ICS 11.100.10
CCS C 30

NATIONAL STANDARD OF THE PEOPLE'S REPUBLIC OF CHINA

GB/T 45214-2025

Data quality evaluation method of human whole genome sequencing

人全基因组高通量测序数据质量评价方法

Issued on: January 24, 2025 **Implemented on: February 1, 2026**

Issued by: State Administration for Market Regulation;
Standardization Administration of the PRC

Table of Contents

Foreword

This document is in accordance with the provisions of GB/T 1.1-2020 "Guidelines for standardization work Part 1: Structure and drafting rules for standardization documents" Drafting.

Please note that some of the contents of this document may involve patents. The issuing organization of this document does not assume the responsibility for identifying patents.

This document is proposed by the State Food and Drug Administration.

This document is under the jurisdiction of the National Technical Committee for Standardization of Medical Clinical Testing Laboratories and In Vitro Diagnostic Systems (SAC/TC136).

This document was drafted by: China Food and Drug Administration, Shenzhen MGI Intelligent Manufacturing Technology Co., Ltd., Shanghai Sildi Biomedical Science Technology Co., Ltd., Illumina (China) Scientific Instruments Co., Ltd., Beijing Genomics Institute of the Chinese Academy of Sciences (National Center for Bioinformatics Center), Cyna Biotech (Beijing) Co., Ltd., Shenzhen Hypros Biotech Co., Ltd., Beijing PanGenomics Gene Technology Co., Ltd.

Shenzhen Mingyi Intelligent Manufacturing Technology Co., Ltd., Shenzhen BGI Co., Ltd., Beijing Institute of Medical Device Inspection (Beijing Medical Device Inspection Institute Biological Protection Equipment Inspection and Research Center), Shanghai International Human Phenotype Group Research Institute.

The main drafters of this document are: Li Lili, Zhao Xia, He Qingzhong, Liu Fangfang, Ci Weimin, Chen Zitian, Chen Shifu, Hu Yunfu, Wu Jian, Peng Jiguang, Chen Fang, Wang

Ruixia, Shi Leming, Ding Guohui, Huang Jie.

Method for quality assessment of human whole genome high-throughput sequencing data

1 Scope

This document defines the terms and definitions of human whole genome high-throughput sequencing data quality, specifies quality requirements, and describes the corresponding evaluation method.

This document applies to the data quality of whole genome sequencing of human genomic DNA samples using high-throughput gene sequencing technology. evaluate.

This document does not apply to dideoxy chain termination sequencing [Sanger sequencing] technology and single molecule sequencing technology, as well as human samples De novo sequencing, haplotype sequencing, and sequencing of human tumor tissue samples are not applicable to the animal, plant, virus, or bacterial species contained in human samples. Bacteria, parasites, etc.

2 Normative references

This document has no normative references.

3 Terms and definitions

The following terms and definitions apply to this document.

3.1

Perform whole genome sequencing on different human individuals or groups.

Note. Includes 23 pairs of human chromosome nucleic acid sequences and mitochondrial nucleic acid sequences.

3.2 Library

A group of fragment molecules to be sequenced with adapters (DNA fragments of known sequence).

Note. Usually, it is divided into PCR library (a library constructed by a certain number of PCR amplification cycles) according to whether it relies on PCR amplification process.

There are three types of libraries. sequencing library (sequencing library) and PCR-free library (sequencing library constructed directly without relying on PCR amplification process).

3.3

When sequencing multiple samples together, the percentage of sequencing fragments that are correctly identified and assigned labels to the total number of sequencing fragments.

3.4

The percentage of sequenced bases with base recognition quality above the specified threshold to the total number of sequenced bases.

Note. Usually expressed as Q20, Q30, etc.

3.5 GC content

The sum of the number of guanine (G) and cytosine (C) in the sequenced fragment bases accounts for all purine bases [adenine Adenine (A) and Guanine] and the percentage of the total number of pyrimidine bases [thymine (T) and cytosine].