TECHNICAL SPECIFICATION

ISO/TS 20428

First edition
2017-05

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

Informatique de santé — Éléments de données et leurs métadonnées pour décrire l'information structurée de la séquence génomique clinique dans les dossiers de santé électroniques

Reference number
ISO/TS 20428:2017(E)

ISO copyright office
Ch. de Blandonnet 8 • CP 401
CH-1214 Vernier, Geneva, Switzerland
Tel. +41 22 749 01 11
Fax +41 22 749 09 47
copyright@iso.org
www.iso.org

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 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 on 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 the following URL: www.iso.org/iso/foreword.html.

The committee responsible for this document is ISO/TC 215, *Health informatics*.

Introduction

Based on the rapid advancement of sequencing technologies, clinical sequencing has been highlighted as one of methods to realize personalized medicine and precision medicine. There are lots of sequencing data in the public domain with clinical information[1]. In addition, genome-scale clinical sequencing is being adopted broadly in medical practice[2]. 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 is increasingly highlighted to precision medicine, clinical trial and translational research[3].

However, until now, there is no international standard for representing clinical sequencing results with a structured format for electronic health records, in consequence, 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 in 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[4].

To implement a structured clinical sequencing report in electronic health records, all necessary data fields should be defined 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 interoperability of genomic variants, especially for single nucleotide polymorphism (SNP) data[5]. HL7 is also developing a domain analysis model for genomics using HL7 version 3[6] and fast healthcare interoperability resources (FHIR)[7]. Recently, to facilitate genomic information, SMART on FHIR Genomics has been developed[8],[9]. The Clinical Data Interchange Standard Consortium (CDISC) published a study data tabulation model implementation guide: pharmacogenomics/genetics[10]. Several other international organizations such as Global Alliance for Genomics and Health (GA4GH), Actionable Genome Consortium, and Displaying and Integrating Genetic Information Through the EHR (DIGITizE) of Institute of Medicine in 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[11]. 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[12].

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 metadata for a clinical sequencing report for electronic health records will be developed. It 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 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.

Health 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 their necessary metadata to implement a structured clinical genomic sequencing report and their metadata 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 5),

— defines the required data fields and their metadata for a structured clinical sequencing report (see Clause 6),

— defines the optional data (see Clause 7),

— 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. However, the necessary steps such as de-identification or consent from patient should be applied. The 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, and

— 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 terminological databases for use in standardization at the following addresses:

— ISO Online browsing platform: available at http://www.iso.org/obp

— IEC Electropedia: available at http://www.electropedia.org/

3.1
allele
one of several alternate forms of a gene which occur at the same locus on homologous chromosomes and which become separated during meiosis and can be recombined following fusion of gametes

[SOURCE: ISO 16577:2016, 3.6]

3.2
benign
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 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 and phenotypes, with supporting evidence

Note 1 to entry: http://www.ncbi.nlm.nih.gov/clinvar/.

3.7
copy number variation
CNV
variation in the number of copies of one or more sections of the DNA

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

Note 1 to entry: http://cancer.sanger.ac.uk/cosmic.

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

Note 1 to entry: https://www.ncbi.nlm.nih.gov/SNP/.

3.10
deletion
mutation in which a part of a chromosome or a sequence of DNA 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

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

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

3.13
exome
part of the genome formed by exons

3.14
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]

3.15
gene panel
technique for sequencing the targeted genes in a genome

3.16
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

3.17
germline
series of germ cells each descended or developed from earlier cells in the series, regarded as continuing through successive generations of an organism

3.18
indel
insertion (3.19) or/and *deletion* (3.10)

3.19
insertion
addition of one or more nucleotide base pairs into a DNA sequence

3.20
inversion
chromosome rearrangement in which a segment of a chromosome is reversed end to end

3.21
large indel
insertion or deletion up to ~1 kb

3.22
likely benign
alterations with strong evidence against pathogenicity

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

3.23
likely pathogenic
alterations with strong evidence in favor of pathogenicity

3.24
pathogenic
characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes, or pharmacologic responses to a therapeutic intervention

3.25
prenatal/fetal
biomaterial sample of fetuses before birth

Note 1 to entry: Prenatal/fetal DNA sequencing: Reading the DNA of foetuses to diagnose Mendelian disease of unborn child.

