

(19) 日本国特許庁(JP)

## (12) 特許公報(B2)

(11) 特許番号

特許第6371280号  
(P6371280)

(45) 発行日 平成30年8月8日(2018.8.8)

(24) 登録日 平成30年7月20日(2018.7.20)

(51) Int.Cl.

F 1

C 12 Q 1/6869 (2018.01)

C 12 Q 1/6869

Z

C 12 Q 1/6886 (2018.01)

C 12 Q 1/6886

Z

G 06 F 19/18 (2011.01)

G 06 F 19/18

請求項の数 55 (全 68 頁)

(21) 出願番号 特願2015-517896 (P2015-517896)  
 (86) (22) 出願日 平成25年6月14日 (2013.6.14)  
 (65) 公表番号 特表2015-527057 (P2015-527057A)  
 (43) 公表日 平成27年9月17日 (2015.9.17)  
 (86) 國際出願番号 PCT/IB2013/054898  
 (87) 國際公開番号 WO2013/190441  
 (87) 國際公開日 平成25年12月27日 (2013.12.27)  
 審査請求日 平成27年8月10日 (2015.8.10)  
 (31) 優先権主張番号 61/662,878  
 (32) 優先日 平成24年6月21日 (2012.6.21)  
 (33) 優先権主張国 米国(US)  
 (31) 優先権主張番号 61/682,725  
 (32) 優先日 平成24年8月13日 (2012.8.13)  
 (33) 優先権主張国 米国(US)

(73) 特許権者 513154234  
 ザ チャイニーズ ユニバーシティー オ  
 ブ ホンコン  
 中華人民共和国 香港, ニューテリトリ  
 ーズ, シャンティン, ザ ノーレッジ  
 トランスファー オフィス  
 (74) 代理人 100099759  
 弁理士 青木 篤  
 (74) 代理人 100077517  
 弁理士 石田 敏  
 (74) 代理人 100087871  
 弁理士 福本 積  
 (74) 代理人 100087413  
 弁理士 古賀 哲次

最終頁に続く

(54) 【発明の名称】癌検出のための血漿中DNAの突然変異解析

## (57) 【特許請求の範囲】

## 【請求項1】

対象の癌または前癌性変化を検出することを補助する方法であつて、  
 前記対象の構成的ゲノムを取得し；  
 前記対象の生物試料の複数のDNAフラグメントの各々の1つ以上の配列タグをコンピューターシステムに受け入れ、前記生物試料が無細胞性DNAを含み；  
 前記配列タグのゲノム位置をコンピューターシステムによって決定し；  
 第一座位の第一数を決定するために、前記配列タグを、構成的ゲノムとコンピューターシステムによって比較し、ここで、

上記第一座位の各々において、構成的ゲノムと比べた配列異型を有する配列タグ数が 10  
 、カットオフ値より上で、前記カットオフ値が1より大きい；

前記第一座位における配列異型を有する配列タグのカウントに基づいてパラメーターを  
 決定し；および

前記対象の癌のレベルの分類を決定するために、前記パラメーターを、閾値と比較する  
 こと、

を含む、方法。

## 【請求項2】

前記閾値が、1つ以上の他の対象からの1つ以上の試料から決定される、請求項1に記載の方法。

## 【請求項3】

前記座位の前記カットオフ値が、前記座位におけるゲノム位置を有する配列タグの全数に依存する、請求項 1 に記載の方法。

【請求項 4】

異なるカットオフ値が、少なくとも 2 つの前記第一座位に使用される、請求項 1 に記載の方法。

【請求項 5】

前記第一座位の 1 つについて第一カットオフ値を動的に決定し、前記第一座位の 1 つが第一領域内に属する、ことをさらに含む、請求項 4 に記載の方法。

【請求項 6】

前記第一カットオフ値が、前記第一座位の 1 つのシークエンス深度に基づいて決定される、請求項 5 に記載の方法。 10

【請求項 7】

前記第一カットオフ値が、シーケンスエラー率、前記第一領域のシークエンス深度  $D$ 、および前記第一領域のヌクレオチド位置数に依存する偽陽性率に基づいて決定される、請求項 5 に記載の方法。

【請求項 8】

前記第一カットオフ値が、前記第一領域の真陽性数に基づいて決定される、請求項 7 に記載の方法。

【請求項 9】

前記生物試料の、前記第一領域のシークエンス深度  $D$  および腫瘍由来 DNA の分画濃度  $f$  に基づいて、前記第一カットオフ値の真陽性数を算出することをさらに含む、請求項 8 に記載の方法。 20

【請求項 10】

前記真陽性数の算出が、式 :

【数 1】

$$Pb = 1 - \sum_{i=0}^{r-1} Poisson(i, M_p)$$

30

( 式中、  $Pb$  は、真陽性を検出する確率であり、  $r$  は、前記第一カットオフ値であり、  $M_p = D \times f / 2$  である )

に従ったポアソン分布確率を使用する、請求項 9 に記載の方法。

【請求項 11】

前記第一カットオフ値が、次の基準 :

シークエンス深度が 50 より小さいならば、前記第一カットオフ値が 5 である、

シークエンス深度が 50 ~ 110 ならば、前記第一カットオフ値が 6 である、

シークエンス深度が 111 ~ 200 ならば、前記第一カットオフ値が 7 であり、

シークエンス深度が 201 ~ 310 ならば、前記第一カットオフ値が 8 であり、 40

シークエンス深度が 311 ~ 450 ならば、前記第一カットオフ値が 9 であり、

シークエンス深度が 451 ~ 620 ならば、前記第一カットオフ値が 10 であり、

シークエンス深度が 621 ~ 800 ならば、前記第一カットオフ値が 11 である、

のいずれか 1 つを使用して決定される、請求項 5 に記載の方法。

【請求項 12】

前記パラメーターが、第一座位の前記第一数の加重和であり、前記第一座位の各々の分布が、前記第一座位に割り当てる重要な度値に基づいて、重み付けされる、請求項 1 に記載の方法。

【請求項 13】

前記パラメーターが、第一座位の前記第一数における配列異型を有する前記配列タグの 50

和を含む、請求項 1 に記載の方法。

【請求項 1 4】

前記和が加重和であり、1つの前記第一座位が、第二前記第一座位の第二重みより異なる第一重みを有する、請求項 1 3 に記載の方法。

【請求項 1 5】

前記第一重みが、前記第二重みより大きく、1つの第一座位が、癌と関連し、前記第二前記第一座位が、癌と関連しない、請求項 1 4 に記載の方法。

【請求項 1 6】

前記パラメーターが、第一座位の前記第一数である、請求項 1 に記載の方法。

【請求項 1 7】

配列タグのゲノム位置の決定が：

前記配列タグの少なくとも1つの部分を、参照ゲノムにアライメントさせ、前記配列タグのアライメントが、前記配列タグと前記参照ゲノムとの間の1つ以上のミスマッチを許容すること

を含む、請求項 1 に記載の方法。

【請求項 1 8】

前記構成的ゲノムに対する前記配列タグの比較が、

前記参照ゲノムと比べた異型を有する第二座位の第二数を決定するために、前記構成的ゲノムを、前記参照ゲノムと比較し；

前記アライメントに基づいて、第三座位の第三数を決定し：

前記第三座位の各々において、前記参照ゲノムと関する配列異型を有する前記配列タグの数が、カットオフ値より上であり；および

第一座位の前記第一数を得るために、前記第三数と前記第二数の差を得ることを含む、請求項 1 7 に記載の方法。

【請求項 1 9】

前記第三数と前記第二数の前記差を得ることが、前記第一座位を特定する、請求項 1 8 に記載の方法。

【請求項 2 0】

前記パラメーターの決定が、

第一座位の第一数の各座位について、

前記座位にアライメントされ、前記座位に配列異型を有する配列タグをカウントし；および

各カウントに基づいて前記パラメーターを決定することを含む、請求項 1 9 に記載の方法。

【請求項 2 1】

前記構成的ゲノムが、50%以上の構成的DNAを含む前記対象からの構成的試料に由来する、請求項 1 に記載の方法。

【請求項 2 2】

配列タグのゲノム位置の決定が、

前記配列タグの少なくとも1つの部分を、前記構成的ゲノムにアライメントさせ、配列タグの前記アライメントが、前記配列タグと前記構成的ゲノムとの間の1つ以上のミスマッチを許容すること

を含む、請求項 1 に記載の方法。

【請求項 2 3】

前記構成的ゲノムへの前記配列タグの比較が、

前記アライメントに基づいて、前記対象の前記構成的ゲノムと関連する遺伝子位置において、配列異型を有する配列タグを特定し；

各遺伝子位置に、配列異型を示し；

前記遺伝子位置にアライメントされ、前記遺伝子位置において配列異型を有する配列タグの各数をカウントし；および

10

20

30

40

50

前記各数に基づいて、パラメーターを決定することを含む、請求項 2 2 に記載の方法。

【請求項 2 4】

前記各数に基づいて、前記パラメーターが、第一和を得るために、前記各数を加算し；および前記パラメーターを決定するために、前記第一和を使用することを含む、請求項 2 3 に記載の方法。

【請求項 2 5】

前記パラメーターを決定するために、前記第一和の使用が、配列異型を示す遺伝子位置の数を、前記第一和から引くことを含む、請求項 2 4 に記載の方法。 10

【請求項 2 6】

前記パラメーターを決定するために、前記第一和の使用が、アライメントされた配列タグの量に基づいて、前記第一和を正規化することを含む、請求項 2 4 に記載の方法。

【請求項 2 7】

前記 1 つ以上の配列タグが、前記生物試料中の D N A フラグメントのランダムシーケンシングから生成される、請求項 1 に記載の方法。

【請求項 2 8】

前記パラメーターが、腫瘍由来 D N A の分画濃度である、請求項 1 に記載の方法。

【請求項 2 9】

癌のレベルが、癌が存在するか否か、癌のステージ、腫瘍のサイズ、癌の重症度、又は突然変異もしくは複数の突然変異に伴う前癌病態によって分類される、請求項 1 ~ 2 8 のいずれか 1 項に記載の方法。 20

【請求項 3 0】

対象の 1 つ以上の腫瘍中の腫瘍細胞の複数のクローニングの不均一性を解析する方法であつて、

前記対象の構成的ゲノムを取得し；

前記対象の生物試料中の複数の D N A フラグメントの各々の 1 つ以上の配列タグをコンピューターシステムによって受け、前記生物試料が、無細胞性 D N A を含み；

前記配列タグのゲノム位置をコンピューターシステムによって決定し；

第一座位の数を決定するために、前記配列タグを、前記構成的ゲノムとコンピューター 30 システムによって比較し：

各第一座位において、前記構成的ゲノムに関する配列異型を有する前記配列タグの第一数が、カットオフ値より上であり、前記カットオフ値が 1 より大きく；および

前記第一座位の配列タグの前記第一数に基づいて、前記 1 つ以上の腫瘍中の腫瘍細胞の多数のクローニングの不均一性測度を算出すること、

ここで、不均一性測度は、複数のクローニングのすべてによって共有されない突然変異の量を示す、

を含む、方法。

【請求項 3 1】

複数のクローニングの内の少なくとも 2 つが单一の腫瘍内に存在する、請求項 3 0 に記載の方法。 40

【請求項 3 2】

不均一性測度が腫瘍内不均一性測度を含む、請求項 3 0 に記載の方法。

【請求項 3 3】

前記 1 つ以上の腫瘍が複数の腫瘍であり、不均一性測度が腫瘍内不均一性測度を含む、請求項 3 0 に記載の方法。

【請求項 3 4】

不均一性レベルの分類を決定するために、前記不均一性測度を、1 つ以上の閾値と比較することをさらに含む、請求項 3 0 に記載の方法。

【請求項 3 5】

前記 1 つ以上の閾値が、その腫瘍が生検を行われ、不均一性レベル測定のための生検された腫瘍中の突然変異を決定するために解析された 1 つ以上の他の対象から決定され、および 1 つ以上の他の対象の無細胞性 D N A を含む生物試料からの不均一性測度が、閾値を決定するために使用される、請求項 3 4 に記載の方法。

【請求項 3 6】

1 つ以上の閾値に対する前記不均一性測度の比較が、前記不均一性測度に基づいて、不均一性レベルをアウトプットする校正機能に、前記不均一性測度をインプットすることを含む、請求項 3 5 に記載の方法。

【請求項 3 7】

前記不均一性測度は、第一座位の全数を含む、請求項 3 0 に記載の方法。

10

【請求項 3 8】

複数の不均一性測度が算出され、不均一性測度の算出が、各第一座位で、配列変異を有する配列タグの比率を算出し；前記比率の値の、第一座位の数のヒストグラムを作成し；および前記ヒストグラム中の複数のピークを特定することを含む、請求項 3 0 に記載の方法。

【請求項 3 9】

複数の不均一性測度の内の 1 つが、特定したピークの前記数に対応する、請求項 3 8 に記載の方法。

【請求項 4 0】

複数の不均一性測度の内の 1 つが、複数のピークの内の 2 つのピーク高さの比を含む、請求項 3 8 に記載の方法。

20

【請求項 4 1】

前記比率各々が、特定の第 1 座位で測定した腫瘍 D N A の分画濃度を表す、請求項 3 8 に記載の方法。

【請求項 4 2】

前記不均一性測度が、配列変異を有する配列タグの第一特定量を有する第一座位の第 1 比率および配列変異を有する配列タグの第二特定量を有する第一座位の第 2 比率の比に対応する、請求項 3 0 に記載の方法。

【請求項 4 3】

前記第一特定量が、前記第二特定量より小さい、請求項 4 2 に記載の方法。

30

【請求項 4 4】

前記第一特定量が、第一範囲および前記第二特定量が、第二範囲であり、前記第一範囲が、前記第二範囲より下である、請求項 4 3 に記載の方法。

【請求項 4 5】

前記第一特定量および前記第二特定量が、分画濃度または配列変異を有する配列タグの絶対数に対応する、請求項 4 2 に記載の方法。

【請求項 4 6】

前記不均一性測度が、前記第一数の各々に対応する第一座位の数のヒストグラムから決定される、請求項 3 0 に記載の方法。

40

【請求項 4 7】

複数の不均一性測度が算出され、第一座位が、第一サブセットおよび第二サブセットを含み、前記複数の不均一性測度が、前記第一サブセットに対応する前記第一数の第一ヒストグラムおよび前記第二サブセットに対応する前記第一数の第二ヒストグラムを含む、請求項 4 6 に記載の方法。

【請求項 4 8】

前記不均一性測度が、特定値以上の対応する第一数を有する第一座位の比率を含む、請求項 3 0 に記載の方法。

【請求項 4 9】

対象がヒトである、請求項 1 ~ 4 8 のいずれか 1 項に記載の方法。

50

## 【請求項 5 0】

生物試料が、尿、胸膜液、腹水、腹腔液、唾液、涙、脳脊髄液、又は大便試料である、請求項 1 ~ 4 9 のいずれか 1 項に記載の方法。

## 【請求項 5 1】

90%超の構成的 DNA を含む対象の構成的試料を得；

構成的試料中の DNA 断片のランダムシーケンシングを行って、構成的試料中の複数の DNA フラグメントの各々について 1 以上の 第二配列タグを得；

参照ゲノムに 第二配列タグ の少なくとも一部をアライメントさせ、

ここで、第二配列タグのアライメントは、M 以下のゲノム位置で第二配列タグと 参照ゲノムとの間のミスマッチを許容し（ここで、M は、1 以上の整数を示す）；及び 10

第二配列タグ及び当該アライメントに基づいて構成的ゲノムを構築すること、  
を更に含む、請求項 1 ~ 5 0 のいずれか 1 項に記載の方法。

## 【請求項 5 2】

構成的試料が生物試料であり、構成的ゲノムを構築することが、

2 つのアレルを有するホモ接合性座位又はヘテロ接合性座位の決定を含むコンセンサス配列を決定し、

構成的ゲノムでコンセンサス配列を使用すること、  
を含む、請求項 5 1 に記載の方法。

## 【請求項 5 3】

対象の生物試料を準備し、

生物試料中の DNA フラグメントのランダムシーケンシングを行って、生物試料中の複数の DNA フラグメントの各々について 1 以上の配列タグを作製すること、  
を更に含む、請求項 1 ~ 5 2 のいずれか 1 項に記載の方法。

## 【請求項 5 4】

請求項 1 ~ 5 0 のいずれか 1 項の方法を実行するために、コンピューターシステム制御が実行されたときの複数の指示を保存しているコンピューター読み取り可能媒体。

## 【請求項 5 5】

少なくとも 1 つのプロセッサ及びメモリを更に含む、請求項 1 ~ 5 0 のいずれか 1 項に記載の方法を実行するための手段を含むコンピューター装置。

## 【発明の詳細な説明】

## 【背景技術】

## 【0 0 0 1】

## 関連出願

本出願は、2012年7月20日に出願された2012年6月21日に出願された米国特許仮出願第 61/662,878 号、発明の名称「癌検出のための血漿中 DNA の突然変異解析」；2012年8月13日に出願された米国特許仮出願第 61/682,725 号、発明の名称「癌検出のための血漿中 DNA の突然変異解析」；2012年8月31日に出願された米国特許仮出願第 61/695,795 号、発明の名称「癌検出のための血漿中 DNA の突然変異解析」；および2012年10月8日に出願された米国特許仮出願第 61/711,172 号、発明の名称「癌検出のための血漿中 DNA の突然変異解析」の優先権の利益を主張し、これらの出願は、全目的のためにその全文を参照することにより本明細書に組み入れられる。

## 【0 0 0 2】

腫瘍由来 DNA が、癌患者の無細胞血漿 / 血清中に存在することが示されている (Chen XQ et al. Nat Med 1996; 2: 1033-1035)。現行法は、癌に関連していると知られている突然変異の直接的解析に基づいている (Diehl F et al. Proc Natl Acad Sci 2005; 102: 16368-16373; Forshaw T et al. Sci Transl Med 2012; 4: 136ra68)。別の方法で、血漿中 DNA のランダムシーケンシングにより検出される癌関連コピー数変異が研究された (Lo らによる米国特許公開第 2013/0040824 号)。

## 【0 0 0 3】

10

20

30

40

50

経時的に、1つ以上の癌細胞が、増殖優位性を獲得し、娘細胞の複数のクローンを產生することが知られている。最終的に、腫瘍増殖および/またはその転移巣は、クローンの癌細胞群の集合体を含むことになる。この現象は、典型的に、腫瘍不均一性と言われる (Gerlinger M et al. N Engl J Med 2012; 366: 883-892; Yap TA et al. Sci Transl Med 2012; 4: 127ps10)。

#### 【0004】

癌は、高度に不均一、すなわち、同組織型癌突然変異プロファイルが広く変化し得ることが知られている。したがって、特定突然変異の直接的解析は、通常、それらの特異的突然変異と関連すると知られている特定癌型内の場合のサブセットのみを検出され得る。加えて、腫瘍由来DNAは、通常、ヒト血漿中DNAの少数の種であり；血漿中DNAの絶対濃度が低い。したがって、血漿中または血清中の癌関連突然変異の1群または小さい群の直接的検出は、標的突然変異を抱えていると分かっている癌を有する患者の中でさえ、低分析感度しか得られない。さらに、単一腫瘍内でさえ、突然変異の面から、有意に腫瘍内不均一性であることが示されている。突然変異は、腫瘍細胞の亜集団のみに発見され得る。原発性腫瘍と転移性病変との間の突然変異プロファイルの違いは、なおさら大きい。腫瘍内および原発性転移不均一性の一例は、結腸直腸癌を患っている患者のKRAS、BRAFおよびPIK3CA遺伝子である (Baldus et al. Clin Cancer Research 2010. 16:790-9.)。

10

#### 【0005】

患者が、原発性腫瘍 (KRAS突然変異を保因するが、PIK3CAを保因しない) および隠れた転移性病変 (PIK3CAを保因するが、KRAS突然変異を保因しない) を有するシナリオでは、もし、原発性腫瘍のKRAS突然変異の検出に着目したならば、隠れた転移性病変は検出されない。しかしながら、もし、両方の突然変異を分析したならば、原発性腫瘍および隠れた転移性病変の両方が検出され得る。それゆえ、両方の突然変異に関する評価は、残った腫瘍組織の検出に高感度を有することになる。そのような単純な例は、癌をスクリーニングしているとき、および起こり得る突然変異の型がほとんどまたは全く見当がつかないとき、より複雑になる。

20

#### 【発明の概要】

#### 【0006】

従って、癌の広範なスクリーニング、検出、または評価を行うための新規技術を提供することが望まれている。

30

#### 【0007】

##### 概要

実施形態は、同対象の構成的DNAのものと比較したときに、癌のスクリーニングまたはモニターを行う対象の生物試料 (例えば、血漿または血清) の体細胞突然変異の発生頻度を観察し得る。ランダムシークエンシングは、これらの発生頻度を測定するために使用され得る。パラメーターは、これらの発生頻度から算出され、癌レベルの分類決定のために使用され得る。偽陽性は、異型配列リード (タグ) の少なくとも特定数を有するいずれかの異型座位を必要とすることにより選別され、それゆえ、より正確なパラメーターが得られ得る。種々の異型座位の相対的発生頻度は、患者の腫瘍不均一性レベルを決定するために、解析され得る。

40

#### 【0008】

ひとつの実施形態では、該パラメーターは、癌を有しない、または癌のリスクの低い対象群由来の同パラメーターと比較され得る。テスト対象および癌を有しない、または癌のリスクの低い対象群から得られたパラメーターの有意差は、テスト対象が、癌または前癌性状態を有する、または将来癌を発症するかもしれないリスクの増加を示し得る。したがって、ひとつの実施形態では、血漿中DNA解析は、事前の腫瘍ゲノム情報なしで実行され得る。したがって、そのような実施形態は、特に、癌スクリーニングに有用である。

#### 【0009】

別の実施形態では、実施形態は、治療後の癌患者のモニターおよび残存腫瘍がないか、

50

または腫瘍再発がないかを見るためにも使用され得る。例えば、残存腫瘍を有するまたは腫瘍再発した患者は、残存腫瘍がない、または腫瘍再発が観察されない者より、体細胞突然変異が高頻度に発生することになる。該モニターは、体液中または血漿もしくは血清などの無細胞核酸を有する他試料中の腫瘍関連遺伝子異常の時間的变化を確認するための処置に続く、複数時点での癌患者から得られる試料を含み得る。

#### 【0010】

ひとつの実施形態に従えば、方法は、対象の癌または前癌性変化を検出する。該対象の構成的遺伝子が得られる。1つ以上の配列タグは、該対象の生物試料中の複数のDNAフラグメントの各々に対して受け入れられ、該生物試料は、無細胞性DNAを含む。ゲノム位置は、配列タグに対して決定される。該配列タグは、第一座位の第一数を決定するために構成的ゲノムと比較される。各第一座位で、該構成的ゲノムと比べた配列異型を有する該配列タグ数は、カットオフ値より上であり、該カットオフ値は、1より大きい。パラメーターは、該第一座位の配列異型を有する配列タグのカウントに基づいて決定される。該パラメーターは、該対象の癌レベルの分類決定のため、閾値と比較される。

10

#### 【0011】

別の実施形態に従って、方法は、対象の1つ以上の腫瘍不均一性を解析する。該対象の構成的ゲノムが得られる。1つ以上の配列タグは、該対象の生物試料の複数のDNAフラグメントの各々に対して、受け入れられ、該生物試料は、無細胞性DNAを含む。ゲノム位置は、配列タグに対して決定される。配列タグは、第一座位の第一数を決定するために、構成的ゲノムと比較される。各第一座位で、構成的ゲノムと比べた配列異型を有する配列タグ数は、カットオフ値より上であり、該カットオフ値は1より大きい。1対上の腫瘍の不均一性の測定は、第一遺伝子位置セットの各第一数に基づいて、算出される。

20

#### 【0012】

別の実施形態に従えば、方法は、無細胞性DNAを含む生物試料中の腫瘍DNAの分画濃度を決定する。1つ以上の配列タグは、生物試料中の複数のDNAフラグメントの各々に対して受け入れられる。ゲノム位置は、配列タグに対して決定される。複数のゲノム領域の各々に対して、ゲノム領域内のDNAフラグメントの各量は、ゲノム領域内のゲノム位置を有する配列タグから決定される。該各量は、各密度を得るために正規化される。該各密度は、ゲノム領域が、1コピー欠失または1コピー獲得を示しているか確認するために、参考密度と比較される。第一密度は、1コピー欠失を示すと確認された各密度から、または1コピー獲得を示すと確認された各密度から、算出される。該分画濃度は、差を得るために該第一密度を別の密度と比較することにより算出され、該差は参考密度で正規化される。

30

#### 【0013】

他の実施形態は、本明細書に記載の方法と関連したシステムおよびコンピューター読み取り可能媒体に関する。

#### 【0014】

本発明の特徴および優位性は、次の詳細な説明および付随する図を参照することにより、より理解され得る。

#### 【図面の簡単な説明】

40

#### 【0015】

【図1】本発明の実施形態に従った、対象の癌または前癌性変化を検出する方法100のフローチャートである。

【図2】本発明の実施形態に従った、試料ゲノム(SG)を直接、構成的ゲノム(CG)と比較する方法のフローチャートを示す。

【図3】本発明の実施形態に従った、参照ゲノム(RG)を使用して、試料ゲノム(SG)を構成的ゲノム(CG)と比較する方法300のフローチャートを示す。

【図4】試料中の腫瘍由来のDNAの分画濃度が10%であると推定されるとき、突然変異が、本発明の実施形態に従った試料中に存在すると分類する判断基準として、異なる発生数を使用して、正確に確認された癌関連一ヌクレオチド突然変異数を示す表400であ

50

る。

【図5】試料中の腫瘍由来DNAの分画濃度が、5%であると推定されるとき、偽陽性座位の期待数および確認された突然変異期待数を示す表である。

【図6A】10%および20%の腫瘍由来のDNAの血漿中分画濃度を有する血漿中の癌関連突然変異の検出率を示し、潜在的癌関連突然変異を誘発する目安として4および6の発生率( $r$ 値)を使用したグラフ600である。

【図6B】シークエンス深度に対する発生率( $r$ 値)4、5、6および7の基準を使用して、ヌクレオチド変化を有するときに誤って分類されたヌクレオチド配列部位の期待値を示すグラフ650である。

【図7A】試料中の腫瘍由来DNAの分画濃度が、5%であると推定されるときに、眞の癌関連部位および異なるシークエンス深度を有する偽陽性部位の数を示すグラフ700である。

【図7B】全ゲノム(WG)および全エクソンの解析を伴う偽陽性部位の予測数を示すグラフ750である。

【図8】本発明の実施形態に従った、血漿中腫瘍由来DNAの分画濃度を含む、治療前後の4HCC患者の結果を示す表800である。

【図9】本発明の実施形態に従った、16の健康な対照群のHCC関連SNVの検出を示す表900である。

【図10A】本発明の実施形態に従った、HCC患者の腫瘍試料の配列リード密度の分布プロットを示す。

【図10B】本発明の実施形態に従った、HCC患者の血漿中全ビンのzスコアの分布プロット1050を示す。

【図11】本発明の実施形態に従った、HCC患者の血漿のzスコアの分布プロット1100を示す。

【図12】本発明の実施形態に従った、無細胞性DNAを含む生物試料の腫瘍DNAの分画濃度を測定する方法1200のフローチャートである。

【図13A】本発明の実施形態に従った、診断時の卵巣癌および乳癌を有する患者の血漿中突然変異解析の表1300を示す。

【図13B】本発明の実施形態に従った、腫瘍切除後の両側性卵巣癌および乳癌を有する患者の血漿中突然変異解析の表1350を示す。

【図14A】HCC1の血漿中DNAの-ヌクレオチド変異の検出を示す表1400である。

【図14B】HCC2の血漿中DNAの-ヌクレオチド変異の検出を示す表1450である。

【図15A】HCC3の血漿中DNAの-ヌクレオチド変異の検出を示す表1500である。

【図15B】HCC4の血漿中DNAの-ヌクレオチド変異の検出を示す表1550である。

【図16】卵巣癌(および乳癌)を有する患者の血漿中DNAの-ヌクレオチド変異の検出を示す表1600である。

【図17】発生率およびシークエンス深度の異なる必要予測感度を示す表1700である。

【図18】異なるカットオフおよび異なるシークエンス深度の偽陽性座位の予測数を示す表1800である。

【図19】異なる腫瘍部位で検出された突然変異数を図解する3つのダイアグラムを示す。

【図20】治療前および治療後の血漿試料の腫瘍由来突然変異を保因するフラグメント数を示す表2000である。

【図21】単一腫瘍部位で検出された突然変異および全4腫瘍部位で検出された突然変異の血漿中発生率分布を示すグラフ2100である。

10

20

30

40

50

【図22】不均一腫瘍から発生する突然変異の血漿中発生率予測分布を示すグラフ2200である。

【図23】16の健康な対照群の実施形態の特異性が回復されることを示す。

【図24】本発明の実施形態に従って、対象の1つ以上の腫瘍の不均一性を解析する方法2400のフローチャートである。

【図25】本発明の実施形態に従ったシステムおよび方法で有用な例のコンピューターシステム2500のブロックダイアグラムを示す。

【発明を実施するための形態】

【0016】

定義

10

本明細書において使用される場合、「座位（locus）」またはその複数形の「座位（loci）」という用語は、ゲノム上で変異を有することがあり得るあらゆる長さのヌクレオチド（または塩基対）の位置またはアドレスである。「BIN」はゲノム中の所定の長さの領域である。複数のBINが同一の第1の長さ（解像度）を有してもよく、異なる複数のBINが同一の第2の長さを有してもよい。1つの実施形態において、BINは互いに重複しない。

