遗传 Hereditas (Beijing) 2021 年 10 月, 43(10): 921－923
www.chinagene.cn

中国人类遗传资源与生物大数据

方向东[1,2,3]，朱波峰[4]

1. 中国科学院北京基因组研究所(国家生物信息中心)，中国科学院基因组科学与信息重点实验室，北京 100101
2. 中国科学院大学，北京 100049
3. 中国科学院北京基因组研究所(国家生物信息中心)，基因组与精准医学检测技术北京市重点实验室，北京 100101
4. 南方医科大学，法医学院，广州 510515

人类遗传资源是指含有人体基因组、基因及其产物(RNA 和蛋白质)的器官、组织、细胞、血液、制备物、DNA 构建体等人类遗传资源材料及利用人类遗传资源材料产生的数据等人类遗传资源信息。中国人类遗传资源是国家自然科技资源的重要组成部分，是维护公众健康、国家安全和社会公共利益的重要战略资源。2019 年 3 月 20 日，《中华人民共和国人类遗传资源管理条例》在国务院第 41 次常务会议上通过，自 2019 年 7 月 1 日起开始施行。中国人类遗传资源在人类进化、种族溯源、法医学和生物医学研究中发挥着不可或缺的重大作用。

队列研究可通过对某一特定人群分组进行随访或者纵向观察监测，比较各组危险因素和暴露程度与人群结果的联系，发掘疾病产生原因以及揭示疾病发展进程，为各类慢性疾病的预防和诊断提供有效的科学依据。大型人群队列已然成为流行病学研究和重大复杂性疾病研究的主要方法之一。目前国内外均有很多典型的大型队列研究，例如欧洲 10 国的 European Prospective Investigation into Cancer and Nutrition (EPIC，52.1 万个体)、美国的 NIH-AARP Diet and Health Study (NIH-AARP，56.6 万个体)、中国慢性病前瞻性研究项目(51.3 万个体)、中国泰州人群健康跟踪调查(20 万个体)以及近年来采集组学信息的精准医学研究队列(如英国十万基因组计划、美国精准医疗、中国十万人基因组计划)等。来自北京大学公共卫生学院流行病与卫生统计学系的王文秀等[1]在《基于“中国慢性病前瞻性研究”的遗传资源建设与应用》一文中对中国医学科学院、北京大学和英国牛津大学联合开展的“中国慢性病前瞻性研究”项目(China Kadoorie Biobank, CKB)进行了针对性介绍，重点展示了该项目资源的采集与管理以及近年来获得的遗传学研究成果。来自复旦大学人类表型组研究院的陈兴栋等[2]在《大型人群队列遗传资源建设与利用》一文中以“泰州队列”为例展示了大型人群队列遗传资源在建设过程中各环节的原则、方法、标准体系和具体实践经验等，为今后我国大型人群队列遗传资源建设工作提供科学参考。《人类遗传资源管理条例》中指出人类遗传资源不仅包括人类遗传资源材料，还包括人类遗传资源信息，明确了人类遗传资源数据的重要性，这些数据不仅关乎个人的健康，还与国家安全相关，需要统一的平台保存管理。来自中国科学院北京基因组研究所(国家生物信息中心)国家基因组科学数据中心的张思思等[3]在《GSA-Human：人类遗传资源数据管理的公共系统》一文中详细介绍了人类遗传资源数据库系统 GSA- Human 的数据库性能、运作模式、国内外影响力，并深入思考了其面临的挑战以及未来的发展方向，共同为丰富国家人类遗传资源库数据尽一份力，推进遗传资源成为国家重要战略资源。

法医学是人类遗传资源利用的重要领域之一，典型的人类遗传标记如单核苷酸多态性(single nucleotide polymorphism, SNP)、短串联重复序列

收稿日期: 2021−10−14

作者简介: 方向东，博士，研究员，研究方向：医学遗传学、精准医学大数据。E-mail: fangxd@big.ac.cn
朱波峰，博士，教授，研究方向：法医遗传学。E-mail: zhubofeng7372@126.com

