TECHNICAL SPECIFICATION

ISO/TS 22693

First edition
2021-06

Genomics informatics — Structured clinical gene fusion report in electronic health records

Informatique génomique — Rapport de fusion de gènes clinique structuré pour les dossiers de santé électroniques

Reference number
ISO/TS 22693:2021(E)

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 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.

Introduction

With the rapid advancement of next generation sequencing technologies, clinical sequencing has been applied to realize precision medicine. ISO/TS 20428[1] aims at standardizing the clinical sequencing reports in electronic health records but focuses on only DNA variations. However, the importance of transcriptome information has been increased. The transcriptome is the complete set of all messenger RNA molecules, which encode for the amino acid sequence of proteins. RNA sequencing gives us a large amount of information on gene expression and RNA alterations in disease status. From a molecular diagnostic standpoint, RNA-based measurements have the potential for broad application across diverse areas of human health, including disease diagnosis, prognosis, and therapeutic selection.

A fusion gene is a hybrid gene made by the combination of two or more genes that had previously existed independently. It is known to occur due to structural abnormalities of chromosomes such as insertion, deletion, translocation, and inversion. Fluorescence in situ hybridization (FISH) has been used as a gold standard as a method of detecting gene fusion in clinical practice but advances in technology have enabled RNA-based detection of fusion genes that directly affect protein coding. One of the most widely applied RNA-based technologies is qRT-PCR (Quantitative Reverse Transcription-Polymerase Chain Reaction). The relatively inexpensive NGS (Next generation sequencing) method is actively used in clinical practice as it detects many genes at once. There are DNA-based and RNA-based methods for detecting fusion using NGS, but it is recognized that using RNA-based is more accurate in terms of detection sensitivity.

Technological advancements have continually shaped the way that RNA-based (transcriptome) measurements are used in the clinic. There are several commercially available RNA-based clinical tests.[2] In order to complement ISO/TS 20428, the RNA sequencing report is necessary. Among driver RNA sequencing results, the most prevalent gene fusion was chosen as the first step. This document will aid in developing other clinical RNA sequencing or whole transcriptome sequencing reports.

In this document, the data elements and their standardized metadata for gene fusion report using RNA sequencing in electronic health records will be described. A structured clinical report for the fusion gene will provide pertinent information on bioinformatics analysis to help clinical decisions.

Genomics informatics — Structured clinical gene fusion report in electronic health records

1 Scope

The document defines the data elements and their necessary metadata to implement a structured clinical gene fusion report whose data are generated by next generation sequencing technologies.

This document

— describes the reporting guideline for RNA sequencing approaches focusing on detecting novel and known fusion partners,

— defines the required data fields and their metadata for a structured clinical gene fusion report,

— defines the optional data fields and their metadata,

— covers the fusion gene from human specimen using whole transcriptome sequencing by next generation sequencing technologies for clinical practice and translational research,

— does not cover the fusion gene detection using DNA sequencing methods,

— does not cover the basic research and other scientific areas,

— does not cover the other biological species,

— does not cover the Sanger sequencing methods, and

— does not cover the other structural variations.

This document only defines the data elements and their metadata for the structured clinical sequencing report in electronic health records. Therefore, its layout can be designed based on the institutional decision if all elements are included as in this document.

2 Normative references

The following documents are referred to in the text in such a way that some or all of their content constitutes requirements of this document. For dated references, only the edition cited applies. For undated references, the latest edition of the referenced document (including any amendments) applies.

ISO 8601 (all parts), *Date and time — Representations for information interchange*

ISO 20397-2:2021, *Biotechnology — Massively parallel sequencing — Part 2: Quality evaluation of sequencing data*

ISO/TS 22220:2011, *Health informatics — Identification of subjects of health care*

ISO/TS 22692:2020, *Genomics informatics— Quality control metrics for DNA sequencing*

ISO/TS 27527:2010, *Health informatics — Provider identification*

HGNC:Bruford E.A., Braschi B., Denny P. et al. , *Guidelines for human gene nomenclature*. Nat Genet **52**, 754–758 (2020). https://doi.org/10.1038/s41588-020-0669-3

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 https://www.iso.org/obp

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

3.1
benign
alterations with very strong evidence against pathogenicity

[SOURCE: ISO/TS 20428:2017, 3.2]

3.2
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

[SOURCE: ISO/TS 20428:2017, 3.34]

3.3
chromosome
structure that comprises discrete packages of *DNA* (3.5) and proteins that carries genetic information which condense to form characteristically shaped bodies during nuclear division

[SOURCE: ISO 19238:2014, 2.7]

3.4
clinical sequencing
next generation sequencing or later sequencing technologies with human *specimens* (3.2) for clinical practice and clinical trials

[SOURCE: ISO/TS 20428:2017, 3.5, modified — "samples" was changed to "specimens".]

