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)

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

3.40
whole genome sequencing
WGS
technique that determines the complete *DNA* (3.11) 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
ENUM	Enumerated type
EHR	Electronic Health Record
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

5 Use case scenario

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

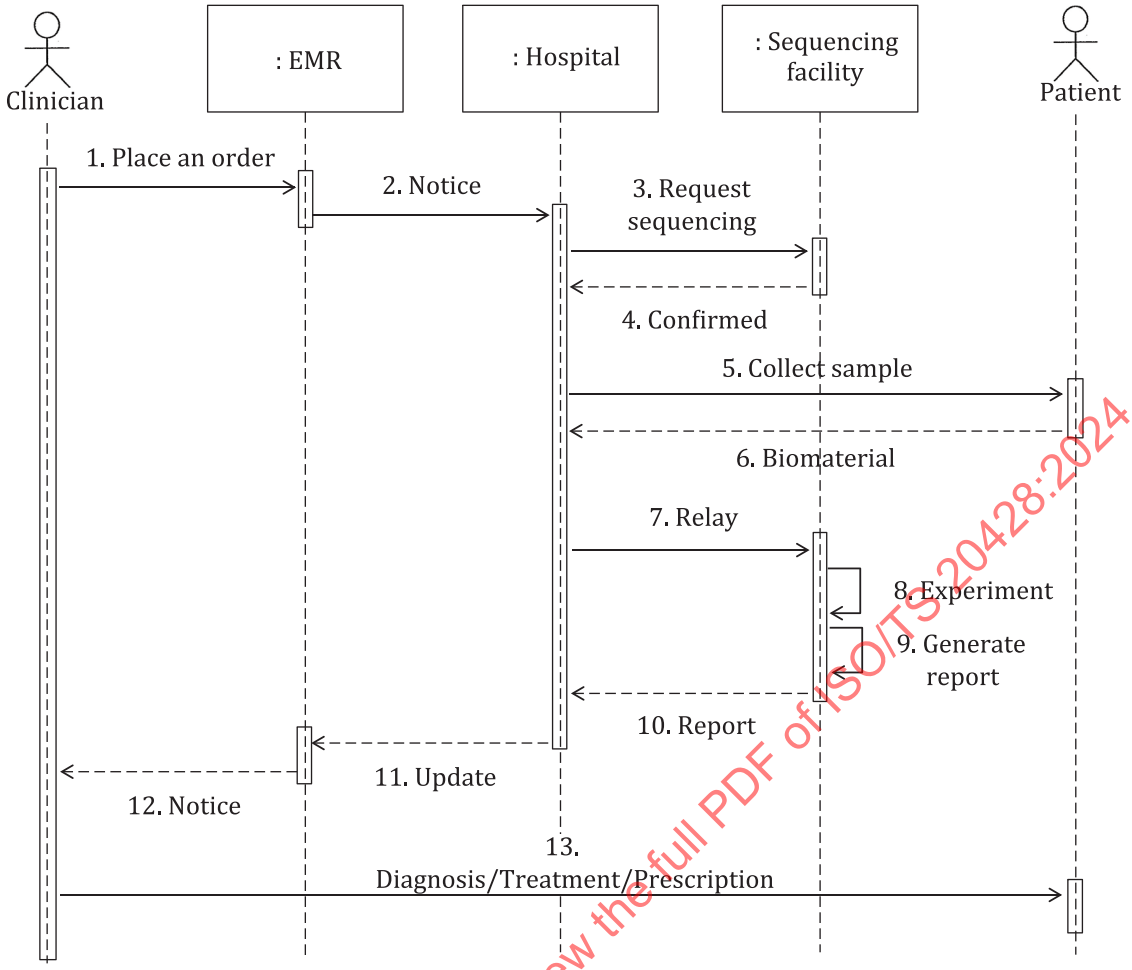

Figure 1 — Abstracted flow of generating clinical genomic sequencing report

The possible use cases for the clinical sequencing orders are well described in Reference [17].

6 Composition of a clinical sequencing report

6.1 General

The structured clinical sequencing report may mainly consist of two parts: the summary part and the detailed contents part as in Figure 2. The summary part can include the subset of required fields to help clinicians quickly overview the most important findings concisely.[24],[25] The detailed content part should contain all required fields and the selected optional fields. Annex A contains an example of the summary part and the detailed contents of a clinical sequencing report.

