



中华人民共和国国家标准

GB/T XXXXX—XXXX

医用高通量基因测序数据共享要求

Medical high throughput gene sequencing data sharing requirements

征求意见稿

本稿完成时间：2024 年 7 月 30 日

在提交反馈意见时，请将您知道的相关专利连同支持性文件一并附上。

XXXX – XX – XX 发布

XXXX – XX – XX 实施

国家市场监督管理总局
国家标准化管理委员会 发布

目 次

前 言 II

1 范围 1

2 规范性引用文件 1

3 术语和定义 1

4 基本原则 3

5 共享内容 3

6 共享场景和方式 4

7 共享管理 6

参 考 文 献 8

前 言

本文件按照GB/T 1.1—2020《标准化工作导则 第1部分：标准化文件的结构和起草规则》的规定起草。

请注意本文件的某些内容可能涉及专利。本文件的发布机构不承担识别专利的责任。

本文件由国家食品药品监督管理局提出并归口。

本文件起草单位：

本文件主要起草人：

医用高通量基因测序数据共享要求

1 范围

本文件规定了医用高通量基因测序数据共享的基本原则、共享内容、共享场景和方式、共享管理。

本文件适用于医用高通量基因测序数据相关方之间进行基因数据共享活动的管理，也适用于监管部门、第三方评估机构对医用高通量基因测序数据共享活动进行的监督、管理、评估。

本文件不适用于涉及国家秘密的医用高通量基因测序数据共享活动。

2 规范性引用文件

下列文件中的内容通过文中的规范性引用而构成本文件必不可少的条款。其中，注日期的引用文件，仅该日期对应的版本适用于本文件；不注日期的引用文件，其最新版本（包括所有的修改单）适用于本文件。

- GB/T 25069 信息安全技术 术语
- GB/T 30534 科技报告保密等级代码与标识
- GB/T 35273 信息安全技术 个人信息安全规范
- GB/T 37964 信息安全技术 个人信息去标识化指南
- GB/T 37988 信息安全技术 数据安全能力成熟度模型
- GB/T 39335 信息安全技术 个人信息安全影响评估指南
- GB/T 39725 信息安全技术 健康医疗数据安全指南
- GB/T 41806 信息安全技术 基因识别数据安全要求
- GB/T 43697 数据安全技术 数据分类分级规则
- GB/T XXXXX 数据安全技术 数据安全保护要求

3 术语和定义

GB/T 25069、GB/T 35273、GB/T 37964、GB/T 37988和GB/T 41806界定的以及下列术语和定义适用于本文件。

3.1

医用高通量基因测序数据 clinical high throughput gene sequencing data
医学用途的高通量测序产生的基因数据及元数据。

3.2

医用高通量基因测序数据资源 clinical high throughput gene sequencing data resource
在医用场景下，医用高通量基因测序数据生产、共享和处理过程中产生的所有数据及元数据。

3.3

主数据 master data

关于业务实体的数据。

3.4

元数据 metadata

定义和描述其他数据的数据。

[来源：GB/T 18391.1—2009，3.2.16]

3.5

统一资源标识符 uniform resource identifier;URI

包含了名称或地址,指向Web上某一对象的短数据串。

[来源：GB/T 25069—2022,3.602]

3.6

相关方 interested party

为完成某个目的而进行的行为过程中的参与者。不论是主动参与还是被动参与，皆为相关方。

3.7

知情同意 informed consent

医用高通量基因测序数据处理者和数据主体或数据主体的法定监护人之间关于医用高通量基因测序数据处理边界的约定。

3.8

数据处理 data processing

对数据进行存储、加工、检索、统计和维护等操作的过程。

3.9

完全公开共享 completely public sharing

数据一旦发布，很难召回，一般通过互联网直接公开发布。

[来源：GB/T 37964—2019，3.12]

3.10

受控公开共享 controlled public sharing

通过数据使用协议对数据的使用进行约束。

[来源：GB/T 37964—2019，3.13]

3.11

领地公开共享 enclave public sharing

在物理或者虚拟的领地范围内共享，数据不能流出到领地范围外。

[来源：GB/T 37964—2019，3.14]

