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

adaptive evolution, gene fusions can have devastating consequences, often leading to genomic disorders or cancer."[8]

The clinical use cases of gene fusion are well explained in NCCN guidelines.[9] For example, Reference [9] demonstrates how the identification of a fusion gene (ALK rearrangement) can guide therapeutic decision. Several commercially RNA-based tests are available, but most of them use microarray or qRT-PCR (Quantitative Reverse Transcription-Polymerase Chain Reaction). The first commercially available RNA-based test using NGS technology was the gene fusion report.

6 Composition of clinical gene fusion report

6.1 General

The structured RNA fusion report may mainly consist of two parts, as in ISO/TS 20428: the summary section and the detailed section. As defined in ISO/TS 20428, the summary section should include the subset of required fields to help clinicians quickly overview the most significant findings. The detailed section should contain all required fields (see Clause 7) and the selected optional fields (see Clause 8).

6.2 Summary section

The summary section should report the overall interpretation of a test based on clinical implication. It should contain identified gene fusion information. The additional clinical significance with the relevant therapies can be included. In addition, the summary section can include the subset of the detailed section, as necessary.

6.3 Detailed section

The detailed section should contain all the required fields and the selected optional fields. The required fields mainly focus on helping clinicians by providing the necessary gene fusion information, interpretation results, and related treatments. They include all necessary fields for clinical practice. The information that can be only described in the clinical sequencing report is included in the required fields to minimize the length of clinical sequencing report. The optional fields provide more detailed information to clinicians. They can also facilitate translational research with the necessary steps such as de-identification or consents from the subject of care.

Annex A demonstrates the composition of the clinical gene fusion report by exemplifying the informative values.

7 Fields and their nomenclature of required data

7.1 General

Data elements, their metadata, and cardinality are summarized in Table 1. The fields are chosen for mainly clinical practice using gene fusion reports from RNA sequencing or whole transcriptome sequencing. Cardinality represents that this data element shall appear once in the report (One) or multiple times in the report (Many). Table A.1 demonstrates the example of required fields.

Table 1 — Data elements, their metadata, and cardinality for required fields

Data elements			Metadata (Primary)	Cardinality
Clinical sequencing order	Clinical sequencing order code	Order code	Institutional Coding System	One
		Information on sequencing order	TEXT	One
	Date and time	Order date	ISO 8601 (all parts)	One
		Specimen collection date		One
		Order received date		One
		Report date		One
		Addendum creation date		Many
Information on subject of care	Identifier		ISO/TS 22220:2011	One
	Name			One
	Birth date		ISO 8601 (all parts)	One
	Sex		ISO/TS 22220:2011	One
	Ancestry		HL7® v3 Code System Ethnicity	One
	Referring diagnosis		ICD	One
Information of legally authorized person ordering clinical sequencing			ISO/TS 27527:2010	One
Performing laboratory	Basic information		TEXT	One
	Information of report generator		TEXT	One
	Information of legally confirmed person on sequencing report		ISO/TS 27527:2010	One
Specimen information	Type of specimen		SPREC	One
Fusion gene information	Gene and its partner gene name		HGNC	Many
	Chromosome			Many
	Break point		NUMERIC	Many
	Validation		ENUM (“RT-PCR confirmed”, “Not confirmed”) [a]	Many
Classification of variants	Tier		ENUM (“Tier 1”, “Tier 2”, “Tier 3”, “Tier 4”)[a]	Many
Recommended treatment	Medication		ISO 11615	Many
	Clinical trial information		Clinical trial ID	Many
	Other recommendation		TEXT	Many
	Supporting information		TEXT	Many

[a] ENUM represents that the contents shall be chosen among the given categories.

7.2 Clinical sequencing order

7.2.1 General

The clinical sequencing order consists of clinical sequencing order code (see 7.2.2) and its ordering date and time (see 7.2.3).

7.2.2 Clinical sequencing order code

7.2.2.1 Order code

The relevant clinical sequencing orders should be represented in accordance with relevant international, national or institutional coding systems.

