

(19)日本国特許庁(JP)

## (12)特許公報(B2)

(11)特許番号  
特許第7679306号  
(P7679306)

(45)発行日 令和7年5月19日(2025.5.19)

(24)登録日 令和7年5月9日(2025.5.9)

(51)国際特許分類

C 1 2 Q	1/6869(2018.01)	C 1 2 Q	1/6869	Z Z N A
C 1 2 Q	1/6827(2018.01)	C 1 2 Q	1/6827	Z
C 1 2 Q	1/6876(2018.01)	C 1 2 Q	1/6876	Z
C 1 2 Q	1/6837(2018.01)	C 1 2 Q	1/6837	Z
C 1 2 Q	1/6888(2018.01)	C 1 2 Q	1/6888	Z

請求項の数 16 (全55頁) 最終頁に続く

(21)出願番号 特願2021-555073(P2021-555073)  
 (86)(22)出願日 令和2年3月12日(2020.3.12)  
 (65)公表番号 特表2022-524208(P2022-524208  
 A)  
 (43)公表日 令和4年4月28日(2022.4.28)  
 (86)国際出願番号 PCT/CN2020/079067  
 (87)国際公開番号 WO2020/182193  
 (87)国際公開日 令和2年9月17日(2020.9.17)  
 審査請求日 令和5年2月24日(2023.2.24)  
 (31)優先権主張番号 PCT/CN2019/077750  
 (32)優先日 平成31年3月12日(2019.3.12)  
 (33)優先権主張国・地域又は機関  
 中国(CN)

(73)特許権者 521412504  
 クラウン バイオサイエンス (スーチョウ  
 ウ) インコーポレイテッド  
 中華人民共和国 215000 ジанс  
 ー, スーチョウ, インダストリアル パ  
 ーク, シンフー ストリート ナンバー 2  
 18, ビルディング エ-6, ルーム 3  
 03  
 (74)代理人 100149076  
 弁理士 梅田 慎介  
 (74)代理人 100119183  
 弁理士 松任谷 優子  
 (74)代理人 100173185  
 弁理士 森田 裕  
 (74)代理人 100162503

最終頁に続く

(54)【発明の名称】 腫瘍モデルの同定のための方法および組成物

## (57)【特許請求の範囲】

## 【請求項1】

ヒトまたはマウス試料を認証するための方法であって、

前記ヒトまたはマウス試料から核酸を得ること；

(i)前記ヒト試料の、配列番号1～237で示されるヒト単一ヌクレオチド多型(SNP)

)遺伝子座の全てにおける遺伝子型、または(ii)前記マウス試料の、配列番号238～

436で示されるマウスSNP遺伝子座の全てにおける遺伝子型を検出すること；

(i)前記ヒト試料の遺伝子型を、既知の正体を有するヒト参照試料のライブラリー中の、

配列番号1～237で示されるヒトSNPの全てについての参照遺伝子型と比較すること

、または(ii)前記マウス試料の遺伝子型を、既知の正体を有するマウス参照試料ライ

ブリー中の、配列番号238～436で示されるマウスSNPの全てについての参照遺伝

子型と比較すること；および

(i)前記ヒト試料の認証を、前記ヒト参照試料のライブラリー中の前記ヒト参照試料の遺

伝子型との類似性によって決定すること、または(ii)前記マウス試料の認証を、前記マ

ウス参照試料ライブラリー中の前記マウス参照試料の遺伝子型との類似性によって決定す

ること

を含む、方法。

## 【請求項2】

前記ヒトまたはマウス試料が、細胞、組織、オルガノイド、またはそれらの組合せである、請求項1に記載の方法。

**【請求項 3】**

前記ヒトまたはマウス試料が、細胞株または腫瘍組織である、請求項 2 に記載の方法。

**【請求項 4】**

前記ヒトまたはマウス試料が夾雜物を含み、前記方法が、前記試料中の夾雜物のパーセンテージを決定することをさらに含む、請求項 1 に記載の方法。

**【請求項 5】**

前記ヒト試料が夾雜物を含み、前記方法が、夾雜物の正体を決定することをさらに含む、請求項 1 に記載の方法。

**【請求項 6】**

前記検出するステップが、次世代シークエンシング (NGS) を使用する、請求項 1 に記載の方法。 10

**【請求項 7】**

前記核酸がバーコード化される、請求項 1 に記載の方法。

**【請求項 8】**

表 3 に示される群から選択される性染色体 SNP を検出することによって、前記ヒトまたはマウス試料を得た対象の性別を同定することをさらに含む、請求項 1 に記載の方法。

**【請求項 9】**

前記ヒト試料を得た対象の民族性を同定することをさらに含む、請求項 1 に記載の方法。

**【請求項 10】**

前記ヒトまたはマウス試料中のウイルスまたはマイコプラズマの存在を検出することをさらに含む、請求項 1 に記載の方法 (ただし、ヒトの診断方法を除く。)。 20

**【請求項 11】**

前記マウス試料が、4T1、A20、B16-BL6、B16-F0、B16-F1、B16-F10、C1498、Colon26、CT26WT、E.G7-Ova、EL4、EMT6、H22、Hepa1-6、J558、J774A1、JC、KLN205、L1210、L5178-R、LLC、MBT2、MC38、MPC-11、Neuro-2a、P388D1、P815、Pan02、Renca、RM1、S91およびWEHI164からなる群から選択されるマウス腫瘍モデルの試料である、請求項 1 に記載の方法。

**【請求項 12】**

ヒトまたはマウス試料を認証するためのキットであって、 30

(i) ヒト試料中の配列番号 1 ~ 237 で示されるヒト SNP の全て、または (ii) マウス試料中の配列番号 238 ~ 436 で示されるマウス SNP の全てを検出するためのプライマー；および

ヒトまたはマウス SNP を含有する DNA 断片を、前記プライマーを使用して増幅するための薬剤

を含む、キット。

**【請求項 13】**

表 3 に示される群から選択される性染色体 SNP を検出するためのプライマーをさらに含む、請求項 12 に記載のキット。

**【請求項 14】**

前記試料中のウイルス感染またはマイコプラズマ夾雜を検出するためのプライマーをさらに含む、請求項 13 に記載のキット。

**【請求項 15】**

ヒトまたはマウス試料を同定するためのマイクロアレイであって、

(i) ヒト試料中の配列番号 1 ~ 237 で示されるヒト SNP の全て、または (ii) マウス試料中の配列番号 238 ~ 436 で示されるマウス SNP の全てを検出するためのプローブを含む、マイクロアレイ。

**【請求項 16】**

指示が記憶された非一過性のコンピュータ可読媒体であって、前記指示がプロセッサに

10

20

30

40

50

よって実行されるとき、前記指示がプロセッサに、

(i) ヒト試料中の配列番号 1 ~ 2 3 7 で示されるヒト S N P 遺伝子座の全て、または(ii)マウス試料中の配列番号 2 3 8 ~ 4 3 6 で示されるマウス S N P 遺伝子座の全てにおける遺伝子型の検索；

(i) 前記ヒト試料の遺伝子型の、既知の正体を有するヒト参照試料のライブラリー中の、配列番号 1 ~ 2 3 7 で示される前記ヒト S N P の全てについてのヒト参照遺伝子型との比較、または(ii) 前記マウス試料の遺伝子型の、既知の正体を有するマウス参照試料のライブラリー中の、配列番号 2 3 8 ~ 4 3 6 で示される前記マウス S N P の全てについてのマウス参照遺伝子型との比較；および

(i) 前記ヒト試料の認証の、前記ヒト参照試料のライブラリー中の前記ヒト参照試料の遺伝子型との類似性による決定、または(ii) 前記マウス試料の認証の、前記マウス参照試料ライブラリー中の前記マウス参照試料の遺伝子型との類似性による決定

を行わせる、コンピュータ可読媒体。

【発明の詳細な説明】

【技術分野】

【0001】

関連出願の相互参照

本出願は、2019年3月12日出願の出願 P C T / C N 2 0 1 9 / 0 7 7 7 5 0 号に基づく優先権を主張するものであり、その開示を本明細書に参考として組み込む。

【0002】

本発明は全般に、分子生物学、がん生物学および動物モデルに関する。

【背景技術】

【0003】

細胞株、オルガノイド、異種移植および同種移植モデルは、腫瘍学およびその他の生物医学的研究において有用なモデルシステムである。モデルの認証および特徴づけは、それらの適切な利用を助け、誤同定および誤用、交差夾雜、誤ったがん分類、長期培養および遺伝的浮動に起因するゲノム変化などの一連の問題を軽減するが、これらは全て、それらの一般的な使用に起因して、細胞株において特に十分に注目されたものである。例えば、様々な研究が、細胞株バンクについて、約 10 ~ 40 % の誤同定 / 夾雜率を報告している。

【0004】

細胞形態学試験、イソ酵素学 [isoenzymology] 、細胞遺伝学的分析（核型分析および F I S H ）、ヒトリンパ球抗原 ( H L A ) タイプ分け、ショートタンデムリピート ( S T R ) プロファイリング、单一ヌクレオチド多型 ( S N P ) タイプ分け、D N A および R N A シークエンシングを含む、細胞株を認証するための様々な方法が存在している ( Freedman, L. P. et al. Biotechniques 59, 189 - 90, 192 ( 2015 ) ; Nims, R. W. & Reid, Y. In Vitro Cell Dev Biol Anim 53, 880 - 887 ( 2017 ) )。これらの技術のうち、 S T R プロファイリングは最も広く使用されており、ヒト細胞株を認証する際のその適用を指南する基準 ( A S N - 0 0 0 2 ) が存在する ( Almeida, J. L. , Cole, K. D. & Plant, A. L. P L o S Biol 14, e 1 0 0 2 4 7 6 ( 2 0 1 6 ) )。マウス細胞株については、19の S T R マーカーのパネルも開発されている ( Zaaijer, S. et al. Elife 6 ( 2 0 1 7 ) )。夾雜物を検出するための S T R アッセイの感度は、約 5 ~ 10 % である ( Yu, M. et al. Nature 520, 307 - 11 ( 2 0 1 5 ) )。近年、 S N P タイプ分けは、その改善された精度、感度および低減されたコストのおかげで、細胞株および生体試料認証のためにますます使用されるようになってきている。 S N P は、 P C R 、ならびにトランスクリプトームシークエンシングまたは R N A - seq 、全エキソームシークエンシング ( W E S ) および全ゲノムシークエンシング ( W G S ) を含む次世代シークエンシング ( N G S ) によってプロファイリングすることができる。現行の S N P アッセイは、約 3 ~ 5 % の検出感度を有する。それらの認証および特徴づけを促進する、 S T R 、 S N P および細胞株についての

10

20

30

40

50

その他の情報を含むデータベースも存在する。

#### 【0005】

細胞株に加えて、オルガノイドおよびマウス腫瘍モデルが、腫瘍学研究および薬物開発において広く使用されている。オルガノイドは、幹細胞、初代および操作された腫瘍試料、ならびに多くの生物構造および機能を維持する異種移植されたヒト腫瘍に由来する *in vitro* 三次元培養物である。マウス腫瘍モデルは、患者由来異種移植 (PDX) 、細胞株由来異種移植 (CDX) 、同系またはマウス細胞株由来モデル、マウス同種移植モデルなどを含む *in vivo* システムである。PDXなど、これらのモデルの一部は、細胞株よりも、原発性腫瘍についての病理組織学およびゲノミクスを忠実に捕捉することができる。細胞株と同様、これらの腫瘍モデルには、類似の品質管理上の問題があるが、さらなる問題が存在する。異種移植モデルでは、腫瘍は、ヒト腫瘍細胞およびマウス間質細胞を含有し、後者は、モデルの継代の間にヒトの対応物を徐々に置き換え、これがゲノム不均一性、移植部位の違い (皮下および同所) 、増殖の変動および切開のランダムさと複合されると、腫瘍のヒト - マウス遺伝子組成は、同じ PDX からのものであっても、一部の試料がほとんど純粋なヒト内容またはマウス内容である程度まで、かなり異なることになる。このような腫瘍 - 宿主混合および干渉は、全ての移植された腫瘍モデルに生じ、STRマークーおよびSNPについての対立遺伝子頻度のゆらぎを引き起こし、したがって、従来のSTRおよびSNPベースの認証方法に有害な影響を与える。大規模試料認証はまた、多くの種類の *in vitro* および *in vivo* モデルが同時に維持され使用されるバイオバンクについては特に、ロジスティック上の負荷があり、エラーブローンである。したがって、腫瘍モデルを同定および認証するための新たなSNPベースのアッセイを開発する必要がある。

10

#### 【発明の概要】

#### 【0006】

一態様では、本開示は、試料を同定または認証するための方法を提供する。一実施形態では、この方法は、試料から核酸を得ること；複数のヒト単一ヌクレオチド多型 (SNP) 遺伝子座または複数のマウス SNP 遺伝子座において、試料について遺伝子型を検出すること；試料についての遺伝子型を参照遺伝子型と比較すること；および試料の同定を決定することを含む。ある種の実施形態では、ヒト SNP は、表 1 に示される群から選択される。ある種の実施形態では、マウス SNP は、表 2 に示される群から選択される。

20

#### 【0007】

ある種の実施形態では、試料は、細胞、組織、オルガノイド、またはそれらの組合せである。ある種の実施形態では、試料は、細胞株または腫瘍組織である。ある種の実施形態では、試料は、異種移植または同種移植腫瘍モデルに由来する。ある種の実施形態では、試料は、患者由来異種移植 (PDX) 、細胞株由来異種移植 (CDX) 、同系またはマウス細胞株由来モデル、マウス同種移植モデルに由来する。

30

#### 【0008】

ある種の実施形態では、試料は夾雜物を含み、方法は、試料中の夾雜物のパーセンテージを決定することをさらに含む。ある種の実施形態では、この方法は、夾雜物の正体を決定することをさらに含む。

40

#### 【0009】

ある種の実施形態では、検出するステップは、次世代シーケンシング (NGS) またはシーケンシングベースの SNP アレイを使用する。ある種の実施形態では、核酸は、バーコード化される。

#### 【0010】

ある種の実施形態では、この方法は、試料を得た対象の性別を同定することをさらに含む。ある種の実施形態では、この方法は、試料を得た対象の民族性を同定することをさらに含む。ある種の実施形態では、この方法は、試料中のウイルスまたはマイコプラズマの存在を検出することをさらに含む。ある種の実施形態では、この方法は、試料を得た免疫不全マウスの系統を決定することをさらに含む。

50

## 【0011】

別の態様では、本開示は、ヒト構成成分およびマウス構成成分を含む試料を認証する方法を提供する。ある種の実施形態では、この方法は、試料から核酸を得ること；100個以上のマウスゲノム遺伝子座において、試料の遺伝子型を検出すること、ここで、マウスゲノム遺伝子座の各々が、対応する相同的なヒトゲノム遺伝子座を有し、マウスゲノム遺伝子座の各々および対応する相同的なヒトゲノム遺伝子座が、同一な隣接ヌクレオチド配列を有する；ならびに遺伝子型に基づいて、試料中のマウス構成成分の比率を決定すること、を含む。ある種の実施形態では、マウスゲノム遺伝子座は、表6から選択される。

## 【0012】

別の態様では、本開示は、試料を同定するためのキットを提供する。ある種の実施形態では、このキットは、試料中の一群のヒトS N P遺伝子座または一群のマウスS N P遺伝子座を検出するためのプライマーを含む。ある種の実施形態では、このキットは、ヒトまたはマウスS N Pを含有するD N A断片を、これらのプライマーを使用して増幅するための薬剤をさらに含む。

10

## 【0013】

別の態様では、本開示は、ヒトまたはマウス試料を同定するためのマイクロアレイを提供する。ある種の実施形態では、このマイクロアレイは、一群のヒトまたはマウスS N P遺伝子座において試料の遺伝子型を検出するためのプローブを含む。

## 【0014】

さらに別の態様では、本開示は、指示が記憶された非一過性のコンピュータ可読媒体であって、指示がプロセッサによって実行されるとき、指示がプロセッサに、一群のヒトまたはマウスS N P遺伝子座における試料の遺伝子型の検索；試料の遺伝子型の、参照遺伝子型との比較；および試料の同定の決定、を行わせる、コンピュータ可読媒体を提供する。

20

## 【0015】

さらに別の態様では、本開示は、主要構成成分および微量構成成分を含む試料を認証するための方法を提供する。ある種の実施形態では、この方法は、100個以上のS N P遺伝子座において、試料の遺伝子型を検出すること；表11に示される式に従って、S N P遺伝子座の各々についてS N P不均一性比率を決定すること；遺伝子型をモデル化する混合ガウス分布を使用して、S N P遺伝子座についてのS N P不均一性比率に基づいて試料不均一性比率を決定すること；および試料の遺伝子型を、参照試料において各々検出された一群の参照遺伝子型と比較すること、試料の遺伝子型に対して最も高い同一性を有する参照遺伝子型を有する参照試料を同定すること、(i)参照遺伝子型が試料の遺伝子型に対して90%よりも高く同一であり、かつ試料不均一性比率が10%未満であるときに、または(ii)参照遺伝子型が試料の遺伝子型に対して80%よりも高く同一であり、かつ試料不均一性比率が10%よりも高いときに、主要構成成分が参照試料であることを決定することによって、試料の主要構成成分を決定すること、を含む。

30

## 【0016】

ある種の実施形態では、この方法は、試料の微量構成成分を決定することをさらに含む。ある種の実施形態では、この方法は、試料中の主要構成成分および微量構成成分のパーセンテージを決定することをさらに含む。

40

## 【0017】

以下の図面は本明細書の一部を構成し、本開示のある特定の態様をさらに示すために含まれる。本開示は、本明細書で提示した特定の実施形態の詳細な説明と組み合わせてこれらの図面の1つまたは複数を参照することによって、よりよく理解することができる。

## 【図面の簡単な説明】

## 【0018】

【図1A】図1A～1Cは、細胞株の認証および試料の遺伝的不均一性を示す。図1A：無関係/不一致、同一な、および密接に関係する細胞株対についての遺伝子型類似性。

【図1B】図1B：118の夾雜なしの細胞株、220のP D Xおよび31のP D X Oモデルにおける不均一性比率。

50

【図1C】図1C：不均一性比率は、PDXモデルでは、マウス比率と正に相関する。

【図2】図2A～2Dは、不均一性比率が夾雑を検出および定量するために使用できることを示している。図2A～2D：後者について5%、2.5%、1.25%および0.625%の細胞比率での、細胞株MV-4-11(MV411)およびLNCaPクローニング(FGC(LNCAPCLONEFGC))の連続混合、図2E：純粋なLNCaPクローニング(FGC)細胞株、図2F：純粋なMV-4-11細胞株。横軸上の各刻みは、対応するSNP不均一性比率を有するインフォーマティブなSNP部位を示す。確率密度は、2/3成分混合ガウスを仮定することによって推定した。試料の通し番号は、括弧内の主要構成成分細胞株と共に、右上のボックス中に表示される。試料不均一性比率は下に示される。

【図3】図3A～3Fは、夾雑検出、夾雑物推論および夾雑比率推定を示す。図3A：試料19R58129は、夾雑している微量細胞株LNCaPクローニング(FGC(LNCAPCLONEFGC))と混合されたMV411である。LNCAPCLONEFGCは、1.41%の夾雑比率で、夾雑物として正確に同定された( $p$ 値=5.01E-17)。LNCaP-C4-2(C42)およびLNCAPCLONEFGCは共に、LNCaPに由来し、高い遺伝的同一性を共有している。分位数-分位数プロットにおいて、各ドットは、参照細胞株であり、理論的分位数および試料分位数を、MV411参照細胞株と1055参照細胞株との間の遺伝子型類似性にフィットさせたベータ分布から計算した。99%信頼帯は影付きである。図3B：異なる不均一性比率下で細胞株中に夾雑している第2の細胞株を推論する精度。既知の第2の細胞株が夾雑している合計94の細胞株試料を試験し、試料を、不均一性比率によってピニングした。図3C：細胞株「G-292クローニングA141B1」は、7.62%の試料不均一性比率を有し、SNP不均一性比率の確率密度における区別できる右ピークを伴うが、これは、この細胞株が夾雑していたことを示している。図3D：OCI-AML-2は、6.21%の夾雑比率で、細胞株「G-292クローニングA141B1」中の夾雑物として推論された( $p$ 値=1.58E-07)。図3E：シミュレートされた細胞株混合物における、推定夾雑比率と既知の夾雑比率との間の完璧に近い相関。図3F：既知の夾雑を有する細胞株試料についての不均一性比率と夾雑比率との間の高い相関。

