

(19) 日本国特許庁 (JP)

(12) 特 許 公 報 (B2)

(11) 特許番号

特許第6196157号
(P6196157)

(45) 発行日 平成29年9月20日 (2017.9.20)

(24) 登録日 平成29年8月25日 (2017.8.25)

(51) Int. Cl.

F I

C 1 2 Q 1/68 (2006.01)
G O 6 F 19/18 (2011.01)C 1 2 Q 1/68 Z N A Z
G O 6 F 19/18

請求項の数 34 (全 54 頁)

(21) 出願番号 特願2013-541152 (P2013-541152)
 (86) (22) 出願日 平成23年11月30日 (2011.11.30)
 (65) 公表番号 特表2014-502155 (P2014-502155A)
 (43) 公表日 平成26年1月30日 (2014.1.30)
 (86) 国際出願番号 PCT/AU2011/001562
 (87) 国際公開番号 W02012/071621
 (87) 国際公開日 平成24年6月7日 (2012.6.7)
 審査請求日 平成26年10月20日 (2014.10.20)
 (31) 優先権主張番号 61/418,391
 (32) 優先日 平成22年11月30日 (2010.11.30)
 (33) 優先権主張国 米国 (US)
 (31) 優先権主張番号 61/529,877
 (32) 優先日 平成23年8月31日 (2011.8.31)
 (33) 優先権主張国 米国 (US)

(73) 特許権者 513134487
 ザ チャイニーズ ユニバーシティ オブ
 ホンコン
 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、ピ チウ
 ビルディング、ルーム 301、ノウレッ
 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 敬
 (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】 癌関連の遺伝子または分子異常の検出

(57) 【特許請求の範囲】

【請求項 1】

生物の生体試料における癌関連の染色体の欠失もしくは増幅の分析方法であって、
 前記生体試料が正常細胞、および潜在的に癌関連細胞に由来する核酸分子を含み、前記
 核酸分子の少なくとも一部は前記生体試料中においてセルフリーであり、以下のステップ
 :

第1染色体領域の前記生物の正常細胞の第1および第2ハプロタイプを決定すること
 、ここで、前記第1染色体領域は第1の複数の遺伝子座を含み、前記第1および第2ハ
 プロタイプは、前記第1の複数の遺伝子座のそれぞれで異型接合する；

前記生体試料中の複数の前記核酸分子のそれぞれに対して、前記生物の基準ゲノム中
 の前記核酸分子の位置を同定し、前記核酸分子のそれぞれの対立遺伝子を決定すること；

前記同定した位置および決定した対立遺伝子に基づき核酸分子の第1群を前記第1ハ
 プロタイプ由来と同定すること、ここで、前記第1群は前記第1の複数の遺伝子座のそれ
 ぞれで位置する少なくとも1つの核酸分子を含む；

前記同定した位置および決定した対立遺伝子に基づき核酸分子の第2群を前記第2ハ
 プロタイプ由来と同定すること、ここで、前記第2群は前記第1の複数の遺伝子座のそれ
 ぞれで位置する少なくとも1つの核酸分子を含む；

核酸分子の前記第1群の第1値をコンピュータシステムで算出すること、ここで、前
 記第1値は前記第1群の前記核酸分子の特性を決定する；

核酸分子の前記第2群の第2値を前記コンピュータシステムで算出すること、ここで

10

20

、前記第 2 値は前記第 2 群の前記核酸分子の特性を決定する；

前記第 1 値を前記第 2 値と比較して、前記第 1 染色体領域がいずれかの癌関連細胞中に欠失または増幅を示すかどうかの分類を評価することを補助すること、を含む、方法。

【請求項 2】

前記第 1 染色体領域が癌関連細胞中の欠失または増幅を示すとして分類され、前記癌関連細胞が悪性腫瘍および／または前悪性病巣を含む、請求項 1 に記載の方法。

【請求項 3】

前記第 1 の複数の遺伝子座の各遺伝子座が、前記第 1 の複数の遺伝子座の別の遺伝子座と少なくとも 500 塩基離れている、請求項 1 または 2 に記載の方法。

10

【請求項 4】

前記比較が前記第 1 値と前記第 2 値との間の差または割合を決定することを含み、前記分類が前記差または前記割合を少なくとも 1 つの閾値と比較することを含む、請求項 1 ～ 3 のいずれか 1 項に記載の方法。

【請求項 5】

前記少なくとも 1 つの閾値が、健常な生物由来であるかまたは欠失もしくは増幅を有しない領域由来であり、逐次確率比検定、t 検定、またはカイ二乗検定が前記少なくとも 1 つの閾値を決定するために使用される、請求項 4 に記載の方法。

【請求項 6】

前記第 1 値が前記第 1 群の前記核酸分子のサイズ分布の統計値に対応し、前記第 2 値が前記第 2 群の前記核酸分子のサイズ分布の統計値に対応する、請求項 1 ～ 5 のいずれか 1 項に記載の方法。

20

【請求項 7】

前記第 1 値が前記第 1 群の前記核酸分子の平均サイズであり、前記第 2 値が前記第 2 群の前記核酸分子の平均サイズである、請求項 6 に記載の方法。

【請求項 8】

前記第 1 値 Q_{HapI} がカットオフサイズより短い前記第 1 群中の核酸分子の部分であり、第 2 値 Q_{HapII} がカットオフサイズより短い第 2 群中の核酸分子の部分である、請求項 6 に記載の方法。

【請求項 9】

30

前記比較が前記第 1 値と前記第 2 値との間の差を決定することを含み、前記差が $Q = Q_{HapI} - Q_{HapII}$ であり、閾値を超す正の Q 値は、前記第 2 ハプロタイプが前記生物中の腫瘍組織の欠失を含むかまたは前記第 1 ハプロタイプが前記生物中の腫瘍組織の増幅を含むことを示す、請求項 8 に記載の方法。

【請求項 10】

前記比較が前記第 1 値と前記第 2 値との間の差の決定を含み、前記差が $Q = Q_{HapI} - Q_{HapII}$ であり、 Q 値がほぼゼロであることが、前記第 1 染色体領域中に欠失または増幅が存在しないことを示す、請求項 8 に記載の方法。

【請求項 11】

前記第 1 値 F_{HapI} および第 2 値 F_{HapII} がそれぞれのハプロタイプに対して $F = \frac{w}{N} \text{長}$ として定義され、式中 w 長はカットオフサイズ w 以下の長さに対応する群の核酸分子長の合計を示し、かつ N 長は N 塩基以下の長さに対応する群の核酸分子長の合計を示し、式中 N が w より大きい、請求項 6 に記載の方法。

40

【請求項 12】

前記比較が前記第 1 値と前記第 2 値との間の差を決定することを含み、前記差が $F = F_{HapI} - F_{HapII}$ であり、閾値より大きい F の正值は、前記第 2 ハプロタイプが前記生物中に腫瘍組織の欠失を含むことを示す、請求項 8 に記載の方法。

【請求項 13】

前記比較が前記第 1 値と前記第 2 値との間の差を決定することを含み、前記差が $F = F_{HapI} - F_{HapII}$ であり、閾値より大きい F の負値は、前記第 2 ハプロタイプ

50

が前記生物中に腫瘍組織の増幅を含むことを示す、請求項 8 に記載の方法。

【請求項 14】

前記第 1 群の前記第 1 値が、前記第 1 の複数の遺伝子座に位置する核酸分子数に対応し、前記第 2 群の前記第 2 値が、前記第 1 の複数の遺伝子座に位置する核酸分子数に対応する、請求項 1 ～ 5 のいずれか 1 項に記載の方法。

【請求項 15】

前記生体試料中の癌 DNA の分画濃度を決定するために、前記第 1 値と前記第 2 値との比率を算出することをさらに含む、請求項 14 に記載の方法。

【請求項 16】

前記生体試料中の癌 DNA の分画濃度を複数時点で決定すること、
前記生物中の癌レベルを診断、病期決定、予見、もしくは進行をモニタリングすることを補助するために前記複数時点で前記分画濃度を使用すること、
をさらに含む、請求項 15 に記載の方法。

【請求項 17】

複数の他の染色体領域のそれぞれに対して、
請求項 1 に記載の方法を反復すること、
欠失もしくは増幅を示す染色体領域の第 1 番号を決定すること、
前記第 1 番号を 1 つもしくは複数の閾値と比較して前記生物中の癌レベルを決定することを補助すること、をさらに含む、請求項 1 ～ 16 のいずれか 1 項に記載の方法。

【請求項 18】

欠失もしくは増幅を示すものとして同定された各染色体領域に対する欠失もしくは増幅量を決定することを補助することと、
前記量を 1 つもしくは複数の閾値と比較して前記生物中の癌レベルを決定することと、
をさらに含む、請求項 17 に記載の方法。

【請求項 19】

請求項 17 に記載の方法を複数時点で反復すること、
前記生物中の前記癌レベルを診断、病期決定、予見、もしくは進行をモニタリングすることを補助するために前記第 1 番号を前記複数時点で使用すること、
をさらに含む、請求項 17 に記載の方法。

【請求項 20】

前記生物中の前記癌レベルを診断、病期決定、予見、または進行をモニタリングすることを補助するために前記第 1 番号を前記複数時点で使用することが、前悪性状態の前記存在または進行を決定することを補助することを含む、請求項 19 に記載の方法。

【請求項 21】

前記各染色体領域が所定の長さである、請求項 17 に記載の方法。

【請求項 22】

生物の生体試料を分析する方法であって、
前記生体試料が正常細胞、および潜在的に癌関連の細胞に由来する核酸分子を含み、前記核酸分子の少なくとも一部は前記生体試料中においてセルフリーであり、以下のステップ：

各染色体領域が複数の遺伝子座を含む、前記生物の複数の重複しない染色体領域を同定することと、

前記生物の生体試料中の複数の核酸分子のそれぞれに対して、

前記生物の基準ゲノムにおける前記核酸分子の位置を同定すること；

多数の非重複染色体領域の各々に対して、核酸分子の各群を、前記同定した位置に基づき前記染色体領域由来であると同定すること、ここで、前記核酸分子の各群は、前記染色体領域の前記複数の遺伝子座のそれぞれに位置する少なくとも 1 つの核酸分子を含む；

核酸分子の前記各群の各値をコンピュータシステムで算出すること、ここで、前記各値は前記各群の核酸分子の特性を定義する；

前記各値を基準値と比較し、染色体領域が欠失または増幅を示すかどうかの分類を決

10

20

30

40

50

定すること；

欠失もしくは増幅を示すとして分類される染色体領域の量を決定すること、

前記量を閾値と比較して、前記生物が癌を有するかどうかの分類を評価することを補助すること、
を含む、方法。

【請求項 2 3】

各染色体領域が所定の長さである、請求項 2 2 に記載の方法。

【請求項 2 4】

前記複数の重複しない染色体領域が前記生物の前記ゲノムにわたる、請求項 2 2 または 2 3 に記載の方法。

10

【請求項 2 5】

染色体領域が、基準値を超える、核酸分子の同定された群の対応する各値をそれぞれ有する少なくとも特定の数の連続した染色体領域に含まれる時にのみ、前記染色体領域が欠失又は増幅を示すとして分類され、前記特定の数が 2 つ以上である、請求項 2 2 ~ 2 4 のいずれか 1 項に記載の方法。

【請求項 2 6】

前記閾値が少なくとも 1 つの値を有し、該値を前記閾値と比較することが、該値が前記閾値よりも大きいかな否かを決定することを含む、請求項 2 2 ~ 2 5 のいずれか 1 項に記載の方法。

【請求項 2 7】

20

正常細胞、および潜在的に癌関連細胞に由来する核酸分子を含む生体試料を用いて生物の染色体異常の進行を決定することを補助する方法であって、

前記核酸分子の少なくとも一部は前記生体試料中においてセルフリーであり、以下のステップ：

前記生物の基準ゲノムにおける 1 つまたは複数の重複しない染色体領域を同定すること、ここで、各染色体領域が複数の遺伝子座を含み、

複数時点のそれぞれで、前記生物の生体試料中の複数の核酸分子のそれぞれに対して、前記基準ゲノムにおける前記核酸分子の位置を同定すること；

1 つまたは複数の非重複染色体領域の各々に対して、核酸分子の各群を、前記同定した位置に基づき前記染色体領域由来であると同定すること、ここで、前記核酸分子の各群は、前記染色体領域の前記複数の遺伝子座のそれぞれに位置する少なくとも 1 つの核酸分子を含む；

30

核酸分子の前記各群の各値をコンピュータシステムで算出すること、ここで、前記各値は前記各群の核酸分子の特性を定義する；

前記各値を基準値と比較し、第 1 染色体領域が欠失または増幅を示すかどうかの分類を決定すること；

複数時点で前記 1 つまたは複数の非重複染色体領域の 1 つまたは複数の分類を使用して、生物の染色体異常の進行を決定することを補助すること、ここで、前記決定は、時間ごとに分類の数値変化を計算し、数値範囲を閾値と比較することによって行われる、
を含む、方法。

40

【請求項 2 8】

前記分類が、各時点で各染色体領域において前記各値と前記基準値との間の差を反映する数値である、請求項 2 7 に記載の方法。

【請求項 2 9】

1 つまたは複数の重複しない染色体領域が複数の染色体領域である、請求項 2 7 または 2 8 に記載の方法。

【請求項 3 0】

前記複数時点のそれぞれで、

欠失もしくは増幅を示すとして分類される染色体領域量を決定すること、

前記量を使用して前記生物の前記癌の進行をモニタリングすることを補助すること、

50

をさらに含む、請求項 29 に記載の方法。

【請求項 31】

前記基準値が、前記核酸分子が得られた時点で欠失または増幅を示さない基準染色体領域由来と同定される核酸分子群から得られる、請求項 27 ~ 30 のいずれか 1 項に記載の方法。

【請求項 32】

1 つまたは複数の重複しない染色体領域の第 1 染色体領域での前記生物の正常細胞に対する第 1 および第 2 ハプロタイプを決定することをさらに含み、ここで、前記第 1 染色体領域は、第 1 の複数の遺伝子座を含み、前記 2 ハプロタイプが前記第 1 の複数の遺伝子座のそれぞれで異型接合しており、

10

前記第 1 の染色体領域の前記各値が前記第 1 ハプロタイプ由来と同定された核酸分子の第 1 群から得られ、

前記第 1 の染色体領域の前記基準値が前記第 2 ハプロタイプ由来と同定された核酸分子の第 2 群から得られる、請求項 27 ~ 30 のいずれか 1 項に記載の方法。

【請求項 33】

請求項 1 ~ 32 のいずれか 1 項に記載の方法に対する作動を実施する処理装置を制御するための複数の命令を記憶するコンピュータ読み取り可能媒体であって、前記命令が請求項 1 ~ 32 のいずれか 1 項で定義された工程を含む、コンピュータ読み取り可能媒体。

【請求項 34】

請求項 1 ~ 32 のいずれか 1 項で定義された工程を実行するための手段を含むコンピュータシステムであって、少なくともプロセッサ及びメモリを含む、コンピュータシステム。

20

【発明の詳細な説明】

【背景技術】

【0001】

関連出願の相互参照

本願は、2010 年 11 月 30 日に出願された「癌関連の遺伝子異常の検出 (DETECTION OF GENETIC ABERRATIONS ASSOCIATED WITH CANCER)」と題された、米国仮特許出願第 61/418,391 号、および 2011 年 8 月 31 日に出願された「癌関連の遺伝子または分子異常の検出 (DETECTION OF GENETIC OR MOLECULAR ABERRATIONS ASSOCIATED WITH CANCER)」と題された、米国仮特許出願第 61/529,877 号の優先権を主張し、かつ本願はその非仮特許出願であり、これらの全内容は、すべての目的で参照することによって本明細書に組み込まれる。

30

【0002】

本願は、同一出願人が所有する、2010 年 11 月 5 日に L o らによって出願された (代理人整理番号 80015-794101/006610US) 「サイズベースのゲノム分析 (Size-Based Genomic Analysis)」と題された、米国特許出願第 12/940,992 号 (米国公開特許第 2011/0276277 号)、および 2010 年 11 月 5 日に L o らによって出願された (代理人整理番号 80015-794103/006710US) 「母体生体試料由来の胎児ゲノム分析 (Fetal Genomic Analysis From A Maternal Biological Sample)」と題された、米国特許出願第 12/940,993 号 (米国公開特許第 2011/0105353 号) に関し、これらの開示は、参照することによってそれらの全体が組み込まれる。

40

【0003】

背景

癌は、多くの人に影響する一般的な疾患である。癌は、重度の症状が現れるまで同定されないことが多い。一般的な癌タイプにおいては、癌を有する場合がある患者を同定するスクリーニング技術がある。しかしながら、かかる技術は、信頼できないか、患者を放射

50

線に曝することが多い。他の多くの癌タイプにおいては、効果的なスクリーニング技術はない。

【0004】

ヘテロ接合性の欠失 (LOH) は、肺および頭頸部癌患者の循環 DNA の特定の遺伝子座において検出されている (Chen XQ, et al. Nat Med 1996; 2: 1033-5; Nawroz H, et al. Nat Med 1996; 2: 1035-7)。しかしながら、かかる技術は、特定の遺伝子座を試験することによって検出可能である LOH は比較的少量であることによって妨げられている。デジタル PCR を使用する場合でも、これらの方法は依然として、少量の LOH を検出することができないことから問題を抱えている。さらに、かかる技術は、特定の癌タイプにおいて起こると知られている特定の遺伝子座の調査に依然として限定される。したがって、一般的に癌スクリーニングは、可能ではないかまたは効果的ではない。

10

【0005】

癌の存在のスクリーニングの他、当該技術はまた、癌患者の予後および治療効果 (例えば術後回復、または化学療法、免疫療法、もしくは標的療法) のモニタリングを提供しない。かかる技術は、高額 (例えば造影技術)、不正確、非効率、非感受性であるかまたは造影技術に使用される放射線に患者を曝し得ることが多い。

【0006】

したがって、癌のスクリーニング、予測、および患者のモニタリングの新規技術を提供することが所望される。

20

【発明の概要】

【0007】

簡単な概要

実施形態は、癌関連の遺伝子異常を決定するシステム、装置、および方法を提供する。セルフリーの DNA 断片を含む生体試料を分析して、例えば、腫瘍中の欠失および/または増幅による染色体領域の不平衡を同定する。複数の遺伝子座を伴う染色体領域を使用して、有効性および/または精度を高めることができる。次いで、かかる不平衡を使用して、患者の癌を診断もしくはスクリーニングし、ならびに癌患者を予見することができる。不平衡の重症度ならびに不平衡を示す領域数も使用することができる。加えて、患者を経時的に検査して1つもしくは複数の染色体領域のそれぞれの重症度ならびに染色体領域数を追跡してスクリーニングおよび予測を可能とし、ならびに (例えば治療後に) 進行をモニタリングすることができる。

30

【0008】

1つの実施形態によって、生物の生体試料における癌関連染色体の欠失または増幅の分析方法を提供する。生体試料は、正常細胞に由来し、潜在的に癌関連細胞に由来する核酸分子を含む。核酸分子の少なくとも一部は試料中においてセルフリーである。第1および第2ハプロタイプは、生物の正常細胞を第1染色体領域において決定する。第1染色体領域は、第1の複数の異型接合の遺伝子座を含む。試料中の複数の各核酸分子は、同定された生物の基準ゲノムにおける位置および決定されたそれぞれの対立遺伝子を有する。位置および決定された対立遺伝子は、第1ハプロタイプ由来の核酸分子の第1群および第2ハプロタイプ由来の第2群を決定するために使用される。コンピュータシステムは、第1群の第1値および第2群の第2値を算出する。各値は、核酸分子の各群の特性 (例えば群中の分子の平均サイズまたは数) を定義した。第1値を第2値と比較し、第1染色体領域がいずれかの癌関連細胞内で欠失または増幅を示すかどうかの分類を決定する。

40

【0009】

別の実施形態によって、生物の生体試料の分析方法を提供する。生体試料は、正常細胞に由来し、潜在的に癌関連細胞に由来する核酸分子を含む。核酸分子の少なくとも一部は試料中においてセルフリーである。生物の複数の重複しない染色体領域を同定する。各染色体領域は、複数の遺伝子座を含む。試料中の複数の核酸分子はそれぞれ、同定された生物の基準ゲノムにおける位置を有する。各染色体領域において、核酸分子の各群を、同定

50

した位置に基づき染色体領域由来と同定する。各群はそれぞれ、染色体領域の複数の遺伝子座のそれぞれに位置する少なくとも1つの核酸分子を含む。コンピュータシステムは、各群の各値を算出し、各値は、各群の核酸分子の特性を定義する。各値を基準値と比較し、染色体領域が欠失または増幅を示すかどうかの分類を決定する。次いで、欠失または増幅を示すとして分類される染色体領域量を決定する。

【0010】

別の実施形態によって、正常細胞に由来し、潜在的に癌関連細胞に由来する核酸分子を含む生体試料を用いて生物の染色体異常の進行を決定する方法を提供する。核酸分子の少なくとも一部は生体試料中においてセルフリーである。1つまたは複数の重複しない染色体領域を生物の基準ゲノムにおいて同定する。各染色体領域は、複数の遺伝子座を含む。生物から異なる時点で採取した試料を分析して、進行を決定する。試料において、試料中の複数の核酸分子はそれぞれ、同定された生物の基準ゲノムにおける位置を有する。各染色体領域において、核酸分子の各群を、同定した位置に基づき染色体領域由来と同定する。少なくとも1つの核酸分子を含む各群は、染色体領域の複数の遺伝子座のそれぞれに位置する。コンピュータシステムは、核酸分子の各群の各値を算出する。各値は、各群の核酸分子の特性を定義する。各値を基準値と比較して、第1染色体領域が欠失または増幅を示すかどうかの分類を決定する。次いで、生物の染色体異常の進行を決定するために複数時点での各染色体領域分類を使用する。

【0011】

本発明の他の実施形態は、本明細書に記載の方法に関連したシステム、携帯型消費者装置、およびコンピュータ読み取り可能媒体に関する。

【0012】

以下の発明を実施するための形態および添付の図面を参照して、本発明の本質および利点のより良い理解が得られ得る。

【図面の簡単な説明】

【0013】

【図1】欠失異常を示す癌細胞の染色体領域を図示する。正常細胞及び癌細胞のHap I = 配列番号1；正常細胞Hap II = 配列番号2。

【図2】増幅異常を示す癌細胞の染色体領域を図示する。正常細胞及び癌細胞のHap I = 配列番号1；正常細胞及び癌細胞のHap II = 配列番号2。

【図3】異なる癌タイプおよび関連領域およびそれらの対応する異常を例証する表300を示す。

【図4】本発明の実施形態による、血漿中で行なわれた測定値と共に異常を示さない癌細胞内の染色体領域を図示する。正常細胞及び癌細胞のLOH領域のHap I = 配列番号1；正常細胞Hap II = 配列番号2。

【図5】本発明の実施形態による、欠失領域を決定するため血漿中で行なわれた測定値と共に癌細胞内の染色体領域510の欠失を図示する。正常細胞及び癌細胞のLOH領域のHap I = 配列番号1；正常細胞Hap II = 配列番号2。

【図6】本発明の実施形態による、増幅領域を決定するため血漿中で行なわれた測定値と共に癌細胞内の染色体領域610の増幅を図示する。正常細胞及び癌細胞のHap I = 配列番号1；正常細胞及び癌細胞のHap II = 配列番号2。

【図7】本発明の実施形態による、腫瘍組織中の単対立遺伝子の増幅を示す染色体1pに位置する断片の肝細胞癌(HCC)患者の血漿DNAのRHDO分析を示す。

【図8】本発明の実施形態による、欠失を含む腫瘍が存在する場合、染色体領域の2ハプロタイプにおける断片サイズ分布変化を示す。

【図9】本発明の実施形態による、増幅を含む腫瘍が存在する場合、染色体領域の2ハプロタイプにおける断片サイズ分布変化を示す。

【図10】本発明の実施形態による、染色体領域が欠失または増幅を示すかどうかを決定する生物の生体試料のハプロタイプの分析方法を例証するフローチャートである。

【図11】本発明の実施形態による、欠失領域を決定するための血漿中で行なわれた測定

10

20

30

40

50

値と共に癌細胞内で欠失している副領域 1 1 3 0 を伴う領域 1 1 1 0 を示す。正常細胞及び癌細胞の L O H 領域の H a p I = 配列番号 1 ; 正常細胞 H a p I I = 配列番号 2。

【図 1 2】本発明の実施形態による、R H D O 分析を用いた異常位置のマッピング方法を示す。

【図 1 3】本発明の実施形態による、別方向から開始した R H D O 分類を示す。

【図 1 4】本発明の実施形態による、複数の染色体領域を用いて生物の生体試料を分析する方法 1 4 0 0 のフローチャートである。

【図 1 5】本発明の実施形態による、腫瘍由来断片の様々な数の断片および分画濃度に必要な深さを例証する表 1 5 0 0 を示す。図 1 5 は、試料中の癌由来 D N A の分画濃度の異なるパーセンテージを分析する推定分子数を提供する。

