

精准医疗背景下生物样本信息的价值初探

王弋波

(中国科学技术信息研究所, 北京 100038)

摘要: 精准医疗的产业化进程赋予生物样本信息新的价值, 同时使生物样本信息的管理面临全新局面。基于国内外政策制度归纳生物样本信息的定义, 通过对精准医疗背景下生物样本生命周期各环节信息活动的解析, 梳理生物样本信息的具体内容, 构建生命周期中的信息流, 结合信息活动主体在信息流中的关系, 说明生物样本信息管理的需求与实现, 揭示生物样本信息的价值定位, 提出制订数据标准、建立云平台、深化理论研究等建议。

关键词: 生物样本; 信息管理; 精准医疗; 价值定位

中图分类号: G203

文献标识码: A

DOI: 10.3772/j.issn.1674-1544.2017.04.003

Value Proposition of Bio Specimen Information Under Precise Medical Background

WANG Yibo

(Institute of Scientific & Technical Information of China, Beijing 100038)

Abstract: Bio specimen information had been given new value by the industrialization process of precision medical, and the information management of Bio specimen is facing a new situation. By inducing the domestic and foreign policies, the definition of Bio specimen information had been clear. A lifecycle of Bio specimen under Precise Medical background was described to analysis the information active and content in its stages. Through the construction of information flow and subjects relationship, the demand of active subjects and its realization in the lifecycle was revealed, and 4 features was proposed to indicate the value proposition of Bio specimen information. 3 proposals was given, involving data standards, cloud platform and theoretical research.

Keywords: bio specimen, information management, precise medical, value proposition

1 引言

2015年, 美国率先提出“精准医疗”发展计划^[1], 2016年我国颁布的《“十三五”国家科技创新规划》将“精准医疗”以及“精准医疗关键技术”纳入国家重点部署工作^[2]。中美两国都对“精准医疗”的关注预示其即将到来的产业化进程。分子生物学、遗传学、转化医学等相关学

科的理论发展和技术突破为精准医疗产业提供了技术前景^[3], 然而在后互联网时代的浪潮中, 与大数据、云计算、人工智能等信息技术的充分结合将为精准医疗产业发展提供必要条件^[4]。

精准医疗技术研发和临床治疗的核心工作是“通过组学技术, 对大样本人群与特定疾病类型进行生物标记物的分析与鉴定、验证与应用”和“通过精确寻找靶点、精确亚分类等方式实施个

作者简介: 王弋波 (1985—), 男, 中国科学技术信息研究所助理研究员, 研究方向: 科技信息管理。

基金项目: 国家重点研发计划项目“中国人类遗传资源样本库建设”之课题“中国人类遗传资源样本库信息管理平台建设”(2016YFC1201702)。

收稿日期: 2017年6月25日。

性化精准治疗”^[5]。无论是分析生物标记物还是研发靶向药物，都必须完成生物样本的采集、制备、储存等环节，最终依靠基因测序等技术手段获取相关的组学数据。但仅凭组学数据并不足以发现基因、环境和疾病之间错综复杂的关联。这种关联必须通过组学数据和生物样本信息的对照分析来建立。因此，生物样本是精准医疗产业运行的物质基础，生物样本信息则贯穿整个产业链，发挥着十分关键且无法替代的作用。

生物样本信息的建设水平决定了精准医疗产业的信息化和智能化程度，也决定了产业的发展速度和未来可能承担的风险^[6]。目前，国内外有关生物样本信息的学术研究和管理探索并不鲜见^[7-10]，但大多局限于生物样本管理信息系统或相关数据库建设层面。尤其是国内研究多关注于样本库管理制度、样本制备和保存测试等问题^[11]，对生物样本信息的供需关系、产业应用、内容分化及其关联结构等问题缺乏关注。随着生物样本信息的重要性日渐突出，相关行业标准和技术规范已经在国际科技竞争中呈现出较高的战略意义^[12-14]。为此，本文对精准医疗背景下生物样本信息的价值进行初步探讨。