DOI: 10.16288/j.yczz.21-355

(short tandem repeat, STR)和微单倍型(microhaplotype, MH)等被应用于族群、系谱、体貌等特征刻画，亲子鉴定以及群体遗传结构等相关法医学研究。来自四川大学华西基础医学与法医学院的王浩宇等[4]在《基于全基因组数据 AI-SNPs 筛选及大陆次级区域内群体遗传结构差异研究》一文中使用千人基因组计划中东亚的五个群体，建立了基于 F_{ST} 的祖源信息单核苷酸多态性遗传标记(ancestry-informative single nucleotide polymorphism, AI-SNP)筛选法，对于减小大陆次级区域内群体遗传结构差异对群体相关医学研究的影响具有实际应用价值。来自中山大学中山医学院法医学系的刘志勇和北京警察学院的任贺等[5]在《基于有限突变模型和大规模数据的 19 个常染色体 STR 的实际突变率研究》一文中使用 Slooten 与 Ricciardi 提出的有限突变模型和大规模数据，对 28,313 例中国北京汉族已确认亲生关系的亲子鉴定案的 19 个常染色体 STR 基因座突变现象进行了深入研究。其团队重视隐形突变和一些少见的突变，得到了更为接近真实情况的新突变率结果。为科学解释 STR 亲子鉴定结果，优化法医学亲子鉴定和个体识别方法提供了重要依据和理论支撑。来自四川大学华西基础医学与法医学院李茜等[6]在《微单倍型遗传标记的法医基因组学研究》一文中基于千人基因组计划中 105 个中国南方汉族个体的全基因组测序数据，构建了迄今为止最全面的 MH 数据集，并且提出了构建该 MH 数据库的设计方案。在法医实践案件侦查中，办案人员面对的对象主要为尸体/人体、毛发、骨、血液、精液及其斑痕等人体生物检材，在研究过程中伴随着一系列的社会伦理道德和生物安全问题。来自中山大学中山医学院法医学系的刘志勇等[7]在《法医遗传学研究和鉴定中的伦理问题》一文中提及了国际上相关的伦理性规范文件，并对样本/检材收集、法医 DNA 表型分析、法医遗传系谱学分析、亲子鉴定、数据交流共享等具体方面所涉及的伦理问题进行了深入探讨，并针对我国法医遗传学工作中出现的伦理问题提出三大建议。同样来自西安交通大学口腔医院的郭瑜鑫等[8]在《生物安全视野下的法医学研究》一文中整理了法医学研究中涉及到的生物安全内容，同时点明了法医学研究今后面临的挑战与机遇、提倡各从业人员重视生物安全问题，遵循生物安全原则，净化法医学研究环境。

近年来，随着测序技术的发展和组学新技术的不断涌现，基因组、转录组、表观组、蛋白质组、代谢组等不同种类的组学数据指数级增长，而对多组学数据的整合分析，也成为科学家探索生命机制和疾病、肿瘤等发生发展规律的新方向。来自中国科学院北京基因组研究所(国家生物信息中心)的王昕玥等[9]在《组学大数据和医学人工智能》一文中举例介绍了近年来多组学和人工智能在医学领域各自的应用进展，以及两者结合应用相对于单组学或非人工智能的优势，最后简单阐述多组学分析和人工智能在现阶段面临的挑战，为未来精准医学这一必然趋势提供可行的研究思路。多组学研究技术的更新迭代以及信息时代的迅猛发展，伴随着数据量的爆炸式增长，也带领生物医学开始进入大数据时代，来自中国科学院上海营养与健康研究所生物医学大数据中心的郑广勇等[10]在《前沿信息技术在生物医学大数据中的应用及展望》一文中介绍了云计算、区块链、人工智能等前沿信息技术在生物医学大数据中的应用与展望。

未来，应用人工智能等前沿信息技术，探索生物医学多组学大数据，揭示人类遗传资源所蕴含的重要信息与规律将成为国家人口健康与安全领域的重要研究方向。我国人类遗传资源进入管理化、规范化道路，必将提高我国面对生物安全问题的防范能力，完善道德伦理规范，为我国人类遗传资源的保护、开发、管理和利用建立良好的研究环境。

参考文献(References):

[1] Wang WX, Huang T, Li LM. Construction and application of human genetic resources in the China Kadoorie Biobank. *Hereditas (Beijing)*, 2021, 43(10): 972–979.
王文秀, 黄涛, 李立明. 基于“中国慢性病前瞻性研究”的遗传资源建设与应用. 遗传, 2021, 43(10): 972–979. [DOI]

