



(12) 发明专利申请

(10) 申请公布号 CN 117120080 A

(43) 申请公布日 2023. 11. 24

(21) 申请号 202280026453.4

陈丰春

(22) 申请日 2022.04.01

(74) 专利代理机构 北京集佳知识产权代理有限公司 11227

(30) 优先权数据

专利代理师 张福誉 韩晓帆

2104715.4 2021.04.01 GB

(85) PCT国际申请进入国家阶段日

(51) Int.Cl.

2023.09.28

A61K 39/00 (2006.01)

(86) PCT国际申请的申请数据

PCT/EP2022/058793 2022.04.01

(87) PCT国际申请的公布数据

W02022/207925 EN 2022.10.06

(71) 申请人 阿基里斯治疗英国有限公司

地址 英国大伦敦

(72) 发明人 安德鲁·罗思

马克西米利安·普林茨·祖·萨尔
姆-霍斯特马尔

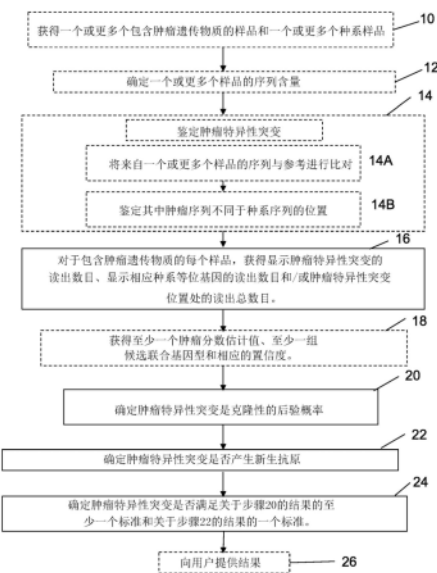
权利要求书3页 说明书38页 附图11页

(54) 发明名称

克隆性新抗原的鉴定及其用途

(57) 摘要

提供了在对象中确定肿瘤特异性突变是否可能是克隆性的方法。所述方法包括提供来自所述对象的包含肿瘤遗传物质的一个或多个样品的序列数据,所述序列数据包含针对所述一个或多个样品中的每一个,以及作为后验概率确定肿瘤特异性突变为克隆性的可能性,该后验概率取决于:所述突变为克隆性的先验概率,以及在所述肿瘤特异性突变是(i)克隆性的或(ii)非克隆性的情况下,鉴于所述一个或多个样品中每一个的肿瘤分数和一个或多个候选联合基因型而观察到该序列数据的概率。还描述了相关的方法、系统和产品。



1. 在对象中确定肿瘤特异性突变是否可能是克隆性的方法,所述方法包括:

向处理器提供来自所述对象的包含肿瘤遗传物质的一个或更多个样品的序列数据,所述序列数据针对所述一个或更多个样品中的每一个包含以下的至少两者:所述样品中显示所述肿瘤特异性突变的读出数目(d_b)、所述样品中显示相应种系等位基因的读出数目和在所述肿瘤特异性突变的位置的读出总数目(d),以及

通过所述处理器作为后验概率确定所述肿瘤特异性突变为克隆性的可能性,该后验概率取决于:

所述突变为克隆性的先验概率,和

在所述肿瘤特异性突变是(i)克隆性的或(ii)非克隆性的情况下,鉴于所述一个或更多个样品中每一个的肿瘤分数和一个或更多个候选联合基因型而观察到所述序列数据的概率,所述一个或更多个候选联合基因型各自包含以下群体在所述肿瘤特异性突变位置的基因型:正常群体、不包含所述肿瘤特异性突变的参考肿瘤群体和包含所述肿瘤特异性突变的变体肿瘤细胞群体。

2. 权利要求1所述的方法,其中鉴于每个样品中的肿瘤分数和一个或更多个候选联合基因型而观察到所述序列数据的概率取决于鉴于肿瘤分数、癌细胞分数和一个或更多个候选联合基因型而观察到所述序列数据的概率($\Pr(d, db | \pi, \phi, t)$),任选地其中鉴于肿瘤分数、癌细胞分数和一个或更多个候选联合基因型而观察到所述序列数据的概率是鉴于肿瘤分数、癌细胞分数和所述一个或更多个候选联合基因型中的每一个而观察到所述序列数据的概率的加权和。

3. 权利要求1或权利要求2所述的方法,其中鉴于每个样品中的肿瘤分数和一个或更多个候选联合基因型而观察到所述序列数据的概率是对每个样品中癌细胞分数的所有可能值进行积分而获得的,其中所述癌细胞分数是包含所述肿瘤特异性突变的肿瘤细胞的比例。

4. 前述权利要求中任一项所述的方法,其中获得来自多个样品的序列数据,并且其中鉴于所述多个样品中每一个的肿瘤分数和一个或更多个候选联合基因型而观察到所述序列数据的概率是作为鉴于相应样品中的肿瘤分数和所述一个或更多个候选联合基因型而观察到每个样品的所述序列数据的概率的乘积而获得的。

5. 前述权利要求中任一项所述的方法,其还包括通过所述处理器针对每个样品获得肿瘤分数的至少一个估计值,和一个或更多个候选联合基因型的至少一个相应组。

6. 前述权利要求中任一项所述的方法,其中观察到所述序列数据的概率合并了鉴于至少一个样品或所述一个或更多个样品的各自的肿瘤分数和一个或更多个候选联合基因型的相应组而观察到所述序列数据的多个概率,任选地其中所述方法包括通过所述处理器针对至少一个样品或所述一个或更多个样品获得肿瘤分数的多个估计值,和一个或更多个候选联合基因型的多个相应组。

7. 前述权利要求中任一项所述的方法,其中所述突变为克隆性的先验概率被设置为中性先验,或被设置为来源于先验数据和/或专业知识的值。

8. 前述权利要求中任一项所述的方法,其还包括针对在所述对象中鉴定的多个肿瘤特异性突变而重复所述方法,并且任选地还包括至少部分基于所确定的所述多个肿瘤特异性突变在该对象中为克隆性的可能性来对所述多个肿瘤特异性突变进行排序或以其他方式

划分优先顺序。

9. 前述权利要求中任一项所述的方法, 其还包括通过所述处理器在所述对象中鉴定一个或更多个肿瘤特异性突变, 所述鉴定任选地使用来自所述对象的包含肿瘤遗传物质的一个或更多个样品和来自所述对象的一个或更多个种系样品的序列数据来进行。

10. 前述权利要求中任一项所述的方法, 其还包括向用户提供所确定的所述肿瘤特异性突变为克隆性的概率和/或者来源于其或与其相关的值, 例如通过用户界面提供。

11. 在对象中鉴定一种或更多种克隆性新抗原的方法, 所述方法包括:

在所述对象中鉴定多个肿瘤特异性突变;

使用前述权利要求中任一项所述的方法在所述对象中确定一个或更多个所述肿瘤特异性突变是否可能是克隆性的; 以及

确定一个或更多个所述肿瘤特异性突变是否可能产生新抗原,

其中, 克隆性新抗原是这样的肿瘤特异性突变, 其满足所述肿瘤特异性突变是否可能是克隆性的一个或更多个预定标准和所述肿瘤特异性突变是否可能产生新抗原的一个或更多个标准。

12. 权利要求11所述的方法, 其中克隆性新抗原是这样的肿瘤特异性突变, 其满足选自以下的至少一个标准: 具有高于预定阈值的为克隆性的概率; 具有高于阈值的为克隆性的概率, 该阈值适应性地为在已确定概率的肿瘤特异性突变中选出克隆性概率最高的预定数目的肿瘤特异性突变; 以及具有高于阈值的为克隆性的概率, 该阈值适应性地为在已确定概率的肿瘤特异性突变中选出预定最高百分位数的肿瘤特异性突变。

13. 权利要求11或权利要求12所述的方法, 其中克隆性新抗原是这样的肿瘤特异性突变, 其满足选自以下的至少一个标准: 与在肿瘤细胞中表达的表达产物相关、被预测为产生不在所述对象的正常细胞中表达的蛋白质或肽、被预测为产生至少一种可能由MHC分子呈递的肽和被预测为产生免疫原性的蛋白质或肽, 其中所述MHC分子优选是已知存在于所述对象中的MHC等位基因。

14. 为已诊断为患有癌症的对象提供预后的方法, 所述方法包括在来自所述对象的一个或更多个样品中鉴定多个肿瘤特异性突变以及使用权利要求1至9中任一项所述的方法来确定所述肿瘤特异性突变中的每一个为克隆性的可能性。

15. 为已诊断为患有癌症的对象提供免疫治疗的方法, 所述方法包括:

使用权利要求11至13中任一项所述的方法鉴定一种或更多种克隆性新抗原; 以及

设计靶向一种或更多种经鉴定的克隆性新抗原的免疫治疗。

16. 权利要求15所述的方法, 其中所述靶向一种或更多种所述克隆性新抗原的免疫治疗是免疫原性组合物, 其是包含免疫细胞或治疗性抗体的组合物。

17. 权利要求16所述的方法, 其中所述免疫治疗是包含T细胞的组合物, 所述T细胞识别所述一种或更多种经鉴定的克隆性新抗原中的至少一种。

18. 权利要求17所述的方法, 其中所述组合物富集靶向所述一种或更多种经鉴定的克隆性新抗原中至少一种的T细胞, 任选地其中所述方法包括获得T细胞群体以及扩增该T细胞群体以提高靶向所述一种或更多种经鉴定的克隆性新抗原中至少一种的T细胞的数量或相对比例。

19. 组合物, 其包含通过根据权利要求18所述的方法获得的或可获得的T细胞群体。

20. 治疗已诊断为患有癌症的对象的方法,所述方法包括施用使用权利要求15至18所述方法提供的免疫治疗,或权利要求19所述的组合物。

21. 系统,其包括:

处理器;和

计算机可读介质,所述计算机可读介质包含这样的指令,当所述指令由所述处理器执行时,导致所述处理器进行权利要求1至18中任一项所述方法的步骤。

22. 一个或多个非暂态计算机可读介质,其包含这样的指令,当所述指令由一个或多个处理器执行时,导致所述一个或多个处理器进行权利要求1至18中任一项所述方法的步骤。

克隆性新抗原的鉴定及其用途

技术领域

[0001] 本公开内容涉及用于确定肿瘤特异性突变是否可能是克隆性的,以及用于鉴定来源于在肿瘤中存在的肿瘤特异性突变的克隆性新抗原的方法。本公开内容还涉及利用或靶向经鉴定的克隆性新抗原的用于治疗癌症的方法和组合物。

背景技术

[0002] 癌细胞已知可获得突变,其中的一些突变可赋予进化优势。作为结果,肿瘤通常包含多个基因型不同的相关群体(或克隆)。表征肿瘤的克隆性组成在治疗背景下是特别重要的。事实上,靶向仅在肿瘤细胞群体的亚群中存在的突变(也称为“亚克隆性”突变)可与有限的临床益处(因为该治疗仅靶向部分群体)和复发或转移的高可能性(因为未受影响的克隆仍然能够增殖)相关。相反,越来越认为靶向克隆性新抗原(作为突变存在的结果表达的抗原,其存在于所有肿瘤细胞中)或组合多种靶向治疗可以是有效控制肿瘤所必需的(McGranahan et al., 2015)。另外,克隆性新抗原负荷已知与至少一些癌症中的预后,以及对用检查点抑制剂治疗的敏感性相关(McGranahan et al., 2016; Litchfield et al., 2021)。

发明内容

[0003] 本发明人已开发了用于确定肿瘤特异性突变是否可能是克隆性的新方法,所述方法使用来自包含肿瘤细胞或来源于其的遗传物质的一个或更多个样品的序列数据,所述方法解决了现有技术方法的一个或更多个问题。该方法在鉴定克隆性新抗原中具有特别用途,例如出于癌症治疗或预后的目的。所述方法使用了严格的统计框架以将个体突变分类为克隆性的,并且提供了反映分配中置信度的概率。所述方法是快速的、灵活的、稳健的和可复制的,依赖于可解释的假设,可灵活并入体细胞拷贝数畸变数据,并且可在提供其预测时考虑由于多重倍性/纯度解导致的拷贝数调用中的不确定性。

[0004] 因此,根据第一方面,提供了在对象中确定肿瘤特异性突变是否可能是克隆性的方法,所述方法包括:提供或获得来自对象的包含肿瘤遗传物质的一个或更多个样品的序列数据,所述序列数据针对一个或更多个样品中的每一个包含以下中的至少两者:所述样品中显示所述肿瘤特异性突变的读出数目(d_b)、所述样品中显示相应种系等位基因的读出数目、和在所述肿瘤特异性突变的位置的读出总数目(d);以及作为后验概率确定所述肿瘤特异性突变为克隆性的可能性,该后验概率取决于:所述突变为克隆性的先验概率,和在所述肿瘤特异性突变是(i)克隆性的或(ii)非克隆性的情况下,鉴于一个或更多个样品中每一个的肿瘤分数和一个或更多个候选联合基因型而观察到所述序列数据的概率,所述一个或更多个候选联合基因型各自包含以下群体在所述肿瘤特异性突变位置的基因型:正常群体、不包含所述肿瘤特异性突变的参考肿瘤群体和包含所述肿瘤特异性突变的变体肿瘤细胞群体。

[0005] 所述方法获得了突变为克隆性的概率($P(Z=1)$)作为后验概率($p(Z=1 | d_b, d, \pi,$

$t, \rho)$), 其取决于突变为克隆性的先验概率 (ρ) 和观察到所述序列数据的概率 (也称为观察到所述序列数据的“可能性”, 或简称为所述序列数据的“可能性”)。因此, 其可针对任何突变简单地使用与正在研究的突变相关的所述序列数据来单独获得, 并产生依赖于明确假设的容易解释的输出值 (即突变为克隆性的先验概率, 以及鉴于群体结构的明确模型而观察到所述序列数据的可能性)。换句话说, 这样的概率取决于可通过由可检查的严格假设组强调的贝叶斯 (Bayesian) 框架从以下二者中获得的数据: 包含肿瘤遗传物质的一个或更多个样品和任何可获得的先验知识。该输出值可用于比较任何数目的突变, 例如将任何数目的突变划分优先顺序, 这不需要覆盖整个基因组。该输出值还具有组合来自包含肿瘤遗传物质的多个样品的证据的能力, 但可使用包含肿瘤遗传物质的单个样品来等同地确定。

[0006] 本方面的方法可具有以下特征中的一个或更多个。

[0007] 所述方法可以是计算机实现的。因此, 获得序列数据的步骤可通过处理器进行, 并且确定肿瘤特异性突变为克隆性的可能性的步骤可由所述处理器进行。获得序列数据的步骤可包括接收包含来自对象的一个或更多个样品的序列读出的序列数据, 以及从所述序列读出中确定以下中的至少两者: 样品中显示肿瘤特异性突变的读出数目 (d_b)、样品中显示相应种系等位基因的读出数目和在肿瘤特异性突变位置的读出总数目 (d)。至少确定肿瘤特异性突变为克隆性的可能性的步骤可以是计算机实现的。确定肿瘤特异性突变为克隆性的可能性的步骤可包括以下步骤: 数值积分, 以获得后验概率。特别地, 该步骤可包括确定突变为克隆性的后验概率, 这鉴于突变为克隆性的先验概率, 和在肿瘤特异性突变是 (i) 克隆性的或 (ii) 非克隆性的情况下, 观察到序列数据的概率, 这是通过求解多个一维积分 (例如, 如针对每个样品的积分对, 分别表示突变为克隆性和非克隆性的假设) 将在 0 和 1 之间的所有可能的癌细胞分数内的所观察的序列数据的概率积分进行的。这些数值积分可针对每个样品和每个突变独立 (例如, 如并行) 求解。提供的步骤可包括一个或更多个步骤, 其中全部或一些步骤是计算机实现的。

[0008] 肿瘤特异性突变为克隆性的概率可取决于突变为克隆性的先验概率 (ρ), 其通过以下: 在鉴于突变为克隆性的先验概率的情况下, 突变被分配至克隆性类别的先验概率 ($P(Z=1|\rho) = \rho$); 以及在突变为克隆性的先验概率的情况下, 突变被分配至非克隆性类别的先验概率 ($P(Z=0|\rho) = (1-\rho)$)。如果肿瘤特异性突变为克隆性的 (鉴于一个或更多个样品中每一个的肿瘤分数和一个或更多个候选联合基因型), 则观察到序列数据的概率可在癌细胞分数内被边缘化。类似地, 如果肿瘤特异性突变不是克隆性的, 鉴于一个或更多个样品中每一个的肿瘤分数和一个或更多个候选联合基因型, 则观察到序列数据的概率可在癌细胞分数内被边缘化。

[0009] 肿瘤特异性突变为克隆性的概率可取决于: 在突变为克隆性的先验概率的情况下, 突变被分配至克隆性类别的先验概率 ($P(Z=1|\rho) = \rho$), 乘以在每个样品中鉴于肿瘤分数和一个或更多个候选联合基因型而观察到序列数据的概率, 如果突变为克隆性的 (其可被计算为 ψ_1 , 每个样品中序列数据的可能性, 在癌细胞分数内被边缘化); 以及在突变是非克隆性的先验概率的情况下, 突变被分配至非克隆性类别的先验概率 ($P(Z=0|\rho) = 1-\rho$) 乘以鉴于肿瘤分数和一个或更多个候选联合基因型在每个样品中观察到序列数据的概率, 如果突变为非克隆性的 (其可被计算为 ψ_0 , 每个样品中序列数据的可能性, 在癌细胞分数内被边缘化)。

[0010] 肿瘤特异性突变为克隆性的概率可作为以下之比获得：(i) 在突变为克隆性的先验概率的情况下，突变被分配至克隆性类别的先验概率乘以在每个样品中鉴于肿瘤分数和一个或多个候选联合基因型而观察到序列数据的概率，如果突变为克隆性的 ($p(d_b, d, Z=1 | \pi, t, \rho)$)，其可被表示为 $\rho\psi_1$)，除以 (ii) 以下的总和：(i) (即 $p(d_b, d, Z=1 | \pi, t, \rho)$) 和在突变是亚克隆性的先验概率的情况下，突变被分配至非克隆性类别的先验概率乘以鉴于在每个样品中的肿瘤分数和一个或多个候选联合基因型在每个样品中观察到序列数据的概率，如果突变为非克隆性的 ($p(d_b, d, Z=0 | \pi, t, \rho)$)，其可被表示为 $(1-\rho)\psi_0$)。

[0011] 突变为克隆性的概率可使用等式 (11a) 来获得。在等式 (11a) 中，项 $\Pr(d_b, d | \pi, \varphi, t)$ 可由等式 (3)、(4)、(3a)、(4a)、(3b) 或 (4b) 中的任一个给出。在等式 (11) 中，项 $p(\varphi | Z=0)$ 和 $p(\varphi | Z=1)$ 可由等式 (6) 给出。

[0012] 克隆性突变可以是存在于来自对象的包含肿瘤遗传物质的一个或多个样品中的所有或基本上所有肿瘤细胞中 (或者一个或多个样品中的所有肿瘤遗传物质中) 的突变。这样的突变可在对象的所有肿瘤细胞中存在或可假设在对象的所有肿瘤细胞中存在 (因为这方面的完全确定性可与对象中所有肿瘤细胞的测序相关，但在一个或多个样品中的基本上所有细胞中的存在可用作这方面的指示)。

[0013] 鉴于每个样品中的肿瘤分数和一个或多个候选联合基因型而观察到序列数据的概率可取决于鉴于肿瘤分数、癌细胞分数和一个或多个候选联合基因型而观察到序列数据的概率 ($\Pr(d, db | \pi, \varphi, t)$)。鉴于肿瘤分数、癌细胞分数和一个或多个候选联合基因型而观察到序列数据的概率可以是鉴于肿瘤分数、癌细胞分数、和一个或多个候选联合基因型中的每一个而观察到序列数据的概率的加权和。

[0014] 有利地，观察到序列数据的概率 (序列数据的可能性) 可在多个候选基因型上计算 (例如，作为包含针对每个候选基因型的项的概率总和，参见例如等式 (3a)、(3b))，其贡献可以被加权，例如以反映关于候选基因型的相对概率的先验知识 (例如一些基因型是否比另一些基因型更可能发生的任何先验知识)。当这样的先验知识不是可用的或期望的时，每个候选基因型的概率可被同等地加权。考虑的各个候选基因型的权重合适地相加为 1，使得总概率反映所考虑的不同候选联合基因型的相对贡献。当使用单个候选联合基因型时，可将其权重分配为 1 (即不可获得总和)。

[0015] 鉴于肿瘤分数、癌细胞分数和特定候选联合基因型 (G_i) 而观察到序列数据的概率 (其可以被计算为 ψ_z ，每个样品中序列数据的可能性，在癌细胞分数内边缘化) 可使用具有参数 d_b 和 $\xi(G_i, \varphi, t)$ 的二项分布获得。或者，鉴于肿瘤分数、癌细胞分数和特定候选联合基因型而观察到序列数据的概率可使用具有参数 d_b 、 $\xi(G_i, \varphi, t)$ 、和 γ 的 β 二项分布获得。在两种情况下 (即无论使用二项分布还是 β 二项分布)，假设特定基因型 G_i 、癌细胞分数 φ 和肿瘤纯度 t ， $\xi(G_i, \varphi, t)$ 可表示用变体等位基因对读出进行取样的概率。概率 $\xi(G_i, \varphi, t)$ 可作为以下的函数获得：针对正常基因型、变体基因型和参考基因型中的每种的总拷贝数，用来自具有基因型 G_i 的群体中的变体对读出进行取样的概率，其鉴于以下：基因型中的基因座处是变体的等位基因的比例和测序错误率，样品中的肿瘤分数，和突变的癌细胞分数。

[0016] 鉴于每个样品中的肿瘤分数和一个或多个候选联合基因型而观察到序列数据的概率可对每个样品中癌细胞分数的所有可能值进行积分来获得，其中癌细胞分数是包含肿瘤特异性突变的肿瘤细胞的比例。因此，确定肿瘤特异性突变为克隆性的可能性的步骤

可包括使用处理器以将所述积分进行数值积分。

[0017] 癌细胞分数(ϕ)可取0和1之间的值。换句话说,在突变为克隆性的或非克隆性的情况下,鉴于每个样品中的肿瘤分数和一个或多个候选联合基因型而观察到序列数据的概率可通过在癌细胞分数的所有可能值上将取决于癌细胞分数的值进行积分(即,在癌细胞分数内边缘化)来获得。取决于癌细胞分数的值可表示为 $\Pr(d_b, d | \pi, \phi, t) = p(\phi | Z = Z)$,其中第一项是鉴于肿瘤分数、癌细胞分数和一个或多个候选联合基因型而观察到序列数据的概率,并且在突变被分类为克隆性的或非克隆性的(分别为 $Z=1$ 或 $Z=0$)的情况下,第二项是癌细胞分数的先验概率(即基于癌细胞分数对克隆性/非克隆性突变应如何表现的假设的概率)。因此,鉴于每个样品中的肿瘤分数和一个或多个候选联合基因型而观察到序列数据的概率可作为 $\int_0^1 \Pr(d_b, d | \pi, \phi, t) p(\phi | Z = z) d\phi$ 获得。