【図4】図4A～4Dは、ヒト-マウス混合物中のマウス比率の推定を示す。図4A：90%、80%、70%、50%、30%、20%、10%、7%、5%および0%のマウス比率を有する一連のヒト-マウスDNA混合物における、ディープNGSシークエンシングによるマウス比率の正確な推定。図4B～4C：3つのアプローチによって220のPDXおよび31のPDX由来オルガノイドモデルにおいて推定し、各モデルについて同じ試料についてアッセイしたマウス比率。図4D：220のPDXモデルにおける、ディープNGSシークエンシングによって推定したマウス比率とWESによって推定したマウス比率との間の二次関係。

#### 【発明を実施するための形態】

##### 【0019】

本開示をより詳細に説明する前に、本開示は記載した特定の実施形態に限定されず、それ自体もちろん変化することができるることを理解されたい。本明細書で使用した専門用語は、特定の実施形態を説明するためだけのものであって、限定を意味するものではなく、なぜならば、本開示の範囲は添付の特許請求の範囲によってのみ限定されるからであることも理解されたい。

##### 【0020】

他に規定しなければ、本明細書で使用した技術用語および科学用語は全て、本開示が属する業界の当業者によって通常理解されるのと同じ意味を有する。本明細書で記載したものと類似の、または同等のいかなる方法および材料も本開示の実践または試験において使用することができるが、好ましい方法および材料をここで説明する。

##### 【0021】

本明細書で引用した刊行物および特許は全て、それぞれ個々の刊行物または特許が具体的かつ個々に参考として組み込まれることが指示されているように参考として本明細書に

10

20

30

40

50

組み込まれており、刊行物が引用されたものに関連して方法および／または材料を開示して記載するために、参考として本明細書に組み込まれている。いかなる刊行物の引用も、その開示が本出願日より前であるためであり、本開示は先行開示によりこのような刊行物に先行する権利がないことの承認と解釈されるべきではない。さらに、提供された刊行物の日付は、個別に確認する必要があり得る実際の刊行日とは異なっていてもよい。

#### 【0022】

本開示を読めば当業者には明らかなように、本明細書で記載し例示した個々の実施形態のそれぞれは、本開示の範囲または精神を逸脱することなく、その他のいくつかの実施形態のいずれかの特性から容易に分離するか、または一緒にすることができる別個の構成成分および特性を有する。任意の引用した方法は、引用した事象の順番で、または論理的に可能な任意の他の順番で実施することができる。

10

#### 【0023】

##### 定義

以下の定義は読者を支援するために提供される。他に規定しなければ、本明細書で使用した技術用語、表記およびその他の科学的用語または医学用語または専門用語は全て、化学および医学業界の当技術者によって通常理解される意味を有するものとする。場合によつては、通常理解される意味を有する用語は、明瞭になるように、および／または容易に参照できるように本明細書で定義されており、本明細書にこのような定義を含めるのは、当業界で一般的に理解されるような用語の定義をめぐる実質的な違いを表すためのものと必ずしも解釈されるべきではない。

20

#### 【0024】

本明細書で使用する場合、単数形「a」、「a n」および「t h e」は、文脈が明確に他のことを指示していない限り、複数の参照物を含む。

#### 【0025】

用語「対立遺伝子」は、特定の多型遺伝子座の2つ以上の既存の遺伝的バリエントのうち1つを意味する。

#### 【0026】

用語「量」または「レベル」は、試料中に存在する目的のポリヌクレオチドまたは目的のポリペプチドの量を意味する。このような量は、絶対的に、すなわち、試料中のポリヌクレオチドまたはポリペプチドの全量で表現されているか、あるいは相対的に、すなわち、試料中のポリヌクレオチドまたはポリペプチドの濃度で表現されていてもよい。

30

#### 【0027】

本明細書で使用する場合、用語「がん」または「腫瘍」は、異常な細胞増殖が関与する任意の疾患を意味し、体内的任意の組織、器官または細胞に影響を及ぼす疾患の全病期および全形態を含む。この用語には、悪性、良性、軟組織または固形物として特徴づけられる公知のがんおよび新生物状態の全て、ならびに転移前および転移後のがんを含む全病期およびグレードのがんが含まれる。全般的に、がんは、がんが位置する、またはがんが発生した組織または器官ならびに癌性組織および細胞の形態によって分類することができる。本明細書で使用する場合、がんの種類には、限定されるものではないが、急性リンパ性白血病（A L L）、急性骨髓性白血病、副腎皮質癌、肛門がん、星状細胞腫、小児小脳または大脳基底細胞癌、胆管がん、膀胱がん、骨腫瘍、脳がん、小脳星状細胞腫、大脳星状細胞腫／悪性神経膠腫、上衣腫、髄芽腫、テント上原始神経外胚葉性腫瘍、視覚路および視床下部神経膠腫、乳がん、バーキットリンパ腫、子宮頸がん、慢性リンパ性白血病、慢性骨髓性白血病、結腸がん、肺気腫、子宮内膜がん、上衣腫、食道がん、ユーイング肉腫、網膜芽細胞腫、胃〔gastric〕（胃〔stomach〕）がん、神経膠腫、頭部および頸部がん、心臓がん、ホジキンリンパ腫、脾島細胞癌（脾内分泌部）、カポジ肉腫、腎臓がん（腎細胞がん）、咽頭がん、白血病、肝臓がん、肺がん、神経芽細胞腫、非ホジキンリンパ腫、卵巣がん、脾臓がん、咽頭がん、前立腺がん、直腸がん、腎細胞癌（腎臓がん）、網膜芽細胞腫、ユーイングファミリーの腫瘍、皮膚がん、胃がん、精巣がん、咽頭がん、甲状腺がん、腫がんが含まれる。

40

50

## 【0028】

本明細書で使用する「細胞」は、原核細胞であっても真核細胞であってもよい。原核細胞には、例えば、細菌が含まれる。真核細胞には、例えば、真菌、植物細胞および動物細胞が含まれる。動物細胞（例えば、哺乳類細胞またはヒト細胞）の種類には、例えば、循環／免疫系または器官の細胞（例えば、B細胞、T細胞（細胞傷害性T細胞、ナチュラルキラーT細胞、制御性T細胞、ヘルパーT細胞）、ナチュラルキラー細胞、顆粒球（例えば、好塩基性顆粒球、好酸性顆粒球、好中性顆粒球および過分葉好中球）、単核球またはマクロファージ、赤血球細胞（例えば、網状赤血球）、肥満細胞、血小板または巨核球、および樹状細胞）、内分泌系または器官の細胞（例えば、甲状腺細胞（例えば、甲状腺上皮細胞、傍濾胞細胞）、副甲状腺細胞（例えば、副甲状腺主細胞、好酸性細胞）、副腎細胞（例えば、クロム親和性細胞）および松果体の細胞（例えば、松果体細胞））、神経系または器官の細胞（例えば、神経膠芽細胞（例えば、星状膠細胞および乏突起膠細胞）、小膠細胞、巨細胞性神経分泌細胞、星状細胞、ベッセル細胞および下垂体細胞（例えば、性腺刺激ホルモン分泌細胞、副腎皮質刺激ホルモン分泌細胞、甲状腺刺激ホルモン分泌細胞、成長ホルモン分泌細胞および乳腺刺激ホルモン分泌細胞））、呼吸系または器官の細胞（例えば、肺胞細胞（I型肺胞細胞およびII型肺胞細胞）、クララ細胞、杯状細胞、肺胞マクロファージ）、循環系または器官の細胞（例えば、心筋細胞および周辺細胞）、消化系または器官の細胞（例えば、胃主細胞、壁細胞、杯状細胞、パネット細胞、G細胞、D細胞、ECL細胞、I細胞、K細胞、S細胞、腸内分泌細胞、腸クロム親和性細胞、APUD細胞、肝臓の細胞（例えば、肝実質細胞およびクッパー細胞））、外被系または器官の細胞（例えば、骨の細胞（例えば、骨芽細胞、骨細胞および破骨細胞）、歯の細胞（例えば、セメント芽細胞およびエナメル芽細胞）、軟骨の細胞（例えば、軟骨芽細胞および軟骨細胞）、皮膚／髪の細胞（例えば、毛胞、角化細胞およびメラニン形成細胞（母斑細胞））、筋肉細胞（例えば、筋細胞）、脂肪細胞、線維芽細胞および腱細胞）、泌尿器系または器官の細胞（例えば、有足細胞、傍糸球体細胞、糸球体内メサンギウム細胞、糸球体外メサンギウム細胞、腎臓近位尿細管刷子縁細胞および緻密斑細胞）ならびに生殖器系または器官の細胞（例えば、精子、セルトリ細胞、ライディッヒ細胞、卵子、卵母細胞）が含まれる。細胞は、正常で健康な細胞であるか、または病的で不健康な細胞（例えば、がん細胞）であってもよい。細胞にはさらに、哺乳類接合子または胚性幹細胞、胎児幹細胞、人工多能性幹細胞および成人幹細胞を含む幹細胞が含まれる。幹細胞とは、未分化状態を維持しながら細胞分裂のサイクルを経て、特殊化した細胞種に分化することができる細胞である。幹細胞は、全能性幹細胞、多能性幹細胞、複能性幹細胞、寡能性幹細胞および単能性幹細胞であってもよく、いずれも体細胞から誘導することができる。幹細胞はまた、がん性幹細胞を含むことができる。哺乳類細胞は、齧歯類細胞、例えば、マウス、ラット、ハムスター細胞であってもよい。哺乳類細胞は、ウサギ目細胞、例えば、ウサギ細胞であってもよい。哺乳類細胞はまた、靈長類細胞、例えば、ヒト細胞であってもよい。ある特定の例では、細胞は、マスバイオプロダクションで使用される細胞、例えば、CHO細胞である。

## 【0029】

用語「相補性」は、伝統的なワトソン・クリックまたはその他の非伝統的な方法のいずれかによって、別の核酸配列と水素結合を形成する核酸の能力を意味する。パーセント相補性は、第2の核酸配列と水素結合（例えば、ワトソン・クリック塩基対）を形成することができる核酸分子中の残基のパーセンテージを指示する（例えば、10のうち5、6、7、8、9、10は50%、60%>、70%>、80%>、90%および100%相補性である）。「完全に相補性である」は、核酸配列の連続した残基が全て、第2の核酸配列中の同じ数の連続した残基と水素結合することを意味する。本明細書で使用する「実質的に相補性である」は、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、30、35、40、45、50、またはそれ以上のヌクレオチドの領域に亘って少なくとも60%、65%、70%、75%、80%、85%、90%、95%、97%、98%、99%または100%である相補性

10

20

30

40

50

の程度を意味するか、あるいはストリンジェント条件下でハイブリダイズする2つの核酸を意味する。

【0030】

本開示では、「含む〔comprises〕」、「含んだ〔comprised〕」、「含んでいる〔comprising〕」、「含有する〔contains〕」、「含有している〔containing〕」などの用語は、米国特許法にある意味を有し、これらは包括的またはオープンエンドであり、追加的な、引用されていない要素または方法ステップを排除しないことに注意されたい。「から本質的になっている〔consisting essentially of〕」および「から本質的になる〔consists essentially of〕」などの用語は、米国特許法にある意味を有し、これらは請求された発明の基本的で新規な特質に著しい影響を及ぼさない追加的な成分またはステップを含むことを許可する。用語「からなる〔consists of〕」および「からなっている〔consisting of〕」は米国特許法にあるものとみなされる意味を有し、すなわち、これらの用語はクローズエンドである。

10

【0031】

本明細書で使用する場合、用語「夾雜物」は、試料中の主要構成成分とは異なる、または試料の不純物もしくはその他の望ましくない影響、例えば、だめにすること〔spoiling〕、汚染、感染を引き起こす、試料中に存在する構成成分を意味する。

【0032】

用語「決定する」、「評価する」、「アッセイする」、「測定する」および「検出する」は同義に使用することができ、定量的および半定量的決定の両方を意味する。定量的および半定量的決定のいずれかを意味する場合、目的のポリヌクレオチドまたはポリペプチドの「レベルを決定する」または目的のポリヌクレオチドまたはポリペプチドを「検出する」という語句を使用することができる。

20

【0033】

用語「ゲノム」は、その染色体の完全なDNA配列によって示される、個々の生物または細胞が運搬する総遺伝情報を意味する。

【0034】

用語「ハイブリダイジング」は、ストリンジェント条件下で特定のヌクレオチド配列に対して優先的な核酸分子の結合、二重鎖形成またはハイブリダイジングを意味する。用語「ストリンジェント条件」は、混合した集団（例えば、組織生検からの細胞溶解物またはDNA調製物）中において、プローブがその標的部分配列に優先的にハイブリダイズし、他の配列に対しては比較的少ない程度でハイブリダイズするか、または全くハイブリダイズしない条件を意味する。核酸ハイブリダイゼーションの場合における（例えば、アレイ、マイクロアレイ、サザンもしくはノザンハイブリダイゼーションにおけるような）「ストリンジェントハイブリダイゼーション」および「ストリンジェントハイブリダイゼーション洗浄条件」は、配列に依存し、異なる環境パラメータ下では異なる。核酸のハイブリダイゼーションの広範な指針は、例えば、*Tijssen Laboratory Techniques in Biochemistry and Molecular Biology - Hybridization with Nucleic Acid Probes part I, Ch. 2, "Overview of principles of Hybridization and the strategy of Nucleic acid probe assays,"* (1993) Elsevier, N.Y. に見出される。全般的に、高ストリンジェントハイブリダイゼーションおよび洗浄条件は、規定されたイオン強度およびpHでの特定の配列の融点（T<sub>m</sub>）より約5℃低くなるように選択される。T<sub>m</sub>は、（規定されたイオン強度およびpHで）標的配列の50%が完全に一致したプローブにハイブリダイズする温度である。非常にストリンジェントな条件は、特定のプローブのT<sub>m</sub>に等しくなるように選択される。サザンまたはノザンプロットにおいてアレイまたはフィルター上に100個を上回る相補的残基を有する相補的核酸のハイブリダイゼーションのためのストリンジェントハイブリダイゼーション条件の一例は、標準的なハイブリダイゼーション溶液を使用して42℃である（例えば、*Sambrook*

30

40

50

k and Russell Molecular Cloning: A Laboratory Manual (3rd ed.) Vol. 1 - 3 (2001) Cold Spring Harbor Laboratory, Cold Spring Harbor Press, NYを参照のこと)。高ストリンジエント洗浄条件の一例は、NaCl 0.15Mで72で約15分間である。ストリンジエント洗浄条件の一例は、0.2×SSC洗浄で65で約15分間である。しばしば、バックグラウンドプローブシグナルを除去するために、高ストリンジエント洗浄に先行して低ストリンジエント洗浄を行う。例えば、100ヌクレオチドを上回る二本鎖のための中程度ストリンジエント洗浄の一例は、1×SSCで45で15分間である。例えば、100ヌクレオチドを上回る二本鎖のための低ストリンジエント洗浄の一例は、4×SSCから6×SSCで40で15分間である。

10

#### 【0035】

用語「遺伝子座」は、生物学的機能にかかわらず、当業界で公知の参照ゲノム中の染色体座標によって規定される、ゲノム中のDNA配列の任意のセグメントを意味する。DNA遺伝子座は、複数の遺伝子を含有してもよく、遺伝子を含有しなくてもよい; DNA遺伝子座は、単一の塩基対または数百万の塩基対であり得る。

#### 【0036】

用語「核酸」および「ポリヌクレオチド」は同義に使用され、デオキシリボヌクレオチドもしくはリボヌクレオチドのいずれかの任意の長さのヌクレオチドの多量体型、またはそれらの類似体を意味する。ポリヌクレオチドは、任意の三次元構造を有することができ、公知のまたは公知ではない任意の機能を果たすことができる。ポリヌクレオチドの非限定的な例には、遺伝子、遺伝子断片、エキソン、イントロン、メッセンジャーRNA(mRNA)、トランスクレーバーRNA、リボソームRNA、リボザイム、cDNA、sRNA、一本鎖の短いかまたは長いRNA、組換えポリヌクレオチド、枝分かれポリヌクレオチド、プラスミド、ベクター、任意の配列の単離されたDNA、調節領域、任意の配列の単離されたRNA、核酸プローブおよびプライマーが含まれる。核酸分子は、直鎖状または環状であってもよい。

20

#### 【0037】

用語「オリゴヌクレオチド」は、少なくとも約5ヌクレオチド～約500ヌクレオチド(例えば、5、6、7、8、9、10、12、15、18、20、21、22、25、30、35、40、45、50、55、60、65、70、75、80、85、90、100、125、150、175、200、250、300、350、400、450または500ヌクレオチド)の核酸配列を意味する。一部の実施形態では、例えば、オリゴヌクレオチドは、例えば、ポリメラーゼ連鎖反応(PCR)増幅アッセイにおけるプライマーとして、および/またはハイブリダイゼーションアッセイもしくはマイクロアレイにおけるプローブとして使用され得る、約15ヌクレオチド～約30ヌクレオチド、または約20ヌクレオチド～約25ヌクレオチドであり得る。本発明のオリゴヌクレオチドは、当業界で周知のように、天然または合成の、例えば、DNA、RNA、PNA、LNA、修飾された骨格などであり得る。

30

#### 【0038】

用語「多型遺伝子座」は、2つ以上の対立遺伝子が同定されているゲノム遺伝子座を意味する。

40

#### 【0039】

用語「参照遺伝子型」は、本明細書で使用する場合、参照試料、例えば、既知の正体を有する試料中に存在する、1つ以上のゲノム遺伝子座の予め決定された遺伝子型を意味する。参照遺伝子型は、試験試料中に存在する特定のゲノム遺伝子座の遺伝子型を比較するための基礎として機能するように、本発明の方法の使用のために適切である。参照遺伝子型は、試料の性質ならびにこのような参照試料を確立する基となつた対象の性別、年齢、民族性などのその他の要素に応じて変化し得る。

#### 【0040】

50

本明細書で使用する用語「試料」または「生物学的試料」は、任意の細胞、組織、オルガノイド、または目的の1つ以上の核酸分子を含有する任意のその他の試料を意味する。ある種の実施形態では、試料は、細胞（例えば、正常細胞、がん細胞、細胞株）、組織（例えば、正常組織、がん組織、異種移植または同種異系移植組織）、オルガノイドなどである。

【0041】

用語「単一ヌクレオチド多型」または「SNP」は、2つ以上の代替的対立遺伝子が集団内に感知できる頻度、例えば>1%で存在する、ゲノム配列中の単一のヌクレオチド位置を意味する。SNPは、遺伝子のコード配列内、遺伝子の非コード領域内、および/または遺伝子の遺伝子間（例えば、イントロン）領域中に存在し得る。タンパク質コード領域中に存在するわけではないSNPもなお、遺伝子スプライシング、転写因子結合および/または非コードRNAの配列に対して影響を有し得る。本明細書で提供したSNP命名法は、GenBank（登録商標）データベースにおいて入手可能な、National Center for Biotechnological Information（NCBI）によって各固有のSNPに割り当てられた公式のReference SNP（rs）同定番号を意味する。

10

【0042】

本明細書で使用する場合、用語「対象」は、ヒトまたは任意の非ヒト動物（例えば、マウス、ラット、ウサギ、イヌ、ネコ、ウシ、ブタ、ヒツジ、ウマまたは靈長類）を意味する。ヒトには、出生前および出生後の形態が含まれる。多くの実施形態では、対象は人間である。対象は患者であってもよく、これは疾患の診断または処置のために医療機関に来院した人を意味する。用語「対象」は、本明細書では「個体」または「患者」と同義に使用される。対象は、疾患または障害に罹っているか、または罹りやすくてよいが、疾患または障害の症状を表していてもいなくてもよい。

20

【0043】

用語「基材」は、アレイの文脈で使用する場合、関連するアッセイ構成成分（例えば、アッセイ領域、細胞、試験化合物など）を支持することができる材料を意味する。基材の例には、限定されるものではないが、ガラス、Siベースの材料、官能化ポリスチレン、官能化ポリエチレングリコール、官能化有機ポリマー、ニトロセルロースまたはナイロンメンブレン、紙、綿、および合成に適切な材料が含まれる。基材は、平らである必要はなく、球状形状（例えば、ビーズ）を含む任意の型の形状が含まれる。基材に結合される材料は、基材の任意の部分に結合され得る（例えば、多孔性基材材料の内側部分に結合され得る）。本発明の技術の好ましい実施形態は、基材に結合した核酸プローブを有する。核酸プローブは、非ランダムな化学的または物理的相互作用を介して基材と関連する場合、基材に「結合」されている。一部の好ましい実施形態では、結合は、例えばリンカーによって提供されるような、共有結合を介したものである。

30

【0044】

用語「腫瘍モデル」は、本明細書で使用する場合、がんの発達および増悪を研究するため、ならびにヒトに与えられる前に処置を試験するために使用される、細胞、組織または動物を意味する。

40

【0045】

用語「腫瘍試料」には、1つまたは複数の腫瘍細胞を含有する生物学的試料または生物源からの試料が含まれる。生物学的試料には、体液、例えば、血液、血漿、血清もしくは尿の試料、または、例えば、生検によって、細胞、組織もしくは器官、好ましくはがん細胞を含むか、もしくは本質的にはがん細胞からなることが疑われる腫瘍組織から得られた試料が含まれる。