3.12

数据安全 data security

保护在存储、通信及处理过程中的数据不被破坏、泄露、篡改或非授权使用而采取的措施及技术。

3.13

风险分析 risk analysis

在风险识别的基础上，通过对所收集的大量的详细资料加以分析，评估和预测风险发生的概率和损失程度的过程。

3.14

数据管理委员会 committee of data management

为确保机构业务运转的连续性，保障基因数据及关联信息的保密性、完整性、可用性，在基因数据处理者内部设立的专业机构。

注：数据管理委员会负责制定各项数据管理制度，实施数据安全技术防护，开展业务活动过程中的数据安全风险评估和合规审计，制定数据安全事件应急预案，及时处置安全事件，组织数据安全教育及培训等，确保开展业务活动过程中的数据安全、个人信息隐私得到保护。

3.15

数据共享过程审计 process audit of data sharing

对机构数据共享行为的效率、效果、执行情况进行监督、评价和鉴证的行为。

4 基本原则

医用高通量基因测序数据共享应遵循以下原则：

- a) 合法、正当、必要的原则，符合人类遗传资源管理和数据安全等相关规定；
- b) 尊重个体和群体，符合医学伦理，保护个人隐私信息，保障信息主体所享有的法定权利，不与信息主体的基本权利和自由相冲突；
- c) 公开开放，推动医学发展，数据共享的相关方形成共识，共同组成“技术安全、评估合规、过程可控”的数据共享管理机制。

5 共享内容

5.1 概述

医用高通量基因测序数据资源见图 1，包括医用高通量基因测序产生的核酸序列数据（主数据）及相关元数据、模型数据。

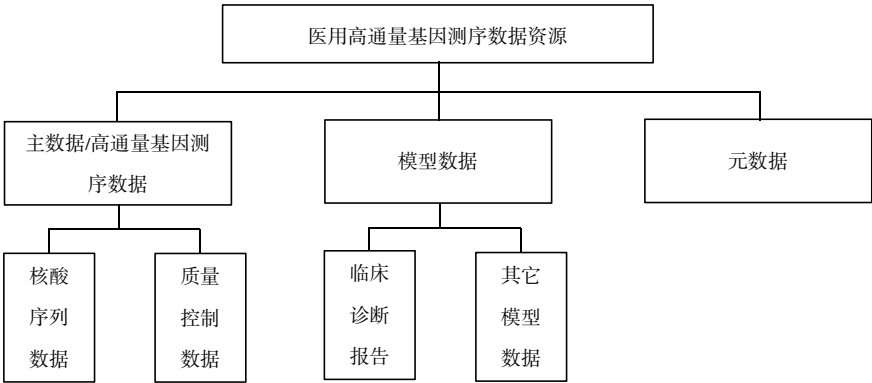


图 1 医用高通量测序基因测序数据资源

5.2 主数据

医用高通量基因测序数据资源的主数据是医用高通量基因测序产生的数据，包括核酸数据和对应的质量数据。

5.3 模型数据

模型数据是基于医用高通量基因测序数据处理后的数据，包括但不限于临床诊断报告、分析过程数据。

5.4 元数据

医用高通量基因测序元数据是描述主数据和模型数据的数据，包括但不限于知情同意书、实验过程、样本信息、样本对应患者、数据处理方法、数据处理相关标准等其他信息。

6 共享场景和方式

6.1 共享场景

6.1.1 在临床服务场景中，医用高通量基因测序数据首先用于对于患者治疗和康复过程有利的临床目的的达成。该场景下，利益相关方包括但不限于数据主体（患者）、医疗机构、检验机构、研究机构、相关企业和监管部门。对于数据共享过程中发生的委托服务过程，应在知情同意和协议约定的情况下符合统一的共享要求。