【図 1 6】本発明の実施形態による、相対的ハプロタイプ用量 (R H D O) 分析による血漿中の腫瘍由来 D N A の分画濃度を測定する原理を示す。H a p I と H a p I I は、本発明の実施形態による、非腫瘍組織における 2 ハプロタイプを示す。非腫瘍組織及び癌細胞の H a p I = 配列番号 3 ; 非腫瘍組織 H a p I I = 配列番号 4。

【図 1 7】本発明の実施形態による、核酸分子を含む生体試料を用いて、生物の染色体異常の進行を決定する方法を例証するフローチャートである。

【図 1 8 A】癌患者の染色体 4 の q 群の断片の R H D O 分析用の S P R T 曲線を示す。点は、各異型接合の遺伝子座後の累積数の比率を表す。

【図 1 8 B】は、治療後の患者の染色体 4 の q 群の断片の R H D O 分析用の S P R T 曲線を示す。

【図 1 9】H C C 中に見られる一般的な染色体異常を示す。

【図 2 0 A】標的的分析を用いて H C C と健常患者の正規化タグ数比率の結果を示す。

【図 2 0 B】H C C 患者 3 例および健常対照対象 4 例における対象の富化および大規模並列シーケンシング後のサイズ分析の結果を示す。

【図 2 1】本発明の実施形態による、血漿 D N A を計数する配列したタグからのデータを示す H C C 患者の円状プロットを示す。

【図 2 2】本発明の実施形態による、H C C を呈さない慢性 B 型肝炎ウイルス (H B V) 保因者の血漿試料における配列したタグ計数分析を示す。

【図 2 3】本発明の実施形態による、第 3 期鼻咽頭癌 (N P C) 患者の血漿試料における配列したタグ計数分析を示す。

【図 2 4】本発明の実施形態による、第 4 期 N P C 患者の血漿試料における配列したタグ計数分析を示す。

【図 2 5】本発明の実施形態による、腫瘍組織中のヘテロ接合の欠失 (L O H) を示す領域サイズに対する血漿 D N A の累積頻度プロットを示す。

【図 2 6】L O H 領域における配列した血漿 D N A のサイズに対する Q を示す。本発明の実施形態による、サイズ 1 3 0 b p で Q は 0 . 2 に達する。

【図 2 7】本発明の実施形態による、腫瘍組織中の染色体の複製を伴う領域サイズに対する血漿 D N A の累積頻度プロットを示す。

【図 2 8】本発明の実施形態による、増幅した領域における配列した血漿 D N A サイズに対する Q を示す。

【図 2 9】本発明の実施形態によるシステムおよび方法と共に使用可能な例のコンピュータシステム 9 0 0 のブロック図を示す。

【発明を実施するための形態】

【 0 0 1 4 】

定義

本明細書において使用される用語「生体試料」とは、対象（例えば、ヒト、癌患者、癌患者の疑いがあるもの、または他の生物）から採取される任意の試料を指し、1 つまたは複数の目的の核酸分子を含む。

【 0 0 1 5 】

用語「核酸」または「ポリヌクレオチド」とは、デオキシリボ核酸 (D N A) またはリ

10

20

30

40

50

が核酸 (RNA)、および一本または二本鎖のいずれかの形態であるそれらのポリマーを指す。特に限定されない限り、当該用語は、基準核酸と同程度の結合特性を有する、既知の天然ヌクレオチドアナログを含む核酸を含み、そして天然に存在するヌクレオチドに同様の方法で代謝される。他に示されない限り、特定の核酸配列はまた、その保存的に改変された改変体 (例えば、縮重コドン置換)、対立遺伝子、オルソログ、単一のヌクレオチド多型 (SNP)、コピー数変異体、および相補的配列、ならびに明確に示された配列を暗黙的に含む。具体的には、縮重コドン置換は、1つまたは複数の選択された (またはすべての) コドンの第三の位置が混合塩基および/またはデオキシノシン残基で置換されている配列を生じさせることによって達成され得る (Batzer et al., Nucleic Acid Res. 19:5081 (1991); Ohtsuka et al., J. Biol. Chem. 260:2605-2608 (1985); および Rossolini et al., Mol. Cell. Probes 8:91-98 (1994))。核酸という用語は、遺伝子、cDNA、mRNA、低分子ノンコードRNA、マイクロRNA (miRNA)、Piwi相互作用RNA、ならびに遺伝子もしくは遺伝子座によってコードされたショートヘアピンRNA (shRNA) を含むが、これらに限定されない。

【0016】

用語「遺伝子」とは、ポリペプチド鎖または転写RNA生成物を産生することに関与するDNA断片を意味する。それは、コード領域に先行する、およびコード領域に続く領域 (リーダーおよびトレイラー)、ならびに個々のコード断片 (エキソン) 間の介在配列 (イントロン) を含み得る。

【0017】

本明細書において使用される用語「臨床関連核酸配列」または「臨床関連染色体領域」 (または検査される領域/断片) とは、その潜在的不平衡が検査される、より大きなゲノム配列断片に対応するポリヌクレオチド配列のことか、またはそのより大きなゲノム配列自体を指すことができる。例としては、欠失もしくは増幅した、または潜在的に欠失もしくは増幅したゲノム断片 (単純な複製を含む)、または断片の副領域を含むより大きい領域が挙げられる。いくつかの実施形態では、複数の臨床関連核酸配列、または臨床関連核酸配列の同等の複数のマーカーを、領域の不平衡を検出するためのデータを提供するために使用することができる。例えば、染色体上の5つの非連続的配列からのデータは、可能性のある不平衡の決定のための付加的な方法において、効果的に必要とされる試料の容量を1/5へと減少させるために使用することができる。

【0018】

本明細書において使用される用語「基準核酸配列」または「基準染色体領域」とは、試験領域と比較するために定量プロファイルまたはサイズ分布を使用する核酸配列を指す。基準核酸配列の例としては、欠失または増幅を含まない染色体領域、完全ゲノム (例えば配列したタグ総数による正規化を介して)、正常であると知られている1つまたは複数の試料由来の領域 (試験される試料の同一領域である可能性がある)、または染色体領域の特定のハプロタイプが挙げられる。かかる基準核酸配列は、試料中に内因的に存在していてもよく、試料プロセスまたは分析中に外因的に添加されてもよい。いくつかの実施形態では、基準染色体領域は、疾患を伴わない健常状態を代表するサイズプロファイルを示す。さらに他の実施形態では、基準染色体領域は、疾患を伴わない健常状態を代表する定量プロファイルを示す。

【0019】

本明細書において使用される用語「に基づく」とは、「少なくとも一部に基づく」ことを意味し、そして、1つの方法の入力とその方法の出力との関係において生じる、別の値の決定において使用される1つの値 (または結果) を指す。本明細書において使用される用語「導く」とは、1つの方法の入力とその方法の出力との関係を指し、例えば誘導が公式算出であるときに起こる。

【0020】

本明細書において使用される用語「変数」とは、定量的データのセットを特徴づける数値、および/または定量的データセット間における数的関係を意味する。例えば、第1の核酸配列の第1の量と第2の核酸配列の第2の量との間における比率（または比率の関数）は変数である。

【0021】

本明細書において使用される用語「遺伝子座 (locus)」またはその複数形「遺伝子座 (loci)」は、ゲノム全体で多型を有する場合がある任意の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。

【0022】

本明細書において使用される用語「配列不平衡」または「異常」とは、臨床関連染色体領域における少なくとも1つのカットオフ値によって定義される、基準量からの任意の有意な偏差を意味する。配列不平衡は、染色体用量不平衡、対立遺伝子の不平衡、変異用量不平衡、コピー数不平衡、ハプロタイプ用量不平衡および他の類似した不平衡を含むことができる。例として、対立遺伝子の不平衡は、腫瘍が、遺伝子の欠失した1つの対立遺伝子または遺伝子の増幅した1つの対立遺伝子またはそのゲノム中の2つの対立遺伝子の差動増幅を有し、それによって、試料中の特定の遺伝子座の不平衡を生成する場合に起こる可能性がある。別の例では、患者は、腫瘍抑制遺伝子中の遺伝性変異を有する可能性がある。次いで、患者は、腫瘍抑制遺伝子の非変異性対立遺伝子が欠失している腫瘍発現に進展する可能性がある。したがって、腫瘍内、変異用量不平衡がある。腫瘍がそのDNAを患者の血漿中へ放出する場合、腫瘍DNAは、血漿中で患者の構成上DNA（正常細胞由来）と混合される。本明細書に記載の方法の使用を介して、血漿中のこのDNA混合物の変異用量不平衡を検出することができる。

【0023】

本明細書において使用される用語「ハプロタイプ」とは、同一染色体または染色体領域上に共に伝達される複数の遺伝子座での対立遺伝子の組み合わせを指す。ハプロタイプは、わずか1つの遺伝子座対もしくは染色体領域、または完全染色体を指し得る。用語「対立遺伝子」とは、同一物理的ゲノム遺伝子座での代替DNA配列を指し、これは異なる表現型の特性をもたらしてももたらさなくてもよい。（男性ヒト対象の性染色体を除く）各染色体の2コピーを伴う任意の特定の二倍体生物において、各遺伝子の遺伝子型は、同型接合体中で同じであり、異型接合体中で異なる遺伝子座に存在する対立遺伝子対を含む。生物の集団または種は、典型的には、様々な個体間の各遺伝子座で複数の対立遺伝子を含む。集団中に複数の対立遺伝子が見出されるゲノム遺伝子座は、多形部位と称する。遺伝子座の対立遺伝子多型は、集団中の対立遺伝子数（すなわち、多型度）の存在、または異型接合体率（すなわち、ヘテロ接合率）として測定可能である。本明細書において使用される用語「多型」とは、その頻度に関わらず、ヒトゲノム中の任意の個人間多型を指す。かかる多型の例としては、単一のヌクレオチド多型、単純な直列反復多型、挿入欠失多型、変異（疾患原因であり得る）およびコピー数の多型が挙げられるが、これらに限定されない。

【0024】

用語「配列したタグ」とは、核酸分子、例えば、DNA断片のすべてまたは一部から決定する配列を指す。断片の片端のみ、例えば、約30bpを配列することが多い。次いで、配列したタグは基準ゲノムに整列できる。あるいは、断片の両端は、2つの配列したタグを配列して生成することができ、これは、アラインメントのより高い精度を得ることができ、断片長も得ることができる。

【0025】

用語「ユニバーサルシーケンシング」とは、アダプターを断片末端に加え、シーケンシング用のプライマーをアダプターに結合するシーケンシングを指す。したがって、任意の断片を同一プライマーと配列でき、したがってシーケンシングはランダムとすることができる。

【0026】

用語「サイズ分布」とは、特定の群（例えば特定のハプロタイプ由来または特定の染色体領域由来の断片）に対応する分子の長さ、質量、重量、または他のサイズ測定値を示す任意の1つの値または一連の値を指す。様々な実施形態において、様々なサイズ分布を使用することができる。いくつかの実施形態では、サイズ分布は、1つの染色体断片サイズ（例えば、平均（average）、中央値、または平均（mean））を他の染色体断片と比較したランキングに関する。他の実施形態では、サイズ分布は、染色体断片の実サイズの統計値に関することができる。1つの遂行では、統計値は、染色体断片の任意の平均（average）、平均（mean）、または中央値サイズを含むことができる。別の遂行において、統計値は、カットオフ値未満の断片の全長を含むことができ、これは、全断片または少なくとも大きなカットオフ値未満の断片の全長で割ってもよい。

10

【0027】

本明細書において使用される用語「分類」とは、試料の特定の特性に関連する任意の数（1つまたは複数）または他の特性（1つまたは複数）を指す。例えば、「+」記号（または「正」という言葉）は、欠失または増幅を有するとして分類される試料を表すことができる。分類は、2進（例えば、正または負）とすることも、より多くの分類のレベル（例えば、1～10または0～1段階）を有することもできる。用語「カットオフ」および「閾値」は、操作に使用される所定数を指す。例えば、カットオフサイズは、これを超えると断片が除外されるサイズを指すことができる。閾値は、これを超えるまたはこれ未満のものに特定の分類が適用される値であってもよい。これらの用語はいずれも、これら文脈のいずれにおいても使用することができる。

20

【0028】

用語「癌レベル」は、癌の有無、癌病期、腫瘍サイズ、どれだけの染色体領域の欠失もしくは増幅が関与したか（例えば二倍または三倍）、ならびに/または癌重症度の他の測定値を指すことができる。癌レベルは、数であっても他の特性であってもよい。レベルは0である可能性がある。癌レベルはまた、欠失または増幅に関する前悪性状態または前癌状態も含む。

【0029】

詳細な説明

癌組織（腫瘍）は、異常（染色体領域の欠失または増幅等）を有する可能性がある。腫瘍は、DNA断片を体液中に放出する可能性がある。実施形態では、DNA断片を分析して染色体領域におけるDNAの正常（予想）値と比較して異常を同定することによって、腫瘍を同定することができる。

30

【0030】

欠失もしくは増幅の厳密なサイズならびに位置は変わる可能性がある。一般的に癌または特定の癌タイプにおいて特定領域が異常を示す（それによって、特定の癌の診断に至る）と知られている時点があり得る。特定領域が未知である場合、ゲノム（例えば欠失または増幅した塩基数）全体を通して分散する場合があります、かつ全体サイズが変化する、異常領域を検出するために完全ゲノムまたは大部分のゲノムの系統的分析方法を適用してもよい。染色体領域（1つまたは複数）を経時的に追跡し、異常の重症度または異常を示す領域数の変化を同定することができる。この追跡により、腫瘍のスクリーニング、予測、およびモニタリングに関する重要な情報を（例えば治療後、または再発もしくは腫瘍進行を検出するために）得ることができる。

40

【0031】

この詳細は、まず癌中の染色体異常の例から始まる。次いで、生体試料中においてセルフリーであるDNAを検出および分析することによって染色体異常の検出方法の例について論じる。1つの染色体領域における異常の検出方法を確立後、多くの染色体領域における異常の検出方法を系統的方法に使用して、患者をスクリーニング（診断）および予見する方法について記載する。この詳細は、経時的な1つまたは複数の領域における染色体異常の検査から得られた数値指標を追跡して、患者のスクリーニング、予測、およびモニタリングを提供する方法についても記載する。次いで、実施例について論じる。

50

【 0 0 3 2 】

I . 癌中の染色体異常の例

染色体異常は、一般に癌細胞内で検出される。さらに、染色体異常の特徴パターンは、選択された癌タイプに見出すことができる。例えば、染色体群 1 p、1 q、7 q、15 q、16 p、17 q および 20 q における DNA 獲得ならびに 3 p、4 q、9 p および 11 q における DNA 喪失は、一般に肝細胞癌 (HCC) に検出される。先行研究は、かかる遺伝子異常は癌患者の循環 DNA 中にも検出することができることによって立証されている。例えば、ヘテロ接合性の欠失 (LOH) は、肺および頭頸部癌患者の循環 DNA 分子の特定の遺伝子座に対して検出されている (Chen XQ, et al. Nat Med 1996; 2: 1033 - 5; Nawroz H, et al. Nat Med 1996; 2: 1035 - 7)。血漿または血清中に検出された遺伝子改変は、腫瘍組織中に見出されるものと同一である。しかしながら、腫瘍由来 DNA は総循環のセルフリーの DNA のわずかな部分のみの寄与となるため、通常、腫瘍細胞の LOH に起因する対立遺伝子の不平衡は小さい。循環 DNA 分子間の異なる遺伝子座の対立遺伝子の正確な定量化 (Chang HW, et al. J Natl Cancer Inst. 2002; 94: 1697 - 703) において、多数の治験担当医がデジタルポリメラーゼ連鎖反応 (PCR) 技術を開発している (Vogelstein B, Kinzler K W. Proc Natl Acad Sci USA. 1999; 96: 9236 - 41; Zhou W, et al. Nat Biotechnol 2001; 19: 78 - 81; Zhou W, et al. Lancet. 2002; 359: 219 - 25)。腫瘍 DNA 中の特定の遺伝子座における LOH に起因する小さな対立遺伝子の不平衡の検出においてデジタル PCR は、リアルタイム PCR または他の DNA 定量化方法よりはるかに感受性がある。しかしながら、デジタル PCR は、特定の遺伝子座における非常に小さな対立遺伝子の不平衡の同定に依然として困難を有する可能性があり、したがって、本明細書に記載の実施形態では、集合的な形式で染色体領域について分析する。

【 0 0 3 3 】

本明細書に記載の技術はまた、前悪性状態または前癌状態検出における適応も有する。かかる状態の例としては、肝硬変および頸部上皮内癌が挙げられる。前者の状態は、肝細胞癌における前悪性状態であるのに対し、後者の状態は、頸部癌における前悪性状態である。かかる前悪性状態は、それらの発生においていくつかの分子改変をすでに所有し、悪性腫瘍となることが報告されている。例えば、染色体群 1 p、4 q、13 q、18 q における LOH の存在、および 3 つを超す遺伝子座における併用欠失は、肝硬変患者における HCC 出現リスクの増大に關与する (Roncalli M et al. Hepatology 2000; 31: 846 - 50)。かかる前悪性病巣はまた、低濃度の可能性が高いが、循環中に DNA を放出もする。当該技術によって、血漿中の DNA 断片を分析して血漿中の循環前悪性 DNA の濃度 (分画濃度を含む) を測定することによって欠失または増幅を検出することができる。かかる異常は容易に検出され (例えばかかる変化のシーケンシング深さまたは数が検出される)、濃度は、末期癌状態へ進行する可能性または迅速さを予測することになる。

【 0 0 3 4 】

A . 染色体領域の欠失

図 1 は、欠失異常を示す癌細胞の染色体領域を例証する。正常細胞は、2 ハプロタイプ、Hap I と Hap II で示される。図示するように、Hap I と Hap II の両方は、複数の異型接合の遺伝子座 110 (単一のヌクレオチド多型 SNP と呼ばれる) のそれぞれで配列を有する。癌関連細胞内において、Hap II は染色体領域 120 が欠失している。例として、癌関連細胞は、腫瘍 (例えば、悪性腫瘍)、腫瘍の転移巣 (例えば局所リンパ節内、または遠位臓器内) 由来であることも、前癌または前悪性病巣 (例えば、上述) 由来であることもできる。

【 0 0 3 5 】

2 相同ハプロタイプの 1 つが欠失している癌細胞の染色体領域 120 において、異型接

合のSNP 110はすべて、対応する欠失した相同染色体上の他の対立遺伝子の欠失による同種接合を示すことになる。したがって、この染色体異常タイプは、ヘテロ接合性の欠失（LOH）と呼ばれる。領域120において、これらのSNPの欠失していない対立遺伝子は、正常組織に見出すことができる2ハプロタイプのうち1つを表す。図1に示す実施例において、LOH領域120でハプロタイプI（Hap I）は、腫瘍組織の遺伝子型を特定することによって決定することができる。他のハプロタイプ（Hap II）は、正常組織の明らかな遺伝子型と癌組織とを比較することによって決定することができる。Hap IIは、欠失した対立遺伝子をすべて結合することによって構築することができる。すなわち、癌細胞の領域120内に現れない領域120の正常細胞内の対立遺伝子のすべてが、同一ハプロタイプ、すなわちHap I上で決定される。この分析を介して、患者（例えば、肝細胞癌HCC患者）のハプロタイプは、腫瘍組織中LOHを示すすべての染色体領域において決定することができる。かかる方法は、患者が癌細胞を有する場合にのみ有用であり、領域120内のハプロタイプを決定するためにのみ役に立つが、欠失した染色体領域の良好な例証を提供する。

10

【0036】

B．染色体領域の増幅

図2は、増幅異常を示す癌細胞の染色体領域を例証する。正常細胞は、2ハプロタイプ、Hap IとHap IIで示される。図示するように、Hap IとHap IIの両方は、複数の異型接合の遺伝子座210のそれぞれの配列を有する。腫瘍細胞内のHap IIは2倍（複製）増幅した染色体領域220を有する。

20

【0037】

同様に、腫瘍組織中の単対立遺伝子増幅を伴う領域において、マイクロアレー分析等の方法によってSNP 210で増幅した対立遺伝子を検出することができる。2ハプロタイプのうち1つ（図2に示す実施例のHap II）は、染色体領域220におけるすべての増幅した対立遺伝子を結合することによって決定することができる。特定の遺伝子座で増幅した対立遺伝子は、遺伝子座の各対立遺伝子数を比較することによって決定することができる。次いで、他のハプロタイプ（Hap I）は、増幅されていない対立遺伝子を結合することによって決定することができる。かかる方法は、患者が癌細胞を有する場合にのみ有用であり、領域220内のハプロタイプを決定するためにのみ役に立つが、増幅した染色体領域の良好な例証を提供する。

30

【0038】

増幅は、2つを超す染色体を有することに起因しても、1染色体中に遺伝子を反復することに起因してもよい。1領域は直列的に複製されていても、領域の1つまたは複数のコピーを含む微小染色体であってもよい。増幅はまた、コピーされ、かつ異なる染色体または同一染色体における異なる領域に挿入された1染色体の遺伝子からもたらされてもよい。かかる挿入は増幅タイプである。

【0039】

II．染色体領域の選択

癌組織はこれらのセルフリーのDNA（および潜在的細胞DNA）の少なくとも一部に寄与することになるため、癌組織のゲノム異常は、試料（血漿および血清等）中に検出することができる。異常検出の問題は、腫瘍または癌が非常に小さいことによって癌細胞から得られるDNAが比較的少ない場合がある点である。したがって、異常を伴う循環DNA量は非常に少なく、それによって、検出が非常に困難となる。異常を検出する上で十分なDNAがゲノム中の単一遺伝子座にない場合がある。本明細書に記載の方法は、複数の遺伝子座（ハプロタイプ）を含む染色体領域でDNAを分析し、したがって、ハプロタイプ上で凝集時に1遺伝子座での小さな多型を認知可能な差に変換することによって、この困難を克服することができる。したがって、領域の複数遺伝子座を分析することによって、より高い精度を得ることができ、偽陽性および偽陰性を低減することができる。

40

【0040】

さらに、異常領域が非常に小さいことによって異常の同定が困難となる場合がある。1

50

遺伝子座または特定の遺伝子座のみを使用する場合、それらの遺伝子座にない異常は見逃されることになる。本明細書に記載の一部の方法によって、一部の領域の異常を見つけるために全領域を調査できる。分析する領域がゲノムをスパンする場合、各種長さおよび位置の異常を見つけるために以下に詳述するように完全ゲノムを分析することができる。

【0041】

これらの点を例証するため、上に示すように、領域は異常を有することができる。しかしながら、領域は、分析用に選択しなければならない。領域の長さおよび位置は、結果を変え、したがって分析に影響を及ぼす可能性がある。例えば、図1の第1領域を分析する場合、異常が検出されないことになる。第2領域を分析する場合、例えば、本明細書に記載の方法を用いて異常を検出することができる。第1領域と第2領域の両方を含む広い領域を分析する場合、あるものは広い領域の一部のみが異常を有するという困難に直面し、これは、いかなる異常を同定することもより困難にし得、ならびに異常の正確な位置および長さを同定する問題に直面する。様々な実施形態は、これらの困難の一部および/またはすべてに対処することができる。領域選択の説明は、同じ染色体領域のハプロタイプを使用するか、または2つの異なる染色体領域を使用する方法に等しく適応可能である。

10

【0042】

A. 特定の染色体領域の選択

1つの実施形態では、特定領域は癌または患者の知識に基づき選択できる。例えば、領域は、一般に多くの癌または特定の癌において異常を示すことが既知である可能性がある。領域の正確な長さおよび位置は、癌タイプまたは特定のリスク要因を有する患者に関して何が周知であるかに関する文献を参照することによって決定することができる。加えて、上記のように患者の腫瘍組織を得て分析し、異常領域を同定することができる。かかる技術には癌細胞を得ることを必要であるが（これは、診断されたばかりの患者において実用的でない場合がある）、かかる技術は、（例えば、癌組織を除去する術後、または化学療法もしくは免疫療法もしくは標的療法後、または腫瘍再発もしくは進行を検出するため）同患者において経時的にモニタリングする領域を同定するために使用することができる。

20

【0043】

あるものは、複数の特定領域を同定することができる。かかる領域のそれぞれ1つの分析を独立して使用することも、異なる領域を集合的に分析することもできる。加えて、異常の位置付けにおいてより高い精度を得るために領域を再分割してもよい。

30

【0044】

図3は、異なる癌タイプおよび関連領域、ならびにそれらの対応する異常を例証する表300を示す。列310は異なる癌タイプを列挙する。本明細書に記載の実施形態は、異常に関連する任意のタイプの癌に使用することができ、したがって、このリストは、例にすぎない。列320は、獲得（増幅）が同じ列の特定の癌と関連する領域（例えば、7pなどの広い領域、または17q25）を示す。列330は、喪失（欠失）を見出すことができる領域を示す。列340は、これらの領域と特定の癌との関係を論じる参照を列挙する。

