



National Standard of the People's Republic of China

GB/T XXXX— XXXX

General technical requirements for whole genome sequencing
of severe acute respiratory syndrome coronavirus 2 (SARS–
CoV-2)

新型冠状病毒全基因组测序通用技术要求

(Draft for approval)

Issue date: 202X–XX–XX

Implementation date: 202X–XX–XX

Issued by State Administration for Market Regulation

Standardization Administration of the People's Republic of China

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.

General technical requirements for whole genome sequencing of SARS-CoV-2

1. Scope

This document specifies the requirements for whole genome sequencing methods of SARS-CoV-2 and describes the corresponding test methods.

This document is applicable to whole genome sequencing of SARS-CoV-2 in samples such as oropharyngeal swabs, nasopharyngeal swabs, airway aspirates, sputum, bronchoalveolar lavage fluid, other respiratory secretions, or viral cultures, including metatranscriptomic sequencing, multiplex polymerase chain reaction (PCR) sequencing, and hybrid capture sequencing.

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 19489–2008, General requirements for laboratory biosafety

GB/T 19495.2–2004, *Laboratory technical requirements for detection of genetically modified products*

3. Terms and definitions

The following terms and definitions are listed in this document.

3.1 gene Sequencing

Determination of different base types in nucleic acid molecules, i.e., determining the composition or sequence of bases such as adenine (A), guanine (G), cytosine (C), and thymine (T) or uracil (U) that make up nucleic acid molecules. [Source: YY/T 1723–2020, 3.1]

3.2 whole Genome Sequencing of SARS-CoV-2

Perform whole genome sequencing of the SARS-CoV-2 to obtain complete genomic information.

3.3 metatranscriptomic Sequencing

The entire community in a specific environment is taken as the research subject, and all RNA is directly extracted without isolation and culturing, and then gene sequencing is performed after the complementary DNA is synthesized through reverse transcription. [Source: GB/T 40226–2021, 3.3, modified]

3.4 multiplex Polymerase Chain Reaction (Multiplex PCR)

Multiplex polymerase chain reaction is an amplification reaction in which more than two pairs of primers are used to simultaneously amplify multiple nucleic acid fragments in the same amplification reaction system. [Source: GB/T 19915.5–2005, 3.2, modified]

3.5 hybridization Capture

The process of customizing target gene region-specific probes for one or more gene nucleotide sequences, hybridizing with genomic nucleotides, and enriching target gene fragments. [Source: GB/T 37872–2019, 3.1.6, modified]

3.6 cycle threshold (Ct), crossing Point (Cp)

The number of cycles experienced when the fluorescence signal in the reaction tube reaches exponential

amplification during real-time monitoring of the amplification process. The main calculation method is to use the 10-fold standard deviation of the fluorescence values of the first 3 to 15 cycles of the amplification process as the threshold, and the amplification cycle number at the intersection of the amplification curve and the threshold line is the cycle threshold (Ct). [Source: YY/T 1182—2020, 3.4]

3.7 reads

Sequence fragments containing base sequences and quality values produced by high-throughput sequencing platforms. [Source: GB/T 35890—2018, 3.2]

3.8 read length of gene sequencing

The length of quality-qualified sequence fragments that can be read in a single run, usually expressed in the number of bases. [Source: YY/T 1723—2020, 3.4]

3.9 coverage rate of sequencing

The proportion of the nucleotide sequence detection results of the sample to be tested that is covered on the reference sequence (sequencing coverage rate = length of covered area ÷ total length of reference sequence). [Source: GB/T 30989—2014, 3.30, modified]

3.10 depth of sequencing

The number of times a specific nucleotide is detected in the sample to be tested. [Source: GB/T 30989—2014, 3.31, modified]

3.11 barcode, index

A characteristic short fragment of deoxynucleotide, serves as a unique identifier for the source of a specific sample during mixed sequencing of multiple samples. [Source: GA/T 1693—2020, 3.10, modified]

4. Requirements

4.1 Biosafety

Laboratory instruments and equipment must comply with the provisions of GB 19489 and GB/T 19495.2. Laboratories shall comply with relevant biosafety regulations. Operations involving uncultured infectious materials, such as nucleic acid extraction and clinical sample inactivation before reliable inactivation methods are applied, shall be conducted in a biosafety level 2 laboratory (BSL-2), with employing personal protection for biosafety level 3 laboratory (BSL-3). Nucleic acid detection on infectious materials or live viruses after reliable inactivation methods shall be conducted in BSL-2. Subsequent operations on extracted nucleic acids or prepared cDNA can be conducted in a biosafety level 1 laboratory (BSL-1).

4.2 Samples

The collection, packaging, transportation, storage, inactivation methods, and pretreatment methods of samples shall be validated.

4.3 Nucleic Acid Extraction and Purification

The efficiency of nucleic acid extraction and purification, such as yield, purity, and integrity, shall be validated.

