



(12)发明专利申请

(10)申请公布号 CN 107704727 A

(43)申请公布日 2018.02.16

(21)申请号 201711071334.8

(22)申请日 2017.11.03

(71)申请人 杭州风起智能科技有限公司  
地址 310019 浙江省杭州市江干区九环路9号3幢2楼301-3室

(72)发明人 刘琦 周驰 刘洪马 刘峰 陈珂 马骏

(74)专利代理机构 杭州永曙知识产权代理事务所(特殊普通合伙) 33280  
代理人 曹康华

(51)Int.Cl.  
G06F 19/18(2011.01)  
G01N 33/53(2006.01)

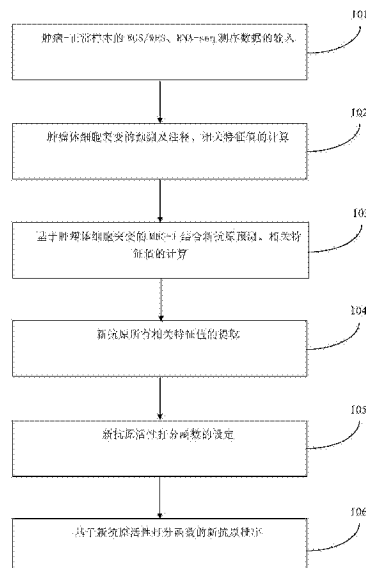
权利要求书2页 说明书6页 附图1页

(54)发明名称

基于肿瘤新抗原特征值的新抗原活性预测和排序方法

(57)摘要

本发明公开了一种基于肿瘤新抗原特征值的新抗原免疫活性打分和排序方法,包括以下步骤:肿瘤-正常样本的WGS/WES、RNA-seq测序数据的输入,肿瘤体细胞突变的预测及注释、相关特征值计算;基于肿瘤体细胞突变的MHC-I结合新抗原预测、相关特征值计算;新抗原相关特征值的提取;新抗原活性打分函数的设定;基于新抗原活性打分函数的新抗原排序。本发明首先分析计算出肿瘤体细胞的突变并完成突变注释,计算出部分特征值,然后预测MHC-I结合新抗原,计算出部分特征值;提取肿瘤新抗原的所有相关特征值,进而设定新抗原活性打分函数,最后通过新抗原活性打分函数对新抗原进行排序。本方法与传统的筛选方法相比更为高效精准,对于肿瘤免疫疗法具有重要的应用价值。



1. 基于肿瘤新抗原特征值的新抗原免疫活性预测和排序方法,其特征在于,包括以下步骤:

(1), 肿瘤-正常样本的WGS/WES、RNA-SEQ测序数据的输入:输入肿瘤-正常样本的全基因组测序数据WGS或全外显子组测序数据WES、转录组测序数据RNA-SEQ;

(2), 肿瘤体细胞突变的预测及注释、相关特征值的计算:基于步骤(1)输入的测序数据,调用Varscan或Mutect工具分析计算出肿瘤体细胞突变,调用VEP (Variant Effect Prediction) 工具完成突变注释,调用PyClone、Kallisto、Varscan或Mutect工具计算出如下特征值:突变基因克隆比、突变基因表达值TPM、等位基因突变频率VAF;

(3) 基于肿瘤体细胞突变的MHC-1结合新抗原预测、相关特征值的计算:基于步骤(2)中的肿瘤体细胞突变及注释数据,调用NetMHCpan、Netchop、OptiType工具预测MHC-1结合新抗原,并计算出如下特征值:突变肽段与MHC亲和力排序百分比、未突变肽段与MHC亲和力排序百分比、肽段剪切呈递效率;

(4) 新抗原所有相关特征值的提取:针对步骤(3)中预测的MHC-1结合新抗原,提取出肿瘤新抗原的所有相关特征值;

(5) 新抗原活性打分函数的设定:针对步骤(4)中提取的新抗原特征值,设定新抗原活性打分函数;

(6) 基于新抗原活性打分函数的新抗原排序:通过新抗原活性打分函数对新抗原进行排序。

2. 根据权利要求1所述的方法,其特征是:

步骤(4)中,新抗原相关特征值包括 $R_m$ 、 $A$ 、 $R_n$ 、 $E$ 、 $NC$ 、 $CL$ ,其中:

$R_m$ —突变肽段与MHC亲和力排序百分比,由NetMHCpan计算得出;

$A$ —等位基因突变频率VAF,由Varscan/Mutect/Strelka2计算得到;

$R_n$ —未突变肽段与MHC亲和力排序百分比,由NetMHCpan计算得出;

$E$ —突变基因表达值TPM,由Kallisto计算;

$NC$ —肽段剪切呈递效率,由netchop计算得出;

$CL$ —突变基因克隆比,由pyclone计算得出。

3. 根据权利要求2所述的方法,其特征是:

步骤(5)中,提出的新抗原活性预测打分函数为:

$$\text{Neo\_Score} = \text{abundance} \cdot \text{dissimilarity} \cdot \text{clonality};$$

$$\text{clonality} = NC \cdot CL;$$

$$\text{abundance} = L(R_m) \cdot A \cdot \tanh(E/k);$$

$$\text{dissimilarity} = (1 - L(R_n) / 2);$$

其中: $L(x) = 1 / (1 + e^{5(x-2)})$ ,  $\tanh(x)$  为双曲正切函数;

$k$ 为转录本表达丰度阈值,默认值为1。

4. 根据权利要求3所述的方法,其特征是:步骤(6)中,通过新抗原活性预测函数对新抗原进行排序算法过程如下:

a), 对于所有预测的MHC-1结合新抗原,调用新抗原活性预测函数Neo\_score计算出新抗原活性的预测值;

b), 基于新抗原活性的预测值,采用快速排序算法对新抗原进行排序;

c), 输出新抗原排序结果。

## 基于肿瘤新抗原特征值的新抗原活性预测和排序方法

### 技术领域

[0001] 本发明涉及肿瘤免疫治疗领域,具体地,涉及一种基于肿瘤新抗原特征值的新抗原活性打分和排序方法。

### 背景技术

[0002] 近年来,肿瘤免疫治疗大放异彩,临床试验不断取得特破,治愈率和有效缓解率持续提升。肿瘤新抗原的高效精准筛选在肿瘤免疫疗法中是极其重要而基础的工作,特别是对TCR-T/TIL、个体化疫苗等肿瘤免疫疗法尤为重要。

[0003] 目前,对于肿瘤新抗原的筛选方法目前通行的方案是两个步骤:步骤一、基于肿瘤-正常组织的WGS/WES数据,调用Mutect/VarScan等工具计算肿瘤细胞的基因突变;步骤二、调用NetMHCpan等算法预测MHC-1结合新抗原。

[0004] 目前还没有有效的方法基于抗原的活性进行排序,以提升抗原筛选效率。上述方案中由于没有对预测得到的MHC-1结合新抗原进行活性排序,因此会给实验验证的带来巨大的工作量,造成肿瘤新抗原的筛选效率的低下。

### 发明内容

[0005] 本发明针对上述现有技术中存在的不足,提供了一种基于肿瘤新抗原特征值的新抗原活性打分和排序方法,可大大降低实验验证的工作量,并进一步实现肿瘤新抗原的高效精准筛选。

[0006] 本发明的技术方案是这样实现的:

[0007] 基于肿瘤新抗原特征值的新抗原免疫活性预测和排序方法,其特征在于,包括以下步骤:

[0008] (1),肿瘤-正常样本的WGS/WES、RNA-SEQ测序数据的输入:输入肿瘤-正常样本的全基因组测序数据WGS或全外显子组测序数据WES、转录组测序数据RNA-SEQ;

[0009] (2),肿瘤体细胞突变的预测及注释、相关特征值的计算:基于步骤(1)输入的测序数据,调用VarScan或Mutect工具分析计算出肿瘤体细胞突变,调用VEP (Variant Effect Prediction) 工具完成突变注释,调用PyClone、Kallisto、VarScan或Mutect工具计算出如下特征值:突变基因克隆比、突变基因表达值TPM、等位基因突变频率VAF;

