ISO/TS 23357:####(X:2023(E)

ISO TC 215/ SC 1 / WG 1

Secretariat: KATS

Genomics informatics -— Clinical genomics data sharing specification for next-generation sequencing

CD stage

Warning for WDs and CDs

This document is not an ISO International Standard. It is distributed for review and comment. It is subject to change without notice and may not be referred to as an International Standard.

Recipients of this draft are invited to submit, with their comments, notification of any relevant patent rights of which they are aware and to provide supporting documentation.

To help you, this guide on writing standards was produced by the ISO/TMB and is available at

A model manuscript of a draft International Standard (known as "The Rice Model") is available at

Date: 2023-05-04

Contents

Foreword

ISO (the International Organization for Standardization ~~(ISO)~~) is a worldwide federation of national standards bodies (ISO member bodies). The work of preparing International Standards is ~~typically conducted by~~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 ~~participate~~take part in the work. ~~The~~ 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~~document 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~~ISO draws attention to the possibility that ~~some of~~ the ~~elements~~implementation of this document may ~~be~~involve the ~~subject~~use of (a) patent(s). ISO takes no position concerning the evidence, validity or applicability of any claimed patent rights in respect thereof. As of the date of publication of this document, ISO had not received notice of (a) patent(s) which may be required to implement this document. However, implementers are cautioned that this may not represent the latest information, which may be obtained from the patent database available at www.iso.org/patents~~.~~. 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 in the ISO list of patent declarations received (see).~~

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, ~~definitions~~the meaning of ISO specific terms and expressions related to conformity assessment, as well as information about ~~the~~ISO's adherence ~~of ISO~~ to the World Trade Organization (WTO) principles in the Technical Barriers to Trade (TBT~~)~~, see ~~the following URL.~~ www.iso.org/iso/foreword.html.

Field Code Changed

This document was prepared by Technical Committee ISO/~~TC215~~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.

Introduction

Owing to the rapid advancement of next-generation sequencing technologies, the human genome is being adopted in clinical settings to realize precision medicine ~~[1].~~[7] Massive parallel sequencing or next-generation sequencing (NGS) is any of several high-throughput approaches to DNA sequencing using the concept of massively parallel processing. These technologies use miniaturized and parallelized platforms for sequencing of 1 million to 43 billion short reads (50-400 bases each) per instrument run. The data obtained in a clinical setting should be shared with another institution when patients move or shared with the patient if requested.

The clinical application steps based on clinical sequence information consist of:

~~(1)~~ a) specimen collection, processing~~,~~ and storage;

~~(2)~~ b) DNA extraction;

~~(3)~~ c) DNA processing and library preparation;

~~(4)~~ d) generation of sequence reads and base calling;

~~(5)~~ e) sequencing alignment/mapping;

~~(6)~~ f) variant calling;

~~(7)~~ g) variant annotation and filtering;

~~(8)~~ h) variant evaluation and assertion;

~~(9)~~ i) generation of test report ~~[2]~~[8].

It is required to share clinical sequencing information at a level that can reproduce the results of the institution that obtained the initial clinical sequencing information. In addition, the shared clinical genomic sequencing data should be interoperable.

This document proposes a data specification to integrate multi-layered sequencing files and related parameters and clinical data for achieving the reproducibility of genomic data in clinical practice.

This document will assist health IT companies by proposing new system requirements to deal with genomic data.

This document can be used to store and share clinical genomic data in electronic health records. In addition, it will be helpful in translational research, which requires genomic and clinical data from multiple institutes.

Genomics informatics — Clinical genomic data sharing specification for next-generation sequencing

1 ~~1.~~ Scope

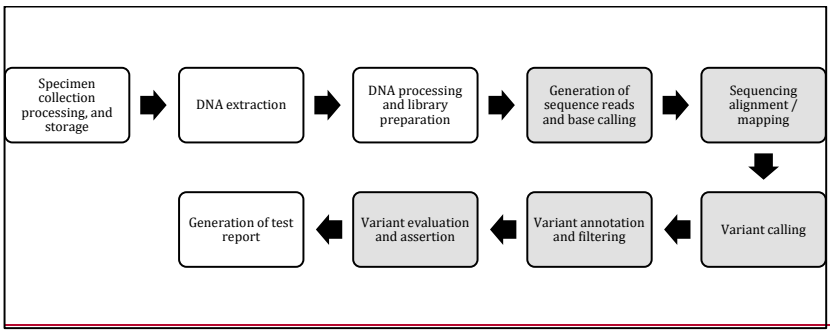
This document specifies clinical sequencing information generated by ~~massively~~massive parallel sequencing technology for sharing health information via massively parallel sequencing. This document covers the data fields and their metadata from ~~steps~~the generation of sequence reads and base calling to variant evaluation and assertion for archiving reproducibility during health information exchange of clinical sequence information. However, the specimen collection, processing~~,~~ and storage, DNA extraction~~,~~ and DNA processing and library preparation, and the generation of test report are not in the scope of this document.

