

首次揭示汉族人群可显著分成7个亚群

瑞金医院发布全自主中国人最大规模基因研究成果

2020年05月12日

作者：陶婷婷 李东

4月30日，瑞金医院内分泌科领衔的ChinaMAP（中国代谢解析计划）联盟携全国29家研究机构和医院，在中科院上海生命科学研究院主办的《Cell Research》杂志首次报道了对覆盖全国27个省份和直辖市，8个民族，超过1万人的高深度全基因组测序数据和表型的系统性分析。上海交通大学医学院附属瑞金医院、国家代谢性疾病临床医学研究中心王卫庆教授和毕宇芳是论文的共同通讯作者，曹亚南、李林和徐敏研究员等为共同第一作者。

此次发表意味着，通过中国人自己的仪器、平台、分析方法，对中国人的基因组特征进行深入、广泛的研究，具有前所未有的重要意义。

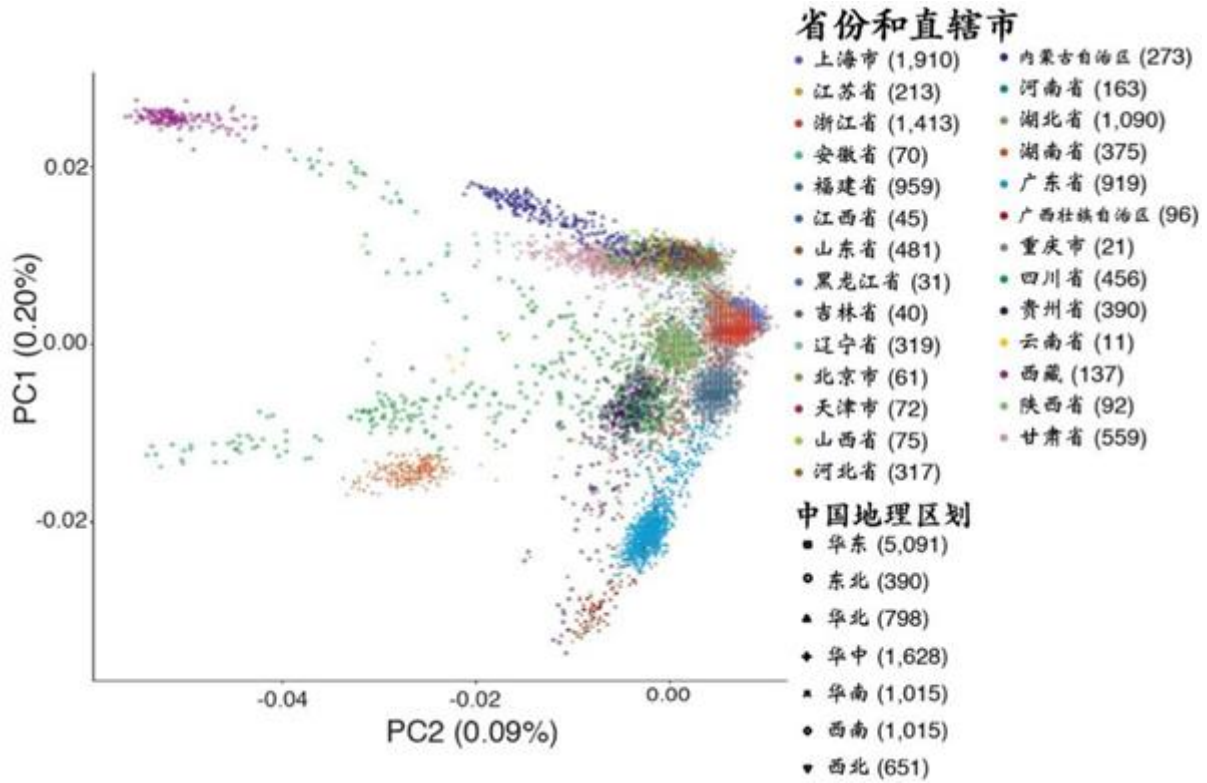
打造中国人自有的基因组数据库

长期以来，中国人的很多遗传疾病研究，都直接应用外国人的数据和结论。但由于不同地域人群和种族之间的历史渊源和遗传背景存在着巨大差异，如果把具有其他人群偏向性的知识和结论直接拿来做为中国人的疾病风险评估、遗传咨询或诊断治疗依据，是并不完善和可靠的。