2 生物样本信息的内涵

在学术界和产业界，生物样本的基本定义已经形成基本共识。但由于基因组、蛋白质组、代谢组等组学技术长期处于精准医疗相关研究的核心地位，生物样本信息的定义随研究内容变化呈现较大差异，例如过于宽泛的“生物信息”^[3]或过于局限的“临床信息”^[15]等，相对尚存争议。现从政府的政策和管理来看生物样本信息的定义则更加清晰。

目前，国内外对生物样本通常从“人类遗传资源”和“医学研究样本”两方面实施，前者强调人类生物样本的遗传价值和长期保存，后者注重医学伦理相关制度。例如，经济合作与发展组织（OECD）于2009年发布的《人体生物银行和遗传研究数据库指南》中，对生物样本库管理对象的界定为“用于遗传研究的结构化资源包括：

(a)人类遗传材料和/或对这些遗传材料进行分析生成的信息；(b)相关联的信息”^[16]。此处尤其突出“人类遗传材料”，以区别于OECD对包括动植物、细菌病毒等更广泛“生物遗传资源”的管理。类似地，美国癌症研究院（NCI）对生物银行界定为“生物银行是指用于研究的人类生物样本、相关数据、数据存储的实体以及相关流程和政策的集合”^[17]。虽然两者表述的对象相同，但OECD更关注实体的“遗传材料”，并突出生物样本与其关联信息的关系，即按照信息来源将其关联信息区分为“分析生成的”和“相关联的”两类；NCI主要强调样本信息的数据形态、管理流程和政策环境，突出医学研究对生物样本管理的需求，即对样本实体、关联数据和伦理法规认可的需求。

在国内政策法规中，国务院于1998年颁布的《人类遗传资源管理暂行办法》是我国目前唯一全面管理人类遗传资源的法规文件，将人类遗传资源界定为“人类遗传资源是指含有人体基因组、基因及其产物的器官、组织、细胞、血液、制备物、重组脱氧核糖核酸（DNA）构建体等遗传材料及相关的信息资料”^[18]。此处主要明确“遗传材料”的范畴，但并未展开阐述“相关信息资料”的内涵。在医疗行业层面，中国医药生物技术协会将生物样本库的管理对象界定为“包含任何包含人体生物信息的生物物质，包括人体组织、血液、分泌物、排泄物及其衍生物”^[19]。此处并未涉及生物样本的信息，可见国内医疗行业对相关领域的认识尚有不足。

基于管理目的和行业背景的因素，虽然各政策主体对生物样本信息的描述方式或有差异，但表达的判断依据基本一致。综合两种管理口径可以看到，生物样本是指以实物形式存在的“遗传材料”或“生物物质”，而与生物样本相关联的信息均可纳入生物样本信息的范畴。但生物样本只是转化医学研究中承载人体生物信息的载体，在样本采集和制备中产生，在样本使用中被消耗。生物样本信息的描述对象可以是生物样本实体，也可以是生物样本在生命周期各环节中的

活动。因此，考虑到生物样本信息描述对象的差异，归纳相关政策规范可以得到生物样本信息的如下定义：狭义的生物样本信息是指生物样本的实物形态及其保存状态相关的信息；广义的生物样本信息是指在生物样本的采集、处理、存储、利用等过程中所生产和记录的相关信息。

下面再以生物样本的生命周期来看生物样本信息的内涵。生物样本的生命周期及其相关信息内容如图1所示。其中，围绕生物样本实体的产生和消耗，可以概略地将生物样本生命周期分解为采集、储存、使用3个主要环节；考虑对生物样本的组织管理，还可以将其生命周期向上追溯到科研课题管理环节；考虑生物样本所承载生物信息的分析研究，还可以将生物样本生命周期向下延伸至生物医药研发环节。

据此可以看到，狭义的生物样本信息仅包括生物样本储存环节所产生的信息，广义的生物样本信息应当涵盖其整个生命周期。虽然生物样本生命周期各环节的最终目标均是实现转化医学研究或临床治疗，但工作内容和专业技术差异较大，各环节产生信息的内容也各不相同。

(1) 科研课题管理环节是指相关领域研究人

员在设计研究方案、组织课题实施的过程中所实施的管理工作，例如研究机构的科技项目管理或企业的产品（技术）研发管理。科研课题管理环节产生的信息主要涉及两个方面：一是资助机构信息，这决定了后续环节产生的大部分信息的权属关系；二是研究内容信息，这部分信息是对课题所生成生物样本的整体描述，并为生物样本的采集、处理、储存、使用、分析以及质量控制等后续环节提供依据。