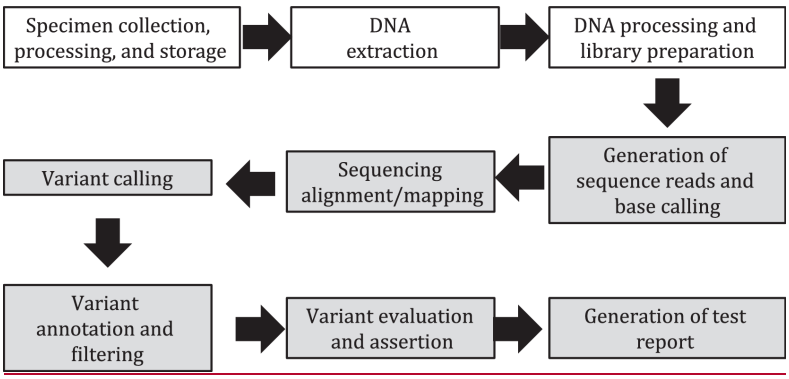
This document hence defines the data types, relationship, optionality, cardinalities~~,~~ and bindings of terminology of the data.

In essence, this document specifies:

— the required data fields and their metadata from generation of sequence reads and base calling to variant evaluation and assertion for sharing clinical genomic sequencing data files generated by massively parallel sequencing technology, as shown in Figure 1~~.~~;

— the sequencing information from human samples using DNA sequencing by massively parallel sequencing technologies for clinical practice.





NOTE The grey shaded text indicates the scope of this document.

Figure ~~The~~ 1 — Clinical application processes based on next-generation sequencing (NGS) data~~.~~
~~The gray filled box is the scope of this document.~~

2 ~~2.~~ Normative references

There are no normative references in this document.

43 ~~3.~~ Terms and definitions

For the purposes of this document, the ~~following~~ terms and definitions given in [external document reference xxx] and the following apply~~:~~.

~~The~~ ISO and ~~the~~ IEC maintain ~~terminological~~terminology 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
~~**chromosome**~~
~~structures consisting of or containing DNA, which carries the genetic information essential to the cell.~~

~~[SOURCE: ISO 19238:2014, 2.7]~~

~~**3.2**~~
clinical sequencing
next ~~-~~generation sequencing or future sequencing technologies using human samples for clinical practice and clinical trials

[SOURCE: ISO/TS 20428:2017, 3.5~~]~~

~~**3.3**~~
~~**deletion**~~

~~mutation~~, modified — "later" has been replaced with "future" in ~~which a part of a chromosome or a sequence of DNA is lost during DNA replication~~the definition.]

~~[SOURCE: ISO 20428:2017, 3.10]~~

~~3.4~~
3.2
deoxyribonucleic acid
DNA
molecule that encodes the genetic information in the nucleus of cells

[SOURCE: ISO 25720:2009, 4.7]

3.~~5~~3
DNA sequencing
~~four~~determination of the order of nucleotide bases (adenine, guanine, cytosine, and thymine) ~~are the four nucleotide bases that make up~~in a ~~DNA~~ molecule~~.~~ of *DNA* (3.4)

Note 1 to entry: Sequence is generally described from the 5' end.

[SOURCE: ISO~~/TS 178221:2014~~ 17822:2020, 3.~~20~~19]

3.~~6~~4
exome
part of the genome that corresponds to the complete complement of the exons of a cell

~~[SOURCE: ISO 20428:2017, 3.13]~~

3.7
FASTQ
text-based format for storing both the biological sequence (typically nucleotide sequence) and its corresponding quality scores

3.8
gene
category of nucleic acid sequences that functions as a unit of heredity and codes for the basic instructions for the development, reproduction, and maintenance of organisms

~~[SOURCE: ISO 11238:2012, 2.1.16]~~

3.9
germline
series of germ cells, each descended or developed from earlier cells in the series, regarded as continuing through successive generations of an organism

[SOURCE: ISO/TS 20428:2017, 3.17]

3.10
indel
insertion (3.15) or/and *deletion* (3.7)

[SOURCE: ISO/TS 20428:2017, 3.18]