



(12) 发明专利申请

(10) 申请公布号 CN 118932031 A

(43) 申请公布日 2024. 11. 12

(21) 申请号 202410868756.1

(22) 申请日 2019.11.12

(30) 优先权数据

62/760,743 2018.11.13 US

62/929,554 2019.11.01 US

(62) 分案原申请数据

201980079987.1 2019.11.12

(71) 申请人 麦利亚德基因公司

地址 美国犹他州

(72) 发明人 A·扎尔基赫 K·蒂姆斯

M·佩里 A·古丁

(74) 专利代理机构 北京纪凯知识产权代理有限公司

11245

专利代理师 丁秀云

(51) Int. Cl.

C12Q 1/6869 (2018.01)

C12Q 1/6827 (2018.01)

C12Q 1/6886 (2018.01)

权利要求书1页 说明书30页 附图7页

(54) 发明名称

用于体细胞突变的方法和系统及其用途

(57) 摘要

本发明的发明名称为用于体细胞突变的方法和系统及其用途。本发明提供了用于检测癌细胞中的体细胞突变的方法和组合物。所述方法可以用于测量肿瘤突变负荷。提供了用于鉴定和治疗受益于用如免疫检查点抑制剂等抗癌剂进行的治疗的受试者的方法、用于治疗受试者的癌症的方法以及用于对患有癌症的受试者进行监测和预测的方法。

序列读段映射	S101
序列读段过滤	S103
碱基调用计数	S105
位置过滤	S107
生成变体/SNP碱基调用矩阵	S109
体细胞得分计算	S111
TMB计算	S113

1. 一种用于检测体细胞变体的方法,所述方法包括:
  - (a) 对样品的细胞进行测序;
  - (b) 鉴定一组杂合SNP位置,其中每个SNP具有等位基因B和A;
  - (c) 检测两个种系等位基因配对的SNP位置和在所述SNP位置附近的位置的变体,其中所述两个种系等位基因配对是(i)等位基因B和第一变体等位基因以及(ii)等位基因A和第二变体等位基因,所述第二变体等位基因可以与所述第一变体等位基因相同或不同;以及
  - (d) 检测第三等位基因配对,所述第三等位基因配对是(iii)等位基因B和第三变体等位基因,所述第三变体等位基因不同于所述第一变体等位基因。
2. 根据权利要求1所述的方法,其中所述等位基因配对各自在含有所述SNP位置之一的连续核酸序列中检测到,使得变体位置位于所述SNP位置的一个检测长度内。
3. 根据权利要求2所述的方法,其中所述连续核酸序列的读段长度为约100到5000个碱基。
4. 根据权利要求2所述的方法,其中所述检测长度为所述SNP位置的每个侧翼上的200到1000个连续碱基位置。
5. 根据权利要求1所述的方法,其中所述方法不利用单独的种系比较物样品。
6. 根据权利要求1所述的方法,其中所述样品是癌组织样品、肿瘤细胞样品或肿瘤样品。
7. 根据权利要求1所述的方法,其中使所述样品中的非肿瘤细胞的量最小化。
8. 根据权利要求1所述的方法,其中所述肿瘤样品含有非肿瘤细胞。
9. 根据权利要求1所述的方法,其中通过大规模平行测序、通过杂交或用扩增来检测所述等位基因配对。
10. 根据权利要求1所述的方法,其中所述一组杂合SNP位置为至少5000个SNP位置或至少100,000个SNP位置或至少500,000个SNP位置或至少1,000,000个SNP位置或至少2,000,000个SNP位置。

## 用于体细胞突变的方法和系统及其用途

[0001] 本申请是分案申请,原申请的申请日是2019年11月12日,申请号为2019800799871 (PCT/US2019/061036),发明名称为“用于体细胞突变的方法和系统及其用途”。

### 技术领域

[0002] 本发明涉及用于通过核酸测序检测癌细胞中的体细胞突变的方法、组合物、试剂盒和系统。更具体地,本公开提供了用于测量肿瘤突变负荷、用于鉴定和治疗受益于用如免疫检查点抑制剂等抗癌剂进行治疗的受试者以及用于治疗受试者的癌症和用于对患有癌症的受试者进行监测和预测的方法。

### 背景技术

[0003] 细胞中癌症的标志之一是基因组中体细胞变体的存在。参见例如,Theodor Boveri,《细胞科学杂志(J.Cell Sci.)》(2008) 121:1-84。体细胞变体可以用作癌症的生物标志物,特别是当可以准确地检测和记录变体的频率时。然而,难以定量地检测体细胞变体。

[0004] 癌细胞中的体细胞变体的频率范围可以为低于0.1直至数百每Mb。用于检测体细胞变体的方法的缺点包含由于变体的低频率出现而引起的低灵敏度。尝试在低频率下鉴定和计数体细胞变体可能无法克服高通量核酸测序方法中的噪声水平。

[0005] 进一步地,在需要参考基因组的核酸测序方法中,参考基因组中各种等位基因的不充分表示可能由于群体或种族偏差而导致不准确。

[0006] 一些常规测序方法的显著缺点是需要非癌症种系比较物样品用于将种系变体与在癌症样品中检测到的变体区分开。非癌症种系比较物样品可以提供待从在癌细胞中检测到的体细胞变体中减去的基线。事实上,在许多情况下,这种比较物样品甚至可能不可用。

[0007] 所需要的是用于以高灵敏度检测体细胞变体的方法、组合物和系统。还期望改进测序方法以准确地检测和计数体细胞变体。

[0008] 迫切需要用于治疗癌症和鉴定受益于治疗的受试者的方法。所需要的是不需要非癌症比较物样品以及来自患有癌症的受试者的肿瘤或组织样品的方法和系统。

[0009] 长期以来,需要通过涉及直接检测变体以减少误差的方法来实现这些目标。

### 发明内容

[0010] 本发明提供了用于检测癌细胞中的体细胞突变、用于鉴定和治疗受益于用如免疫检查点抑制剂等抗癌剂进行治疗的受试者、用于测量肿瘤突变负荷、用于治疗受试者的癌症以及用于对患有癌症的受试者进行监测和预测的方法、组合物、试剂盒和系统。

[0011] 体细胞突变的测量结果可以提供癌症的治疗方法、诊断方法和预后方法。

[0012] 在一些方面,本发明提供了用于选择和鉴定受益于治疗(如使用抗癌剂治疗癌症)的受试者的方法。对于这些受试者,可以选择治疗方式来治疗癌症。

[0013] 在另外的方面,本发明提供了用于对癌细胞中的肿瘤突变频率进行测量和评分的

方法。得分可以用于计算来自受试者的样品的肿瘤突变负荷。肿瘤突变负荷可以充当疾病，如癌症的生物标志物。

[0014] 体细胞变体可以与受试者对使用某些药物的治疗的应答相关。例如，高肿瘤突变负荷值可以与患有癌症的受试者对施用免疫检查点抑制剂药物的有利应答相关。

[0015] 本发明的实施例包含：

[0016] 一种用于检测体细胞变体的方法，所述方法包括：

[0017] (a) 对样品的细胞进行测序；

[0018] (b) 鉴定一组杂合SNP位置，其中每个SNP具有等位基因B和A；

[0019] (c) 检测两个种系等位基因配对的SNP位置和在所述SNP位置附近的位置的变体，其中所述两个种系等位基因配对是 (i) 等位基因B和第一变体等位基因以及 (ii) 等位基因A和第二变体等位基因，所述第二变体等位基因可以与所述第一变体等位基因相同或不同；以及

[0020] (d) 检测第三等位基因配对，所述第三等位基因配对是 (iii) 等位基因B和第三变体等位基因，所述第三变体等位基因不同于所述第一变体等位基因。所述等位基因配对各自可以在含有所述SNP位置之一的连续核酸序列中检测到，使得变体位置位于所述SNP位置的一个检测长度内。所述连续核酸序列的读段长度可以为约100到5000个碱基。所述检测长度可以为所述SNP位置的每个侧翼上的200到1000个连续碱基位置。所述方法不利用单独的种系比较物样品。所述样品可以是癌组织样品、肿瘤细胞样品或肿瘤样品。可以使样品中的非肿瘤细胞的量最小化。样品可以含有非肿瘤细胞。可以通过大规模平行测序、通过杂交或用扩增来检测等位基因配对。所述一组杂合SNP位置可以为至少500个SNP位置或至少1000个SNP位置或至少5000个SNP位置。所述方法可以以0.1每Mb或0.3每Mb或0.7每Mb的最低水平检测到体细胞变体。检测可以用靶向的SNP小组获得。所述检测可以是通过使用人类参考基因组的片段化测序获得的。

[0021] 一种用于检测体细胞变体的方法，所述方法包括：

[0022] (a) 对肿瘤样品的细胞进行测序；

[0023] (b) 使用大规模平行核酸测序方法从所述样品获得序列读段，其中所述序列读段具有读段长度；

[0024] (c) 将所述序列读段映射到参考基因组；

[0025] (d) 组装映射到所述参考基因组的杂合SNP位置的序列读段的体细胞变体计数矩阵，其中所述计数矩阵具有第一元件和第二元件，所述第一元件和所述第二元件分别对SNP等位基因B和A与变体等位基因的等位基因配对进行计数，并且其中所述计数矩阵具有第三元件，所述第三元件对来自与所述第一元件中不同的变体等位基因配对的SNP等位基因B的读段序列进行计数；以及

[0026] (e) 计算所述第三元件的体细胞突变显著性得分(S)。所述方法不利用单独的种系比较物样品。所述样品可以是癌组织样品、肿瘤细胞样品或肿瘤样品。所述方法可以以0.1每Mb或0.3每Mb或0.7每Mb的最低水平检测到体细胞变体。序列读段可以用靶向的SNP小组获得。所述读段长度可以为100到5000个或200到1000个连续碱基位置。对于参考基因组的被覆盖的部分，平均读段深度可以为至少50x或100x。所述参考基因组可以是人类基因组。可以对序列读段进行误差过滤和位置过滤。

[0027] 所述体细胞突变显著性得分(S)由公式I给出:

[0028]  $S = (C(Z,P)^2 / (C(Z,P) + C(X,P)) + (C(Z,P) - E)^2 / E) / 2 * 10$  公式I

[0029] 其中C(Z,P)是第三元件计数,C(X,P)是第一元件计数,并且E是针对所有SNP区域从所述矩阵中的除最高的三个计数之外的所有其它计数的平均值计算的误差率。

[0030] 一种用于鉴定患有癌症的受试者受益于治疗的方法,所述方法包括:

[0031] (a) 对来自所述受试者的肿瘤样品的细胞进行测序;

[0032] (b) 鉴定一组杂合SNP位置,其中每个SNP具有等位基因B和A;

[0033] (c) 检测两个种系等位基因配对的SNP位置和在所述SNP位置附近的位置的变体,其中所述两个种系等位基因配对是(i)等位基因B和第一变体等位基因以及(ii)等位基因A和第二变体等位基因,所述第二变体等位基因可以与所述第一变体等位基因相同或不同;以及

[0034] (d) 检测第三等位基因配对,所述第三等位基因配对是(iii)等位基因B和第三变体等位基因,所述第三变体等位基因不同于所述第一变体等位基因,其中所述第三等位基因配对源于体细胞变体;

[0035] (f) 计算来自从所述等位基因配对检测到的所述体细胞变体的肿瘤突变负荷的值;以及

[0036] (g) 如果所述肿瘤突变负荷大于参考水平,则鉴定所述患有癌症的受试者受益于治疗。

[0037] 一种用于鉴定患有癌症的受试者受益于治疗的方法,所述方法包括:

[0038] (a) 对来自所述受试者的肿瘤样品的细胞进行测序;

[0039] (b) 使用大规模平行核酸测序方法从所述样品获得序列读段,其中所述序列读段具有读段长度;

[0040] (c) 将所述序列读段映射到参考基因组;

[0041] (d) 组装映射到所述参考基因组的杂合SNP位置的序列读段的体细胞变体计数矩阵,其中所述计数矩阵具有第一元件和第二元件,所述第一元件和所述第二元件分别对SNP等位基因B和A与变体等位基因的等位基因配对进行计数,并且其中所述计数矩阵具有第三元件,所述第三元件对来自与所述第一元件中不同的变体等位基因配对的SNP等位基因B的读段序列进行计数;

[0042] (e) 通过以下步骤计算所述样品的肿瘤突变负荷的值:

[0043] (i) 计算所述第三元件的体细胞突变显著性得分(S);以及

[0044] (ii) 从体细胞突变显著性得分高于阈值的体细胞变体的数量计算所述肿瘤突变负荷的所述值,所述数量通过杂合SNP区域中的位置的总数进行归一化;以及

[0045] (f) 如果所述肿瘤突变负荷大于体细胞突变参考水平,则鉴定所述患有癌症的受试者受益于治疗。所述参考基因组中的杂合SNP的数量可以为约100直至所述参考基因组中的杂合SNP的总数。所述体细胞突变参考水平可以是所述受试者将受益于所述治疗的水平。所述体细胞突变参考水平可以是所述参考基因组的平均肿瘤突变负荷。所述体细胞突变参考水平可以是患有与所述受试者相同种类的癌症的参考群体的平均肿瘤突变负荷。所述体细胞突变参考水平可以是未患有癌症的参考群体的平均肿瘤突变负荷。所述体细胞突变参考水平可以是未受益于所述治疗的参考群体的平均肿瘤突变负荷。所述体细胞突变参考水

平可以用来自所述受试者的不同样品获得。所述肿瘤突变负荷阈值可以为15或20或30或40,并且所述肿瘤突变负荷由公式II给出:

[0046]  $TMB = N(S > \text{阈值}) / (N(\text{HomHet}) + N(\text{HetHet})) * 1000000$  公式II

[0047] 其中N是体细胞突变显著性得分高于所述阈值的体细胞变体的数量,所述数量通过杂合SNP区域中的位置的总数(N(HomHet)+N(HetHet))进行归一化。

[0048] 一种用于治疗有需要的受试者的癌症的方法,所述方法包括:

[0049] (a) 对来自所述受试者的肿瘤样品的细胞进行测序;

[0050] (b) 鉴定一组杂合SNP位置,其中每个SNP具有等位基因B和A;

[0051] (c) 检测两个种系等位基因配对的SNP位置和在所述SNP位置附近的位置的变体,其中所述两个种系等位基因配对是(i)等位基因B和第一变体等位基因以及(ii)等位基因A和第二变体等位基因,所述第二变体等位基因可以与所述第一变体等位基因相同或不同;以及

[0052] (d) 检测第三等位基因配对,所述第三等位基因配对是(iii)等位基因B和第三变体等位基因,所述第三变体等位基因不同于所述第一变体等位基因,其中所述第三等位基因配对源于体细胞变体;