4.4 Viral Nucleic Acid Quality Control

4.4.1 Nucleic acid quantification

The nucleic acid load of SARS-CoV-2 in the sample shall be determined as follows:

- a) The viral nucleic acid quantification methods, such as real-time fluorescent quantitative PCR, shall be

validated;

b) There shall be clear regulations on nucleic acid concentration. For example, when using real-time fluorescent quantitative PCR to determine the viral load in a sample, whole-genome sequencing is conducted when Ct value ≤ 32 ; when Ct value > 32 , the nucleic acid concentration may not meet the requirements for high-throughput sequencing, but sequencing can be performed after optimizing experimental conditions.

4.4.2 Fragment analysis

It is recommended to validate the viral nucleic acid fragment analysis methods.

4.5 Sequencing Library Preparation

The sequencing library preparation methods shall be validated. Library preparation methods include but are not limited to metatranscriptomic sequencing, multiplex PCR sequencing, and hybrid capture sequencing, as follows.

a) Metatranscriptomic sequencing: The main steps include removing host/other pathogen RNA (optional), reverse transcription, adding adapters, library amplification (optional), library product purification and nucleic acid quantification, and library product homogenization and mixing.

b) The main steps of the multiplex PCR sequencing method include: reverse transcription of the RNA from the sample nucleic acid, enrichment amplification using primers covering the SARS-CoV-2 genome, addition of adapters, library amplification (optional), purification and quantification of library products, and normalization mixing of library products. The targeted amplification primer pool shall be updated in a timely manner according to changes in SARS-CoV-2 variant sites.

c) The main steps of the hybrid capture sequencing method include: reverse transcription of RNA in the sample nucleic acid, addition of adapters, library amplification (optional), enrichment capture using hybrid probes specific to the SARS-CoV-2 genome, library amplification (optional), purification and quantification of library products, and normalization mixing of library products.

4.6 Sequencing Library Quality Control

Quality control shall be performed on the sequencing library, including but not limited to tag jumping contamination rate, library amplification uniformity, library product purification efficiency, fragment length, nucleic acid concentration, and the quantity of normalized mixing.

4.7 Sequencing

Sequencing shall be performed on the sequencing library, and the sequencing method shall be validated. Sequencing methods include reversible terminator sequencing, joint probe anchored polymerization sequencing, and single-molecule nanopore strand sequencing.

4.8 Sequencing Data Quality Control

4.8.1 Sequencing read length and effective data volume

The sequencing read length and data volume of the sequencing library shall be validated. Under the specified sequencing read length mode, the effective data volume shall meet the requirements specified by the laboratory user.

4.8.2 Base recognition quality percentage

The percentage of base or sequencing fragment recognition quality above the specified threshold in a single sequencing run shall be calculated, and the average base recognition quality percentage shall not be lower than the requirements specified by the laboratory user.

4.9 Viral Genome Alignment

4.9.1 Virus reference genome

The reference strain NC_045512.2 on GenBank shall be used as the reference genome for the SARS-CoV-2.

NOTE Reference genome information: https://www.ncbi.nlm.nih.gov/nuccore/NC_045512.2

4.9.2 Sequencing coverage and sequencing depth

The sequencing coverage of the entire SARS-CoV-2 genome shall be $\geq 96\%$ with a complete S gene, and the average sequencing depth shall be $\geq 10 \times$. The single nucleotide variant (SNV) filtering threshold shall meet the requirements specified by the laboratory user.

4.10 Virus genome assembly and typing

4.10.1 Virus genome assembly

Under the specified sequencing coverage and sequencing depth, the sequencing results of the SARS-CoV-2 shall be assembled, including reference-based assembly or de novo assembly.

The virus genome assembly method shall be validated.

4.10.2 Virus genome typing

The assembled SARS-CoV-2 genome shall be typed using methods such as the "Pangolin method," "GISAID method," and "Nextstrain method."

The virus genome typing method shall be validated.

4.10.3 Typing accuracy

The typing accuracy shall be validated. Under the specified sequencing coverage and sequencing depth, the detected typing results shall be consistent with the standard strain type of the SARS-CoV-2.

4.10.4 Mutation Analysis

Analyze mutation types such as single nucleotide variants (SNV) and insertions/deletions (Indel), and annotate the corresponding amino acid changes.

5. Test method

5.1 Biosafety

The laboratory shall be zoned according to different work contents to avoid cross-contamination, including reagent storage and preparation area, sample preparation area, nucleic acid amplification area, library preparation area, high-throughput sequencing area, data analysis area, etc. The centralized layout of the laboratory shall follow the principle of "independent zones, unidirectional flow (pay attention to wind direction, pressure gradient direction), adapt to local conditions, and facilitate work".

5.2 Samples

The collection, packaging, transportation, storage, inactivation, and pretreatment of samples shall be carried out according to the technical guidelines in the "SARS-COV-2 Prevention and Control Plan (Ninth Edition)"

5.3 Nucleic Acid Extraction and Purification

Nucleic acid extraction and purification methods include guanidine isothiocyanate-phenol method, spin column purification method, magnetic bead method, etc., and shall be tested according to the extraction reagent instructions.

