

肿瘤新抗原预测方法研究进展

吴静成, 陈枢青*, 周展* (浙江大学药学院, 浙江省抗肿瘤药物临床前研究重点实验室, 杭州 310058)

摘要: 肿瘤新抗原是一类肿瘤特异性抗原, 是潜在的肿瘤免疫治疗理想靶标。由于肿瘤新抗原产生及刺激 T 细胞应答的生理过程复杂, 如何高效地发现与鉴定肿瘤新抗原仍是一个巨大的挑战。随着肿瘤免疫基因组学数据的积累和人工智能预测方法的深入研究, 研究者能够借助人工智能新算法进行肿瘤新抗原预测方法的研究, 为基于肿瘤新抗原的肿瘤免疫治疗奠定坚实的基础。本文围绕肿瘤新抗原胞内处理、抗原提呈和 T 细胞识别等生理过程系统综述了肿瘤新抗原预测方法的进展。

关键词: 肿瘤新抗原; 肿瘤免疫治疗; 抗原提呈; 预测方法; 人工智能

中图分类号: R966 文献标志码: A 文章编号: 1007-7693(2022)21-2879-08

DOI: 10.13748/j.cnki.issn1007-7693.2022.21.026

引用本文: 吴静成, 陈枢青, 周展. 肿瘤新抗原预测方法研究进展[J]. 中国现代应用药学, 2022, 39(21): 2879-2886.

Advances in Tumor Neoantigen Prediction Methods

WU Jingcheng, CHEN Shuqing*, ZHOU Zhan* (Zhejiang University, College of Pharmaceutical Sciences, Zhejiang Key Laboratory of Preclinical Research on Antitumor Drugs, Hangzhou 310058, China)

ABSTRACT: Neoantigen is tumor specific antigen and a promising target for tumor immunotherapy. However, due to the complex physiological process of neoantigen generating and stimulating T cell response, it's still a huge challenge to discover and identify neoantigen efficiently. With the accumulation of tumor immunogenomics data and the in-depth study of artificial intelligence prediction methods, researchers could take advantage of artificial intelligence algorithms for neoantigen prediction methods, laying a solid foundation for tumor immunotherapy based on neoantigens. The research reviewed the advances in tumor neoantigen prediction methods for antigen processing and presenting, as well as antigen recognition by T cells.

KEYWORDS: tumor neoantigen; tumor immunotherapy; antigen presenting; prediction method; artificial intelligence

肿瘤在发生、发展过程中会累积大量的体细胞突变, 这些突变包括点突变、插入缺失、染色体易位等。突变所在的基因有可能表达有异于正常蛋白的突变蛋白, 这些突变蛋白在胞内经过一系列生理过程后降解为突变多肽, 然后与主要组织相容性复合体(major histocompatibility complex, MHC)结合, 提呈到细胞表面并刺激 T 细胞应答。这种来自肿瘤特异性蛋白编码区突变, 能够刺激 T 细胞应答的突变多肽与 MHC 复合物(pMHC)被称为肿瘤新抗原^[1]。长期以来, 肿瘤新抗原由于其高度的肿瘤特异性, 被视为极具前景的肿瘤免疫治疗靶标, 引起了基础及临床研究的广泛关注。临床研究表明, 基于肿瘤新抗原的个体化疫苗能在黑色素瘤和胶质母细胞瘤中产生新抗原特异性 T 细胞^[2-5], 可以有效降低黑色素瘤的复发率^[3], 并诱导黑色素瘤消退^[6]。基于肿瘤新抗原的过继性疗法不仅在多种原位肿瘤中发挥作用^[7-10], 同时能对转移性黑色素瘤患者产生治疗效果^[11-12]。此外, 在这些靶向新

抗原的疗法中, 极少见到针对野生型非突变肽的脱靶免疫反应^[13-15], 表明突变肽与野生肽之间不存在交叉反应^[16-20]。因此, 基于肿瘤新抗原的免疫疗法既有效又安全^[21-22]。

1 肿瘤新抗原预测全流程软件

传统鉴定肿瘤新抗原的方法通量低, 操作复杂, 因此研究者们借助生物信息学的方法以提高鉴定肿瘤新抗原的效率, 降低鉴定成本。目前已开发的肿瘤新抗原全流程预测软件包括 pVAC-seq^[23]、INTERGTATE-Neo^[24]、TSNAD^[25-26]、MuPeXI^[27]、CloudNeo^[28]、Timiner^[29]、ScanNeo^[30]、Neopepsee^[31]、pTuneos^[32]、NeoFuse^[33]、ASNEO^[34]、pVACtools^[35]等。

这些软件的基本思路一致。首先, 通过一系列的生物信息学软件(如 Trimmomatic^[36]、BWA^[37]、Samtools^[38]、GATK^[38]等)从二代测序的原始结果中鉴定出肿瘤特有的 DNA 突变, 并通过 Annovar^[38]或 VEP^[39]等注释软件得到对应的氨基酸突变; 其

基金项目: 国家自然科学基金项目(U20A20409, 31971371)

作者简介: 吴静成, 男, 博士, 助理研究员 E-mail: 11819049@zju.edu.cn *通信作者: 周展, 男, 博士, 副教授 E-mail: zhanzhou@zju.edu.cn 陈枢青, 男, 博士, 教授 E-mail: chenshuqing@zju.edu.cn

次,从测序数据通过 SOAP-HLA^[40]等软件鉴定出患者特有的 MHC 分子;随后,软件会提取出所有包含突变氨基酸的 8 到 11 肽,将其与 MHC 分子组合,通过多肽-HLA 结合预测软件(如 NetMHCpan^[41]、MHCflurry^[42]、DeepHLApan^[43]等)进行亲和力的预测。最后,不同的软件会根据不同的标准对预测得到的结果进行筛选,得到一系列潜在的肿瘤新抗原。