[0010] (3)基于肿瘤体细胞突变的MHC-1结合新抗原预测、相关特征值的计算:基于步骤(2)中的肿瘤体细胞突变及注释数据,调用NetMHCpan、NetChop、OptiType工具预测MHC-1结合新抗原,并计算出如下特征值:突变肽段与MHC亲和力排序百分比、未突变肽段与MHC亲和力排序百分比、肽段剪切呈递效率;

[0011] (4)新抗原所有相关特征值的提取:针对步骤(3)中预测的MHC-1结合新抗原,提取出肿瘤新抗原的所有相关特征值;

[0012] (5)新抗原活性打分函数的设定:针对步骤(4)中提取的新抗原特征值,设定新抗原活性打分函数;

[0013] (6) 基于新抗原活性打分函数的新抗原排序:通过新抗原活性打分函数对新抗原进行排序。

[0014] 作为优选,步骤(4)中,新抗原相关特征值包括 $R_m$ 、 $A$ 、 $R_n$ 、 $E$ 、 $NC$ 、 $CL$ ,其中:

[0015]  $R_m$ —突变肽段与MHC亲和力排序百分比,由NetMHCpan计算得出;

[0016]  $A$ —等位基因突变频率VAF,由Varscan/Mutect/Strelka2计算得到;

[0017]  $R_n$ —未突变肽段与MHC亲和力排序百分比,由NetMHCpan计算得出;

[0018]  $E$ —突变基因表达值TPM,由Kallisto计算;

[0019]  $NC$ —肽段剪切呈递效率,由netchop计算得出;

[0020]  $CL$ —突变基因克隆比,由pyclone计算得出。

[0021] 作为优选,步骤(5)中,提出的新抗原活性预测打分函数为:

[0022]  $Neo\_Score = abundance \cdot dissimilarity \cdot clonality$ ;

[0023]  $clonality = NC \cdot CL$ ;

[0024]  $abundance = L(R_m) \cdot A \cdot \tanh(E/k)$ ;

[0025]  $dissimilarity = (1 - L(R_n)/2)$ ;

[0026] 其中: $L(x) = 1/(1 + e^{5(x-2)})$ ,  $\tanh(x)$  为双曲正切函数;

[0027]  $k$  为转录本表达丰度阈值,默认值为1。

[0028] 作为优选,步骤(6)中,通过新抗原活性预测函数对新抗原进行排序算法过程如下:

[0029] a),对于所有预测的MHC-1结合新抗原,调用新抗原活性预测函数 $Neo\_score$ 计算出新抗原活性的预测值;

[0030] b),基于新抗原活性的预测值,采用快速排序算法对新抗原进行排序;

[0031] c),输出新抗原排序结果。

[0032] 采用了上述技术方案的本发明的设计思想及有益效果是:

[0033] 本发明的技术方案,提出了一种肿瘤新抗原活性预测打分函数,基于新抗原活性预测函数对所预测的MHC-1结合新抗原进行活性排序,从而实现高效精准的肿瘤新抗原筛选。

[0034] 此函数基于肿瘤新抗原生成、切割转运、新抗原与MHC结合这个完整过程设计,打分函数分为3个部分,其中,克隆性( $clonality$ )衡量新抗原由突变至短肽的效率 $NC$ 及新抗原在所有肿瘤细胞中分布的比例 $CL$ ,是影响肿瘤疫苗疗效的重要因素;丰度( $abundance$ )衡量新抗原表达量及新抗原与MHC-1结合并形成pMHC复合物的效率,新抗原突变基因表达量 $E$ 越高,等位基因突变频率 $A$ 越高,复合物间结合亲和力 $R_m$ 越高( $IC_{50}$ 值越小),免疫原性则越强;不相似度( $dissimilarity$ )衡量突变肽段与对应正常肽段亲和力的差异 $R_n$ ,由于中枢耐受机制存在,二者差异越大,新抗原的特异性就越强,用于治疗副作用也越小。函数中两个映射函数用于归一化(0到1)计算值, $L(x)$ 中阈值2为肽段-MHC结合亲和力筛选阈值, $\tanh(x/k)$ 保证新抗原表达丰度超过设定阈值 $k$ 时,函数值变化趋于平缓。

