



(12)发明专利

(10)授权公告号 CN 109682978 B

(45)授权公告日 2020.07.03

(21)申请号 201811448569.9

(22)申请日 2018.11.30

(65)同一申请的已公布的文献号  
申请公布号 CN 109682978 A

(43)申请公布日 2019.04.26

(66)本国优先权数据  
201711242932.7 2017.11.30 CN

(73)专利权人 四川康德赛医疗科技有限公司  
地址 610093 四川省成都市高新区科园南路88号12栋201号

(72)发明人 丁平

(74)专利代理机构 北京中政联科专利代理事务所(普通合伙) 11489

代理人 陈超

(51)Int.Cl.  
C12Q 1/6886(2018.01)

(56)对比文件  
CN 103180730 A,2013.06.26,  
WO 2017/106638 A1,2017.06.22,  
Jasreet Hundal,et al.pVAC-Seq: A

genome-guided in silico approach to identifying tumor neoantigens.《Genome Medicine》.2016,

Xiuli Zhang,et al.Breast cancer neoantigens can induce CD8+ T-cell responses and antitumor immunity.《Cancer Immunol Res.》.2017,第5卷(第7期),

Malachi Griffith,et al.Genome Modeling System :AKnowledge Management Platform for Genomics.《PLOS Computational Biology》.2015,第1-21页.

Mette Voldby Larsen,et al.An integrative approach to CTL epitope prediction: A combined algorithmintegrating MHC classI binding TAP transport efficiency and proteasomal cleavage predictions.《Eur.J.Immunol.》.2005,第35卷

(续)

审查员 刘迎鸣

权利要求书2页 说明书10页  
序列表2页 附图6页

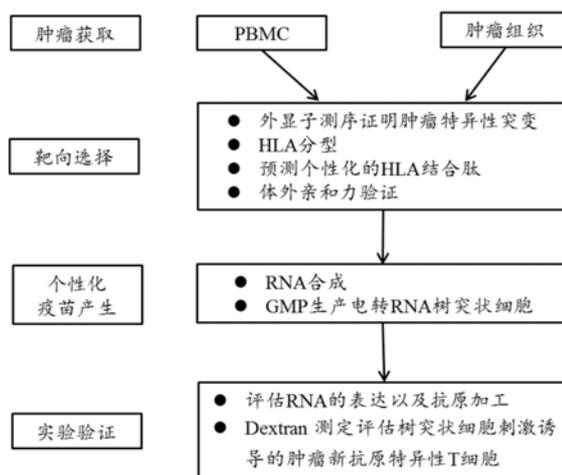
(54)发明名称

一种肿瘤突变肽MHC亲和力预测方法及其应用

(57)摘要

本发明提出了一种肿瘤突变肽MHC亲和力的预测方法,包括如下步骤:(一)肿瘤特异性突变序列的获取;(二)对突变肽MHC亲和力预测,以所述突变序列翻译成预定长度的氨基酸FASTA序列,获得测试肽;对所述测试肽分别进行与MHCI分子、MHCII分子的亲和力评估,按照确定的等级标准筛选突变肽;(三)对步骤(二)中筛选出的具有高亲和力所述突变肽,进行蛋白酶体切割预测,筛选出能够被蛋白酶体有效切割的所述突变肽;(四)实验验证实验合成步骤(三)验证突变肽与患者特有MHC的亲和力。本发明还提出了一种

基于肿瘤突变肽MHC亲和力的预测方法在筛选肿瘤新生抗原及制备个性化肿瘤疫苗中的用途。



CN 109682978 B

[接上页]

(56)对比文件

Vanessa Jurtz et al.NetMHCpan 4.0:  
Improved peptide-MHC class I interaction  
predictions integrating eluted ligand and  
peptide binding affinity data.  
《J.Immunol》.2017,第199卷(第9期),第3360-

3368页.

Jasreet Hundal et al.Cancer  
Immunogenomics: Computational Neoantigen  
Identification and Vaccine Design.《Cold  
Spring Harb Symp Quant Biol》.2017,第81卷  
第105-111页.

1. 一种肿瘤突变肽MHC亲和力预测方法,包括如下步骤:

(一) 肿瘤特异性突变序列的获取

对肿瘤组织及正常组织分别抽取DNA进行外显子深度测序,采用GMS分析,获得突变序列;

(二) 对突变肽MHC亲和力预测

以所述突变序列翻译成19个氨基酸的FASTA序列,替换氨基酸每侧9个氨基酸,获得测试肽;

对所述测试肽分别进行与MHC I分子、MHC II分子的亲和力评估,按照确定的等级标准筛选突变肽;

其中,所述亲和力评估是分别利用NetMHCpan预测与MHC I类分子结合的所述测试肽链,NetMHCIIpan预测与MHC II类分子结合的所述测试肽链;

其中,所述等级标准包括:

(1) 包含预测的结合表位的新开放阅读框 (novel open reading frames, neoORFs);

(2) 由于锚定残基突变导致小于150nM的高亲和力的体细胞核苷酸突变;

(3) 由于非锚定残基其他位置的突变导致小于150nM的高亲和力的体细单核苷酸突变;

(4) 没有预测结合表位的新开放阅读框;

(5) 较低于(2)和(3)亲和力的新开放阅读框;

在进行所述等级标准筛选后,在每个排名组中,

首先,突变肽MHC亲和力强,WT序列肽MHC亲和力弱优先;

其次,原癌基因突变优先;

再次,以突变肽亲和力的不同进行排序;

最后,基于亲和力相同的情况下,按照突变的等位基因频率用作进一步排序;

(三) 蛋白酶体切割预测

对步骤(二)中筛选出的具有高亲和力所述突变肽,进行蛋白酶体切割预测,筛选出能够被蛋白酶体有效切割的所述突变肽;

(四) 实验验证

实验合成步骤(三)筛选出的所述突变肽,与已知高亲和力结合MHC的肽进行MHC竞争结合实验,验证突变肽与患者特有MHC的亲和力。

2. 如权利要求1所述的方法,其特征在于:所述GMS分析相关突变包括如下步骤:

1) 构建一个正常对照和两个肿瘤样品的180-280bp文库;

2) Aaglient SureSelect4 Human All Exon V6捕获文库,并使用picard-tools去除PCR扩增所形成的重复;

3) Illumina HiSeq4000平台上生成外显子序列数据,FastQ数据质控;

4) BWA对外显子读数比对,联合多重设置过滤假阳性。

3. 如权利要求2所述的方法,其特征在于:所述联合多重设置过滤假阳性包括如下步骤:

1) Local realignment将由indel导致错配的区域进行重新比对,降低indel附近的比对错误率;

2) unidifiedGenotyper中的stand\_call\_conf:在变异检测过程中,用于区分低质量变

异位点和高质量变异位点的阈值；

3) VarScan Somatic版本2.2.6过滤假阳性。

4. 如权利要求1-3任一所述的方法,其特征在于:

所述步骤(二)中,还包括通过PCR-rSSO对肿瘤患者的HLA I类和II类分子类型进行测定,并通过PCR-SSP进行进一步验证。

5. 如权利要求4所述的方法,其特征在于:

所述蛋白酶体切割预测是采用NetChop C-term 3.0预测能被蛋白酶体C-末端切割位点的肽。

6. 如权利要求5所述的方法,其特征在于:

采用NetChop C-term 3.0预测能被蛋白酶体C-末端切割位点的肽的预测阈值是0.5,以区分肽C末端切割位点和肽表位内的切割位点。

7. 一种如权利要求1-6任一所述的预测方法在筛选肿瘤新生抗原及制备个性化肿瘤疫苗中的用途。

## 一种肿瘤突变肽MHC亲和力预测方法及其应用

### 技术领域

[0001] 本发明涉及生物技术领域,特别涉及一种MHC亲和力预测方法、用途及基于亲和力预测的个性化肿瘤疫苗制备方法。

### 背景技术

[0002] 随着NGS测序技术的发展,筛选肿瘤特异性抗原实现了技术上的突破。2013年,Rosenberg团队率先利用外显子技术,在肿瘤细胞系上发现新生抗原(neoantigens),并验证了其免疫反应。通过使用NGS技术和构建算法模型,外显子测序和转录组测序能准确表征肿瘤细胞的DNA和RNA,找出可能引起免疫细胞识别的肿瘤突变,生物信息学工具的发展则提高了肿瘤新生抗原的筛选能力,基因组大数据和计算机算法加速了肿瘤表位预测以及MHC(主要组织相容性复合体)亲和力预测,推动了癌症个性化免疫治疗(personal vaccine against cancer)的发展。