7.2.2.2 Information on sequencing order

Since the order code cannot fully describe the purpose of the clinical sequencing to detect gene fusion, the detailed description of the order can be given as free text.

7.2.3 Date and time

7.2.3.1 General

All date and time in the report shall be represented in accordance with by ISO 8601 (all parts). ISO 8601 (all parts) remove unambiguity on day-date conventions.

EXAMPLE Date represented as 'YYYY-MM-DD.'

7.2.3.2 Order date

The order date is the date when a clinician ordered the necessary sequencing test to detect gene fusion.

7.2.3.3 Order received date

The order received date is the date when the performing laboratory received and confirmed the clinical sequencing order. Order date (see 7.2.3.2) and order received date could be different.

7.2.3.4 Specimen collection date

Specimen collection date is the date when a specimen is taken from subject of care or tissue. Since the specimen could be collected in diverse ways such as surgery and biopsy in the different days, the date shall be reported.

7.2.3.5 Report date

The report date is the date when the performing laboratory generates the requested gene fusion report. It might be divided into report generation date and confirmation date of legally confirmed by the authorized personnel.

7.2.3.6 Addendum creation date

The addendum creation date is the date when the performing laboratory creates the addendum of the previous report based on up-to-date information or re-analysis. The performing laboratory shall create the addendum of the existing sequencing report based on the clinician's request or law enforcement.

If there is no addendum, this field can be omitted.

7.3 Information on subject of care

7.3.1 General

The patient (subject of care) information shall be represented in accordance with ISO/TS 22220:2011. ISO/TS 22220:2011 indicates the data elements and structure suited to accurate and procedurally appropriate and sensitive identification of individuals in health care supported by computer technology,

or through interactions between computer systems. It provides guidelines for improving the positive identification of subjects of care within and between health care organizations[11].

7.3.2 Subject of care identifier

Subject of care identifier indicates the unique identifier of patient or recipient of care.

7.3.3 Subject of care name

Subject of care name indicates the name of patient or recipient of care.

7.3.4 Subject of care birth date

The birth date shall be represented in accordance with ISO 8601 (all parts) as in 7.2.3.

7.3.5 Subject of care sex

The subject of care's sex shall be represented in accordance with ISO/TS 22220:2011, 7.4.

EXAMPLE Male represented by '1' or 'M'.

7.3.6 Subject of care ancestry

The ancestry of the subject of care shall be notified to represent his or her genetic origin. The ancestry information should be represented in accordance with the HL7®[1] v3 Code System Race (https://www.hl7.org/fhir/v3/Race/cs.html). Alternatively, if there are national standards, those coding systems can be used, for example, HL7® FHIR US Core Implementation Guide, Race & Ethnicity – CDC (https://www.hl7.org/fhir/us/core/CodeSystem-cdcrec.html).

7.3.7 Referring diagnosis

The referring diagnosis indicates the existing or suspicious diagnosis of the subject of care. Based on this information and the sequencing results, the report can include the relevant therapies to the subject of care.

7.4 Information on legally authorized person ordering clinical sequencing

7.4.1 General

Information on legally authorized person who ordered gene fusion shall be represented in accordance with ISO/TS 27527:2010. ISO/TS 27527:2010 provides a framework for improving the positive identification of providers.

The detailed items of this field, such as the name of ordering physician, his/her medical specialty, or contact numbers, can be chosen by the implementing hospitals or laboratories.

7.5 Performing laboratory

7.5.1 General

This field includes the information laboratory that performs the sequencing.

1) HL7® is a trademark of a product supplied by Health Level Seven International. This information is given for the convenience of users of this document and does not constitute an endorsement by ISO of the product named. Equivalent products may be used if they can be shown to lead to the same result.

7.5.2 Basic information on performing laboratory

The name of the performing laboratory and contact points such as phone numbers, emails, or address shall be indicated. This information can be given in free text.

7.5.3 Information on report generator

Information on subject of provider who generated a report of sequencing results shall be represented in accordance with ISO/TS 27527:2010 or by free text.

7.5.4 Information of legally confirmed person on sequencing report

Information of legally confirmed physician shall be given in accordance with ISO/TS 27527:2010, since the qualified physician shall confirm the contents of the final report.

