

# 宏基因组二代测序技术临床应用中的 数据合规性研究与实践

郑国栋<sup>1</sup> 苏磊<sup>1,2</sup>

(1. 南方医科大学第一临床医学院, 广东广州 510630;

2. 中国人民解放军南部战区总医院, 广东广州 510010)

**摘要:** 数据已经成为重要的生产要素, 宏基因组测序标本采集涉及面广, 宏基因组二代测序技术数据是医疗健康领域重要的基础性战略资源。结合《网络安全法》和《个人信息安全规范》《数据安全法》《个人信息保护法(草案)》等相关规定, 提出宏基因组二代测序数据合规性框架和数据合规要点, 为医疗机构在利用宏基因组二代测序数据开展医疗诊治活动的同时, 切实保证测序机构不出现数据合规风险或法律纠纷, 并建议强化内部管理, 兼顾社会效益, 承担相应的社会责任, 积极开展数据合规培训和教育。

**关键词:** 宏基因组二代测序; 数据合规性框架; 数据合规风险

**DOI:** 10.3772/j.issn.1674-1544.2022.04.007

**CSTR:** 15994.14.issn.1674.1544.2022.04.007

中图分类号: TP309.2

文献标识码: A

## Research and Practice of data Compliance in Clinical Applications of Metagenomic Second-generation Sequencing Technology

ZHENG Guodong<sup>1</sup>, SU Lei<sup>1,2</sup>

(1. The First School of Clinical Medicine, Southern Medical University, Guangzhou 510630; 2. General Hospital of Southern War Zone, Chinese People's Liberation Army, Guangzhou 510010)

**Abstract:** Data has become an important factor of production. The collection of metagenome sequencing specimens involves a wide range of specimen. Metagenome second-generation sequencing technology data is an important basic strategic resource in the field of medical and health. Combined with the relevant provisions of the network security law, personal information security code, data security law and personal information protection law (Draft), this paper tentatively puts forward the compliance framework and key points of metagenome second-generation sequencing data, so as to provide a reference for medical institutions to carry out medical diagnosis and treatment activities while using metagenome second-generation sequencing data. It is suggested to strengthen internal management, give consideration to social benefits, assume corresponding social responsibilities, and actively carry out data compliance training and education.

**Keywords:** metagenome second generation sequencing, data compliance framework, data compliance risk

**作者简介:** 郑国栋 (1979—), 男, 南方医科大学第一临床医学院副主任医师, 研究方向为感染学; 苏磊 (1960—), 中国人民解放军南部战区总医院主任医师, 研究方向脓毒症和中暑 (通信作者)。

**收稿时间:** 2022年2月25日。

## 0 引言

数据已经成为重要的生产要素，在数据的收集、使用和流动过程中面临着数据合规挑战。基因测序机构在利用数据实现经济效益的同时，应承担相应的社会责任，积极开展数据合规。

对病原体快速准确的诊断是感染性疾病精准治疗的关键。宏基因组二代测序 (Metagenomic next generation sequencing, mNGS) 技术适用于各种病原体的检测，其具有无需培养、检测迅速便捷、准确性高以及标本来源广等优势。与一般个人信息相比，医疗数据一旦泄露就会造成更加严重的负面影响。合规利用宏基因组二代测序数据不仅可以促进医疗服务模式的改进，而且对整个社会的经济发展都有着重要的作用，是国家重要的基础性战略资源。但是，目前还存在以下问题：临床对 NGS 应用的适应人群认识不足，送检标本存在较大的盲目性和随意性；操作过程复杂，缺乏数据交互规范；提供检验的主体多元，检测质量良莠不齐，缺乏对检验主体的合规性审查；结果解释缺乏可信的标准，误导临床的案例屡见不鲜，限制了该技术的临床应用和普及。鉴于此，本文侧重讨论宏基因组二代测序数据合规的内涵和要点，以期对宏基因组二代测序技术全流程的合规框架和实践有更全面的认识，为提升临床抗感染诊治能力提供帮助。

## 1 数据合规相关研究进展

### 1.1 关于医疗数据合规政策

2018年5月25日，欧盟《一般数据保护条例》(GDPR) 正式实施<sup>[1]</sup>。GDPR 法案要求不论数据控制者、处理者及其处理行为在欧盟境内还是境外，只要处理的是欧盟境内居民的数据，均适用此法案，对数据实施长臂管理。目前，全球已有近 100 个国家和地区制定了数据安全保护的法律，数据安全保护专项立法已成为国际惯例。