[0018] 如果突变被分类为克隆性的,则特定癌细胞分数的先验概率可定义为具有参数 α (设置为 >1 的值,例如99,但也可使用任何其他值)和 $\beta=1$ 的 β 分布($\beta(\phi | \alpha, 1)$)。如果突变被分类为非克隆性的,则特定癌细胞分数的先验概率可定义为具有参数 $\alpha=1$ 和 $\beta=1$ 的 β 分布($\beta(\phi | 1, 1)$)。

[0019] 可获得来自多个样品的序列数据并且鉴于多个样品中每一个的肿瘤分数和一个或多个候选联合基因型而观察到序列数据的概率可作为鉴于相应样品中的肿瘤分数和所述一个或多个候选联合基因型而观察到序列数据的概率的乘积而获得。

[0020] 有利地,本方法能够无缝集成获自多个样品的支持/反对(for/against)突变克隆性的证据,如果这些证据是可用的。这是特别有利的,因为可利用多区域测序数据的用于推断肿瘤细胞群体的克隆性结构的方法已表明在基准研究中表现得特别好(Farahani et al., 2017)。

[0021] 所述方法还可包括针对每个样品获得或提供肿瘤分数的至少一个估计值,以及一个或多个候选联合基因型的至少一个相应组。肿瘤分数估计值可使用用于确定包含肿瘤和正常细胞的混合物的样品中的等位基因特异性拷贝数谱的方法来获得。使用测序或阵列数据来进行其的方法在本领域中是已知的,例如通过将等位基因特异性数据表达为参数(包括等位基因特异性拷贝数、肿瘤非整倍体和肿瘤细胞分数)的函数,并确定最佳拟合所有数据的这些参数的值来进行。这样的方法的实例包括例如ASCAT(Van Loo et al., 2010)等。或者,可实验性地确定肿瘤分数估计值。因此,所述方法还可包括获得针对一个或多个样品中每一个的肿瘤分数估计值。特别地,所述方法可包括通过处理器获得针对每个样品的肿瘤分数的至少一个估计值,其包括使用序列数据确定等位基因特异性拷贝数和肿瘤分数的估计值的处理器,以及通过所述处理器确定的与所述等位基因特异性拷贝数相关的一个或多个候选联合基因型的组。

[0022] 一个或多个候选基因型的组可使用混合样品中针对肿瘤细胞的等位基因特异性拷贝数或来源于其的变量(或者相反地,这样的等位基因特异性拷贝数可来源于的变量,例如B等位基因分数和 $\log R$)来获得。混合样品中肿瘤细胞的等位基因特异性拷贝数可使用用于确定包含肿瘤和正常细胞的混合物的样品中等位基因特异性拷贝数谱的方法例如,如ASCAT(Van Loo et al., 2010)或ascatNgs(Raine et al., 2016)等来获得。

[0023] 因此,所述方法还可包括针对一个或多个样品的每个获得以下中至少两者的估计值:样品中肿瘤细胞中主要等位基因的拷贝数、样品中肿瘤细胞中次要等位基因的拷贝

数和样品中肿瘤细胞中肿瘤特异性突变位置处的总拷贝数。样品中肿瘤细胞中拷贝数的估计值可表示样品中在肿瘤细胞的整个群体中汇总的(例如平均值)估计值。

[0024] 一个或更多个候选联合基因型的组可作为与以下假设相容的候选联合基因型获得:仅包含正常等位基因A的正常群体(即 $G_H = AA$ 或A,例如,如果基因座在性染色体上);不包含变体等位基因B的参考群体(即 $G_R = (A) * n$);以及包含至少一个拷贝的变体等位基因B的变体群体(即 $G_V = (A) * m(B) * |$)。

[0025] 有利地,候选基因型的组可包括还与以下假设之一相容的候选联合基因型:(i) 参考群体基因型与正常群体基因型匹配,并且变体群体的拷贝数等于该位置的总拷贝数并且最高至变体等位基因的主要拷贝数(major copy number);或(ii) 参考群体的拷贝数等于该位置的总拷贝数,并且变体群体具有1个变体等位基因并且变体群体的拷贝数等于该位置总拷贝数(“主要拷贝数优先”)。该方法有利地在考虑群体的基因型中的不确定性而不考虑太多情况之间取得了良好的平衡。

[0026] 相反或除此之外,一个或更多个候选联合基因型的组可包括与以下假设相容的任何候选联合基因型:每个突变都是二倍体和杂合的(即 $G_V = AB, G_R = AA$) (“AB优先”)。相反或除此之外,一个或更多个候选联合基因型的组可包括与以下假设相容的任何候选联合基因型:每个突变都是二倍体和纯合的(即 $G_V = BB, G_R = AA$) (“BB优先”)。相反或除此之外,一个或更多个候选联合基因型的组可包括与以下假设相容的任何候选联合基因型:变体群体的基因型在突变区域具有预测的总拷贝数,仅具有一个突变体等位基因(即 $G_V = (A) * mB$,其中 $m = \text{总拷贝数} - 1$) (“无接合性优先(no zygosity prior)”)。相反或除此之外,一个或更多个候选联合基因型的组可包括与以下假设相容的任何候选联合基因型:变体群体的基因型在突变区域具有预测的总拷贝数,具有至少一个突变体等位基因,并且参考群体是AA或基因型的拷贝数等于预测的总拷贝数并且没有变体等位基因(即 $G_R = (A) * n$,其中n是总拷贝数, $G_V = (A) * m(B) * 1$,其中 $m + I = n$ 且 $I > 1$) (“总拷贝数优先”)。相反或除此之外,一个或更多个候选联合基因型的组可包括与以下假设相容的任何候选联合基因型:变体群体的基因型之突变体等位基因的数目对应于主要拷贝数或次要拷贝数(“亲本模式”)。

[0027] 观察到序列数据的概率可与鉴于一个或更多个样品中至少一个样品的各自的肿瘤分数和一个或更多个候选联合基因型的相应组而观察到序列数据的多个概率合并,任选地,其中所述方法包括针对至少一个或者一个或更多个样品获得肿瘤分数的多个估计值和一个或更多个候选联合基因型的多个相应组。因此,所述方法可包括针对至少一个样品获得肿瘤分数的多个估计值。这可包括通过处理器确定肿瘤分数和与序列数据相容的相应多个等位基因特异性拷贝数的多个估计值,以及通过处理器确定与所述多个等位基因特异性拷贝数相关的一个或更多个候选联合基因型的多个组。

[0028] 本方法有利地能够确定突变为克隆性的概率,这考虑了多个可能的肿瘤分数和候选联合基因型的相应组。换句话说,本方法能够获得突变为克隆性的概率,这在可获得肿瘤分数和候选联合基因型的多个拷贝数解内进行积分。相比之下,现有技术方法通常依赖于以下的单一估计值:肿瘤纯度和等位基因特异性拷贝数(可从中获得候选联合基因型),其通常是专家定义的最佳标准人工选择的。选择被认为是最佳的拷贝数解的步骤是高度易错的,并且依赖于单一解的方法的输出值可能根据解显著变化。

[0029] 因此,有利地,观察到序列数据的概率(序列数据的可能性)可在多组候选基因型

和相应的肿瘤分数估计值中计算(例如,作为包含针对每个拷贝数解的项的概率的总和,参见等式(3b)、(4b)),其贡献可被加权,例如以反映获得肿瘤分数估计值和候选基因型的组之拷贝数解中的置信度。考虑的拷贝数解的贡献的权重合适地相加为1,使得总概率反映考虑的不同拷贝数解的相对贡献。当使用单一拷贝数解时,可将其权重分配为1(即不可获得总和)。

[0030] 突变为克隆性的先验概率可被设置为中性先验,或者被设置为来源于先验数据和/或专业知识的值。有利地,本文中所述的方法能够在用于确定突变为克隆性的概率的贝叶斯框架中考虑关于突变的先验知识,如果这样的知识是可用的。然而,即使在不存在这样的知识的情况下,其也能够提供突变为克隆性的可能性。

[0031] 针对突变为克隆性的先验概率的值可取决于对象、肿瘤、突变或这些的组合。例如,值可使用先前获得的关于相关组群的患者(例如,如患有相同类型癌症或癌症亚型的患者)的数据来确定。例如,在这样的组群中的克隆性vs.亚克隆性的突变的比例的知识可用于设置本文中使用的先验概率。或者,值可基于关于癌症类型或突变的先验知识来任意设置。例如,已在多个癌症样品中发现并且已在这些样品中被鉴定为通常是克隆性的特定突变可被分配高于0.5的概率。

[0032] 所述方法还可包括重复用于在对象中鉴定多个肿瘤特异性突变的方法。所述方法还可包括至少部分基于所确定的多个肿瘤特异性突变在对象中为克隆性的可能性来对所述多个肿瘤特异性突变进行排序或以其他方式划分优先顺序。

[0033] 所述方法还可包括在对象中鉴定一个或更多个肿瘤特异性突变。在对象中鉴定一个或更多个肿瘤特异性突变可使用来自对象的包含肿瘤遗传物质的一个或更多个样品的序列数据和来自对象的一个或更多个种系样品的序列数据例如通过比较所述序列数据来进行。在对象中鉴定一个或更多个肿瘤特异性突变可包括将来自至少一个包含肿瘤遗传物质的样品的序列数据与参考序列比对以及鉴定样品的序列不同于参考序列的位置。所述方法还可包括将来自至少一个种系的样品的序列数据与参考序列比对以及鉴定包含肿瘤遗传物质的样品的序列不同于种系样品的位置。

[0034] 提供来自对象的一个或更多个样品的序列数据的步骤可包括从用户(例如通过用户界面)、从一个或更多个计算装置或者从一个或更多个数据存储器或数据库接收序列数据,或由从用户(例如通过用户界面)、从一个或更多个计算装置或者从一个或更多个数据存储器或数据库接收序列数据组成。

[0035] 提供序列数据的步骤还可包括对来自对象的包含肿瘤遗传物质的一个或更多个样品进行测序(或以其他方式确定样品中存在的基因组物质的序列组成)。

[0036] 所述方法还可包括对来自对象的一个或更多个种系样品进行测序(或以其他方式确定样品中存在的基因组物质的序列组成)。

[0037] 所述方法还可包括从对象获得一个或更多个包含肿瘤遗传物质的样品以及任选的一个或更多个种系样品。

[0038] 所述方法还可包括向用户提供所确定的肿瘤特异性突变为克隆性的概率和/或者来源于其或与其相关的值,例如通过用户界面提供。例如,所述方法可包括基于所确定的肿瘤特异性突变为克隆性的概率来提供“克隆性状态”标记(flag)或值。作为另一个实例,所述方法可包括提供鉴定突变的信息(例如,如突变的序列和其基因组位置)。

[0039] 根据另一个方面,提供了在对象中鉴定一种或更多种克隆性新抗原的方法,所述方法包括:在对象中鉴定多个肿瘤特异性突变;使用前述方面的任一实施方案的方法在对象中确定一个或更多个所述肿瘤特异性突变是否可能是克隆性的;以及确定一个或更多个所述肿瘤特异性突变是否可能产生新抗原,其中克隆性新抗原是这样的肿瘤特异性突变,其满足所述肿瘤特异性突变是否可能是克隆性的一个或更多个预定标准和所述肿瘤特异性突变是否可能产生新抗原的一个或更多个标准。根据本方面还描述了在对象中鉴定一种或更多种克隆性新抗原的方法,所述方法包括:通过处理器使用来自所述对象的一个或更多个样品的序列数据在对象中鉴定多个肿瘤特异性突变;通过处理器使用前述权利要求中任一项所述的方法在对象中确定一个或更多个肿瘤特异性突变是否可能是克隆性的;以及通过所述处理器选择一个或更多个肿瘤特异性突变作为候选克隆性新抗原,其中候选克隆性新抗原是这样的肿瘤特异性突变,其满足肿瘤特异性突变是否可能是克隆性的至少一个或更多个预定标准和任选地满足肿瘤特异性突变是否可能产生新抗原的一个或更多个标准。

[0040] 本方面的方法可具有以下特征中的任意一个或更多个。

[0041] 克隆性新抗原可以是这样的肿瘤特异性突变,其满足选自以下的至少一个标准:具有高于预定阈值的为克隆性的概率;具有高于阈值的为克隆性的概率,该阈值适应性地设置为在已确定概率的肿瘤特异性突变中选出克隆性概率最高的预定数目的肿瘤特异性突变;以及具有高于阈值的为克隆性的概率,所述阈值适应性地设置为在已确定概率的肿瘤特异性突变中选出预定最高百分位数的肿瘤特异性突变。因此,关于肿瘤特异性突变是否可能是克隆性的一个或更多个预定标准可选自:具有高于预定阈值的为克隆性的可能性的突变、具有高于适应性地设置为在可能性被确定的肿瘤特异性突变中选出预定数目的克隆性可能性最高的突变和具有适应性地设置为在可能性被确定的肿瘤特异性突变中选出预定最高百分位数的肿瘤特异性突变的阈值的为克隆性的可能性。

[0042] 克隆性新抗原可以是这样的肿瘤特异性突变,其满足选自以下的至少一个标准:与在肿瘤细胞中表达的表达产物相关、被预测为产生不在所述对象的正常细胞中表达的蛋白质或肽、被预测为产生至少一种可能由MHC分子呈递的肽、被预测为产生至少一种可能由已知存在于对象中的MHC等位基因呈递的肽、和被预测为产生免疫原性的蛋白质或肽。例如,克隆性新抗原可以是肿瘤特异性突变,其满足以下标准:其被预测为导致蛋白质序列变化(例如,因为其是编码的、因为其影响剪接位点、因为其导致截短的肽等),因此产生了不可在对象的正常细胞中表达的蛋白质或肽。是否是这种情况可例如通过与对象的预测的正常蛋白质组比较来进一步证实。因此,关于肿瘤特异性突变是否可能产生新抗原的一个或更多个标准可选自:突变与在肿瘤细胞中表达的表达产物相关、突变被预测为产生不在对象的正常细胞中表达的蛋白质或肽、突变被预测为产生至少一种可能由MHC分子呈递的肽、突变被预测为产生至少一种可能由已知存在于对象中的MHC等位基因呈递的肽以及突变被预测为产生免疫原性的蛋白质或肽。

[0043] 所述方法还可包括鉴定与一种或更多种克隆性新抗原相关的一种或更多种肽(即由于存在肿瘤特异性突变而被预测为存在于肿瘤细胞中的一种或者更多种肽序列,其中肿瘤特异性突变满足如上所述的一个或更多个标准(与克隆性的可能性和产生克隆性新抗原的可能性有关))。

[0044] 如技术人员理解的,本文中所述操作的复杂性(至少由于获得需要如本文所述的数值积分的后验概率的复杂性,和通常通过对基因组DNA进行测序产生的数据量)使得其超出心理活动的范围。因此,除非上下文另外指明(例如在描述了样品制备或采集步骤的情况下),否则本文中所述方法的所有步骤都是计算机实现的。

[0045] 根据另一个方面,提供了为已诊断为患有癌症的对象提供预后的方法,所述方法包括在来自对象的一个或更多个样品中鉴定多个肿瘤特异性突变以及使用第一方面的任何实施方案的方法来确定每个肿瘤特异性突变为克隆性的可能性。

[0046] 所述方法还可包括将对象分类为具有高克隆性新抗原负荷与低克隆性新抗原负荷,其至少部分取决于具有高于预定阈值的为克隆性的概率的肿瘤特异性突变的比例,其中与具有低克隆性新抗原负荷的对象相比,具有高克隆性新抗原负荷的对象具有改善的预后。

[0047] 根据另一个方面,提供了为已诊断为患有癌症的对象提供免疫治疗的方法,所述方法包括:使用如本文中所述的方法(例如根据第二方面的任何实施方案的方法)来鉴定一种或更多种克隆性新抗原;以及设计靶向一种或更多种经鉴定的克隆性新抗原的免疫治疗。

[0048] 所述方法可具有以下特征中的任意一个或更多个。

[0049] 靶向一种或更多种克隆性新抗原的免疫治疗可以是免疫原性组合物,一种包含免疫细胞或治疗性抗体的组合物。免疫原性组合物可包含经鉴定的克隆性新抗原(例如,如新抗原肽或蛋白质或者显示出新抗原的细胞)的一个或更多个克隆,或足以表达一种或者更多种经鉴定的克隆性新抗原的物质(例如,编码新抗原的DNA或RNA分子)。包含免疫细胞的组合物可包含T细胞、B细胞和/或树突细胞。包含治疗性抗体的组合物可包含一种或更多种这样的抗体,所述抗体识别一种或更多种经鉴定的克隆性新抗原中的至少一种。抗体可以是单克隆抗体。

[0050] 在任何方面的任何实施方案中,癌症可选自膀胱癌、胃癌、食管癌、乳腺癌、结直肠癌、宫颈癌、卵巢癌、子宫内膜癌、肾癌(肾细胞癌)、肺癌(小细胞癌、非小细胞癌和间皮瘤)、脑癌(胶质瘤、星形细胞瘤、胶质母细胞瘤)、黑素瘤、淋巴瘤、小肠癌(十二指肠癌和空肠癌)、白血病、胰腺癌、肝胆肿瘤、生殖细胞癌、前列腺癌、头颈部癌、甲状腺癌和肉瘤。癌症可以是肺癌。癌症可以是黑素瘤。癌症可以是膀胱癌。癌症可以是头颈部癌。

[0051] 在任何方面的任何实施方案中,对象可以是人。

[0052] 设计靶向一种或更多种经鉴定的克隆性新抗原的免疫治疗可包括针对一种或更多种所靶向的克隆性新抗原中的每种来设计一种或更多种候选肽,每种肽包含所靶向的克隆性新抗原的至少一部分。

[0053] 所述方法还可包括获得一种或更多种候选肽。所述方法还可包括针对一种或更多种特性来测试一种或更多种候选肽。测试可在体外或在计算机上进行。例如,可针对免疫原性、由MHC分子显示的倾向性(任选地由特定的MHC分子等位基因显示,其中等位基因可根据由对象表达的MHC等位基因选择)、引起免疫细胞群体增殖的能力等来测试一种或更多种肽。

[0054] 所述方法还可包括产生免疫治疗。所述方法还可包括获得已经用一种或更多种候选肽脉冲的树突细胞群体。免疫治疗可以是这样的组合物,所述组合物包含识别一种或更

多种经鉴定的克隆性新抗原中至少一种的T细胞。所述组合物可富集靶向一种或更多种经鉴定的克隆性新抗原中至少一种的T细胞。所述方法可包括获得T细胞群体并且扩增所述T细胞群体,以提高靶向一种或更多种经鉴定的克隆性新抗原中至少一种的T细胞的数量或相对比例。

[0055] 所述方法还可包括获得T细胞群体。T细胞群体可从对象中分离,例如从获自对象的一个或更多个肿瘤样品、或从对象的外周血样品、或从对象的来自其他组织的样品中分离。T细胞群体可包括肿瘤浸润性淋巴细胞。T细胞可使用本领域公知的方法分离。例如,可从基于CD3、CD4或CD8的表达的样品产生的单细胞悬液中纯化T细胞。T细胞可由通过Ficoll-paque梯度从样品中富集。

[0056] 所述方法还可包括扩增T细胞群体。例如,T细胞可通过在已知为T细胞提供有丝分裂刺激的条件下进行离体培养来扩增。举例来说,T细胞可用细胞因子(例如IL-2)或用有丝分裂抗体(例如抗CD3和/或CD28)进行培养。T细胞可与抗原呈递细胞(antigen-presenting cell,APC)(其可以是经辐照的)共培养。APC可以是树突细胞或B细胞。树突细胞可用包含一种或更多种经鉴定的新抗原的肽作为单一刺激剂或作为刺激新抗原肽的合并物进行脉冲处理。T细胞的扩增可使用本领域已知的方法,包括例如使用人工抗原呈递细胞(artificial antigen presenting cell,aAPC)(其提供另外的共刺激信号)和呈递合适的肽的自体PBMC来进行。自体PBMC可用如本文中所讨论的包含新抗原的肽作为单一刺激剂或者作为替代地作为刺激新抗原的合并物进行脉冲。

[0057] 根据另一个方面,提供了用于扩增T细胞群体以用于在对象中治疗癌症的方法,所述方法包括:使用如本文所述的方法,例如根据第二方面的任何实施方案的方法,来鉴定一种或更多种克隆性新抗原;获得T细胞群体,所述T细胞群体包含能够特异性识别经鉴定的克隆性新抗原中的一种的T细胞;以及将T细胞群体与包含经鉴定的克隆性新抗原的组合物共培养。

[0058] 所述方法可具有以下特征中的一个或更多个。

[0059] 所获得的T细胞群体可被假设为包含能够特异性识别经鉴定的克隆性新抗原中的一种的T细胞。所述方法优选地包括鉴定多种克隆性新抗原。T细胞群体可包含多种T细胞,每种T细胞能够特异性识别多种经鉴定的克隆性新抗原中的一种,并将T细胞群体与包含多种经鉴定的克隆性新抗原的组合物进行共培养。共培养可导致特异性识别一种或更多种新抗原的T细胞群体的扩增。扩增可通过将T细胞与新抗原和抗原呈递细胞共培养进行。抗原呈递细胞可以是树突细胞。因此,扩增可以是对新抗原具有特异性的T细胞的选择性扩增。扩增还可包括一个或更多个非选择性扩增步骤。

[0060] 根据另一个方面,提供了这样的组合物,所述组合物包含通过根据前述方面的任何实施方案的方法获得的或可获得的T细胞群体。

[0061] 根据另一个方面,提供了这样的组合物,所述组合物包含新抗原、新抗原特异性免疫细胞、或识别新抗原的抗体,其用于在对象中治疗或预防癌症,其中所述新抗原已使用本文所述的方法鉴定为克隆性新抗原。

[0062] 根据另一个方面,提供了这样的组合物,所述组合物包含新抗原、新抗原特异性免疫细胞、或识别新抗原的抗体,其中所述新抗原已使用本文所述的方法鉴定为克隆性新抗原。

[0063] 根据另一个方面,提供了这样的细胞或细胞群体,所述细胞或细胞群体在其表面表达新抗原,其中所述新抗原已使用本文中所述的方法确定为克隆性新抗原。

[0064] 根据另一个方面,提供了新抗原、识别新抗原的免疫细胞或识别新抗原的抗体,其用于在对象中治疗或预防癌症,其中所述新抗原已使用本文中所述的方法鉴定为克隆性新抗原。

[0065] 根据另一个方面,提供了新抗原、识别新抗原的免疫细胞或识别新抗原的抗体在制备用于在对象中治疗或预防癌症的药物中的用途,其中所述新抗原已使用本文中所述的方法鉴定为克隆性新抗原。

[0066] 根据另一个方面,提供了治疗已诊断为患有癌症的患者的方法,所述方法包括施用使用本文中所述的方法提供的免疫治疗或如本文中所述的组合物。

[0067] 根据另一个方面,提供了系统,其包括:处理器;和包含这样的指令的计算机可读介质,当所述指令由处理器执行时,导致处理器进行本文中所述的任意方法例如根据上述第一、第二、第三或第四方面的任何实施方案的方法的步骤。

[0068] 根据另一个方面,提供了一个或多个非暂态计算机可读介质,其包含这样的指令,当所述指令由一个或多个处理器执行时,导致一个或多个处理器进行本文中所述的任意方法例如根据上述第一、第二、第三或第四方面的任何实施方案的方法的步骤。