【0045】

潜在的染色体異常を有するこれらの領域は、本明細書に記載の方法によって分析用の染色体領域（1つまたは複数）として使用することができる。癌中の改変された他のゲノム領域の例は、Cancer Genome Anatomy Projectデータベース（cgap.nci.nih.gov/Chromosomes/RecurrentAberrations）およびAtlas of Genetics and Cytogenetics in Oncology and Haematology（atlasgeneticsoncology.org/Tumors/TumorListe.html）に見ることができる。

40

【0046】

見ることができるように、同定した領域は非常に広い可能性があるのに対し、他はより

50

具体的である可能性がある。異常は、表中に同定した完全領域を含まない場合がある。したがって、異常タイプに関するかかる指標は、特定の患者において異常が実在する位置を厳密に突き止めないが、分析用の広い領域に関する概略ガイドとしてより頻繁に使用し得る。かかる広い領域は、個別にならびに集合的に分析する多くの副領域（均一サイズであり得る）を広い領域内に含み得る（詳細については本明細書に記載）。したがって、実施形態は、次に記載のように、検査する特定の癌に基づき広い領域を選択する態様と併合し得るが、より一般的な技術（例えば、副領域の検査）も使用し得る。

【 0 0 4 7 】

B．恣意的な染色体領域の選択

別の実施形態では、分析する染色体領域を恣意的に選択する。ゲノムは、例えば、1メガベース（Mb）長、または他の所定の断片長（500Kbまたは2Mb等）の領域に分離することができる。半数体ヒトゲノムには約30億の塩基があるため、領域が1Mbである場合は、ヒトゲノム中に約3,000領域があることになる。これらの領域は、次いで、後に詳細に論じるようにそれぞれ分析することができる。

【 0 0 4 8 】

かかる領域は、癌または患者のいかなる知識にも基づかず、ゲノムの分析する領域への系統的な分割に基づき決定されてもよい。1つの遂行では、染色体が複数の所定の断片長を有しない（例えば、百万塩基によって割ることができない）場合、染色体の最終領域が所定の長さ未満（例えば1Mb未満）である可能性がある。別の遂行において、染色体全長および作製する断片数（典型的には染色体間で異なる）に基づき、各染色体を等長（またはほぼ等しく、丸め誤差内で）の領域に分離することができる。かかる遂行では、各染色体断片長は異なる可能性がある。

【 0 0 4 9 】

上述のように、検査する特定の癌に基づき特定領域を同定することができるが、次いで特定の領域を、より狭い領域（例えば、より大きい領域にわたる等サイズの副領域）に再分割することができる。このような形で、異常を突き止め得る。以下の論説において、染色体領域に関する一般的な言及はいずれも、具体的に同定された領域であっても、恣意的に選択された領域であっても、または両方の組み合わせであってもよい。

【 0 0 5 0 】

III．特定のハプロタイプにおける異常の検出

この項目において、セルフフリーのDNAを含む生体試料を分析することによる、単一染色体領域における異常の検出方法を記載する。この項目のある実施形態では、単一染色体領域は、領域の複数の遺伝子座での異型接合（異なる対立遺伝子）であり、それによって、所定の遺伝子座での特定の対立遺伝子を知ることによって区別することができる2ハプロタイプを提供する。したがって、所与の核酸分子（例えば、セルフフリーのDNA断片）は、2ハプロタイプの特定の1つ由来と同定することができる。例えば、断片を配列して、染色体領域に整列された配列タグを得ることができ、次いで、対立遺伝子が属する異型接合の遺伝子座でのハプロタイプを同定することができる。2つの一般的技術タイプが、特定のハプロタイプ（Hap）での異常の決定、具体的にタグ計数およびサイズ分析について以下に記載する。

【 0 0 5 1 】

A．ハプロタイプの決定

2ハプロタイプを識別するため、染色体領域の2ハプロタイプをまず決定する。例えば、図1の正常細胞に示される2ハプロタイプHapIとHapIIを決定することができる。図1において、ハプロタイプは、異型接合しており2ハプロタイプを識別することができるようにする第1の複数の遺伝子座110を含む。この第1の複数の遺伝子座は、分析する染色体領域にわたる。異なる異型接合の遺伝子座（異性）上の対立遺伝子をまず決定して、次いで段階的に患者のハプロタイプを決定することができる。

【 0 0 5 2 】

SNP対立遺伝子のハプロタイプは、単分子分析方法によって決定することができる。

かかる方法の例は、Fan et al (Nat Biotechnol. 2011; 29: 51 - 7), Yang et al (Proc Natl Acad Sci USA. 2011; 108: 12 - 7) および Kitzman et al (Nat Biotechnol. 2011 Jan; 29: 59 - 63) によって記載されている。あるいは、個体のハプロタイプは、家族（例えば両親、兄弟姉妹、および子供）の遺伝子型の分析によって決定することができる。例として、Roach et al (Am J Hum Genet. 2011; 89(3): 382 - 97) および Lo et al (Sci Transl Med. 2010; 2: 61ra91) によって記載されている方法が挙げられる。さらに別の実施形態では、個体のハプロタイプは、腫瘍組織およびゲノムDNAの遺伝子型結果を比較することによって決定することができる。これらの対象の遺伝子型は、tを使用するなどのマイクロアレー分析によって実施することができる。

10

【0053】

ハプロタイプは、当業者に周知である他の方法によっても構築できる。かかる方法の例としては、単分子分析ベースのもの（デジタルPCR等）（Ding C and Cantor CR. Proc Natl Acad Sci USA 2003; 100: 7449 - 7453; Ruano G et al. Proc Natl Acad Sci USA 1990; 87: 6296 - 6300）、染色体選別もしくは分離（Yang H et al. Proc Natl Acad Sci USA 2011; 108: 12 - 17; Fan HC et al. Nat Biotechnol 2011; 29: 51 - 57）、精子ハプロタイプ（Lien S et al. Curr Protoc Hum Genet 2002; Chapter 1: Unit 1.6）ならびに造影技術（Xiao M et al. Hum Mutat 2007; 28: 913 - 921）が挙げられる。他の方法としては、対立遺伝子特異性PCRベースのもの（Michalatos - Beloin S et al. Nucleic Acids Res 1996; 24: 4841 - 4843; Lo YMD et al. Nucleic Acids Res 1993; 21: 3561 - 3567）、クローニングおよび制限酵素消化（Smirnova AS et al. Immunogenetics 2007; 59: 93 - 8）等が挙げられる。さらに他の方法は、対象のハプロタイプを統計的評価から推定する集団中のハプロタイプブロック不平衡構造の分布および結合に基づく（Clark AG. Mol Biol Evol 1990; 7: 111 - 22; 10: 13 - 9; Salem RM et al. Hum Genomics 2005; 2: 39 - 66）。

20

30

【0054】

LOH領域のハプロタイプの別の決定方法は、腫瘍組織が入手可能である場合、対象の正常組織および腫瘍組織の遺伝子型を特定することによる。LOHの存在下、腫瘍細胞の非常に高い分画濃度を伴う腫瘍組織は、LOHを示す領域内のすべてのSNP遺伝子座において明らかな同型接合性を示すことになる。これらのSNP遺伝子座の遺伝子型は1ハプロタイプ（図1のLOH領域のHap I）を含むことになる。他方、正常組織は、対象がLOH領域内のSNP遺伝子座において異型接合していることを示すことになる。正常組織に存在するが腫瘍組織には存在しない対立遺伝子は、もう一方のハプロタイプ（図1のLOH領域のHap II）を含むことになる。

40

B. 相対的ハプロタイプ用量（RHDO）分析

【0055】

上述のように、染色体領域の1ハプロタイプの増幅または欠失を伴う染色体異常は、腫瘍組織中の染色体領域の2ハプロタイプの用量不平衡に至ることになる。腫瘍増殖患者の血漿中の循環DNAの一部は、腫瘍細胞に由来する。癌患者の血漿中の腫瘍由来DNAの存在に起因し、かかる不平衡はそれらの血漿にも存在することになる。2ハプロタイプの用量不平衡は、各ハプロタイプ由来の分子数計数を介して検出することができる。

【0056】

LOHが腫瘍組織中に観察された染色体領域（例えば図1の領域120）において、H

50

a p I は、腫瘍組織からの H a p I I の寄与の欠如のため、H a p I I と比較して循環 DNA 分子（断片）間で過剰発現する。腫瘍組織中にコピー数増幅が観察される染色体領域において、腫瘍組織からさらなる H a p I I 用量が放出されるために、H a p I I の単対立遺伝子増幅による影響を受ける領域において H a p I I は H a p I と比較して過剰発現する。過剰発現か不足発現かを決定するため、試料中のある DNA 断片は、H a p I または H a p I I 由来と判定され、これは様々な方法、例えばユニバーサルシーケンシングを実施すること、ならびにデジタル PCR および配列特異性プローブを整列するかもしくは使用することによって行なうことができる。

【 0 0 5 7 】

癌患者の血漿（または他の生体試料）由来の複数の DNA 断片をシーケンシングして配列したタグを生成後、2 ハプロタイプ上の対立遺伝子に対応する配列したタグを同定して計数することができる。次いで、2 ハプロタイプのそれぞれに対応する配列したタグ数を比較して、2 ハプロタイプが血漿中で等しく発現しているかどうかを決定することができる。1 つの実施形態では、血漿中の 2 ハプロタイプ発現が有意に異なるかどうかを決定するために逐次確率比検定（S P R T）を使用することができる。統計的有意差によって、分析する染色体領域での染色体異常の存在が示唆される。さらに、血漿中の 2 ハプロタイプの定量差は、下記のように血漿中の腫瘍由来 DNA の分画濃度推定に使用することができる。

【 0 0 5 8 】

本願に記載の DNA 断片の同一性（例えばヒトゲノムにおけるその位置）決定のための診断アプローチは、検出基盤として大規模並列シーケンシング使用に限定されない。これらの診断アプローチは、例えば、マイクロフルイディクスデジタル PCR（例えば、F l u i d i g m デジタルアレイシステム、マイクロ液滴デジタル PCR システム（例えば R a i n D a n c e、および Q u a n t a L i f e 製）、B E A M i n g システム（すなわちビーズ、乳剤 PCR、増幅、および磁気）（D i e h l e t a l . P r o c N a t l A c a d S c i U S A 2 0 0 5 ; 1 0 2 : 1 6 3 6 8 - 1 6 3 7 3）、リアルタイム PCR、質量分析ベースのシステム（例えば S e q u e n o m M a s s A r r a y システム）および多重結紮依存プローブ増幅（M L P A）分析にも適用できるが、これらに限定されない。

【 0 0 5 9 】

正常領域

図 4 は、本発明の実施形態による、血漿中で行なわれた測定値と共に異常を示さない癌細胞内の染色体領域を例証する。染色体領域 4 1 0 は、任意の方法によって、例えば、検査する特定の癌に基づき、またはゲノムの広い領域にわたる所定の断片を使用する一般的スクリーニングに基づき選択されてもよい。2 ハプロタイプを識別するため、2 ハプロタイプをまず決定する。図 4 は、染色体領域 4 1 0 における正常細胞の 2 ハプロタイプ（H a p I と H a p I I）を示す。ハプロタイプは、第 1 の複数の遺伝子座 4 2 0 を含む。この第 1 の複数の遺伝子座 4 2 0 は、分析する染色体領域 4 1 0 にわたる。図示するように、これらの遺伝子座は、正常細胞内で異型接合している。癌細胞の 2 ハプロタイプも示される。癌細胞内において、欠失または増幅する領域はない。

【 0 0 6 0 】

図 4 はまた、各遺伝子座 4 2 0 の各ハプロタイプ上の対立遺伝子数を示す。染色体領域 4 1 0 のある副領域において総累積も得られる。対立遺伝子数は、それぞれの特定の遺伝子座で特定のハプロタイプに対応する DNA 断片数に対応する。例えば、第 1 遺伝子座 4 2 1 を含み、対立遺伝子 A を有する DNA 断片は H a p I に向かって計数されることになる。そして、対立遺伝子 T を有する DNA 断片は、H a p I I に向かって計数されることになる。断片の整列位置（すなわち特定の遺伝子座を含むかどうか）および含有対立遺伝子の決定は、本明細書に記載のように様々な方法で決定することができる。2 ハプロタイプ上の数の比率は、統計的有意差が存在するかどうかを決定するために使用し得る。この比率は、本明細書においてオッズ比と呼ばれる。2 値間の差も使用されてもよく、この差

10

20

30

40

50

は、断片の総数によって正規化し得る。比率および差（およびその関数）は、異常の有無の分類を決定する閾値と比較する変数例である。

【0061】

RHDO分析は、同一ハプロタイプ上のすべての対立遺伝子（例えば累積数）を使用して、例えば、上述のLoの特許出願第12/940,992号および同第12/940,993号に記載の母体血漿中に行なうことができる、血漿中の2ハプロタイプの何らかの不平衡が存在するかどうかを決定することができる。この方法は、何らかの不平衡が存在するかどうかを決定するために使用されるDNA分子数を有意に増大でき、それゆえ、癌または前悪性状態の不在下において対立遺伝子数の確率的な分布由来の癌の存在に起因する不平衡を識別するより良い統計的検出力をもたらす。複数のSNP遺伝子座の別々の分析と対照的に、RHDOアプローチは、同一染色体上に位置する対立遺伝子を共に分析することができるように、2つの染色体（ハプロタイプ情報）上の対立遺伝子の相対位置を使用することができる。ハプロタイプ情報の不在下、異なるSNP遺伝子座の対立遺伝子数は、血漿中のハプロタイプが過剰発現であるか不足発現であるかどうかを統計的に決定するために共に添加することができない。対立遺伝子数の定量化は、大規模並列シーケンシング（例えば、合成システムによるイルミナシーケンシング、Life Technologiesによる結紮技術（SOLiD）によるシーケンシング、Ion Torrent、およびLife TechnologiesによるIon Torrent シーケンシングシステム、ナノポアシーケンシング（nanoporetech.com）、および454シーケンシング技術（Roche）、デジタルPCR（例えばマイクロフルイディクスデジタルPCR（例えば、Fluidigm（fluidigm.com））またはBEAMing（ビーズ、乳剤PCR、増幅、磁気（inostics.com））または液滴PCR（例えばQuantaLife（quantalife.com）およびRainDance（raindancetechologies.com）製）およびリアルタイムPCRを用いて実施することができるが、これらに限定されない。技術の他の遂行において、溶液中捕獲を用いて（例えば、Agilent SureSelectシステム、Illumina TruSeq Custom Enrichment Kit（illumina.com/applications/sequencing/targeted_resequencing.ilmn）、またはMyGenostics GenCap Custom Enrichmentシステム（mygenostics.com/））を用いて）またはアレーベースの捕獲（例えば、Roche NimbleGenシステムを用いて）富化された標的シーケンシングを使用することができる。

【0062】

図4に示す実施例において、わずかな対立遺伝子の不平衡が第1の2つのSNP遺伝子座において観察された（第1SNPにおいて24対26および第2SNPにおいて18対20）。しかしながら、対立遺伝子数は、実際の対立遺伝子の不平衡が存在するかどうかを決定する上で統計的に十分ではない。したがって、染色体領域410において2ハプロタイプ間に対立遺伝子の不平衡が存在しないと統計的に結論する上で2ハプロタイプの対立遺伝子の累積数が十分になるまで、同一ハプロタイプ上の対立遺伝子数を共に添加する（本実施例における第5SNP）。統計的有意な分類に達した後、累積数を初期化する（本実施例における第6SNP）。次いで、領域410の特定の副領域において2ハプロタイプ間に対立遺伝子の不平衡が存在しないと統計的に結論する上で再び2ハプロタイプの対立遺伝子の累積数が十分になるまで累積数を決定する。総累積数は、完全領域にも使用することができるが、先の方法によって、完全領域410に対向して異常の位置を決定する上で精度を高めるために提供する異なる副領域（すなわち副領域）を検査することができる。実際の対立遺伝子の不平衡の有無を決定するための統計的検査の例としては、逐次確率比検定（Zhou W, et al. Nat Biotechnol 2001; 19:78-81; Zhou W, et al. Lancet. 2002; 359:219-25）、t検定、およびカイ二乗検定が挙げられるが、これらに限定されない。

10

20

30

40

50

【0063】

欠失の検出

図5は、本発明の実施形態による、欠失領域を決定するため血漿中で行なわれた測定値と共に癌細胞内の染色体領域510の欠失を例証する。図5は、染色体領域510における正常細胞の2ハプロタイプ(Hap IとHap II)を示す。ハプロタイプは、分析する染色体領域510にわたる第1の複数の異型接合の遺伝子座520を含む。癌細胞の2ハプロタイプも示される。癌細胞内において、領域510はHap IIにおいて欠失している。図4、図5には、さらに各遺伝子座520の対立遺伝子数も示す。総累積はまた、染色体領域510内のある副領域に対しても維持される。

【0064】

腫瘍組織は典型的には腫瘍細胞と非腫瘍細胞の混合物を含むため、LOHは領域510内の遺伝子座の2つの対立遺伝子量の比率の非対称性によって明らかにされ得る。かかる状況において、領域510内のハプロタイプHap II欠失は、正常組織上の対応する遺伝子座と比較してDNA断片量の相対的低下を示す遺伝子座520の組み合わせによって決定することができる。より頻繁に現れる断片とのハプロタイプはHap Iであり、これは腫瘍細胞内に保持される。ある実施形態では、欠失したハプロタイプおよび保持されたハプロタイプがより容易に決定されるように、腫瘍試料中の腫瘍細胞率を富化する手順を実施することが所望され得る。かかる手順の1例は、(手動またはレーザー捕獲技術のいずれかによる)顕微解剖である。

【0065】

理論的に、Hap I上の各対立遺伝子は腫瘍組織中のLOHを示す染色体領域における循環DNA中で過剰発現し、対立遺伝子の不平衡度は血漿中の腫瘍DNAの分画濃度に依存する。しかしながら、同時に、任意の循環DNA試料中の2つの対立遺伝子の相対的豊富さも、ポアソン分布によっても支配されることになる。統計的分析を実施して、観察された対立遺伝子の不平衡が癌組織中LOHの存在によるかまたは偶然によるかを判定することができる。癌中LOHに関する実際の対立遺伝子の不平衡の検出力は、分析する循環DNA分子数、腫瘍DNAの分画濃度に依存する。腫瘍DNAのより高い分画濃度および分析するより多数の分子によって、実際の対立遺伝子の不平衡を検出するための感受性および特異性がより高まる。

【0066】

図5に示す実施例において、わずかな対立遺伝子の不平衡が第1の2つのSNP遺伝子座において観察された(第1SNPにおいて24対22、および第2SNPにおいて18対15)。しかしながら、対立遺伝子数は、実際の対立遺伝子の不平衡が存在するかどうかを決定する上で統計的に十分ではない。したがって、領域510において2ハプロタイプ間に対立遺伝子の不平衡が存在しないと統計的に結論する上で2ハプロタイプの対立遺伝子の累積数が十分になるまで、同一ハプロタイプ上の対立遺伝子数を共に添加する(本実施例における第5SNP)。いくつかの実施形態では、不平衡のみ既知であり、具体的なタイプ(欠失または増幅)は決定されない。次いで、領域510の特定の副領域において2ハプロタイプ間に対立遺伝子の不平衡が存在しないと統計的に結論する上で再び2ハプロタイプの対立遺伝子の累積数が十分になるまで累積数を決定する。総累積数は、完全領域にも使用することができ、ならびに本明細書に記載の任意の方法で行ない得る。

【0067】

染色体領域の増幅の検出

図6は、本発明の実施形態による、増幅領域を決定するため血漿中で行なわれた測定と共に癌細胞内の染色体領域610の増幅を例証する。LOHに加えて、染色体領域の増幅も癌組織中に頻繁に観察される。図6に示す実施例において、染色体領域610内のHap IIは、癌細胞内の3つのコピーに増幅される。図示するように、領域610は、先行の図に示すより長い領域に対向する6つの異型接合の遺伝子座のみを含む。増幅は、過剰発現が統計的に有意であると決定される第6遺伝子座において統計的に有意として検出される。いくつかの実施形態では、不平衡のみ既知であり、具体的なタイプ(欠失または増幅)

10

20

30

40

50

は決定されない。他の実施形態では、癌細胞を得て分析し得る。かかる分析によって、不平衡が欠失による（癌細胞が欠失領域において同種接合している）か、または増幅による（癌細胞が増幅領域において異型接合している）かに関する情報を得ることができる。他の実施では、欠失または増幅が存在するかどうかを、I V項の方法を用いて決定し、領域全体（すなわち個別にハプロタイプではない）を分析することができる。領域が過剰発現の場合、異常は増幅であり；および領域が不足発現である場合、異常は欠失である。領域620もまた分析して、累積数は不平衡が存在しないことを確認する。

【0068】

血漿RHD O分析のSPRT分析

異型接合の遺伝子座を有する任意の染色体領域において、RHD O分析は、血漿中に2ハプロタイプの何らかの用量不平衡があるかどうかを決定するために使用することができる。これらの領域において、血漿中のハプロタイプ用量不平衡の存在は、血漿試料中の腫瘍由来DNAの存在を示唆する。1つの実施形態では、SPRT分析は、Hap IおよびHap IIの配列したリード数差が統計的有意であるかどうかを判定するために使用することができる。このSPRT分析例において、我々はまず2ハプロタイプのそれぞれに由来する配列したリード数を決定する。次いで、我々は潜在的に過剰発現したハプロタイプ（例えば、1ハプロタイプのリード数を他のハプロタイプのリード数で割った部分）によって寄与された、配列したリードの比例的量を示す変数（例えば部分）を決定することができる。潜在的に過剰発現したハプロタイプは、LOHのシナリオにおいて欠失していないハプロタイプ、および染色体領域の単対立遺伝子増幅のシナリオにおいて増幅したハプロタイプである。次いで、この部分は、帰無仮説、すなわちハプロタイプ用量不平衡の不在、および代替的な仮説、すなわちハプロタイプ用量不平衡の存在に基づき構築される2つの閾値（上限および下限の閾値）と比較する。部分が上限の閾値を超す場合、血漿中の2ハプロタイプの統計的有意な不平衡の存在を示す。部分が下限の閾値未満の場合、2ハプロタイプの統計的有意な不平衡の不在を示す。部分が上限の閾値と下限の閾値との間の場合、結論を下す上で十分な統計的検出力がないことを示す。分析する領域の異型接合の遺伝子座数の漸増は、順調なSPRT分類を作製できるまで実施され得る。

【0069】

SPRTの上限の境界および下限の境界を算出する方程式は、

上限の閾値 = $[(\ln 8) / N - \ln] / \ln$; 下限の閾値 = $[(\ln 1 / 8) / N - \ln] / \ln$ であり、式中、

【0070】

【数1】

$$\delta = (1 - \theta_1) / (1 - \theta_2)$$

【0071】

および

【0072】

【数2】

$$\gamma = \frac{\theta_1(1 - \theta_2)}{\theta_2(1 - \theta_1)}$$

【0073】

であり、 θ_1 は対立遺伝子の不平衡が血漿に存在する場合に潜在的に過剰発現したハプロタイプからの配列したタグの推定部分であり、 θ_2 は対立遺伝子の不平衡が存在しない場合の任意の2ハプロタイプの推定部分、すなわち0.5であり、NはHap IとHap I

I の配列したタグ総数であり、 \ln は自然対数、すなわち \log_e を表す数学記号である。 f_1 は、血漿試料に存在するものと予期される（または存在することが既知である）腫瘍由来 DNA の分画濃度（F）に依存する。

【0074】

LOH のシナリオにおいて、 $f_1 = 1 / (2 - F)$ である。単対立遺伝子増幅のシナリオにおいて、 $f_1 = (1 + zF) / (2 + zF)$ である（式中 z は、腫瘍中で増幅する染色体領域の余剰コピー数を示す）。例えば、1つの染色体が複製される場合、特定の染色体の1余剰コピーがあることになる。次いで、 z は1と等しい。

【0075】

図7は、本発明の実施形態による、腫瘍組織中の単対立遺伝子の増幅を示す染色体1pに位置する断片に対するHCC患者の血漿DNAのRHDO分析を示す。緑色三角は患者データを表す。分析するSNP数が増大するにつれ、配列したリード総数は増大した。腫瘍中の増幅したハプロタイプからの配列したリードの総数の部分は、分析する配列したリード総数が増加するにつれて変わり、最終的に上限の閾値を超す値に達した。これは、有意なハプロタイプ用量不平衡を示し、それゆえ、血漿中にこの癌関連染色体異常が存在することを支持する。

