



个性化肿瘤新抗原疫苗中抗原肽预测研究进展*

王广志¹⁾ 李雨雨¹⁾ 谢 鹭^{1,2)**}⁽¹⁾ 上海海洋大学食品学院, 上海 201306; ⁽²⁾ 上海生物信息技术研究中心, 上海 201203)

摘要 失控的突变导致肿瘤的发生, 其中某些非同义突变(错义、移码、融合)多肽, 被蛋白酶体降解成短肽后被抗原提呈细胞(antigen-presenting cells, APCs)识别, 呈递至引流淋巴结, 符合主要组织相容性复合物(major histocompatibility complex, MHC)结合基序的短肽, 继而被T细胞表面因子捕获而产生免疫反应, 引发肿瘤的消退, 我们称之为“新抗原”(neoantigens). 这类抗原由于未受胸腺的阴性筛选, 被T细胞识别为“异类”, 不易受免疫耐受机制的影响, 从而可作为免疫介导肿瘤治疗的有效靶点. 新一代测序技术极大推动了新抗原疫苗的可行性, 但从测序识别肿瘤的体细胞突变到TCR(T-cell receptor)识别新抗原产生免疫反应, 中间存在着大量的候选假阳性新抗原多肽, 这对于针对新抗原而设计疫苗无疑是难以跨越的障碍. 一套有效合理的筛选方法, 是新抗原疫苗制备过程中不可或缺的一环. 然而国内未见相关综述报道, 本文调研了目前新抗原免疫治疗过程中的新抗原肽预测及筛选研究进展.

关键词 肿瘤, 新抗原肽, 预测, 疫苗**中图分类号** Q7, Q81**DOI:** 10.16476/j.pibb.2019.0019

疫苗因能够激活免疫系统以预防或治疗感染和一些其他疾病而对人类健康有着重要的影响, 然而预防性的癌症疫苗目前仅对病毒来源的癌症有效, 如人类乳头瘤病毒介导的宫颈癌^[1-2]. Sipuleucel-T疫苗是美国食品药品监督管理局(FDA)首个批准的治疗性的癌症疫苗, 对前列腺癌的治疗也有一定的效果^[3]. 与其他的免疫疗法, 如免疫检查点抑制(immune checkpoint blockade, CPB)、过继T细胞疗法(adoptive cell transfer, ACT)以及抗原嵌合受体疗法(chimeric antigen receptors, CARs)相比, 大多数疫苗都不能产生显著的临床效果^[4]. 新抗原疫苗是最近开发出的一种有效的癌症疫苗, 也是目前最前沿的抗肿瘤疫苗技术之一, 接种新抗原疫苗既可以扩大已有的新抗原特异性T细胞群, 又可以在癌症患者中诱导更广泛的新型T细胞特异性, 增强肿瘤抑制^[5], 同时新抗原疫苗与其他免疫疗法如PD-1/PD-L1联合治疗效果更显著^[6-7].

传统上, 癌症疫苗被设计针对肿瘤中那些过表达的肿瘤相关抗原(tumor associated antigens, TAAs)或胚胎抗原(carcinoembryonic antigen,

CTAs), 然而TAAs由于来源于人体正常组织, 经过胸腺发育的阴性选择, 因而可能引发中央和外周免疫耐受机制, 从而导致疫苗接种的低效率以及产生自体免疫^[8]. 肿瘤特异性抗原(tumor-specific antigen, TSAs)又称“新抗原”, 是由肿瘤内部随机的体细胞突变产生的, 不同于正常组织^[9-10], 可以引发有效、安全和高度特异性的抗肿瘤免疫反应. 2017年, Ott^[11]和Sahin^[12]团队均证实, 筛选来自黑色素瘤中的“新抗原”制成的多肽以及RNA疫苗, 均可刺激CD8⁺以及CD4⁺T细胞的增殖从而导致肿瘤的消退. 值得注意的是, 本来针对CD8⁺T细胞设计的疫苗, 却介导了更多的CD4⁺T细胞的增多, 这引发了人们对主要组织相容性复合物MHC-II类分子抗肿瘤作用的思考. 个性化肿瘤新抗原疫苗正受到越来越多肿瘤患者的关注. 虽然个性化新抗原疫苗目前被认为是一种极其有效且安全的免疫治疗, 但是其制备复杂耗时、专业性高, 尤

* 国家自然科学基金资助项目(31870829)资助.