[0053] (e) 计算来自检测到的所述体细胞变体的肿瘤突变负荷的值;

[0054] (f) 如果所述肿瘤突变负荷大于参考水平,则鉴定患有癌症的所述受试者受益于治疗;以及

[0055] (g) 施用癌症治疗。

[0056] 一种用于治疗有需要的受试者的癌症的方法,所述方法包括:

[0057] (a) 对来自所述受试者的肿瘤样品的细胞进行测序;

[0058] (b) 使用大规模平行核酸测序方法从所述样品获得序列读段,其中所述序列读段具有读段长度;

[0059] (c) 将所述序列读段映射到参考基因组;

[0060] (d) 组装映射到所述参考基因组的杂合SNP位置的序列读段的体细胞变体计数矩阵,其中所述计数矩阵具有第一元件和第二元件,所述第一元件和所述第二元件分别对SNP等位基因B和A与变体等位基因的等位基因配对进行计数,并且其中所述计数矩阵具有第三元件,所述第三元件对来自与所述第一元件中不同的变体等位基因配对的SNP等位基因B的读段序列进行计数;

[0061] (e) 通过以下步骤计算所述样品的肿瘤突变负荷的值:

[0062] (i) 针对每个体细胞变体计算所述第三元件的体细胞突变显著性得分(S);以及

[0063] (ii) 从体细胞突变显著性得分高于阈值的体细胞变体的数量计算所述肿瘤突变负荷的所述值,所述数量通过杂合SNP区域中的位置的总数进行归一化;

[0064] (f) 如果所述肿瘤突变负荷大于体细胞突变参考水平,则鉴定患有癌症的所述受试者将受益于治疗;以及

[0065] (g) 施用癌症治疗。所述癌症治疗可以包括施用免疫检查点抑制剂药物。

[0066] 一种用于治疗有需要的受试者的癌症的方法,所述方法包括:

[0067] (a) 对来自所述受试者的肿瘤样品的细胞进行测序;

[0068] (b) 使用大规模平行核酸测序方法从所述样品获得序列读段,其中所述序列读段

具有读段长度；

[0069] (c) 将所述序列读段映射到参考基因组；

[0070] (d) 组装映射到所述参考基因组的杂合SNP位置的序列读段的体细胞变体计数矩阵,其中所述计数矩阵具有第一元件和第二元件,所述第一元件和所述第二元件分别对SNP等位基因B和A与变体等位基因的等位基因配对进行计数,并且其中所述计数矩阵具有第三元件,所述第三元件对来自与所述第一元件中不同的变体等位基因配对的SNP等位基因B的读段序列进行计数；

[0071] (e) 通过以下步骤计算所述样品的肿瘤突变负荷的值：

[0072] (i) 针对每个体细胞变体计算所述第三元件的体细胞突变显著性得分(S)；以及

[0073] (ii) 从体细胞突变显著性得分高于阈值的体细胞变体的数量计算所述肿瘤突变负荷的所述值,所述数量通过杂合SNP区域中的位置的总数进行归一化；

[0074] (f) 如果所述肿瘤突变负荷大于体细胞突变参考水平,则鉴定患有癌症的受试者将受益于治疗；

[0075] (g) 在一段时间内监测所述受试者的癌症体征和症状；以及

[0076] (h) 施用癌症治疗。所述治疗可以是施用免疫检查点抑制剂。

[0077] 一种用于监测患有癌症的受试者对治疗的应答的方法,所述方法包括：

[0078] (a) 对来自所述受试者的肿瘤样品的细胞进行测序；

[0079] (b) 鉴定一组杂合SNP位置,其中每个SNP具有等位基因B和A；

[0080] (c) 检测两个种系等位基因配对的SNP位置和在所述SNP位置附近的位置的变体,其中所述两个种系等位基因配对是(i)等位基因B和第一变体等位基因以及(ii)等位基因A和第二变体等位基因,所述第二变体等位基因可以与所述第一变体等位基因相同或不同；以及

[0081] (d) 检测第三等位基因配对,所述第三等位基因配对是(iii)等位基因B和第三变体等位基因,所述第三变体等位基因不同于所述第一变体等位基因,其中所述第三等位基因配对源于体细胞变体；

[0082] (e) 计算来自检测到的所述体细胞变体的肿瘤突变负荷的值。

[0083] 一种用于监测患有癌症的受试者对治疗的应答的方法,所述方法包括：

[0084] (a) 对来自所述受试者的肿瘤样品的细胞进行测序；

[0085] (b) 使用大规模平行核酸测序方法从所述样品获得序列读段,其中所述序列读段具有读段长度；

[0086] (c) 将所述序列读段映射到参考基因组；

[0087] (d) 组装映射到所述参考基因组的杂合SNP位置的序列读段的体细胞变体计数矩阵,其中所述计数矩阵具有第一元件和第二元件,所述第一元件和所述第二元件分别对SNP等位基因B和A与变体等位基因的等位基因配对进行计数,并且其中所述计数矩阵具有第三元件,所述第三元件对来自与所述第一元件中不同的变体等位基因配对的SNP等位基因B的读段序列进行计数；

[0088] (e) 通过以下步骤计算所述样品的肿瘤突变负荷的值：

[0089] (i) 针对每个体细胞变体计算所述第三元件的体细胞突变显著性得分(S)；以及

[0090] (ii) 从体细胞突变显著性得分高于阈值的体细胞变体的数量计算所述肿瘤突变

负荷的所述值,所述数量通过杂合SNP区域中的位置的总数进行归一化。

[0091] 一种用于对患有癌症的受试者进行预后的方法,所述方法包括:

[0092] (a) 对来自所述受试者的肿瘤样品的细胞进行测序;

[0093] (b) 鉴定一组杂合SNP位置,其中每个SNP具有等位基因B和A;

[0094] (c) 检测两个种系等位基因配对的SNP位置和在所述SNP位置附近的位置的变体,其中所述两个种系等位基因配对是(i)等位基因B和第一变体等位基因以及(ii)等位基因A和第二变体等位基因,所述第二变体等位基因可以与所述第一变体等位基因相同或不同;以及

[0095] (d) 检测第三等位基因配对,所述第三等位基因配对是(iii)等位基因B和第三变体等位基因,所述第三变体等位基因不同于所述第一变体等位基因,其中所述第三等位基因配对源于体细胞变体;

[0096] (e) 计算来自检测到的所述体细胞变体的肿瘤突变负荷的值;以及

[0097] (f) 如果所述肿瘤突变负荷大于TMB参考水平,则将所述受试者预后为具有不良预后。

[0098] 一种用于对患有癌症的受试者进行预后的方法,所述方法包括:

[0099] (a) 对来自所述受试者的肿瘤样品的细胞进行测序;

[0100] (b) 使用大规模平行核酸测序方法从所述样品获得序列读段,其中所述序列读段具有读段长度;

[0101] (c) 将所述序列读段映射到参考基因组;

[0102] (d) 组装映射到所述参考基因组的杂合SNP位置的序列读段的体细胞变体计数矩阵,其中所述计数矩阵具有第一元件和第二元件,所述第一元件和所述第二元件分别对SNP等位基因B和A与变体等位基因的等位基因配对进行计数,并且其中所述计数矩阵具有第三元件,所述第三元件对来自与所述第一元件中不同的变体等位基因配对的SNP等位基因B的读段序列进行计数;

[0103] (e) 通过以下步骤计算所述样品的肿瘤突变负荷的值:

[0104] (i) 针对每个体细胞变体计算所述第三元件的体细胞突变显著性得分(S);以及

[0105] (ii) 从体细胞突变显著性得分高于阈值的体细胞变体的数量计算所述肿瘤突变负荷的所述值,所述数量通过杂合SNP区域中的位置的总数进行归一化;

[0106] (f) 如果所述肿瘤突变负荷大于TMB参考水平,则将所述受试者预后为具有不良预后;以及

[0107] (g) 施用癌症治疗。

[0108] 一种用于鉴定患有癌症的受试者受益于治疗的试剂盒,所述试剂盒包括:

[0109] (a) 用于从来自所述受试者的样品获得序列读段的试剂,其中所述序列读段可以用于获得所述样品的肿瘤突变负荷的值;以及

[0110] (b) 使用所述试剂用于获得所述序列读段并且使用肿瘤突变负荷的所述值用于鉴定所述受试者的说明书。

[0111] 一种用于检测体细胞变体的系统,所述系统包括:

[0112] 用于从样品接收、富集和扩增核酸的装置,其中所述样品含有癌细胞和非癌细胞;

[0113] 用于从所述核酸合成文库的装置;



- [0114] 用于使所述文库与测序芯片接触的装置；
- [0115] 用于检测所述文库中的序列并且将序列数据传送到处理器的装置；
- [0116] 一个或多个处理器,所述一个或多个处理器用于执行以下步骤:
- [0117] (a) 提供含有癌细胞和非癌细胞的样品;
- [0118] (b) 使用大规模平行核酸测序方法从所述样品获得序列读段,其中所述序列读段具有读段长度;
- [0119] (c) 将所述序列读段映射到参考基因组;
- [0120] (d) 组装映射到所述参考基因组的杂合SNP位置的序列读段的体细胞变体计数矩阵,其中所述计数矩阵具有第一元件和第二元件,所述第一元件和所述第二元件分别对SNP等位基因B和A与变体等位基因的等位基因配对进行计数,并且其中所述计数矩阵具有第三元件,所述第三元件对来自与所述第一元件中不同的变体等位基因配对的SNP等位基因B的读段序列进行计数;
- [0121] (e) 通过以下步骤计算所述样品的肿瘤突变负荷的值:
- [0122] (i) 针对每个体细胞变体计算所述第三元件的体细胞突变显著性得分(S);以及
- [0123] (ii) 从体细胞突变显著性得分高于阈值的体细胞变体的数量计算所述肿瘤突变负荷的所述值,所述数量通过杂合SNP区域中的位置的总数进行归一化;以及
- [0124] 显示器,所述显示器用于对序列信息进行显示、制图和报告。
- [0125] 一种非暂时性机器可读存储介质,其中存储有用于由处理器执行的指令,所述指令使所述处理器执行用于检测体细胞变体的方法的步骤,所述方法包括:
- [0126] (a) 提供含有癌细胞和非癌细胞的样品;
- [0127] (b) 使用大规模平行核酸测序方法从所述样品获得序列读段,其中所述序列读段具有读段长度;
- [0128] (c) 将所述序列读段映射到参考基因组;
- [0129] (d) 组装映射到所述参考基因组的杂合SNP位置的序列读段的体细胞变体计数矩阵,其中所述计数矩阵具有第一元件和第二元件,所述第一元件和所述第二元件分别对SNP等位基因B和A与变体等位基因的等位基因配对进行计数,并且其中所述计数矩阵具有第三元件,所述第三元件对来自与所述第一元件中不同的变体等位基因配对的SNP等位基因B的读段序列进行计数;
- [0130] (e) 通过以下步骤计算所述样品的肿瘤突变负荷的值:
- [0131] (i) 针对每个体细胞变体计算所述第三元件的体细胞突变显著性得分(S);以及
- [0132] (ii) 从体细胞突变显著性得分高于阈值的体细胞变体的数量计算所述肿瘤突变负荷的所述值,所述数量通过杂合SNP区域中的位置的总数进行归一化;以及
- [0133] (f) 对来自所述样品的序列信息进行显示、制图和报告。

## 附图说明

- [0134] 图1:用于通过核酸测序检测和评估肿瘤突变负荷的方法和步骤的图示。
- [0135] 图2:种系等位基因和种系变体的图示。(顶部)杂合变体V/W的种系等位基因,其位于杂合SNP B/A附近。每个SNP等位基因仅与一个变体等位基因相关,并且对于覆盖SNP和VAR位置两者的读段,预期仅两个独特的序列读段BV和AW。(底部)纯合变体W/W的种系等位

基因,其位于杂合SNP B/A附近。每个SNP等位基因仅与一个变体等位基因相关,并且对于覆盖SNP和VAR位置两者的读段,预期仅两个独特的序列读段BW和AW。

[0136] 图3:体细胞等位基因和体细胞变体的图示。(顶部)观察到杂合变体V/W的等位基因,其位于杂合SNP B/A附近。对于覆盖SNP和VAR位置两者的读段,预期两个正常等位基因配对BV和AW有两个独特的序列读段。然而,SNP等位基因B与两个变体等位基因BV和BW相关。因此,BW表示新生突变。这些读段的矩阵示出了BV和AW的大(L)计数,以及可能较小的BW的(s)计数。(底部)观察到纯合变体W/W的等位基因,其位于杂合SNP B/A附近。对于覆盖SNP和VAR位置两者的读段,预期两个正常等位基因配对BW和AW有两个独特的序列读段。然而,SNP等位基因B与两个变体等位基因BV和BW相关。因此,BV表示新生突变。这些读段的矩阵示出了BW和AW的大(L)计数,以及可能较小的BV的(s)计数。

[0137] 图4:用于通过核酸测序检测和评估肿瘤突变负荷的方法的示例实施例。对于位于杂合SNP (Hom/Het)附近的纯合体细胞变体,如所示将序列读段堆叠映射到参考基因组(WT)。组装示出了等位基因配对GA(计数55)、AA(计数32)和AG(计数23)的计数的计数矩阵。第三最大计数AG(计数23)的出现源于一些癌细胞中的体细胞突变。

[0138] 图5:用于通过核酸测序检测和评估肿瘤突变负荷的方法的示例实施例。对于位于杂合SNP (Het/Het)附近的杂合体细胞变体,组装示出了等位基因CG(计数39)、GT(计数34)和GG(计数7)的计数的计数矩阵。第三最大计数GG(计数7)的出现源于一些癌细胞中的体细胞突变。

[0139] 图6:来自结肠癌样品的测序数据的图示。每条曲线通过等位基因比率%(X轴)表示变体位置的数量(Y轴)。一个样品示出了表示高TMB样品的大峰。在非常低的等位基因比率值(小于10%)下位于左侧的高峰反映了被忽略的测序误差。为了对TMB值进行计数,对于大于30的得分(Y轴),TMB值可以计算为等位基因比率在约15%到约65%的范围内的曲线下面积。

[0140] 图7:与涉及从种系比较物样品中减去数据或种系过滤的常规方法相比,来自用于通过核酸测序检测和评估结肠癌和乳腺癌样品中的肿瘤突变负荷的本发明的基于SNP的方法的数据的曲线图。使用仅具有肿瘤样品而没有第二种系比较物样品的本发明的直接SNP分析方法(实心圆),获得了对令人惊讶地优于常规方法的肿瘤突变负荷的评估。与常规方法相比,本发明的基于SNP的方法(实心圆)的灵敏度令人惊讶地增加。更具体地,本发明的基于SNP的方法(实心圆)令人惊讶地比用于使用已知种系变体的数据库并且过滤常见变体以尝试去除种系背景(空心圆)来评估肿瘤突变负荷的核酸测序的方法更准确。

