



National Standard of the People’s Republic of China

GB/T XXXXX—XXXX

Requirements for medical high throughput gene sequencing data
sharing
医用高通量基因测序数据共享要求

(Draft for approval)

Issue date:

Implementation date:

Issued by	State Administration for Market Regulation Standardization Administration of the People’s Republic of China
-----------	--

Content

Foreword 错误！未定义书签。

1 Scope2

2 Normative references2

3 Terms and Definitions2

4 Basic principles4

5 Shared content5

6 Sharing scenarios and methods 6

7 Sharing management 7

Bibliography10

Foreword

SAC/TC 136 is in charge of this English translation. In case of any doubt about the contents of English translation, the Chinese original shall be considered authoritative.

This document is drafted in accordance with the rules given in GB/T 1.1—2020 Directives for standardization - Part 1: Rules for the structure and drafting of standardizing documents.

Attention is drawn to the possibility that some of the contents of this standard may be the subject of patent rights. The issuing body of this document shall not be held responsible for identifying any or all such patent rights.

This document was proposed by National Medical Products Administration.

This document was prepared by SAC/TC 136.

This document was drafted by the following organizations: International Human Phenome Institutes(Shanghai), International Institutes for Food and Drug Control, China Electronics Standardization Institute, Peking Union Medical College of Chinese Academy of Medical Sciences, Fudan University, Beijing Institute of Medical Device Testing (Beijing Medical Bio-protection Equipment Testing and Research Center), Greater Bay Area Institute of Precision Medicine (Guangzhou), BGI Genomics Co., Ltd. (Shenzhen), Shanghai Medical Device and Cosmetics Evaluation and Verification Center, Shenzhen Longgang District People's Hospital, C&M Information Technologies Co., Ltd.

Requirements for medical high throughput gene sequencing data sharing

1 Scope

This document specifies the basic principles, content, scenarios, methods, and management for medical high throughput gene sequencing data sharing.

This document is applicable to the management of gene data sharing activities among relevant parties involved in medical high throughput gene sequencing data, and is also suitable for supervision, administration, and evaluation of medical high throughput gene sequencing data sharing activities by regulatory authorities and third-party assessment organizations.

2 Normative references

The following documents are referred to in the text in such a way that some or all of their content constitutes requirements of this document. For dated references, only the edition cited applies. For undated references, the latest edition of the referenced document(including any amendments)applies.

GB/T 30534 Codes and identification of secret level for scientific and technical reports

GB/T 35273 Information security technology—Personal information security specification

GB/T 37964 Information security technology—Guide for de-identifying personal information

GB/T 39335 Information security technology—Guidance for personal information security impact assessment

GB/T 39725 Information security technology—Guide for health data security

GB/T 41806—2022 Information security technology—Security requirements of genetic recognition data

GB/T 43697 Data security technology — Rules for data classification and grading

3 Terms and definitions

选择一项。 For the purposes of this document, the terms and definitions given in GB/T 25069, GB/T 35273, GB/T 37964, GB/T 37988, and GB/T 41806, and the following apply.

3.1

medical high throughput gene sequencing data

Gene data and metadata generated by high throughput sequencing for medical purposes.

3.2

medical high throughput gene sequencing data resource

All data and metadata generated in the production, sharing, and processing of medical high throughput gene sequencing data in medical scenarios.

3.3

medical high throughput gene sequencing data subject

Natural persons identified or associated with medical high throughput gene sequencing data.

3.4

master data

Data about business entities.

3.5

metadata

Data that defines and describes other data.

[Source: GB/T 18391.1—2009, 3.2.16]

3.6

interested party

Participants in a process carried out for a particular purpose.

Note: Whether actively or passively involved, they are all relevant parties.

3.7

informed consent

An agreement between the medical high throughput gene sequencing data processor and the medical high throughput gene sequencing data subject or its legal guardian regarding the boundaries of medical high throughput gene sequencing data processing.

3.8

data processing

Process of storing, processing, retrieving, analyzing, and maintaining data.

3.9

completely public sharing

Once data is released, it is difficult to recall and is generally published directly through the Internet.

[Source: GB/T 37964—2019, 3.12]

3.10

controlled public sharing

Restrict the use of data through a data usage agreement.

[Source: GB/T 37964—2019, 3.13]

3.11

enclave public sharing

Shared within a physical or virtual territorial scope, data must not flow outside that scope. [Source: GB/T 37964—2019, 3.14]

3.12

data security

Measures and technologies that are taken to protect data from being damaged, leaked, tampered with, or used without authorization during storage, communication, and processing.

3.13

risk analysis

Systematic use of available information to identify hazards and estimate risks.

[Source: GB/T 42062-2022, 3.19]

3.14

data security risk

The probability of the occurrence of data security incidents and their implications for national security, public interest, as well as the lawful rights and interests of organizations and individuals.