** 通讯联系人.

Tel: 021-20283705, E-mail: luxie2017@outlook.com

收稿日期: 2019-01-25, 接受日期: 2019-04-01

其是疫苗制备的第一步新抗原的筛选尤其困难. 本文将就目前的新抗原多肽筛选的研究方法进行综述, 为相关领域的研究者 and 应用者们提供参考.

1 抗原的提呈

新抗原产生于自体, 属于内源性抗原亦称肿瘤特异性抗原 (TSAs). 准确理解抗原的提呈对精准预测新抗原是至关重要的. 抗原的呈递是抗原识别中最具选择性的一步, 新抗原亦不例外, 抗原多肽只有高亲和结合到 MHC 分子继而被提呈到肿瘤细胞的表面, T 细胞才有可能与之接触识别. 抗原处理和提呈的步骤包括: 内源性蛋白质被多催化蛋白酶降解成肽段群, 在胞浆中被其他蛋白酶进一步修剪, 其中一部分被抗原处理复合体转运子 (transporter associated with antigen processing, TAP) 选择性地转运到内质网, 在那里被内质网氨基肽酶进一步修剪. 最后肽段在一系列蛋白质集合下被绑定到 MHC 分子, MHC-肽复合体最终由高尔基体输出至肿瘤细胞表面, 经 TCR 判断是否开

启免疫应答^[13].

2 新抗原肽提呈过程研究中常用的生物信息学软件

洗脱实验发现, 无论 I 类还是 II 类 MHC 分子, 都有固定的抗原肽氨基酸残基序列——称为“common motif”, 即锚定位点. MHC 分子在人类中被叫做人白细胞抗原 (human leucocyte antigen, HLA), 主要向 CD4⁺T 细胞提呈外源性抗原, 其肽结合凹槽两端呈开放状态, 能够容纳约 10~30 个较长肽段氨基酸残基, 同时有着更复杂的结合基序. 由于新抗原是内源性的, 且 MHC- I 分子结合的多肽长度比较固定, 常为 8~11 肽, 加上基序较之 MHC- II 分子相对简单, 目前的研究大都集中在 MHC- I 分子提呈多肽的预测. 但随着新抗原疫苗临床实验证实很多的 CD4⁺T 细胞被同样介导而增殖, 人们也开始越来越关注 MHC- II 分子的提呈作用. 基于这些锚定位点信息, 结合机器学习算法, 国际上已开发出多种预测表位肽与 MHC 结合的软件工具^[14-15], 常用的 MHC-肽结合工具见表 1.

Table 1 MHC binding affinity prediction tools

表1 MHC-肽结合预测工具

软件/模型	发表/更新时间	pMHC- I 类结合	pMHC- II 类结合	预测算法
NetMHC ^[16]	2008.08.	√		神经网络
PSSMHCpan ^[17]	2017.05.	√		位置特异性打分矩阵
NetMHCpan ^[18]	2017.10.	√		神经网络
MHCflurry ^[19]	2018.07.	√		神经网络
PRORED ^[20]	2001.12.		√	打分矩阵
CONSENSUS ^[21]	2010.11.		√	打分矩阵
NetMHCIIpan ^[22]	2018.02.		√	神经网络
BOTA ^[23]	2018.10.		√	深度神经网络

NetMHCpan-4.0 就是利用人工神经网络 (ANN) 预测肽与已知序列的 MHC- I 分子结合, 该方法在超过 180 000 个定量结合数据和 MS 衍生的 MHC 洗脱配体的组合上进行训练. 结合亲和力数据涵盖来自人 (HLA-A, B, C, E)、小鼠 (H-2)、牛 (BoLA)、灵长类动物 (Patr, Mamu, Gogo) 和猪 (SLA) 的 172 个 MHC 分子. MS 洗脱的配体数据涵盖 55 个 HLA 和小鼠等位基因, 用 NetMHCpan 软件算法进行肿瘤新抗原预测, 计算 MHC- I 类分子与预测肽的结合力. 预测的结果中, 若 %rank<0.5, 认为该短肽是 MHC- I 类分子的强结合; 0.5<%rank<2, 认为该短肽是 MHC- I 类分

子的弱结合; %rank>2, 认为该肽不能与 MHC- I 分子结合. 若被预测的肽段与 MHC- I 类分子的结合力越强, 其成为新抗原表位的可能性越大. NetMHCIIpan 是典型的 MHC- II 类分子结合肽预测工具, 类似于 NetMHCpan, 可用于预测人的 MHC- II 类 HLA-DR、HLA-DP 和 HLA-DQ, 以及小鼠分子 (H-2) 结合肽的预测, 阈值在 2<%rank<10. 即便 NetMHCpan 是目前广泛认可的 MHC-肽结合预测软件, 其预测结果中依然存在着大量的假阳性候选新抗原多肽, 仅仅约 1% 的预测结合肽能被 TCR 识别产生免疫原性^[24]. 因为预测出的绑定多肽可能不被蛋白酶体水解产生, 或者不能有效地刺激

