態が既知の対象から採取されたものであっても健康状態が不明の対象から採取されたものであってもよい。一部の態様では、対象の健康状態が、例えば実験試料の解析、生検、身体検査、検査所見、目視検査または遺伝子解析によって調べられ得る。実験試料によって調べられ得る対象の健康状態は、罹病状態、疾患リスクあり、または健常であり得る。

40

【0108】

製造物品

また、生物学的試料中、例えば便試料中の選択された便由来真核生物RNAバイオマーカーを検出および定量するためのキットを提供する。したがって、パッケージ製品（例えば、貯蔵、搬送または販売のための濃縮濃度または使用準備済濃度でパッケージングされた1種または複数種の本明細書に記載の組成物が内蔵された滅菌容器）およびキットもまた本発明の範囲内である。該製品は、1種または複数種の本発明の組成物が内蔵された容器

50

(例えば、バイアル、ビン、ボトル、バッグ、マイクロプレート、マイクロチップまたはビーズ)を含み得る。また、該製造物品に、例えば包装材、使用のための使用説明書、シリンジ、送達デバイス、バッファーまたは他の対照試薬をさらに含めてもよい。

【0109】

キットには、生物学的試料中の2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30種またはそれより多くの便由来真核生物RNAバイオマーカーに対応するRNA、例えば表1もしくは表2または表1と表2の組合せあるいは表3または表4から選択される便由来真核生物RNAバイオマーカーに対応するRNAを検出することができる化合物または薬剤;および標準物質;ならびに任意で、検出、定量または増幅を行なうために必要な1種または複数種の試薬が含まれ得る。一部の態様では、キットに、生物学的試料中の2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30種またはそれより多くの便由来真核生物RNAバイオマーカーに対応するRNA、例えば表1もしくは表2または表1と表2の組合せあるいは表3または表4から選択される便由来真核生物RNAバイオマーカーに対応するRNAを検出することができる化合物または薬剤;および標準物質;ならびに任意で、検出、定量または増幅を行なうために必要な1種または複数種の試薬が含まれ得る。該化合物、薬剤および/または試薬は適当な容器内にパッケージングされ得る。キットに、核酸を検出および定量するために該キットを使用するための使用説明書をさらに含めてもよい。また、キットに、アッセイされ、内蔵された試験試料と比較され得る対照または一連の対照も含めてもよい。キットの各成分は個々の容器内に封入され得、種々の該容器のすべてが単一のパッケージ内に、該キットを用いて行なわれたアッセイの結果を解釈するための使用説明書とともに含められ得る。一部の態様では、キットに、1種または複数種の対照マーカーに特異的なプライマーまたはオリゴヌクレオチドプローブが含まれ得る。一部の態様では、キットに、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30種またはそれより多くの便由来真核生物RNAバイオマーカー、例えば表1もしくは表2または表1と表2の組合せあるいは表3または表4から選択される便由来真核生物RNAバイオマーカーの定量に特異的な試薬が含まれる。

【0110】

一部の態様では、キットに、細菌細胞および便中の他の成分からの真核生物細胞の分離用の、ならびに患者の便試料由来、例えばヒト患者の便試料由来の便由来真核生物RNAの抽出用の試薬が含まれ得る。したがって、キットには、バッファー、エマルジョンビーズ、シリカビーズ、安定化試薬ならびに種々のフィルターおよび遠心分離用容器が含まれ得る。また、キットには、試料の混入物を最小限にするため、および便試料中の便由来真核生物RNAの安定性を確実にするための便の取り扱いのための使用説明書も含められ得る。また、キットには、試料の保存を確実にするためのもの、例えば安定化バッファー、保冷剤またはヒートパックも含めてもよい。一部の態様では、キットに便収集デバイスが含まれ得る。

【0111】

また、該製品に、説明書き(例えば、印刷ラベルまたは添付文書または製品の使用が説明された他の媒体(例えば、音声-またはビデオテープまたはコンピュータ可読媒体))を含めてもよい。説明書きを容器と一体にしてもよく(例えば、容器に貼る)、試薬が使用され得る様式を記載してもよい。試薬は、使用可能状態であり得(例えば、適切なユニット内に存在する)、1種または複数種のさらなる佐剤、担体または他の希釈剤が挙げられ得る。あるいはまた、試薬濃縮形態で、希釈剤および希釈のための使用説明書とともに提供してもよい。

【実施例】

【0112】

実施例1: ヒト便試料の入手

ヒトの便の収集: 患者に、便座に取り付けたバケツ内に排便するように依頼し、得られ

10

20

30

40

50

た試料をフリーザー内に、ハリコフ (Kharkiv) 国立医科大学 (ハリコフ, ウクライナ) に輸送されるまで保存した。便を50 mL容コニカルチューブ内にアリコートに分け、-80℃で保存した。この試料をハリコフ国立医科大学からドライアイス下でCapital Biosciences (ゲーサーズバーグ, MD) に搬送し、即座に-80℃のフリーザーに移した。そこから、試料をドライアイス下でBioGenerator Labs (セントルイス, MO) に搬送し、ここで、試料を、抽出まで-80℃のフリーザー内に保存した。

【0113】

ヒト試料のタイプ: 便試料を、結腸直腸がん (ステージI~IV) を有する195人の患者、前がん性腺腫を有する126人の患者、良性ポリープを有する8人の患者および結腸内視鏡検査で陰性所見を有した125人の患者から入手し、454例の集合サンプルとした。健康個体は、結腸直腸がん、炎症性腸疾患、セリアック病、過敏性腸症候群、直近の20日以内の下痢または任意の他の胃腸疾患の病歴がない患者とした。良性ポリープ患者は結腸内視鏡検査を受ける前に便試料を提供し、この内視鏡検査時に医師が検出したポリープは、その後の生検および組織学的評価によって良性であると判断されたものであった。罹病状態の個体は、結腸直腸がんまたは前がん性腺腫を有すると診断された患者であった。結腸直腸がん患者は直近の1ヶ月以内の結腸内視鏡検査およびその後の生検によってステージI-ステージIVの結腸直腸がんを有すると診断されており、化学療法、放射線および/または手術が挙げられ得るなんらの生検後処置もまだ受けていなかった。前がん性腺腫の患者は結腸内視鏡検査を受ける前に便試料を提供し、この内視鏡検査時に医師が検出したポリープは、その後の生検および組織学的評価によって前がん性であると判断されたものであった。健康個体および良性ポリープ個体は、ジェンダーならびに年齢層 (50~60歳、60~70歳、70~80歳および80~90歳) に基づいて腺腫患者およびがん患者とマッチさせた。この収集に用いられた患者はCapital Biosciencesによる同意を承諾した。Schulman内部審査委員会がこの収集に関する倫理面の監視を提供した。

【0114】

実施例2: ヒト核酸の抽出

全核酸の抽出: 各便試料を50 mL容コニカルチューブ内に入れた。およそ1,000~25,000 mgの便を各チューブに添加した。さらに20~40 mLの溶液を各チューブに添加した。この溶液は、ハンクス平衡塩溶液 (HBSS) (Sigma-Aldrich) と0.05%のTween-20 (Sigma-Aldrich) および0.0002%のRNase Inhibitor (Sigma-Aldrich) との混合物を含有するものであった。便を溶液中に懸濁させ、およそ0~10℃で0~10分間回転させた。この溶液を1000 rpmで4℃にて10分間遠心分離し、上清みを廃棄した。およそ4~10 mLのEasyMag (登録商標) Lysis Buffer (bioMerieux) をベレットに添加し、このベレットを溶液中に再懸濁させた。この溶液を2500~3500 rpmで20~25℃にて10~15分間遠心分離した。分画遠心分離中、溶液は3層に分離した。底層には固形細胞内デブリが含まれており、中間層はヒト核酸が濃縮された親水性層であり、上層は疎水性の脂質層であった。上部の2つの層を新たな15 mL容コニカルチューブに移し、この溶液を2500 rpmで20~25℃にて10分間、再度遠心分離した。この遠心分離工程の結果、3層への分離がみられた: 底層は固形細胞内デブリであり、中間層はヒト核酸が濃縮された親水性層であり、上層は疎水性の脂質層であった。この溶液からの大型デブリをスクリーニングするため、20uL容ピペットチップを1mL容ピペットチップ上に配置し、2mLの親水性層を15mL容チューブからピペッティングし、EasyMag (登録商標) Disposableカートリッジ (bioMerieux) に移した。さらに、60 uLのEasyMag (登録商標) Magnetic Silica (bioMerieux) をカートリッジに添加した。ビーズを溶液中に0.5~1分間、ピペットを用いて混合した。ビーズに結合された核酸をバッファー溶液中に、Specific A Protocolを製造業者の指示書に従って用いて溶出させた。溶出された核酸の容量は70 uLであった。この核酸溶液を1.5mL容チューブ内にピペッティングし、氷上に置いた。次いで、先の工程で使用したのと同じEasyMag (登録商標) Disposableカートリッジ (bioMerieux) に、先で使用した15mL容チューブ内の同じ溶液のさらに2mLの親水性層を同じ手法を用いて再ロードし、大型デブリをスクリーニングして取り出した。さらに20 uLのEasyMag (登録商標) Magnetic Silica (bioMerieux) をカートリッジに

添加した。ビーズを溶液中に0.5～1分間、ピペットを用いて混合した。上記のとおり、ビーズに結合された核酸をバッファー溶液中に、Specific A Protocolを製造業者の指示書に従って用いて溶出させた。溶出された核酸の容量は70 uLであった。この核酸溶液を、既に最初の70 uLの溶出液が入っている元の1.5mL容チューブ内にピペッティングし、合わせた溶液を氷上に置いた。