7.6 Biospecimen information

7.6.1 General

This field describes the information on the specimen or biological specimen from subject of care.

7.6.2 Type of specimen

Types of specimen can be represented in accordance with the SPREC of the International Society for Biological and Environmental Repositories.[13] Currently, SPREC Version 3.0 is the up-to-date version.

EXAMPLE BLD (Blood), BUF (buffy coated), non-blood tissue (CEN), semen (SEM).

7.7 Fusion gene information

7.7.1 General

The information on gene fusion shall be described using HGNC or transcript ID (https://asia.ensembl.org/Help/View?id=151).

7.7.2 Gene and its partner gene information

The gene symbols and names of the target gene and its partner gene to form a fusion gene shall be represented in accordance with HGNC. The HGNC approved gene symbol and HGNC ID can be used. The symbol is case-insensitive. If a suitable gene is not in the list of HGNC, transcript ID from NCBI or Ensemble can be used.

The target gene is the gene located in 5' position of the fusion gene. The partner gene is located in 3' position of the fusion gene.

7.7.3 Chromosome information

The chromosome number of each gene shall be given using the notation of HGNC.

7.7.4 Breakpoints information

Gene position and/or exon number at the breakpoint shall be given for both involved genes.

7.8 Classification of variants

The classification of variants can follow A Joint Consensus Recommendation of the Association for Molecular Pathology, American Society of Clinical Oncology, and College of American Pathologists [14].

Though this recommendation focuses on cancer, it is one of the most up-to-date and reliable recommendation for other diseases.

The recommendation proposes the four-tiered system to categorize variations based on their clinical significances: tier I, variants with strong clinical significance; tier II, variants with potential clinical significance; tier III, variants of unknown clinical significance; and tier IV, variants deemed benign or likely benign. This recommendation also covers the ACMG guideline, which categorizes the variations into five categories including pathogenic, likely pathogenic, uncertain significance, likely benign, and benign.

However, if institution or country has its own classification of variants, its own classification may be used with references such as guidelines or journal articles.

7.9 Recommend treatment

7.9.1 General

The recommend treatment such as medication, clinical guidelines, or clinical trials can be reported to help clinicians. If the found gene fusion is included in the established clinical guidelines, clinical trials, or academic articles, the relevant therapy shall be provided with the supporting information.

7.9.2 Medication

The associated medication can be represented in accordance with MPID (Medicinal Product IDentifier) or IMPID (Investigational MPID) of ISO 11615 (IDMP, Identification of Medicinal Product) Standard. Alternatively, the Anatomical Therapeutic Chemical (ATC) classification, maintained by the WHO Collaborating Centre for Drug Statistics Methodology, can be used to represent the classification of active ingredients of drugs. International Nonproprietary Names (INN) it can be alternatively used as a drug name, since INN facilitates the identification of pharmaceutical substances or active pharmaceutical ingredients. National or local standards can also be used.

7.9.3 Clinical trial information

The clinical trial information can be given to help clinicians. It should be represented in accordance with UTN (Universal Trial Number) used in WHO International Clinical Trials Registry Platform. Alternatively, US ClinicalTrials.gov ID, EU EudraCT number, or another domestic registry ID can be used.

The detailed clinical trial information, which helps the clinicians, can be included in the report.

7.9.4 Other recommendation

The other helpful information with supporting information can be given. For example, the experimental results from academic papers can be given.

7.9.5 Supporting information

The supporting information on the provided treatment can be described using free text.

EXAMPLE NCCN guideline version and relevant information, clinical trial phase or other information.

8 Fields and their nomenclature of optional data

8.1 General

The data elements for optional fields may be used to understand the RNA sequencing process and the interpretation results. These fields can also be applied to clinical trial and translational research. The fields that are not listed in this document can be used based on the institutional decision.

Table 2 summarizes the optional data fields, their nomenclature, and cardinality. Cardinality represents that this data element shall appear once in the report (One) or multiple times in the report (Many). Table A.2 demonstrates the example of optional fields.