(2) 生物样本采集环节是指研究人员通过临床检验或人群调查等方式，采集捐献人的生物物质，并记录相关信息的活动。采集环节不仅为生物样本的生产提供了原始材料，而且是收集捐献人表型信息和环境信息的重要过程。仅从信息传播层面看，在采集环节将捐献人的全部动态活动转变为静态的生物样本实体和数据记录。从而转化为医学研究所关注的基因、表型和环境等信息。因此，采集环节产生的信息包括对两个主体的描述：一是对捐献人的描述，具体包括人口学信息和表型信息，其中表型信息又因临床研究和遗传学研究的差异，分别以病历信息和样本采集调查表的形式体现（精准医疗领域通常为病历

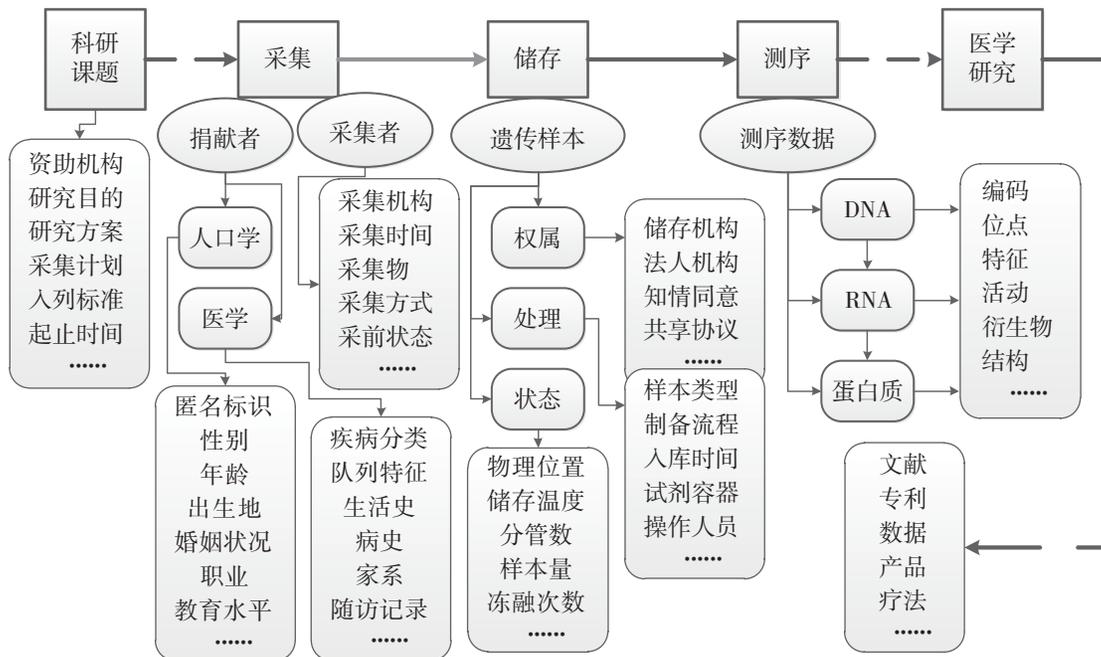


图1 生物样本生命周期及信息内容

信息)；二是对采集人和采集过程的描述，即采集机构、采集时间、采集方式、捐献者采前状态等。

(3) 生物样本储存环节是指研究人员将采集到的原始材料处理成为可长期保存的形态，并为生物样本保存提供妥善环境和相应管理的活动。原始材料的处理过程对生物样本所保留生物信息的影响较大，且不同研究内容均有其严格对应的处理方式和保存方式。此外，随着生物样本集中保存机制的广泛应用，专业化的管理部门还需要明确生物样本的权属信息。因此，储存环节产生的信息包括3个部分：一是生物样本的权属信息，即储存机构、样本所有人、捐献人知情同意书、使用条件等；二是处理过程信息，即样本内容物的类型标注、标准操作流程(SOP)及制备过程记录、入库时间、使用的试剂和容器等；三是保存状态信息，即生物样本的物理位置、储存温度、冻融记录等。