【0115】

DNase処理: この140 uLの溶液を、Baseline-Zero-DNase (Epicenter) で35～40℃にて20～40分間処理した。EasyMag (登録商標) Lysis Bufferの1～2 mLのアリコートをし、DNaseで処理した溶液に添加し、試料を新たなEasyMag (登録商標) Disposableカートリッジに移した。全溶液を、この新たなカートリッジに、60 uLのEasyMag (登録商標) Magnetic Silicaとともに添加した。ビーズに結合された核酸をバッファー溶液中に、EasyMag (登録商標) Generic Protocolを製造業者の指示書に従って用いて溶出させた。溶出された核酸の容量は25 uLであった。この核酸溶液を1.5mL容チューブ内にピペッティングし、0～6℃で保存した。

【0116】

実施例3: ヒト便試料中のヒト核酸レベルの測定

抽出結果: 1～2 uLの上記で抽出した各試料を、全核酸およびRNA完全性について、Agilent 2100 Bioanalyzerを用いて評価した。この試料を定性的および定量的に解析した。電気泳動解析を用いて、抽出されたRNAの品質を確認した。電気泳動ファイルを、各試料のバンドをRNAラダーのサイズマーカーによって示されたバンドと比較し、18Sおよび28SのリボソームRNA (rRNA) バンドを確認することにより読み取った。このrRNAバンドは、標準化ラダーにおける2,000ヌクレオチドマーカー付近の2つの太い顕著なバンドである。定性的には、十分なバンドイングおよび濃いバンド強度により、さらなる解析、例えばマイクロアレイシーケンシング、ポリメラーゼ連鎖反応 (PCR)、核酸シーケンシング、分子バーコーディングまたはプローブ-キャプチャーに十分なインタクトな核酸が入手可能であることが示された。電気泳動分離図は、RNA完全性数値 (RIN)、全RNA質量および全rRNA質量の定量を有する各電気泳動ファイルのグラフ表示である。定量的には、RINが大きいほど、全RNA質量が多く、かつ、全rRNA質量が多いほど、試料が、さらなる解析、例えばマイクロアレイシーケンシング、ポリメラーゼ連鎖反応 (PCR)、核酸シーケンシング、分子バーコーディングまたはプローブ-キャプチャーに有用であろう可能性が高い。

【0117】

図1Aは、6つの試料での電気泳動ファイルおよび文献に記載された方法に基づいて抽出した1つの試料の電気泳動分離図である。図1Bは、6つの試料での電気泳動ファイルおよび上記のようにして抽出した1つの試料の電気泳動分離図である。上記のようにして抽出した試料の方が大きなRINおよび多くの真核生物質量がもたらされた。また、より明瞭なりボソームRNAバンド (18Sおよび28S)、ならびに18Sバンドより下のバンドイングが最小限であることによって明示されるより少ない細菌ノイズによって、上記のようにして抽出したseRNAの品質の方が高いことも示された。

【0118】

実施例4: 安定化バッファー中でのseRNAのインキュベーション

11例の試料を、安定化バッファーを用いた試験の実施に選択した。これらの試料を5グラムのアリコートに分割し、3つのコホート: コホート1 (n=11)、コホート2 (n=11) およびコホート3 (n=8) にした。コホート1の試料は、上記の方法を用いて即座に抽出した (図2A)。コホート2の試料は安定化バッファー中でインキュベートし、室温で24時間保存した後、上記の方法を用いて抽出した (図2B)。コホート3の試料は安定化バッファー中でインキュベートし、室温で48時間保存した後、上記の方法を用いて抽出した (図2C)。

【0119】

抽出後、すべての試料 (n=30) を、Agilent Bioanalyzerを用いて定性的に解析した。はっきりした明確なインタクトなりボソーム18Sおよび28Sバンドが、すべての試料の単離

されたRNAで検出される。真核生物RNAの量を推定するために使用され得る18Sおよび28Sの強度は、安定化バッファー中でのインキュベーションにより増大した。さらに、18Sより下のバンドニングによって示される細菌ノイズは、安定化バッファー中でのインキュベーションにより減少した。

【0120】

また、Agilent BioanalyzerによるデータでもRNA完全性数値（RIN）および真核生物質量の定量が可能であった。RINは、すべての試料で充分であった。各コホートの全RINは、安定化バッファー中でのインキュベーションにより増大し、平均RINはコホート1、コホート2およびコホート3で、それぞれ4.6、5.9および7.1であった。真核生物質量は、すべての試料において充分であった。全真核生物質量は、安定化バッファー中でのインキュベーションにより増大し、平均質量はコホート1、コホート2およびコホート3で、それぞれ11.1ng、39.7ngおよび78.4ngであった。

【0121】

実施例5:RNA転写物の解析

330例の試料を、Affymetrix GeneChip（商標）Human Transcriptome Array 2.0（Santa Clara, CA）を用いた解析に選出した。およそ100 ngのDNaseフリー糞便中RNAを、Affymetrix GeneChip（商標）Human Transcriptome Array 2.0との後続ハイブリダイゼーションを伴うAmbio WT-picoキットを製造業者のプロトコルのとおりに用いて増幅した。すべての試料を、Signal Space Transformation-Robust Multiarray Analysis（SST-RMA）を用いてAffymetrix Expression Console（商標）で正規化した。

【0122】

Affymetrix Microarray内の70,523の転写物クラスターのうち、3,977個の遺伝子に対応する5,149の転写物クラスターのサブセットを予備選択し、差次的発現について評価した。この初期選択によって偽発見率を低減させ、がんの発症および進行において既知の機能を有していない遺伝子をフィルタリングにより排除した。

【0123】

この330例の個体を265例の個体のトレーニングセットと65例の個体のテストセットに分割した。トレーニングセットは、差次的発現する遺伝子を同定するため、および計算モデルを構築するために使用し、一方、テストセットは、計算モデルの検出精度を調べるために使用した。標準的なLIMMAパッケージを使用し、前がん性腺腫またはCRCのいずれかを有する個体と結腸内視鏡検査で所見なしの個体との間で差次的発現されるRNA転写物クラスターのサブセットを同定した。すべてのバイオマーカーを、対数オッズスコアに従ってランク付けし、上位200位にランク付けされたバイオマーカー（ $p < 0.05$ ）を機械学習モデルの構築におけるフィーチャーとして使用した。RBFカーネルを伴うSupport Vector Machine Model（ ν -SVM）をモデル開発のために選出した。このカーネル関数により、フィーチャーを、明示的にコンピューティングされないより高次元の空間に拡張することによる個体間の距離の計算が可能である。SVMにより、ラベル群を分離する最大マージン超平面が見出される。パラメータ ν は、最大マージンを調べるために使用される個体の割合の下限を規定する。SVMモデルを、トレーニングセットの265例すべての個体に由来する200種の転写物の発現レベルを用いてトレーニングした。SVMの内部検証により0.776の総ROC AUCが得られた。このモデルでは、CRCおよび腺腫を評価した場合、それぞれ0.829および0.788のROC AUCが得られた（図3A）。

【0124】

また、このマルチターゲットRNAバイオマーカーアルゴリズムを、独立したテストセット内の65例の個体でも使用した。このモデルは、スクリーニング結腸内視鏡検査で陽性所見を有した全個体のうち79%（43例のうち34例）、前がん性腺腫を有する個体のうち95%、およびがんを有する個体のうち65%を正しく同定した。CRCでのモデル感度は、 $>4\text{cm}$ の直径の腫瘍のうち72%が正確に検出されたというように、サイズと直接的に相関した。腺腫でのサイズに対するモデル感度は不可知であり、小型（ $<5\text{mm}$ ）病変と大型（ $>1\text{cm}$ ）病変の両方で100%の予測精度であった（図3B）。

【0125】

実施例6:seRNA発現シグネチャーを用いたCRC分子サブタイピング

Affymetrix Microarray内の70,523の転写物クラスターのうち、274個の遺伝子に対応する転写物クラスターのサブセットを選択し、結腸直腸がんを有すると診断された個体由来の患者試料を、大腸癌サブタイピングコンソーシアム（CRCSC）によって規定されたCRCのコンセンサス分子サブタイプ（CMS）を用いてアノテーションした（図4A）。CRCSC分類器は、分子サブタイプクラス分類の精度を向上させる能力に関する各遺伝子の重要度に基づいて組織化されている。転写物クラスターの発現を遺伝子レベルで、各遺伝子と関連している転写物クラスターの発光量の中央値を用いてまとめた。遺伝子発現データを遺伝子レベルおよび全コホート間で、発現レベルの中央値を用いて正規化した。正規化データを、R Package CMS Classifierにおいて規定されたランダムフォレスト分類器への入力として使用し、コンセンサス分子サブタイプをラベリングした。

10

【0126】

CMS Classifierからの出力は4つの値を含み、各々は、試料がどのくらいCMS1~4と関連していそうかの事後確率である。CMS1は、高度のマイクロサテライト不安定性（MSI-H）および免疫浸潤と関連しているシグネチャーを有する腫瘍を含む。図4Bは、結腸直腸がんサブタイプCMS1の同定に有用な25個の例示的な結腸直腸新生物分子サブタイプバイオマーカー遺伝子を示す。CMS2~4は、それぞれ標準型、代謝型または間葉型の遺伝子発現シグネチャーと関連している。CMS分類器に基づき、117例のうち14例（12%）の個体がCMS1とクラス分類され、117例のうち100例（85%）がCMS2~4（標準型、代謝型および間葉型）とクラス分類され、117例のうち3例（3%）がCMS1/CMS2混合型とクラス分類された（図5）。