【0017】

20

「ランダムシークエンシング」という用語は、本明細書において使用される場合、配列決定する核酸断片であって、シークエンシング術の前に具体的に特定または予め決定されていない核酸断片を配列決定することを指す。特定の遺伝子の座位を標的とする配列特異的プライマーは必要ではない。「ユニバーサルシークエンシング」という用語は、あらゆる断片に対して開始することができるシークエンシングを指す。1つの実施形態において、アダプターを断片の末端に付加し、シークエンシング用のプライマーがそれらのアダプターに結合する。したがって、あらゆる断片を同一のプライマーにより配列決定することができ、したがって、そのシークエンシングはランダムであり得る。

【0018】

30

「配列タグ」（配列リードとも呼ぶ）という用語は、本明細書において使用される場合、核酸分子のいずれかの部分または全体に由来する、配列決定されたヌクレオチド鎖を指す。例えば、配列決定されたタグは核酸断片から配列決定された（例えば、約30個の）ヌクレオチドからなる短鎖、核酸断片の両端部分のヌクレオチドからなる短鎖、または生物学的試料に存在する核酸断片全体の配列決定であり得る。核酸断片はより大きい核酸分子のいずれかの部分である。断片（例えば、遺伝子）はより大きい核酸分子の他の部分と別れて（すなわち、結合されずに）存在し得る。

【0019】

40

「構成的ゲノム」（CGとも呼ぶ）という用語はゲノム中の座位におけるコンセンサスヌクレオチドから構成され、したがって、コンセンサス配列と見なされ得る。CGは対象のゲノム全体（例えば、ヒトゲノム）、またはゲノムの部分のみを範囲とすることができます。細胞のDNAならびに（例えば、血漿中に見出され得るような）無細胞性DNAから構成的ゲノム（CG）を獲得することができる。理想的には、コンセンサスヌクレオチドは、ある座位が1つのアレルについてホモ接合性であり、または2つのアレルについてヘテロ接合性であることを示す方がよい。ヘテロ接合性座位は通常はある遺伝的多型のメンバーである2つのアレルを含有する。例として、ある座位に整列させられたリードのうちに2つのアレルがそれぞれ少なくとも所定のパーセンテージ（例えば、30%または40%）で現れる閾値が、その座位がヘテロ接合性であるか決定するための判定基準であり得る。1種類のヌクレオチドが充分なパーセンテージ（例えば、70%以上）で現れる場合、その座位はCG中でホモ接合性であると判断され得る。1つの健常細胞のゲノムは細胞分裂の間に自然に起こる無作為突然変異のために別の健常細胞のゲノムと異なることがあり得るが、このようなコンセンサスが使用されるとき、CGは変化しないはずである。幾つかの細胞はゲノム再構成したゲノムを有することがあり得、例えば、抗体およびT細胞受容体遺伝子に関係するBリンパ球およびTリンパ球がそうである。そのような大規模な

50

差異はそれでも血液中の有核細胞集団のうちの比較的に小さい集団であり、したがって、血液細胞の充分な試料採取（例えば、シークエンス深度）により、そのような再構成が構成的ゲノムの決定に影響を与えることはない。口腔細胞、皮膚細胞、毛包、または様々な正常な体組織の生検組織を含む他の細胞種も CG 源として役立ち得る。

【 0 0 2 0 】

「構成的 DNA」という用語は、対象が持つて生まれた遺伝的構成を反映するあらゆる起源の DNA を指す。対象について、構成的 DNA を得ることができる「構成的試料」の例には健常な血液細胞 DNA、口腔細胞 DNA および毛根 DNA が含まれる。これらの健常細胞に由来する DNA が対象の CG を定義する。それらの細胞は様々な方法で、例えば、ある人が癌を有していないと知られているとき、または癌細胞もしくは前腫瘍細胞（例えば、肝臓癌が疑われるときの毛根 DNA）を含む可能性が無い組織から試料を得ることができるように健常であると特定され得る。別の例として、患者に癌が無いときに血漿試料を得ることができ、決定された構成的 DNA をその後の（例えば、1 年以上後の）血漿試料から得られた結果と比較する。別の実施形態において、50% 未満の腫瘍 DNA を含有する単一の生物試料を構成的ゲノムと腫瘍関連遺伝的変化の推測のために使用することができる。そのような試料では、腫瘍関連一ヌクレオチド突然変異の濃度は CG 中のヘテロ接合性 SNP の各アレルの濃度よりも低い。そのような試料は下に記載される試料ゲノムの決定のために使用される生物試料と同一のものであり得る。

【 0 0 2 1 】

「生物試料」という用語は、本明細書において使用される場合、対象（例えば、ヒト、癌を有する人、癌を有すると疑われる人、または他の生物）から採られ、且つ、1 種類以上の目的の無細胞性核酸分子を含有するあらゆる試料を指す。生物試料には無細胞性 DNA が含まれる場合があり得、それらのうちの幾らかは健常細胞に、幾らかは腫瘍細胞に起源を有し得る。例えば、腫瘍 DNA は血液または他の液体、例えば、尿、胸膜液、腹水、腹腔液、唾液、涙または脳脊髄液の中に見出され得る。非液体例は大便試料であり、それは下痢状の液体と混ぜ合わされていてもよい。そのような試料のうちの幾つかについて、非侵襲的に生物試料を得ることができる。いくつかの実施形態において、生物試料を構成的試料として使用することができる。

【 0 0 2 2 】

「試料ゲノム」（SG とも呼ぶ）という用語はゲノム（例えば、ヒトゲノム）の位置に対して整列させられた配列リードのコレクションである。試料ゲノム（SG）はコンセンサス配列ではないが、充分な数のリード（例えば、少なくとも 2 または 3、またはそれより高いカットオフ値）でのみ現れ得るヌクレオチドを含む。アレルが充分な回数で現れず、CG の一部ではない（すなわち、コンセンサス配列の一部ではない）場合、そのアレルは「一ヌクレオチド突然変異」（SNM とも呼ぶ）を表し得る。本発明を用いて、例えば、（例えば、マイクロサテライト中のタンデムリピート単位の数または単なるタンデムリピートの多型に影響する）2 つ以上のヌクレオチドが関わる突然変異、（染色体内または染色体間であり得る）染色体転座、および配列逆位を含む他の種類の突然変異も検出され得る。

【 0 0 2 3 】

「基準ゲノム」（RG とも呼ぶ）という用語は生物試料に由来する配列リードおよび構成的試料を整列させることができ、比較することができる半数体ゲノムまたは二倍体ゲノムを指す。半数体ゲノムについて、各座位にただ 1 つのヌクレオチドが存在する。二倍体ゲノムについて、ヘテロ接合性座位が特定され得ることがあり得、そのような座位は 2 つのアレルを有し、どちらかのアレルが座位へのアラインメントで合致し得る。

【 0 0 2 4 】

「癌のレベル」という用語は、癌が存在するかどうか、癌のステージ、腫瘍のサイズ、および / または癌の重症度についての他の尺度を指し得る。癌のレベルは数または他の記号であり得る。そのレベルはゼロであり得る。癌のレベルは突然変異または多数の突然変異に伴う前腫瘍または前癌病態（状態）も含む。癌のレベルは様々な方法で用いられ得る

10

20

30

40

50

。例えば、癌を有していると以前は知られていない者に癌が存在するか、スクリーニングによりチェックすることができる。癌を有していると診断された者を評価して検査することができる。検出は「スクリーニング」を意味することがあり得、または癌を思わせる特徴（例えば、症状または他の陽性試験）により誰かが癌を有しているかチェックすることを意味することがあり得る。

#### 詳細な説明

##### 【0025】

腫瘍から直接的に採取されておらず、且つ、無細胞性核酸を含む生物試料（例えば、血液血漿／血清試料）の分析による癌の検出について実施形態が提供される。無細胞性核酸は体中の様々な種類の組織に起源を有し得る。この方法では、様々な癌の検出のための広範な分析が実施され得る。

10

##### 【0026】

（一ヌクレオチド突然変異、欠失、増幅、および再構成を含む）遺伝的変化は癌の発生の際に腫瘍細胞において蓄積する。実施形態において、癌を検出およびモニターするためには大量平行シークエンシングを用いて体液（例えば、血漿、血清、唾液、腹水、胸膜液および脳脊髄液）の中で一ヌクレオチド変異（SNV）とも呼ばれる一ヌクレオチド突然変異（SNM）を検出および定量することができる。SNM（または他の種類の突然変異）数の定量はスクリーニング試験の一部として初期ステージの癌を特定するための機序を提供することができる。様々な実施形態において、（例えば、複数のSNM、例えば、少なくとも3、4、または5事例のSNMが特定の座位において特定されることを必要とすることにより）シークエンシングエラーを区別し、健常細胞において生じる自然突然変異を識別するために注意が払われる。

20

##### 【0027】

幾つかの実施形態により、同一の腫瘍内の細胞に関わることがあり得る腫瘍の不均一性（すなわち、腫瘍内の不均一性）または体の（同一部位または異なる部位の）異なる腫瘍に由来する細胞に関わることがあり得る腫瘍の不均一性を分析するための非侵襲的方法も提供される。例えば、各突然変異を含有する相対的腫瘍細胞量の推定を含む、そのように不均一な腫瘍のクローニング構造の非侵襲的な分析を行うことができる。高い相対的濃度で存在する突然変異ほど体内でより多くの数の悪性腫瘍細胞、例えば、体の中でも他の悪性腫瘍細胞と比べて腫瘍形成過程中的早い時期に生じた細胞に存在する（Welch JS et al. Cell 2012; 150: 264-278）。そのような突然変異は、比較的に多く存在するため、比較的に少ない突然変異よりも癌DNAを検出するより高い診断感度を示すことが予期される。突然変異の相対的存在量の変化を連続的にモニタリングすれば、疾患の進行により自然に生じるか、治療へ応答して生じる腫瘍のクローニング構造の変化を非侵襲的にモニターすることが可能になる。そのような情報は予後の評価または治療に対する腫瘍耐性の早期検出に有用になる。

30

##### 【0028】

#### I. 序論

突然変異はDNA複製および／またはDNA修復のエラーのため、細胞分裂中に生じ得る。1つの種類のそのような突然変異は一ヌクレオチドの変化を伴い、その変化はゲノムの様々な部分の複数の配列に関わり得る。癌は増殖優位性を獲得した単一癌細胞のクローニング増殖に起因すると一般に考えられている。このクローニング増殖は祖先癌細胞を起源とする全ての癌細胞において突然変異（例えば、一ヌクレオチド突然変異）の蓄積をもたらすことになる。これらの子孫腫瘍細胞は一組の突然変異（例えば、一ヌクレオチド突然変異）を共有する。本明細書に記載されるように、癌関連一ヌクレオチド突然変異は癌患者の血漿／血清において検出可能である。

40

##### 【0029】

幾つかの実施形態により生物試料（例えば、血漿または血清）における全ての突然変異を効果的にスクリーニングすることができる。突然変異の数が固定されていない（様々な亜集団の腫瘍細胞に由来する数百、数千、または数百万の癌関連突然変異が検出され得る

50

) ので、実施形態により特定の突然変異を検出する技術よりも良い感度が提供され得る。突然変異の数を用いて癌を検出することができる。

#### 【 0 0 3 0 】

多数または全ての突然変異のそのようなスクリーニングを実施するため、実施形態により、腫瘍由来DNAを含有している可能性がある生物試料（例えば、血漿および血清を含む体液）中における遺伝的変異の検索（例えば、無作為検索）を実施することができる。試料、例えば血漿の使用は腫瘍または癌の侵襲的生検を実施する必要性を不要にする。また、スクリーニングがゲノムの全領域または広い領域を範囲とすることができますとき、そのスクリーニングは列挙可能であり、且つ、公知のどのような突然変異にも限定されず、あらゆる突然変異の存在を用いることができる。また、突然変異の数はゲノムの全領域または広い領域にわたって合計されるので、より高い感度を得ることができる。

10

#### 【 0 0 3 1 】

しかしながら、ヒトゲノムには一ヌクレオチド多型（SNP）を含む多型性部位が存在し、その部位は突然変異に数えられるべきではない。実施形態により、検出された遺伝的変異が癌関連突然変異である可能性があるのか、またはゲノム中の多型であるのか確認することができる。例えば、癌関連突然変異とゲノム中の多型の間の決定の一部として、実施形態により構成的ゲノムを決定することができ、その構成的ゲノムは多型を含むことがあり得る。構成的ゲノム（CG）の多型はシークエンシングデータ中に充分に高いパーセンテージ（例えば、30～40%）で示される多型に限定され得る。

20

#### 【 0 0 3 2 】

そこで、生物試料から得られた配列を構成的ゲノムに整列させることができ、一ヌクレオチド突然変異（SNM）または他の種類の突然変異である変異を特定することができる。これらのSNMは公知の多型に含まれない変異であり、したがって、癌関連であり、構成的ゲノムの一部ではないと表示され得る。健常な人は、例えば、細胞分裂の間に生じた健常な細胞での無作為突然変異のためにある特定の数のSNMを有することがあり得るが、癌を有する人であればより多数のSNMを有するであろう。

30

#### 【 0 0 3 3 】

例えば、癌を有する人にとっては、体液において検出可能なSNMの数は同一人物の構成的ゲノムに存在する多型よりも多いことになる。腫瘍由来DNAを含有する体液試料と主として構成的DNAを含有するDNA試料において検出される変異の量の間で比較を行うことができる。1つの実施形態において、「主として」という用語は90%超を意味することになる。別の好ましい実施形態において、「主として」という用語は95超、97%超、98%超、または99%超を意味するだろう。体液における変異の量が主として構成的DNAの試料の変異の量を超えるとき、体液が腫瘍由来DNAを含有する可能性の上昇が存在する。

#### 【 0 0 3 4 】

DNA試料における変異を無作為に検索するために用いられ得る一つの方法は（例えば、大量平行シークエンシングを用いる）ランダムシークエンシングまたはショットガンシークエンシングである。シークエンシング・バイ・ライゲーションプラットフォーム（例えば、Life Technologies社のSOLIDプラットフォーム）、Ion Torrent/Ion Proton、半導体シークエンシング、Roche 454、一分子シークエンシングプラットフォーム（例えば、Helicos社、Pacific Biosciences社およびナノポア社）を含むあらゆる大量平行シークエンシングプラットフォームを用いることができる。しかし、シークエンシングエラーが生じることがあり得、それが構成的DNA中の変異として、または腫瘍DNAに由来する突然変異として誤解釈され得ることが知られている。したがって、我々の提唱するアプローチの特異度を改善するため、例えば、ある座位で検出される少なくとも特定の数（例えば、2または3）のアレルがSNMとして数えられることを必要とすることと共に適切なシークエンス深度を用いることによってシークエンシングエラーまたは分析エラーの他の要素が起こる見込みを考慮することができる。

40

50

## 【0035】

本明細書に記載されるように、無作為に検出された試料中に存在する遺伝的変異の量が構成的DNAと分析エラー（例えば、シークエンシングエラー）のために不注意にも検出されることがあり得る変異について予期される遺伝的变化の量を超えるとき、実施形態により、生物試料（例えば、体液）中の腫瘍由来DNAの存在についての証拠が提供され得る。その情報を癌のスクリーニング、診断、予後予測およびモニタリングのために用いることもできる。以降の節において、我々は、血漿／血清または他の試料（例えば、体液）における一ヌクレオチド突然変異の検出のために使用され得る分析ステップを説明する。体液には血漿、血清、脳脊髄液、胸膜液、腹水、乳頭分泌物、唾液、気管支肺胞洗浄液、痰、涙、汗、および尿が含まれ得ることになる。大便試料は大腸癌由来の腫瘍DNAを含有することが示されているので(Berger BM, Ahlquist DA. Pathology 2012; 44: 80-88)、その技術を体液に加えて大便試料にも適用することができる。

10

## 【0036】

## I I . 一般的なスクリーニング法

図1は本発明の実施形態に従って対象において癌または前癌変化を検出するための方法100のフローチャートである。実施形態により、対象に由来する生物試料中の無細胞性DNAを分析して腫瘍の結果生じる可能性がある無細胞性DNA中の変異を検出することができる。その分析は、健常細胞の一部である多型を考慮するために対象の構成的ゲノムを使用することができ、シークエンシングエラーを考慮することができる。1種類以上のプロセッサーを含むコンピューターシステムを使用して方法100と本明細書に記載される方法のうちのいずれかを全体的または部分的に実行することができる。

20

## 【0037】

ステップ110では、対象の構成的ゲノムを得る。試験対象の構成的DNAから構成的ゲノム(CG)を決定することができる。様々な実施形態において、メモリーからCGを読むことができ、または、例えば、無細胞性DNAを含む試料に由来する細胞中にあり得る構成的DNAの配列リードを分析することにより能動的にCGを決定することができる。例えば、非血液腫瘍が疑われるときに血液細胞を分析して対象の構成的DNAを決定することができる。

## 【0038】

様々な実施形態において、大量平行シークエンシング、アレイベース・ハイブリダイゼーション、プローブベース・インソリューション・ハイブリダイゼーション、ライゲーションベース・アッセイ、プライマー伸長反応アッセイ、およびマススペクトロメトリーを用いると構成的DNAの分析を実施することができる。1つの実施形態において、対象の生涯のある時点、例えば、生誕時、または出生前期においてさえ（それは胎児細胞を用いて、または無細胞性DNA断片を介して実施され得ることになる。米国特許出願公開第2011/0105353号明細書を参照のこと）CGを決定することができ、その後、CGは対象の生涯の他の時点で体液または他の試料を得たときのものに対して参照され得る。したがって、コンピューターメモリーからCGを単に読んでもよい。構成的ゲノムが基準ゲノムと異なる座位のリストとして構成的ゲノムを読むことができる。

30

## 【0039】

ステップ120では、対象の生物試料中の複数のDNA断片のそれぞれについて1種類以上の配列タグが受領され、その場合、その生物試料は無細胞性DNAを含む。1つの実施形態において、生物試料中のDNA断片のランダムシークエンシングから1種類以上の配列タグを作製する。ペアドエンド(paired-end)シークエンシングが実施されると1つより多くの配列タグを得ることができる。1種類のタグはDNA断片の各末端に対応することになる。

40

## 【0040】

試料（例えば、血漿、血清または他の体液）中の無細胞性DNAを分析して遺伝的変異を検索することができる。構成的DNAを分析するのに使用したのと同じ分析プラットフォームを用いて無細胞性DNAを分析することができる。あるいは、異なる分析プラット

50

フォームを用いることだってできる。例えば、大量平行シークエンシングを用いて無細胞性DNA試料の配列を決定することができ、または大量平行シークエンシングの前にゲノムの部分を捕捉または濃縮することだってできる。濃縮を用いる場合、例えば、選択された部分のゲノムの溶液相捕捉または固相捕捉を用いることだってできる。その後、捕捉したDNAについて大量平行シークエンシングを行うことができる。

#### 【0041】

ステップ130では、配列タグのゲノム上の位置を決定する。1つの実施形態において、配列タグが、1種類以上の他の対象から得られる基準ゲノムに対して整列させられる。別の実施形態において、ゲノム配列タグが試験対象の構成的ゲノムに対して整列させられる。当業者に知られている技法を用いて、例えば、基礎局所的アラインメント検索ツール(BLAST)を使用してアラインメントを実施することができる。

10

#### 【0042】

ステップ140では、少なくともN個の配列タグが構成的ゲノム(CG)と比べて配列異型を有する場合に第1座位数が決定される。Nは2以上である。下でより詳細に考察するように、2、3、4、5、またはそれより多くのNを有することにより、シークエンシングエラーならびに細胞中で(例えば、細胞分裂のために)無作為に生じる体細胞突然変異を除去することができる。異型を有するが、(例えば、ちょうど1種類の異型配列タグのように)1種類以上の判定基準を満たすことがない座位は潜在的または仮想突然変異と呼ばれるが、1種類以上の特定の判定基準を満たす座位を突然変異(異型)または突然変異座位(異型座位)として特定することができる。配列異型はちょうど1つのヌクレオチドまたは複数のヌクレオチドの異型であることだってあり得る。

20

#### 【0043】

Nは、絶対値と対照的に、ある座位の全タグのパーセンテージとして決定され得る。例えば、異型リードから推定される腫瘍DNAの分画濃度が10%(または他のあるパーセンテージ)以上であると判断されると、異型座位が特定され得る。言い換えると、200配列リードで座位がカバーされるとき、異型アレルを示す少なくとも10個の配列リードという判定基準がその異型を突然変異として定義するために必要とされ得る。10配列リードの異型アレルと190リードの野生型アレルであれば10%( $2 \times 10 / (10 + 190)$ )という腫瘍DNAの分画濃度を示す。

30

#### 【0044】

1つの実施形態において、配列タグ(集合的に試料ゲノムと呼ばれる)を直接的にCGと比較して異型を決定することができる。別の実施形態において、基準ゲノム(RG)を介して試料ゲノム(SG)をCGと比較して異型を決定する。例えば、CGとSGの両方をRGと比較して異型を示す座位のそれぞれの数(例えば、セット)を決定することができ、その後、第1座位数を得るために差分をとることができる。その第1数を数として簡単に得ることができ、またはその第1数は座位の特定のセットに対応することができ、その第1数は第1座位における配列タグからパラメーターを決定するためにさらに分析され得る。

30

#### 【0045】

1つの実施形態において、構成的DNAと血漿中DNAのシークエンシングの結果を比較して一ヌクレオチド突然変異が血漿中DNAに存在するか決定する。構成的DNAがホモ接合性である領域を分析することができる。例示目的に、特定の座位の遺伝子型が構成的DNAにおいてホモ接合性であり、AAであると仮定しよう。その場合、血漿中にA以外のアレルが存在すると特定の座位に一ヌクレオチド突然変異(SNM)が存在する可能性を示す。SNMの存在の可能性を示す座位はステップ140において第1座位数を形成し得る。

40

#### 【0046】

1つの実施形態において、特定の種類の癌または特定の集団のサブセットで突然変異を特に受けやすいと知られているゲノムの部分を標的にすることが有用であることがあり得る。後者の態様との関連で、実施形態により、特定の人種で特に普遍的な種々の突然変異

50

、例えば、(肝臓癌について) B型肝炎ウイルスまたは(子宮頸部癌について)ヒトパピローマウイルスを担持する対象、または体細胞突然変異を生じやすい遺伝的素質を有する対象、またはDNAミスマッチ修復遺伝子に生殖細胞系突然変異を有する対象に特に共通する突然変異を探し求めることができる。その技術はB R C A 1 突然変異またはB R C A 2 突然変異を有する対象における卵巣癌および乳癌の中の突然変異をスクリーニングすることにも有用であるだろう。その技術はA P C 突然変異を有する対象における大腸癌の中の突然変異をスクリーニングすることにも同様に有用であるだろう。

#### 【0047】

ステップ150では、第1座位に配列異型を有する配列タグの数に基づいてパラメーターを決定する。一例では、そのパラメーターは、少なくともN個のDNA断片がある座位において構成的ゲノムと比べて配列異型を有する場合の第1座位数である。したがって、ある座位がその第1数に含まれる前に特定されたNコピーより多くの特定の異型を有することを確実にするためにその配列タグの数を単純に使用することができる。別の実施形態において、そのパラメーターは第1座位において構成的ゲノムと比べて配列異型を有する配列タグの総数であり得る、またはその総数を含み得る。

10

#### 【0048】

ステップ160では、対象についてのパラメーターを(例えば、1つ以上の他の対象に由来する)閾値と比べてその対象におけるある分類の癌のレベルを決定する。癌のレベルの例には対象が癌または前癌状態を有しているか、または癌を発生する可能性が上昇しているかが含まれる。1つの実施形態において、その閾値はその対象から以前に得られた試料から決定され得る。

20

#### 【0049】

別の実施形態において、前記の1つ以上の他の対象は癌を有していない、または癌の低いリスクを有すると判断され得る。したがって、閾値は正常値、正常範囲であり得、または正常値または正常範囲からの統計的に有意な偏差を示し得る。例えば、試験対象において検出された突然変異の数が正常であるか決定するために、癌を有しないまたは癌の低いリスクを有する対象の血漿において検出可能である、特定の対象のCGと比べた突然変異の数を正常範囲として使用することができる。別の実施形態において、前期の他の対象は癌を有していると知られていてもよく、したがって、同様の数の突然変異が癌を表し得る。

30

#### 【0050】

1つの実施形態において、前期の他の対象は、試験対象の臨床的特徴、例えば、性別、年齢、食習慣、喫煙習慣、薬歴、既往症、家族健康歴、選択されたゲノム座位の遺伝子型、ウイルス感染(例えば、B型肝炎ウイルスまたはC型肝炎ウイルスまたはヒトパピローマウイルスまたはヒト免疫不全ウイルスまたはエプスタイン・バールウイルスの感染)または(細菌(例えば、ヘリコバクター・ピロリ)および寄生生物(例えば、肝臓ジストマ(Clonorchis sinensis)などのような)他の感染因子の感染の状態に合致する臨床的特徴を有するように選択され得る。例えば、B型肝炎ウイルスまたはC型肝炎ウイルスを保持する対象は肝細胞癌を発生する上昇したリスクを有する。したがって、B型肝炎またはC型肝炎と同様の突然変異の数またはパターンを有する試験対象は肝細胞癌を発生する上昇したリスクを有するとみなされ得る。一方、別の肝炎患者よりも多くの突然変異を示すB型肝炎またはC型肝炎の患者は、適切なベースライン(すなわち、別の肝炎患者と比べたもの)を使用するので、より高い分類の癌のレベルを有すると適切に特定され得る。同様に、ヒトパピローマウイルス感染を保持する対象は子宮頸部癌および頭頸部癌の上昇したリスクを有する。エプスタイン・バールウイルスの感染は鼻咽頭癌、胃癌、ホジキンリンパ腫および非ホジキンリンパ腫と関連している。ヘリコバクター・ピロリの感染は胃癌と関連している。肝臓ジストマの感染は胆管癌と関連している。

40

#### 【0051】

癌の進行と治療応答のモニタリングに様々な時点における突然変異の数の変化のモニタリングを用いることができる。前癌状態の進行、または対象が癌を発生するリスクの変化

50

を詳細に記録するためにそのようなモニタリングを用いることもできる。

【0052】

モニタリングに変異を示す配列タグの量を用いることもできる。例えば、ある座位における異型リードの分画濃度を使用することができる。1つの実施形態において、連続的モニタリング中の試料における腫瘍関連遺伝的变化の分画濃度の上昇は、疾患の進行または切迫した疾患の再発を表し得る。同様に、連続的モニタリング中の試料における腫瘍関連遺伝的变化の分画濃度の低下は、治療に対する応答、および/または疾患の寛解、および/または良好な予後を表し得る。

【0053】

III. ゲノムの決定

10

以下に、上で考察した様々なゲノムをさらに詳細に説明する。例えば、基準ゲノム、構成的ゲノム、および試料ゲノムを考察する。

【0054】

A. 基準ゲノム

基準ゲノム (RG) は対象の半数体ゲノムまたは二倍体ゲノムまたは集団のコンセンサスを指す。基準ゲノムは公知であり、したがって、新しい患者に由来するシークエンシングリードを比較するために使用され得る。患者の試料に由来する配列リードを整列および比較して RG に由来するリード中の変異を特定することができる。半数体ゲノムについては各座位にたった1つのヌクレオチドが存在し、したがって、各座位をヘミ接合性とみなすことができる。二倍体ゲノムについてはヘテロ接合性座位が特定され得、そのような座位は2つのアレルを有し、どちらかのアレルがアラインメントで前記の座位に合致し得る。

20

【0055】

ある集団の対象の間で基準ゲノムが同一であり得る。患者の分類（例えば、癌を有するか否か）のために使用される適切な閾値を決定するためにこの同一の基準ゲノムを健常対象に使用することができる。しかしながら、異なる集団、例えば、異なる民族に、または異なる家族にも異なる基準ゲノムを使用することができる。

【0056】

B. 構成的ゲノム

30

対象（例えば、ヒトまたは他の二倍体生物）の構成的ゲノム (CG) はその対象の二倍体ゲノムを指す。CG は、第1アレルが第1ハプロタイプに由来し、異なる第2アレルが第2ハプロタイプに由来する場合にヘテロ接合性座位を表し得る。2つのヘテロ接合性座位をカバーする2つのハプロタイプの構造は公知である必要がないこと、すなわち、一方のヘテロ接合性座位上のどのアレルが別のヘテロ接合性座位のアレルと同一のハプロタイプであるか公知である必要がないことに留意されたい。各ヘテロ接合性座位における2つのアレルの存在そのものが充分であり得る。