【0076】

腫瘍組織中で増幅および欠失を示すHCC患者のすべての染色体領域において、SPRTを用いたRHDO分析を実施した。結果は、LOHを有することが既知である922断片および増幅を有することが既知である105断片において以下のとおりである。LOHについて、922断片をSPRTに分類し、血漿中にハプロタイプ用量不平衡を有するものとして断片921を正確に同定し、精度99.99%を得た。単対立遺伝子増幅において、105断片をSPRTに分類し、血漿中にハプロタイプ用量不平衡を有するものとして断片105を正確に同定し、精度100%を得た。

【0077】

C. 相対的ハプロタイプサイズ分析

2ハプロタイプに整列された断片用量の代替的な計数法として、それぞれのハプロタイプの断片サイズを使用することができる。例えば、特定の染色体領域において、あるハプロタイプ由来のDNA断片サイズを他のハプロタイプのDNA断片サイズと比較することができる。あるものは、領域の第1ハプロタイプの異型接合の遺伝子座で任意の対立遺伝子に対応するDNA断片サイズ分布を分析して、それを第2ハプロタイプの異型接合の遺伝子座で任意の対立遺伝子に対応するDNA断片サイズ分布と比較することができる。サイズ分布の統計的有意差は、計数でできる方法と同様に、異常を同定するために使用することができる。

【0078】

総（すなわち腫瘍＋非腫瘍）血漿DNAのサイズ分布は、癌患者中で増大することが報告されている（Wang BG, et al Cancer Res. 2003; 63: 3966-8）。しかしながら、あるものが（総（すなわち腫瘍＋非腫瘍）DNA量の代わりに）特に腫瘍由来DNAについて研究している場合、腫瘍由来DNA分子のサイズ分布は、非腫瘍細胞に由来する分子でより短いことが観察されている（Diehl et al. Proc Natl Acad Sci USA. 2005; 102: 16368-73）。したがって、循環DNAのサイズ分布は、癌関連染色体異常が存在するかどうかを決定するために使用することができる。サイズ分析の原理を図8に示す。

【0079】

図8は、本発明の実施形態による、欠失を含む腫瘍が存在する場合の染色体領域の2ハプロタイプにおける断片サイズ分布変化を示す。図8に例証するように、T対立遺伝子は、腫瘍組織中で欠失している。その結果、腫瘍組織は、血漿中へA対立遺伝子の短い分子のみ放出する。腫瘍由来の短いDNA分子は、血漿中のA対立遺伝子のためのサイズ分布の全体的な短縮に至り、それゆえ、血漿中のT対立遺伝子と比較してA対立遺伝子のより短いサイズ分布に至る。先の項目で論じたように、同一ハプロタイプ上に位置する対立遺

10

20

30

40

50

伝子はすべて共に分析することができる。換言すれば、あるハプロタイプ上に位置する対立遺伝子を保因するDNA分子のサイズ分布は、他のハプロタイプ上の対立遺伝子を保因するDNA分子のサイズ分布と比較することができる。腫瘍組織中の欠失ハプロタイプは、血漿中のより長いサイズ分布を示す。

【0080】

サイズ分析はまた、癌関連染色体領域の増幅を検出するためにも適用できる。図9は、本発明の実施形態による、増幅を含む腫瘍が存在する場合、染色体領域の2ハプロタイプにおける断片サイズ分布変化を示す。図9に示す実施例において、対立遺伝子Tを保因する染色体領域が腫瘍中に複製される。その結果、T対立遺伝子を保因する増量した短いDNA分子は血漿中に放出され、それゆえT対立遺伝子のサイズ分布は、結果的にA対立遺伝子のサイズ分布と比較して全体的に短縮される。サイズ分析を適用して、同一ハプロタイプ上に位置する対立遺伝子をすべてまとめることができる。換言すれば、腫瘍組織中で増幅したハプロタイプのサイズ分布は、腫瘍中で増幅しないハプロタイプのサイズ分布より短いことになる。

【0081】

循環DNAのサイズ分布短縮の検出

2ハプロタイプ、すなわちHapIおよびHapIIから生じるDNA断片サイズは、対末端大規模並列シーケンシングによって決定することができるが、これに限定されない。DNA断片の端のシーケンシング後、配列したリード（タグ）を基準ヒトゲノムに整列することができる。配列したDNA分子サイズは、各端の最外ヌクレオチド配位から推定することができる。分子の配列したタグは、配列したDNA断片がHapIまたはHapIIから生じたかどうかを決定するために使用することができる。例えば、配列したタグの1つは、分析する染色体領域に異型接合の遺伝子座を含み得る。

【0082】

したがって、配列した分子のそれぞれにおいて、我々はサイズと、HapIまたはHapIIから生じているかどうかとの両方を決定することができる。各ハプロタイプに整列された断片サイズに基づき、コンピュータシステムは、HapIおよびHapIIの両サイズ分布プロファイル（例えば平均断片サイズ）を算出できる。HapIとHapII由来のDNA断片サイズ分布は、異常を同定する上で十分にサイズ分布が異なる時点を決断するために適切な統計的分析を用いて比較することができる。対末端大規模並列シーケンシングとは別に、DNA断片サイズを決定するために、全DNA断片のシーケンシング、質量分析、および標準と観察されたDNA分子長を観察して比較する視覚的な方法が挙げられるが、これらに限定されない他の方法を使用することができる。

【0083】

次に、我々は腫瘍の遺伝子異常に関連する短い循環DNAの2つの検出方法例を紹介する。これら2方法は、DNA断片の2集団のサイズ分布差の定量的測定値を得ることを目的とする。DNA断片の2集団は、HapIとHapIIに対応するDNA分子を指す。

【0084】

短いDNA断片の部分差

1つの遂行では、短いDNA断片の部分を使用する。あるものは、カットオフサイズ（w）を短いDNA分子を定義するものとして設定する。カットオフサイズは変わることができ、異なる診断目的に適合するように選択することができる。コンピュータシステムは、サイズカットオフ以下の分子数を決定することができる。次いで、DNA断片の部分（Q）は、短いDNA数をDNA断片総数で割ることによって算出できる。Q値は、DNA分子集団のサイズ分布による影響を受ける。短い全体のサイズ分布は、DNA分子の高い割合が短い断片であり、したがって、高いQ値が得られることを示す。

【0085】

次いで、HapIとHapII間の短いDNA断片の部分の差を使用することができる。HapI由来とHapII由来とのDNA断片サイズ分布差は、HapIとHapIIとにおける短い断片の部分の差（Q）によって反映することができる。 $Q = Q_{HapI}$

I - $Q_{Hap I I}$ (式中、 $Q_{Hap I}$ は $Hap I$ DNA 断片の短い断片の部分であり、かつ $Q_{Hap I I}$ は、 $Hap I I$ DNA 断片の短い断片の部分である)。 $Q_{Hap I}$ および $Q_{Hap I I}$ は、各ハプロタイプ由来の断片サイズ分布の 2 群の統計値例である。

【0086】

先の項目における例証として、腫瘍組織中 $Hap I I$ が欠失している場合、 $Hap I$ DNA 断片サイズ分布は、 $Hap I I$ DNA 断片より短いことになる。その結果、正の Q 値が観察されることになる。正の Q 値は、閾値と比較して、欠失が存在しているとみなす上で十分に Q が大きいかどうかを決定することができる。 $Hap I$ の増幅はまた、正の Q 値も示す。腫瘍組織中に $Hap I I$ の複製がある場合、 $Hap I I$ DNA 断片サイズ分布は、 $Hap I$ DNA 断片より短いことになる。それゆえ、 Q 値は負となることになる。染色体異常の不在下、血漿 / 血清中の $Hap I$ と $Hap I I$ DNA 断片サイズ分布は類似する。それゆえ、 Q 値はほぼ 0 である。

【0087】

患者の Q は、正常個体と比較して値が正常であるかどうかを決定することができる。さらにまたはあるいは、患者の Q 値は、類似した癌の患者から得られた値と比較して値が異常であるかどうかを決定することができる。かかる比較は、本明細書に記載の閾値との比較 (1 つまたは複数) に関与する可能性がある。疾患モニタリングの状況において、

Q 値は、経時的に連続してモニタリングできる。 Q 値の変化は、血漿 / 血清中の腫瘍 DNA の増大した分画濃度を示し得る。この技術の選択された遂行では、腫瘍 DNA の分画濃度は、腫瘍病期、疾患の前兆および進行と相関する可能性がある。異なる時点での測定値を用いたかかる遂行について、後に詳細に論じる。

【0088】

短い DNA 断片に寄与する全長の部分差

この遂行では、短い DNA 断片によって寄与される全長の部分を使用する。コンピュータシステムは、(例えば所与の領域の特定のハプロタイプ由来の断片または所与の領域由来のみである) 試料中 DNA 断片群の全長を決定することができる。これ未満は「短い断片」として定義される DNA 断片カットオフサイズ (w) を選択することができる。カットオフサイズは変わる可能性があり、異なる診断目的に適合するように選択することができる。次いで、コンピュータシステムは、カットオフサイズ以下の DNA 断片のランダムな選択の長さを積算することによって短い DNA 断片の全長を決定することができる。短い DNA 断片によって寄与された全長の部分は、次いで、 $F = \frac{w \text{ 長さ}}{N \text{ 長さ}}$ (式中、 w 長さは、長さ w (bp) 以下の DNA 断片の合計した長さを示し、かつ N 長は、所定の長さ N 以下の DNA 断片の合計した長さを示す) に従い算出できる。1 つの実施形態では、 N は 600 塩基である。しかしながら、「全長」を算出するために、他のサイズ限界、例えば 150 塩基、180 塩基、200 塩基、250 塩基、300 塩基、400 塩基、500 塩基および 700 塩基を使用することができる。

【0089】

イルミナゲノム分析器システムは、600 塩基より長い DNA 断片の増幅およびシーケンシングに効果的ではないため、600 塩基値以下を選択し得る。さらに、分析を 600 塩基未満の DNA 断片に限定することはまた、ゲノムの構造的多型から生じるバイアスを回避することもできる。サイズが基準ゲノムに DNA 断片端をマッピングすることによって生物情報工学的に推定される場合、構造的多型、例えば、再配列 (Kidd et al., Nature 2008; 453: 56-64) の存在下で DNA 断片サイズは過大評価する可能性がある。さらに、成功裏に配列して基準ゲノムにマッピングした全 DNA 断片の > 99.9% が 600 塩基未満であるため、600 塩基以下の長さの断片をすべて含むことによって、試料中の DNA 断片サイズ分布の代表的な推定が得られる。

【0090】

したがって、 $Hap I$ と $Hap I I$ との間の短い DNA 断片によって寄与された全長の部分差を使用することができる。 $Hap I$ および $Hap I I$ DNA 断片間のサイズ分布の摂動化は、それらの F 値の差によって反映することができる。ここで我々は、 F_{Hap}

I および F_{HapI} をそれぞれ $HapI$ および $HapII$ に対する短い DNA 断片によって寄与された全長の部分として定義する。 $HapI$ と $HapII$ との間の短い DNA 断片の全長の部分差 (F) は、 $F = F_{HapI} - F_{HapII}$ として算出できる。 F_{HapI} および F_{HapII} は、各ハプロタイプ由来の断片サイズ分布 2 群の統計値例である。

【0091】

先の項目に例証した実施形態に類似して、腫瘍組織中の $HapII$ の欠失は、 $HapI$ DNA 断片サイズ分布の $HapII$ DNA 断片と比較した明らかな短縮につながる。これは F の正值につながる。 $HapII$ が複製される場合、負の F 値が観察されることになる。染色体異常の不在下、 F 値はほぼ 0 である。

10

【0092】

患者の F 値は、正常個体と比較して、値が正常であるかどうかを決定することができる。患者の F 値は、類似した癌の患者から得られた値と比較して、値が異常であるかどうかを決定することができる。かかる比較は、本明細書に記載の閾値との比較 (1 つまたは複数) に関与する可能性がある。疾患モニタリングの状況において、 Q 値は、連続的にモニタリングできる。 F 値の変化は、血漿 / 血清中の腫瘍 DNA の増大した分画濃度を示し得る。

【0093】

D. 一般的方法

図 10 は、本発明の実施形態による、染色体領域が欠失または増幅を示すかどうかを決定する生物の生体試料のハプロタイプの分析方法を例証するフローチャートである。生体試料は、正常細胞に由来し、潜在的に癌関連細胞に由来する核酸分子 (断片とも呼ばれる) を含む。これらの分子は、試料中においてセルフリーの場合がある。生物は、複数の染色体コピー、すなわち、少なくとも二倍体生物を有する任意のタイプとすることができるが、より高倍数体の生物を含むことができる。

20

【0094】

この方法および本明細書に記載の他の任意の方法の 1 つの実施形態では、生体試料は、セルフリーの DNA 断片を含む。本願に記載の異なる方法を例証するために血漿 DNA 分析を使用しているが、これらの方法は、正常および腫瘍由来 DNA の混合物を含む試料中の腫瘍関連染色体異常を検出するためにも適用できる。他の試料タイプとしては、唾液、裂傷、胸膜流体、腹水、胆液、尿、血清、脾液、大便および子宮頸管スミア試料が挙げられる。

30

【0095】

1010 工程では、第 1 および第 2 ハプロタイプは、生物の正常細胞の第 1 染色体領域において決定される。ハプロタイプは、任意の適切な方法 (本明細書に記載の方法等) によって決定することができる。染色体領域は、任意の方法、例えば、本明細書に記載の方法によって選択し得る。第 1 染色体領域は、異型接合している第 1 の複数の遺伝子座 (例えば、領域 410 の遺伝子座 420) を含む。異型接合の遺伝子座 (異性) は互いに離れていてもよく、例えば、遺伝子座は、第 1 の複数の遺伝子座の別の遺伝子座と 500 または 1000 塩基 (以上) 離れていることができる。他の異性は、第 1 染色体領域に存在し得るが、使用されない。

40

【0096】

1020 工程では、生体試料中の複数の核酸分子は、各分子の位置および対立遺伝子に関して特性決定される。例えば、生物の基準ゲノムにおける核酸分子の位置を同定することができる。この位置付けは、分子シーケンシングを実施して (例えばユニバーサルシーケンシングを介して)、分子の 1 つまたは 2 つ (対末端) の配列したタグを得て、次いで配列したタグ (1 つまたは複数) を基準ゲノムに整列することを含む様々な方法で実施することができる。かかる整列は、塩基局所整列検索ツール (BLAST) 等のツールを用いて実施することができる。位置は、染色体群中の番号で同定することができる。断片が由来するハプロタイプがどれかを決定するために、1 つの異型接合の遺伝子座 (異性

50

）の対立遺伝子を使用することができる。

【 0 0 9 7 】

1 0 3 0 工程では、核酸分子の第 1 群を、同定した位置および決定した対立遺伝子に基づき第 1 ハプロタイプ由来と同定する。例えば、対立遺伝子 A を有する図 4 の遺伝子座 4 2 1 を含む断片を、H a p I 由来と同定することになる。第 1 群は、第 1 の複数の遺伝子座のそれぞれで位置する少なくとも 1 つの核酸分子を含むことによって、第 1 染色体領域にわたることができる。

【 0 0 9 8 】

1 0 4 0 工程では、核酸分子の第 2 群を、同定した位置および決定した対立遺伝子に基づき第 2 ハプロタイプ由来と同定する。例えば、対立遺伝子 T を有する図 4 の遺伝子座 4 2 1 を含む断片を、H a p I I 由来と同定することになる。第 2 群は、第 1 の複数の遺伝子座のそれぞれで位置する少なくとも 1 つの核酸分子を含む。

【 0 0 9 9 】

1 0 5 0 工程では、コンピュータシステムは、核酸分子の第 1 群の第 1 値を算出する。第 1 値は第 1 群の核酸分子の特性を決定する。第 1 値の例としては、第 1 群における分子数のタグ数および第 1 群における分子のサイズ分布が挙げられる。

【 0 1 0 0 】

1 0 6 0 工程では、コンピュータシステムは、核酸分子の第 2 群の第 2 値を算出する。第 2 値は第 2 群の核酸分子の特性を決定する。

【 0 1 0 1 】

1 0 7 0 工程では、第 1 値を第 2 値と比較し、第 1 染色体領域が欠失または増幅を示すかどうかの分類を決定する。欠失または増幅の存在の分類によって、癌関連細胞を有する生物に関する情報を得ることができる。比較例としては、本明細書に記載のように、2 値の差もしくは比率の取得、ならびに結果と 1 つもしくは複数の閾値との比較が挙げられる。例えば、比率は、S P R T 分析における閾値と比較することができる。分類例は、正（すなわち検出した増幅または欠失）、負、および未分類、ならびに各種の正負度（例えば、1 ~ 1 0 の整数、または 0 ~ 1 の実数の使用）を含むことができる。増幅は、単純な複製を含むことができる。かかる方法は、腫瘍 D N A、および前癌病巣、すなわち癌の前駆体由来の D N A を含む癌関連核酸の存在を検出することができる。

【 0 1 0 2 】

E . 深さ

分析の深さとは、特定の精度内で分類または他の決定を得るために分析する必要がある分子量を指す。1 つの実施形態では、深さは、既知の異常に基づき算出し得、次いで、当該深さを有する測定および分析を実施し得る。別の実施形態では、分析は、分類するまで継続し得、分類する深さは、癌レベル（例えば、癌病期または腫瘍サイズ）を決定するために使用することができる。以下に、深さに関与する一部の算出例を提供する。

【 0 1 0 3 】

偏差は、本明細書に記載の任意の差または比率を指すことができる。例として、偏差は、第 1 値と第 2 値との間であることも、または本明細書に記載の閾値もしくは腫瘍濃度由来の変数とすることもできる。偏差が二重の場合、測定する必要がある断片数は $1 / 4$ 低下する。より一般に、偏差が N 倍増大する場合、測定する必要がある断片数は $1 / N^2$ である。命題として、偏差が $1 / N$ 低下する場合、試験する断片数は N^2 増加する。 N は実数とすることも整数とすることもできる。

【 0 1 0 4 】

腫瘍 D N A が試料（例えば血漿）1 0 % である症例を想定すると、一千万個の断片のシーケンシングから統計的有意差が見られると推定される。例えば、ここで、富化手順は、試料中に 2 0 % の腫瘍 D N A があり、次いで、必要な断片数が 2 , 5 0 0 , 0 0 0 断片となるように実施する。このような形で、深さは、試料中の腫瘍 D N A パーセンテージに相関することができる。

【 0 1 0 5 】

増幅量はまた、深さにも影響を及ぼす。当該領域の2倍のコピー量（例えば正常2個に対して4個）を有する領域において、X断片数を分析する必要があるものと仮定する。領域が正常コピーの4倍量を有する場合、この領域は、 $X/4$ 量の断片を必要とする。

【0106】

F. 閾値

標準値からの変数の偏差量（例えば各ハプロタイプ値の差または比率）は、上記のように診断を得るために使用することができる。例えば、偏差は、領域の一方のハプロタイプ断片の平均サイズと、もう一方のハプロタイプ断片の平均サイズとの差であり得る。偏差がある量（例えば、正常試料および/または領域から決定される閾値）を超える場合、欠失または増幅が同定される。しかしながら、閾値を超える範囲も参考にすることができ、これはそれぞれ異なる癌レベルに対応する複数の閾値として使用につながる。例えば、正常より高い偏差からどの癌病期にあるかを得ることができる（例えば第4病期の不平衡度は第3病期の不平衡度より高い）。より高い偏差はまた、大型であるために多くの断片を放出するおよび/または領域が何倍も増幅している腫瘍に起因している可能性もある。

【0107】

異なる癌レベルが得られる他に、各種閾値によって、異常を伴う領域または特定の領域を効率的に検出することもできる。例えば、あるものは、主に3倍以上の増幅を追及する高閾値を設定し、1ハプロタイプの欠失より大きな不平衡を得ることができる。領域の2コピーの欠失も検出することができる。また、低閾値は、異常を有し得る領域を同定するために使用することができ、次いで、これらの領域をさらに分析して、異常の有無および位置を確認することができる。例えば、二分探索（または八分木等の高分探索）を、高閾値を用いた階層の低層で実施することができる。

【0108】

図11は、本発明の実施形態による、欠失領域を決定するための血漿中で行なわれた測定値と共に癌細胞内で欠失している副領域1130を伴う領域1110を示す。染色体領域1110は、本明細書に記載の任意の方法（ゲノムを均一サイズの断片に分割する等）によって選択し得る。図11はさらに、各遺伝子座1120の対立遺伝子数を示す。領域1140（正常領域）および領域1130（欠失領域）のそれぞれにおいて総累積も維持される。

【0109】

領域1110を分析用に選択する場合、累積数はHap Iに対して258、Hap IIに対して240であり、11遺伝子座において18差となる。かかる差は、総数パーセンテージとして、欠失した副領域1130のみを分析する場合より小さい。これは、領域1110の約半分が正常であり、癌細胞内の副領域1130はすべて欠失しているため道理にかなっている。したがって、使用する閾値に応じて領域1110内の異常が見逃される可能性がある。

【0110】

副領域の欠失を検出するため、実施形態は、比較的広い領域の低閾値を使用することができる（本実施例において、領域1110は、同定する欠失領域サイズより比較的大きいと考えられる）。低閾値はより多くの領域を同定し、これは一部の偽陽性を含むが、偽陰性を低減する。ここで、偽陽性は、さらなる分析を介して除去することができ、これはまた異常を突き止めることもできる。

【0111】

領域をさらなる分析用にフラグ化後、さらなる分析用に領域を副領域に分割することができる。図11において、あるものは11個の遺伝子座を（例えば、二分木を用いて）半分に分けて、6つの遺伝子座の副領域1140および5つの遺伝子座をとまなう副領域1130を得ることができる。これらの領域は、同一閾値以上の厳しい閾値で分析できる。次いで、本実施例において、副領域1140は正常と同定し、副領域1130は欠失または増幅を含むと同定する。このような形で、異常を有しないものとしてより大きい領域を却下することができ、高信頼で（例えば、より高い閾値を用いて）異常を示す副領域を

同定するために疑わしい領域（低閾値を超える領域）のさらなる分析に時間を費やすことができる。本明細書においてR H D Oを使用した、サイズ技術も等しく適応可能である。

【 0 1 1 2 】

第1のレベルの探索のための領域サイズ（および低分木レベルの副領域サイズ）を、検出する異常サイズに基づき選択することができる。癌は、10MB長の異常を伴う10領域を示すことが見出されている。患者はまた、異常を示す100MB領域も有している。癌の後期は、より長い異常部分を有し得る。

【 0 1 1 3 】

G . 領域内の異常位置の精練

前の項目において、探索木に基づく領域の副領域への分割を論じた。ここで、我々は副領域を分析する他の方法について論じ、領域内の異常を突き止める。

10

【 0 1 1 4 】

図12は、本発明の実施形態による、R H D O分析を用いた異常位置のマッピング方法を示す。染色体領域は、H a p IとH a p I Iで標識した非癌細胞のハプロタイプで水平に示す。癌細胞内のH a p I Iの欠失領域はL O Hで標識する。

【 0 1 1 5 】

図示のように、R H D O分析は、仮説染色体領域1202の左側から開始して右側方向に行なわれる。各矢印は、断片のR H D O分類を示す。各断片は、断片自体の領域、具体的には、より広い領域の異性の一部を有する副領域とみなすことができる。R H D O分類断片サイズは、分類を決定できる前の遺伝子座の数（および遺伝子座位）に依存する。各R H D O断片に含まれる遺伝子座の数は、各断片において分析する分子数、所望精度（例えばS P R T分析におけるオッズ比）、および試料中の腫瘍由来DNAの分画濃度に依存する。図4および図5に例証した実施例として2ハプロタイプ間に統計的有意差が存在するかどうかを決定する上で分子数が適している場合に、分類を行なう。

20

【 0 1 1 6 】

各色つきの水平矢印は、DNA試料中にハプロタイプ用量不平衡がないことを示すR H D O分類断片を示す。腫瘍中L O Hを伴わない領域内で、6つのR H D O分類を行ない、それぞれハプロタイプ用量不平衡がないことを示す。次のR H D O分類断片1210は、L O Hを伴うおよび伴わない領域間1205接合と交差する。図12の下部にて、R H D O断片1210のS P R T曲線を示す。黒色垂直矢印は、L O Hを伴う領域と伴わない領域との間の接合を示す。L O Hを伴う領域からの増大データが蓄積すると共に、この断片のR H D Oの分類はハプロタイプ用量不平衡の存在を示す。

30

【 0 1 1 7 】

各白色水平矢印は、ハプロタイプ用量不平衡が存在することを示すR H D O分類断片を示す。右側の連続する4つのR H D Oはまた、DNA試料中のハプロタイプ用量不平衡の存在も示す。L O Hを伴うおよび伴わない領域間の接合の位置は、R H D O分類の変化、すなわちハプロタイプ用量不平衡の存在から不在またはその逆変化を示す第1R H D O断片以内と推定することができる。