20

【0127】

実施例7:ヒト便試料の入手、抽出および測定

ヒトの便の収集:患者に、便座に取り付けたバケツ内に排便するように依頼し、得られた試料をクーリエ便に引き渡し、ワシントン大学医学部の消化器疾患研究コアセンター（Digestive Diseases Research Core Center）（セントルイス,ミズーリ州）に輸送した。便を50 mL容コニカルチューブ内にアリコートに分け、-80℃で保存した。そこから、試料をドライアイス下でBioGenerator Labs（セントルイス,MO）に輸送し、ここで、試料を、抽出まで-80℃のフリーザー内に保存した。この収集に用いられた患者はワシントン大学医学部による同意を承諾した。また、ワシントン大学医学部の内部審査委員会がこの収集に関する倫理面の監視を提供した。

30

【0128】

ヒト試料のタイプ:便試料を、結腸直腸がん（ステージI~IV）を有する6人の患者、前がん性腺腫を有する4人の患者および結腸内視鏡検査で陰性所見を有した14人の患者から入手し、24例の集合サンプルとした。これらの試料は、ハリコフ国立医科大学およびワシントン大学医学部の両者のHuman Stool Collection由来のものであった。試料ラベルを確認し、先のヒト試料のタイプで概要を示した基準と整合する様式でマッチさせた。

【0129】

全核酸の抽出:seRNAを試料から、先のDNase処理を含む全核酸の抽出で概要を示した方法と整合する様式で抽出し、seRNAの品質を、抽出結果で概要を示した方法と整合する様式で解析した。

40

【0130】

実施例8:RNA転写物の解析

ライブラリーの調製:seRNAのライブラリーを、398種のカスタムアンプリコンからなるIllumina Targeted RNA Custom Panelを用いて作製した。ライブラリーの調製は、ProtoScript II Reverse Transcriptase（Illumina）を用いた最初のcDNA合成工程、オリゴプールの標的化対象seRNAとのハイブリダイゼーション工程、Illumina試薬（AM1、ELM4、RSB、UB1）を用いたオリゴ体の伸長工程、およびポリメラーゼ連鎖反応（PCR）による増幅工程に依存した。全質量入力は200~400ngの範囲であり、使用したPCRサイクルの回数は26

50

～28回の範囲であった。ライブラリーの増幅後、cDNA捕捉物を、Illumina試薬（RSB, AMPure, XPビーズ EtOH）を用いて洗浄した。ライブラリー調製物を量および品質について、Agilent BioAnalyzerおよびQubit Fluorometric Quantitation（Thermo Fisher）を用いて解析した。この解析で記載した試料はすべて、初期品質チェックに合格し、次世代シーケンシングの候補であった。

【0131】

シーケンシング: 個々の試料に対してユニークな指標を使用し、ライブラリー調製物のプーリングおよびIllumina NextSeq Systemの同じフローセルでのすべての試料のマルチプレックス化を可能にした。24例すべての試料を1つのレーンに中出力フローセル（Illumina）でプールした。リードの各末端の最初の150塩基対をシーケンシングし（2×150）、シーケンシングされたリードを付加し、FASTQファイルを出力した。FASTQファイルの品質チェックにより、19例の試料がバイオインフォマティクス解析に十分な総リードおよび十分な品質を有することが示された。

10

【0132】

アラインメント: シーケンシング後、配列からカスタムプライマー配列をトリミングし、トリミングしたリードを最新の参照ゲノム（GRCh38）に対してアラインメントした。転写物の発現量を、遺伝子座における平均カバー率を算出することにより取得した。転写物の発現量を、2つのハウスキーピング遺伝子（GAPDHおよびACTB）での平均カバー率によって正規化した。

20

【0133】

実施例9: 種々のプラットフォームでの生物学的反復実施

4つの試料をマイクロアレイおよびシーケンシングの両方において評価した。複数のプラットフォームでの398種の転写物の線形回帰により、中程度の再現性（ピアソンの r の範囲=0.48～0.63）が示された。シーケンシングでは、低発光の転写物のシグナル範囲によって明示されるように、マイクロアレイと比べて増加した分解度が示された（図6）。

【0134】

実施例10: seRNAを用いた階層的クラスタリング解析

教師なし主成分分析（PCA）を、13例すべてのユニーク試料のRNAシーケンシングデータに対して行なった。クラスタリングは、CRCを有する患者、腺腫を有する患者および新生物所見のない患者の間で観察された。がんを有する患者由来の試料で、最も大きな偏差および他の患者集団との分離が示され、一方、新生物所見のない患者由来の試料で、より狭いクラスタリングが示された（図7）。

30

【0135】

実施例11: seRNAを用いたシーケンシングでのバリエーションの評価

バリエーションコーリングおよびアノテーション: Integrative Genomics Viewerを使用し、CRC腫瘍形成に参与するバリエーションを同定した。アンプリコンパネルは、398個の捕捉遺伝子のゲノム空間の約3%をカバーした。例示的なドライバー突然変異を図9に示す。図8に示すように、本発明者らは、いくつかの潜在的ドライバー突然変異を同定した。このような突然変異には、高リスク腺腫を有する患者でのAPCにおけるミスセンス突然変異（13%のバリエーションアレル頻度（VAF））、高リスク腺腫を有する患者でのSMAD4におけるミスセンス突然変異（17%のVAF）、ステージIのCRCを有する患者でのMAPK3の調節領域における3'欠失（7%のVAF）、結腸内視鏡検査で所見なしの患者でのPIK3CAにおけるミスセンス突然変異（12%のVAF）、高リスク腺腫を有する患者でのKRASにおけるミスセンス突然変異（3%のVAF）、および高リスク腺腫を有する患者でのCDH1におけるミスセンス突然変異（2%のVAF）が含まれた（図8）。

40

【0136】

実施例12: ヒト便試料の入手

ヒトの便の収集: 便試料を、ワシントン大学医学部の消化器疾患研究コアセンター（DDRCC）（St. Louis, MO）で入手した。すべての患者に便試料収集キットが郵便で送付され、キットはクーリエ便によってDDRCCに送り返された。臨床データ（例えば、個体群統計情

50

報、結腸内視鏡検査結果など)はDDRCCによって収集された。各試料を便中血液について市販の免疫化学的便検査(FIT)(Polymedco,OC-Light S FIT)を用いて試験した後、-80℃で凍結させた。この試験のために集められた各患者には結腸内視鏡検査が行なわれ、陽性所見を有した患者には生検およびその後、新生物のクラス分類を決定するための組織病理学的審査が行なわれた。腺腫のクラス分類は、組織病理学検査(良性に対して前がん性)、ポリープの数、ポリープのサイズおよび分化度に基づいて階層化した。がんのクラス分類は、American Joint Committee on Cancer(AJCC)7 TNMシステムに基づいて階層化した。患者が結腸内視鏡検査において所見なしであった場合は健常とラベリングした。

【0137】

ヒト試料のタイプ:合計で、275例の個体由来の便試料をこの試験のために収集した。シーケンシングデータ、FIT、個体群統計情報(すなわち、性別、年齢、民族、喫煙状態および家族歴)ならびに組織病理学検査情報を伴う結腸内視鏡検査結果は、該当する場合、すべての患者について取得した。この試験において、11人の患者がCRC(ステージI~IV)を有し、26人の患者が高リスク腺腫(HRA)を有し、37人の患者が中リスク腺腫(MRA)を有し、61人の患者が低リスク腺腫(LRA)を有し、50人の患者が良性ポリープを有し、90人の患者が結腸内視鏡検査で所見なしであった。患者の型、個体群統計および加工処理情報を図10にまとめる。健常個体は、結腸内視鏡検査で所見なしであり、結腸直腸がん、炎症性腸疾患、セリアック病、過敏性腸症候群、直近の20日以内の下痢または任意の他の胃腸疾患の病歴がない患者であった。良性ポリープ患者は結腸内視鏡検査を受ける前に便試料を提供し、この内視鏡検査時に医師が検出したポリープは、その後の生検および組織学的評価によって良性であると判断されたものであった。罹病状態の個体は、結腸直腸がんまたは前がん性腺腫を有すると診断された患者であった。結腸直腸がん患者は直近の1ヶ月以内の結腸内視鏡検査およびその後の生検によってステージI-ステージIVの結腸直腸がんを有すると診断されており、化学療法、放射線および/または手術が挙げられ得るなんらの生検後処置もまだ受けていなかった。前がん性腺腫の患者(HRA、MRAおよびLRA)は結腸内視鏡検査を受ける前に便試料を提供し、この内視鏡検査時に医師が検出したポリープは、その後の生検および組織学的評価によって前がん性であると判断されたものであった。腺腫リスクの階層化は、ポリープのサイズ、ポリープの数、異形成の程度および細胞の形態構造に基づいた。患者集団の結腸直腸がん患者を多くしたが、残りの試料は無症候スクリーニング集団に代表的であった。この収集に用いられた患者はワシントン大学医学部による同意を承諾した。ワシントン大学医学部の内部審査委員会がこの収集に関する倫理面の監視を提供した(IRB #20111107)。

