



罕见病临床样本库的建立与应用

郭丹^{1,2,3}, 金晔^{2,3}, 刘炜达^{2,3}, 张雅然^{1,2,3}, 王安琪^{1,2,3}, 张抒扬^{3*}

1. 中国医学科学院&北京协和医学院, 北京协和医院临床生物样本中心, 北京 100730;

2. 中国医学科学院&北京协和医学院, 北京协和医院医学科学研究中心, 北京100730;

3. 中国医学科学院&北京协和医学院, 北京协和医院疑难重症及罕见病重点实验室, 北京 100730

* 联系人, E-mail: shuyangzhang103@nrdrs.org

收稿日期: 2023-03-06; 接受日期: 2023-06-25; 网络版发表日期: 2024-03-22

中央高水平医院临床科研专项(批准号: 2023-PUMCH-F-004)、中国医学科学院医学与健康科技创新工程项目(批准号: 2021-I2M-1-053)、中央高水平医院临床研究专项(批准号: 2022-PUMCH-D-002)和国家重点研发计划(批准号: 2022YFC2703100)资助

摘要 罕见病是发病率或患病率极低的一类疾病的总称, 因其种类繁多、诊断困难、治疗手段有限, 一直是困扰全球医学界的难题。随着转化医学的飞速发展, 近年来各国逐渐在罕见病的研究、诊疗和药物研发方面加大投入, 生物样本库在常见病领域已有诸多成功的应用, 在罕见病领域也发挥着重要作用。罕见病生物样本库将生物样本信息与患者临床信息、组学数据等进行整合, 促进疾病精准表型和发病机制研究, 支持开展高水平、大规模的队列研究, 加速罕见病新药、新治疗技术研发及疗效评价, 支撑从基础研究到临床转化的全链条研究模式。本文通过对临床样本库在罕见病领域的建立与应用进行回顾与展望, 旨在促进罕见病样本库标准化、规范化、信息化、多元化和可共享的模式建立, 推动全球范围内的协作和转化应用, 这对于罕见病诊疗能力的提升和医学科学的发展具有重要意义。

关键词 罕见病, 生物样本库, 标准化, 信息化, 转化与共享

罕见病(rare disease, RD)是指发病率或患病率极低的一类疾病的总称^[1,2]。目前已发现的罕见病有6000~8000种, 其中大约80%是遗传性的, 50%~75%是儿童发病^[3~5], 通常伴随终生且使人衰弱, 并具有较高的致畸率和致死率, 同时由于个体及病种间异质性较大且复杂罕见, 因此很难被诊断。目前, 由于罕见病在全球各地区的发病率和患病率具有显著的差异性, 所以针对罕见病的定义尚未有全球性的共识和标准。世界卫生组织(World Health Organization, WHO)将其定义为发病率在0.65%~1%之间的疾病^[6], 欧盟将其定义

为流行率为5/10000, 且威胁生命或使人持续衰弱的疾病; 在美国, 罕见病被定义为流行率低于7.5/10000且影响少于20万名患者的疾病^[7]。中国通过专家共识将患病率<1/500000, 或新生儿发病率<1/10000的疾病定义为罕见病。2018年, 中国国家卫生健康委等5部委发布《第一批罕见病目录》, 首次通过该目录界定了中国的罕见病^[8,9], 在全球范围内发出了中国的声音。

虽然罕见病单一病种患病人数较少, 但由于种类繁多, 其影响范围广泛, 研究显示, 全世界约有2.63亿~4.46亿罕见病患者^[10], 相当于世界第三大国的人口数

引用格式: 郭丹, 金晔, 刘炜达, 等. 罕见病临床样本库的建立与应用. 中国科学: 生命科学, 2024, 54: 1041–1049

Guo D, Jin Y, Liu W D, et al. Development and application of rare diseases biobank (in Chinese). Sci Sin Vitae, 2024, 54: 1041–1049, doi: 10.1360/SSV-2023-0038

量. 如进一步涵盖患者及其家庭成员, 预计在全球范围内约有10.5亿~14.0亿人与罕见病息息相关^[5]. 此外, 漫长的诊断过程、极少的有效治疗手段、昂贵的医疗费用始终困扰着罕见病患者^[11,12]. 在欧洲, 25%的罕见病患者确诊需要经历5~30年, 40%的患者曾经历多次误诊, 导致无效和不必要的医疗救治^[13]. 在另一项对英国和美国的罕见病家系调查中, 患者经历多次转诊或误诊, 平均确诊时间长达5.6~7.6年^[14]. 同样在我国根据《2019中国罕见病患者综合社会调查》显示, 罕见病的误诊比例高达51.2%, 平均确诊时间约为4年. 然而, 尽管经过漫长而曲折的确诊后, 仅有不到3%的罕见病患者能够接受合适的药物治疗^[15,16]; 在全球范围内统计, 也仅有不到十分之一的患者能够接受治疗^[17]. 这样的情况在中国也尤为显著, 罕见病药物缺乏、可及性差使其治疗方法非常有限(图1), 而由疾病引起的各种残疾以及诊断和治疗的不确定性, 均对患者家庭生活的健康、社会心理和经济方面产生不利影响^[18].