为了规范数据的生成、采集、存储、加工、分析、服务等处理，我国出台了多项法律法规及政策文件。2021年6月10日，《中华人民共和

国数据安全法》(以下简称“《数据安全法》”) 正式发布(2021年9月1日实施)<sup>[2]</sup>；2021年8月20日，《中华人民共和国个人信息保护法》(以下简称“《个人信息保护法》”) 正式发布(2021年11月1日实施)<sup>[3]</sup>。这两部法律同《中华人民共和国网络安全法》一起共同构建了我国的数据治理立法框架，共同维护网络安全和数据安全。

2020年12月14日，国家标准化管理委员会(简称“国标委”)发布了GB/T 39725-2020《信息安全技术 健康医疗数据安全指南》(简称“《指南》”) <sup>[4]</sup>，并于2021年7月1日正式实施。《指南》由全国信息安全标准化技术委员会(简称“安标委”)信安标委提出并归口，健康医疗数据生命周期内可能涉及各主体都参与了起草工作。其中包括北京协和医院、上海市儿童医院等各城市三甲医院，东软集团、零氦科技等第三方数据服务商，各保险公司、大学以及与网络安全保护相关的中国信息安全测评中心等。此前，国标委于2018年12月发布了《指南》的征求意见稿。作为推荐性国家标准，《指南》并不具有强制法律效力。各企业在合规过程中，可将《指南》作为参考，结合《网络安全法》《个人信息安全规范》《国家健康医疗大数据标准、安全和服务管理办法(试行)》及《人口健康信息管理办法(试行)》等法规及标准以保证全面性。

### 1.2 关于宏基因组测序技术

二代测序的病原学诊断技术(NGS)，又称为高通量测序技术。该技术首先对采集样本进行核酸提取，然后将测序后的基因片段与系统中的病原体进行对比，从而确定样本所含病原体。该技术测序流程主要包括标本核酸提取建库、DNA和/或RNA富集、上机测序、生物学信息分析等。NGS大体上分为宏基因组测序(metagenomic next generation sequencing, mNGS)和靶向扩增子测序(targeted amplicon sequencing, TAS)<sup>[5]</sup>。mNGS检测迅速、快捷，准确性高，适用于细菌、真菌、病毒等各种病原微生物的检测<sup>[6]</sup>，并可发现新的未知病原体<sup>[6]</sup>，因此在临床传染性疾病的诊疗过程中迅速得到认可，被

广泛应用于临床及科研领域<sup>[7]</sup>。

宏基因组测序标本采集范围包括外周血标本、骨髓标本、肺泡灌洗液标本、痰液标本、尿液标本、大便标本、脑脊液标本、导管尖标本、穿刺液标本等。2018年9月发布的《国家健康医疗大数据标准、安全和服务管理办法（试行）》（简称“《服务管理办法》”）第四条首次明确提出“健康医疗大数据，是指在人们疾病防治、健康管理等过程中产生的与健康医疗相关的数据”。这个概念几乎全面涵盖了人们进行健康活动所涉及到的数据。

在宏基因组测序医疗活动中，可以更详细地分为个人属性数据、健康状况数据、医疗应用数据和卫生资源数据等。①个人属性数据包括：人口统计信息，主要有姓名、年龄、性别、民族、国籍、职业、住址、工作单位、家庭成员信息、联系人信息、收入等；个人身份信息，主要有姓名、身份证、工作证、居住证、社保卡、可识别个人的影像图像、健康卡号、住院号、各类检查检验相关单号等；个人通讯信息，主要有个人电话号码、邮箱、账号及关联信息等；个人生物识别信息，主要有基因、指纹、声纹、掌纹、耳廓、虹膜、面部特征等；个人健康监测传感设备ID等。②健康状况数据包括主诉、现病史、既往病史、体格检查（体征）、家族史、症状、健康体检数据、遗传咨询数据、可穿戴设备采集的健康相关信息、生活方式等。③医疗应用数据包括知情告知信息、基因测序、人体微生物检测。④卫生资源数据包括医院基本数据、医院运营数据、医院公卫数据等。