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

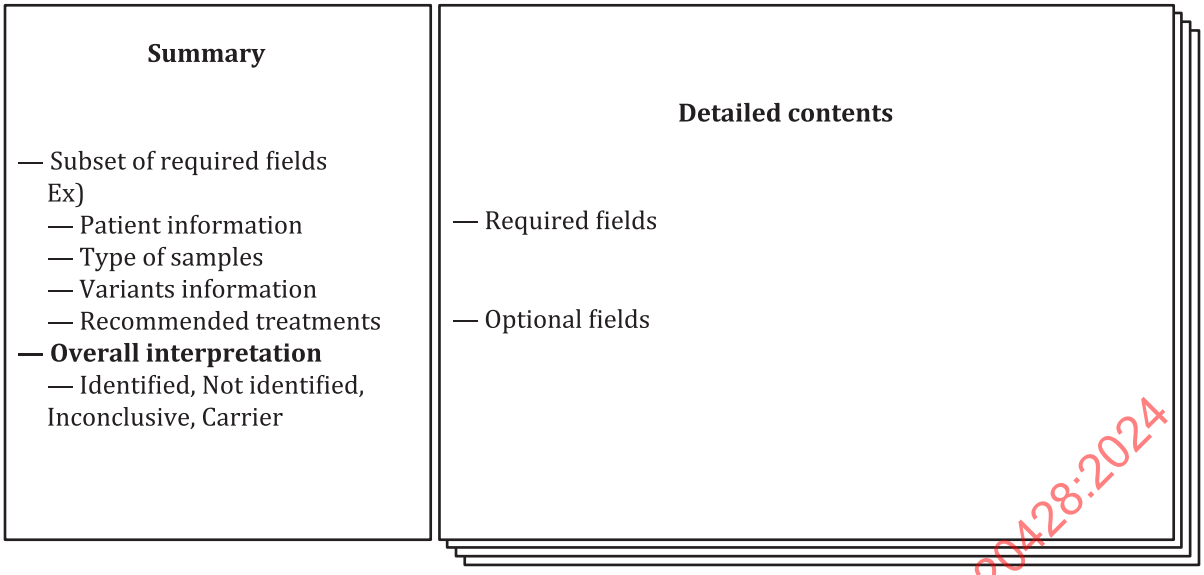

Figure 2 — Composition of a clinical sequencing report

6.2 Overall interpretation in summary

The summary part should report the overall interpretation of a genomic test with a succinct description: identified, not identified, inconclusive, or carrier.[22] Table 1 summarizes each interpretation.

Table 1 — Overall interpretation in the summary part

Interpretation	Remarks
Identified	detection of a variant that explains a patient's condition
Not identified	no variant identified of likely relevance to the diagnostic indication
Inconclusive[a]	a clear explanation of the patient's condition was not found
Carrier	identification of variants of recessive carrier screening tests

[a] Inconclusive can result from variants of unknown significance being identified, or from a single heterozygous variant being identified for a recessive condition.

6.3 Detailed contents

The detailed contents consist of two parts: the required and optional fields. The detailed content part should contain all required fields and the selected optional fields. HL7 Implementation Guide for CDA Release 2 Genetic Testing Report[33] can be a good example. However, HL7 Genetic Testing Report only focuses on the single gene test and does not cover next-generation sequencing technology.

The required fields mainly focus on helping clinicians by providing the necessary genomic information, interpretation results and the related treatments. They include all necessary fields for clinical practice. The optional fields give more detailed information to clinicians. They also can facilitate translational research with the necessary steps such as de-identification or consent from the patient. In Reference [24], the data fields for the molecular genetic report templates were categorized into "required", "optional", "possible" and "not necessary" based on the survey. The data fields in this document were re-categorized to implement a structured clinical sequencing report by reviewing the existing free-text format sequencing report.

7 Fields and their nomenclature of required data

7.1 General

The required fields are chosen for clinical practice. The information which can be only described in the clinical sequencing report is included in the required fields to minimize the length of clinical sequencing report. Cardinality indicates that this data element shall appear once in the report (1..1), one to many (1..N), or zero to many (0..N). The other relevant information can be included in the optional fields or other clinical reports in the EHRs. The summary of data elements and their metadata is shown in Table 2.