## 具体实施方式

[0141] 本发明提供了用于检测癌细胞中的体细胞突变的方法、组合物、试剂盒和系统。体细胞突变的测量结果可以提供癌症的治疗方法、诊断方法和预后方法。

[0142] 在一些方面,本发明提供了用于选择和鉴定受益于治疗(如使用抗癌剂治疗癌症)的受试者的方法。对于这些受试者,可以选择治疗方式来治疗癌症。

[0143] 在另外的方面,本发明提供了用于对癌细胞中的肿瘤突变频率进行测量和评分的方法。得分可以用于计算来自受试者的样品的肿瘤突变负荷。肿瘤突变负荷可以充当疾病,例如癌症的生物标志物。

[0144] 体细胞变体可以与受试者对使用某些药物的治疗的应答相关。例如,高肿瘤突变负荷值可以与患有癌症的受试者对施用免疫检查点抑制剂药物的有利应答相关。

[0145] 如本文所使用的,与体细胞变体的频率有关的量可以被定义为“肿瘤突变负荷”(TMB)。TMB可以计算为相对于在确定体细胞变体的计数中测定的基因组位置的总数归一化的癌症样品中的体细胞变体的计数。TMB可以表示为每兆碱基DNA的突变数。

[0146] TMB也可以从RNA进行测量并且表达为每兆碱基RNA的突变数。

[0147] 可以获得TMB的量度作为在一组基因组位置中的体细胞变体的量度。所述一组基因组位置可以是所述基因组的一组SNP区域。

[0148] 在一些实施例中,可以使用测序数据或测序读段来鉴定一组杂合SNP位置。

[0149] 在一些实施例中,可以使用已知的人类SNP位置来鉴定一组杂合SNP位置。

[0150] 本发明的TMB的量度可以是基因组的体细胞突变负荷的替代物。本发明的TMB的量度可以提供直接反映基因组的体细胞突变数量的数值水平。本发明的TMB的量度可以提供数值水平,所述数值水平可以是基因组的总突变负荷的有效估计。本发明的TMB的量度可以不同于在其它文献中标记为“TMB”的量。

[0151] 在一些方面,本发明提供了用于检测体细胞突变并确定突变水平的方法和系统。突变负荷可以从涵盖检测基因组中体细胞突变的独特算法获得,其中体细胞突变各自位于基因组中SNP位置阵列中的SNP位置附近。

[0152] 在某些方面,本发明的TMB的量度可以从涵盖检测基因组中的一部分体细胞突变的独特算法获得,其中体细胞突变各自位于基因组中SNP位置阵列中的SNP位置附近。

[0153] 在另外的方面,本发明的TMB的量度可以提供直接反映基因组的体细胞突变数量的数值水平,其中突变可以影响基因组中位置的功能。

[0154] 在另外的方面,本发明的用于测量TMB的方法可以利用通过提供感兴趣的基因座的多个独立读段的任何测序技术获得的数据。在各个实施例中,可以利用桑格序列方法(Sanger sequence method)。

[0155] 在另外的方面,本发明的用于测量TMB的方法可以与任何SNP组、全外显子组/基因组测序和其中可以对SNP进行测序的基因组一起利用。

[0156] 在一些实施例中,可以使用HRD(麦利亚德基因公司(Myriad Genetics, Inc.))测序,其是基于杂交捕获的基因小组,所述基因小组也从整个基因组中取样SNP。HRD测定可以利用SNP来重建可以从中推导出HRD得分的肿瘤-CN/LOH概况。HRD测定可以用于对大量SNP基因座进行测序。

[0157] 在某些实施例中,可以使用具有足够数量的SNP(包含两侧上的侧翼区域)的任何测序数据。

[0158] 在另外的方面,任何基于序列的NGS测定可以用于本发明的用于测量TMB的方法中。

[0159] 在另外的方面,本发明的实施例提供了用于治疗患有癌症的受试者的方法。可以通过对来自受试者的样品中的肿瘤突变负荷进行评估来选择和鉴定患有癌症的受试者。可以用抗癌剂(如有效量的免疫检查点抑制剂)治疗受试者。

[0160] 本发明的方面包含用于以有利地优越的灵敏度检测样品中的体细胞变体的方法、组合物和系统,包含本发明的TMB的量度。

[0161] 本发明可以进一步提供用于对样品的核酸进行测序的改进方法。本发明的改进的测序方法可以用于准确地检测和计数体细胞变体。

[0162] 本公开中描述的实施例包含用于治疗癌症以及鉴定受益于治疗的受试者的方法。本发明的独特方法可以用来自受试者的单一样品而无需非癌症比较物样品进行。本公开的方法提供体细胞变体的直接量度,其可以用于确定体细胞变体得分和肿瘤突变负荷的值。来自受试者的样品,如来自患有癌症的受试者的肿瘤或组织样品中的体细胞突变的直接测量结果和肿瘤突变负荷的评估可以提供疾病的准确生物标志物。

[0163] 本发明的另外的方面包含用于直接检测体细胞变体的方法,其可以减少由于种族偏差而引起的误差。本公开的方法可以通过对可以仅归因于癌细胞的序列读段进行计数来从单个测试样品中检测体细胞变体。在这些方法中,可以确定与个体相关并且较少受群体或种族偏差影响的肿瘤突变负荷。

[0164] 通过本发明的方法确定的肿瘤突变负荷在某些癌症中可以是特别预测的。肿瘤突变负荷可以用于检测和诊断癌症,以及确定预后。

[0165] 癌症的实例包含前列腺癌、黑色素瘤、膀胱癌、乳腺癌、血液癌、间皮瘤、肺癌和实体瘤。

[0166] 在一些实施例中,本发明提供了用于评估肿瘤突变负荷的方法,其中异常状态可能指示不良预后。

[0167] 在另外的实施例中,用于评估肿瘤突变负荷的方法可以与一个或多个临床参数组合以诊断癌症和/或对癌症进行预后。

[0168] 临床参数的实例包含例如临床列线图。

[0169] 在某些实施例中,高水平的肿瘤突变负荷可以指示癌症的存在。

[0170] 在另外的实施例中,高水平的肿瘤突变负荷可以指示受试者的癌症复发或进展的风险增加,对于所述受试者,临床列线图得分指示相对低的复发或进展风险。

[0171] 例如,高水平的肿瘤突变负荷可以示出与肿瘤分级或分期无关或与列线图得分无关的癌症复发或进展的风险增加。因此,高水平的肿瘤突变负荷可以检测出单独使用临床参数未检测到的增加的风险。

[0172] 在一些方面,本公开提供了包括确定癌症患者的至少一个临床参数并且确定从患者获得的样品中的肿瘤突变负荷的体外诊断方法。

[0173] 在一些实施例中,肿瘤突变负荷的异常状态可以指示癌症复发或进展的可能性增加。

[0174] 在某些实施例中,一个或多个临床参数与对肿瘤突变负荷的评估的组合可以改进关于癌症的预测能力。在一些实施例中,可以对多于一个临床参数进行评估并且将其与肿瘤突变负荷的评估组合。

[0175] 在另外的方面,本发明包含包括确定患者的至少一个临床参数或列线图得分并且评估患者的肿瘤突变负荷的体外诊断方法。

[0176] 本发明的方面包含通过对来自受试者的组织或细胞样品,更具体地肿瘤样品中的肿瘤突变负荷进行评估来分类癌症的方法。

[0177] 本公开的肿瘤样品可以含有癌症和非癌症正常细胞的混合物。可以获得本公开的肿瘤样品,以便使样品中的非癌症或非肿瘤含量最小化。例如,可以通过在活检中仅切除肿

瘤组织或通过仅去除没有或具有最小正常组织边缘的病变使样品中的非肿瘤含量最小化。

[0178] 在某些实施例中,优选的是使样品中的非肿瘤含量最小化,使得所测量的体细胞突变可以与肿瘤突变负荷的量有关。肿瘤突变负荷量可以用于表征肿瘤中的新生突变或体细胞突变的水平。

[0179] 在另外的实施例中,即使当样品含有一些非肿瘤内容物时,所测量的体细胞突变也可以与肿瘤突变负荷的量有关。肿瘤突变负荷量可以用于表征肿瘤样品中的新生突变或体细胞突变的水平,以便分析受试者的临床状态。

[0180] 本发明的实施例可以有利地在用于检测体细胞突变而无需种系减除的方法中利用含有癌症和非癌细胞的样品。本发明的用于检测体细胞突变而无需种系减除的方法可以对仅存在于肿瘤中的突变数进行计数,甚至在含有癌症和非癌症正常细胞的混合物的样品中也是如此。本发明的用于检测体细胞突变而无需种系减除的方法可以鉴定哪些突变存在于正常细胞中和哪些突变存在于肿瘤细胞中,并且仅对存在于肿瘤中的突变计数。

[0181] 在一些实施例中,可以获得本公开的肿瘤样品以便使样品中的非癌症含量最小化,从而使得可以以增加的准确性和/或精确度检测体细胞突变。

[0182] 在某些实施例中,本发明的方法可以有利地检测癌细胞中的体细胞突变而无需种系减除,甚至在含有癌症和非癌细胞的样品中也是如此。

[0183] 关于肿瘤突变负荷的参考值可以表示具有类似结果的多个训练患者(例如,癌症患者)的平均TMB水平,所述结果的临床数据和随访数据可用并且足以根据疾病结果(例如,复发或预后)对患者进行定义和分类。

[0184] TMB的参考值可以是在已经用抗癌剂治疗的患有癌症的受试者的群体的TMB水平。在一些实施例中,所述群体可以包括已经用特定抗癌剂治疗的一组受试者和已经用不同抗癌剂治疗的另一组受试者。

[0185] TMB的参考值可以是对用抗癌剂的治疗无反应的患有癌症的受试者群体的TMB水平。

[0186] 在一些实施例中,TMB值可以在对用抗癌剂的治疗具有不同反应性的受试者之间进行区分。在某些实施例中,TMB值可以将总存活期增加或用抗癌剂治疗之后无进展存活期的受试者与存活期未增加的受试者进行区分。在另外的实施例中,TMB值可以鉴定受益于治疗性治疗或对治疗性治疗有反应的群体的受试者。

[0187] 可以从表征为具有“良好结果”的多个训练癌症患者中生成“良好预后值”,例如,在一段时间(如在初始治疗后五年或十年或更久)内癌症未复发的患者或在初始诊断后五年或十年或更久的时间内没有癌症进展的患者。

[0188] 可以从定义为具有“不良结果”的多个训练癌症患者中生成“不良预后值”,例如在初始治疗后五年或十年或更久的时间内具有癌症复发的患者或在初始诊断后五年或十年或更久的时间内具有癌症进展的患者。

[0189] 因此,良好预后值可以表示具有“良好结果”的患者的TMB的平均水平,而不良预后值可以表示具有“不良结果”的患者的TMB的平均水平。

[0190] 在一些实施例中,当TMB的值增加时,受试者可能具有不良预后。

[0191] 在某些实施例中,TMB的值可能增加超过正常值或阈值量。

[0192] 在各个实施例中,TMB的值可以比良好预后值更接近不良预后值,这可以指示受试

者的不良预后。

[0193] 在其它实施例中,TMB的值可以比不良预后值更接近良好预后值,这可以指示受试者的良好预后。

[0194] 在另外的实施例中,TMB的值可以通过将患者分配到风险组来确定,并且可以针对TMB平均值设置阈值。

[0195] 阈值可以基于接受者工作特征(ROC)曲线来选择,所述接受者工作特征曲线绘制了灵敏度与{1减去特异性}。

[0196] 在一些实施例中,TMB参考水平可以为约1到约30或约2到约30或约3到约30或约4到约30或约5到约30或约6到约30或约7到约30或约8到约30或约9到约30或约10到约30或约10到约20个突变每Mb。

[0197] 在一些实施例中,TMB参考水平可以为约5到约300或约10到约300或约30到约300或约50到约300个突变每Mb。

[0198] 在一些实施例中,TMB参考水平可以为约1或约2或约3或约4或约5或约6或约7或约8或约9或约10或约20个突变每Mb。

[0199] 在一些实施例中,TMB参考值可以为约30或约50个突变每Mb。

[0200] 通常,可以通过确定癌症的一个或多个临床相关特征和/或确定患有癌症的患者的特定预后来对癌症进行分类。因此,“对癌症进行分类”可以包含:(i)评估转移潜力、转移到具体器官的潜力、复发风险和/或肿瘤进程;(ii)评估肿瘤分期;(iii)在不存在癌症治疗的情况下确定患者预后;(iv)确定患者对治疗(例如,化学疗法、放射疗法、切除肿瘤的外科手术等)的应答(例如,肿瘤收缩或无进展生存期)的预后;(v)诊断患者对当前治疗和/或过去治疗的实际反应;(vi)确定患者的优选治疗进程;(vii)治疗后患者复发的预后(一般治疗或一些特定治疗);(viii)患者预期寿命的预后(例如,总生存期的预后)。

[0201] “阴性分类”是指癌症的不利临床特征(例如,不良预后)。实例包含(i)增加的转移潜力、转移到具体器官的潜力和/或复发风险;(ii)晚期肿瘤分期;(iii)在不存在癌症治疗的情况下的不良患者预后;(iv)患者对特定治疗(例如,化学疗法、放射疗法、切除肿瘤的外科手术等)的应答(例如,肿瘤收缩或无进展生存期)的不良预后;(v)治疗后患者复发的不良预后(一般治疗或一些特定治疗);(vi)患者预期寿命的不良预后(例如,总生存期的预后)。

[0202] 在一些实施例中,复发相关临床参数(或高列线图得分)和增加的TMB可以指示癌症的阴性分类(例如,复发或进展的可能性增加)。

[0203] 通常,TMB的值升高可能伴随癌细胞的快速增殖,这可能指示更具侵袭性的癌症。具有升高的TMB值的受试者在治疗后复发的可能性可能会增加。具有升高的TMB值的受试者可能具有增加的癌症进展或更快速进展的可能性,其中快速增殖的细胞可以引起肿瘤快速生长、增加毒力和/或转移。具有升高的TMB值的受试者可能需要相对更积极的治疗。