[0035] 此活性预测函数考虑了新抗原产生过程中的综合影响因素,排序后的抗原更有意义与应用价值。

## 附图说明

[0036] 图1是本发明实施例所述的基于肿瘤新抗原特征值的新抗原活性方法和排序方法示意图。

### 具体实施方式

[0037] 以下结合附图对本发明的优选实施例进行说明,应当理解,此处所描述的优选实施例仅用于说明和解释本发明,并不用于限定本发明。

[0038] 实施例一:

[0039] 如图1所示,一种基于肿瘤新抗原特征值的新抗原活性方法和排序方法,包括以下步骤:

[0040] 步骤101:肿瘤-正常样本的WGS/WES、RNA-seq测序数据的输入(使用黑色素瘤病人样本—mel\_21,Science 2015:Carreno B M,Magrini V,Beckerhapak M,et al.Cancer immunotherapy.A dendritic cell vaccine increases the breadth and diversity of melanoma neoantigen-specific T cells.[J].Science,2015,348(6236):803-8.)

[0041] 步骤102:肿瘤体细胞突变的预测及注释、相关特征值的计算:基于肿瘤-正常样本的WGS/WES、RNA-seq测序数据,调用Varscan/Mutect等工具分析计算出肿瘤体细胞突变,调用VEP (Variant Effect Prediction) 工具完成突变注释,调用PyClone、Kallisto、Varscan/Mutect工具计算出如下特征值:突变基因克隆比CL、突变基因表达值TPM、等位基因突变频率VAF。以文献中有活性肽段对应的特征值为例,病人样本—mel\_21的3个肽段对应计算出的E (TPM) 和A (VAF) 值如下:

[0042]

Neoantigen	CL	E (TPM)	A (VAF)
KMIGNHLWV (gene:CDKN2A;amino acid change:E153K;normal peptide:EMJGNHLWV)	0.9836	1.71	0.55
CLNEYHLFL (gene:NDC1;amino acid change:F169L;normal peptide:CLNEYHLFF)	0.9827	14.1	0.40
AMFWSVPTS (gene:TKT;amino acid change:R446W;normal peptide:AMFRSVPTS)	0.9835	61.3	0.47

[0043] 表1

[0044] 步骤103:基于肿瘤体细胞突变的MHC-I结合新抗原预测、相关特征值的计算:基于步骤(1)中的肿瘤体细胞突变及注释数据,调用NetMHCpan、Netchop、OptiType工具预测MHC-I结合新抗原,并计算出如下特征值:突变肽段与MHC亲和力排序百分比 $R_m$ 、未突变肽段与MHC亲和力排序百分比 $R_n$ 、肽段剪切呈递效率NC。以文献中有活性肽段对应的特征值为例,病人样本—mel\_21的3个肽段对应计算出的 $R_m$ 、 $R_n$ 和NC值如下:

[0045]

Neoantigen	$R_m$	$R_n$	NC
KMIGNHLWV	0.17	2.5	0.92496

[0046]

(gene:CDKN2A; amino acid change:E153K; normal peptide:EMIGNHLWV)			
CLNEYHLFL (gene:NDC1; amino acid change:F169L; normal peptide:CLNEYHLFF)	0.3	5	0.96394
AMFWSVPTS (gene:TKT; amino acid change:R446W; normal peptide:AMFRSVPTS)	1.7	4	0.45193

[0047] 表2

[0048] 步骤104:新抗原相关特征值的提取:针对步骤103中预测的MHC-1结合新抗原,提取出肿瘤新抗原的所有相关特征值,病人样本-me1\_21的新抗原及特征值见表1、表2;

[0049] 步骤105:新抗原活性打分函数的设定:针对步骤104中提取的新抗原特征值,设定新抗原活性打分函数;

[0050] 步骤106:基于新抗原活性打分函数的新抗原排序:通过新抗原活性打分函数对新抗原进行排序,表3给出了病人样本-me1\_21中经PMHC活性实验验证的新抗原的打分(Neo\_Score)以及在全集中的排序(Rank)。