CD8⁺T. 此外免疫表位数据库 (the immune epitope database, IEDB) [25] 整合了目前比较常用的一些新抗原筛选的可用预测软件, 涉及有蛋白酶体切割 NetChop、TAP 转运、MHC 结合预测等. 相关的软件还有 HLA 亚型预测: 例如 HLAMiner [26]、Athlates [27]、Polysolver [28]、Optitype [29]、seq2HLA [30] 等. 基于 WGS/WES 测序数据预测患者的 HLA 分型, 其预测精度可达 90%~99% 不等.

3 新抗原肽筛选流程及常用的管道 (pipeline)

3.1 基于MHC-多肽结合预测的新抗原筛选流程

目前的个性化新抗原多肽筛选流程总体可分为4步：a. 对病人外周血单核细胞和自身肿瘤组织进

行全外显子组 (whole exome sequencing, WES) 测序找出肿瘤特异性突变多肽, 利用外周血细胞 RNA-seq 或者 DNA-seq 分析患者 HLA 分型 .

b. MHC-肽结合预测软件预测 MHC 呈递的多肽 .

c. 进一步筛选可信度高的新抗原, 降低阳性肽条目 .

d. 人工合成新抗原多肽, 进行体外 T 细胞识别验证实验 . 常规的新抗原预测流程见图 1.

准确鉴定患者的新抗原是困难的，目前主流的新生抗原预测的主要依赖体细胞错义突变来源的多肽，例如 SNV 和 indel，由于这些方法没有考虑当两个基因在基因组中重排异常转录产生的融合基因，相关研究者也开发了关于融合基因的新抗原预测流程 INTEGRATE-neo，详细的基于 MHC-肽结合预测的新抗原筛选管道见表 2。