5.4 Viral Nucleic Acid Quality Control

The quality control methods for viral nucleic acid include nucleic acid concentration measurement and fragment analysis, as follows:

- a) Viral nucleic acid quantification methods include real-time fluorescent quantitative PCR method, fluorescent dye method, etc., and shall be tested according to the steps in the detection reagent instructions;
- b) If necessary, fragment analysis can be performed using a fragment analyzer, capillary electrophoresis, etc.

5.5 Sequencing Library Preparation

Library preparation methods include but are not limited to metatranscriptomic sequencing, multiplex PCR sequencing, and hybrid capture sequencing, and shall be tested according to the library preparation instructions. Prepare libraries for SARS-COV-2 standards and determine if the results meet the requirements as given in 4.5.

5.6 Sequencing Library Quality Control

Library quality control includes clearly labeled added tag adapters, detecting library concentration, detecting library fragment size, etc., and shall be tested according to the methods specified or recommended in the library preparation instructions. Normalize and mix qualified library products, with concentration based on the requirements of sequencing reagents and instruments. Perform library quality control on SARS-COV-2 standards and determine if the results meet the requirements as given in 4.6.

5.7 Sequencing Methods

Sequencing methods include reversible terminator sequencing, probe-anchored polymerization sequencing, and single-molecule nanopore sequencing, and shall be tested according to the methods in the sequencing reagent instructions and sequencing instrument instructions. Sequence SARS-COV-2 standards and determine if the results meet the requirements as given in 4.7.

5.8 Sequencing Data Quality Control

Sequencing data quality control includes removing sequences with adapters or low quality, host genome sequences, etc., and removing primer sequences from multiplex PCR sequencing data. After sequencing, the quality values of bases or sequencing fragments are analyzed and statistically evaluated. Perform sequencing data quality control on SARS-COV-2 standards and determine if the results meet the requirements as given in 4.8.

5.9 Viral Genome Alignment

Align the quality-controlled data with the reference sequence to obtain the bam file. Calculate the sequencing coverage value and average sequencing depth value according to formula (1) and formula (2), respectively. Determine if the results meet the requirements as given in 4.9.2.

$$CVR = \frac{CVB}{GB} \leftrightarrow 100\% \dots\dots\dots(1)$$

In the formula:

CVR: Sequencing coverage rate;

CVB: Number of bases in the covered genome sequence;

GB: Total number of bases in the genome.

$$DP = \frac{MB}{CVB} \dots\dots\dots(2)$$

In the formula:

DP: Average sequencing depth;

MB: Total number of aligned bases;

CVB: Number of bases in the covered genome sequence.

5.10 Virus genome assembly and typing

Detect mutation sites from the bam file aligned with the reference sequence, and filter SNVs that do not meet the requirements, and use bioinformatics tools and algorithms to assemble and correct the virus genome to obtain a consensus sequence, and then use "Pangolin method", "GISAI method", and "Nextstrain method" to perform typing identification on the assembled genome sequence to determine whether the typing accuracy of the SARS-CoV-2 standard is in line with the requirements as given in 4.10. Use bioinformatics tools to determine new mutation sites by comparing with the reference genome and typing reference strains, and annotate the corresponding amino acid changes, etc.

Bibliography

- [1] GB 19489–2008, General requirements for laboratory biosafety
- [2] GB/T 19495.2–2004 Detection of genetically modified organism and derived products-General requirements for laboratories
- [3] GB/T 19915.5–2005 Protocol of multiplex PCR identification of *Streptococcus suis* type 2
- [4] GB/T 30989–2014 Technical regulation of high-throughput gene sequencing
- [5] GB/T 35537–2017 Requirements of the high-throughput gene sequencing result evaluation
- [6] GB/T 35890–2018 Technical specification of high throughput sequencing data format
- [7] GB/T 37872–2019 Guidelines for validation of next-generation target region sequencing
- [8] GB/T 40226–2021 Detection of environmental microbial metagenome-High-throughput sequencing
- [9] GB/T 40982–2021 Quality assessment requirements for severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) nucleic acid detection kit
- [10] GA/T 1693–2020 Forensic science-Specifications for second generation sequencing - based DNA examination
- [11] YY/T 1182–2020 Nucleic acids amplification test reagents(kits)
- [12] YY/T 1723–2020 High-throughput gene sequencer
- [13] DB32/T 3762.11–2020 Technical specifications for SARS-CoV-2 detection Part 11: Whole genome high-throughput sequencing
- [14] T/SZGIA 2–2018 The quality standard of Using high-throughput sequencing data to interpretate human whole genome variation
- [15] National Health Commission of the People's Republic of China, Laboratory biosafety guide for 2019-nCoV (Second Edition) [R] 2020
- [16] National Health Commission of the People's Republic of China, Protocol for Prevention and Control of COVID-19 (Ninth Edition) [R] 2022