青岛能源所开发出基于微生物组大数据的疾病检测方法

作者：writer 来源：中国科学院

本文原地址：https://www.iikx.com/news/progress/8739.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！

微生物组具有服务疾病诊治与生态监控的巨大潜力，但是其影响因素错综复杂。如何通过菌群检测实现快速精准的疾病诊断呢？中国科学院青岛生物能源与过程研究所单细胞中心发明了基于菌群大数据搜索的疾病检测方法，为此共性问题提供了原创的解决方案。该工作于3月17日在线发表于*mSystems*（美国微生物学会会刊）。

作为与生俱来、无处不在的“小伙伴”，微生物组与人体、环境的健康有着千丝万缕的关联。同时，菌群检测具有非侵入性、可量化、可预警等优势。因此，如何用菌群来判断和识别人体或环境的健康状态，一直是精准医学和大健康的热点问题之一。但是，作为一种疾病诊断手段，菌群面临着三个关键的挑战。

首先，“漏诊”问题：现有的检测手段通常只针对特定疾病，并依赖于该疾病已知的标识微生物来构建检测模型。然而很多疾病尚无明确标识物，而且同一种疾病在不同人群中的生物标识物种经常不尽一致，因此难以普适性地判断待检样本是否健康。其次，“误诊”问题：同一种生物标识经常与多种疾病相关联，导致难以准确地分辨具体的疾病类型。此外，由于元基因组数据的极端高通量、高度异质性、多种测序平台共存等特点，菌群诊断模型的构建、训练和更新，均具有突出的挑战性。

针对上述关键问题，副研究员苏晓泉带领的单细胞中心组群生物信息研究组，利用前期开发的微生物组搜索引擎MSE (Microbiome Search Engine; http://mse.ac.cn
)，发明了基于菌群大数据搜索的疾病检测新策略（图1）。与现有基于模型训练的机器学习手段相比，该方法不再依赖于疾病相关的特定标识微生物，而是利用待测菌群元基因组与已有菌群元基因组大数据的整体相似度来实现检测。首先，计算待测样本相对于数据库中所有健康样本的异常程度，即可评估其是否健康。其次，与数据库中的多种疾病的参照样本进行比对，即可识别具体的患病种类。基于超过3000例肠道菌群样本的测试表明，针对炎症性肠病（IBD）、结直肠癌（CRC）、艾滋病毒感染（HIV）和肠腹泻病（EDD）等，MSE在回答“是否健康”和“哪种疾病”这两个问题上的准确率均超过80%，显著高于目前常用的机器学习算法，从而有效地降低了“漏诊”和“误诊”的可能。另一方面，与常规手段相比，该方法在跨研究、跨测试人群、跨测序平台、样品污染等因素影响下的适应性和抗干扰能力等方面，也具有显著优势。

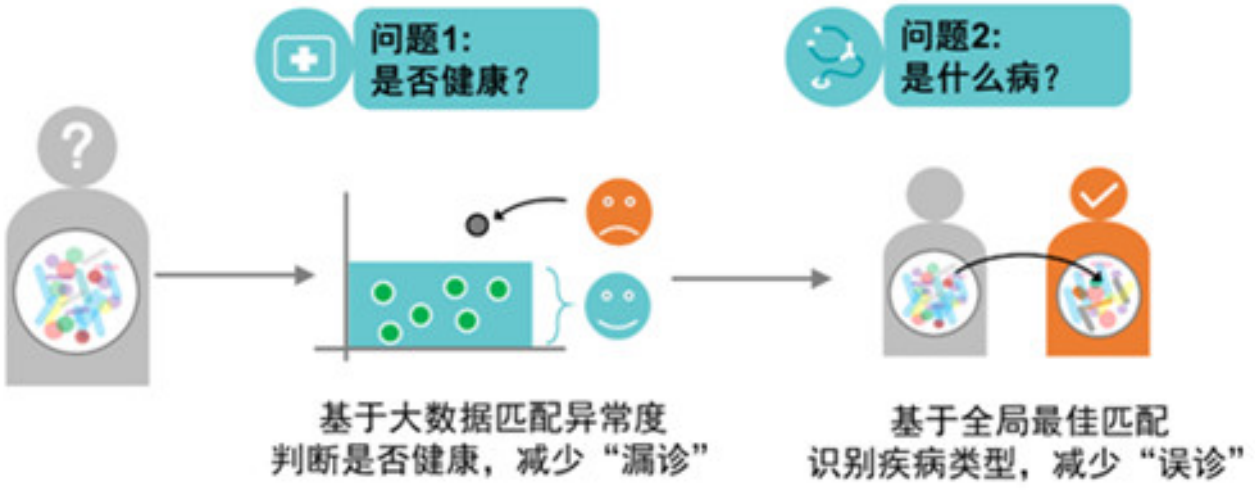
此外，得益于其强大的高通量菌群结构搜索能力，MSE的使用与百度、谷歌一样简单，通常在0.5秒内即能完成诊断。同时，MSE能够通过其菌群大数据系统的实时更新，来提升检测的可靠性以及拓宽其适用范围。因此，随着微生物组测序数据的爆发式增长，该方法在实际应用中的便捷

性、可延展性等特色将愈发突出。

单细胞中心开发的MSE是目前国内外最大的微生物组数据科学研究体系之一，其参照数据库包括来自肠道、皮肤、口腔、室内环境、海洋、土壤等广袤生态系统的超过23万例高质量、标准化、可比较的元基因组样本。与本研究提出的诊断手段相结合，MSE为建立针对各种慢性疾病或生态灾害的菌群诊断系统和干预手段效能评价体系奠定了一个共性的方法学基础。

该研究由单细胞中心与加州大学圣迭戈分校微生物组创新中心合作完成，并得到了中科院微生物组计划、山东省自然科学基金重大基础研究项目、国家自然科学基金等的支持。

论文链接



图：基于菌群大数据搜索引擎实现疾病检测的原理

研究团队单位：青岛生物能源与过程研究所

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转发