尽管国际日益重视并有政策倾斜, 国际罕见病组织2019年发表声明, 呼吁应给予罕见病患者全民健康保险(universal health coverage, UHC)政策^[19], 同时联合国提出罕见病患者在医疗规划中应予以特殊考虑^[20], 但罕见病的诊断与治疗仍充满挑战. 接下来需要根据疾病类型进行细化, 规范化开展临床队列研究, 并通过建立罕见病临床数据和生物样本库等, 研究疾病精准表型和自然病程的演变, 并系统地将临床表型、基因组数据、各类生物样本和临床研究的数据整合并开展精准化诊疗^[21], 不断提高对罕见病发病机制和治疗方式的认知, 为个体化诊疗和新药研发提供基础.

1 临床生物样本库的国内外发展现状

根据美国国立癌症研究所(National Cancer Institute, NCI)的定义, 生物样本库是用于研究的生物样本、相关数据、储存实体、相关流程和政策的集合^[22]. 近年来, 随着生物科技和转化医学的飞速发展, 大规模、高质量生物样本及相关信息的需求快速增长, 生物样本库为基础医学、临床医学和药物研发提供宝贵的生物资源, 在全球医学研究中的重要性日益突出.

在国际上, 最具有代表性的英国生物银行UK Biobank通过采集捐赠者丰富的表型和健康相关信息, 包

括生物测量、生活方式指标、血液和尿液中的生物标记以及身体和大脑成像的数据, 描述并集中分析遗传数据及基因型质量、群体结构属性和遗传数据的相关性, 探索人类遗传变异与人类生物学和疾病之间的联系, 以及它们与广泛的环境和生活方式因素之间的联系^[23]. 泛欧洲生物样本库与生物分子资源研究中心-欧洲研究中心联盟(Biobanking and BioMolecular Resources Research Infrastructure-European Research Infrastructure Consortium, BBMRI-ERIC)是建立在欧洲生物银行基础上的一种新模式, 覆盖了全欧洲515个生物样本库, 超过6000万份生物样本. 其中包括136个临床和特定疾病的样本库和189个人群的样本库^[24]. NCI早在1989年成立人体组织样本库(Cooperative Human Tissue Network, CHTN), 由6所大学及医疗机构参与, 鼓励存储和利用人体组织用于癌症在分子生物学、免疫学和遗传学领域的研究, 这些机构每年会为数百名研究者提供8万份研究样本^[25].

随后十几年的发展中, 生物医学研究在信息化的需求呈指数级增长, 信息化的样本库成为连接临床数据和分子机制的桥梁. 2003年NCI在CHTN的基础上建立了美国癌症生物医学信息学网格(cancer Biomedical Informatics Grid, caBIG, <http://cabig.nci.nih.gov>), 通过开发信息化网络设施, 使全球的科学家和科研机构可以共享样本和分析数据, 从而有效地形成一个全球合作的癌症研究社区, 支持肿瘤学的转化和临床研究. 截止到2007年, 已经有900余位临床医生和科学家参与其中^[26]. 日本建立国家中心生物样本库网络(National Center Biobank Network, NCBN), 由六个国家医学研究中心的生物样本库通过网络化形式联合, 针对肿瘤、心血管、精神疾病、神经疾病和肌肉、慢性病和传染病、儿科和老年疾病等收集者临床信息和样本衍生数据并形成协作网, 截至2022年, 在NCBN上登记的生物资源总数超过400000个, 登记的患者总数为120081^[27].

我国在生物样本库建设方面有着极大的优势, 首先是人口基数大, 病例资源丰富, 有利于快速高效地进行大样本转化研究; 其次作为亚洲人群的显著代表之一, 临床及生物样本资源对于验证全球转化成果具有重要价值, 同时我国的医疗行业管理体制高效统一, 兼具多样性和地方特色, 为建设标准化、有价值的生物样本库奠定了坚实基础. 因此, 我国从“十三五”开始全



图 1 中国罕见病诊疗所面临的三大难题: 漏诊误诊率高, 诊疗能力有限, 缺少可及的药物(部分引自《2019中国罕见病患者综合社会调查》)

Figure 1 The three major challenges of Rare Disease in China: high rate of misdiagnosis, limited diagnostic and treatment capacity, and lack of accessible drugs (partly quoted from the 2019 Comprehensive Social Survey of Rare Disease Patients in China)

面建设、集中部署人类样本库和各类疾病队列样本库, 并相继涌现出了一批特色库、学科群库、医院库、区域库等, 例如中华民族永生细胞库、江苏泰州自然人群队列、国家干细胞资源库、北京重大疾病临床数据和样本资源库、“十二五”新药创制重大专项“临床标本资源库”、上海张江生物银行、深圳国家基因库等一批国内具有代表性的疾病和人群队列样本库^[28-32]。

