

生物样本库信息化建设的现状分析与展望

李海欣,何娜,陈可欣

(天津医科大学肿瘤医院,国家肿瘤临床医学研究中心,天津市“肿瘤防治”重点实验室,天津 300060)

摘要:随着后基因组时代的来临和信息技术的迅猛发展,传统的样本信息收集已不能满足当今生物样本库发展的需要,信息化的发展和应用迅速提升了样本信息采集、处理的速度,提高了工作效率及样本利用率,能够满足大规模、高效收集、管理和利用样本信息、临床信息和流行病学信息的需要。同时,为转化医学及个体化治疗提供信息基础。全文就样本库信息化建设的内容、产生的问题、前景等方面进行阐述。

关键词:生物样本库;信息化;整合平台

中图分类号:R197 文献标识码:B 文章编号:1004-0242(2015)04-0262-06

doi:10.11735/j.issn.1004-0242.2015.04.A003

Present Situation and Prospects of Information Construction of Biobank

LI Hai-xin, HE Na, CHENG Ke-xin

(National Clinical Research Center of Cancer, Tianjin Medical University Cancer Institute and Hospital, Tianjin 300060, China)

Abstract: With the advent of post-genome era and the rapid development of information technology, the traditional biobank which only store and collect samples cannot satisfy the needs of the development of the biobank. Development of information and applying informatization management have rapidly improved the speed of sample information collection and can meet the needs of large-scale and efficiency collection, management and utilization of sample, sample clinical and epidemiological information. Meanwhile, application of information set a foundation for translation medicine researches and personalized medicine. The contents, problems and challenges were comprehensive illustrated in this article.

Key words: biobank; information; integration platform

生物样本库是遵照国际标准化指南 [诸如国际生物与环境样本库协会 (ISBER)、美国国家癌症研究所 (NCI) 及其所辖的生物样本库和生物样本研究办公室 (OBRR) 制定的生物标本处理与质量控制的最佳实践方案] 采集、处理、运输和储存生物样本以及伴随这些生物样本相关的临床、随访等信息的应用系统^[1]。自 20 世纪 90 年代末以来生物样本库已成为医学研究的重要资源, 支持多种类型的科学研究 (如基因组学与个体化医疗)^[2]。

随着后基因组时代的来临和信息技术的发展, 生物数据的积累出现了前所未有的飞跃, 数据的本

质出现了从生理生化数据向遗传信息的飞跃, 生物样本库也从传统的储存样本向信息化方向发展, 如何收集和利用生物样本蕴含的海量生物学信息是摆在我们面前的关键问题。而建立信息化的生物样本库可以解决这一难题。生物样本库的信息化建设不仅能够更为全面有效的搜集、保存、检索、分析和管理工作, 还能挖掘和利用生物样本及其相关信息。生物样本的信息化建设主要包括以下三个方面: (1) 生物样本的信息化管理 (采集、处理、储存和运输, 管理 QA/QC 程序过程中信息的追踪与处理)。(2) 样本相关临床、随访信息的信息化管理。(3) 生物样本库信息整合平台的建设。本文将从以上三个方面来展开综述并对样本库信息化建设当中产生的问

收稿日期: 2014-10-20; 修回日期: 2014-12-24

通讯作者: 李海欣, E-mail: lihaixin@tjmuch.com

题进行反思与展望。

1 生物样本的信息化管理

生物样本资源属于不可再生资源,高质量的生物样本对于生物医学研究起着越发重要的作用,信息化管理系统能够为建设高质量、高标准的生物样本库提供有力的支持。样本的信息化管理需要满足以下条件:(1)生物样本应有唯一的标识,能够被识别并追踪。(2)应能记录样本质量,如样本的储存温度环境和冻融次数等。样本标识的种类多种多样,目前还没有统一的金标准。根据2012年国际生物和环境样本库协会(ISBER)对于合作机构的统计报道,51%的合作机构已经使用二维码,二维码能够有效的节省空间、快速准确地对日常样本进行管理^[3,4]。法国生物样本库开发了短信服务系统,致力于通过手机短信建立样本贡献者与样本库之间的联系,通过短信系统生成唯一的样本编号,患者可以通过与个人信息相关的(生日、样本号等组合)认证密码确认或更改知情同意信息,获得信息反馈,同时给了患者更多的知晓并保护个人遗传信息的空间,并有助于匿名数据管理,平衡了研究需要和病人隐私之间的关系^[5]。2004年美国国家癌症研究所开发了名为caTissue的生物样本信息资源管理的网络软件,该软件是目前世界上最大的生物样本信息资源交流平台^[6],制定了一整套高效的临床样本管理策略,在全美多家癌症中心进行运行和推广。我国生物样本库信息化建设虽然起步较晚,但是各个样本库也越来越注重信息化管理与整合。目前有多多个规模较大的样本库逐步采用以二维码平台为基础的样本管理系统(如美国RURO公司的FreezePro 2012样本管理系统,上海万格样本管理系统等)。这些信息系统的应用带动了样本库的精细化管理,不仅可以快速追踪样本信息和记录,而且对于生物样本库的科学收集策略进行了科学规范化管理,有效的节约时间和空间,增加样本管理的准确度。