Table 2 — Data elements and their metadata for optional fields

Data elements			Metadata (Primary)	Cardinality
Associated diseases and phenotypes			ICD	Many
Reference genome version			Genome Reference Consortium Human Genome release ID	One
Genomic information related to race			TEXT	Many
Fusion gene information	Fusion gene ID		Relevant database ID	Many
	HGVS version		TEXT	Many
	Fusion transcript image		IMAGE	Many
	Read counts for evidence reads (or supporting reads)		NUMERIC	Many
	Count of total fusion events			Many
	Type of fusion		HGVS	Many
	Validation		ENUM (“RT-PCR confirmed”, “Not confirmed”) [a]	Many
Detailed sequencing information	Clinical sequencing date		ISO 8601 (all parts)	One
	Quality control metrics		TEXT	Many
	Sequencing platform information	Type of sequencers	TEXT	One
		Library preparation methods		One
		Target capture methods (if used)		One
		Read type	ENUM (“single-end”, “paired-end”) [a]	One
		Read length	NUMERIC	One
	Analysis platform information	Alignment tools	TEXT	Many
		Fusion gene caller		Many
		Annotation tools and databases		Many
References			TEXT	Many

[a] ENUM represents that the contents shall be chosen among the given category.

8.2 Associated disease and phenotypes

If possible, associated diseases and phenotypes of the detect fusion gene should be included using ICD codes.[11] Other phenotypes that are not classified by ICD can be represented in accordance with

SNOMED-CT®[2], Human Phenotype Ontology (http://human-phenotype-ontology.github.io/) or other well-known ontologies.

Since the detect fusion gene can have multiple associated phenotypes including diseases, the reported diseases or phenotypes should be listed based on the confidence level of the chosen database.

In this field, only diseases or phenotypes that are associated with the found variants should be indicated. The previously known diseases of the subject of care, which are not associated with the variants, should be excluded in this field.

8.3 Reference genome version

Reference sequences are the baseline from which variation is reported. If different reference sequences are used, the variant calls are also different. The reference sequence should be represented in accordance with Genome Reference Consortium Human Genome release ID or Locus Reference Genomic ID.

Reference sequences should be represented in accordance with Genome Reference Consortium Human Genome release ID (https://www.ncbi.nlm.nih.gov/projects/genome/guide/human/index.shtml). If there is an update, the patch number should be appended.

EXAMPLE GRCh38.p13 (GRCh38 Patch Release 13).

LRG (Locus Reference Genomic) ID (http://www.lrg-sequence.org/), which is maintained by EBI, can be used as well. RefSeqGene (https://www.ncbi.nlm.nih.gov/refseq/rsg/) in NCBI is part of the LRG project.

EXAMPLE LRG_1.

8.4 Genomic information related to race

When genomic information related to race was reported, the reference data set for racial information should be notified using a well-known database such as HapMap (http://hapmap.ncbi.nlm.nih.gov/), 1000 Genomes (http://www.1000genomes.org/), Genome Aggregation Database (gnomAD) (https://gnomad.broadinstitute.org/), or other databases which can give racial information.

When using the human genetic database, the certified public database, i.e., FDA-recognized database, should be firstly applied[1].

8.5 Fusion gene information

The fusion gene information, which is not reported in the required section due to VUS or no recommended treatment, can be listed in the optional section. The information should be reported by 7.7. These fields in the optional section include more diverse information for clinicians or researchers that is not included in the required section.

8.5.1 Fusion gene ID

This field indicates the identifier of the database, which contains fusion genes, i.e. COMSIC fusion ID.

EXAMPLE COSF1271.

2) SNOMED-CT® is a trademark of a product supplied by the International Health Terminology Standards Development Organisation. This information is given for the convenience of users of this document and does not constitute an endorsement by ISO of the product named. Equivalent products may be used if they can be shown to lead to the same result.

8.5.2 HGVS version

HGVS version might be notified using free text[12].

EXAMPLE HGVS v19.01 or previous version.

8.5.3 Fusion transcript image

To give a graphical explanation, an abstracted image of detected fusion gene might be included as in Figure 1. Figure 1 is the example images from Figure 4 of Reference [5]. Any kind of informative image could be used.