我国的生物样本库在飞速发展, 但同时也存在一些瓶颈和弊端, 一方面各类样本库的容量和规模不断增大, 样本面临共享难、利用率低等问题, 有待规范建设、整体布局和统筹协调; 另一方面面临建设周期长、投入多但转化慢、产出少, 可持续发展能力较为不足等困境。因此, 需要通过加强资源共享网络建设, 整合资源、建立共享机制、联动产学研用、促进转化产出, 加速样本库领域的高质量发展, 抓紧时机赶超国际发展。此外, 目前生物样本库在常见病领域已有诸多成功的应用, 标准化、规范化、信息化的发展也日趋成熟, 共享和转化的机制亟待完善, 这些经验在罕见病样本库的建立中亦发挥着重要作用。

2 罕见病生物样本库在全球范围内的建立与应用

罕见病缺乏诊断标准、资源分散、信息不全, 导致诸多研究结果无法横向比较, 罕见病生物样本库汇

集临床样本和诊疗数据资源, 能够从根本上解决这些问题。目前世界已有一些国家建立了临床生物样本库支持罕见病相关研究, 或者建立罕见病生物样本库, 通过多种方式支持罕见病机制与调控、临床诊断、新药研发等基础研究和转化应用^[33,34]。

2009年, 日本在国家生物医学创新研究所(National Institute of Biomedical Innovation, NIBIO)建立了针对罕见病的样本库^[35], 采用规范化的方式从全国各地医疗机构收集罕见病患者生物样本以及临床信息, 并在全国范围内进行共享, 推动日本和全世界对罕见病的研究、治疗及预防。截至2017年8月, NIBIO已收集了4147份罕见病患者的DNA、血清、血浆和细胞样本, 涵盖马方综合征、多发性硬化、视神经脊髓炎、色素性视网膜炎等102种罕见病, 并已向各研究机构提供了来自2850例的13686份生物样本^[36]。2013年, 意大利建立基因样本库协作网(Telethon Network of Genetic Biobanks, TNGB)专注于罕见病研究, 该协作网具有750多种不同遗传缺陷罕见病的75900份生物样本, 侧重原发性心肌病、心血管、眼科、血液学、内分泌系统、神经系统、颅面障碍、运动障碍、罕见的肿瘤等21个系统类型^[37]。2019年, 西北大学人类代谢组学中心(Centre for Human Metabolomics, CHM)在南非和非洲启动了罕见疾病(罕见病)生物样本库计划^[38], 其通过罕见先天性疾病的样本和信息收集, 阐述罕见病自然史和病理生理学, 用以开发并优化诊断

方法及潜在的治疗方法。

此外, 部分早期建立的罕见病样本库陆续开展支撑罕见病的诊疗及转化研究^[39-41]。遗传性肾小管病是一组具有重大社会影响的罕见疾病, 因为通常需要终身治疗, 且没有治疗方法可以阻止肾损害的进展。通过建立肾源性肾小管样本库(Urine-Derived Kidney Tubuloid Biobank), 基于RNA、蛋白质、功能和形态等层面比较和评价致病性未知的遗传变异^[42], 为阐明该类疾病的病理生理机制研究提供巨大的可能性, 并为队列研究提供支持。此外, Tubuloid Biobank以不同的方式促进药物开发, 将潜在的治疗性化合物在特定的患者亚组人群上进行测试, 促进遗传性肾小管病的新药研发。肾病胱氨酸病是一种罕见的遗传性疾病, 在出生后的几个月内发病, 导致继发于溶酶体胱氨酸积累的终末期肾病和其他肾外并发症。研究人员基于意大利基因样本库协作网TNGB, 通过将药物基因组和RNA-seq表达谱进行比较和分析, 确定了与肾病型胱氨酸病病理生理学相关的潜在药物和代谢通路^[43], 体现了样本库对于药物研发转化的支撑性作用。

色素失禁症是因神经外胚层缺陷可导致残疾。目前, 既没有明确的治愈方法, 尚未有任何可靠的数据来预测疾病的严重程度。2015年启动了色素失禁症遗传生物样本库(Incontinentia Pigmenti Genetic Biobank, IPGB)项目, 以建立大规模的生物样本库, 提供诊断为色素失禁症患儿的详细临床信息, 促进其基因型-表型的相关性研究, 并根据社会人口学数据提供色素失禁症的流行病学数据^[44]。

1型肺动脉高压(pulmonary arterial hypertension, PAH)是一种罕见疾病, 尽管近年来治疗取得了进展, 但病死率高。既往研究表明, BMPR2(bone morphogenetic protein receptor type 2)和其他风险基因的突变使患者易患该病^[45,46], 但绝大多数非家族性病例在遗传学上仍未明确。美国国家生物样本和数据资源库(National Biological Sample and Data Repository for PAH, PAH Biobank)在识别PAH新的风险基因上起到了关键作用。通过对PAH Biobank中2572例PAH患者的生物样本进行基于病例对照基因的关联分析, 表明组织激肽释放酶1(tissue kallikrein 1, KLK1)和 γ -谷氨酰羧化酶(gamma glutamyl carboxylase, GGCX)为特发性肺动脉高压(idiopathic pulmonary arterial hypertension, IPAH)候选风险因子, 并进一步证实T-box 4(TBX4)与儿童发