2 样本相关信息的信息化管理

样本相关信息(见附录1)包括临床信息(既往史、家族史、诊断结果、病理信息、影像学信息、治疗

相关信息等)、随访信息(预后信息、生存信息等)和流行病学信息(人口信息、家谱信息、基因组学、蛋白组学、表型信息、生活方式、环境暴露因素等信息)。人类基因组计划的发展使得生物医学研究向着探索疾病多因素(包括研究表观遗传学、生活方式和环境暴露的交互作用)方向发展。因此,仅有样本信息是不够的,只有全面收集样本衍生出来的相关信息,才能全面深入探索疾病生物学网络,指导未来的个体化医疗。

2.1 临床和随访信息的管理

传统的电子病历已经不能满足当今医学研究的发展,整合临床和科研应用的电子病历系统进行长期、动态、连续、大规模的数据积累,将丰富的临床病历资源转化为医学研究资源是当前的发展趋势。目前国外研发的结构化电子病历能够实现快速浏览,将电子病历与临床随访病历合为一体并能够按照研究需求进行文本检索和数据挖掘,其在线功能还能在很大程度上解决不同机构在病历获取时的沟通问题,使得研究过程得到显著优化,研究效率、研究质量和研究能力得到大幅度的提升^[7-10]。因此,科研电子病历系统为生物样本库信息平台的整合奠定了基础。

2.2 流行病学信息的管理

生物医学正在面临解密基因与环境对人类健康的威胁的重大挑战,为了解决这个问题,基于人群的前瞻性的生物样本信息资源就显得尤为重要。全面而可靠地收集流行病学信息(问卷调查)能够深入的挖掘疾病与环境或基因的关系。流行病学信息包括基于不同疾病的危险因素和暴露因素,人口学信息、生活方式、家族史、治疗史、物理措施等。其中具有代表性的英国生物样本库是基于人群的生物样本信息资源库,收集年龄在40~69岁之间的样本502 682份,流行病学等信息在建库时通过问卷采集,其中如饮食和运动等难以用普通问卷来准确计量的因素采用更为详尽的网络调查问卷系统。研究者可根据不同研究需要获取流行病学问卷调查结果,通过收集大量信息探究基因与环境的交互作用,从而获得治疗与预防疾病的方法,现已发表多篇基于队列研究的文章^[11-21]。我国广州生物样本库联合香港大学公共卫生学院和英国伯明翰大学医学院创建了基于人群的生物样本库,主要通过建立中老年队列,研究慢性重大疾病与其流行病学因素的关系,从而扩展研

究的广度与深度,已经研究发现多种生活及饮食因素与慢性病的发生有关^[22-30]。由此,全面系统的掌握流行病学信息,对于未来的多学科交互研究具有重大意义。

3 生物样本库信息整合平台的建设

随着医疗模式逐步向个体化医疗发展,将基础研究信息与临床信息等的整合也变得越发重要。现代化的基因组扩增平台,例如高通量的基因测序平台,已经生成了上百个基因数据库。通过整合病人的临床数据、生物样本以及基因信息,能够发现新的生物标志物及分子靶点、加强为病人选择合适的治疗方案或加入临床试验的能力、提供高质量的数据以支持临床决策并能够跟踪病人病情、治疗方案、治疗结果及相关费用(附录2)。与此同时,生命科学的研究正在向生物样本库之间信息共享的国际多学科合作方向发展。但问题在于不同的样本库使用不同的程序,而且从地区到全球规模的研究合作的进程中,必然带来异质性增长问题,目前还没有统一的标准。因此,迫切需要建立标准化的生物样本库信息整合平台来解决这些问题。