[0069] 根据另一个方面,提供了这样的计算机程序,所述计算机程序包含这样的代码,当所述代码在计算机上执行时,导致计算机进行本文中所述的任何方法例如根据上述第一、第二、第三或第四方面的任何实施方案的方法的步骤。

附图说明

[0070] 图1是示意性地示出了确定肿瘤特异性突变是否可能是克隆性的方法及其在鉴定克隆性新抗原中的用途的流程图。

[0071] 图2是示意性地示出了提供免疫治疗的方法的流程图。

[0072] 图3示出了用于确定肿瘤特异性突变是否可能是克隆性的和/或用于鉴定克隆性新抗原和/或用于提供免疫治疗的系统的一个实施方案。

[0073] 图4示意性地示出了由本文中公开的方法所假设构造的群体。

[0074] 图5示出了使用合成数据来鉴定本文中所述克隆性突变的方法的结果。每幅图示出了针对不同最大拷贝数和样品数的ROC曲线(接收者操作特征,真阳性率($TPR = TP / (TP + FN)$),其中 TP = 真阳性数, FN = 真阴性数;灵敏度)作为假阳性率($FPR = FP / (FP + TN)$),其中 FP = 假阳性数, TN = 真阴性数; $1 - \text{特异性}$)的函数,其中3条独立的曲线各自示出了不同的模拟读出深度。A.用于模拟和鉴定的二项式散发模型(binomial emission model)。B.用于模拟的 β 二项式散发模型,用于鉴定的二项式模型。C.用于模拟和鉴定的 β 二项式散发模型。

[0075] 图6示出了鉴定本文所述的克隆性突变的方法的结果,所述方法使用了来自二倍体细胞系(A)和非整倍体细胞系(B)的细胞系混合实验的数据。每幅图的标题表示哪个样品被包含在分析中的何处。每幅图示出了使用如本文中所述方法的ROC曲线(标记为“ACE”),以及示出在相同数据上运行PyClone(Roth et al., 2014)时获得的TPR和FPR的单个点。

[0076] 图7示出了鉴定本文中所述的克隆性突变的方法的结果,所述方法使用了来自

TRACERx项目的数据。A. 分析的数据概述。针对每个样品,示出了使用TRACERx流水线鉴定的克隆性和亚克隆性SNV的数目(人工整理的PyClone结果)。B. 由如本文中所述的方法预测的普遍概率的分布。(左图)所有ASCAT (Van Loo et al., 2010) 倍性/纯度解考虑为按概率加权的结果。(右图) 当选择ASCAT中最可能的解时的结果。C. ROC曲线示出了如本文中所述方法的预测性能(考虑或没有考虑按概率加权的来自ASCAT的所有可能解),该曲线使用人工整理的PyClone (Roth et al., 2014) 调用作为基本事实。

[0077] 图8示出了如本文中所述的方法与比较方法之间的运行时间比较的结果。比较方法(PyClone VI)已针对计算效率进行了高度优化,但不可并行。本文所述的方法(ACE)还没有成为广泛的计算效率优化的对象,但通过设计(显示为用2个、3个或4个CPU实现)变得更有效且可并行。

具体实施方式

[0078] 在癌症患者中鉴定克隆性突变和新抗原的方法具有明确的临床应用,这是由于考虑到其在癌症治疗和预后中的重要性。已提出了多种方法来尝试重建肿瘤样品的克隆性结构(Schwartz and Schaeffer, 2017)。然而,所有这些方法都具有多个限制其临床效用的缺点。因此,发明人已确定,仍然需要改善的方法用于确定肿瘤特异性突变是否可能是克隆性的。在癌症中鉴定克隆性突变(存在于所有癌细胞中的突变)的问题远非容易解决的。事实上,原代样品包含细胞的混合物(包括正常细胞和癌细胞),并且癌症基因组通常包含广泛和异质性的拷贝数变化。因此,突变的等位基因发生率是多种因素组合的结果,所述因素包括正常细胞的比例、具有突变的肿瘤细胞的比例以及这些肿瘤细胞中每一个的突变的等位基因拷贝数。与任何实验方案相关的技术噪声的来源(例如测序错误)、那些方案的限制(例如短读出比对不确定性)以及初始样品(primary sample)表示部分快照的癌症状态的固有限制进一步加重了这一点。Roth et al. (2014) 试图通过提出一种贝叶斯聚类方法(称为“PyClone”)来应对这一挑战,所述方法将深度测序的体细胞突变的组分组合为共有相同进化历史的簇。这是通过估计其细胞患病率和说明由区段拷贝数变化和正常细胞污染引入的等位基因失衡来完成的。

[0079] 虽然该方法有利地通过清晰的统计框架得到强调,但本发明人确定了由于多个原因其不太适合在非研究环境中使用。首先,该方法相对缓慢,并且在突变组(例如深度测序突变的大量组,或者最近的完整基因组)而不是单个突变的水平下进行操作。这两个缺点均是这样的事实的结果:该方法推断整个突变组上的聚类结构。因此,该方法解决了需要计算模型参数的后验密度的复杂问题,所述模型参数包括在来自癌症的多个样品中鉴定的所有突变组中每一个突变的细胞患病率,以及在该突变组上的聚类结构。本发明人已确定这是不期望的,这是由于至少两个原因:如果在整个突变组上的聚类结构不是目的的,则该方法不必要地缓慢,并且该方法阻止了通过独立于同时分析的另一一些突变的临床流水线的任何突变的可追溯性。事实上,任何可能是克隆性的突变的鉴定均与一起分析的另一一些突变具有内在联系。因此,待分析的突变组中的任何变化需要重新运行该流水线,并可能产生不同的结果。相比之下,由本发明人提出的方法独立地分析每个突变,并且不推断依赖于突变群体的任何结构。这为计算机技术提供了明显的改进,因为由本发明人开发的算法的计算复杂度比常规技术(PyClone)的计算复杂度更低,因此总体上和对于每个突变所需的计算资

源(处理器、存储器和/或网络资源)较低。此外,对突变组的处理是完全可并行的,因为突变之间没有依赖性。这导致复杂性进一步降低、处理速度提高并且提高了易处理性,因为对于单个突变在运行中的任何问题不影响另一些突变的运行。

[0080] 另外,现有技术方法需要专家手动干预以通过分析聚类结构并选择假设为克隆性的突变簇,以及对所选择的簇内的突变优先排序,来将输出值转化为克隆性的指示。最后,该方法还需要专家手动干预以选择合适的拷贝数估计值,因为错误的拷贝数解极大地影响了聚类结果,以及缺乏自动设置具有所需可靠性水平的拷贝数的方法。这些多个水平的手动干预是有问题的,不仅因为它们需要(人)专家的存在以进行由机器更有效地完成的工作。事实上,由专家在选择拷贝数解、选择假设为克隆性的突变簇以及对该簇内的突变进行优先排序中应用的过程无法可靠地自动化。该过程需要专业知识,并且依赖于专家使用他们的判断以做出至少在某种程度上最终是主观的选择。这不能自动化,并且也不能可靠地再现,因为该过程涉及主观性。相比之下,根据本发明的方法,这些手动干预均不是必要的。多个拷贝数解及其置信度可以无缝集成(使得不需要被手动或自动选择),并且单独地对每个突变预测克隆性的可能性(使得不需要选择突变簇或在该簇内优先排序,并且可基于预测的可能性来定义自然的、可预测的和可再现的优先顺序)。因此,克隆性突变比先前可能的突变更可靠地被鉴定,使得其使用能够超出纯粹的研究并进入治疗开发的领域。对个体突变的克隆性之可能性的可靠估计值的可用性代表了对癌症治疗以及癌症监测和管理领域的明显改进,因为虽然假设了这样的克隆性突变在这些领域中的潜在用途,但是缺乏可靠鉴定这些用于整合到临床流水线中的突变的工具阻碍了其在实践中的实际使用。

[0081] 在本公开内容中,将使用以下术语,并且旨在如以下所指示进行定义。

[0082] 如本文所使用的“样品”可以是细胞或组织样品、生物流体、提取物(例如从对象获得的DNA提取物),从中可以获得基因组物质用于基因组分析例如基因组测序(例如全基因组测序、全外显子组测序)。样品可以是获自对象的细胞、组织或生物流体样品(例如活检物)。这样的样品可以被称为“对象样品”。特别地,样品可以是血液样品、或肿瘤样品、或来源于其的样品。样品可以是来自对象新鲜获得的样品,或者可以是在基因组分析之前已处理和/或储存(例如冷冻、固定或者经过一个或更多个纯化、富集或提取步骤)的样品。样品可以是细胞或组织培养物样品。因此,如本文中所述的样品可以指包含来源于其的细胞或基因组物质的任何类型的样品,无论是来自从对象获得的生物样品,或来自从例如细胞系获得的样品。在一些实施方案中,样品是从对象(例如人对象)获得的样品。样品优选来自哺乳动物(例如,如哺乳动物细胞样品或来自哺乳动物对象的样品,所述哺乳动物例如猫、狗、马、驴、绵羊、猪、山羊、牛、小鼠、大鼠、兔或豚鼠),优选来自人(例如,如人细胞样品或来自人对象的样品)。另外,样品可以被运输和/或储存,并且收集可以在远离基因组序列数据采集(例如测序)位置的位置处进行,和/或本文中所述的任何计算机实施的方法步骤可以在远离样品收集位置和/或远离基因组数据采集(例如测序)位置的位置处进行(例如计算机实施的方法步骤可借助于联网计算机,例如借助于“云”提供者进行)。

[0083] “混合样品”是指假设包含多个细胞类型或来源于多个细胞类型的遗传物质的样品。在本公开内容的上下文中,混合样品通常是包含肿瘤细胞、或假设(预期)包含肿瘤细胞或来源于肿瘤细胞的遗传物质的样品。从对象获得的样品(例如,如肿瘤样品)通常是混合样品(除非它们经过一个或更多个纯化和/或分离步骤)。通常来说,样品包含肿瘤细胞和至

少一种另外的细胞类型(和/或来源于其的遗传物质)。例如,混合样品可以是肿瘤样品。“肿瘤样品”是指来源于肿瘤或从肿瘤获得的样品。这些样品可包含肿瘤细胞和正常(非肿瘤)细胞。正常细胞可包含免疫细胞(例如,如淋巴细胞)和/或另外的正常(非肿瘤)细胞。在这样的混合样品中的淋巴细胞可被称为“肿瘤浸润淋巴细胞”(tumour-infiltrating lymphocyte, TIL)。肿瘤可以是实体瘤或者非实体或血液学肿瘤。肿瘤样品可以是原发性肿瘤样品、肿瘤相关的淋巴结样品或来自对象的转移部位的样品。包含肿瘤细胞或来源于肿瘤细胞的遗传物质的样品可以是体液样品。因此,来源于肿瘤细胞的遗传物质可以是循环肿瘤DNA或外排体中的肿瘤DNA。替代地或除此之外,样品可包含循环肿瘤细胞。混合样品可以是已被处理以提取遗传物质的细胞、组织或体液的样品。用于从生物样品中提取遗传物质的方法是本领域已知的。混合样品可以已经历了可改变样品中多个细胞类型或来源于多个细胞类型的遗传物质的比例的一个或更多个处理步骤。例如,包含肿瘤细胞的混合样品可以已被处理以富集肿瘤细胞中的样品。因此,在可存在少量其他类型的细胞的基础上,经纯化的肿瘤细胞的样品可被称为“混合样品”,即使出于特定目的可以假设该样品是纯的(即肿瘤分数为1或100%)。

[0084] 术语“肿瘤分数”(有时也称为“肿瘤纯度”或简称“纯度”,或异常细胞分数(aberrant cell fraction, ACF)是指混合样品中含有DNA的细胞(所述细胞是肿瘤细胞)的比例,或者是指假设产生样品中来自肿瘤细胞和非肿瘤细胞的遗传物质的特定混合物的当量比例。用于在样品中确定肿瘤分数的方法是本领域已知的。例如,在细胞或组织样品的情况下,肿瘤分数可通过分析病理切片(例如经苏木精和伊红(hematoxylin and eosin, H&E)染色的切片或其他组织化学或免疫组织化学切片,通过对样品的一个或更多个代表性区域中的肿瘤细胞进行计数来进行),或使用高通量测定(例如流式细胞术)进行估计。在包含遗传物质的样品的情况下,肿瘤分数可使用尝试对肿瘤和种系基因组进行解卷积的序列分析过程(例如如ASCAT(Van Loo et al., 2010), ABSOLUTE(Carter et al., 2012)或ichorCNA(Adalsteinsson et al. 2017))进行估计。

[0085] “正常样品”或“种系样品”是指假设不包含肿瘤细胞或来源于肿瘤细胞的遗传物质的样品。种系样品可以是血液样品、组织样品或经纯化的样品(例如来自对象的外周血单个核细胞样品)。类似地,当提及序列或基因型时,术语“正常”、“种系”或“野生型”是指除肿瘤细胞之外的细胞的序列/基因型。种系样品可包含小比例的肿瘤细胞或来源于其的遗传物质,并且然而出于实际目的,可以假设不包含所述细胞或遗传物质。换言之,所有细胞或遗传物质均可被假设为正常和/或者与该假设不兼容的序列数据可被忽略。

[0086] 术语“序列数据”是指指示样品中具有特定序列的基因组物质的存在且优选地还指示样品中具有特定序列的基因组物质的量的信息。这样的信息可使用以下获得:测序技术(例如,如下一代测序(next generation sequencing, NGS),例如全外显子组测序(whole exome sequencing, WES)、全基因组测序(whole genome sequencing, WGS)或捕获的基因组基因座的测序(靶向或组测序)),或使用阵列技术(例如,如拷贝数变异阵列,或其他分子计数测定)。当使用NGS技术时,序列数据可包括具有特定序列的测序读出的数量的计数。当使用非数字技术(例如阵列技术)时,序列数据可包括指示样品中具有特定序列的序列数量的信号(例如强度值),例如通过与适当的对照比较。使用本领域已知的方法(例如,如Bowtie(Langmead et al., 2009))将序列数据映射到参考序列(例如参考基因组)。因此,测序读出

或等效非数字信号的计数可与特定的基因组位置相关(其中“基因组位置”是指序列数据映射到的参考基因组中的位置)。此外,基因组位置可包含突变,在这种情况下,测序读出或等效非数字信号的计数可与特定基因组位置处的每个可能的变体(也称为“等位基因”)相关。鉴定在样品中特定位置处的突变的存在的过程被称为“变体调用”并且可以使用本领域已知的方法(例如,如GATK HaplotypeCaller,<http://gatk.broadinstitute.org/hc/en-us/articles/360037225632-HaplotypeCaller>)来进行。例如,序列数据可包含与特定基因组位置处的种系(有时也称为“参考”)等位基因匹配的读出(或等效的非数字信号)的数量的计数,以及与基因组位置处的突变(有时也称为“替代”)等位基因相匹配的读出(或等效的非数字信号)的数量的计数。

[0087] 此外,使用本领域已知的方法可将序列数据用于推断沿基因组的拷贝数谱。拷贝数谱可以是等位基因特异性的。在本公开内容的上下文中,拷贝数谱优选是等位基因特异性的和肿瘤/正常样品特异性的。换言之,本公开内容中使用的拷贝数谱优选使用旨在进行以下的方法来获得:分析包含肿瘤和正常细胞的混合物的样品,并且产生针对样品中肿瘤细胞和正常细胞的等位基因特异性拷贝数谱。混合样品的等位基因特异性拷贝数谱可使用例如ASCAT (Van Loo et al., 2010) 从序列数据(例如使用如上所述的读出计数)获得。另一些方法是已知的并且同样合适。优选地,在本公开内容的上下文中,用于获得等位基因特异性拷贝数谱的方法是报道多个可能的拷贝数解和相关的质量(quality)/置信度量度的方法。例如,ASCAT为倍性(整个肿瘤样品的倍性,不是片段特异性的)值和纯度值的每个组合输出拟合优度量度,其中评价了相应的等位基因特异性拷贝数谱。注意,通过这样的方法生成的肿瘤特异性拷贝数谱代表整个肿瘤细胞群的平均值或汇总(即,其不考虑在肿瘤群体内的异质性,这是本文中所述的新发展的目的)。

[0088] 术语“总拷贝数”是指样品中基因组区域的拷贝的总数。术语“主要拷贝数”是指样品中最普遍的等位基因的拷贝数。相反,术语“次要拷贝数”是指样品中除了最普遍的等位基因之外的等位基因的拷贝数。除非另有说明,否则这些术语是指所推断的主要拷贝数和所推断的肿瘤拷贝数谱的主要拷贝数(和总拷贝数)。术语“正常拷贝数”或“正常总拷贝数”是指样品中正常细胞中基因组区域的拷贝数。正常细胞通常具有每条染色体的两个拷贝(除非该细胞是遗传上雄性的并且该染色体是性染色体),并因此在实施方案中可假设正常拷贝数等于2(除非基因组区域在X或Y染色体上,并且在分析下的样品是来自雄性对象的,在这种情况下,可假设正常拷贝数等于1)。或者,特定基因组区域的正常拷贝数可使用正常样品来确定。

[0089] 术语“log R值”(有时称为“logR”、“logRR”、“LLR”)是指经归一化总信号强度的量度,定量了基因组基因座处的总拷贝数。在本公开内容的上下文中,该术语通常是指包含肿瘤遗传物质的样品的log R值,并且归一化通常通过参考正常样品(其优选为所匹配的正常样品,但还可以是处理匹配的正常样品或另一些合适的正常参考样品)来进行。例如,当使用NGS时,logR可以作为读出深度的归一化log变换($\log(\text{肿瘤读出深度}/\text{正常读出深度})$)来获得。术语“平均B等位基因频率”(mean B allele frequency, MBAF,有时也称为“B等位基因频率”(B allele frequency, BAF))是在基因组位置处的经归一化等位基因强度比率的量度。在本公开内容的上下文中,该术语通常是指包含肿瘤遗传物质的样品的BAF值,并且归一化通常通过参考正常样品(其优选为所匹配的正常样品,但还可以是处理匹配的正常

样品或其他合适的正常参考样品)来进行。例如,BAF可以作为肿瘤等位基因与正常等位基因的等位基因频率的比率来获得。拷贝数谱通常包含在称为“区段”的基因组区域上的拷贝数估计值。因此,与基因组位置相关的BAF和logR可以指与特定基因组位置(例如,如突变的基因组位置)重叠的区段的BAF和logR。此外,BAF和logR可用于获得相应的主要和次要拷贝数。在一些实施方案中,即使仅使用肿瘤拷贝数谱,也可为肿瘤拷贝数谱估计值和正常拷贝数谱估计值二者提供拷贝数度量的值。

[0090] 术语“肿瘤特异性突变”、“体细胞突变”或简称“突变”可互换使用,并且是指与来自同一对象的健康细胞相比,肿瘤细胞中核苷酸序列(例如DNA或RNA)中的差异。核苷酸序列中的差异可导致来自同一对象的健康细胞不表达的蛋白质的表达。例如,突变可以是单核苷酸变体(single nucleotide variant,SNV)、多核苷酸变体(multiple nucleotide variant,MNV)、缺失突变、插入突变、易位、错义突变、易位、融合、剪接位点突变或肿瘤细胞遗传物质的任何其他变化。突变可导致来自同一对象的健康细胞中不存在的蛋白质或肽的表达。突变可通过以下来鉴定:外显子组测序、RNA测序、全基因组测序和/或靶向基因组测序和/或单基因的常规Sanger测序,随后进行序列比对并将来自肿瘤样品的DNA和/或RNA序列与来自参考样品或参考序列(例如种系DNA和/或RNA序列,或来自数据库的参考序列)的DNA和/或RNA序列进行比较。合适的方法是本领域已知的。

[0091] “插入缺失突变”是指生物体的核苷酸序列(例如DNA或RNA)中碱基的插入和/或缺失。通常来说,插入缺失突变发生在生物体的DNA,优选基因组DNA中。在一些实施方案中,插入缺失可以是1至100个碱基,例如1至90、1至50、1至23或1至10个碱基。插入缺失突变可以是移码插入缺失突变。移码插入缺失突变是由一个或多个核苷酸的插入或缺失引起的核苷酸序列阅读框的变化。这样的移码插入缺失突变可以产生新的开放阅读框,所述开放阅读框通常与由对象中相应的健康细胞中的非突变DNA/RNA编码的多肽高度不同。

[0092] “新抗原”(或“新-抗原”)是作为癌细胞内突变的结果而产生的抗原。因此,新抗原不被正常(即非肿瘤)细胞表达(或以显著更低的水平表达)。新抗原可以被处理以产生不同的肽,所述肽当在MHC分子的情况下存在时可以被T细胞识别。如本文中所述,新抗原可用作癌症免疫治疗的基础。本文中提及的“新抗原”旨在也包括来源于新抗原的肽。如本文所使用的术语“新抗原”旨在涵盖具有免疫原性的新抗原的任何部分。如本文所指的“抗原”分子是当以适当的方式呈递至免疫系统或免疫细胞时,其自身或其一部分能够刺激免疫应答的分子。新抗原与特定MHC分子(由特定HLA等位基因编码)的结合可以使用本领域已知的方法来预测。用于预测MHC结合的方法的一些实例包括由Lundegaard et al.,O'Donnell et al.,和Bullik-Sullivan et al.描述的那些。例如,新抗原的MHC结合可以使用netMHC-3(Lundegaard et al.)和netMHCpan4(Jurtz et al.)算法进行预测。由此预测已预测与特定MHC分子结合的新抗原将由所述MHC分子呈递在细胞表面上。

[0093] “克隆性新抗原”(有时也称为“躯干新抗原”)是由存在于来自对象的一个或多个样品中的基本上每个肿瘤细胞中(或者可以假设存在于样品中肿瘤遗传物质所来源的基本上每个肿瘤细胞中)的突变产生的新抗原。类似地,“克隆性突变”(有时称为“躯干突变”)是存在于来自对象的一个或多个样品中的基本上每个肿瘤细胞中(或者可以假设存在于样品中肿瘤遗传物质所来源的基本上每个肿瘤细胞中)的突变。因此,克隆性突变可以是存在于来自对象的一个或多个样品中的每个肿瘤细胞中的突变。“亚克隆性”新抗原是由存

在于来自对象的一个或更多个肿瘤样品中的细胞的亚群或一部分中(或者可以假设存在于样品中肿瘤遗传物质所来源的肿瘤细胞的亚群中)的突变产生的新抗原。类似地,“亚克隆性”突变是存在于来自对象的一个或更多个肿瘤样品中的细胞的亚群或一部分中(或者可以假设存在于样品中肿瘤遗传物质所来源的肿瘤细胞的亚群中)的突变。新抗原或突变在来自对象的一个或更多个样品的背景下可以是克隆性的,而在可存在于对象中的整个肿瘤细胞群(例如包括原发性肿瘤和转移的所有区域)的背景下不是真正克隆性的。因此,在克隆性突变是存在于对象中的基本上每个肿瘤细胞(即所有肿瘤细胞)中的突变的意义上,克隆性突变可以是“真正克隆性”的。这是因为一个或更多个样品可以不代表存在于对象中的细胞的每一个亚群。因此,在本公开内容的上下文中,“克隆性新抗原”或“克隆性突变”还可称为“普遍存在的(ubiquitous)新抗原”或“普遍存在的突变”,以指示新抗原存在于已被分析的基本上所有肿瘤细胞中,但可以不存在于对象中可存在的所有肿瘤细胞中。术语“克隆性”和“普遍存在的”可互换使用,除非上下文指示旨在提及“真正克隆性”。与一个或更多个样品或者对象相关的措词“基本上每个肿瘤细胞”可指一个或更多个样品或者对象中至少70%、至少75%、至少80%、至少85%、至少90%、至少91%、至少92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少98%或至少99%的肿瘤细胞。