病PAH的相关性^[47]。这些研究表明, 利用罕见病样本库的资源优势, 有利于开展大规模表型与基因型关联及分析研究, 有力支撑罕见病致病机制及治疗靶点等研究。

除了罕见病样本库, 常见疾病及人群队列样本库也在利用资源和优势支持罕见病的研究, 例如, UK Biobank针对167246名欧洲血统个体的外显子组测序数据进行深入分析, 以确定每种疾病潜在的罕见致病性变异, 发现并鉴定了23575人患有420种罕见疾病^[48], 体现了人口生物库有利于描绘疾病图谱并阐明罕见疾病的人口统计学和遗传关联。2015年启动的摩洛哥生物样本库(Eastern Morocco biobank, BRO biobank), 截至2020年纳入的2446位参与者中, 55%为患有罕见病(血红蛋白病、智力障碍、性别分化障碍、肌病等)^[49], 将摩洛哥和北非少数民族纳入国际生物医学及罕见病研究。

综上所述, 生物样本库在常见病领域已经有了诸多成熟的建设和应用的经验, 然而在罕见病领域还需要建设和发展。同时由于罕见病病种繁多, 漏诊误诊率高, 药物可及性差, 疾病负担极重, 亟待提高对疾病致病机制的认识。以罕见病生物样本库为基础可支撑开展临床前发病机制和调控研究, 提高对疾病致病机制的认识, 支撑开展高水平、大规模的队列研究, 促进疾病精准表型和自然病程的演变研究, 推动罕见病新药、新治疗技术研发及疗效评价研究。因此, 一方面加快更多罕见病特色样本库的建立, 注重标准化、规范化、信息化, 另一方面促进建设中的罕见病样本库建立多元化、促转化、可共享的模式, 不断投入到多区域、全球化的协同中, 促进全球范围内的科研协作和转化应用。

3 罕见病生物样本库的全球范围内的协作与共享

由于罕见病患者的稀缺, 单一地域或国家的罕见病样本库不足以支持罕见病的研究发展, 罕见病样本库的建立正呈现全球协作的趋势, 在国际上人们所熟知的欧洲生物银行(EuroBioBank)、意大利基因生物样本库协作网络(TNGB)、Orphanet等生物样本库聚焦在罕见病领域并建立多元化、多病种可共享的样本库, 在欧洲及全球范围内有力地促进罕见病研究^[50]。

这种高层次、多区域、全球范围内的协作,需要在罕见病样本库建设的过程中注重标准化、规范化、信息化、多元化、可共享的模式建立。

首先,罕见病研究队列的高通量测序技术和生物信息学分析使罕见致病基因的发现在近二十年间增加了四倍,同时人类基因组计划和HapMap(Haplotype Map)计划的启动,结合GWAS(genome-wide association studies)和测序研究,已支持识别数千种人类疾病和特征的6万多个遗传关联^[51]。英国UK Biobank通过建立标准化的临床检测分析加速解决未知罕见病的病因及机制研究,针对捐献者进行全基因组WGS(whole genome sequencing)检测分析,以确定定量表型及罕见变异关联^[33],UK Biobank正是在标准化、规范化、信息化的基础上通过调控变异的探索进行罕见病未来研究和临床干预,因此标准化、规范化、信息化的罕见病生物样本库已成为罕见病研究基础中的基础。

其次,虽然基因组测序数据生产已经成为常规,但分析和解释仍然是一项具有挑战性的工作^[52],并且针对罕见病的各项研究单依靠基因变异发现、预测变异效应和基因分型还不足以全面阐释疾病发病的机制,需要结合细胞、类器官和动物模型和人类样本,因此建立包含人、模式动物等细胞、核酸、类器官模型等不同类型的多元化罕见病生物样本库,对于疾病精准表型和自然病程的演变研究、新药研发及疗效评价和转化研究有着重要的作用。

再次,罕见病生物样本库国际化和联盟化的优势也不断凸显,近年来已经成立了几个联盟,以促进研究和最大限度地获取罕见疾病资源,其中包括欧洲生物样本库网络和西班牙国家罕见疾病登记处(Spanish Rare Diseases Registries Research Network, SpainRDR)和生物样本库(Biobanco Nacional de Enfermedades Raras, BioNER)等。欧洲生物银行成立于2001年,由11个国家的25个罕见病生物库组成,已有超过50万份样本,为全球科学家进行罕见疾病研究储存和提供高质量的DNA、细胞和组织样本。此外,疑难罕见病联盟(RD-Connect)是欧洲2012年启动的另一项全球基础设施项目,将21个罕见病生物库的数据汇集后向学术界和工业界的研究人员开放共享,将罕见病基因组数据与患者登记处、临床生物样本库和生物信息学工具整合,为罕见病的研究提供集中式、可共享的优势资源,以加速研发和转化^[50]。2003年,欧洲生物银行逐步与