(4) 生物样本使用环节通常是指研究人员将保存的生物样本用于基因测序、蛋白质检测及生物标志物筛查等检验检测活动。而在精准医疗的部分特殊场景下，生物样本使用环节也可以是医疗人员将其直接用于临床治疗活动中，例如干细胞移植治疗。检验检测活动会消耗生物样本，使其失去实体形态，但产生的各类组学数据则是相关研究的核心信息。因此，按照组学信息描述对

象的差异，使用环节产生的数据分别对应DNA、RNA和蛋白质三大类，各类下细分的检验项目所包含的参数又存在一些差异。

(5) 生物医药研发(医学研究)环节是指科研人员综合组学、表型和环境等信息开展研究，最终产生科技文献、科学数据、技术专利、新产品和新疗法等各类成果。无论其商业价值如何，研究成果与生物样本之间的关联关系都具有重要的科研参考价值，所以记录这种关系的信息则需要在生物样本管理中妥善保存。

3 生物样本信息流与信息管理

虽然生物样本生命周期各环节的工作内容不同，专业技术存在很大差异，但各环节产生的生物样本信息并非孤立存在。仅从各环节信息内容的描述对象区分，可以将生物样本信息内容区分为“实体信息管理”“权属信息管理”和“科研信息管理”3个层面。各环节具体工作所处层面以及相应信息流如图2所示。

(1) 实体信息管理是对生物样本从原始材料的采集到处理为可长期保存的形态后入库保藏直至在检验检测中被消耗的过程描述，主要强调生物样本实体形态的准确描述和活动留痕。实体层中的各项活动按照先后次序逐一发生，每一项活动都必须建立在前一项产生的信息基础上。

(2) 权属信息管理是对参与生物样本采集和

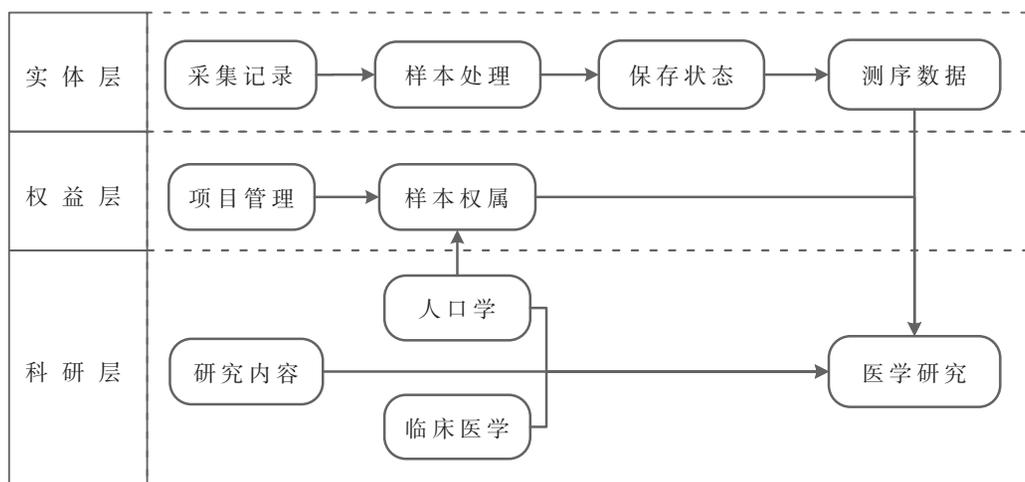


图2 生物样本生命周期中的信息流

保存活动中的捐献人、资助机构、保存机构、科研机构等主体权责的说明和相关文件的记录，主要强调各类主体权责范围的明确阐述和相关证明文件的完备性。权属层的各项活动是以严格保障各类主体的权益为目的，必须通过实体层和科研层的部分活动获取信息，并最终体现在医学研究的成果管理上。

(3) 科研信息管理包括研究方案、临床医学信息与人口学信息的保存和传递，并对研究方案的实现过程和相应数据分析的结果给予完整记录，主要强调多源信息集成和信息质量控制。科研层存在大量并行发生的信息活动，例如：“研究内容”虽然指导整个科研项目的进程，但并不直接决定其他信息活动的结果；“人口学信息”和“临床医学信息”的采集通常并行发生，最终共同服务于“医学研究”。