【0046】

腫瘍試料の同定のためのSNP

バイオバンク試料（例えば、細胞株）の誤同定および夾雜は、生物医学的研究を悩ませてきた。ショートタンデムリピート（STR）および単一ヌクレオチド多型（SNP）ア

50

ツセイは、生体試料を認証するために広く使用されており、それぞれ 5 ~ 10 % および 3 ~ 5 % の感度で夾雜を検出することができる。本開示は、一態様では、夾雜を検出するのに 1 % の感度を有する方法を提供する。この方法はさらに、混合細胞株試料について、夾雜物を同定でき、夾雜比率を推定できる。この方法は、細胞株認証について報告された、すばぬけて最も高感度かつ正確な方法である。ある種の実施形態では、この方法は、異種移植腫瘍などのヒト - マウス混合試料における種間夾雜を検出することもでき、マウス比率を正確に推定することもできる。ある種の実施形態では、マイコプラズマおよびモリキュートも、研究標的の中にある。ある種の実施形態では、この多機能方法は、ヒト試料の集団構造および性別を同時に推論する。ある種の実施形態では、DNA バーコード化技術のおかげで、本明細書で開示した方法は、従来の STR アッセイと匹敵する試料当たりのコストで、単一の実行で 100 ~ 200 の試料をプロファイリングすることができ、それにより、高品質のバイオバンクを維持するための真にハイスループットかつ低コストのツールになっている。

#### 【 0047 】

本明細書で記載した方法および組成物は、腫瘍モデルから得られた試料を同定および認証するために使用され得る一群の SNP 遺伝子座の発見に一部基づく。ある種の実施形態では、腫瘍モデルは、原発性ヒト腫瘍、患者由来異種移植 (PDX) 、ヒト腫瘍細胞株、ヒト細胞株由来異種移植およびヒトオルガノイドを含むヒト腫瘍モデルである。ある種の実施形態では、SNP は、いくつかのヒト腫瘍モデルの RNAseq または全エキソームシークエンシング (WES) データに基づいて、ヒト SNP から選択される。選択されたヒト SNP は、22 個の常染色体に亘って主に非連鎖不平衡 (非 LD) ブロック中に位置する、高度に発現される遺伝子のエキソン領域中に位置する。したがって、各ヒト腫瘍モデルは、選択されたヒト SNP 遺伝子座において、固有の遺伝子型 (すなわち、SNP フィンガープリント) を有する。

#### 【 0048 】

ある種の実施形態では、選択されたヒト SNP 遺伝子座は、マウスゲノムにおいて相同性を有する。このようなヒト SNP 遺伝子座を標的化するプライマーを使用して試料が増幅される場合、試料がマウスの細胞または組織と混合されていれば、対応するマウス遺伝子座のヌクレオチド配列が生成され得る。このようなヒト SNP は、例えば、これらの SNP のマウスおよびヒト読み取りの数に基づいて、ヒトおよびマウスの細胞 / 組織の混合物中のマウス含量のパーセンテージを推定するために使用され得る。

#### 【 0049 】

ある種の実施形態では、本明細書で使用するヒト SNP は、表 1 に示される群から選択される。

#### 【 0050 】

ある種の実施形態では、SNP には、マウス腫瘍細胞株などのマウス腫瘍モデルを同定および認証するためのマウス SNP が含まれる。一部の実施形態では、本明細書で使用するマウス SNP は、表 2 に示される群から選択される。

#### 【 0051 】

ある種の実施形態では、SNP には、試料を得た対象の性別を決定するための、性染色体 (X 染色体および Y 染色体) 中のヒト SNP がさらに含まれる。ある種の実施形態では、性染色体 SNP は、表 3 に示される群から選択される。

#### 【 0052 】

ある種の実施形態では、SNP には、試料を得た免疫不全マウスの系統を決定するために使用され得るマウス SNP がさらに含まれる。一部の実施形態では、SNP は、表 4 に示される。

#### 【 0053 】

##### 方法

一態様では、本開示は、試料を同定および認証するための方法を提供する。

#### 【 0054 】

10

20

30

40

50

ある種の実施形態では、本明細書で開示した方法は、試料を参照（例えば、標準的ながん細胞株）と一致させるためのものである。従来のSTRおよびSNPアッセイは、遺伝子型ベースのTanabe-Mastersアルゴリズムおよびそのバリエーションを主に使用していた。STRアッセイは、たくさんのマーカーについて類似のシグナルを生じる。SNPアッセイは、しばしばさらに多くのSNPを遺伝子型決定する。したがって、2つの試料を一致と呼ぶためにSNPアッセイによって使用される類似性閾値は、より高い場合が多い。しかし、従来のアッセイの一一致力〔matching power〕は、約100のSNPを用いた場合であっても、夾雜された試料についてひどく損なわれ得る。ある種の実施形態では、本明細書で開示した方法は、ヒト試料について237のSNP部位の高深度（3000×）シークエンシングを実施し、試料、または夾雜された試料の主要構成成分を同定することにおいて100%の精度を示した。

#### 【0055】

ある種の実施形態では、本明細書で開示した方法は、生物学的試料における夾雜を検出するためのものである。細胞株における夾雜を検出するための感度は、STRアッセイについては約5～10%、SNPアッセイについては3～5%である。しかし、性能はむしろ、>20%の夾雜ですら96-SNPアッセイによって2つの無関係の細胞株の混合物中で検出されなかつた程度まで、不安定であり得る（Liang-Chu, M.M. et al. PLoS One 10, e0116218 (2015)）。ある種の実施形態では、本明細書で開示した方法は一貫して、不均一性比率だけを使用する場合、その値および区別できる二／三峰性分布の両方によって、2%の感度に到達する。感度は、SNPフィンガープリントを有する参照試料のライブラリー中に夾雜物が存在する場合には、1%に達する。夾雜なしの細胞株は、多クローニ性およびシークエンシングエラーに起因して、約1%の夾雜を有する細胞株試料と匹敵するレベルの遺伝的不均一性を示すので、このような感度は事実上、理論的検出限界である。

#### 【0056】

ある種の実施形態では、本明細書で開示した方法は、夾雜物を同定するためのものである。細胞株の交差夾雜は、バイオバンクでは一般的である。夾雜された培養物の組成は、細胞株の異なる増殖速度に起因して、経時的に変化する。細胞株は、遺伝子変異など、ゲノミクスが異なっており、薬物処置に異なって応答して、薬物スクリーニングにおいて誤った結果を引き起こし得る。本開示の発明者らは、1000を超えるがん細胞株についてSNPフィンガープリントライブラリーを構築し、それにより、夾雜している細胞株は、一義的に同定され得る。さらに、夾雜比率が正確に推定され得る。細胞株の品質をチェックすることに加えて、この能力は、生物学的または化学的な干渉の下での2つの細胞株の動的組成をモニターすることなどの、その他の利用を有し得る。

#### 【0057】

種内夾雜に加えて、ある種の実施形態では、本明細書で開示した方法は、ヒトとマウスとの間の種間夾雜を正確に検出および定量することができる。ある種の実施形態では、本明細書で開示した方法は、SNPではなく、2つの種間で異なっているが、同一な隣接ヌクレオチド配列を有する、108の相同的DNAセグメントを使用し、したがって、共通プライマーを、ヒトおよびマウスDNAセグメントの偏りのない増幅のために設計することができる。このアプローチは、一連のマウス-ヒトDNA混合物ベンチマーク試料において、完璧な性能を示した。相同性ベースの原理は、その他の種間夾雜を検出するために使用され得る。

#### 【0058】

ある種の実施形態では、本明細書で開示した方法の力は、いくつかの新規の特性から来ている。第1は、ディープNGSシークエンシングであり、これは、SNPの遺伝子型およびヌクレオチド頻度の両方を得るものであるが、従来のSTRおよびSNPアッセイは、SNP遺伝子型をプロファイリングするだけである。第2に、SNPプロファイリングとは別に、本明細書で開示した方法は、マイクロラズマ夾雜を検出し、マウス-ヒト混合比率を推定するために、標的化シークエンシングを実施する。第3に、一そろいの統計モ

10

20

30

40

50

デルおよびアルゴリズムが、ディープNGSシークエンシングデータを利用するため開発されており、認証方法を自動的でロバストかつ客観的なものにしている。最後に、DNAバーコード技術が、100～200の試料の同時の並列シークエンシングを可能にするために使用され、コストを大幅に低減させる。

【0059】

本明細書で開示したハイスループットで低成本の方法は、認証済みの高品質の試料を維持するために、バイオバンクによって慣用的に使用され得る。この方法は、その他の種およびさらにはマイクロバイオームからの試料に広く適応され得、任意のNGSシークエンシングプラットフォーム上で実現され得る。

【0060】

一実施形態では、この方法は、試料から核酸を得ること；本明細書で開示した複数のヒトまたはマウスの単一のSNP遺伝子座において、試料についての遺伝子型を検出すること；試料についての遺伝子型を、参照試料において検出された参照遺伝子型と比較すること；および試料の同定を決定することを含む。ある種の実施形態では、50、60、70、80、90、100、110、120、130、140、150、160、170、180、190、200またはそれ以上のSNP遺伝子座における遺伝子型が検出される。

【0061】

試料から得られた核酸は、RNAまたはDNAであり得る。ある種の実施形態では、試料から得られた核酸は、試料から単離されたゲノムDNAである。ある種の実施形態では、試料から得られた核酸は、ゲノムDNAであり、試料から単離された総RNAまたはmRNAである。ある種の実施形態では、試料から得られた核酸は、例えば、PCR反応または逆転写後のPCRによって増幅される。

【0062】

SNP遺伝子座における試料についての遺伝子型は、当業界で公知の任意の適切な方法、例えば、限定されるものではないが、シークエンシングベースの方法およびハイブリダイゼーションベースの方法に基づいて検出することができる。

【0063】

ある種の実施形態では、検出するステップは、増幅ステップを含む。このような場合、検出剤は、SNP遺伝子座を含有するゲノム領域にハイブリダイズし、ポリメラーゼの存在下でそのSNP遺伝子座を包囲するポリヌクレオチド配列を増幅することができる、少なくとも一対のプライマーを含む。SNPを含有するゲノム領域を増幅するために使用されるプライマー対は、プライマーまたはプローブが、ゲノム領域またはその相補鎖に特異的にハイブリダイズできるように、ゲノム領域の少なくとも一部分に対して十分な同一性または相補性を有する。「特異的にハイブリダイズする」は、本明細書で使用する場合、プライマーまたはプローブが、ストリンジエント条件下で意図した配列にハイブリダイズできることを意味する。「ストリンジエント条件」は、本明細書で使用する場合、5×SSPE、5×デンハルト液、0.5%SDSおよび100ug/ml変性サケ精子DNAからなる溶液中で42でハイブリダイズし、次いで、0.5×SSCおよび0.1%SDSを含む溶液で42で洗浄することを意味する。

【0064】

PCRなどの適切な核酸増幅方法による増幅後、増幅生成物中の配列またはSNPが検出される。ある種の実施形態では、増幅生成物は、50bp～500bpの長さを有する。ある種の実施形態では、増幅生成物中のSNPの配列は、シークエンシングベースの方法、例えば、次世代シークエンシング(NGS)方法を使用して検出される。ある種の実施形態では、NGS方法は、多数のSNP遺伝子座における配列を決定するために使用される。ある種の実施形態では、NGS方法は、各試料から得られた核酸をバーコード化することによって、多くの試料からのSNP遺伝子座の配列を同時に決定するために使用され得る。

【0065】

試料から得られた核酸がRNAである場合、増幅ステップは、試料中のRNAのcDNA

10

20

30

40

50

Aを產生するための逆転写ステップを任意選択で含み得る。次いで、cDNAは、SNPの存在の検出を可能にするために、プライマーを使用して増幅される。

【0066】

一部の実施形態では、例えば、マイクロアレイが、核酸中のSNPを検出するために使用される。マイクロアレイは、固相支持体に結合した捕捉プローブの再現性のあるパターンから構成される。標識されたRNAまたはDNAは、アレイ上の相補的プローブにハイブリダイズし、その後レーザースキャナによって検出される。SNPの存在は、アレイ上の特異的プローブに結合する標識されたRNAまたはDNAの強度を測定することによって検出され得る。

【0067】

機械的合成法を使用したこれらのアレイの合成技術は、例えば、米国特許第5,384,261号に記載されている。平面アレイ表面が使用されることが多いが、アレイは実際にはいかなる形状の表面に製造されていてもよく、多重の表面に製造されてもよい。アレイはまた、ビーズ、ゲル、ポリマー表面、光ファイバーなどの纖維、ガラスまたは任意のその他の適切な基材上の核酸であってもよく、米国特許第5,770,358号、第5,789,162号、第5,708,153号、第6,040,193号および第5,800,992号を参照のこと。アレイは、診断法または包括的装置のその他の操作を可能にするような方法で包装されていてもよい。

【0068】

本発明を実践するために必要なプローブおよびプライマーは、周知の技術を使用して合成するかまたは標識することができる。プローブおよびプライマーとして使用するオリゴヌクレオチドは、Beaucage and Caruthers, Tetrahedron Letts. (1981) 22:1859-1862によって最初に記載された固相ホスホラミダイトリエステル法によって、Needham-Van Devanter et al, Nucleic Acids Res. (1984) 12:6159-6168に記載されているように、自動合成機を使用して化学合成することができる。

【0069】

ある種の実施形態では、この方法は、例えば、表3に示される群から選択される性染色体SNPを検出することによって、試料を得た対象の性別を同定することをさらに含む。ある種の実施形態では、この方法は、試料を得た対象の民族性を同定することをさらに含む。ある種の実施形態では、この方法は、例えば、表4に示される供給業者のSNPを検出することによって、試料を得た免疫不全マウスの系統を決定することをさらに含む。

【0070】

ある種の実施形態では、本明細書で開示した方法は、腫瘍モデルにおける、A/B/C型肝炎ウイルス(HAV/HBV/HCV)、ヒト免疫不全ウイルス(HIV)、エブスタイン・バーウィルス(EBV)およびヒトパピローマウイルス(HPV)を含む一般的なウイルス感染およびマイコプラズマ夾雜を検出することをさらに含む。ある種の実施形態では、ウイルス感染およびマイコプラズマ夾雜を検出するために使用されるマーカーは、表5に示される。

【0071】

ある種の実施形態では、本明細書で開示した方法は、主要構成成分および微量構成成分を含む試料を認証するために使用され得る。ある種の実施形態では、この方法は、不均一性比率を推定すること；試料の主要構成成分を決定すること；試料の微量構成成分を決定すること；ならびに主要構成成分および微量構成成分の混合比率を推定することを含む。

【0072】

ある種の実施形態では、不均一性比率は、以下のように推定することができる。ディープNGSシークエンシングデータから不均一性比率を推定するために使用され得る6つのインフォーマティブな遺伝子型組合せが存在する(表11)。これらは、4つの区別できるヌクレオチド頻度パターンを示す。組合せ1および2は、同じパターンを生じ、本発明者らは、微量構成成分S2のパーセンテージ、すなわち不均一性比率を計算するために、

10

20

30

40

50

平均式を使用する。この式は、S N P の数が大きい場合に接近して近似されるはずのシナリオである、2つの組合せが等しい頻度で生じる場合に、比率の正確な推定を生じる。類似の平均化アプローチが、組合せ4および5に使用される。不均一性比率が低い場合、シークエンシングエラーが、不均一性比率の推論を妨害し得る。これを軽減するために、2段階の統計手法が使用され得る。シークエンシングエラーを  $e = 0.001$  と仮定し、所与の S N P 部位におけるシークエンシング深度を  $n$  ( $n \geq 500$ 、 $n < 500$  の S N P はいずれも破棄される) と仮定すると、 $k$  個の誤ったヌクレオチドを観測する確率は、パラメータ  $n$  および  $e$  による二項分布に従う。

【0073】

【数1】

10

$$f(k, n, e) = \binom{n}{k} e^k (1 - e)^{n-k}$$

【0074】

各  $n$  について、累積密度関数が、 $n$  個のヌクレオチドのうち  $h$  個よりも多くの誤ったヌクレオチドを観測する確率が  $0.01$  よりも小さくなるような閾値  $h$  を得るために計算され得る。シークエンシングデータにおいて、対応する閾値  $h$  よりも小さい数の読み取りを有する低頻度ヌクレオチドはいずれも破棄される。次いで、期待値最大化アルゴリズム (R用パッケージ `mclust`、バージョン `3.5.3`) が、最大不均一性 (この研究では全ての試料について `0.2` を使用した) よりも小さいヌクレオチド頻度の分布をモデル化する混合ガウス ( $1 \sim 3$  つの成分を有する) のパラメータを推定するために使用される。単一のガウス成分、または全てのデータポイントの  $60\%$  よりも多くのを占める最小平均を有するガウス成分のみが存在する場合、全てのデータポイントの中央値を試料不均一性比率とし、さもなくば、その他のガウス成分中のデータポイントの中央値を試料不均一性比率とする。

20

【0075】

試料中の主要構成成分を決定するために、S N P 部位における遺伝子型が、参照試料については  $10\%$ 、夾雜されている可能性がある試験試料については  $25\%$  である閾値よりも大きい対立遺伝子頻度を有するヌクレオチドのみを使用して決定される。参照試料と試験試料との間の遺伝子型類似性は、試験試料中の  $500$  未満のシークエンシング深度を有する S N P を除いた、同一な遺伝子型を有する S N P のパーセンテージである。試験試料の主要構成成分は、最も高い遺伝子型類似性を有する参照試料であり、この類似性は、試験試料の不均一性比率が  $< 10\%$  (または  $> 10\%$ ) である場合には、 $90\%$  (または  $80\%$ ) よりも高くなければならない。さもなくば、主要構成成分はコールされない。

30

【0076】

不均一性比率の推定および主要構成成分の決定の後、試験試料の微量構成成分が決定され得る。主要構成成分とその他の参照試料 (例えば、ゲノムデータがある全ての細胞株) のうち 1 つとの混合物について、おそらくは  $1 \sim 4$  のヌクレオチドを有するキメラ遺伝子型が、全ての S N P 部位において得られ得る。ヌクレオチドの頻度が、不均一性比率を使用して計算される。同様に、試験試料のキメラ遺伝子型が得られる。2つのキメラ遺伝子型は、それらが同じヌクレオチドを保有し、各ヌクレオチドの頻度が 3 倍以内である場合、同一とみなされる。次いで、試験試料と、主要構成成分と組み合わせた各参照試料との間の、遺伝子型類似性が計算される。次いで、全てのペアワイズ遺伝子型類似性のセットに、パラメータ ( $\alpha, \beta$ ) を用いるベータ分布がフィットされる。

40

【0077】

【数2】

$$f(x, \alpha, \beta) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1} (1-x)^{\beta-1}$$

50

## 【0078】

等式中、 $\Gamma(\cdot)$  はガンマ関数であり、 $\times$  は遺伝子型類似性である。そのパラメータは、R用パッケージ `fitdistplus` (バージョン 3.5.3) によって推定される。フィットさせたベータ分布から、特定の値よりも大きい任意の遺伝子型類似性を観測する確率が計算される。99%信頼帯を伴う分位数 - 分位数グラフが、視覚化のために、全ての観測された遺伝子型類似性についてプロットされる。参照試料は、(1) それが最も高い遺伝子型類似性を有する場合、(2) その遺伝子型類似性が、分位数 - 分位数グラフ中の 99% 信頼上限を上回る場合、および(3) フィットさせたベータ分布におけるその  $p$  値  $< 1.0 \times 10^{-6}$  である場合、微量構成成分とみなされる。

## 【0079】

2つの参照試料についての混合比率は、以下のように推定することができる。2つの構成成分  $S_1$  および  $S_2$  を、 $S_1$  について、 $S_2$  について (1 -  $\theta$ ) の比率で混合すると仮定し、このとき、 $0 < \theta < 1$  である。ディープNGSシークエンシングデータから、両方の構成成分中の  $n$  個全ての SNP のヌクレオチド頻度が、正確に推定され得る。SNPについて、その4つのヌクレオチド頻度が、構成成分  $S_1$  については  $\{A_1, T_1, G_1, C_1\}$ 、構成成分  $S_2$  については  $\{A_2, T_2, G_2, C_2\}$  として示され、これらは合計すると 1 になる。原理上、頻度のうち 1 つは、SNP がホモ接合性である場合には 1 に近く、2つの頻度は共に、SNP がヘテロ接合性である場合には 0.5 に近い。実際のデータは、シークエンシングエラーおよびランダムさ、ならびに細胞株の多クローニング性に起因して、いくらかの偏差を有し得る。

## 【0080】

混合試料のシークエンシングデータから、4つのヌクレオチドの実際の発生率は、 $\times = \{n_A, n_T, n_G, n_C\}$  として示される。このような観測の尤度は、