在这些软件中, pVAC-seq 是首个将肿瘤新抗原预测流程化的软件,主要提供点突变来源的肿瘤新抗原预测。INTERGTATE-Neo 则提供基因融合来源的肿瘤新抗原预测,它的主要思路是把自己前期开发的用于鉴定基因融合的软件 INTERGRATE 与后续预测新抗原的流程整合在一起。TSNAD 能支持全基因组/全外显子组,转录组数据的分析,并得到包含点突变、插入缺失、融合基因等不同突变来源的肿瘤新抗原。此外,其提供了网络在线工具和本地下载工具,方便研究者们使用。之后开发的一些软件中, CloudNeo 是第一个提供肿瘤新抗原预测云平台的工具, ScanNeo 提供了 INDEL 来源的肿瘤新抗原预测, ASNEO 提供了可变剪切来源的肿瘤新抗原预测。

目前开发的肿瘤新抗原预测软件在实际应用中效果并不理想。有研究表明现有软件预测出来的肿瘤新抗原中,只有很少一部分(1%~2%)能引起免疫反应^[1]。这就促使研究者们对肿瘤新抗原的生理过程进行更深入的研究。

2 肿瘤新抗原产生的生理过程相关研究进展

肿瘤新抗原从突变开始到引起 T 细胞免疫反应的过程见图 1。突变基因转录翻译后形成突变蛋白,其在胞内由蛋白酶体水解成多肽片段,其中包含突变的多肽片段不存在于正常细胞,是肿瘤新抗原的前体;该突变多肽由抗原加工相关转运蛋白(transporter associated with antigen processing, TAP)转运到内质网上,进一步被剪切成更短的短肽(只在氨基端被剪切,羧基端保持完整^[44-45]),之后与 MHC 分子结合,途径高尔基体,提呈到肿瘤细胞的表面;提呈的 pMHC 复合体会与 T 细胞表面受体(T cell receptor, TCR)相互识别,引起 T 细胞的级联反应,产生免疫应答。然而,突变蛋白在蛋白酶体水解下,会形成怎样的多肽片段;TAP 转运蛋白能转运怎样的多肽;多肽与 MHC 分子相互作用的特异性及强度;pMHC 复合物与 TCR 的结合有怎样的规律特征等问题仍是尚未解决的难

题,仍需研究者们坚持不懈的研究。

目前,肿瘤新抗原各个生理过程的研究取得了不同程度的进展,并开发了相应的方法或工具,结果见表 1。

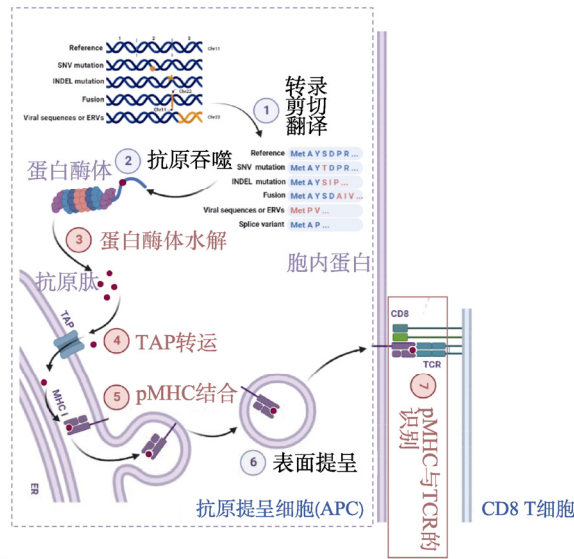


图 1 肿瘤新抗原产生及免疫识别的生理过程

Fig. 1 Physiological process of tumor neoantigen production and immune recognition

表 1 不同生理过程相关预测工具

Tab. 1 Prediction tools related to different physiological processes

生理过程	预测工具	模型方法
蛋白酶体水解	NetChop ^[46]	人工神经网络
	Cascleave ^[47-48]	支持向量回归
	PROSPER ^[49]	支持向量回归
	PROSPEROUS ^[50]	逻辑回归
	iProt-Sub ^[51]	支持向量回归
	DeepCalpain ^[52]	深度神经网络
	DeepCleave ^[53]	卷积神经网络、迁移学习
	Procleave ^[54]	条件随机场
	Pepsickle ^[55]	决策树、深度神经网络
	MPSC ^[56]	循环神经网络、迁移学习
	DeepDigest ^[57]	卷积神经网络、迁移学习
	3CLP ^[58]	随机森林
	SMM ^[59]	稳定矩阵法
	PredTAP ^[60]	人工神经网络、隐马尔可夫
TAP 转运	TAPPred ^[61]	支持向量机
	TAPREG ^[62]	支持向量机
	DeepTAP	循环神经网络
多肽与 MHC 结合	NetMHCpan ^[41,63-65]	人工神经网络
	NetMHCIIpan ^[66]	人工神经网络
	MixMHCpred ^[67]	解卷积神经网络
	PSSMHCpan ^[68]	位置特异性打分矩阵
	HLA-CNN ^[69]	卷积神经网络
	MHCflurry ^[42,70]	深度神经网络
	DeepHLApan ^[43]	循环神经网络
	EDGE ^[71]	深度神经网络
	MixMHC2pred ^[72]	解卷积神经网络
	MARIA ^[73]	多模态循环神经网络
TCR-pMHC 相互作用	DeepTCR ^[74]	变分自编码器
	NetTCR ^[75]	卷积神经网络
	pMTnet ^[76]	循环神经网络