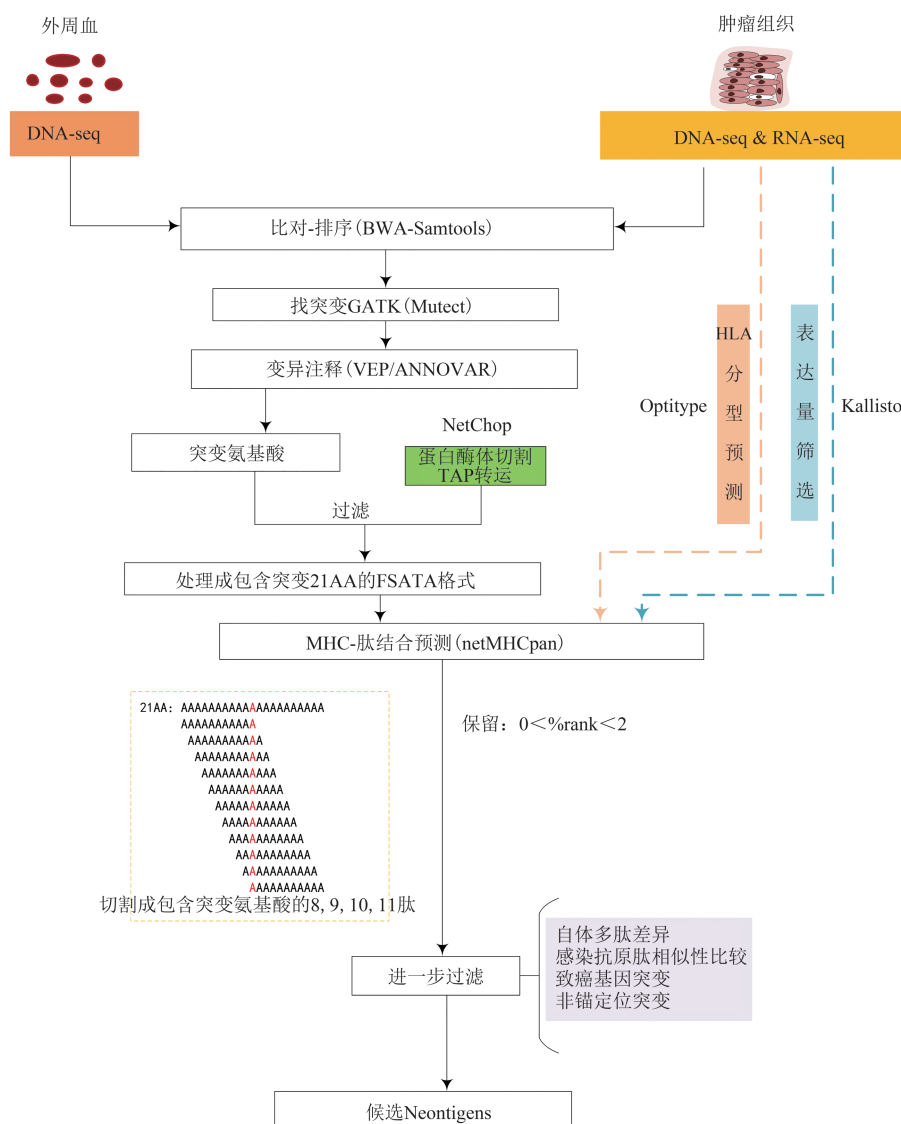


Fig. 1 Flow chart of neoantigen prediction base on MHC binding affinity prediction

图1 基于MHC-多肽结合预测的新抗原筛选流程示意图

Table 2 Pipelines for neoantigen prediction

表2 新抗原预测管道

软件	发表时间	突变来源	特点	HLA分型预测	MHC-多肽结合预测
pVAC-seq ^[31]	2016.12.	SNV, indel	整合肿瘤突变和表达数据	HLAminer/Athlates	NetMHC3.4
TSNAD ^[32]	2017.04.	SNV, indel	增加了膜蛋白的胞外突变	SOAP-HLA2.2	NetMHCpan2.8
Cloud-Neo ^[33]	2017.06.	SNV, indel	首款云计算 workflow 管道	HLAminer/Polysolver	NetMHCpan3.0
TIminer ^[34]	2017.10.	SNV, indel	免疫浸润+新抗原分析	Optitype	NetMHCpan2.8
Neopepsee ^[35]	2018.04.	SNV, indel	基于机器学习	-	Netmhcpa
INTEGRATE-neo ^[36]	2016.10.	SV	从融合基因中预测新抗原	HLAminer	NetMHC3.4

pVAC-seq, 通过整合肿瘤突变和表达数据 (DNA-seq 和 RNA-seq) 识别特异性肿瘤突变. 开发者通过对 4 例转移的黑色素瘤患者使用 pVAC-seq 预测管道最终鉴定到能够引起免疫原性的新抗原数目分别为: 3/16 (19%)、3/14 (21%)、3/18 (17%)、4/12 (33%). TSNAD 是国内浙江大学整合开发的一款预测新抗原管道, 预测的突变不仅限于 MHC 分子提呈的多肽, 而且增加了膜蛋白的胞外突变特性. Cloud-Neo 是首款云计算 workflow 预测新抗原的管道, 对 23 例黑色素瘤样本分析预测到的新抗原平均数量分别为: HLAminer, 107.89; Polysolver, 133.53; TIminer, 结合 RNA-seq 数据使用 Kallisto 定量筛选确定表达的基因, 预测新抗原. Neopepsee, 通过输入 RNA-seq 原始数据以及体细胞突变 VCF 文件, 运用机器学习的方法从 14 个新抗原相关特征中挑选 9 个强相关的特征预测新抗原. 相比于传统基于 IC50/%rank 值的预测方法, PR 曲线 (以查准率为纵轴, 查全率为横轴绘制的曲线, 曲线下面积越大性能越好) 性能改进了 2~3 倍. 在 224 个公开的胃腺癌数据集上应用 Neopepsee, 平均每个病人预测到 7 个新抗原, 且预测的新抗原负荷与患者的预后显著关联. INTEGRATE-neo, 从融合基因出发预测新抗原. 目前新抗原预测的 pipeline “核心” 基本一致, 均定位于外显子区域确定有表达的蛋白质.

3.2 基于MHC结合肽的质谱新抗原筛选

基于 HLA 结合肽免疫纯化的质谱数据新抗原预测, 能极大提高预测准确率^[37-39], 因为基于质谱 (MS) 的方法可以精准捕获大量已加工肽和已呈递肽的描述, 从而避免了抗原呈递预测所导致的误差. 基于质谱策略的新抗原筛选步骤主要包括: a. HLA 结合肽的免疫纯化; b. 液相串联质谱鉴定 (LC-MS/MS); c. 用户自定义蛋白质数据库构建 (SNP-calling 突变蛋白序列以及相应的 Uniport 人类标准蛋白质序列); d. 使用蛋白质组的定性定量分

析工具, 如 Peaks^[40]、MaxQuant^[41]、Mascot^[42]、SEQUEST^[43], 对 MS 数据进行搜库来鉴定真实 MHC 结合的突变多肽; e. 筛选候选新抗原, 体外验证免疫原性.