## 【0081】

## 【数3】

$$\mathcal{L}(\theta|x) = P_\theta(x) = \text{const} \times \prod_{M \in \{A, T, G, C\}} (\theta M_1 + (1 - \theta) M_2)^{n_M}$$

である。

## 【0082】

尤度  $P(x_i)$  は、観測されたデータ  $x_i$  を用いて、任意の SNP  $i$  ( $1, 2, \dots, n$ ) について計算され得、全ての SNP についてデータ  $X = \{x_1, x_2, \dots, x_n\}$  を観測する尤度は、

## 【0083】

## 【数4】

$$\mathcal{L}(\theta|X) = \text{const} \times \prod_{i=1}^n P_\theta(x_i)$$

である。

## 【0084】

したがって、log-尤度は、

## 【0085】

## 【数5】

$$\log \mathcal{L}(\theta|X) = \sum_{i=1}^n \log P_\theta(x_i)$$

10

20

30

40

50

である。

【0086】

尤度を最大化する は、 の段階的增加によって解かれ得る。

【0087】

キットおよびマイクロアレイ

別の態様では、本開示は前述の方法で使用するためのキットを提供する。このキットは、本明細書で記載した方法を実施するための試薬のいずれかまたは全てを含むことができる。ある種の実施形態では、このキットは、試料中の一群のヒトSNP遺伝子座または一群のマウスSNP遺伝子座を検出するためのプライマーを含む。ある種の実施形態では、このキットは、試料を得た対象の性別を同定するために性染色体SNPを検出するためのプライマーをさらに含む。ある種の実施形態では、このキットは、試料を得た対象の民族性を同定するために民族性SNPを検出するためのプライマーをさらに含む。ある種の実施形態では、このキットは、試料を得た免疫不全マウスの系統を決定するために供給業者のSNPを検出するためのプライマーをさらに含む。ある種の実施形態では、このキットは、試料中のウイルス感染またはマイコプラズマ夾雜を検出するためのプライマーをさらに含む。

10

【0088】

ある種の実施形態では、このキットは、ヒトまたはマウスSNPを含有するDNA断片を、これらのプライマーを使用して增幅するための薬剤をさらに含む。さらに、キットは、本明細書で提供した方法の実践のための指示（すなわち、プロトコル）を含有する指示資料を含んでいてもよい。指示資料は典型的に文書または印刷された資料を含むが、これだけに限定されない。このような指示を記憶することができ、末端使用者がこのような指示と連絡することができる任意の媒体が本発明では検討される。このような媒体には、限定されるものではないが、電子記憶媒体（例えば、磁気ディスク、テープ、カートリッジ、チップ）、光学媒体（例えば、CD-ROM）などが含まれる。このような媒体には、このような指示資料を提供するインターネットサイトへのアドレスを含めることができる。

20

【0089】

別の態様では、本開示は、例えば、Eds., Bowtell and Sambrook DNA Microarrays: A Molecular Cloning Manual (2003) Cold Spring Harbor Laboratory Pressに記載されるようなアレイスライドまたはチップなどの固相支持体に結合したオリゴヌクレオチドプローブを提供する。このような装置の構造は、例えば、米国特許および特許公開、米国特許第5,837,832号、PCT出願第WO95/11995号、米国特許第5,807,522号；米国特許第7,157,229号、第7,083,975号、第6,444,175号、第6,375,903号、第6,315,958号、第6,295,153号および第5,143,854号、第2007/0037274号、第2007/0140906号、第2004/0126757号、第2004/0110号、第2004/0110211号、第2003/0143550号、第2003/0003032号および第2002/0041420号に記載されたように、当業界では周知である。核酸アレイはまた、以下の参考文献に総括されている：Biotechnol Annu Rev (2002) 8: 85-101; Sosnowski et al. Psychiatr Genet (2002) 12(4): 181-92; Heller, Annu Rev Biomed Eng (2002) 4: 129-53; Kolchinsky et al., Hum. Mutat (2002) 19(4): 343-60; およびMcGail et al., Adv Biochem Eng Biotechnol (2002) 77: 21-42。

30

【0090】

マイクロアレイは、固相支持体に固定された、通常は合成アンチセンスポリヌクレオチドまたはcDNAの断片のいずれかである多数の固有の一本鎖ポリヌクレオチドから構成され得る。典型的なポリヌクレオチドは好ましくは約6~60ヌクレオチド長、より好ま

40

50

しくは約15～30ヌクレオチド長、最も好ましくは約18～25ヌクレオチド長である。ある特定の種類のアレイまたはその他の検出キット／系のために、ほんの約7～20ヌクレオチド長であるオリゴヌクレオチドを使用することが好ましいことがある。化学ルミネセンス検出技術と併せて使用したアレイなどの他の種類のアレイでは、好ましいプローブ長は、例えば、約15～80ヌクレオチド長、好ましくは約50～70ヌクレオチド長、より好ましくは約55～65ヌクレオチド長、最も好ましくは約60ヌクレオチド長であってもよい。

#### 【0091】

##### コンピュータが実現する方法、システムおよび装置

本明細書で記載した方法のいずれも、全体的にまたは部分的に、ステップを実施するために構成することができる1つまたは複数のプロセッサを含むコンピュータシステムで実施することができる。したがって、実施形態は、それぞれのステップまたはステップのそれぞれの群を実施する異なる構成成分を有する可能性のある、本明細書で記載した方法のいずれかのステップを実施するために構成されたコンピュータシステムを対象とする。番号付けしたステップとして提示されているが、本明細書の方法のステップは、同時に、または異なる順番で実施することができる。さらに、これらのステップの一部は、その他の方法の他のステップの一部と共に使用することができる。また、ステップの全部または一部は任意選択であってもよい。どの方法のどのステップも、これらのステップを実施するためにモジュール、回路またはその他の手段で実施することができる。

#### 【0092】

本明細書で言及したコンピュータシステムのいずれも、任意の適切な数のサブシステムを利用することができる。一部の実施形態では、コンピュータシステムは、サブシステムがコンピュータ機器の構成成分であり得る単一のコンピュータ機器を含む。その他の実施形態では、コンピュータシステムは、内部構成成分を有する、それぞれがサブシステムである多数のコンピュータ機器を含むことができる。サブシステムはシステムバスを介して相互接続することができる。さらなるサブシステムは、例えば、プリンター、キーボード、記憶装置、ディスプレイアダプターに結合されるモニターおよびその他を含む。I/O調節器と結合する周辺装置および入力／出力(I/O)装置は、シリアルポートなどの当業界で公知の任意の数の手段によってコンピュータシステムに連結することができる。例えば、シリアルポートまたは外部インターフェース(例えば、イーサネット登録商標、Wi-Fiなど)は、コンピュータシステムをインターネットなどの広域ネットワーク、マウス入力装置またはスキャナーに連結するために使用することができる。システムバスを介した相互接続は、中央処理装置が各サブシステムと連絡し、システムメモリまたは記憶装置(例えば、ハードドライブもしくは光学ディスクなどの固定ディスク)からの指示の実行を制御するのを可能にし、サブシステム間の情報交換を可能にする。システムメモリおよび／または記憶装置はコンピュータ可読媒体を組み入れることができる。本明細書で言及したデータのいずれも、1構成成分から別の構成成分への出力であってもよく、使用者への出力であってもよい。

#### 【0093】

コンピュータシステムは、例えば、外部インターフェースによって、または内部インターフェースによって一緒に連結した複数の同じ構成成分またはサブシステムを含むことができる。一部の実施形態では、コンピュータシステム、サブシステムまたは機器は、ネットワークと連絡することができる。このような場合、一コンピュータをクライアント、別のコンピュータをサーバーとみなすことができ、それぞれが同じコンピュータシステムの一部であってもよい。クライアントおよびサーバーはそれぞれ、多数のシステム、サブシステムまたは構成成分を含むことができる。

#### 【0094】

本開示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もしく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／またはモジュラー型もしくは一体型の一般的なプログラマブルプロセッサを有するコンピュータソフトウェアを使

10

20

30

40

50

用して、制御論理の形態で実現することができることを理解されたい。本明細書で使用する場合、プロセッサは、同じ集積チップ上にマルチコアプロセッサを含むか、または単一の回路基板上に多数の処理装置を含むか、またはネットワークされた多数の処理装置を含む。本明細書で提供した開示および教示に基づいて、当業者ならば、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合せを使用して、本開示の実施形態を実現するためのその他のやり方および／または方法に気づき、理解するだろう。

#### 【0095】

本出願で記載したソフトウェア構成成分または機能のいずれも、例えば、Java 登録商標、C++ または Perl などの任意の適切なコンピュータ言語を使用し、例えば、従来技術またはオブジェクト指向技術を使用して、プロセッサによって実行されるソフトウェアコードとして実現することができる。ソフトウェアコードは、記憶および／または伝達のためのコンピュータ可読媒体上の一連の指示または命令として記憶することができ、適切な媒体には、ランダムアクセスメモリ (RAM)、リードオンリメモリ (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、フラッシュメモリなどが含まれる。コンピュータ可読媒体はこのような記憶もしくは伝達装置の任意の組合せであってもよい。

#### 【0096】

このようなプログラムはまた、インターネットを含む様々なプロトコルに適合する有線、光学および／または無線ネットワークを介した伝達に適応したキャリア信号を使用して記号化され、伝達され得る。したがって、本発明の実施形態によるコンピュータ可読媒体は、このようなプログラムで記号化されたデータ信号を使用して作製することができる。プログラムコードで記号化されたコンピュータ可読媒体は、互換性がある装置にパッケージ化されてもよく、または（例えば、インターネットダウンロードを介して）その他の装置から別々に提供されてもよい。任意のこのようなコンピュータ可読媒体は、単一のコンピュータ製品（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、または内に存在していてもよく、システムまたはネットワーク内の異なるコンピュータ製品上または内に存在していてもよい。コンピュータシステムは、本明細書で言及した結果のいずれかを使用者に提供するために、モニター、プリンターまたはその他の適切なディスプレイを含んでいてもよい。

#### 【実施例】

#### 【0097】

以下の実施例は、特許請求した発明をよりよく例示するために提供され、本発明の範囲を限定しないものと解釈される。以下に記載した特定の組成物、材料および方法は全て、全体または一部が本発明の範囲内にある。これらの特定の組成物、材料および方法は、本発明を限定するものではなく、本発明の範囲内にある特定の実施形態を例示するに過ぎない。当業者は、創作能力を発揮することなく、かつ本発明の範囲を逸脱することなく、同等の組成物、材料および方法を開発することができる。本発明の範囲内にありながら、記載した本明細書の手法において多くの変更を行うことができることを理解されたい。このような変更が本発明の範囲内に含まれることは、本発明者らの意図である。

#### 【0098】

#### 【実施例1】

材料および方法

核酸抽出

細胞、PDX および PDXO からのゲノム DNA を、製造業者の指示に従って DNeasy Blood & Tissue Kit (QIAGEN, Cat. 69506, CA) を使用して精製した。DNA の完全性を、2100 Bioanalyzer (Agilent) によって決定し、NanoDrop (Thermo Scientific) を使用して定量した。高品質の DNA 試料 (OD260 / 280 = 1.8 ~ 2.0, OD260 / 230 = 2.0, > 1 μg) の 1 つのアリコートを、ディープNGS シーケンシン

10

20

30

40

50

グおよびW E S シークエンシングに使用した。細胞、P D X およびP D X O からの総 R N A を、製造業者の指示に従ってR Ne a s y M i n i K i t ( Q I A G E N 、 C a t . 7 4 1 0 6 、 C A ) を使用して精製した。総 R N A の完全性を、2 1 0 0 B i o a n a l y s e r ( A g i l e n t ) によって決定し、N a n o D r o p ( T h e r m o S c i e n t i f i c ) を使用して定量した。高品質の R N A 試料 ( O D 2 6 0 / 2 8 0 = 1 . 8 ~ 2 . 2 、 O D 2 6 0 / 2 3 0 = 2 . 0 、 R I N = 8 . 0 、 > 1  $\mu$  g ) の1つのアリコートを、ディープN G S シークエンシングおよびR N A s e q シークエンシングに使用した。

#### 【 0 0 9 9 】

##### 細胞株混合物の調製

細胞株混合物を、2つの細胞株からの細胞を所与の比率で混合することによって調製した。細胞増殖速度に基づいて、細胞を、T 7 5 中で1 5 m l の培地中に播種して、細胞のコンフルエンスを6 0 % ~ 8 0 % に到達させ、その後、C O <sub>2</sub> W a t e r J a c k e t e d I n c u b a t o r ( S A N Y O ) で一晩インキュベートした。細胞を対数増殖期間の間に回収し、血球計 ( C h o n g g u a n g ) でカウントして濃度を計算した。次いで、2つの細胞株からの細胞を、予め規定された比率に従って混合して、細胞株混合物を作製し、これを引き続いて、3, 0 0 0 r p m で5分間遠心分離した。上清を吸引し、細胞ペレットをD N A 抽出のために- 2 0 で貯蔵した。

#### 【 0 1 0 0 】

##### ヒト - マウスD N A 混合物の調製

一連のマウス - ヒトD N A 混合物ベンチマーク試料を、マウス脾臓D N A およびヒトゲノムD N A ( T h e r m o S c i e n t i f i c 、 C a t . 4 3 1 2 6 6 0 ) を混合することによって調製した。マウス脾臓D N A を、製造業者の指示に従ってD Ne a s y B l o o d & T i s s u e K i t ( Q I A G E N 、 C a t . 6 9 5 0 6 、 C A ) を使用して精製し、N a n o D r o p ( T h e r m o S c i e n t i f i c ) を使用して定量した。マウス脾臓D N A およびヒトゲノムD N A を2 0 0 n g /  $\mu$  L に希釈し、次いで、予め規定された比率で混合した。D N A 混合物を、後にディープN G S シークエンシングに使用した。

#### 【 0 1 0 1 】

##### ディープN G S シークエンシングをバーコード化する

多重P C R を使用して、1 5 0 b p のペアードエンド読み取り長さ ( p E 1 5 0 ) を用いるI l l u m i n a シークエンサーのための標的シークエンシングライブラリーを調製した。N G S ディープシークエンシングは、6 3 0 のアンプリコンをカバーし、それらのサイズは、1 6 0 b p ~ 2 6 0 b p の範囲であった。ゲノムD N A を、I G T - E M 8 0 8 ポリメラーゼ混合物 ( i G e n e T e c h B i o s c i e n c e C o . , L t d . , 9 5 で3分30秒間、9 8 で20秒間および6 0 で8分間のインキュベーション1 8 サイクル、7 2 で5分間維持) を使用して増幅し、次いで、A M P u r e X P ビーズ ( B e c k m a n 、 C a t . A 6 3 8 8 1 ) によって精製した。

#### 【 0 1 0 2 】

バーコード化を、第2ラウンドの増幅によって実行した。簡潔に述べると、精製された標的アンプリコンを鋳型とし、P C R 反応のための上流のI G T - I 5 インデックス ( 1 0  $\mu$  M ) 、下流のI G T - I 7 インデックス ( 1 0  $\mu$  M ) およびポリメラーゼ混合物と共に添加した。次いで、この混合物を、以下の設定での増幅のためにサーマルサイクラー中に置いた：9 5 で3分30秒間、9 8 で20秒間、5 8 で1分間および7 2 で3 0 秒間のインキュベーション9サイクル、7 2 で5分間固定。次いで、バーコード化されたライブラリーを、A M P u r e X P ビーズ ( B e c k m a n 、 C a t . A 6 3 8 8 1 ) を使用して精製した。

#### 【 0 1 0 3 】

ライブラリー構築後、Q u b i t 3 . 0 f l u o r o m e t e r d s D N A H S A s s a y ( T h e r m o F i s h e r S c i e n t i f i c ) を使用して、得られた

10

20

30

40

50

シークエンシングライブラーの濃度を定量した。Agilent BioAnalyzer 2100 (Agilent) を使用して、280 bp ~ 420 bp の範囲のサイズ分布を分析した。ペアードエンドシークエンシングを、2 × 150 bp ペアードエンドシークエンシングのための Illumina 提供のプロトコルに従って Illumina システムを使用して実施した。

#### 【0104】

##### RNAseq および WES シークエンシング

RNAseq シークエンシングでは、mRNA に焦点を当てたシークエンシングライブラーを、総 RNA から構築した。ポリ - A mRNA を、オリゴ - dT が結合した磁気ビーズを使用して総 RNA から精製し、次いで、断片化緩衝剤によって断片化した。短い断片を鋳型として使用して、第 1 鎖 cDNA を、逆転写酵素およびランダムプライマーを使用して合成し、その後、第 2 鎖 cDNA を合成した。次いで、合成された cDNA を、ライブラー構築プロトコルに従って、末端修復、リン酸化および「A」塩基付加に供した。次いで、シークエンシングアダプターを、cDNA 断片の両方の末端に付加した。cDNA 断片についての PCR 増幅後、標的化された 250 ~ 350 bp の断片を浄化した。ライブラー構築後、Qubit 3.0 fluorometer dsDNA HS Assay (Thermo Fisher Scientific) を使用して、得られたシークエンシングライブラーの濃度を定量し、サイズ分布を、Agilent BioAnalyzer 2100 (Agilent) を使用して分析した。ライブラーの検証後、HiSeq PE Cluster Kits (Illumina) と併せて Illumina cBOT クラスター生成システムを使用して、クラスターを生成した。ペアードエンドシークエンシングを、2 × 150 ペアードエンドシークエンシングのための Illumina 提供のプロトコルに従って Illumina システムを使用して実施した。

#### 【0105】

WES は、Wuxi Nextcode Co. Ltd. (Shanghai, China) が実施した。簡潔に述べると、ゲノム DNA を抽出し、180 ~ 280 bp の平均サイズに断片化した。DNA ライブラーを、Illumina の製造業者のペアードエンドプロトコルによって生成した。エキソンを、Agilent SureSelect Human All Exon V6 によって捕捉し、引き続いて、Illumina NovaSeq プラットフォーム (Illumina Inc., San Diego, CA, USA) によってシークエンシングして、150 bp のペアードエンド読み取りを生成した。

#### 【0106】

##### SNP の選択およびプロファイリング

本発明者らは、以下のいくつかの基準によって、ヒト試料認証のためのパネル SNP を選択した：1) SNP がエキソン中にある、2) 大きい染色体セグメントの欠失および重複を含む染色体の異常性が腫瘍において共通しているので、SNP が、22 個全ての常染色体上に位置し、互いに十分に離れている、3) SNP が、高度に発現される遺伝子中にある、4) SNP のマイナー対立遺伝子頻度 (MAF) が、International HapMap Project の 3 つの参照集団、すなわち、中国の漢族 (CHB)、ナイジェリアのヨルバ族 (YRI) ならびに CEPH コレクションからの北および西ヨーロッパ人祖先を持つユタ州住民 (CEU) において、0.5 に近い。

#### 【0107】

##### ベンチマーク試料およびデータ

2 つの細胞株ベンチマーク試料セットを調製した。第 1 のセットは、PANC-1 および RT4、MV-4-11 および「LNCaP クローン FGC」、CAL27 および Raji を含む 3 対の細胞株について、78 の試料を有する。各対は、純粋な 2 つの細胞株と、細胞カウントによる 8 つの混合比率についての 3 つの反復とを含む、26 の試料を有する (補足の表 S2)。第 2 のセットは、大抵は小さいが特定されない比率で既知の第 2 の細胞株が各々夾雜した 22 の細胞株を有する (補足の表 S3)。

10

20

30

40

50

## 【0108】

不均一性比率を推定する

ディープNGSシークエンシングデータから不均一性比率を推定するために使用され得る6つのインフォーマティブな遺伝子型組合せが存在する(表11)。これらは、4つの区別できるヌクレオチド頻度パターンを示す。組合せ1および2は、同じパターンを生じ、我々は、微量構成成分S2のパーセンテージ、すなわち不均一性比率を計算するために、平均式を使用する。この式は、SNPの数が大きい場合に接近して近似されるはずのシナリオである、2つの組合せが等しい頻度で生じる場合に、比率の正確な推定を生じる。類似の平均化アプローチが、組合せ4および5に使用される。不均一性比率が低い場合、シークエンシングエラーが、不均一性比率の推論を妨害し得る。これを軽減するために、我々は、2段階の統計手法を使用する。シークエンシングエラーを $e = 0.001$ と仮定し、所与のSNP部位におけるシークエンシング深度を $n$ ( $n > 500$ 、 $n < 500$ のSNPはいずれも破棄される)と仮定すると、 $k$ 個の誤ったヌクレオチドを観測する確率は、パラメータ $n$ および $e$ による二項分布に従う。