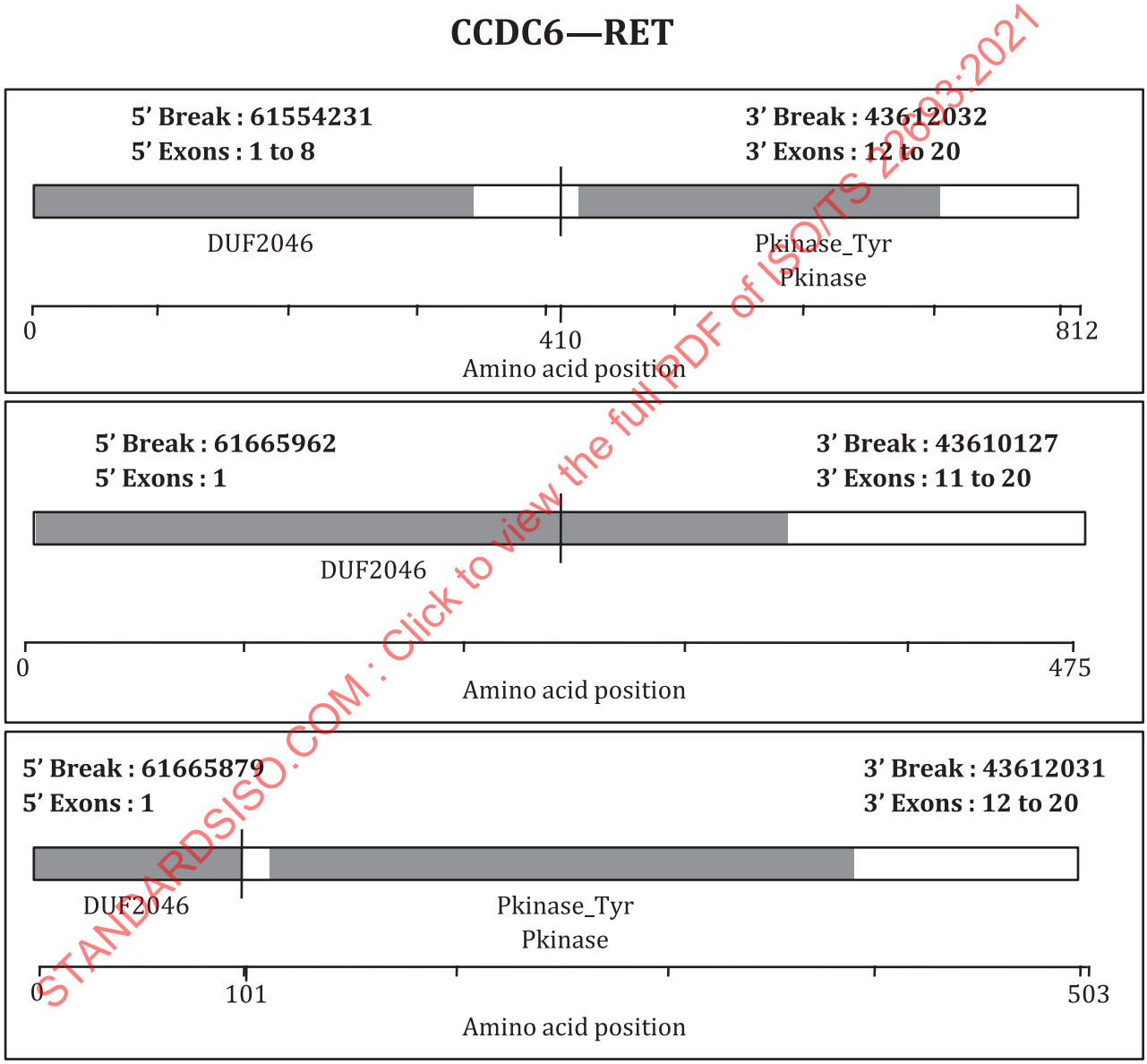

Figure 1 — Fusion transcript example image (Reproduced from Reference [5])

8.5.4 Read counts for evidence reads (or supporting reads)

This field indicates the number of the evidence reads (sequencing read) to support detecting the breakpoint of the fusion gene. The counts of the evidence reads or supporting reads can be reported in the optional field. This read count indicates the confidentiality of the reported gene fusion.