[0003] 鉴于癌症预后与免疫反应成正相关,因而特定的免疫细胞被作为免疫治疗的潜在靶标。免疫系统中最有效且最主要的抗原呈递细胞是树突状细胞(dendritic cell,DC),因此这类细胞也成为个性化癌症免疫治疗策略之一。Sébastien Anguille等利用含有WT1RNA自体树突状细胞免疫AML患者,II期临床试验结果显示可以预防或延缓化疗后43%的AML患者的复发,改善的总体生存期(Blood Clinical Trials and Observations, Aug.23, 2017)。

[0004] 癌症个性化免疫治疗另一个新的热点利用高通量测序识别肿瘤特异性的突变蛋白。基因组测序的最新进展表明,在癌症起始和进展的过程中,存在数以万计的不同的体细胞突变。大多数突变(过客突变)不赋予肿瘤生长优势,但会导致肿瘤基因组的不稳定。只有较少数量的癌症突变(驱动突变)干扰正常细胞调节并有助于癌症生长和对靶向治疗产生抗性。迄今为止,已经鉴定出约140个可以驱动肿瘤发生的基因。然而,驱动突变和过客突变都可能改变氨基酸编码序列,统称为非同义突变,形成肿瘤表达的突变蛋白,这些突变蛋白不被正常细胞表达。这些异常蛋白质序列在胞内被蛋白酶体加工成短肽,然后再被主要组织相容性复合物(MHC,在人类中也称为人类白细胞抗原(HLA))结合,呈递到细胞表面上,从而作为外来抗原被T细胞识别。

[0005] 2017年7月,来自美国波士顿Dana-Farber癌症中心(Neon Therapeutics公司支持的临床项目)和德国美因茨大学的两个研究团队(Biopharmaceutical New Technologies, BioN Tech公司支持的临床项目)最近展示了高通量测序在肿瘤“个性化疫苗”在临床上取得的重大突破,并分别在《Nature》上发表了题为“An immunogenic personal neoantigen vaccine for patients with melanoma”(Nature 22991, July 5, 2017)和“Personalized RNA mutanome vaccines mobilize poly-specific therapeutic immunity against cancer”(Nature 23003, July 5, 2017)的成果文章。两个临床项目都是首先对患者肿瘤组织样本进行测序,并分别使用独特算法预测哪些突变最有可能引起免疫反应,并基于预测结果,分别开发出了以多肽片段和RNA为基础的疫苗,在晚期黑色素瘤3期或4期的中晚期复

发高危人群身上进行临床实验。

[0006] 在来自Dana Farbar研究所的Wu博士带领的团队中,研究人员为每位患者量身定制了含有13-20种不同新生抗原的多肽疫苗。常规临床数据显示该类晚期肿瘤的复发率高达50%,而结果显示6位患者中,4位患者在接受疫苗两年后没有出现复发迹象。另外2名患者出现了复发迹象,但是在接受了PD-1抗体药物治疗后获得了完全缓解。

[0007] 类似结果在另一组研究中也被发现。由Ugur Sahin教授率领的美因茨大学团队为每位患者研发了含有10种不同肿瘤的新生抗原的两个多聚表位RNA疫苗。13位受试者中,8位在接受疫苗后一年内没有出现复发迹象,其他5名患者在接受疫苗时肿瘤已经出现扩散,其中2名在接受疫苗后肿瘤缩小,另外1名患者在接受PD-1抗体药物后得到完全缓解。

[0008] 该两项研究证实了在癌症治疗中利用高通量测序预测的个性化肿瘤新生抗原允许患者免疫系统更加高效精准的识别癌症抗原,这为开发癌症个性化免疫治疗方案提供了极其关键的参考。癌症个性化免疫治疗法和CAR-T在技术原理上有相识之处:首先筛查个性化的癌症特异性抗原,然后通过基因工程在体外放大肿瘤特异性抗原,从而激活T淋巴细胞,产生主动免疫。但癌症个性化免疫治疗法比CAR-T的技术流程更简洁,因此临床应用前景更好。

[0009] 一般而言,肿瘤新生抗原的预测主要分为两个步骤:

[0010] 首先,要鉴定出肿瘤细胞特有的非同义体细胞突变,这些突变是新生抗原的可能来源。

[0011] 然后,要基于非同义体细胞突变筛选出可被T细胞识别的新生抗原,筛选标准主要依据该新生抗原的表达、加工处理、被HLA呈递到肿瘤细胞表面以及被T细胞识别。

[0012] 肿瘤新生抗原的预测过程实际上是多组学结合、多算法融合的生物信息学过程。基于成对样本的全外显子组测序的结果,可以准确、全面地鉴定出肿瘤特异非同义体细胞突变,高深度的RNA-seq测序,筛选出表达的体细胞突变。最后,结合鉴定出的患者特异的HLA,通过各种算法预测患者特定HLA与体细胞突变相应新生抗原之间的亲和性,采用陪审团策略筛选出与HLA分子亲和性强的新生抗原,这些预测得到的新生抗原将会给患者进行靶向肿瘤特异的T细胞诱导或改造的靶点。

[0013] 对于个性化肿瘤疫苗技术的突破,关键的一步是采用合适的算法,能够根据肿瘤样本的测序结果,对新生抗原表达,加工递呈以及被MHC和TCR识别进行预测。肿瘤细胞的体细胞突变是随机的,相同病理类型肿瘤的不同患者之间出现相同新抗原的几率低于1%,不可能预制新抗原。所以每一个肿瘤患者接受治疗前,都必须检测肿瘤突变,并分析和预测可能作为治疗靶点的新生抗原。这也是为什么肿瘤新生抗原的研究在上个世纪前半叶就已经开启,但受限于检测和分析技术,临床应用非常缓慢。

[0014] 目前在对新生抗原的预测上虽然取得一定的突破,但是,准确预测新抗原依旧是构建个性化肿瘤疫苗的最主要挑战,其核心是算法,每个实验室都有一套自己的算法和流程,标准不统一,而且由于分析复杂度的问题,业界认为目前的算法准确性预计<40%。其中涉及基因数据预处理、突变检测、HLA分型鉴定、表达定量、蛋白裂解、TAP转运、MHC亲和力预测以及TCR预测等,而MHC亲和力预测,是最为关键的一环。

## 发明内容

[0015] 针对新生抗原的识别、筛选中存在的上述问题,本发明提出了一种MHC亲和力预测方法,包括如下步骤:

[0016] (一) 肿瘤特异性突变序列的获取

[0017] 对肿瘤组织及正常组织分别抽取DNA进行外显子深度测序,采用GMS分析,获得突变序列;

[0018] (二) MHC对突变表位亲和力预测

[0019] 以所述突变翻译成19个氨基酸的FASTA序列,理想的状态是突变导致的替换氨基酸每侧9个氨基酸,获得测试肽。;

[0020] 对所述测试肽链分别进行与MHCI分子、MHCII分子的亲和力评估,按照确定的等级标准筛选突变肽;

[0021] (三) 蛋白酶体切割预测

[0022] 对步骤(二)中筛选出的具有高亲和力所述突变肽链,进行蛋白酶体切割预测,筛选出能够被蛋白酶体有效切割的所述突变肽链;

[0023] (四) 实验验证

[0024] 实验合成步骤(三)筛选出的所述突变肽链,与已知高亲和力结合MHC的肽进行MHC竞争结合。

[0025] 如上所述的方法,其特征在于:所述GMS分析相关突变包括如下步骤:

[0026] 1) 构建一个正常对照和两个肿瘤样品的180-280bp文库;

[0027] 2) Aglient SureSelect4Human All Exon V6捕获文库,并通过picard-tools去除PCR扩增所形成的重复;