10

## 【0109】

【数6】

$$f(k, n, e) = \binom{n}{k} e^k (1 - e)^{n-k}$$

20

## 【0110】

各 $n$ について、我々は、累積密度関数を計算し、 $n$ 個のヌクレオチドのうち $h$ 個よりも多くの誤ったヌクレオチドを観測する確率が0.01よりも小さくなるような閾値 $h$ を得る。シークエンシングデータにおいて、対応する閾値 $h$ よりも小さい数の読み取りを有する低頻度ヌクレオチドはいずれも破棄される。次いで、我々は、最大不均一性(この研究では全ての試料について0.2を使用した)よりも小さいヌクレオチド頻度の分布をモデル化する混合ガウス(1~3つの成分を有する)のパラメータを推定するために、期待値最大化アルゴリズム(R用パッケージmclust、バージョン3.5.3(Team, R.C.R: A language and environment for statistical computing. 3.5.3 edn(R Foundation for Statistical Computing, Vienna, Austria.、2018)))を使用する。単一のガウス成分、または全てのデータポイントの60%よりも多くを占める最小平均を有するガウス成分のみが存在する場合、全てのデータポイントの中央値を試料不均一性比率とし、さもなくば、その他のガウス成分中のデータポイントの中央値を試料不均一性比率とする。

30

## 【0111】

試料の主要構成成分を決定する

SNP部位における遺伝子型が、参照試料については10%、夾雜されている可能性がある試験試料については25%である閾値よりも大きい対立遺伝子頻度を有するヌクレオチドのみを使用して決定される。参照試料と試験試料との間の遺伝子型類似性は、試験試料中の500未満のシークエンシング深度を有するSNPを除いた、同一な遺伝子型を有するSNPのパーセンテージである。試験試料の主要構成成分は、最も高い遺伝子型類似性を有する参照試料であり、この類似性は、試験試料の不均一性比率が<10%(または>10%)である場合には、90%(または80%)よりも高くなければならない。さもなくば、主要構成成分はコールされない。

40

## 【0112】

試料の微量構成成分を決定する

不均一性比率の推定および主要構成成分の決定の後、我々は、試験試料の微量構成成分を決定する。主要構成成分とその他の参照試料(例えば、ゲノムデータがある全ての細胞株)のうち1つとの混合物について、我々は、おそらくは1~4のヌクレオチドを有する

50

キメラ遺伝子型を、全ての SNP 部位において得る。ヌクレオチドの頻度が、不均一性比率を使用して計算される。同様に、我々は、試験試料のキメラ遺伝子型を得る。2つのキメラ遺伝子型は、それらが同じヌクレオチドを保有し、各ヌクレオチドの頻度が3倍以内である場合、同一とみなされる。次いで、我々は、試験試料と、主要構成成分と組み合わせた各参照試料との間の、遺伝子型類似性を計算する。次いで、全てのペアワイズ遺伝子型類似性のセットに、パラメータ( , )を用いるベータ分布がフィットされる。

【0113】

【数7】

$$f(x, \alpha, \beta) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1} (1-x)^{\beta-1}$$

10

【0114】

等式中、( )はガンマ関数であり、xは遺伝子型類似性である。そのパラメータを、R用パッケージ `fitdistplus` (バージョン3.5.3) によって推定した。次いで、フィットさせたベータ分布から、我々は、特定の値よりも大きい任意の遺伝子型類似性を観測する確率を計算した。99%信頼帯を伴う分位数-分位数グラフを、視覚化のために、全ての観測された遺伝子型類似性についてプロットした。参照試料は、(1)それが最も高い遺伝子型類似性を有する場合、(2)その遺伝子型類似性が、分位数-分位数グラフ中の99%信頼上限を上回る場合、および(3)フィットさせたベータ分布におけるそのp値 < 1.0E-6 である場合、微量構成成分とみなした。

20

【0115】

2つの細胞株の混合比率を推定する

細胞株が、2つの参照試料についての混合比率の推定を説明するために使用される。2つの細胞株 S1 および S2 を、S1 について、S2 について(1 - )の比率で混合すると仮定し、このとき、0 1 である。ディープNGSシークエンシングデータから、両方の細胞株中のn個全てのSNPのヌクレオチド頻度が、正確に推定され得る。SNPについて、その4つのヌクレオチド頻度が、細胞株 S1 については {A1, T1, G1, C1}、細胞株 S2 については {A2, T2, G2, C2} として示され、これらは合計すると 1 になる。原理上、頻度のうち 1 つは、SNP がホモ接合性である場合には 1 に近く、2つの頻度は共に、SNP がヘテロ接合性である場合には 0.5 に近い。実際のデータは、シークエンシングエラーおよびランダムさ、ならびに細胞株の多クローン性に起因して、いくらかの偏差を有し得る。

30

【0116】

混合試料のシークエンシングデータから、4つのヌクレオチドの実際の発生率は、 $x = \{n_A, n_T, n_G, n_C\}$  として示される。このような観測の尤度は、

【0117】

【数8】

$$\mathcal{L}(\theta|x) = P_\theta(x) = \text{const} \times \prod_{M \in \{A, T, G, C\}} (\theta M_1 + (1-\theta)M_2)^{n_M}$$

40

である。

【0118】

尤度  $P(x_i)$  は、観測されたデータ  $x_i$  を用いて、任意の SNP  $i$  (1, 2, ..., n) について計算され得、全ての SNP についてデータ  $X = \{x_1, x_2, \dots, x_n\}$  を観測する尤度は、

【0119】

【数9】

50

$$\mathcal{L}(\theta|X) = \text{const} \times \prod_{i=1}^n P_\theta(x_i)$$

である。

【0120】

したがって、log - 尤度は、

【0121】

【数10】

10

$$\log \mathcal{L}(\theta|X) = \sum_{i=1}^n \log P_\theta(x_i)$$

である。

【0122】

次いで、尤度を最大化する が、 の段階的増加によって解かれ得る。上記手法は、任意の 2 つのヒト試料の混合物についても同様に使用され得る。

【0123】

20

夾雑物検出のための細胞株混合物のシミュレーション

シミュレーションを、PANC-1 および RT4、MV-4-11 および「LNCaP クローンFGC」、CAL27 および Raji を含む 3 つの細胞株対について実施した。6 つ全ての細胞株を、ディープNGS シークエンシングによってプロファイリングして、それらの SNP フィンガープリントを得た。対になった 2 つの細胞株を、in silico で混合したが、第一の細胞株の比率は  $r$  であり、 $r$  は、以下の値をとる：0.15%、0.30%、0.625%、1.25%、2.5%、5%、10%、15% および 20%。各 SNP 部位について、 $r \times n$  個のヌクレオチドを第 1 の細胞株から得、式中、 $n$  は、500 ~ 5000 のランダムな整数であり、 $r \times n$  を、第 1 の細胞株におけるそれらの頻度に従って、4 つのヌクレオチド (A、T、G、C) にさらに振り分けた。同様に、(1 -  $r$ )  $\times n$  個のヌクレオチドを、第 2 の細胞株から得た。次いで、比率を逆転させ、そして、対称サンプリングを、第 2 の細胞株について比率  $r$  で実施した。

30

【0124】

RNAseq および WES データセットからマウス比率を推定する

シークエンシング読み取りを、デフォルトパラメータを用いて、RNAseq データについてはマッピングツール STAR (Dobin, A. et al. STAR: ultrafast alignment of short reads to large reference genomes using Burrows-Wheeler transform. Bioinformatics 29, 15-21 (2013)) を使用し、WES データについては BWA (Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25, 1754-60 (2009)) を使用して、ヒト (hg19) およびマウス (mm10) ゲノムにマッピングした。読み取りがヒトゲノムのみにマッピングされた場合、またはマウスゲノムに対してよりもヒトゲノムに対しての不一致がより少なかった場合、その読み取りは、ヒト読み取りとして分類した。マウス読み取りも同様に割り当てた。読み取りが、多くても 2 ずれた近い数の不一致で両方のゲノムにマッピングされた場合、その読み取りは分類不能であり、破棄した。マウス比率は、全ての保持した読み取りのうちのマウス読み取りの割合であった。

40

【0125】

[実施例 2]

この実施例は、ヒト試料認証および夾雑検出を例示する。

50

## 【0126】

## S N P プロファイリングおよびフィンガープリント

S N P のパネルを、細胞株、異種移植片およびオルガノイドを含むヒト試料を認証するために選択した（表1）。S N P を、3 0 0 0 の平均深度を用いるディープN G S シークエンシングによってプロファイリングした。各試料は、全てのS N P について、ヌクレオチド同一性および頻度の両方からなる固有のS N P フィンガープリントを有する。細胞株は、遺伝的浮動および不均一性に起因して、継代間およびバイオバンク間でゆらぐS N P フィンガープリントを有し得、したがって、現在のS N P フィンガープリントが、より良いキュレーションのためにプロファイリングされ得ることが強調される。S N P フィンガープリントは、比較的低深度のN G S データによって、低減された正確さで生成され得る。この実施例では、本発明者らは、本発明者らおよびC C L E がプロファイリングしたR N A s e q データから、1 0 5 0 の細胞株についてのS N P フィンガープリントを生成した。これは、参照として機能する。

## 【0127】

本発明者らは、2 1 7 の細胞株試料、2 2 0 のP D X および3 1 のP D X 由来オルガノイド（P D X O ）試料についてのディープN G S シークエンシングからのS N P プロファイリングデータを使用した、認証、特徴づけ、種内および種間夾雜検出を例示した。細胞株試料について、本発明者らは、連続希釈からの既知の混合比率の2つの細胞株の混合物および6つの対応する純粋な細胞株（表7）、未知の混合比率の2つの細胞株の混合物（表8）、ならびに1 1 7 の未混合の細胞株（表9）を試験した。

## 【0128】

## ヒト試料の認証

試料の正体、または夾雜された試料の主要構成成分を、参照試料のライブラリーに対するその遺伝子型類似性によって決定した。2 1 7 の試験した細胞株試料において、同じ細胞株間の遺伝子型類似性は、9 8 . 6 % の平均で常に > 9 0 % であり、最低は、J E G - 3 の1 6 . 7 % の夾雜を伴った、A - 8 7 5 細胞培養物についての9 1 . 7 % であった（図1 A、表8）。対照的に、無関係の細胞株間の遺伝子型類似性は、ほぼ常に5 0 % を下回った。表示ミス、夾雜、同じ患者に由来すること、1つの細胞株が別の細胞株の親であることなどを含む様々な理由によって、密接に関係するまたは同じ同義の群中にある細胞株がなおも存在した。例えば、H C T - 1 5 およびH C T - 8 は、同じ患者に由来した可能性が高い；Q G Y - 7 7 0 1 が夾雜し、これはH e L a 誘導体である。データセット中の1 6 のこのような細胞株対についての遺伝子型類似性は、8 4 % ~ 9 6 % の範囲である（表10）。これらの細胞株対は、H L E およびH L F などのほぼ同一なものを除き、区別することができる。同じモデル間の遺伝子型類似性は、平均して、2 2 0 のP D X および3 1 のP D X 由来オルガノイド（P D X O ）試料については9 8 . 0 % ( 8 7 . 2 ~ 1 0 0 % ) であり、ほぼ全てが、異なるモデル間で5 0 % を下回る。

## 【0129】

## 遺伝的不均一性の推定

試料が夾雜なしであり、純粋に単クローニ性の二倍体である場合、S N P 部位は、ホモ接合性またはヘテロ接合性のいずれかであり、観測されたヌクレオチド頻度は、ディープN G S シークエンシングデータにおいて1または0 . 5 に近く、この違いは、シークエンシングにおけるエラーおよびランダムさのみから来ている。実際には、細胞株は、微量クローニを有している可能性があり、異数体であり、または夾雜されており（夾雜物）、したがって、本発明者らは、S N P 部位において、0 . 5 および1からかけ離れた頻度を観測しただけでなく、3または4のヌクレオチドもまた観測した。このような情報は、試料の遺伝的不均一性を推定するために使用することができる。

## 【0130】

支配的なクローニは、試料の主要構成成分であり、微量クローニおよび夾雜物は、微量構成成分である。4つの観測されたヌクレオチド頻度パターンに基づいて、S N P 不均一性比率を推定するために使用され得る主要構成成分および微量構成成分の6つのインフォ

10

20

30

40

50

ーマティブな遺伝子型組合せが存在する（表11）。S N P部位は、4つのパターンのうち1つを生じる場合、インフォーマティブである。引き続いて、試料不均一性比率を、統計的モデル化アプローチによって、個々のS N P不均一性比率から推定する（実施例1を参照のこと）。試験試料を使用して、本発明者らは、夾雜なしの細胞株が、平均して、107のインフォーマティブなS N P部位を有するが、夾雜された細胞株が、わずかに多い112を有することを見出した。平均して、P D XおよびP D X Oモデルは、それぞれ、156および111のインフォーマティブなS N P部位を有し、これは、P D Xモデルにおけるより高い遺伝的不均一性および／またはマウス夾雜を反映している。

### 【0131】

#### 夾雜の検出および定量

本発明者らは、3つの分析を組み合わせることによって、試料夾雜を検出した。第1に、夾雜された試料は、高い不均一性比率を有し得るが、夾雜なしの試料はそれを有さない。試験試料において、118の推定上夾雜なしの細胞株のうち115（97.5%）が、不均一性比率<2%を有し、全てが<3%を有した（図1B）。対照的に、本発明者らは、夾雜された細胞株について高い不均一性比率を観測し、例えば、J E G - 3細胞と混合されたA - 875細胞培養物は、15.5%の不均一性比率を有した（表8）。上で示したように、不均一性比率は、夾雜比率（夾雜物のパーセンテージ）に比例し、したがって、夾雜の良好な指標である。P D Xモデルから切開されたヒト腫瘍は、マウス間質を含有し、実際、本発明者らは、マウス夾雜（図1C）によって引き起こされる、P D X腫瘍におけるより高い不均一性比率（図1B）を観測した。P D X Oは、P D Xのin vitro培養物として、はるかに小さいおよびしばしば痕跡量のみのマウス細胞に起因して、有意に小さい不均一性比率を有する（図1B）。

10

### 【0132】

夾雜は、試料について、S N P不均一性比率の確率密度における区別できる右ピークによって示された（図2A～2F）。夾雜および不均一性比率が増加するにつれて、ピークは右にシフトし、時には2つのピークに分かれる。二／三峰性分布は、消滅したか、または夾雜なしの細胞株もしくは非常に低い夾雜比率（<1%）および不均一性比率（<2%）の細胞株についてわずかに出現したに過ぎない。

20

### 【0133】

最後に、夾雜物は、直感的な視覚化および厳密な確率論的測定を与える統計的モデル化によって直接検出することができる（実施例1、図3Aを参照のこと）。各々別の細胞株と混合された94の細胞株試料において、本発明者らは、不均一性比率が2%である場合、細胞株中の微量夾雜物細胞株を常に正確に推論することができる（図3B）。不均一性比率が1～2%および<1%であるとき、精度は、約80%および50%まで下がる。8つの不首尾に終わった試料について、7つの試料がクリーンであると特徴づけられ、1つだけが、誤った夾雜している細胞株と評価された。もちろん、このような推論は、夾雜している細胞株も既知のS N Pフィンガープリントを有するものである場合にもっともらしいに過ぎない。本発明者らは、我々のバイオバンクにおいていくつかの夾雜された細胞株を検出し、1つの例は、細胞株「G - 292 クローンA 141 B 1」であり、これは、7.62%の高い不均一性比率を有し（図3C）、6.21%のO C I - A M L - 2が夾雜していた（図3D）。

30

### 【0134】

夾雜している細胞株を同定した後で、本発明者らは、最尤アプローチを使用して、夾雜比率（すなわち、第2の細胞株のパーセンテージ）を推定することができる（実施例1を参照のこと）。シミュレーション研究により、推定された夾雜比率が、既知の比率と極めて近いことが示された（図3E）。本発明者らは、不均一性比率と夾雜比率との間の緊密な線形の相関を観測した（図3F）。したがって、以前に議論したように、不均一性比率は、夾雜の良好な推定子であり、夾雜物が標準的な細胞株でない場合に特に有用である。やはり夾雜された試料においてであるが、夾雜物は、時には遺伝的不均一性の大部分に寄与するとはいえ、その一部に寄与するだけであり、その結果、夾雜比率は、対応する不均

40

50

一性比率よりも概して小さく（表8を参照のこと）、わずかな逸脱が、データ処理方法によって引き起こされた。

【0135】

要約すると、その値および分布による不均一性比率は、ヒト試料についての信頼性のある夾雑尺度である。不均一性比率 2 %を有する細胞株試料は、夾雑されている可能性が非常に高く、夾雑物がこれもまた SNP フィンガープリント情報を有する別の細胞株である場合、その正体が推論され得、夾雑比率は、細胞またはDNA混合比率によって測定すると、1 %で、前例のない感度で推定され得る（表7および8）。

【0136】

【実施例3】

この実施例は、マウス腫瘍モデル認証を例示する。

【0137】

マウスSNPのパネル（表2を参照のこと）を、4T1、A20、B16-BL6、B16-F0、B16-F1、B16-F10、C1498、Colon26、CT26WT、E.G7-Ova、EL4、EMT6、H22、Hepa1-6、J558、J774A1、JC、KLN205、L1210、L5178-R、LLC、MBT2、MC38、MPC-11、Neuro-2a、P388D1、P815、Pan02、Renca、RM1、S91およびWEHI164を含む、前臨床免疫調節薬物開発において一般に使用される32の同系マウス腫瘍モデルを認証するために選択した。ほとんどのモデルは、6つの固有のSNPを有する。Colon26およびCT26WTは、BALB/cマウス系統に起源するマウス結腸腺癌モデルであり、各々、12のSNPを6つの共通のSNPと併せて有し、合計18の固有のSNPを有する。B16-BL6、B16-F0、B16-F1およびB16-F10は、C57BL/6マウス系統におけるマウス黒色腫細胞株であり、全てB16に由来し、したがって、高い遺伝的類似性を共有する。具体的には、B16は、B16-F0の親株であり、順に、B16-F0は、B16-F1の親株である。B16-F10は、B16-F0の10代目の連続継代であり、B16-BL6<sup>46</sup>の親株である。本発明者らは、7つの共通のSNPを使用して、試験細胞株をこの群に最初に割り当て、次いで、6つの固有のSNPを各々が有するB16-BL6、B16-F0およびB16-F10に割り当て、18のSNPのいずれも観測されない場合には、試験細胞株にはB16-F1が割り当てられる。これらのモデルについての認証は、100%の精度を達成した。

10

20

30

40

【0138】

【実施例4】

この実施例は、ヒト-マウス種間夾雑検出を例示する。

【0139】

本発明者らは、ヒトhg19およびマウスマム10ゲノムを比較し、一群の100~300bpのセグメントを同定した（表3を参照のこと）が、その結果、各セグメントは、ヒトとマウスとの間で、挿入、欠失および点変異によって有意に異なっているが（31~97%の配列類似性）、共通のプライマー対が設計できるような同一な隣接配列をなおも有する。NGSシークエンシングの後、本発明者らは、ヒト読み取りとマウス読み取りとを分離し、全てのセグメントについてマウス比率を計算し、これらの比率の中央値を、ヒト-マウス混合試料中のマウス比率とした。この方法は、マウスおよびヒトのDNAが連続希釈によって混合されたベンチマーク試料のセットにおいて極めて高い精度を実証した（図4A）。本発明者らは、RNAseqおよびWESデータからマウス含量を推定する方法もまた開発した（実施例1を参照のこと）。本発明者らは、220のPDXおよび31のPDXOモデルにおいてマウス比率を推定することにおいて、3つの方法を比較した（図4B~C）。DNA（WESおよびディープNGSシークエンシングのため）およびRNA（RNAseqのため）を抽出し、モデルの同じ試料からシークエンシングして、試料の変動を除去した。PDXOモデルは概して、低いマウス含量を有した。PDXモデルでは、ディープNGSシークエンシングデータから正確に推定されたマウス比率は最も

50

高く、次がRNAseqからのものであり、次がWESからのものであった。これは、WESにおいて使用したエキソン捕捉キットがヒトエキソンを富化するように設計されたものであり、相同的なマウスエキソンに対する低いハイブリダイゼーション親和性を有したこととに主に起因する。RNAseqは、種を優先しないポリA富化プロトコルを使用したが、遺伝子発現は、PDXのヒト腫瘍およびマウス間質において、大きな時空間的可変性を有する。実際、本発明者らは、ディープNGSシークエンシングデータとWESデータとの間に、マウス比率について非常に強い二次関係を観測したが( $R = 0.96$ 、図4D)、ディープシークエンシングデータとRNAseqデータとの間には、はるかに弱い線形の相関を観測した( $R = 0.62$ )。

【0140】

10

[実施例5]

