INTERNATIONAL STANDARD

ISO 20397-1

First edition
2022-03

Biotechnology — Massively parallel sequencing —

Part 1: Nucleic acid and library preparation

Biotechnologie — Séquençage parallèle massif —

Partie 1: Acides nucléiques et préparation des collections

Reference number
ISO 20397-1:2022(E)

ISO copyright office
CP 401 • Ch. de Blandonnet 8
CH-1214 Vernier, Geneva
Phone: +41 22 749 01 11
Email: copyright@iso.org
Website: www.iso.org

Published in Switzerland

Contents

Page

Foreword

ISO (the International Organization for Standardization) is a worldwide federation of national standards bodies (ISO member bodies). The work of preparing International Standards is normally carried out through ISO technical committees. Each member body interested in a subject for which a technical committee has been established has the right to be represented on that committee. International organizations, governmental and non-governmental, in liaison with ISO, also take part in the work. ISO collaborates closely with the International Electrotechnical Commission (IEC) on all matters of electrotechnical standardization.

The procedures used to develop this document and those intended for its further maintenance are described in the ISO/IEC Directives, Part 1. In particular, the different approval criteria needed for the different types of ISO documents should be noted. This document was drafted in accordance with the editorial rules of the ISO/IEC Directives, Part 2 (see www.iso.org/directives).

Attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights. ISO shall not be held responsible for identifying any or all such patent rights. Details of any patent rights identified during the development of the document will be in the Introduction and/or on the ISO list of patent declarations received (see www.iso.org/patents).

Any trade name used in this document is information given for the convenience of users and does not constitute an endorsement.

For an explanation of the voluntary nature of standards, the meaning of ISO specific terms and expressions related to conformity assessment, as well as information about ISO's adherence to the World Trade Organization (WTO) principles in the Technical Barriers to Trade (TBT), see www.iso.org/iso/foreword.html.

This document was prepared by Technical Committee ISO/TC 276, *Biotechnology.*

A list of all parts in the ISO 20397 series can be found on the ISO website.

Any feedback or questions on this document should be directed to the user's national standards body. A complete listing of these bodies can be found at www.iso.org/members.html.

Introduction

Massively parallel sequencing (MPS) is a high throughput analytical technology for nucleic acid sequencing. MPS methods can process thousands to billions of nucleotide sequence reads simultaneously in a single run, allowing whole genomes, transcriptomes and specific nucleic acid targets from different organisms to be analysed in a relatively short time.

MPS is used in many life science disciplines permitting determination and high throughput analysis of millions of nucleotide bases. The biological variability of deoxyribonucleic and ribonucleic acid polymers from living organisms provides challenges in accurately determining their sequences. The quality of sequence determination by MPS depends on many factors including, but not limited to, sample quality, library preparation, and sequencing data quality.

The quality of nucleic acids and libraries prepared for MPS is critical to obtaining high quality sequence data. Controlling the upstream processing steps of MPS and evaluating nucleic acid samples and libraries for their suitability for sequencing significantly improves MPS results, downstream analyses and ultimately conclusions dependent upon the MPS data.

Biotechnology — Massively parallel sequencing —

Part 1: Nucleic acid and library preparation

1 Scope

This document specifies the general requirements for and gives guidance on quality assessments of nucleic acid samples. It specifies general guidelines for library preparations and library quality assessments prior to sequencing and data generation.

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.

ISO 20395:2019, *Biotechnology — Requirements for evaluating the performance of quantification methods for nucleic acid target sequences — qPCR and dPCR*

3 Terms and definitions

For the purposes of this document, the terms and definitions given in ISO 20395:2019 and the following apply.

ISO and IEC maintain terminological databases for use in standardization at the following addresses:

— ISO Online browsing platform: available at https://www.iso.org/obp

— IEC Electropedia: available at https://www.electropedia.org/

3.1
adapter
oligonucleotides of known sequence that are enzymatically added (e.g. ligase or polymerase chain reaction) to the end(s) of a DNA/cDNA fragment

3.2
barcode
index
short sequence of typically six or more nucleotides that serve as a way to identify or label individual samples when they are sequenced in parallel on a single sequencing lane, chip or both

Note 1 to entry: Barcodes are typically located within the sequencing *adapters* (3.1).

3.3
barcoding
indexing
unique DNA sequence identification
method that enables multiple samples to be pooled for sequencing

Note 1 to entry: Each sample is identified by a unique *barcode* (3.2), which enables identification of results during the parallel analysis.