【0057】

多型のために CG は RG と異なり得る。例えば、RG 上の座位は T でホモ接合性であり得るが、CG は T / A でヘテロ接合性である。したがって、CG はこの座位において変異を示すことになる。（例えば、家族内で受け継がれる）遺伝性突然変異または（胎児において生じるが、その親には存在しない）新規突然変異のために CG は RG と異なることもあります。遺伝性突然変異は通常は「生殖細胞系突然変異」と呼ばれる。そのような突然変異のうちの幾つかは家族内で受け継がれる B R C A 1 突然変異のような癌体质と関連する。そのような突然変異は人の生存期間中の細胞分裂に起因して生じることがあり、癌になる途中で細胞とその子孫を押しやることがある「体細胞突然変異」とは異なる。

40

【0058】

CG 決定の目標は、体細胞突然変異を特定するためにそのような生殖細胞系突然変異と新規突然変異を試料ゲノム (SG) の突然変異から除去することである。その後、対象における癌の可能性を評価するために SG における体細胞突然変異の量を使用することができる。シークエンシングエラーを除去し、可能性としては滅多に起こらない（例えば、1

50

リードしか異型を示さない) 体細胞突然変異であって、したがって癌に関連しそうにない体細胞突然変異を除去するために、これらの体細胞突然変異をさらにフィルターにかけることができる。

【0059】

1つの実施形態において、細胞(バフィーコートDNA)を使用してCGを決定することができる。しかしながら、無細胞性DNA(例えば、血漿または血清)から同様にCGを決定することもできる。細胞の大半が非悪性腫瘍である種類の試料、例えば、健常対象に由来するバフィーコートについて、大部分のゲノムまたはコンセンサスゲノムがCGである。そのCGでは、各ゲノム座位は、試料抽出された組織中の細胞の大部分によって保持されるDNA配列からなる。シークエンス深度は、構成的ゲノム内のヘテロ接合性部位を解明するのに充分であるべきである。

10

【0060】

別の例として、CGを決定するための構成的試料として血漿を使用することができる。例えば、血漿中の腫瘍DNAが50%未満であり、SNMがヘテロ接合性状態である、例えば、突然変異が新しいアレルの付加である場合では、その新しいアレルは25%未満の濃度を有し得る。一方、CG中のSNPのヘテロ接合性アレルの濃度は約50%に達するはずである。こうして、CGの体細胞突然変異と多型の間で区別をつけることができる。1つの実施形態において、適切なカットオフは、血漿、または著しい腫瘍濃度を有する他の混合物を使用するときに多型から体細胞突然変異を決定するために30~40%の間であり得る。血漿中の腫瘍DNAが50%未満であることを確実にするために腫瘍DNA濃度の測定が有用であり得る。腫瘍DNA濃度の決定の例は本明細書に記載されている。

20

【0061】

C. 試料ゲノム

試料ゲノム(SG)はRGとCGの場合のように単に半数体ゲノムまたは二倍体ゲノムというわけではない。SGは試料に由来するリードのコレクションであり、CGに相当する構成的DNAに由来するリード、腫瘍DNAに由来するリード、(例えば、細胞分裂により生じる突然変異に起因する)CGと比べた無作為突然変異を示す健常細胞に由来するリード、およびシークエンシングエラーを示す健常細胞に由来するリードを含み得る。どのリードがSGに含まれているのか正確に管理するために様々なパラメーターを使用することができる。例えば、あるアレルが少なくとも5リードを示すことを必要とすることにより、SG中に存在するシークエンシングエラーを減少させることができ、ならびに無作為突然変異に起因するリードを減少させることができる。

30

【0062】

例として、対象が健康であり、すなわち、癌を有しないと仮定しよう。例示目的に、1000細胞に由来するDNAがこの対象から得られた1mlの血漿中にある(すなわち、1000ゲノム等量のDNA)。血漿DNAは通常は約150bpのDNA断片からなる。ヒトゲノムは $3 \times 10^9$ bpであるので、半数体ゲノム当たり約 $2 \times 10^7$ DNA断片が存在することになる。ヒトゲノムは二倍体であるので、1mlの血漿当たり約 $4 \times 10^7$ DNA断片が存在することになる。

40

【0063】

数百万~数十億の細胞が単位時間当たりに血漿中にこれらのDNAを放出しており、これらの細胞に由来する断片が循環中に混合することから、 $4 \times 10^7$ DNA断片が $4 \times 10^7$ 個の異なる細胞に由来することができただろう。これらの細胞が互いに(遠い、例えば、元の接合子と対照的な)最近のクローン関係を保持しない場合(すなわち、これらの細胞が最近の祖先細胞を共有しない場合)、これらの断片の間で1回よりも多く突然変異が見られることが無い可能性が統計的に存在する。

【0064】

一方、1mlの血漿中DNA当たり1000ゲノム等量の中に最近の祖先細胞を共有するある特定のパーセンテージの細胞が存在する場合(すなわち、これらの細胞がクローニング的に互いに関連する場合)、その血漿中DNAにおいて優先的に代表される(例えば、血

50

漿中のクローン性突然変異プロファイルを示す)このクローンに由来する突然変異を見ることがあるだろう。そのようなクローン的に関連する細胞は癌細胞であり得、または癌になる途中であるが、まだそこまで至っていない(すなわち、前癌状態の)細胞であり得る。したがって、突然変異が1回より多く現れることを必要とすることにより、試料中に特定された「突然変異」の中のこの自然分散を取り除くことができ、それにより癌細胞または前癌細胞に関連するより多くの突然変異を残すことができ、そして癌または前癌状態の検出、特に早期検出を可能にする。

#### 【0065】

おおまかにいうと、各細胞分裂の後に平均的に1事例の突然変異がゲノムに蓄積することができる述べられている。これまでの研究により、血漿中DNAの大半が造血細胞に由来することが示されている(Lui YY et al. Clin Chem 2002; 48: 421-427)。造血幹細胞は25~50週毎に1回複製すると推定されている(Catlin SN, et al. Blood 2011; 117: 4460-4466)。したがって、単純化しておおまかにいうと、健康な40歳の対象であれば造血幹細胞当たり約40~80事例の突然変異を蓄積しているだろう。

10

#### 【0066】

この人物の血漿中に1ml当たり1000ゲノム等量が存在し、且つ、これらの細胞のそれぞれが異なる造血幹細胞に由来する場合、 $4 \times 10^{10}$  DNA断片(すなわち、ゲノム当たり $4 \times 10^7$  DNA断片、および1mlの血漿当たり1000ゲノム等量)の中に40,000~80,000事例の突然変異が期待され得る。しかしながら、各突然変異が1回見られるので、(例えば、カットオフ値Nが1よりも大きい場合に)各突然変異はまだ検出限界未満であり得、したがって、これらの突然変異がフィルタリングで除外されることがあります、それによって癌状態の結果である可能性がより高い突然変異に分析を集中することが可能になる。カットオフ値は1よりも大きいあらゆる値(整数または非整数)であり得、且つ、異なる座位および領域で変動し得る。シークエンス深度と腫瘍DNAの分画濃度も癌細胞または前癌細胞から突然変異を検出する感度(例えば、検出可能な突然変異のパーセンテージ)に影響し得る。

20

#### 【0067】

##### I V . CGとの直接比較

幾つかの実施形態により、CGはホモ接合性であるがSG中の少数の種(すなわち、腫瘍DNA)はヘテロ接合性であるヌクレオチド位置を特定することができる。高い深度(例えば、50倍を超えるカバー度)で位置をシークエンシングすると、健常細胞および癌細胞のDNA混合物においてその位置に1または2つのアレルが存在するか検出することができる。2つのアレルがそこで検出されるとき、(1) CGがヘテロ接合性であるか、または(2) CGがホモ接合性であるが、SGがヘテロ接合性であるかのどちらかである。これらの2つのシナリオは主要なアレルと少数派のアレルの相対的な数を調べることによって区別され得る。前者のシナリオであれば、それらの2つのアレルは同様なカウント数であるが、後者のシナリオであれば、それらのカウント数に大きな差が存在する。試験試料に由来するリードの相対的アレル数のこの比較は構成的ゲノムに対して配列タグを比較する1つの実施形態である。方法100の第1座位は、アレルの数が上限閾値(CGにおける多型に対応する閾値)よりも小さく、下限閾値(エラー、および癌状態と関連しないほど充分に低い割合で生じる体細胞突然変異に対応する閾値)よりも大きい場合の座位として決定され得る。したがって、構成的ゲノムと第1座位は同時に決定され得る。

30

#### 【0068】

別の実施形態において、突然変異を特定するための過程により、CGを最初に決定し、次にCGと比べて充分な数の突然変異を有する座位を決定することができる。CGは試験試料と異なる構成的試料から決定され得る。

40

#### 【0069】

図2は、本発明の実施形態に従って試料ゲノム(SG)を直接的に構成的ゲノム(CG)と比較する方法200のフローチャートを示す。プロック210において、対象の構成的ゲノムを獲得する。構成的ゲノムは、例えば、以前の時点で採取された試料または方法

50

200が実施される直前に入手および分析される構成的試料から獲得され得る。

【0070】

ブロック220において、対象の生物試料中の複数のDNA断片のそれぞれについて1種類以上の配列タグが受領される。本明細書において記述されるような様々な技法を用いてシークエンシングを実施することができる。配列タグは、断片の配列がそうであると信じられているものの測定値である。しかし、配列タグの1種類以上の塩基がエラーであり得る。

【0071】

ブロック230において、配列タグの少なくとも一部を構成的ゲノムと整列させる。そのアラインメントはCGが様々な座位においてヘテロ接合性であることを考慮することができ得る。そのアラインメントは完全一致を必要としないので、異型が検出されるだろ  
う。

10

【0072】

ブロック240において、構成的ゲノムと比べてある座位において配列異型を有する配列タグが特定される。配列タグは1つよりも多くの異型を有することができただろう。各座位と各配列タグについての異型を追跡することができる。異型は、CG中に無いあらゆるアレルであり得る。例えば、CGはA/Tでヘテロ接合性であり、異型はGまたはCであり得る。

【0073】

ブロック250において、コンピューターシステムは異型を有する各座位について、その座位に整列し、且つ、その座位において配列異型を有する配列タグの各第1数を数えることができる。したがって、各座位はその座位において見られる異型の関連するカウント数を有し得る。典型的には、例えば、50%未満である腫瘍DNA濃度のためにある座位において、CGに対応する配列タグと比べて、より少ない数の異型が見られる。しかしながら、幾つかの試料は50%よりも高い腫瘍DNAの濃度を有し得る。

20

【0074】

ブロック260において、各第1数に基づいてパラメーターを決定する。1つの実施形態において、各数がカットオフ値よりも大きい（例えば、2よりも大きい）場合、その各数を合計に加えることができ、その合計がパラメーターである、またはそのパラメーターを決定するために使用される。別の実施形態において、カットオフ値よりも大きい各数を有する座位の数をパラメーターとして使用する。

30

【0075】

ブロック270において、癌のレベルを分類するためにパラメーターを閾値と比較する。上に記載したように、他の対象に由来する試料の分析から閾値を決定することができる。これらの他の対象の健常状態または癌状態に応じて分類を決定することができる。例えば、前期の他の対象がステージ4の癌を有していて、現在のパラメーターが前期の他の対象から得られたパラメーターの値に近かった（例えば、特定の範囲内にあった）場合、現在の対象はステージ4の癌を有すると分類された可能性がある。しかしながら、パラメーターが閾値を超える（すなわち、どのようにパラメーターが定義されるかに応じて、閾値よりも大きい、または小さい）場合、その分類はステージ4未満であると特定され得る。前期の他の対象が癌を有していないときに同様の分析を行うことができる。

40

【0076】

各閾値が異なるセットの対象から決定される場合に複数の閾値を使用して分類を決定することができる。各セットの対象は共通する癌のレベルを有することがあり得る。したがって、現在のパラメーターを各セットの対象の値と比較することができ、それらのセットのうちの1つに対しての一致を提供することができ、またはある範囲を提供することができる。例えば、そのパラメーターは前癌状態またはステージ2の対象について得られたパラメーターとほぼ等しい可能性がある。別の例として、現在のパラメーターは幾つかの異なる癌のレベルにおそらく一致し得る範囲にあり得る。したがって、その分類は1よりも高い癌のレベルを含み得る。

50

## 【0077】

## V. 基準ゲノムの使用

構成的DNAと生物試料由来のDNAの両方のゲノム配列をヒト基準ゲノムと比較することができる。基準ゲノムと比べて構成的DNAよりも多くの変化が血漿試料に存在するとき、癌が生じるより高い見込みがある。1つの実施形態において、基準ゲノム中のホモ接合性座位を試験する。構成的DNAと生物試料由来のDNAの両方におけるヘテロ接合性座位の量を比較する。生物試料のDNAから検出されたヘテロ接合性部位の量が構成的DNAから検出されたヘテロ接合性部位の量を超えるとき、癌が生じるより高い見込みがある。

## 【0078】

10

その分析は、CGにおいてホモ接合性である座位に限定されることもあり得る。SNMはヘテロ接合性座位についても定義され得るが、これは一般に第3の異型の生成を必要とすることになる。言い換えると、ヘテロ接合性座位がA/Tである場合、新しい異型はCまたはGのどちらかであることになる。ホモ接合性座位についてSNMを特定することが一般的により簡単である。

## 【0079】

構成的DNAと比べた生物試料DNAにおけるヘテロ接合性座位の量の増加の程度は、健常対象において見られる変化の割合と比較されると、癌状態または前癌状態を示唆し得る。例えば、そのような部位における増加の程度が健常対象において観察される増加の程度をある特定の閾値で超える場合、癌状態または前癌状態を示唆するとそのデータをみなすことができる。1つの実施形態において、癌を有しない対象における突然変異の分布が確認され、ある特定の数の標準偏差（例えば、2倍または3倍の標準偏差）として閾値を探ることができる。

20

## 【0080】

1つの実施形態は、ある座位における少なくとも特定の数の異型をその座位が数えられる前に必要とし得る。別の実施形態は一度変化を見ることに基づくデータに関する試験でさえ提供する。例えば、血漿中に見られる変異（エラー+本当の突然変異または多型）の総数が構成的DNAにおける変異の総数よりも統計的に有意に大きいとき、癌の証拠が存在する。

## 【0081】

30

図3は、本発明の実施形態に従って基準ゲノム（RG）を使用して試料ゲノム（SG）を構成的ゲノム（CG）と比較する方法300のフローチャートを示す。方法300は、RGが既に獲得されていると考え、且つ、生物試料についての配列タグが既に受領されているものと考える。

## 【0082】

プロック310において、配列タグの少なくとも一部を基準ゲノムに整列させる。そのアラインメントは、変異が検出されているとミスマッチを認めることができる。基準ゲノムは対象と類似の集団に由来し得る。整列させられた配列タグは試料ゲノム（SG）を有效地に含む。

## 【0083】

40

プロック320において、可能性がある異型、例えば、一ヌクレオチド突然変異（SNM）の第1数（A）が特定される。それらの可能性があるSNMは、SGの配列タグがRGと異なるヌクレオチドを示す座位である。例えば、変異を示す配列タグの数がカットオフ値よりも大きくなくてはならないこと、およびRGにおいてある座位がホモ接合性であるかどうかということのよう他の判定基準を使用することができる。特定の座位がメモリーの中にそれらの座位を保存することにより特定され、追跡されるときに可能性があるSNMのセットをセットAと表すことができる。それらの特定の座位を決定することができ、または単にそのようなSNMの数を決定することができる。

## 【0084】

プロック330において、構成的試料に由来するDNA断片を配列決定することによつ

50

て得られた配列タグを基準ゲノムに整列させることにより構成的ゲノムを決定する。以前のあらゆる時点で、および以前のあらゆる時点で得られた構成的試料を用いてこのステップを実施することができただろう。アラインメントを以前に実施した場合は、メモリーから CG を単に読むことができる。1つの実施形態において、構成的試料は血液細胞であり得る。

#### 【0085】

ブロック 340において、CG の整列させられた配列タグが基準ゲノムと比べてある座位において異型（例えば、SNM）を有する場合の座位の第 2 数（B）が特定される。一組の座位が特異的に追跡される場合、B はただの数と対照的にそのセットを表し得る。

#### 【0086】

ブロック 350において、試料ゲノム中に存在するが CG には存在しない異型（SNM）を特定するためにセット A からセット B を減算する。1つの実施形態において、CG がホモ接合性であるヌクレオチド位置に SNM のセットが限定され得る。このフィルタリングを達成するために、CG がホモ接合性である特定の座位がセット C において特定され得る。別の実施形態において、CG がある座位においてホモ接合性ではない場合にその座位は第 1 数 A または第 2 数 B の中に数えられない。別の実施形態において、あらゆる公知の多型を（例えば、SNP データベースにおけるその存在により）フィルタリングで除外することができる。

#### 【0087】

1つの実施形態において、ブロック 350における減算は単に数の減算であり得、したがって、特定の潜在的な SNM は除去されず、単に値が減算される。別の実施形態において、その減算は（例えば、セット B がセット A のサブセットである場合に）セット A とセット B の間の差をとってセット B に存在しない特定の SNM を特定する。論理値では、これは [ A AND NOT (B) ] と表現され得る。結果生じた特定された異型のセットを C とラベル付けすることができる。パラメーターを数 C として決定することができ、またはセット C から決定することができる。

#### 【0088】

いくつかの実施形態において、突然変異の性質を考慮することができ、異なる荷重が異なる種類の突然変異によるものとされる。例えば、一般に癌と関連する突然変異により大きな加重（座位の相対的加重について言及するとき、重要度値とも呼ばれる）を与えることができる。そのような突然変異は腫瘍関連突然変異のデータベース、例えば、癌における体細胞突然変異カタログ（COSMIC）（[www.sanger.ac.uk/genetics/CGP/cosmic/](http://www.sanger.ac.uk/genetics/CGP/cosmic/)）の中に見出され得る。別の例として、非同義変化に関連する突然変異により大きな加重を与えることができる。

#### 【0089】

したがって、ある座位において異型を示すタグの数が別の座位におけるタグの数と異なる加重を有し得る場合にその第 1 数 A は加重された合計として決定され得るだろう。その第 1 数 A はこの加重合計を反映し得る。同様の計算を B に実施することができ、したがって、数 C とパラメーターがこの加重を反映し得る。別の実施形態において、特定の座位のセット C が決定されるときにこれらの加重が考慮される。例えば、加重合計がセット C の座位の数について決定され得る。他の本明細書に記載される方法にそのような加重を使用することができる。

#### 【0090】

よって、癌のレベルの分類を決定するために閾値と比較されるパラメーターは RG と比べた SG と CG の変異を示す座位の数であり得る。他の実施形態において、変異を示す（配列タグを介して数えられた）DNA 断片の総数を数えることができる。他の実施形態において、パラメーターを得るために別の処理法でそのような数を使用することができる。

#### 【0091】

1つの実施形態において、各座位における異型の濃度がパラメーターであり得、閾値と比較され得る。ある座位が（異型を示す特定の数のリードのカットオフに加えて）潜在的

10

20

30

40

50

異型座位であるか、数えられるべき座位を有しているか決定するためにこの閾値を使用することができる。その濃度もSNMの合計における加重係数として使用され得るだろう。

#### 【0092】

##### V I . カットオフ値を使用する偽陽性の減少

上で述べたように、前記のアプローチの感度を改善するために大きなゲノム領域（例えば、ゲノム全体）または多数のゲノム領域について多数の無細胞性DNA断片（例えば、血漿中の循環DNA）の中で一ヌクレオチド突然変異を調査することができる。しかしながら、シークエンシングエラーなどの分析エラーがこのアプローチの実現性、正確性、および特異性に影響し得る。ここで、我々はシークエンシングエラーの重要性を例示するために例として大量平行シークエンシングプラットフォームを用いる。Illuminaシークエンシング・バイ・シンセシス（sequencing-by-synthesis）プラットフォームのシークエンシングエラー率は配列決定されたシークエンシングヌクレオチド当たり約0.1%～0.3%である（Minoche et al. Genome Biol 2011, 12:R112）。シークエンシング・バイ・ライゲーションプラットフォーム（例えば、Life Technologies社のSOLIDプラットフォーム）、Ion Torrent/Ion Proton、半導体シークエンシング、Roche 454、一分子シークエンシングプラットフォーム（例えば、Helicos社、Pacific Biosciences社およびnanopore社）を含むあらゆる大量平行シークエンシングプラットフォームを用いることができる。

#### 【0093】

肝細胞癌についての以前の研究において、癌ゲノム全体について約3,000の一ヌクレオチド突然変異が存在することが示された（Tao Y et al. 2011 Proc Natl Acad Sci USA; 108: 12042-12047）。循環している全DNAのうちのわずかに10%が腫瘍細胞由来し、且つ、我々は平均して1倍の半数体ゲノムカバー度のシークエンス深度で血漿中DNAを配列決定すると仮定すると、我々はシークエンシングエラーのために900万（ $3 \times 10^9 \times 0.3\%$ ）の一ヌクレオチド変異（SNV）に出会うことになる。しかしながら、それらの一ヌクレオチド突然変異の大半は2本の相同染色体のうちのわずかに1本において生じることが予期される。100%腫瘍DNAを有する試料について1倍の半数体ゲノムカバー度のシークエンス深度で、我々は3,000突然変異のうちのわずかに半分、すなわち、1,500突然変異を検出すると予期する。我々が1倍の半数体ゲノムカバー度で10%の腫瘍由来DNAを含有する血漿試料を配列決定するとき、我々はわずかに150（ $1,500 \times 10\%$ ）癌関連一ヌクレオチド突然変異を検出すると予期する。したがって、癌関連突然変異の検出についてのシグナルノイズ比は60,000中の1である。この非常に低いシグナルノイズ比は、正常事例と癌事例を区別するためにこのアプローチを用いることの的確性は、我々が生物試料（例えば、血漿）における全ての一ヌクレオチド変化をパラメーターとして単純に使用した場合に非常に低いことを示唆する。

#### 【0094】

シークエンシング技術の進歩により、継続的なシークエンシングエラー率の低下があると予期される。1種類よりも多くのシークエンシングプラットフォームを用い、且つ、相互プラットフォームシークエンシングの結果の比較を介して同一の試料を分析し、シークエンシングエラーに影響されている可能性があるリードを正確に示すこともできる。別のアプローチは、同一の対象から異なる時点で採られた2つの試料を分析することである。しかしながら、そのようなアプローチは時間がかかる。

#### 【0095】

1つの実施形態において、癌患者の血漿中の一ヌクレオチド突然変異の検出においてシグナルノイズ比を増強する1つの方法は、試料中において同一の突然変異が複数回生じている場合にのみその突然変異を数えることである。選択されたシークエンシングプラットフォームにおいて、特定のヌクレオチド置換を伴うシークエンシングエラーはより一般的である可能性があり、試験対象と対照被検者の両方の試験試料と構成的DNA試料のシークエンシングの結果に影響することになる。しかしながら、一般に、シークエンシングエ

10

20

30

40

50

ラーは無作為に生じる。

【0096】

シークエンシングエラーを有する機会は、複数のDNA断片中の同一のヌクレオチド位置において同一の変化を観察するとき、指數関数的に低くなる。一方、試料中において本当の癌関連突然変異性変化を検出する機会はシークエンス深度と試料中の腫瘍DNAの分画濃度によって影響を受ける。複数のDNA断片においてその突然変異を観察する機会はシークエンス深度と腫瘍DNAの分画濃度と共に増加することになる。(血漿中のもののような)無細胞性腫瘍DNAを有する試料を使用する様々な実施形態において、その分画濃度は5%、10%、20%、および30%であり得る。1つの実施形態において、その分画濃度は50%未満である。

10

【0097】

図4は、本発明の実施形態に従って突然変異を試料中に存在するものと分類するための判定基準として異なる出現回数を使用して正確に特定された癌関連一ヌクレオチド突然変異の数を示す表400である。同一の分類基準に基づいて、シークエンシングエラーのために突然変異を有していると誤って特定されたヌクレオチド位置の数も示されている。シークエンシングエラー率は0.1%であると考えられる(Minoche et al. Genome Bio 2011, 12:R112)。試料中の腫瘍由来DNAの分画濃度は10%であると考えられる。

【0098】

図4は、試料中の腫瘍由来DNAの分画濃度が10%であると考えられるときに血漿中において検出された癌関連突然変異の数と偽陽性コールの数との間の比率が、試料において同一の変化が見られる回数であって、突然変異を定義するための回数の増加と共に指數関数的に上昇することになることを示している。言い換えると、癌突然変異検出の感度と特異度の両方が改善することになる。さらに、癌関連突然変異の検出感度はシークエンス深度によって影響を受ける。100倍の半数体ゲノムカバー度のシークエンシングにより、3,000突然変異のうちの2,205(73.5%)が、試料中の少なくとも4DNA断片における特定の突然変異の出現頻度の判定基準を用いても検出され得る。最小数の断片についての他の値、例えば、3、5、8、10、および10よりも大きい数を使用してもよい。

20

【0099】

図5は、試料中の腫瘍由来DNAの分画濃度が5%であると考えられるときに特定された偽陽性座位の予期される数と突然変異の予期される数を示す表500である。試料中の腫瘍由来DNAの分画濃度が低いほど癌関連突然変異の同一の検出感度を達成するために高いシークエンス深度が必要とされることになる。特異度を維持するためにより厳しい判定基準が必要とされることになる。例えば、10%の腫瘍DNA分画の状況で試料中に少なくとも4回出現するという判定基準の代わりに少なくとも5DNA断片において特定の突然変異が出現するという判定基準を用いることが必要とされることになる。表400と表500は、本明細書に記載されるように仮定または測定され得るカバー度の倍率と腫瘍DNA濃度の前提で使用するためのカットオフ値についてのガイダンスを提供する。

30

【0100】

突然変異を定義するために一ヌクレオチド変化を1回より多く検出するという判定基準を用いることの別の利点は、この基準によって非悪性腫瘍組織における一ヌクレオチド変化に起因する偽陽性の検出を最小にすることが予期されることである。正常細胞の体細胞分裂の際にヌクレオチド変化が生じ得るので、体の中の各健常細胞は多数の一ヌクレオチド変化を保持することがあり得る。これらの変化が偽陽性の結果になる可能性があり得る。しかしながら、細胞のそれらの変化は細胞が死んだときに血漿/血清中に存在することになる。異なる正常細胞が異なるセットの突然変異を保持することが予期される一方、1つの細胞で生じた突然変異が血漿/血清中において多数のコピーで存在する可能性はない。このことは、腫瘍増殖は本質的にクローン性であるので複数のコピーが血漿/血清中で見られることが予期される腫瘍細胞内での突然変異と対照的である。したがって、あるクローンに由来する複数の細胞が死に、それらのクローンを表すシグネチャ突然変異を放出

40

50

することになる。

【0101】

1つの実施形態において、シークエンシングの前に特定のゲノム領域について標的濃縮を実施することができる。この標的濃縮ステップは、実施されるシークエンシングの総量は同一で目的の領域のシークエンス深度を増大させることができる。さらに別の実施形態において、比較的に低いシークエンス深度を用いる1ラウンドのシークエンシングを最初に実施することができる。次に、少なくとも1つの一ヌクレオチド変化を示す領域を、より高い倍率のカバー度を有する第2ラウンドのシークエンシングのために濃縮することができる。その後、標的濃縮を用いたシークエンシングの結果について突然変異を定義するために複数回の出現という判定基準を適用することができる。

10

【0102】

VII. ダイナミック・カットオフ

上に記載したように、ある座位が数えられるべき突然変異（例えば、SNM）として適格であるか決定するために異型（可能性がある突然変異）を裏付けるリードの数に対するカットオフ値Nを使用することができる。そのようなカットオフを使用することにより偽陽性を減らすことができる。以下の考察は異なる座位に対するカットオフの選択の方法を提供する。以下の実施形態において、我々は1つの優勢な癌クローニングが存在すると仮定する。様々な量の腫瘍DNAを血漿に放出する複数の癌細胞クローニングを含むシナリオのために同様の分析を実施することができる。

【0103】

A. 血漿中で検出される癌関連突然変異の数

血漿中で検出可能な癌関連突然変異の数は、多数のパラメーター、例えば、（1）腫瘍組織中に存在する突然変異の総数が患者の血漿中で検出可能な腫瘍関連突然変異の最大数である、腫瘍組織中の突然変異の数（ $N_T$ ）；（2）血漿中の腫瘍由来DNAの分画濃度が高いほど血漿中で腫瘍関連突然変異を検出する機会が高くなる、血漿中の腫瘍由来DNAの分画濃度（ $f$ ）；（3）シークエンス深度とは、配列決定された領域が配列リードによってカバーされる回数のことであり、例えば、10倍の平均シークエンス深度とは配列決定された領域内の各ヌクレオチドが平均して10配列リードによってカバーされるという意味であり、シークエンス深度が増加すると癌関連突然変異を検出する機会が増大することになる、シークエンス深度（ $D$ ）；および（4）本当の癌関連突然変異とシークエンシングエラーを区別するために使用されるカットオフ値である、潜在的癌関連突然変異としてヌクレオチド変化を定義するために血漿中において検出されるそのヌクレオチド変化の最小の回数（ $r$ ）によって影響を受けることがあり得る。