欧洲生物分子和生物分子资源研究基础设施(BBMRI)、疑难罕见病联盟(RD-Connect)、运动神经元病等协作组织(Translational Research in Europe for the Assessment and Treatment of Neuromuscular Disorders, TREAT-NMD)等形成研究网络,共同建立了国际高水准的疑难罕见病的生物样本库和患者注册登记平台以及多组学数据平台。自欧洲生物银行网络成立以来,平均每年新增18800个样本并出库7000个样本进行多组学和基因学研究^[53],已有超过13万份罕见病样本用于共享及支撑各项研究并发表了400余篇论文。此外,西班牙RDR和BioNER是为促进罕见病研究和健康相关事项而创建的,将BioNER捐赠的生物样本与RDR收集的临床信息相关联,这种协调方式为罕见病研究提供了很好的范例^[54]。

我国罕见病样本库的建设也正在快速发展,并且我国人口众多,疑难性罕见病资源丰富,研究历史悠久,早年研究多集中在病例报道。近年来疑难性罕见病队列和注册登记研究逐渐形成。目前全球在美国临床试验数据库(<https://clinicaltrials.gov>)网站注册的疾病注册登记研究中,我国牵头和参与的疑难性罕见病的注册登记研究也在逐年显著增加,如淋巴管肌瘤病、戈谢病、庞贝病、结节性硬化症等,截至2023年2月,淋巴管肌瘤病的临床试验研究全球有550项,中国有32项;戈谢病的临床试验研究全球有143项,中国有6项。目前,罕见病已经初步形成了辐射全国的多学科协作网络,如北京协和医院依托中国国家罕见病注册系统(National Rare Diseases Registry System of China, NRDRS)^[55]和全国罕见病诊疗协作网络及罕见病联盟,历经六年建成国际最大的罕见病资源库之一,涵盖190多个罕见病注册队列近7万患者遗传诊断、生物样本信息等,包括约1.7万例生物样本2.4万例基因数据,并形成了全国多中心的协作网络,为并且在日趋向标准化、规范化、信息化、多元化、可共享的方向发展。

近年来,越来越多的罕见病生物样本资源库与病历注册登记体系等多体系的协同和联合,形成国际化和联盟化的基础支撑资源(表1)。加强沟通和协作,使罕见病研究人员能够更好地共享样本和相关数据,同时保护数据安全和患者及家庭成员的隐私,并遵守国际道德和法律要求,对于合规、合理、高效、快速地推动罕见病的研究具有重要作用。

表 1 全球罕见病样本库、数据库及病例注册系统^{a)}

Table 1 Global rare disease biobank and case registration system^{a)}

成立时间	名称	创办地区	共享范围	链接	参考文献
2000	欧洲孤儿药物及罕见疾病信息网 (Orphanet)	法国	国际性	https://www.orpha.net/consor/cgi-bin/index.php?lng=EN	[56]
2001	欧洲生物样本库(EuroBioBank)	欧洲	国际性	http://www.eurobiobank.org/	[53]
2005	国家生物医学创新研究所(NIBIO)	日本	国际性	https://www.nibiohn.go.jp/nibio/english/part/bioresources/detail.html	[35]
2007	意大利Telethon基因生物样本库协作网 (TNGB)	意大利	国际性	http://biobanknetwork.telethon.it/	[37]
2010	NIH罕见病登记处计划(NIH Rare Diseases Registry, RaDaR)	美国	国际性	https://registries.ncats.nih.gov/about-radar/	[57]
2012	罕见病注册登记系统和罕见病 (RD-Connect Registry & Biobank Finder)	欧盟	国际性	https://rd-connect.eu/what-we-do/biosamples-data/rb-finder-for-biobanks/	[50]
2013	肺动脉高压生物样本库(PAH Biobank)	美国	国际性	www.pahbiobank.org	[47]
2013	西班牙国家罕见病生物库(The Spanish National Rare Disease Biobank)	西班牙	国际性	https://bioner.isciii.es/home/	[54]
2015	色素失禁症遗传生物样本库(The Incontinentia Pigmenti Genetic Biobank, IPGB)	意大利	国际性	http://www.igb.cnr.it/ipgb	[44]
2015	TREAT-NMD DMD全球数据库(UMD TREAT-NMD DMD mutations database)	美洲	国际性	http://umd.be/TREAT_DMD/	[58]
2015	摩洛哥生物样本库(BRO biobank)	摩洛哥	国际性	NA	[49]
2016	中国国家罕见病注册系统(NRDRS)	中国	国际性	https://www.nrdrs.org.cn/	[55]
2019	罕见代谢性疾病生物样本库(Rare disease Biobank of Center for Human Metabolomics, CHMRD)	南非	国际性	https://natural-sciences.nwu.ac.za/centre-human-metabolomics/bio-bank	[38]

a) 1. 以上样本库官网均显示数据共享; 2. UMD: Universal Mutation Database; DMD: Duchenne muscular dystrophy

4 展望

在过去十年中, 各国政府、研究机构和生物医药企业对罕见病患者的高度重视, 大大推动了罕见病生物样本库的建设、快速发展和转化应用, 将其视为发展生物医药领域核心竞争力和提升国民健康的战略举措, 促使罕见病生物样本库的活动规模、样本数量、涵盖的疾病类型和支持的研究活动范围不断增长和扩展, 表明罕见病生物样本库全面快速发展的时代已然到来. 一方面药物基因组学的发展和人类全基因组测序技术的日新月异, 使个体化医疗取得了进展, 罕见病生物样本库作为基础性支撑的重要性也日益凸显, 其提供的各种基因信息、遗传学、家庭及个人疾病史、治疗史和生活方式、健康记录等可提高对罕见疾病总体负担的了解^[48]; 另一方面通过将遗传学与广泛和深入的表型分析结合, 随着样本量的不断增加, 结合罕见疾病的致病变异广度分析和深度挖掘, 促进疾

病精准表型和自然病程的演变研究, 与健康记录相关联, 来指导疾病的预测与预防、卫生经济学评估、医疗健康政策制定等, 实现罕见病生物样本库研究的巨大价值.