[0051]

Neoantigen	Rank	$R_m$	$R_a$	E(TPM)	NC	CL	A(VAF)	Neo_Score
KMIGNHLWV (gene:CDKN2A; amino acid change:E153K; normal peptide:EMIGNHLWV)	1	0.17	2.5	1.71	0.92496	0.9836	0.55	0.455
CLNEYHLFL (gene:NDC1; amino acid change:F169L; normal peptide:CLNEYHLFF)	3	0.3	5	14.1	0.96394	0.9827	0.40	0.385
AMFWSVPTS (gene:TKT; amino acid change:R446W; normal peptide:AMFRSVPTS)	33	1.7	4	61.3	0.45193	0.9835	0.47	0.173

[0052] 表3

[0053] 在3个鉴定出的新抗原中,CLNEYHLFL为使用肿瘤疫苗刺激DC细胞之前就表现出可激活CD8+T细胞的免疫活性,其余2个则是在使用疫苗增强免疫系统能力后,具备不同程度的免疫活性。我们看到,KMIGNHLWV和CLNEYHLFL处于排序Top 3(总候选新抗原数94)位置。而AMFWSVPTS无论是在体外实验还是人体肿瘤微环境中,由于表达量偏低,决定其免疫原性和免疫应答偏弱,在我们的排序结果中处于33位,实验结果与我们的预测排序相当吻合。

[0054] 实施例二:

[0055] 如图1所示,一种基于肿瘤新抗原特征值的新抗原活性方法和排序方法,包括以下步骤:

[0056] 步骤101:肿瘤-正常样本的WGS/WES、RNA-seq测序数据的输入(使用黑色素瘤病人

样本二me1\_38, Science 2015: Carreno B M, Magrini V, Beckerhapak M, et al. Cancer immunotherapy. A dendritic cell vaccine increases the breadth and diversity of melanoma neoantigen-specific T cells. [J]. Science, 2015, 348 (6236): 803-8.

[0057] 步骤102: 肿瘤体细胞突变的预测及注释、相关特征值的计算: 基于肿瘤-正常样本的WGS/WES、RNA-seq测序数据, 调用Varscan/Mutect等工具分析计算出肿瘤体细胞突变, 调用VEP (Variant Effect Prediction) 工具完成突变注释, 调用PyClone、Kallisto、Varscan/Mutect工具计算出如下特征值: 突变基因克隆比CL、突变基因表达值TPM、等位基因突变频率VAF; 以文献中有活性肽段对应的特征值为例, 病人样本二me1\_38的3个肽段对应计算出的E (TPM) 和A (VAF) 值如下:

[0058]

Neoantigen	CL	E (TPM)	A (VAF)
QLSCISTYV (gene:OR8B3; amino acid change:T190I; normal peptide:QLSCTSTYV)	0.978	0.743	1
FLYNLLTRVY (gene:SEC24A; amino acid change:P469L; normal peptide:FLYNPLTRVY)	0.961	5.570	0.478
KLMNIQQKL (gene:AKAP13; amino acid change:Q285K; normal peptide:KLMNIQQQL)	0.904	0.459	0.285

[0059] 表4

[0060] 步骤103: 基于肿瘤体细胞突变的MHC-1结合新抗原预测、相关特征值的计算: 基于步骤(1)中的肿瘤体细胞突变及注释数据, 调用NetMHCpan、Netchop、OptiType工具预测MHC-1结合新抗原, 并计算出如下特征值: 突变肽段与MHC亲和力排序百分比 $R_m$ 、未突变肽段与MHC亲和力排序百分比 $R_n$ 、肽段剪切呈递效率NC; 以文献中有活性肽段对应的特征值为例, 病人样本二me1\_38的3个肽段对应计算出的 $R_m$ 、 $R_n$ 和NC值如下:

[0061]