“用中国人自己的数据组建中国人自己的精准医疗体系。”在这样的初心之下，国家代谢性疾病临床医学研究中心（上海）基于上海交通大学医学院附属瑞金医院牵头开展的多项覆盖全国的队列研究，依托转化医学国家重大科技基础设施(上海)和医学基因组学国家重点实验室，实施了中国代谢解析计划ChinaMAP (China Metabolic Analytics Project)。

不同地域人群的变异特征与人口迁徙相关

王卫庆教授介绍，研究团队对队列中代表中国不同地区和民族的10588人DNA样本进行了40×深度全基因组测序，完成了高质量的中国人遗传变异数据构建、中国人群体结构分析、基因组特征比较以及变异频谱和致病性变异解析。在ChinaMAP一期数据库中，包含1.36亿个基因多态性位点SNP和1千万个插入或缺失位点（NDEL），其中一半是在国际通用的多个数据库中均没有的新位点。



研究覆盖中国7大地理区域，包括了人口排名前10的汉族、壮族、回族、满族、苗族、彝族、藏族和蒙古族，显示了中华民族跨地理区域人群遗传背景的多样性和复杂性。研究团队首次揭示了汉族人群可显著分成7个亚群：北方汉族（北京、天津、河南、河北、山东、辽宁、吉林、黑龙江、山西），西北汉族（甘肃、陕西），东部汉族（江苏、浙江、上海、安徽），中部汉族（湖北），南方汉族（贵州、四川、重庆、湖南、云南、江西），东南汉族（福建）和岭南汉族（广东、广西）。少数民族中，藏族、彝族、蒙古族、苗族和壮族都有着独特的人群聚类，而满族和北方汉族相近，回族和西北、北方汉族相近。不同地域人群的变异特征也与中国历史上的人口迁徙和变迁相关，例如河西走廊是丝绸之路中不同民族迁徙的交通要道，历史上包括粟特人等许多民族曾在此经商生活。

中国人遗传特征与非洲差距最大，与东亚相似

研究揭示，现代河西走廊地区的人们具有的基因多态性位点数量更多更复杂。与全世界其他人群相比，中国人的遗传特征与欧洲、非洲、南亚和拉丁美洲人群之间存在着巨大差异，与非洲人群差距最大，而与东亚人群非常相似。研究者在成分分析中发现，日本人群与中国北方汉族人群聚类完全重叠。

国人先天性甲状腺功能减低症比欧美人群高10倍

众所周知，不少遗传性疾病是写在基因密码当中的。曹亚南介绍，中国人群中先天性甲状腺功能减低症、慢性胰腺炎、遗传性掌跖角化症等疾病的致病基因变异携带者较欧美人群显著更多，且具有地域分布特征。

疾病相关变异位点在中国人群与欧美人群中的频率差异，说明我国的遗传咨询和解读，对重要性不确定的基因变异（VUS）的研究，以及相关临床指南和路径制定，需要依据中国人自己的大样本和高质量数据。

喝酒脸红与特定基因有关，降脂药物谨慎用

此外，ChinaMAP对中国人的营养代谢和药物代谢相关遗传特征也进行了分析比较。

对受到广泛关注的酒精代谢能力，从总体上看，北方人比南方人酒量相对更好，藏族、蒙古族、彝族人民和河南人酒量居于全国前列，山东人酒量并不突出，福建人和广东人排在最后。

ChinaMAP证实，导致喝酒脸红和酒精代谢能力差的乙醛脱氢酶2 基因rs671变异是东亚人特异性的，在中国人群中的携带者（纯合子比例4.50%，杂合子比例34.27%）远高于全球其他人群。rs671变异也是诱发食管癌发生的重要风险因素，所以喝酒脸红应少喝酒。

