FINAL DRAFT
Technical Specification

ISO/DTS 20428

Genomics Informatics — Data elements and their metadata for describing structured clinical genomic sequence information in electronic health records

Informatique génomique — Éléments de données et leurs métadonnées pour décrire les informations structurées de la séquence génomique clinique dans les dossiers de santé électroniques

ISO/TC **215**/SC **1**

Secretariat: **KATS**

Voting begins on:
2024-03-07

Voting terminates on:
2024-05-02

Reference number
ISO/DTS 20428:2024(en)

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 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).

ISO draws attention to the possibility that the implementation of this document may involve the 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.

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*.

This second edition cancels and replaces the first edition (ISO/TS 20428:2017), which has been technically revised.

The main changes are as follows:

— title was updated;

— contents were enhanced to reflect advances in bioinformatics techniques and to cover more broad clinical applications;

— terminology was refined for neural expression and elucidating content;

— Table 1 and Figure 1 were updated;

— Annex B was removed.

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

Based on the rapid advancement of sequencing technologies, clinical sequencing has been highlighted as one of methods to realize genomic medicine, personalized medicine and precision medicine. There are lots of sequencing data in the public domain with clinical information.[13] In addition, genome-scale clinical sequencing is being adopted broadly in medical practice.[14] Many hospitals have started to sequence patients' whole genome, whole exome, or targeted genes using the next-generation sequencing technologies. These genomic data obtained by next-generation sequencing technologies can be used for both clinical purposes, to diagnose patients and choose the right medications, and research purposes. Therefore, the management of genomic and clinical data are increasingly highlighted in precision medicine, clinical trial, and translational research.[15]

However, until now, there is no international standard for representing clinical sequencing results with a structured format for electronic health records. Consequently, the necessary genomic test results are not efficiently delivered to the clinicians. There are a few related standards for modelling genetic testing results (i.e. ISO 25720 and several HL7 documents from HL7 clinical genomics working group). However, these standards or drafts are mainly focused on the traditional genetic testing results for a single gene test. Based on the rapid development and adoption of next-generation sequencing techniques which can detect diverse genetic variants at the genome level, there is, therefore, still a need to develop a standard to present clinical sequencing data in such a way they become useful for clinicians.[16]

To implement a structured clinical sequencing report in electronic health records, all necessary data fields and the metadata for each chosen field should be defined. For example, it needs to be determined which vocabulary, in particular gene descriptions and/or disease codes, can be applied in particular fields. In ISO TC 215, GSVML (Genomic Sequence Variation Markup Language) was proposed for the interoperability of genomic variants, especially for single nucleotide polymorphism (SNP) data.[11] HL7 is also developing a domain analysis model for genomics using HL7 version 3[17] and fast healthcare interoperability resources (FHIR).[18] Recently, to facilitate genomic information, SMART on FHIR Genomics has been developed.[19],[20] The Clinical Data Interchange Standard Consortium (CDISC) published a study data tabulation model implementation guide: pharmacogenomics/genetics.[21] Several other international organizations, such as the Global Alliance for Genomics and Health (GA4GH), Actionable Genome Consortium, and Displaying and Integrating Genetic Information Through the EHR (DIGITizE) of the Institute of Medicine in the US, tried to develop the similar standards. The working group of the American College of Medical Genetics and Genomics Laboratory Quality Assurance Committee published the ACMG clinical laboratory standards for next-generation sequencing.[22] In addition, web-based tools become available that link genotypic information to phenotypic information, and exchanging information and using it in personalized medicine can be very helpful.[23]

In this document, to enable the standard use of patient genomic data from clinical sequencing for healthcare purposes as well as for clinical trials and research, the data elements and their metadata for a clinical sequencing report for electronic health records are developed. This document further explains how and where particular appropriate terminological systems that describe the genomes and/or diseases can be applied in these fields. By defining the necessary fields with a structured format based on coded data that adhere themselves to terminological principles such as concept representation and governance, this document can help implement clinical decision support service.