在面临机遇、快速发展的同时, 罕见病生物样本库也在隐私和伦理管理、国家和国际层面的合作共享, 生物样本收集、储存和加工的标准化, IT、质量管理体系及可持续性发展等领域面临着不同的挑战^[59], 需要在发展中不断探索和创新.

总之, 罕见病生物样本库对于疾病的预测、预防和个性化医疗具有重要的作用, 建立标准化、大规模、信息化、多元化的区域库、国家库及国际库联盟是发展趋势, 也是进行罕见病机制、临床与转化研究合作的基础. 此外, 面向医疗机构、高校、研究院所、研发企业等各级科研机构开放共享, 形成高层次、区域化协同共享机制和支撑模式, 将对于改善罕见病诊疗能力、提升全民健康具有重大意义.

参考文献

- 1 Groft S C, Posada M, Taruscio D. Progress, challenges and global approaches to rare diseases. *Acta Paediatr*, 2021, 110: 2711–2716
- 2 Richter T, Nestler-Parr S, Babela R, et al. Rare disease terminology and definitions-a systematic global review: report of the ISPOR rare disease special interest group. *Value Health*, 2015, 18: 906–914
- 3 Boycott K M, Vanstone M R, Bulman D E, et al. Rare-disease genetics in the era of next-generation sequencing: discovery to translation. *Nat Rev Genet*, 2013, 14: 681–691
- 4 Chung C C Y, Chu A T W, Chung B H Y. Rare disease emerging as a global public health priority. *Front Public Health*, 2022, 10: 1028545
- 5 Groft S C, Posada de la paz M. Rare diseases: joining mainstream research and treatment based on reliable epidemiological data. *Adv Exp Med Biol*, 2017, 1031: 3–21
- 6 Abozaid G M, Kerr K, McKnight A, et al. Criteria to define rare diseases and orphan drugs: a systematic review protocol. *BMJ Open*, 2022, 12: e062126
- 7 Ferreira C R. The burden of rare diseases. *Am J Med Genet*, 2019, 179: 885–892
- 8 Liu P, Gong M, Li J, et al. Innovation in informatics to improve clinical care and drug accessibility for rare diseases in China. *Front Pharmacol*, 2021, 12: 719415
- 9 Lu Y, Gao Q, Ren X, et al. Incidence and prevalence of 121 rare diseases in China: Current status and challenges: 2022 revision. *Intractable Rare Dis Res*, 2022, 11: 96–104
- 10 Nguengang Wakap S, Lambert D M, Olry A, et al. Estimating cumulative point prevalence of rare diseases: analysis of the Orphanet database. *Eur J Hum Genet*, 2020, 28: 165–173
- 11 Ahmed M A, Okour M, Brundage R, et al. Orphan drug development: the increasing role of clinical pharmacology. *J Pharmacokinet Pharmacodyn*, 2019, 46: 395–409
- 12 Gruhn S, Witte J, Greiner W, et al. Epidemiology and economic burden of meningococcal disease in Germany: A systematic review. *Vaccine*, 2022, 40: 1932–1947
- 13 EURORDIS. Survey of the Delay in Diagnosis for 8 Rare Diseases in Europe ('Eurordiscare 2'). 2007. Available from URL: <https://www.eurordis.org/IMG/pdf/>
- 14 The Lancet Diabetes Endocrinology . Spotlight on rare diseases. *Lancet Diabetes Endocrinol*, 2019, 7: 75
- 15 Czech M, Baran-Kooiker A, Atikeler K, et al. A review of rare disease policies and orphan drug reimbursement systems in 12 Eurasian countries. *Front Public Health*, 2020, 7: 416
- 16 Makarova E V, Krysanov I S, Valilyeva T P, et al. Evaluation of orphan diseases global burden. *Eur J Transl Myol*, 2021, 31: 9610
- 17 Chan A Y L, Chan V K Y, Olsson S, et al. Access and unmet needs of orphan drugs in 194 countries and 6 areas: a global policy review with content analysis. *Value Health*, 2020, 23: 1580–1591
- 18 Pearson I, Rothwell B, Olaye A, et al. Economic modeling considerations for rare diseases. *Value Health*, 2018, 21: 515–524
- 19 Group L. The National Economic Burden of Rare Disease Study. 2021. Available from URL: <https://everylifefoundation.org/wpcontent/uploads/2021/2002/>
- 20 Rare Diseases International. Rare Diseases: Leaving no one Behind in Universal Health Coverage. 2019. Available from URL: <https://d254mlohi254u805.cloudfront.net/rdi/2019/>
- 21 Brookes A J, Robinson P N. Human genotype-phenotype databases: aims, challenges and opportunities. *Nat Rev Genet*, 2015, 16: 702–715
- 22 National Cancer Institute. NCI Best Practices for Biospecimen Resources. 2016. Available from URL: <https://biospecimens.cancer.gov/bestpractices/2016-NCIBestPractices.pdf>
- 23 Backman J D, Li A H, Marcetta A, et al. Exome sequencing and analysis of 454,787 UK Biobank participants. *Nature*, 2021, 599: 628–634
- 24 Holub P, Swertz M, Reihs R, et al. BBMRI-ERIC directory: 515 biobanks with over 60 million biological samples. *Biopreserv Biobank*, 2016, 14: 559–562
- 25 Qualman S J, France M, Grizzle W E, et al. Establishing a tumour bank: banking, informatics and ethics. *Br J Cancer*, 2004, 90: 1115–1119
- 26 Workspace C S P. The cancer Biomedical Informatics Grid (caBIG): infrastructure and applications for a worldwide research community. *Stud Health Technol Inform*, 2007, 129: 330–334
- 27 Omae Y, Goto Y, Tokunaga K. National center biobank network. *Hum Genome Var*, 2022, 9: 38