[0204] 在一些实施例中,本发明提供了通过评估肿瘤突变负荷来对癌症进行分类的方法,其中异常状态指示复发或进展的可能性增加。

[0205] 在另外的实施例中,本发明提供了通过评估肿瘤突变负荷来确定受试者的癌症的预后的方法,其中升高的TMB可以指示癌症复发或进展的可能性增加。

[0206] 在另外的实施例中,可以在癌症外科手术之前进行评估,例如使用活检样品。在其

它实施例中,可以在癌症外科手术之后进行评估,例如使用切除的癌症样品。

[0207] 在某些实施例中,可以在治疗之前、期间或之后从癌症患者获得一种或多种细胞的样品。

[0208] 癌症治疗的实例包含外科手术切除受影响的器官、放射疗法、激素疗法(例如,使用GnRH拮抗剂、GnRH激动剂、抗雄激素)、化学疗法和高强度聚焦超声。

[0209] 癌症受试者的主动监视包含观察和定期监测而无需进行侵入性治疗。如果出现症状或者如果有迹象表明癌症生长正在进行或加速,则可以在监视期间或之后开始积极治疗。

[0210] 主动监视可能涉及癌症转移风险增加。监视可以会持续一个或多个或一年或多年或更久。

[0211] 本发明可以提供用于治疗癌症患者或提供用于选择患者的治疗的指导的方法。在所述方法中,可以确定对TMB和一个或多个复发相关临床参数的评估。如果来自患者的样品具有升高的TMB并且患者具有一个或多个复发相关临床参数,则可以建议、启动或继续进行积极治疗。如果患者既没有升高的TMB,也没有复发相关临床参数,则可以建议或启动或继续进行主动监视。在某些实施例中,TMB或TMB和一个或多个临床参数可以指示建议积极治疗或建议特定的积极治疗或建议积极治疗。

[0212] 通常,可以建议辅助疗法(例如,在前列腺切除术或放射疗法之后的化学疗法、放射疗法、HIFU、激素疗法等)用于侵袭性疾病。

[0213] 用于检测体细胞突变的方法

[0214] 参考图1,本公开包含用于通过核酸测序检测体细胞突变并且评估基因组的肿瘤突变负荷的方法。

[0215] 在用于检测体细胞变体的方法中,在步骤S101中,可以使用大规模平行核酸测序方法从含有癌细胞和非癌细胞的样品获得序列读段。序列读段的读段长度的范围可以为约50直至约5000个核苷酸。序列读段可以映射到参考基因组。可以在步骤S103中对序列读段进行误差过滤。可以在步骤S105中对核苷酸的碱基调用(base call)进行计数,并且可以在步骤S107中执行位置过滤。可以在步骤S109中组装体细胞变体-SNP序列读段碱基调用计数矩阵。计数矩阵可以使用参考基因组的一组杂合SNP区域。对于每个杂合SNP位置,计数矩阵具有第一元件和第二元件,所述第一元件和第二元件仅对具有位于杂合SNP位置的一个读段长度内的至少第一变体的读段序列进行计数;以及第三元件,所述第三元件仅对来自癌细胞的具有位于杂合SNP位置的一个读段长度内的至少第二变体的读段序列进行计数。在步骤S111中,可以针对位于杂合SNP位置的一个读段长度内的每个体细胞变体计算第三元件的体细胞突变显著性得分(S)。在步骤S113中,可以基于体细胞突变显著性得分计算样品的肿瘤突变负荷。

[0216] 一组杂合SNP区域可以基于与患者无关的一组个体来鉴定。

[0217] 在某些实施例中,可以对位置进行彻底过滤以去除多态性位置。在多于一个样品中具有变体的位置可以被认为是多态性的。有关个体的存在可能会复制变异并且创建错误的多态性位置。因此,在鉴定多态性之前,可以使用一组非相关个体。

[0218] SNP位置组可以是预定的。如果位置是非重复的、非多形态性的和不倾向于高误差率,则这些位置可能是合格的。这可以从基于例如先前分析的约100个或更多个非相关个体

或约50个或更多个非相关个体或约20个或更多个非相关个体或约10个或更多个非相关个体的统计来估计。

[0219] 在某些实施例中,用于计算TMB的合格位置的数量可以为1000个或更多个或5000个或更多个或100,000个或更多个或300,000个或更多个或500,000个或更多个或1,000,000个或更多个或1,500,000个或更多个或1,700,000个或更多个或1,900,000个或更多个或2,000,000个或更多个。

[0220] 在一些实施例中,用于计算TMB的合格位置的数量可以为至少1000个或至少5000或至少100,000或至少300,000或至少500,000或至少1,000,000或至少1,500,000或至少1,700,000或至少1,900,000或至少2,000,000。

[0221] 在一些实施例中,用于计算TMB的合格位置的数量可以为1000到3,000,000或5000到2,500,000、100,000到2,500,000或500,000到2,500,000。

[0222] 在一些实施例中,对于参考基因组的被覆盖的部分,平均读段深度可以为至少50x或100x。

[0223] 样品可以含有癌细胞和非癌细胞。样品中癌细胞和非癌细胞的存在可以允许本发明的方法检测体细胞突变,以及将体细胞突变与种系突变进行区分,而无需使用比较物样品,如种系比较物样品。

[0224] 通常,可以存在癌细胞,因为样品可以取自患有癌症的受试者,并且样品可以含有取自癌症部位的组织或细胞。在一些实施例中,样品可以是来自肿瘤去除的组织或细胞。在某些实施例中,样品可以是来自恶性肿瘤去除的组织或细胞。在另外的实施例中,样品可以是来自肿瘤去除的组织或细胞,所述肿瘤包含非肿瘤组织或细胞的边缘。

[0225] 本发明的实施例包含独特的算法,所述算法用于仅使用来自受试者的单一样品直接检测体细胞突变并且评估肿瘤突变负荷的方法中,而无需用于减去从比较物样品获得的种系量的步骤。

[0226] 图2示出了种系等位基因和种系变体的图示。在图2中,顶部示出了种系细胞中具有等位基因V和W的杂合变体位置的核酸序列,其位于具有等位基因B和A的杂合SNP附近。每个SNP等位基因仅与一个变体等位基因,即BV和AW相关。在检测这些等位基因配对时,预期仅两个独特的序列检测,BV和AW。在通过片段化进行的测序中,对于覆盖SNP和VAR位置两者的读段长度,预期仅两个独特的序列读段,BV和AW。

[0227] 在图2顶部可以注意到,具有与B相关的两个变体等位基因V和W的概率极小到零。

[0228] 在图2中,底部示出了种系细胞中具有等位基因W和W的纯合变体位置的核酸序列,其位于具有等位基因B和A的杂合SNP附近。每个SNP等位基因与同一变体等位基因,即BW和AW相关。在检测这些等位基因配对时,预期仅两个独特的序列检测,BW和AW。在通过片段化进行的测序中,对于覆盖SNP和VAR位置两者的读段长度,预期仅两个独特的序列读段,BW和AW。

[0229] 图3示出了体细胞等位基因和体细胞变体的图示。

[0230] 在图3中,顶部示出了样品细胞中具有等位基因V和W的杂合变体位置的核酸序列,其位于具有等位基因B和A的杂合SNP附近。在没有体细胞突变变体的细胞中,每个SNP等位基因将仅与一个变体等位基因,例如BV和AW相关。在检测这些等位基因配对时,预期仅两个独特的序列检测,BV和AW。在通过片段化进行的测序中,对于覆盖SNP和VAR位置两者的读段



长度,预期仅两个独特的序列读段,BV和AW。因此,对于两个正常预期的等位基因配对BV和AW,将存在相对较大的读段计数 $L_1$ 和 $L_2$ 。在具有体细胞突变变体的癌细胞中,SNP等位基因将与第二变体等位基因,例如BW相关。因此,对于新的等位基因配对BW,将存在相对较小的读段计数s。s的非零计数的存在指示SNP等位基因B被发现或与两个不同的变体等位基因V和W相关。因此,V或W可以被视为新生突变,并且更具体地体细胞突变。s的非零计数指示BW通过体细胞突变源于癌细胞。

[0231] 在图3中,顶部示出了具有等位基因V和W的杂合变体位置的Het-Het计数矩阵,其位于具有等位基因B和A的杂合SNP附近。在不存在癌细胞的情况下或者在不存在体细胞突变的情况下,s为零,并且图3顶部变得等同于图2顶部。

[0232] 本发明的实施例考虑了作为体细胞突变的等位基因比率的特征。等位基因比率可以定义为非野生型碱基的比率,并且可以在0到100%之间变化。

[0233] 通常,等位基因比率描述了变体等位基因相对于WT参考等位基因的分数,并且可以在0到100%之间变化。

[0234] 通常,如果不存在含有体细胞突变的癌细胞,则可以发现等位基因比率为零。通常,等位基因比率为100%将指示体细胞突变以高水平存在。

[0235] 在图3中,底部示出了样品细胞中具有等位基因W和W的纯合变体位置的核酸序列,其位于具有等位基因B和A的杂合SNP附近。在没有体细胞突变变体的细胞中,每个SNP等位基因将仅与一个变体等位基因,例如BW和AW相关。在检测这些等位基因配对时,预期仅两个独特的序列检测,BW和AW。在通过片段化进行的测序中,对于覆盖SNP和VAR位置两者的读段长度,预期仅两个独特的序列读段,BW和AW。因此,对于两个正常预期的等位基因配对BW和AW,将存在相对较大的读段计数 $L_1$ 和 $L_2$ 。在具有体细胞突变变体的癌细胞中,SNP等位基因将与第二变体等位基因,例如BV相关。因此,对于新的等位基因配对BV,将存在相对较小的读段计数s。s的非零计数的存在指示SNP等位基因B被发现或与两个不同的变体等位基因V和W相关。因此,V或W可以被视为新生突变,并且更具体地体细胞突变。s的非零计数指示BV通过体细胞突变源于癌细胞。

[0236] 在图3中,底部示出了具有等位基因W和W的纯合变体位置的Hom-Het计数矩阵,其位于具有等位基因B和A的杂合SNP附近。在不存在癌细胞的情况下或者在不存在体细胞突变的情况下,s为零,并且图3底部变得等同于图2底部。

[0237] 非零s的存在指示SNP等位基因B被发现或与两个不同的变体等位基因V和W相关,并且因此鉴定存在新生突变。

[0238] 在一些实施例中,对于位于杂合SNP附近的变体,在噪声水平以上可检测的第三非零读段计数仅可以源于癌细胞中的体细胞突变。可以在存在非癌细胞的情况下获得第三显著读段计数,而无需减去从第二种系比较物样品获得的任何种系量。事实上,在这种独特的算法中不需要第二种系比较物样品。

[0239] 肿瘤突变负荷

[0240] 在不希望受任何特定理论的束缚的情况下,下文阐述了一种用于评估体细胞突变得分和肿瘤突变负荷(TMB)的方法。

[0241] 根据本发明的TMB值可以使用测序数据来计算,所述测序数据是使用不需要种系减除的本发明的独特算法来自受试者的单一样品获得的。测序数据可以通过本领域已知

的多种方法获得,所述方法包含微电泳法、杂交测序、单分子实时观察和循环阵列测序。

[0242] TMB值可以使用片段化测序数据来计算,所述片段化测序数据是使用不需要种系减除的本发明的独特算法从来自受试者的单一样品获得的。只有长度跨越变体和SNP位置两者的序列读段值可以包含在计数矩阵的组装中。通常,读段应覆盖SNP和待计数的位置。使用比较物样品进行种系减除不是必需的。可以使用一组SNP位置来获得测序数据。可以将SNP的等位基因频率与变体进行比较以确定所述变体是种系还是体细胞的。

[0243] 可以使用约一个读段长度的SNP区域来检测SNP位置附近的变体。读段长度可以足以覆盖SNP位置和变体位置。一组SNP区域可以提供检测体细胞变体并定量样品的TMB值所需的测序数据。

[0244] 如本文所使用的,当变体在SNP位置的约一个测序读段长度内时,所述变体可以在SNP位置“附近”。SNP区域可以是关于SNP位置的 $\pm 1$ 个读段长度。

[0245] 本领域已知的人类SNP位置组的实例包含SNP阵列6.0(昂飞公司(Affymetrix))。

[0246] 对于包含变体位置的SNP区域,可以计算计数矩阵,其中计数矩阵的每个元件C(X1,X2)可以是具有非SNP调用 $X1 = (T, C, G \text{ 或 } A)$ 和SNP调用 $X2 = (T, C, G \text{ 或 } A)$ 的映射读段的数量。

[0247] 量X,Y和P,Q分别与图2和3中的实例V,W和B,A相对应。

[0248] 在这个矩阵中的两个最大计数, $C(X,P) \geq C(Y,Q)$ ,可以归因于四个位置等位基因条件之一:

[0249] HomHom: $C(Y,Q) \leq 3$ 仅留下一个显著计数, $C(X,P)$ ,这表明非SNP和SNP位置两者均是纯合的;

[0250] HetHom: $X \neq Y$ 且 $P=Q$ ,这表明非SNP位置是杂合的并且SNP位置是纯合的;

[0251] HomHet: $X=Y$ 且 $P \neq Q$ ,这表明非SNP位置是纯合的并且SNP位置是杂合的;并且

[0252] HetHet: $X \neq Y$ 且 $P \neq Q$ ,这表明非SNP位置和SNP位置两者均是杂合的。

[0253] 具有杂合SNP位置的HomHet和HetHet条件可以用于将可归因于体细胞突变的读段计数与可归因于正常种系等位基因配对的读段计数进行区分。对于来自患有癌症的受试者的样品,体细胞突变可以归因于癌细胞的存在。这可以在不从单独的样品中单独地获得种系比较物数据的情况下完成。

[0254] 对于上述计数矩阵,矩阵中的第三最大计数 $C(Z,P)$ 或 $C(Z,Q)$ 的存在可以归因于癌细胞的体细胞突变。

[0255] 当计数显著高于背景测序误差率时,第三最大计数可以用于检测体细胞突变。平均误差率E可以从除了最高三个计数之外的所有其它计数来计算。在某些实施例中,平均误差率E可以从矩阵中除了最高三个计数之外的所有其它计数的平均值来计算。

[0256] 体细胞突变的Phred样显著性得分(其是具有一个自由度的卡方概率)可以用公式I来计算:

$$[0257] \quad S = (C(Z,P)^2 / (C(Z,P) + C(X,P)) + (C(Z,P) - E)^2 / E) / 2 * 10$$

[0258] 公式I

[0259] 其中 $C(Z,P)$ 是第三元件计数, $C(X,P)$ 是第一元件计数,并且E是针对所有SNP区域从所述矩阵中的除最高的三个计数之外的所有其它计数的平均值计算的误差率。

[0260] 误差率E的值可以计算为所有位置上的平均值,并且通常为约1或更小。

[0261] TMB水平可以作为具有 $S>30$ 的位置数,所述数量用 $M$ 碱基中的杂合SNP区域中的位置的总数 $\{N(\text{HomHet})+N(\text{HetHet})\}$ 进行归一化,如公式II所示:

[0262]  $\text{TMB} = N(S>30) / (N(\text{HomHet}) + N(\text{HetHet})) * 1000000$

[0263] 公式II

[0264] 在不希望受任何特定理论的束缚的情况下,下文阐述了一种基于以上描述用于确定肿瘤突变负荷(TMB)的的方法。

[0265] TMB值可以使用片段化测序数据来计算,所述片段化测序数据是使用不需要种系减除的本发明的独特算法来自受试者的单一样品获得的。使用比较物样品进行种系减除不是必需的。可以使用一组SNP位置。

[0266] 可以绘制来自一组SNP区域的测序数据,以示出变体位置的数量(y轴)与等位基因比率(x轴)。曲线下面积可以是体细胞变体的存在的估计。使用测序数据的这种排列,通过对曲线下面积进行积分,可以获得被鉴定为体细胞变体的变体总数的值。被鉴定为体细胞变体的变体总数的值可以是TMB的量度。因此,TMB的量度可以作为从约15%的等位基因比率直至约85%的等位基因比率或直至约65%的等位基因比率的曲线下面积获得,其中所述曲线绘制了一组SNP区域中变体位置的数量(y轴)相对于变体的等位基因比率(x轴)的关系。

[0267] 在一些实施例中,TMB的量度可以作为变体计数(y轴)等位基因比率(x轴)曲线下面积(从约15%的等位基因比率直至约50%的等位基因比率、或从约15%的等位基因比率直至约55%的等位基因比率、或从约15%的等位基因比率直至约60%的等位基因比率、或从约15%的等位基因比率直至约65%的等位基因比率、或从约15%的等位基因比率直至约75%的等位基因比率、或从约15%的等位基因比率直至约85%的等位基因比率)获得。

[0268] 通常,在具有非野生型碱基的位置中的体细胞突变的发生可能是罕见的,因此高等位基因比率值的误差可能不太可靠。因此,变体计数(y轴)等位基因比率(x轴)曲线下面积可以优选地取自约15%的等位基因比率直至约65%的等位基因比率,以减少误差。

[0269] 在一些实施例中,平均误差率 $E$ 的量度可以作为变体计数(y轴)等位基因比率(x轴)曲线在等位基因比率为约10-15%下的值来获得。

[0270] 系统

[0271] 在本发明的系统中,样品分析的结果可以以可传输的形式传达给医师、看护者、基因顾问、患者和其它人,所述可传输的形式可以传达或传输给任何上述各方中的任何一方。这种形式可以变化并且可以是有形的或无形的。结果可以体现在描述性陈述、示意图、照片、图表、图像或任何其它可显示形式中。陈述和视觉形式可以记录在有形介质(如纸)、计算机可读介质(如软盘、压缩盘等)上,或者记录在无形介质(例如,电子邮件或互联网或内联网上的网站形式的电子介质)上。另外,结果还可以以声音形式记录,并且通过电话、传真、无线移动电话、互联网电话等通过任何合适的介质(例如,模拟或数字电缆线、光纤电缆等)进行传输。

[0272] 在本发明的系统中,测试结果的信息和数据可以在任何地方产生,并且被传送到不同的位置。本发明进一步涵盖用于产生至少一个患者样品的可传输形式的测试信息的方法。

[0273] 基于计算机的分析功能可以以任何合适的语言和/或浏览器来实施。例如,可以用

C语言并且优选地使用面向对象的高级编程语言(如,Visual Basic、SmallTalk、C++等)来实施。应用程序可以被编写成适合多种环境,如包含Windows™ 98、Windows™ 2000、Windows™ NT等的Microsoft Windows™环境。另外,还可以为MacIntosh™、SUN™、UNIX或LINUX环境编写应用程序。另外,所述功能步骤还可以使用通用的或平台无关的编程语言来实施。此类多平台编程语言的实例包含但不限于超文本标记语言(HTML)、JAVATM、JavaScript™、Flash编程语言、公共网关接口/结构化查询语言(CGI/SQL)、实用提取报告语言(PERL)、AppleScript™和其它系统脚本语言、编程语言/结构化查询语言(PL/SQL)等。可以使用支持Java™或JavaScript™的浏览器,如HotJava™、Microsoft™、Explorer™或Netscape™。当使用活动内容网页时,其可以包含Java™小应用程序或ActiveX™控件或其它活动内容技术。

[0274] 分析功能还可以体现在计算机程序产品中并且用于上述系统或其它基于计算机或互联网的系统。因此,本发明的另一方面涉及一种计算机程序产品,其包括计算机可用介质,所述计算机可用介质具有在其上体现的计算机可读程序代码或指令,用于使得处理器能够执行体细胞突变得分和/或TMB分析。这些计算机程序指令可以加载到计算机或其它可编程设备上以产生机器,使得在计算机或其它可编程设备上执行的指令创建用于实施上述功能或步骤的装置。这些计算机程序指令还可以存储在可以指导计算机或其它可编程设备以特定方式运行的计算机可读存储器或介质中,使得存储在计算机可读存储器或介质中的指令产生包含实施分析的指令装置的制品。计算机程序指令还可以加载到计算机或其它可编程设备上,以使得在计算机或其它可编程设备上执行一系列操作步骤以产生计算机实施的过程,使得在计算机或其它可编程设备上执行的指令提供用于实施上述功能或步骤的步骤。

[0275] 本发明的实施例可以提供一种非暂时性机器可读存储介质,其中存储有由处理器执行的指令,所述指令使处理器执行用于确定和计算TMB的方法的步骤。

[0276] 非易失性、非暂时性机器可读存储介质的实例包含各种类型的只读存储器(ROM)、硬盘驱动器、固态存储器装置、闪存驱动器、压缩盘只读存储器(CD-ROM)、DVD、光盘、磁盘或可以用于携带或存储具有计算机可执行指令或数据结构的程序代码的任何其它存储介质。所述介质可以由通用或专用计算机(如处理器)访问。

[0277] 本发明的实施例可以提供一种计算系统,其可以具有一个或多个处理器、一个或多个存储器装置、文件系统、通信模块、操作系统和/或用户接口,所述一个或多个处理器、一个或多个存储器装置、文件系统、通信模块、操作系统和/或用户接口中的每一个可以通信地耦合。

[0278] 计算系统可以具有操作系统,所述操作系统可以被布置成利用各种硬件和软件资源。操作系统可以被布置成接收并执行用于系统的其它组件的指令。

[0279] 计算系统的实例包含膝上型计算机、台式计算机、服务器计算机、移动电话或智能手机、平板电脑和其它便携式计算系统。

[0280] 计算系统的实例包含处理器、专用或通用计算机。

[0281] 处理器可以被布置成执行存储在机器可读存储介质上的指令。处理器可以包含一个或多个微处理器、各种控制器、数字信号处理器或专用集成电路,并且可以接收和/或传送数据,以及执行所存储的指令以转换数据。在一些实施例中,处理器可以接收、解释和执

行来自程序代码或各种介质的指令。处理器可以接收和转换数据,以及将数据存储于存储器或文件中。在某些实施例中,处理器可以从存储器或文件中取出指令并且将指令接收到存储器中。

[0282] 机器可读存储介质可以是非易失性的。存储器或介质可以在文件系统中存储指令或数据文件,并且可以包含机器可读存储介质。机器可读存储介质可以是非暂时性的。机器可读存储介质可以在其中存储可以由处理器执行的指令。

[0283] 通信装置可以是能够传送和/或接收数据的任何设备、系统或组件的组合。可以通过网络或通信线路传送和/或接收数据。通信装置可以通信地链接到其它组件。

[0284] 通信装置的实例包含网卡、调制解调器、天线、红外或可见通信组件、蓝牙组件、通信芯片组、广域网、WiFi组件、802.6或更高级装置以及蜂窝通信装置。通信装置可以通过线路、电线或网络与其它组件、装置或系统交换数据。

[0285] 本公开的系统可以包含一个或多个处理器、一个或多个非暂时性机器可读存储介质、一个或多个文件系统、一个或多个存储器装置、操作系统、一个或多个通信模块和一个或多个用户接口,所述一个或多个处理器、一个或多个非暂时性机器可读存储介质、一个或多个文件系统、一个或多个存储器装置、操作系统、一个或多个通信模块和一个或多个用户接口中的每一个可以通信地链接。

[0286] 在以下文献中描述了一些计算生物学方法:例如Setubal等人,《计算生物学方法导论(Introduction To Computational Biology Methods)》(1997);Salzberg等人,《分子生物学计算方法(Computational Methods In Molecular Biology)》(1998);Rashidi和Buehler,《生物信息学基础:在生物科学和医学中的应用(Bioinformatics Basics: Application In Biological Science And Medicine)》(2000);Ouelette和Bzevanis,《生物信息学:基因和蛋白质分析的实用指南(Bioinformatics:A Practical Guide For Analysis Of Gene And Proteins)》(2001)。

[0287] 抗癌剂

[0288] 免疫检查点抑制剂药物可以释放T细胞,以杀死受试者的癌细胞。这些药物可以阻断使癌细胞能够逃避免疫系统并且改进存活率的蛋白质。

[0289] 免疫检查点抑制剂是可以预防或抑制免疫细胞和/或免疫应答被旨在被杀死的非常癌细胞关闭或下调或抑制的治疗剂。

[0290] 通常,免疫检查点抑制剂药物对于少于13%患有癌症的受试者是有效的。因此,能够选择和鉴定受益于用此类药物治疗的受试者是有用的。

[0291] 免疫检查点抑制剂的实例包含PD1抑制剂、易普利姆玛(ipilimumab)(参见例如,Gulley和Dahut,《自然临床实践肿瘤学(Nat.Clin.Practice Oncol.)》(2007)4:136-137)、替西利姆单抗(tremelimumab)(参见例如,Ribas等人,《肿瘤学家(Oncologist)》(2007)12:873-883)和表1中所列出的药剂。

[0292] 表1:检查点抑制剂

	药物	靶标	用途
	易普利姆玛 (Yervoy) (易普利姆玛、MDX-010、MDX-101) (百时美施贵宝 (Bristol-Myers Squibb))	CTLA4	黑色素瘤、NSCLC、SCLC、膀胱癌、前列腺癌
	替西利姆单抗 (替西利姆单抗 (ticilimumab), CP-675,206) (阿斯利康 (AstraZeneca))	CTLA4	间皮瘤
	纳武单抗 (Opdivo) (纳武单抗 (nivolumab)) (百时美施贵宝)	PD1	恶性黑色素瘤
	派姆单抗 (Keytruda) (派姆单抗 (pembrolizumab), 兰伯丽珠单抗 (lambrolizumab), MK-3475) (默克公司 (Merck))	PD1	恶性黑色素瘤
	MEDI4736 (阿斯利康)	PDL1	NSCLC
	MPDL3280A (罗氏公司 (Roche) / 基因泰克 (Genentech))	PDL1	尿路膀胱癌或 NSCLC
	匹地利珠单抗 (Pidilizumab) (CT-011) (治疗技术公司 (CureTech))	PD1	血液学或实体瘤
[0293]	利瑞鲁单抗 (lirilumab) (BMS-986015) (百时美施贵宝)	KIR	血液学或实体瘤
	吡啶莫德 (Indoximod) (NLG-9189) (纽琳基因公司 (Newlink Genetics))	IDO1	乳腺癌
	INCB024360 (因赛特公司 (Incyte))	IDO1	实体瘤
	MEDI0680 (AMP-514) (阿斯利康)	PD1	实体瘤
	MSB-0010718C (德国默克集团 (Merck KGaA))	PDL1	实体瘤
	PF-05082566 (辉瑞公司 (Pfizer))	4-1BB (CD137)	血液学或实体瘤
	MEDI6469 (阿斯利康)	OX40 (CD134)	实体瘤
	BMS-986016 (百时美施贵宝)	LAG3	血液学或实体瘤
	NLG-919 (纽琳基因公司)	IDO1	实体瘤
	乌瑞芦单抗 (Urelumab) (BMS-663513) (百时美施贵宝)	4-1BB (CD137)	血液学或实体瘤

[0294] 另外的定义

[0295] 提供以下术语或定义仅用于帮助理解本公开。

[0296] 除非在本文中具体定义, 否则本文所使用的所有术语具有与本公开领域的技术人员将理解的相同的含义。

[0297] 在 Sambrook 等人,《分子克隆:实验室手册 (Molecular Cloning: A Laboratory Manual)》, 第2版, 冷泉港实验室出版社 (Cold Spring Harbor Press), 普莱恩维尤, 纽约 (1989); 以及 Ausubel 等人,《当代分子生物学实验指南 (Current Protocols in Molecular Biology)》(增刊47), 约翰·威利父子出版公司 (John Wiley & Sons), 纽约 (1999) 中给出了一些方法。

[0298] 除非本文另外明确定义, 否则本文所使用的术语不应被解释为具有小于本领域的

普通技术人员所理解的范围。

[0299] 如本文所使用的,“单核苷酸多态性”(SNP)或“SNP基因座”是具有在单个碱基上不同的等位基因的基因座,其中较罕见的等位基因在群体中的频率为至少1%。

[0300] 如本文所使用的,基因座处的“等位基因”是在群体中在所述基因座处出现的所有基因变体的集合,每个变体是单个“等位基因”。例如,在SNP基因座处通常仅存在两个等位基因。

[0301] 如本文所使用的,“变体”是测试基因序列与参考基因序列之间的差异。变体可以在单个碱基上不同,或者变体可以在多于一个碱基上不同。变体还包含插入和缺失。

[0302] 如本文所使用的,如果第一变体和第二变体两者均位于同一染色体(母体或父体)DNA链上,则第一变体与第二变体“连锁”。“连锁”是指两个或更多个变体被连锁的状态。

[0303] “位置等位基因模型”是表示测试基因座处的等位基因与SNP基因座处的等位基因之间的连锁的模型。在种系中,位置等位基因模型通常将描述测试基因座处的父体等位基因与SNP基因座处的父体等位基因之间的连锁,以及测试基因座处的母体等位基因与SNP基因座处的母体等位基因之间的连锁。在测试基因座处存在体细胞变体(即,测试基因座处的第三可能等位基因)的情况下,位置等位基因模型将另外描述测试基因座处的第三等位基因与SNP基因座处的母体或父体等位基因之间的连锁。

[0304] 如本文所使用的,“突变”在下面详细描述,但通常是指与受试者的种系相比在体细胞组织中获得核苷酸变化。“突变负荷”在下面详细描述,但通常是指所分析的含有突变的基因座的数量或比例,其中“高突变负荷”或“HML”通常是指超过某个参考值或阈值的数量或比例、或由其推导出的得分。