综合3个层面的信息流可以看到：(1)“医学研究”是生物样本相关所有信息活动的最终服务对象，是生物样本信息管理的核心需求；(2)实体层的信息活动相对独立，生物样本的实体信息很容易形成独立的数据集，但依靠实体层内部管理难以保障相关联信息的完整；(3)权属层的信息涉及科研机构和捐献人两方面，项目管理方需要对伦理审查、样本管理规范和相关信息安全机制提供保障；(4)科研层的信息活动是提供生物样本表型信息和环境信息的主要来源。这些信息是“样本权属”的部分信息来源，而与组学信息（如“测序数据”）之间的关联关系是开展“医学研究”的必要条件。

生物样本在其生命周期中经历众多主体的管理，在精准医疗背景下各信息活动相应的主体如图3所示。

由于“医学研究”在信息流中的核心地位，转化医学研究部门的需求成为生物样本信息价值主张的主要依据。对于转化医学研究部门，其获取生物样本信息的目的是开展高质量的研究并确保研究的可持续性。因此，其需求可以归纳为4个主要目标，即充分获取信息、保障信息质量、符合伦理法规、提升可拓展性。

对照生物样本生命周期各环节，其他参与主体对以上4个目标的实现如图4所示。

“充分获取信息”目标主要依靠组学实验室、医院信息中心和样本采集部门实现。组学实验室提供组学信息；医院信息中心提供表型信息；医院信息中心和样本采集部门共同提供环境信息。

“保障信息质量”目标的实现涉及生物样本库、生化实验室、组学实验室和样本采集部门，需要所有涉及样本实体活动的主体制定合理的操作规范并严格实行，同时详细记录生物样本的采集加工过程。

“符合伦理法规”目标主要依靠样本捐献者和项目管理部门实现。样本捐献者和项目管理部门共同提供知情同意信息；项目管理部门还需要预先了解项目开展地域的相关伦理和法律，并监督其实施过程的合法性。此外，项目管理部门还要明确项目实施中各参与主体的相关权益，避免

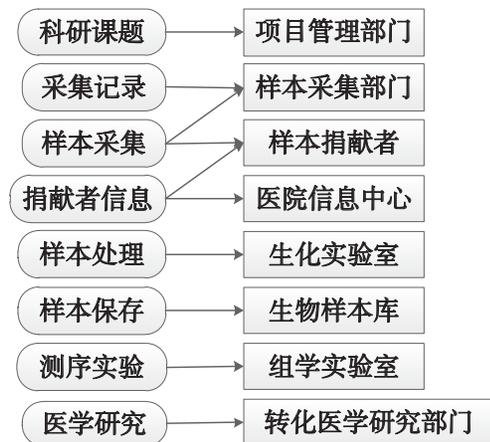


图3 生物样本信息生产加工各环节对应主体

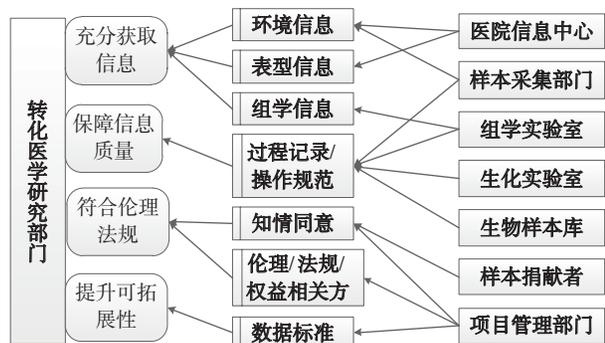


图4 参与主体对核心目标的实现

任何主体的权益受到侵害。

“提升可扩展性”目标主要依靠项目管理部门实现，项目管理部门为生物样本管理所制定数据标准体系的系统性和科学性决定其科研成果的可读性、可信度和应用范围等。

4 精准医疗运行的支撑

从精准医疗背景下生物样本信息管理的需求分析可以看到，参与生物样本生命周期的各类主体通过复杂的信息生产和传递路径为转化医学研究提供所需信息资源。生物样本信息在采集加工的过程中呈现出多种形态，为精准医疗运行提供重要支撑。