【 0 1 1 8 】

図13は、本発明の実施形態による、別方向から開始したR H D O分類を示す。図13において、両方向からのR H D O分類を示す。左側から開始するR H D O分析から、L O Hを伴うおよび伴わない領域間の接合は、ハプロタイプ用量不平衡の存在を示す第1R H D O断片1310以内と推定することができる。右側から開始するR H D O分析から、接合は、ハプロタイプ用量不平衡は存在しないと示唆される第1R H D O断片1320以内と推定することができる。2方向に実行されるR H D O分析の情報を統合し、L O Hを伴うおよび伴わない領域間の接合位置1330を推定することができる。

40

【 0 1 1 9 】

I V . 異常の非特異的ハプロタイプ検出

R H D O方法は、異型接合の遺伝子座の使用に依存する。ここで、二倍体生物の染色体は、2ハプロタイプをもたらす一部の差を有するが、異型接合の遺伝子座の数は変わる可

50

能性がある。一部の個体は、比較的少ない異型接合の遺伝子座を有し得る。この項目に記載の実施形態は、2領域を比較して同種接合であり、同一領域の2ハプロタイプではない遺伝子座にも使用することができる。したがって、いくつかの欠点が2つの異なる染色体領域との比較から存在し得るが、より多くのデータポイントが得られ得る。

【0120】

相対的染色体領域用量方法において、（例えば、当該領域に整列された配列したタグの計数によって決定した）1染色体領域由来の断片数を、（基準染色体領域由来であっても健常と知られている別試料の同一領域であってもよい）予想値と比較する。このような形で、配列したタグがどのハプロタイプに由来するかに関わらず断片の染色体領域を算出することになる。したがって、異性を含まない配列したタグを依然として使用することができる。比較を行なうため、実施形態は、比較前にタグ数を正規化できる。各領域は、（互いに離れた）少なくとも2つの遺伝子座によって定義し、これら遺伝子座での断片を使用して領域に関する集合的な値を得ることができる。

10

【0121】

特定領域における配列したリード（タグ）の正規化した値は、その領域に整列された配列したリード数を完全ゲノムに整列可能な配列したリード総数で割ることによって算出することができる。この正規化したタグ数によって、1つの試料からの結果を別の試料の結果と比較できる。例えば、正規化された値は、上述のように、特定領域由来であると予想される配列したリード率（例えば、パーセンテージまたは分数）とすることができる。しかしながら、当業者に明らかであるように他の多くの正規化が可能である。例えば、あるものは、1領域数を基準領域数で割ることによって標準化できる（上の症例では、基準領域は、単に完全ゲノムである）。次いで、この正規化タグ数は、癌を示さない1つまたは複数の基準試料から決定し得る閾値と比較されてもよい。

20

【0122】

次いで、研究症例の正規化したタグ数を、1つまたは複数の基準対象、例えば癌を呈さない者の正規化したタグ数と比較する。1つの実施形態では、比較は、特定の染色体領域における症例の z -スコアを算出することによって行なう。 z -スコアは、以下の方程式： z -スコア = (症例の正規化したタグ数 - 平均) / $S \cdot D$ 。（式中、「平均」は、基準試料において特定の染色体領域に整列された平均の正規化したタグ数であり；および $S \cdot D$ は基準試料において特定領域に整列された正規化したタグ数の標準偏差である）を用いて算出する。それゆえ、 z -スコアは、研究症例の染色体領域の正規化したタグ数が1つまたは複数の基準対象の同染色体領域において平均の正規化したタグ数から離れている標準偏差の数である。

30

【0123】

検査した生物が癌を有する状況において、腫瘍組織中で増幅される染色体領域は、血漿DNA中で過剰発現することになる。これは、 z -スコア値に正値をもたらすことになる。他方、腫瘍組織中に欠失している染色体領域は、血漿DNA中で不足発現する。これは、 z -スコア値に負値をもたらすことになる。 z -スコアの等級は、いくつかの要因によって決定される。

【0124】

1要因は、生体試料（例えば血漿）中の腫瘍由来DNAの分画濃度である。試料（例えば血漿）中の腫瘍由来DNAの分画濃度が高いほど、研究症例の正規化したタグ数と基準症例間の差が開く。それゆえ、 z -スコアの等級はより大きくなる。

40

【0125】

別の要因は、1つまたは複数の基準症例における正規化したタグ数の多型である。研究症例の生体試料（例えば血漿）中の染色体領域の同程度の過剰発現と共に、基準群の正規化したタグ数の多型が小さい（すなわち標準偏差が小さい）と、より高い z -スコアをもたらすことになる。同様に、研究症例の生体試料（例えば血漿）中の染色体領域の同程度の不足発現と共に、基準群の正規化したタグ数の標準偏差が小さいほど、より負の z -スコアをもたらすことになる。

50

【0126】

別の要因は、腫瘍組織中の染色体異常の等級である。染色体異常の等級とは、特定の染色体領域（獲得または喪失のいずれか）におけるコピー数変化を指す。腫瘍組織中のコピー数変化が高いほど、血漿DNA中の特定の染色体領域の過剰発現度または不足発現度は高い。例えば、染色体の両コピーの喪失は、2つの染色体コピーのうちの1つの喪失より血漿DNA中の染色体領域の高い不足発現をもたらす、それゆえ、より負のz-スコアをもたらす。典型的には、癌中に複数の染色体異常がある。各癌における染色体異常はさらにその本質（すなわち増幅または欠失）、程度（単一または複数のコピーの獲得または喪失）および範囲（染色体の長さの点から見た異常サイズ）によって変わる可能性がある。

【0127】

正規化したタグ数の測定精度は、分析する分子数による影響を受ける。我々は、分画濃度が約12.5%、6.3%および3.2%である場合、1つのコピー変化（獲得または喪失のいずれか）を伴う染色体異常を検出するためにそれぞれ15,000、60,000および240,000分子を分析する必要があるものと予測する。異なる染色体領域における癌の検出のためのタグ計数に関するさらなる詳細については、Loらによる「大規模並列ゲノムシーケンシングを用いた胎児染色体の異数性診断（Diagnosing Fetal Chromosomal Aneuploidy Using Massively Parallel Genomic Sequencing）」と題された米国特許公開第2009/0029377号（その内容全体は、すべての目的で参照によって本明細書に組み込まれる）に記載されている。

【0128】

実施形態は、タグ計数法の代わりにサイズ分析も使用することができる。正規化したタグ数の代わりに、サイズ分析も使用し得る。サイズ分析は、本明細書および米国特許第12/940,992号に記載されるように様々な変数を使用することができる。例えば、上記Q値を使用してもF値を使用してもよい。これらの値はリード数で測らないため、かかるサイズ値は他の領域からの計数による正規化を必要としない。ハプロタイプ特異的方法の技術は、非特異的方法にも同様に使用することができる。例えば、領域の深さおよび精練に關与する技術を使用し得る。いくつかの実施形態では、2領域を比較時に特定領域のGCバイアスを考慮することができる。RHDO方法は同じ領域を用いるため、かかる補正は必要ではない。

【0129】

V. 複数の領域

ある癌は典型的には特定の染色体領域に異常を伴い存在する可能性があるが、かかる癌は必ずしも同一領域のみに存在しない。例えば、追加の染色体領域は、異常を示す可能性があり、かかる追加領域の位置は未知である場合がある。さらに、早期癌を同定するために患者をスクリーニングする場合、あるものは、広範囲の癌を同定することを望む場合があり、これはゲノム全体に存在する異常を示す可能性がある。これらの状態に対応するため、実施形態は、どの領域が異常を示すかを決定する系統的な形式で複数の領域を分析することができる。異常の数およびそれらの位置（例えばそれらが連続的であるかどうか）を使用して、例えば、異常を確認し、癌病期を決定し、癌の診断を行ない（例えば数が閾値を超す場合）、ならびに異常を示す様々な領域の数および位置に基づき予後を得ることができる。

【0130】

したがって、実施形態は、異常を示す領域数に基づき生物が癌を有するかどうかを同定することができる。したがって、あるものは複数（例えば、3000）の領域を試験して異常を示すいくつかの領域を同定することができる。領域は、完全ゲノムを網羅してもよいし、一部のみのゲノム、例えば、非反復領域を網羅してもよい。

【0131】

図14は、本発明の実施形態による、複数の染色体領域を用いて生物の生体試料を分析する方法1400のフローチャートである。生体試料は、核酸分子（断片とも呼ばれる）

10

20

30

40

50

を含む。

【0132】

1410工程では、生物の複数の重複しない染色体領域を同定する。各染色体領域は、複数の遺伝子座を含む。上述のように、領域は、1Mbサイズであることも、他の一部の同等サイズであることもできる。次いで、完全ゲノムは、それぞれ所定のサイズおよび位置の約3,000領域を含むことができる。また、上述のように、かかる所定の領域は、使用する特定の染色体の長さもしくは特定数の領域、ならびに本明細書に記載の他の任意の基準に適するように変更することができる。領域が異なる長さを有する場合、かかる長さは、例えば、本明細書に記載のように、結果を正規化するために使用することができる。

10

【0133】

1420工程では、生物の基準ゲノムにおける核酸分子の位置を、複数の核酸分子のそれぞれに対して同定する。位置は、本明細書に記載の任意の方法、例えば、配列したタグを得るための断片のシーケンシングおよび基準ゲノムへの配列したタグの整列によって決定し得る。分子の特定のハプロタイプは、ハプロタイプ特異的方法においても決定することができる。

【0134】

1430～1450工程は、各染色体領域において実施される。1430工程では、核酸分子の各群を、同定した位置に基づき染色体領域由来と同定する。各群は、染色体領域の複数の遺伝子座のそれぞれで位置する少なくとも1つの核酸分子を含む。1つの実施形態では、例えば、上記のRHDO方法のように、群は、染色体領域の特定のハプロタイプに整列された断片とすることができる。別の実施形態では、群は、IV項に記載の方法のように染色体領域に整列された任意の断片とすることができる。

20

【0135】

1440工程では、コンピュータシステムは、核酸分子の各群の各値を算出する。各値は、各群の核酸分子の特性を定義する。各値は、本明細書に記載の任意の値とすることができる。例えば、値は、群の断片数または群の断片サイズ分布の統計値とすることができる。各値はまた、正規化した値、例えば、試料のタグ数の総番号または基準領域のタグ数番号のために分割した領域のタグ数とすることもできる。各値は、別の値（例えば、RHDOにおいて）からの差とすることも、または比率とすることもでき、それによって、領域差の特性を提供することができる。

30

【0136】

1450工程では、各値を基準値と比較して、第1染色体領域が欠失を示すか、または増幅を示すかの分類を決定する。この基準値は、本明細書に記載の任意の閾値または基準値とすることができる。例えば、基準値は正常試料のために決定された閾値とすることができる。RHDOにおいて、各値は、2ハプロタイプのタグ数の差または比率とすることができ、基準値は、統計的有意な偏差が存在することを決定する閾値とすることができる。別の例では、基準値は、別のハプロタイプまたは領域のタグ数またはサイズ値とすることができ、比較は、差または比率（またはかかる関数）を取り、次いで、差または比率が閾値を超すかどうかの決定を含むことができる。

40

【0137】

基準値は、他の領域の結果に基づき変わる可能性がある。例えば、隣接領域も偏差を示す場合（1つの閾値、例えば、z-スコア3と比較して小さい）、低い閾値を使用することができる。例えば、3つの連続領域がすべて第1の閾値を超す場合、癌の可能性が高い場合がある。したがって、この第1の閾値は、断続的領域から癌を同定する必要がある別の閾値より低い場合がある。さらに偏差の小さな3つの（または3つを超える）領域を有することは、感受性および特異性を保持できる確率効果の十分に低い確率を有することができる。

【0138】

1460工程では、欠失または増幅を示すとして分類される染色体領域量を決定する。

50

算出する染色体領域は、制限を設けることができる。例えば、少なくとも1つの他の領域と隣接した領域のみを算出し得る（または連続領域は、あるサイズ、例えば、4つ以上の領域を必要とする可能性がある）。領域が等しくない実施形態において、数は、またそれぞれの長さも考慮できる（例えば、数は、異常領域の全長である可能性がある）。

【0139】

1470工程では、量は、試料の分類を決定する閾値量と比較する。例として、分類は、生物が癌を有するかどうか、癌病期、および癌の予後とすることができる。1つの実施形態では、領域が現れる位置に関わらず、全異常領域を計数し、単一閾値を使用する。別の実施形態では、閾値は、計数された領域の位置およびサイズに基づき変わる可能性がある。例えば、特定の染色体または染色体群上の領域量は、特定の染色体（または群）の閾値と比較し得る。複数の閾値を使用し得る。例えば、特定の染色体（または群）上の異常領域量は、第1の閾値より高くなければならず、ゲノム中の異常領域の総量は、第2の閾値より高くなければならない。

【0140】

この領域量の閾値はまた、計数された領域の不均衡の強度にも依存する可能性がある。例えば、癌分類を決定する閾値として使用される領域量は、各領域の異常を検出するために使用される特異性および感受性（異常閾値）に依存する可能性がある。例えば、異常閾値が低い（例えば z -スコア2である）場合、閾値量は、高く（例えば、150）選択することができる。しかしながら、異常閾値が高い（例えば、 z -スコア3である）場合、閾値量は、より低（例えば、50）い場合がある。異常を示す領域量は、加重値とすることもでき、例えば、高不均衡を示す1つの領域は、小さな不均衡のみを示す領域より加重であることができる（すなわち異常の単なる正負以上の分類がある）。

【0141】

したがって、正規化したタグ数（または群の特性の他の各値）の有意な過剰発現または不足発現を示す染色体領域量（数および/またはサイズを含み得る）は、疾患重症度を反映するために使用することができる。異常正規化したタグ数を伴う染色体領域量は、2つの要因、すなわち腫瘍組織中の染色体異常の数（またはサイズ）および生体試料（例えば血漿）中の腫瘍由来DNAの分画濃度によって決定することができる。癌が進行性であるほど、より多くの（およびより大きな）染色体異常を示す傾向がある。それゆえ、より多くの癌関連染色体異常が試料（例えば血漿）中に潜在的に検出可能である。より進行性の癌患者であるほど、より高い腫瘍負荷が血漿中の腫瘍由来DNAのより高い分画濃度につながることになる。その結果、腫瘍関連染色体異常は、血漿試料中でより容易に検出される。

【0142】

癌スクリーニングまたは検出の状況において、正規化したタグ数（または他の値）の過剰発現または不足発現を示す染色体領域量を、検査された対象が癌を有する可能性を決定するために使用することができる。カットオフ ± 2 （すなわち z -スコア > 2 または < -2 ）を用いて、研究領域の約5%が、確率のみによって対照対象の平均から有意に離れた z -スコアを提供すると予想される。完全ゲノムを1Mb断片に分割する場合、完全ゲノムに対して約3,000断片である。したがって、約150断片が z -スコア > 2 または < -2 を有すると予想される。

【0143】

したがって、癌の有無を決定するために z -スコア > 2 または < -2 の断片の数に対してカットオフ値（閾値）150を使用することができる。診断目的に適合するように、異常 z -スコアの断片数に対して他のカットオフ値（例えば、100、125、175、200、250および300）を選択することができる。低カットオフ値、例えば100は、より感受性が高い検査をもたらすことになるが、特異性はより低く、より高いカットオフ値はより特異的だが感受性は低い。偽陽性分類数は、 z -スコアのカットオフ値を増大することによって低減できる。例えば、カットオフ値が3に増大する場合、断片0.3%のみが偽陽性である。この状況において、異常 z -スコアを有する4つ以上の断片を癌の

10

20

30

40

50

存在を示すために使用することができる。他のカットオフ値、例えば 1、2、4、5、10、20 および 30 も、異なる診断目的に適合するように選択できる。しかしながら、癌関連染色体異常を検出する感受性は低下することになり、診断するために必要な異常断片数は増大する。

【0144】

特異性を損なわずに感受性を改善する 1 つの可能なアプローチは、隣接する染色体断片の結果を考慮する。1 つの実施形態では、 z -スコアのカットオフは > 2 および < -2 であり続ける。しかしながら、染色体領域は、2 連続断片が同一タイプの異常を示す（例えば両断片の z -スコアが > 2 である）場合のみに潜在的に異常であるとして分類される。正規化したタグ数の偏差がランダムな誤差である場合、同一方向に偽陽性である 2 つの連続断片を有する確率は 0.125% ($5\% \times 5\% / 2$) である。他方、染色体異常が 2 つの連続断片を包含する場合、低カットオフ値は、血漿試料中の断片の過剰発現または不足発現の検出の感受性をより高くすることになる。対照対象の平均からの正規化したタグ数（または他の値）の偏差はランダムな誤差によらないため、連続的分類の必要性は、感受性に対して有意な有害な影響を有しない。他の実施形態では、隣接断片の z -スコアは、より高いカットオフ値を用いて共に添加できる。例えば、3 つの連続断片の z -スコアを合計してカットオフ値 5 を使用することができる。この概念は、4 つ以上の連続断片に拡大することができる。

10

【0145】

量と異常閾値の組み合わせはまた分析目的、および任意の生物についての先行する知識（またはその欠如）にも依存する可能性がある。例えば、正常健常集団の癌をスクリーニングする場合、典型的にはあるものは、領域が異常を有すると同定される場合に、潜在的領域量（すなわち領域数の高閾値）と異常閾値の両方での、高特異性を使用する。しかしながら、高リスク患者（例えば、しこりまたは家族歴、喫煙、HPV ウイルス、肝炎ウイルス、または他のウイルスを訴える患者）において、より高い感受性を有する（偽陰性が低い）ために閾値はより低くする可能性がある。

20

【0146】

1 つの実施形態では、あるものが染色体異常を検出するために 1 Mb 消散および低検出限界 6.3% の腫瘍由来 DNA を使用する場合、各 1 Mb 断片の分子数は 60,000 である必要があることになる。これは、完全ゲノムに対して約 1 億 8 千万 ($60,000$ リード / Mb $\times 3,000$ Mb) の整列可能なリードに変換される。

30

【0147】

図 15 は、本発明の実施形態による、腫瘍由来断片の様々な数の断片および分画濃度に必要な深さを例証する表 1500 を示す。列 1510 は、試料のための腫瘍細胞由来の断片濃度を提供する。濃度が高いほど異常の検出が容易であるため、分析に必要な分子数が少ない。列 1520 は、断片ごとに必要な推定分子数を提供し、これは、深さに関する上記の項目に記載の方法を介して算出し得る。

【0148】

断片サイズが小さいほど、より小さな染色体異常を検出するより高い消散が得られる。しかしながら、これは全体で分析する分子数の必要性を増大することになる。断片サイズが大きいほど、消散支出の分析に必要な分子数は低下する。したがって、大きな異常のみ検出することができる。1 つの遂行では、より広い領域を使用することができ、異常を示す断片を再分割してこれらの副領域を（例えば、上記のように）分析して、より良い消散が得られる可能性がある。列 1530 は、各断片サイズを提供する。値が小さいほど、より多くの領域が使用される。列 1540 は、完全ゲノムについて分析される分子数を示す。したがって、あるものが推定値（または検出する最小濃度）を有する場合、分析する分子数を決定することができる。

40

【0149】

VI. 経時的進行

腫瘍が進行するにつれ、腫瘍はより多くの DNA 断片を放出するため（例えば、腫瘍増

50

殖、より多くの壊死、より高い血管分布によって)、腫瘍断片量は増大する。腫瘍組織から血漿中へのDNA断片が増えるほど、血漿中の不平衡度は増大する(例えば、RHDO中の2ハプロタイプ間のタグ数差は増大する)。加えて、腫瘍断片数が増大するため、異常が存在する領域数はより容易に検出することができる。例えば、領域の腫瘍DNA量は少なすぎて異常を検出することができない可能性がある。なぜなら腫瘍が小さく少量の癌DNA断片が放出される場合は十分な断片が分析されず、統計的有意差を確立することができないためである。小さい腫瘍でも多くの断片を分析できるが、大量の試料(例えば多くの血漿)を必要とする可能性がある。

【0150】

癌進行の追跡は、1つまたは複数の領域における異常量(例えば、不平衡または必要な深さによって)または異常を示す染色体領域量(数および/またはサイズ)を使用することができる。1例では、1つの領域(またはいくつかの領域)の異常量が他の領域の異常より速く増大する場合、当該領域(1つまたは複数)を、癌をモニタリングする好ましいマーカーとして使用することができる。この増大は、大きいことから断片を多く放出し、かつ/または領域が何倍にも増幅している腫瘍によってもたらされる可能性がある。あるものはまた、術後の異常値(例えば異常量もしくは異常を示す領域数、またはその組み合わせ)をモニタリングして、腫瘍が適切に除去されていることを確認することもできる。

【0151】

該技術の様々な遂行において、腫瘍DNAの分画濃度の決定は、癌病期、前兆、または進行モニタリングに使用される。進行度の測定によって、現時点での癌病期ならびに癌の増殖もしくは拡大速度に関する情報を得ることができる。癌の「病期」は、以下、すなわち、腫瘍サイズ、組織学的外観、リンパ節合併症の存在/不在、および遠位転移の存在/不在のすべてまたは一部に関する。癌の「前兆」は、疾患進行の機会および/または癌からの生存の機会の推定に関与する。これはまた、患者の臨床的進行の無いことになる時間または生存期間の推定にも関与する可能性がある。癌の「モニタリング」は、癌が進行しているか(例えば、サイズ増大、リンパ節合併症の増大、または遠位臓器への拡大、すなわち転移)の確認にも関与することになる。モニタリングはまた、腫瘍が治療によって制御されているかどうかの確認にも関与する可能性がある。例えば、治療が効果的である場合、あるものは、腫瘍サイズの低下、転移またはリンパ節合併症の退化、患者の一般的な健康な状態の改善(例えば体重増加)を確認することができる。

【0152】

A. 癌DNAの分画濃度の決定

1つまたは複数の領域の異常増量の追跡方法の1つは、領域(1つまたは複数)に対する癌DNAの分画濃度の決定である。次いで、癌DNAの分画濃度の変化を腫瘍の経時的追跡に使用することができる。この追跡は診断に使用することができる、例えば、第1測定値は、背景レベル(人の一般的異常レベルに対応し得る)を提供することができ、後の測定値によって、腫瘍増殖(したがって、癌)を示唆する変化を確認することができる。癌DNAの分画濃度の変化はまた、治療がいかに良く行なわれているかの予見にも使用することができる。この技術の他の遂行では、血漿中の腫瘍DNAの分画濃度増大は、患者の予後不良、または腫瘍負荷の増大を示すことになる。

【0153】

癌DNAの分画濃度は、様々な方法で決定することができる。例えば、あるハプロタイプを別のハプロタイプと比較した(または1つの領域を別の領域と比較した)タグ数の差である。別の方法は、統計的有意差が見られる前の深さ(すなわち分析する断片数)である。先の例において、ハプロタイプ用量の差は、ヘテロ接合の欠失を有する染色体領域を分析することによって、生体試料(例えば血漿)中の腫瘍由来DNAの分画濃度を決定するために使用することができる。

【0154】

腫瘍由来DNA量は、癌患者における腫瘍負荷と正相関することが示されている(L o e t a l . C a n c e r R e s . 1 9 9 9 ; 5 9 : 5 4 5 2 - 5 . および C h a n

10

20

30

40

50

et al. Clin Chem. 2005; 51: 2192-5)。したがって、R H D O分析による生体試料（例えば血漿試料）中の腫瘍由来DNAの分画濃度の連続モニタリングを使用して患者の疾患進行をモニタリングすることができる。例えば、治療後に連続収集した試料（例えば血漿）中の腫瘍由来DNAの分画濃度のモニタリングを、治療の成功を決定するために使用することができる。

【0155】

図16は、本発明の実施形態による、R H D O分析による血漿中の腫瘍由来DNAの分画濃度を測定する原理を示す。2ハプロタイプ間の不平衡を決定し、この不平衡度を、試料中の腫瘍DNAの分画濃度を決定するために使用することができる。

【0156】

H a p IとH a p I Iは、非腫瘍組織中の2ハプロタイプを示す。H a p I Iは、副領域1610内の腫瘍組織中で部分的に欠失している。したがって、血漿中に検出された欠失領域1610に対応するH a p I I関連断片は、非腫瘍組織によって寄与されている。他方、H a p I中の領域1610は、腫瘍組織と非腫瘍組織の両方に存在する。したがって、H a p IとH a p I Iとのリード数差は、血漿中の腫瘍由来DNA量を表す。

【0157】

腫瘍由来DNAの分画濃度(F)は、以下の式： $F = (N_{H a p I} - N_{H a p I I}) / N_{H a p I} \times 100\%$ （式中、 $N_{H a p I}$ は、L O Hによる影響を受ける染色体領域に位置する異型接合のS N PにおけるH a p I上の対立遺伝子に対応する配列したリード数であり、かつ $N_{H a p I I}$ は、L O Hによる影響を受ける染色体領域1610に位置する異型接合のS N PにおけるH a p I I上の対立遺伝子に対応する配列したリード数である）を用いて、L O Hによる影響を受ける染色体領域における欠失しているおよび欠失していない染色体由来の配列リード（タグ）数から算出できる。