[0305] 如本文所使用的,“下一代测序”或“NGS”是指使测序过程并行化,同时产生数千或数百万个序列的各种高通量测序过程和技术。NGS通常通过以下步骤进行:首先,通过体外PCR克隆扩增生成DNA测序文库;第二,通过合成对DNA进行测序,使得通过向互补链添加核苷酸而不是通过桑格测序的典型的链终止化学来确定DNA序列;第三,空间分离的扩增的DNA模板在大规模平行过程中同时测序,通常无需物理分离步骤。测序反应的NGS并行化可以在单个仪器运行中生成数百兆碱基到千兆碱基的核苷酸序列读段。与通常报告分子聚集体集合的平均基因型的常规测序技术(如桑格测序)不同,NGS技术通常将许多单个DNA片段的序列(下面详细讨论的序列读段)数字化制表,使得可以检测低频率变体(例如,在核酸分子的异质群体中以低于约10%、5%或1%频率存在的变体)。术语“大规模平行”还可以用于指代通过NGS从许多不同模板分子同时生成序列信息。

[0306] NGS策略可以包含若干种方法,包括但不限于:(i)微电泳法;(ii)杂交测序;(iii)单分子的实时观察;以及(iv)循环阵列测序。循环阵列测序是指通过模板延伸和基于成像的数据收集的迭代循环获得密集DNA阵列的序列的技术。可商购获得的循环阵列测序技术包括但不限于例如在454基因组测序仪(罗氏应用科学(Roche Applied Science);巴塞尔)中使用的454测序;例如在依诺米那基因组分析仪(Illumina Genome Analyzer)、依诺米那HiSeq、MiSeq和NextSeq(圣地亚哥,加利福尼亚州)、SOLiD平台(应用生物系统公司(Applied Biosystems);福斯特城,加利福尼亚州)、Polonator(多佛/哈佛)中使用的Solexa技术和HeliScope单分子测序仪技术(Helicos;剑桥,马萨诸塞州)。其它NGS方法包含单分子实时测序(例如,太平洋生物公司(Pacific Bio))和离子半导体测序(例如,离子

激流测序公司(Ion Torrent sequencing))。对于NGS测序技术的更详细讨论,参见例如,Shendure和Ji,下一代DNA测序(Next Generation DNA Sequencing),《自然生物技术(Nat.Biotech.)》(2008)26:1135-1145。

[0307] 如本文所使用的,“患者”或“个体”或“受试者”是指人。患者、个体或受试者可以是男性或女性。患者、个体或受试者可以是已经经历或正在经历疾病的治疗干预的患者、个体或受试者。患者、个体或受试者还可以是先前未诊断患有疾病的患者、个体或受试者。

[0308] 如本文所使用的,“样品”或“生物样品”是指样品,如活检或组织样品、冷冻样品、血液和血液部分或产品(例如,血清、血小板、红细胞等)、肿瘤样品、唾液、支气管肺泡灌洗液、培养的细胞(例如,原代培养物)、外植体以及转化的细胞、粪便、尿液等。

[0309] “活检”是指去除组织样品以进行诊断或预后评估的过程,并且是指组织样本本身。各种活检技术可以应用于本公开的方法。所应用的活检技术将取决于待评估的组织类型(例如,肺等)、肿瘤的大小和类型以及其它因素。代表性的活检技术包含但不限于切除活检、切口活检、针刺活检、外科手术活检和骨髓活检。“切除活检”是指去除整个肿瘤块,其周围有少量正常组织。“切口活检”是指去除包含肿瘤的横截面直径的楔形组织。通过内窥镜检查或荧光检查做出的诊断可能需要“芯针活检”或“细针抽吸活检”,其通常会从靶组织内获得细胞悬浮液。

[0310] “体液”包含从哺乳动物体内获得的所有流体,无论是经加工的(例如,血清)还是未经加工的,其可以包含例如血液、血浆、尿液、淋巴液、胃液、胆汁、血清、唾液、汗液以及脊髓液和脑液。生物样品通常从受试者获得。

[0311] 如本文所使用的,“癌细胞样品”或“肿瘤样品”意指包括至少一种癌细胞或源自其的生物分子的样本。癌症的实例包含肺癌(例如,非小细胞肺癌(NSCLC))、卵巢癌、结肠直肠癌、乳腺癌、子宫内膜癌和前列腺癌。此类生物分子的非限制性实例包含核酸和蛋白质。“源自”癌细胞样品的生物分子包含位于样品内或从样品提取的分子以及此类生物分子的人工合成的拷贝或版本。此类人工合成的分子的一个说明性的非限制性实例包含PCR扩增产物,其中来自样品的核酸充当PCR模板。癌细胞样品的“核酸”包含位于癌细胞中的核酸或源自癌细胞的生物分子。

[0312] 如本文所使用的,“得分”意指所选的一个值或一组值,以便提供受试者的病状的变量或特性或样品中的突变负荷程度的定量量度,和/或区分、区别或以其它方式表征突变负荷。包括得分的一个或多个值可以基于例如导致从受试者获得的所测得量的一种或多种样品成分的定量数据。在某些实施例中,所述得分可以源自单一成分、参数或评估,而在其它实施例中,所述得分源自多种成分、参数和/或评估。得分可以基于解释函数或者从解释函数推导出;例如,使用各种统计算法中的任何统计算法从特定预测模型推导出的解释函数。“得分变化”可以是指例如从一个时间点到下一个时间点的得分的绝对变化或得分的百分比变化或每单位时间得分的变化(即,得分变化速率)。

[0313] 如本文所使用的,“测试基因座”是基因组基因座(例如,在染色体内的指定位置处的单个核苷酸),其序列或基因型根据本公开进行评估,其中在这种基因座处的突变(例如,与参考基因型或序列相比)潜在地在突变负荷的测量结果中进行计数。

[0314] 如本文所使用的,术语“治疗(treatment)”或“疗法(therapy)”或“治疗方案(therapeutic regimen)”包含受试者的所有临床管理和旨在维持、改善、改进或以其它方



式改变受试者的病状的干预,无论是生物的、化学的、物理的或其组合。这些术语可以在本文中同义地使用。治疗包含但不限于施用预防性药物或治疗化合物(包含小分子和生物药物)、运动方案、物理疗法、饮食调整和/或补充、肥胖外科手术干预、施用治疗化合物(处方或非处方)和在预防HML表征的疾病、延迟所述疾病的发作或改善所述疾病中有效的任何其它治疗。“对治疗的应答”包含受试者对任何上述治疗的应答,无论是生物的、化学的、物理的还是前述的组合。“治疗进程”涉及特定治疗或治疗方案的剂量、持续时间、程度等。本文所使用的初始治疗方案是第一线治疗。

[0315] 本公开的另外的方面

[0316] 本公开的各方面包含以下:

[0317] 用于检测样品中的测试基因座处的体细胞变体的存在的方法,所述方法包括:在来自所述样品的第一连续核酸链上检测单核苷酸多态性(“SNP”)基因座处的第一等位基因和所述测试基因座处的第二等位基因;在来自所述样品的第二连续核酸链上检测所述SNP基因座处的第三等位基因和所述测试基因座处的第四等位基因;以及在来自所述样品的第三连续核酸链上检测所述SNP基因座处的所述第三等位基因和所述测试基因座处的第五等位基因,其中所述第一等位基因和所述第三等位基因是不同的等位基因,并且所述第四等位基因和所述第五等位基因是不同的等位基因。

[0318] 在一些实施例中,第二等位基因和第四等位基因是相同或不同的等位基因。核酸可以是脱氧核糖核酸(DNA)。可以通过测序来检测一个或多个等位基因。可以通过杂交来检测一个或多个等位基因。可以通过聚合酶链反应(PCR)扩增来检测一个或多个等位基因。样品可以包括在测试基因座处具有体细胞变体的细胞和在测试基因座处不具有体细胞变体的细胞。样品可以是组织样品。样品可以是肿瘤样品。

[0319] 用于检测样品中的体细胞变体的方法,所述方法包括:检测个体是杂合的SNP基因座;在所述SNP基因座周围的连续区域内的测试位置处检测与所述SNP基因座处的第一SNP等位基因连锁的第一测试等位基因;以及在所述SNP基因座周围的所述连续区域内的所述测试位置处检测与所述SNP基因座处的所述第一SNP等位基因连锁的第二测试等位基因,其中所述第一测试等位基因和所述第二测试等位基因是不同的等位基因。在一些实施例中,进一步包括在所述SNP基因座周围的所述连续区域内的所述测试位置处鉴定与所述SNP基因座处的第二SNP等位基因连锁的第三测试等位基因,其中所述第一SNP等位基因和所述第二SNP等位基因是不同的等位基因。第一测试等位基因和第三测试等位基因可以是相同的等位基因。第一测试等位基因和第三测试等位基因可以是不同的等位基因。可以通过测序、杂交或通过聚合酶链反应扩增来检测一个或多个等位基因。样品可以包括在测试基因座处具有体细胞变体的细胞和在测试基因座处不具有体细胞变体的细胞。样品可以是组织样品。样品可以是肿瘤样品。

[0320] 用于测量样品中的体细胞变体的频率的方法,所述方法包括:检测所述样品是杂合的多个SNP基因座;在部分a中鉴定的每个SNP基因座周围的连续区域内,测定多个测试基因座以检测所述多个测试基因座的每个测试基因座的与每个SNP等位基因连锁的多个测试等位基因;以及确定变体频率,所述变体频率包括其中检测到的与SNP等位基因连锁的测试等位基因的数量大于一的测试基因座的数量,所述变体频率相对于所测定的测试基因座的总数进行归一化。可以通过测序、通过杂交或通过聚合酶链反应扩增来检测一个或多个等

位基因。样品可以包括在测试基因座处具有体细胞变体的细胞和在测试基因座处不具有体细胞变体的细胞。样品可以是组织样品或肿瘤样品。

[0321] 用于检测体细胞突变的系统,所述系统包括多个传感器,所述多个传感器用于测量一组预定的SNP中的每一个周围的区域中的每个位置的位置等位基因模型编号。

[0322] 用免疫检查点抑制剂治疗个体的方法,所述方法包括:检测所述个体是杂合的多个SNP基因座;在部分a中鉴定的每个SNP基因座周围的连续区域内,测定多个测试基因座以检测所述多个测试基因座的每个测试基因座的与每个SNP等位基因连锁的多个测试等位基因;确定变体频率,所述变体频率包括其中检测到的与SNP等位基因连锁的测试等位基因的数量大于一的测试基因座的数量,所述变体频率相对于所测定的测试基因座的总数进行归一化;并且当所述变体频率超过预定阈值时,向所述个体施用治疗有效量的免疫检查点抑制剂。可以通过测序、通过杂交或通过聚合酶链反应扩增来检测一个或多个等位基因。样品可以包括在测试基因座处具有体细胞变体的细胞和在测试基因座处不具有体细胞变体的细胞。样品可以是组织样品或肿瘤样品。

[0323] 本文具体提及的所有出版物、专利和文献均出于所有目的特此通过引用整体并入。

[0324] 除非另外定义,否则本文所使用的所有技术术语和科学术语的含义与本发明涉及的领域的普通技术人员通常理解的含义相同。尽管类似或等同于本文所述的方法和材料的方法和材料可以用于实践或测试本发明,但下面描述了合适的方法和材料。另外,本文的材料、方法和实例仅是说明性的,并且不旨在是限制性的。

[0325] 尽管出于清楚理解的目的,已经通过说明和举例的方式相当详细地描述了前述公开内容,但是本领域的技术人员将理解,可以在本发明和所附权利要求书的范围内实践各种改变和修改。

[0326] 实例

[0327] 实例1:图4示出了用于通过核酸测序检测和评估肿瘤突变负荷的方法的结果。对于包括位于杂合SNP (Hom/Het)附近的纯合体细胞变体的模型,如所示将序列读段堆叠映射到参考基因组(WT)。组装示出了等位基因配对GA (55)、AA (32)和AG (23)的检测的计数矩阵。第三最大计数AG (23)的出现源于癌细胞中的体细胞突变。

[0328] 等位基因比率计算为VAR位置中不同等位基因的比率。在此Hom-Het实例中,等位基因比率 =  $(23+1) / (32+55+23+1) * 100 = 21.6\%$ 。

[0329] SNP是杂合的,其中等位基因比率为  $(32+23) / \{ (32+23) + (55+1) \} * 100 = 49.5\%$  (A/G55:56)。

[0330] 如图4所示,误差率E为约1.0。因此,S的值为约

[0331]  $S = ((23 \times 23 / (23+55)) + (23-E) (23-E) / E) / 2 \times 10 = 2679$ 。E的值计算为所有位置上的平均值,并且通常为约1.0或更小。

[0332] 对于此示例位置,在图6中样品为306926,具有高TMB。

[0333] 实例2:图5示出了用于通过核酸测序检测和评估肿瘤突变负荷的方法的结果。

[0334] 在此特定实例中,读段长度为100bp,并且总SNP窗口为 $100 * 2 - 1 = 199$ bp。对于此示例位置,在图6中样品为306926,具有高TMB。

[0335] 对于包括位于杂合SNP (Het/Het)附近的杂合体细胞变体的模型,组装示出了等位

基因CG (39)、GT (34) 和GG (7) 的检测的计数矩阵。第三最大计数GG (7) 的出现源于癌细胞中的体细胞突变。

[0336] 等位基因比率计算为VAR位置中不同等位基因的比率。在此Het-Het实例中, 等位基因比率 =  $39 / (34 + 7 + 39) * 100 = 48.8\%$ 。

[0337] SNP作为T/G是杂合的。

[0338] 实例3: 图6示出了来自结肠癌样品的测序数据。每条曲线通过等位基因比率% (X轴) 表示变体位置的数量 (Y轴)。一个样品示出了表示高TMB样品的大峰。在非常低的等位基因比率值 (小于10%) 下位于左侧的高峰反映了被忽略的测序误差。为了对TMB得分进行计数, 将TMB计数视为等位基因比率在15%到65%的范围内的曲线下面积。来自图6的数据示出于表2中。表2的最后两列示出了每1Mb的合格位置的总数和TMB值 (绝对值和归一化值)。样品306926的TMB为417每Mb, 并且样品306932的TMB为32.7每Mb。

[0339] 表2: 结肠癌样品的TMB (每Mb)

[0340]

样品标签	样品ID	覆盖范围	位置总数	MutPos	每Mb
CTCAATGA	306926	100.3	1720440	717	416.8
TCCGTCTA	306927	119.9	2019276	40	19.8
AGGCTAAC	306928	110.8	1856679	32	17.2
CCATCCTC	306929	104.7	1830688	36	19.7
AGATGTAC	306930	106.1	1913312	56	29.3
TCTTCACA	306931	96.4	1459685	13	8.9
CCGAAGTA	306932	113.7	1926863	63	32.7
CGCATACA	306933	100.0	1706073	49	28.7
AATGTTGC	306934	128.8	2076785	23	11.1
TGAAGAGA	306935	115.8	1904586	52	27.3
AGATCGCA	306936	97.3	1774434	29	16.3
AAGAGATC	306937	124.3	2087068	44	21.1
CAACCACA	306938	139.7	2174624	44	20.2
TGGAACAA	306939	155.4	2123021	30	14.1
CCTCTATC	306940	133.8	2152846	16	7.4
ACAGATTC	306941	118.9	2049170	55	26.8