近年来,生物样本库已经从收集组织和血液样本为中心向收集样本所有信息(包括社会、临床和病理资料与基因组谱)的方向演变。生物样本库需要全面的管理系统来追踪样本相关的所有活动和信息。因此,生物样本信息资源库需要采取一种理想的元数据模型来处理异构数据集的集成。目前,为了推进生物样本库信息系统整合建设,国外的很多机构已经研发出了多种软件系统。其中泛欧洲生物样本库与分子生物资源研究中心(BBMRI)开发了一个最小的数据集用于多中心共享生物样本^[31]。欧洲生物信息研究所(European Bioinformatics Institute)开发了SIMBioMS系统,在生物医学研究中起到收集、储存、管理和检索研究课题、生物样本和高通量的实验数据信息的作用(包括高通量实验获得的实验数据)^[32]。国立生物医学计算中心(National Center for Biomedical Computing)研发了i2b2系统为临床和转化医学研究提供临床和基因组数据的整合平台^[33]。意大利热那亚大学生物信息学和机械系统工程学院(DIB-RIS)建立了XTENS平台进行多学科多中心的合作,

有效追踪病人的临床记录,样本管理信息和基因组数据^[34]。韩国国立生物样本库(NBK)研发了CHIS系统整合各医院收集的生物样本资源和研究需要的临床信息^[35]。意大利锡耶纳大学分子生物学系建立了RTT数据库鼓励来自全球临床医生和研究人员的合作项目,并提供有助于更好地了解Rett综合征潜在发病机制的重要生物资源^[36]。虽然国际上已经开发了多种信息整合系统,但还没有全球通用、标准化的系统平台诞生,在一定程度上阻碍了各国间的合作。

4 我国样本库建设中存在的问题与展望

随着当今医学生物技术和医学信息学的高速发展使得建立多学科多中心(包括基础、流病、临床)研究之间的桥梁至关重要。而生物样本库的信息化建设正是建立这座桥梁的关键平台。目前,关于样本库的信息化各国都在积极建设中,但这个过程也存在着一一些问题,针对我国样本库信息化建设现状作者认为存在以下问题:

(1)样本信息编码没有统一的标准:我国各个研究机构的编码方式良莠不齐,使得非本机构的专业人员无法识别其他机构的样本编码,造成研究机构之间交流的障碍,样本信息因此不能得到有效的利用。目前,我国还缺乏行之有效的样本编码准则。

(2)缺乏科研电子病历系统:我国从1995年开始,各地陆续开展了电子病历的研究工作,但到目前为止我国的电子病历还处于半结构化状态仅是完成文本录入,不能将其以字段化的方式与样本库等其他信息系统相连接,如与科研相结合还需二次录入,严重阻碍了样本库信息化建设的脚步。

(3)缺乏全国性的生物样本资源库网络:世界各国都在积极建设生物样本库信息交流平台,其中美国国家癌症研究机构研发的caTissue生物样本资源库管理软件是目前全球最大的生物样本资源交流平台。我国还缺乏全国性的生物样本资源库网络,样本信息还无法与临床等其他信息整合,使得样本信息资源无法共享,阻碍了我国生物样本库之间及在世界范围内的交流。

(4)生物样本库信息平台建设缺乏统一的管理、使用、共享等标准化流程及规范:我国信息化管理水

平比较落后,受到样本来源、科研体制及管理机制等因素等的限制,缺乏统一的运行标准,尚不能与国际的标准体系接轨。导致样本相关资料收集不全、流失及不能充分利用的情况发生,大大降低了样本的科研应用价值。

(5) 尚不存在完善的分类别的生物样本库信息系统建设。随着转化医学的发展,我国的一些科研机构认识到生物样本库信息系统建设的重要性,建立起了以科研为主的干细胞库,肿瘤组织库,出生队列库等,但是缺乏更为全面的分类别的信息系统(如按疾病分:代谢性疾病、肾脏疾病、肿瘤类等)。还没有通过分类别进行标准化、信息化储存与管理,不能随时供各医院调用,降低了不同类别生物样本的应用价值。