6.1.2 医用高通量基因测序数据除应用于临床服务外，还将应用于学术研究场景、医疗创新相关的转化场景和药械监管相关的监管场景。

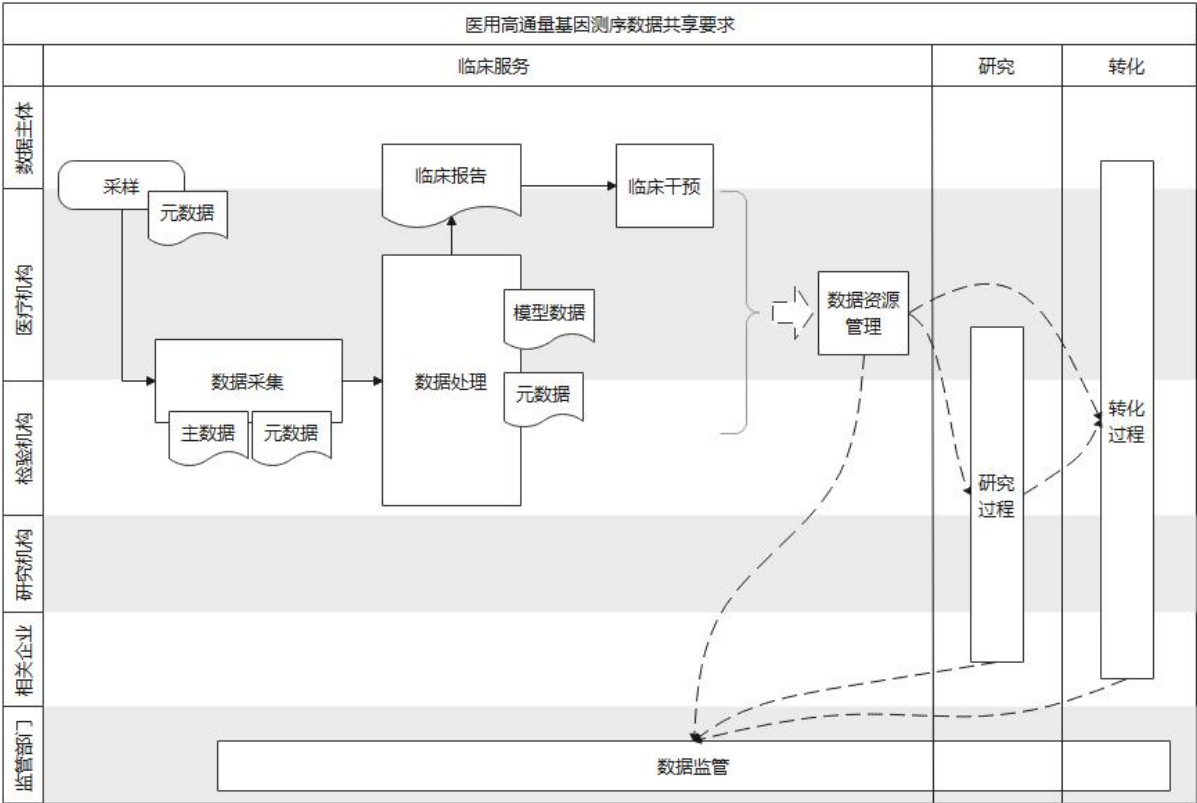


图 2 医用高通量基因测序数据共享场景

6.2 共享方式

6.2.1 数据公开共享类型可划分为完全公开共享、受控公开共享、领地公开共享，不同公开共享类型的数据去标识化应按照 GB/T 37964 和 GB/T 39725 的相关要求执行。

注：常见的数据公开共享形式包括接口共享和批量共享。

6.2.2 数据公开共享形式包括网站公开、文件共享、API 接入、在线查询、数据分析平台和安全计算环境下的共享。

6.2.3 在临床服务过程中，应在数据主体、医疗机构和检验机构三方共享数据。如需在临床服务之外进行数据共享活动，应通过医疗机构专门的数据资源管理部门或政府授权的国家数据设施实施。

6.2.4 本文件鼓励医用高通量基因测序数据用于科学研究。在科学研究情况下，医用高通量基因测序数据集给予统一资源标识符，其共享活动应遵循科学数据共享的有关规定。如表 1 所示。

6.2.5 在药械监管场景下，对于临床参比基因组标准数据集的使用，应在可控、可监管的安全计算环境下进行。

表 1 数据共享方式

数据类型	数据共享场景	数据公开共享类型	数据公开共享形式
个人基因组数据	临床服务	受控公开共享	API接口或文件共享
	科学研究	完全公开共享或领地公开共享	网站共享或在线查询
临床参比基因组标准数据集	药械监管	领地公开共享	安全计算环境

7 共享管理

7.1 组织

医用高通量基因测序数据共享的相关方,除数据主体外,其他机构的组织设计应符合GB/T 41806的第五章要求,包括设立伦理委员会及数据管理委员会,制定个人信息保护制度,建立数据安全管理体系,形成完整的信息安全管理体系,并应针对医用高通量基因测序数据进行数据安全风险分析和评估、数据安全风险监测预警、数据访问控制,以保障医用高通量基因测序数据全生命周期的安全性及数据主体的合法权益。