3.26
sequence read
read
fragmented nucleotide sequences which are used to reconstruct the original sequence for next generation sequencing technologies

3.27
read type
type of run in the sequencing instrument

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

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

Note 3 to entry: Paired-end: Paired end runs read from one end to the other end, and then start another round of reading from the opposite end.

3.28
reference sequence
digital nucleic acid sequence database, assembled by scientists as a representative example of human genome

3.29
ribonucleic acid
RNA
polymer of ribonucleotides occurring in a double-stranded or single-stranded form

[SOURCE: ISO 22174:2005, 3.1.3]

3.30
sequence variation
DNA sequence variation
variation
differences of DNA sequence among individuals in a population

Note 1 to entry: Variation implies *CNV* (3.7), *deletion* (3.10), *insertion* (3.19), *indel* (3.18), *small indel* (3.32), *large indel* (3.20), and *SNP* (3.31).

[SOURCE: ISO 25720:2009, 4.8]

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

Note 1 to entry: Pronounced “snip”.

[SOURCE: ISO 25720:2009, 4.23]

3.32
small indel
insertion or deletion of 2 ~100 nucleotides

3.33
somatic cell
cells of the body in contrast to the germ line cells

3.34
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

3.35
subject of care
any person who uses, or is a potential user of, a health care service

[SOURCE: ISO/TS 22220:2011, 3.2]

3.36
target capture
method to capture genomic regions of interest from a DNA sample prior to sequencing

3.37
uncertain significance
alterations with limited and/or conflicting evidence regarding pathogenicity

3.38
whole exome sequencing
WES
technique for sequencing all the protein-coding genes in a genome

3.39
whole genome sequencing
WGS
technique that determines the complete DNA sequence of an organism's genome at a single time

4 Abbreviated terms

This list of abbreviated terms includes all abbreviations used in this document.

ACMG	the American College of Medical Genetics and Genomics
COSMIC	the Catalogue of Somatic Mutations in Cancer
CPIC	the Clinical Pharmacogenetics Implementation Consortium
EBI	the European Bioinformatics Institute
FHIR	Fast Healthcare Interoperability Resources
HGNC	the HUGO Gene Nomenclature Committee
HGVS	the Human Genome Variation Society
HUGO	the Human Genome Organization
IARC	International Agency for Research on Cancer
LOINC	Logical Observation Identifiers Names and Codes

NCBI	National Center for Biotechnology Information
NCCN	National Comprehensive Cancer Network
NGS	Next Generation Sequencing
SNP	Single Nucleotide Polymorphism
SPREC	Standard Preanalytical Code
WES	Whole Exome Sequencing
WGS	Whole Genome Sequencing

5 Use case scenario

The abstracted use case for generating a clinical genomic sequencing report is demonstrated in Figure 1. At first, the clinician will place a clinical sequencing order using the electronic health records system (step 1 in Figure 1). After the order, a responsible department will request DNA sequencing to the sequencing facility (step 2). This sequencing facility can be located inside of the hospital or be an independent sequencing facility outside the hospital (step 3). When confirming the order, the sequencing facility will request a sample from the patient (step 4). The hospital will collect a sample from the patient (steps 5 and 6). The pre-collected samples, i.e. biobank sample, the samples acquired by a previous laboratory or pathology orders, can be used as well. The biomaterial from the patients will be delivered to the sequencing facility (step 7). After receipt, the sequencing facility will perform a sequencing analysis (step 8) and generate the report (step 9). This report will be sent to the requested hospital (step 10), and the report will be updated in the electronic health record system (step 11). The ordering clinician will be notified of the completion of the sequencing order (step 12). Finally, the ordering clinician will make a diagnosis or give a proper treatment (step 13). A patient can have a copy of final report.