3.4
GC content
GC
percentage of guanine and cytosine in one or more nucleic acid sequence(s)

Note 1 to entry: The amount of guanine and cytosine in a polynucleotide is usually expressed as fraction (or percentage) of total nitrogenous bases. Total nitrogenous bases comprise the total number of nucleotide bases of reads from one or more MPS run.

[SOURCE: ISO 20397-2:2021, 3.15]

3.5
library
sequencing library
DNA, cDNA or RNA that has been prepared for massively parallel sequencing within a specific size range and typically containing *adapters* (3.1) and/or identifiers recognised for sequence specific priming, sequence capture, and/or identification of specific extracts

Note 1 to entry: Libraries can be DNA or cDNA. cDNA libraries are prepared for RNA sequencing on most sequencers. Some instruments can directly sequence RNA.

3.6
library preparation
sequencing library preparation
set of procedures used to prepare DNA or RNA fragments containing tags, and sequencing primer binding regions for massively parallel sequencing (MPS)

3.7
spike-in control
spike-in process control
target sequence often of defined sequence identity and concentration that are spiked in the sample at various steps of the massively parallel sequencing protocol

Note 1 to entry: Process controls can be used to evaluate any protocol step but are typically applied as nucleic acids controls prior to *library preparation* (3.6).

3.8
Q score
measure of the sequencing quality of a given nucleotide base

[SOURCE: ISO 20397-2:2021, 3.32, modified — The notes to entry have been deleted.]

4 Nucleic acid sample quality evaluation

4.1 General

The laboratory shall establish, implement and document a workflow for nucleic acid quantification that ensures accurate and reproducible results. Requirements for nucleic acid sample quantity and quality can vary between MPS methods. Nucleic acid purification methods can also affect the quality of nucleic acids used for library preparation.

A quality control procedure shall be developed to clearly define nucleic acid quality and library composition. This procedure shall be verified, implemented and documented, and permit accurate quantification of nucleic acid at the minimum amount of nucleic acid required for the MPS performed. The measurement uncertainty and sensitivity of the procedure used for this determination shall be determined. The quantification allows appropriate adjustment of the nucleic acid concentration for input into the MPS sequencer.

Annex A provides a quality control checklist. Quantity, purity and integrity are major quality indicators for the prepared samples. Additional general considerations of sample quality regarding multiplex molecular testing including NGS are available in ISO 21474-1:2020.

4.2 Sample quantification

A range of methods for nucleic acid quantification are provided in ISO 20395:2019, 5.2. Other methods (e.g. electrophoresis) can also be used for the quantification.

Optimal sample amounts and concentrations appropriate for different MPS applications are listed in Table B.2.

4.3 Sample purity

Nucleic acid sample purity analysis shall be conducted in accordance with methods specified in ISO 20395:2019, 5.4.

4.4 Sample integrity

4.4.1 General

A range of methods used for assessing sample integrity is described in ISO 20395:2019, Annex B.

Gel electrophoresis and microfluidic analysis system can be used to evaluate nucleic acid sample integrity.

4.4.2 Agarose gel electrophoresis

Agarose gel electrophoresis can be used as a method for separating and isolating different sized nucleic acid molecules. It can also be used to determine nucleotide acid integrity. For example, optimally, genomic DNA (gDNA) samples have a strong main band of high molecular mass (greater than 20 kbp[1)] in size) with minimal band dispersion.

4.4.3 Capillary gel electrophoresis

Capillary gel electrophoresis can also be used to assess nucleic acid integrity.

4.4.4 Microfluidic analysis system

Microfluidic analysis system can be used to assess the integrity of genomic DNA or RNA extracted from various materials.

NOTE 1 DNA or RNA integrity number is commonly used as a numerical quality assessment criterion. The higher the value, the better the quality.

NOTE 2 Specifies the appropriate threshold depending on the type of devices.

4.4.5 PCR method

A PCR method can be used for integrity evaluation. High quality samples can generate data that are more useful than data generated from degraded samples.

NOTE Formalin-fixed and paraffin-embedded (FFPE) samples can cause challenges for some DNA applications. Further guidance is given in the ISO 20166 series.

1) kbp = kilo base pairs.

5 Nucleic acid library preparation

5.1 General

The laboratory shall establish, implement and document each procedure for nucleic acid library preparation that ensures accurate and reproducible results.