Genomics Informatics — Data elements and their metadata for describing structured clinical genomic sequence information in electronic health records

1 Scope

The document defines the data elements and the requisite metadata essential for implementing a structured clinical genomic sequencing report in electronic health records, particularly focusing on the genomic data generated by next-generation sequencing technology.

This document:

— defines the composition of a structured clinical sequencing report (see Clause 6);

— defines the required data fields and their metadata for a structured clinical sequencing report (see Clause 7);

— defines the optional data (see Clause 8);

— covers the DNA-level variation from human samples using whole genome sequencing, whole exome sequencing, and targeted sequencing (disease-targeted gene panels) by next-generation sequencing technologies (though whole transcriptome sequencing and other technologies are important to provide better patient care and enable precision medicine, this document only deals with DNA-level changes);

— covers mainly clinical applications and clinical research such as clinical trials and translational research which uses clinical data (basic research and other scientific areas are outside the scope of this document);

— does not cover the other biological species, i.e. genomes of viruses and microbes;

— does not cover the Sanger sequencing methods.

2 Normative references

There are no normative references in this document.

3 Terms and definitions

For the purposes of this document, the following terms and definitions apply.

ISO and IEC maintain 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
allele
one of several alternate forms of a *gene* (3.15) which occur at the same locus on homologous *chromosomes* (3.4) and which become separated during meiosis and can be recombined following fusion of gametes

[SOURCE: ISO 16577:2016, 3.4]

3.2
benign
benign variant
alterations with very strong evidence against pathogenicity

3.3
biomaterial
materials taken from the human body such as tissue, blood, plasma, or urine

3.4
chromosome
structure that comprises discrete packages of *DNA* (3.11) and proteins that carries genetic information which condense to form characteristically shaped bodies during nuclear division

[SOURCE: ISO 19238:2014, 2.7]

3.5
clinical sequencing
next-generation sequencing or later sequencing technologies with human samples for clinical practice and clinical trials

3.6
ClinVar
freely accessible, public archive of reports of the relationships among human *variations* (3.31) and phenotypes, with supporting evidence *variant* (3.31)

Note 1 to entry: ClinVar is available at https://www.ncbi.nlm.nih.gov/clinvar/.

3.7
copy number variation
CNV
variation (3.31) in the number of copies of one or more sections of the *DNA* (3.11)

3.8
Catalogue of Somatic Mutations in Cancer
COSMIC
online database of somatically acquired mutations found in human cancer

Note 1 to entry: COSMIC is available at http://cancer.sanger.ac.uk/cosmic.

3.9
dbSNP
database of *SNPs* (3.32) provided by the US National Center for Biotechnology Information (NCBI)

Note 1 to entry: dbSNP is available at https://www.ncbi.nlm.nih.gov/SNP/.

3.10
deletion
variant (3.31) in which a part of a *chromosome* (3.4) or a sequence of *DNA* (3.11) is lost during DNA replication

3.11
deoxyribonucleic acid
DNA
molecule that encodes genetic information in the nucleus of cells

[SOURCE: ISO 25720:2009, 4.7]

3.12
DNA sequencing
determining the order of nucleotide bases (adenine, guanine, cytosine and thymine) in a molecule of *DNA* (3.11)

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

[SOURCE: ISO/TS 17822-1:2014, 3.20]

3.13
electronic medical record
EMR
electronic health record
EHR
electronic record derived from a computerized system used primarily for delivering patient care in a clinical setting

[SOURCE: ISO/TR 24291:2021, 3.3, modified — The preferred term "electronic health record" and its abbreviation "EHR" have been added.]