【0138】

トレーニングセットとテストセットへの分離:154例の前向き収集便試料をトレーニングセットとして使用し、110例の前向き収集便試料をホールドアウトテストセットとして使用した。また、CRC患者由来の11例の後向き収集便試料もホールドアウトテストセットに含めた。トレーニングセットおよびホールドアウトテストセットをカテゴリー差、個体群統計差および取り扱いの差についてt検定(母集団の平均)またはz検定(母集団の度数)を用いて評価し、p値が0.05未満である場合、有意性を示すとした。トレーニングセットの特徴とホールドアウトテストセットの特徴との間に2つの統計学的有意差がみられた。第1に、トレーニングセットには後向き収集試料(すなわち、CRCを有する患者由来の試料)を含めなかった。第2に、ホールドアウトテストセットは、トレーニングセットと比べて異なる加工処理品質を有した。具体的には、便由来真核生物RNA抽出に使用した平均便入力量の低減がみられ(12.9グラム対12.0グラム;p値=0.03)、平均便由来真核生物RNA濃度の低減がみられ(168.6 ng/uL対56.1 ng/uL;p値<0.01)、平均ライブラリー調製物断片サイズの低減がみられた(200.6塩基対192.2塩基対;p値<0.01)(図10)。

【0139】

実施例13:カスタムキャプチャーパネルの開発

パネル転写物:639種のアンプリコンのカスタムキャプチャーパネルをライブラリー調製のためにIllumina DesignStudioにおいて開発した。カスタムキャプチャープローブは、

以前に実施された研究および文献を用いて選択した408種の転写物と関連していた。

【0140】

マイクロアレイ転写物: 転写物をマイクロアレイ実験に基づいて選択した。この実験では、全seRNAを便試料から抽出し、発現を、Affymetrix Human Transcriptome Array 2.0 (Thermo Fisher Scientific, Waltham, MA) を用いて評価した。CRCまたは前がん性腺腫を有する177人の患者（罹病状態コホート）で得られたマイクロアレイ発現プロファイルを、結腸内視鏡検査で所見なしの88人の患者（健常コホート）の発現プロファイルと比較した。214種の転写物が差次的発現されていると同定され ($p < 0.03$)、キャプチャーパネルに選択した。

【0141】

NanoString転写物: 転写物をNanoString実験に基づいて選択した。この実験では、全seRNAを便試料から抽出し、発現を、nCounter（登録商標）PanCancer Pathways Panel (NanoString, Seattle, WA) およびnCounter（登録商標）PanCancer Progression Panel (NanoString, Seattle, WA) を用いて評価した。CRCまたは前がん性腺腫を有する59人の患者（罹病状態コホート）で得られたNanoString発現プロファイルを、結腸内視鏡検査で所見なしの26人の患者（健常コホート）の発現プロファイルと比較した。123種の転写物が差次的発現されていると同定され、キャプチャーパネルに選択した。

【0142】

他の転写物: 文献を、CRCに関与しているさらなる転写物について評価した。これには、GeneCards、ClinVar、Catalogue of Somatic Mutations in Cancer (COSMIC)、Clinical Interpretations of Variants in Cancer (CIViC)、大腸癌サブタイピングコンソーシアム分類器および他の直接関係のある研究の検索を含めた。これらの文献を用いて、71種の転写物をカスタムキャプチャーパネルに選択した。

【0143】

実施例14: ヒト核酸の抽出

全核酸の抽出: 各便試料を50 mL容コニカルチューブ内に入れた。およそ6,000~25,000 mgの便を各チューブに添加した。さらに20~40 mLの溶液を各チューブに添加した。この溶液は、10mMのトリズマ塩基 (Sigma-Aldrich, St. Louis, MO)、0.05%のTween-20 (Sigma-Aldrich) を含む1mM EDTA (Sigma Aldrich) および0.0002%のRNase Inhibitor (Sigma-Aldrich) の混合物を含有するpH7.5のものであった。この溶液を1000 rpmで4℃にて10分間遠心分離し、上清みを廃棄した。およそ4~10 mLのEasyMag（登録商標）Lysis Buffer (bioMerieux, Durham, NC) をペレットに添加し、このペレットを溶液中に再懸濁させた。この溶液を2500~3500 rpmで20~25℃にて10~15分間遠心分離した。分画遠心分離中、溶液は3層に分離した。底層には固形細胞内デブリが含まれており、中間層はヒト核酸が濃縮された親水性層であり、上層は疎水性の脂質層であった。上部の2つの層を新たな15 mL容コニカルチューブに移し、この溶液を2500 rpmで20~25℃にて15分間、再度遠心分離した。この遠心分離工程の結果、3層への分離がみられた: 底層は固形細胞内デブリであり、中間層はヒト核酸が濃縮された親水性層であり、上層は疎水性の脂質層であった。この溶液からの大型デブリをスクリーニングするため、10 uL容ピペットチップを1 mL容ピペットチップ上に配置し、2 mLの親水性層を15 mL容チューブからピペッティングし、EasyMag（登録商標）Disposableカートリッジ (bioMerieux) に移した。さらに、50 uLのEasyMag（登録商標）Magnetic Silica (bioMerieux) をカートリッジに添加した。ビーズを溶液中に0.5~1分間、ピペットを用いて混合した。ビーズに結合された核酸をバッファー溶液中に、Specific A Protocolを製造業者の指示書に従って用いて溶出させた。溶出された核酸の容量は70 uLであった。この核酸溶液を1.5 mL容チューブ内にピペッティングし、氷上に置いた。次いで、先の工程で使用したものと同一EasyMag（登録商標）Disposableカートリッジ (bioMerieux) に、先で使用した15 mL容チューブ内の同じ溶液のさらに2 mLの親水性層を同じ手法を用いて再ロードし、大型デブリをスクリーニングして取り出した。さらに20 uLのEasyMag（登録商標）Magnetic Silica (bioMerieux) をカートリッジに添加した。ビーズを溶液中に0.5~1分間、ピペットを用いて混合した。上記のとおり、

10

20

30

40

50

ビーズに結合された核酸をバッファー溶液中に、Specific A Protocolを製造業者の指示書に従って用いて溶出させた。溶出された核酸の容量は70 uLであった。この核酸溶液を、既に最初の70 uLの溶出液が入っている元の1.5 mL容チューブ内にピペッティングし、合わせた溶液を氷上に置いた。先で使用した同じ15 mLの溶液からのさらに2 mLの親水性層を、新たなEasyMag（登録商標）Disposableカートリッジ（bioMerieux）に同じ手法を用いて添加し、大型デブリをスクリーニングして取り出した。さらに、20 uLのEasyMag（登録商標）Magnetic Silica（bioMerieux）をカートリッジに添加した。ビーズを溶液中に0.5～1分間、ピペットを用いて混合した。ビーズに結合された核酸をバッファー溶液中に、Specific A Protocolを製造業者の指示書に従って用いて溶出させた。溶出された核酸の容量は70 uLであった。この核酸溶液を、最初の2回の溶出液が入っている1.5 mL容チューブ内にピペッティングし、合わせた溶液を氷上に置いた。次いで、先の工程で使用したのと同じEasyMag（登録商標）Disposableカートリッジ（bioMerieux）に、先で使用した15 mL容チューブ内の同じ溶液のさらに2 mLの親水性層を同じ手法を用いて再ロードし、大型デブリをスクリーニングして取り出した。さらに20 uLのEasyMag（登録商標）Magnetic Silica（bioMerieux）をカートリッジに添加した。ビーズを溶液中に0.5～1分間、ピペットを用いて混合した。上記のとおり、ビーズに結合された核酸をバッファー溶液中に、Specific A Protocolを製造業者の指示書に従って用いて溶出させた。溶出された核酸の容量は70 uLであった。この核酸溶液を、既に最初の3回の70 uLの溶出液が入っている元の1.5 mL容チューブ内にピペッティングし、合わせた溶液を氷上に置いた。

【0144】

DNase処理: この280 uLの溶液を、Baseline-Zero-DNase（Epicenter）で35～40℃にて20～40分間処理した。EasyMag（登録商標）Lysis Bufferの1～2 mLのアリコートをし、DNaseで処理した溶液に添加し、試料を新たなEasyMag（登録商標）Disposableカートリッジに移した。全溶液を、この新たなカートリッジに、85 uLのEasyMag（登録商標）Magnetic Silicaとともに添加した。ビーズに結合された核酸をバッファー溶液中に、EasyMag（登録商標）Generic Protocolを製造業者の指示書に従って用いて溶出させた。溶出された核酸の容量は25 uLであった。この核酸溶液を1.5 mL容チューブ内にピペッティングし、-80℃で保存した。