[0028] 3) Illumina HiSeq4000平台上生成外显子序列数据,FastQ数据质控;

[0029] 4) BWA (版本0.5.9) 对外显子读数比对,联合多重设置过滤假阳性。

[0030] 如上所述的方法,其特征在于:所述联合多重设置过滤假阳性包括如下步骤:

[0031] 1) Local realignment将由indel导致错配的区域进行重新比对,降低indel附近的比对错误率;

[0032] 2) unifiedGenotyper中的stand\_call\_conf:在变异检测过程中,用于区分低质量变异位点和高质量变异位点的阈值;

[0033] 3) VarScan Somatic版本2.2.6过滤假阳性。

[0034] 如上任一所述的方法,其特征在于:

[0035] 所述步骤(二)中,还包括通过PCR-rSSO对肿瘤患者的HLA I类和II类分子类型进行测定,并通过PCR-SSP进行进一步验证。

[0036] 如上所述的方法,其特征在于:

[0037] 所述测试肽是由突变导致的替换氨基酸两侧各4-10个氨基酸构成,优选的是左右各9个氨基酸。

[0038] 如上所述的方法,其特征在于:所述亲和力评估是分别利用NetMHCpan预测与MHC I类分子结合的所述测试肽链,NetMHCIIPan预测与MHC II类分子结合的所述测试肽链。如上任一所述的方法,其特征在于:所述等级标准包括如下五个等级:

[0039] (1) 包含预测的结合表位的新开放阅读框 (novel open reading frames,

neoORFs) ;

[0040] (2) 由于锚定残基突变导致小于150nM的高亲和力的体细胞核苷酸突变;

[0041] (3) 由于非锚定残基其他位置的突变导致小于150nM的高亲和力的体细单核苷酸突变;

[0042] (4) 没有预测结合表位的新开放阅读框;

[0043] (5) 较低与(2)和(3)亲和力的新开放阅读框。

[0044] 如上所述的方法,其特征在于:在进行所述等级标准筛选后,在每个排名组中,

[0045] 首先,突变肽MHC亲和力强,WT序列肽MHC亲和力弱优先;

[0046] 其次,原癌基因突变优先;

[0047] 再次,以突变肽亲和力的不同进行排序;

[0048] 最后,基于亲和力相同的情况下,按照突变的等位基因频率用作进一步排序

[0049] 如上任一所述的方法,其特征在于:

[0050] 所述蛋白酶体切割预测是采用NetChop C-term 3.0预测能被蛋白酶体C-末端切割位点的肽。

[0051] 如上所述的方法,其特征在于:

[0052] 采用NetChop C-term 3.0预测能被蛋白酶体C-末端切割位点的肽的预测阈值是0.5,以区分肽C末端切割位点和肽表位内的切割位点。

[0053] 本发明还提出了一种基于肿瘤突变肽MHC亲和力的预测方法在筛选肿瘤新生抗原及制备个性化肿瘤疫苗中的用途。

## 附图说明

[0054] 图1个性化肿瘤疫苗概念图

[0055] 图2个性化肿瘤疫苗流程框图

[0056] 图3不同解剖位置肿瘤中鉴定的错义突变在不同评估鉴定阶段的分布情况

[0057] 图4其中A为相关RNA变性PAGE胶电泳图,B为相关RNA的northern印迹分析。

[0058] 图5其中A为30批次树突状细胞表型分析结果;其中B为其中一例树突状细胞表型流式图。

[0059] 图6其中A为RNA表达流式分析图;其中B为10例不同电压下RNA表达统计图;其中C为10例不同时间下RNA表达统计图。

[0060] 图7其中A为20例不同量肿瘤RNA表达统计图;其中B为20例不同量肿瘤RNA和定量CD40L表达统计图。

[0061] 图8其中A为20例不同量肿瘤RNA和定量CD40L导入树突状细胞分泌IL-12统计图;其中B为20例不同时间段IL-12分泌统计图。

[0062] 图9树突状细胞刺激诱导的肿瘤新抗原特异性T细胞

[0063] 具体实施方法

[0064] 如图1所示的从肿瘤组织到个性化肿瘤疫苗的概念图,包括获取肿瘤组织,鉴定肿瘤新生抗原,基于新生抗原体外合成RNA以及辅助DC激活T细胞的CD40LRNA,将合成的RNA电转到成熟的DC细胞,回输到肿瘤患者体内,肿瘤新生抗原RNA在DC细胞表达并递呈到细胞表面,DC归巢到淋巴结,递呈抗原被T细胞识别,激活T细胞,T细胞迁移到肿瘤组织,杀灭肿瘤

细胞。

[0065] 基于本发明的方法,基于上述概念图,如图2所示的是本发明的流程图。

[0066] 下面结合具体实施例对本发明的流程做进一步的说明。

[0067] 实施例1 肿瘤新生抗原选择

[0068] (一) 肿瘤突变筛选

[0069] 利用高通量测序,获得肿瘤特异性编码序列突变(tumor specific coding sequence mutations),包括单个、两个以及三个核苷酸突变导致单个氨基酸错义突变和小片段核苷酸插入/缺失。获取病人多个部位肿瘤组织以及外周血单个核细胞(Peripheral blood mononuclear cells,PBMC,正常组织对照),抽提DNA进行外显子测序。所有肿瘤组织来源于快速冷冻保存组织,也可来源于甲醛固定的石蜡包埋组织;正常组织对照PBMC离心后以细胞沉淀形式冷冻保存。采用我们自己的基因组建模系统(Genome Modeling System,GMS)分析相关突变,详细流程如下。

[0070] 对于每个患者,使用500ng肿瘤/PBMC(正常对照)基因组DNA样品,构建一个正常和两个肿瘤180-280bp文库,Aglient SureSelect4Human All Exon V6捕获文库,为去除这一过程中由于PCR扩增所形成的重复,我们使用picard-tools设置REMOVE\_DUPLICATES=FALSE标识这些序列,方便GATK的识别。Illumina HiSeq4000平台上生成外显子序列数据,FastQ数据质控,这一过程中我们设置-bqsrBAQGOP 40,并计算过重,考虑reads两端的最小质量值为2,小于该值将不考虑,对bam文件里reads的碱基质量值进行重新校正,使最后输出的bam文件中reads中碱基的质量值能够更加接近真实的与参考基因组之间错配的概率。BWA(版本0.5.9)对外显子读数比对,联合多重设置过滤假阳性:(1)Local realignment将由indel导致错配的区域进行重新比对,降低indel附近的比对错误率。我们利用两个已知的可靠的indel位点(-known Mills\_and\_1000G\_gold\_standard.indels.hg19.vcf-known1000G\_phase1.indels.hg19.vcf),重新比对这些位点。(2)unifiedGenotyper中的stand\_call\_conf:在变异检测过程中,用于区分低质量变异位点和高质量变异位点的阈值。只有质量值高于这个阈值的位点才会被视为高质量的。低于这个质量值的变异位点会在输出结果中标注LowQual。(3)VarScan Somatic版本2.2.6过滤假阳性。

[0071] (二) MHC结合肽预测

[0072] 人类主要组织相容性复合物(MHC,人类也称为HLA)基因组是非常多态的,包含数千个等位基因,每个等位基因编码不同的MHC分子。MHC分子在细胞介导的免疫反应中起重要作用。肽与MHC分子的结合是肽作为免疫原的先决条件。MHCII分子结合的肽,被CD4<sup>+</sup>T细胞识别;MHCI分子结合的肽,被CD8<sup>+</sup>T细胞识别。通过反向序列特异性寡核苷酸探针(PCR-rSSO)测定病人HLA I类和II类分子类型,其具有通过序列特异性引物(PCR-SSP)进一步验证。下述用于预测与病人特异MHC分子结合的肽方法依托大而精确的数据集,这些数据来源于免疫表位数据库(Immune Epitope DataBase,IEDB)。