3.15

data security risk assessment

The entire process of conducting information research, risk identification, risk analysis, and risk assessment for the security of data and data processing activities.

3.16

process audit of data sharing

The act of supervising, evaluating, and attesting the efficiency, effectiveness, and implementation status of an organization's data-sharing behavior.

4 Basic principles

The sharing of medical high throughput gene sequencing data should adhere to the following principles:

- a) The principle of legality, legitimacy, and necessity in compliance with relevant regulations on the management of human genetic resources and data security.
- b) Respect for individuals and groups, in line with medical ethics, protecting personal privacy information, safeguarding the legal rights of subjects of medical high throughput gene sequencing data, and not conflicting with the fundamental rights and freedoms of the subjects.
- c) Openness and transparency, promoting medical development, with relevant parties reaching a consensus and jointly establishing a data sharing management mechanism that ensures "data availability, security compliance, and process control".

5 Shared content

5.1 Overview

As shown in Figure 1, the resources of medical high throughput gene sequencing data include the nucleic acid sequence data (primary data) generated by medical high throughput gene sequencing, as well as related metadata and model data.

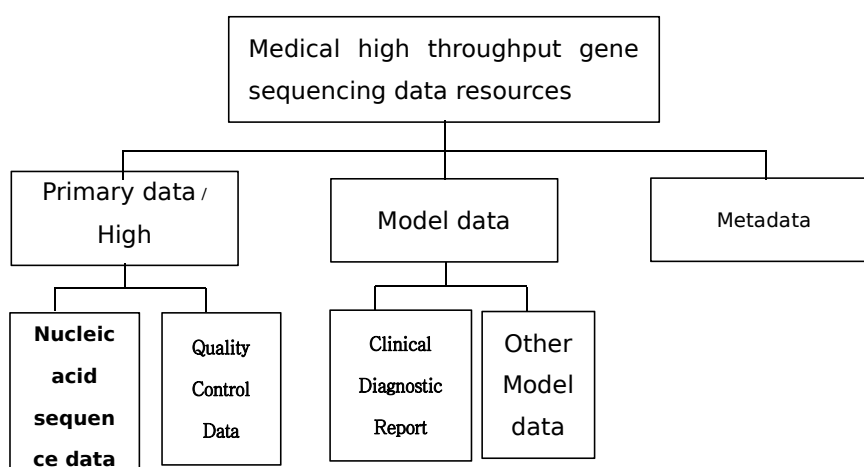


Figure 1 Medical high-throughput gene sequencing data resources

5.2 Master data

The master data of medical high throughput gene sequencing data resources are those generated by medical high throughput gene sequencing, including nucleic acid sequence data and corresponding quality control data.

5.3 Model data

Model data are processed data based on medical high throughput gene sequencing data, including but not limited to clinical diagnosis reports and analysis process data.

5.4 Metadata

The metadata of medical high throughput gene sequencing describes the master data and model data, including but not limited to informed consent forms, experimental procedures, sample information, correspondence between samples and patients, data processing methods, and other relevant standards for data processing.

6 Sharing scenarios and methods

6.1 Sharing scenarios

6.1.1 In clinical service scenarios, medical high throughput gene sequencing data are first used to achieve clinical purposes that benefit patients' treatment and rehabilitation process. In this scenario, stakeholders include but are not limited to subjects of medical high throughput gene sequencing data (patients), medical institutions, testing institutions, research institutions, related enterprises, and regulatory authorities.

6.1.2 In addition to being applied in clinical service scenarios, medical high throughput gene sequencing data may also be applied in academic research scenarios, transformation scenarios related to medical innovation, and regulatory scenarios related to drug and device supervision. See Figure 2 for the sharing scenarios of medical high throughput gene sequencing data.