目前质谱策略存在的一个困难是患者多个 HLA 等位基因的共同表达使得鉴定到的多肽难以区分 HLA 亚型. 幸运的是, Wu 研究小组^[38]通过使用表达单个 HLA 分型的细胞系很好地解决了这一难题, 并用自开发的 MSIntrinsicEC 神经网络算法基于质谱多肽数据联合基因的表达量筛选, 其预测到的新抗原能够达到 50% 以上, 被认为是新抗原预测的极佳流程. 其研究于 2017 年发表在 *Immunity*, 但相关的算法并未公开. 最近 Bulik-Sullivan 等^[39]利用 74 例肿瘤患者 HLA 质谱数据构建的深度学习模型 EDGE 同样取得了惊人的效果, 使用该模型分析 9 例经过 PD-1 或 PD-L1 治疗的转移性非小细胞肺癌患者, 取预测排名前 20 的新抗原与 T 细胞体外培养 2 w, 结果 5/9 (56%) 患者识别到新抗原特异性 T 细胞反应.

4 新抗原特征及对预测新抗原进一步筛选的依据

TsAs 存在与自体多肽的差异, 与 TsAs 受到中枢耐受机制的影响不同, 因而 T 细胞可以很好地识别, 但在其呈递过程中又受到多种因素的影响, 导致预测结果较不理想, 为了能够筛选到能真正具备激活 T 细胞免疫潜力的新抗原, 研究人员在 MHC-多肽结合性预测的基础上又做了许多筛选, 如: 与自体多肽的差异比较^[44]、与感染抗原的相似性比较^[45]、基于基因表达量的筛选^[38]、考察突变是否位于驱动性致癌基因^[46]等. 此外基于 T 细胞更可能作用于抗原肽的非锚定区域, 如 9 肽的第 3~7 位^[47], 因此非锚定位点的突变被认为更能促进 T 细胞的识别. 另有关于 20 000 例黑色素瘤中新抗原的研究发现, 98% 的新抗原被发现在非驱动突变,

仅有2%的新抗原来源于驱动突变^[48]. 同时新抗原也很少存在共享的情况, 仅仅 H3.3 K27M、MYD88 L265P、KRAS G12D 在很少的患者中被同时发现^[49]. 目前对 TCR 识别这一过程的机制还不明朗, 所以即使是被 MHC 提呈的多肽, 相当多的肽依旧不能引起免疫应答. 因此基于新抗原疫苗原理的免疫疗法也只能是个性化的.

5 “非编码区筛选新抗原”——新的希望

最近蒙特利尔大学免疫学和癌症研究所 (IRIC) 报道了他们的惊人发现^[50]: 传统的基于 WES 的筛选忽略了很多 TSAs, 那些来自非编码区的未突变但异常表达的 TSA (aberrantly expressed TSAs, aeTSAs) 也应该被考虑是新抗原. 该团队建立了一种蛋白质基因组学的方法通过转录组测序以及质谱筛选从 2 个小鼠癌细胞系和 7 个人类原发性肿瘤中, 鉴定到 40 个 aeTSAs, 发现其中 90% 的 TSAs 来自于非编码区, 相关的研究也证实相对于突变的 TSA (mutated tumor-specific antigens, mTSAs) 的个体特异性, aeTSAs 可以在肿瘤之间共享, 因而 aeTSAs 是制备疫苗的首选. 同时 99% 的癌症突变来源于非编码区域^[51], 外显子区域只占人类蛋白质编码序列的 2%, 而高达 75% 的基因组是能够被潜在转录和翻译的. 这有可能间接解释了为什么目前基于 WES 新抗原预测的真阳性率那么低, 不过尽管这一发现堪称领域的突破, 它还需要更多的实验研究来证实. 如果这种情况广泛存在, 相信当研究人员把基因非编码区也纳入新抗原预测范围, 普适的新抗原疫苗将不再是梦想.