2.1 肿瘤新抗原胞内处理过程相关研究

2.1.1 蛋白酶体水解研究 目前蛋白酶体水解相关的算法包括 NetChop^[46]、CasCleave^[47-48]、PROSPER^[49]、PROSPERous^[50]、iProt-Sub^[51]、DeepCalpain^[52]、DeepCleave^[53]、Procleave^[54]、Pepsickle^[55]、MPSC^[56]、DeepDigest^[57]、3CLP^[58]等,其中 NetChop 曾与预测其他肿瘤新抗原生理过程的软件一起结合用于预测肿瘤新抗原,但结果表明 NetChop 的加入会降低整体的效果^[77],其原因可能在于训练模型所用的数据稀少导致其泛化性差。MEROPS^[78]数据库是蛋白酶体相关算法进行训练时最常用的数据源,其包含了人体大部分的肽酶,以及这些肽酶能水解的底物、水解位点和水解产物。但蛋白酶体与肽酶之间并不完全等同, MEROPS 数据库里的数据在进行模型训练前需要进行预处理。此外,有研究表明蛋白质在蛋白酶体中被水解成的片段可能发生拼接,从而形成与原来蛋白质不一样的拼接多肽^[79]。这些多肽形成的 pMHC 复合物也能够引起肿瘤患者的免疫应答^[80],说明了拼接肽也是潜在的肿瘤新抗原前体。研究者进一步发现在能被 MHC I 型分子结合并呈送到细胞表面的多肽中,有 1/3 是拼接形成的多肽^[81]。虽然在随后的一项研究中, Mylonas 等^[82]认为如此高比例的拼接多肽可能是在分析质谱数据的过程中出现了偏差,将自然水解的多肽当成了拼接多肽。然而拼接多肽的存在,不管比例是高是低,仍然让蛋白酶体水解这一问题越发复杂,也更需要从 MEROPS 数据库之外获取更多可靠的数据,如 Specht 等^[83]提供的一个同时包含拼接肽和自然水解肽的数据集。

2.1.2 TAP 转运研究 用传统的实验方法来测量 TAP 与多肽的结合非常耗时且昂贵,随着生物信息学和计算机技术的发展,目前已经开发了部分预测多肽和 TAP 结合的计算工具。如 Zhang 等^[60]开发了一种结合人工神经网络和隐马尔可夫模型预测与人 TAP 结合的肽的工具 PredTAP; Bhasin 和 Raghava 利用基于支持向量机和级联支持向量机的方法来预测 TAP 配体的结合亲和力,生成了在线预测工具 TAPPred^[61]; Peters 等^[59]使用稳定矩阵法来预测 TAP 结合肽; Diez-Rivero 等^[62]建立了一种基于支持向量机的预测可变长度的 TAP 结合肽亲和力的在线工具 TAPREG 等;其中稳定矩阵法^[59]在 2005 年时由 Larsen 等^[84]稍加改进整合进

NetCTL 中用于表位的预测,遗憾的是整合的软件并未展现出更优越的性能^[77]。然而,由于 TAP 结合数据量少,数据质量不高,TAP 转运相关研究进展缓慢,仍有待深入研究。

2.2 多肽和 MHC 相互作用预测研究

在这一系列的生理过程中,多肽和 MHC 分子相互作用的研究最为深入。更准确地说,I 型 MHC 分子与多肽相互作用的研究较多,II 型 MHC 分子的研究相对较少。这是因为多肽和 MHC I 型相互作用的实验数据较多,不仅包含大量两者是否结合的数据,还有为数不少的两者之间亲和力的数据,这为构建预测多肽和 MHC 分子相互作用的预测模型提供了坚实的基础。同时,由于 I 型 MHC 分子能够结合的多肽长度在 8~11 肽,大部分集中在 9 肽,比 II 型 MHC 分子能结合的多肽长度更短,复杂性更低,得到了很多研究者的青睐。目前已经开发的相关算法包括但不限于 NetMHCpan^[41,63-65]、PSSMHCpan^[68]、HLA-CNN^[69]、DeepHLApan^[43]、MHCflurry^[42,70]、EDGE^[71]、MixMHCpred^[67]等。在这些算法中,NetMHCpan 开发最早,使用者最多,前文提及的全流程肿瘤新抗原预测软件中很多都将其作为多肽与 MHC 分型相互作用预测的工具,而这也是新抗原预测软件最核心的部分; DeepHLApan 同时考虑了多肽和 MHC 的相互作用与 pMHC 复合物的免疫原性,通过深度学习构建了肿瘤新抗原预测算法。该算法包含 2 个模型,分别为结合模型和免疫原性模型,其结合模型可以在未训练的 HLA 等位基因上获得良好的性能,并且在最新的 IEDB 基准数据集和独立的质谱数据集上具有与其他公认的工具相当的性能,而免疫原性模型得到的免疫原性评分可以显著提高新抗原的预测精度,见图 2。DeepHLApan 的性能表现表明了深度学习在肿瘤新抗原预测方法构建上的优越性。II 型 MHC 分子与多肽相互作用的软件较为稀少,包括 NetMHCIIpan^[66]、MixMHC2pred^[72]和 MARIA^[73],除 NetMHCIIpan 外,其余软件均为近期开发,这一方面说明了相关研究的难度较高,另一方面也表明此方面的研究仍大有可为。另外,还有研究者另辟蹊径,从 pMHC 复合物的稳定性出发对肿瘤新抗原的预测进行了研究^[85]。

2.3 TCR 和 pMHC 相互作用预测研究

由 TCR-pMHC 相互作用引起的 T 细胞激活是一个多参数的过程,它不仅受到结合常数的影响,

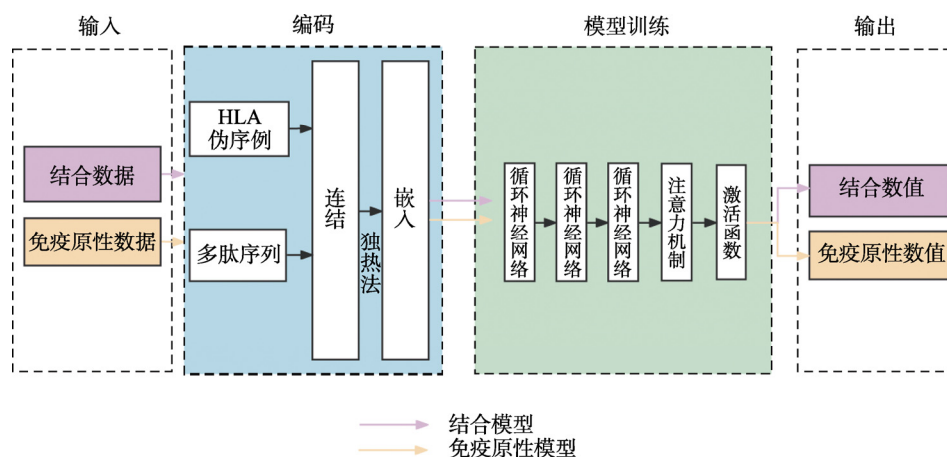


图2 基于深度学习的肿瘤新抗原预测算法 DeepHLApan 框架图

Fig. 2 Framework of DeepHLApan, a tumor neoantigen prediction algorithm based on deep learning

