



## 分子细胞科学卓越创新中心吴立刚组开发深度学习模型CapHLA预测肿瘤新生抗原

发布时间: 2024-12-27 【字体: 大 中 小】 [打印](#) [关闭](#)

12月23日，国际学术期刊*Briefings in Bioinformatics*在线发表了中国科学院分子细胞科学卓越创新中心(生物化学与细胞生物学研究所)吴立刚研究组的最新研究成果“CapHLA: a comprehensive tool to predict peptide presentation and binding to HLA class I and class II”。该研究开发了深度学习模型CapHLA及新生抗原(neoantigen)质量评价指标PAE，为肿瘤新生抗原疫苗的设计开发提供了重要工具。

肿瘤细胞的基因组或转录组发生变异而产生的带有特异性氨基酸序列变异的多肽被称为新生抗原。新生抗原是一类肿瘤特异性抗原(TSAs)，并不存在于正常的组织器官当中。新生抗原与人类白细胞抗原(HLA)分子结合形成peptide-HLA(pHLA)复合物，并呈递到细胞表面，这一过程是T细胞特异性识别肿瘤细胞并引发特异性免疫反应的关键。HLA通常分为HLA-I和HLA-II两大类。HLA-I结合细胞内8-15个氨基酸长度的多肽，之后被CD8+ T识别。HLA-II结合12-20个氨基酸长度的多肽，随后被CD4+ T细胞识别。CD4+和CD8+ T细胞均在肿瘤杀伤中发挥了重要作用。肿瘤新生抗原疫苗目前已经在胰腺癌、黑色素瘤、小细胞肺癌等多种肿瘤的临床实验中展现出优异的药效，而新生抗原选择的准确性是决定肿瘤疫苗药效的关键因素。但目前已发表的多种新生抗原预测算法存在诸多不足：准确度不够、对长度有限制、只能预测HLA-I类或HLA-II类等。