8.5.5 Count of total fusion events

This field represents the number of total fusion events in the single clinical sequencing order.

8.5.6 Type of fusion

Types of chromosome rearrangement causing gene fusion might be included. The type of fusion could be inversion, deletion, tandem duplication, or translocation. This information should be reported using HGVS.

8.5.7 Validation using standard method (RT-PCR)

The detected gene fusion should be validated using the current standard method, such as RT-PCR (Reverse Transcription-Polymerase Chain Reaction). The value of this field might be chosen between "RT-PCR confirmed" and "Not confirmed".

8.6 Detailed sequencing information

8.6.1 Clinical sequencing date

The clinical sequencing date is the date when the performing laboratory generates the fusion gene results using the received specimen. The date should be represented in accordance with ISO 8601 (all parts) as other date information.

8.6.2 Quality control metrics

The necessary quality control information can be included in these fields. ISO/TS 22692:2020 or ISO 20397-2:2021 shall be used to indicate the quality of sequencing results. ISO 20397-2:2021 specifies general requirements and recommendations for quality assessments and control of next generation sequencing (NGS) data. It covers post raw data generation procedures, sequencing alignments, and variant calling. ISO/TS 22692:2020 identifies quality metrics for the detection of DNA variants using NGS technology. It also defines the data types, relationships, optionality, cardinalities and terminology bindings of the data.

8.6.3 Sequencing platform information

8.6.3.1 General

Information on sequencing techniques and data, including sequencing platform, capture method, and alignment algorithm, should be given in the text.

8.6.3.2 Type of sequencers

The specific sequencer that performs the sequencing might be given in fee text.

EXAMPLE Illumina MiSeq, NextSeq 550, NoVaSeq 6000, X10, Thermo Fisher Ion Torrent, PacBio Sequel.

8.6.3.3 Library preparation methods

The sequencing library preparation methods might be given in free text.

EXAMPLE Sequencer compatible library kit (Illumina TruSeq RNA library kit, Ion Total RNA-Seq Kit, ArcherDx fusion flex series, or any kinds of used RNA seq library prep kits).

8.6.3.4 Target capture methods

The exome or targeted region capture methods for targeted RNA sequencing should be notified.

EXAMPLE Amplicon, probe capture or none.

8.6.3.5 Read type

Sequencing read type might be given using free text.

EXAMPLE Single-end, Paired-end.

8.6.3.6 Read length

The sequencing read length information might be given using free text

EXAMPLE 101 bp, 35-250 bp.

8.6.4 Analysis platform information

8.6.4.1 General

The primary, secondary (if applicable tertiary) analysis pipelines might be included in free text. The parameter setting for each pipeline should also be mentioned to confirm the reliability of the variant call and reproducibility.

EXAMPLE GATK 4.1.8.0.

8.6.4.2 Alignment tools

The name of the alignment tool and its version might be notified in free text.

EXAMPLE BWA-MEM 0.7.17.

8.6.4.3 Fusion gene caller

The name of the fusion gene caller and its version might be notified in free text.

EXAMPLE defuse v0.8.0, TopHat-Fusion, STAR-Fusion.

8.6.4.4 Annotation tools and databases

The name of the annotation tools and source of databases that is publicly available or private might be reported in free text.

EXAMPLE SOAPfuse, FusionCatcher, JAFFA, EricScript, chimerascan, PRADA, defuse, FusionMap, TopHat-Fusion, MapSplice, BreakFusion, SnowShoes-FTD, FusionQ, FusionHunter, ShortFuse.

8.7 References

All necessary information should be cited by the proper references. References can be the published articles, curated databases, or websites. Any consistent reference format can be used.

Annex A
(informative)

Example structure of clinical sequencing report

Tables A.1 and A.2 show the composition of the clinical gene fusion report with example values.