同时也会受到激活链的长度和动力学, 负反馈, 结合时引起的信号是持续还是有限的, pMHC 的浓度高低等因素的影响。例如, TCR 与 pMHC 复合物的解离时间(τ)一般与 T 细胞激活呈正相关, 即时间越长, T 细胞激活程度越高^[86-88], 但有研究表明这种正相关在 pMHC 含量很高时会消失^[89]。而激活 T 细胞的 pMHC 半数有效浓度(EC_{50})与 TCR-pMHC 的解离常数(K_d)息息相关^[90], 最大有效浓度则和解离率($k_{off}=1/\tau$)呈负相关^[91]。T 细胞激活与 TCR-pMHC 亲和力之间复杂的关系促使研究者们构建了不同的数学模型来解释发现的实验现象。这些数学模型基于不同的生理学假设, 推导相应的公式, 根据假设的不同, 其所能解释的能力也有所区别。最简单的模型是占有模型^[92], 它认为 T 细胞的激活程度与被 pMHC 结合的 TCR 的数量呈正相关, 一旦 TCR 与 pMHC 结合, 他们就会产生足够的信号促使 T 细胞激活。这个模型可以解释实验中发现的 EC_{50} 与 K_d 之间的强相关^[93]。而且该模型预测出 T 细胞的最大应答与 TCR-pMHC 结合参数无关, 与实验数据一致^[90]。其次是动态矫正模型^[94], 它假设 T 细胞的激活并不是在 pMHC 与 TCR 结合的瞬间就产生的, 而是需要结合一段时间, 因此它认为 T 细胞激活的程度是正比于被 pMHC 结合达到一定时间的 TCR 的比例。相比而言, 动态矫正模型更合理, 且支持占有模型的实验结果都支持该模型, 所以后续的模式优化都是以之为基础。之后的模型在动态矫正模型的基础上加上有限的信号^[95]、持久的信号^[89]、负反馈^[96]、pMHC 重结合^[97]、激活链的稳定性^[98]、有限信号与不相关前馈链的结合^[99]等各种假设,

进一步探讨了各种情形下 T 细胞应答与 TCR-pMHC 亲和力之间的相关性。然而, 由于构建这些表型模型时所用的数据有限, 有些实验数据甚至相互矛盾, 所以目前仍无法在 T 细胞应答和 TCR-pMHC 亲和力之间建立完全的联系^[92], 而临床试验表明, 在过继性细胞疗法中, 一味地增强两者的亲和力带来的不一定是更好的疗效, 而是严重的不良反应^[100-101]。因此, 需要更进一步对 TCR-pMHC 之间的相互作用在 T 细胞激活中的作用进行研究。

除了用于解释潜在机制的表位模型外, 有研究者基于现有的结构数据、序列数据利用分子动力学模拟、机器学习等方法构建用于预测 TCR-pMHC 相互作用的模型。如 PAComplex 可同时预测 pMHC 和 pTCR 之间相互作用的界面^[102]; TCRdist^[103]、GLIPH^[104]和 iSMART^[105]等工具通过聚类的方法来研究 TCR 的特异性, 即通过研究能与特定多肽或 pMHC 相互作用的 TCR 的序列特征, 来判断某个 TCR 是否会与特定多肽或 pMHC 相互识别; NetTCR^[75]、DeepTCR^[74]、pMTnet^[76]等则通过深度学习的方法对现有的 TCR-pMHC 数据进行训练, 利用深度学习强大的拟合能力, 进行 TCR-pMHC 相互作用模型的构建。然而, 目前所构建的这些模型只是一些初步的尝试, 离真正解决 TCR-pMHC 相互作用仍有一定的距离。

3 总结

肿瘤新抗原在肿瘤免疫治疗中起着非常重要的作用, 基于新抗原的多种免疫治疗方案在临床上均表现出优异的治疗效果。然而肿瘤新抗原在体内的生理过程复杂, 目前大部分肿瘤新抗原软

件预测出的结果在临床应用中仍存在较高的假阳性。这个问题一方面归因于生理过程数据的质量与数量，另一方面归因于训练模型的适用性。近些年来，得益于各种高通量技术的发展，肿瘤新抗原预测研究所需要的相关数据在迅速增加，这为进行肿瘤新抗原预测模型的构建提供了良好的基础；其次，由于深度学习强大的拟合能力，其在生物学领域得到了广泛应用，也使得利用深度学习的方法进行肿瘤新抗原预测成为未来研究的趋势。

基于生理过程构建的肿瘤新抗原预测算法涉及了底层的生物学机制，相应构建算法的思路不仅在肿瘤免疫靶标的发现上适用，在其他重大疾病如传染病、糖尿病、心脑血管疾病等重大疾病的免疫靶标发现上同样适用。对新抗原预测算法的持续研究有望为肿瘤及其他重大疾病的免疫治疗提供坚实的基础。