生物样本库信息化建设的最终目的就是将样本信息与临床信息、随访信息、流行病学信息相结合,对样本信息和样本相关信息进行综合管理和有机整合。把生物样本库建设成为以样本库信息系统为核心,临床信息系统、随访系统、流行病学信息系统及其他系统相辅助的信息管理与交流平台。

我国样本资源丰富,蕴含着大量遗传信息,如何有效的挖掘和利用样本库这座遗传信息的“金矿”,对于我国样本信息资源库参与国际化的竞争与合作有着重大战略意义。因此,迫切需要加强科普工作,争取政府和民众对于生物样本库建设工作的理解和支持,增加参与度;增加政府对于样本库长期而稳定的支持和投入;提高基础数据服务的水平通过样本数据库信息管理平台实现样本信息数字化及远程共享;制定统一的样本库信息资源管理标准(采用全国通用的样本信息资源库管理系统),促进样本信息与临床等信息资源的整合;增加生物样本库的疾病种类并分类进行标准化、信息化储存与管理,充分发挥各类生物样本的应用价值建立国家生物样本库信息资源共享网络,有助于促进地区及国家间的交流与合作。相信建立标准化,信息资源整合、共享机制的生物样本库能够将我国丰富的样本资源通过科研成果转化为集预防、诊断、治疗、为一体的个体化医疗,为人类的健康贡献力量。

参考文献:

[1] Vaught J, Rogers J, Myers K, et al. An NCI perspective on

creating sustainable biospecimen resources[J]. J Natl Cancer Inst Monogr, 2011, 2011(42): 1-7.

[2] Greely HT. The uneasy ethical and legal underpinnings of large-scale genomic biobanks [J]. Annu Rev Genomics Hum Genet, 2007, 8: 343-364.

[3] Fearn P, Michels C, Meagher K, et al. 2012 International Society for Biological and Environmental Repositories Informatics Working Group: survey results and conclusions [J]. Biopreserv Biobank, 2013, 11(1): 64-66.

[4] Nussbeck SY, Skrowny D, O'Donoghue S, et al. How to design biospecimen identifiers and integrate relevant functionalities into your biospecimen management system [J]. Biopreserv Biobank, 2014, 12(3): 199-205.

[5] Tebbakha R. Biobank-short message service for linking patients and samples[J]. Telemed J E Health, 2013, 19(9): 717-721.

[6] The caBIG Strategic Planning Workspace. The Cancer Biomedical Informatics Grid (caBIG): infrastructure and applications for a worldwide research community [J]. Stud Health Technol Inform, 2007, 129(Pt 1): 330-334.

[7] Luder S. Electronic medical record. Fiction of a safe digital identity in health care [J]. MMW Fortschr Med, 2014, 156(3): 17.

[8] Onieva-Garcia MA, Molina-Rueda MJ, Almagro-Lopez D, et al. Epidemiological survey versus the electronic medical record in the investigation of a Cryptosporidium outbreak in a day school[J]. Gac Sanit, 2015, 29(1): 76-77.

[9] Matsumura Y, Hattori A, Manabe S, et al. Interconnection of Electronic Medical Record with Clinical Data Management System by CDISC ODM[J]. Stud Health Technol Inform, 2014, 205: 868-872.

[10] Hall JB, Dumitrescu L, Dilks HH, et al. Accuracy of administratively-assigned ancestry for diverse populations in an electronic medical record-linked biobank [J]. PLoS One, 2014, 9(6): e99161.