【0145】

実施例15: ヒト便試料中のヒト核酸レベルの測定

抽出結果: 1～2 uLの上記で抽出した各試料を、全核酸およびRNA完全性について、Agilent 2100 Bioanalyzerを用いて評価した。この試料を定性的および定量的に解析した。電気泳動解析を用いて、抽出されたRNAの品質を確認した。電気泳動ファイルを、各試料のバンドをRNAラダーのサイズマーカーによって示されたバンドと比較し、18Sおよび28SのリボソームRNA（rRNA）バンドを確認することにより読み取った。このrRNAバンドは、標準化ラダーにおける2,000ヌクレオチドマーカー付近の2つの太い顕著なバンドである。定性的には、十分なバンドングおよび濃いバンド強度により、さらなる解析、例えばマイクロアレイシーケンシング、ポリメラーゼ連鎖反応（PCR）、核酸シーケンシング、分子バーコーディング、アンプリコンシーケンシングまたはプローブ-キャプチャーに十分なインタクトな核酸が入手可能であることが示された。電気泳動分離図は、RNA完全性数値（RIN）、全RNA質量および全rRNA質量の定量を有する各電気泳動ファイルのグラフ表示である。定量的には、RINが大きいほど、全RNA質量が多く、かつ、全rRNA質量が多いほど、試料が、さらなる解析、例えばマイクロアレイシーケンシング、ポリメラーゼ連鎖反応（PCR）、核酸シーケンシング、分子バーコーディング、アンプリコンシーケンシングまたはプローブ-キャプチャーに有用であろう可能性が高い。また、試料をRNA濃度についてもQubit 4.0 Fluorometerを用いて評価した。RNA濃度は、溶出液中に存在するRNA選択的に結合するQubitアッセイ内の構成要素によって生成される蛍光の定量によって求められる。定量的には、RNA濃度が高いほど、試料が、さらなる解析、例えばマイクロアレイシーケンシング、ポリメラーゼ連鎖反応（PCR）、核酸シーケンシング、分子バーコーディング、アンプリコンシーケンシングまたはプローブ-キャプチャーに有用であろう可能性が

高い。

【0146】

実施例16:RNA転写物の解析

ライブラリーの調製:seRNAのライブラリーを、639種のカスタムアンプリコンからなるIllumina Targeted RNA Custom Panelを用いて作製した。ライブラリーの調製は、ProtoScript II Reverse Transcriptase (Illumina, San Diego, CA) を用いた最初のcDNA合成工程、オリゴプールの標的化対象seRNAとのハイブリダイゼーション工程、Illumina試薬 (AM1、ELM4、RSB、UB1) を用いたオリゴ体の伸長工程、およびポリメラーゼ連鎖反応 (PCR) による増幅工程に依存した。全質量入力は200~400 ngの範囲であり、使用したPCRサイクルの回数は28回~30回の範囲であった。ライブラリーの増幅後、cDNA捕捉物を、Illumina試薬 (RSB, AMPure, XPビーズ EtOH) を用いて洗浄した。ライブラリー調製物を量および品質について、Agilent 2100 BioAnalyzerおよびQubit 4.0 Fluorometer (Thermo Fisher) を用いて解析した。この解析で記載した試料はすべて、初期品質チェックに合格し、下流解析の候補であった。

10

【0147】

シーケンシング解析:個々の試料に対してユニークな指標を使用し、ライブラリー調製物のプーリングおよびIllumina NextSeq 550 Systemのフローセルでの試料のマルチプレックス化を可能にした。PhiXスパイクインを品質管理のために使用した。275例の試料を8つの個々の高出力フローセルランにおいてプーリングした (Illumina)。リードの各末端の150個までの塩基対をシーケンシングし (2×150)、シーケンシングされたリードを付加し、FASTQファイルを出力した。FASTQファイルの品質チェックにより、275例すべての試料がバイオインフォマティクス解析に十分な総リード (>100,000) および十分な品質を有することが示された。

20

【0148】

アラインメント:シーケンシング後、リードからカスタムプライマー配列をトリミングし、最新の参照ゲノム (GRCh38) に対してHISAT2.3.0によってアラインメントした。転写物の発現量を、遺伝子座における平均カバー率を算出することにより取得した。各転写物について、未加工アンプリコン発現量を、報告された発現が、マッピングされたGAPDHリード100万個あたりのアンプリコンのリード数と等しくなるように内部ハウスキーピング遺伝子であるGAPDHに対して正規化した。

30

【0149】

転写物の選択:639種のアンプリコンの正規化発現を、トレーニングセットのすべての試料 (n=154例の試料) について評価した。これらの639種のアンプリコンのうち、48種のアンプリコンはいずれの試料においても発現されず、さらに71種のアンプリコンはすべての試料の>95%において発現されなかった;これらのアンプリコンは解析から除外した。残りのアンプリコンについて、トレーニングセットを100の異なる9:1スプリットに分割することによってブートストラップ解析を行ない、それにより、各スプリットを情報提供的アンプリコンについて評価した。アンプリコンは、絶対log₂倍数変化が両対比群 (HRA対LRA、良性ポリープ、結腸内視鏡検査で所見なし;MRA対LRA、良性ポリープ、結腸内視鏡検査で所見なし) において1より大きい場合、および対比群間のANOVAがp値<0.05を有する場合、情報提供的とみなした。転写物選択プロセスを図11Aにさらに図示する。合計で、29個の遺伝子由来の40種のアンプリコンが100のスプリットのうち少なくとも1つにおいて情報提供的であると同定された (図11B)。アンプリコンが全ブートストラップスプリットの少なくとも33%において情報提供的であると判断された場合、これは差次的発現されているとみなされ、モデル開発のためのフィーチャーの候補にした。差次的発現されていると同定されたアンプリコンが10種類あった (100のスプリットのうち少なくとも33のスプリットにおいて情報提供的) (図11B)。未加工GAPDH値を各試料中の全真核生物RNAの測定値とみなす。未加工GAPDH値は、MRA、HRAおよびCRCを有する患者では健常患者と比べて高いことが観察された (図12)。また、個体群統計フィーチャー (年齢、喫煙状態、これまでの家族歴、民族およびジェンダー) もモデル開発に考慮した。最終的に、10種の差

40

50

次的発現転写物、未加工GAPDH値ならびに2つの個体群統計識別子（年齢および喫煙状態）をモデル開発のためのフィーチャーの候補とした。

【0150】

実施例17:ランダムフォレストモデルの開発

ランダムフォレストモデルを、154人の患者のトレーニングセットと13個の全候補フィーチャーを用いて構築した。5,000の決定木をブートストラップトレーニング試料から構築し;各ノードスプリットをGini重要度によって最適化し;各ツリーは最大深さに到達するまで構築した。具体的な態様を本明細書で論考しているが、より多くの数および/またはより少ない数の決定木、より多くの数および/またはより少ない数の候補フィーチャーなどを用いる任意の適当なモデル、例えばランダムフォレストモデルが作成され得ることは認識されよう。さらに、他の型のモデル、例えばディープラーニングモデルまたはサポートベクターモデルも、多様なパラメータを用いて使用され得よう。このランダムフォレストモデルでは差次的発現転写物、未加工GAPDH値、年齢および喫煙状態などの候補フィーチャーを使用した。具体的な態様を本明細書で論考しているが、すべての情報提供的フィーチャーおよび/または情報提供的フィーチャーの選択されたサブセットを用いる任意の適当なモデル、例えばランダムフォレストモデルが作成され得ることは認識されよう。

【0151】

このモデルによる出力を、大きな数ほど新生物所見または陽性所見における高い信頼度を反映する0~1の予測が得られるように構成した。一部の態様では新生物所見または陽性所見における信頼度を改変するために免疫化学的便検査（FIT）を使用した。例えば、FIT陽性試料では、予測スコアは1まで高くなるであろう。3分割内部交差検証を使用し、トレーニングモデルのパフォーマンスを評価した。3分割内部交差検証では3つの異なる2:1スプリットを使用し、それにより、大きい方のスプリットを用いてモデルを構築し、小さい方のスプリットにおいて使用した。モデルの予測値を用いて受信者操作特性（ROC）曲線を作成し、曲線下面積（AUC）を用いてモデルのパフォーマンスを測定した。3つのスプリットのROC曲線の中央値を使用し、交差検証パフォーマンスを近似した。ROC曲線をFITフィーチャーの組込みありおよびなしでプロットした。FITフィーチャーありでプロットしたROC曲線では、FIT強制陽性モデル予測値は1に等しい。示した本実施例において、FITフィーチャーなしでの内部交差検証では、HRA対他のすべてのカテゴリー（MRA、LRA、良性ポリープおよび結腸内視鏡検査で所見なし）で0.65のROC AUCが得られた。示した本実施例において、FITフィーチャーありでの内部交差検証では、HRA対他のすべてのカテゴリー（MRA、LRA、良性ポリープおよび結腸内視鏡検査で所見なし）で0.70のROC AUCが得られた（図13）。

【0152】

ホールドアウトテストセット:最終のランダムフォレストモデルを、トレーニングセット内の154例すべての試料を用いて構築した。作成されたモデルについて、Gini重要度によって測定された最も影響力のあるフィーチャーはACY1およびTNFRSF10Bであり（Gini重要度 ≥ 0.13 ）、最も重要度が低いフィーチャーはPER3であった（Gini重要度 < 0.05 ）。未加工GAPDH値は、ランダムフォレストモデルの構築における重要度が4位のフィーチャーであった（図14）。このモデルをホールドアウトテストセットの110例の前向き収集便試料において使用した。ROC曲線をFITフィーチャーありおよびなしでプロットし、曲線下面積（AUC）を用いてモデルのパフォーマンスを測定した。このモデルでは、FITフィーチャーなしで0.67のROC AUCおよびFITフィーチャーありで0.78のROC AUCが得られた（図15）。

【0153】

モデル予測値:ホールドアウトテストセットにおけるモデル予測値を疾患の重症度と関連させた（図16）。疾患の重症度とのモデル出力の相関は生物学の直接的反映であり、モデルの一部として特別にトレーニングしなかった。前述の態様では、フィーチャー選択およびモデル入力に3つのカテゴリー（HRA、MRAおよびその他すべて）の使用を含めたが、疾患サブタイプ（例えば、HRAのサブセット）および疾患の程度（例えば、HRAはMRAより重度である）はモデルトレーニングのフィーチャーとして使用しなかった。モデル出力が