REFERENCES

- LEE C H, YELENSKY R, JOOSS K, et al. Update on tumor neoantigens and their utility: Why it is good to be different[J]. Trends Immunol, 2018, 39(7): 536-548.
- ŁUKSZA M, RIAZ N, MAKAROV V, et al. A neoantigen fitness model predicts tumour response to checkpoint blockade immunotherapy[J]. Nature, 2017, 551(7681): 517-520.
- OTT P A, HU Z T, KESKIN D B, et al. An immunogenic personal neoantigen vaccine for patients with melanoma[J]. Nature, 2017, 547(7662): 217-221.
- KESKIN D B, ANANDAPPA A J, SUN J, et al. Neoantigen vaccine generates intratumoral T cell responses in phase Ib glioblastoma trial[J]. Nature, 2019, 565(7738): 234-239.
- HILF N, KUTTRUFF-COQUI S, FRENZEL K, et al. Actively personalized vaccination trial for newly diagnosed glioblastoma[J]. Nature, 2019, 565(7738): 240-245.
- SAHIN U, DERHOVANESEAN E, MILLER M, et al. Personalized RNA mutanome vaccines mobilize poly-specific therapeutic immunity against cancer[J]. Nature, 2017, 547(7662): 222-226.
- TRAN E, TURCOTTE S, GROS A, et al. Cancer immunotherapy based on mutation-specific CD4⁺ T cells in a patient with epithelial cancer[J]. Science, 2014, 344(6184): 641-645.
- TRAN E, ROBBINS P F, LU Y C, et al. T-cell transfer therapy targeting mutant KRAS in cancer[J]. N Engl J Med, 2016, 375(23): 2255-2262.
- YEE C, THOMPSON J A, BYRD D, et al. Adoptive T cell therapy using antigen-specific CD8⁺ T cell clones for the treatment of patients with metastatic melanoma: *in vivo* persistence, migration, and antitumor effect of transferred T cells[J]. Proc Natl Acad Sci USA, 2002, 99(25): 16168-16173.
- ZACHARAKIS N, CHINNASAMY H, BLACK M, et al. Immune recognition of somatic mutations leading to complete durable regression in metastatic breast cancer[J]. Nat Med, 2018, 24(6): 724-730.
- GOFF S L, DUDLEY M E, CITRIN D E, et al. Randomized, prospective evaluation comparing intensity of lymphodepletion before adoptive transfer of tumor-infiltrating lymphocytes for patients with metastatic melanoma[J]. J Clin Oncol, 2016, 34(20): 2389-2397.
- ROSENBERG S A, YANG J C, SHERRY R M, et al. Durable complete responses in heavily pretreated patients with metastatic melanoma using T-cell transfer immunotherapy[J]. Clin Cancer Res, 2011, 17(13): 4550-4557.
- SCHUMACHER T N, SCHEPER W, KVISTBORG P. Cancer neoantigens[J]. Annu Rev Immunol, 2019(37): 173-200.
- STRØNEN E, TOEBES M, KELDERMAN S, et al. Targeting of cancer neoantigens with donor-derived T cell receptor repertoires[J]. Science, 2016, 352(6291): 1337-1341.
- TRAN E, ROBBINS P F, ROSENBERG S A. 'Final common pathway' of human cancer immunotherapy: Targeting random somatic mutations[J]. Nat Immunol, 2017, 18(3): 255-262.
- ALSPACH E, LUSSIER D M, MICELI A P, et al. MHC-II neoantigens shape tumour immunity and response to immunotherapy[J]. Nature, 2019, 574(7780): 696-701.
- GUBIN M M, ZHANG X L, SCHUSTER H, et al. Checkpoint blockade cancer immunotherapy targets tumour-specific mutant antigens[J]. Nature, 2014, 515(7528): 577-581.
- GUBIN M M, ARTYOMOV M N, MARDIS E R, et al. Tumor neoantigens: Building a framework for personalized cancer immunotherapy[J]. J Clin Invest, 2015, 125(9): 3413-3421.
- KREITER S, VORMEHR M, VAN DE ROEMER N, et al. Mutant MHC class II epitopes drive therapeutic immune responses to cancer[J]. Nature, 2015, 520(7549): 692-696.
- MATSUSHITA H, VESELY M D, KOBOLDT D C, et al. Cancer exome analysis reveals a T-cell-dependent mechanism of cancer immunoediting[J]. Nature, 2012, 482(7385): 400-404.
- SAHIN U, TÜRECI Ö. Personalized vaccines for cancer immunotherapy[J]. Science, 2018, 359(6382): 1355-1360.
- YAMAMOTO T N, KISHTON R J, RESTIFO N P. Developing neoantigen-targeted T cell-based treatments for solid tumors[J]. Nat Med, 2019, 25(10): 1488-1499.
- HUNDAL J, CARRENO B M, PETTI A A, et al. pVAC-Seq: A genome-guided in silico approach to identifying tumor neoantigens[J]. Genome Med, 2016, 8(1): 11.
- ZHANG J, MARDIS E R, MAHER C A. INTEGRATE-neo: a pipeline for personalized gene fusion neoantigen discovery[J]. Bioinformatics, 2017, 33(4): 555-557.
- ZHOU Z, LYU X Z, WU J C, et al. TSNAD: An integrated software for cancer somatic mutation and tumour-specific neoantigen detection[J]. R Soc Open Sci, 2017, 4(4): 170050.
- ZHOU Z, WU J C, REN J N, et al. TSNAD v2.0: A one-stop software solution for tumor-specific neoantigen detection[J]. Comput Struct Biotechnol J, 2021(19): 4510-4516.
- BJERREGAARD A M, NIELSEN M, HADRUP S R, et al. MuPeXI: prediction of neo-epitopes from tumor sequencing data[J]. Cancer Immunol Immunother, 2017, 66(9): 1157-1168.