The quality of the MPS library is determined by the following procedures including, but not limited to:

a) fragmentation;

b) addition of universal sequences;

c) size selection;

d) amplification;

e) purification and clean up;

f) library quantification;

g) library qualification.

5.2 Fragmentation

5.2.1 General

Some sequencing methods (e.g. short read sequencing) require the template DNA, cDNA or RNA to be fragmented as a first step prior to library preparation.

Fragmentation can be performed either mechanically or enzymatically to produce the DNA or RNA size range that is required for the particular method and sequencing platform. Chemical fragmentation is typically reserved for long RNA fragments.

The selection of a fragmentation method should take into account the impact of the specific approach on evenness of coverage in the final libraries, e.g. to avoid the introduction of a GC bias.

The amount of starting material available and potential sample loss resulting from each approach should also be considered.

5.2.2 Mechanical fragmentation

Mechanical shearing can be performed using focused acoustic shearing devices. The resulting fragment sizes (150 bp[2)] to 5 000 bp) can be controlled by varying the intensity and duration of ultrasonic acoustic waves.

Hydrodynamic shearing can be used to produce larger fragments (typically 1 kbp to 75 kbp), but requires large DNA input amounts (>1 µg) and the throughput is low.

Nebulization is another alternative, which uses compressed air to force DNA or RNA through a small hole. Although fragment size can be controlled to an extent, large amounts of input DNA (microgram quantities) are required and the method is only suitable for small sample sizes.

5.2.3 Enzymatic fragmentation

Enzymatic fragmentation methods (e.g. using fragmentases, transposases or endonucleases) have the advantage of higher throughput compared to mechanical methods and typically result in a lower

2) bp = base pairs.

sample loss. The disadvantage of enzymatic approaches is that they can typically result in sequence bias since many enzymes have specific recognition sequences or sequence preferences.

5.2.4 Chemical fragmentation

Chemical fragmentation is typically reserved for breaking up long RNA fragments. Chemical fragmentation is performed by heating RNA with a divalent metal cation (magnesium or zinc). The length of the resulting products ranges from 115 base nucleotides to 350 base nucleotides and can be adjusted by increasing or decreasing the time of incubation.

5.2.5 Fragmented nucleic acid sample quantity

Certain library preparation protocols require the fragmented DNA or RNA to be quantified (as some material can be lost during the process). This can be performed using any of the methods described in 4.2, but most typically this is done with spectrophotometry or intercalating fluorescent dyes.

5.2.6 Fragmented nucleic acid sample purity

If there is a risk that impurities from the fragmentation process (e.g. components of enzymatic fragmentation reactions) can be carried over into the purified products, the sample purity can be assessed using the methods described in 4.3.

5.2.7 Fragmented nucleic acid size distribution

The fragmented nucleic acid should be checked to determine whether the appropriate size range has been achieved. This can be done using the methods described for monitoring sample integrity in 4.4.

5.2.8 Fragmented nucleic acid purification using gel electrophoresis

Purification of fragmented nucleic acids can be done by separating nucleic acids with a specific size from those with other sizes before downstream library preparation and sequencing steps. The purification can be achieved by using capillary electrophoresis, bead-based methods or other electrophoresis methods. The purified nucleic acids can be quantified as described in 4.2.

5.3 Addition of universal sequences

5.3.1 Repair

Because damage can rise from fragmentation, the nucleic acid sample shall be repaired after this process to improve efficiency in subsequent preparation steps.

NOTE The following conditions can warrant a repair procedure: abasic sites, nicks, thymine dimers, blocked 3′-ends, oxidized guanines or pyrimidines, deaminated cytosine.

The end of the fragment shall be polished (i.e. the addition of 5′-PO_4 or 3′-OH) to make it suitable for ligation.

5.3.2 Ligation of adapter

Evaluation of the adapter sequence can include, but is not limited to:

a) length;

b) the design of the adapter;

c) the ratio of adapter to DNA.

The ratio of adapter is critical and requires optimization.

5.3.3 Barcoding/indexing

Libraries can be single or dual indexed. Unique dual indexing can be used with MPS platforms that use patterned flow cells to mitigate the effects of incorrect assignment of libraries from the expected index to a different index.

The sequence of barcodes set should be as unique and heterologous as possible.

Nucleotide sequences chosen for barcodes/indexes should be differentiable within the sequencing project.