20

【0104】

1つの実施形態において、血漿中において検出される癌関連突然変異の数を予想するためにポアソン分布を使用する。ある突然変異が2本の相同染色体のうちの1本の上のヌクレオチド位置に存在すると考えると、 $D$ というシークエンス深度で突然変異が血漿に存在する予期される回数（ $M_p$ ）は、 $M_p = D \times f / 2$ として計算される。

30

【0105】

特定の突然変異部位において血漿中でその突然変異を検出する確率（ $P_b$ ）は、  
【数1】

40

$$P_b = 1 - \sum_{i=0}^{r-1} \text{Poisson}(i, M_p)$$

として計算され、その式では、 $r$ （カットオフ値）は、潜在的腫瘍関連突然変異としてヌクレオチド変化を定義するために血漿中において見られるそのヌクレオチド変化の回数であり；Poisson（ $i$ ,  $M_p$ ）は平均で $M_p$ 回検出する場合に $i$ 回検出するときのポア

50

ソン分布確率である。

【0106】

血漿中において検出されることが予期される癌関連突然変異の総数 ( $N_P$ ) は、 $N_P = N_T \times P_b$  として計算され得、その式では、 $N_T$  は腫瘍組織中に存在する突然変異の数である。以下のグラフは、潜在的突然変異を呼び出すための出現回数 ( $r$ ) の様々な判定基準および様々なシークエンス深度を用いて血漿中において検出されることが予期される腫瘍関連突然変異のパーセンテージを示す。

【0107】

図 6 A は、腫瘍由来 DNA の 10 % および 20 % の血漿分画濃度を有する血漿における、可能性がある癌関連突然変異を呼び出すための判定基準として 4 回および 6 回の出現回数 ( $r$ ) を使用した癌関連突然変異の検出率を示すグラフ 600 である。同一の  $r$  を使用すると、血漿中の腫瘍由来 DNA の分画濃度が高いほど血漿中で検出可能な癌関連突然変異の数が大きくなるという結果になる。血漿中の腫瘍由来 DNA の同一の分画濃度を使用すると、 $r$  が大きくなるほど検出される突然変異の数が小さくなるという結果になる。

10

【0108】

B. エラーのために検出される偽陽性ヌクレオチド変化の数

血漿中 DNA シークエンシングデータ中のヌクレオチド変化はシークエンシングエラーとアラインメントエラーに起因して生じ得る。偽陽性ヌクレオチド変化を有するヌクレオチド位置の数は二項分布に基づいて数学的に予想され得る。偽陽性部位 ( $N_{FP}$ ) の数に影響するパラメーターには、(1) シークエンシングエラー率が不正確な配列決定されたヌクレオチドの割合として定義される、シークエンシングエラー率 ( $E$ ) ; (2) 高いシークエンス深度を用いるほど、シークエンシングエラーを示すヌクレオチド位置の数が増加することになる、シークエンス深度 ( $D$ ) ; (3) 潜在的癌関連突然変異を定義するための同一のヌクレオチド変化の最小の出現回数 ( $r$ ) ; および (4) 目的の領域内のヌクレオチド位置の総数 ( $N_l$ ) が含まれ得る。

20

【0109】

突然変異の発生は無作為な過程として一般にみなされ得る。それ故、潜在的突然変異を定義するための出現の判定基準が上昇すると、偽陽性ヌクレオチド位置の数が  $r$  と共に指數関数的に減少することになる。既存のシークエンシングプラットフォームのうちの幾つかでは、ある特定の配列関係はよりシークエンシングエラーを有する傾向にある。そのようなシークエンシング関係の例には GGC モチーフ、ホモポリマー (例えば、AAAAA A A)、および単純リピート (例えば、ATATATATAT) が含まれる。これらの配列関係はヌクレオチド変化または挿入 / 欠失人為産物を実質的に増加させる (Nakamura K et al. Nucleic Acids Res 2011;39:e90 および Minoche AE et al. Genome Biol 2011;12,R112)。さらに、反復配列、例えばホモポリマーおよび単純リピートは計算的にはアラインメントにあいまいさを持ち込むことになり、したがって、ヌクレオチド変異について偽陽性の結果になる。

30

【0110】

目的の領域が大きいほど、より大きい数の偽陽性ヌクレオチド位置が観察されることになる。ゲノム全体において突然変異を探している場合、目的の領域はゲノム全体になり、関係するヌクレオチドの数は 30 億になる。一方、エクソンに焦点を当てている場合、エクソンをコードするヌクレオチドの数、すなわち、約 4500 万が目的の領域になる。

40

【0111】

シークエンシングエラーに関係する偽陽性ヌクレオチド位置の数は次の計算に基づいて決定され得る。シークエンシングエラーのために同一の位置において同一のヌクレオチド変化を有する確率 ( $P_{Er}$ ) は、

【数2】

$$P_{Er} = C(D, r) E \left( \frac{E}{3} \right)^{r-1}$$

として計算され得、その式では、

【数3】

$$C(D, r)$$

10

は総計で D 要素から r 要素を選択するための可能な組合せの数であり、r は潜在的突然変異を定義するための出現回数であり、D はシークエンス深度であり、および E はシークエンシングエラー率である。

【数4】

$$C(D, r)$$

20

は、

【数5】

$$C(D, r) = \frac{D!}{r! (D-r)!}$$

として計算され得る。

【0112】

突然変異について偽陽性であるヌクレオチド位置 (NFP) の数は、

30

【数6】

$$N_{FP} = N_I P_{Er}$$

として計算され得、その式では、N<sub>I</sub> は目的の領域におけるヌクレオチド位置の総数である。

【0113】

図 6 B は、4 回、5 回、6 回および 7 回の出現 (r) という判定基準とシークエンス深度を用いてヌクレオチド変化を有していると誤って分類されるヌクレオチド位置の予期される数を示すグラフ 650 である。目的の領域はこの計算ではゲノム全体 (30 億のヌクレオチド位置) であると考えられる。シークエンシングエラー率は配列決定されたヌクレオチドのうちの 0.3 % であると考えられる。理解できるように、r の値は偽陽性に対して大きな影響を有する。しかし、図 6 A からわかるように、少なくとも有意により高いシークエンス深度が使用されるまで r の値が大きいほど検出される突然変異の数も減る。

40

【0114】

C. 最小出現数 (r) の選択

上で考察したように、本当の癌関連突然変異部位とシークエンシングエラーに起因する偽陽性部位の数はシークエンス深度と共に増加することになる。しかしながら、それらの増加率は異なる。それ故、偽陽性部位の数を低い値に保ちつつ本当の癌関連突然変異の検

50

出を最大化するためにシークエンス深度と  $r$  の値の選択を利用することが可能である。

【0115】

図7Aは、異なるシークエンス深度での本当の癌関連突然変異部位と偽陽性部位の数を示すグラフ700である。腫瘍組織における癌関連突然変異の総数は3,000であると考えられ、血漿中における腫瘍由来DNAの分画濃度は10%であると考えられる。シークエンシングエラー率は0.3%であると考えられる。図の説明において、TPは対応する突然変異が腫瘍組織に存在する真の陽性部位を表し、FPは対応する突然変異が腫瘍組織に存在せず、シークエンシングデータ中に存在するヌクレオチド変化がシークエンシングエラーに起因する偽陽性部位を表す。

【0116】

グラフ700より、我々が潜在的突然変異部位を定義するために血漿中における6という最小出現回数( $r = 6$ )を判定基準として用いる場合、110倍のシークエンス深度で約1,410事例の本当の癌関連突然変異が検出されることになる。この判定基準を用いると、わずかに約20偽陽性部位が検出されることになる。我々が潜在的突然変異を定義するために7回という最小出現回数( $r = 7$ )を判定基準として用いる場合、検出され得る癌関連突然変異の数は470～約940だけ減少することになる。それ故、 $r = 6$ という判定基準は血漿における癌関連突然変異の検出感度をより高くすることになる。

【0117】

一方、可能性がある突然変異を定義するために我々が6および7という最小出現回数( $r$ )の判定基準を用いる場合、200倍のシークエンス深度で検出される本当の癌関連突然変異の数はそれぞれ約2,800と2,600になる。これらの2つの $r$ の値を用いると、偽陽性部位の数はそれぞれ約740と20になる。それ故、200倍のシークエンス深度では潜在的突然変異を定義するために $r = 7$ というより厳しい判定基準を用いることにより、本当の癌関連突然変異の検出感度に著しく悪影響を与えることなく偽陽性部位の数を大いに減少させることができる。

【0118】

D. 血漿中において可能性がある突然変異を定義するためのシークエンシングデータのダイナミック・カットオフ

目的の領域内において各ヌクレオチドのシークエンス深度は異なる。我々が血漿中において潜在的突然変異を定義するためのヌクレオチド変化の出現回数に固定カットオフ値を適用する場合、より多くの配列リード(すなわち、より高いシークエンス深度)によりカバーされるヌクレオチドが、より低いシークエンス深度を有するヌクレオチドと比較して、腫瘍組織にそのような変化が無い状態でシークエンシングエラーのためにヌクレオチド変異を有すると誤って分類されるより高い確率を有することになる。この問題を克服するための1つの実施形態は、特定のヌクレオチド位置の実際のシークエンス深度および偽陽性変異を呼び出す確率の所望の上限に応じて $r$ のダイナミック・カットオフ値を異なるヌクレオチド位置に適用することである。

【0119】

1つの実施形態において、最大許容可能偽陽性率は $1.5 \times 10^{-8}$ ヌクレオチド位置中の1位置に固定され得る。この最大許容可能偽陽性率を用いると、ゲノム全体において特定される偽陽性部位の総数は20未満になる。図6Bにおいて示される曲線に従って異なるシークエンス深度に対する $r$ の値を決定することができ、且つ、これらのカットオフは表1に示されている。他の実施形態において、他の異なる最大許容可能偽陽性率、例えば、 $3 \times 10^{-8}$ 分の1、 $10^{-8}$ 分の1または $6 \times 10^{-7}$ 分の1を使用することができる。偽陽性部位の対応する総数はそれぞれ10未満、30未満および50未満になる。

【0120】

10

20

30

40

## 【表1】

表1. 特定のヌクレオチド位置の異なるシークエンス深度に対する潜在的突然変異を定義するための血漿中に存在するヌクレオチド変化の最小出現回数( $r$ )。最大偽陽性率は $1.5 \times 10^{-8}$ ヌクレオチド中の1ヌクレオチドに固定されている。

特定のヌクレオチド位置のシークエンス深度	潜在的突然変異を定義するための血漿中DNAシークエンシングデータ中に存在するヌクレオチド変化の最大出現回数
50未満	5
50~110	6
111~200	7
201~310	8
311~450	9
451~620	10
621~800	11

10

20

30

40

## 【0121】

## E. 標的濃縮シークエンシング

図7Aに示されるように、より大きい $r$ の値を用いることを許容することにより、シークエンス深度が高いほど偽陽性部位の数を低く保ちつつ癌関連突然変異の検出感度がより良くなり得る。例えば、110倍のシークエンス深度では6という $r$ 値を用いて1,410事例の本当の癌関連突然変異が血漿中において検出され得るが、一方、シークエンス深度が200倍まで増加し、7という $r$ 値が適用されると、検出される本当の癌関連突然変異の数は2,600になる。2セットのデータは約20という偽陽性部位の期待数を示すことになる。

## 【0122】

200倍の深度までのゲノム全体のシークエンシングは現時点では比較的に費用がかかるが、そのようなシークエンス深度を達成する一つの可能な方法はより小さい目的の領域に集中することである。例えば、限定されないが、ハイブリダイゼーションにより目的のゲノム領域を捕捉するためにDNAバイトまたはRNAバイトを使用することによって標的領域の分析を達成することができる。次に、捕捉した領域を、例えば、磁気的手段によりプルダウンし、その後にシークエンシングの対象とする。例えば、Agilent Sure Select 標的濃縮システム、Roche Nimblegen 標的濃縮システムおよび Illumina 標的化再シークエンシングシステムを使用してそのような標的捕捉を実施することができる。別のアプローチは、標的領域のPCR增幅を行い、その後にシークエンシングを実施することである。1つの実施形態において、目的の領域エクソンである。そのような実施形態において、血漿中DNAに対して全てのエクソンの標的捕捉を実施することができ、次にエクソン領域について濃縮された血漿中DNAを配列決定することができる。

## 【0123】

より高いシークエンス深度を有することに加え、ゲノム全体を分析する代わりに特定の

50

領域に集中することでサーチ領域内のヌクレオチド位置の数を著しく減少させることになり、且つ、同一のシークエンシングエラー率の前提で偽陽性部位の数を減少させることになる。

#### 【 0 1 2 4 】

図 7 B は、ゲノム全体 ( W G ) と全てのエクソンの分析を伴う偽陽性部位の予想数を示すグラフ 7 5 0 である。各種の分析について、 $r$  の 2 つの異なる値である 5 と 6 を使用する。200 倍のシークエンス深度では、血漿中において突然変異を定義するために  $r = 5$  を用いる場合、偽陽性部位の予想数はゲノム全体と全てのエクソンについてそれぞれ約 23,000 と 230 である。血漿中において突然変異を定義するために  $r = 6$  を用いる場合、偽陽性部位の予想数はそれぞれ 750 と 7 である。それ故、目的の領域におけるヌクレオチドの数の限度により、血漿突然変異分析における偽陽性の数を著しく減少させることができる。10

#### 【 0 1 2 5 】

エクソン捕捉シークエンシングにおいて、またはエクソーム捕捉シークエンシングにおいてさえも、サーチ領域内におけるヌクレオチドの数が減少する。それ故、我々が癌関連突然変異の検出のより高い偽陽性率を許容しても、偽陽性部位の絶対数は比較的に低いレベルに保たれ得る。より高い偽陽性率を許容することで血漿中において一ヌクレオチド変異を定義するための最小出現回数 ( $r$ ) について厳しさが少ない判定基準を用いることを許すことになる。これにより、本当の癌関連突然変異を検出するためのより高い感度が生じることになる。20

#### 【 0 1 2 6 】

1 つの実施形態において、我々は  $1.5 \times 10^6$  という最大許容可能偽陽性率を用いることができる。この偽陽性率により、標的とされたエクソン内の偽陽性部位の総数はわずかに 20 になる。 $1.5 \times 10^6$  という最大許容可能偽陽性率を用いる異なるシークエンス深度に対する  $r$  の値が表 2 に示されている。他の実施形態において、他の異なる最大許容可能偽陽性率、例えば、 $3 \times 10^6$  分の 1、 $10^6$  分の 1 または  $6 \times 10^5$  分の 1 を用いることができる。偽陽性部位の対応する総数はそれぞれ 10 未満、30 未満および 50 未満になる。1 つの実施形態において、上に記載されたように異なる加重は異なる部類の突然変異に帰するものであり得る。

#### 【 0 1 2 7 】

30

## 【表2】

表2. 特定のヌクレオチド位置の異なるシークエンス深度に対する潜在的突然変異を定義するための血漿中に存在するヌクレオチド変化の最小出現回数 ( $r$ )。最大偽陽性率  $1.5 \times 10^{-6}$  ヌクレオチド中の1ヌクレオチドに固定されている。

特定のヌクレオチド位置のシークエンス深度	潜在的突然変異を定義するための血漿中DNAシークエンシングデータ中に存在するヌクレオチド変化の最大出現回数	10
50未満	4	
50~125	5	
126~235	6	
236~380	7	
381~560	8	20
561~760	9	

## 【0128】

## V III. 癌の検出

上で述べたように、異型座位における配列タグの数を様々な方法で用いてパラメーターを決定することができ、そのパラメーターを閾値と比較して癌のレベルを分類する。ある座位または多数の座位における全てのリードに対する異型リードの分画濃度が、使用することができる別のパラメーターである。以下にパラメーターと閾値を計算する幾つかの例がある。

## 【0129】

## A. パラメーターの決定

特定の座位においてCGが第1アレルについてホモ接合性であり、異型アレルが生物試料（例えば、血漿）において見られる場合、分画濃度は  $2p / (p + q)$  として計算され得、その式では、pは異型アレルを有する配列タグの数であり、qはCGの第1アレルを有する配列タグの数である。この式は、腫瘍のハプロタイプのうちのわずかに1つが、典型的な事例であるだろうことに、異型を有するということを仮定する。したがって、各ホモ接合性座位について分画濃度を計算することができる。分画濃度の平均をとることができる。別の実施形態において、分画濃度を決定するために数pはそれらの座位の全てについての配列タグの数を含むことができ、数qについても同様である。次に例を説明する。

## 【0130】

4人のHCC患者の血漿における腫瘍由来-ヌクレオチド異型(SNV)のゲノムワイド検出を調査した。我々は腫瘍DNAとバフィーコートDNAをそれぞれ29.5倍（27倍~33倍の範囲）および43倍（39倍~46倍の範囲）の半数体ゲノムカバー度の平均深度でシークエンシングした。4人のHCC患者のそれぞれに由来する腫瘍DNAとバフィーコートDNAからのMPSデータを比較し、腫瘍DNA中には存在するが、バフィーコートDNA中には存在しないSNVを厳しいバイオインフォマティクスアルゴリズムで掘り当てた。このアルゴリズムは仮想SNVが少なくとも閾値数の、それが本当のSNVとして分類される前に配列決定された腫瘍DNA断片の中（すなわち、対応する配列決定されたタグの中）に存在することを必要とした。例えば、本明細書に記載されるよう

30

40

50

な特定のスクレオチドのシークエンス深度とシークエンシングエラー率を考慮することにより閾値数が決定された。

【0131】

図8は、本発明の実施形態に従って血漿中における腫瘍由来DNAの分画濃度を含む、治療前後の4人のHCC患者の結果を示す表800である。腫瘍関連SNVの数は4人のHCCの事例で1,334から3,171までの範囲であった。血漿中において検出可能であったそのようなSNVの割合は治療前後に記載される。治療前に腫瘍関連SNVのうちの15%~94%が血漿中において検出された。治療後に、そのパーセンテージは1.5%~5.5%の間であった。したがって、検出されるSNVの数は癌のレベルとまさに相関する。このことは、癌のレベルを分類するためのパラメーターとしてSNVの数を用いることができる事を示している。

10

【0132】

血漿中における腫瘍由来DNAの分画濃度が全(すなわち、突然変異型および野生型)配列に対する突然変異型の分画数により決定された。その式は $2p / (p + q)$ であり、その式では、2は腫瘍で突然変異形成されるただ1つのハプロタイプに対応する。これらの分画濃度はゲノムワイド凝集性アレル喪失(GAAL)分析によって決定された分画濃度とよく相関し(Chan KC et al. Clin Chem 2013;59:211-24)、且つ、手術の後に低下した。したがって、その分画濃度も癌のレベルの決定に有用なパラメーターであることが示される。

20

【0133】

SNV分析に由来する分画濃度は腫瘍負荷量を伝えることができる。より高い腫瘍負荷量(例えば、より高い推定分画濃度)を有する癌患者がより低い腫瘍負荷量を有する者よりも高い頻度の体細胞突然変異を有する。したがって、実施形態を予後予測に用いることもできる。一般に、より高い腫瘍負荷量を有する癌患者はより低い腫瘍負荷量を有する者よりも悪い予後を有する。前者の群は、したがって、その疾患のために死亡するより高い見込みを有することになる。いくつかの実施形態において、(例えば、リアルタイムPCRまたは蛍光分析法を用いて)生物試料、例えば、血漿におけるDNAの絶対濃度を決定することができる場合、腫瘍関連遺伝的変化の絶対濃度を決定することができ、臨床的検出および/またはモニタリングおよび/または予後予測に用いることができる。

30

【0134】

B. 閾値の決定

閾値を決定するために表800を使用することができる。上で述べたように、SNV分析により決定されたSNVの数と分画濃度は癌のレベルに相関する。個別にその閾値を決定することができる。例えば、閾値を決定するために治療前の値を用いることができる。様々な実施形態において、閾値は絶対値の治療前からの相対的变化であり得る。適切な閾値はSNVの数または分画濃度が50%減少した値であり得る。そのような閾値は、表800中の事例のそれについてより低い癌のレベルの分類を提供することになる。そのような閾値はシークエンス深度に左右され得ることに留意されたい。

【0135】

1つの実施形態において、複数の試料にある閾値を使用することだってでき、そのパラメーターの治療前の値を考慮してもしなくてもよい。例えば、癌を有しない、または低い癌のレベルを有すると対象を分類するために100個のSNVという閾値を使用することだってできる。100個のSNVというこの閾値は表800の中の4つの事例のそれによって満たされる。パラメーターとして分画濃度が使用された場合は、1.0%という閾値がHCC1~HCC3を実際に0レベルの癌と分類することになり、1.5%という第2閾値がHCC4を低い癌のレベルと分類することになる。したがって、1より大きい閾値を使用して2つより多くの分類を得ることができる。

40

【0136】

他の可能な閾値を例示するため、我々は腫瘍関連SNVについて健常対照の血漿を分析した。複数回の測定を健常対象に行って構成的ゲノムと比べて生物試料からどれくらい多

50

くの範囲の変異が予想されるか決定することができる。

【0137】

図9は、本発明の実施形態に従う16人健常な対照被検者におけるHCC関連SNVの検出を示す表900である。SNV分析アプローチの特異度を推定するために表900を用いることができる。16人の健常対照は異なる行として記載されている。列は特定のHCC患者について検出されたSNVを調べ、異型アレルを有する異型座位における配列リードの数と野生型アレル（すなわち、CGに由来するアレル）を有する配列リードの数を示す。例えば、HCC1について、対照C01はそのような異型座位において40異型リードを有したが、31, 261リードの野生型アレルを有した。最後の列はHCC1患者のSNVの全てにまたがる総分画濃度を示す。HCC関連SNVはHCC患者に特異的であったので、HCC関連SNVの存在は偽陽性を表す。あるカットオフ値が本明細書に記載されるようにこれらの見かけの配列異型に適応される場合、これらの偽陽性の全てがフィルタリングで除外されることになる。

【0138】

16人の健常対照の血漿における少数のこれらの仮想腫瘍関連突然変異の存在がこの方法の「確率的ノイズ」を表し、シークエンシングエラーに起因するようであった。そのようなノイズから推定される平均分画濃度は0.38%であった。これらの値は健常対象についての範囲を示す。したがって、最も高い分画濃度が0.43%であったので、HCCについての0レベルの癌の分類の閾値は約0.5%であり得た。したがって、全ての癌細胞がHCC患者から除去される場合、これらの低い分画濃度が予期されることになる。

【0139】

表800を参照し直して、0.5%を0レベルの癌の閾値として使用した場合、HCC1とHCC3の治療後血漿データがSNV分析に基づいて0レベルを有すると決定されることになる。HCC2は0よりも高いあるレベルとして分類される可能性がある。HCC4も0よりも高いあるレベルまたはいくらかより高いレベルとして分類される可能性があるが、治療前試料と比較して未だに比較的に低いレベルに分類される可能性がある。

【0140】

パラメーターが異型座位の数に対応する1つの実施形態において、閾値は0であり得る（すなわち、1つの異型座位が非0レベルの癌を示すことがあり得る）。しかしながら、（例えば、深度の）多数の設定を用いると、閾値はより高くなり、例えば、5または10の絶対値になる。治療後に人がモニターされる1つの実施形態において、閾値は使用中に現れる（腫瘍を直接的に分析することにより特定された）ある特定のパーセンテージのSNVであり得る。ある座位において必要とされる異型リードの数のカットオフ値が充分に大きい場合、1つの異型座位を有することだけで非0レベルの癌を示す可能性がある。

【0141】

したがって、生物試料（例えば、血漿）に由来するDNA中の変異（例えば、一ヌクレオチド変異）の定量的分析を癌の診断、モニタリングおよび予後予測に用いることができる。癌の検出のために試験対象の血漿中において検出される一ヌクレオチド変異の数を一群の健常対象のそれと比較することができる。健常対象では、血漿中における見かけの一ヌクレオチド変異はシークエンシングエラー、血液細胞に由来する非クローニ性突然変異、および他の器官に由来する非クローニ性突然変異に起因し得る。表900に示されるように、正常な健常対象の細胞が少数の突然変異を担持することが示されている（Conrad D F et al. Nat Genet 2011;43:712-4）。したがって、試験した患者が非0レベルの癌に対応する異常に多数の一ヌクレオチド変異を血漿中に有するか決定するための基準範囲として一群の見たところ健康な対象の血漿における見かけの一ヌクレオチド変異の総数を用いることができる。

【0142】

基準範囲を構築するために使用される健常対象は年齢と性別に関して試験対象と一致し得る。これまでの研究において、体細胞における突然変異の数は年齢と共に増加することが示されている（Cheung NK et al, JAMA 2012;307:1062-71）。したがって、我々は歳をと

るにつれ、大半の時間で細胞が比較的に良性であっても、それらの細胞のクローンを蓄積することが人にとって「正常」になり、または臨床的に重要になるのに非常に長い時間がかかることになる。1つの実施形態において、異なる対象群、例えば、異なる年齢、性別、民族、および他のパラメーター（例えば、喫煙の状況、肝炎の状況、アルコール歴、薬歴）について基準レベルを作成することができる。

#### 【0143】

基準範囲は使用するカットオフ値（すなわち、ある座位において必要とされる異型配列タグの数）、ならびに考えられる偽陽性率と他の変数（例えば、年齢）に基づいて変わり得る。したがって、特定のセットの1種類以上の判定基準について基準範囲を決定することができ、同一の判定基準を用いてある試料のパラメーターが決定されることになる。その後、同一の判定基準を用いて両方が決定されたのでそのパラメーターを基準範囲と比較することができる。

10

#### 【0144】

上で述べたように、実施形態は、癌のレベルを決定するために複数の閾値を用いることができる。例えば、第1レベルは、閾値よりも低いパラメーターについて癌の兆候を判定することはできず、且つ、前癌レベルであり得る少なくとも第1癌のレベルを判定することができた。他のレベルは異なるステージの癌に対応することができた。

#### 【0145】

##### C. 実験変数に対する依存度

シークエンシングの深度は少数派（例えば、腫瘍）ゲノムの最小検出閾値の構築にとって重要であり得る。例えば、10半数体ゲノムのシークエンス深度を使用する場合、シークエンシング技術を用いてもエラー無しで検出できるだろう最小腫瘍DNA濃度は1/5、すなわち、20%である。一方、100半数体ゲノムのシークエンス深度を使用する場合、2%にまで落ちることがあり得る。この分析は、たった1つの突然変異座位が分析されているシナリオを指している。しかしながら、より多くの突然変異座位が分析されているときに最小腫瘍DNA濃度はより低いことがあり得、且つ、二項確率関数によって管理される。例えば、シークエンス深度が10倍であり、且つ、腫瘍DNAの分画濃度が20%である場合、突然変異を検出する機会は10%である。しかしながら、我々が10事例の突然変異を有する場合、少なくとも1事例の突然変異を検出する機会は $1 - (1 - 10\%)^{10} = 65\%$ であるだろう。

20

#### 【0146】

幾つかの効果がシークエンス深度の増大について存在する。シークエンス深度が高いほど、より多くのシークエンシングエラーが見られることになる。図4および5を参照のこと。しかしながら、シークエンシングエラーはゲノム中で無作為に生じるが、突然変異は所与の集団の細胞について同一の位置で生じるので、より高いシークエンス深度ほど、シークエンシングエラーを細胞（例えば、癌細胞）の亜集団のクローン増殖に起因する突然変異と区別することができる。

#### 【0147】

シークエンス深度が高いほど、「健常細胞」に由来するより多くの突然変異を特定することになる。しかしながら、これらの健常細胞のクローン増殖が無く、且つ、それらの突然変異プロファイルが異なるとき、これらの健常細胞における突然変異は血漿中におけるそれらの出現頻度によって前記の突然変異と区別され得る（例えば、突然変異を示す所望の数のリードに対するカットオフNを用いることにより、例えば2、3、4、5、またはそれより多くのNを有することにより）。

40

#### 【0148】

上で述べたように、閾値は、クローン増殖し、したがって、他の機構ではフィルタリングで除外され得ないだろう健常細胞における突然変異の量に依存し得る。予期されるこの分散は健常対象を分析することにより獲得され得る。クローン増殖が時間と共に起こるので、患者の年齢が健常対象において見られる分散に影響することができ、したがって、閾値は年齢に依存し得る。

50

## 【0149】

## D. 標的化アプローチとの組合せ

いくつかの実施形態において、標的化アプローチと組合せてランダムシークエンシングを用いることができる。例えば、癌患者を提示して血漿試料のランダムシークエンシングを実施することができる。コピー数の変化およびSNVについて血漿中DNAのシークエンシングデータを分析することができる。変化（例えば、増幅／欠失または高密度のSNV）を示す領域は連続的モニタリング目的のために標的とされ得る。一つの方法として効果的にそのモニタリングを一定の期間にわたって行うことができ、またはランダムシークエンシングの直後に行うことができる。標的化分析については、非侵襲的出生前診断用に血漿中DNAを濃縮するために溶液相ハイブリダイゼーションベースの捕捉アプローチをうまく利用した (Liao GJ et al. Clin Chem 2011;57:92-101)。そのような技術が上述されている。したがって、癌の検出とモニタリングのために標的化アプローチとランダムアプローチを併用することができる。