Table 2 — Data elements and their metadata for required fields

Data elements			Metadata (primary)	Cardinality
Clinical sequencing orders	Clinical sequencing order code	Order code	LOINC	1..1
		Information on sequencing order	TEXT	
	Date and time	Order date	ISO 8601	1..1
		Specimen collection date		
		Order received date		
		Report date		
		Addendum creation date		0..N
	Specimen information		ISO/TS 22220:2011	1..1
Information on subject of care	Identifiers		ISO/TS 22220:2011	1..1
	Name			
	Birth date		ISO 8601	
	Sex		ISO/TS 22220:2011	
	Population		HL7 v3 Code System Race	
Information of legally authorized person ordering clinical sequencing			ISO/TS 27527:2010	1..1
Performing laboratory	Basic information		TEXT	1..1
	Information of report generator		TEXT	
	Information of legally confirmed person on sequencing report		ISO/TS 27527:2010	
Associated diseases and phenotypes			ICD code with the version used	0..N
Biomaterial information	Type of sample		SPREC	0..N
	Genomic source class in biomaterial		LOINC	
	Conditions of specimen		TEXT	
Genetic variations	Gene symbols and names		HGNC	0..N
	Sequence variation information	Notation	HGVS	
		Effects of variants	TEXT	
		Sequence variant ID	(Name/version of the reference database) Database unique ID	

[a] ENUM indicates the contents should be chosen among the given category.

Table 2 *(continued)*

Data elements		Metadata (primary)	Cardinality
Classification of variants	Pathogeny	ENUM ("Pathogenic", "Likely pathogenic", "Unknown significance", "Likely benign", "Benign")[a]	0..N
	Clinical relevance	ENUM ("Identified", "Likely identified", "Uncertain", "Not identified")	
Recommended treatment	Medication	ISO 11615	
	Clinical trial information	Clinical trial ID	
	Known protocols related to a variant	TEXT	
	Other recommendation	TEXT	

[a] ENUM indicates the contents should be chosen among the given category.

7.2 Clinical sequencing orders

7.2.1 General

When a clinician orders a clinical genomic sequencing, the order code and the required date information should be given.

7.2.2 Clinical sequencing order code

7.2.2.1 Order code

The relevant clinical sequencing orders should be represented following guidance established in LOINC (Logical Observation Identifiers Names and Codes) with the used LOINC coding system version. Unfortunately, there are no LOINC codes for the given clinical sequencing orders, but only molecular genetic codes until now. When LOINC codes become available, this can be considered. In the meantime, other international, national or institutional coding system can be alternatively used, for example CPT (Current Procedure Terminology) in the US or NPU (Nomenclature for Properties and Units) in the Scandinavian countries.

7.2.2.2 Information on sequencing order

If the order code cannot fully describe the purpose of the clinical sequencing, the detailed description of the sequencing can be added as free text, for example the sequencing panel name with gene information.

7.2.3 Date and time

7.2.3.1 General

For a clinical sequencing report, diverse dates and times should be reported due to time delay. All dates and times in the report should be represented following guidance established in ISO 8601.

7.2.3.2 Order date

Order date is the date on which the clinician ordered the necessary clinical genomic sequencing.

7.2.3.3 Order received date

Order received date is the date that the performing laboratory received and confirmed the clinical genomic sequencing order.

7.2.3.4 Specimen collection date

Specimen collection date indicates the date when the specimen was taken from the patient or tissue.

7.2.3.5 Report date

Report date is the date when the performing laboratory generates the sequencing report.

7.2.3.6 Addendum creation date

Addendum creation date is the date when the performing laboratory creates the addendum of the previous report using up-to-date information. The analysis pipelines or referred database is updated due to technology advancement. The reference sequence is regularly updated. In addition, when updating, the performing laboratory should create the addendum of the existing sequencing report based on the clinician's request.

7.2.4 Specimen information

The specimen information can be represented by subject of care identifier type code from ISO 22220:2011.

EXAMPLE 13-S-048435_A1 – Pathology Number: ISO 22220:2011 [SOC identifier designation: 13-S-048435_A1, SOC identifier geographic area: 1 (local), SOC identifier issuer: AMC (ABC Medical Center), 02 (speciality number – pathology)].