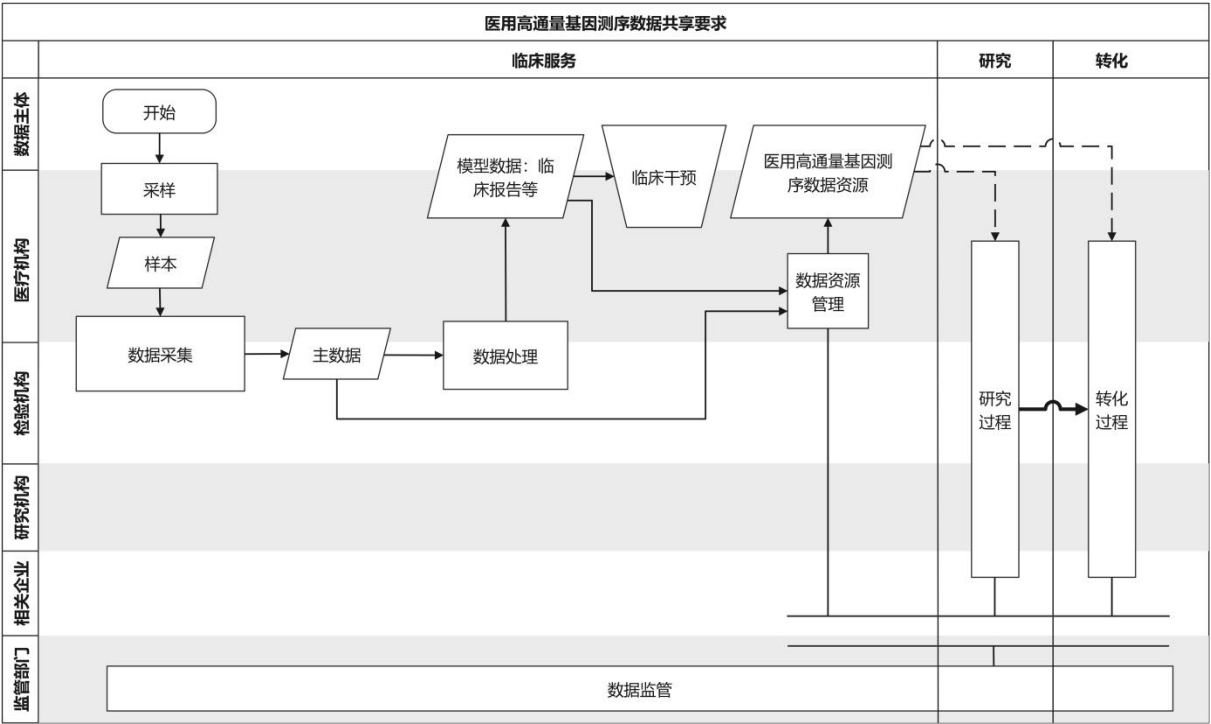


Figure 2 Sharing scenarios of medical high throughput gene sequencing data

6.2 Sharing methods

6.2.1 The types of data sharing may be divided into entirely public sharing, controlled

public sharing, and territorial public sharing. The data deidentification of different kinds of public sharing shall follow the relevant requirements of GB/T 37964 and GB/T 39725.

6.2.2 The forms of data sharing include website publishing, file sharing, application programming interface, online query, data analysis platform, and sharing in a secure computing environment.

6.2.3 In clinical service scenarios, data shall be shared among subjects of medical high throughput gene sequencing data, medical institutions, and testing institutions. If data sharing activities need to be carried out outside clinical service scenarios, they shall be implemented by specialized data resource management departments of medical institutions or national data infrastructure authorized by the government.

6.2.4 This document encourages the use of medical high throughput gene sequencing data for scientific research. In the case of scientific research, medical high throughput gene sequencing datasets shall be given a unified resource identifier, and their sharing activities shall follow the relevant provisions on scientific data sharing. As shown in Table 1.

6.2.5 In the scenario of drug and device supervision, the use of clinical reference genome standard datasets shall be carried out in a controllable and supervised secure computing environment.

Table 1 Data sharing methods

Data type	Data sharing scenarios	Type of data public sharing	Form of data public sharing
Personal genomic data	Clinical Services	Controlled public sharing	Application programming interface or file sharing, etc.
	Scientific research	Full public sharing or territorial public sharing	Website sharing or online querying, etc.
Clinical reference genomic standard dataset	Medical device and pharmaceutical regulation	Territorial public Sharing	Sharing within a secure computing environment

7 Sharing management

7.1 Organization

医用高通量基因测序数据共享的相关方，除医用高通量基因测序数据主体外，其他机构的组织设计应符合GB/T 41806—2022中第5章要求，包括明确数据安全负责人和管理机构、设立伦理管理制度、制定个人信息保护制度、建立数据安全管理机制、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医用高通量基因测序数据主体的合法权益。

In addition to the subjects of medical high throughput gene sequencing data, the organizational design of other institutions involved in sharing medical high throughput gene sequencing data shall meet the requirements of Chapter 5 in GB/T 41806—2022. This includes appointing a person responsible for data security and a managing institution, establishing an ethics management system, formulating a personal information protection system, establishing a data security management mechanism, and forming a complete information security management system. Moreover, it is necessary to conduct data security risk assessment, data security risk monitoring and early warning, and data access control for medical high throughput gene sequencing data to ensure the security of medical high throughput gene sequencing data throughout its life cycle and protect the lawful rights and interests of the subjects of medical high throughput gene sequencing data.

7.2 Personnel

The relevant personnel engaged in the sharing of medical high throughput gene sequencing data shall receive training related to data security to enhance their professional competence.

7.3 Processes

7.3.1 Planning

7.3.1.1 The confidentiality level of data shall be determined. The confidentiality levels of medical high throughput gene sequencing data shall be divided into four levels according to GB/T 30534, namely "public level", "restricted level", "secret level" and "confidential level". Determining and changing the confidentiality level for secret-level and confidential-level data shall follow relevant national confidentiality regulations; the data controller shall evaluate and modify the restricted-level data.

