Technical Specification

ISO/TS 20428

Second edition
2024-06

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

Reference number
ISO/TS 20428:2024(en)

COPYRIGHT PROTECTED DOCUMENT

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 ..... **v**
Introduction ..... **vi**
1 Scope **1**
2 Normative references **1**
3 Terms and definitions **1**
4 Abbreviated terms **5**
5 Use case scenario **6**
6 Composition of a clinical sequencing report **7**
6.1 General 7
6.2 Overall interpretation in summary 8
6.3 Detailed contents 8
7 Fields and their nomenclature of required data **9**
7.1 General 9
7.2 Clinical sequencing orders 10
7.2.1 General 10
7.2.2 Clinical sequencing order code 10
7.2.3 Date and time 10
7.2.4 Specimen information 11
7.3 Information on subject of care 11
7.3.1 General 11
7.3.2 Subject of care identifiers 11
7.3.3 Subject of care name 11
7.3.4 Subject of care birth date 11
7.3.5 Subject of care sex 11
7.3.6 Subject of care population 11
7.4 Information on legally authorized person ordering clinical sequencing 12
7.5 Performing laboratory 12
7.5.1 General 12
7.5.2 Basic information on performing laboratory 12
7.5.3 Information on report generator 12
7.5.4 Information of legally confirmed person on sequencing report 12
7.6 Associated diseases and phenotypes 12
7.7 Biomaterial information 12
7.7.1 General 12
7.7.2 Types of sample 12
7.7.3 Genomic source class in biomaterial 13
7.7.4 Conditions of specimen that can limit adequacy of testing 13
7.8 Genetic variations 13
7.8.1 General 13
7.8.2 Gene symbols and names 13
7.8.3 Sequence variation information 13
7.9 Classification of variants 14
7.9.1 General 14
7.9.2 Classification of variants based on the pathogeny 15
7.9.3 Classification of variants based on clinical relevance 15
7.10 Recommended treatment 15
7.10.1 General 15
7.10.2 Classification of variants based on clinical relevance 15
7.10.3 Clinical trial information 15
7.10.4 Known protocols related to a variant 16
7.10.5 Other recommendation 16
7.11 Addendum 16

8 Fields and their nomenclature of optional data **16**
8.1 General 16
8.2 Subject of care population 17
8.3 Medical history 17
8.4 Family history/Pedigree information 17
8.5 Reference genome version 18
8.6 Populational genomic information 18
8.7 Karyotopic sex 18
8.8 Genetic variation 18
8.9 Classification of variants 18
8.9.1 General 18
8.9.2 Classification of variants as secondary finding 18
8.10 Detailed sequencing information 18
8.10.1 General 18
8.10.2 Clinical sequencing date 18
8.10.3 Quality control metrics 19
8.10.4 Base calling information 19
8.10.5 Sequencing platform information 19
8.10.6 Analysis platform information 20
8.11 References 20
Annex A (informative) **Example structure of clinical sequencing report** **21**
Bibliography **27**

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]