この実施例は、試料中のマイコプラズマの検出を例示する。

【0141】

本発明者らは、有効性が証明された、全てのマイコプラズマ種の検出のための一対のユニバーサルプライマー、ならびに*A. laidlawii*、*M. arginine*、*M. fermentans*、*M. genitalium*、*M. hominis*、*M. hyo rhinis*、*M. orale*、*M. pneumonia*、*M. salivarium*および*U. urealyticum*を含む11のモリキュートを検出するための11対のプライマーを使用した(Molla Kazemiha, V. et al. Cytotechnology 61, 117-24 (2009))。本発明者らは、ディープNGSシークエンシング方法によって、バイオバンクにおいて1つのマイコプラズマ夾雜された細胞株を同定し、引き続いて、マイコプラズマ検出キットによってそれを検証した。

20

【0142】

[実施例6]

この実施例は、集団構造分析および性別決定を例示する。

【0143】

ヒト試料認証に使用したSNPのパネルのうち、143は、International HapMap Project (International HapMap, C. The International HapMap Project. Nature 426, 789-96 (2003))によって特徴づけられたものである。本発明者らは、fastSTRUCTURE(Raj, A., Stephens, M. & Pritchard, J. K. Genetics 197, 573-89 (2014))を使用して、以下の3つの参照集団の集団構造分析を実施した：中国の漢族(CHB)、ナイジェリアのヨルバ族(YRI)ならびにCEPHコレクションからの北および西ヨーロッパ人祖先を持つユタ州住民(CEU)。406人全ての個体を、高い確率で一義的に割り当てた。次いで、本発明者らは、東アジア人患者に由来する423のPDXモデルおよび米国の西洋人患者に由来する634のPDXモデルをプロファイリングした。全ての東アジア人PDXモデルは、1つの例外だけを除いて、支配的なCHB組成を有している。西洋人PDXモデルの大部分は、CEU組成を支配的に有し、残りは、主要なCHBもしくはYRI組成、または参照集団のうち2つもしくは3つの混合物を有している。本発明者らは、性別推論のためにY染色体にある3つのSNPも使用し(表3)、これは、Y染色体が失われた腫瘍試料を除き、常に正確であった。

30

【0144】

40

本開示は、特定の実施形態(そのうちのいくつかは好ましい実施形態である)を参照しながら特に示し説明したが、当業者は、本明細書で開示した本開示の精神および範囲を逸脱することなく形態および詳細の様々な変化を行うことができることを理解されたい。

【0145】

50

参考文献

1. Identity crisis. Nature 457, 935-6 (2009).

2. American Type Culture Collection Standards Development Organization Wo

- rkgroup, A.S.N. Cell line misidentification: the beginning of the end. *Nat Rev Cancer* 10, 441-8 (2010).
3. Capes-Davis, A. et al. Match criteria for human cell line authentication: where do we draw the line? *Int J Cancer* 132, 2510-9 (2013).
4. Gartler, S.M. Apparent Hela cell contamination of human heteroploid cell lines. *Nature* 217, 750-1 (1968).
5. Lacroix, M. Persistent use of "false" cell lines. *Int J Cancer* 122, 1-4 (2008).
6. Lorsch, J.R., Collins, F.S. & Lippincott-Schwartz, J. Cell Biology. Fixing problems with cell lines. *Science* 346, 1452-3 (2014).
7. Fusenig, N.E., Capes-Davis, A., Bianchini, F., Sundell, S. & Lichter, P. The need for a worldwide consensus for cell line authentication: Experience implementing a mandatory requirement at the International Journal of Cancer. *PLoS Biol* 15, e2001438 (2017). 10
8. Yu, M. et al. A resource for cell line authentication, annotation and quality control. *Nature* 520, 307-11 (2015).
9. Bian, X., Yang, Z., Feng, H., Sun, H. & Liu, Y. A Combination of Species Identification and STR Profiling Identifies Cross-contaminated Cells from 482 Human Tumor Cell Lines. *Sci Rep* 7, 9774 (2017).
10. Horbach, S. & Halffman, W. The ghosts of HeLa: How cell line misidentification contaminates the scientific literature. *PLoS One* 12, e0186281 (2017). 20
11. de Maagd, R.A. et al. Identification of *Bacillus thuringiensis* delta-endotoxin Cry1C domain III amino acid residues involved in insect specificity. *Appl Environ Microbiol* 65, 4369-74 (1999).
12. Azari, S., Ahmadi, N., Tehrani, M.J. & Shokri, F. Profiling and authentication of human cell lines using short tandem repeat (STR) loci: Report from the National Cell Bank of Iran. *Biologicals* 35, 195-202 (2007).
13. Wu, M.L. et al. A 2-yr service report of cell line authentication. *In Vitro Cell Dev Biol Anim* 49, 743-5 (2013).
14. Masters, J.R. HeLa cells 50 years on: the good, the bad and the ugly. *Nat Rev Cancer* 2, 315-9 (2002). 30
15. MacLeod, R.A. et al. Widespread intraspecies cross-contamination of human tumor cell lines arising at source. *Int J Cancer* 83, 555-63 (1999).
16. Cosme, B. et al. Are your results valid? Cellular authentication a need from the past, an emergency on the present. *In Vitro Cell Dev Biol Anim* 53, 430-434 (2017).
17. Ye, F., Chen, C., Qin, J., Liu, J. & Zheng, C. Genetic profiling reveals an alarming rate of cross-contamination among human cell lines used in China. *FASEB J* 29, 4268-72 (2015).
18. Freedman, L.P. et al. The culture of cell culture practices and authentication--Results from a 2015 Survey. *Biotechniques* 59, 189-90, 192 (2015). 40
19. Nims, R.W. & Reid, Y. Best practices for authenticating cell lines. *In Vitro Cell Dev Biol Anim* 53, 880-887 (2017).
20. Almeida, J.L., Cole, K.D. & Plant, A.L. Standards for Cell Line Authentication and Beyond. *PLoS Biol* 14, e1002476 (2016).
21. Almeida, J.L. et al. Interlaboratory study to validate a STR profiling method for intraspecies identification of mouse cell lines. *PLoS One* 14, e0218412 (2019).
22. Zaaijer, S. et al. Rapid re-identification of human samples using portable DNA sequencing. *Elife* 6(2017).
23. Yousefi, S. et al. A SNP panel for identification of DNA and RNA specimen 50

- s. *BMC Genomics* 19, 90 (2018).
24. Jobling, M.A. & Gill, P. Encoded evidence: DNA in forensic analysis. *Nat Rev Genet* 5, 739-51 (2004).
25. Sanchez, J.J. et al. A multiplex assay with 52 single nucleotide polymorphisms for human identification. *Electrophoresis* 27, 1713-24 (2006).
26. Didion, J.P. et al. SNP array profiling of mouse cell lines identifies their strains of origin and reveals cross-contamination and widespread aneuploidy. *BMC Genomics* 15, 847 (2014).
27. Liang-Chu, M.M. et al. Human biosample authentication using the high-throughput, cost-effective SNPtrace(TM) system. *PLoS One* 10, e0116218 (2015). 10
28. Pengelly, R.J. et al. A SNP profiling panel for sample tracking in whole-exome sequencing studies. *Genome Med* 5, 89 (2013).
29. Morgan, A.P. et al. The Mouse Universal Genotyping Array: From Substrains to Subspecies. *G3 (Bethesda)* 6, 263-79 (2015).
30. Castro, F. et al. High-throughput SNP-based authentication of human cell lines. *Int J Cancer* 132, 308-14 (2013).
31. El-Hoss, J. et al. A single nucleotide polymorphism genotyping platform for the authentication of patient derived xenografts. *Oncotarget* 7, 60475-60490 (2016).
32. Ruitberg, C.M., Reeder, D.J. & Butler, J.M. STRBase: a short tandem repeat DNA database for the human identity testing community. *Nucleic Acids Res* 29, 320-2 (2001). 20
33. van der Meer, D. et al. Cell Model Passports-a hub for clinical, genetic and functional datasets of preclinical cancer models. *Nucleic Acids Res* 47, D923-D929 (2019).
34. Tuveson, D. & Clevers, H. Cancer modeling meets human organoid technology. *Science* 364, 952-955 (2019).
35. Day, C.P., Merlino, G. & Van Dyke, T. Preclinical mouse cancer models: a maze of opportunities and challenges. *Cell* 163, 39-53 (2015).
36. Guo, S. et al. Molecular Pathology of Patient Tumors, Patient-Derived Xenografts, and Cancer Cell Lines. *Cancer Res* 76, 4619-26 (2016). 30
37. Khaled, W.T. & Liu, P. Cancer mouse models: past, present and future. *Semin Cell Dev Biol* 27, 54-60 (2014).
38. Li, Q.X., Feuer, G., Ouyang, X. & An, X. Experimental animal modeling for immuno-oncology. *Pharmacol Ther* 173, 34-46 (2017).
39. Chao, C. et al. Patient-derived Xenografts from Colorectal Carcinoma: A Temporal and Hierarchical Study of Murine Stromal Cell Replacement. *Anticancer Res* 37, 3405-3412 (2017).
40. Fasterius, E. & Al-Khalili Szigyarto, C. Analysis of public RNA-sequencing data reveals biological consequences of genetic heterogeneity in cell line populations. *Sci Rep* 8, 11226 (2018).
41. Ghandi, M. et al. Next-generation characterization of the Cancer Cell Line Encyclopedia. *Nature* 569, 503-508 (2019).
42. Vermeulen, S.J. et al. Did the four human cancer cell lines DLD-1, HCT-15, HCT-8, and HRT-18 originate from one and the same patient? *Cancer Genet Cytogenet* 107, 76-9 (1998).
43. Rebouissou, S., Zucman-Rossi, J., Moreau, R., Qiu, Z. & Hui, L. Note of caution: Contaminations of hepatocellular cell lines. *J Hepatol* 67, 896-897 (2017).
44. Barretina, J. et al. The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature* 483, 603-7 (2012). 50

45. Bairoch, A. The Cellosaurus, a Cell-Line Knowledge Resource. *J Biomol Tech* 29, 25-38 (2018).
46. Molla Kazemiha, V. et al. PCR-based detection and eradication of mycoplasmal infections from various mammalian cell lines: a local experience. *Cytotechnology* 61, 117-24 (2009).
47. International HapMap, C. The International HapMap Project. *Nature* 426, 789-96 (2003).
48. Raj, A., Stephens, M. & Pritchard, J.K. fastSTRUCTURE: variational inference of population structure in large SNP data sets. *Genetics* 197, 573-89 (2014).
49. Masters, J.R. et al. Short tandem repeat profiling provides an international reference standard for human cell lines. *Proc Natl Acad Sci U S A* 98, 8012-7 (2001).
50. Hideyuki Tanabe, Y.T., Daisuke Minegishi, Miharu Kurematsu, Tohru Masu, Hiroshi Mizusawa. Cell line individualization by STR multiplex system in the cell bank found cross-contamination between ECV304 and EJ-1/T24. *Tiss. Cult. Res. Commun.* 18, 329-338 (1999).
51. Team, R.C. R: A language and environment for statistical computing. 3.5.3 edn (R Foundation for Statistical Computing, Vienna, Austria., 2018).
52. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29, 15-21 (2013).
53. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25, 1754-60 (2009).

【 0 1 4 6 】

10

20

30

40

50

【表 1 - 1】

表 1. ヒト SNP

遺伝子座	Rs 番号	配列番号	遺伝子座	Rs 番号	配列番号
chr1_10473196	rs2229687	1	chr11_118967758	rs643788	123
chr1_1222267	rs11260579	2	chr11_120107411	rs882856	124
chr1_1249187	rs12142199	3	chr11_122928622	rs4802	125
chr1_151733335	rs8480	4	chr11_125479363	rs2241502	126
chr1_169345868	rs1028180	5	chr11_3028140	rs729662	127
chr1_169519112	rs6020	6	chr11_4159457	rs9937	128
chr1_201969082	rs1130790	7	chr11_47188411	rs3740691	129
chr1_201973565	rs3820439	8	chr11_65632262	rs558114	130
chr1_202304868	rs14451	9	chr11_6630833	rs1043390	131
chr1_20977000	rs1043424	10	chr11_67200819	rs4930427	132
chr1_20982631	rs4704	11	chr11_82973004	rs8789	133
chr1_220154768	rs1061160	12	chr12_104341103	rs7645	134
chr1_226352498	rs2306120	13	chr12_109994870	rs9593	135
chr1_230829139	rs1051038	14	chr12_112037000	rs695871	136
chr1_234529570	rs2175593	15	chr12_118473054	rs9788041	137
chr1_234614390	rs117698521	16	chr12_12967127	rs1051374	138
chr1_46095272	rs1135812	17	chr12_2968169	rs3742076	139
chr1_46660295	rs2292487	18	chr12_2997397	rs2907608	140
chr1_52264064	rs1770791	19	chr12_49230035	rs1057908	141
chr1_54683856	rs15921	20	chr12_50529736	rs3741562	142
chr2_10712278	rs3732114	21	chr12_6638116	rs740850	143
chr2_109513601	rs3827760	22	chr12_6647109	rs1803621	144
chr2_109543883	rs922452	23	chr12_67706466	rs1060350	145
chr2_109552936	rs3215127	24	chr13_111298392	rs436462	146
chr2_109580638	rs260691	25	chr13_115004914	rs2296971	147
chr2_207006676	rs1801318	26	chr13_25000617	rs7571	148
chr2_220046975	rs3731900	27	chr13_28239970	rs14105	149
chr2_232326417	rs1131171	28	chr14_102514227	rs13749	150
chr2_238672703	rs3739038	29	chr14_105222037	rs1132975	151
chr2_242572846	rs3208142	30	chr14_106208082	rs11621259	152
chr2_242618050	rs1131195	31	chr14_106208086	rs1045853	153
chr2_27260469	rs1124649	32	chr14_106236128	rs12890621	154
chr2_27550967	rs1049817	33	chr14_21967916	rs1139130	155
chr2_33226533	rs4952330	34	chr14_24615435	rs4575	156
chr2_3504687	rs9950	35	chr14_24736027	rs14193	157
chr2_68388823	rs1137930	36	chr14_49294391	rs34609389	158
chr2_69659126	rs4453725	37	chr14_51716188	rs7161242	159
chr2_71365676	rs357756	38	chr14_64908845	rs2236225	160
chr2_85769711	rs1078004	39	chr14_75359670	rs2230237	161
chr2_99995804	rs1376443	40	chr15_40328665	rs8208	162
chr3_108188993	rs9868484	41	chr15_44038899	rs2411284	163
chr3_121500699	rs17849995	42	chr15_48384907	rs2250072	164
chr3_14174427	rs4685076	43	chr15_48426484	rs1426654	165
chr3_183560195	rs13091	44	chr15_48485926	rs2413887	166

10

20

30

40

50

【表 1 - 2】

表 1. ヒト SNP

遺伝子座	Rs 番号	配列番号	遺伝子座	Rs 番号	配列番号
chr3_183861243	rs843358	45	chr15_52901433	rs12915981	167
chr3_186509517	rs187868	46	chr15_63937209	rs2229749	168
chr3_193374964	rs9851685	47	chr15_63988357	rs2255243	169
chr3_33907945	rs3183987	48	chr15_75189930	rs1130741	170
chr3_50334231	rs2269432	49	chr15_75650836	rs1128933	171
chr3_50355730	rs35455589	50	chr15_75932129	rs13737	172
chr3_50378176	rs4688725	51	chr15_77344793	rs11737	173
chr3_52727257	rs2289247	52	chr15_89858602	rs1138465	174
chr3_58154327	rs8640	53	chr15_91525197	rs2301826	175
chr4_10099340	rs13441	54	chr16_11773662	rs3190321	176
chr4_1330759	rs1128427	55	chr16_15129970	rs7200543	177
chr4_164435265	rs2304802	56	chr16_2049640	rs2286469	178
chr4_1737502	rs11248073	57	chr16_2285357	rs26840	179
chr4_183815688	rs4742	58	chr16_27238110	rs1127228	180
chr4_186097045	rs6855305	59	chr16_70515355	rs11054	181
chr4_25419283	rs9174	60	chr16_70602221	rs12909	182
chr4_39458051	rs2125313	61	chr16_718514	rs7204542	183
chr4_57889677	rs1056364	62	chr16_75590092	rs3743601	184
chr4_83795806	rs10025654	63	chr16_75646576	rs3743599	185
chr5_137902339	rs1042665	64	chr16_81010073	rs1127390	186
chr5_140048209	rs2251860	65	chr17_19247075	rs4924987	187
chr5_141014494	rs2530223	66	chr17_38179492	rs2302777	188
chr5_176940384	rs335438	67	chr17_40722029	rs665268	189
chr5_179290845	rs30386	68	chr17_5294976	rs14231	190
chr5_311478	rs2244029	69	chr17_61908556	rs13030	191
chr5_31409252	rs2241337	70	chr17_7217463	rs2292064	192
chr5_33951693	rs16891982	71	chr17_73016621	rs1044228	193
chr5_33964210	rs183671	72	chr17_74056413	rs2665998	194
chr5_61694379	rs247264	73	chr17_80008392	rs9916764	195
chr5_72798845	rs14010	74	chr17_80039481	rs1140616	196
chr5_74069863	rs6874609	75	chr18_12351342	rs11080572	197
chr6_105726036	rs1051484	76	chr18_33750046	rs8299	198
chr6_150114745	rs4816	77	chr18_77805856	rs3744872	199
chr6_26545632	rs4871	78	chr19_10226256	rs7710	200
chr6_2745352	rs6927195	79	chr19_1110829	rs2302109	201
chr6_42992825	rs3749903	80	chr19_13885484	rs10104	202
chr6_49403282	rs8589	81	chr19_17628587	rs6743	203
chr6_70407465	rs12648	82	chr19_19023853	rs3177137	204
chr6_89793894	rs1130809	83	chr19_1997363	rs1610045	205
chr6_90039670	rs10502	84	chr19_2762585	rs2302491	206
chr6_97339078	rs6684	85	chr19_39196745	rs3745859	207
chr7_116528240	rs4808	86	chr19_39322087	rs9419	208
chr7_127721507	rs322825	87	chr19_39926521	rs17626	209
chr7_128607384	rs8043	88	chr19_4362691	rs243261	210

10

20

30

40

50

【表 1 - 3】

表 1. ヒト SNP

遺伝子座	Rs 番号	配列番号	遺伝子座	Rs 番号	配列番号
chr7_140159721	rs10243155	89	chr19_4454000	rs11909	211
chr7_150916228	rs6949587	90	chr19_45490570	rs3786505	212
chr7_158536267	rs2305473	91	chr19_49513273	rs1062708	213
chr7_2577781	rs1043291	92	chr19_51301456	rs4802741	214
chr7_6063283	rs4560	93	chr19_580665	rs4682	215
chr7_6066461	rs2639	94	chr19_8468337	rs2230876	216
chr7_75959188	rs2072435	95	chr20_25260931	rs2227890	217
chr7_897492	rs10950789	96	chr20_2638579	rs6753	218
chr7_99047978	rs883403	97	chr20_2996497	rs1178016	219
chr7_99747130	rs12878	98	chr20_31427635	rs2070090	220
chr8_11710888	rs12338	99	chr20_3193978	rs8362	221
chr8_125498547	rs3812471	100	chr20_391170	rs7059	222
chr8_144662353	rs1062391	101	chr20_43530234	rs4931	223
chr8_144697041	rs1049832	102	chr20_568696	rs6053171	224
chr8_95877787	rs713113	103	chr21_38568308	rs6579	225
chr9_100823084	rs13049	104	chr21_46271452	rs235314	226
chr9_124914613	rs4679	105	chr22_32795641	rs5749426	227
chr9_127177161	rs4574	106	chr22_32887150	rs9726	228
chr9_131397479	rs4837291	107	chr22_38273749	rs9466	229
chr9_131767668	rs2287363	108	chr22_39134715	rs1062687	230
chr9_135974100	rs886017	109	chr22_42276742	rs2228314	231
chr9_139298593	rs1051957	110	chr22_42912106	rs1812240	232
chr9_26978259	rs11555693	111	chr22_42970032	rs137055	233
chr9_33026572	rs20583	112	chr22_43195147	rs1128013	234
chr9_86278817	rs7866234	113	chr22_43610207	rs138993	235
chr9_96238578	rs10821135	114	chr22_50885775	rs1053744	236
chr10_102746503	rs2863095	115	chrX_75004529	rs1343879	237
chr10_1046712	rs2306409	116			
chr10_120920588	rs10749291	117			
chr10_12209752	rs6686	118			
chr10_16796919	rs1049632	119			
chr10_27434483	rs2274634	120			
chr10_72179746	rs1043098	121			
chr10_99219885	rs2152092	122			

