
科学家利用机器学习让耐药检测更高效

作者：writer 来源：中国科学院

本文原地址：<https://www.iikx.com/news/progress/6048.html>

本文仅供学习交流之用，版权归原作者所有，请勿用于商业用途！

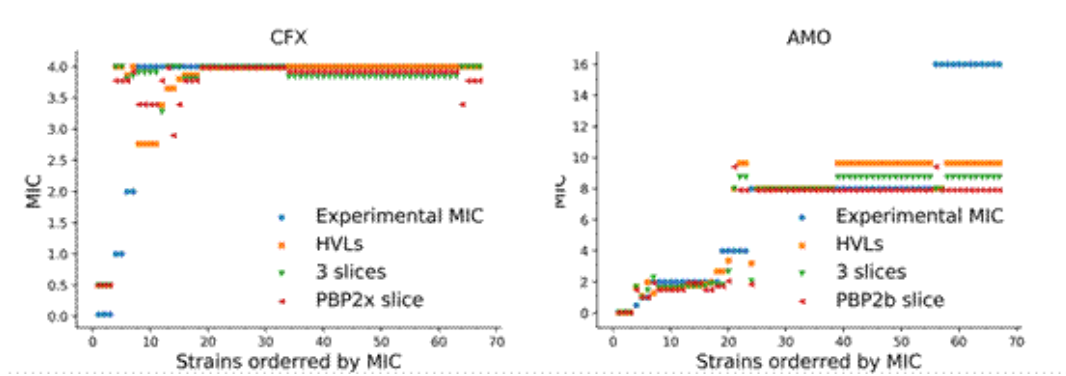
科学家利用机器学习让耐药检测更高效。细菌耐药已成为影响全人类健康的重大问题，引起了全世界广泛的关注。世界卫生组织提出的解决耐药措施之一是研发耐药快速准确的新型诊断技术和相关试剂。传统的检测方法基于细菌培养，周期长，易导致漏诊、误诊，延误最佳治疗时机。而基于基因的检测技术，如具有灵敏、高效、快捷特点的基因芯片、数字PCR等技术，是公认的快速检测技术。然而，到目前为止，由于耐药基因型与表型结果的不一致，使得基因检测只能作为培养法的辅助手段用于耐药的检测。

中国科学院微生物研究所冯婕研究组等针对肺炎链球菌β-内酰胺耐药这一重要临床问题，采用机器学习的方法挖掘耐药相关数据的规律，建立了基因型和表型之间的联系，使得基因检测不再是一个辅助手段，而有望成为一种主要的耐药快速检测技术。

肺炎链球菌β-内酰胺耐药的主要机制是三种青霉素结合蛋白(PBP1a, PBP2b和PBP2x)的转肽酶结构域(TPD)的改变。由于不同临床肺炎链球菌分离株PBPs的高度变异性，以及链球菌间重组导致的嵌合结构，使得PBPs极具多样化，导致了很难将PBPs的突变与临床耐药性联系起来。冯婕组研究人员首先将NCBI数据库已公布的PBPs序列通过类别方差法计算，得到了139个与耐药高度相关的HVLs (highly variant amino acid)。再以4300株肺炎链球菌的转肽酶结构域(TPD)序列以及对应头孢唑啉、阿莫西林的耐药表型作为数据库，将其中80%的数据作为训练集，20%的数据作为检验集，用HVLs去预测头孢唑啉和阿莫西林的耐药水平，结果发现与用PBPs蛋白的TPD序列预测效果一样好。进一步分析发现，HVLs与PBPs的某些区域的序列有很强的相关性。因此，分别使用来自pbp2x (2253 bp)的750 bp片段和来自pbp2b (2058 bp)的750 bp片段可以很好地预测头孢唑啉和阿莫西林的耐药性。这种长度只需要一个Sanger测序反应即可，不仅使检测操作更加简单，也降低了成本。此外，通过对人工构建的突变体和来自更多临床分离的菌株的耐药表型的检测，进一步确认了机器学习法能精确预测耐药表型。应用该预测方法，研究人员分析了NCBI数据库中已测序的8138株肺炎链球菌，进而建立了耐药表型、血清型以及ST型之间的关联，促进了对肺炎链球菌的流行病学的认识。

该研究成果在线发表于Briefings in Bioinformatics杂志，冯婕与南方科技大学教授杨亮为共同通讯作者。该研究得到国家自然科学基金和北京市科学技术委员会的资助。

论文链接



机器学习预测实验菌株的耐药水平

更多 科学进展 请访问 <https://www.iikx.com/news/progress/>

本文版权归原作者所有，请勿用于商业用途，[爱科学iikx.com](https://www.iikx.com)转发