- 28 Chu J Y, Xu J J, Fu S B, et al. The establishment of the immortalized cell bank of different Chinese ethnic groups (in Chinese). *Int J Genet*, 2008, 31: 241–247+301 [褚嘉祐, 徐玖瑾, 傅松滨, 等. 中华民族永生细胞库的建立. *国际遗传学杂志*, 2008, 31: 241–247+301]
- 29 Wang X, Lu M, Qian J, et al. Rationales, design and recruitment of the Taizhou Longitudinal Study. *BMC Public Health*, 2009, 9: 223
- 30 Li H Y, Zhang L, Zhang X J, et al. Preliminary exploration of results of Beijing biobank of clinical resources (in Chinese). *Chin J Hosp Admin*, 2013, 29: 863–865 [李海燕, 张雷, 张雪娇, 等. 北京重大疾病临床数据和样本资源库建设成果初探. *中华医院管理杂志*, 2013, 29: 863–865]
- 31 Du L L, Gao H J. Discussion on the sustainable development of biobank (in Chinese). *Transl Med J*, 2019, 8: 274–276+281 [杜莉莉, 郜恒骏. 生物样本库可持续性发展的探讨. *转化医学杂志*, 2019, 8: 274–276+281]
- 32 Gao H J, Du L L, Zhang X Y, et al. Status, opportunities and challenges of biobanks (in Chinese). *Med J Peking Union Med Coll Hosp*, 2018, 9: 172–176 [郜恒骏, 杜莉莉, 张小燕, 等. 生物样本库发展的现状、机遇与挑战. *协和医学杂志*, 2018, 9: 172–176]
- 33 Turro E, Astle W J, Megy K, et al. Whole-genome sequencing of patients with rare diseases in a national health system. *Nature*, 2020, 583: 96–102
- 34 Garcia M, Downs J, Russell A, et al. Impact of biobanks on research outcomes in rare diseases: a systematic review. *Orphanet J Rare Dis*, 2018, 13: 202
- 35 A network of bioresource facilities in Japan: the human bioresource consortium technical chapter (Japanese association for human bio-resource research). *Biopreserv Biobank*, 2013, 11: 57–63
- 36 Tada M, Hirata M, Sasaki M, et al. The Rare Disease Bank of Japan: establishment, current status and future challenges. *Hum Cell*, 2018, 31: 183–188
- 37 Filocamo M, Baldo C, Goldwurm S, et al. Telethon Network of Genetic Biobanks: a key service for diagnosis and research on rare diseases. *Orphanet J Rare Dis*, 2013, 8: 129
- 38 Conradie E H, Malherbe H, Hendriks C J, et al. An overview of benefits and challenges of rare disease biobanking in Africa, focusing on South Africa. *Biopreserv Biobank*, 2021, 19: 143–150
- 39 Kopper O, de Witte C J, Löhmußsaar K, et al. An organoid platform for ovarian cancer captures intra- and interpatient heterogeneity. *Nat Med*, 2019, 25: 838–849
- 40 Sachs N, de Ligt J, Kopper O, et al. A living biobank of breast cancer organoids captures disease heterogeneity. *Cell*, 2018, 172: 373–386.e10
- 41 Van de Wetering M, Francies H E, Francis J M, et al. Prospective derivation of a living organoid biobank of colorectal cancer patients. *Cell*, 2015, 161: 933–945
- 42 Schutgens F, Rookmaaker M B, Verhaar M C. A perspective on a urine-derived kidney tubuloid biobank from patients with hereditary tubulopathies. *Tissue Eng Part C Methods*, 2021, 27: 177–182
- 43 Bellomo F, De Leo E, Taranta A, et al. Drug repurposing in rare diseases: an integrative study of drug screening and transcriptomic analysis in nephropathic cystinosis. *Int J Mol Sci*, 2021, 22: 12829
- 44 Fusco F, Valente V, Fergola D, et al. The Incontinentia Pigmenti Genetic Biobank: study design and cohort profile to facilitate research into a rare disease worldwide. *Eur J Hum Genet*, 2019, 27: 1509–1518
- 45 Morrell N W, Aldred M A, Chung W K, et al. Genetics and genomics of pulmonary arterial hypertension. *Eur Respir J*, 2019, 53: 1801899
- 46 Walsh R, Jurgens S J, Erdmann J, et al. Genome-wide association studies of cardiovascular disease. *Physiol Rev*, 2023, 103: 2039–2055
- 47 Zhu N, Pauciulo M W, Welch C L, et al. Novel risk genes and mechanisms implicated by exome sequencing of 2572 individuals with pulmonary arterial hypertension. *Genome Med*, 2019, 11: 69
- 48 Patrick M T, Bardhi R, Zhou W, et al. Enhanced rare disease mapping for phenome-wide genetic association in the UK Biobank. *Genome Med*, 2022, 14: 85
- 49 Lhousni S, Belmokhtar K Y, Belmokhtar I, et al. Morocco's first biobank: establishment, ethical issues, biomedical research opportunities, and challenges. *Biomed Res Int*, 2020, 2020: 8812609
- 50 Gainotti S, Torreri P, Wang C M, et al. The RD-Connect Registry & Biobank Finder: a tool for sharing aggregated data and metadata among rare disease researchers. *Eur J Hum Genet*, 2018, 26: 631–643
- 51 Claussnitzer M, Cho J H, Collins R, et al. A brief history of human disease genetics. *Nature*, 2020, 577: 179–189
- 52 Lappalainen T, Scott A J, Brandt M, et al. Genomic analysis in the age of human genome sequencing. *Cell*, 2019, 177: 70–84
- 53 Mora M, Angelini C, Bignami F, et al. The EuroBioBank Network: 10 years of hands-on experience of collaborative, transnational biobanking for rare diseases. *Eur J Hum Genet*, 2015, 23: 1116–1123