10

20

30

40

【0147】

【表 2 - 1】

表 2. マウス SNP

遺伝子座	遺伝子	配列番号	細胞株
chr1_91387260	Ilkap	238	4T1
chr5_136026554	Dtx2	239	4T1
chr11_100695233	Dhx58	240	4T1
chr11_69740416	Polr2a	241	4T1
chr12_110649884	Dync1hl	242	4T1
chr19_47898131	Itpr1p	243	4T1
10			
chr2_29843149	Urm1	244	A20
chr2_122052069	Eif3j1	245	A20
chr10_117355993	Cpsf6	246	A20
chr10_127064881	Cdk4	247	A20
chr11_101288969	Becn1	248	A20
chr11_69589186	Trp53	249	A20 および MC38
chr11_69589202	Trp53	250	A20 および MC38
20			
chr5_90215024	Cox18	251	B16BL6
chr13_37985317	Ssr1	252	B16BL6
chr14_117186703	Gpc6	253	B16BL6
chr14_55745832	Nop9	254	B16BL6
chr15_80929461	Tnrc6b	255	B16BL6
chr18_60692027	Ndst1	256	B16BL6
30			
chr3_100144344	Wdr3	257	B16BL6; B16F0; B16F1; B16F10
chr4_109883738	Faf1	258	B16BL6; B16F0; B16F1; B16F10
chr5_108087329	Mtf2	259	B16BL6; B16F0; B16F1; B16F10
chr15_97790681	Slc48a1	260	B16BL6; B16F0; B16F1; B16F10
chr16_57166188	Nit2	261	B16BL6; B16F0; B16F1; B16F10
chr19_32799998	Pten	262	B16BL6; B16F0; B16F1; B16F10
chr19_5716287	Ehbp111	263	B16BL6; B16F0; B16F1; B16F10
40			
chr2_10091850	Kin	264	B16F0
chr2_5053146	Optn	265	B16F0
chr3_122275655	Dnttip2	266	B16F0
chr4_129639489	Txlna	267	B16F0
chr7_131362539	2310057M21Rik	268	B16F0
chr9_109100255	Plxnb1	269	B16F0
50			
chr5_115985533	Cit	270	B16F10
chr8_119599093	Taf1c	271	B16F10
chr9_86564707	Pgm3	272	B16F10
chr13_21445463	Zscan26	273	B16F10
chr18_12197207	Npc1	274	B16F10
chrX_169313612	Hccs	275	B16F10

【表 2 - 2】

表 2. マウス SNP

遺伝子座	遺伝子	配列番号	細胞株
chr3_69030400	Smc4	276	C1498
chr4_116074925	Uqcrh	277	C1498
chr11_101411430	Aarsd1	278	C1498
chr11_69588367	Trp53	279	C1498; JC
chr11_69588367	Trp53	280	C1498; JC
chr17_84712802	Lrpprc	281	C1498
chr18_24638470	Elp2	282	C1498
chr8_83571885	Tecr	283	Colon26
chr12_84609021	Abcd4	284	Colon26
chr13_4135258	Akr1c18	285	Colon26
chr13_93880550	Arsb	286	Colon26
chr18_80197758	Rbfa	287	Colon26
chrX_41825559	Thoc2	288	Colon26
chr3_95659269	Mc11	289	CT26WT
chr6_124749315	Atn1	290	CT26WT
chr14_56693508	Mphosph8	291	CT26WT
chr17_35881376	Dhx16	292	CT26WT
chr18_60550220	Dctn4	293	CT26WT
chrX_164071728	Siah1b	294	CT26WT
chr13_74667993	Erap1	295	CT26WT Colon26
chr14_101695962	Uchl3	296	CT26WT Colon26
chr14_34343682	Glud1	297	CT26WT Colon26
chr15_12924098	Drosha	298	CT26WT Colon26
chr15_99404465	Tmbim6	299	CT26WT Colon26
chrX_157454879	Sms	300	CT26WT Colon26
chr2_105013312	Eif3m	301	EG7Ova
chr3_32728469	Mrpl47	302	EG7Ova
chr7_19514493	Trappc6a	303	EG7Ova
chr7_48803057	Zdhhc13	304	EG7Ova
chr15_103243005	Hnrnpa1	305	EG7Ova
chrX_73788891	Ssr4	306	EG7Ova
chr3_97168031	Acp6	307	EL4
chr3_14557226	Lrrcc1	308	EL4
chr4_98934494	Usp1	309	EL4
chr8_122890806	Ankrd11	310	EL4
chr14_14114149	Psmd6	311	EL4
chr17_56424099	Ptprs	312	EL4
chr5_33643749	Slbp	313	EMT6

10

20

30

40

50

【表 2 - 3】

表 2. マウス SNP

遺伝子座	遺伝子	配列番号	細胞株
chr9_3430504	Cwf19l2	314	EMT6
chr10_50849489	Ascc3	315	EMT6
chr11_70014640	Acadv1	316	EMT6
chr19_8770295	Nxfl	317	EMT6
chrX_13158898	Usp9x	318	EMT6
chr5_149623543	Hspf1	319	H22
chr6_131370348	Ybx3	320	H22
chr7_135698360	Mki67	321	H22
chr9_109842576	Nm6	322	H22
chr9_21757780	Spc24	323	H22
chr12_24711691	Rrm2	324	H22
chr3_145578132	Znhit6	325	Hepa16
chr9_72617951	Rfx7	326	Hepa16
chr13_3573709	BC016423	327	Hepa16
chr15_31594403	Cct5	328	Hepa16
chr17_23676012	Tnfrsf12a	329	Hepa16
chr19_15956304	Cep78	330	Hepa16
chr4_55378242	Rad23b	331	J558
chr7_126371216	Spns1	332	J558
chr9_111230959	Mlh1	333	J558
chr11_50210648	Sqstm1	334	J558
chr11_69588647	Trp53	335	J558 および Renca
chr11_69588703	Trp53	336	J558 および Renca
chrX_101293807	Med12	337	J558
chr5_3559131	Fam133b	338	J774A1
chr8_72255808	Ap1ml	339	J774A1
chr10_26872682	Arhgap18	340	J774A1
chr13_114826176	Mocs2	341	J774A1
chr17_35835434	Tubb5	342	J774A1
chr19_37387069	Kif11	343	J774A1
chr8_72739404	Sin3b	344	JC
chr9_119918477	Wdr48	345	JC
chr12_17277245	Pdia6	346	JC
chr18_12189845	3110002H16Rik	347	JC
chr19_8831307	Hnrnpul2	348	JC
chr2_25372768	Sapcd2	349	KLN205
chr3_145596142	Znhit6	350	KLN205
chr5_145132963	Pdap1	351	KLN205

10

20

30

40

50

【表 2 - 4】

表 2. マウス SNP

遺伝子座	遺伝子	配列番号	細胞株
chr8_122571980	Cdt1	352	KLN205
chr17_84706019	Lrpprc	353	KLN205
chrX_140472073	Prps1	354	KLN205
chr5_38234081	Lyar	355	L1210
chr5_69566389	Guf1	356	L1210
chr6_135023351	Ddx47	357	L1210
chr10_40850958	Cdc40	358	L1210
chr12_54768043	Snx6	359	L1210
chrX_94078824	Zfx	360	L1210
chr1_55080340	Hspd1	361	L5178R
chr3_96579869	Polr3g1	362	L5178R
chr9_57256682	1700017B05Rik	363	L5178R
chr10_116498369	Cnot2	364	L5178R
chr11_109436637	Amz2	365	L5178R
chrX_105874791	Atrx	366	L5178R
chr4_127047898	Zmym1	367	LLC
chr4_129008072	Ak2	368	LLC
chr4_155855159	Dvl1	369	LLC
chr5_145244535	Zfp655	370	LLC
chr6_145246772	Kras	371	LLC
chr14_86866528	Diap3	372	LLC
chr1_161074777	Cenpl	373	MBT2
chr2_112406248	Katnb1l	374	MBT2
chr2_69194469	Spc25	375	MBT2
chr7_19006050	Irf2bp1	376	MBT2
chr13_104144156	Trappc13	377	MBT2
chr16_48999045	C330027C09Rik	378	MBT2
chr7_41625342	2610021A01Rik	379	MC38
chr8_70180548	Tmem161a	380	MC38
chr9_22013055	Prkcsb	381	MC38
chr13_74646321	Erap1	382	MC38
chr15_34485603	Hrsp12	383	MC38
chr5_145144973	Bud31	384	MPC11
chr6_145232109	Kras	385	MPC11
chr13_3575438	BC016423	386	MPC11
chr15_61989534	Myc	387	MPC11
chr17_35016227	Vars	388	MPC11
chr19_46076132	Nolc1	389	MPC11

10

20

30

40

50

【表 2 - 5】

表 2. マウス SNP

遺伝子座	遺伝子	配列番号	細胞株
chr2_170515838	Pfdn4	390	Neuro2a
chr7_123428178	Lcmt1	391	Neuro2a
chr11_96911133	Cdk5rap3	392	Neuro2a
chr13_97191232	Hexb	393	Neuro2a
chr15_31598022	Cct5	394	Neuro2a
chr16_20680966	Eif4g1	395	Neuro2a
chr8_122482698	Piezol	396	P388D1
chr8_70296436	Ddx49	397	P388D1
chr9_24424805	Dpy19l1	398	P388D1
chr11_98694175	Psmd3	399	P388D1
chr12_73982520	Snapc1	400	P388D1
chr13_69811634	Med10	401	P388D1
chr1_43983175	Tpp2	402	P815
chr7_105636932	Arfip2	403	P815
chr9_36759241	Stt3a	404	P815
chr11_94634572	Lrrc59	405	P815
chr12_108812956	Yy1	406	P815
chr13_104811305	Cwc27	407	P815
chr1_63152796	Ndufs1	408	Pan02
chr7_127972166	Fus	409	Pan02
chr7_45156316	Pih1d1	410	Pan02
chr16_16983639	Mapk1	411	Pan02
chr17_75538733	Fam98a	412	Pan02
chr19_6920138	Esrra	413	Pan02
chr4_147941100	2510039O18Rik	414	Renca
chr12_69579944	Mettl21d	415	Renca
chr13_19376528	Stard3nl	416	Renca
chr17_71517626	Ndc80	417	Renca
chrX_93420657	Polal	418	Renca
chr1_24711551	Lmbrd1	419	RM1
chr4_140702160	Rcc2	420	RM1
chr11_120720063	Lrrc45	421	RM1
chr11_69589607	Trp53	422	RM1
chr15_12890119	Drosha	423	RM1
chr17_46648312	Mrp12	424	RM1
chr3_142810708	Pkn2	425	S91-P1-150414
chr3_94864330	Pogz	426	S91-P1-150414

10

20

30

40

50

【表 2 - 6】

表 2. マウス SNP

遺伝子座	遺伝子	配列番号	細胞株
chr4_131865081	Mecr	427	S91-P1-150414
chr9_15308558	Taf1d	428	S91-P1-150414
chr17_33925530	Tapbp	429	S91-P1-150414
chr17_35668634	Gtf2h4	430	S91-P1-150414
chr5_29441373	Nom1	431	WEHI164
chr13_106947227	Dimtl	432	WEHI164
chr16_56029611	Pcnp	433	WEHI164
chr17_45419631	Cdc51	434	WEHI164
chr18_35572424	Matr3	435	WEHI164
chr19_23676211	Gm6563	436	WEHI164

10

【0148】

【表 3】

表 3. ヒト Y 染色体 SNP

遺伝子座	Rs 番号
chrY_14832620	rs7067496
chrY_15467824	rs2032654
chrY_15591537	rs2032653

20

【0149】

【表 4】

表 4

染色体	位置
第 11 染色体	78371203
第 16 染色体	15839180

30

【0150】

【表 5】

表 5

ウイルス	ゲノム	配列
EBV	NC_009334	86-249
EBV	NC_009334	549-765
EBV	NC_009334	1037-1189
EBV	NC_009334	2571-2732
HBV	NC_003977	304-489
HBV	NC_003977	1393-1618
HPV16	NC_001526	7152-7271
HPV16	NC_001526	7402-7901
HPV16	NC_001526	86-406
HPV18	NC_001357	30-1774
HIV	NC_001802	20-177
HIV	NC_001802	8443-8949
マイコプラズマ	CP029295.1	505718-506180

40

50

【0 1 5 1】

【表 6 - 1】

表 6. ヒトと相同なマウスゲノム配列

遺伝子座	配列番号	遺伝子座	配列番号
chr1:105664142-105664584	437	chr11:101189377-101189820	488
chr1:131599847-131600265	438	chr11:101277344-101277788	489
chr1:133620651-133621057	439	chr11:101867409-101867715	490
chr1:38175197-38175663	440	chr11:102509968-102510377	491
chr1:42255546-42255968	441	chr11:114183205-114183653	492
chr1:43553833-43554279	442	chr11:115849863-115850303	493
chr1:55148075-55148405	443	chr12:101040300-101040687	494
chr1:55914041-55914488	444	chr12:107638060-107638501	495
chr1:55987400-55987857	445	chr12:66469994-66470265	496
chr2:114047027-114047475	446	chr13:31911885-31912294	497
chr2:114049076-114049523	447	chr13:38196485-38196783	498
chr2:114734704-114735308	448	chr13:39523526-39523952	499
chr2:114938795-114939209	449	chr13:43200327-43200776	500
chr2:116075697-116076099	450	chr13:43200696-43201123	501
chr2:119326991-119327420	451	chr13:44317005-44317340	502
chr2:119411547-119411942	452	chr13:44375738-44376192	503
chr2:140659041-140659312	453	chr14:100461399-100461810	504
chr3:144089441-144089832	454	chr14:100950247-100950697	505
chr3:34504194-34504633	455	chr14:100978309-100978722	506
chr3:36986772-36987094	456	chr14:103095102-103095512	507
chr3:37025884-37026323	457	chr14:105815614-105816038	508
chr3:6002716-6003025	458	chr14:111681168-111681625	509
chr4:100853347-100853768	459	chr14:114547663-114548108	510
chr4:102760274-102760632	460	chr14:123186854-123187271	511
chr4:41519808-41520239	461	chr14:52463385-52463656	512
chr4:43443358-43443775	462	chr15:102430819-102431236	513
chr4:43445287-43445715	463	chr15:102811273-102811660	514
chr4:76039612-76039883	464	chr15:102966510-102966781	515
chr5:106322349-106322777	465	chr15:103298047-103298653	516
chr5:122861456-122861856	466	chr15:103524801-103525249	517
chr5:122988599-122989028	467	chr15:34141703-34142120	518
chr6:108664368-108664777	468	chr16:29666774-29667232	519
chr7:102097926-102098361	469	chr16:29875939-29876394	520
chr7:102428566-102429007	470	chr16:6057738-6058045	521
chr7:102698517-102698966	471	chr16:6776352-6776786	522
chr7:105384072-105384436	472	chr16:76321297-76321735	523
chr7:105635729-105636145	473	chr16:78941222-78941652	524
chr7:105740673-105740944	474	chr16:80265994-80266442	525
chr7:107665615-107666028	475	chr16:80434685-80435068	526
chr7:108754667-108755104	476	chr16:87128020-87128361	527
chr8:103447835-103448254	477	chr16:87319732-87320189	528
chr8:115428668-115429087	478	chr16:91114059-91114450	529

10

20

30

40

50

【表 6 - 2】

表 6. ヒトと相同なマウスゲノム配列

遺伝子座	配列番号	遺伝子座	配列番号
chr8:123892100-123892548	479	chr17:15370365-15370824	530
chr9:119495272-119495684	480	chr17:26742420-26742783	531
chr9:120929616-120930041	481	chr17:26935067-26935511	532
chr9:124124306-124124697	482	chr17:27876446-27876890	533
chr9:24974244-24974660	483	chr17:30292965-30293355	534
chr9:82866266-82866706	484	chr18:19963182-19963628	535
chr9:84973186-84973624	485	chr18:25632724-25633184	536
chr10:29698549-29699004	486	chr18:34606137-34606581	537
chr10:75061569-75061989	487	chr18:34641045-34641374	538
		chr18:34759289-34759725	539
		chr18:34863942-34864380	540
		chr19:41962522-41962964	541
		chr19:46061929-46062390	542
		chr19:46251931-46252398	543
		chr19:46306521-46306978	544

10

20

【0 1 5 2】

30

40

50

【表 7 - 1】

表 7. 連続希釈を用いた細胞株に対する認証および夾雑物検出\*

細胞株混合物	主要構成成分	微量構成成分 (夾雑物)	微量構成率 (ペーセンテージ)*	イソフオーマティブな SNP の数	不均一性比率(ペーセンテージ)**	推論された主要構成成分***	推論された微量構成成分*** (夾雑物)***	夾雑比率(ペーセンテージ)***	P 値****
PANC1:RT4	PANC1	-	-	118	1.4 (99.46%)	PANC1	-	-	-
	RT4	PANC1	5	86	3.3 RT4 (97.78%)	PANC1	-	-	-
	RT4	PANC1	2.5	80	1.85 RT4 (97.93%)	PANC1	-	-	-
	RT4	PANC1	1.25	81	1.1 RT4 (97.76%)	PANC1	-	-	-
	RT4	PANC1	0.625	80	1.06 RT4 (97.76%)	PANC1	-	-	-
	PANC1	RT4	5	131	8.09 (99.50%)	PANC1	-	-	-
	PANC1	RT4	2.5	128	4.06 (99.35%)	PANC1	-	-	-
	PANC1	RT4	1.25	132	2.75 (99.49%)	PANC1	-	-	-
	PANC1	RT4	0.625	139	2.62 (99.47%)	PANC1	-	-	-
	LNCAPCLONE FGC	-	-	97	1.005 LNCAPCLONE FGC (99.03%)	LNCAPCLONE FGC (99.03%)	-	-	-
LNCAPCLONE FGC: MV411	MV411	-	-	93	0.965 (99.03%)	MV411	-	-	-
	LNCAPCLONE FGC	5	99	9 (99.45%)	MV411	LNCAPCLONE FGC (96.83%)	9.08	5.01E-17	
	MV411	LNCAPCLONE FGC	2.5	111	4.51 (99.50%)	LNCAPCLONE FGC (98.34%)	4.14	5.01E-17	
	MV411	LNCAPCLONE FGC	1.25	117	2.18 (99.18%)	LNCAPCLONE FGC (90.32%)	2.01	1.67E-09	
	MV411	LNCAPCLONE FGC	0.625	112	1.58 (99.00%)	LNCAPCLONE FGC (89.44%)	0.0148	1.67E-09	

【表 7 - 2】

細胞株混合物	主要構成成分	微量構成成分(夾雜物)	微量構成成分比率(ペーセンテージ)*	イソフオーマティープなSNPの数	不均一性比率(ペーセンテージ)**	推論された主要構成成分***	微量構成成分(夾雜物)****	推論された微量構成成分(夾雜物)****	夾雜比率(ペーセンテージ)****	P値
LNCAPCLONE FGC	MV411	5	102	2.35	LNCAPCLONE FGC (98.99%)	LNCAPCLONE FGC (94.57%)	MV411	MV411	2.14	5.01E-17
LNCAPCLONE FGC	MV411	2.5	101	1.67	LNCAPCLONE FGC (99.04%)	LNCAPCLONE FGC (91.58%)	MV411	MV411	0.88	2.37E-11
LNCAPCLONE FGC	MV411	1.25	98	1.49	LNCAPCLONE FGC (99.03%)	LNCAPCLONE FGC (87.77%)	MV411	MV411	0.71	2.76E-11
LNCAPCLONE FGC	MV411	0.625	105	1.36	LNCAPCLONE FGC (99.3%)	LNCAPCLONE FGC (99.3%)	-	-	-	-
CAL27 RAJI	-	-	39	1.39	CAL27 RAJI (97.39%)	CAL27 RAJI (98.56%)	-	-	-	-
RAJI	CAL27	5	116	5.36	RAJI (98.66%)	RAJI (99.12%)	CAL27	CAL27	5.54	5.01E-17
RAJI	CAL27	2.5	127	4.17	RAJI (98.32%)	RAJI (94.70%)	CAL27	CAL27	4.01	8.37E-11
RAJI	CAL27	1.25	121	1.84	RAJI (98.51%)	RAJI (90.83%)	CAL27	CAL27	2.21	8.37E-06
RAJI	CAL27	0.625	116	2.49	RAJI (98.50%)	RAJI (90.43%)	CAL27	CAL27	1.67E-07	
CAL28:RAJI	RAJI	5	121	7.11	CAL27 RAJI (99.17%)	CAL27 RAJI (99.51%)	RAJI (99.17%)	RAJI (99.51%)	5.41	5.01E-17
CAL27	RAJI	2.5	113	3.79	CAL27 RAJI (98.94%)	CAL27 RAJI (92.06%)	RAJI (98.94%)	RAJI (92.06%)	2.41	4.30E-13
CAL27	RAJI	1.25	112	2.14	CAL27 RAJI (98.61%)	CAL27 RAJI (83.42%)	RAJI (98.61%)	RAJI (83.42%)	1.21	4.20E-07
CAL27	RAJI	0.625	112	1.4	CAL27 RAJI (98.49%)	CAL27 RAJI (83.75%)	RAJI (98.49%)	RAJI (83.75%)	0.61	1.59E-07