Neoantigen	$R_m$	$R_n$	NC
QLSCISTYV (gene:OR8B3; amino acid change:T190I; normal peptide:QLSCTSTYV)	0.12	0.6	0.969
FLYNLLTRVY (gene:SEC24A; amino acid change:P469L; normal peptide:FLYNPLTRVY)	0.7	1.1	0.961
KLMNIQQKL (gene:AKAP13; amino acid change:Q285K; normal peptide:KLMNIQQQL)	0.5	0.25	0.972

[0062] 表5

[0063] 步骤104: 新抗原相关特征值的提取: 针对步骤(2)中预测的MHC-1结合新抗原, 提取出肿瘤新抗原的所有相关特征值, 病人样本二me1\_38的新抗原及特征值见表4、表5;

[0064] 步骤105:新抗原活性打分函数的设定:针对步骤(3)中提取的新抗原特征值,设定新抗原活性打分函数;

[0065] 步骤106:基于新抗原活性打分函数的新抗原排序:通过新抗原活性打分函数对新抗原进行排序,表6给出了病人样本二me1\_38中经PMHC活性实验验证的新抗原的打分(Neo\_Score)以及在全集中的排序(Rank)

[0066]

Neoantigen	Rank	$R_m$	$R_n$	E (TPM)	NC	CL	A (VAF)	Neo_Score
QLSCISTYV (gene:OR8B3;amino acid change:T190I;normal	8	0.12	0.6	0.743	0.969	0.978	1	0.299

[0067]

peptide:QLSCTSTYV)								
FLYNLLTRVY (gene:SEC24A;amino acid change:P469L;normal peptide:FLYNPLTRVY)	20	0.7	1.1	5.570	0.961	0.961	0.478	0.223
KLMNIQQKL (gene:AKAP13;amino acid change:Q285K;normal peptide:KLMNIQQQL)	66	0.5	5	0.459	0.972	0.904	0.285	0.053

[0068] 表6

[0069] 在3个鉴定出的新抗原中,FLYNLLTRVY为使用肿瘤疫苗刺激DC细胞之前就表现出可激活CD8+T细胞的免疫活性,其余2个则是在使用疫苗增强免疫系统能力后,具备不同程度的免疫活性。我们看到QLSCISTYV和FLYNLLTRVY处于Top 20(总候选新抗原数117)。而KLMNIQQKL无论是在体外实验还是人体肿瘤微环境中,由于表达量偏低,决定其免疫原性和免疫应答偏弱,在我们的排序结果中处于66位,实验结果与我们的预测排序相当吻合。

