

This is a preview of "ISO 20397-2:2021". [Click here to purchase the full version from the ANSI store.](#)

First edition  
2021-02

---

---

# Biotechnology — Massively parallel sequencing —

## Part 2: Quality evaluation of sequencing data

*Biotechnologie — Séquençage massivement parallèle —*

*Partie 2: Évaluation de la qualité des données de séquençage*



Reference number  
ISO 20397-2:2021(E)

© ISO 2021



**COPYRIGHT PROTECTED DOCUMENT**

© ISO 2021

All rights reserved. Unless otherwise specified, or required in the context of its implementation, no part of this publication may be reproduced or utilized otherwise in any form or by any means, electronic or mechanical, including photocopying, or posting on the internet or an intranet, without prior written permission. Permission can be requested from either ISO at the address below or ISO's member body in the country of the requester.

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

Published in Switzerland

This is a preview of "ISO 20397-2:2021". Click here to purchase the full version from the ANSI store.

## Contents

	Page
<b>Foreword</b> .....	<b>iv</b>
<b>Introduction</b> .....	<b>v</b>
<b>1 Scope</b> .....	<b>1</b>
<b>2 Normative references</b> .....	<b>1</b>
<b>3 Terms and definitions</b> .....	<b>1</b>
<b>4 Raw data</b> .....	<b>6</b>
4.1 General.....	6
4.2 Raw data file.....	6
4.3 Quality assessment of raw data.....	6
4.3.1 General.....	6
4.3.2 Basic statistics.....	7
4.3.3 Quality metrics.....	7
4.4 Raw data pre-processing.....	8
<b>5 Sequence alignment and mapping</b> .....	<b>8</b>
5.1 General.....	8
5.2 Alignment and mapping file format.....	9
5.3 Quality control of sequencing alignment and mapping.....	9
5.3.1 Basic alignment statistics.....	9
5.3.2 Quality indicators.....	10
5.3.3 Methods for alignment and mapping quality assessment.....	11
5.4 Alignment post-processing.....	11
<b>6 Variant calling</b> .....	<b>11</b>
6.1 General.....	11
6.2 Data file for variant calling.....	11
6.3 Quality metrics in the variant calling.....	12
6.4 Processing of false positive variants.....	12
6.5 Sequence annotation.....	12
<b>7 Validation</b> .....	<b>12</b>
7.1 General.....	12
7.2 Validation of quality metrics.....	13
<b>8 Documentation</b> .....	<b>14</b>
<b>Annex A (informative) Quality metrics for specific example MPS platforms</b> .....	<b>15</b>
<b>Annex B (informative) Coverage and read recommendations by applications</b> .....	<b>16</b>
<b>Annex C (informative) Software for sequence alignment and mapping</b> .....	<b>18</b>
<b>Bibliography</b> .....	<b>19</b>

## 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](http://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](http://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](http://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](http://www.iso.org/members.html).

This is a preview of "ISO 20397-2:2021". [Click here to purchase the full version from the ANSI store.](#)

## Introduction

Massively parallel sequencing (MPS) is a high-throughput analytical approach to nucleic acid sequencing utilizing massively parallel processing, that allows whole genomes, transcriptomes and specific nucleic acid targets from different organisms to be investigated in a relatively short time.

MPS is used in many life science disciplines permitting determination and high throughput analysis of millions and thousands of millions of nucleotide bases. The biological variability of deoxyribonucleic and ribonucleic acid polymers from living organisms results in 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, platform selection, and sequencing data quality.

The analysis of sequencing data poses significant bioinformatics challenges in various areas such as data storage, computation time and variant detection accuracy. One of the major challenges associated with sequencing data that is sometimes easily overlooked is monitoring quality control metrics over all stages of the data processing pipeline. Knowledge of data quality is essential for downstream analysis of sequences. Quality control for nucleic acid sequencing data handling and analysis can be separated into three stages: raw data, alignment and variant calling. This document provides a list of considerations for quality evaluation of MPS sequencing data, and the specific recommendations for different MPS platforms.