【0158】

上の式は、欠失を含まない染色体領域（H a p I）上に位置する異型接合の遺伝子座における累積タグ数としてのp、および欠失を含む染色体領域（H a p I I）1610の累積タグ数としてのqの決定と同等であり、試料中の腫瘍DNAの分画濃度(F)を $F = 1 - q / p$ として算出する。図11に例証した例において、腫瘍DNAの分画濃度は14%（ $1 - 104 / 121$ ）である。

【0159】

H C C患者の血漿試料中の腫瘍由来DNAの分画濃度を腫瘍切除の前後に収集した。腫瘍切除前、所与の染色体領域の第1ハプロタイプにおける $N_{H a p I}$ は30,443であり、染色体領域の第2ハプロタイプにおける $N_{H a p I I}$ は16,221であり、これによってF46.7%が得られた。腫瘍切除後、 $N_{H a p I}$ は31,534であり、 $N_{H a p I I}$ は31,089であり、これによってF1.4%が得られた。このモニタリングによって、腫瘍切除が成功したことが示される。

【0160】

循環DNAサイズプロファイル変化度は、分画濃度を決定するためにも使用することができる。1つの遂行では、腫瘍組織と非腫瘍組織の両方に由来する血漿DNAの正確なサイズ分布を決定することができ、次いで、2つの既知の分布間に収まる測定したサイズ分布によって分画濃度を得ることができる（例えば、腫瘍組織と非腫瘍組織のサイズ分布の2つの統計値の間の線形モデルを用いて）。あるいは、サイズ変化の連続モニタリングを使用することができる。1つの態様では、サイズ分布変化は、血漿中の腫瘍DNAの分画濃度に比例するものとして決定される。

【0161】

異なる領域間の差も類似した形で、すなわち、上記の非特異的ハプロタイプ検出方法において使用することができる。タグ計数法において、いくつかの変数を疾患進行のモニタリングに使用することができる。例えば、染色体異常を示す領域のz-スコアの等級は、生体試料（例えば血漿）中の腫瘍由来DNAの分画濃度を反映するために使用することができる。特定領域の過剰発現度または不足発現度は、試料中の腫瘍由来DNAの分画濃度

10

20

30

40

50

および腫瘍組織中のコピー範囲またはコピー数の変化数に比例する。z - スコアの等級は、対照対象と比較した試料中の特定の染色体領域の過剰発現度または不足発現度の測定値である。したがって、z - スコアの等級は、試料中の腫瘍DNAの分画濃度、および、ひいては患者の腫瘍負荷を反映することができる。

【0162】

B．領域数の追跡

上述のように、染色体異常を示す領域数を癌スクリーニングに使用することができ、同様にモニタリングおよび予見するためにも使用することができる。例として、モニタリングは、現時点での癌病期、癌が再発したかどうか、および治療が効いたかどうかを決定するために使用することができる。腫瘍が進行するにつれ、腫瘍のゲノム構成はより分解する。この継続した分解を同定するため、領域の数（例えば、前定義した1 Mbの領域）を追跡する方法を、腫瘍の進行を同定するために使用することができる。癌病期がより進行した腫瘍は、異常を示す領域をより多く有する。

10

【0163】

C．方法

図17は、本発明の実施形態による、核酸分子を含む生体試料を用いて、生物の染色体異常の進行を決定する方法を例証するフローチャートである。1つの実施形態では、核酸分子の少なくとも一部はセルフリーである。例として、染色体異常は、悪性腫瘍であっても前悪性病巣由来であってもよい。また、異常増大は、経時的に染色体異常を含む細胞をより多く有する生物に起因する可能性があり、または細胞当たり増大した量の異常を含む細胞率を有する生物に起因する可能性がある。低下の例として、治療（例えば手術または化学療法）は、癌関連細胞の除去または低減を生じる可能性がある。

20

【0164】

1710工程では、生物の1つまたは複数の重複しない染色体領域を同定する。各染色体領域は複数の遺伝子座を含む。領域は、任意の適切な方法、例えば、本明細書に記載の方法によって同定することができる。

【0165】

1720～1750工程は、複数時点のそれぞれで実施される。各時間は、試料を生物から採取時の異なる時間と対応する。この試料は、所与の期間において分析する試料である。例えば、試料は、6ヶ月の間、毎月採取され得、試料の採取直後に分析することができる。あるいは、いくつかの測定値を複数期間にわたって採取後に分析することもできる。

30

【0166】

1720工程では、生物のこの生体試料を分析して、生物の核酸分子の基準ゲノムにおける位置を同定する。位置は、本明細書に記載の任意の方法、例えば、配列したタグを得るための断片のシーケンシングおよび基準ゲノムへの配列したタグの整列によって決定し得る。分子の特定のハプロタイプは、ハプロタイプ特異的方法に対しても決定することができる。

【0167】

1730～1750工程は、1つまたは複数の染色体領域のそれぞれに対して実施する。複数の領域を使用する場合、V項の実施形態を使用し得る。1730工程では、核酸分子の各群を、同定した位置に基づき染色体領域由来として同定する。各群は、染色体領域の複数の遺伝子座のそれぞれで位置する少なくとも1つの核酸分子を含む。1つの実施形態では、例えば、上記のRHDO方法のように、群は、染色体領域の特定のハプロタイプに整列された断片とすることができる。別の実施形態では、群は、I V項に記載の方法のように染色体領域に整列された任意の断片とすることができる。

40

【0168】

1740工程では、コンピュータシステムは、核酸分子の各群の各値を算出する。各値は、各群の核酸分子の特性を定義する。各値は、本明細書に記載の任意の値とすることができる。例えば、値は、群内の断片数、または群内の断片サイズ分布の統計値とすること

50

ができる。各値はまた、正規化した値、例えば、試料のタグ数の総番号または基準領域に対するタグ数の番号を分割した領域のタグ数とすることもできる。各値は、別の値（例えば、R H D Oにおける）からの差、または比率とすることもでき、それによって、領域差の特性を得ることができる。

【 0 1 6 9 】

1 7 5 0 工程では、各値を基準値と比較して、第 1 染色体領域が欠失を示すか、または増幅を示すかの分類を決定する。この基準値は、本明細書に記載の任意の閾値または基準値とすることができる。例えば、基準値は正常試料を決定する閾値とすることができる。R H D Oに対して、各値は、2 ハプロタイプのタグ数差またはタグ数比率とすることができる。基準値は、統計的有意な偏差が存在することを決定する閾値とすることができる。別の例では、基準値は、別のハプロタイプまたは領域のタグ数またはサイズ値とすることができる。比較は、差または比率（またはかかる関数）を取り、次いで、差または比率が閾値を超すかどうか決定することを含むことができる。基準値は、例えば、本明細書に記載の任意の適切な方法および基準に応じて決定することができる。

10

【 0 1 7 0 】

1 7 6 0 工程では、生物の染色体異常の進行を決定するために複数時点での各染色体領域の分類を使用する。進行は、生物が癌を有するかどうか、癌病期、および癌の予後を決定するために使用することができる。これらの決定はそれぞれ、本明細書に記載の癌分類に関連する可能性がある。

【 0 1 7 1 】

20

この癌分類は、様々な方法で実施することができる。例えば、異常領域量を計数して、閾値と比較することができる。領域の分類は数値（例えば、異なるハプロタイプまたは異なる領域の各値および基準値を伴う腫瘍濃度）とすることができ、濃度変化を決定することができる。濃度変化は閾値と比較して有意な上昇が生じたかどうか、したがって腫瘍の存在を示唆するかどうかを決定することができる。

【 0 1 7 2 】

V I I . 実施例

A . S P R T を用いた R H D O

この項目では、我々は、肝細胞癌（H C C）患者において S P R T を用いた相対的ハプロタイプ用量（R H D O）分析の使用例を示す。この患者の腫瘍組織中、2 つの染色体 4 の 1 つの欠失が観察された。これは染色体 4 上の S N P においてヘテロ接合の欠失をもたらす。この患者のハプロタイピングにおいて、患者、患者の妻および息子のゲノム D N A を分析して、個体 3 例の遺伝子型を決定した。次いで、患者の構成上ハプロタイプは、それらの遺伝子型に由来した。大規模並列シーケンシングを実施して、染色体 4 の 2 ハプロタイプに対応する S N P 対立遺伝子で配列したリードを同定して、計数した。

30

【 0 1 7 3 】

R H D O および S P R T の方程式および原理を上述した。1 つの実施形態では、R H D O 分析は、例えば、2 ハプロタイプのうち 1 つが増幅または欠失している場合に腫瘍由来 D N A 1 0 % の存在に対応する D N A 試料中ハプロタイプ用量差 1 0 % を検出するためにプログラムされる。他の実施形態では、R H D O 分析の感受性は、D N A 試料中の腫瘍由来 D N A 2 %、5 %、1 5 %、2 0 %、2 5 %、3 0 %、4 0 % および 5 0 % 等を検出するように設定することができる。R H D O 分析の感受性は、S P R T 分類曲線の上限閾値および下限の算出のための変数において調節することができる。調節可能な変数は、所望の検出限界レベル（例えば、分析する分子数に影響する検出可能であるべき腫瘍濃度パーセンテージ）および分類閾値、例えば、オッズ比（あるハプロタイプのタグ数を他のハプロタイプのタグ数と比較した比率）の使用とすることができる。

40

【 0 1 7 4 】

この R H D O 分析において、帰無仮説では、同一用量で存在する染色体 4 における 2 ハプロタイプである。代替仮説では、2 ハプロタイプ用量は、生体試料（例えば血漿）中 1 0 % 超異なる。2 ハプロタイプに対応する S N P 対立遺伝子を用いて配列したリード数は

50

、異なるSNP累積からのデータとしての2仮説と統計的に比較した。SPRT分類は、累積したデータが2ハプロタイプ用量が等量か、または少なくとも10%統計的に異なる量で存在するかを決定する上で十分である場合に行なわれる。染色体4のq群の典型的SPRT分類ブロックを図18Aに示す。本明細書において例証目的としてのみ閾値10%を使用する。他の程度差(例えば0.1%、1%、2%、5%、15%または20%)も検出できる。一般に、検出を望む差が小さいほど、分析する必要があるDNA分子数は多い。逆に言えば、検出を望む差が大きいほど、分析する必要があるDNA分子数は少なく、なおかつ統計的に有意な結果に達する。この分析において、SPRTにオッズ比を使用するが、他の変数(z-スコアまたはp値等)も使用し得る。

【0175】

10

HCC患者から診断時に採取したこの血漿試料では、染色体4のp群およびq群には成功的なRHDO分類がそれぞれ76件および148件あった。RHDO分類はすべて、診断時に採取した血漿試料中にはハプロタイプ用量不平衡があったことを示している。比較として、腫瘍の外科的切除後に採取した患者の血漿試料も図18Bに示すように分析した。治療後の試料において、染色体4のp群およびq群は成功的なRHDO分類がそれぞれ4件および9件あった。4つのRHDO分類はすべて、血漿試料中に>10%の観察可能なハプロタイプ用量不平衡が存在しなかったことを示している。染色体4qの9つのRHDO分類中、7つはハプロタイプ用量不平衡が存在しないことを示し、2つは不平衡が存在することを示している。RHDOブロック数は、>10%の用量不平衡は腫瘍切除後に有意に低下したことを示し、これは、>10%の用量不平衡を示す染色体領域サイズは、治療後試料中では治療前試料中より有意に小さいことを示している。これらの結果によって、血漿中腫瘍DNAの分画濃度は、腫瘍の外科的切除後に低下したことが示唆される。

20

【0176】

非ハプロタイプ特異的方法と比較して、RHDO分析によって腫瘍DNAの分画濃度をより正確に推定することができ、特に疾患進行のモニタリングに有用である。したがって、あるものは、疾患進行をともしなう症例は血漿中の腫瘍DNAの分画濃度増大を示し、対して、疾患の安定したか、または腫瘍の退行もしくは腫瘍サイズの縮小した症例は、血漿中の腫瘍DNAの分画濃度の低下を有することになることを予期する。

B. 標的分析

【0177】

30

選択された実施形態では、DNA断片のユニバーサルシーケンシングは、以下の標的富化アプローチを実施することができる。本明細書において、かかるアプローチは、富化された標的シーケンシングとも呼ばれる。かかるアプローチの1つの実施形態は、溶液中捕獲システム(例えばAgilent SureSelectシステム、イルミナTruSeq Custom Enrichment Kit(illumina.com/applications/sequencing/targeted_resequencing.ilmn))を用いたか、またはMyGenostics GenCap Custom Enrichmentシステム(mygenostics.com/))もしくはマイクロアレーベースの捕獲システム(例えばRoche NimbleGenシステム)による断片の前選択である。他の何らかの領域も捕獲することができるが、ある領域が選択的に捕獲される。かかる方法は、分析されるかかる領域をより深く(例えば、デジタルPCRを用いてより多くの断片を配列化または分析することができる)、かつ/またはより安価で分析することを可能にすることができる。深さが深いほど、領域の感受性は増大できる。他の富化方法は、断片サイズおよびメチル化パターンに基づき実施することができる。

40

【0178】

したがって、ゲノム形式でのDNA試料分析の代替は、一般的な染色体異常を検出するために目的の領域を標的化することである。分析的プロセスは主に染色体異常が潜在的に存在する領域か、または特に特定の腫瘍タイプの特徴である変化、もしくは特に臨床的に重要となる変化を伴う領域に焦点を当てるため、標的アプローチは、このアプローチの費

50

用効果を潜在的に改善できる。後者の例としては、特定の癌タイプの腫瘍形成において早期に起こる変化（例えば1 qおよび8 qの増幅存在、ならびに8 qの欠失は、HCCにおける早期染色体変化である - van Malensteyn et al. Eur J Cancer 2011; 47: 1789 - 97）、または良好予後もしくは不良予後に関連する変化（例えば6 qおよび17 q獲得、ならびに6 pおよび9 p喪失が腫瘍進行中に観察され、18 q、8 pおよび17 pのLOH存在は大腸癌患者の生存不良に関連する - Westra et al. Clin Colorectal Cancer 2004; 4: 252 - 9）、または治療応答の予測（例えば7 pでの獲得の存在は、上皮成長因子受容体変異患者におけるチロシンキナーゼ阻害剤への応答を予測する - Yuan et al. J Clin Oncol 2011; 29: 3435 - 42）が挙げられる。

癌中で改変されたゲノム領域の他の例は、いくつかのオンラインデータベース（例えばCancer Genome Anatomy Project データベース（cgap.nci.nih.gov/Chromosomes/RecurrentAberrations）およびAtlas of Genetics and Cytogenetics in Oncology and Haematology（atlasgeneticsoncology.org/Tumors/TumorListe.html））に見ることができる。対して、非標的ゲノムアプローチでは、潜在的に異常を伴う領域と同程度で染色体異常が生じる可能性が低い領域を分析する。

【0179】

我々は、標的富化戦略を適用して、HCC患者3例および健常対照対象4例由来の血漿試料を分析した。Agilent製SureSelect捕獲システム（Gnirke et al. Nat. Biotechnol 2009. 27: 182 - 9）を用いて、標的富化を実施した。可能な標的富化技術の例としてSureSelectシステムを選択した。他の溶液相（IlluminaTruSeq Custom Enrichmentシステム）、または固相（例えばRoche-Nimblegenシステム）標的捕獲システム、およびアンプリコンベースの標的富化システム（例えばQuantalife systemおよびRainDanceシステム）も使用することができる。捕獲プローブは、HCC中の一般に異常を示す、およびめったに異常を示さない染色体領域上に位置するように設計される。次いで、標的捕獲後、各DNA試料をIlluminaGAIIx分析器上の流れ細胞の1列によって配列した。増幅および欠失がめったに起こらない領域は、増幅および欠失がより一般的に存在する領域との比較参照用に使用される。

【0180】

図19において、HCC中に見られる一般的な染色体異常を示す（図は、Wongら（Am J Pathol 1999; 154: 37 - 43）から適応した）。染色体表意文字の右側の線は個々の患者試料中の染色体の獲得を表し、左側の線は染色体の喪失を表す。太線は、高レベル獲得を表す。長方形は、標的捕獲プローブの位置を表す。

【0181】

標的タグ計数分析

染色体異常の検出において、我々はまず潜在的異常を伴う領域および基準領域の正規化したタグ数を算出した。次いで、Chenら（PLoS One 2011; 6: e21791）によって既に記載されているように、正規化したタグ数を領域のGC含量用に補正した。この実施例において、染色体8のp-群を潜在的異常領域として選択し、染色体9のq-群を基準領域として選択した。染色体異常用のAffymetrix SNP 6.0アレーを用いてHCC患者3例の腫瘍組織を分析した。患者3例の腫瘍組織中8 pおよび9 qにおける染色体用量の変化を以下に示す。HCC013患者は、8 p喪失を有し、9 qは不変であった。HCC027患者は、8 p獲得を有し、9 qは不変であった。HCC023患者は、8 p喪失を有し、9 qは不変であった。

【0182】

次いで、標的分析を用いて、HCC患者3例および健常対照対象4例のchr8 pと9 qとの間の正規化したタグ数比を算出した。図20Aは、HCCと健常患者との正規化し

たタグ数比の結果を示す。HCC013およびHCC023症例において、8pと9qとの間の正規化したタグ数比の低下が観察された。これは、腫瘍組織中の染色体8p喪失の所見と一致する。HCC027症例において、増大した比率が観察され、これはこの症例の腫瘍組織中の染色体8p獲得と一致する。点線は、4例の正常症例の平均値の2つの標準偏差を伴う領域を表す。

【0183】

標的サイズ分析

先の項目において、我々は、癌患者における血漿DNA断片サイズプロファイルを決定することによって、癌関連変化の検出原理について記載している。サイズ変化も、標的富化アプローチを用いて検出することができる。HCC症例3例（HCC013、HCC027およびHCC023）において、配列したリードを基準ヒトゲノムに整列後に配列したDNA断片の各サイズを決定した。両端の最外ヌクレオチド配位から配列したDNA断片サイズを推定した。他の実施形態では、完全DNA断片を配列して、次いで、配列した長さから断片サイズを直接決定することができる。染色体8pに整列されたDNA断片サイズ分布を染色体9qに整列されたDNA断片サイズ分布と比較した。2つのDNA集団のサイズ分布差の検出において、150bpより短いDNA断片の割合をこの実施例の各集団においてまず決定した。他の実施形態では、他のサイズカットオフ値、例えば80bp、110bp、100bp、110bp、120bp、130bp、140bp、160bpおよび170bpを使用することができる。次いで、Q値を2つの割合の差として決定した。 $Q = Q_{8p} - Q_{9q}$ （式中、 Q_{8p} は、150bpより短い染色体8pに整列されたDNA断片の割合であり、かつ Q_{9q} は、150bpより短い染色体9qに整列されたDNA断片の割合である）。

【0184】

DNA断片のサイズ分布がより短いとカットオフ値（すなわちこの実施例において150bp）より短いDNAの割合がより高くなるため、より高い（より正側の）F値は、染色体9qに整列されたものと相対的に染色体8pに整列されたDNA断片のより短い分布を表す。反対に、より小さな（より負側の）結果は、染色体9qに整列されたものと比較して染色体8pに整列されたDNA断片のより長いサイズ分布を示すことになる。

【0185】

図20Bは、HCC患者3例および健常対照対象4例における対象の富化および大規模並列シーケンシング後のサイズ分析の結果を示す。健常対照対象4例の正のQ値は、染色体9qに整列されたものと比較して染色体8pに整列されたDNA断片のわずかにより短いサイズ分布を示す。点線は、対照対象4例の平均からの2つの標準偏差内のQ間隔を表す。HCC013およびHCC023症例のQ値は、対照対象の平均値未満の2つを超す標準偏差であった。これら2症例は、腫瘍組織中の染色体8pの欠失を有していた。腫瘍中8pの欠失は、この染色体領域における血漿に対する腫瘍由来DNAの寄与の低下をもたらす。循環中の腫瘍由来DNAは非腫瘍組織に由来するDNAより短いため、これは、染色体8pに整列された血漿DNA断片より明らかに長いサイズ分布につながる。これは、これら2症例のより低い（より負側の）Q値と一致する。対して、HCC027症例における8pの増幅は、この領域に整列されたDNA断片の明らかにより短い分布につながる。したがって、8pに整列された血漿DNA断片のより高い割合は短いとみなされる。これは、HCC027のQ値は、健常対照対象より正側であるという観察と一致する。

【0186】

C．腫瘍由来染色体異常の検出のための複数領域

染色体異常（ある染色体領域の欠失および増幅を含む）は、一般に腫瘍組織中に検出される。染色体異常の特徴的パターンは、異なる癌タイプにおいて観察される。ここで、我々は、いくつかの例を用いて、癌患者の血漿中のこれら癌関連染色体異常を検出する異なるアプローチを例証する。我々のアプローチはまた、癌スクリーニングならびに疾患進行および治療応答のモニタリングにも有用である。HCC患者1例および鼻咽頭（NPC）

患者2例由来の試料を分析した。HCC患者では、腫瘍の外科的切除前後に静脈血試料を収集した。NPC患者2例では、診断時に静脈血試料を収集した。加えて、慢性B型肝炎保因者1例の血漿試料および血漿中に検出可能なエプスタイン・パール・ウイルスDNAを保有する対象1例を分析した。これら対象2例はいずれの癌も呈さなかった。

【0187】

腫瘍由来染色体異常の検出をマイクロアレー分析を用いて実施した。具体的には、Affymetrix SNP 6.0マイクロアレーシステムを用いてHCC患者の血球および腫瘍試料から抽出したDNAを分析した。血球および腫瘍組織の遺伝子型は、Affymetrix Genotyping Console v4.0を用いて決定した。獲得および喪失を含む染色体異常は、SNPの異なる対立遺伝子の強度およびマイクロアレー上のコピー数多型(CNV)プローブに基づくBirdseed v2アルゴリズムを用いて決定した。

計数ベース分析

【0188】

血漿中の配列したタグ計数分析を実施するため、静脈血10ミリリットルを各対象から収集した。各血液試料において、遠心分離後の試料から血漿を単離した。QIAmp血液ミニキット(Qiagen)を用いて、DNAを4~6mLの血漿から抽出した。既に記載されているように(Lo YMD, Sci Transl Med 2010, 2: 61ra91)、血漿DNAライブラリーを構築してから、Illumina Genome Analyzer基盤を用いて、大規模並列シーケンシングに供した。血漿DNA分子の対末端シーケンシング実施した。各分子を50bpの2つの各端で配列し、したがって、計100bp/分子とした。SOAP2プログラム(soap.genomics.org.cn/)(Li Ret al. Bioinformatics 2009, 25: 1966-7)を用いて、各配列の2つの端を非反復マスクヒトゲノム(UCSC genome.ucsc.eduよりダウンロードしたHg18NCBI.36)に整列した。

【0189】

次いで、ゲノムを複数の1メガベース(1Mb)断片に分割し、各1Mb断片に整列された配列したリード数を決定した。次いで、各ピンのタグ数を、各ピンのGC含量による局所重み付き散布図平滑化(LOESS)回帰に基づくアルゴリズムを用いて補正した(Chen E et al. PLoS One 2011, 6: e21791)。この補正は、異なるゲノム断片間のGC含量差のために生じるシーケンシングに関する定量バイアスを最小限にすることを目的とする。1Mb断片への上記分割は、例証目的で使用する。他の断片サイズ、例えば2Mb、10Mb、25Mb、または50Mb等も使用することができる。また、特定の患者の特定の腫瘍および全般的な腫瘍の特定タイプのゲノム特徴に基づき断片サイズを選択することが可能である。さらに、例えば、Helicosシステム(www.helicosbio.com)、またはPacific Biosciences Single Molecular Real-Timeシステム(www.pacificbiosciences.com)等の単分子シーケンシング技術に対する、シーケンシングプロセスによってGCバイアスが低いことを示すことができる場合、GC補正工程は省略できる。

【0190】

先行研究において、我々は、癌を有しない対象由来の血漿試料57件を配列した。これらの血漿シーケンシング結果を、各1Mb断片に対してタグ数の基準範囲を決定するために使用した。各1Mb断片に対して、個体57例のタグ数の平均および標準偏差を決定した。次いで、研究対象の結果は、以下の方程式: $z\text{-スコア} = (\text{症例の配列したタグ数} - \text{平均}) / S.D.$ (式中、「平均」は、基準試料において特定の1Mb断片に整列された配列したタグ数の平均であり、かつS.D.は、基準試料において特定の1Mb断片に整列された配列したタグ数の標準偏差である)を用いて算出したz-スコアとして表した。