[0073] 应用两个独立方法对突变表位进行筛选:1)NetMHCpan用于预测与MHC I类分子结合的患者特异性突变表位;2)NetMHCIIpan用于预测与MHC II类分子结合的患者特异性突变表位。两个方法都是泛特异性预测器,涵盖具有已知序列的所有HLA类分子。将导致的氨基酸替换(Amino acid substitutions,AAS)每个错义突变翻译成19个氨基酸的FASTA序列,理想的状态是替换氨基酸每侧9个氨基酸。然后通过上述两种HLA肽结合算法评估每个

19个氨基酸序列,根据以下等级标准选择表位:(1)包含预测的结合表位的新开放阅读框(novel open reading frames, neoORFs);(2)由于锚定残基突变导致的高预测亲和力(<150nM)的体细胞核苷酸突变;(3)由于非锚定残基其他位置的突变导致高亲和力(<150nM)的体单核苷酸突变;(4)没有预测结合表位的新开放阅读框;(5)较低与(2)和(3)亲和力(<150-500nM)的新开放阅读框。每个排名组中考虑突变肽与WT序列肽结合亲和力的差异,其次原癌基因突变得优先考虑,否则通过突变肽亲和力的不同对表位进行排序,相同的亲和力的表位,突变的等位基因频率用作进一步区分。另外,还需考虑可能影响肽合成或溶解度的各种生物化学性质(疏水性以及存在多个半胱氨酸等)。

[0074] (三)MHC结合肽蛋白酶体切割预测

[0075] 内源性抗原加工处理第一步是胞内蛋白酶体切割肽的C端,形成6-30个氨基酸的肽段。因此,天然的HLA I型肽是体内蛋白酶体切割预测的间接证据。NetChop C-term 3.0由1260个MHC I型肽组成的数据库。使用该数据库预测有蛋白酶体C-末端切割位点的肽,阈值0.5用于区分肽C末端切割位点和肽表位内的切割位点。

[0076] (四)验证MHC结合的肽

[0077] 合成含特异性突变的冻干肽,溶解在含10%DMSO无菌水中,测试其无菌性、纯度、内毒素和残留的有机物。通过10nM FITC标记特异性突变肽和不同浓度未标记已知高亲和力结合HLA的肽,在过量的 $\beta$ 2m条件下与40ug/ml可溶性的HLA竞争结合,用抑制50%未标记肽结合的浓度(IC50)来表示特异性突变肽与HLA结合的亲和力。亲和力标准为:高亲和力: $\log(\text{IC}_{50}\text{nM}) < 3.7$ ;中亲和力: $\log(\text{IC}_{50}\text{nM}) 3.7-4.7$ ;低亲和力: $\log(\text{IC}_{50}\text{nM}) 4.7-5.5$ ;非常低的亲和力: $\log(\text{IC}_{50}\text{nM}) \geq 6.0$ 。

[0078] 基于上述步骤,我们利用外显子组测序,检测不同解剖位置肿瘤样品中的肿瘤特异性错义突变,使用三种算法进行并集鉴定。BRAF等位基因频率作为肿瘤变异等位基因上限,以评估其他错义突变编码基因的频率。每种编码错义突变对应的氨基酸被翻译19个氨基酸序列,并通过相关HLA软件和NetChop C-term 3.0软件,评估能与HLA结合且能被蛋白酶体切割候选肽。体外实验验证,基于预测的突变肽和WT序列肽与HLA结合亲和力的差异和/或预测的突变肽与HLA亲和力高低进一步筛选候选物,通过荧光偏振测定确定能与HLA结合肽。图3显示患者5不同解剖位置肿瘤中鉴定的错义突变在不同评估鉴定阶段的分布情况,根据相应标准患者5选择了5个肿瘤特异性错义突变作为候选疫苗。

[0079] 实施例2 体外合成RNA

[0080] (1)合成肿瘤新生抗原(Neoantigen) RNA

[0081] 根据预测结果合成肿瘤新抗原RNA,如果突变多( $\geq 10$ 个),合成串联基因结构(Tandem mini-gene constructs, TMC)的RNA;如果突变少(<10个),就合成多条含单突变的RNA。

[0082] 1)合成串联基因结构的RNA

[0083] 合成肿瘤新抗原DNA片段,通过非免疫原性甘氨酸/丝氨酸连接子(起始连接子GGSGGGGSGG,中间连接子GGSGGGGGGGG和末端连接子GGSLGGGGG)连接的7个预测的表位DNA片段(每个表位由19个氨基酸组成,突变位于第10位。)克隆到起始载体中,载体中含有优化HLA I类和II类途径的SP和MITD结构域(SP, MRVTAPRTLILLLSGALALTETWAGS; MITD, IVGIVAGLAVLAVVVIGAVVATVMCRKSSGGKGGSYSQAASSDSAQGSVDVSLTA)和提高RNA稳定性和翻译效

率的骨架序列元件。将该DNA线性化, nanodrop超微量紫外分光光度计定量, 在7.5mM ATP, CTP, UTP, GTP和3mM $\beta$ -S-ACA (D1) 帽类似物的条件下用T7RNA聚合酶进行体外转录。磁珠纯化RNA, 并通过凝胶电泳和northern印迹分析评估RNA完整性, 进一步分析包括测定浓度, 外观, pH值, 渗透压浓度, 内毒素水平和无菌性。

[0084] 对RNA表达以及RNA转入到树突状细胞的稳定性进行优化, 包括RNA分别在7个位置表达情况、2-7个RNA重复表达对特异性T细胞效率影响以及RNA之间连接子氨基酸密码子优化对RNA稳定性的影响。构建在不同位置和2-7重复串联的MART-1 RNA, 然后电转到树突状细胞中, 通过流式细胞仪检测RNA表达量以及MHC Dextran测定特异性T细胞效率影响进行评价。构建进行了氨基酸密码子优化的不同MART-1 RNA, 并标记GFP荧光标签, 然后电转到树突状细胞中, 通过流式细胞仪检测1-24h内GFP+树突状细胞的百分率来评价RNA的稳定性, 从而反应树突状细胞递呈抗原给免疫细胞持续时间。根据优化结果显示, 7个位置RNA表达没有差异, 重复表达对特异性T细胞效率不产生明显影响, 优化的最佳连接子为: 起始连接子为GGSGGGGSGG, 中间连接子为GGSGGGGGGGG和末端连接子为GGSLGGGGG。

[0085] 2) 合成单突变的RNA

[0086] 直接体外合成仅单个突变的RNA, 含有70个腺苷残基的polyA尾和I型帽结构, 并标记荧光蛋白标签。

[0087] 根据实施例1预测和体外验证结果, 基于上述实验步骤, 体外合成患者5肿瘤新生抗原5个RNA。图4中A为相关RNA变性PAGE胶电泳图, 图4中显示northern印迹分析RNA的质量, 结果显示我们合成的RNA质量和大小是符合标准的。

[0088] (2) MART-1 RNA和CD40L RNA合成

[0089] MART-1 RNA作为一个标准品, 用于后续评价树突状细胞诱导特异性CTL效应; CD40L RNA转入树突状细胞中, 增加DC IL-12表达, 促进特异性CTL的分化。

[0090] 1) MART-1 RNA

[0091] 从SK-Me1-28细胞中抽提总RNA, 然后使用Powerscript逆转录酶逆转录产生总cDNA, 接着在PFU酶 (Stratagene) 作用下使用MART-1特异性引物: MART-1 forward: 5'-CCACCATGCCAAGAGAAG-3' 和MART-1 reverse: 5'-TTAAGGTG AATAAGGTGG-3', 获得编码MART-1的cDNA。将PCR产物克隆到pGem4Z 64T载体中。为了在MART-1开放阅读框27位获得突变, 通过将密码子GCC变为CTC即将丙氨酸转化为亮氨酸, 使用QuickChange方法 (Stratagene) 进行定点诱变。所得PCR片段携带T7启动子, CD40L 5'-UTR序列, HA标签, 随后是特异性抗原编码序列, 以及T64尾。这些PCR片段作为体外转录模板产生含特异性抗原RNA包含CD40L 5'-UTR, HA标签和64个腺苷残基的polyA尾。根据生产商的说明书, 使用Cellscript的AmpliScribe T7-Flash Transcription Kit进行体外转录。然后使用CellscriptScriptCap m7G Capping System和ScriptCap 2'-O-甲基转移酶对未加帽的RNA加帽, 以产生I型帽结构。