10

## 【0150】

したがって、突然変異形成した可能性があることが上述した非標的化ゲノムワイドアプローチを用いてわかった座位の標的化シークエンシングを実施することができる。そのような標的化シークエンシングは、（例えば、Agilent Sure Select、NimbleGen Sequence Capture、または Illumina 標的化再シークエンシングシステムを用いる）溶液相ハイブリダイゼーション技術または固相ハイブリダイゼーション技術とそれに続く大量平行シークエンシングを用いて実施され得る。別のアプローチは標的化シークエンシングのために（例えば、PCRベースの）増幅システムを実行することである (Forshew T et al. Sci Transl Med 2012; 4: 135ra68)。

20

## 【0151】

## IX. 分画濃度

ある座位が突然変異として特定される前にその座位における所望の数の変異に対するカットオフ値を決定するために腫瘍DNAの分画濃度を用いることができる。例えば、分画濃度が比較的に高いと知られていた場合、本当のSNVについて比較的に大きい数の異型リードが存在するはずであることが知られているので、高カットオフを用いてより多くの偽陽性をフィルタリングで除外することができるだろう。一方、分画濃度が低かった場合、より低いカットオフが必要とされることがあり得、それで幾つかのSNVを見落とすことはない。この場合、分画濃度がパラメーターとして用いられるSNV分析と異なる方法によって分画濃度が決定されることになる。

30

## 【0152】

分画濃度の決定に様々な技術を用いることができ、それらのうちの幾つかは本明細書に記載されている。混合物、例えば、腫瘍細胞と非腫瘍細胞の混合物を含有する生検試料、または腫瘍細胞から放出されたDNAと非腫瘍細胞から放出されたDNAを含有する癌患者由来の血漿試料における腫瘍由来DNAの分画濃度を決定するためにこれらの技術を用いることができる。

40

## 【0153】

## A. G A A L

ゲノムワイド凝集性アレル喪失 (G A A L) はヘテロ接合性を失った座位を分析する (Chan KC et al. Clin Chem 2013;59:211-24)。ヘテロ接合性である構成的ゲノムCGの部位に関して、腫瘍はアレルのうちの一方の欠質を有する座位を多くの場合に有する。したがって、そのような座位の配列リードは別のアレルよりも一方のアレルからより多くのものを示し、その場合にその差異は試料中の腫瘍DNAの分画濃度に比例する。そのような計算の例が以下に示される。

## 【0154】

HCC患者のバフィーコートと腫瘍組織から抽出されたDNAについてAffymetrixゲノムワイドヒトSNPアレイ6.0システムを用いて遺伝子型を決定した。Aff

50

fy metrix ジェノタイピングコンソール第4.1版を用いてマイクロアレイデータを処理した。Birddseed第2版アルゴリズムを用いて遺伝子型分析と一ヌクレオチド多型(SNP)コーリングを実施した。ヘテロ接合性喪失(LOH)領域の特定とコピー数分析の実施のためにバフィーコートと腫瘍組織の遺伝子型解析データを用いた。前記のジェノタイピングコンソールを Affymetrixからの初期パラメーターおよび100bpという最小ゲノム断片サイズおよびその断片内の最小で5遺伝子マーカーと共に用いてコピー数分析を実施した。

## 【0155】

LOHを有する領域が、腫瘍組織では1コピーを有し、且つ、バフィーコートでは2コピーを有する領域として特定され、これらの領域内のSNPはバフィーコートではヘテロ接合性であり、腫瘍組織ではホモ接合性であった。腫瘍組織においてLOHを示すゲノム領域について、バフィーコートでは存在するが腫瘍組織では存在しないか、強度が低下しているSNPアレルは染色体領域の欠質した断片上のアレルであると考えられた。バフィーコートと腫瘍組織の両方で存在するアレルは染色体領域の欠質していない断片に由来するとみなされた。腫瘍において1コピー喪失を有する全ての染色体領域について欠質アレルと非欠質アレルを担持する配列リードの総数を計数した。これらの2つの値の差異を用い、次の等式を用いて試料中の腫瘍由来DNAの分画濃度( $F_{GAAL}$ )を推測した：

## 【数7】

$$F_{GAAL} = \frac{N_{non-del} - N_{del}}{N_{non-del}}$$

20

式中、 $N_{non-del}$ は非欠質アレルを担持する配列リードの総数を表し、 $N_{del}$ は欠質アレルを担持する配列リードの総数を表す。

## 【0156】

## B. ゲノム表現を用いる推定

GAAL技術の問題は、特定の座位(すなわち、LOHを示す座位)が特定され、そのような座位に整列する配列リードが使用されることである。そのような要求が追加のステップと、したがって、追加コストを加える場合があり得る。コピー数、例えば、配列リード密度だけを使用する実施形態を今から説明する。

30

## 【0157】

染色体変化、例えば、増幅および欠失は癌ゲノムにおいて頻繁に観察される。癌組織において観察されるこれらの染色体変化は通常は染色体小領域に関係し、これらの変化は1Mbよりも短いことがあり得る。そして、癌関連染色体変化異なる患者において異質であり、したがって、異なる患者で異なる領域が影響を受けることがあり得る。数十、数百、または数千ものコピー数の変化が癌ゲノムに見つかることも珍しくはない。これらの要因の全てが腫瘍DNA濃度の決定を困難なものにしている。

## 【0158】

40

実施形態は、腫瘍関連染色体変化の結果生じる定量的变化の分析を伴う。1つの実施形態において、例えば、Illumina HiSeq 2000シークエンシングプラットフォームによる大量平行シークエンシングを用いて癌細胞と正常細胞に由来するDNAを含有するDNA試料のシークエンシングを行う。得られたDNAは血漿または他の適切な生物試料中の無細胞性DNAであり得る。

## 【0159】

腫瘍組織において増幅される染色体領域は増加した配列決定される確率を有し、腫瘍組織において欠質される領域は減少した配列決定される確率を有することになる。結果として、増幅領域に整列する配列リードの密度が増加することになり、欠質領域へのその整列が減少することになる。変異の程度はDNA混合物における腫瘍由来DNAの分画濃度に

50

比例する。腫瘍組織に由来するDNAの割合が高いほど、染色体変化によってより大きい変化が引き起こされることになる。

#### 【0160】

##### 1. 高い腫瘍濃度を有する試料における評価

4人の肝細胞癌患者の腫瘍組織からDNAを抽出した。Covaria DNA超音波処理システムを用いてそのDNAを断片化し、記載される(Chan KC et al. Clin Chem 2013;59:211-24)ように Illumina HiSeq 2000プラットフォームを用いてそのDNAの配列を決定した。それらの配列リードをヒト基準ゲノム(hg18)に対して整列させた。その後、ゲノムを1Mbの bin(領域)に分割し、記載される(Chen EZ et al. PLoS One. 2011;6:e21791)ようにGCバイアスを調整した後に各binについて配列リード密度を計算した。

10

#### 【0161】

配列リードを基準ゲノムに対して整列させた後に様々な領域について配列リード密度を計算することができる。1つの実施形態において、配列リード密度は、基準ゲノムに(例えば、基準ゲノム中の唯一の位置に)に対して整列させることができる総配列リード数で除算された特定のbin(例えば、1Mb領域)にマップされるリードの数として決定される割合である。腫瘍組織において増幅される染色体領域と重なるbinはそのような重複を有しないbinのものよりも高い配列リード密度を有することが予期される。一方、欠質した染色体領域と重なるbinはそのような重複を有しないbinのものよりも低い配列リード密度を有することが予期される。染色体変化を有する領域と有しない領域の間の配列リード密度の差の大きさは試料中の腫瘍由来DNAの割合と腫瘍細胞における増幅/欠失の程度によって主に影響を受ける。

20

#### 【0162】

異なる種類の染色体変化に対応する配列リード密度を有するbinを特定するために様々な統計的モデルを用いることができる。1つの実施形態において、正規混合モデル(Finite mixture models、2004年中のMcLachlan G and Peel D.のMultivariate normal mixtures. 81~116頁。John Wiley & Sons Press)を用いることができる。他の統計的モデル、例えば、二項混合モデルおよびポアソン回帰モデル(Finite mixture models、2004年中のMcLachlan G and Peel D. Mixtures with non-normal components. 135~174頁。John Wiley & Sons Press)も用いることができる。

30

#### 【0163】

バフィーコートDNAのシークエンシングから決定されたのと同一のbinの配列リード密度を用いてあるbinについての配列リード密度を正規化することができる。異なるbinの配列リード密度は特定の染色体領域の配列関係によって影響を受けることがあり得、したがって、正規化は変化を示す領域をより正確に特定することに役立ち得る。例えば、異なる染色体領域のマップ可能性(ある配列をその元の位置に戻って配列させる確率を指す)は異なることがあり得る。さらに、コピー数の多型(すなわち、コピー数変異)もこれらのbinの配列リード密度に影響することになる。それ故、バフィーコートDNAを用いる正規化は、異なる染色体領域間の配列関係の差異に関連する変異を最小にする可能性があり得る。

40

#### 【0164】

図10Aは、本発明の実施形態に従うHCC患者の腫瘍試料の配列リード密度の分布プロット1000を示す。HCC患者からの外科的切除の後に腫瘍組織を得た。x軸は患者の腫瘍組織とバフィーコートの間の配列リード密度の比率(R)の $\log_2$ を表す。y軸はbinの数を表す。

#### 【0165】

欠失を有する領域、増幅を有する領域、および染色体変化を有しない領域を表すために正規混合モデルを用いてピークを分布曲線にフィットさせることができる。1つの実施形態において、異なる推算値にまたがって赤池情報量基準(AIC)によりピークの数を決定することができる。 $\log_2 R = 0$ (すなわち、R = 1)での中央ピークはどのような

50

染色体変化も有しない領域を表す。左側のピーク（中央ピークと比べて）は1コピー喪失を有する領域を表す。右側のピーク（中央ピークと比べて）は1コピー増幅を有する領域を表す。

【0166】

腫瘍由来DNAの分画濃度は増幅領域と欠質領域を表すピーク間の距離によって反映され得る。その距離が大きいほど試料中の腫瘍由来DNAの分画濃度が高くなる。試料中の腫瘍由来DNAの分画濃度は次の等式を用いて $F_{GR}$ として表されるこのゲノム表現アプローチによって決定され得る：

【数8】

10

$$F_{GR} = R_{right} - R_{left}$$

式中、 $R_{right}$ は右側ピークのR値であり、 $R_{left}$ は左側ピークのR値である。最大の差異は、100%に対応する1である。HCC患者から得られた腫瘍試料中の腫瘍由来DNAの分画濃度は66%であると推定され、その場合に $R_{right}$ と $R_{left}$ の値はそれぞれ1.376と0.712である。

【0167】

この結果を検証するためにゲノムワイド凝集性アレル喪失（GAAL）分析を用いる別 の方法も用いて腫瘍DNAの割合の分画濃度を独自に決定した（Chan KC et al. Clin Chem 2013;59:211-24）。表3は4人のHCC患者の腫瘍組織における腫瘍由来DNAのゲノム表現（ $F_{GR}$ ）とGAAL（ $F_{GAAL}$ ）アプローチを用いた分画濃度を示す。これらの2つの異なるアプローチによって決定された値は互いによく合致する。

20

【0168】

【表3】

表3. GAALとゲノム表現（GR）によって決定された分画濃度を示す。

HCC腫瘍	$F_{GAAL}$	$F_{GR}$
1	60.0%	66.5%
2	60.0%	61.4%
3	58.0%	58.9%
4	45.7%	42.2%

30

【0169】

2. 低腫瘍濃度を有する試料における評価

試料DNAのうちの50%より多くが腫瘍由来であるとき、すなわち、腫瘍DNAが大部分であるとき、腫瘍DNAの分画濃度を測定するために我々のゲノム表現方法を用いることができる事が上記の分析により示された。以前の分析において我々は、腫瘍由来DNAが小さな割合（すなわち、50%未満）を表す試料にこの方法を適用することもできる事を示した。小さな割合の腫瘍DNAを含有し得る試料には癌患者の血液、血漿、血清、尿、胸膜液、脳脊髄液、涙、唾液、腹水および糞便が含まれるが、これらに限定されない。幾つかの試料では腫瘍由来DNAの分画濃度は49%、40%、30%、20%、10%、5%、2%、1%、0.5%、0.1%またはそれ未満であり得る。

40

【0170】

そのような試料にとって、増幅および欠失を有する領域を表す配列リード密度のピークは、上で例示したように比較的に高濃度の腫瘍由来DNAを含有する試料におけるほど明

50

確ではないことがあり得る。1つの実施形態において、癌細胞中の染色体変化を有する領域は、癌DNAを含有しないことが知られている基準試料との比較をおこなうことにより特定され得る。例えば、癌を有しない対象の血漿を基準として用いてそれらの染色体領域の配列リード密度の規範的範囲を決定することができる。試験対象の配列リード密度を基準群の値と比較することができる。1つの実施形態において、配列リード密度の平均値と標準偏差(SD)を決定することができる。各ビンについて、試験対象の配列リード密度を基準群の平均値と比較して次の式を用いてzスコア(z-score)を決定する：

【数9】

$$z-score = \frac{(GR_{test} - \overline{GR}_{ref})}{SD_{ref}},$$

10

式中、GR<sub>test</sub>は癌患者の配列リード密度を表し、

【数10】

$$\overline{GR}_{ref}$$

20

は基準対象の平均配列リード密度を表し、且つ、

【数11】

$$SD_{ref}$$

は基準対象の配列リード密度のSDを表す。

【0171】

3以下のzスコアを有する領域は癌患者における特定のビンの配列リード密度の著しい出現不足を表し、腫瘍組織における欠失の存在を示唆する。3より大きいzスコアを有する領域は癌患者における特定のビンの配列リード密度の過剰出現を意味し、腫瘍組織における増幅の存在を示唆する。

30

【0172】

次に、全てのビンのzスコアの分布を構築して異なる数のコピー獲得とコピー喪失、例えば、1または2コピーの染色体の欠失、および1コピー、2コピー、3コピーおよび4コピー追加の染色体という結果になる増幅に関係する領域を特定することができる。幾つかの事例では、1コピーより多くの染色体または1コピーより多くの染色体の領域が関わり得る。

【0173】

図10Bは、本発明の実施形態に従うHCC患者の血漿における全てのビンのzスコアの分布プロット1050を示す。(左から右へ)1コピー喪失、コピー変化無し、1コピー獲得、および2コピー獲得を表すピークをzスコア分布にフィットさせる。次に、例えば、上に記載されたように正規混合モデルを用いて異なる種類の染色体変化を有する領域が特定され得る。

40

【0174】

次に、1コピー獲得または1コピー喪失を示すビンの配列リード密度から試料中の癌DNAの分画濃度(F)を推測することができる。特定のビンについて決定される分画濃度は、

【数12】

$$F = \frac{|GR_{test} - \overline{GR}_{ref}| \times 2}{GR_{ref}} \times 100\%$$

として計算され得る。これを、

【数13】

10

$$F = \frac{|z - score \times SD_{ref}|}{\overline{GR}_{ref}} \times 2$$

と表現することもでき、それを、

【数14】

20

$$F = |z - score| \times CV \times 2$$

と書き直すこともでき、式中、 $CV$  は基準対象の配列リード密度の測定のための変異の係数であり、

【数15】

$$CV = \frac{SD_{ref}}{\overline{GR}_{ref}}$$

30

である。

【0175】

1つの実施形態において、それらの bin に由来する結果を組み合わせる。例えば、1 コピー獲得を示す bin の  $z$  スコアの平均を計算することができ、または結果生じる  $F$  値の平均を計算することができる。別の実施形態において、 $F$  を推測するために用いられる  $z$  スコアの値が統計的モデルにより決定され、且つ、図 10B および図 11 において示されるピークにより表される。1 コピー獲得を示す領域の分画濃度を決定するために、例えば、右側のピークの  $z$  スコアを用いることができる。

40

【0176】

別の実施形態において、3 以下の  $z$  スコアと 3 より大きい  $z$  スコアを有する全ての bin はそれぞれ 1 コピー喪失および 1 コピー獲得を有する領域に属することがあり得る。なぜなら、これらの 2 種類の染色体変化は最も一般的だからである。染色体変化を有する bin の数が比較的に小さく、且つ、正規分布のフィッティングが正確ではないことがあり得るとき、この近似は非常に有用である。

【0177】

図 11 は、本発明の実施形態に従う HCC 患者の血漿の  $z$  スコアの分布プロット 1100 を示す。染色体変化と重なる bin の数は比較的に小さいが、3 以下の  $z$  スコアと 3 より大きい  $z$  スコアを有する全ての bin をそれぞれ 1 コピー喪失と 1 コピー獲得の正規分布に

50

フィットさせた。

【0178】

G A A L 分析とこの G R ベースのアプローチを用いて 4 人の H C C 患者の血漿における腫瘍由来 D N A の分画濃度を決定した。結果が表 4 に示されている。理解できるように、推定分画表示は G A A L 分析と G R 分析の間でよく相関する。

【0179】

【表 4】

表 4. 染色体変化の分析により推定される血漿中の腫瘍由来 D N A の分画濃度

10

試料	血漿中の腫瘍由来 D N A の分画濃度	
	G A A L 分析	G R 分析
事例 11	4. 3 %	4. 5 %
事例 13	5 %	5. 5 %
事例 23	5. 2 %	6. 2 %
事例 27	7. 6 %	6. 1 %

20

【0180】

C. 分画濃度の決定方法

図 12 は、無細胞性 D N A を含む生物試料中の腫瘍 D N A の分画濃度を本発明の実施形態に従って決定する方法 1200 のフローチャートである。上記の実施形態を含む様々な実施形態を通して方法 1200 を実施することができる。

【0181】

ブロック 1210において、1種類以上の配列タグが生物試料中の複数の D N A 断片のそれぞれについて受領される。他の方法に対して本明細書に記載されるようにブロック 1210 を実施することができる。例えば、血漿試料から D N A 断片の一端の配列を決定することができる。別の実施形態において、D N A 断片の両端の配列を決定することができ、それによってその断片の長さを推定することができる。

30

【0182】

ブロック 1220において、配列タグについてゲノム上の位置を決定する。それらのゲノム上の位置は、例えば、本明細書に記載されるように、それらの配列タグを基準ゲノムに対して整列させることにより決定され得る。断片の両端の配列が決定される場合、対になったタグを、特定の距離、例えば、500 塩基または 1,000 塩基より短くさせられている 2 つのタグの間の距離を有する対として整列させることができる。

【0183】

ブロック 1230において、複数のゲノム領域のそれぞれについて、ゲノム領域内にゲノム上の位置を有する配列タグからゲノム領域内の D N A 断片のそれぞれの量を決定する。それらのゲノム領域は基準ゲノムにおける等しい長さの非重複性ビンであり得る。1つの実施形態において、あるビンに整列する多数のタグを計数することができる。したがって、各ビンは整列したタグの対応する数を有することがあり得る。ビンがある特定の数の整列したタグを有する頻度を示すヒストグラムを計算することができる。方法 1200 は、それぞれ同じ長さを有するゲノム領域（例えば、1 M b のビン）であって、非重複性の領域に対して実施され得る。他の実施形態において、考慮され得る、異なる長さを用いることができ、それらの領域は重なってもよい。

40

【0184】

ブロック 1240において、それぞれの密度を得るために前記のそれぞれの量が正規化

50

される。1つの実施形態において、それぞれの密度を得るために前記のそれぞれの量を正規化することは、前記のそれぞれの密度と基準密度を決定するために同一の整列した基準タグの総数を用いることを含む。別の実施形態において、前記のそれぞれの量を整列した基準タグの総数で除算することができる。

#### 【0185】

ブロック1250において、ゲノム領域が1コピー喪失を示すのか、または1コピー獲得を示すのか特定するために前記のそれぞれの密度を基準密度と比較する。1つの実施形態において、前記のそれぞれの密度と基準密度の間で（例えば、zスコアの決定の一部として）差を計算し、その差をカットオフ値と比較する。様々な実施形態において、健常細胞の（例えば、バフィーコートに由来する）試料から、または前記のそれぞれの量それら自体（例えば、大半の領域が喪失または獲得を示さないという仮定で中央値または平均値をとることにより）から基準密度を得ることができる。

10

#### 【0186】

ブロック1260において、1コピー喪失を示すと特定された1つ以上のそれぞれの密度から、または1コピー獲得を示すと特定された1つ以上のそれぞれの密度から第1密度を計算する。第1密度はちょうど1つのゲノム領域に対応し得る、または複数のゲノム領域の密度から決定され得る。例えば、第1密度は1コピー喪失を有するそれぞれの密度から計算され得る。それらのそれぞれの密度は、腫瘍濃度を与えられて、腫瘍における領域の欠質の結果生じる密度の差異の量の測定値を提供する。同様に、第1密度が1コピー獲得を有するそれぞれの密度に由来するものである場合、腫瘍における領域の二倍化の結果生じる密度の差異の量の測定値を得ることができる。上記の節は、第1密度に使用される平均密度を決定するために複数の領域の密度をどのように用いることができるかということとの様々な例を説明している。

20

#### 【0187】

ブロック1270において、差異を得るために第1密度を別の密度と比較することにより分画濃度を計算する。その差異を基準密度で正規化し、それはブロック1270において実行され得る。例えば、差異を基準密度で除算することにより差異を基準密度で正規化することができる。別の実施形態において、それより前のブロックにおいて差異を正規化することができる。

#### 【0188】

30

1つの実施形態において、前記の別の密度は、例えば、2つ上の節で述べたような基準密度である。したがって、分画濃度の計算は前記の差異を2で乗算することを含むことがあり得る。別の実施形態において、別の密度は、例えば、1つ上の節で述べたような、1コピー喪失を示すと特定されたそれぞれの密度から計算される第2密度である（その場合に、第1密度は、1コピー獲得を示すと特定されたそれぞれの密度を用いて計算される）。この場合、第1密度と基準密度の第1比率（例えば、 $R_{right}$ ）を計算すること、および第2密度と基準密度の第2比率（ $R_{left}$ ）を計算することによって正規化された差異を決定することができ、その場合に、差異は第1比率と第2比率の間にある。上に記載したように、前記のそれぞれの密度のヒストグラムの分布曲線にピークをフィットさせることによって1コピー喪失または1コピー獲得を示すゲノム領域の特定を実行することができる。

40

#### 【0189】

まとめると、実施形態は、腫瘍組織において染色体領域が増幅または欠質されているか、および、領域が増幅または欠質されているか同時に決定するために異なる染色体領域における血漿中DNAのゲノム表現を分析して腫瘍由来DNAの分画濃度を推定するためにそのゲノム表現を用いることができる。幾つかの実施形態は、異なる種類の変化、すなわち、1コピー、2コピー、3コピーまたは4コピーの獲得、および1コピーまたは2コピーの喪失に関連するゲノム表現を決定するために、異なる領域のゲノム表現の包括的分布を分析する正規混合モデルを用いる。

#### 【0190】

50

実施形態は他の方法、例えば、ゲノムワイド凝集性アレル喪失 (G A A L) アプローチ (米国特許出願公開第13/308,473号明細書; Chan KC et al. Clin Chem 2013;59:211-24) および腫瘍関連一ヌクレオチド突然変異の分析 (Forshew T et al. Sci Transl Med. 2012;4:136ra68) に対して幾つかの利点を有する。染色体変化を有する領域に位置する全ての配列リードを、その領域の配列リード密度を決定するために用いることができ、したがって、それらの配列リードは腫瘍DNAの分画濃度に関して情報を与える。一方、G A A L分析では、個体においてヘテロ接合性であり、且つ、染色体獲得または染色体喪失を有する染色体領域内に位置する一ヌクレオチドをカバーする配列リードだけが情報を与えることになる。同様に、癌関連突然変異の分析にとって、突然変異をカバーする配列リードだけが腫瘍DNA濃度の推定に有用になる。したがって、他のアプローチと比較したときに腫瘍由来DNAの分画濃度の評価において同一の程度の正確性を達成するために比較的に少ないシークエンシングリードが必要とされ得るので、実施形態がシークエンシングデータのより費用効果がある使用法を許容することがあり得る。10

#### 【0191】

##### X. 代替方法

特定の突然変異が、座位が真の突然変異であると確認されるための基準 (それによって、陽性予測値を調製する) として配列タグ上で見られる時の数の使用は別として、癌性突然変異が確認される高予測値を提供するカットオフ値の使用的代わりに、またはそれに加えて、他の技術を採用できる。例えば、シークエンスデータをプロセシングするとき、例えば、配列決定したヌクレオチドの品質スコアを考慮に入れることにより、異なるストリージェンシーのバイオインフォマティクスのフィルターを使用し得る。ひとつの実施形態では、DNAシークエンサーおよび種々のシークエンスエラーブロファイルを用いたシークエンシング化学を使用できる。低シークエンスエラー比率を用いたシークエンサーおよび化学は、高陽性予測値を得ることになる。シークエンシング精度を上げるために、同じDNAフラグメントのシークエンシングの反復も使用され得る。ひとつの可能な解析戦略は、パシフィックバイオサイエンスのサーチュラーコンセンサスシークエンシング戦略である。20

#### 【0192】

別の実施形態では、配列決定したフラグメントのサイズ情報を、データ解析に組み入れることができる。腫瘍由来DNAは、血漿中非腫瘍由来DNAより短いので (米国特許出願第13/308,473号参照) 、潜在的腫瘍由来突然変異を含むより短い血漿中DNAフラグメントの陽性予測値は、より長い血漿中DNAフラグメントより高いであろう。サイズデータは、もし、血漿中DNAのペアエンドシークエンシングを実施したならば、容易に得られるだろう。別の方法として、長いリード長を用いたDNAシークエンサーを使用し、それゆえ、血漿中DNAフラグメントの完全長を得られ得る。DNAシークエンシング前に、血漿中DNA試料のサイズ分画も実施できる。サイズ分画を使用できる方法の例としては、ゲル電気泳動法、マイクロフルイディクス法 (例えば、Capillary Lab Chip XTシステム) およびサイズ排除スピンカラムの使用が挙げられる。30

#### 【0193】

さらに別の実施形態では、非造血系癌を有する患者の血漿中腫瘍関連突然変異の分画濃度は、もし、血漿中のより短いDNAフラグメントに注目するならば、増加すると予想されることになる。ひとつの実施では、2つ以上の異なるサイズ分布のDNAフラグメントの血漿中腫瘍関連突然変異の分画濃度を比較し得る。非血液学的癌を有する患者は、より長いフラグメントと比較したとき、より短いフラグメントの腫瘍関連突然変異のより高分画濃度を有するだろう。40

#### 【0194】

いくつかの実施形態では、2つ以上の同血液試料アリコットから、または同じ機会または異なる機会に採取した2つ以上の血液試料からのシークエンス結果を1つに合わせることができる。1つ以上のアリコットまたは試料に見られる潜在的突然変異は、腫瘍関連突然変異の高陽性予測値を有することになる。該陽性予測値は、そのような突然変異を示す50

試料数に比例して増加することになる。異なる時点で採取した血漿試料中に存在する潜在的突然変異は、潜在的突然変異と見なされ得る。

#### 【実施例】

##### 【0195】

###### X I . 実施例

次は、技術およびデータの例であり、本発明の実施形態に限定すると考えるべきでない。

###### A . 材料および方法

試料収集に関して、肝細胞癌（HCC）患者、慢性B型肝炎保因者、および乳癌および卵巣癌を同時に患っている患者を募集した。全HCC患者は、バルセロナクリニック肝臓癌ステージA1の疾患有していた。全被験者からの末梢血液を、EDTA含有チューブに収集した。HCC患者の腫瘍組織を、その癌切除手術中に得た。

##### 【0196】

末梢血液試料を、4で10分間、1,600gで遠心分離した。血漿部を、4で10分間、16,000gで再遠心分離して、それから、80で貯蔵した。4.8mLの血漿から無細胞性DNA分子を、QIAamp DSP DNA B100dミニキット（キヤゲン社）の血液および体液プロトコルに従って抽出した。血漿中DNAを、SpeedVac濃縮装置（Savant DNA 120；サーモサイエンティフィック社）を用いて、DNAシークエンスライブラーの後に続く調製のための容器当たり40μlの最終容量に濃縮した。

##### 【0197】

ゲノムDNAを、QIAamp DSP DNA B100dミニキットの血液および体液プロトコルに従って、患者のバフィーコート試料から抽出した。DNAを、QIAamp DNAミニキット（キヤゲン社）を用いて、腫瘍組織から抽出した。

##### 【0198】

ゲノムDNA試料のシークエンスライブラーは、製造者の指示書に従って、ペアエンド試料調製キット（イルミナ社）で構築した。手短に言えば、ゲノムDNAの1~5マイクログラムを、コバリス社S220超音波破碎機を用いて、200bpフラグメントへ、最初に剪断した。その後、DNA分子を、T4DNAポリメラーゼおよびクレノウポリメラーゼで末端修復し；それから、T4ポリヌクレオチドキナーゼを、その5'末端をリン酸化するために使用した。3'オーバーハングを、3'末端から5'末端方向のエキソヌクレアーゼ欠乏のクレノウフラグメントを用いて生成した。イルミナ社アダプターオリゴヌクレオチドを、粘着末端にライゲートした。アダプターをライゲートしたDNAを、12サイクルのPCRで濃縮した。血漿中DNA分子が短いフラグメントであり、血漿試料の全DNA量が、比較的小さかったので、断片化工程を省略し、血漿試料からDNAライブラーを構築するとき、15サイクルのPCRを使用した。