【 0 1 9 1 】

図 2 1 ~ 2 4 は、研究対象 4 例の配列したタグ計数分析の結果を示す。1 M b 断片は、プロット端に示される。ヒト染色体数および表意文字（最外環）は、時計回り方向の p t e r から q t e r に配向される（動原体は黄色で示す）。図 2 1 において、内環 2 1 0 1 は、腫瘍分析から決定した異常（欠失または増幅）領域を示す。内環 2 1 0 1 は 5 段階で示される。当該段階は - 2（最内線）~ + 2（最外線）である。 - 2 値は対応する領域の両染色体コピーの喪失を示す。 - 1 値は 2 つの染色体コピーのうち 1 つの喪失を示す。 0 値は染色体の獲得も喪失も示さない。 + 1 値は 1 つの染色体コピーの獲得を示し、 + 2 値は 2 つの染色体コピーの獲得を示す。

【 0 1 9 2 】

中央環 2 1 0 2 は血漿分析結果を示す。見られるように、結果は内環に反映する。中央環 2 1 0 2 は尺度のより多くの線であるが、進行は同一である。外環 2 1 0 3 は治療後の血漿分析からのデータポイントを示し、これらのデータポイントは灰色である（過剰 / 不足発現なし - 異常なしを確認）。

【 0 1 9 3 】

血漿中の配列したタグの過剰発現を伴う染色体領域（ z - スコア > 3 ）を緑色点 2 1 1 0 で表す。血漿中の配列したタグの不足発現を伴う領域（ z - スコア < - 3 ）を赤色点 2 1 2 0 で表す。血漿中で有意な染色体異常が検出されない領域（ z - スコアが、 - 3 ~ 3 ）を灰色点で表す。過剰 / 不足発現を計数の総数によって正規化する。シーケンシング前の増幅において、正規化には G C バイアスを考慮し得る。

【 0 1 9 4 】

図 2 1 は、本発明の実施形態による、血漿 D N A を計数する配列したタグからのデータを示す H C C 患者の円状プロットを示す。内側から外側への追跡：マイクロアレー分析により検出した腫瘍組織の染色体異常（赤色および緑色は、それぞれ欠失および増幅を表す）；腫瘍の外科的切除前および切除後 1 ヶ月目に得た血漿試料の z - スコア分析。腫瘍切除前、マイクロアレー分析により腫瘍組織中に同定したものと十分に相関する染色体異常を血漿中に検出した。腫瘍切除後、ほとんどの癌関連染色体異常は血漿中に消失した。これらのデータは、疾患進行および治療効果のモニタリングのかかるアプローチ値を反映する。

【 0 1 9 5 】

図 2 2 は、本発明の実施形態による、 H C C を有しない慢性 H B V 保因者の血漿試料における配列したタグ計数分析を示す。 H C C 患者（図 2 1 ）と対照的に、この患者の血漿中の癌関連染色体異常は検出されなかった。これらのデータは、癌のスクリーニング、診断、およびモニタリングのアプローチ値を反映する。

【 0 1 9 6 】

図 2 3 は、本発明の実施形態による、第 3 期 N P C 患者の血漿試料における配列したタグ計数分析を示す。治療前に採取した血漿試料中の染色体異常を検出した。具体的には、染色体 1、3、7、9、および 14 において有意な異常を同定した。

【 0 1 9 7 】

図 2 4 は、本発明の実施形態による、第 4 期 N P C 患者の血漿試料における配列したタグ計数分析を示す。治療前に採取した血漿試料中の染色体異常を検出した。第 3 期の疾患患者（図 2 3 ）と比べてより多くの染色体異常を検出した。配列したタグ数も対照平均からより離れ、すなわち z - スコアはゼロから（正または負のいずれか側に）より離れた。対照と比較して増加した染色体異常数および配列したタグ数の高い偏差度は、より進行した疾患の病期においてゲノム改変度がより深いことを反映し、それゆえ、病期、前兆および癌のモニタリングのかかるアプローチ値を反映する。

【 0 1 9 8 】

サイズベース分析

先行研究において、腫瘍組織に由来する D N A のサイズ分布は、非腫瘍組織に由来するサイズ分布より短いことが示されている（ D i e h l F e t a l . P r o c N a

10

20

30

40

50

t l A c a d S c i U S A 2 0 0 5 , 1 0 2 (4 5) : 1 6 3 6 8 - 7 3) 。 先の項目において、我々は、血漿DNAのサイズ分析によって血漿ハプロタイプ不平衡を検出するアプローチを概説した。ここで、我々は、このアプローチをさらに例証するために、HCC患者のシーケンシングデータを使用した。

【0199】

例証目的において、我々はサイズ分析用の2領域を同定した。1つの領域(染色体1(chr1);配位:159,935,347~167,219,158)において、2つの相同染色体の1つの複製を腫瘍組織中に検出した。他の領域(染色体10(chr10);配位:100,137,050~101,907,356)において、2つの相同染色体の1つの欠失(すなわちLOH)を腫瘍組織中に検出した。さらに、配列した断片がどのハプロタイプに由来するかの決定に加えて、配列した断片サイズも、基準ゲノムにおける配列した断片の最外ヌクレオチド配位を用いて生物情報工学的に決定された。次いで、2ハプロタイプのそれぞれから断片サイズ分布を決定した。

【0200】

Chr10のLOH領域において、1ハプロタイプが腫瘍組織中で欠失していた(欠失ハプロタイプ)。したがって、この欠失したハプロタイプに整列された血漿DNA断片はすべて非癌組織に由来した。他方、腫瘍組織中で欠失していなかったハプロタイプ(非欠失ハプロタイプ)に整列された断片は腫瘍組織に由来することも非腫瘍組織に由来することもできる。腫瘍由来DNAのサイズ分布はより短いため、我々は、欠失していないハプロタイプ由来の断片に対して、欠失したハプロタイプ由来のものと比較してより短いサイズ分布を予期することになる。2つのサイズ分布差は、断片の累積頻度をDNA断片サイズに対してプロットすることによって決定することができる。短いサイズ分布のDNA集団は、短いDNAをより豊富に有することになり、それゆえ、サイズスペクトルの短い末端で累積頻度がより急速に増大する。

【0201】

図25は、本発明の実施形態による、腫瘍組織中のLOHを示す領域サイズに対する血漿DNAの累積頻度プロットを示す。X軸は、塩基対の断片サイズである。Y軸は、X軸上の値未満のサイズを有する断片パーセンテージである。欠失していないハプロタイプ由来の配列は、欠失したハプロタイプ由来の配列と比較すると、より急速に増大し、サイズ170bp未満の累積頻度がより高い。これは、欠失していないハプロタイプ由来の短いDNA断片がより豊富であったことを示す。これは、欠失していないハプロタイプ由来の腫瘍由来の短いDNAの寄与となるため、上の予測と一致する。

【0202】

1つの実施形態では、サイズ分布差は、2つのDNA分子集団の累積頻度差によって定量化することができる。我々は、 Q を2つの集団の累積頻度差として定義する。 $Q = Q_{\text{非欠失}} - Q_{\text{欠失}}$ (式中、 $Q_{\text{非欠失}}$ は、欠失していないハプロタイプ由来の配列したDNA断片の累積頻度を示し、かつ $Q_{\text{欠失}}$ は、欠失したハプロタイプ由来の配列したDNA断片の累積頻度を示す)。

【0203】

図26は、LOH領域における配列した血漿DNAサイズに対する Q を示す。本発明の実施形態による、サイズ130bpで Q は0.2に達する。これは、短いDNAを決定するカットオフとして130bpを使用することが上の方程式における使用のために最適であることを示す。このカットオフを用いて、短いDNA分子は、欠失していないハプロタイプ由来の集団中で、欠失したハプロタイプ由来の集団中と比較して、20%より豊富である。次いで、このパーセンテージ差(または同様に由来した値)は、癌を呈さない個体に由来する閾値と比較することができる。

【0204】

染色体の増幅を伴う領域において、1ハプロタイプを腫瘍組織中で複製した(増幅したハプロタイプ)。この増幅したハプロタイプ由来の腫瘍由来の短いDNA分子の余剰量が血漿中へ放出されるため、増幅したハプロタイプ由来の断片サイズ分布は、増幅されてい

10

20

30

40

50

ないハプロタイプ由来の断片におけるサイズ分布より短い。LOHシナリオに類似して、サイズ分布差は、断片の累積頻度をDNA断片サイズに対してプロットすることによって決定することができる。短いサイズ分布のDNA集団は、短いDNAをより豊富に有することになり、それゆえ、サイズスペクトルの短い末端で累積頻度がより急速に増大する。

【0205】

図27は、本発明の実施形態による、腫瘍組織中の染色体の複製を伴う領域サイズに対する血漿DNAの累積頻度プロットを示す。増幅したハプロタイプ由来の配列は、増幅されていないハプロタイプ由来の配列と比較すると、より急速に増大し、サイズ170bp未満の累積頻度がより高い。これは、増幅したハプロタイプ由来の短いDNA断片がより豊富であったことを示す。これは、腫瘍由来の短いDNAのより多数が増幅したハプロタイプに由来したため、以下に示す予測と一致する。

10

【0206】

LOHシナリオに類似して、サイズ分布差は、2つのDNA分子集団の累積頻度差によって定量化することができる。我々は、 Q を2つの集団の累積頻度差として定義する。

$Q = Q_{\text{増幅}} - Q_{\text{非増幅}}$ （式中、 $Q_{\text{増幅}}$ は、増幅したハプロタイプ由来の配列したDNA断片の累積頻度を示し、かつ $Q_{\text{非増幅}}$ は、増幅されていないハプロタイプ由来の配列したDNA断片の累積頻度を示す）。

【0207】

図28は、本発明の実施形態による、増幅した領域における配列した血漿DNAサイズに対する Q を示す。 Q は、本発明の実施形態による、126bpサイズで、0.08に達する。これは、短いDNAを決定するカットオフとして126bpを使用し、短いDNA分子は、増幅したハプロタイプ集団中で増幅されていないハプロタイプからの集団と比較して、8%より豊富であることを示す。

20

【0208】

D. 追加の技術

他の実施形態では、配列特異性技術を使用し得る。例えば、オリゴヌクレオチドは、特定領域の断片のハイブリッド形成に設計し得る。次いで、オリゴヌクレオチドは、配列したタグ数と類似した形で計数することができる。この方法は、特定の異常を示す癌に使用し得る。

【0209】

VIII. コンピュータシステム

本明細書に記載の任意のコンピュータシステムは、任意の適切な数のサブシステムを使用し得る。かかるサブシステムの例を、図9でコンピュータ装置900に示す。いくつかの実施形態では、コンピュータシステムは、サブシステムをコンピュータ装置の構成部分とすることができる単独コンピュータ装置を含む。他の実施形態では、コンピュータシステムは、内部構成部分と共にそれぞれサブシステムである複数のコンピュータ装置を含むことができる。

30

【0210】

図29に示されたサブシステムは、システムバス2975を通じて相互接続される。さらなるサブシステム、例えばプリンター2974、キーボード2978、固定ディスク2979、ディスプレイアダプター2982と接続されたモニター2976等が示される。I/Oコントローラー2971と接続する、周辺機器および入力/出力(I/O)装置は、当技術分野において既知である任意の数の手段、例えばシリアルポート2977によってコンピュータシステムと接続することができる。例えば、シリアルポート2977または外部インターフェース2981(例えばイーサネット、ワイファイ等)は、コンピュータシステム2900をインターネットのような広い領域ネットワーク、マウス入力装置、またはスキャナーと接続するために使用することができる。システムバス2975経由の相互接続によって、セントラルプロセッサ2973を個々のサブシステムと通信させ、そしてシステムメモリー2972または固定ディスク2979からの指示の実行、およびサブシステム間の情報の交換をさせることが可能である。システムメモリー2972および

40

50

／または固定ディスク 2979 は、コンピュータ読み取り可能媒体を具現化し得る。本明細書に記載のいずれの値も、1 構成部分から別の構成部分に出力することができ、ユーザーに提供することができる。

【0211】

コンピュータシステムは、例えば、外部インターフェース 2981 によって、または内部インターフェースによって共に接続されている複数の同一構成部分またはサブシステムを含むことができる。いくつかの実施形態では、コンピュータシステム、サブシステム、または装置は、ネットワーク上で伝達できる。かかる例では、1つのコンピュータはクライアントとみなすことができ、別のコンピュータはサーバーとみなすことができ、それぞれ同一コンピュータシステムの一部とすることができる。クライアントおよびサーバーは、それぞれ、複数のシステム、サブシステム、または構成部分を含むことができる。

10

【0212】

本発明の任意の実施形態は、モジュール式または統合した様式で、ハードウェアを用いて、および／またはコンピュータソフトを用いて、制御ロジック式に実行できることが理解されるべきである。本明細書に提供する本開示および教示に基づき、ハードウェアおよびハードウェアとソフトウェアの組み合わせを用いて本発明の実施形態を実行する他のやり方 (ways) および／または方法 (methods) について当業者は知っており理解しているであろう。

【0213】

本用途に記載される任意のソフトウェア・コンポーネントまたはソフトウェア機能は、任意の適切なコンピュータ言語、例えば Java、C++ または Perl 等を使用して、例えば従来の、またはオブジェクト指向の技術を使用して処理装置によって実行される、ソフトウェアコードとして実装されてもよい。ソフトウェアコードは、一連の指示またはコマンドとして、記憶および／または伝達のためのコンピュータ読み取り可能媒体に記憶されてもよく、適切な媒体としては、ランダムアクセスメモリー (RAM)、読み取り専用メモリー (ROM)、ハードドライブもしく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) もしくは DVD (デジタル多用途ディスク) などの光学媒体、フラッシュメモリー等が挙げられる。コンピュータ読み取り可能媒体は、かかる記憶装置または伝達装置の任意の組み合わせであり得る。

20

【0214】

かかるプログラムはまた、コード化され、インターネットを含む種々のプロトコールに従う有線、光、および／または無線ネットワークを経由した伝達のために適合されるキャリア信号を使用して伝達され得る。したがって、本発明の実施形態のコンピュータ読み取り可能媒体は、かかるプログラムでコード化されたデータ信号を使用して作製し得る。プログラムコードでコード化されたコンピュータ読み取り可能媒体は、互換性のある装置と共にパッケージ化されてもよいし、他の装置とは別に (例えばインターネットダウンロードを通じて) 提供されてもよい。任意のかかるコンピュータ読み取り可能媒体は、単一のプログラム製品 (例えばハードドライブ、CD、または全コンピュータシステム) 上にまたはその中に装備されてもよく、システムまたはネットワーク内における異なるコンピュータプログラム製品上またはその中に存在してもよい。コンピュータシステムは、ユーザーに本明細書において言及された任意の結果を提供するためのモニター、プリンター、または他の好適なディスプレイを含んでもよい。

30

40

【0215】

本明細書に記載の方法はいずれも、工程を実施するように構成することができる処理装置を含むコンピュータシステムを用いて、全体として実施してもよいし部分的に実施してもよい。したがって、実施形態は、本明細書に記載の任意の方法の工程を、潜在的に各工程または工程の各群を実施する異なる構成部分と共に実施するように構成されたコンピュータシステムに關することができ、番号付けした工程を提示するが、本明細書の方法の工程は、同時に実施しても異なる順で実施してもよい。加えて、これらの工程の一部は、他の方法からの他の工程の一部とともに使用してもよい。また、工程のすべてまたは一部

50

は任意選択的であり得る。加えて、モジュール、回路、またはこれらの工程を実施する他の手段を用いて任意の方法の任意の工程を実施することができる。

【0216】

特定の実施形態の具体的な詳細は、本発明の実施形態の真意および範囲から逸脱することなく任意の適切な形で併合し得る。しかしながら、本発明の他の実施形態は、それぞれ個々の態様、またはこれら個々の態様の具体的な組み合わせに関する具体的な実施形態に関し得る。

【0217】

上述の本発明の例示的な実施形態の記述は、例証および説明目的で提示している。記載した厳密な形に本発明を包括または限定することは意図しておらず、多くの改変および変形が上の教示に照らして可能である。本発明の原理およびその実用的な応用を最適に説明し、それにより企図される特定の使用に適するように様々な改変を伴い様々な実施形態で本発明を他の当業者が最適に使用することを可能にするために、実施形態を選択および記載した。

【0218】

「a」、「an」または「the」の列挙は、これに反すると別段の記載のない限り、「1つまたは複数」を意味するものとする。上述したすべての特許、特許出願、刊行物、および上述の説明は、すべての目的で参照することによってそれらの全体が本明細書に組み込まれる。先行技術として認められるものはない。

10

【図1】

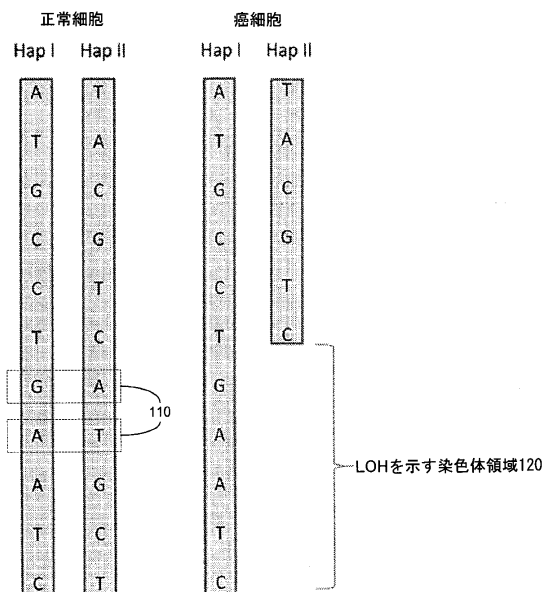


FIG. 1

【図2】

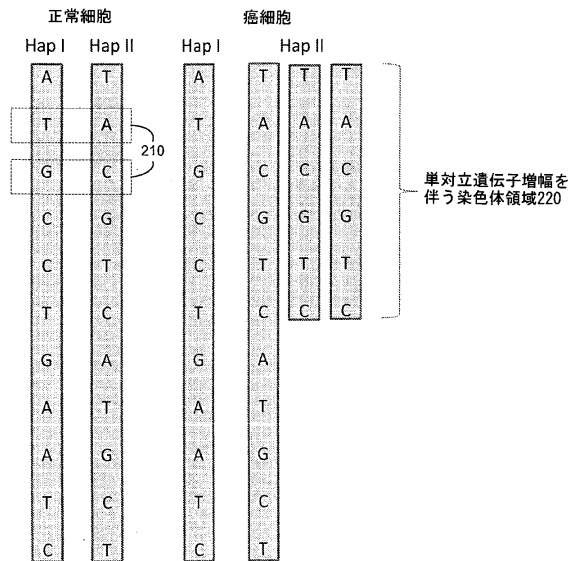


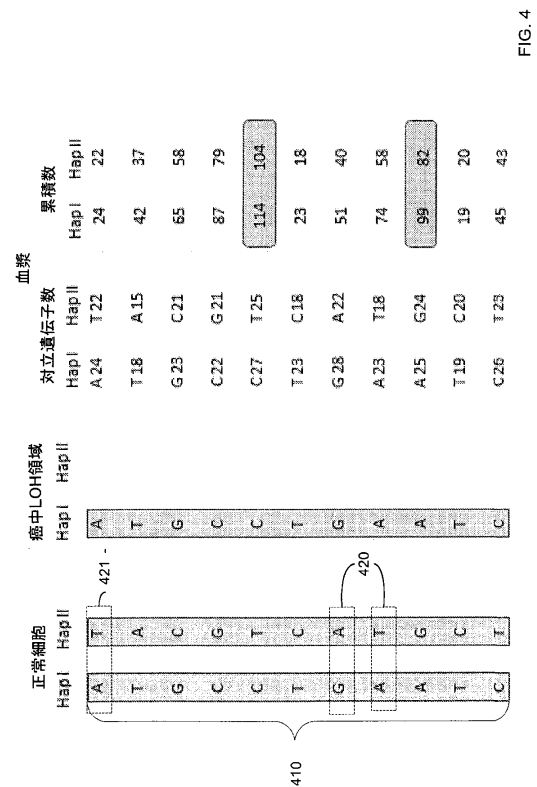
FIG. 2

【図3】

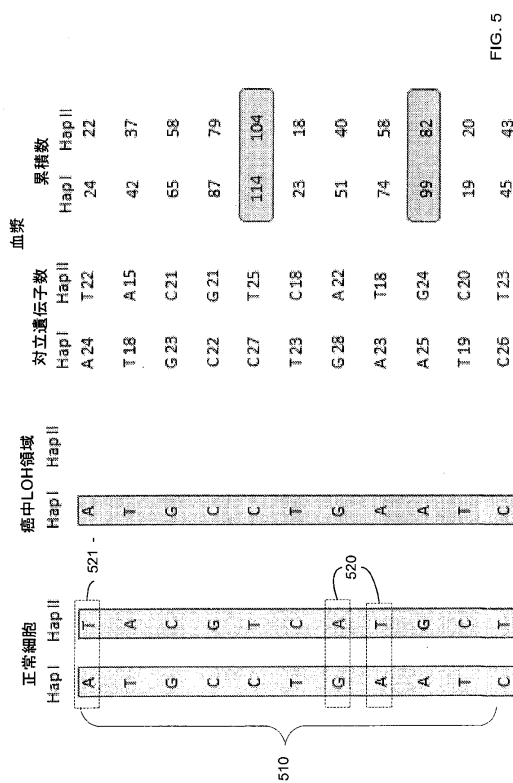
310	320	330	340
癌タイプ	獲得	喪失	参考文献
甲状腺癌	7p, 8q, および 9q	22	Hemmer et al. Am J Pathol, 1999;154:1539-47.
胃癌	2q37, 3p21, 5q34-35, 7q34-36, 11q13, 11q23-25, 12q24, 15q23-25, 17q21-25, および 20q12-13	4q, 13q, 5q, 6q, および 18p	Noguchi et al. Am J Clin Pathol 2001; 115:828-34
前立腺癌	1q32, 3q26, 4q26, 7q21, 8q22, 9q33, 17q25, および Xq21	2q22, 4q27-4q28, 5q15, 6q15, 8p21, 10q23, 12p13, 13q21, 15q23, 16q22, および 18q21-22	Sun et al. Prostate 2007;67:692-700.
小細胞肺癌	3q26-29, 5p12-13, および 8q23-24	3p13-14, 4q32-35, 5q32-35, 8p21-22, 10q25, 13q13-14, および 17p12-13	Balsara et al. Oncogene 2002;21:6877-83.
非小細胞肺癌	1q31, 3q25-27, 5p13-14, および 8q23-24	3p21, 8p22, 9p21-22, 13q22, および 17p12-13	Balsara et al. Oncogene 2002;21:6877-83.
鼻咽頭癌	1p34, 3q26, 6q25, および 3q26	3p, 9p, 9q, 11q, 13q, および 14q	Lo et al. Semin Cancer Biol 2002;12:451-62
膀胱癌	1q, 5p, 6p, 8q, 11q, 17q, および 20q	3p, 4q, 4q, 6q, 8p, 9p, および 18q	El-Rifai et al. Am J Pathol 2000;156:871-8
結腸直腸癌	13q および 20q	4q および 18q	De Angelis et al. Int J Colorectal Dis 2001;16:38-45.
頭頸部癌	3q26 および 11q13	3p, 9p, および 17p	Smeets et al. Oncogene 2006;25:2558-64
黒色腫	1q, 2, 6p, 7, 8, 17, および 20	6q, 8p, 9, および 10	Bastian et al. Cancer Res 1998;58:2170-5.
リンパ腫	1q, 3, 6p, 7, 11, 12, 18, および X	1p, 8p, および X	Monni et al. Blood 1996;87:5269-78