疾患の重症度と相関していることを考慮すると、これにより、モデル出力と用いた特定のサブタイプおよび疾患の重症度の前向き特定が可能となる。さらに、モデルに疾患の重症度情報を提供するためにモデルのパラメータを変更することにより陽性所見と陰性所見の階層化が改善した。

【0154】

ダウンサンプリング解析:モデルトレーニングの程度を理解するため、トレーニングセットの154例の試料のうちのダウンサンプリング分を選択し、ホールドアウトテストセットを用いてパフォーマンスを評価した。ダウンサンプリング分は10%増分で30%~100%の範囲であった。各ダウンサンプリング分について、ブートストラップ法を用いてフィーチャー選択を行ない、候補フィーチャーを用いてランダムフォレストモデルをトレーニングし、モデルをホールドアウトテストセットに対して使用した。ホールドアウトテストセットのROC AUCを用いてモデルのパフォーマンスを評価した。このプロセスを各ダウンサンプリング分について10回繰り返してサブサンプリングにおける選択バイアスを低減させ、モデルのパフォーマンスをFITフィーチャーの組込みありおよびなしで評価した。ダウンサンプリング解析により、トレーニングに使用した試料の総数とホールドアウトテストセットにおけるパフォーマンス間の直接的な関係が示された。FITフィーチャーを除外した場合、HRA対他のすべてのカテゴリーのROC AUCの中央値は0.55（トレーニングデータの30%）から0.67（トレーニングデータの100%）に増大した（図17A）。FITフィーチャーを含めた場合、HRA対他のすべてのカテゴリーのROC AUCの中央値は0.72（トレーニングデータの30%）から0.78（トレーニングデータの100%）に増大した（図17B）。

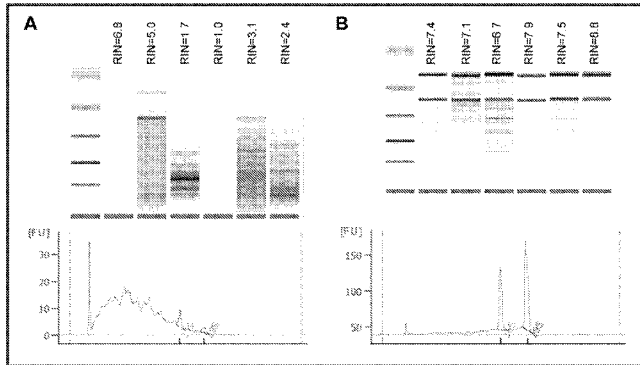
【0155】

最終精度:上記の態様を継続し、ランダムフォレストモデルを、CRC患者由来の11例の後向き収集便試料に対しても使用した。このモデルによる出力では、0~1の予測および1に等しいFIT強制陽性モデル予測値が示された。免疫化学的便検査陽性（FIT+）または陽性モデル予測値（Model+）を有する試料を陽性とみなし、他のすべての試料を陰性とみなした。CRC試料のみを陽性とみなし、他のカテゴリー（HRA、MRA、LRA、良性ポリープおよび結腸内視鏡検査で所見なし）を陰性とみなすROC曲線をプロットした。この補完ホールドアウトテストセットの121例すべての試料を使用すると、このモデルで0.94のROC AUCが得られた。CRCおよびHRAの試料を陽性とみなし、他のすべてのカテゴリー（MRA、LRA、良性ポリープおよび結腸内視鏡検査で所見なし）を陰性とみなす別個のROC曲線をプロットした。この補完ホールドアウトテストセットの121例すべての試料を使用すると、このモデルで0.87のROC AUCが得られた（図18）。ROC曲線上の精度が最大となる点を選択し、感度と特異性を算出した。この点において、モデルは89%の特異性（n=99例の試料）でCRCでは91%の感度（n=11例の試料）およびHRAでは73%の感度（n=11例の試料）を示した（図18）。

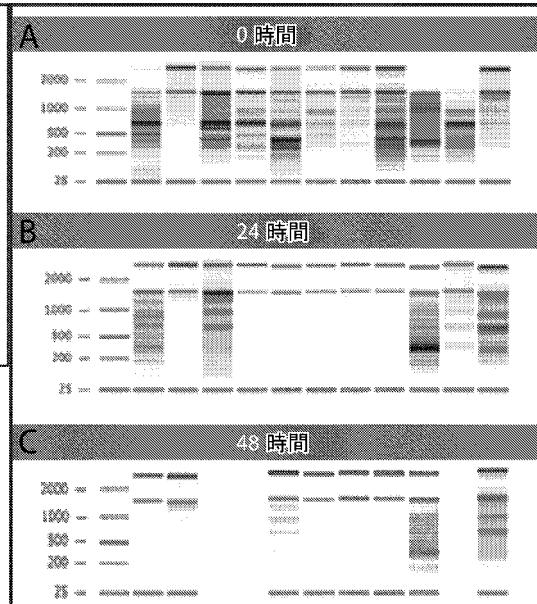
【0156】

スクリーニング集団に対する外挿:よりよい近似の究極のモデルのパフォーマンスを得るため、補完ホールドアウトテストセットで観察された精度プロファイルを前向きスクリーニング集団で予測された相対度数に外挿した。上記のROC曲線をプロットし、モデルのパフォーマンスを示した。前向きスクリーニング集団における予測罹病率に対してがんおよびHRAを加重した場合、このモデルにおいてCRCおよびHRAの試料対他のすべてのカテゴリーで0.80のROC AUCが得られた（図19）。また、前向きスクリーニング集団への結果の外挿により、新生物所見、陰性的中率（NPV）および陽性的中率（PPV）の混合感度の計算も可能になる。この外挿精度プロファイルでは、CRCとHRAの混合感度74%、陽性的中率37%および陰性的中率98%が示された（図19）。