[0341] 位置总数 = 覆盖范围为50或更大的所选位置的数量

[0342] MutPos = 得分为30或更高的变体位置的数量

[0343] 每Mb =  $MutPos * 1000000 / 位置总数$

[0344] 通常, 每Mb具有10个突变的TMB相对较高, 并且当外推到全基因组时与总共超过32,000个体细胞突变相对应。

[0345] 参考图6, TMB是从突变得分为30或更大并且等位基因比率在15-65%的范围内的位置计算出的, 并且通过Mb中的合格位置的总数进行计数和归一化。参考图6, 数据曲线示出了具有所需得分的变体位置 (Y轴) 的数量。

[0346] 实例4: 图7示出了与涉及从种系比较物样品中减去数据或种系过滤的常规方法相比, 使用用于通过核酸测序检测和评估结肠癌和乳腺癌样品中的肿瘤突变负荷的本发明的

基于SNP的方法获得的数据的曲线图。来自图7的数据概括于表3中。

[0347] 结肠癌样品是结肠微卫星 (Colon Micro-Satellite)。乳腺癌样品是一组44个患者样品,其是钼敏感性乳腺肿瘤。

[0348] 表3:本发明的TMB分析与常规方法的比较

编号	样品 d	群组	X 轴	Y 轴 (空心圆)	Y 轴 (实心圆)
1	172326	乳腺	0	8.85	0.433243
2	172327	乳腺	0.5	12.85	4.927275
[0349] 3	172328	乳腺	1.1	9.05	1.353341
4	172332	乳腺	0.9	7.95	1.295587
5	172333	乳腺	0.4	12.2	1.032044
6	172336	乳腺	0.6	7.7	1.142761

[0350]

7	172337	乳腺	1.1	10.55	2.612515
8	172339	乳腺	3.1	12.35	5.639995
9	172340	乳腺	0.1	7.85	0.475758
10	172341	乳腺	0.1	6.8	0.159636
11	172342	乳腺	1.7	10.7	1.649034
12	172345	乳腺	1.8	9.5	2.091111
13	172346	乳腺	1.6	11.35	1.014355
14	172347	乳腺	0.4	21.65	0.573091
15	172349	乳腺	0.2	9	0.834013
16	172350	乳腺	1.9	10.55	2.945048
17	172351	乳腺	0.3	7.4	0.31697
18	172352	乳腺	0.2	9.05	0.421089
19	172353	乳腺	0.7	8.4	0.419443
20	172354	乳腺	0.6	13.45	0.418599
21	172355	乳腺	0.5	9.75	0.569258
22	172356	乳腺	1	6.65	1.125821
23	172357	乳腺	1.6	11.1	3.386773
24	172358	乳腺	1.4	13.75	1.146581
25	172359	乳腺	1.4	8.35	1.268059
26	172360	乳腺	0.7	10.65	1.379488
27	172712	乳腺	3.8	10.55	3.698196
28	172713	乳腺	0.6	4.85	1.254093
29	172716	乳腺	15.1	19.425	4.567614
30	172719	乳腺	1.2	13.65	2.66069
31	172720	乳腺	0	8.2	0
32	172721	乳腺	1.3	14.65	0.890209
33	172722	乳腺	0.8	10.9	1.226617
34	172723	乳腺	2.7	13.35	4.582397
35	172724	乳腺	0	10.6	0
36	172727	乳腺	0.6	9.8	0.965028
37	172728	乳腺	2.4	10.7	3.881554
38	172729	乳腺	0.5	8.525	0
39	172730	乳腺	1.9	8.2	2.296721
40	173206	乳腺	1.4	12.925	2.432384
41	173207	乳腺	2.9	10.325	5.095719
42	173208	乳腺	1.3	9.975	1.652989
43	173210	乳腺	1.1	12.45	2.850926
44	175917	乳腺	1.3	8.9	0.767679
45	193406	结肠	4.173179	27.86667	8.897859
46	193411	结肠	59.46998	132.8667	123.6433
47	193412	结肠	2.884223	14.55	5.940877
48	193413	结肠	1.538395	7.7	1.260531
49	193415	结肠	10.2934	25.1	17.85718

[0351]	50	193416	结肠	27.47211	38.96	24.94902
	51	193417	结肠	19.32901	33.43333	17.20717
	52	193418	结肠	15.11196	24.95	17.73474
	53	193419	结肠	29.84983	48.05	34.01409
	54	193420	结肠	16.15368	35.62	27.02036
	55	271207	结肠	0.719131	12.8	0
	56	271208	结肠	43.15642	79.3	36.93433

[0352] 使用仅具有肿瘤样品而没有第二种系比较物样品的本发明的直接基于SNP的方法(图7,实心圆),获得了对令人惊讶地优于常规方法的肿瘤突变负荷的评估。与常规方法相比,本发明的基于SNP的方法(图7,实心圆)的灵敏度令人惊讶地增加。

[0353] 在图7中,相同x轴位置处的空心圆和实心圆表示与种系过滤(图7,空心圆)相比,通过本发明的方法(图7,实心圆)对同一患者样品的测量结果。

[0354] 在图7中,X轴表示通过全外显子组测序评估的TMB值,其中使用每名患者的基于血液的种系参考样品减去种系变体。关于本发明的方法(图7,实心圆)和种系过滤的方法(图7,空心圆),相同的样品用于全外显子组测序。这种方法被认为是常规的“黄金标准”,基于血液的减除去除了种系变体。

[0355] 在图7中,Y轴示出了本发明的方法(图7,实心圆)和种系过滤的方法(图7,空心圆)与常规的“黄金标准”方法相比的方式。由使用HRD测定获得的数据确定Y轴值。

[0356] 更具体地,本发明的基于SNP的方法(图7,实心圆)令人惊讶地比用于使用已知种系变体的数据库并且过滤常见变体以尝试去除种系背景(图7,空心圆)来评估肿瘤突变负荷的核酸测序的方法更准确。使用已知种系变体的数据库通过核酸测序并且过滤常见变体以尝试去除种系背景(图7,空心圆)来检测和评估肿瘤突变负荷的这种常规方法提供了不准确的肿瘤突变负荷水平。因此,本发明的独特且直接的基于SNP的方法(图7,实心圆)的准确性和灵敏度与尝试减去种系量的方法(图7,空心圆)相比令人惊讶地增加并且出乎意料地有利。

[0357] 进一步地,本发明的直接的基于SNP的方法令人惊讶地优于在从0.1个突变每Mb直至100个突变每Mb(增加1000倍)的广泛突变频率范围内用种系减除进行常规全外显子组测序,因为本发明的直接的基于SNP的方法不需要种系减除样品和改善的灵敏度。更具体地,本发明的基于SNP的方法(图7,实心圆)不利用并且不需要配对的肿瘤和种系比较物样品来减去种系量。本发明的基于SNP的方法(图7,实心圆)仅利用肿瘤样品。本发明的基于SNP的方法,仅使用肿瘤样品,令人惊讶地从种系量检测、鉴定和分离体细胞突变。

[0358] 更具体地,图7示出了,本发明的基于SNP的方法(图7,实心圆)比种系过滤(图7,空心圆)为全外显子组测序(表示为x轴)提供了更一致的结果。如图7所示,在约10TMB每兆碱基或约20TMB每兆碱基下,种系过滤方法(图7,空心圆)是不准确的(偏离线)。因此,种系过滤无法准确地评估低于约10每兆碱基,或甚至低于约20每兆碱基的TMB值。

[0359] 实例5:本发明的方法使用用于直接检测体细胞突变并且仅使用来自患有癌症的受试者的第一单个样品评估肿瘤突变负荷的独特算法而没有用于减去种系量的步骤,所述方法与使用配对的肿瘤和种系比较物样品减去种系量的全外显子组测序(WES)的方法进行比较。本发明的方法进一步与减去种系比较物的MYCHOICE HRD-PLUS方法进行比较。

[0360] 对来自44个乳腺和12个结肠肿瘤的匹配的肿瘤和正常DNA进行WES和MYCHOICE

HRD-PLUS方法中的每一种。MYCHOICE HRD-PLUS测定将同源重组缺陷分析与108个基因的重新测序和MSI分析组合。

[0361] 对于一个比较,通过鉴定配对样品中的所有变体并且减去种系变体而从WES计算出TMB量度。

[0362] 为了不同的比较,使用了MYCHOICE HRD-PLUS。这种测定靶向跨基因组分布的约27,000个SNP。将约100bp的序列读段映射到所述一组SNP区段,其中每个SNP周围有±400个碱基的窗口,并且其中有最多7个错配。

[0363] 将若干误差滤波器应用于映射的序列,以减少突变调用的潜在模糊性:

[0364] 忽略具有多个映射定位的读段;

[0365] 读段末端可能倾向于测序误差,因此忽略了每个读段中的碱基1-10和>86;

[0366] 如果同一插入物的正向(F)读段和反向(R)读段两者被映射,则所述正向读段和反向读段的映射定位必须与50-500bp的插入片段大小相对应;

[0367] F或R读段必须重叠SNP位置;

[0368] 如果F和R读段重叠,则所述读段的调用被组合,并且在这种情况下,SNP调用必须相同;

[0369] 忽略重叠中的具有不同碱基调用的位置(可鉴定的测序误差)。

[0370] TMB值是使用MYCHOICE HRD-PLUS数据以两种方式计算的。首先,具有种系数量的减除。在这种方法中,观察到与每个SNP相邻的400bp序列。在这些序列区域内鉴定变体,并且然后使用配对的样品进行种系减除。

[0371] 在第二个实验中,仅使用来自患有癌症的受试者的第一单个样品和不需要种系减除的本发明的独特算法,计算MYCHOICE HRD-PLUS数据的TMB值。

[0372] 在第二个实验中,仅跨越变体和SNP两者的序列读段包含在计数矩阵的组装中。将SNP的等位基因频率与变体进行比较,以确定所述变体是种系还是体细胞的。不使用种系减除。

[0373] 在此第二实验中,对于所有剩余位置,计算计数矩阵,其中每个元件 $C(X1, X2)$ 是具有非SNP调用 $X1 = (T, C, G \text{ 或 } A)$ 和SNP调用 $X2 = (T, C, G \text{ 或 } A)$ 的映射读段的数量。在这个矩阵中的两个最大计数, $C(X, P) \geq C(Y, Q)$ ,归因于四个位置等位基因条件之一:

[0374] HomHom:  $C(Y, Q) \leq 3$ 仅留下一个显著计数, $C(X, P)$ ,这意味着非SNP和SNP位置两者均是纯合的;

[0375] HetHom:  $X \neq Y$ 且 $P = Q$ ,即,非SNP位置是杂合的并且SNP位置是纯合的;

[0376] HomHet:  $X = Y$ 且 $P \neq Q$ ,即,非SNP位置是纯合的并且SNP位置是杂合的;

[0377] HetHet:  $X \neq Y$ 且 $P \neq Q$ ,即,非SNP和SNP位置两者均是杂合的。

[0378] 使用具有杂合SNP位置的HomHet和HetHet条件来区分来自癌症和非癌细胞的读段。对于这些条件,矩阵的第三最大计数 $C(Z, P)$ 或 $C(Z, Q)$ 可以归因于癌细胞的体细胞突变。

[0379] 当计数显著高于背景测序误差率时,第三最大计数可以用于检测体细胞突变。平均误差率 $E$ 从除了最高三个计数之外的所有其它计数来计算。

[0380] 体细胞突变的Phred样显著性得分(其是具有一个自由度的卡方概率)用公式I来计算:

[0381] 
$$S = (C(Z, P)^2 / (C(Z, P) + C(X, P)) + (C(Z, P) - E)^2 / E) / 2 * 10$$

[0382] 公式I

[0383] TMB水平是具有 $S>30$ 的位置数,所述实例用M碱基中的杂合SNP区域中的位置总数 $\{N(\text{HomHet})+N(\text{HetHet})\}$ 进行归一化,如公式II所示:

$$[0384] \quad \text{TMB} = N(S>30) / (N(\text{HomHet}) + N(\text{HetHet})) * 1000000$$

[0385] 公式II

[0386] 用于计算TMB的中位序列长度对于WES为9.7Mb,对于具有种系减除的MYCHOICE HRD-PLUS为4.6Mb,并且对于本发明的不需要种系减除的独特算法为1.9Mb。

[0387] 比较了用于确定TMB的三种不同方法的结果。比较显示,不需要种系减除的本发明独特算法提供了令人惊讶的准确的TMB值。TMB结果的比较示出于表4中。

[0388] 表4:使用和不使用种系减除获得的TMB水平的比较

	具有种系减除的 WES	具有种系减除的 MYCHOICE HRD-PLUS	没有种系减除的本发明
[0389] 具有种系减除的 WES	-----	1.6** $p = 4.6 \times 10^{-6}$	1.5** $p = 1.2 \times 10^{-5}$
具有种系减除的 MYCHOICE HRD- PLUS	0.895*	-----	0.04** $p = 0.88$
没有种系减除的本发明	0.908*	0.834*	-----

[0390] \*相关系数。

[0391] \*\*变体每Mb的平均差(具有p值)。

[0392] 表4中的相关系数显示,与具有种系减除的基于WES的常规方法以及具有种系减除的MYCHOICE HRD-PLUS相比,使用不需要种系减除的独特算法的本发明方法提供了令人惊讶地准确的TMB值。