- 1123-1130.
- [28] BAIS P, NAMBURI S, GATTI D M, et al. CloudNeo: A cloud pipeline for identifying patient-specific tumor neoantigens[J]. *Bioinformatics*, 2017, 33(19): 3110-3112.
 - [29] TAPPEINER E, FINOTELLO F, CHAROENTONG P, et al. TIminer: NGS data mining pipeline for cancer immunology and immunotherapy[J]. *Bioinformatics*, 2017, 33(19): 3140-3141.
 - [30] WANG T Y, WANG L, ALAM S K, et al. ScanNeo: identifying indel-derived neoantigens using RNA-Seq data[J]. *Bioinformatics*, 2019, 35(20): 4159-4161.
 - [31] KIM S, KIM H S, KIM E, et al. Neopepsee: Accurate genome-level prediction of neoantigens by harnessing sequence and amino acid immunogenicity information[J]. *Ann Oncol*, 2018, 29(4): 1030-1036.
 - [32] ZHOU C, WEI Z, ZHANG Z, et al. pTuneos: Prioritizing tumor neo antigens from next-generation sequencing data[J]. *Genome Med*, 2019(11): 67.
 - [33] FOTAKIS G, RIEDER D, HAIDER M, et al. NeoFuse: Predicting fusion neoantigens from RNA sequencing data[J]. *Bioinformatics*, 2020, 36(7): 2260-2261.
 - [34] ZHANG Z B, ZHOU C, TANG L H, et al. ASNEO: Identification of personalized alternative splicing based neoantigens with RNA-seq[J]. *Aging*, 2020, 12(14): 14633-14648.
 - [35] HUNDAL J, KIWALA S, MCMICHAEL J, et al. pVACtools: A computational toolkit to identify and visualize cancer neoantigens[J]. *Cancer Immunol Res*, 2020, 8(3): 409-420.
 - [36] BOLGER A M, LOHSE M, USADEL B. Trimmomatic: A flexible trimmer for Illumina sequence data[J]. *Bioinformatics*, 2014, 30(15): 2114-2120.
 - [37] LI H, DURBIN R. Fast and accurate long-read alignment with Burrows-Wheeler transform[J]. *Bioinformatics*, 2010, 26(5): 589-595.
 - [38] LI H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data[J]. *Bioinformatics*, 2011, 27(21): 2987-2993.
 - [39] MCLAREN W, GIL L, HUNT S E, et al. The ensembl variant effect predictor[J]. *Genome Biol*, 2016, 17(1): 122.
 - [40] LI R Q, YU C, LI Y R, et al. SOAP2: An improved ultrafast tool for short read alignment[J]. *Bioinformatics*, 2009, 25(15): 1966-1967.
 - [41] 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[J]. *J Immunol*, 2017, 199(9): 3360-3368.
 - [42] O'DONNELL T J, RUBINSTEYN A, BONSAK M, et al. MHCflurry: Open-source class I MHC binding affinity prediction[J]. *Cell Syst*, 2018, 7(1): 129-132.
 - [43] WU J C, WANG W Z, ZHANG J C, et al. DeepHLApan: A deep learning approach for neoantigen prediction considering both HLA-peptide binding and immunogenicity[J]. *Front Immunol*, 2019(10): 2559.
 - [44] CRAIU A, AKOPIAN T, GOLDBERG A, et al. Two distinct proteolytic processes in the generation of a major histocompatibility complex class I-presented peptide[J]. *Proc Natl Acad Sci USA*, 1997, 94(20): 10850-10855.
 - [45] SERWOLD T, GAW S, SHASTRI N. ER aminopeptidases generate a unique pool of peptides for MHC class I molecules[J]. *Nat Immunol*, 2001, 2(7): 644-651.
 - [46] KEŞMİR C, NUSSBAUM A K, SCHILD H, et al. Prediction of proteasome cleavage motifs by neural networks[J]. *Protein Eng*, 2002, 15(4): 287-296.
 - [47] SONG J N, TAN H, SHEN H B, et al. Cascleave: Towards more accurate prediction of caspase substrate cleavage sites[J]. *Bioinformatics*, 2010, 26(6): 752-760.
 - [48] WANG M J, ZHAO X M, TAN H, et al. Cascleave 2.0, a new approach for predicting caspase and granzyme cleavage targets[J]. *Bioinformatics*, 2014, 30(1): 71-80.
 - [49] SONG J N, TAN H, PERRY A J, et al. PROSPER: An integrated feature-based tool for predicting protease substrate cleavage sites[J]. *PLoS One*, 2012, 7(11): e50300.
 - [50] SONG J N, LI F Y, LEIER A, et al. PROSPERous: High-throughput prediction of substrate cleavage sites for 90 proteases with improved accuracy[J]. *Bioinformatics*, 2018, 34(4): 684-687.
 - [51] SONG J N, WANG Y N, LI F Y, et al. iProt-Sub: A comprehensive package for accurately mapping and predicting protease-specific substrates and cleavage sites[J]. *Brief Bioinform*, 2019, 20(2): 638-658.
 - [52] LIU Z X, YU K, DONG J S, et al. Precise prediction of calpain cleavage sites and their aberrance caused by mutations in cancer[J]. *Front Genet*, 2019(10): 715.
 - [53] LI F Y, CHEN J X, LEIER A, et al. DeepCleave: A deep learning predictor for caspase and matrix metalloprotease substrates and cleavage sites[J]. *Bioinformatics*, 2020, 36(4): 1057-1065.
 - [54] LI F Y, LEIER A, LIU Q Z, et al. Procleave: Predicting protease-specific substrate cleavage sites by combining sequence and structural information[J]. *Genomics Proteomics Bioinformatics*, 2020, 18(1): 52-64.
 - [55] WEEDER B R, WOOD M A, LI E, et al. Pepsickle rapidly and accurately predicts proteasomal cleavage sites for improved neoantigen identification[J]. 2021. Doi: 10.1093/bioinformatics/btab628.
 - [56] OZOLS M, ECKERSLEY A, PLATT C I, et al. Predicting proteolysis in complex proteomes using deep learning[J]. *Int J Mol Sci*, 2021, 22(6): 3071.
 - [57] YANG J H, GAO Z Q, REN X H, et al. DeepDigest: Prediction of protein proteolytic digestion with deep learning[J]. *Anal Chem*, 2021, 93(15): 6094-6103.
 - [58] CHEN H T, ZHU Z Z, QIU Y, et al. Prediction of coronavirus 3C-like protease cleavage sites using machine-learning algorithms[J]. *Viol Sin*, 2022, 37(3): 437-444.
 - [59] PETERS B, BULIK S, TAMPE R, et al. Identifying MHC class I epitopes by predicting the TAP transport efficiency of epitope precursors[J]. *J Immunol*, 2003, 171(4): 1741-1749.
 - [60] ZHANG G L, PETROVSKY N, KWON C K, et al. PRED(TAP): A system for prediction of peptide binding to the human transporter associated with antigen processing[J]. *Immunome Res*, 2006(2): 3.
 - [61] BHASIN M, LATA S, RAGHAVA G P S. TAPPred prediction of TAP-binding peptides in antigens[J]. *Methods*