Barcode length should be as short as possible. See Table B.1 for example references for this quality metric.

The barcode should be designed to minimize adapter-dimer/primer-dimer artefacts.

The sequence of the barcode/index normally consists of a 6 bp to 12 bp. Each barcode/index in a DNA sequencing library should have a unique sequence that is easily differentiable from all of the other barcodes/indexes in that DNA sequencing library.

Barcodes are normally located next to the adapters.

The set of barcode sequences can be evaluated using the following parameters:

a) diversity of barcode sequences (relevant to some MPS platforms);

b) hamming distance (number of bases difference between each index combination).

If b) is used for the evaluation, the hammering distance should be > 3 bp.

5.4 Size selection

The optimal length of nucleotides in the library should be determined for the specific application. This is typically achieved by using the following methods including, but not limited to:

a) bead-based methods;

b) electrophoresis.

5.5 Amplification

Bias of polymerase can cause a failure of the amplification.

Polymerase can introduce errors; this can be minimized by increasing the sequence coverage or technical replicates.

NOTE 1 Over amplification produces PCR duplicates.

NOTE 2 High-fidelity amplification enzyme can be used to reduce errors.

5.6 Purification and clean up procedures

Before MPS is performed, the PCR-amplified pool of fragments or unfragmented nucleic acid should be cleaned to remove excess adapters or other contaminants. Techniques can include, but are not limited to:

a) bead clean up;

b) spin column;

c) enzymatic clean up;

d) gel electrophoresis-based clean up.

5.7 Library quantification

5.7.1 Library quantification method

Libraries can be quantified by, but are not limited to, the following methods:

a) nucleic acid fluorescent dyes or spectrophotometric assays;

b) electrophoresis;

c) real-time quantitative PCR assays;

d) digital PCR assays;

e) methods for high molecular weight nucleic acid.

5.7.2 Selection of quantification method

Quantification method selection depends on several factors including, but not limited to:

a) the size distribution of the libraries (e.g. fluorometric methods is typically used for selection, but not appropriate for libraries with a wide distribution in fragment sizes);

b) the type of library preparation method used (e.g. qPCR quantification methods typically use primers that bind to the library adapters and therefore are most accurate for library preparation methods in which there can be large amounts of DNA that does not contain sequencing adapter).

5.8 Library qualification

5.8.1 General

Library qualification shall be performed to detect potential problems such as high percentage of short DNA fragments or adapter dimers, and to determine whether they are of the appropriate length (taking into account the additional length of the sequencing adapters).

The laboratory shall use validated protocols to obtain fragment sizes of the specific molecular mass range for the library preparation.

Specific measure(s) should be taken to minimize primer dimers, adapter dimers, and broader bands of higher molecular mass. Primer dimers are usually minimized by using magnetic beads and do not constitute a significant problem unless they dominate the reaction.

5.8.2 Methods

Electrophoretic methods are generally used to assess the overall range of the DNA fragment size constituting the library.

6 Validation

A validation for intended use protocol shall be developed, implemented and documented for the entire workflow, before accumulating validation data.

Assay validation should be performed using samples of the type intended for the assay, so that the test performance is representative of the larger population of samples. However, MPS of multiple genes cannot be validated as if it was a single-analyte test. There is high level of variation in the types of samples, types of variants, allele burden, and targeted exons or regions. The validation process should systematically identify the potential failures throughout the library preparation. Errors at each step

and their source can be addressed and evaluated at different levels, such as assay design, method validation and/or quality control.

The performance requirements for the method can be verified during the validation procedure, and the same specifications shall be used to monitor the performance of the assay each time a sample is processed. The fit for purpose assessment of the whole measurement system shall be taken into account to establish the performance requirements.

Given the inherent differences among platforms, specific applications and informatics tools, specific recommendations for ranges and thresholds cannot be offered on all circumstances, and each laboratory shall define the criteria and means to monitor all quality metrics to ensure optimal analytical performance based on the examination manufacturer's instructions. Quality metrics for validation based on exemplary MPS platforms are listed in Table B.1.

7 Reference materials or controls

7.1 General

Control samples can be used to readily detect sources of error. Control samples should be used to monitor steps of the method for which validation and verification data have shown a potential for variability or error. Cell lines, DNA/genome fragments, microbial organisms, etc. can be used as reference materials for quality control purpose.

The use of controls shall be specified in a range of minimum and maximum nucleic acid detection levels. Every batch of prepared control samples should be verified before use as a control and documented.