[0092] 2) CD40L RNA

[0093] 从PMA激活的正常志愿者T细胞中抽提总RNA, 然后使用Gene Amp Gold试剂盒 (Applied Bioscience) 反转录成总cDNA, 接着使用CD40L特异性引物: CD40L forward: 5'-GCATGATCGAAACATACAACC-3' 和CD40L reverse: 5'-GTATTATGAAGACTCCCAGCG-3', 获得0.8kb CD40L DNA片段。将纯化的PCR片段亚克隆到pCR2.1 (pCR2.1CD40L  $\Delta$  XE-MET1) 载体中。该

CD40L cDNA模板,使用修饰的CD40L 5'-UTR为获得最佳转录起始,并删除开放阅读框中的第一个ATG从而产生单一的CD40L RNA产物以获得最佳的CD40L蛋白表达。根据生产商的说明书,使用Cellscript的AmpliScribe T7-Flash Transcription Kit进行体外转录。然后使用CellscriptScriptCap m7G Capping System和ScriptCap2'-O-甲基转移酶对未加帽的RNA加帽,以产生I型帽结构。

[0094] 实施例3 树突状细胞(dendritic cell,DC)培养以及成熟

[0095] 1) 单核细胞诱导产生不成熟的树突状细胞

[0096] 通过Ficoll-histopaque密度离心从健康志愿者的白细胞富集物中分离外周血单核细胞(PBMC),磷酸盐缓冲液(PBS)室温洗涤4次,30mL AIM-V培养基(Invitrogen)重悬 $2 \times 10^8$  PBMC,加入到T150培养瓶中,37°C,5%CO<sub>2</sub>, $\geq 75\%$  RH条件下培养2小时,去除非贴壁细胞,重新加入含有800U/ml GM-CSF(R&D)和800U/mL IL-4(R&D)的X-VIVO 15培养基,37°C,5%CO<sub>2</sub>, $\geq 75\%$  RH条件下培养5天。

[0097] 2) 细胞因子刺激树突状细胞成熟

[0098] 第5天,更换含有TNF- $\alpha$  (20ng/ml)、IFN- $\gamma$  (800U/ml)和PGE<sub>2</sub> (2ug/ml)的X-Vivo 15培养基,37°C,5%CO<sub>2</sub>, $\geq 75\%$  RH条件下培养过夜,第六天收获成熟的DC。

[0099] 图5中A显示30批次诱导培养树突状细胞表型分析结果;图5中B显示其中一例树突状细胞表型流式分析图,结果显示我们能获得表型稳定成熟树突状细胞。

[0100] 实施例4 RNA导入树突状细胞

[0101] 使用电转的方法将RNA导入树突状细胞。在电转前,收获DC,PBS洗涤,然后以 $4 \times 10^7$ /mL的浓度重悬于冰冷的costorsol中,置于冰上。DC与mRNA混合,置于4mm电转杯中,使用Biorad GenePulserXcell Eukaryotic System电转仪电转(电压300V,电容200uf,电阻150ohms,时间:10ms)。电转后的DC,立即用含有800U/ml GM-CSF(R&D)和800U/mL IL-4(R&D)的X-VIVO 15培养基按 $1 \times 10^6$ /ml浓度,在6孔板或T75培养瓶中,37°C,5%CO<sub>2</sub>, $\geq 75\%$  RH条件下培养4小时,然后DC以 $3 \times 10^6$ /ml浓度冻存在10%二甲基亚砷,10%葡萄糖(50%溶液)和80%自体血浆中。

[0102] 实施例5 RNA电转参数优化

[0103] 对RNA导入树突状细胞的电转电压以及时间进行优化,通过细胞存活率,细胞生长状态和RNA表达效率来评估。

[0104] 电压是电转染的重要因素,细胞在电场的作用下,膜通透性增加或是形成小孔,以完成转染过程。大多数哺乳动物细胞电转电压为400-900V/cm为佳,因此使用电极距离为4mm电转杯,那么电压范围为:160V-360V。细胞直径也影响电压,直径越小的细胞需要越高的电压,直径越大的细胞需要越小的电压。因此以100V电压为起始,每次增加50V进行优化。

[0105] 哺乳动物细胞电转染时选择方波脉冲,具有较高的转染效率和细胞存活率。在方波脉冲中,电转时间可直接设定。多数哺乳动物细胞电转时间为10-40msec。在参数优化中,参照增加电压应当降低电转时间,而减小电压则应当增大电转时间原则。因此以45msec为起始,每次减少5msec进行优化。

[0106] 我们以CD40L RNA导入树突状细胞对电转电压以及时间进行优化,通过细胞存活率、细胞生长状态和RNA表达效率来评估。结果显示在300V电压和10ms电击时间下细胞存活率和RNA表达效率最好(图6)。

[0107] 实施例6 多条RNA转入树突状细胞条件优化

[0108] 在优化好的电转条件下,对一次导入 $1 \times 10^6$ DC的编码肿瘤新抗原的RNA量进行优化,通过细胞存活率、细胞生长状态和TCRm<sup>+</sup>/GFP<sup>+</sup>细胞百分率来评估。结果显示,在仅电转肿瘤RNA情况下,随着RNA量增加情况,细胞的存活率下降,且以10-8ug RNA量最为明显,会导致80%的细胞死亡,综合评价电转7ug RNA/ $10^6$ 细胞量,TCRm<sup>+</sup>/GFP<sup>+</sup>细胞百分率最佳(图7A);为了促进树突状细胞细胞因子分泌,我们混合电转一定量的CD40L RNA,同样RNA量增加,细胞的存活率下降,同时通过TCRm<sup>+</sup>细胞和CD40L<sup>+</sup>细胞百分率分析显示电转7ug RNA/ $10^6$ 细胞,肿瘤RNA/CD40L为1:2,为最佳肿瘤RNA电转条件(图7B)。

[0109] 实施例7 评估RNA的表达以及抗原加工

[0110] 1) 肿瘤新抗原的RNA的表达以及抗原加工

[0111] RNA合成时加入荧光蛋白和SVG9标签,将RNA电转到树突状细胞中,使用流式细胞术检测,通过TCRm<sup>+</sup>荧光蛋白+细胞来分析RNA表达和抗原肽加工情况。TCRm为T细胞受体模拟物单克隆抗体,可以检测到表达细胞表面的HLA/SVG9肽复合物,从而检测RNA表达情况,荧光蛋白可以检测RNA转入效率。

[0112] 2) CD40L RNA表达

[0113] 复苏DC细胞,加入到流式管中,1500rpm离心5min,弃上清,加入2ul Dead Cell Discriminator混匀,白光照射10min,加入2ml FBS染色缓冲液,离心弃上清,加入5ul Discriminator Stop Reagent和250ul BD cytotoxic混匀,4度孵育10min,加入2ml染色缓冲液,离心弃上清,然后再用2ml 1X Perm wash洗涤细胞两次,弃上清,加入20ul APC-CD40L抗体,室温避光染色30min,2ml 1X Perm wash洗涤细胞两次,离心弃上清,300ul FBS染色缓冲液重悬细胞,流式细胞仪检测CD40L表达。

[0114] 实施例8 对DC分泌IL-12p70的评价

[0115] 树突状细胞IL-12的表达,能够促进特异性CTL的分化,因此通过检测DC分泌IL-12p70评估树突状细胞的效应。复苏DC细胞,X-VIVO15培养基重悬细胞为 $1 \times 10^6$ cell/ml,加入1ml细胞到24孔培养板中( $1 \times 10^6$ cell/well),37℃培养过夜,收集培养上清用于检测IL-12p70。使用Human Soluble Protein Master Buffer Kit(BD bioscience)和Human IL-12p70 Flex Set(BD bioscience)测定DC分泌的IL-12p70。稀释IL-12标准品和测试的上清样品,加入捕获磁珠,室温孵育1小时,然后加入PE标记的检测试剂,室温孵育2小时,加入1ml洗涤缓冲液,200g离心5分钟,300ul缓冲液重悬细胞,BD流式细胞仪检测,FCAP软件分析数据。