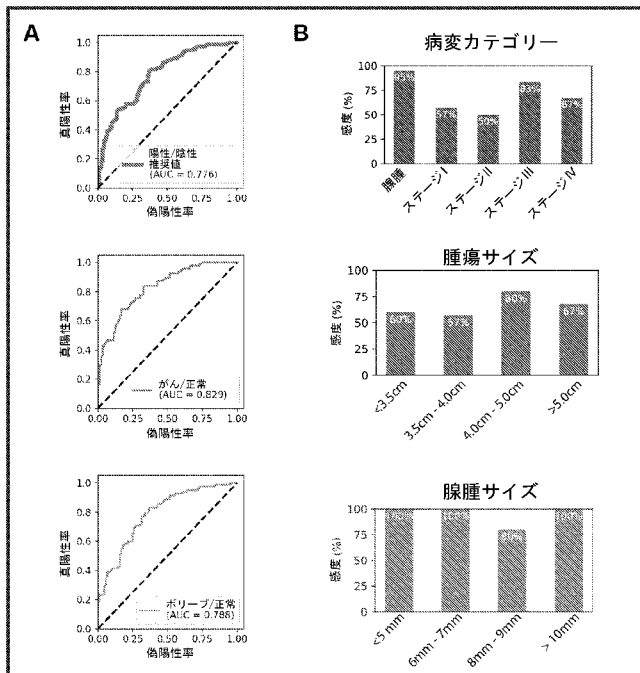
【図 1】



【図 2】



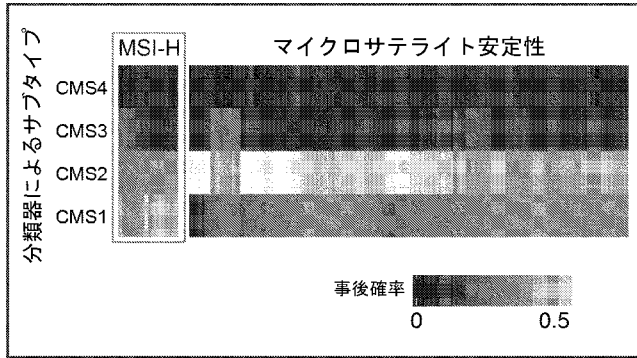
【図 3】



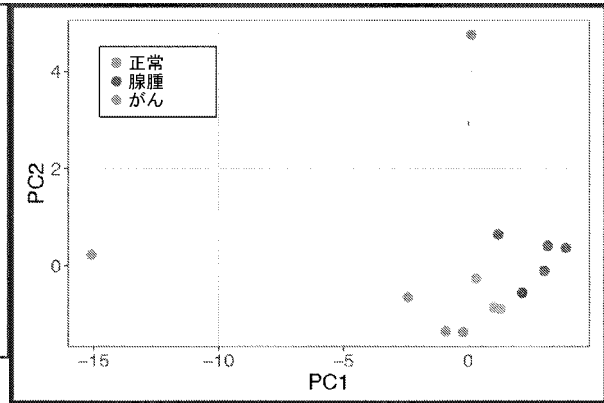
【図 4】



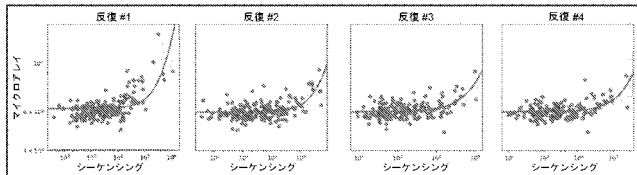
【図 5】



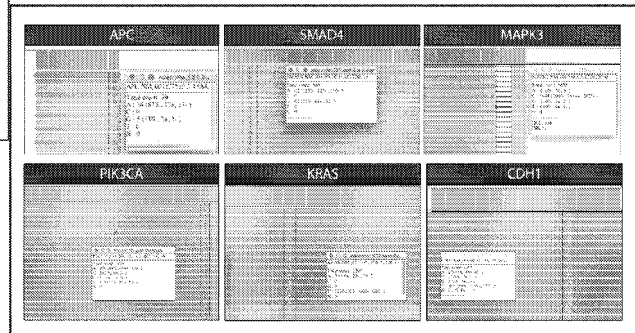
【図 7】



【図 6】



【図 8】



【図 9】

APC	SMAD4	CDKN2A	BRAF
KRAS	MLH1	CDH1	MAKP3
TP53	CTNNB1	PIK3CA	NRAS
BMP3	EGFR	PTEN	
NDRG4	BRCA1	VEGFA	

【図 10】

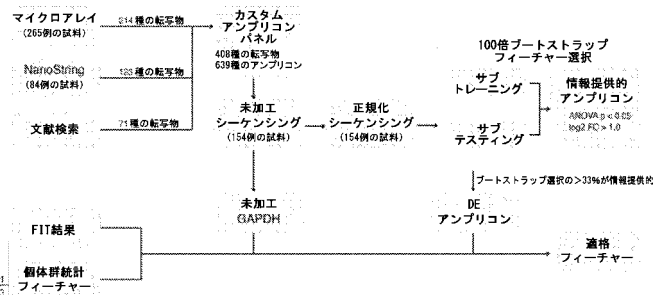
カテゴリ	記	前向き トレーニング セット	前向き ホールドアウト テストセット	後向き ホールドアウト テストセット	合計
がん (13)	ステージI~IV結腸直腸がん	0	0	11	11
高リスク腺腫 (2)	上皮下/高度異形成を伴う腺腫、任意のサイズ	1	2	0	3
高リスク腺腫 (2)	腺腫、種々の増殖パターン (≥25%)、任意のサイズ	4	2	0	6
高リスク腺腫 (2)	腺腫、≥1.0cmのサイズ	5	5	0	10
高リスク腺腫 (4)	腺腫状病変、≥1.0cmのサイズ	5	4	0	9
中リスク腺腫 (3)	1つまたは2つの腺腫、≥5mmのサイズかつ<1.0cmのサイズ、非進行型	11	9	0	20
中リスク腺腫 (4)	2つ以上の腺腫、<1.0cmのサイズ、非進行型	19	7	0	26
低リスク腺腫 (4)	1つまたは2つの腺腫、≤5mmのサイズ、非進行型	27	24	0	51
良性ポリープ (6)	過形成性、良性ポリープ (1つまたは複数)	25	24	0	50
陰性 (5)	結腸内視鏡検査で所見なし	55	35	0	90
合計		154	110	11	275

カテゴリ	記	トレーニング セット	完全ホールドアウト テストセット	p値*
性別 (女性対男性)		63.0%	62.0%	0.87
年齢		52.9±6.4	54.3±7.4	0.97
民族的背景 (白人対その他)		54.1%	44.6%	0.10
喫煙状態 (喫煙未経験者対その他)		70.1%	66.9%	0.57
家族歴 (腸癌家族歴対その他)		76.1%	79.9%	0.29

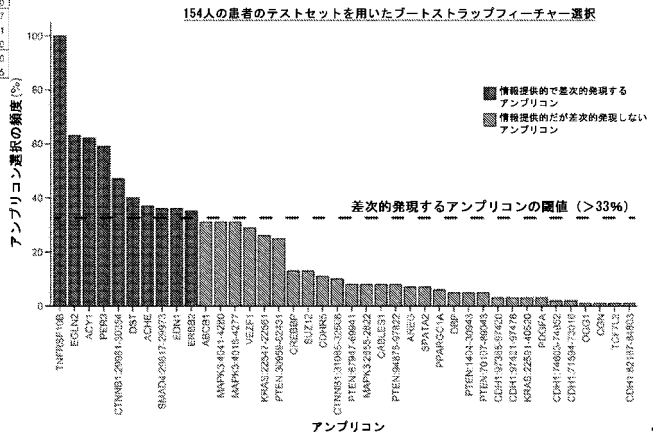
カテゴリ	記	トレーニング セット	完全ホールドアウト テストセット	p値*
平均便量 (グラム)		122.9±3.2	122.9±4.4	0.90
平均RNA安全性値 (RIN)		4.6±1.2	4.4±1.6	0.18
平均RNA濃度 - Oubit (ng/μL)		153.6±139.9	56.1±79.3	<0.01
平均断片サイズ (bp)		200.6±9.6	192.2±16.8	<0.01
全ライブラリ量 (ng)		29.7±29.5	32.9±4.1	0.50

* 頻度の有意な変化を両側カイニ乗検定を用いて計算した
* 平均の有意な変化を両側t検定を用いて計算した

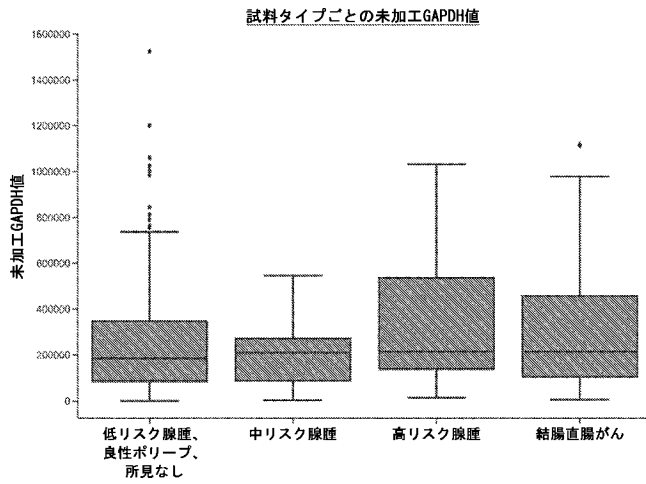
【図 11 A】



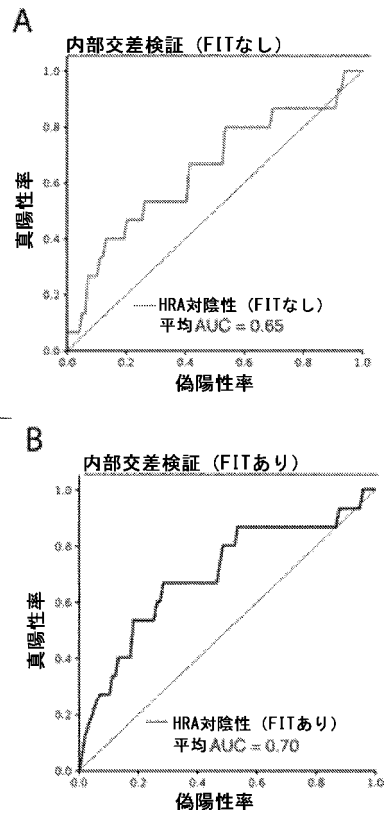
【図 11 B】



【図 1 2】



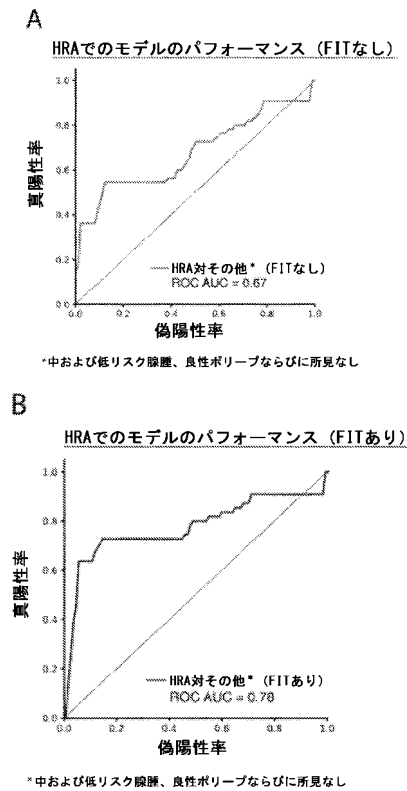
【図 1 3】



【図 1 4】

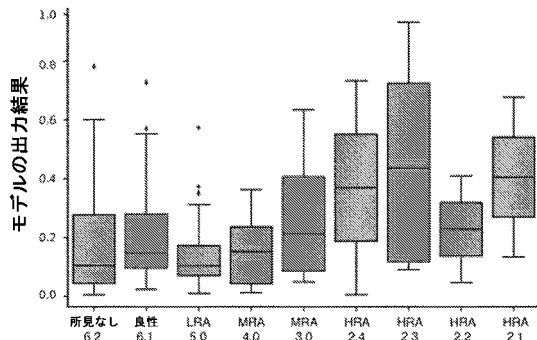
フィーチャー	フィーチャータイプ	重要度
ACY1	正規化発現量	0.134
TNFRSF10B	正規化発現量	0.130
DST	正規化発現量	0.091
GAPDH	未加工発現量	0.087
EDN1	正規化発現量	0.079
EGLN2	正規化発現量	0.074
年齢	個体群 統計情報	0.073
ACHE	正規化発現量	0.068
喫煙状態	個体群 統計情報	0.060
ERBB2	正規化発現量	0.055
SMAD4	正規化発現量	0.054
CTNNB1	正規化発現量	0.050
PER3	正規化発現量	0.045