##### 【0199】

アジレント2100バイオアナライザー（アジレントテクノロジーズ社）を、アダプターをライゲートしたDNAライブラーの品質およびサイズを確認するために使用した。それから、DNAライブラーを、製造者の指示書に従って、カパライブラークオリフィケーションキット（カパバイオシステムズ社）により、測定した。DNAライブラーを希釈し、ペアエンドシークエンシングフロー細胞にハイブリダイズした。DNAクラスターを、Truseq PEクラスター形成キットバージョン2（イルミナ社）を用いて、cBotクラスター形成システム（イルミナ社）上に形成し、Truseq SBSキットバージョン2（イルミナ社）を用いて、HiSeq 2000システム（イルミナ社）上に、51\_2サイクルまたは76\_2サイクルを配列決定した。

##### 【0200】

ペアエンドシークエンスデータを、短いオリゴヌクレオチドアライメントプログラム2（SOAP2）の方法により、ペアエンド法で解析した。各ペアエンドリードのため、各末端から50bpまたは70bpを、非反復マスクした参照ヒトゲノム（hg18）にア

10

20

30

40

50

ライメントした。2ヌクレオチドまでのミスマッチを、各末端のアライメントで許容した。それから、2末端のこれらの可能なアライメントのゲノム座標を、いずれもの組み合わせが、正しい配向性を有する、600bpより小さいまたは等しい挿入サイズにまたがっている、および参照ヒトゲノムの単一の位置にマッピングされている同じ染色体にアライメントされる2末端を可能にするかどうかを確認するために、解析した。複製リードは、インサートDNA分子がヒトゲノム中の特定された開始および終止位置を示すペアエンドリードとして定義され；該複製リードは、前述したように、除去された(Lo et al. Sci Transl Med 2010; 2: 61ra91)。

#### 【0201】

いくつかの実施形態では、対の腫瘍および構成的DNA試料を、腫瘍関連一ヌクレオチド異型(SNV)を確認するために、配列決定した。いくつかの実施では、構成的DNA(この例では、バフィコートDNAである)のヘテロ接合性部位におけるSNVの発生に着目した。原則的に、構成的DNA中にはないが、腫瘍組織のシークエンスデータで検出されたいずれものヌクレオチド変異は、潜在的突然変異(すなわち、SNV)になり得た。しかしながら、シークエンスエラー(配列決定されたヌクレオチドの0.1%~0.3%)のため、多くの偽陽性は、もし、腫瘍組織のシークエンスデータのいずれかのヌクレオチド変化の単一発生が腫瘍関連SNVと見なされたならば、ゲノム中に確認されることになる。偽陽性の数を減じるひとつ的方法は、腫瘍関連SNVが誘発する前に、腫瘍組織中のシークエンスデータの同じヌクレオチド変化の複数の発生の観察の判断基準を策定することであろう。

10

#### 【0202】

シークエンスエラーの発生が確率過程であるので、シークエンスエラーによる偽陽性の数は、腫瘍関連SNVと見なされる観察されたSNVに必要な発生数の増加に対して、指數関数的に比例して減少することになる。他方では、偽陽性数は、シークエンス深度の増加に比例して増加することになる。これらの関係性を、ポアソンおよび二項式分布関数で予測できる。実施形態は、観察されたSNVを関連する腫瘍と見なすための発生の動的カットオフを決定し得る。実施形態は、腫瘍シークエンスデータ中の特定ヌクレオチドの実際範囲、シークエンスエラー比率、可能な最大偽陽性比率、および所望の突然変異検出感度を考慮し得る。

20

#### 【0203】

いくつかの例では、偽陽性を低減する非常にストリンジエントな判断基準を指定した。例えば、突然変異は、構成的DNAシークエンシング中に完全に存在しないことが必要であり得、特定のヌクレオチド位置のシークエンス深度は、20倍である必要があった。いくつかの実施形態では、発生のカットオフは、 $10^{-7}$ 以下の偽陽性検出率を達成した。いくつかの例では、アライメントの人為的結果のため、偽陽性を最小限にするセントロメアの、テロメアの、および低複雑度領域内であるSNVを除去もした。加えて、dbSNPビルド135データベースの既知SNPへの仮想SNVマッピングも、除去された。

30

#### B. 切除前および切除後

#### 【0204】

図13Aは、本発明の実施形態に従った診断時の卵巣癌および乳癌を有する患者の血漿中突然変異解析の表1300を示す。ここに、両側性卵巣癌および乳癌を有する患者の例を示す。血漿のシークエンスデータを、該患者(バフィコート)の構成的DNAのシークエンシング結果と比較した。血漿中に存在するが、構成的DNA中に存在しない一ヌクレオチド変化を、潜在的突然変異と見なした。該患者の右側および左側の卵巣癌を各々、2つの部位で試料採取し、全4つの腫瘍試料を作成した。腫瘍突然変異は、4つの異なる部位における全4つの卵巣腫瘍組織で検出した突然変異であった。

40

#### 【0205】

3.6百万個以上の一ヌクレオチド変化を、シークエンシングにより、少なくとも1回の血漿中で検出した。これらの変化内で、2,064個のみも、0.06%の陽性予測値を得る腫瘍組織で検出した。血漿中少なくとも2回検出される判断基準を使用して、潜在

50

的突然変異の数は、有意に、99.5%減少し、18,885となった。腫瘍突然変異の数は、3%減少したのみで、2,003となり、陽性予測値は、増加して、11%となった。

#### 【0206】

血漿中、少なくとも5回検出の判断基準を使用して、2,572潜在的突然変異を検出し、それらの中で、1,814が、全腫瘍組織中で検出された突然変異であり、従って、71%の陽性予測値を得た。発生数（例えば、2、3、4、6、7、8、9、10、その他）の他の判断基準は、感度および必要な陽性予測値に依存する潜在的突然変異を定義するために使用され得る。より高い発生数が、基準として使用されるほど、感度低減で、陽性予測値がより高くなることになる。

10

#### 【0207】

図13Bは、本発明の実施形態に従った腫瘍切除後の両側性卵巣癌および乳癌を有する患者の血漿中突然変異解析の表1350を示す。患者の切除手術を行った。血液試料を、卵巣腫瘍および乳癌の切除後1日目で採取した。それから、血漿中DNAを、配列決定した。この試料では、卵巣癌からの突然変異のみ解析した。3百万以上の潜在的突然変異を、血漿中試料で、少なくとも1回検出した。しかしながら、少なくとも5回発生する判断基準を使用して、潜在的突然変異数は、238に減少した。有意な減少が、診断で採取した試料の潜在的突然変異数と比較したとき、および5つの突然変異の同基準を使用したとき、観察された。

#### 【0208】

ひとつの実施形態では、血漿中で検出されたヌクレオチド変化数は、検出、モニターおよび癌患者の予後判定のためのパラメーターとして使用され得る。発生の異なる数は、所望の感度および特異度を達成するための判断基準として使用され得る。より高い腫瘍量、それゆえより悪い予後を有する患者は、血漿中に見られるより高い突然変異量を有すると予想されるだろう。

20

#### 【0209】

そのような解析では、癌の異なる型の突然変異量プロファイルを確認できる。モニターの目的では、治療に応答する患者の血漿中の突然変異量が減少することを知ることになる。もし、腫瘍が再発したならば、例えば、再発の間に、突然変異量が、増加することが予想されるだろう。そのようなモニターは、患者の選択された治療方法の有効性のモニター、および特定治療に対する耐性発現の検知を可能にするだろう。

30

#### 【0210】

血漿中DNAシークエンシングの結果中に見ることができる特定突然変異解析を通して、感度を予測する標的（例えば、上皮成長因子受容体遺伝子の突然変異およびチロシンキナーゼ阻害剤治療の応答）および特定標的治療耐性（例えば、結腸直腸癌のKRAS突然変異およびパニツムマブおよびセツキシマブによる治療耐性）を確認もでき、治療レジメ計画を導くことができる。

#### 【0211】

上記例は、両側性卵巣癌のものである。乳癌の突然変異についても同解析が実行でき、その結果、血漿中のこれらの癌型の両方の突然変異を追跡できるはずである。原発性癌の突然変異およびその転移（単数または複数）を追跡するため、同様な戦略も使用し得る。

40

#### 【0212】

実施形態は、明らかに健康な対象または特定のリスク要因を有する対象（例えば、喫煙状態、ウイルス状態（例えば、肝炎ウイルス保因者、ヒトパピローマウイルス感染対象など））の癌スクリーニングに有用であるはずである。そのような対象の血漿中に見ることができる突然変異量は、対象が特定時間枠内に症候性癌を発症するリスクを示すはずである。したがって、より高い血漿中突然変異量を有する対象は、より低い突然変異量を有するものより、より高いリスクを有すると予想されることになる。さらに、そのような血漿中突然変異量の一過性プロファイルは、リスクの強力な指標にもなるだろう。例えば、もし、対象が、各年に実行した1つの血漿中突然変異量を有し、且つ、もし、該突然変異量

50

が、だんだんと増加しているならば、その結果、この対象は、追加の癌スクリーニング法、例えば、胸部X線、超音波、コンピューター断層撮影、磁気共鳴画像法またはポジトロン放出断層撮影を勧められるべきである。

C. 血漿シークエンシングから突然変異を推定するための動的カットオフ

【0213】

肝細胞癌（HCC）を有する4人の患者および卵巣癌および乳癌を有する1人の患者を、この研究のために募集した。後者の患者では、卵巣癌の解析に着目した。血液試料を、腫瘍切除手術の前後に、各患者から収集した。切除した腫瘍組織も集めた。腫瘍組織から抽出したDNA、術前血液試料の白血球および手術前および手術後血漿試料を、HiSeq 2000シークエンシングシステム（イルミナ社）を用いて、配列決定した。シークエンスデータを、短いオリゴヌクレオチド解析パッケージ2（SOAP2）（Li R et al. Bioinformatics 2009; 25: 1966-1967）を用いて、参照ヒトゲノム配列（hg18）にアライメントした。白血球細胞のDNA配列を、各研究対象の構成的DNAと見なした。

10

【0214】

この例では、腫瘍関連SNMを、血漿中DNAシークエンスデータおよび腫瘍組織を参照しないCGから、最初推定した。それから、血漿からの推定結果を、該推定結果の正確度を確認するために、該腫瘍組織（客観的基準として）から得たシークエンスデータと比較した。その際、該客観的基準は、該腫瘍組織の突然変異を決定するために、該腫瘍組織からのシークエンスデータおよび該構成的配列を比較することにより作成した。この解析で、研究対象の構成的DNAが、ホモ接合性であるヌクレオチド位置に着目した。

20

1. 非標的全ゲノム解析

【0215】

白血球のシークエンス深度、腫瘍組織および各患者の血漿中DNAを、表5に示す。

【0216】

【表5】

表5. 4件のHCC症例の種々の試料のシークエンス深度メディアン

症例	シークエンス深度のメディアン（倍）			
	白血球細胞	腫瘍組織	術前血漿	術後血漿
HCC1	3.9	2.9	2.3	2.4
HCC2	3.9	2.9	2.5	2.8
HCC3	4.6	3.3	1.8	2.1
HCC4	4.6	2.7	2.0	2.3
卵巣癌患者	4.4	5.3	3.7	2.8

30

【0217】

表1に示された血漿突然変異定義の最小発生率の動的カットオフ（ $r$ ）は、各患者の血漿中の突然変異を確定するために使用される。各座位のシークエンス深度が変わり得るので、これは、該カットオフは変わり得、座位のリード全数に対するカットオフの依存度を、効果的に提供する。例えば、該深度メディアンが50未満であるが（表5）個別の座位のシークエンス深度は、大きく変わり、100回以上にわたり得る。

【0218】

シークエンスエラーに加え、エラーの別原因は、アライメントエラーであろう。このタイプのエラーを最小限にするために、突然変異を有する配列リードは、ボウタイアライメントプログラム（Langmead B et al. Genome Biol 2009, 10:R25）を使用して、参照ゲノ

40

50

ムに再アライメントされた。S O A P 2 により参照ゲノムの独自位置にアライメントされ得たリードのみおよびボウタイを、血漿突然変異のダウンストリーム解析に使用した。異なるアルゴリズムに基づいて、アライメントソフトウェアパッケージの他の組み合わせも使用した。

#### 【 0 2 1 9 】

実際のシークエンスデータのシークエンスエラーおよびアライメントエラーをさらに最小化するために、配列リードの－ヌクレオチド変異を示すヌクレオチド位置を呼び出す2つの追加フィルタリングアルゴリズムを適用した：(1)突然変異を有する 70% の配列リードは、マッピング品質 Q 2 0 (すなわち、ミスアライメント確率 < 1%) を有するボウタイを用いて、同じゲノム座標に再アライメントし得た；(2)突然変異を有する

70% の配列リードは、該配列リードの両末端 (すなわち、5'末端および3'末端) の 5 bp 以内でなかった。シークエンスエラーが配列リードの両末端でより高頻度に見られるので、このフィルタリング規則を研究した。

#### 【 0 2 2 0 】

腫瘍ゲノムの予備知識なしで、腫瘍の推定に影響する因子も研究した。そのようなひとつのパラメーターは、血漿中腫瘍由来DNAの分画濃度であった。このパラメーターを、別の客観的標準パラメーターと見なし、G A A L を使用して、該腫瘍ゲノムの予備知識で参照目的のため推定できる。

#### 【 0 2 2 1 】

表 6 は、治療前および治療に渡って、血漿中に検出されたヌクレオチド変化を示す。H C C 1 では、該腫瘍ゲノムの予備知識なしで、全 9 6 1 －ヌクレオチド変異を検出した。血漿中に検出されたこれらのヌクレオチド変化の中で、8 2 8 は、癌関連突然変異であった。H C C の切除手術後、ヌクレオチド変化の全数を、4 3 に減少し、癌関連突然変異は、それの中になかった。

#### 【 0 2 2 2 】

参照目的では、術前血漿中腫瘍由来DNAの分画濃度は、5 3 % であり、該腫瘍ゲノムの予備知識で推定した。H C C 2 、H C C 3 およびH C C 4 では、該腫瘍ゲノムの予備知識なしで、血漿中－ヌクレオチド変異の数は、術前血漿試料で 2 7 ~ 3 2 の範囲と推定した。これらの結果は、約 2 0 倍のシークエンス深度を用いて、癌関連突然変異の非常に低いパーセンテージを血漿中で検出でき、血漿中で検出されたほとんどの配列変化がシークエンスエラーのためであるという、数学的予測に適合している。腫瘍切除後、検出された配列変化の数に有意な変化はなかった。参照目的では、血漿中腫瘍由来DNAの分画濃度を、2 . 1 % ~ 5 % の範囲と推定し、該腫瘍ゲノムの予備知識で推定した。

#### 【 0 2 2 3 】

10

20

30

【表6】

表6. 血漿中検出されたヌクレオチド変化

	術前血漿			術後血漿		
	腫瘍由来 DNAの 分画濃度	一ヌクレ オチド変 異全数	特定され た癌関連 突然変異 の数	腫瘍由来 DNAの 分画濃度	一ヌクレ オチド変 異全数	特定され た癌関連 突然変異 の数
HCC 1	5.3%	961	828	0.4%	43	0
HCC 2	5%	32	0	0.6%	49	0
HCC 3	2.1%	29	0	0.2%	32	0
HCC 4	2.6%	27	0	1.3%	35	1
卵巣癌 (および 乳癌) 患 者	4.6%	1718	1502	0.2%	2	0

## 2. エクソンの標的濃縮

## 【0224】

上記の通り、解析対象領域のシークエンス深度の増加は、血漿中癌関連突然変異を特定するための感度および特異度の両方を増加させ、それゆえ、癌患者と非癌患者との間の識別力を増加させ得る。全ゲノムのシークエンス深度を増加すると、今もなお非常に費用がかかるが、代替方法は、シークエンシングの特定領域を強化する。ひとつの実施形態では、選択されたエクソンまたは本当に全エクソンを、シークエンシングするため、多くを標的化し得る。この方法は、配列リードの全量を増加することなく、標的領域のシークエンス深度を有意に増加し得る。

## 【0225】

HCC患者および卵巣癌(および乳癌)を有する患者の血漿中DNAのシークエンスライブラリーを、エクソンの標的エンリッチメントのアジレント Sure Select 全エキソンキットを使用して、捕獲した。それから、エクソンを多く含むシークエンスライブラリーを、HiSeq 2000 シークエンシングシステムを使用して、配列決定した。配列リードを、ヒト参照ゲノム(hg18)にアライメントした。アライメント後、エクソンに一意的にマッピングした配列リードを、一ヌクレオチド変異に関して解析した。エクソームキャプチャー解析で血漿中一ヌクレオチド変異を同定するため、表2に示した動的カットオフ値を使用した。

## 【0226】

図14Aは、HCC1の血漿中DNAの一ヌクレオチド変異の検出を示す表1400である。腫瘍ゲノムの予備知識なしで、標的化シークエンスデータから、血漿中の57個の一ヌクレオチド変異の全てを推定した。腫瘍組織から得られたシークエンスデータから、引き続き検証するとき、55個は、真の腫瘍関連突然変異であると分かった。前述のように、術前血漿中腫瘍由来DNAの分画濃度は、5.3%であった。腫瘍切除後、一ヌクレオチド変異は、血漿から得られた標的化シークエンスデータで検出されなかった。これらの

10

20

30

40

50

結果は、血漿中一ヌクレオチド変異数の定量的解析が、癌患者の疾病進行のモニターに使用され得ることを示している。

#### 【0227】

図14Bは、HCC2の血漿中DNAの一ヌクレオチド変異の検出を示す表1450である。腫瘍ゲノムの予備知識なしで、標的化シークエンスデータから、血漿中の18個の一ヌクレオチド変異の全てを推定した。これらの突然変異全ては、腫瘍組織で発見された。前述のように、術前血漿中腫瘍由来DNAの分画濃度は、5%であった。腫瘍切除後、一ヌクレオチド変異は、血漿から得られた標的化シークエンスデータで検出されなかった。血漿中腫瘍由来DNAの高分画濃度を有するHCC1と比較して、HCC2を有する症例の血漿中では、ほとんど一ヌクレオチド変異を検出しなかった。これらの結果は、血漿中腫瘍由来DNA濃度が、腫瘍量と正の相関関係があることが分かっている(Chan KC et al. Clin Chem 2005;51:2192-5)ので、血漿中一ヌクレオチド変異数が、血漿中腫瘍由来DNAの分画濃度、それゆえ、患者の腫瘍量に反映するパラメーターとして使用され得ることを示唆している。

#### 【0228】

図15Aは、HCC3の血漿中DNAの一ヌクレオチド変異の検出を示す表1500である。腫瘍ゲノムの予備知識なしで、標的化シークエンスデータから、切除前および切除後両方の血漿試料中に一ヌクレオチド変異を観察しなかった。これは、たぶん、この患者の血漿中腫瘍由来DNAの比較的低い分画濃度(2.1%)のせいであると思われる。シークエンス深度のさらなる増加は、腫瘍由来DNAの低分画濃度を有する症例での癌関連突然変異を検出するための感度の改善を予測される。

#### 【0229】

図15Bは、HCC4の血漿中DNAの一ヌクレオチド変異の検出を示す表1550である。腫瘍ゲノムの予備知識なしで、標的化シークエンスデータから、血漿中の3個の一ヌクレオチド変異の全てを推定した。これらの突然変異全ては、腫瘍組織で発見された。血漿中腫瘍由来DNAの高分画濃度を有するHCC1およびHCC2と比較して、2%の血漿中分画腫瘍DNAを有するHCC4症例の血漿中で、ほとんど一ヌクレオチド変異は検出されなかった。これらの結果は、血漿中一ヌクレオチド変異数が、血漿中腫瘍由来DNAの分画濃度および患者の腫瘍量を反映するパラメーターとして使用され得ることを示唆している。

#### 【0230】

図16は、卵巣癌(および乳癌)を有する患者の血漿中DNAの一ヌクレオチド変異の検出を示す表1600である。腫瘍ゲノムの予備知識なしで、血漿中の標的化シークエンスデータから、64個の一ヌクレオチド変異の全てを推定した。これらの中で、59個が卵巣腫瘍組織で発見された。血漿中卵巣腫瘍由来DNAの推定分画濃度は、64%であった。一ヌクレオチド変異の全数の有意な減少は、卵巣癌の切除後血漿中で検出された。

#### 【0231】

SureSelect 標的エンリッチメントシステム(アジレント社)の使用に加えて、シークエンシングのため、エクソンから配列を濃縮するために、ニンブルジェンSeqCap EZエクソーム+UTR 標的エンリッチメントシステム(ロシェ社)も使用した。ニンブルジェンSeqCapシステムは、5'および3'非翻訳領域だけでなく、ゲノムのエクソン領域を対象にする。4人のHCC患者の治療前血漿試料、2人の健康な対照群および2人の癌を有しない慢性B型肝炎保因者を解析した(表7)。他の実施形態では、他の標的エンリッチメントシステム、これに限定されないが、溶液相または固相ハイブリダイゼーションを使用するものが、使用され得る。

#### 【0232】

10

20

30

40

## 【表7】

表7. 配列キャプチャーにシンプルジェンSeqCap EZエクソーム+UTR標的エンリッチメントシステムを使用した、4人のHCC患者(HCC1~4)のエキソームシークエンス結果。HCC3の治療前血漿のシークエンス解析を、PCR重複リードの高パーセンテージのため、最適以下であった。

	治療前血漿			治療後血漿	
	G A A L解析による血漿中腫瘍由来DNAの分画濃度	動的カットオフを満たす血漿中で検出された配列変化数	対応する腫瘍組織中に検出された突然変異と重複する配列変化数	動的カットオフを満たす血漿中で検出された配列変化数	対応する腫瘍組織中に検出された突然変異と重複する配列変化数
HCC1	5.3%	6.9	6.4	1	1
HCC2	5%	5.1	4.7	3	0
HCC3	2.1%	0	0	1	0
HCC4	2.6%	8	7	0	0

## 【0233】

2人の癌を有しない慢性B型肝炎保因者および2人の健康な対照群では、動的カットオフ基準値を満たす1つ以下のヌクレオチド変異を検出した(表8)。4人のHCC患者のうち3人では、動的カットオフ要求を満たす血漿中に検出された配列変化数は、少なくとも8個であった。HCC3では、動的カットオフを満たすSNVは、検出されなかった。この試料では、重複のない配列決定されたリード数が低く導かれる配列決定されたリード中のPCR重複リードが高比率であった。血漿中で検出可能なSNVの著しい減少を、腫瘍切除手術後に観察した。

## 【0234】

10

20

30

## 【表8】

表8.配列キャプチャーにニンブルジェンSeqCap EZエクソーム+UTR  
標的エンリッチメントシステムを使用した、2人の慢性B型肝炎保因者(HBV1  
およびHBV2)および2人の健康な対照群(Ctr11およびCtr12)のエ  
キソームシークエンス結果

	動的カットオフを満たす血漿中で検出された配列変化数
HBV1	0
HBV2	1
Ctr11	1
Ctr12	1

## XII.腫瘍不均一性

## 【0235】

生物試料(例えば、血漿/血清)中の一ヌクレオチド突然変異の定量も、腫瘍不均一性、腫瘍内および腫瘍間不均一性の解析に有用である。腫瘍内不均一性は、同じ腫瘍内の腫瘍細胞の複数のクローニーの存在に関係する。腫瘍間不均一性は、同組織型だが、異なる部位(同器官あるいは異器官)に存在する2つ以上の腫瘍の腫瘍細胞の複数のクローニーの存在に関係する。特定型腫瘍では、腫瘍不均一性の存在は、悪い予後指標である(Yoon HH et al. J Clin Oncol 2012; 30: 3932-3938; Merlo LMF et al. Cancer Prev Res 2010; 3: 1388-1397)。特定の型の腫瘍では、腫瘍不均一性程度が高いほど、腫瘍進行の機会または標的治療に続く耐性クローニーの発達がより高くなるはずである。

## 【0236】

癌は、1つの腫瘍細胞のクローニー増殖から生じると思われているが、癌の増殖および進化は、癌の異なる部分での新規の様々な突然変異の集積をもたらすはずである。例えば、癌患者が、転移を発症するとき、最初の器官にある腫瘍および転移性腫瘍は、多くの突然変異を共有するはずである。しかしながら、2つの部位の癌細胞も、他腫瘍部位に存在しない独自突然変異セットを保因するはずである。該2つの部位により共有される突然変異は、1つの腫瘍部位で観察されるだけの突然変異よりも、高濃度に存在すると予測される。

## A.実施例

## 【0237】

両側性卵巣癌および乳癌を有する患者の血漿を解析した。両方の卵巣腫瘍は、重篤な腺癌であった。最長寸法で、左側は6cm、右側は12cmであった。結腸および網に複数の転移性病変もあった。白血球から抽出したDNAを、平均44倍ハプロイドゲノム包括度まで、イルミナ社の合成時解読プラットフォームを使用して、配列決定した。1つのアレルのみ示す、すなわちホモ接合性のヌクレオチド位置を、さらに、血漿の一ヌクレオチド突然変異で解析した。

## 【0238】

DNAを、左側および右側腫瘍の4つの異なる部位から抽出し、イルミナ社シークエンシングプラットフォームを使用して、配列決定した。2つの部位(部位AおよびB)は、右腫瘍から、他の2つの部位(部位CおよびD)は、左腫瘍からであった。部位AおよびBは、ほぼ4cm離れていた。部位CおよびD間の距離も、ほぼ4cmであった。血漿試料を、卵巣腫瘍切除手術前後で、患者から採取した。それから、DNAを、患者の血漿から抽出した。血漿試料だけでなく、部位A、B、CおよびDからの腫瘍のシークエンス深

10

20

30

40

50

度を表9に示す。

【0239】

【表9】

表9. 部位A、B、CおよびDからの腫瘍のシークエンス深度

試料	元シークエンシン グリード数	アライメントさ れたリード数	ハプロイドゲ ノム包括度の 倍数
バフィーコートからの 構成的DNA	1,091,250,072	876,269,922	43.81
右卵巣腫瘍（部位A）	1,374,495,256	1,067,277,229	53.36
右卵巣腫瘍（部位B）	934,518,588	803,007,464	40.15
左卵巣腫瘍（部位C）	1,313,051,122	1,036,643,946	51.83
左卵巣腫瘍（部位D）	1,159,091,833	974,823,207	48.74
前採取血漿試料	988,697,457	741,982,535	37.10
後採取血漿試料	957,295,879	564,623,127	28.23

【0240】

現在の例では、単一腫瘍関連-ヌクレオチド突然変異を定義するために、ヌクレオチド位置を、少なくとも、腫瘍組織で20回、構成的DNAで30回、配列決定する。他の実施形態では、他のシークエンス深度は、例えば、35倍、40倍、45倍、50倍、60倍、70倍、80倍、90倍、100倍および>100倍で、使用され得る。シークエンシング費用低下は、もっとより容易に実行される深度の増加を可能にすることになる。ヌクレオチド位置は、構成的DNAで、ホモ接合性であるのに対して、ヌクレオチド変化は、腫瘍組織で観察される。腫瘍組織中のヌクレオチド変化発生率の基準は、腫瘍組織中の特定のヌクレオチド位置の全シークエンス深度に依存する。20~30倍のヌクレオチド包括度では、ヌクレオチド変化発生率（カットオフ値）は、少なくとも5倍である。31~50倍の包括度では、ヌクレオチド変化発生率は、少なくとも6倍である。51~70倍の包括度では、発生率は、少なくとも7倍である。これらの基準は、真の突然変異検出感度の予測およびポアソン分布を使用した偽陽性座位の予測数から導かれる。

【0241】

図17は、発生率およびシークエンス深度の異なる必要予測感度を示す表1700である。感度は、特定のカットオフを使用して特定倍の深度で検出した真変異数に対応するはずである。より高いシークエンス深度ほど、より多くの突然変異配列リードが得られるよう、変異が、所与のカットオフで、より検出され易い。より高いカットオフ値では、判定基準が、より厳密であるので、突然変異体は検出されにくくなる。

【0242】

図18は、異なるカットオフおよび異なるシークエンス深度の偽陽性座位の予測数を示す表1800である。偽陽性数は、より多くの配列リードが得られるので、シークエンス深度増加とともに増加する。しかしながら、偽陽性は、70までのシークエンス深度でさえ、5以上のカットオフ値で予測されない。他の実施形態では、発生率の異なる基準は、所望の感度および特異性を達成するために使用され得る。