(1) 生物样本信息是精准医疗产业数据资源的核心集成体。如前所述，生物样本信息包括组学信息、表型信息、环境信息、采集加工过程记录、操作规范、知情同意书、伦理法规、利益相关方、数据标准等众多信息，是多维数据的集成体。不同于单纯的医学研究，在精准医疗产业化进程中，出于项目管理、资产管理和产品工程化等目的，这些信息已经成为支撑精准医疗产业发展的必要数据资源。因此，以信息集成体的形式系统地整合和管理生物样本信息是精准医疗产业应当广泛普及的基本认识。

(2) 生物样本信息是精准医疗产业深度信息化的数据资源保障。生物样本的采集加工过程记录、操作规范和数据标准是评估生物样本信息质量的主要依据，也是借助“互联网+”和大数据技术推动相关科研成果广泛应用的必要条件。精准医疗产业的深度信息化必须建立在完善的生物样本信息资源基础上，通过对生物样本信息的深入挖掘和增值利用实现。因此，完善相关行业规范、提升生物样本信息管理系统建设水平是精准医疗产业信息化的必要工作。

(3) 生物样本信息是转化医学研究过程的主要信息组织形式。即使暂不考虑知识产权、伦理法规等问题，生物样本信息中用于转化医学研究的组学信息、表型信息、环境信息、实验过程记录以及相应操作规范和数据标准等信息分别来自

不同参与主体，其技术领域和管理口径均有较大差异。为避免多源数据关联性的缺失或损坏，以生物样本为纽带组织相关信息已成为转化医学研究的普遍工作方式。因此，进一步深化生物样本信息管理的理论与实践研究将对转化医学发展将产生深远的影响。

(4) 生物样本信息是产业化需求与医学研究的数据接口。精准医疗产业是医药产业、临床医学与基础医学研究的结合，是诸多技术领域和学科方向的融合发展。生物样本信息作为贯穿其业务主线的信息资源，在多领域数据间建立起紧密的关联关系，将医学研究成果与其产品之间的联系清晰呈现。因此，规范化的生物样本信息将使产业化需求与医学研究的对应关系更加明朗，不仅为医学研究成果的快速产业化提供便利，也为精准医疗产业的发展方向提供参考和指导。

5 结语与建议

本文从精准医疗产业发展的信息化需求入手，厘清样本信息的概念范畴和价值定位。转化医学研究发展快速提升了组学数据的商业价值，而精准医疗的产业化进程则将生物样本信息的资源价值提升到新的水平。由于产业化和信息化带来对信息规模和信息质量的需求，生物样本信息从“遗传资源保藏”背景下的档案形态转化为精准医疗背景下的资源形态，成为精准医疗产业的重要资源和资产。相对地，面对产业化需求在可读性、可靠性、可扩展性的更高要求，现有管理体系下高水平生物样本信息的稀缺现状也被凸显出来。依据对生物样本信息的信息流和需求分析，提出如下建议。

(1) 研究并制订生物样本信息数据标准和技术标准，降低多维数据集成管理成本，提升数据可扩展性。虽然生物样本信息管理可以实现多维数据集成、组织多源数据并管控数据质量，但以项目或机构为主导的信息管理难以形成稳定的管理机制，并且数据标准和管理信息系统的重复建设，跨机构、跨学科数据标准的兼容问题等都会增加生物样本信息的管理成本。而在行业层面研

究并制订相关数据标准和技术标准不仅能有效降低信息管理成本,同时对提升生物样本信息可读性、可靠性和可拓展性具有良好的促进作用。

(2) 建立开放式的生物样本信息管理云平台及其数据安全保障机制,快速提升产业信息化水平。生物样本的采集加工需要多主体的共同参与,而生物样本管理信息系统的重复建设不仅为相关研究带来巨大的经费压力,摊薄的建设经费也限制了各主体的管理信息化水平。因此,建立开放式的生物样本信息管理云平台,为医院、企业、研究所提供便捷和相对廉价的信息管理服务,是快速提升产业信息化水平的有效方式。但由于医疗信息的敏感性,其云平台必须具有完善的数据安全保障机制,并兼顾安全性和数据有效性。

