
汉族10万人基因组数据库及在线分析平台发布

作者：writer 来源：中国科学院

本文原地址：<https://www.iikx.com/news/progress/8050.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！

1月8日，国际学术期刊《核酸研究》（Nucleic Acids Research

）正式发表了中国科学院上海营养与健康研究所/马普计算生物学研究所徐书华课题组与生物医学大数据中心的研究成果“PGG.Han: the Han Chinese genome database and analysis platform”。该研究作为十万人汉族基因组计划一期成果，收集并分析了11万汉族人基因组单核苷酸变异数据，并发布了汉族人群的遗传数据库及在线分析平台——“PGG.Han”（<http://www.pgghan.org>），此举填补了中国人群在大规模可用参考基因组数据集上的空白，并提供了免费在线计算分析平台。

众所周知，不同人群和个体由于其遗传背景不同，大多数性状或疾病的遗传基础和分子机制存在不同程度的差异。随着研究的深入和规模的增长，对大型人群参考数据集的需求越来越迫切。近年来，世界各国都在着手建立针对性的人群特异性参考数据集。早在2012年，英国就启动了万人基因组计划（UK10K），并取得了显著的成果。紧随其后的美国精准医学项目（TOPMed）对个体化医疗的发展起到了支撑性作用。其他地区包括亚洲不少国家也都先后启动了类似的国家计划。近年来的研究也逐步凸显了大规模人群特异性参考数据集在复杂疾病的深入解析以及罕见病相关基因突变的判定等方面的重要性。但是，我国作为人口大国以及汉族作为世界上人口最多的族群，长期以来尚未建立适合国情的大型参照数据集，这个状况直到我国的精准医学计划启动三年以后也未得到改善。

经过多年的努力和积累，徐书华研究组联合多家研究单位，促成10万人汉族基因组计划（The Han100K

Initiative）的启动和

实施。Han100K计划的成员单位和合作

者可以通过链接（

<https://www.hanchinese genomes.org/HCGD/about>

）查

询，并且

处于持续更新中。

作为该计划一期成果，汉族基因组数

据库及在线分析平台（PGG.Han）肩负着重要使命。PGG

.Han是目前唯一针对汉族人群的、也是最大规模的、集数据存储、发布与在线分析为一体的参考基因组数据平台。当前上线的PGG

.Han版本主要基于对早前基因组数据的收集和分析整理，更专注于提供平台和构架，为后期持续

纳入的高质量深度测序数据奠定基础。目前PGG

.Han的汉族样本

在地理分布上共覆盖了33个省级

行政地区及部分海外汉族群体。PGG

.Han不仅提供了汉族人群的精细人群遗传结构图谱和可视化等位基因频率地图，更重要的是还整合了三个实用性强的在线分析工具；包括祖源推断（AncestryInference）、基因型填补（Genotype Imputation）和全基因组关联分析（GWAS）。这些分析流程由徐书华研究组设计与开发，由生物学大数据中心进行软件封装，并首次提供免费公开访问和使用。用户可以上传自己的样本数据，充分利用数据库中存储的汉族自然人群参考数据集进行在线分析，最终得到可供下载的分析报告。所有上传数据均受到严格的保护，同时考虑了非计算生物学背景人员的可操作性，并在持续改善用户体验。目前正力图通过广泛合作持续纳入更多高质量深度测序数据，以及提升该数据库所依赖的高性能计算分析能力。

该工作由徐书华研究组高扬（上海科技大学）、张超、王晓骥、刘畅、潘雨闻等人，以及生物学大数据中心张国庆、袁力赟、凌鋈超等人合作完成，得到中科院先导专项、国家自然科学基金委、上海市科委和国家重点研发计划等多项基金的资助。

[论文链接](#)

研究团队单位：上海营养与健康研究所

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](http://www.iikx.com)转发