[0116] 我们评估肿瘤RNA/CD40L为1:2时,电转不同量RNA和不同时间段树突状细胞分泌IL-12情况。结果显示电转7ug RNA/ $10^6$ 细胞,24h树突状细胞分泌IL-12量最佳(图8)。

[0117] 实施例9 Dextran测定评估树突状细胞刺激诱导的肿瘤新抗原特异性T细胞

[0118] (1) 扩增肿瘤新抗原特异性CTL

[0119] 使用CD8<sup>+</sup>T细胞分离试剂盒(stem cell)从PBMC中纯化的CD8<sup>+</sup>T细胞。6孔板(corning)中,按照1:10的比例加入 $5 \times 10^5$ 导入RNA的DCs和 $5 \times 10^6$ CD8<sup>+</sup>T细胞,5ml的R-10培养基(含0.2U/ml IL-12和10ng/ml IL-7)共培养3天,第3天更换20U/ml IL-2和10ng/ml IL-7的R-10培养基再培养3天,在第7天复苏导入RNA的DCs,收获共培养细胞,再按导入RNA的DCs和CD8为1:10比例开始新一轮共培养。共进行3轮共培养,最后一轮共培养刺激3天后,检测

共培养物中的肿瘤新抗原特异性CTL及其相应功能。

[0120] (2) 流式细胞仪检测共培养物中的肿瘤新抗原特异性CTL及其相应功能(胞内细胞因子检测)

[0121] 第17天(第3轮共培养刺激3天后),收获共培养细胞,复苏导入RNA的DCs,DCs和CD8按1:10比例共培养4小时,培养同时加入4ul FITC CD107a抗体、0.25ul Brefeldin A和0.16ul Monensin(BD Biosciences)。培养结束后,加入1ml Aquadye孵育20分钟后,洗涤,染色APC-neoantigene Dextramer,然后加入PE-cy7 CD8,BV605CD45RA和PECy5CD28进行表面染色。为了进一步测定CTL胞内细胞因子变化,进行胞内染色。Cytotfix buffer(BD Biosciences)固定细胞,perm/wash buffer和cytofix/cytoperm buffer(BD Biosciences)进一步固定和破膜细胞,然后加入eflour 450IFN-r、AF700 TNF-A和PE IL-2染色,FACS缓冲液洗涤,使用TruCOUNT计数珠(BD Biosciences)绝对细胞计数。BD流式细胞仪收集数据,FlowJo软件分析数据,定义Aquadye-/CD8+/neoantigene+/CD28+/CD45RA-为肿瘤新抗原特异性CTL,进一步分析CD107,IFN-r,TNF- $\alpha$ 和IL-2阳性细胞,分析CTL功能。

[0122] 结果显示我们导入患者5肿瘤RNA的树突状细胞能够诱导7%左右肿瘤特异性CD8<sup>+</sup>T淋巴细胞,同时这些细胞高表达TNF- $\alpha$ 和IFN-r,且具有细胞毒性功能(图9)。

[0123] 需要说明的是:以上仅用以说明而非限制本发明的技术方案,尽管参照上述实施例对本发明进行了详细说明,本领域的普通技术人员应当理解:依然可以对本发明进行修改或者等同替换,而不脱离本发明的精神和范围的任何修改或局部替换,均应涵盖在本发明的权利要求范围当中。

[0001]	SEQUENCE LISTING		
[0002]	<110>	丁平	
[0003]	<120>	一种肿瘤突变肽MHC亲和力预测方法及其应用	
[0004]	<130>	2019	
[0005]	<150>	CN201711242932.7	
[0006]	<151>	2017-11-30	
[0007]	<160>	9	
[0008]	<170>	PatentIn version 3.3	
[0009]	<210>	1	
[0010]	<211>	10	
[0011]	<212>	PRT	
[0012]	<213>	人工序列(artificial sequence)	
[0013]	<400>	1	
[0014]		Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly	
[0015]		1 5 10	
[0016]	<210>	2	
[0017]	<211>	12	
[0018]	<212>	PRT	
[0019]	<213>	人工序列(artificial sequence)	
[0020]	<400>	2	
[0021]		Gly Gly Ser Gly Gly Gly Gly Gly Gly Gly Gly Gly	
[0022]		1 5 10	
[0023]	<210>	3	
[0024]	<211>	10	
[0025]	<212>	PRT	
[0026]	<213>	人工序列(artificial sequence)	
[0027]	<400>	3	
[0028]		Gly Gly Ser Leu Gly Gly Gly Gly Gly Gly	
[0029]		1 5 10	
[0030]	<210>	4	
[0031]	<211>	26	
[0032]	<212>	PRT	
[0033]	<213>	人工序列(artificial sequence)	
[0034]	<400>	4	
[0035]		Met Arg Val Thr Ala Pro Arg Thr Leu Ile Leu Leu Leu Ser Gly Ala	
[0036]		1 5 10 15	
[0037]		Leu Ala Leu Thr Glu Thr Trp Ala Gly Ser	
[0038]		20 25	