7.2 Control samples

There are various reference samples that can be used in quality benchmarks or validation experiments or both. These should be used when they are available. A list of reference material and provider examples is provided in Annex C.

Control samples generally fall into three categories:

a) well characterized cell lines;

NOTE 1 The most widely used are HapMap cell lines or Personal Genome Project cell lines.

b) well characterized DNA/genome fragments;

NOTE 2 Have particular advantages because they can be designed to incorporate specific sequence variants at known positions.

c) microbial reference materials that cover a wide range in % GC content are commonly used to evaluate GC bias.

7.3 Positive control

DNA or RNA from a well characterized species or a known source should be used as a positive control to monitor the sequence quality on different runs to identify problems with the sequencing chemistry.

The same DNA or RNA extract should be consistently used for similar sources or applications.

The positive control shall have fragments in a range that is suitable for the technology being used, and the results of sequencing shall be consistent between runs.

The frequency of use of the positive control for monitoring quality over time should be established, implemented and documented.

7.4 Negative control

A negative control can be used during the sample preparation. In some applications, such as target sequencing/gene enrichment applications, specific DNA or RNA template negative controls can be used for evaluation.

NOTE A negative control is used to evaluate cross-contamination occurring during library preparation and the general specificity of a procedure.

The negative control shall have no detectable peaks and minimal sequencing reads associated with the negative control.

A negative control is not required for every run. The frequency of use of the negative control for monitoring quality over time should be established, implemented and documented.

7.5 No-template control

No-template controls should be performed during library preparation to monitor contamination/ non-specific amplification during PCR steps, and these can potentially also be sequenced with certain protocols to confirm the identity of such products.

7.6 Spike-in control

A spike-in control can be used to verify target sequencing and instrument performance.

This control is carried along the whole process and undergoes the same handling steps as the investigated sample, from initial quantification to final downstream processing. If a sequence error is observed in the reference control, the same error occurred in all likelihood in the main sample.

Spike-in control can be used only in the library construction.

A controlled library should be included in each run.

In high concentrations (>1 %), the spike-in control can be used in some short-read sequencing platforms to boost sequence diversity in low-diversity libraries.

Reference materials can be any useful stable samples (such as cell lines) or maintained by third parties (e.g. national standardization bodies). A list of reference material and provider examples is provided in Annex C.

7.7 Reference materials

Commercially available materials can be used for specific applications.

8 Contaminations

8.1 General

Special attention should be paid to cross-contamination in the processing procedure, especially during pre-PCR steps including the sample pre-analytical phase (e.g. processing of original sample, DNA fragmentation and nucleic acid library construction). Contamination can be monitored through the use of appropriate controls (e.g. no-template controls in PCR steps).

8.2 Primary sample evaluation

The primary sample should be evaluated to ensure that there is no contamination by exogenous DNA, protein, RNA or other potential impurities. This can be achieved using the methods described in 4.4. If required, the sample purity can be improved through purification methods such as column-based kits or magnetic beads.

8.3 Protocol and operation procedure

At appropriate intervals, technical operator competence should be evaluated with respect to the defined procedure and requirements to ensure proper method performance and avoid cross-contamination between different samples or nucleic acid libraries. Run performance can be assessed by analysing quality control metrics such as error rates, read counts and Q scores.

NOTE Performing simultaneous detection with different libraries using a single gel electrophoresis experiment can lead to compromised results.

Appropriate extraction methods and tools shall be used to extract target DNA fragment from gel.

Annex A
(informative)

Checklist for sample quality assessment before library construction

Points of considerations for sample quality assessment:

a) What is the source of the nucleic acid sample (bacteria, tissues, blood, etc.)?

NOTE 1 Knowing this information can be critical in upfront quality control steps.

EXAMPLE If the sample is from plants known to have high polysaccharides, they can be co-extracted, and a clean-up procedure can be necessary to remove them from the sample.

b) What methods were used for DNA isolation? Does the isolation method properly remove salts or detergents?

c) Was sample quantitation performed based on fluorometry?

d) Was there enough DNA to perform library preparation?

e) Were gels run to assess quality of the gDNA?

NOTE 2 It is important to know whether you are working with RNA, DNA or degraded DNA.

f) Were necessary pre-shearing clean-up methods performed?

g) Is the sample fit for purpose? Are the primary sample/sample pre-analytical conditions used within the range of the intended use?