[0094] 然而,如本文中所述鉴定为可能是克隆性(或“普遍存在的”)的新抗原/突变可能是真正克隆性的,或者与鉴定为不可能是克隆性的新抗原/突变相比至少更有可能是真正克隆性的。此外,当用于鉴定克隆性新抗原/突变的样品捕获肿瘤遗传多样性的更完整图像时(例如通过包含来自对象的多个样品(例如如来自肿瘤不同区域的样品),和/或通过包含固有地捕获肿瘤细胞多样性的样品(例如如ctDNA样品)),本文中鉴定的克隆性新抗原/突变是真正克隆性的可能性的置信度提高。相反,如本文中所述鉴定为不可能是克隆性的新抗原/突变不可能是真正克隆性的,因为新抗原/突变不可能是克隆性的鉴定指示即使在由取样过程提供的限制性观察中,也有新抗原/突变不存在于所有肿瘤细胞中的证据。因此,鉴定克隆性新抗原/突变的过程可被视为基于对可从一个或更多个样品获得的对象肿瘤的克隆性结构的限制性观察而对最有可能是克隆性的候选新抗原/突变进行优先。

[0095] 术语“癌细胞分数”(或“CCF”)是指含有突变(例如如导致特定新抗原的突变)的肿瘤细胞的比例。在本公开内容的上下文中,癌细胞分数可基于一个或更多个样品来估计,并因此可不等于对象(如上所述)中的真实癌细胞分数。然而,基于一个或更多个样品估计的癌细胞分数可提供可能的真实癌细胞分数的可用指示。此外,如上所述,当用于估计癌细胞分数的样品捕获肿瘤遗传多样性的更完整图像时,这样的估计的准确性可以提高。基因组数据中的另外的噪声源和混杂因素意味着从一个或更多个样品确定的癌细胞分数代表了估计值。因此,虽然真正克隆性突变/新抗原应具有 $CCF=1$,但实际上,与预期与更低的CCF估计值相关的不太可能是克隆性的突变相比,预期更有可能是克隆性的突变/新抗原与更高的CCF估计值(其可能不等于1)相关。

[0096] 例如,癌细胞分数估计值可通过整合变体等位基因频率与拷贝数和纯度估计值来获得,如Landau et al (2013)所述。这样的CCF估计值还可用于鉴定可能是克隆性的突变。例如,克隆性突变可被限定为具有估计癌细胞分数(cancer cell fraction, CCF) ≥ 0.75 (例如 $CCF \geq 0.80$ 、 0.85 、 0.90 、 0.95 或 1.0)的突变。亚克隆性突变可被限定为具有 $CCF < 0.95$ 、 0.90 、 0.85 、 0.80 或 0.75 的突变。此外,CCF估计值可与使概率与0至1的CCF的多个可能值中

的每个相关的分布相关(例如来源于该分布),从该分布可获得置信度的统计学估计值。例如,如果95% CCF置信区间为 ≥ 0.75 (即估计的CCF的95%置信区间的上限大于或等于0.75),则可将突变限定为可能是克隆性突变。换言之,如果存在具有下限L和上限H的CCF区间,其使得 $P(L < \text{CCF} < H) = 95\%$ 且 $H \geq 0.75$,则可将突变限定为可能是克隆性突变。或者,如果存在突变的癌细胞分数(CCF)达到或超出如上所限定的所需值的超过50%机会或概率(例如0.75或0.95,例如50%、55%、60%、65%、70%、75%、80%、85%、90%、95%或更高的机会或概率),则可将突变鉴定为克隆性的。换言之,如果 $P(\text{CCF} > 0.75) \geq 0.5$,则可将突变鉴定为克隆性的。例如,基于突变的CCF超出0.95(或0.75,或任何其他所选择的阈值)的后验概率是否分别大于或小于0.5,可将突变分类为可能是克隆性的或亚克隆性的。

[0097] 根据本公开内容的方法,如下文将进一步描述的,获得了突变是克隆性的可能性。这等同于 $P(\text{CCF} = 1)$ 。在这种背景下,如下文将进一步解释的,如果 $P(\text{CCF} = 1)$ 超出阈值,则可将突变鉴定为可能是克隆性的。阈值可以是固定的。例如,如果 $P(\text{CCF} = 1) > 0.05$,则可将突变鉴定为可能是克隆性的。或者,可以针对所研究的一组特定突变来确定阈值。在一些实施方案中,可以基于具有已知克隆性/非克隆性状态的基准数据集来设置阈值,以达到预定的精度和/或召回率。基准数据集可使用合成数据和/或使用从具有已知克隆性结构的群体获得的数据集(例如细胞系混合数据)来获得。例如,如果 $P(\text{CCF} = 1) > t$,其中 t 是使得鉴定出基准数据集中真正克隆性突变的95%(或任何其他值,例如如50%、55%、60%、65%、70%、75%、80%、85%、90%)(即假阴性率为至多5%)的最大值,则可将突变鉴定为可能是克隆性的。作为另一个实例,如果 $P(\text{CCF} = 1) > t$,其中 t 是使得基准数据集中超出阈值的突变的至少50%(或任何其他值,例如如50%、55%、60%、65%、70%、75%、80%、85%、90%)是真正克隆性突变(即真阳性率为至少50%)的最小值,则可将突变鉴定为可能是克隆性的。或者,可以设置阈值,使得与具有满足上述标准的置信区间(例如,使得估计CCF的95%置信区间的上限大于或等于0.75)的估计CCF相关的任何突变(或一定%的突变)被选择为可能是克隆性的。或者,可以设置阈值,使得与具有满足上述标准的后验概率分布(例如,其CCF超出0.95(或0.75,或任何其他所选择的阈值)的后验概率大于0.5)的估计CCF相关的任何突变(或一定%的突变)被选择为可能是克隆性的。

[0098] 癌症免疫治疗(或简称“免疫治疗”)是指包括向对象施用免疫原性组合物(例如疫苗)、包含免疫细胞的组合物或免疫活性药物(例如如治疗性抗体)的治疗方法。术语“免疫治疗”还可以指治疗组合物本身。在本公开内容的上下文中,免疫治疗通常靶向新抗原。例如,免疫原性组合物或疫苗可包含新抗原、新抗原呈递细胞或表达新抗原所必需的物质。作为另一个实例,包含免疫细胞的组合物可包含识别新抗原的T和/或B细胞。免疫细胞可从肿瘤或其他组织(包括但不限于淋巴结、血液或腹水)中分离,离体或体外扩增并重新施用于对象(称为“过继细胞治疗”的过程)。替代地或除此之外,T细胞可从对象中分离,并经改造以靶向新抗原(例如,通过插入与新抗原结合的嵌合抗原受体)并重新施用于对象。作为另一个实例,治疗性抗体可以是识别新抗原的抗体。本领域技术人员将理解,如果新抗原是细胞表面抗原,则如本文所提及的抗体将识别该新抗原。当新抗原是胞内抗原时,抗体将识别新抗原肽-MHC复合物。如本文所提及的,“识别”新抗原的抗体涵盖这两种可能性。此外,免疫治疗可靶向多种新抗原。例如,免疫原性组合物可包含多种新抗原、呈递多种新抗原的细胞或表达多种新抗原所必需的物质。作为另一个实例,组合物可包含识别多种新抗原的免

疫细胞。类似地,组合物可包含识别相同新抗原的多种免疫细胞。作为另一个实例,组合物可包含识别多种新抗原的多种治疗性抗体。类似地,组合物可包含识别相同新抗原的多种治疗性抗体。

[0099] 如本文中所述的组合物可以是另外包含可药用载体、稀释剂或赋形剂的药物组合物。药物组合物可任选地包含一种或更多种另外的药学活性多肽和/或化合物。这样的制剂可以例如是适合于静脉内输注的形式。

[0100] 提及“免疫细胞”旨在涵盖免疫系统的细胞,例如T细胞、NK细胞、NKT细胞、B细胞和树突细胞。在一个优选的实施方案中,免疫细胞是T细胞。识别新抗原的免疫细胞可以是经改造T细胞。新抗原特异性T细胞可表达嵌合抗原受体(chimeric antigen receptor, CAR)或者特异性结合新抗原或新抗原肽的T细胞受体(T cell receptor, TCR)、或者特异性结合新抗原或新抗原肽的亲和力增强的T细胞受体(TCR)(如下文中进一步讨论的)。例如,T细胞可表达嵌合抗原受体(CAR)或者特异性结合新抗原或新抗原肽的T细胞受体(TCR)(例如特异性结合新抗原或新抗原肽的亲和力增强的T细胞受体(TCR))。或者,识别新抗原的免疫细胞群可以是来自患有肿瘤的对象中分离的T细胞群。例如,T细胞群可以由从对象中分离的样品(例如如肿瘤样品、外周血样品或来自对象其他组织的样品)中的T细胞产生。T细胞群可以由来自其中鉴定出新抗原的肿瘤的样品产生。换言之,T细胞群可以来源于待治疗患者的肿瘤的样品中分离,其中新抗原也来自所述肿瘤的样品中鉴定出。T细胞群可包含肿瘤浸润淋巴细胞(tumour infiltrating lymphocyte, TIL)。

[0101] 术语“抗体”(Antibody, Ab)包括单克隆抗体、多克隆抗体、多特异性抗体(例如,双特异性抗体)和表现出期望的生物活性的抗体片段。术语“免疫球蛋白”(immunoglobulin, Ig)可与“抗体”互换使用。一旦例如通过根据本公开内容的方法已经鉴定出合适的新抗原,则可以使用本领域已知的方法来产生抗体。

[0102] “免疫原性组合物”是能够在对象中诱导免疫应答的组合物。该术语与术语“疫苗”可互换使用。本文中所述的免疫原性组合物或疫苗可导致对象中产生免疫应答。可产生的“免疫应答”可以是体液和/或细胞介导的免疫(例如刺激抗体产生,或者刺激细胞毒性或杀伤细胞),其可以识别并破坏(或以其他方式消除)在细胞表面表达对应于疫苗中抗原的抗原的细胞。免疫原性组合物可包含一种或更多种新抗原,或者表达一种或更多种新抗原所必需的物质。另外,新抗原可以以细胞(例如抗原呈递细胞,例如树突细胞)的形式递送。抗原呈递细胞(例如树突细胞)可用新抗原或新抗原肽进行脉冲或负载,或者进行遗传修饰(通过DNA或RNA转移)以表达一种、两种或更多种新抗原或新抗原肽,例如2、3、4、5、6、7、8、9或10种新抗原或新抗原肽。制备树突细胞免疫原性组合物或疫苗的方法是本领域已知的。

[0103] 新抗原肽可使用本领域已知的方法合成。术语“肽”以正常意义使用以意指通常通过相邻氨基酸的 α -氨基和羧基之间的肽键相互连接的一系列残基(通常是L-氨基酸)。该术语包括经修饰的肽和合成的肽类似物。新抗原肽可以在肽内的任何残基位置处包含癌细胞特异性突变(例如由单核苷酸变体(SNV)编码的非沉默氨基酸替换)。举例来说,能够与MHC I类分子结合的肽的长度通常为7至13个氨基酸。因此,氨基酸替换可存在于包含十三个氨基酸的肽中的第1、2、3、4、5、6、7、8、9、10、11、12或13位处。在一些实施方案中,可使用较长的肽(例如21至31-聚体),并且突变可以在任何位置处,例如在肽的中心处,例如在第10、11、12、13、14、15或16位处。这样的肽还可用于刺激CD4和CD8细胞二者以识别新抗原。

[0104] 如本文所使用的“治疗”是指相对于治疗之前的症状,减少、减轻或消除正在治疗的疾病的一种或更多种症状。“预防”(或预防(prophylaxis))是指延迟或预防疾病症状的发作。预防可以是绝对的(使得不发生疾病)或者可仅在一些个体中或对有限的时间量是有效的。

[0105] 如本文所使用的,术语“计算机系统”包括用于实现根据上述实施方案的系统或实施根据上述实施方案的方法的硬件、软件和数据存储装置。例如,计算机系统可包含中央处理单元(central processing unit,CPU)、输入装置(mean)、输出装置和数据存储器,其可作为一个或更多个连接的计算装置来实现。优选地,计算机系统具有显示器或包含具有显示器的计算装置,以提供视觉输出显示(例如在业务过程(business process)的设计中)。数据存储器可包括RAM、磁盘驱动器或其他非暂时性计算机可读介质。计算机系统可包含通过网络连接并且能够在该网络内相互通信的多个计算装置。明确设想了计算机系统可以由云计算机组成或包含云计算机。

[0106] 如本文所使用的,术语“计算机可读介质”包括但不限于任何非暂时性介质或者可由计算机或计算机系统直接读出和访问的介质。所述介质可包括但不限于磁存储介质(例如软盘、硬盘存储介质和磁带)、光存储介质(例如光盘或CD-ROM)、电存储介质(例如存储器,包括RAM、ROM和闪存),以及上述的混合和组合(例如磁/光存储介质)。

[0107] 克隆性突变的鉴定

[0108] 本公开内容提供了用于使用来自一个或更多个包含肿瘤细胞或来源于其的遗传物质的样品的序列数据来确定肿瘤特异性突变是否可能是克隆性的方法。本公开内容还提供了用于鉴定克隆性新抗原的方法,所述方法包括确定一个或更多个肿瘤特异性突变是否是/可能是克隆性的。将通过参考图1来描述举例说明性方法。在任选的步骤10下,可从对象获得包含来自肿瘤的基因组物质的样品。样品通常是包含来自多种细胞类型(包括肿瘤细胞)的基因组物质的混合样品。优选地,可以获得或者可以先前已获得所匹配的样品,其不包含来自肿瘤细胞的基因组物质或者可从其中提取来自正常细胞的基因组物质。所匹配的样品是从与肿瘤样品相同的对象获得的样品。使用所匹配的正常样品提高了调用(call)体细胞(肿瘤特异性)突变的准确性,因为可以将肿瘤样品中鉴定的任何变体位置与所匹配的正常样品中的变体位置进行比较以排除种系变体。相同的所匹配的正常样品可用于分析来自对象的多个肿瘤样品。此外,所匹配的样品和一个或更多个肿瘤样品可在不同时间获得。例如,可在诊断或切除肿瘤的时间时获得第一肿瘤样品和所匹配的样品,并且可以在稍后的时间点时获得另外的肿瘤样品并与初始所匹配的样品一起分析。当所匹配的样品不可用时,可使用参考样品或包含常见体细胞变体的基因组。或者,可使用经处理的所匹配的正常样品,其可以不是从同一对象获得的,或者可从对象池获得。

[0109] 在任选的步骤12下,可以例如通过使用全外显子组测序或全基因组测序之一对样品中的基因组物质进行测序来确定一个或更多个混合样品和任选地所匹配的样品的序列内容。尽管测序方法是优选的,但可以使用替代方法例如如等位基因特异性拷贝数阵列,因为它们产生代表样品中每个特定序列的数目的数字输出。在任选的步骤14下,可以分析序列数据以鉴定可能存在于肿瘤细胞中但不存在于非癌细胞中的一个或更多个突变。这些代表肿瘤特异性突变,并且可用作候选新抗原。这可包括以下步骤:比对来自一个或更多个样品(即,混合样品和种系样品,如果可用的话)的序列,以及鉴定其中肿瘤序列与种系序列不

同或可以假设与种系序列不同(例如,如果对象的种系序列不可用的话)的基因组位置。

[0110] 在步骤16下,获得了混合样品在候选肿瘤特异性突变的基因组位置处的序列数据,包括支持突变等位基因(也称为“非参考等位基因”)的读出计数、支持基因组位置处的种系等位基因(A,如果基因座在种系群体中是杂合的,则统称为“种系等位基因”,也称为“参考”、“野生型”或“正常”等位基因)的读出计数、和/或候选肿瘤特异性突变的基因组位置处的读出总计数。仅需要获得这些度量中的两个,因为第三个度量可以从这些的任意两个中推导出。替代地或除此之外,序列数据可包括从其中可获得计数的读出数据或强度数据。在任选的步骤18下,可获得关于与包含肿瘤遗传物质的每个样品相容的至少一种拷贝数解的信息。该信息可包含样品肿瘤分数的等位基因特异性拷贝数度量,其选自主要拷贝数、次要拷贝数、总拷贝数、平均B等位基因频率、log R值和肿瘤倍性以及正常拷贝数,或者来源于这些度量的信息,例如与这些等位基因特异性拷贝数度量相容的一组候选联合基因型。并非所有这样的等位基因特异性拷贝数度量都是必要的,因为一些度量包含冗余信息和/或可与合适的默认值相关。例如,正常拷贝数可以与合适的默认值相关,如上所述。此外,仅需要主要拷贝数、总拷贝数和次要拷贝数中的两个以推断第三个。类似地,这三个值可以从MBAF和logR值推断出(反之亦然)。任选地,拷贝数解可与相应的置信度度量相关。当这样的度量不可用时,可以假设每个拷贝数解具有同等可能性。每个候选联合基因型包含以下群体在肿瘤特异性突变位置的基因型:正常群体、不含肿瘤特异性突变的参考肿瘤群体和含肿瘤特异性突变的变体肿瘤细胞群体。

[0111] 在步骤20下,作为后验概率确定肿瘤特异性突变是克隆性的概率,该后验概率取决于:突变是克隆性的先验概率,以及在肿瘤特异性突变是(i)克隆性的或(ii)非克隆性的情况下,鉴于一个或多个样品中每一个的肿瘤分数和一个或多个候选联合基因型而观察到序列数据的概率。先验概率是表示在考虑一些证据之前关于数量的信念的概率。在本上下文中,突变是克隆性的先验概率可表示在肿瘤中突变是克隆性的概率,其基于现有知识或假设,并且不考虑来自混合样品的序列数据。在步骤22下,确定肿瘤特异性突变是否可能产生新抗原。例如,可以确定突变是否可能导致通过种系细胞(其基因组不包含突变)不表达的肽或蛋白质。该步骤可以在步骤14之后的任何点处进行,并且特别地不需要在步骤16至20之后进行。例如,在确定肿瘤特异性突变是否可能是克隆性的之前,可以根据候选肿瘤特异性突变是否可能产生新抗原来过滤候选肿瘤特异性突变。在步骤24下,可以鉴定满足应用于步骤20的结果的一个或多个标准和应用于步骤22的结果的一个或多个标准的肿瘤特异性突变。这些可被认为是代表候选克隆性新抗原。在任选的步骤26下,可以例如通过用户界面向用户提供前述步骤(并且特别是步骤20至24)中任一者的结果。这些结果可用于例如为对象提供免疫治疗或预后,如下文将进一步描述的。

[0112] 应用

[0113] 上述方法在癌症诊断、预后和治疗方法的背景下存在应用。特别地,上述方法可用于提供靶向克隆性新抗原的免疫治疗。因此,本文中描述还为对象提供免疫治疗的方法,所述方法包括从来自对象的一个或多个样品中鉴定一种或更多种克隆性新抗原。

[0114] 图2示意性地示出了提供免疫治疗的示例性方法。在任选的步骤210下,从对象获得了一个或多个包含肿瘤遗传物质的样品和一个或多个种系样品。对象可以是已被诊断为患有癌症的对象,并且可以是(但不需要是)为其提供免疫治疗的同一对象。在步骤212

下,使用本文中所述的方法(例如通过参考图1)获得了候选克隆性新抗原的列表。列表可包含单一新抗原或多种新抗原。优选地,列表包含多种新抗原。在步骤214下,设计了靶向至少一种(并且任选地多种)候选新抗原的免疫治疗。设计这样的免疫治疗可包括鉴定候选克隆性新抗原中每种的一种或更多种候选肽(步骤214A)。例如,可以针对候选克隆性新抗原中的至少一种设计多种肽,其与相应的种系肽相比在其长度和/或表征新抗原的序列变体的位置方面不同。在步骤214B下,可以在体外和/或在计算机中测试一种或更多种经鉴定的肽,以评价一个或多个特性,例如其免疫原性、被MHC分子展示的可能性等。在任选的步骤214C下,可例如基于步骤214B的结果来选择肽中的一种或更多种。

[0115] 在步骤216下,可获得所选择的肽。具有所选择的序列的肽可以使用本领域已知的任何方法(例如如使用表达系统或通过直接合成)来获得。在步骤218下,可以使用一种或更多种候选肽来产生免疫治疗。免疫治疗可包含一种或更多种候选肽或足以表达其的物质(例如,在免疫原性组合物或疫苗的情况下),或者可包含已使用候选肽获得的分子或细胞(例如,在选择性结合候选肽的治疗性抗体、或者特异性识别候选肽的免疫细胞的情况下)。在任选的步骤220下,免疫治疗可以施用于对象,所述对象优选是已从其获得用于鉴定克隆性新抗原的样品的对象。将描述产生免疫治疗的实例,所述免疫治疗包含选择性富集的具有识别一种或更多种克隆性新抗原的T细胞的T细胞群。在步骤218A下,可获得T细胞群。T细胞可从待治疗的对象获得,但并非必须如此。T细胞可从肿瘤样品、从血液样品、或从任何其他组织样品获得。在步骤218B下,可获得树突细胞群。例如,树突细胞群可来源于来自待治疗对象的单个核细胞(例如外周血单个核细胞(peripheral blood mononuclear cell), PBMC)。在步骤218C下,树突细胞群可以是用候选肽进行脉冲。在步骤218D下,可使用经脉冲的树突细胞群来选择性地扩增T细胞群。可使用另外的扩增因子,例如如细胞因子或刺激性抗体。

[0116] 因此,本公开内容还提供了T细胞组合物,其包含选择性富集的具有识别一种或更多种克隆性新抗原的T细胞的T细胞群,其中所述一种或更多种克隆性新抗原已使用本文中所述的方法中的任一者进行鉴定。

[0117] 在如本文中所述的T细胞组合物中,扩增的新抗原反应性T细胞群可以比没有扩增的T细胞群具有更高的活性,如通过T细胞群对新抗原肽再刺激的响应而测量的。活性可以通过细胞因子产生来测量,并且其中较高的活性是活性提高5至10倍或更高。

[0118] 提及多种克隆性新抗原可以是指多种肽或蛋白质,每种肽或蛋白质包含产生新抗原的不同肿瘤特异性突变。所述多种可以是2至250、3至200、4至150或5至100个肿瘤特异性突变,例如5至75或10至50个肿瘤特异性突变。每个肿瘤特异性突变可由一种或更多种克隆性新抗原肽代表。换言之,多种克隆性新抗原可包含多种不同的肽,其中一些包含含有相同肿瘤特异性突变的序列(例如在肽序列内的不同位置处,或在不同长度的肽内)。

[0119] 根据本公开内容产生的T细胞群将具有提高数目或比例的T细胞,所述T细胞靶向一种或更多种预测为克隆性的新抗原。也就是说,T细胞群的组成将不同于“天然”T细胞群(即未经历本文中所讨论的扩增步骤的群体)的组成,因为靶向预测为克隆性的新抗原的T细胞的百分比或比例将提高。根据本公开内容的T细胞群可具有至少约0.2、0.3、0.4、0.5、0.6、0.7、0.8、0.9、1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、25、30、35、40、45、50、55、60、65、70、75、80、85、90、95或100%靶向预测为克隆性的新抗原的T细胞。