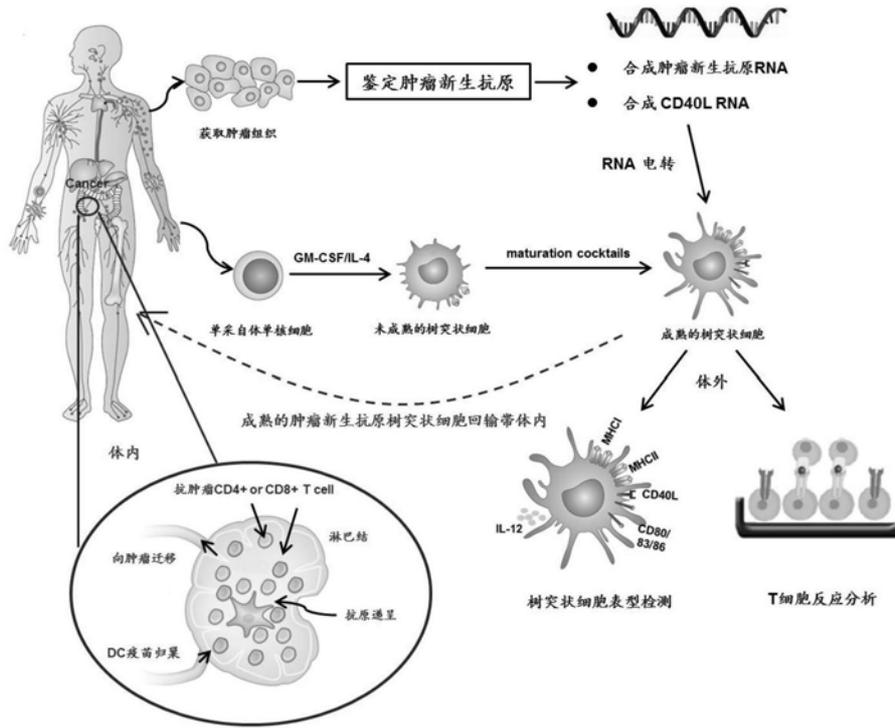


图1

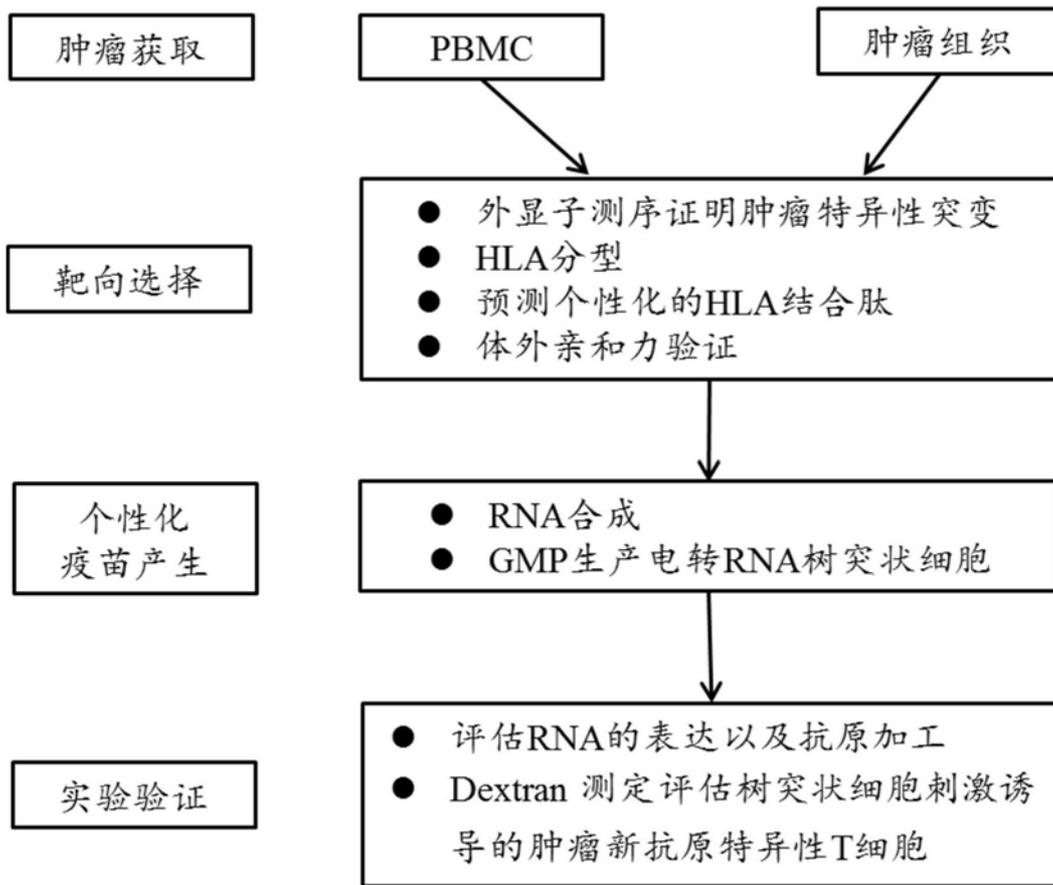


图2

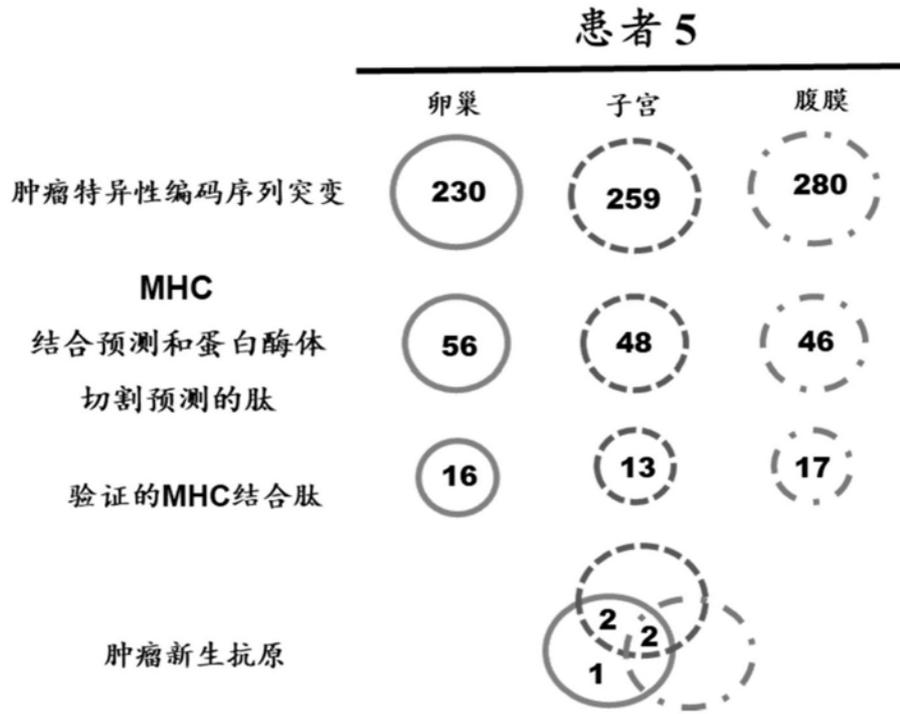


图3

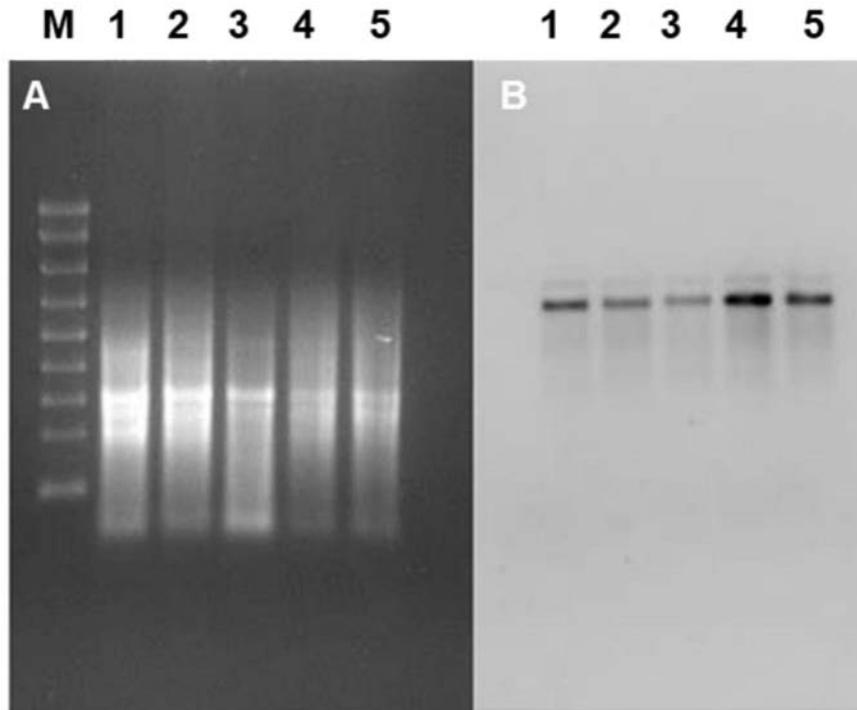


图4

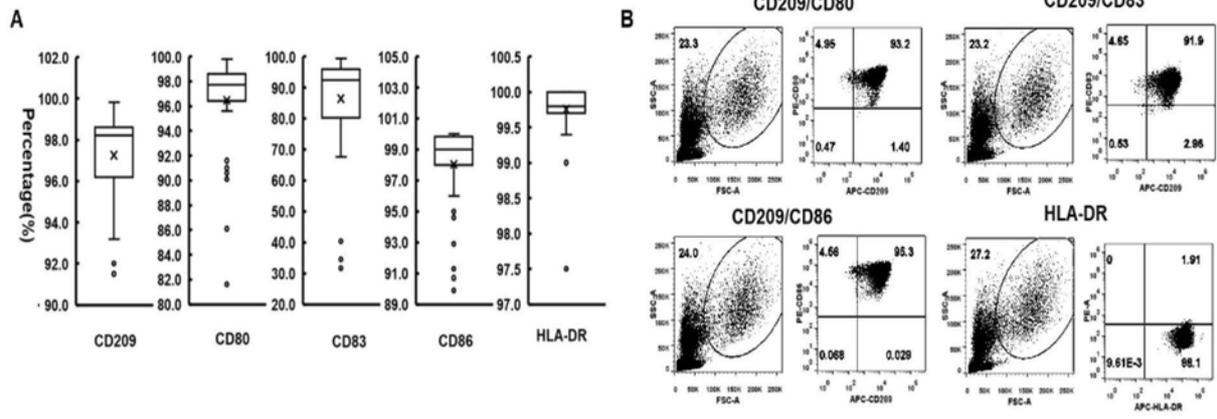