- 54 Rubinstein Y R, Posada de la paz M, More M. Rare disease biospecimens and patient registries: interoperability for research promotion, a European example: EuroBioBank and SpainRDR-BioNER. *Adv Exp Med Biol*, 2017, 1031: 141–147
- 55 Guo J, Liu P, Chen L, et al. National Rare Diseases Registry System (NRDRS): China's first nation-wide rare diseases demographic analyses. *Orphanet J Rare Dis*, 2021, 16: 515
- 56 Orphanet Report Series-Rare Disease Registries in Europe. 2021. Available from URL: <http://www.orpha.net/orphacom/cahiers/docs/GB/Registries.pdf>
- 57 Rubinstein Y R, Groft S C, Bartek R, et al. Creating a global rare disease patient registry linked to a rare diseases biorepository database: Rare Disease-HUB (RD-HUB). *Contemp Clin Trials*, 2010, 31: 394–404
- 58 Bladen C L, Salgado D, Monges S, et al. The TREAT-NMD DMD Global Database: analysis of more than 7,000 duchenne muscular dystrophy mutations. *Hum Mutat*, 2015, 36: 395–402
- 59 Vaught J. Biobanking comes of age: the transition to biospecimen science. *Annu Rev Pharmacol Toxicol*, 2016, 56: 211–228

Development and application of rare diseases biobank

GUO Dan^{1,2,3}, JIN Ye^{2,3}, LIU WeiDa^{2,3}, ZHANG YaRan^{1,2,3}, WANG AnQi^{1,2,3} & ZHANG ShuYang³

1 Clinical Biobank, Peking Union Medical College Hospital, Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing 100730, China;

2 Medical Research Center, Peking Union Medical College Hospital, Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing 100730, China;

3 State Key Laboratory of Complex Severe and Rare Diseases, Peking Union Medical College Hospital, Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing 100730, China

Rare diseases (RD) are a general term for a group of diseases with extremely low incidence or prevalence and have been a challenge for the global medical community due to their diverse nature, difficulty in diagnosis, and limited treatment options. With the rapid development of translational medicine, countries have gradually increased investment in the research, diagnosis, and treatment of RD and drug development in recent years. Biobanks have been successfully applied in the field of common diseases and also play an important role in the field of RD. RD biobank integrates biological sample information with patient clinical information and multi-omics data to promote accurate phenotype and pathogenesis research of diseases, support high-level and large-scale cohort studies, accelerate the development and efficacy evaluation of new drugs and new treatment technologies, and support the whole chain of research modes from basic research to clinical transformation. This article reviews the development and application of clinical biobank in the field of RD, aiming to promote the establishment of a standardized, normalized, informationized, diversified and shareable model of RD biobank, and promote global collaboration and translational application, which is of great significance for the improvement of RD diagnosis and treatment capabilities and the development of medical science.

rare disease, biobank, standardization, informationization, transformation and sharing

doi: [10.1360/SSV-2023-0038](https://doi.org/10.1360/SSV-2023-0038)