【0243】

10

20

30

40

50

図19は、異なる腫瘍部位で検出された突然変異数を図解する3つのダイアグラムを示す。該突然変異を、直接腫瘍をシークエンシングすることにより決定した。部位Aは、その腫瘍に特有の71個の突然変異を有し、部位Bは、4cmしか離れていないにもかかわらず、122個の部位特有の突然変異を有する。部位AおよびBの両方で、10個の突然変異が見られた。部位Cは、その腫瘍に特有の168個の突然変異を有し、部位Dは、4cmしか離れていないにもかかわらず、248個の部位特有の突然変異を有する。部位CおよびDの両方で、12個の突然変異が見られた。該異なる腫瘍部位の突然変異プロファイルに、有意な不均一性がある。例えば、248個の突然変異は、部位Dでのみ検出されたが、他の3つの腫瘍部位では検出されなかった。総数2,129個の突然変異が、全部位に渡って見られた。したがって、多くの突然変異が、異なる腫瘍の中で共有された。したがって、7つのSNV群があった。コピー数異常に關して、これらの4領域の中で違いは観察されなかった。

#### 【0244】

図20は、治療前および治療後の血漿試料の腫瘍由来突然変異を保因するフラグメント数を示す表2000である。各突然変異を保因する腫瘍由来DNAの推定分画濃度も示した。突然変異カテゴリーは、突然変異が検出される腫瘍部位を表す。例えば、カテゴリーA突然変異は、部位Aにのみ存在する突然変異を表すが、カテゴリーABC'D突然変異は、4つの腫瘍部位の全てで存在する突然変異を表す。

#### 【0245】

全4腫瘍部位に存在する該2,129個の突然変異では、2,105個(98.9%)が、少なくとも1つの血漿中DNAフラグメント中で検出可能であった。他方で、4腫瘍部位の1つのみに存在する該609個の突然変異では、77個(12.6%)のみが、少なくとも1つの血漿中DNAフラグメント中で検出可能であった。したがって、血漿中一ヌクレオチド突然変異の定量は、腫瘍組織中の相対的に大量のこれらの突然変異を反映するために使用され得る。この情報は、癌不均一性の研究に有用であるはずである。この例では、潜在的突然変異は、一旦、シークエンスデータで見られたら、コールされる。

#### 【0246】

循環腫瘍DNAの分画濃度を、各SNV群で決定した。術前および術後の血漿中腫瘍DNAの分画濃度は、全4領域(すなわち群ABC'D)により共有されたSNVにより決定されるとき、それぞれ、46%および0.18%であった。これらの後者のパーセンテージは、GAAAL解析で得られたものとよく相関があり、46%および0.66%であった。全4領域(すなわち群ABC'D)により共有された突然変異は、最も高い腫瘍由来DNA分画濃度を、血漿に提供した。

#### 【0247】

AB群およびCD群からのSNVで決定した術前血漿中腫瘍由来DNAの分画濃度は、それぞれ、9.5%および1.1%であった。これらの濃度は、右側および左側卵巣腫瘍の相対的サイズと一致した。領域特有のSNV(すなわち、A、B、C、およびD群のもの)で決定された腫瘍由来DNAの分画濃度は、概して低かった。これらのデータは、癌患者の全腫瘍量の正確な測定のために、全ゲノムショットガン法の使用が、標的特異性腫瘍関連突然変異のより従来型の方法と比較して、より代表的な実態を提供し得ることを示唆している。後者のアプローチでは、もし、腫瘍細胞のサブセットのみが標的化された突然変異を有するならば、標的化された突然変異を有しない腫瘍細胞が原因で、切迫した再発または疾病進行に関する重要な情報を見逃し得、または治療耐性クローン発生を見逃し得る。

#### 【0248】

図21は、単一腫瘍部位で検出された突然変異および全4腫瘍部位で検出された突然変異の血漿中発生率分布を示すグラフ2100である。棒グラフ2100は、2つのタイプの突然変異：(1)1つの部位のみに検出される突然変異および(2)全4腫瘍部位に検出される突然変異のデータを示す。横軸は、突然変異が血漿中に検出される回数である。縦軸は、横軸の特定値に対応する突然変異%を示す。例えば、タイプ(1)突然変異の約

10

20

30

40

50

88%が、血漿中に1回のみ見られた。ご覧の通り、1つの部位で見られる突然変異は、ほとんど1回検出され、4回以上検出されなかつた。単一腫瘍部位に存在する突然変異は、全4腫瘍部位に存在する突然変異と比較して、血漿中で、ずっと少なくしか検出されなかつた。

#### 【0249】

この技術の1つの応用は、臨床医が、種々の分類の突然変異を保因する腫瘍細胞量を予測することを可能にするはずである。これらの突然変異の比率は、潜在的に、標的薬剤を用いて治療可能であるはずである。腫瘍細胞のより高比率で保因する突然変異を標的とする薬剤は、より顕著な治療効果を有すると期待されるだろう。

#### 【0250】

図22は、不均一腫瘍から発生する突然変異の血漿中発生率予測分布を示すグラフ2200である。該腫瘍は、2群の突然変異を含む。ひとつの群の突然変異は、全腫瘍細胞に存在し、他群の突然変異は、2部位が各卵巣腫瘍の代表的である近似に基づいて、腫瘍細胞の4分の1にのみ存在する。血漿中腫瘍由来DNAの全分画濃度を、40%であると仮定する。血漿試料を、ヌクレオチド位置当たり50回の平均深度まで配列決定されると仮定する。この血漿中発生率予測分布に従って、全腫瘍組織に存在する突然変異は、血漿中それらの発生率による4分の1の腫瘍細胞に存在するのみの突然変異と、差別化され得る。例えば、6回の発生は、カットオフとして使用され得る。全腫瘍細胞に存在する突然変異では、92.3%の突然変異は、少なくとも6回、血漿中に存在することになる。対照的に、4分の1の腫瘍細胞に存在する突然変異では、12.4%のみの突然変異は、少なくとも6回、血漿中に存在することになる。

10

#### 【0251】

図23は、16の健康な対照群の実施形態の特異性が回復されることを示す。それらの血漿中DNA試料を、30倍包括度のメティアンまで配列決定した。上記卵巣癌患者の血漿中に存在する突然変異の検出を、これらの健康な対象中で実行した。卵巣癌患者の腫瘍に存在する突然変異は、健康な対照群の血漿中シーケンスデータにあまり頻繁に検出されず、突然変異カテゴリーが、>1%の明白な分画濃度を有するものはなかつた。これらの結果は、この検出方法が高度に特異性であることを示す。

20

#### B. 方法

#### 【0252】

30

図24は、本発明の実施形態に従って、対象の1つ以上の腫瘍の不均一性を解析する方法2400のフローチャートである。方法2400の特定の工程は、本明細書に記載された通りに実行され得る。

#### 【0253】

ブロック2400で、対象の構成的ゲノムが得られる。ブロック2420で、1つ以上の配列タグは、対象の生物試料の複数のDNAフラグメントの各々のために受けられ、該生物試料は、無細胞性DNAを含む。ブロック2430で、ゲノム位置は、配列タグのために決定される。ブロック2440で、該配列タグは、第一座位の第一数を決定するために、該構成的ゲノムと比較される。各第一座位で、該構成的ゲノムと比べた配列異型を有する配列タグの数は、カットオフ値より上であり、該カットオフ値は、1より大きい。

40

#### 【0254】

ブロック2450で、1つ以上の腫瘍の不均一性の測度は、第一遺伝子位置のセットの各第一数に基づいて算出される。ひとつの態様では、該測度は、腫瘍により共有されない突然変異数に比較して、腫瘍により共有される突然変異数を表す値を提供し得る。ここで、様々な腫瘍は、該物内に異なる腫瘍を有する单一物として存在し得、これは、通常、腫瘍内不均一性と呼ばれるものを表し得る。該測度はまた、いくつかの突然変異が、多くのまたはほとんどの腫瘍中にある突然変異と比較した1つまたは数個の腫瘍中にかかるかどうかも関係し得る。不均一性の1つ以上の測度は、算出され得る。

#### 【0255】

ブロック2460で、該不均一性測度は、不均一性レベルの分類を決定する閾値と比較

50

され得る。測定した 1 つ以上は、様々な方法で使用され得る。例えば、1 つ以上の不均一性測度の手段は、腫瘍進行の変化を予測するために使用され得る。いくつかの腫瘍では、より不均一性であるほど、進行の機会がより高く、治療（例えば、標的治療）後の耐性クローンの発生機会が高い。

### C. 腫瘍不均一性測度

#### 【0256】

不均一性測度の 1 例は、血漿中突然変異の異なる群の「濃縮バンド」の数である。例えば、もし、患者内の 2 つの優位型腫瘍クローンがあるならば、且つ、もし、これらのクローンが、異なる濃度で存在するならば、そのとき、血漿中に異なる濃度を有する 2 つの異なる突然変異を見ることを期待するはずである。これらの異なる値は、各セットが腫瘍の 1 つに対応する、異なる突然変異セットの分画濃度を決定することにより、コンピューターで計算され得る。

10

#### 【0257】

これらの濃度の各々は、「濃縮バンド」または「濃縮クラス」と呼ばれ得る。もし、患者がより多くのクローンを有しているならば、そのとき、より多くの濃縮バンド／クラスが見られるだろう。したがって、バンドがより多いほど、不均一性は、より大きい。濃縮バンド数を、様々な突然変異の分画濃度をプロットすることにより見ることができる。ヒストグラムを、様々な濃度で作成でき、異なるピークが、異なる腫瘍（または 1 つの腫瘍の異なるクローン）に対応する。大きなピークは、おそらく、全てまたはいくつかの腫瘍（または腫瘍クローン）により、共有される突然変異のものであろう。これらのピークは、大きなピークを決定するために、どのより小さなピークが合わされているかを決定するために解析され得る。例えば、図 10B および図 11 のフィッティング手順と同様な、フィッティング手順が使用され得る。

20

#### 【0258】

ひとつの実行では、該ヒストグラムは、Y 軸に、座位の量（例えば、数または比率）を、X 軸に、分画濃度をプロットしている。全てまたはいくつかの腫瘍により共有される突然変異は、高分画濃度をもたらすはずである。該ピークサイズは、特定の分画濃度増加につながる座位の量を表すことになる。低濃度および高濃度でのピークの相対的サイズは、腫瘍（または腫瘍クローン）の不均一性程度を反映するはずである。高濃度での大きなピークは、ほとんどの腫瘍が、ほとんどまたは全ての腫瘍（または腫瘍クローン）により共有されることを反映し、腫瘍不均一性程度が低いことを示している。もし、低濃度でのピークが大きいならば、そのとき、ほとんどの突然変異は、少数の腫瘍（または少数の腫瘍クローン）により共有される。これは、腫瘍不均一性程度が高いことを示していることになる。

30

#### 【0259】

存在するピークがより多いほど、より多くの部位特有突然変異がある。各ピークは、突然変異の異なるセットに対応し得、該突然変異セットは、腫瘍のサブセット（例えば、上記例示のように、ちょうど 1 つか 2 つの腫瘍）からのものである。図 19 の例では、おそらく最小濃度を有する 4 部位のみのピーク（腫瘍の相対的サイズに依存する）、A B 部位および C D 部位の 2 ピーク、および全部位に共有される突然変異のピークで、全 7 ピークがあり得る。

40

#### 【0260】

ピーク位置は、腫瘍の相対的サイズも提供し得る。大きな濃度は、大きな腫瘍が、より多くの腫瘍 DNA を試料中に、例えば、血漿中に放出するので、大きな腫瘍に相関があるはずである。

#### 【0261】

不均一性測度の別の例では、相対的に高い異型リード（例えば、9 ~ 13）を有する突然変異リード比率と比較して、相対的に少数の異型リード（例えば、4、5、または 6）を有する突然変異部位比率である。図 22 に戻って参照すれば、種雄特有突然変異が、ほとんど異型リードを有しない（結果として、より小さい分画濃度にもなる）ことを知り得

50

る。共有突然変異は、より多くの異型リード（結果として、より大きい分画濃度にもなる）を有する。第一比率6（より少数）を第二比率10（より多数）により割った比は、不均一性測度を伝える。もし、該比が小さいならば、そのとき、部位特有である突然変異はほとんどなく、従って、不均一性レベルは低い。もし、該比が大きい（または少なくとも、既知試料から校正した値より大きい）ならば、そのとき、不均一性レベルはより大きい。

#### D. 閾値決定

##### 【0262】

閾値は、不均一性レベルを直接決定するために、その腫瘍が生検される（例えば、上記のように）対象から決定され得る。該レベルは、共有される突然変異に対する部位特有突然変異の比など、様々な方法で定義され得る。生体試料（例えば、血漿試料）は、それから、不均一性測度を決定するために解析され得、生体試料からの不均一性測度は、直接腫瘍細胞を解析することにより決定した不均一性レベルに関連付けされ得る。

10

##### 【0263】

そのような手順は、不均一性レベルに関係する閾値の校正を提供し得る。もし、テストの不均一性測度が、2つの閾値の間に位置するならば、そのとき、不均一性レベルは、該閾値に対応するレベル間にあると予測され得る。

##### 【0264】

ひとつの実施形態では、校正曲線は、生検から決定された不均一性レベルと血漿試料（または他試料）から決定された対応する不均一性測度との間で算出され得る。そのような例では、該不均一性レベルは、数値で表され、これらの数値レベルは、種々の分類に対応し得る。数値レベルの異なる範囲は、異なる診断、例えば、異なる癌状態などに対応し得る。

20

#### E. ゲノム表現から分画濃度を使用する方法

##### 【0265】

腫瘍不均一性は、分画濃度、例えば、方法1200の実施形態を使用して決定したものなどを使用しても、解析され得る。1つのコピー欠失を示すゲノム領域は、異なる腫瘍に由来し得る。したがって、様々なゲノム領域で決定された分画濃度は、増幅（または1コピー欠失の削除）が1つだけの腫瘍または複数の腫瘍中に存在するかどうかに依存して、異なり得る。したがって、同じ不均一性測度は、方法1200の実施形態を経て決定された分画濃度のために使用され得る。

30

##### 【0266】

例えば、1つのゲノム領域は、1コピー欠失に対応すると見なされ得、分画濃度は、そのゲノム領域での各密度（該各密度は、分画濃度として使用できる）からだけ決定され得る。ヒストグラムは、様々な密度を有する領域数をカウントすることにより、様々な各密度から決定され得る。もし、1つだけの腫瘍または1つの腫瘍クローニまたは1つの腫瘍沈着物が、特定領域中に増加を有するならば、そのとき、その領域の密度は、複数の腫瘍または複数の腫瘍クローニまたは複数の腫瘍沈着物中に増加を有する領域の密度（すなわち、共有領域中の腫瘍DNAの分画濃度が、部位特有領域より大きい）より小さいことになる。したがって、上記不均一性測度は、ちょうど、異なる部位の分画濃度が、分画濃度分布を示すように、様々な領域中のコピー数獲得または欠失を使用して、同定されるピークに応用され得る。

40

##### 【0267】

ひとつの実行では、もし、各密度がヒストグラムに使用されるならば、別々になって獲得および欠失を有することになる。獲得を示す領域は、獲得だけのヒストグラムを作成することにより別々に解析でき、別のヒストグラムは、欠失だけのために作成され得る。もし、分画濃度が使用されるならば、そのとき、欠失および獲得のピークは、一緒に解析され得る。例えば、分画濃度は、参照密度に対する差（例えば、絶対値）を使用し、従って、獲得および欠失の分画濃度は、同じピークに寄与する。

## 【0268】

本明細書中で述べたいずれのコンピューターシステムも、いずれかの適當な数のサブシステムを利用し得る。そのようなサブシステムの例を、コンピューター装置 2500 の図 25 に示す。いくつかの実施形態では、コンピューターシステムは、ひとつのコンピューター装置を含み、サブシステムは、コンピューター装置のコンポーネントであり得る。他の実施形態では、コンピューターシステムは、複数のコンピューター装置を含み得、各々は、内部コンポーネントを有するサブシステムである。

## 【0269】

図 25 に示すサブシステムは、システムバス 2575 を介して相互接続されている。プリンター 2574、キーボード 2578、固定ディスク 2579、ディスプレイアダプタ 10 - 2582 に接続されたモニター 2576、および他などの追加サブシステムを示す。周辺機器および I/O コントローラー 2571 に接続したインプット / アウトプット (I/O) 機器を、シリアルポート 2577 などの、当業者に周知のいくつもの手段によりコンピューターシステムに結合し得る。例えば、シリアルポート 2577 または外部インターフェース 2581 (例えば、イーサネット、Wi-Fi、他) を、コンピューターシステム 2500 を、インターネットなどの広域ネットワークに接続するために使用し得る。システムバス 2575 を介する相互接続は、中央演算装置 2573 を、サブシステム間の情報交換だけでなく、各サブシステムと通信およびシステムメモリー 2572 または固定ディスク 2579 からの実行命令を制御することを可能にする。システムメモリー 2572 および / または固定ディスク 2579 は、コンピューター読み取り可能媒体を収納し得る。本明細書で述べた値のいずれも、1 つのコンポーネントから別のコンポーネントにアウトプットし得、使用者にアウトプットし得る。

10

20

30

## 【0270】

コンピューターシステムは、例えば、外部インターフェース 2581 または内部インターフェースにより一緒に接続した複数の同じコンポーネントまたはサブシステムを含み得る。いくつかの実施形態では、コンピューターシステム、サブシステム、または装置は、ネットワーク上で通信し得る。そのような場合には、1 つのコンピューターを、クライアントと考え、別のコンピューターをサーバーと考え得、各々は、同じコンピューターシステムの部分であり得る。クライアントおよびサーバーは、各々、複数のシステム、サブシステム、またはコンポーネントを含み得る。

30

## 【0271】

本発明の実施形態のいずれも、ハードウェア (例えば、特定用途向け集積回路またはフィールドプログラマブルゲートアレイ) を使用しておよび / または、一般的に、モデュールまたは一体でプログラマブルプロセッサを用いたコンピューターソフトウェアを使用して、制御論理の形態で実行し得ると理解すべきである。本明細書で使用するとき、プロセッサーは、同じ集積チップ上のマルチコアプロセッサまたはひとつの回路基板上またはネットワーク上の複数の処理装置を含む。本明細書で提供する開示および教示に基づいて、当業者は、ハードウェアならびにハードウェアおよびソフトウェアの組み合わせを使用して、本発明の実施形態を実行するための他の手段および / または方法を知り、認識するだろう。

40

## 【0272】

本出願に記載のいずれものソフトウェアコンポーネントまたは機能は、例えば、従来技術またはオブジェクト指向技術を使用する、例えば、Java、C++ または Perl などのいずれかの適當なコンピューター言語を使用して、プロセッサーにより実行されるソフトウェアコードとして実行され得る。該ソフトウェアコードは、一連の命令またはコマンドとして、記憶および / または伝送用コンピューター読み取り可能媒体上に記憶され得、適當な媒体としては、ランダムアクセスメモリー (RAM)、読み出し専用メモリー (ROM)、ハードディスクまたはフロッピーディスクなどの磁気媒体、またはコンパクトディスク (CD) または DVD (デジタルバーサタイルディスク) などの光学媒体、フラッシュメモリー、および同様のものが挙げられる。コンピューター読み取り可能媒体は、

50

そのような記憶装置または伝送装置のいずれかの組み合わせであり得る。

【0273】

10  
 そのようなプログラムは、インターネットを含む、様々なプロトコルに従う有線、光、および／または無線ネットワークを介して伝送に適合したキャリア信号を使用して、コード化および送信もされ得る。それ自体として、本発明の実施形態に従ったコンピューター読み取り可能媒体は、そのようなプログラムでコードしたデータ信号を使用して、作成され得る。プログラムコードでコードされたコンピューター読み取り可能媒体は、互換デバイスでパッケージ化され得、他のデバイス（例えば、インターネットダウンロードを介して）から別々に提供され得る。そのようなコンピューター読み取り可能媒体はいずれも、ひとつのコンピュータープログラム製品（例えば、ハードドライブ、CD、または全コンピューターシステム）上または内に存在し得、システムまたはネットワーク内の異なるコンピュータープログラム製品上または内に存在し得る。コンピューターシステムは、本明細書で述べたいずれの結果も、使用者に提供するために、モニター、プリンター、または他の適当なディスプレイを含み得る。

【0274】

20  
 本明細書に記載の方法のいずれも、1つ以上のプロセッサーを有するコンピューターシステムを用いて全部または部分的に実行され得、これは、工程実行のため、構成され得る。したがって、実施形態は、潜在的に、各工程または各工程群を実行する種々のコンポーネントを有する、本明細書に記載のいずれもの方法の工程を実行するために構成されたコンピューターシステムに関し得る。番号が付された工程として表されているが、本明細書中の方法の工程を、同時に、または異なる順番で実施し得る。加えて、これらの工程の部分を、他の方法から他の工程の部分と一緒に使用し得る。また、ある工程の全てまたは部分を任意に選択もし得る。加えて、いずれもの方法の工程のいずれも、モジュール、回路、またはこれらの工程を実行するための他手段で実行し得る。

【0275】

特定の実施形態の具体的な詳細は、本発明の実施形態の精神と範囲から逸脱することなく、いずれもの適切な方法で組み合わせ得る。しかしながら、本発明の他の実施形態は、各個々の態様、またはこれらの個々の態様の特定の組み合わせに関係する特定の実施形態に関し得る。

【0276】

30  
 本発明の例示となる実施形態の上記説明は、例示および説明の目的のために表されている。本発明を全て網羅しているわけではなく、記載されたそのままの形態に本発明を限定するものではなく、上記教示に照らして、多くの修正および変化は可能である。実施形態は、本発明の原理およびその実際的用途を最も良く説明し、それにより、他の当業者が、様々な実施形態で、および企図される特定用途に適するように様々な修正を行って、本発明を最良に利用できるように、選択され説明された。

【0277】

「ひとつの(a)」、「ひとつの(an)」または「該(the)」は、特にそれとは反対の指示がない限り、「1つ以上の」を意味することを意図している。

【0278】

40  
 全ての特許、特許出願、文献および本明細書で述べた説明は、全ての目的で、その全文を参照することにより組み入れられる。先行技術と認められるものはない。

【図1】

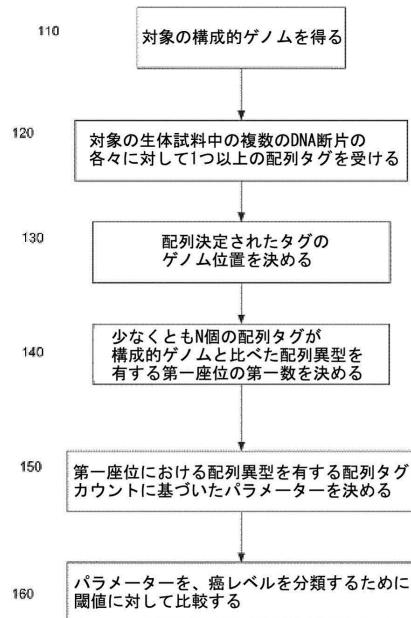


図1

【図2】

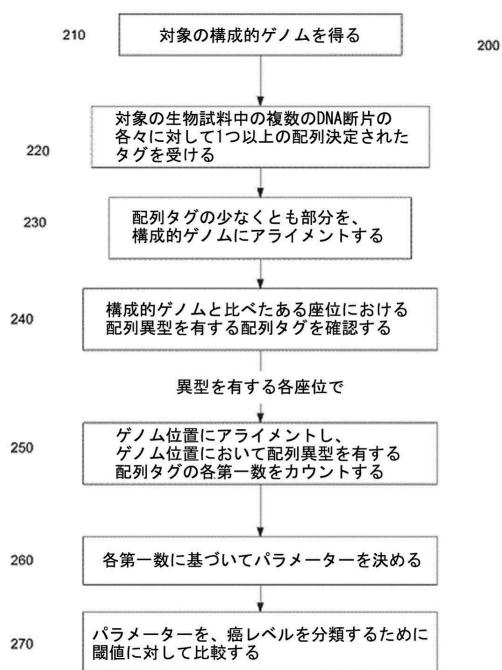


図2

【図3】

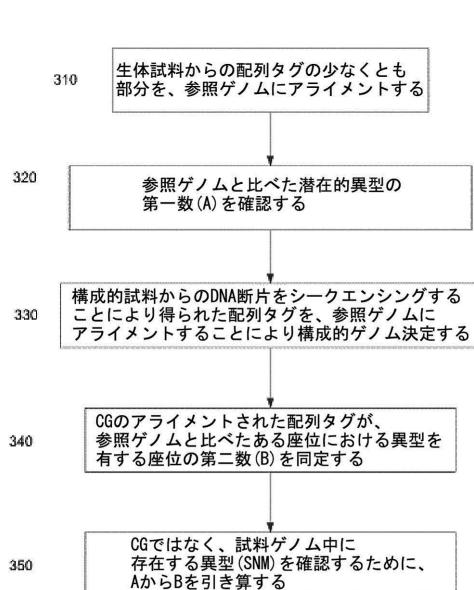


図3

【図4】

観察される回数	シーケンスエラーの深度	シーケンス確認された突然変異の部立	シーケンス確認された突然変異の部立の期待数	シーケンス確認された突然変異の部立の期待数の期待数	シーケンス確認された突然変異の部立の期待数の期待数	シーケンス確認された突然変異の部立の期待数の期待数	シーケンス確認された突然変異の部立の期待数の期待数	シーケンス確認された突然変異の部立の期待数の期待数	シーケンス確認された突然変異の部立の期待数の期待数
1	29,891,000	1,099	1,099	0	0	0	0	0	0
10	29,891,000	1,130	45,000	210	40	41	0	5	0
20	59,403,959	1,896	1,800,000	793	380	311	1	57	0
40	11,531,683	2,594	780,000	1,392	3,203	970	10	159	0
60	174,766,399	2,851	1,170,000	2,046	11,407	1,753	54	1,058	0
80	236,659,961	2,845	3,600,000	2,725	21,387	2,386	176	1,209	1
100	285,187,746	2,840	4,950,000	2,879	53,900	2,635	436	2,203	3
120	338,238,692	2,893	7,400,000	2,968	93,615	2,814	913	2,545	7
140	591,225,294	2,997	9,750,000	2,978	149,013	2,911	1,708	2,753	15
160	441,488,633	2,899	12,720,000	2,961	253,207	2,959	2,922	2,873	30
180	494,189,566	3,000	16,110,000	2,986	318,620	2,981	4,700	2,956	55
200	543,267,741	3,000	19,940,000	2,969	457,980	2,992	7,187	2,969	94
									2,912
									1,299

図4

【図5】

繰り返される回数	シーケンス深度	シーケンス確認された突然変異の部位期待数	シーケンス深度	シーケンス確認された突然変異の部位期待数	シーケンス深度	シーケンス確認された突然変異の部位期待数	シーケンス深度	シーケンス確認された突然変異の部位期待数
1	2,099,500	74	1,499	1	0	0	0	0
16	29,830,499	654	45,000	79	46	6	0	0
19	59,403,590	1,180	190,000	271	380	43	1	0
40	117,611,683	1,896	780,000	793	1,203	10	57	0
60	174,765,399	2,334	1,770,000	1,527	11,407	53	197	0
80	284,650,651	24,594	3,160,000	1,782	27,387	970	176	429
100	386,485,746	27,554	4,950,000	2,138	53,900	1,369	436	727
120	386,318,690	2,851	7,140,000	2,003	9,613	1,750	913	1,058
140	391,925,294	2,969	9,700,000	2,926	10,935	2,687	1,903	1,390
160	443,566,635	27,945	12,720,000	2,725	23,307	2,886	2,922	1,760
180	494,189,566	24,567	16,180,000	2,417	318,610	2,479	4,700	1,973
200	543,307,741	2,900	19,900,000	2,759	437,900	2,636	7,187	2,265