图5

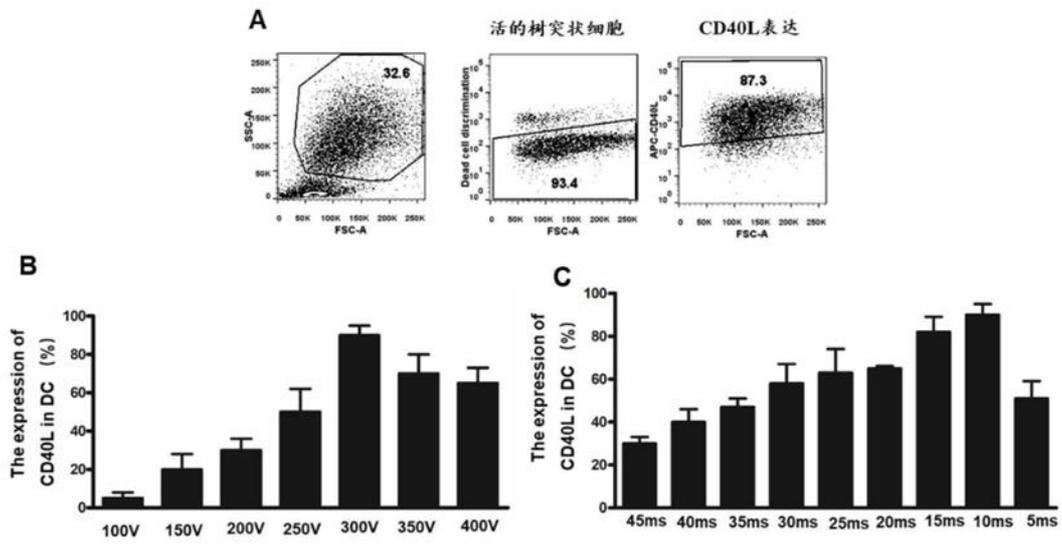


图6

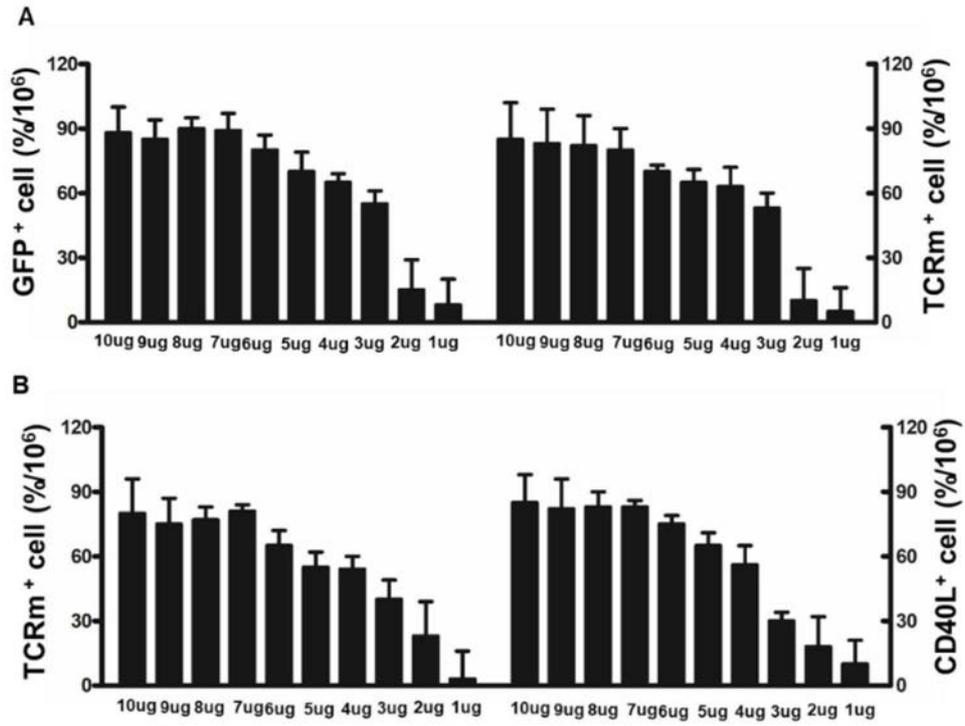


图7

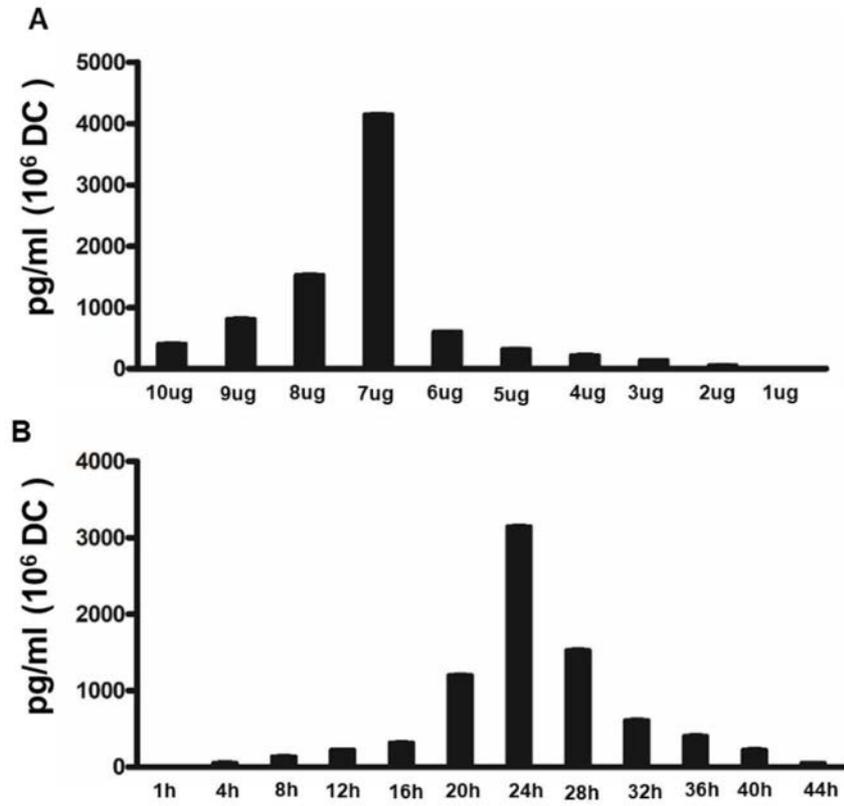


图8

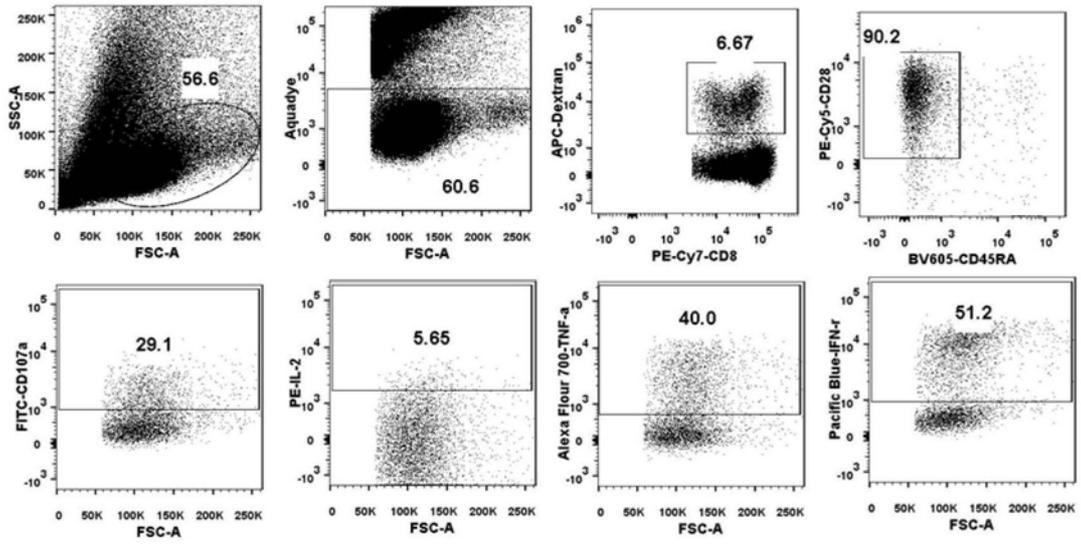


图9