6 目前新抗原预测存在的问题

理想情况下产生的新抗原是能够被 CD8⁺T 细胞识别并清除的, 然而肿瘤能够通过激活各种检查点机制来抑制这种免疫反应. 目前新抗原预测存在的问题主要有: a. 无法百分百的预测真实多肽呈递过程 (质谱检测可以很好地改善这一问题). b. 新抗原肽引发免疫原性的机制不明, 我们需要更多的实验研究去发掘免疫肽的特性. c. 人类相关新抗原数据量目前还太少. d. 新抗原的个体特异性. e. 基于非编码区表达异常产生的新抗原预测流程还不够完善. f. 目前新抗原预测中未知因素高达 45%, 其他已知因素如 MHC 多肽复合物之间的亲和力因素占 28%、基因的表达因素占 18%、蛋白酶体的切割约占 8%, 意味着穷极目前的已有知识对于新抗原

的预测结果也达不到 60%. g. 存在一种可能, TCR 识别是一个随机的过程, 如果那样的话, 从肿瘤中透视的信息可能永远无法精准地预测新抗原.

7 展 望

虽然目前国内新抗原疫苗的发展较之国外还有很大的不足, 也未见新抗原疫苗治疗取得成功的报道, 但国内越来越多的学者开始关注新抗原, 许多临床的实验也都在积极准备进行之中. 相信不久的将来, 个性化肿瘤新抗原疫苗, 会给肿瘤治疗带来新的希望. 同时通过对 TCR 的测序以及非编码区新抗原的揭示, 新抗原治疗的边界和可应用性可能进一步拓展, 真正成为肿瘤免疫治疗的重要范畴.

参 考 文 献

- [1] Van Der Burg S H, Arens R, Ossendorp F, *et al.* Vaccines for established cancer: overcoming the challenges posed by immune evasion. *Nature Reviews Cancer*, 2016, **16**(4): 219-233
- [2] Aldous A R, Dong J Z. Personalized neoantigen vaccines: A new approach to cancer immunotherapy. *Bioorganic & Medicinal Chemistry*, 2018, **26**(10): 2842-2849
- [3] Melero I, Gaudernack G, Gerritsen W, *et al.* Therapeutic vaccines for cancer: an overview of clinical trials. *Nature Reviews Clinical Oncology*, 2014, **11**(9): 509-524
- [4] Yarchoan M, Johnson B A, 3rd, Lutz E R, *et al.* Targeting neoantigens to augment antitumour immunity. *Nature Reviews Cancer*, 2017, **17**(4): 209-222
- [5] Hu Z, Ott P A, Wu C J. Towards personalized, tumour-specific, therapeutic vaccines for cancer. *Nature Reviews Immunology*, 2018, **18**(3): 168-182
- [6] Brown S D, Warren R L, Gibb E A, *et al.* Neo-antigens predicted by tumor genome meta-analysis correlate with increased patient survival. *Genome Research*, 2014, **24**(5): 743-750
- [7] Van Allen E M, Miao D, Schilling B, *et al.* Genomic correlates of response to CTLA-4 blockade in metastatic melanoma. *Science (New York, NY)*, 2015, **350**(6257): 207-211
- [8] Efremova M, Finotello F, Rieder D, *et al.* Neoantigens generated by individual mutations and their role in cancer immunity and immunotherapy. *Frontiers in Immunology*, 2017, **8**: 1679
- [9] Linette G P, Carreno B M. Neoantigen vaccines pass the immunogenicity test. *Trends in Molecular Medicine*, 2017, **23**(10): 869-871
- [10] Hellmann M D, Snyder A. Making it personal: neoantigen vaccines in metastatic melanoma. *Immunity*, 2017, **47**(2): 221-223
- [11] Ott P A, Hu Z, Keskin D B, *et al.* An immunogenic personal neoantigen vaccine for patients with melanoma. *Nature*, 2017, **547**(7662): 217-221
- [12] Sahin U, Derhovanessian E, Miller M, *et al.* Personalized RNA mutanome vaccines mobilize poly-specific therapeutic immunity