3.5
deoxyribonucleic acid
DNA
molecule that encodes genetic information in the nucleus of cells

[SOURCE: ISO 25720:2009, 4.7]

3.6
fusion gene
gene (3.7) that is made by joining parts of two different genes that can occur naturally in the genome by transferring *DNA* (3.5) between *chromosomes* (3.3)

3.7
gene
basic unit of hereditary material that encodes and controls the expression of a protein or protein subunit

3.8
gene fusion
genetic recombination of the parts of two or more *genes* (3.7) resulting in a gene with different or additional regulatory regions, or a new chimeric gene product

3.9
likely benign
alterations with strong evidence against pathogenicity

[SOURCE: ISO/TS 20428:2017, 3.22, modified — Note 1 to entry removed.]

3.10
likely pathogenic
alterations with strong evidence in favour of pathogenicity

[SOURCE: ISO/TS 20428:2017, 3.23]

3.11
pathogenic
characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes, or pharmacologic responses to a therapeutic intervention

[SOURCE: ISO/TS 20428:2017, 3.24]

3.12
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-end read runs the sequencing instrument reads from one end of a fragment to the other end.

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

[SOURCE: ISO/TS 20428:2017, 3.27]

3.13
sequencing read
read
fragmented nucleotide sequences that are used to reconstruct the original sequence for next generation sequencing technologies

[SOURCE: ISO/TS 20428:2017, 3.26]

3.14
reference sequence
nucleic acid sequence with biological relevance

Note 1 to entry: Each reference sequence is indexed by a one-dimensional integer coordinate system whereby each integer within range identifies a single nucleotide. Coordinate values can only be equal to or larger than zero. The coordinate system in the context of this document is zero-based (i.e., the first nucleotide has coordinate 0 and it is said to be at position 0) and linearly increasing within the string from left to right.

[SOURCE: ISO/IEC 23092-1:2020, 3.22]

3.15
ribonucleic acid
RNA
polymer of ribonucleotides occurring in a double-stranded or single-stranded form

[SOURCE: ISO 22174:2005, 3.1.3]

3.16
RNA sequencing
RNA-seq
technique that determines the complete or partial *RNA* (3.15) sequence of an organism's genome

3.17
subject of care
person who uses, or is a potential user of, a health care service

[SOURCE: ISO/TS 22220:2011, 3.2, modified — Admitted term and Note 1 to entry removed.]

3.18
target capture
method to capture genomic regions of interest from a *DNA* (3.5) *specimen* (3.2) prior to sequencing

[SOURCE: ISO/TS 20428:2017, 3.36, modified — "sample" was changed to "specimen."]

3.19
targeted RNA sequencing
technique that determines the *RNA* (3.15) sequence of interest in an organism's genome

3.20
variant of unknown significance
VUS
variation in a genetic sequence for which the association with disease risk is unclear

3.21
whole transcriptome sequencing
technique that determines the complete *RNA* (3.15) sequence of an organism's genome at a single time

4 Abbreviated terms

ACMG	American College of Medical Genetics and Genomics
COSMIC	Catalogue of Somatic Mutations in Cancer
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
NCBI	National Center for Biotechnology Information
NCCN	National Comprehensive Cancer Network
NGS	Next Generation Sequencing
SPREC	Standard Preanalytical Code
WHO	World Health Organization

5 Gene fusion

Gene fusion is a widespread phenomenon and "has been observed across all domains of life. Comparative genomics studies reveal high and persistent incidence of gene fusions and identify lineage-specific factors that promote or hinder the formation of chimeric genes. Studies of recent gene fusions expose the mechanisms of their origin and the diversity of functional changes that accompany their formation. Gene fusions prominently contribute to evolutionary change by providing a continuous source of new genes. Gene duplications often precede gene fusions, permitting the evolution of chimeric genes, but at the same time preserving the original functions. Despite the reputation of gene fusions as drivers of