- Mol Biol, 2007(409): 381-386.
- [62] DIEZ-RIVERO C M, CHENLO B, ZULUAGA P, et al. Quantitative modeling of peptide binding to TAP using support vector machine[J]. *Proteins*, 2010, 78(1): 63-72.
 - [63] NIELSEN M, LUNDEGAARD C, BLICHER T, et al. NetMHCpan, a method for quantitative predictions of peptide binding to any HLA-A and-B locus protein of known sequence[J]. *PLoS One*, 2007, 2(8): e796.
 - [64] HOOFF I, PETERS B, SIDNEY J, et al. NetMHCpan, a method for MHC class I binding prediction beyond humans[J]. *Immunogenetics*, 2009, 61(1): 1-13.
 - [65] NIELSEN M, ANDREATTA M. NetMHCpan-3.0; improved prediction of binding to MHC class I molecules integrating information from multiple receptor and peptide length datasets[J]. *Genome Med*, 2016, 8(1): 33.
 - [66] KAROSIENE E, RASMUSSEN M, BLICHER T, et al. NetMHCIIpan-3.0, a common pan-specific MHC class II prediction method including all three human MHC class II isotypes, HLA-DR, HLA-DP and HLA-DQ[J]. *Immunogenetics*, 2013, 65(10): 711-724.
 - [67] BASSANI-STERMBERG M, CHONG C, GUILLAUME P, et al. Deciphering HLA-I motifs across HLA peptidomes improves neo-antigen predictions and identifies allosteric regulating HLA specificity[J]. *PLoS Comput Biol*, 2017, 13(8): e1005725.
 - [68] LIU G, LI D L, LI Z, et al. PSSMHCPan: A novel PSSM-based software for predicting class I peptide-HLA binding affinity[J]. *GigaScience*, 2017, 6(5): 1-11.
 - [69] VANG Y S, XIE X H. HLA class I binding prediction via convolutional neural networks[J]. *Bioinformatics*, 2017, 33(17): 2658-2665.
 - [70] O'DONNELL T J, RUBINSTEYN A, LASERSON U. MHCflurry 2.0: Improved pan-allele prediction of MHC class I-presented peptides by incorporating antigen processing[J]. *Cell Syst*, 2020, 11(4): 418-419.
 - [71] BULIK-SULLIVAN B, BUSBY J, PALMER C D, et al. Deep learning using tumor HLA peptide mass spectrometry datasets improves neoantigen identification[J]. *Nat Biotechnol*, 2018. Doi: 10.1038/nbt.4313.
 - [72] RACLE J, MICHAUX J, ROCKINGER G A, et al. Robust prediction of HLA class II epitopes by deep motif deconvolution of immunopeptidomes[J]. *Nat Biotechnol*, 2019, 37(11): 1283-1286.
 - [73] CHEN B B, KHODADOUST M S, OLSSON N, et al. Predicting HLA class II antigen presentation through integrated deep learning[J]. *Nat Biotechnol*, 2019, 37(11): 1332-1343.
 - [74] SIDHOM J W, LARMAN H B, PARDOLL D M, et al. DeepTCR is a deep learning framework for revealing sequence concepts within T-cell repertoires[J]. *Nat Commun*, 2021, 12(1): 1605.
 - [75] MONTEMURRO A, SCHUSTER V, POVLSEN H R, et al. NetTCR-2.0 enables accurate prediction of TCR-peptide binding by using paired TCR α and β sequence data[J]. *Commun Biol*, 2021, 4(1): 1060.
 - [76] LU T S, ZHANG Z, ZHU J, et al. Deep learning-based prediction of the T cell receptor-antigen binding specificity[J]. *Nat Mach Intell*, 2021, 3(10): 864-875.
 - [77] KOŞALOĞLU-YALÇIN Z, LANKA M, FRENTZEN A, et al. Predicting T cell recognition of MHC class I restricted neoepitopes[J]. *Oncoimmunology*, 2018, 7(11): e1492508.
 - [78] RAWLINGS N D, BARRETT A J, THOMAS P D, et al. The MEROPS database of proteolytic enzymes, their substrates and inhibitors in 2017 and a comparison with peptidases in the PANTHER database[J]. *Nucleic Acids Res*, 2018, 46(D1): D624-D632.
 - [79] LIEPE J, MISHTO M, TEXTORIS-TAUBE K, et al. The 20S proteasome splicing activity discovered by SpliceMet[J]. *PLoS Comput Biol*, 2010, 6(6): e1000830.
 - [80] MISHTO M, MANSURKHODZHAEV A, YING G, et al. An in silico-in vitro pipeline identifying an HLA-A*02: 01 + KRAS G12V+ spliced epitope candidate for a broad tumor-immune response in cancer patients[J]. *Front Immunol*, 2019, 10: 2572.
 - [81] LIEPE J, MARINO F, SIDNEY J, et al. A large fraction of HLA class I ligands are proteasome-generated spliced peptides[J]. *Science*, 2016, 354(6310): 354-358.
 - [82] MYLONAS R, BEER I, ISELI C, et al. Estimating the contribution of proteasomal spliced peptides to the HLA-I ligandome[J]. *Mol Cell Proteomics*, 2018, 17(12): 2347-2357.
 - [83] SPECHT G, ROETSCHKE H P, MANSURKHODZHAEV A, et al. Large database for the analysis and prediction of spliced and non-spliced peptide generation by proteasomes[J]. *Sci Data*, 2020, 7(1): 146.
 - [84] LARSEN M, LUNDEGAARD C, LAMBERTH K, et al. An integrative approach to CTL epitope prediction: A combined algorithm integrating MHC class I binding, TAP transport efficiency, and proteasomal cleavage predictions[J]. *Eur J Immunol*, 2005, 35(8): 2295-2303.
 - [85] JØRGENSEN K W, RASMUSSEN M, BUUS S, et al. NetMHCstab-predicting stability of peptide-MHC-I complexes; impacts for cytotoxic T lymphocyte epitope discovery[J]. *Immunology*, 2014, 141(1): 18-26.
 - [86] KALERGIS A M, BOUCHERON N, DOUCEY M A, et al. Efficient T cell activation requires an optimal dwell-time of interaction between the TCR and the pMHC complex[J]. *Nat Immunol*, 2001, 2(3): 229-234.
 - [87] COOMBS D, KALERGIS A M, NATHENSON S G, et al. Activated TCRs remain marked for internalization after dissociation from pMHC[J]. *Nat Immunol*, 2002, 3(10): 926-931.
 - [88] IRVING M, ZOETE V, HEBEISEN M, et al. Interplay between T cell receptor binding kinetics and the level of cognate peptide presented by major histocompatibility complexes governs CD8+ T cell responsiveness[J]. *J Biol Chem*, 2012, 287(27): 23068-23078.
 - [89] GONZÁLEZ P A, CARREÑO L J, COOMBS D, et al. T cell receptor binding kinetics required for T cell activation depend on the density of cognate ligand on the antigen-presenting cell[J]. *Proc Natl Acad Sci USA*, 2005, 102(13): 4824-4829.
 - [90] DUSHEK O, ALEKSIC M, WHEELER R J, et al. Antigen potency and maximal efficacy reveal a mechanism of efficient T cell activation[J]. *Sci Signal*, 2011, 4(176): ra39.
 - [91] ZHONG S, MALECEK K, JOHNSON L A, et al. T-cell