[0393] 因此,使用不需要种系减除的独特算法的本发明方法是出乎意料地有利的,因为所述方法不需要种系比较物样品并且可以对含有癌症和非癌细胞的任何样品进行。

[0394] 使用不需要种系减除的独特算法的本发明方法是一种有效的工具,因为对于待评估的每种疾病或群体,可以确定TMB水平的阈值或参考值。



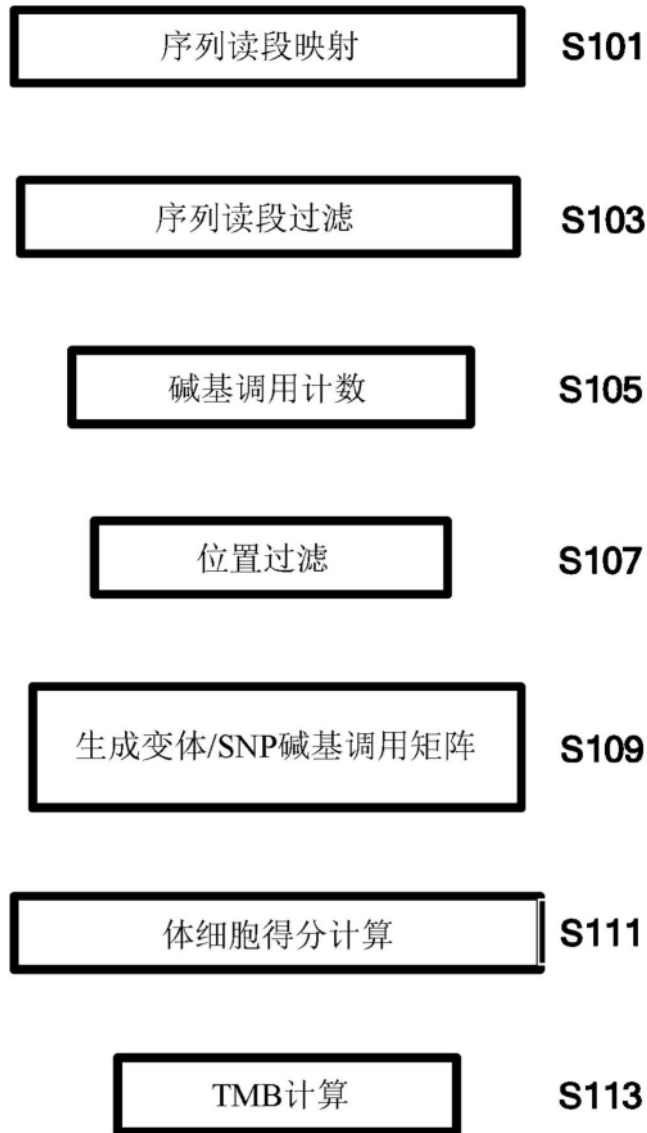


图1

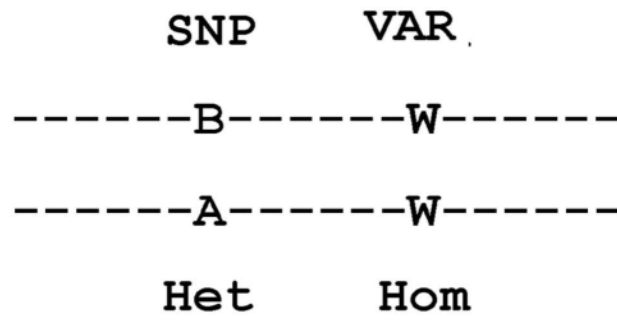
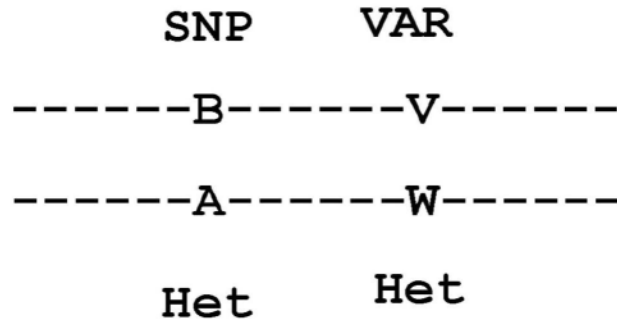


图2

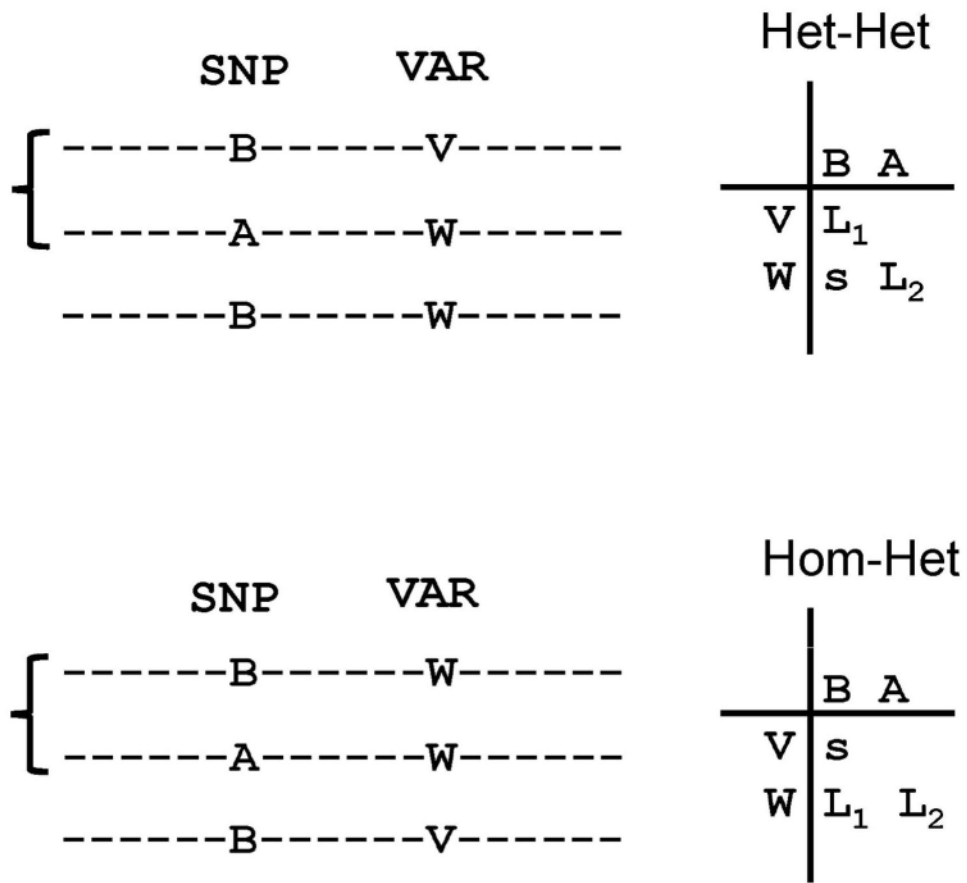


图3

纯合VAR、杂合SNP (Hom/Het)

VAR位置56031

SNP位置56015

	SNP	VAR	
SNP			
WT:	TTTGGTGCTATTCTGAGACCTAATGTGCCTAACCCAGCAA		
VAR			
	TTTGGTGCTATTCTGAGACCTAATGTGCCTAACCCAGCA		A/G
	GGTGCTATTCTAAGACCTAATGTGCCTAACCCAGCAAGC		A/A
	GCTATTCTGAGACCTAATGTGCCTAACCCAGCAAGCGC		A/G
	ATTCTGAGACCTAATGTGCCTAACCCAGCAAGCGCCG		A/G
	TCTAAGACCTAATGTGCCTGACCCAGCAAGCGCCGAG		G/A
	TGAGACCTAATGTGCCTAACCCAGCAAGCGCCGAGAC		A/G
	AAGACCTAATGTGCCTAACCCAGCAAGCGCCGAGACAC		A/A

计数矩阵

SNP	A	T	G	C	
VAR	A	32	0	55	0
	T	0	0	0	0
	G	23	0	1	0
	C	0	0	0	0

第一最大计数 = 55, A/G  
 第二最大计数 = 32, A/A  
 第三最大计数 = 23, G/A

----- (G) -----A---  
 ----- (A) -----A---  
 ----- (A) -----G---

图4

杂合VAR、杂合SNP (Het/Het)

VAR位置246177

SNP位置246231

	VAR	SNP	
SNP			
WT:	CCATGAGCCTGTGTGCCCTGTGA...	ACTCTGTGACCCGTGAGCCTGTGACCC	
VAR			
	CCATGAGCCTGTGTCCCCTGTGA...	ACTCTGTGACCCGTGA	C/G
	ATGAGCCTGTGTGCCCTGTGA...	ACTCTGTGACCCGTTAG	G/T
	TGAGCCTGTGTCCCCTGTGA...	ACTCTGTGACCCGTGAGCC	C/G
	AGCCTGTGTGCCCTGTGA...	ACTCTGTGACCCGTGAGCCTG	G/G
	CCTGTGTGCCCTGTGA...	ACTCTGTGACCCGTGAGCCTGT	G/G
	CTGTGTCCCCTGTGA...	ACTCTGTGACCCGTGAGCCTGTGA	C/G
	TGTGCCCTGTGA...	ACTCTGTGACCCGTTAGCCTGTGACCC	G/T

计数矩阵

SNP	A	T	G	C
VAR	A	0	0	0
	T	0	0	1
	G	0	34	7
	C	0	0	39

第一最大计数 = 39, C/G,  
 第二最大计数 = 34, G/T,  
 第三最大计数 = 7, G/G,

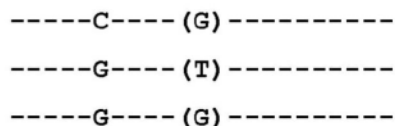
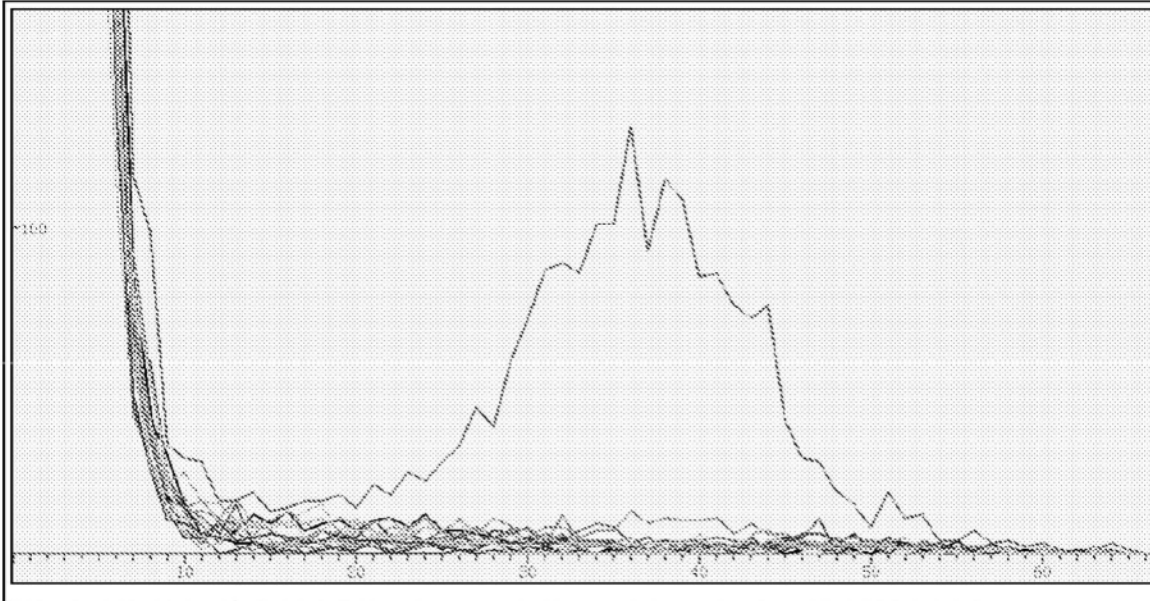


图5

选择运行/泳道: 2814\_AHY2V2BCK2\_1.2 流式细胞搜索: 搜索



样品条	样品Id	位置数/每Mb	Chi(HML)	Chi(非HML)	位置总数	新增:位置总数	新增:位置数/每Mb	
CTCAATGA	306926	1375 / 483.9	24.89	3.57	100.6	4081114	1744566	732 / 419.8
TCCGTCIA	306927	135 / 28.7	61.63	6.87	120.1	4697611	2945495	41 / 20
AGGCTAAC	306928	134 / 30.2	32.79	1.65	111.1	4432873	1831031	33 / 17.5
CCATCCTC	306929	121 / 28.1	21.84	1.55	104.9	4308931	1851782	36 / 19.4
AGATGTAC	306930	172 / 38.9	24.95	1.74	106.5	4425275	1937150	56 / 28.9
TCTTACA	306931	93 / 22.7	21.49	2.25	96.6	4096347	1490012	13 / 8.8
CCGAAGTA	306932	206 / 44.8	32.15	1.62	113.9	4397392	1950120	84 / 32.8
CCCATACA	306933	138 / 31.6	112.36	2.38	100.6	4068431	1728153	48 / 28.4

图6

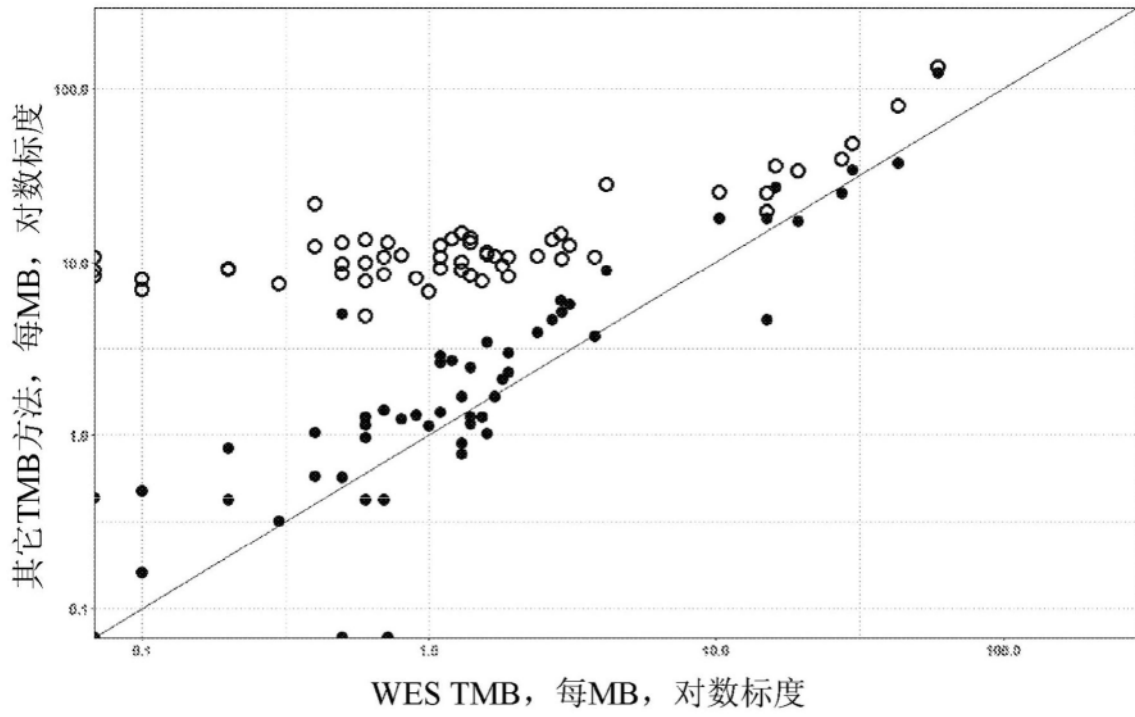


图7