为提高 NGS 技术应用于感染性疾病病原诊断的质量，有效实施质量管理，中华医学会检验医学分会临床微生物学组、中华医学会微生物学与免疫学分会临床微生物学组、中国医疗保健国际交流促进会临床微生物与感染分会组织专家编写制定《宏基因组高通量测序技术应用于感染性疾病病原检测中国专家共识》，促进了该技术规范和发展。

## 2 数据全流程合规框架与数据合规要点

### 2.1 合规框架

宏基因组二代测序业务流程包括三个阶段、八个环节，具体如表 1 所示。数据全流程合规框架围绕宏基因组二代测序业务流程，梳理出包括内外两部分的 6 个合规点，外部主要针对第三方开展合规性审查，内部则主要包括采集、存储、连通、报告和应用 5 个合规点。具体如图 1 所示。

### 2.2 合规要点

#### 2.2.1 数据的采集

宏基因组二代测序数据的采集是指以直接或间接方式获取、存储健康医疗数据的行为。

依据《网络安全法》第四十一条规定，网络运营者收集、使用个人信息应当遵循合法、正当、必要的原则，公开收集、使用规则，明示收集、使用信息的目的、方式和范围，并经被收集者同意；使用网站、应用程序收集时需要制定隐私政策；在间接收集时，应要求个人信息提供方说明个人信息来源，并对其合法性进行确认，同时了解已获取的个人信息处理的授权同意范围。

表 1 宏基因组二代测序业务流程

序号	阶段	步骤
1	准备阶段	提取病原核酸
2		构建测序文库
3	测序阶段	文库加载到测序芯片
4		测序仪测序
5		边合成边测序
6	分析阶段	生成碱基序列
7		数据分析与数据库比对
8		生成电子报告

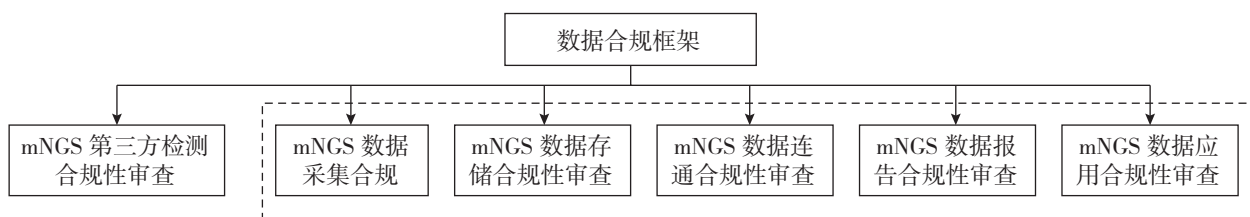


图1 宏基因组二代测序数据合规性框架

2014年5月5日发布的《人口健康信息管理办法(试行)》第六条、第八条规定,采集、利用、管理人口健康信息应当按照法律法规的规定,遵循医学伦理原则,保证信息安全,保护个人隐私。故此,采集信息时应当按照“一数一源、最少够用”的原则,避免重复采集、多头采集。依据前述规定,采集宏基因组二代测序相关数据时,医院必须遵循“个人同意”和“一数一源、最少够用”。

因此,医院在获取数据时必须全方位取得相关单位和个人的有效授权,同时做好数据采集的规范管理。

### 2.2.2 数据的存储

宏基因组二代测序数据存储单位应当采取必要的安全保护措施妥善保管自己掌握的健康医疗数据,避免健康医疗数据泄露、毁损、丢失或者被篡改,也不得非法向他人提供健康医疗数据。

#### (1) 分等级存储

医院在对医生提供患者诊疗数据前应当对数据进行安全影响评估,并对受让方的数据安全保护能力进行评估,再根据评估结果采取相应的有效保密措施,确保数据安全。2011年11月29日发布的《卫生行业信息安全等级保护工作的指导意见》第四条“工作任务”规定,国家信息安全等级保护制度将信息安全保护等级分为五级:第一级为自主保护级,第二级为指导保护级,第三级为监督保护级,第四级为强制保护级,第五级为专控保护级。重要卫生信息系统安全保护等级原则上不低于第三级,如卫生统计网络直报系统、传染性疾病报告系统、卫生监督信息报告系统、突发公共卫生事件应急指挥信息系统等跨省全国联网运行的信息系统;国家、省、地市三级卫生信息平台,新农合、卫生监督、妇幼保健等