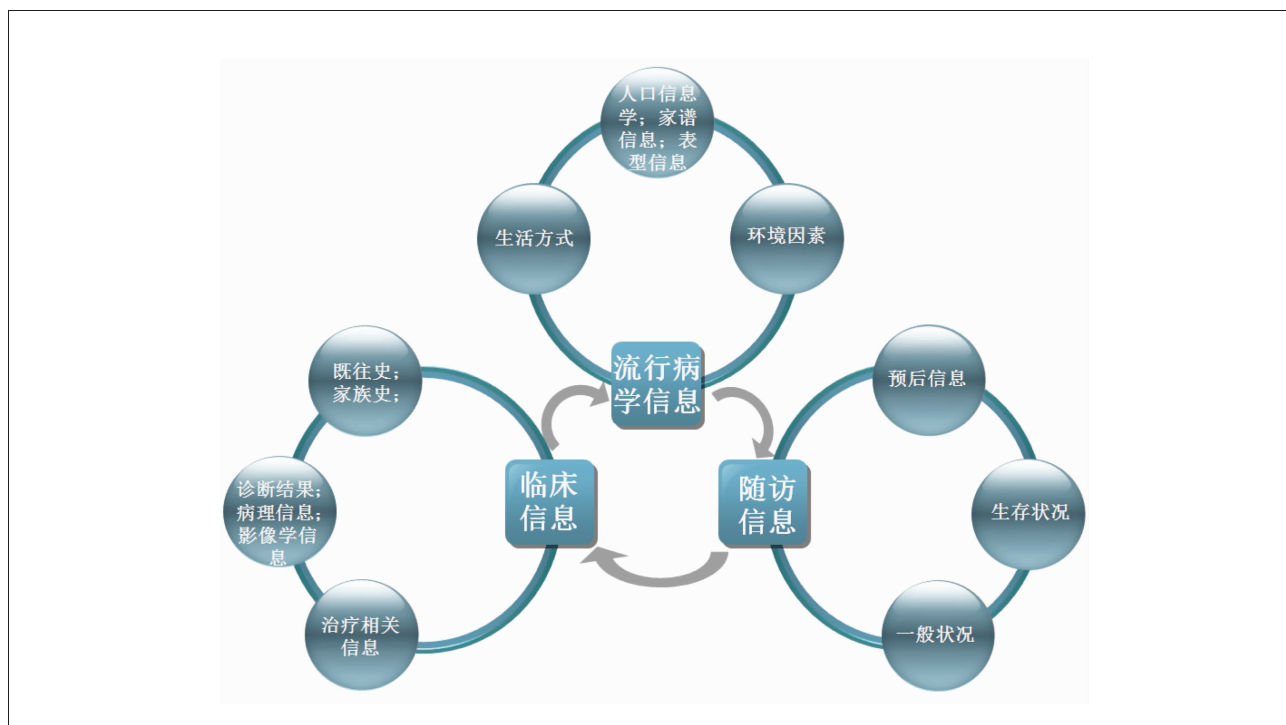
[11] Hutchings S, Ayres J, Cullinan P, et al. Using the UK Biobank study to estimate occupational causes of chronic disease: comparability with the UK national population and adjustment for bias [J]. Occup Environ Med, 2014, 71 Suppl 1: A79.

[12] De Matteis S, Rushton L, Jarvis D, et al. A new, efficient web-based tool to collect and code lifetime job histories in large population-based studies: the COPD project in the UK Biobank cohort[J]. Occup Environ Med, 2014, 71 Suppl 1: A19.