7.3.1.2 Data classification and grading shall be carried out. Medical high throughput gene sequencing data shall be managed through classification and grading following GB/T 43697, identifying important data, and establishing corresponding management measures for important data.

7.3.1.3 Relevant documents for data sharing shall be established, including but not limited to project sharing plans, security guarantee mechanisms, emergency response plans, etc.

7.3.1.4 The project sharing plan shall include but not be limited to the types of data to be shared, sharing objects, sharing methods, and the time limit for sharing.

7.3.1.5 The emergency response plan shall include but not be limited to the conditions for activating the emergency plan, emergency handling procedures, system recovery procedures, incident reporting procedures, and post-incident education and training.

7.3.1.6 The data sharing scope, practical cycle, expected results, and responsible persons for each project shall be clarified.

7.3.1.7 Procedures for identifying and managing conflicts of interest shall be formulated to ensure the fairness and integrity of the data sharing process.

7.3.2 Implementation

7.3.2.1 A dual mechanism of data sharing management and technical support shall be established. For each stage of the data lifecycle, a data sharing security guarantee mechanism shall be established to achieve effective and responsible data sharing.

Note: The stages of the data lifecycle generally include data collection, data processing, data sharing, data destruction, and other data management stages.

7.3.2.2 The key control points of the data sharing process shall be recorded in electronic or paper form, and the objectivity of the records shall be ensured through mechanisms or technical means, specifying the data flow, the responsible parties, and the boundaries of data use.

7.3.3 Inspection

7.3.3.1 Regular inspections of data sharing shall be conducted, and third-party organizations may be commissioned to audit the data sharing process if necessary.

7.3.3.2 In case of deviation from the plan or increased security risks during the data sharing process, timely safety assurance measures and corrective actions shall be taken to prevent more harm.

7.3.4 Improvement

Regular management reviews shall be conducted on data sharing management, technical support mechanisms, and data sharing security assurance mechanisms to enhance the effectiveness, timeliness, and safety of management and protection.

7.4 Security and emergency response

7.4.1 The classification and grading of medical high throughput gene sequencing data shall follow the requirements of GB/T 43697.

7.4.2 Personal information security impact assessments shall be carried out following GB/T 39335.

7.4.3 The management of model data shall comply with the agreed requirements of the informed consent form, carry out important data identification work, and share important data shall follow relevant legal requirements.

7.4.4 Medical high throughput gene sequencing data is considered personal information, and data sharing security shall meet the requirements of GB/T 35273.

Before carrying out medical high throughput gene sequencing data sharing activities, it is necessary to assess the security risks of data sharing and evaluate the effectiveness of response measures in accordance with the provisions of GB/T 39335.

Note: Common security risks mainly include data abuse, data leakage and damage, unauthorized processing of data by trustees, and unauthorized data transfer across borders, etc.

7.4.5 Any abnormalities during the data sharing process shall be promptly addressed and documented.

7.4.6 A complaint-handling mechanism related to data abuse shall be established, non-compliant behaviors should be identified, reported, and managed, and appropriate accountability shall be implemented.

7.4.7 It is advisable to regularly evaluate and revise emergency plans and organize at least one emergency drill each year.

7.4.8 It is advisable to conduct comprehensive assessments based on the security issues identified in inspections, evaluations, monitoring, and early warnings, as well as the results of handling these issues. If necessary, re-identify risks and update security strategies.

7.5 Security incident handling

7.5.1 In the event of a data security incident, it shall be handled in accordance with the emergency response plan. After the incident has been dealt with, a written report shall be promptly submitted to the security protection department, including but not limited to the description of the incident, analysis of its cause and impact, and the measures taken for handling.

7.5.2 Establishing a dedicated emergency data security support team and an expert group is advisable to ensure the timely and effective handling of security incidents.

7.5.3 It is advisable to develop a disaster recovery plan to ensure that the information system can promptly recover from a data security incident and establish a security incident traceability mechanism.

Bibliography

- [1] GB/T 18391.1—2009 Information technology - Metadata registries (MDR) - Part 1: Framework
 - [2] GB/T 42062—2022 Medical devices—Application of risk management to medical devices
 - [3] GB/T 19771—2005 Information technology-Security technology-Public key infrastructure-Minimum interoperability specification for PKI components
 - [4] GB/T 25069 Information security techniques—Terminology
 - [5] GB/T 20984—2022 Information security technology—Risk assessment method for information security
 - [6] GB/T 35890—2018 Technical specification of high throughput sequencing data format
 - [7] GB/T 37988 Information security technology—Data security capability maturity model
 - [8] GB/T 41479—2022 Information security technology—Network data processing security requirements
-