- against cancer. *Nature*, 2017, **547**(7662): 222-226
- [13] Zhang X, Sharma P K, Peter Goedegebuure S, *et al.* Personalized cancer vaccines: targeting the cancer mutanome. *Vaccine*, 2017, **35**(7): 1094-1100
- [14] Gubin M M, Artyomov M N, Mardis E R, *et al.* Tumor neoantigens: building a framework for personalized cancer immunotherapy. *The Journal of Clinical Investigation*, 2015, **125**(9): 3413-3421
- [15] Pritchard A L, Burel J G, Neller M A, *et al.* Exome sequencing to predict neoantigens in melanoma. *Cancer Immunology Research*, 2015, **3**(9): 992-998
- [16] Lundegaard C, Lamberth K, Harndahl M, *et al.* NetMHC-3.0: accurate web accessible predictions of human, mouse and monkey MHC class I affinities for peptides of length 8-11. *Nucleic Acids Research*, 2008, **36**(Web Server issue): W509-512
- [17] Liu G, Li D, Li Z, *et al.* PSSMHCpan: a novel PSSM-based software for predicting class I peptide-HLA binding affinity. *GigaScience*, 2017, **6**(5): 1-11
- [18] Jurtz V, Paul S, Andreatta M, *et al.* NetMHCpan-4.0: improved peptide-MHC class I interaction predictions integrating eluted ligand and peptide binding affinity data. *Journal of Immunology*, 2017, **199**(9): 3360-3368
- [19] O'donnell T J, Rubinsteyn A, Bonsack M, *et al.* MHCflurry: open-source class I MHC binding affinity prediction. *Cell Systems*, 2018, **7**(1): 129-132
- [20] Singh H, Raghava G P. ProPred: prediction of HLA-DR binding sites. *Bioinformatics*, 2001, **17**(12): 1236-1237
- [21] Wang P, Sidney J, Kim Y, *et al.* Peptide binding predictions for HLA DR, DP and DQ molecules. *BMC Bioinformatics*, 2010, **11**:568
- [22] Andreatta M, Karosiene E, Rasmussen M, *et al.* Accurate pan-specific prediction of peptide-MHC class II binding affinity with improved binding core identification. *Immunogenetics*, 2015, **67**(11-12): 641-650
- [23] Graham D B, Luo C, O'connell D J, *et al.* Antigen discovery and specification of immunodominance hierarchies for MHCII-restricted epitopes. *Nature Medicine*, 2018, **24**(11): 1762-1772
- [24] Kristensen V N. The antigenicity of the tumor cell - context matters. *The New England Journal of Medicine*, 2017, **376**(5): 491-493
- [25] Vita R, Overton J A, Greenbaum J A, *et al.* The immune epitope database (IEDB) 3.0. *Nucleic Acids Research*, 2015, **43**(Database issue): D405-412
- [26] Warren R L, Choe G, Freeman D J, *et al.* Derivation of HLA types from shotgun sequence datasets. *Genome Medicine*, 2012, **4**(12): 95
- [27] Liu C, Yang X, Duffy B, *et al.* ATHLATES: accurate typing of human leukocyte antigen through exome sequencing. *Nucleic Acids Research*, 2013, **41**(14): e142
- [28] Shukla S A, Rajasagi M, Dixon P, *et al.* Abstract 1093: sensitive detection of somatic mutations in class I HLA genes reveals enrichment for functional events in cancer. 2015, **75**(15 Supplement): 1093
- [29] Szolek A, Schubert B, Mohr C, *et al.* OptiType: precision HLA typing from next-generation sequencing data. *Bioinformatics*, 2014, **30**(23): 3310-3316
- [30] Boegel S, Lower M, Schafer M, *et al.* HLA typing from RNA-Seq sequence reads. *Genome Medicine*, 2012, **4**(12): 102
- [31] Hundal J, Carreno B M, Petti A A, *et al.* pVAC-Seq: a genome-guided in silico approach to identifying tumor neoantigens. *Genome Medicine*, 2016, **8**(1): 11
- [32] Zhou Z, Lyu X, Wu J, *et al.* TSNAD: an integrated software for cancer somatic mutation and tumour-specific neoantigen detection. *Royal Society Open Science*, 2017, **4**(4): 170050
- [33] Bais P, Namburi S, Gatti D M, *et al.* CloudNeo: a cloud pipeline for identifying patient-specific tumor neoantigens. *Bioinformatics*, 2017, **33**(19): 3110-3112
- [34] Tappeiner E, Finotello F, Charoentong P, *et al.* TIminer: NGS data mining pipeline for cancer immunology and immunotherapy. *Bioinformatics*, 2017, **33**(19): 3140-3141
- [35] Kim S, Kim H S, Kim E, *et al.* Neopepsee: accurate genome-level prediction of neoantigens by harnessing sequence and amino acid immunogenicity information. *Annals of Oncology*, 2018, **29**(4): 1030-1036
- [36] Zhang J, Mardis E R, Maher C A. INTEGRATE-neo: a pipeline for personalized gene fusion neoantigen discovery. *Bioinformatics*, 2017, **33**(4): 555-557
- [37] Yadav M, Jhunjhunwala S, Phung Q T, *et al.* Predicting immunogenic tumour mutations by combining mass spectrometry and exome sequencing. *Nature*, 2014, **515**(7528): 572-576
- [38] Abelin J G, Keskin D B, Sarkizova S, *et al.* Mass spectrometry profiling of HLA-associated peptidomes in mono-allelic cells enables more accurate epitope prediction. *Immunity*, 2017, **46**(2): 315-326
- [39] Bulik-Sullivan B, Busby J, Palmer C D, *et al.* Deep learning using tumor HLA peptide mass spectrometry datasets improves neoantigen identification. *Nature Biotechnology*, 2018[Epublish ahead of print](DOI: 10.1038/nbt.4313)
- [40] Zhang J, Xin L, Shan B, *et al.* PEAKS DB: *de novo* sequencing assisted database search for sensitive and accurate peptide identification. *Molecular & Cellular Proteomics*, 2012, **11**(4): M111.010587
- [41] Tyanova S, Temu T, Cox J. The MaxQuant computational platform for mass spectrometry-based shotgun proteomics. *Nature Protocols*, 2016, **11**(12): 2301-2319
- [42] Perkins D N, Pappin D J, Creasy D M, *et al.* Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis*, 1999, **20**(18): 3551-3567
- [43] Eng J K, McCormack A L, Yates J R. An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *Journal of the American Society for Mass Spectrometry*, 1994, **5**(11): 976-989
- [44] Bjerregaard A M, Nielsen M, Jurtz V, *et al.* An analysis of natural T cell responses to predicted tumor neoepitopes. *Frontiers in*