国家级数据中心;三级甲等医院的核心业务信息系统;卫生部网站系统;其他经过信息安全技术专家委员会评定为第三级以上(含第三级)的信息系统。

《人口健康信息管理办法(试行)》第十一条规定,委托其他机构存储、运维人口健康信息的,委托单位承担人口健康信息的管理和安全管理责任。受委托的存储、运维机构应当严格按照委托协议做好人口健康信息管理的技术支持,禁止超权限采集、开发和利用人口健康信息。

因此,医院存储宏基因组二代测序数据时首先需要依据数据隐私和重要程度进行等级划分,分别设置相应的保密措施。

#### (2) 设定访问权限

《服务管理办法》第二十二条规定,责任单位应当按照《中华人民共和国网络安全法》的要求,严格规范不同等级用户的数据接入和使用权限,并确保数据在授权范围内使用。第二十三条规定,责任单位应当要建立严格的电子实名认证和数据访问控制,规范数据接入、使用和销毁过程的痕迹管理。

2013年11月20日发布的《医疗机构病历管理规定》第六条要求医疗机构及其医务人员应当严格保护患者隐私,禁止以非医疗、教学、研究目的泄露患者的病历资料。电子病历系统应当对操作人员的权限实行分级管理,用户根据权限访问相应保密等级的电子病历资料。授权用户访问电子病历时,自动隐藏保密等级高于用户权限的电子病历资料。

因此,医院采集整理的宏基因组二代测序数据分类数据信息存储时必须设置控制系统,如访问权限的设置,以保证医院存储数据的安全。

#### (3) 补救及报告义务

《网络安全法》第四十二条第二款规定，网络运营者应当采取技术措施和其他必要措施，确保其收集的个人信息安全，防止信息泄露、毁损、丢失。在发生或者可能发生个人信息泄露、毁损、丢失的情况时，应当立即采取补救措施，按照规定及时告知用户并向有关主管部门报告。2018年7月17日发布的《互联网诊疗管理办法（试行）》第二十条与《互联网医院管理办法（试行）》第二十三条均规定，发生患者信息和医疗数据泄露后，医疗机构应当及时向主管的卫生健康行政部门报告，并立即采取有效应对措施。

因此，对于医生调取的患者诊疗数据信息，医院有义务监督其保管使用情况，当发现患者信息泄露时，医院应采取补救措施，及时告知患者和向主管的卫生健康行政部门报告。

#### 2.2.3 数据的利用

宏基因组二代测序数据的利用，是指相关数据的使用、分析、共享、交易等行为。

(1) 使用。医院或者其他单位使用采集的相关数据时，首先要对数据进行脱敏或匿名化。数据脱敏是指采用数据清洗等脱敏技术手段对采集的健康医疗数据进行技术处理以后形成的，不能单独或者通过与其他数据结合识别某一特定自然人的数据。对数据进行脱敏处理后，仍需要采取措施对医疗数据访问设置权限。

(2) 共享。医院等医疗机构为实现跨机构、跨地域的诊疗信息交互、共享和医疗服务协同，需要在各医疗机构、医联体信息平台之间实现数据（电子病历、电子健康档案等）的互联互通与信息共享。在此场景下，医院的医护人员、卫生机构管理人员、医院间联合体及医疗第三方服务机构人员在对相关系统文件、数据库资料等敏感数据进行访问浏览，以及通过内部信息共享交换系统进行文件数据传输、存储等操作时，均可能导致医患隐私等重要信息面临泄露风险。因此，在共享时若只对个人信息进行去标识化（通过对个人信息的技术处理，使其在不借助额外信息的情况下，无法识别个人信息主体的过程）处理，仍需经过患者的明示同意之后方可提供。

同时，对健康医疗数据使用的具体操作人员实行权限管理，对健康医疗数据进行分级；根据健康医疗数据的保密级别授予操作人员相关权限，对登录用户进行身份鉴别，并对登录用户的操作行为进行审计；妥善保管记录健康医疗数据的纸介质、光介质、电磁介质等载体，并采取相应的安全储存措施；对健康医疗数据进行操作的人员、时间、地点、事项等信息，应当留下完整的电子日志加以记录。