[2] Chen XD, Jiang YF, Xu P, Jin L. Construction and utilization of genetic resources in large population cohorts. *Hereditas (Beijing)*, 2021, 43(10): 980–987.
陈兴栋, 蒋艳峰, 徐萍, 金力. 大型人群队列遗传资源建设与利用. 遗传, 2021, 43(10): 980–987. [DOI]

[3] Zhang SS, Chen X, Chen TT, Zhu JW, Tang BX, Wang AK, Dong LL, Zhang ZW, Sun YL, Yu CX, Zhai S, Sun YB, Chen HX, Du ZL, Xiao JF, Zhang Z, Bao YM, Wang YQ, Zhao WM. GSA-Human: Genome Sequence Archive for Human. *Hereditas (Beijing)*, 2021, 43(10): 988–993.
张思思，陈旭，陈婷婷，朱军伟，唐碧霞，王安可，董丽莉，张哲文，孙艳玲，俞彩霞，翟爽，孙玉彬，陈焕新，杜政霖，肖景发，章张，鲍一明，王彦青，赵文明. GSA-Human：人类遗传资源数据管理的公共系统. 遗传, 2021, 43(10): 988–993. [DOI]

[4] Wang HY, Hu YH, Cao YY, Zhu Q, Huang YG, Li X, Zhang J. AI-SNPs screening based on the whole genome data and research on genetic structure differences of subcontinent populations. *Hereditas (Beijing)*, 2021, 43(10): 938–948.
王浩宇，胡渝涵，曹悦岩，朱强，黄雨果，李茜，张霁. 基于全基因组数据的 AI-SNPs 筛选及大陆次级区域内遗传结构差异研究. 遗传, 2021, 43(10): 938–948. [DOI]

[5] Liu ZY, Ren H, Chen C, Zhang JJ, Zhang XM, Shi Y, Shi LY, Chen Y, Cheng F, Jia L, Chen M, Fan QW, Zhang JR, Li WT, Wang MC, Ren ZL, Liu YC, Ni M, Sun HY, Yan JW. Actual mutational research of 19 autosomal STRs based on restricted mutation model and big data. *Hereditas (Beijing)*, 2021, 43(10): 949–961.
刘志勇，任贺，陈冲，张京晶，张晓梦，石妍，石林玉，陈滢，程凤，贾莉，陈曼，范庆炜，张家榕，李万婷，王萌春，任子林，刘雅诚，倪铭，孙宏钰，严江伟. 基于有限突变模型和大规模数据的 19 个常染色体 STR 的实际突变率研究. 遗传, 2021, 43(10): 949–961. [DOI]

[6] Li X, Wang HY, Cao YY, Zhu Q, Shu PY, Hou TY, Wang YT, Zhang J. Forensic genomics research on microhaplotypes. *Hereditas (Beijing)*, 2021, 43(10): 962–971.
李茜，王浩宇，曹悦岩，朱强，舒潘寅，侯婷芸，王雨婷，张霁. 微单倍型遗传标记的法医基因组学研究. 遗传, 2021, 43(10): 962–971. [DOI]

[7] Liu ZY, Wu RG, Li R, Wang QW, Sun HY. Ethical issues of the research and practice in forensic genetics. *Hereditas (Beijing)*, 2021, 43(10): 994–1002.
刘志勇，乌日嘎，李燃，王蔷薇，孙宏钰. 法医遗传学研究和鉴定中的伦理问题. 遗传, 2021, 43(10): 994–1002. [DOI]

[8] Guo YX, Zhao XC. Research in forensic medicine under the view of biosafety. *Hereditas (Beijing)*, 2021, 43(10): 1003–1007.
郭瑜鑫，赵兴春. 生物安全视野下的法医学研究 遗传, 2021, 43(10): 1003–1007. [DOI]

[9] Wang XY, Qu HZ, Fang XD. Omics big data and medical artificial intelligence. *Hereditas (Beijing)*, 2021, 43(10): 930–937.
王昕玥，渠鸿竹，方向东. 组学大数据和医学人工智能 遗传, 2021, 43(10): 930–937. [DOI]

[10] Zheng GY, Zeng T, Li YX. Application and prospect of frontier information technology in biomedical big data. *Hereditas (Beijing)*, 2021, 43(10): 924–929.
郑广勇，曾涛，李亦学. 前沿信息技术在生物医学大数据中的应用及展望 遗传, 2021, 43(10): 924–929. [DOI]