\*: 未混合のものは、3つの技術的反復を用いた各細胞株混合物についての平均値

\*\*: 細胞カウントに基づく微量細胞株のペーセンテージ

\*\*\*: 遺伝子型類似性が括弧内に示される

\*\*\*\*: キメラ遺伝子型類似性が括弧内に示される

\*\*\*\*\*: 推論された微量構成成分が不正確である確率

【0 1 5 3】

【表8-1】

表8.細胞株混合物の認証および夾雜物検出

細胞株混合物*	インフオーマティブなSNPの数	不均一性比率(ペーセンテージ)	推論された主要構成成分** (夾雜物)***	推論された微量構成成分 (夾雜物)***	夾雜比率 (ペーセンテージ)	P値****
ME180:143B	119	6.54	ME180 (98.06%)	143B (97.09%)	7.01	5.01E-17
143B:ME180	135	3.24	143B (98.55%)	ME180 (94.17%)	3.21	5.01E-17
JEG3:A875	104	1.63	JEG3 (98.49%)	A875 (87.94%)	1.21	5.01E-13
A875:JEG3	93	15.50	A875 (91.71%)	JEG3 (99.00%)	16.71	5.01E-17
HT3:C33A	115	3.54	HT3 (100%)	C33A (97.06%)	3.41	5.01E-17
C33A:HT3	90	4.34	C33A (99.01%)	HT3 (100%)	4.61	0
DOTC24510:CASKI	136	5.47	DOTC24510 (98.99%)	CASKI (93.97%)	5.21	5.01E-17
CASKI:DOTC24510	129	4.26	CASKI (98.98%)	DOTC24510 (91.84%)	4.11	5.01E-17
HLE:HCC94	163	2.62	HLE (99.0%), HLF (96.08%)	HCC94 (91.46%)	2.91	5.01E-17
HCC94:HLE	133	10.65	HCC94 (97.6%)	HLE (96.63%), HLF (96.63%)	10.11	5.01E-17
NCIH1993:LS174T	141	3.97	NCIH1993 (98.05%)	LS174T (95.12%), LS180 (95.12%), HM7 (94.63%)	4.21	5.01E-17
LS174T:NCIH1993	114	4.88	LS174T (99.02%), LS180 (99.03%), HM7 (98.54%)	NCIH1993 (97.06%)	4.71	5.01E-17
OSC19:SF763	152	7.03	OSC19 (98.08%)	SF763 (96.15%)	5.71	5.01E-17
SF763:OSC19	133	3.35	SF763 (99.02%)	OSC19 (90.15%)	2.91	2.51E-16
SW626: SJCRH30	155	11.67	SW626 (95.63%)	SJCRH30 (98.54%)	13.21	5.01E-17
SJCRH30:SW626	88	1.79	SJCRH30 (98.55%)	SW626 (94.2%)	2.01	1.58E-16
A875:ME180	115	2.68	A875 (98.56%)	ME180 (95.67%)	2.31	5.01E-17
DOTC24510:CASKI	144	1.75	DOTC24510 (98.5%)	CASKI (86%)	1.71	1.00E-15
OSC19:SF763	130	2.68	OSC19 (98.56%)	SF763 (93.27%)	2.11	5.01E-17

【表 8 - 2】

NOZ:SW626	127	0.82	NOZ (97.56%)	SW626 (82.93%)	0.71	3.98E-11
SNU739:MM1R	121	2.29	SNU739 (99.01%)	MM1R (94.03%), MM1S (94.03%)	1.71	5.01E-17
U251:SR	127	1.09	U251 (98.54%)	SR (89.76%)	1.11	5.01E-17

\*: 主要細胞株:微量/夾雜している細胞株の形式

\*\*: 遺伝子型類似性が括弧内に示される

\*\*\*: キメラ遺伝子型類似性が括弧内に示される

\*\*\*\*: 推論された微量構成成分が誤りである確率

【 0 1 5 4 】

【表 9 - 1】

表 9.細胞株の認証

細胞株	インフォーマティブなSNPの数	不均一性比率(パーセンテージ)	推論された細胞株*
G292CloneA14B1	176	7.62	G292CLONEA141B1(98.53%)
HPAF-II	153	6.63	HPAFII(99.01%)
PL45	135	4.49	PL45(98.42%), PANC1005(98.43%)
HCC827	193	3.46	HCC827(99.51%)
HeLa	138	2.99	HELA(99.52%), HEA229(99.06%)
OCI-AML-2	143	2.77	OCIAML2(98.1%)
K-562	158	2.56	K562(98.52%)
OVCAR-5	154	2.14	OVCAR5(97.52%)
A-427	115	2.05	A427(97.45%)
8505C	158	1.96	8505C(98.05%)
NOZ	143	1.87	NOZ(97.94%)
Hep3B	106	1.77	HEP3B217(95.73%)
SF268	113	1.63	SF268(93.85%)
NCI-H1993	101	1.57	NCI-H1993(96.25%)
NCI-H1793	139	1.56	NCI-H1793(98.55%)
NCI-H1688	120	1.55	NCI-H1688(99.48%)
MX-1	141	1.51	MX1(100%)
NCI-N87	120	1.45	NCI-N87(98.03%)
RBE	110	1.43	RBE(98.94%)
SiHa-579	121	1.43	SIHA(98.84%)
KPL-4	118	1.42	KPL4(96.43%)
EVSA-T	132	1.41	EVSA-T(97.86%)
Ishikawa	90	1.29	ISHIKAWA(100%), ISHIKAWAHERAKLIO02ER(100%)
IM95m	100	1.28	IM95M(98.35%), IM95(98.34%)
MES-SA	99	1.26	MESSA(99.5%)
MHH-CALL-2	189	1.26	MHH-CALL-2(100%)
OZ	107	1.26	OZ(98.95%)
SH-SY5Y	86	1.25	SH-SY5Y(99.47%), SKNSH(99.47%)
PC-9	116	1.24	PC-9(98.48%), PC14(97.99%)
Calu-3	117	1.23	Calu-3(98.02%)
ME-180	89	1.21	ME-180(98.89%)
SUM159PT	89	1.20	SUM159PT(98.94%)
NCI-H322	107	1.18	NCI-H322(99.46%)
LS174T	85	1.18	LS174T(99.44%), HM7(98.89%), LS180(99.45%)
NCI-H292	95	1.18	NCI-H292(100%)
NCI-H1568	153	1.16	NCI-H1568(97.57%)
GTL-16	113	1.15	GTL-16(98.46%), MKN45(97.97%)
KG-1a	104	1.15	KG-1a(97.34%), KG1(97.93%)
HM-7	104	1.15	HM-7(98.38%), LS180(98.92%), LS174T(98.92%)
HCCLM3	119	1.14	HCCLM3(98.92%)
MHCC97-H	75	1.08	MHCC97-H(99.34%)

10

20

30

40

50

【表 9 - 2】

NCI-H1373	137	1.08	NCIH1373(98.16%)
Capan-2	139	1.06	CAPAN2(99.52%)
ML-2	111	1.06	ML2(99.49%)
SW1463	144	1.06	SW1463(93.06%)
NCI-H1395	110	1.06	NCIH1395(98.93%)
A-673	112	1.05	A673(98.54%)
DU-145	149	1.04	DU145(99.52%)
NCI-H1781	154	1.04	NCIH1781(99.51%)
SK-NEP-1	79	1.04	SKNEP1(99.49%)
A-431	75	1.04	A431(98.5%)
HepG2C3A	84	1.02	HEPG2C3A(99.32%)
U251	102	1.01	U251MG(99.45%)
SNU-354	88	1.01	SNU354(94.42%)
Hs445	93	1.00	HS445(99.49%)
JEG-3	94	1.00	JEG3(99.39%)
Z-138	94	1.00	Z138(98.54%)
IHH-4	87	1.00	IHH4(100%)
SF763	84	1.00	SF763(98.44%)
SNU-739-P1	95	1.00	SNU739(98.95%)
HeLa299	78	0.99	HELA(99.43%), HELA229(99.43%)
L-82	106	0.99	L82(97.96%)
MM1R	105	0.98	MM1R(100%), MM1S(99.48%)
HT-3	118	0.97	HT3(99.48%)
SCH	92	0.97	SCH(99.41%)
U118MG	128	0.97	U118MG(96.17%)
OSC-19	118	0.96	OSC19(98.41%)
JVM-13	89	0.94	JVM13(98.97%)
WiDr	113	0.94	HT29(97.14%)
SK-N-SH	78	0.93	SKNSH(99.5%), SHSY5Y(99.50%)
KYSE-410	147	0.92	KYSE410(97.55%)
HBL-1	100	0.92	HBL1(100%)
NAMALWACSN	104	0.91	NAMALWACSN(98.98%), NAMALWA(98.46%)
SNU-368	23	0.91	SNU368(95.89%)
Jurkat	101	0.90	JURKAT(98.34%), JURKATCLONEE61(98.37%)
SK-N-AS	124	0.89	SKNAS(99.04%)
SNU-2535	85	0.88	SNU2535(99.47%)
HCCC-9810	94	0.86	HCCC9810(98.31%)
COLO320DM	96	0.85	COLO320DM(98.47%)
MS751	103	0.85	MS751(97.88%)
SW48	34	0.85	SW48(98.2%)
CCRF-CEM	72	0.84	COClDDP(99.02%), CCRFCEM(99.02%), COC1(99.02%)
YCC-2	121	0.84	YCC2(97.98%)
TJ905	100	0.84	TJ905(97.53%)
CoC1DDP	92	0.83	COClDDP(99.46%), CCRFCEM(99.46%), COC1(99.46%)

10

20

30

40

50

【表9-3】

SK-UT-1	94	0.83	SKUT1(99.03%)
D283	77	0.81	D283MED(98.98%)
DoTc24510	105	0.80	DOTC24510(98.97%)
MCCAR	84	0.80	MCCAR(98.52%)
OCI-LY7	105	0.80	OCILY7(98.04%)
A253	141	0.79	A253(98.06%)
JAR	79	0.79	JAR(98.33%)
SU-DHL-6	134	0.79	SUDHL6(99%)
SR	66	0.79	SR(99.46%)
AN3CA	107	0.78	AN3CA(96.65%)
Hutu80	80	0.78	HUTU80(100%), AZ521(100%)
OCUM-2M	106	0.78	OCUM2M(98.47%)
SW684	74	0.77	SW684(98.98%)
SJCRH30	71	0.76	SJCRH30(98.97%)
SW982	71	0.75	SW982(99.48%)
Y-79	84	0.75	Y79(97.38%)
MKN45	103	0.73	MKN45(98.99%), GTL16(98.47%)
143B	91	0.72	143B(98.96%)
Molt-4	87	0.72	MOLT4(98.98%), MOLT3(99.03%)
RT4	105	0.72	RT4(99.11%)
HCT-8	93	0.71	HCT8(98.5%)
NCI-H460	127	0.71	NCIH460(99.53%)
SW480	141	0.71	SW480(100%)
T.Tn	129	0.71	T.T(95.22%)
YCC-10	99	0.71	YCC10(98.97%)
C-33A	80	0.71	C33A(98.51%)
ASH-3	99	0.67	ASH3(97.06%)
SW626	112	0.66	SW626(97.98%)
SW756	96	0.66	SW756(99.02%)
A-875	80	0.63	A875(98.06%)
CoC1	72	0.61	COCIDDP(98.98%), CCRFCEM(98.98%), COC1(98.98%)
JurkatcloneE6-1	92	0.57	JURKATCLONEE61(98.12%), JURKAT(98.08%)

\*遺伝子型類似性が括弧内に示される

【0155】

10

20

30

40

50

【表 10】

表 10. 関係する細胞株の遺伝子型類似性

試験した細胞株	関係する細胞株	遺伝子型類似性
HeLa	SMMC7721	0.8431
HeLa	BEL7402	0.8529
HeLa	QGY7701	0.8657
HeLa	QGY7703	0.8683
SW480	SW620	0.8852
NCIH1993	NCIH2073	0.8932
OCUM-2M	OCUM2D	0.9
143B	HOS	0.9179
SR	SR786	0.9235
SJCRH30	RH30	0.932
143B	KHOSNP	0.9563
Hep3B	HEP3B217	0.9573
HCT-8	HCT15	0.9608
HepG2C3A	HEPG2	0.9622
HCCLM3	MHCC97H	0.9626

10

20

【0156】

30

40

50

【表 1-1】

組合せ	1	2	3	4	5	6
S1 遺伝子型	AA	AA	AA	AT	AT	AT
S2 遺伝子型	TT	AT	TG	GG	AG	GC
S2 比率 (SNP 不均一性比率)	T/(A+T)**	2T/(A+T)**	(T+G)/(A+T+G)	G/(A+T+G)***	G/(T+G)***	(G+C)/(A+T+G+C)
ヌクレオチド頻度 パターン	大きい A 小さい T	大きい A 小さい T	大きい A 小さい G	大きい A 小さい G	大きい A 小さい G	大きい A 小さい G

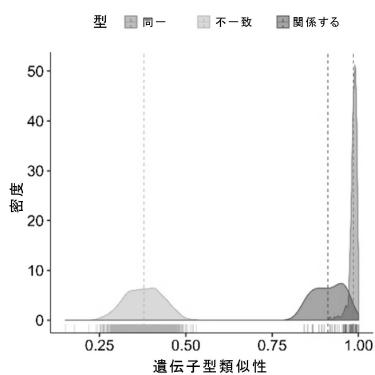
\*: S1 は、混合試料中の主要構成成分であり、S2 は、微量/夾雜している構成成分である。各組合せは、組合せのクラスを示すために特定のヌクレオチドを使用し、例えば、第1 の組合せは、両方がホモ接合性遺伝子型であることを、異なるヌクレオチドを用いて示している。S2 比率を計算するための式では、ヌクレオチドは、NGS シークエンシングデータ中のそのカウント(読み取りの総数)を示している。

\*\*: 組合せ 1 および 2 は、観測された NGS データから区別できないので、1.5T/(A+T) が両方に使用される。

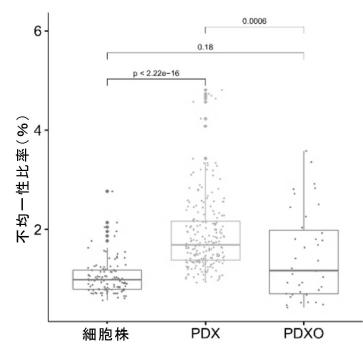
\*\*\*: 組合せ 4 および 5 は、観測された NGS データから区別できないので、1.5G/(A+T+G) が両方に使用される。

## 【図面】

## 【図 1 A】

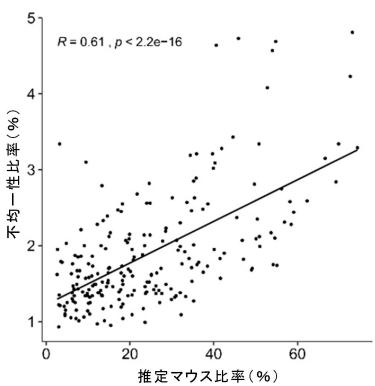


## 【図 1 B】

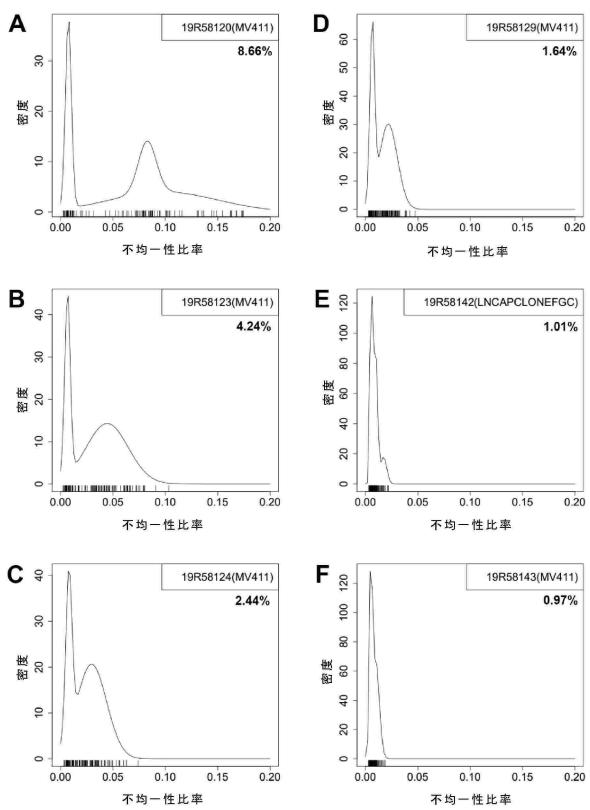


10

## 【図 1 C】



## 【図 2】



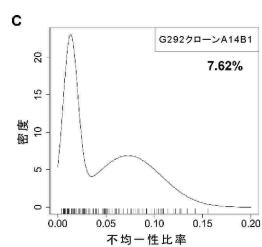
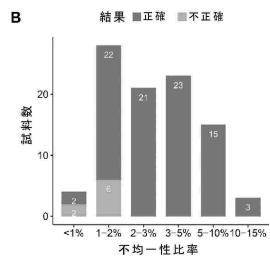
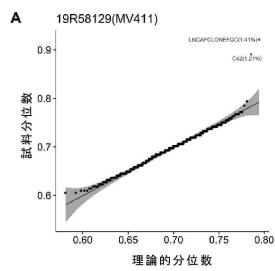
20

30

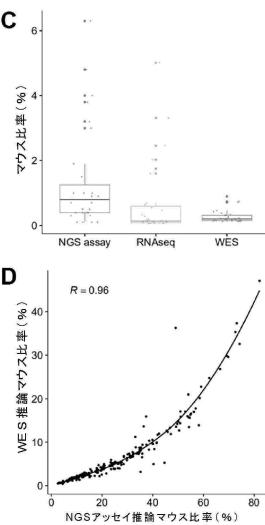
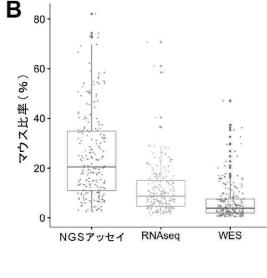
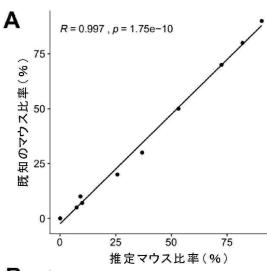
40

50

【図3】

D G292クローニングA14B1  
ODAM128.2%

【図4】



【配列表】

0007679306000001.app

10

20

30

40

50

## フロントページの続き

## (51)国際特許分類

		F I		
C 1 2 Q	1/689(2018.01)	C 1 2 Q	1/689	Z
C 1 2 Q	1/6879(2018.01)	C 1 2 Q	1/6879	Z
C 1 2 N	15/09 (2006.01)	C 1 2 N	15/09	2 0 0

弁理士 今野 智介

(74)代理人 100144794

弁理士 大木 信人

(74)代理人 100204582

弁理士 大栗 由美

(72)発明者 ゲオ, シエン

中華人民共和国 215000 ジアンスー, スーチョウ, インダストリアル パーク, シンフーストリート ナンバー218, ビルディング エ-6, ルーム 303

(72)発明者 チエン, シャオボ

中華人民共和国 215000 ジアンスー, スーチョウ, インダストリアル パーク, シンフーストリート ナンバー218, ビルディング エ-6, ルーム 303

(72)発明者 キアン, ウービン

中華人民共和国 215000 ジアンスー, スーチョウ, インダストリアル パーク, シンフーストリート ナンバー218, ビルディング エ-6, ルーム 303

(72)発明者 リ, ヘンリー

中華人民共和国 215000 ジアンスー, スーチョウ, インダストリアル パーク, シンフーストリート ナンバー218, ビルディング エ-6, ルーム 303

審査官 斎藤 貴子

## (56)参考文献

GUO, F. et al. , Next generation sequencing of SNPs using the HID-Ion AmpliSeq(TM) Identity Panel on the Ion Torrent PGM(TM) platform , Forensic Science International: Genetics , 2016年 , Vol. 25 , P. 73-84 , <https://doi.org/10.1016/j.fsigen.2016.07.021>

DIDION, J. P. et al. , SNP array profiling of mouse cell lines identifies their strains of origin and reveals cross-contamination and widespread aneuploidy , BMC Genomics , 2014年 , Volume 15, article number 847 , P. 1-11

LIANG-CHU, M. M. Y. et al. , Human Biosample Authentication Using the High-Throughput, Cost-Effective SNPtrace(TM) System , PLOS ONE , 2015年 , Vol. 10, No. 2: e0116218 , P. 1-14 , <https://doi.org/10.1371/journal.pone.0116218>

(58)調査した分野 (Int.Cl. , DB名)

C 1 2 Q

C 1 2 N

C A p l u s / M E D L I N E / E M B A S E / B I O S I S ( S T N )