#### 2.2.4 第三方测序机构合规问题清单

宏基因组二代测序要求说明：①测序机构技术、业务及产品（或服务）中涉及到数据采集、清洗、管理、运用的具体环节；不同环节涉及的数据的具体类型，文字、图像、视频等具体情况；②测序机构自身核心技术（如算法的训练、系统的搭建等）是否涉及大量的数据的应用，如是，相关数据的来源及其合规性；③测序机构对外提供的产品（或服务）是否涉及数据的采集运用，如是则说明数据的来源及其合法合规性；④测序机构保证数据采集、清洗、管理、运用等各方面的合规措施；⑤测序机构的数据来源中是否包含向供应商采购，如是则说明是否相关合同中约定数据合规的条款或措施，并结合《民法典》《网络安全法》和《个人信息安全规范》《数据安全法（草案）》《个人信息保护法（草案）》等相关规定，说明相关措施是否能切实保证测序机构不出现数据合规风险或法律纠纷；⑥结合测序机构的产品交付及部署模式，说明测序机构的产品（或服务）中涉及用户的个人数据的情形和场景，以及运用、管理及其合规性；⑦测序机构产品至今是否面临数据合规方面的诉讼或纠纷，并结合相关公开报道，说明测序机构数据的合规性。

### 3 结语

数据已经成为重要的生产要素，宏基因组测序标本采集范围包括外周血标本、骨髓标本、肺泡灌洗液标本、痰液标本、尿液标本、大便标本、脑脊液标本、导管尖标本、穿刺液标本等，  
(下转第82页)

- [22] 章鹰.5G+AIoT时代,三种创新芯片和新型存储器发展方向[EB/OL].[2020-12-17].<http://m.elecfans.com/article/1423074.html>.
- [23] 下一代超低功耗磁随机存储器(MRAM)技术现状[EB/OL].[2021-01-21].<https://news.skynix.com.cn/future-semiconductor-technology-the-present-of-the-next-generation-ultra-low-power-mram-technology/>.
- [24] 清华教授研发忆阻器芯片,可实现存储数据“原地”计算[EB/OL].[2021-01-29].<https://www.nsf.gov.cn/csc/20340/20289/57063/index.htm>.

(上接第55页)

宏基因组二代测序技术数据是医疗健康领域重要的基础性战略资源,因此开展合规性工作,就显得尤为重要。

对于从事宏基因组二代测序的相关厂商和医疗机构,急需了解并掌握《民法典》《网络安全法》和《个人信息安全规范》《数据安全法(草案)》《个人信息保护法(草案)》等相关规定,并结合相关法律法规,制定相关合规性制度文件,切实保证测序机构不出现数据合规风险或法律纠纷。

本文探索性地提出了宏基因组二代测序数据合规性框架和数据合规要点,为医疗机构,在利用宏基因组二代测序数据开展医疗诊治活动的同时,为强化内部内部管理、兼顾社会效益、承担相应的社会责任、积极开展数据合规培训和教育提供了工作指南。

#### 参考文献

[1] 张彭.大数据安全背景下欧盟《通用数据保护条例

(GDPR)》研究[D].上海:华东师范大学,2020.

- [2] 中华人民共和国数据安全法[EB/OL]. [2021-06-10]. <http://www.npc.gov.cn/npc/c30834/202106/7c9af12f51334a73b56d7938f99a788a.shtml>.
- [3] 中华人民共和国个人信息保护法[EB/OL].[2021-08-20]. <http://www.npc.gov.cn/npc/grxxbhfca/grxxbhfca.shtml>.
- [4] 信息安全技术 健康医疗数据安全指南[EB/OL]. [2020-12-14]. <http://www.phic.org.cn/zcyjybzpj/bzypj/bzgf/gjbz/202103/P020210331605989883649.pdf>.
- [5] 何承成,何连福,严桥路. A review of the value of next-generation sequencing in the diagnosis of lower respiratory tract Infection[J]. 临床医学进展, 2021, 11(6): 2521-2525.
- [6] ZHOU P, YANG X L, WANG X G, et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin[J]. Nature, 2020, 579(7798): 270-273.
- [7] LYNCH T, PETKAU A, KNOX N, et al. A Primer on infectious disease bacterial genomics[J]. Clin Microbiol Rev, 2016, 29(4): 881-913.