

This is a preview of ISO/TS 22692:2020. [Click here to purchase the full version from the ANSI store.](#)

First edition
2020-10

Genomics informatic — Quality control metrics for DNA sequencing

Informatique génomique — Mesures de contrôle de la qualité pour le séquençage de l'ADN



Reference number
ISO/TS 22692:2020(E)

© ISO 2020



COPYRIGHT PROTECTED DOCUMENT

© ISO 2020

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
Website: www.iso.org

Published in Switzerland

This is a preview of ISO/TS 22692:2020. Click [here](#) to purchase the full version from the ANSI store.

Contents

	Page
Foreword	v
Introduction	vi
1 Scope	1
2 Normative references	1
3 Terms and definitions	1
4 Abbreviated terms	5
5 Quality control metrics for sample preparation	5
5.1 General.....	5
5.2 Sample sequencing type.....	5
5.2.1 Sequencing type.....	5
5.2.2 Target gene.....	5
5.3 Sample information.....	5
5.3.1 Specimen type.....	5
5.3.2 Sampling date.....	6
5.3.3 Specimen origin.....	6
5.4 Summary of sample preparation related metrics.....	6
6 Quality control metrics for library preparation	6
6.1 General.....	6
6.2 DNA extraction method.....	6
6.2.1 DNA extraction kit.....	6
6.3 DNA quality.....	7
6.3.1 General.....	7
6.3.2 DNA purity.....	7
6.3.3 DNA integrity.....	7
6.4 Library construction.....	7
6.4.1 Library input amount.....	7
6.4.2 Library insert size.....	7
6.4.3 Library construction kit.....	7
6.5 Summary of library preparation related metrics.....	7
7 Quality control metrics for sequencing	8
7.1 General.....	8
7.2 Sequencing information.....	8
7.2.1 Sequencing instrument.....	8
7.2.2 Read length.....	8
7.2.3 Sequencing direction.....	8
7.2.4 Running mode.....	8
7.3 Running quality information.....	9
7.3.1 Error rate.....	9
7.3.2 Percent data quality >Q30.....	9
7.4 Summary of sequencing related metrics.....	9
8 Quality control metrics for data processing	9
8.1 General.....	9
8.2 Data quality measurement.....	9
8.2.1 Total reads.....	9
8.2.2 Mean coverage.....	9
8.2.3 Uniformity.....	10
8.2.4 Duplication rate.....	10
8.2.5 On-target rate.....	10
8.2.6 Q30 rate.....	10
8.3 Sequencing alignment.....	10
8.3.1 Mapping algorithm.....	10

This is a preview of ISO/TS 22692:2020. [Click here to purchase the full version from the ANSI store.](#)

	8.3.2	Local realignment software and version.....	10
8.4		Variant calling.....	10
	8.4.1	Variant calling software and version.....	10
	8.4.2	Variant call quality score.....	10
	8.4.3	Allelic read percentage & ratio	10
8.5		Variant filtering and annotation.....	11
	8.5.1	General.....	11
	8.5.2	Germline filter criteria.....	11
	8.5.3	Mutation and annotation database.....	11
8.6		Summary of data processing related metrics.....	11
Annex A (informative) Example layout of quality control metrics.....			12
Bibliography.....			14

This is a preview of ISO/TS 22692:2020. [Click here to purchase the full version from the ANSI store.](#)

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 215, *Health informatics*, Subcommittee SC 1, *Genomics informatics*.

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.

This is a preview of ISO/TS 22692:2020. Click [here](#) to purchase the full version from the ANSI store.

Introduction

The rapid progress in Next Generation Sequencing (NGS) technology has drastically reduced the cost and time for genomic analysis. A number of research institutions, corporations, and government agencies are competitively collecting a large volume of genomic data through multi-national, multi-institutional projects such as “DiscovEHR”^[9], “gnomAD”^[10] and “UK Biobank”^[11]. The demand for sharing of “high quality” genomic data is growing because large-scale reference data is required for reliable detection of mutation for both industrial and clinical applications.

However, the quality of available genomic data is less than desirable. To establish consistent quality control metrics, details of each stage of NGS process need to be recorded, shared and standardized (processes and data elements collected and coded for each stage and sub-stage). These processes include sample preparation, library preparation, sequencing, and data processing, among others, as shown in [Figure 1](#).

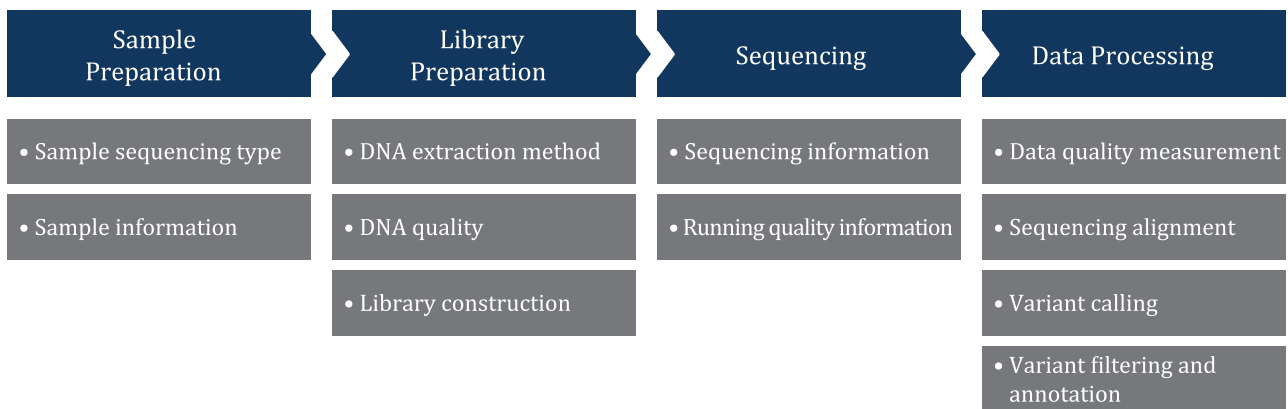


Figure 1 — NGS process