FIG. 3

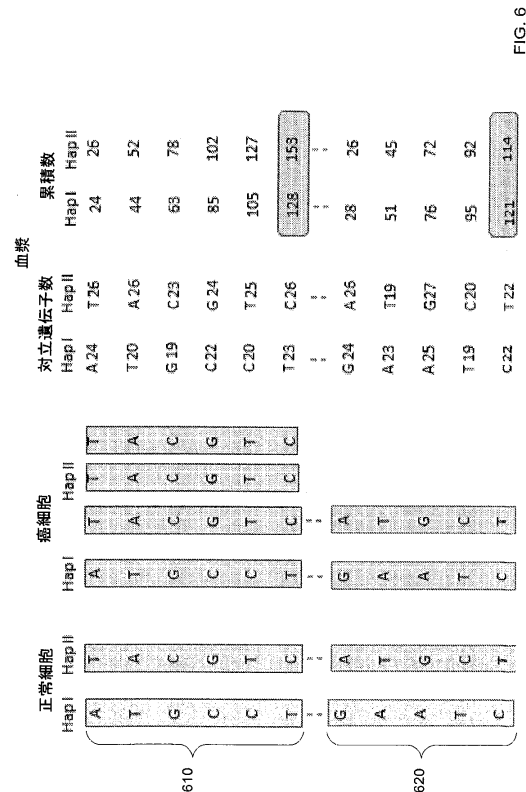
【図4】



【図5】



【図6】



【図 7】

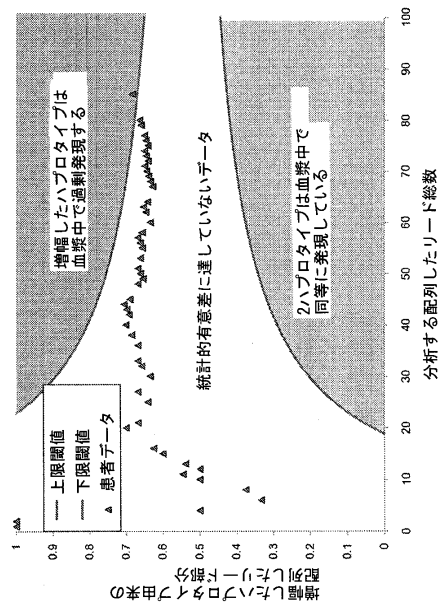


FIG. 7

【図 8】

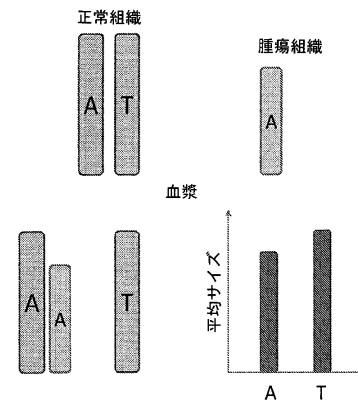


FIG. 8

【図 9】

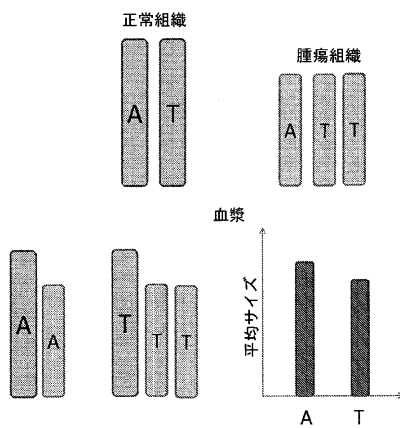


FIG. 9

【図 10】

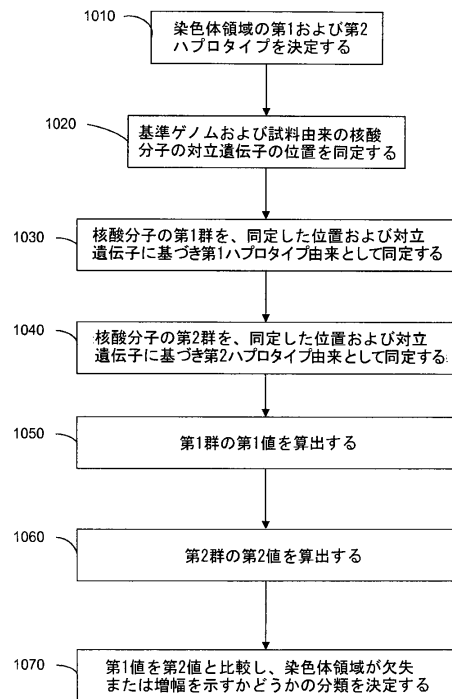


FIG. 10

【図 1 1】

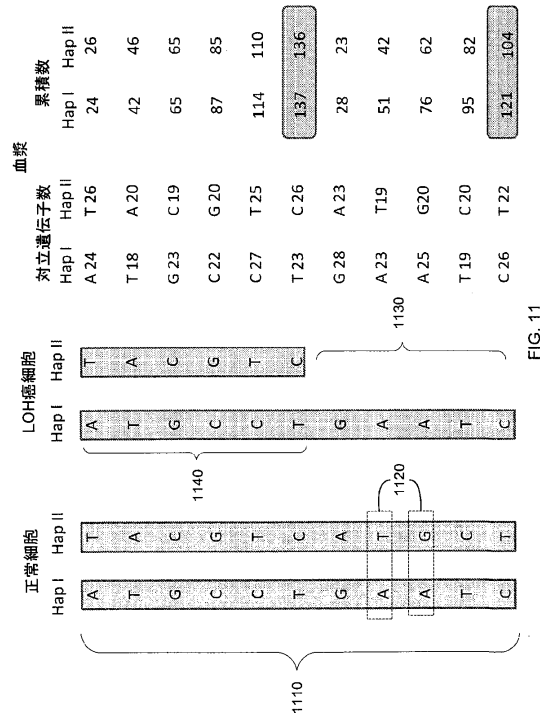


FIG. 11

【図 1 2】

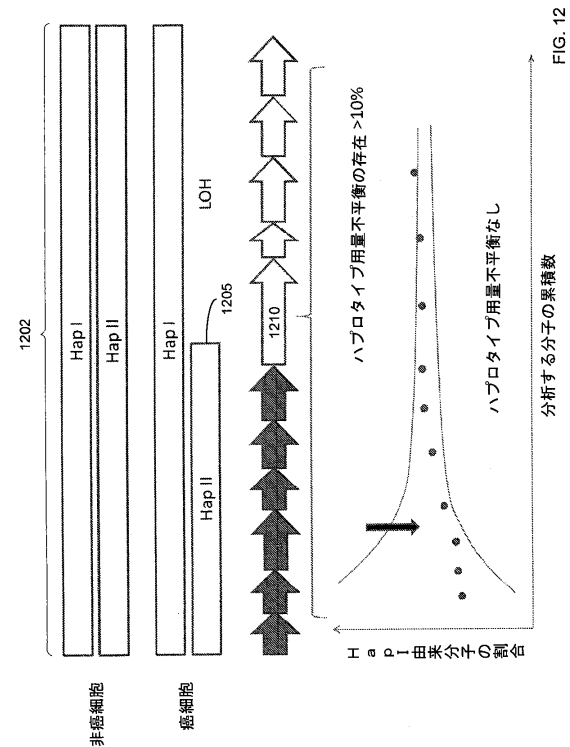


FIG. 12

【図 1 3】

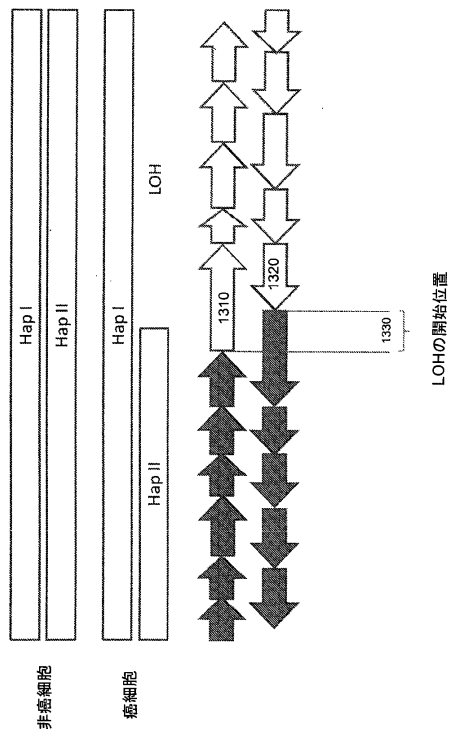


FIG. 13

【図 1 4】

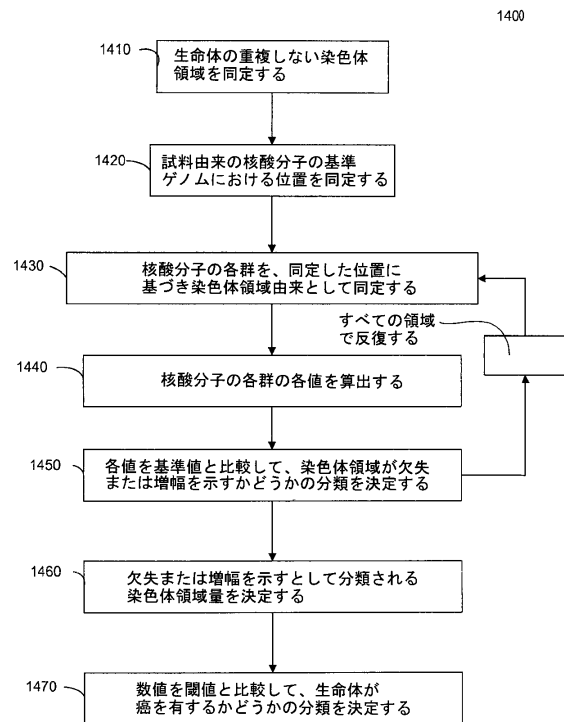


FIG. 14

【図 15】

癌由来DNAの 分画濃度 1510	断片ごとに 必要な推定 分子数 1520	各断片 サイズ (kb) 1530	完全ゲノム に対して 分析する 総分子数 (百万) 1540
50%	950	100	28.5
		1,000	2.85
		5,000	0.570
		10,000	0.285
25%	3,800	100	114
		1,000	11.4
		5,000	2.28
		10,000	1.14
12.5%	15,000	100	450
		1,000	45
		5,000	9
		10,000	4.5
6.3%	60,000	100	1,800
		1,000	180
		5,000	36
		10,000	18
3.2%	240,000	100	7,200
		1,000	720
		5,000	144
		10,000	72

FIG. 15

【図 16】

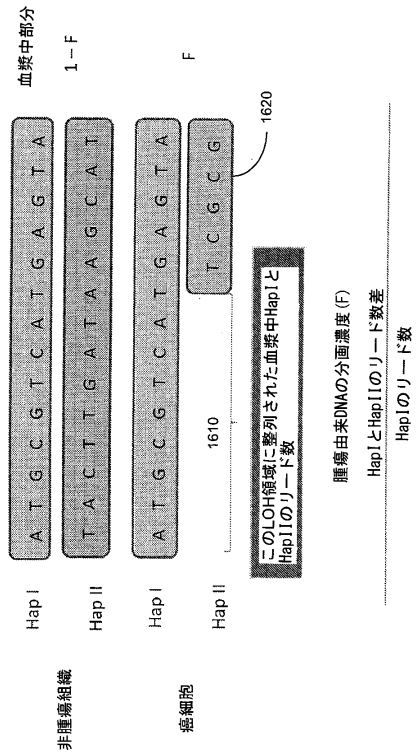


FIG. 16

【図 17】

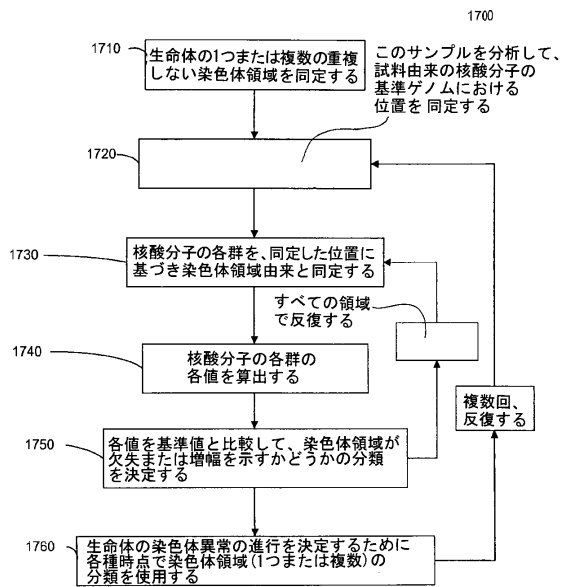


FIG. 17

【図 18 A】

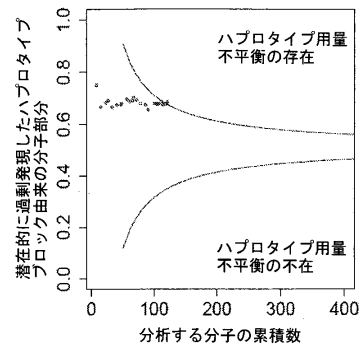


FIG. 18A

【図 18 B】

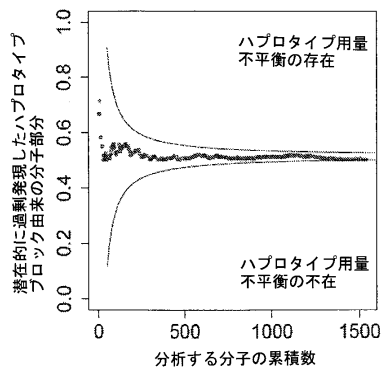
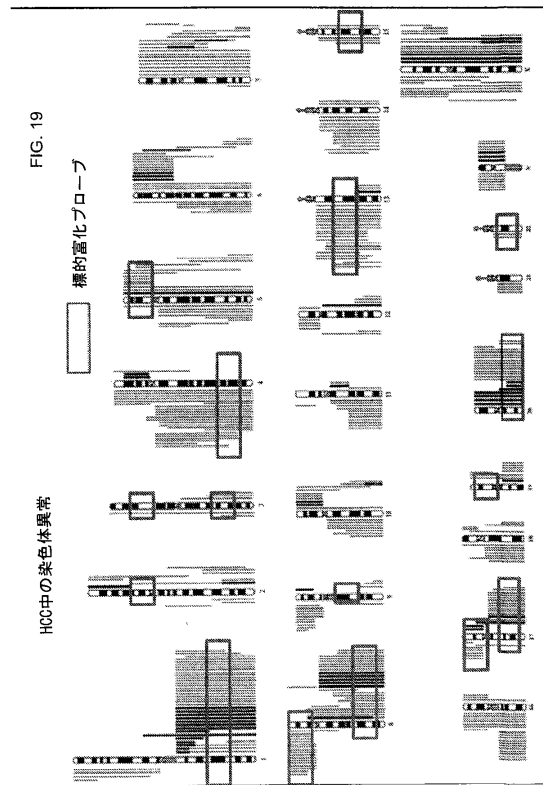


FIG. 18B

【図 19】



【図 20 A】

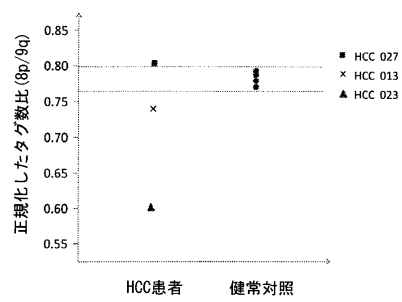


FIG. 20A

【図 21】

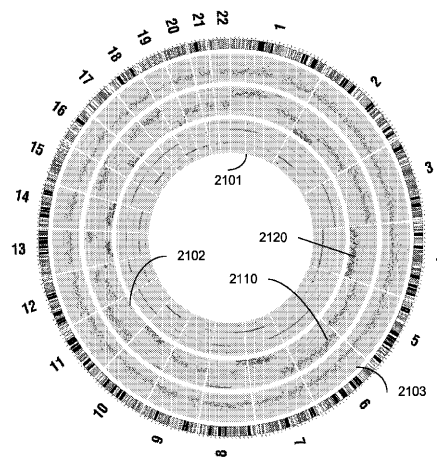


FIG. 21

【図 20 B】

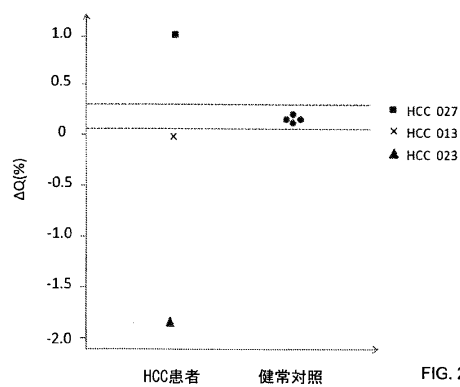


FIG. 20B

【図 22】

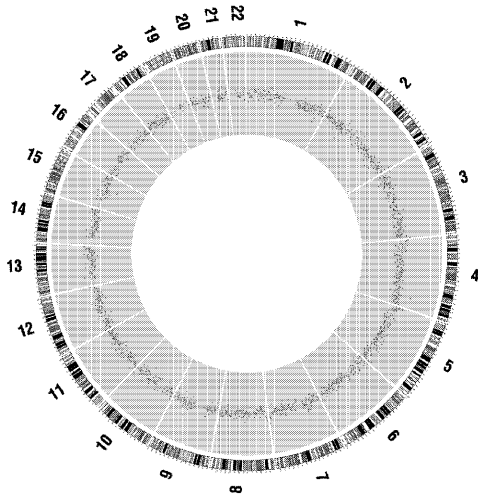


FIG. 22

【図 23】

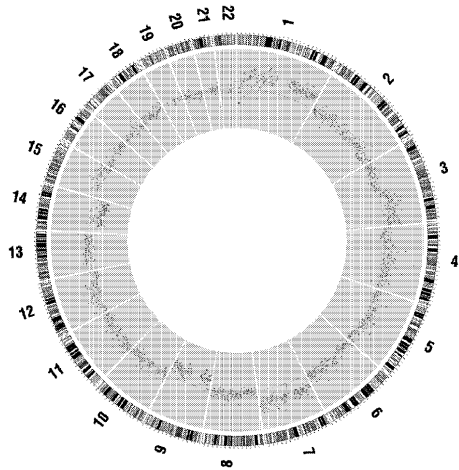


FIG. 23

【図 24】

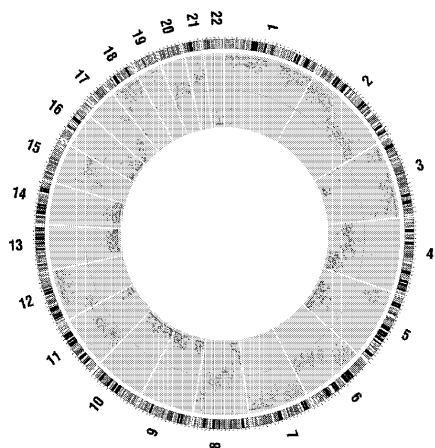


FIG. 24

【図 25】

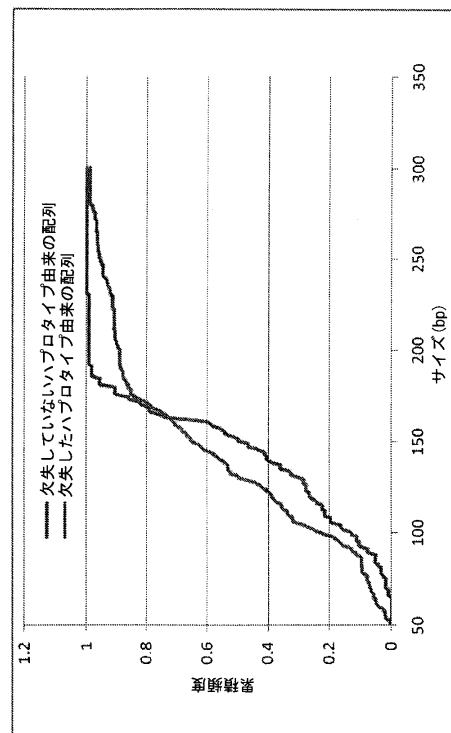


FIG. 25

【図 26】

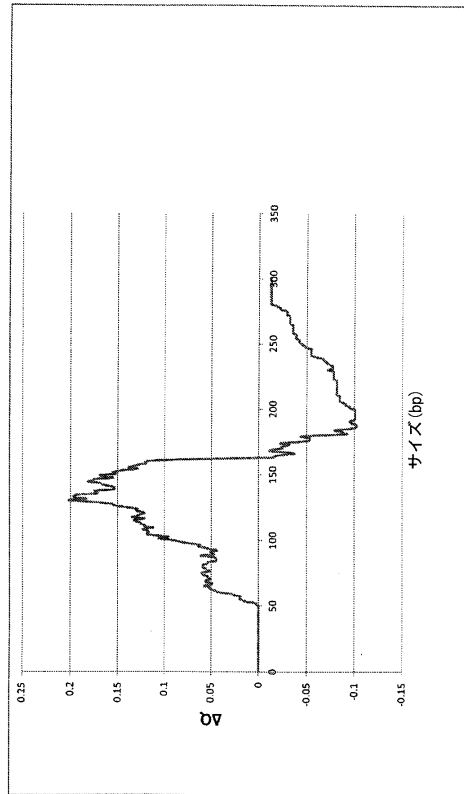


FIG. 26

【図 27】

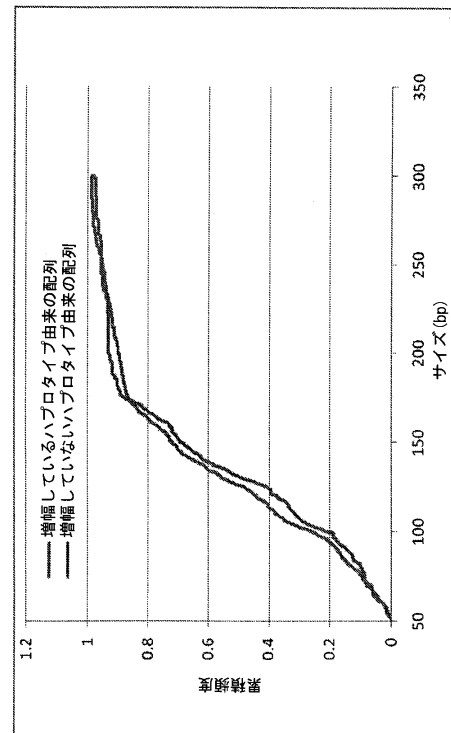


FIG. 27

【図 28】

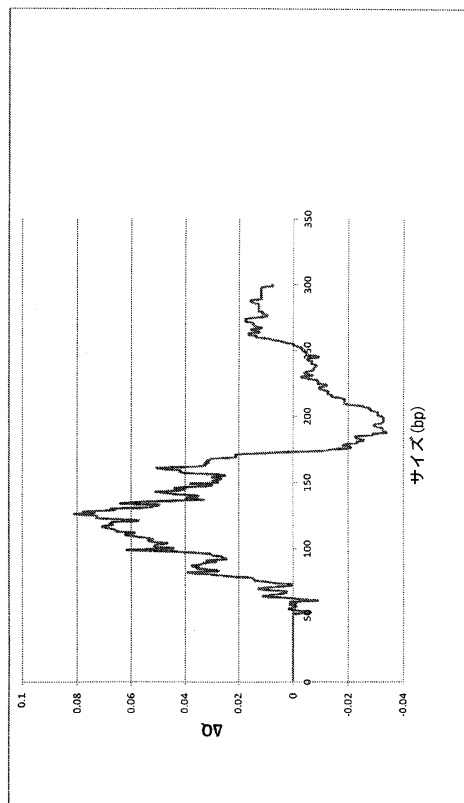


FIG. 28

【図 29】

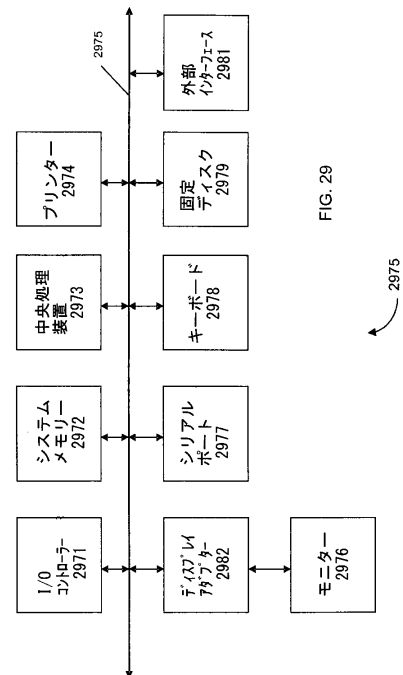


FIG. 29

【配列表】

0006196157000001.app

フロントページの続き

- (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 ユク ミン デニス ロー
中華人民共和国, ホンコン, カウルーン, ホーマンティン, キング タク ストリート 7, フォ
ース フロア
- (72)発明者 チャン クワン チー
中華人民共和国, ホンコン, カウルーン, メイ フー スン チュン, 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 ワイ クン ロッサ チュー
中華人民共和国, ホンコン, ニューテリトリーズ, シャーティン, ダブル ヘブン 52 マロッ
ク パス, ハウス 31
- (72)発明者 チャン ペイヨン
中華人民共和国, ホンコン, ニューテリトリーズ, シャーティン, ガン シン ストリート, 30
- 32, プリンス オブ ウェールズ ホスピタル, スタッ クォーターズ, フラット 110
6 ブロック エー

審査官 西 賢二

- (56)参考文献 特表2010-534068(JP, A)
国際公開第98/008980(WO, A1)
Frank DIEHL, PNAS, 2005年, Vol.102, No.45, Pages 16368-16373
Bret TABACK, Cancer Research, 2001年, Vol.61, Pages 5723-5726
Imke MULLER, Clinical Chemistry, 2008年, Vol.54, No.4, Pages 688-696

(58)調査した分野(Int.Cl., DB名)

C12Q 1/00 - 3/00
C12N 15/00 - 15/90
G01N 33/48 - 33/98
CAplus / MEDLINE / EMBASE / BIOSIS (STN)