3.14
exome
part of the genome formed by exons

3.15
gene
basic unit of hereditary material that encodes and controls the expression of a protein or protein subunit

[SOURCE: ISO 11238:2012, 2.1.16]

3.16
gene panel
technique for sequencing the targeted *genes* (3.15) in a genome

3.17
genomic medicine

medical discipline that involves using genomic information about an individual as part of their clinical care (e.g. for diagnostic or therapeutic decision-making) and the health outcomes and policy implications of that clinical use

3.18
germline
series of germ cells each descended or developed from earlier cells in the series, regarded as continuing through successive generations of an organism

3.19
indel
insertion (3.20) or/and *deletion* (3.10)

3.20
insertion
addition of one or more nucleotide base pairs into a *DNA* (3.11) sequence

3.21
inversion
chromosome (3.4) rearrangement in which a segment of a chromosome is reversed end to end

3.22
large indel
insertion (3.20) or *deletion* (3.10) of greater than 100 nucleotides and less than 1 000 nucleotides

3.23
likely benign
likely benign variant
alterations with strong evidence against pathogenicity

Note 1 to entry: Targeted testing of at-risk family members is not recommended.

3.24
likely pathogenic
likely pathogenic variant
alterations with strong evidence in favour of pathogenicity

3.25
pathogenic
pathogenic variant
genetic alteration that increases an individual's susceptibility or predisposition to a certain disease or disorder

[SOURCE: National Cancer Institute Dictionary of Genetic Terms]

3.26
prenatal
foetal
biomaterial (3.3) sample of foetuses before birth

Note 1 to entry: Prenatal *DNA sequencing* (3.12) is the reading of the *DNA* (3.11) of foetuses to diagnose Mendelian disease of an unborn child.

3.27
sequence read
read
fragmented nucleotide sequences which are used to reconstruct the original sequence for next generation sequencing technologies

3.28
read type
type of *sequence read* (3.27) whose format depends on the sequencing instrument

Note 1 to entry: It can be either single-end or paired-end.

Note 2 to entry: Single-end read are produced when the sequencing instrument reads from one end of a fragment to the other end.

Note 3 to entry: Paired-end reads are produced when the sequencing instrument reads from one end to the other end, and then starts another round of reading from the opposite end.

3.29
reference sequence
digital nucleic acid sequence database, assembled by scientists as a representative example of human genome

3.30
ribonucleic acid
RNA
polymer of ribonucleotides occurring in a double-stranded or single-stranded form

[SOURCE: ISO 22174:2005, 3.1.3]

3.31
sequence variation
DNA sequence variation
variation
variant
differences of *DNA* (3.11) sequence among individuals in a population

Note 1 to entry: Variant implies *CNV* (3.7), *deletion* (3.10), *insertion* (3.20), *indel* (3.19), *small indel* (3.33), *large indel* (3.22), and *SNP* (3.32).

[SOURCE: ISO 25720:2009, 4.8, modified — The preferred terms "sequence variation", "variation" and "variant" have been added; the original note has been deleted and a new Note 1 to entry has been added.]

3.32
single nucleotide polymorphism
SNP
single nucleotide *variation* (3.31) in a genetic sequence that occurs at appreciable frequency in the population

Note 1 to entry: It is pronounced "snip".

[SOURCE: ISO 25720:2009, 4.23, modified — Note 1 to entry has been added.]

3.33
small indel
insertion (3.20) or *deletion* (3.10) of 2 to 100 nucleotides

3.34
somatic variant
variant (3.31) that occurs in the cells of the body that are not germ line cells

3.35
biological specimen
biospecimen
specimen
sample of tissue, body fluid, food, or other substance that is collected or acquired to support the assessment, diagnosis, treatment, mitigation or prevention of a disease, disorder or abnormal physical state, or its symptoms

3.36
subject of care
SOC
person who uses, or is a potential user of, a health care service

[SOURCE: ISO/TS 22220:2011, 3.2, modified — The admitted term "subject of healthcare" and the Note have been removed.]

3.37
target capture
method to capture genomic regions of interest from a *DNA* (3.11) sample prior to sequencing

3.38
uncertain significance
uncertain clinical relevance
variant (3.31) with limited and/or conflicting evidence regarding pathogenicity

3.39
whole exome sequencing
WES
technique for sequencing all the protein-coding genes in a genome