7.2 人员

应对医用高通量基因测序数据共享相关人员进行数据安全相关的培训,提高相关人员的从业能力。

7.3 过程

7.3.1 计划

7.3.1.1 应确定数据的保密等级:医用高通量基因测序数据保密等级应参照GB/T 30534分为4级,即“公开级”、“限制级”、“秘密级”、“机密级”。秘密级、机密级数据密级的确定和变更按国家有关定密规定确定;限制级数据由数据控制者确定和变更。

7.3.1.2 应开展数据分级分类:医用高通量基因测序数据应按照GB/T 43697进行分级分类管理,识别重要数据。重要数据管理应符合GB/T XXXXX—XXXX《数据安全技术 数据安全保护要求》。

7.3.1.3 建立数据共享的相关文件:项目共享方案、安全保障机制、应急预案等。

7.3.1.4 项目共享方案包括但不限于:共享数据类型、共享对象、共享方式、共享时效。

7.3.1.5 应急预案包括但不限于:启动应急预案的条件、应急处理流程、系统恢复流程、事件报告流程、事后教育和培训。

7.3.1.6 应明确每个项目的数据共享边界、实践周期、预期成果、责任人等。

7.3.1.7 应制定识别和管理利益冲突的程序,保证数据共享过程的公正性、诚信性。

7.3.2 实施

7.3.2.1 应建立数据共享管理和技术保障双重机制。针对数据生命周期各阶段,建立数据共享安全保障机制,以实现有效的、负责任的数据共享。

注:数据生命周期各阶段数据生命周期一般包括数据采集、数据处理、数据共享、数据销毁等数据管理阶段。

7.3.2.2 应采用电子或纸质方式记录数据共享过程的关键控制节点,并通过机制或技术手段确保记录的客观性,明确数据流向、责任承担方以及数据使用边界。

7.3.3 检查

7.3.3.1 应定期检查数据共享情况,必要时可以委托第三方机构进行数据共享过程审计。

7.3.3.2 当数据共享过程中出现偏离计划或安全风险增加时,应及时采取安全保障措施和纠正措施,避免造成更大的危害。

7.3.4 改进

应定期对数据共享管理和技术保障机制以及数据共享安全保障机制进行管理评审，以提高管理和保障的有效性、及时性和安全性。

7.4 安全与应急

7.4.1 医用高通量基因测序数据分类分级应符合 GB/T 43697 的要求。

7.4.2 应参考 GB/T 39335 开展个人信息安全影响评估。

7.4.3 模型数据的管理应符合知情同意书的约定要求，开展重要数据识别工作，重要数据的共享应符合相关法定要求。

7.4.4 医用高通量基因测序数据属于个人敏感信息，数据的共享安全应符合 GB/T 35273 的要求。

7.4.5 在开展医用高通量基因测序数据共享活动前，应按照 GB/T 39335 的规定对数据共享的安全风险进行评估，并评估应对措施的有效性。

注：常见的安全风险主要包括数据滥用、数据的泄露和损毁、受托方对数据的未授权处理及未经许可的数据出境等。

7.4.6 应对数据共享过程中的任何异常情况进行及时处理，并形成记录。

7.4.7 应制定数据滥用相关的投诉处理机制，识别、报告和管理违规行为，并实施适当的问责。

7.4.8 宜定期对应急预案进行评估修订，每年至少组织 1 次应急演练。

7.4.9 宜根据检测评估、监测预警中发现的安全问题及处置结果开展综合评估，必要时重新开展风险识别，并更新安全策略。

7.5 安全事件处置

7.5.1 在数据安全事件发生后，应按照应急预案进行处置；事件处置完成后及时按规定向安全保护工作部门书面报告事件情况，内容包括但不限于：事件描述、原因和影响分析、处置方式。

7.5.2 宜组建专门的数据安全应急支撑团队、专家团队，保障安全事件得到及时有效处置。

7.5.3 宜制定灾难恢复计划，确保信息系统及时从数据安全事件中恢复，并建立安全事件追溯机制。

参 考 文 献

- [1] GB/T 18391.1—2009 信息技术 元数据注册系统（MDR） 第1部分：框架
- [2] GB/T 19771-2005 信息技术 安全技术 公钥基础设施 PKI组件最小互操作规范
- [3] GB/T 20984—2022 信息安全技术 信息安全风险评估方法
- [4] GB/T 35890—2018 高通量测序数据序列格式规范
- [5] GB/T 41479 信息安全技术 网络数据处理安全要求