- Immunology, 2017, **8**:1566
- [45] Balachandran V P, Luksza M, Zhao J N, *et al.* Identification of unique neoantigen qualities in long-term survivors of pancreatic cancer. *Nature*, 2017, **551**(7681): 512-516
- [46] Ott P A, Hu Z, Keskin D B, *et al.* An immunogenic personal neoantigen vaccine for patients with melanoma. *Nature*, 2017, **547**(7662): 217-221
- [47] Yadav M, Jhunjhunwala S, Phung Q T, *et al.* Predicting immunogenic tumour mutations by combining mass spectrometry and exome sequencing. *Nature*, 2014, **515**(7528): 572-576
- [48] Schumacher T N, Schreiber R D. Neoantigens in cancer immunotherapy. *Science*, 2015, **348**(6230): 69-74
- [49] Yin Q, Tang J, Zhu X. Next-generation sequencing technologies accelerate advances in T-cell therapy for cancer. *Briefings in Functional Genomics*, 2019, **18**(2): 119-128
- [50] Laumont C M, Vincent K, Hesnard L, *et al.* Noncoding regions are the main source of targetable tumor-specific antigens. *Science Translational Medicine*, 2018, **10**(470): pii eaau5516
- [51] Khurana E, Fu Y, Chakravarty D, *et al.* Role of non-coding sequence variants in cancer. *Nature Reviews Genetics*, 2016, **17**(2): 93-108

Advances in The Prediction of Antigenic Peptides in Personalized Tumor Neoantigen Vaccine*

WANG Guang-Zhi¹⁾, LI Yu-Yu¹⁾, XIE Lu^{1,2)**}

¹⁾College of Food Science, Shanghai Ocean University, Shanghai 201306, China;

²⁾Shanghai Center for Bioinformation Technology, Shanghai 201203, China)

Abstract Runaway mutations cause tumors, some of the non-synonymous mutational (i.e., missense, frameshift, fusion) peptides are degraded into short peptides by proteasome, then APCs (antigen-presenting cells) recognize and present them to draining lymph nodes. These short peptides, in line with the combination of MHC (major histocompatibility complex) motif, are captured by the T cell surface factors and produce an immune response, eventually lead to tumor regression. We call these peptides neoantigens, because these antigens are not negatively screened by thymus gland, they are recognized as "alien" by T cells and are not susceptible to immune tolerance mechanism. Neoantigens can act as effective targets for immune-mediated tumor control. The next generation sequencing technology greatly boosts the feasibility of neoantigen vaccine design, however identifying tumor somatic mutations that can be presented and recognized by TCR (T-cell receptor) is a very demanding yet key step. Many of the predicted "positive" neoantigen peptides are actually "false" and need to be eliminated in further steps to clinical application. Therefore effective screening method is an indispensable part of neoantigen vaccine therapy. However, no related reports have been seen in domestic. In this paper, the computational prediction process of screening methods for potential neoantigen peptides are reviewed.

Key words tumor, neoantigen, prediction, vaccine

DOI: 10.16476/j.pibb.2019.0019

* This work was supported by a grant from The National Natural Science Foundation of China (31870829).

** Corresponding author.

Tel: 86-21-20283705, E-mail: luxix2017@outlook.com

Received: January 25, 2019 Accepted: April 1, 2019