[0120] 本文中所述的免疫治疗可用于治疗癌症。因此,本公开内容还提供了在对象中治疗癌症的方法,其包括向对象施用如本文中所述的免疫治疗组合物。

[0121] 另外,克隆性新抗原的存在已显示出与改善的癌症预后相关。因此,本文中還描述了为已被诊断为患有癌症的对象提供预后的方法,所述方法包括确定来自对象的一个或更多个肿瘤样品中克隆性新抗原的分数和/或数目。

[0122] 适当地,在本文中所述的任意方面的任意实施方案中,癌症可以是卵巢癌、乳腺癌、子宫内膜癌、肾癌(肾细胞)、肺癌(小细胞、非小细胞和间皮瘤)、脑癌(胶质瘤、星形细胞瘤、胶质母细胞瘤)、黑素瘤、梅克尔细胞癌(merkel cell carcinoma)、透明细胞肾细胞癌(clear cell renal cell carcinoma,ccRCC)、淋巴瘤、小肠癌(十二指肠和空肠)、白血病、胰腺癌、肝胆肿瘤、生殖细胞癌、前列腺癌、头颈癌、甲状腺癌和肉瘤。例如,癌症可以是肺癌,例如肺腺癌或肺鳞状细胞癌。作为另一个实例,癌症可以是黑素瘤。在一些实施方案中,癌症可以选自黑素瘤、梅克尔细胞癌、肾癌、非小细胞肺癌(non-small cell lung cancer, NSCLC)、膀胱尿路上皮癌(urothelial carcinoma of the bladder,BLAC)和头颈鳞状细胞癌(head and neck squamous cell carcinoma,HNSC)和高微卫星不稳定性(microsatellite instability,MSI)癌症。在一些实施方案中,癌症是非小细胞肺癌(NSCLC)。在另一些实施方案中,癌症是黑素瘤。

[0123] 使用本公开内容的组合物和方法的治疗还可以涵盖靶向循环肿瘤细胞和/或来源于肿瘤的转移。靶向一种或更多种新抗原的根据本公开内容的治疗可以帮助阻止治疗抗性肿瘤细胞的进化,其可以在标准方法(例如化学治疗、放射治疗或非特异性免疫治疗)的情况下发生。本文中所述的用于治疗癌症的方法和用途可以与另外的癌症治疗组合进行。特别地,本文中所述的T细胞组合物可与免疫检查点干预、共刺激抗体、化学治疗和/或放射治疗、靶向治疗或单克隆抗体治疗组合施用。“组合”可以指在施用如本文中所述的T细胞组合物之前、同时或之后施用另外的治疗。

[0124] 本公开内容还提供了用于产生免疫治疗组合物的方法,所述方法包括鉴定可能是克隆性的新抗原并产生靶向新抗原的免疫治疗组合物。

[0125] 本文中還描述了治疗已被诊断为患有癌症的对象的方法,所述方法包括:通过以下鉴定一种或更多种克隆性新抗原:在对象中鉴定多个肿瘤特异性突变;在对象中确定一个或更多个肿瘤特异性突变是否可能是克隆性的;选择一个或更多个肿瘤特异性突变作为候选克隆性新抗原,其中候选克隆性新抗原是满足关于肿瘤特异性突变是否可能是克隆性的至少一个或更多个预定标准的肿瘤特异性突变;以及用靶向所选候选克隆性新抗原中的一种或更多种的免疫治疗来治疗所述对象;其中在对象中确定肿瘤特异性突变是否可能是克隆性的是使用本文中所述的方法来进行的。特别地,在对象中确定肿瘤特异性突变是否可能是克隆性的可包括:通过处理器从来自包含肿瘤遗传物质的对象的一个或更多个样品中获得序列数据,对于一个或更多个样品中的每个,所述序列数据包含以下中的至少两个:样品中显示肿瘤特异性突变的读出数目(d_b)、样品中显示相应种系等位基因的读出数目、以及在肿瘤特异性突变的位置处的读出总数目(d);并且通过处理器作为后验概率确定肿瘤特异性突变是克隆性的可能性,该后验概率取决于:突变是克隆性的先验概率,以及在肿瘤特异性突变是(i)克隆性的或(ii)非克隆性的情况下,鉴于一个或更多个样品中每一个的肿瘤分数和一个或更多个候选联合基因型而观察到序列数据的概率,每个候选联合基因

型包含以下群体在肿瘤特异性突变位置的基因型：正常群体、不含肿瘤特异性突变的参考肿瘤群体和含肿瘤特异性突变的变体肿瘤细胞群体。

[0126] 候选克隆性新抗原可被选为肿瘤特异性突变，其进一步满足关于肿瘤特异性突变是否可能产生新抗原的至少一个或更多个预定标准。通过所述处理器选择一个或更多个肿瘤特异性突变作为候选克隆性新抗原的步骤可包括，确定一个或更多个肿瘤特异性突变是否满足关于肿瘤特异性突变是否可能产生新抗原的一个或更多个标准，所述标准选自：突变与在肿瘤细胞中表达的表达产物相关，突变被预测为产生在对象的正常细胞中不表达的蛋白质或肽，突变被预测为产生可能由MHC分子呈递的至少一种肽，突变被预测为产生可能由已知存在于对象中的MHC等位基因呈递的至少一种肽，以及突变被预测为产生为免疫原性的蛋白质或肽。通过所述处理器选择一个或更多个肿瘤特异性突变作为候选克隆性新抗原的步骤可包括，通过所述处理器确定一个或更多个肿瘤特异性突变是否满足关于肿瘤特异性突变是否可能是克隆性的一个或更多个预定标准，所述预定标准选自：突变是克隆性的可能性高于预定阈值；突变是克隆性的可能性高于阈值，该阈值适应性地设置为在已确定可能性的肿瘤特异性突变中选出克隆性的可能性最高的预定数目的肿瘤特异性突变；以及突变是克隆性的可能性高于阈值，该阈值适应性地设置为在已确定可能性的肿瘤特异性突变中选出预定最高百分位数的肿瘤特异性突变。

[0127] 靶向所选克隆性新抗原中的一种或更多种的免疫治疗可以是免疫原性组合物、包含免疫细胞的组合物或治疗性抗体。免疫治疗可以是包含识别所鉴定的所选克隆性新抗原中的一种或更多种中的至少一种的T细胞的组合物。组合物可以富集靶向所鉴定的所选克隆性新抗原中的一种或更多种中的至少一种的T细胞。所述方法可包括获得T细胞群，并扩增该T细胞群以提高靶向所鉴定的所选克隆性新抗原中的一种或更多种中的至少一种的T细胞的数目或相对比例。

[0128] 系统

[0129] 图3示出了根据本公开内容用于确定肿瘤特异性突变是否可能是克隆性的和/或鉴定克隆性新抗原和/或用于至少部分基于所鉴定的克隆性新抗原提供预后或提供免疫治疗的系统的一个实施方案。该系统包含计算装置1，所述计算装置1包含处理器101和计算机可读存储器102。在所示的实施方案中，计算装置1还包含被示为屏幕但可包含向用户传递信息的任何其他装置（例如如通过听觉或视觉信号）的用户界面103。计算装置1可通信地（例如如通过网络6）与序列数据采集装置3（例如测序机器）连接，和/或与存储序列数据的一个或更多个数据库2连接。一个或更多个数据库可另外地存储可由计算装置1使用的其他类型的信息，例如如参考序列、参数等。计算装置可以是智能手机、平板电脑、个人计算机或其他计算装置。如本文中所述，计算装置被配置成实现用于确定肿瘤特异性突变是否可能是克隆性的方法。在一些替代实施方案中，计算装置1被配置成与远程计算装置（未示出）通信，远程计算装置本身被配置成实现确定肿瘤特异性突变是否可能是克隆性的方法，如本文中所述。在这样的情况下，远程计算装置还可以被配置成将该方法的结果发送至计算装置。计算装置1与远程计算装置之间的通信可通过有线或无线连接进行，并且可通过本地或公共网络（例如如通过公共互联网或通过WiFi）发生。

[0130] 序列数据采集3装置可以与计算装置1有线连接，或者可以能够通过无线连接（例如如通过网络6）进行通信，如所示的。计算装置1与序列数据采集装置3之间的连接可以是

直接的或间接的(例如如通过远程计算机)。序列数据采集装置3被配置成从核酸样品(例如从细胞和/或组织样品中提取的基因组DNA样品)中获取序列数据。在一些实施方案中,样品可以已经经受过一个或更多个预处理步骤,例如DNA纯化、片段化、文库制备、靶序列捕获(例如如外显子捕获和/或组序列捕获)。优选地,样品没有经受过扩增,或者当其已经经受过扩增时,这是在存在扩增偏倚(bias)控制装置(例如如使用独特的分子标识符)的情况下进行的。适合用于确定基因组拷贝数谱(无论是全基因组还是序列特异性)的任何样品制备方法都可以在本公开内容的上下文中使用。序列数据采集装置优选地是下一代测序仪。序列数据采集装置3可以与一个或更多个数据库2直接或间接连接,序列数据(原始或部分处理的)可以存储在数据库2上。

[0131] 以下内容以实例方式呈现,并且不应被解释为对权利要求书的范围的限制。

[0132] 实施例

[0133] 这些实施例描述了根据本公开内容的鉴定克隆性突变的方法,并使用模拟数据和多种类型的实验数据表明了其用途。

[0134] 方法

[0135] 突变基因型模型

[0136] 该模型的数据是来自 S 个样品($s=1, \dots, S$)的 N 个突变($n=1, \dots, N$)的等位基因计数。为简单起见,并且因为该方法可以分析单个样品和突变,突变的指数 n 和样品的指数 s 将不会明确地包括在本节使用的符号中。如图4所示,该模型假设各突变将被测序的细胞组分成三个亚群:(i)由具有健康种系基因组(可能是突变区域中的二倍体)的细胞组成的正常细胞群;(ii)由没有所讨论突变(在所讨论的突变区域中可以是非整倍体)的癌细胞组成的参考细胞群;以及(iii)由有所讨论突变的癌细胞组成的变体细胞群(在所讨论的突变区域中可以是非整倍体,在所述区域中可不具有与参考群相同的拷贝数)。术语“突变”在此旨在在其最广义上是指在序列数据,并且特别是基因组序列数据中可检测到的任何遗传改变。这特别地包括单核苷酸变体(SNV)、多核苷酸变体(MNV)、插失(indel)等。

[0137] 设 $G = (A, B, AA, AB, AAA, AABB, \dots)$ 是所有基因型的集合,其中 A 和 B 分别代表参考等位基因和变体等位基因。例如, AB 将代表总拷贝数为2的杂合变体(包含一个参考/正常等位基因 A 和一个变体等位基因 B)。在该符号下,在图4中,正常群体具有基因型 AA (其中两个 A 可以相同或不同,即正常群体可以是纯合或杂合的,但两个等位基因都是正常的),参考群体具有基因型 AAA (其中 A 等位基因是选自正常群体的 A 等位基因),并且变体群体具有基因型 $AABB$ (其中 A 等位基因是选自正常群体的 A 等位基因,并且 B 等位基因是任何非参考等位基因)。假设每个亚群内所有细胞的基因型都是恒定的(即参考图4,正常群体中的所有细胞都具有基因型 AA ,参考群体中的所有细胞都具有基因型 AAA ,并且变体群体中的所有细胞都具有基因型 $AABB$)。设 $G = (G_H; G_R; G_V) \in G^3$ 为向量,其中条目分别是正常(健康)、参考和变体群体的基因型(这些个体基因型中的每种将在下文通称为“ G ”)。设 t 为样品中癌细胞的比例。这通常被称为样品的肿瘤含量、肿瘤纯度或细胞性。设 ϕ 为样品中携带突变的癌细胞的比例,即变体群体中癌细胞的相对比例。这通常被称为突变的细胞患病率或癌细胞分数(CCF)。设 ϵ 为假设的测序误差率。限定了以下函数:

[0138] $a(G): G \rightarrow \mathbb{N}$ 是将基因型映射至 A 等位基因数的函数(例如,其中 G 是 AA , $a(G) = 2$);

[0139] $b(G): G \rightarrow \mathcal{N}$ 是将基因型映射至B等位基因数的函数(例如,其中G是AA, $b(G) = 0$);

[0140] $c(G): G \rightarrow \mathcal{N}$ 是将基因型映射至基因座处的总拷贝数的函数(即 $c(G) = a(G) + b(G)$; 例如,其中G是AA, $c(G) = 2$);

[0141] $\mu(G): G \rightarrow \mathcal{N}$ 是将基因型映射至值 $\mu(G) = \text{最小}\{\text{最大}\{(b(G)/c(G)), \epsilon\}, (1-\epsilon)\}$ 的函数,该函数可以被解释为从具有基因型G的群体中取样突变的读出的概率。

[0142] 设 $\xi(G, \phi, t)$ 为取样变体等位基因的读出的概率。假设具有无限的初始细胞群,在测序时对其进行取样,则取样变体等位基因的读出的概率大致与DNA输入库中变体等位基因的拷贝数成比例。更正式地,考虑到测序误差,取样变体等位基因的概率(给定一组基因型G、肿瘤含量t和癌细胞分数 ϕ) 由以下等式(等式(1))给出:

$$[0143] \quad \xi(G, \phi, t) = \frac{1}{T} (1-t)c(G_H)\mu(G_H) + \frac{1}{T} t(1-\phi)c(G_R)\mu(G_R) + \frac{1}{T} t\phi c(G_V)\mu(G_V) \quad (1)$$

[0144] 其中 $T = (1-t)c(G_H) + t(1-\phi)c(G_R) + t\phi c(G_V)$ (2)。

[0145] 变量 $\xi(G, \phi, t)$ 捕获以下之和:来源于每个基因型的变体等位基因的拷贝数乘以取样来自该基因型的突变的读出的概率,通过来源于每个基因型的两个等位基因的总拷贝数之和归一化。

[0146] 变量d是覆盖样品中突变的读出总数目,其中 d_b 包含突变体等位基因。因此,观察到这些读出数目d, d_b ($P(d, d_b | G, \phi, t)$) 的概率可以用具有参数 d_b 和 $\xi(G, \phi, t)$ 的二项式模型(等式(3))表示。这是因为具有参数p的m个伯努利(Bernouilli)随机变量之和遵循具有参数m、 p^2 的二项式分布。可以使用具有平均值 $\xi(G, \phi, t)$ 和精度(方差的倒数) γ 的 β -二项式模型(等式(4))来代替,例如,如果数据具有比由二项式模型可解释的方差更大的方差,则:

$$[0147] \quad P(d, d_b | G, \phi, t) = \text{二项式}(d_b | d, \xi(G, \phi, t)) \quad (3)$$

$$[0148] \quad P(d, d_b | G, \phi, t, \gamma) = \beta\text{二项式}(d_b | d, \xi(G, \phi, t), \gamma) \quad (4)。$$

[0149] 在以下实施例中,将参数 γ 设置为200,但也可可为其他值。迄今为止,假设亚群的基因型是已知的。一般而言,这对于健康群体(例如,来自所匹配的种系样品)可以是真实的,但对于参考和变体群体则是不真实的。替代地,通常观察重叠突变区域的等位基因特异性拷贝数估计值。使用该信息,可以在一组合理基因型内得出先验。将在下一节中解释如何做到这一点。现在假设具有先验概率的向量 π , 其中 π_i 是群体中第i个合理联合基因型 G_i 的先验概率。可以将观察到的数据在所有合理基因型内边缘化的概率写成以下(等式(3a)、(4a)):

$$[0150] \quad P(d, d_b | \pi, \phi, t) = \sum_i \pi_i \text{二项式}(d_b | d, \xi(G_i, \phi, t)) \quad (3a)$$

$$[0151] \quad P(d, d_b | \pi, \phi, t, \gamma) = \sum_i \pi_i \beta\text{二项式}(d_b | d, \xi(G_i, \phi, t), \gamma) \quad (4a)。$$

[0152] 在随后的节中,符号 $\text{Pr}(d, d_b | \pi, \phi, t)$ 将用于等同地指等式(3a)和等式(4a)的表达式。注意, ϕ 和 t 与单个样品相关,因此以上符号分别是 ϕ_s 和 t_s 的简写。

[0153] 引发突变基因型先验

[0154] 上述模型使用已知的联合基因型或先验概率 π , 其中 π_i 是群体中的第i个合理联合基因型 G_i 的先验概率(即 G_i 是健康、变体和参考群体的基因型的一种可能的组合)。可以使用多种方法来设置潜在的基因型先验。

[0155] 例如,一种可能的方法可以称为“主要拷贝数”方法。设 $c_{\text{主要}}$ 和 $c_{\text{次要}}$ 表示肿瘤样品中

重叠突变区域的主要和次要等位基因拷贝数。方法“主要拷贝数方法”考虑两种情况：

[0156] (a) 在第一种情况下,突变在拷贝数事件之前发生。在这种情况下,参考群体基因型与正常群体相匹配。考虑了变体群体的所有可能的突变基因型,该变体群体具有多至 $c_{\text{主要}}$ 染色体包含变体。

[0157] (b) 在第二种情况下,突变在拷贝数事件之后发生。在这种情况下,参考群体具有 $c_{\text{主要}}+c_{\text{次要}}$ 参考等位基因。变体群体具有1个变体等位基因和 $c_{\text{主要}}+c_{\text{次要}}-1$ 参考等位基因。

[0158] 将所有可能的突变基因型的先验权重设置为等同。例如,假设 $c_{\text{主要}}=2$ 和 $c_{\text{次要}}=1$,并且正常拷贝数为2。具有以下可能的基因型:

[0159] $G_1 = (AA, AA, AAB)$

[0160] $G_2 = (AA, AA, ABB)$

[0161] $G_3 = (AA, AAA, AAB)$

[0162] 每个的先验概率都为1/3。注意到,如果等位基因特异性拷贝数不可用,则 $c_{\text{主要}}$ 可设置为总拷贝数并且 $c_{\text{次要}}$ 设置为零。这种方法假设突变仅发生一次,使得如果在变体群体中存在多于一个突变体等位基因拷贝,则发生这种情况是因为突变先于基因座处的拷贝数变化,并随后被扩增。这种方法在考虑群体基因型的不确定性与不考虑太多状态之间达成良好的平衡。

[0163] 可使用替代方法用于设置突变基因型先验。另一种可能的方法是简单地假设每个突变都是二倍体和杂合的(即变体群体中的变体仅发生在两条染色体中的一条上, $G = (G_H = AA, G_R = AA, G_V = AB)$)。这可以被称为“AB先验”。又一个简单的方法是假设每个突变都是二倍体和纯合的(即变体群体中的变体发生在两条染色体二者上, $G = (G_H = AA, G_R = AA, G_V = BB)$)。这可以被称为“BB先验”。又一种可能的简单方法是假设在恰好有一个突变体等位基因的情况下变体群体的基因型在突变区域具有预测的总拷贝数(即假设总拷贝数为3, $G = (G_H = AA, G_R = AA, G_V = AAB)$,即这导致在以上“主要拷贝数”方法中仅考虑 G_1)。这可以被称为“无合子(zygosity)先验”。这些方法在许多情况下可能过于简单,因为它们基本上考虑了单个可能基因型。

[0164] 另一种可能的方法是假设在有至少一个突变体等位基因的情况下变体群体的基因型在突变区域具有预测的总拷贝数,并且参考群体是AA或者拷贝数等于预测的总拷贝数且没有变体等位基因的基因型(具有等同的概率)。这可被称为“总拷贝数先验”,并直观地意指在基因座处的变体群体的基因型具有预测的总拷贝数并且可具有任意数目(>0)的突变体等位基因拷贝(即,假设总拷贝数为3,则在等同概率的情况下,可能的基因型为: $G_1 = (G_H = AA, G_R = AA, G_V = AAB)$, $G_2 = (G_H = AA, G_R = AA, G_V = ABB)$, $G_3 = (G_H = AA, G_R = AA, G_V = BBB)$, $G_4 = (G_H = AA, G_R = AAA, G_V = AAB)$,即,这基本上忽略了主要拷贝数和次要拷贝数的值,并考虑了所有可能的具有n个拷贝的基因型—导致与以上“主要拷贝数”方法相比,考虑了另外的基因型。可使用的又一种方法是“信任”来自拷贝数调用者的主要和次要等位基因的预测数,使得仅考虑具有与主要拷贝数或次要拷贝数相对应的突变体等位基因数的基因型。这可被称为“亲本”模式。例如,如果主要拷贝数=3,次要拷贝数=1,则这种方法将在等同概率的情况下考虑以下可能的基因型: $G_1 = (AA, AA, AAAB)$, $G_2 = (AA, AA, ABBB)$, $G_3 = (AA, AAAA, AAAB)$ (即在变体群体中有1或3个突变等位基因)。相比之下,“主要拷贝数”方法通过考虑1与预测的主要拷贝数之间的所有值来“信任”可能的主要拷贝的范围,而不是其绝对值。在

以上主要拷贝数=3、次要拷贝数=1的实例中,与“亲本”模式相比,这将导致考虑一种更多种基因型,即: $G_1 = (AA, AA, AAAB)$, $G_2 = (AA, AA, ABBB)$, $G_3 = (AA, AA, ABBB)$, $G_4 = (AA, AAAA, AAAB)$ 。因此,“主要拷贝数”方法在考虑来自拷贝数调用的另外的不确定性(与“亲本”方法相比)与不考虑太多不确定性(与“总拷贝数”方法相比)之间达成了良好的平衡。

[0165] 克隆性估计模型

[0166] 本节概述了用于鉴定普遍存在的突变的分层贝叶斯(hierarchical Bayesian)模型。设 Z 为伯努利(Bernoulli)变量,当突变普遍存在(假设是克隆性的)时其为一,否则为零。设 ρ 为突变普遍存在的先验概率。在以下实施例,这被设置为0.5。如上, ϕ 是样品中携带突变的癌细胞的比例。因此,该模型可以表示为:

[0167] $Z|\rho \sim \text{伯努利}(Z|\rho)$ (5)

[0168] $\phi|Z \sim \beta(\phi|\alpha=1, \beta=1)$ 对于 $Z=0$;

[0169] $\beta(\phi|\alpha, \beta=1)$ 对于 $Z=1$ (6)

[0170] $d_b, d|\pi, \phi, t \sim \text{Pr}(d, d_b|\pi, \phi, t)$ (7)

[0171] 其中 α 是 $\phi|Z=1$ 的分布中 >1 的参数。在以下实施例,这被设置为 $\alpha=99$ 。具有参数 $\alpha=99$ 和 $\beta=1$ 的 β 分布向1倾斜,捕获了克隆性突变应富集较高癌细胞分数 ϕ 的假设。参数 α 的其他值是可能的,但捕获该假设的值是优选的。如上所述,等式(7)中的概率由等式(3)/(3a)或(4)/(4a)给出。

[0172] 联合分布可以用以下等式(等式(8))表示:

[0173] $p(d_b, d, \phi, Z=z|\pi, t, \rho) = p(Z=z|\rho) \text{Pr}(d_b, d|\pi, \phi, t) p(\phi|Z=z)$ (8)

[0174] 对于一个样品或对于多个样品:

[0175] $p(d_b, d, \phi, Z=z|\pi, t, \rho) = p(Z=z|\rho) \prod_{s=1}^S \text{Pr}(d_b, d|\pi, \phi, t) p(\phi|Z=z)$ (8a)

[0176] 携带突变的癌细胞的比例(ϕ)是未知的。然而,可以表示:

$$\begin{aligned} p(d_b, d, Z=z|\pi, t, \rho) &= \int_0^1 P(d_b, d, \phi, Z=z|\pi, t, \rho) d\phi \\ [0177] &= P(Z=z|\rho) \int_0^1 \text{Pr}(d_b, d|\pi, \phi, t) p(\phi|Z=z) d\phi \end{aligned} \quad (9)$$

[0178] 对于一个样品或对于多个样品:

[0179] $p(d_b, d, Z=z|\pi, t, \rho) = P(Z=z|\rho) \prod_{s=1}^S \int_0^1 \text{Pr}(d_b, d|\pi, \phi, t) p(\phi|Z=z) d\phi$ (9a)。

[0180] 该量 $\prod_{s=1}^S \int_0^1 \text{Pr}(d_b, d|\pi, \phi, t) p(\phi|Z=z) d\phi$ 可称为 Ψ_z (即,如果突变是非克隆性的和如果突变是克隆性的,则 Ψ_0 和 Ψ_1 分别是指数据的可能性)。当对于 $z=0$, $P(Z=z|\rho) = (1-\rho)$ (即 $Z=0$ 的先验概率,即突变被分类为非克隆性,给出突变是克隆性的先验概率 ρ 等于突变不是克隆性的先验概率),以及对于 $z=1$, $P(Z=z|\rho) = \rho$ (即 $Z=1$ 的先验概率,即突变被分类为克隆性,给出突变是克隆性的先验概率 ρ 等于突变是克隆性的先验概率)时,推论出:

[0181] 对于多个样品(对于单个样品,没有样品内乘积)

$$\begin{aligned}
 p(d_b, d | \pi, t, \rho) &= \sum_{z=0}^1 p(d_b, d, Z = z | \pi, t, \rho) \\
 [0182] \quad &= (1 - \rho) \prod_{s=1}^S \int_0^1 \Pr(d_b, d | \pi, \phi, t) p(\phi | Z = 0) d\phi + \\
 &\quad \rho \prod_{s=1}^S \int_0^1 \Pr(d_b, d | \pi, \phi, t) p(\phi | Z = 1) d\phi \quad (10)
 \end{aligned}$$

[0183] 最终,希望估计的量为突变是克隆性的概率($Z=1$ 的概率),鉴于观察到的读出(d_b, d)、基因型先验(π)、肿瘤分数估计值(t)、和突变是克隆性的先验概率(ρ ,即希望估计 $P(Z=1 | d_b, d, \pi, t, \rho)$)。鉴于以上,这可以表示为:

$$[0184] \quad p(Z = z | d_b, d, \pi, t, \rho) = \frac{p(d_b, d, Z = z | \pi, t, \rho)}{p(d_b, d | \pi, t, \rho)} \quad (11)$$

[0185] 其中 $p(d_b, d | \pi, t, \rho)$ 由等式(10)给出,并且 $p(d_b, d, Z = z | \pi, t, \rho)$ 由等式(9)/(9a)给出。因此,对于 $Z=1$,等式(11)可以写成以下等式(11a):

$$\begin{aligned}
 [0186] \quad &p(Z = 1 | d_b, d, \pi, t, \rho) = \\
 &\frac{\rho \prod_{s=1}^S \int_0^1 \Pr(d_b, d | \pi, \phi, t) p(\phi | Z = 1) d\phi}{(1 - \rho) \prod_{s=1}^S \int_0^1 \Pr(d_b, d | \pi, \phi, t) p(\phi | Z = 0) d\phi + \rho \prod_{s=1}^S \int_0^1 \Pr(d_b, d | \pi, \phi, t) p(\phi | Z = 1) d\phi} \quad (11a)
 \end{aligned}$$

[0187] 其中 ρ 是参数(在以上实施例中设置为0.5), $p(\phi | Z = z)$ 由等式(6)中的 β 分布给出,并且 $\Pr(d_b, d | \pi, \phi, t)$ 由等式(3)/(4)(一个联合基因型)或(3a)/(4a)(具有先验概率 π 的多个候选联合基因型)给出。

[0188] 因此,对 $z=1$ 的等式(11)(即等式(11a))进行估计给出了突变普遍存在的概率(即,鉴于一个或更多个可用样品,假设是克隆性的)。这需要评价 S -维积分(在等式(9)、(10)中,每个样品的积分),这可以使用已知的数值积分有效地进行。本领域已知的任何数值积分算法都可用于此目的。例如,可以使用网格近似。这是有利地简单的,并且足以考虑存在单个参数(ϕ)用于在其上进行积分。

[0189] 鉴于可用的数据,这提供了突变是克隆性的概率的估计值,其可以有效地计算、是易于解释的(鉴于使用明确的清晰假设的严格统计学模型),可以在没有手动输入的情况下获得任何突变、独立于所分析的任何其他突变、可以严格地包括关于突变的现有知识、并且可以用于客观地和自动地优先突变列表(具有伴随的概率)以用于测试和/或使用。

[0190] 对拷贝数预测中的不确定性的计算

[0191] 虽然上述模型已经呈现出许多优点,但是可以通过考虑模型中使用的拷贝数估计值的预测中的不确定性来进一步增强该模型。实际上,上述模型假设拷贝数(例如,用于引出基因型先验的主要/次要/总/拷贝数)是准确预测的。在实践中,这些值中可存在一些不确定性。实际上,肿瘤的等位基因特异性拷贝数分析的问题是复杂的,并且为此已经提出了许多解。一种通常使用的方法是ASCAT(肿瘤的等位基因特异性拷贝数分析, Van Loo et al., 2010),该方法在解释批量拷贝数谱方面考虑了肿瘤细胞的非整倍性和非异常细胞浸润二者,并输出了估计的等位基因特异性拷贝数谱和伴随的肿瘤纯度估计值。总之,基于相关等位基因特异性拷贝数调用应尽可能接近种系杂合单核苷酸多态性(single nucleotide polymorphism, SNP)的非负整数的假设,ASCAT评价了肿瘤倍性和肿瘤分数的多种可能的组合。然后报道了被认为是最优的解(肿瘤和样品的正常部分的估计的肿瘤倍性、肿瘤纯度和等位基因特异性拷贝数调用)以及其拟合优度(基于以上假设)。

[0192] 以上提供的模型可通过修改 π 以包含来自每个所预测拷贝数状态(例如,每个所提出的包含主要和次要拷贝数状态的解)的基因型的条目进行调整以适应多个拷贝数的解及其不确定性,通过与该状态相关的概率进行加权。另外,由于肿瘤纯度估计值可以与这些拷贝数状态一起估计(如例如当使用类似ASCAT的方法时就是这种情况),因此也可以考虑相关肿瘤纯度估计值。注意到,当例如肿瘤纯度被单独估计或测量并且与拷贝数状态估计值不固有地相关时,这可能是不必要的。然而,为了通用性起见,让我们假设有C个可能的拷贝数/肿瘤含量状态的组(例如C个可能的 $c_{\text{主要}}$ 、 $c_{\text{次要}}$ 和 t 的估计值组)。设 π_c 是向量,其中每个条目是每个可能的这样的估计值组的概率。对于每个状态C,可以计算可能基因型的向量 π_{CG} ,如上所述。因此,通过将 π_{CG} 乘以 π_c 中状态C的条目,可以获得最终的基因型向量。这产生了以下稍微修改的等式:

$$[0193] \quad P(d, d_b | \pi, \varphi) = \sum_i \pi_i \text{二项式}(d_b | d, \xi(G_i, \varphi, t_i)) \quad (3b)$$

$$[0194] \quad P(d, d_b | \pi, \varphi, \gamma) = \sum_i \pi_i \beta \text{二项式}(d_b | d, \xi(G_i, \varphi, t_i), Y) \quad (4b)$$

[0195] 其中肿瘤含量 t_i 目前可取决于特定的状态(并且 π_i 是通过将 π_{CG} 乘以 π_c 中状态C的条目而获得的向量 π 的元素)。这些新的密度可以在以上相关等式中被替换。特别地,随后所解决的问题可以表示为求解等式(11a),其中 $\Pr(d_b, d | \pi, \varphi, t)$ 由等式(3b)或等式(4b)给出。如上所述, t_i 、 $c_{\text{主要}}$ 、 $c_{\text{次要}}$ (以及因此根据所使用模型的相容的 π_{CG})和 π_c 的值被提供作为用于进行肿瘤的等位基因特异性拷贝数分析的许多方法的输出,包括但不限于ASCAT。为避免任何疑问,任何产生等位基因特异性拷贝数状态估计值(通常与肿瘤纯度估计值相关)以及置信度或其他度量(可用于对多个解相对于彼此进行加权)的方法都可以用于此目的。

[0196] 实施方式

[0197] 本文中所述的方法可以使用本领域已知的任何编程语言来实施。在以下实施例中,使用了实施上述方法的Python脚本。对于每个突变,将这作为输入:突变标识符、样品标识符、与突变位置处的参考等位基因匹配的读出数目的计数、与突变位置处的替代等位基因匹配的读出数目的计数;以及对于一个或多个拷贝数解的每个,将这作为输入:特定拷贝数解的与突变重叠的主要拷贝数(针对肿瘤)、特定拷贝数解的与突变重叠的次要拷贝数(针对肿瘤)、突变处正常细胞的拷贝数(针对常染色体可设置默认值=2,针对男性对象的性染色体可设置默认值=1)、以及特定拷贝数解的肿瘤纯度值(这也可作为例如ASCAT的输出获得,或可单独获得)。对于特定拷贝数解,肿瘤群体的与突变重叠的主要和次要拷贝数可以直接从ASCAT获得(例如使用ascatNgs, Raine et al., 2016)、或者来源于例如ASCAT的输出,例如使用与突变重叠的拷贝数区段的平均B等位基因频率、与突变重叠的拷贝数区段的log R值、以及解的倍性。例如,使用

$$[0198] \quad \hat{n}_A = (\rho - 1 + 2^r(1 - b)(2(1 - \rho) + \rho\psi))/\rho \text{ 和 } \hat{n}_B = (\rho - 1 + 2^r b(2(1 - \rho) + \rho\psi))/\rho \text{ 可}$$

将位置i处肿瘤的等位基因特异性拷贝数估计值($\hat{n}_A, \hat{n}_B$)表示为以下的函数:位置i处的log R值 r 、位置i处的B等位基因分数值 b 、倍性估计值 Ψ 、肿瘤细胞分数估计值 ρ 和平台依赖的“技术”参数 t (对于下一代测序数据例如WES,其可以设置为 $t=1$)。除可根据对象的性别处理的性染色体上的突变之外,正常群体中的主要和次要拷贝数可以假定为1和1。在提供多个拷贝数解的情况下,可以任选地提供每个解的概率(这也可以从例如证明解的负log可能性的ASCAT的输出中获得)。如果这没有被提供,则多个解中的所有解都可以被视为等同

可能并且接受等同权重。脚本产生输出：突变标识符和突变普遍存在的后验概率。

[0199] 在以下实施例中,无论何时估计拷贝数解,都使用ASCAT (Van Loo et al.,2010) 进行。

[0200] 结果

[0201] 合成数据

[0202] 首先使用合成数据来测试上述方法,所述合成数据是使用上述群体模型模拟的,参考如由等式 (3) (二项式模型) 和 (4) (β -二项式模型) 捕获的图4。对于所有合成数据集,模拟了总共1000个突变。覆盖深度由泊松 (Poisson) 分布模拟,其中平均值为50、100、1000,其与实际覆盖方案相对应。随机模拟拷贝数谱,其中最大总拷贝为2或6。这些设置对应于基因组稳定的基因组 (最大总拷贝数为2) 和高度非整倍体基因组 (最大总拷贝数为6)。模拟了包含一个、两个或四个样品的数据集。所有实施例的肿瘤含量都固定在1.0。

[0203] 图5A示出了当根据二项式模型来模拟数据并且使用二项式模型运行上述方法的结果。对于拷贝数为2的运行,无论深度或样品数如何,性能都近乎完美 (曲线下面积, $AUC = 0.997$ 至 1.0 , 参见下表1)。对于拷贝数为6的运行,随着深度和样品数的提高,性能显著改善 ($AUC = 0.853$ 至 1.0 , 参见下表1)。

[0204] 表1. 使用二项式分布模型进行模拟的合成数据结果。BB-200= β 二项式, $\gamma = 200$ 。

分析中使用的模型	最大拷贝数	平均深度	S (样品数)	AUC
BB-200	2	50	1	0.9975431554
二项式	2	50	1	0.9975431554
BB-200	2	100	1	0.9981312525

[0206]

二项式	2	100	1	0.9981312525
BB-200	2	1000	1	1
二项式	2	1000	1	1
BB-200	2	50	2	1
二项式	2	50	2	1
BB-200	2	100	2	1
二项式	2	100	2	1
BB-200	2	1000	2	1
二项式	2	1000	2	1
BB-200	2	50	4	1
二项式	2	50	4	1
BB-200	2	100	4	1
二项式	2	100	4	1
BB-200	2	1000	4	1
二项式	2	1000	4	1
BB-200	6	50	1	0.852432188
二项式	6	50	1	0.852540223
BB-200	6	100	1	0.9012136777
二项式	6	100	1	0.90078165
BB-200	6	1000	1	0.961997085
二项式	6	1000	1	0.9687399898
BB-200	6	50	2	0.9570949962
二项式	6	50	2	0.9567547662
BB-200	6	100	2	0.9732051405
二项式	6	100	2	0.9741773894
BB-200	6	1000	2	0.9962227762
二项式	6	1000	2	0.9977032559
BB-200	6	50	4	0.9923878205
二项式	6	50	4	0.9924078526
BB-200	6	100	4	0.9974953588
二项式	6	100	4	0.9978354459
BB-200	6	1000	4	0.9995103528
二项式	6	1000	4	0.999959865

[0207] 图5B示出了通过根据精度为100的 β -二项式分布进行模拟并且在使用二项式散发(emission)模型运行上述模型的情况下进行的相同分析。这模拟了其中观察到的数据比所预期的显著更可变的情况。对于拷贝数为2的模拟,结果保持接近完美(AUC 0.997至1.0,表2)。然而,结果(虽然保持良好)不如先前针对拷贝数为6的模拟的分析那样好(AUC 0.846至0.996,表2)。提高覆盖的深度没有如之前一样显著地改善性能。相比之下,提高样品数具有显著作用。

[0208] 图5C示出了当使用精度为100的 β -二项式分布来模拟数据并且用使用精度为200的 β -二项式散发模型的上述方法来分析数据时的结果。性能与先前用使用二项式散发的上述方法进行的分析类似(表2)。与对过度分散的数据使用二项式模型相比,对相同的过度分散的数据使用 β -二项式模型表现稍好,尤其是在较低的测序深度和/或较高的拷贝数下。然而,在该实施例,性能的改善并不显著,并且因此在两种情况下都可以有利地使用较简单的二项式模型(由于不需要设置精度超参数,其参数化更简单)。

[0209] 表2.使用精度为100的 β -二项式散发分布进行模拟的合成数据结果。BB= β 二项式

[0210]

分析中使用的模型	最大拷贝数	平均深度	S (样品数)	AUC
BB-200	2	50	1	0.9975810324
二项式	2	50	1	0.9975810324

BB-200	2	100	1	0.9935050647
二项式	2	100	1	0.9935050647
BB-200	2	1000	1	0.9983754737
二项式	2	1000	1	0.9983754737
BB-200	2	50	2	1
二项式	2	50	2	1
BB-200	2	100	2	0.9999919783
二项式	2	100	2	0.9999919783
BB-200	2	1000	2	1
二项式	2	1000	2	1
BB-200	2	50	4	1
二项式	2	50	4	1
BB-200	2	100	4	1
二项式	2	100	4	1
BB-200	2	1000	4	1
二项式	2	1000	4	1
BB-200	6	50	1	0.8460939595
二项式	6	50	1	0.8455535942
BB-200	6	100	1	0.8628064402
二项式	6	100	1	0.859359738
BB-200	6	1000	1	0.908828000
二项式	6	1000	1	0.899468000
BB-200	6	50	2	0.955679568
二项式	6	50	2	0.9547434743
BB-200	6	100	2	0.9603051358
二项式	6	100	2	0.9590724257
BB-200	6	1000	2	0.9685954382
二项式	6	1000	2	0.9525490196
BB-200	6	50	4	0.9944876472
二项式	6	50	4	0.9943236367
BB-200	6	100	4	0.9937650015
二项式	6	100	4	0.9932320614
BB-200	6	1000	4	0.9990039641
二项式	6	1000	4	0.9961278606

[0211] 细胞系混合数据

[0212] 接下来使用通过混合细胞系产生的数据对该方法进行测试 (Farahani et al., 2017)。数据从两个实验中获得, 其中: 1) 混合了二倍体细胞系和 2) 混合了非整倍体细胞系。如表 3 中所述, 以不同的比例形成了多种混合物。每个实验中存在三类突变: 1) 细胞系 A 特有的突变; 2) 细胞系 B 特有的突变和 3) 由两种细胞系共有的突变。出于测试本文中所述的方法

的目的,将共有的突变视为普遍存在的突变,并将细胞系特异性突变视为亚克隆性突变。等位基因计数数据和拷贝数信息从Farahani et al的补充表S1A和S1B中提取,补充表S1A和S1B从出版商的网站下载。注意到,不可以用该数据来解释拷贝数的不确定性,因为针对每个突变仅给出了预测的主要/次要拷贝。

[0213] 将使用二项式发散模型运行的上述方法与PyClone (Roth et al.,2014) 版本0.13.1进行比较,所述PyClone版本0.13.1运行10,000次迭代,其中1,000次迭代因burning而被丢弃。为了用PyClone鉴定普遍存在的突变,首先鉴定所有样品中具有最高CCF的簇。该簇中的所有突变均被视为普遍存在的,并且所有其他突变均被视为亚克隆性的。注意到,这种策略没有为普遍性(ubiquity)的预测分配置信度值,因此PyClone结果被绘制为ROC曲线上的单点。相比之下,本文中所述的方法为每个突变分配了普遍性的概率,因此可以绘制曲线。因此,也可根据所期望的精度和召回率的平衡来调整用于将突变视为克隆性/亚克隆性的阈值。

[0214] 表3. 细胞系混合比例

[0215]

混合ID	细胞系A	细胞系B
2	0	1
14	0	1
7	0.1	0.9
19	0.1	0.9
6	0.25	0.75
18	0.25	0.75
5	0.5	0.5
17	0.5	0.5
4	0.75	0.25
16	0.75	0.25
3	0.9	0.1
15	0.9	0.1
1	1	0
13	1	0

[0216] 图6A示出了二倍体细胞系混合物的结果。图6A左侧的图来自仅使用单个样品的运行。对于使用其中90%细胞系A和10%细胞系B的混合物S3的两种方法来说,其是最差的。这容易理解,因为来自细胞系A的突变具有接近共有突变的CCF。随着混合物变得更平衡并且细胞系中突变的CCF远离1,性能提高。图6A的右侧示出了来自使用多个样品的运行的结果。图6B示出了来自非整倍体细胞系的混合物的结果。结果通常与二倍体的情况一致,但如所预期的针对两种方法的性能一致较差。

[0217] 该数据表明,如所预期的,随着添加更多样品,性能可能提高。在此分析的模拟数据和人工样品的特定设置中,本文中所述的方法和PyClone的性能对于一些样品是相似的,但是在一些多样品实例中,本文中所述的方法优于PyClone。

[0218] 对每种方法的非正式计时(timing)研究显示,本文中所述的方法每次分析花费约10秒(其中分析包括在患者中确定经鉴定的所有突变的克隆性的可能性)。针对相同的分析

PyClone花费显著更长的时间,大约300至3000秒,这取决于所执行的MCMC迭代次数。因此,即使当患者中经鉴定的所有突变均在同一分析中评估(这不是本方法的要求,因为这些评估是完全独立的)时,本方法也是现有技术方法的至少30倍快。如果对突变的评估是以并联而不是串联的方式进行的(这在使用现有技术方法时是不可能的),差异可能甚至更大。

[0219] 因此,该数据表明,本文中所述的方法在多样品情况下能够至少匹配并优于比较的现有技术方法。这些具有大的临床和实践相关性,因为在可能的情况下获得多个样品被认为在使用可以利用其的方法(例如本文中所述的方法)时可能提高鉴定真正克隆性变体的可能性。与本文中所述的方法相比,PyClone不具有提供突变特异性预测、为每个这样的预测提供置信度估计值、或考虑拷贝数解中的不确定性的能力,并且需要手动输入以及显著更长的计算时间来分析相同的数据。

[0220] 因此,与现有技术的方法相比,本文中所述的方法具有数个优点,其在实际临床背景中特别相关,除了其中现有技术的方法可以以相似的性能使用的特定情况之外(例如以上模拟的和人工的样品)。例如,至少在处理 and 选择单个拷贝数解时需要大量的手动输入意味着在没有这样的专家输入的情况下,本文中所述的方法即使在单个样品的情况下也比现有技术的方法运行的显著更好。此外,即使在专家手动输入可用的情况下,本文中所述的方法在现实生活情境(即,使用来自患者的实际肿瘤样品)下也比现有技术的方法运行的显著更好,其中在现实生活情境下鉴定正确的拷贝数解显著比在以上研究的人工、控制情况下更易出错。

[0221] TRACERx数据

[0222] 本文中所述的方法在来自CRUK TRACERx研究(<http://tracex.co.uk/>)的数据上进行了进一步测试。该数据先前已经使用TRACERx流水线(pipeline)进行了分析(参见McGrath et al., 2016),并且基于PyClone方法的克隆性调用是可用的。另外,倍性/纯度解的ASCAT数据可用于所有样品。图7A示出了数据概述(基于可用的克隆性调用,使用PyClone获得)。因此,考虑到用于推导主要/次要拷贝数的所选倍性/纯度解中的不确定性,可以运行本方法。

[0223] 图7B示出了当1) 考虑所有ASCAT纯度/倍性解、和2) 考虑最大可能性ASCAT解时,从本文中所述的方法预测的普遍存在概率的分布。我们看到,当考虑纯度/倍性不确定性时,本方法系统地普遍存在的突变分配了较高的概率。图7C示出了比较两种方法的ROC曲线。考虑到倍性和纯度的不确定性显著改善了本方法的性能,其中AUC从0.76至0.91。这种分析的一个警告是,经处理的PyClone结果被假定为是真实的(因为与以上分析的模拟数据和人工样品相反,无法获得真实情况)。先前章节表明,针对真实肿瘤样品和多样品情况,本文中所述的方法优于PyClone,因此本文中所述方法的实际性能可能比在此报道的更好。

[0224] 与现有技术方法的运行时间比较

[0225] 本文中所述的方法与PyClone的最新版本(PyClone-VI, Gills&Roth, BMC Bioinformatics volume 21, Article number:571(2020))一起正式计时。该分析的结果在图8中示出(其中“ACE”表示如本文中所述的方法)。该版本的PyClone已显著优化并且因此比上述非正式计时研究中使用的原始版本更快。然而,因为其必须一起分析在癌症中鉴定的所有肿瘤特异性突变,所以运行时间随着突变数目提高而显著提高,并且这不能通过并行化来补救。相比之下,因为本公开内容的方法独立地分析每个突变,所以是完全可并行

的。因此,在其中用于单一癌症的分析可以在多个CPU上运行的所有情况下,与本方法相关的运行时间显著较低(参见图8,其示出了2、3或4个CPU的结果)。作为一个实例,对于具有200至1000个突变(这是临床上现实的范围)的单个患者,使用本方法鉴定每个突变的概率或克隆性的时间可以低至12.8秒至36.5秒(分别针对200和1000个突变)。相比之下,运行PyClone-VI的时间为16.8至99.1秒(分别针对200和1000个突变)。另外,该时间估计不包括对由PyClone产生的簇结构进行手动分析以选择假定为克隆性的突变簇,以及对所选簇内的突变进行优先的时间。相比之下,本文中所述的方法产生了每个突变为克隆性的可直接解释且独立的概率,可直接用于对突变进行优先(即,在没有关于克隆性的进一步评估步骤的情况下)。这代表着资源的显著节约以及工艺易处理性的提高。

[0226] 讨论

[0227] 这些实施例表明了新模型的开发和评估,该模型用于在患者中鉴定癌细胞群中普遍存在的突变,即克隆性突变。这项工作具有数项创新,其与先前的解相比显著改善了性能。特别地,我们将该问题重铸为分类问题并提供预测的概率。这提供了对突变进行排序的能力,而不是先前的返回二进制答案的方法。此外,我们示出了如何解释关于重叠拷贝数的不确定性。在真实世界的的数据中,这导致在鉴定克隆性突变的精度方面性能显著提高。这在实践中意味着通过本文中所述的方法鉴定的克隆性突变比使用现有技术方法鉴定的突变更有可能是真正克隆性的,导致更高的验证率和最终更高的临床成功率。本文中所述的方法的其他益处包括在不改变其他突变评分的情况下添加新突变的能力。此外,本方法与PyClone和可能的其他基于簇的方法相比快数个数量级。

[0228] 参考文献

[0229] Adalsteinsson, V.A., Ha, G., Freeman, S.S. et al. Scalable whole-exome sequencing of cell-free DNA reveals high concordance with metastatic tumors. *Nat Commun* 8, 1324 (2017).

