

生物物理所女蜗基因组资源研究取得进展

作者：writer 来源：中国科学院

本文原地址：<https://www.iikx.com/news/progress/16636.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！

遗传变异图谱是研究人群演化史、医学遗传学、基因型-表型关联的基础。此前，大多数全基因组测序相关研究主要集中在欧洲血统人群。已有研究表明，罕见和低频的变异往往特定于人群或样本，尤其是与疾病相关的变异。针对特定人群的基因组数据可以为全基因组关联研究、区域适应性研究、用药指导等提供更准确的参考。

单倍型参考面板可基于大型人群队列中已知的单倍型信息，对来源于相对稀疏的基因变异芯片或低覆盖率测序样本中缺失的基因型进行推演，是促进全基因组关联研究 (genome-wide association study, GWAS) 有意义且具有成本效益的方法。此前缺乏中国人群特异的参考面板，其他参考面板对中国人群特异的变异推演效果较差，从而导致GWAS中遗漏潜在的表型相关变异。

中国人群全基因组测序资源和单倍型参考面板的缺乏阻碍了世界上最大人群的遗传学与精准医学研究。为此，中国科学院院士、中科院生物物理研究所研究员徐涛团队，研究员何顺民团队合作，在Cell Reports上在线发表了题为NyuWa Genome Resource: A Deep Whole Genome Sequencing-Based Variation Profile and Reference Panel for the Chinese Population

的文章，介绍该团队关于女蜗（NyuWa）中国人群基因组资源库（<http://bigdata.ibp.ac.cn/NyuWa/>）的工作，提供针对中国人群的遗传变异图谱与参考面板基因型推演服务，旨在促进中国人群的遗传学与医学研究。

研究团队分析了2,999个中国人的全基因组深度测序数据（26.2X），并以“女蜗”命名。基于NyuWa数据资源，研究构建了包含7106万SNPs和819万InDels的中国人群遗传变异图谱（图1），并对其进行全面注释。相比其它人群队列，NyuWa数据集包含2501万新变异，其中包括14.9万非同义变异、10.1万有害变异、11493个编码和非编码基因的功能丧失变异、636个癌症相关基因的蛋白截短变异。大量新变异表明，在以往遗传研究中，中国人群的变异代表性不足，NyuWa基因组资源则填补了这一空缺。此外，根据临床相关数据库的注释，研究在NyuWa中发现了1,140个致病变异，以及药物基因组学相关位点（图2A）和癌症风险位点（图2B）上中国人群与世界其他人群的变异频率差异。这些发现有助于中国人群精准医学研究，可能促进新的遗传学和医学进展。

为汉族人群构建一个完整的、大队列的、高质量的参考面板，对汉族的遗传学与医学研究具有参考价值。基于NyuWa数据资源，科研人员构建了包含5804个单倍型和1926万变异的单倍型参考面板，其中325万变异未包含在其它参考面板中，这些NyuWa参考面板特有变异可能会在未来关联研究中带来新发现，是首个数千人级别公开可用的中国人群特异的单倍型参考面板。为评估NyuWa参考面板的基因型推演性能，科研人员使用来自人类基因组多样性计划（the Human Genome

Diversity Project, HGDP) 的亚洲各个人群芯片基因分型数据和高覆盖率WGS数据作为测试数据集。与其他参考面板相比, NyuWa参考面板将汉族人群基因型推演的错误率降低了30%-51%, 在大多数其他东亚和东北亚人群中也有优异表现(图3A-D)。研究进一步比较了不同等位基因频率的推算结果和实际基因型之间的相关性, NyuWa参考面板性能在汉族的所有等位基因频率区间中均具有绝对优势(图3E)。此外, NyuWa参考面板与千人基因组(1KGP3)面板的结合进一步提高了亚洲人群的基因型推演效果。考虑到南北方汉族遗传差异, 科研人员将NyuWa参考面板中的样本分为北方和南方子集, 使用子集样本分别构建北方和南方汉族的参考面板, 通过基因型推演的模拟测试, 证明以NyuWa的人群规模, 一个南北整合的参考面板对中国北方人和南方人均适用(图4)。

综上, 基于中国人群的大型队列深度WGS数据, 研究构建了中国人群的遗传变异图谱和首个数千级别公开可用的中国人群单倍型参考面板, 将所有结果整合为中国人群基因组资源库NyuWa(图5), 有助于中国和亚洲人群的遗传学和精准医学研究。当前有关医学基因组学的知识和指南主要来自以欧洲人群为主的遗传和基因组资源, 可能遗漏有关非欧洲人群的遗传信息。亚洲人群起源、迁徙和融合历史久而复杂, 使得其遗传多样性研究面临挑战和机遇。针对中国人群的全基因组测序工作, 对于扩充世界人群遗传资源多样性、提高中国人群医学研究准确性十分必要, 有助于深入了解亚洲人群结构与人群历史, 并对寻找复杂疾病遗传因素的研究设计以及人口健康指导具有参考价值。

研究工作得到中科院战略性先导科技专项、国家自然科学基金、国家重点研发计划、中科院“十三五”信息化专项、国家基因组科学数据中心的支持。

[论文链接](#)

图1.NyuWa全基因组测序资源的变异数量

图2.NyuWa中的药物基因组学位点（A）与癌症风险基因座（B）变异

图4.南北方汉族测试数据集的基因型推演错误率

图5.NyuWa资源库

研究团队单位：生物物理研究所

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](https://www.iikx.com)转发