- receptor affinity and avidity defines antitumor response and autoimmunity in T-cell immunotherapy[J]. *Proc Natl Acad Sci USA*, 2013, 110(17): 6973-6978.
- [92] LEVER M, MAINI P K, VAN DER MERWE P A, et al. Phenotypic models of T cell activation[J]. *Nat Rev Immunol*, 2014, 14(9): 619-629.
- [93] TIAN S M, MAILE R, COLLINS E J, et al. CD8⁺ T cell activation is governed by TCR-peptide/MHC affinity, not dissociation rate[J]. *J Immunol*, 2007, 179(5): 2952-2960.
- [94] MCKEITHAN T W. Kinetic proofreading in T-cell receptor signal transduction[J]. *Proc Natl Acad Sci USA*, 1995, 92(11): 5042-5046.
- [95] VARMA R, CAMPI G, YOKOSUKA T, et al. T cell receptor-proximal signals are sustained in peripheral microclusters and terminated in the central supramolecular activation cluster[J]. *Immunity*, 2006, 25(1): 117-127.
- [96] FRANÇOIS P, VOISINNE G, SIGGIA E D, et al. Phenotypic model for early T-cell activation displaying sensitivity, specificity, and antagonism[J]. *Proc Natl Acad Sci USA*, 2013, 110(10): E888-E897.
- [97] DUSHEK O, VAN DER MERWE P A. An induced rebinding model of antigen discrimination[J]. *Trends Immunol*, 2014, 35(4): 153-158.
- [98] GÁLVEZ J, GÁLVEZ J J, GARCÍA-PEÑARRUBIA P. TCR/pMHC interaction: Phenotypic model for an unsolved Enigma[J]. *Front Immunol*, 2016(7): 467.
- [99] LEVER M, LIM H S, KRUGER P, et al. Architecture of a minimal signaling pathway explains the T-cell response to a 1 million-fold variation in antigen affinity and dose[J]. *Proc Natl Acad Sci USA*, 2016, 113(43): E6630-E6638.
- [100] MORGAN R A, CHINNASAMY N, ABATE-DAGA D, et al. Cancer regression and neurological toxicity following anti-MAGE-A3 TCR gene therapy[J]. *J Immunother*, 2013, 36(2): 133-151.
- [101] LINETTE G P, STADTMAUER E A, MAUS M V, et al. Cardiovascular toxicity and titin cross-reactivity of affinity-enhanced T cells in myeloma and melanoma[J]. *Blood*, 2013, 122(6): 863-871.
- [102] LIU I H, LO Y S, YANG J M. PA Complex: A web server to infer peptide antigen families and binding models from TCR-pMHC complexes[J]. *Nucleic Acids Res*, 2011, 39(Web Server issue): W254-W260.
- [103] DASH P, FIORE-GARTLAND A J, HERTZ T, et al. Quantifiable predictive features define epitope-specific T cell receptor repertoires[J]. *Nature*, 2017, 547(7661): 89-93.
- [104] GLANVILLE J, HUANG H, NAU A, et al. Identifying specificity groups in the T cell receptor repertoire[J]. *Nature*, 2017, 547(7661): 94-98.
- [105] ZHANG H Y, LIU L C, ZHANG J, et al. Investigation of antigen-specific T-cell receptor clusters in human cancers[J]. *Clin Cancer Res*, 2020, 26(6): 1359-1371.

收稿日期: 2022-08-22

(本文责编: 沈倩)