[13] Cleutjens FA, Spruit MA, Ponds RW, et al. Cognitive func-

- tioning in obstructive lung disease; results from the United Kingdom biobank [J]. *J Am Med Dir Assoc*, 2014, 15(3): 214–219.
- [14] Ntuk UE, Gill JM, Mackay DF, et al. Ethnic-specific obesity cutoffs for diabetes risk; cross-sectional study of 490,288 UK Biobank participants [J]. *Diabetes Care*, 2014, 37(9): 2500–2507.
- [15] Ul-Haq Z, Smith DJ, Nicholl BI, et al. Gender differences in the association between adiposity and probable major depression; a cross-sectional study of 140,564 UK Biobank participants [J]. *BMC Psychiatry*, 2014, 14: 153.
- [16] Ul-Haq Z, Mackay DF, Martin D, et al. Heaviness, health and happiness; a cross-sectional study of 163066 UK Biobank participants [J]. *J Epidemiol Community Health*, 2014, 68(4): 340–348.
- [17] Spruit MA, Sillen MJ, Groenen MT, et al. New normative values for handgrip strength; results from the UK Biobank [J]. *J Am Med Dir Assoc*, 2013, 14(10): 775.
- [18] Harvey NC, Matthews P, Collins R, et al. Osteoporosis epidemiology in UK Biobank; a unique opportunity for international researchers [J]. *Osteoporos Int*, 2013, 24 (12): 2903–2905.
- [19] Tyrrell JS, Yaghootkar H, Freathy RM, et al. Parental diabetes and birthweight in 236 030 individuals in the UK Biobank Study [J]. *Int J Epidemiol*, 2013, 42 (6): 1714–1723.
- [20] Smith DJ, Nicholl BI, Cullen B, et al. Prevalence and characteristics of probable major depression and bipolar disorder within UK biobank; cross-sectional study of 172,751 participants [J]. *PLoS One*, 2013, 8(11): e75362.
- [21] Capps BJ, van der Eijk Y. The tobacco industry, researchers, and ethical access to UK Biobank; using the public interest and public good [J]. *Am J Public Health*, 2014, 104(10): 1833–1839.
- [22] Long MJ, Jiang CQ, Lam TH, et al. Alcohol consumption and electrocardiographic left ventricular hypertrophy and mediation by elevated blood pressure in older Chinese men; the Guangzhou Biobank Cohort Study [J]. *Alcohol*, 2013, 47(6): 473–480.
- [23] Au YS, Jiang C, Cheng KK, et al. Aldehyde dehydrogenase 2-a potential genetic risk factor for lung function among southern Chinese; evidence from the Guangzhou Biobank Cohort Study [J]. *Ann Epidemiol*, 2014, 24(8): 606–611.
- [24] Adab P, Jiang CQ, Rankin E, et al. Breast feeding practice, oral contraceptive use and risk of rheumatoid arthritis among Chinese women; the Guangzhou Biobank Cohort Study [J]. *Rheumatology (Oxford)*, 2014, 53(5): 860–866.
- [25] van Hasselt TJ, Pickles O, Midgley-Hunt A, et al. Effects of tea consumption on renal function in a metropolitan Chinese population; the Guangzhou biobank cohort study [J]. *J Ren Nutr*, 2014, 24(1): 26–31.
- [26] Zhao J, Jiang CQ, Lam TH, et al. Genetically predicted 17beta-estradiol and systemic inflammation in women; a separate-sample Mendelian randomisation analysis in the Guangzhou Biobank Cohort Study [J]. *J Epidemiol Community Health*, 2014, 68(8): 780–785.
- [27] Zhao J, Jiang C, Lam TH, et al. Genetically predicted testosterone and cardiovascular risk factors in men; a Mendelian randomization analysis in the Guangzhou Biobank Cohort Study [J]. *Int J Epidemiol*, 2014, 43 (1): 140–148.
- [28] Zhong Y, Jiang CQ, Cheng KK, et al. Height, its components, and coagulability among older Chinese: The Guangzhou biobank cohort study [J]. *Am J Hum Biol*, 2014, 26(5): 603–608.
- [29] Gui H, Jiang CQ, Cherny SS, et al. Influence of Alzheimer's disease genes on cognitive decline; the Guangzhou Biobank Cohort Study [J]. *Neurobiol Aging*, 2014, 35(10): 2422–2423.
- [30] Sun Y, Jiang CQ, Cheng KK, et al. Leg length is associated with lower values of inflammatory markers in older Chinese: The Guangzhou Biobank Cohort Study [J]. *Ann Hum Biol*, 2015, 42(2): 143–149.
- [31] Norlin L, Fransson MN, Eriksson M, et al. A Minimum Data Set for Sharing Biobank Samples, Information, and Data: MIABIS [J]. *Biopreserv Biobank*, 2012, 10(4): 343–348.
- [32] Krestyaninova M, Zarins A, Viksna J, et al. A System for Information Management in BioMedical Studies--SIM-BioMS [J]. *Bioinformatics*, 2009, 25(20): 2768–2769.
- [33] Abend A, Housman D, Johnson B. Integrating Clinical Data into the i2b2 Repository [J]. *Summit on Translat Bioinforma*, 2009, 2009: 1–5.
- [34] Izzo M, Mortola F, Arnulfo G, et al. A digital repository with an extensible data model for biobanking and genomic analysis management [J]. *BMC Genomics*, 2014, 15 Suppl 3: S3.
- [35] Kim H, Yi BK, Kim IK, et al. Integrating clinical information in National Biobank of Korea [J]. *J Med Syst*, 2011, 35 (4): 647–656.
- [36] Sampieri K, Meloni I, Scala E, et al. Italian Rett database and biobank [J]. *Hum Mutat*, 2007, 28(4): 329–335.

附录 1 样本相关信息示意图



附录 2 生物样本库信息整合平台示意图