[0230] Bulik-Sullivan B, Busby J, Palmer CD, Davis MJ, Murphy T, Clark A, Busby M, Duke F, Yang A, Young L, Ojo NC, Caldwell K, Abhyankar J, Boucher T, Hart MG, Makarov V, Montpreville VT, Mercier O, Chan TA, Scagliotti G, Bironzo P, Novello S, Karachaliou N, Rosell R, Anderson I, Gabrail N, Hrom J, Limvarapuss C, Choquette K, Spira A, Rousseau R, Voong C, Rizvi NA, Fadel E, Frattini M, Jooss K, Skoberne M, Francis J, Yelensky R. Deep learning using tumor HLA peptide mass spectrometry datasets improves neoantigen identification. *Nat Biotechnol*. 2018 Dec 17.

[0231] Carter SL, Cibulskis K, Helman E, McKenna A, Shen H, Zack T, Laird PW, Onofrio RC, Winckler W, Weir BA, Beroukheim R, Pellman D, Levine DA, Lander ES, Meyerson M, Getz G. Absolute quantification of somatic DNA alterations in human cancer. *Nat Biotechnol*. 2012 May; 30(5):413-21.

[0232] Hossein Ferahani, Camila PE de Souza, Raewyn Billings, Damian Yap, Karey Shumansky, Adrian Wan, Daniel Lai, Anne-Marie Mes-Masson, Samuel Aparicio, and Sohrab P Shah. Engineered in-vitro cell line mixtures and robust evaluation of computational methods for clonal decomposition and longitudinal dynamics in cancer. *Scientific Reports*, 7(1):13467, 2017.

- [0233] Vanessa Jurtz, Sinu Paul, Massimo Andreatta, Paolo Marcatili, Bjoern Peters and Morten Nielsen. NetMHCpan-4.0: Improved Peptide-MHC Class I Interaction Predictions Integrating Eluted Ligand and Peptide Binding Affinity Data. *J Immunol* November 1, 2017, 199(9):3360-3368.
- [0234] Langmead, B., Trapnell, C., Pop, M. et al. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* 10, R25 (2009).
- [0235] Lundegaard C, Lamberth K, Herndahl M, Buus S, Lund O, Nielsen M. NetMHC-3.0: accurate web accessible predictions of human, mouse and monkey MHC class I affinities for peptides of length 8-11. *Nucleic Acids Res.* 2008 Jul 1; 36(Web Server issue):W509-12.
- [0236] Nicholas McGranahan, Francesco Favero, Elza C de Bruin, Nicolai Juul Birkbak, Zoltan Szallasi, and Charles Swanton. Clonal status of actionable driver events and the timing of mutational processes in cancer evolution. *Science translational medicine*, 7(283):283re54-283ra54, 2015.
- [0237] McGranahan, N., Furness, A. J., Rosenthal, R., Ramskov, S., Lyngaa, R., Saini, S. K., Jamal-Hanjani, M., Wilson, G. A., Birkbak, N. J., Hiley, C. T., Watkins, T. B., Shafi, S., Murugaesu, N., Mitter, R., Akarca, A. U., Linares, J., Marafioti, T., Hanry, J. Y., Van Allen, E. M., Miao, D., ... Swanton, C. (2016). Clonal neoantigens elicit T cell immunoreactivity and sensitivity to immune checkpoint blockade. *Science* (New York, N.Y.), 351(6280), 1463-1469.
- [0238] Timothy J. O'Donnell, Alex Rubinsteyn, Maria Bonsack, Angelika B. Riemer, Uri Laserson, Jeff Hammerbacher. MHCflurry: Open-Source Class I MHC Binding Affinity Prediction. *Cell Systems* Vol. 7, Issue 1, 129-132, July 25, 2018.
- [0239] Andrew Roth, Jaswinder Khattra, Damian Yap, Adrian Wan, Emma Laks, Justine Biele, Gavin Ha, Samuel Aparicio, Alexandre Bouchard-Côté, and Sohrab P Shah. PyClone: statistical inference of clonal population structure in cancer. *Nature methods*, 11(4):396, 2014.
- [0240] Russell Schwartz and Alejandro A Schäffer. The evolution of tumour phylogenetics: principles and practice. *Nature Reviews Genetics*, 18(4):213, 2017.
- [0241] Van Loo P, Nordgard SH, Lingjaerde OC, Russnes HG, Rye IH, Sun W, Weigman VJ, Marynen P, Zetterberg A, Naume B, Perou CM, Børresen Dale AL, Kristensen VN. Allele-specific copy number analysis of tumors. *Proc Natl Acad Sci USA*. 2010 Sep 28; 107(39):16910-5.
- [0242] Landau DA, Carter SL, Stojanov P, McKenna A, Stevenson K, Lawrence MS, Sougnez C, Stewart C, Sivachenko A, Wang L, Wan Y, Zhang W, Shukla SA, Vartanov A, Fernandes SM, Saksene G, Cibulskis K, Tesar B, Gabriel S, Hacohen N, Meyerson M, Lander ES, Neuberg D, Brown JR, Getz G, Wu CJ. Evolution and impact of subclonal mutations in chronic lymphocytic leukemia. *Cell*. 2013 Feb 14; 152(4):714-26. doi:

10.1016/j.cell.2013.01.019.

[0243] Raine KM, Van Loo P, Wedge DC, Jones D, Menzies A, Butler AP, Teague JW, Tarpey P, Nik-Zainel S, Campbell PJ. *ascatNgs: Identifying Somatically Acquired Copy-Number Alterations from Whole-Genome Sequencing Data*. *Curr Protoc Bioinformatics*. 2016 Dec 8;56:15.9.1-15.9.17. doi:10.1002/cpbi.17.

[0244] Litchfield K, Reading JL, Puttick C, Thakkar K, Abbosh C, Bentham R, Watkins TBK, Rosenthal R, Biswas D, Rowan A, Lim E, Al Bakir M, Turati V, Guerra-Assunção JA, Conde L, Furness AJS, Saini SK, Hadrup SR, Herrero J, Lee SH, Van Loo P, Enver T, Larkin J, Hellmann MD, Turajlic S, Quezada SA, McGranahan N, Swanton C. *Meta-analysis of tumor-and T cell-intrinsic mechanisms of sensitization to checkpoint inhibition*. *Cell*. 2021 Feb 4;184(3):596-614.e14.

[0245] 本文中引用的所有参考文献均通过引用整体并入本文并且用于所有目的,其程度如同每个单独的出版物或专利或专利申请被具体地和单独地指示为通过引用整体并入。

[0246] 本文中描述的具体实施方案是通过示例而非通过限制来提供的。在不脱离所述技术的范围和精神的情况下,所描述的技术的组合物、方法和用途的各种修改和变化对于本领域技术人员将是显而易见的。本文中包括的任何小标题仅是为了方便起见,而不应被解释为以任何方式限制本公开内容。

[0247] 本文中描述的任何实施方案的方法均可作为计算机程序或作为计算机程序产品或携带计算机程序的计算机可读介质来提供,所述计算机程序在计算机上运行时被安排以执行上述方法。

[0248] 除非上下文中另外指示,否则对上述特征的描述和限定不限于本发明的任何特定方面或实施方案,并且同等地适用于所描述的所有方面和实施方案。

[0249] 在整个说明书和权利要求书中,除非上下文另外明确指出,否则以下术语采用与本文明确相关的含义。如本文所用的短语“在一个实施方案中”不一定是指相同的实施方案,但是其可以指相同的实施方案。此外,如本文所用的短语“在另一实施方案中”不一定是指不同的实施方案,但是其可以指不同的实施方案。因此,如下所述,在不脱离本发明的范围或精神的情况下,可以容易地组合本发明的各种实施方案。

[0250] 必须注意的是,除非上下文另外明确指出,否则说明书和所附权利要求书中使用的没有数量词修饰的名词表示一个/种或更多个/种。范围可在本文中表示为从“约”一个特定值和/或至“约”另一特定值。当表示这样的范围时,另一个实施方案包括从一个特定值和/或至另一个特定值。类似地,当通过使用在前的“约”将值表示为近似值时,将理解的是,特定值形成另一个实施方案。与数值相关的术语“约”是任选的,并且意指例如 $\pm 10\%$ 。

[0251] 在本说明书通篇(包括所附的权利要求书),除非上下文另有要求,否则词语“包含”和“包括”以及变化形式将被理解为暗示包含所指出的整体或步骤或者整体或步骤的组,但是不排除任何其他整体或步骤或者整体或步骤的组。

[0252] 除非上下文另外指出,否则本发明的另一些方面和实施方案提供了由术语“由……组成”或“基本上由……组成”替换术语“包含/包括/含有”的上述方面和实施方案。

[0253] 本文中使用的“和/或”被认为具体公开了在具有或不具有另一个的情况下的两个指出的特征或组件中的每一个。例如“A和/或B”被认为具体公开了以下中的每一者:(i) A、

(ii)B、以及 (iii)A和B,就如同每一者在本文中单独陈述一样。

[0254] 在前述说明书中、或在所附权利要求书中、或在附图中公开的以其特定形式或者在用于进行所公开的功能的方式、或者用于获得所公开的结果的方法或工艺方面表示的特征可以适当地分别或以这样的特征的任何组合用于以其多种形式实现本发明。