(3) 深化生物样本信息管理研究,为专业化管理提供理论支撑。生物样本信息管理正在进入一个新的阶段,相比以“遗传资源保藏”为目标的时代,其面向的用户群体、业务架构和制度环境等均发生了较大变化。同时,精准医疗的产业化方兴未艾,新的变化还在不断发生。目前,国内外生物样本库、生物银行、相关研究机构等对生物样本信息的管理仍处于探索阶段,对生物样本信息的认识仍不全面。

(4) 为适应新的用户需求和制度环境,有必要不断推进生物样本信息管理研究,提出支撑专业化管理的系统性管理理论。

参考文献

- [1] COLLINS F S, VARMUS H. A new initiative on precision medicine[J]. *New England Journal of Medicine*, 2015, 372(9): 793.
- [2] 中华人民共和国国务院. 国务院关于印发“十三五”国家科技创新规划的通知[EB/OL]. [2017-07-03]. http://www.most.gov.cn/mostinfo/xinxifenlei/gjkjgh/201608/t20160810_127174.htm.
- [3] 张雷, 李海燕, 范可方, 等. 生物样本库与转化医学研究[J]. *转化医学研究(电子版)*, 2011, 1(2): 44-55.
- [4] LEFF D R, YANG G Z. Big data for precision medicine[J]. *Engineering*, 2015, 1(3): 277-279.
- [5] 中国医药生物技术协会生物样本库标准(试行)
- [6] 张佳星. 推进精准医疗发展助力健康中国建设: 访中国工程院院士、中国医学科学院院长曹雪涛委员[N]. *科技日报*, 2015-03-10(2).
- [7] 王晓民, 张雷. 生物信息资源库建设的现状与未来[J]. *中华肾病研究电子杂志*, 2012, 1(1): 16-21.
- [8] GLIGORIJEVIĆ V, MALODDOGNIN N, PRŽULJ N. Integrative methods for analyzing big data in precision medicine[J]. *Proteomics*, 2016, 16(5): 741.
- [9] 吴曼丽, 王柱清, 罗凯, 等. 临床生物样本库血液样本的可控性采集和质量管理[J]. *转化医学杂志*, 2016, 5(6): 327-332.
- [10] 刘克新, 郑琳, 刘晶, 等. 生物样本库临床信息数据采集系统的需求与设计[J]. *中国病案*, 2015(8): 31-33.
- [11] 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[J]. *中华医院管理杂志*, 2013, 29(11): 863-865.
- [12] 徐世侠, 张在文, 冯博, 等. 我国生物样本库的文献计量分析[J]. *转化医学杂志*, 2014, 3(6): 359-363.
- [13] CHAUSSABEL D, PULENDRAN B. A vision and a prescription for big data-enabled medicine[J]. *Nature Immunology*, 2015, 16(5): 435.
- [14] 李怡, 张焕敬, 刘国彦, 等. 生物样本信息资源库的国内外研究进展[J]. *中国医药生物技术*, 2012, 7(5): 369-372.
- [15] GAINER V S, CAGAN A, CASTRO V M, et al. The biobank portal for partners personalized medicine: a query tool for working with consented biobank samples, genotypes, and phenotypes using i2b2[J]. *Journal of Personalized Medicine*, 2016, 6(1): ???-???
- [16] KARLSON E W, BOUTIN N T, HOFFNAGLE A G, et al. Building the partners healthcare biobank at partners personalized medicine: informed consent, return of research results, recruitment lessons and operational considerations[J]. *Journal of Personalized Medicine*, 2016, 6(1): 2.
- [17] OOFEE COOPERATION DEVELOPMENT. OECD guidelines on human biobanks and genetic research databases[J]. *European Journal of Health Law*, 2010, 17(2): 191.
- [18] 中华人民共和国科学技术部. 人类遗传资源管理暂行办法[EB/OL]. [2017-07-03]. http://www.most.gov.cn/bszn/new/rlyc/wjxz/200512/t20051226_55327.htm.
- [19] 中国医药生物技术协会. 中国医药生物技术协会生物样本库标准(试行)[J]. *中国医药生物技术*, 2011(1): 71-79.