図5

【図6 A】

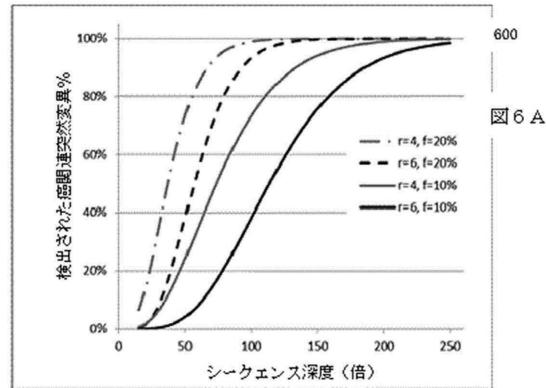


図6 A

【図6 B】

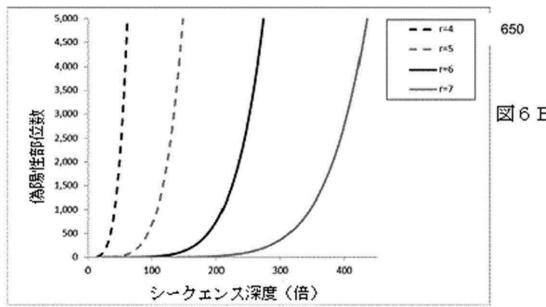


図6 B

【図7 A】

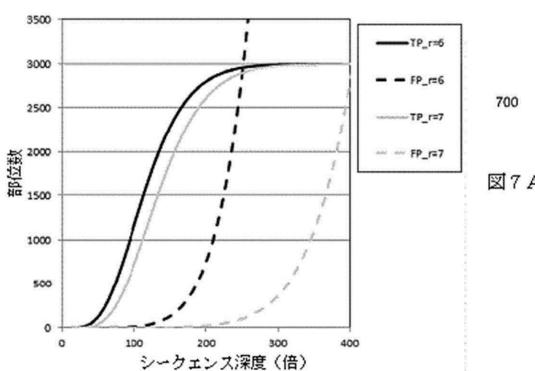


図7 A

【図7 B】

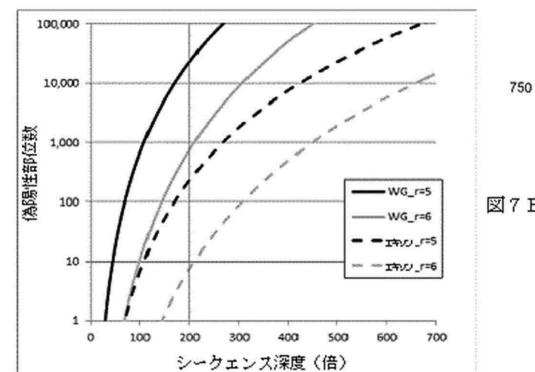


図7 B

【図8】

症例	腫瘍組織中で検出されたSNV数	時点	血漿から配列決定した腫瘍関連SNV数(腫瘍組織中で観察されるSNV%)	SNVを示す血漿中DNAシーケンスリード数(p)	野生型配列DNAシーケンスリード数(q)	SNV解析によるGAA塩基による塩基由来DNAの推定分画濃度( $\frac{2p}{p+q}$ )	GAA塩基によるSNV解析による塩基由来DNAの推定分画濃度( $\frac{2p}{p+q}$ )	
							Pre-Tx	Post-Tx
HCC1	2,840	Pre-Tx	2,569 (94%)	11,389	31,602	53%	52%	52%
		Post-Tx	44 (1.5%)	91	46,888	0.4%	1.4%	1.4%
HCC2	3,105	Pre-Tx	1,897 (35%)	1,490	57,865	5.0%	4.2%	4.2%
		Post-Tx	72 (2.3%)	206	66,692	0.6%	0.5%	0.5%
HCC3	3,171	Pre-Tx	451 (15%)	525	48,866	2.1%	5.6%	5.6%
		Post-Tx	31 (1%)	67	58,862	0.2%	0.5%	0.5%
HCC4	1,334	Pre-Tx	201 (15%)	248	18,527	2.6%	7.0%	7.0%
		Post-Tx	74 (5.5%)	149	22,144	1.3%	2.7%	2.7%

図8

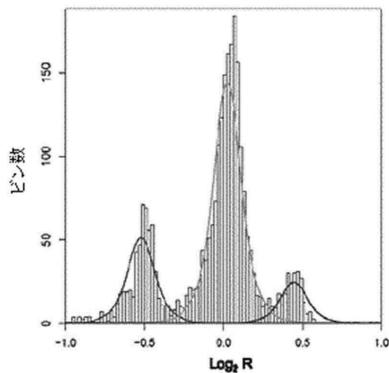
【図 9】

	HCC1 SNV	HCC2 SNV	HCC3 SNV	HCC4 SNV	All SNV
腫瘍関連 野生型 配列を SNVを 示す	腫瘍関連 野生型 配列を SNVを 示す	腫瘍関連 野生型 配列を SNVを 示す	腫瘍関連 野生型 配列を SNVを 示す	腫瘍関連 野生型 配列を SNVを 示す	腫瘍関連 野生型 配列を SNVを 示す
対照	DNA シグナル リード	DNA シグナル リード	DNA シグナル リード	DNA シグナル リード	DNA シグナル リード
C01	40	31,251	66	38,172	30
C02	60	34,368	112	42,660	37
C03	55	37,931	100	46,207	50
C04	41	40,277	37	48,441	39
C05	42	3,655	71	38,787	46
C06	41	34,604	97	43,999	26
C07	52	38,654	118	47,200	61
C08	57	38,514	131	46,464	42
C09	41	34,945	115	43,760	33
C10	44	35,197	89	45,622	51
C11	41	36,693	95	45,108	39
C12	45	39,506	104	50,085	46
C13	36	33,399	71	46,309	39
C14	42	38,199	104	47,832	42
C15	49	36,106	90	45,129	41
C16	38	33,633	66	40,023	41

900

図 9

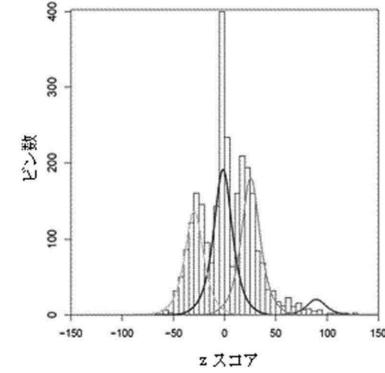
【図 10 A】



1000

図 10 A

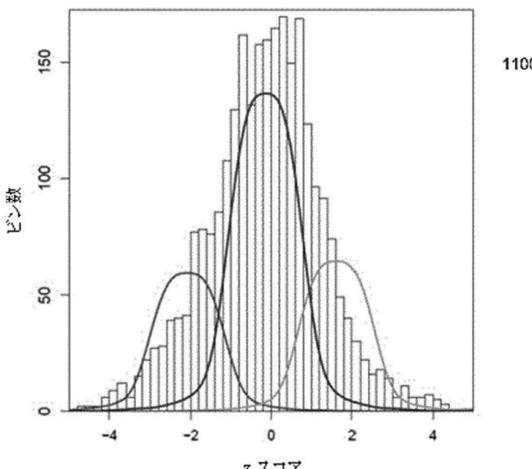
【図 10 B】



1050

図 10 B

【図 11】



1100

図 11

【図 12】

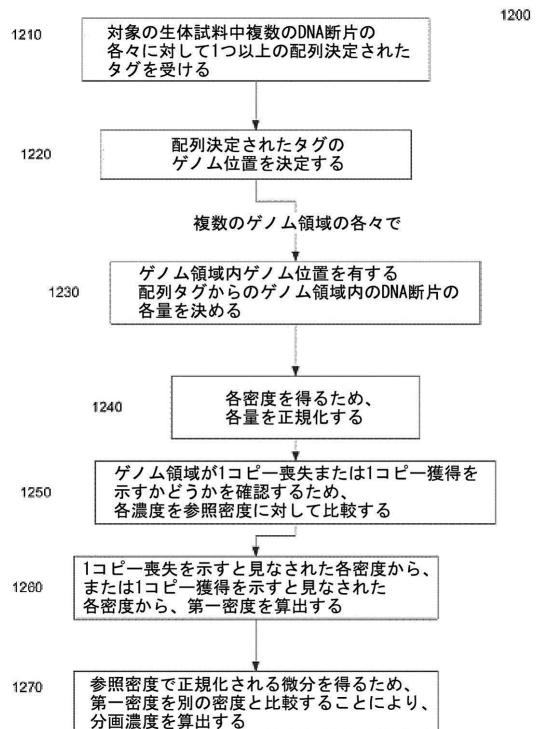


図 12

【図13A】

血漿中検出された回数(N)	血漿中N回検出された潜在的突然変異数	血漿中 $\geq N$ 回検出された潜在的突然変異数	血漿中 $\geq N$ 回検出された腫瘍突然変異	血漿中 $\geq N$ 回検出された腫瘍突然変異	N=1と比較した被検出突然変異%	陽性予測値
1	3,605,085	3,623,970	61	2,064	100%	0.06%
2	13,518	18,885	35	2,003	97%	11%
3	2,029	5,367	71	1,968	95%	37%
4	766	3,338	83	1,897	92%	57%
5	437	2,572	140	1,814	88%	71%
6	313	2,135	161	1,674	81%	78%
7	265	1,822	165	1,513	73%	83%
8	261	1,557	192	1,348	65%	87%
9	231	1,296	185	1,156	56%	89%
10	191	1,065	160	971	47%	91%
11	179	874	165	811	39%	93%
12	148	695	137	646	31%	93%
13	122	547	109	509	25%	93%
14	99	425	92	400	19%	94%
15	74	326	71	308	15%	94%
16	54	252	49	237	11%	94%
17	39	198	39	188	9%	95%
18	35	159	33	149	7%	94%
19	24	124	23	116	6%	94%
20	20	100	20	93	5%	93%
21	18	80	17	73	4%	91%
22	11	62	10	56	3%	90%
23	9	51	9	46	2%	90%
24	8	42	5	37	2%	88%
25	6	34	6	32	3%	94%
26	7	28	5	26	1%	93%
27	8	21	8	21	1%	100%
28	1	13	1	13	1%	100%
29	4	12	4	12	1%	100%
30	8	8	8	8	0%	100%
合計			57	55	1.7%	0

図13A

1300

【図13B】

血漿中検出された回数(N)	血漿中N回検出された潜在的突然変異数	血漿中 $\geq N$ 回検出された潜在的突然変異数	血漿中 $\geq N$ 回検出された腫瘍突然変異	血漿中 $\geq N$ 回検出された腫瘍突然変異	N=1と比較した被検出突然変異%	陽性予測値
1	3,155,634	3,167,144	51	54	100%	<0.01%
2	9,835	11,510	2	3	6%	0.03%
3	1,123	1,675	1	1	2%	0.06%
4	314	552	-	-	-	-
5	132	238	-	-	-	-
6	49	106	-	-	-	-
7	22	57	-	-	-	-
8	11	35	-	-	-	-
9	6	24	-	-	-	-
10	5	18	-	-	-	-
11	4	13	-	-	-	-
12	2	9	-	-	-	-
13	2	7	-	-	-	-
14	3	5	-	-	-	-
15	-	2	-	-	-	-
16	-	2	-	-	-	-
17	1	2	-	-	-	-
18	-	1	-	-	-	-
19	1	1	-	-	-	-
20	-	-	-	-	-	-
21	-	-	-	-	-	-
22	-	-	-	-	-	-
23	-	-	-	-	-	-
24	-	-	-	-	-	-
25	-	-	-	-	-	-
26	-	-	-	-	-	-
27	-	-	-	-	-	-
28	-	-	-	-	-	-
29	-	-	-	-	-	-
30	-	-	-	-	-	-

図13B

1300

【図14A】

特定シーケンス深度	治療前血漿		治療後血漿		腫瘍中の変化を有する部位総数
	血漿中検出された潜在的突然変異を有する部位総数	血漿中N回検出された潜在的突然変異を有する部位総数	血漿中 $\geq N$ 回検出された潜在的突然変異を有する部位総数	血漿中 $\geq N$ 回検出された潜在的突然変異を有する部位総数	
<50	36.3%	9	8	30.5%	0
50 - 125	29.9%	15	15	31.2%	0
126 - 235	18.0%	21	21	20.8%	0
236 - 380	9.4%	7	6	10.9%	0
381 - 560	4.2%	2	2	4.8%	0
561 - 760	1.6%	3	3	1.7%	0
合計		57	55		0

図14A

1400

【図14B】

特定シーケンス深度	治療前血漿		治療後血漿		腫瘍中の変化を有する部位総数
	血漿中検出された潜在的突然変異を有する部位総数	血漿中N回検出された潜在的突然変異を有する部位総数	血漿中 $\geq N$ 回検出された潜在的突然変異を有する部位総数	血漿中 $\geq N$ 回検出された潜在的突然変異を有する部位総数	
<50	30.0%	0	0	27.1%	0
50 - 125	33.6%	2	2	32.5%	0
126 - 235	21.6%	6	6	22.2%	0
236 - 380	10.2%	5	5	11.5%	0
381 - 560	3.7%	4	4	4.9%	0
561 - 760	0.9%	1	1	1.7%	0
合計		18	18	0	0

図14B

1450

【図15A】

シーケンス 深度	治療前血漿		治療後血漿	
	特定 シーケンス 深度での マスクレオチド%	血漿中検出 されたシーケンス 変異を有する マスクレオチド 位置総数	腫瘍中の 対応する 変化を有する 部位置数	血漿中検出 されたシーケンス 変異を有する マスクレオチド 位置総数
<50	34.8%	0	0	18.5%
50 - 125	45.7%	0	0	28.6%
126 - 235	15.2%	0	0	25.1%
236 - 380	0.2%	0	0	16.2%
381 - 560	<0.1%	0	0	3.3%
561 - 760	<0.1%	0	0	3.4%
合計		0	0	0

【図16】

治療前血漿 シーケンス 深さ	治療後血漿				
	血漿中検出 された スクレオチド スクレオチド% 位置総数	腫瘍中の 変化を有する 部位数 位置総数	特定 シーケンス 下での スクレオチド% 位置総数	血漿中検出 された スクレオチド スクレオチド% 位置総数	腫瘍中の 変化を有する 部位数 位置総数
<50	31.1%	7	7	30.8%	0
50 - 125	30.3%	14	14	33.2%	1
126 - 235	20.5%	19	19	21.6%	0
236 - 380	11.3%	13	13	10.3%	0
381 - 560	5.2%	7	4	3.5%	0
561 - 760	1.8%	4	2	0.7%	0
合計		64	39		1
					0

1600

【図15B】

【図17】

シーケンス 深度(倍)	突然変異体が少なくともN回の腫瘍組織シーケンスデータ 中に見られる定義されるときの感度、N=								
	4	5	6	7	8	9	10	11	12
20	73%	56%	38%	24%	13%	7%			
21	77%	60%	43%	28%	16%	9%			
22	80%	64%	47%	31%	19%	11%			
23	83%	68%	51%	35%	22%	13%			
24	85%	71%	55%	39%	26%	15%			
25	87%	75%	59%	43%	29%	18%			
26	89%	78%	63%	47%	33%	21%			
27	90%	80%	67%	51%	36%	24%			
28	92%	83%	70%	55%	40%	27%			
29	93%	85%	73%	59%	44%	30%			
30	94%	87%	76%	62%	48%	34%			
31	95%	88%	78%	66%	51%	37%			
32	96%	90%	81%	69%	55%	41%			
33	96%	91%	83%	72%	58%	44%			
34	97%	93%	85%	74%	61%	48%			
35	97%	94%	87%	77%	65%	51%			
36	98%	95%	88%	79%	68%	54%			
37	98%	95%	90%	82%	70%	58%			
38	99%	96%	91%	84%	73%	61%			
39	99%	97%	92%	85%	76%	64%			
40	99%	97%	93%	87%	78%	67%			
41	99%	98%	94%	88%	80%	69%			
42	99%	98%	95%	90%	82%	72%			
43	99%	98%	96%	91%	84%	75%			
44	100%	98%	96%	92%	86%	77%			
45	100%	99%	97%	93%	87%	79%			
46	100%	99%	97%	94%	89%	81%			
47	100%	99%	98%	95%	90%	83%			
48	100%	99%	98%	95%	91%	84%			
49	100%	99%	98%	96%	92%	86%			
50	100%	99%	99%	97%	93%	88%			
51	100%	100%	99%	97%	94%	89%			
52	100%	100%	99%	97%	95%	90%			
53	100%	100%	99%	98%	95%	91%			
54	100%	100%	99%	98%	96%	92%			
55	100%	100%	99%	98%	96%	93%			
56	100%	100%	99%	99%	97%	94%			
57	100%	100%	100%	99%	97%	95%			
58	100%	100%	100%	99%	98%	95%			
59	100%	100%	100%	99%	98%	96%			
60	100%	100%	100%	99%	98%	96%			
61	100%	100%	100%	99%	98%	97%			
62	100%	100%	100%	99%	99%	97%			
63	100%	100%	100%	100%	99%	98%			
64	100%	100%	100%	100%	99%	98%			
65	100%	100%	100%	100%	99%	98%			
66	100%	100%	100%	100%	99%	98%			
67	100%	100%	100%	100%	99%	99%			
68	100%	100%	100%	100%	100%	99%			
69	100%	100%	100%	100%	100%	100%			
70	100%	100%	100%	100%	100%	100%			

図17

1700

【図18】

シーケンス 深度(倍)	突然変異体が少なくともN回の腫瘍組織シーケンステータ中に 見られると定義されるときの全ゲノム中の偽陽性数、N=								
4	5	6	7	8	9	10	11	12	
20	1	0	0	0	0	0	0	0	0
21	1	0	0	0	0	0	0	0	0
22	1	0	0	0	0	0	0	0	0
23	1	0	0	0	0	0	0	0	0
24	1	0	0	0	0	0	0	0	0
25	1	0	0	0	0	0	0	0	0
26	2	0	0	0	0	0	0	0	0
27	2	0	0	0	0	0	0	0	0
28	2	0	0	0	0	0	0	0	0
29	3	0	0	0	0	0	0	0	0
30	3	0	0	0	0	0	0	0	0
31	3	0	0	0	0	0	0	0	0
32	4	0	0	0	0	0	0	0	0
33	5	0	0	0	0	0	0	0	0
34	5	0	0	0	0	0	0	0	0
35	6	0	0	0	0	0	0	0	0
36	7	0	0	0	0	0	0	0	0
37	7	0	0	0	0	0	0	0	0
38	8	0	0	0	0	0	0	0	0
39	9	0	0	0	0	0	0	0	0
40	10	0	0	0	0	0	0	0	0
41	11	0	0	0	0	0	0	0	0
42	12	0	0	0	0	0	0	0	0
43	14	0	0	0	0	0	0	0	0
44	15	0	0	0	0	0	0	0	0
45	17	0	0	0	0	0	0	0	0
46	18	0	0	0	0	0	0	0	0
47	20	0	0	0	0	0	0	0	0
48	22	0	0	0	0	0	0	0	0
49	24	0	0	0	0	0	0	0	0
50	26	0	0	0	0	0	0	0	0
51	28	0	0	0	0	0	0	0	0
52	30	0	0	0	0	0	0	0	0
53	33	0	0	0	0	0	0	0	0
54	35	0	0	0	0	0	0	0	0
55	38	0	0	0	0	0	0	0	0
56	41	0	0	0	0	0	0	0	0
57	44	0	0	0	0	0	0	0	0
58	47	0	0	0	0	0	0	0	0
59	51	0	0	0	0	0	0	0	0
60	54	0	0	0	0	0	0	0	0
61	58	0	0	0	0	0	0	0	0
62	62	0	0	0	0	0	0	0	0
63	66	0	0	0	0	0	0	0	0
64	71	0	0	0	0	0	0	0	0
65	75	0	0	0	0	0	0	0	0
66	80	0	0	0	0	0	0	0	0
67	85	0	0	0	0	0	0	0	0
68	90	0	0	0	0	0	0	0	0
69	96	0	0	0	0	0	0	0	0
70	102	0	0	0	0	0	0	0	0

【図19】

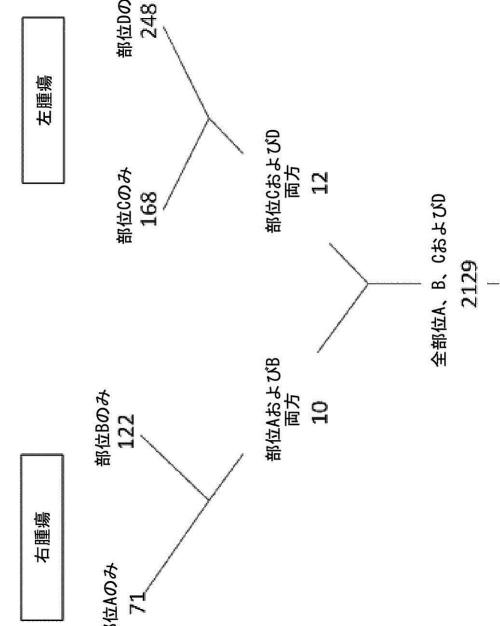


図19

【図20】

カテゴリー	治療前血漿		治療後血漿		
	突然変異を 有する 座位置数	野生型 アレルを 有する DNA断片数	突然変異 アレルを 有する DNA断片数	腫瘍由来DNAの 分画濃度	腫瘍由来DNAの 分画濃度
A	71	3,321	37	2.20%	2,149
B	122	5,633	8	0.28%	3,516
C	168	8,597	51	1.19%	5,438
D	248	10,880	26	0.48%	7,060
AB	10	423	21	9.46%	297
CD	12	569	3	1.05%	333
ABCD	2129	70,940	21,344	46.26%	61,417

図20

【図21】

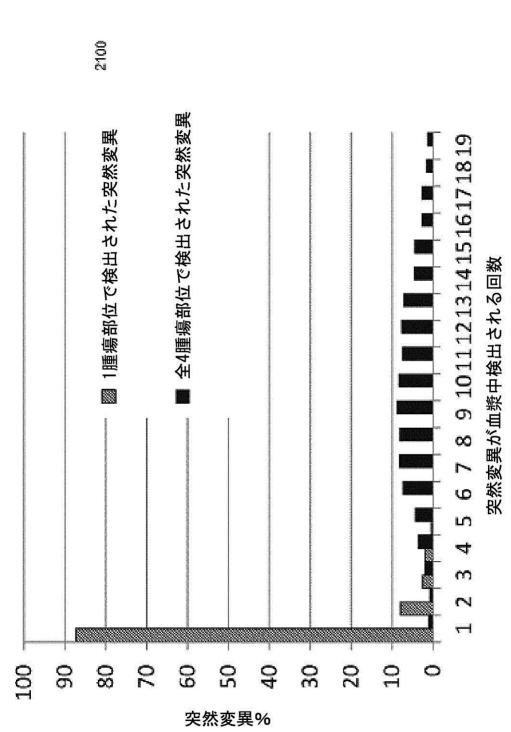
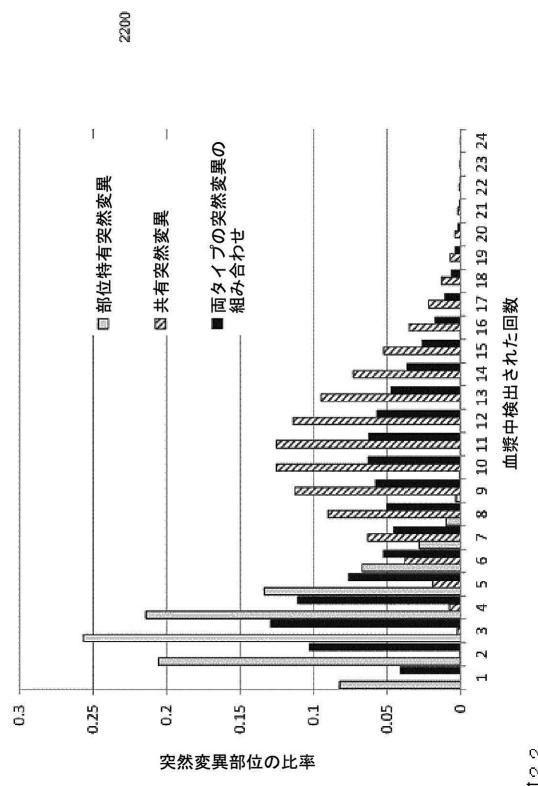


図21

【図22】



【図23】

カギ 数	部位 数	標準化検査用															
		TC1	TC2	TC3	TC4	TC5	TC6	TC7	TC8	TC9	TC10	TC11	TC12	TC13	TC14	TC15	TC16
A	71	0	0.18% (0.18%)	0	0.17% (0.17%)	0	0.17% (0.17%)	0	0.17% (0.17%)	0	0.17% (0.17%)	0	0.17% (0.17%)	0	0.17% (0.17%)	0	0.17% (0.17%)
B	122	0	0	0	0	0	0	0.08% (0.08%)	0	0.11% (0.11%)	0.10% (0.10%)	0.12% (0.12%)	0	0	0	0.11% (0.11%)	
C	168	0	0	0	0	0	0	0.08% (0.08%)	0	0.16% (0.16%)	0	0	0	0	0.07% (0.07%)	0	0
D	248	0	0	0.00% (0.00%)	0	0	0	0	0	0	0.05% (0.05%)	0	0	0	0.06% (0.06%)	0	0
AB	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CD	12	0	0	0	0	0	0	0	0	0	0.08% (0.08%)	0	0	0	0.02% (0.02%)	0	0
ABCD	212	14	0.07% (0.07%)	3	0.02% (0.02%)	2	0.09% (0.09%)	5	0.05% (0.05%)	5	0.08% (0.08%)	4	0.19% (0.19%)	3	0.02% (0.02%)	0	0.00% (0.00%)

図23

【図24】

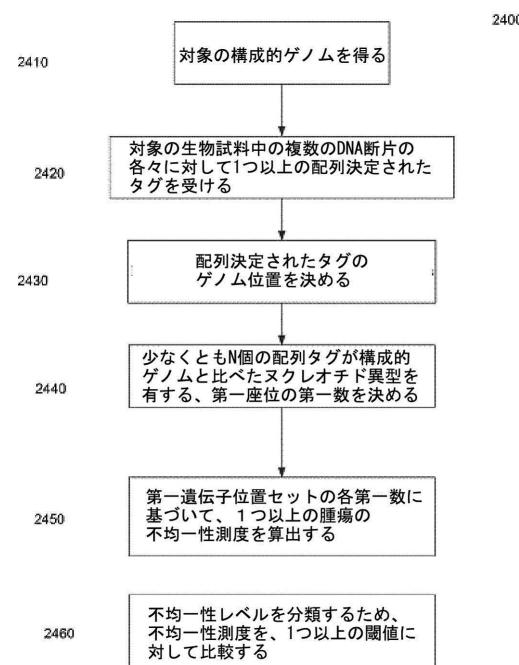


図24

【図25】

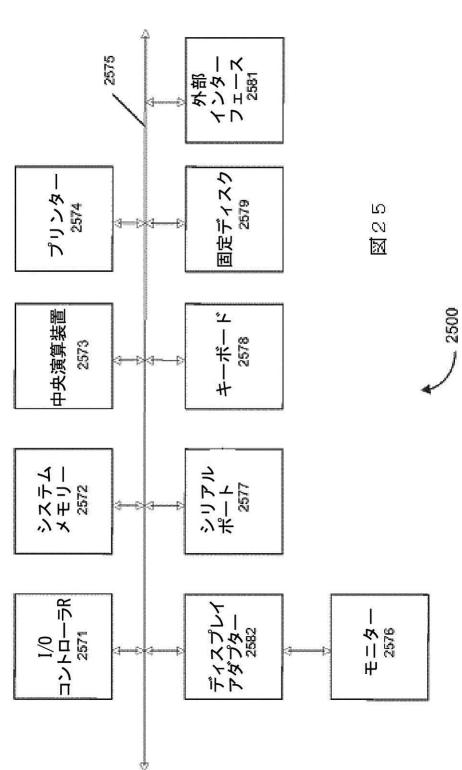


図25

2300

## フロントページの続き

(31) 優先権主張番号 61/695,795  
 (32) 優先日 平成24年8月31日(2012.8.31)  
 (33) 優先権主張国 米国(US)  
 (31) 優先権主張番号 61/711,172  
 (32) 優先日 平成24年10月8日(2012.10.8)  
 (33) 優先権主張国 米国(US)  
 (31) 優先権主張番号 13/801,748  
 (32) 優先日 平成25年3月13日(2013.3.13)  
 (33) 優先権主張国 米国(US)

(74) 代理人 100117019  
 弁理士 渡辺 陽一  
 (74) 代理人 100150810  
 弁理士 武居 良太郎  
 (74) 代理人 100134784  
 弁理士 中村 和美  
 (72) 発明者 ワイ クン ロサ チウ  
 中華人民共和国,香港特別行政区,ニューテリトリーズ,シャーティン,52 マ ロク パス,  
 ダブル ヘブン,ハウス 31  
 (72) 発明者 ロ ユク - ミン デニス  
 中華人民共和国,香港特別行政区,ホマンチン クーロン,キング タック ストリート 7,フ  
 ォース フロア  
 (72) 発明者 チヤン クワン チー  
 中華人民共和国,香港特別行政区,クーロン,メイ フー サン チエン,ブロードウェイ スト  
 リート,ブロック 34,13 / エフ,フラット エー  
 (72) 発明者 ジアーン ペイヨン  
 中華人民共和国,香港特別行政区,ニューテリトリーズ,シャーティン,ノス 62-66 シウ  
 レク ヤン ロード クン ラム コート ファースト フロア オブ ブロック ビー,フ  
 ット 7

審査官 西 賢二

(56) 参考文献 國際公開第2012/071621 (WO, A1)  
 特表2010-534068 (JP, A)  
 特表2010-534069 (JP, A)  
 國際公開第2011/090557 (WO, A1)  
 FORSHEW, T. et al., "Noninvasive identification and monitoring of cancer mutations by targeted deep sequencing of plasma DNA", Sci. Transl. Med., 2012年 5月, Vol. 4, Issues 132-136, pp. 197-208  
 TAO, Y et al., "Rapid growth of a hepatocellular carcinoma and the driving mutations revealed by cell-population genetic analysis of whole-genome data", Proc. Natl. Acad. Sci. U. S. A., 2011年, Vol. 108, pp. 12042-12047  
 GERLINGER, M. et al., "Intratumor heterogeneity and branched evolution revealed by multiregion sequencing", N. Engl. J. Med., 2012年 3月 8日, Vol. 366, pp. 883-892

(58) 調査した分野(Int.Cl., DB名)

C 12 Q 1 / 00 - 3 / 00  
 MEDLINE / EMBASE / BIOSIS (STN)