为啥这些药物我们中国人用副作用大

另外，研究团队也对抗凝药华法林的减量使用、抗血小板药物氯吡格雷的适用人群分类、他汀类降脂药副作用风险人群进行了分析。例如，对于高脂血症常用的降脂药辛伐他汀，中国人中有超过20%的个体存在横纹肌溶解这一不良反应的风险，提示了针对我国人群特征的药物基因组研究和药物基因检测的重要性。

中国人群代谢性疾病风险基因与欧美也有不同

代谢性疾病，特别是II型糖尿病和肥胖，已成为中国和世界范围内发病率最高的重大慢病。但是，欧洲人群中最显著的II型糖尿病遗传风险TCF7L2基因变异（如rs7903146）在中国人中的频率却很低，可见在代谢特征和疾病研究中仅参考和验证欧美人群的结果是不行的。

另外，一个人携带的某个基因变异产生的疾病风险可能并不大，但综合多个基因变异的作用后对个体特征有重大影响。因此，基于特定人群的大规模基因型和表型数据库，通过多基因风险评分(PRS)评估个体的疾病风险是一种比较准确的方法。

在ChinaMAP研究中，研究者对II型糖尿病遗传风险进行了多基因风险评分，以量化评分、年龄和血糖值的排序三维显示了每个人在整个群体中的精确位置。多基因风险评分排名显示了II型糖尿病高风险和低风险的个体之间存在非常显著的血糖差异，高风险个体随着年龄的增加，空腹和餐后2小时血糖都显著高于中风险和低风险者。

此外对比证实，基于东亚人群的基础数据比基于欧洲人群的数据结果更加准确。这些结果提示基于中国人群基础数据对II型糖尿病及其代谢性疾病进行精确风险评估的重要性，对重大慢病的预防、个体化健康管理和公共卫生决策具有价值。

FTO等基因点位对中国人肥胖影响不明显


研究也通过全基因组关联分析探索了中国人群中2型糖尿病和肥胖遗传相关因素。在血糖相关分析中，结果验证了部分已知的2型糖尿病风险高频基因位点，也鉴定和发现了数个新位点，包括控制胰岛β细胞中胰岛素合成和分泌的关键转录因子，以及在以往研究中被证实可调控食欲和能量代谢稳态的位点。

在体重指数BMI相关分析中，研究团队发现了东亚人群新的特异性的CADM2基因位点，CADM2在动物研究中已证实参与调节体重和能量稳态。而FTO等在欧美人群中发现的重要肥胖相关基因位点，在ChinaMAP研究结果中并不显著。

ChinaMAP以覆盖中国各地区的研究队列为基础建立了高质量中国人群数据库，通过高深度全基因组数据和精细表型分析，可为疾病机制研究、预防、遗传咨询和公共卫生管理提供依据。据悉，ChinaMAP在2月下旬报道的新冠病毒受体ACE2相关变异在全球不同人群中的比较分析证实该病毒对所有人种均易感，该成果在Cell Discovery发表，已被访问下载超过20万次。ChinaMAP对汉族和少数民族群体的精确遗传结构分析也为中国人群的精准基因组学研究提供了参考数据。

编辑：liuchun 审核：liuchun

证件信息：沪ICP备10219502号 (<https://beian.miit.gov.cn>)

 沪公网安备 31010102006630号 (<http://www.beian.gov.cn/portal/registerSystemInfo?recordcode=31010102006630>)

中国互联网举报中心 (<https://www.12377.cn/>)

Copyright © 2009-2022

上海科技报社版权所有

上海科荧多媒体发展有限公司技术支持



([//bszs.conac.cn/sitename?method=show&id=5480BDAB3ADF3E3BE053012819ACCD59](http://bszs.conac.cn/sitename?method=show&id=5480BDAB3ADF3E3BE053012819ACCD59))