[0070] 综上所述,我们提出的新抗原免疫活性打分函数,可以很有效的衡量新抗原的免疫活性,为临床实验与肿瘤研究与免疫治疗提供帮助。

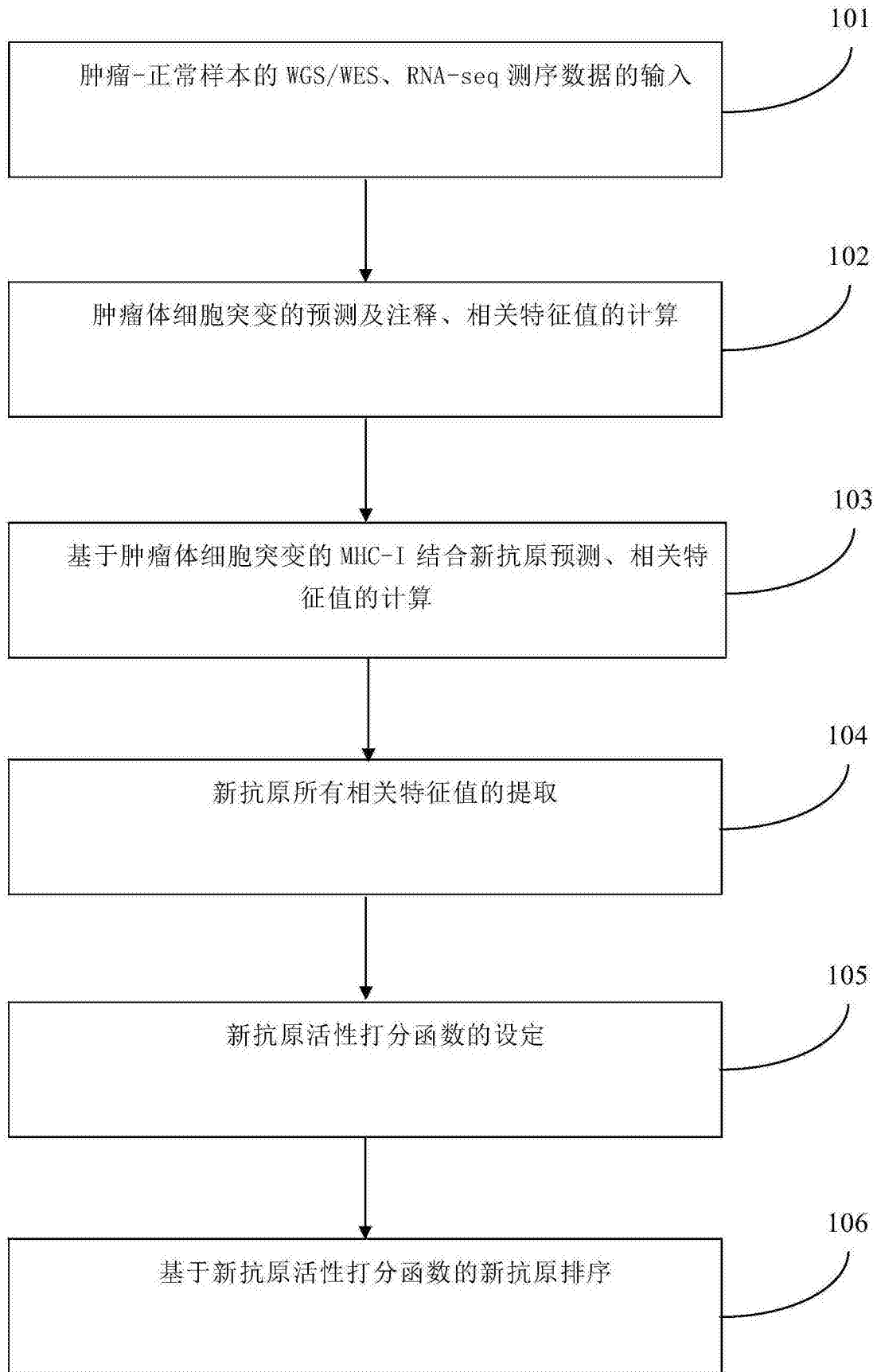


图1

专利名称(译)	基于肿瘤新抗原特征值的新抗原活性预测和排序方法		
公开(公告)号	<a href="#">CN107704727A</a>	公开(公告)日	2018-02-16
申请号	CN201711071334.8	申请日	2017-11-03
[标]发明人	刘琦 周驰 刘洪马 刘峰 陈珂 马骏		
发明人	刘琦 周驰 刘洪马 刘峰 陈珂 马骏		
IPC分类号	G06F19/18 G01N33/53		
CPC分类号	G01N33/53 G16B20/00		
代理人(译)	曹康华		
其他公开文献	CN107704727B		
外部链接	<a href="#">Espacenet</a> <a href="#">SIPO</a>		

摘要(译)

本发明公开了一种基于肿瘤新抗原特征值的新抗原免疫活性打分和排序方法，包括以下步骤：肿瘤-正常样本的WGS/WES、RNA-seq测序数据的输入，肿瘤体细胞突变的预测及注释、相关特征值计算；基于肿瘤体细胞突变的MHC-I结合新抗原预测、相关特征值计算；新抗原相关特征值的提取；新抗原活性打分函数的设定；基于新抗原活性打分函数的新抗原排序。本发明首先分析计算出肿瘤体细胞的突变并完成突变注释，计算出部分特征值，然后预测MHC-I结合新抗原，计算出部分特征值；提取肿瘤新抗原的所有相关特征值，进而设定新抗原活性打分函数，最后通过新抗原活性打分函数对新抗原进行排序。本方法与传统的筛选方法相比更为高效精准，对于肿瘤免疫疗法具有重要的应用价值。