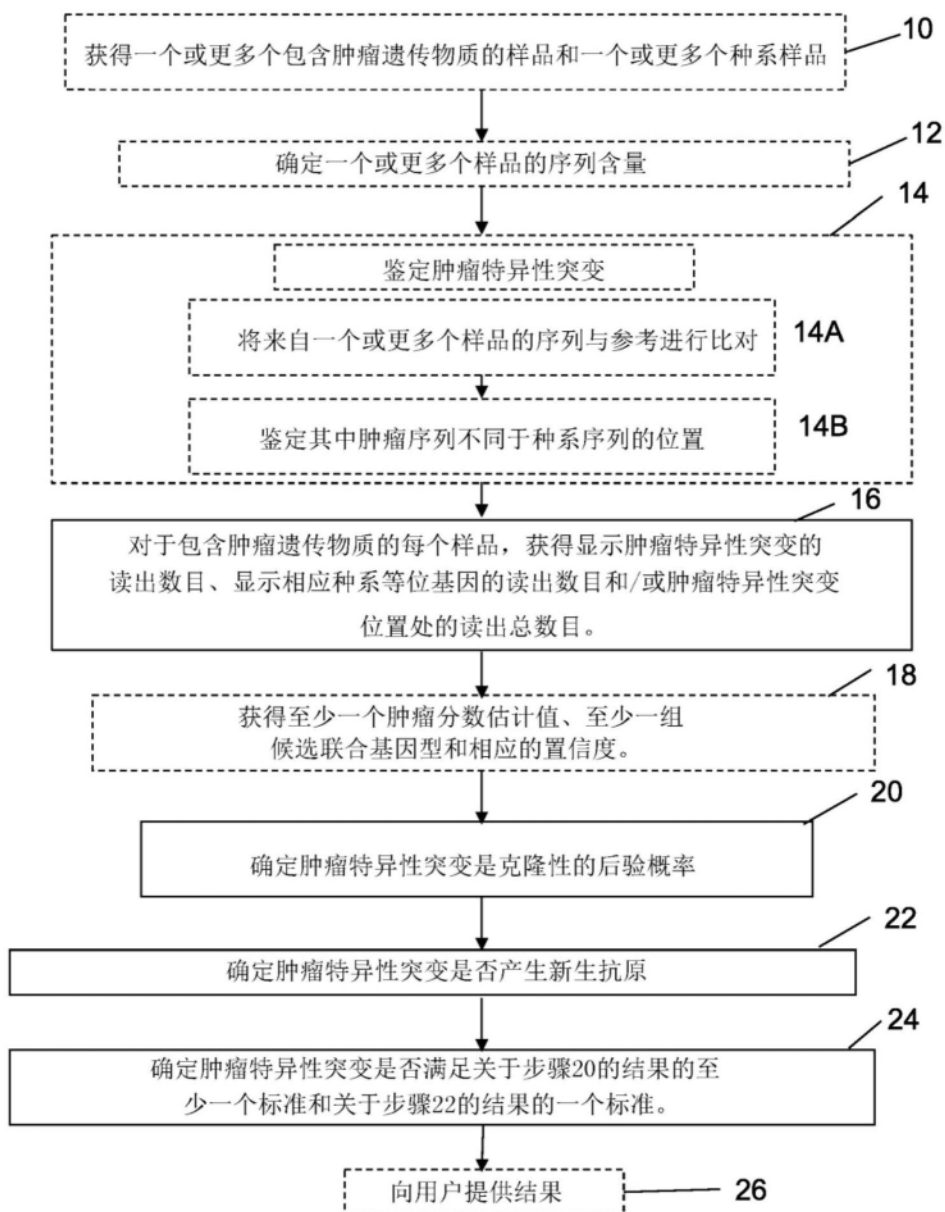


图1

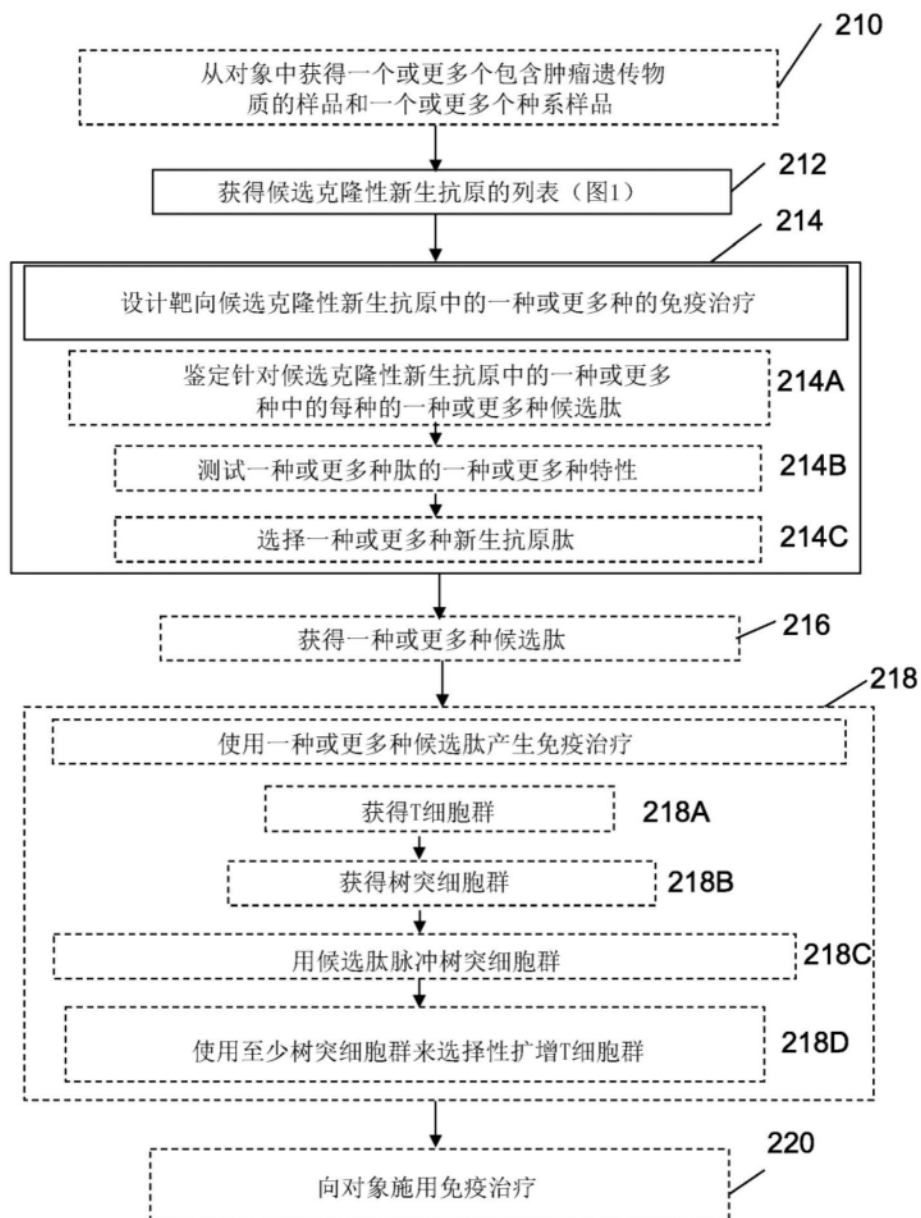


图2

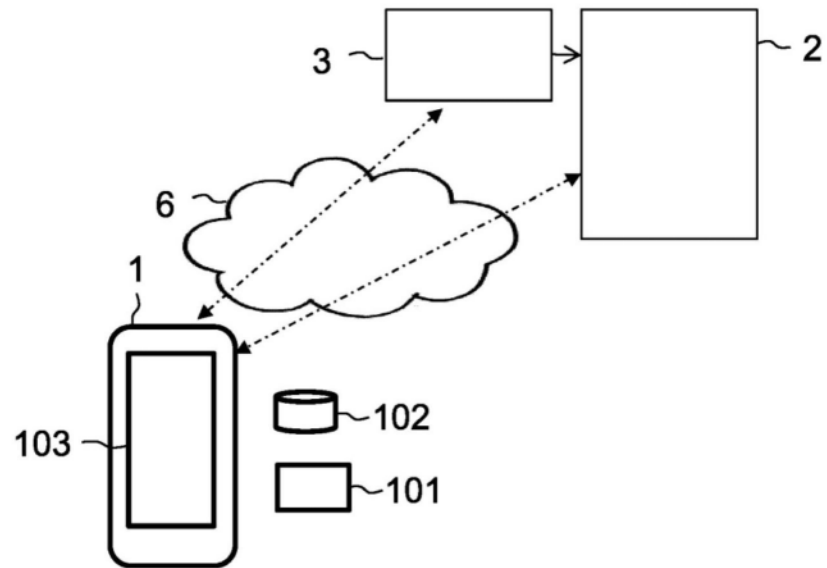


图3

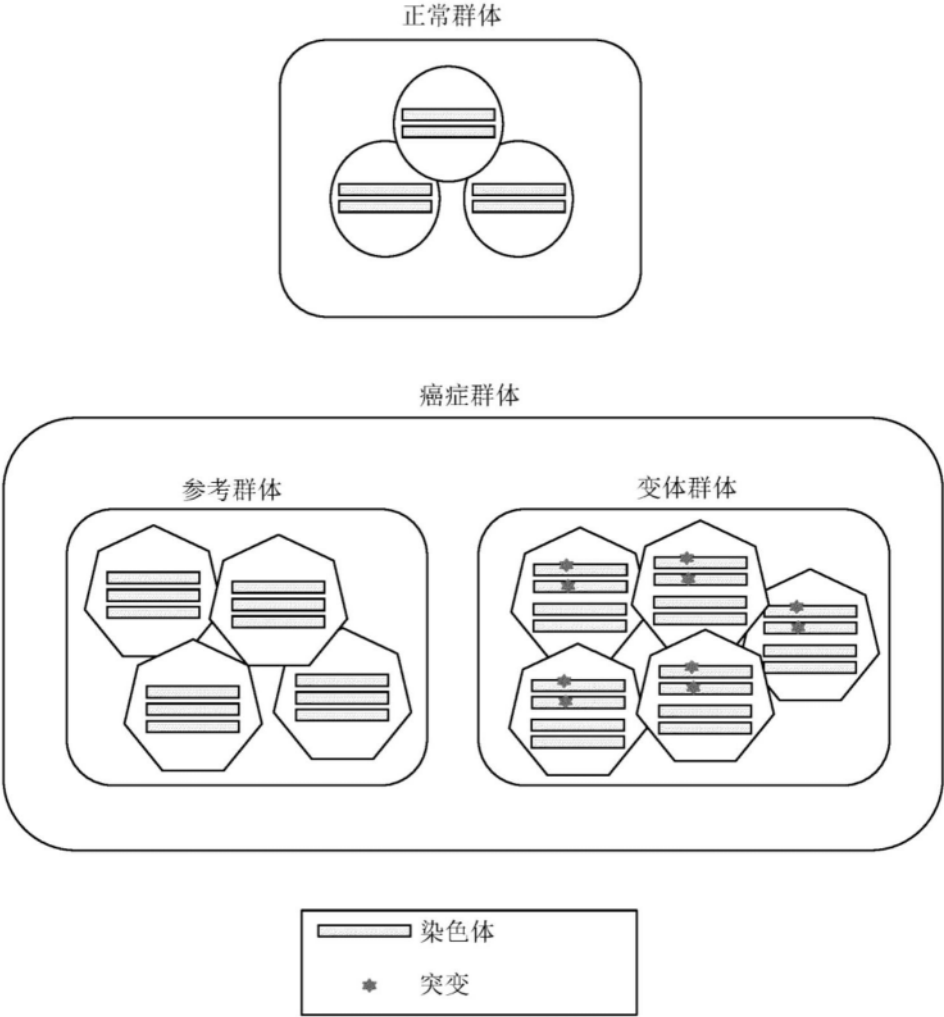


图4

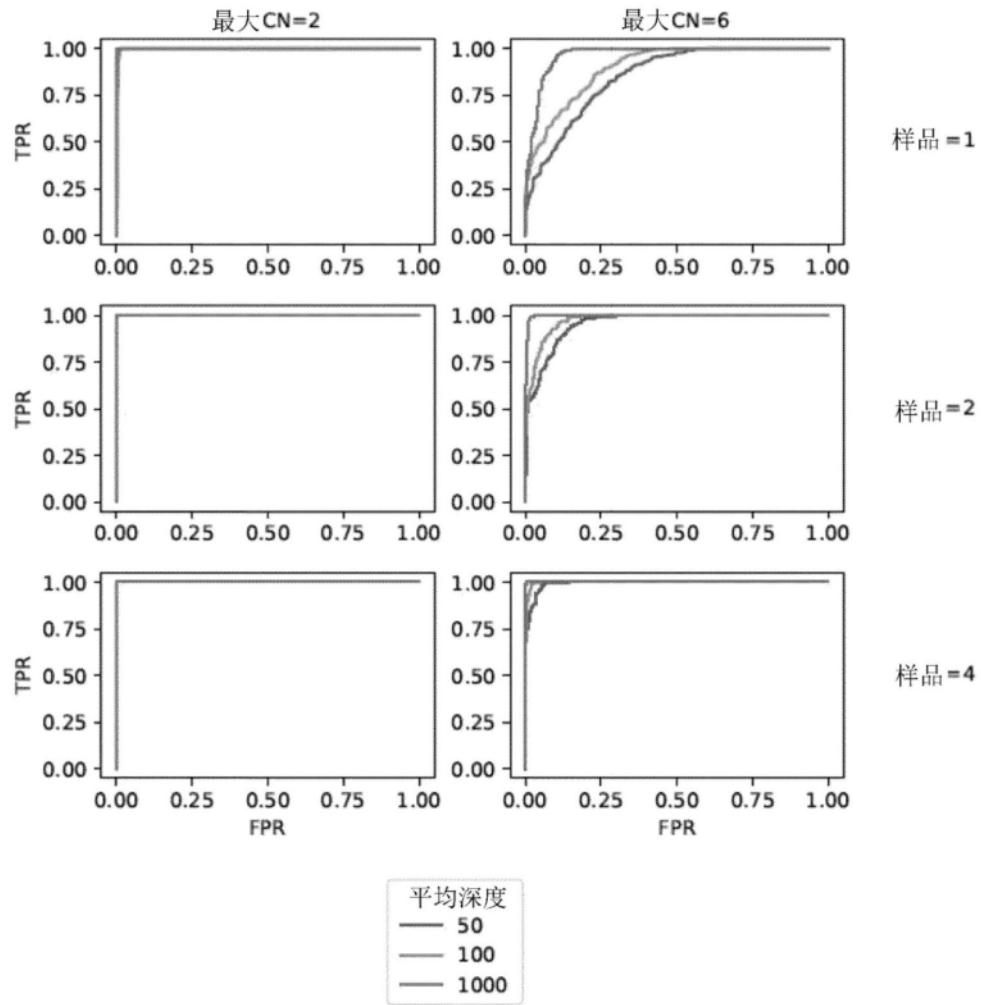


图5A

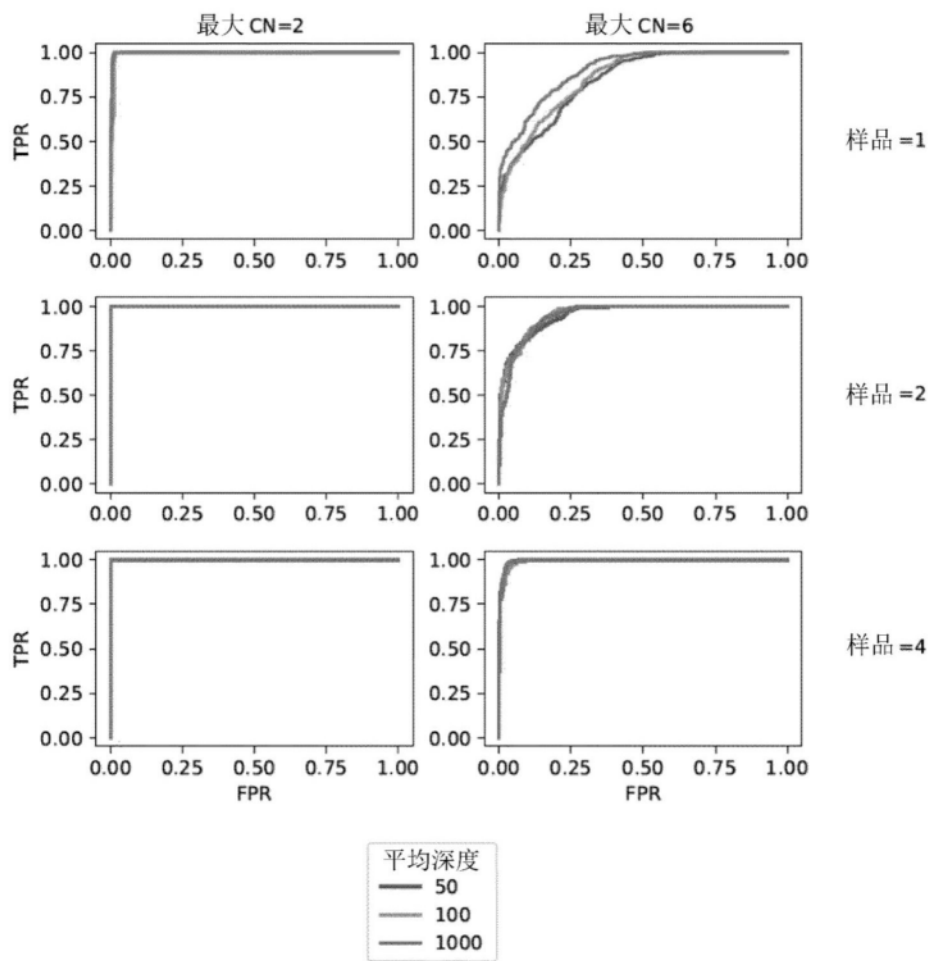


图5B

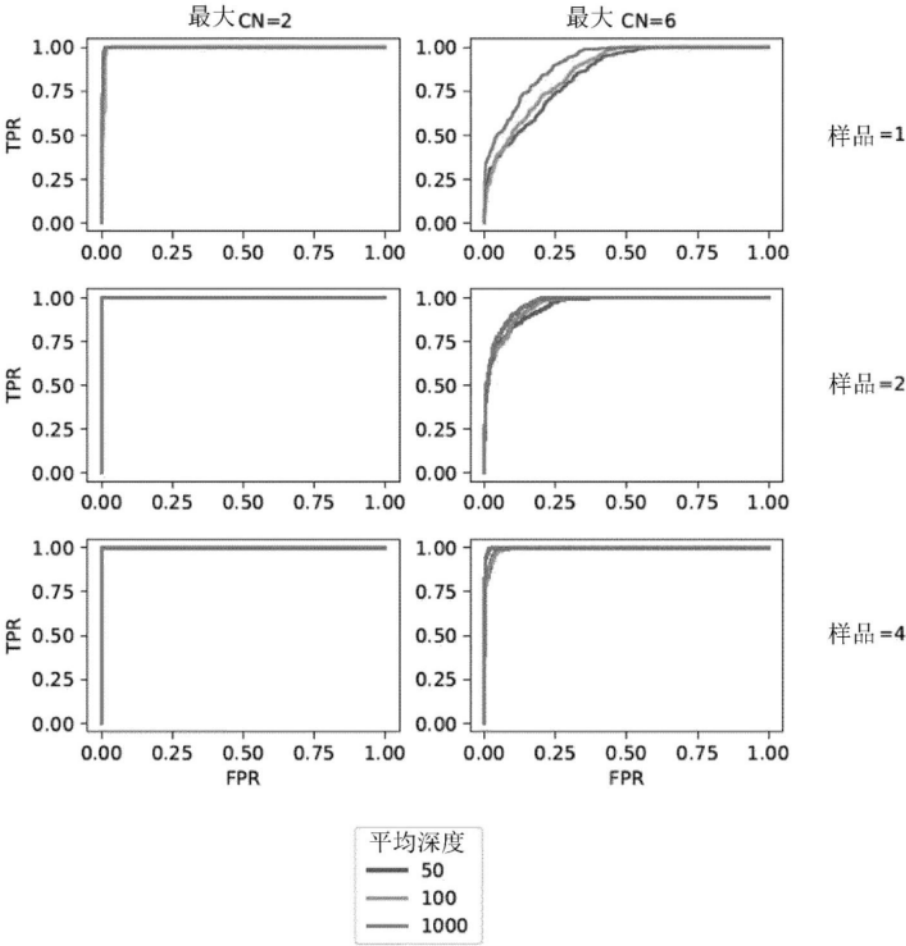


图5C

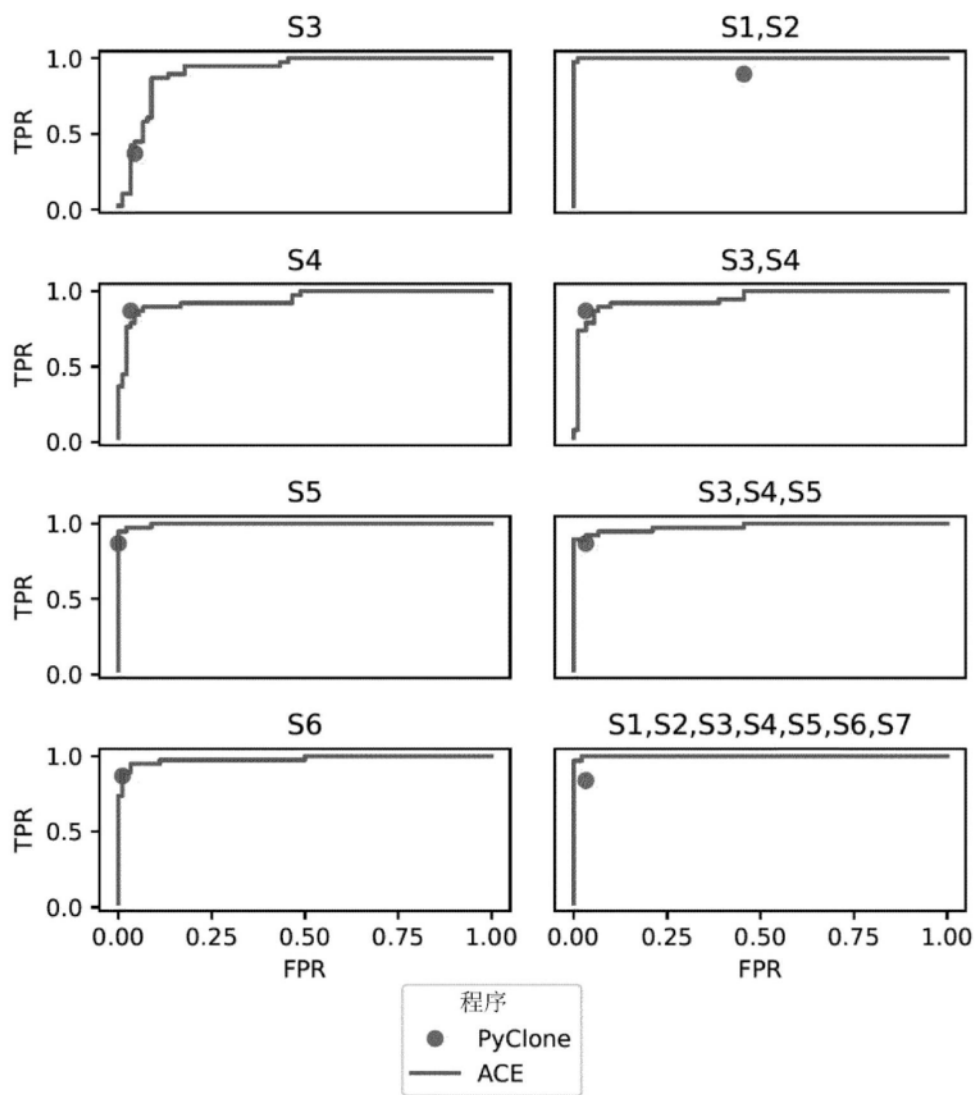


图6A

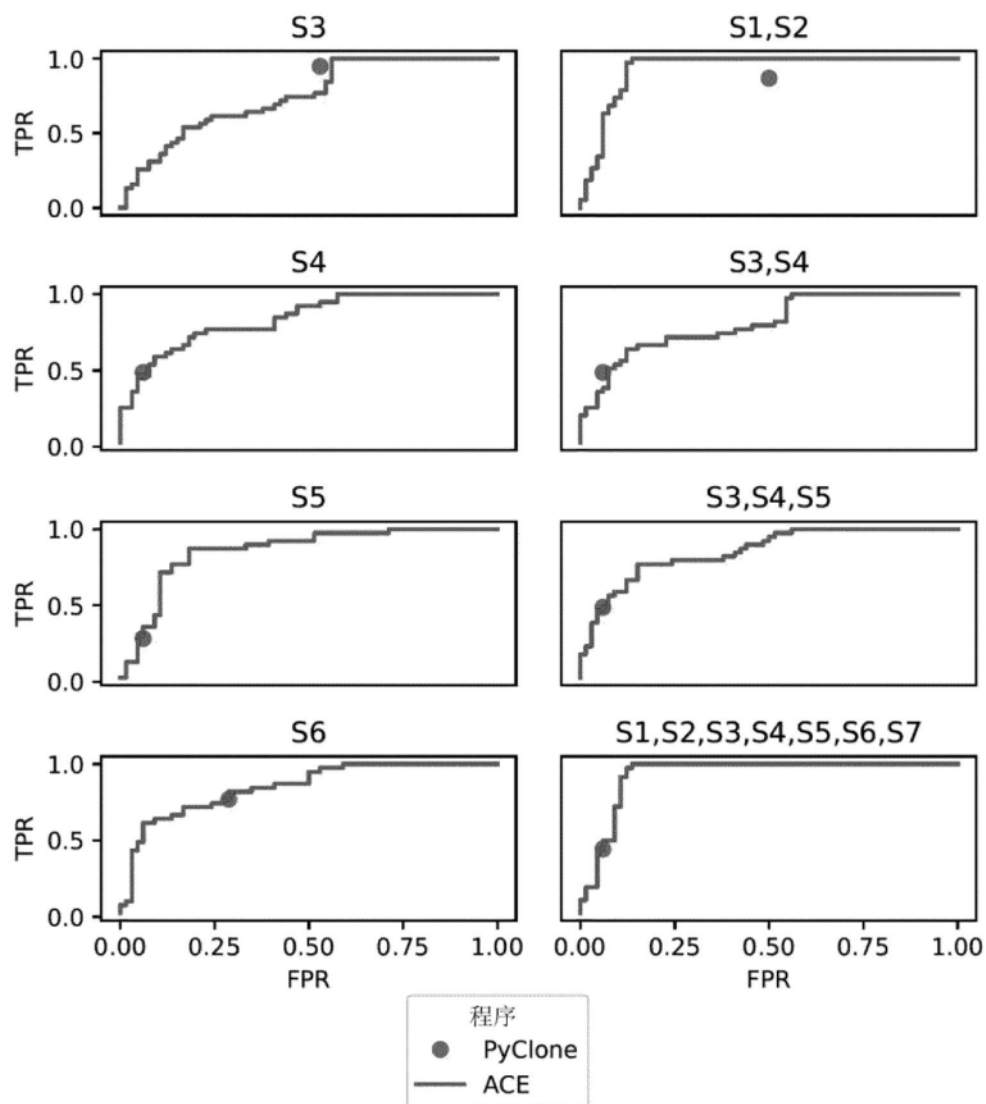


图6B

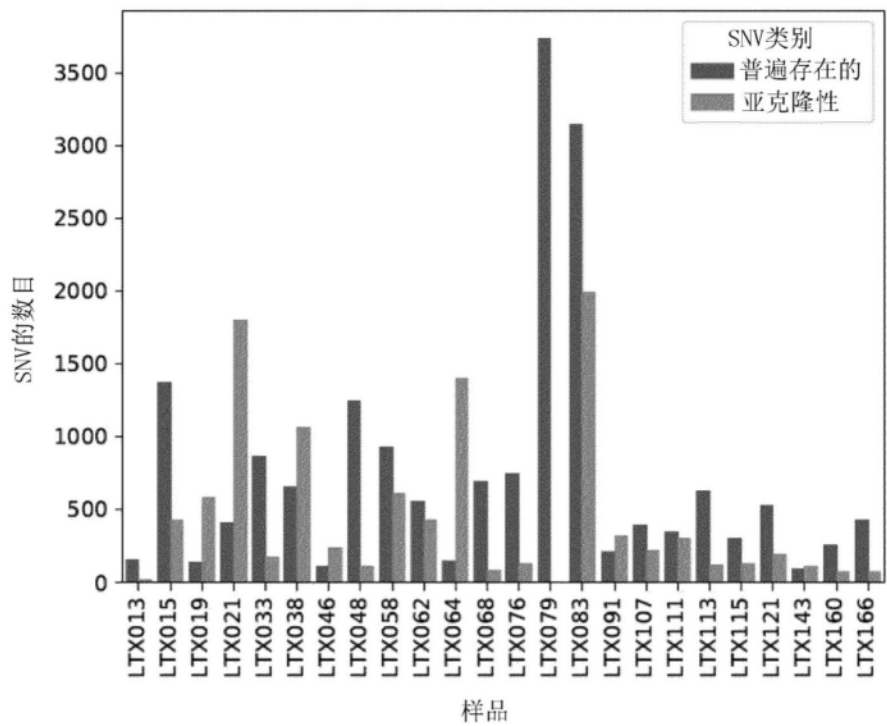


图7A

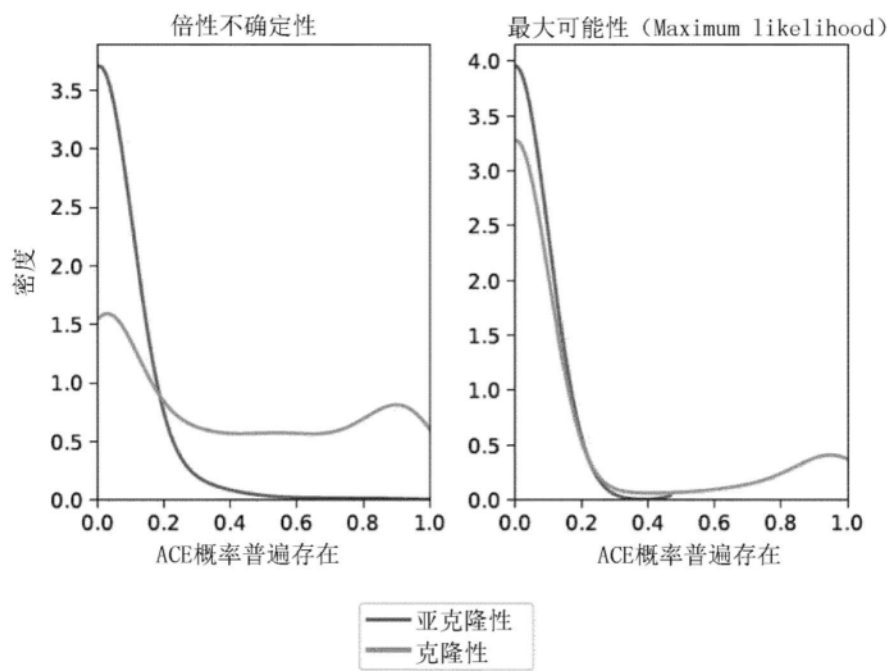


图7B

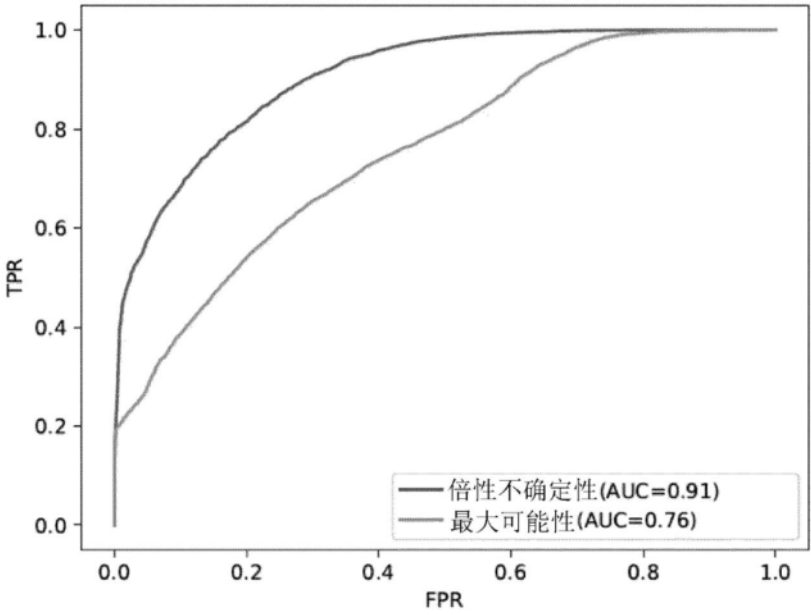


图7C

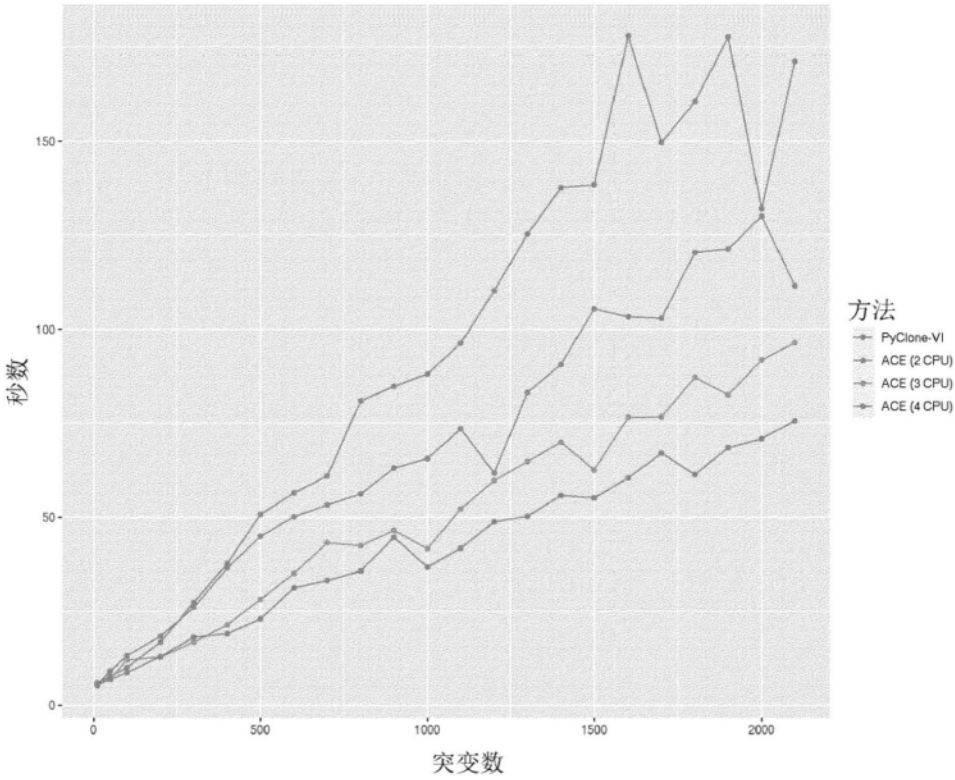


图8