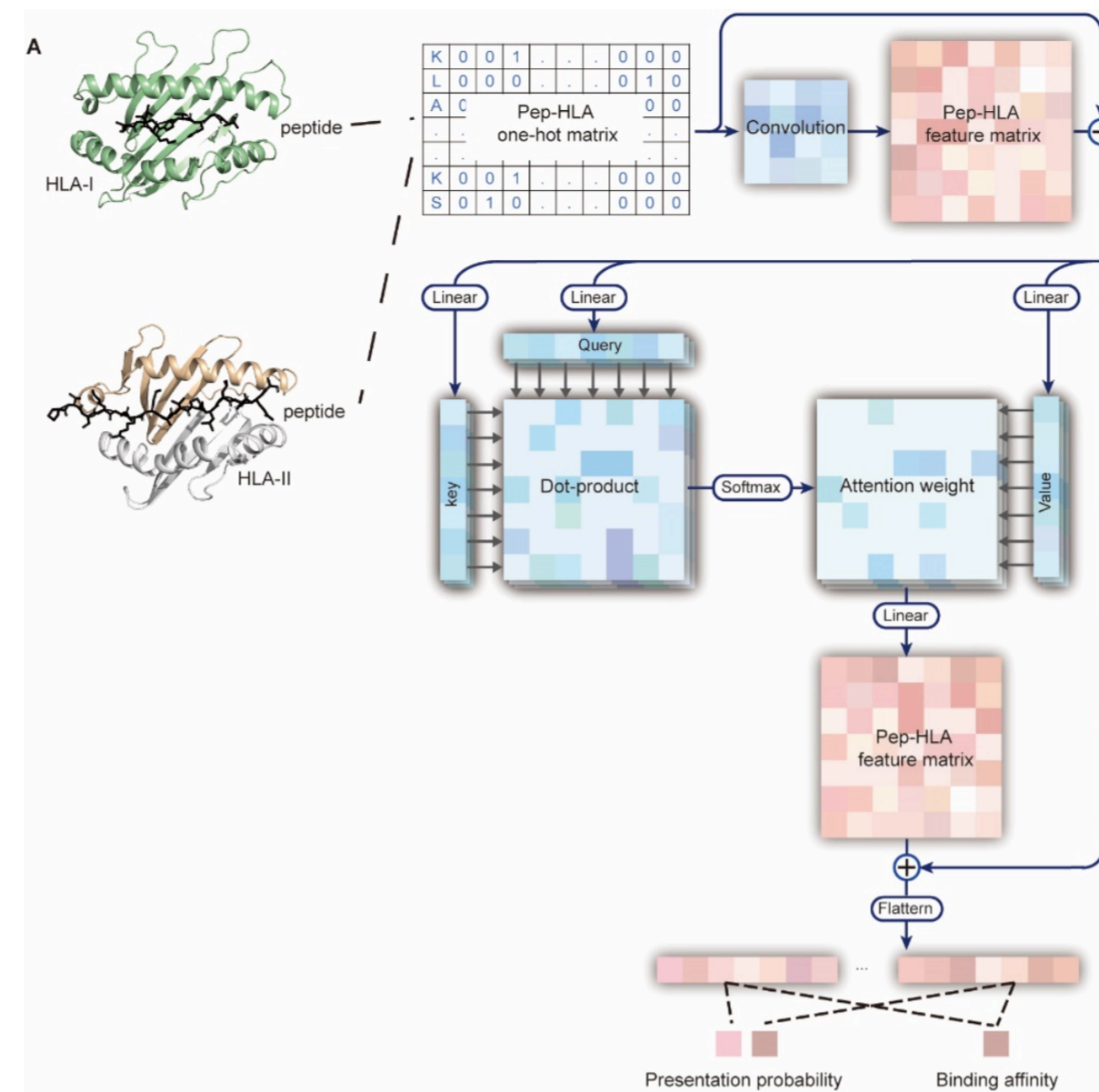
该研究中，研究人员开发了新的深度学习模型CapHLA。CapHLA由四个模块组成：编码层将多肽和HLA结合位点转换为one-hot数字矩阵；卷积层捕捉蛋白酶水解特征和抗原处理相关转运蛋白转运多肽过程中的motif信息；attention层捕捉不同HLA基因型与多肽结合的特异氨基酸互作信息；特征选择层综合处理上述步骤得到的特征并获得分析结果。CapHLA算法可以同时预测各种长度多肽以及HLA-I和HLA-II各种基因型的亲和性及呈递可能性。

研究人员在外部独立验证集上对AUROC、AUPRC、accuracy、MAE等指标进行了分析，CapHLA相较于现有其他预测算法均展现出更准确的预测性能，尤其是在HLA-II和较长多肽的预测能力上实现了大幅提升。研究人员进一步分析经验证具有高免疫原性的新生抗原，发现既需要pHLA结合亲和性高，多肽递呈概率大，还需要多肽表达水平较高。因此研究人员开发了结合这三者参数的新生抗原质量评价指标PAE，取得了更好的预测效果，为肿瘤新生抗原的选择和评价提供了重要工具。研究人员进一步展示了PAE指标还具有预测肿瘤患者免疫治疗疗效的潜力。通过对多组免疫治疗队列数据进行分析，发现大多数队列中PAE值高的病人具有更长的生存期以及更好的免疫治疗响应。在一些癌种中，即使未使用免疫治疗，PAE分数高的患者也有更长的生存期，说明PAE指标可以用于评估肿瘤中新抗原的质量。

分子细胞卓越中心博士研究生常允建为论文第一作者。分子细胞卓越中心吴立刚研究员为该论文的通讯作者。该研究得到分子细胞卓越中心生物信息学平台和高性能计算存储与网络服务中心的大力支持，同时得到了国家自然科学基金委、科技部、中国科学院、上海市的经费支持。

文章链接: <https://academic.oup.com/bib/article/26/1/bbae595/7926360>





Schematic of the CapHLA model

- >> [上一篇: 宁波材料所在有机自旋电子学领域取得进展](#)
- >> [下一篇: 城市环境研究所在入侵水华蓝藻\(拉氏尖头藻\)全球扩散与分布方面取得进展](#)