【図 1 5】

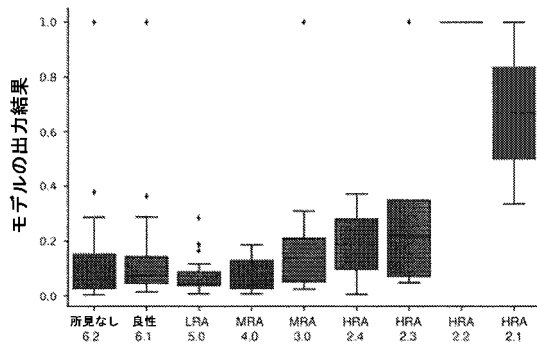


【図 16】

A 試料タイプごとのモデルの出力結果 (FITなし)

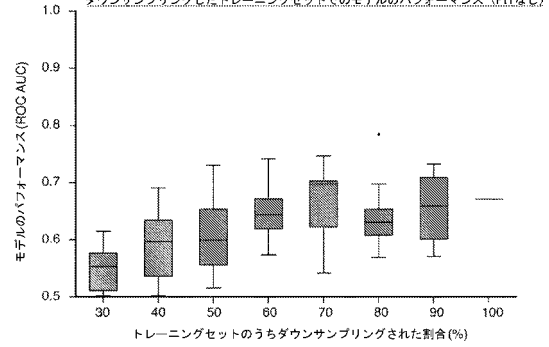


B 試料タイプごとのモデルの出力結果 (FITあり)

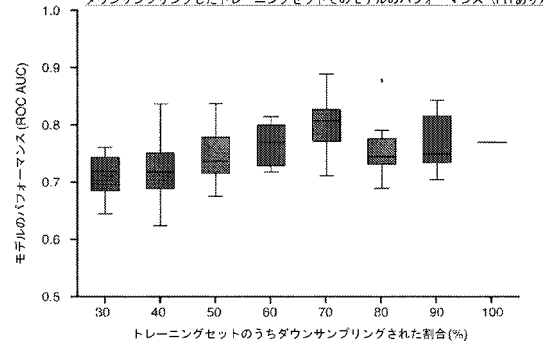


【図 17】

A ダウンサンプリングしたトレーニングセットでのモデルのパフォーマンス (FITなし)

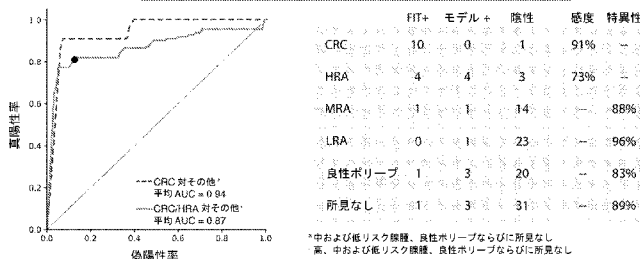


B ダウンサンプリングしたトレーニングセットでのモデルのパフォーマンス (FITあり)



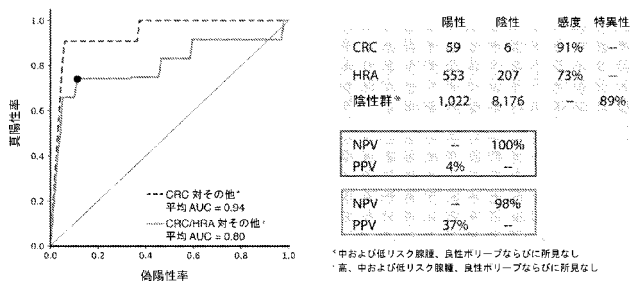
【図 18】

HRAおよびCRCでのホールドアウトテストセットにおけるモデルのパフォーマンス



【図 19】

HRAおよびCRCでの外挿スクリーニング集団におけるモデルのパフォーマンス



【国際調査報告】

INTERNATIONAL SEARCH REPORT		International application No. PCT/US2019/035061
A. CLASSIFICATION OF SUBJECT MATTER C12Q 1/6886(2018.01)i, G01N 33/574(2006.01)i, G01N 33/72(2006.01)i According to International Patent Classification (IPC) or to both national classification and IPC		
B. FIELDS SEARCHED Minimum documentation searched (classification system followed by classification symbols) C12Q 1/6886; C12Q 1/68; G01N 33/48; G01N 33/574; G01N 33/72 Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched Korean utility models and applications for utility models Japanese utility models and applications for utility models Electronic data base consulted during the international search (name of data base and, where practicable, search terms used) cKOMPASS(KIPO internal) & keywords: colorectal cancer, expression level, RNA biomarker, variant allele frequency		
C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 2010-0196889 A1 (BANKAITIS-DAVIS, D. M. et al.) 5 August 2010 See abstract; paragraphs [0039], [0096], [0110], [0151]; tables 1-5; claims 1-6, 13-15.	1-7, 22-25, 27-28
A		37-38
X	US 2013-0296191 A1 (ROEPMAN, P. et al.) 7 November 2013 See paragraphs [0037]-[0039], [0046]-[0048]; tables 1, 4-5; claims 1-10.	37-38
X	US 2010-0112713 A1 (CHAPKIN, R. S. et al.) 6 May 2010 See abstract; claims 1-9.	1-7, 22-25, 27-28
X	US 2014-0221244 A1 (CHAPMAN, K. et al.) 7 August 2014 See abstract; claim 19.	37-38
X	WO 2016-176446 A2 (GENEOSCOPY, LLC) 3 November 2016 See abstract; page 25; paragraphs [0087], [0097], [0099]; tables 1, 3-4; claims 1-14.	1-7, 22-25, 27-28
☐ Further documents are listed in the continuation of Box C. ☒ See patent family annex.		
* Special categories of cited documents: "A" document defining the general state of the art which is not considered to be of particular relevance "D" document cited by the applicant in the international application "E" earlier application or patent but published on or after the international filing date "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) "O" document referring to an oral disclosure, use, exhibition or other means "P" document published prior to the international filing date but later than the priority date claimed "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art "&" document member of the same patent family		
Date of the actual completion of the international search 25 September 2019 (25.09.2019)		Date of mailing of the international search report 25 September 2019 (25.09.2019)
Name and mailing address of the ISA/KR  International Application Division Korean Intellectual Property Office 189 Cheongsu-ro, Seo-gu, Daejeon, 35208, Republic of Korea Facsimile No. +82-42-481-8578		Authorized officer HEO, Joo Hyung  Telephone No. +82-42-481-8150

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US2019/035061

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.: 8-21,26,29-36,39-50
because they relate to subject matter not required to be searched by this Authority, namely:
Claims 8-21,26, 29-36, 39-50 pertain to methods for treatment of the human body by therapy or surgery, as well as diagnostic methods, and thus relate to a subject matter which this International Searching Authority is not required, under PCT Article 17(2)(a)(i) and PCT Rule 39.1(iv), to search.
2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:
3. ☒ Claims Nos.: 32-36,44-48,50
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

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

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

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

Remark on Protest

☐ The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.

☐ The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.

☐ No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No.

PCT/US2019/035061

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 2010-0196889 A1	05/08/2010	AU 2007-322207 A1 CA 2670872 A1 EP 2092082 A2 EP 2402464 A1 US 2014-0024547 A1 WO 2008-063414 A2 WO 2008-063414 A3	29/05/2008 29/05/2008 26/08/2009 04/01/2012 23/01/2014 29/05/2008 29/01/2009
US 2013-0296191 A1	07/11/2013	EP 2655661 A2 EP 2655661 B1 US 10036070 B2 WO 2012-087144 A2 WO 2012-087144 A3	30/10/2013 11/03/2015 31/07/2018 28/06/2012 20/09/2012
US 2010-0112713 A1	06/05/2010	WO 2010-053539 A2 WO 2010-053539 A3	14/05/2010 16/09/2010
US 2014-0221244 A1	07/08/2014	AU 2012-301589 A1 CA 2844822 A1 CN 103907022 A EP 2751561 A2 HK 1199100 A1 JP 2014-525586 A KR 10-2014-0057354 A WO 2013-033629 A2 WO 2013-033629 A3	27/02/2014 07/03/2013 02/07/2014 09/07/2014 19/06/2015 29/09/2014 12/05/2014 07/03/2013 25/04/2013
WO 2016-176446 A2	03/11/2016	US 2018-0282815 A1 WO 2016-176446 A3	04/10/2018 12/01/2017

フロントページの続き

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

(特許庁注：以下のものは登録商標)

1. TRITON
2. ブルートゥース

(74)代理人 100148699

弁理士 佐藤 利光

(74)代理人 100128048

弁理士 新見 浩一

(74)代理人 100129506

弁理士 小林 智彦

(74)代理人 100205707

弁理士 小寺 秀紀

(74)代理人 100114340

弁理士 大関 雅人

(74)代理人 100121072

弁理士 川本 和弥

(72)発明者 バーネル アンドリュウ

アメリカ合衆国 63124 ミズーリ州 セントルイス クレイトン ロード 9300

(72)発明者 バーネル エリカ

アメリカ合衆国 63124 ミズーリ州 セントルイス クレイトン ロード 9300

(72)発明者 カン インミン

アメリカ合衆国 63105 ミズーリ州 クレイトン ウィットバーン ドライブ 8166
アパートメント #5

(72)発明者 ヴェルツラー エリザベス

アメリカ合衆国 63112 ミズーリ州 セントルイス パーシング アベニュー 5390
アパートメント #208

(72)発明者 キャンベル ケイティ

アメリカ合衆国 63110 ミズーリ州 セントルイス ギブソン アベニュー 4220

Fターム(参考) 4B029 AA07 AA23 BB20 FA15

4B063 QA01 QA19 QQ03 QQ52 QR66 QS34 QS36 QX02