7.3 Information on subject of care

7.3.1 General

Information on the subject of care whose specimen was sequenced should be represented following guidance established in ISO 22220:2011, i.e. 12345678 – ISO 22220:2011 [SOC identifier designation: 12345678, SOC identifier geographic area: 1 (local), SOC identifier issuer: AMC (ABC Medical Center), 01 (unique identifier for issuer)].

7.3.2 Subject of care identifiers

The unique identifiers for subjects of care should be included.

7.3.3 Subject of care name

The subject of care's name should be given.

7.3.4 Subject of care birth date

The subject of care's birth date should be given to calculate the patient's age. Birth date should be represented following guidance established in by ISO 8601.

7.3.5 Subject of care sex

The subject of care's sex should be represented following guidance established in ISO 22220:2011.

7.3.6 Subject of care population

The population of the subject of care should be notified to represent his or her genetic origin. The population information should be represented following guidance established in HL7 v3 Code System Race[1)]. Alternatively, if there are national standards, those coding systems can be used, for example the US FDA guidance for Industry – from Collection of Race and Ethnicity data in Clinical Trials.[26]

1) https://terminology.hl7.org/2.0.0/CodeSystem-v3-Race.html

7.4 Information on legally authorized person ordering clinical sequencing

Information on the legally authorized person who ordered the clinical sequencing can be represented by ISO/TS 27527:2010.

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

7.5 Performing laboratory

7.5.1 General

The laboratory that performs the sequencing should be given.

7.5.2 Basic information on performing laboratory

The name of performing laboratory and contact points such as phone numbers or emails should be indicated. This information can be given in free text until the relevant standard is published.

7.5.3 Information on report generator

Information on the provider who generated a report of sequencing results should be reported. It can be represented by ISO/TS 27527:2010 or by free text.

7.5.4 Information of legally confirmed person on sequencing report

Information of legally confirmed physician can be given by ISO/TS 27527:2010.

7.6 Associated diseases and phenotypes

If possible, associated diseases and phenotypes should be included using ICD codes, Human Phenotype Ontology[2], ORPHAcodes[3], or SNOMED CT codes, with the version used.

Since even a single variant can have multiple associated phenotypes including diseases, all possible associated diseases should be listed based on the confidence level.

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

7.7 Biomaterial information

7.7.1 General

The information on the specimen from the patient should be described.

7.7.2 Types of sample

Types of samples can be represented by Standard Preanalytical Code (SPREC) V2.0 of International Society for Biological and Environmental Repositories.[27]

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

2) https://hpo.jax.org/app/

3) https://www.rd-code.eu/introduction/

7.7.3 Genomic source class in biomaterial

The genomic source of the patient's sample should be categorized as follows: 1) germline, 2) somatic, 3) prenatal/foetal, 4) likely germline, 5) likely somatic, 6) likely prenatal/foetal, or 7) unknown genomic origin. This category is based on LOINC answer list LL378-1.

7.7.4 Conditions of specimen that can limit adequacy of testing

The specific conditions that can affect or limit the adequacy of genomic sequencing should be described in text format. For example, tumour purity information shall be given in percentage.

7.8 Genetic variations

7.8.1 General

All found variants should be listed in the report according to their relevance to the patient's indication for testing. However, in the required field, the variants that have the associated treatments should be listed. The other variants can be listed in the optional fields.

NOTE The variants associated with specific treatments will evolve over time as clinical sequencing technologies and treatments advance. Consequently, if new variants with appropriate treatments are identified, an addendum should be created.

As in 7.9, classification of variants can be categorized based on whether the identified variant has a proper treatment. In the clinical setting, the variants without proper treatment have no meaning to the clinicians. Therefore, only variants with treatments should be included in the required field.

The variants information should include the information in 7.8.2 and 7.8.3.

7.8.2 Gene symbols and names

The gene symbol and name which contains the identified variants should be represented following guidance established in HGNC [HUGO (Human Genome Organization) Gene Nomenclature Committee]. The HGNC approved gene symbol and HGNC ID can be used, i.e. HGNC approved symbol: BRCA1, HGNC ID: HGND:1100. The symbol is case-insensitive.

HGNC is a non-profit body which is jointly funded by the US National Human Genome Research Institute (NHGRI) and the Welcome Trust (UK). They operate under the auspices of HUGO.

7.8.3 Sequence variation information

7.8.3.1 General

The sequence variant should be represented following guidance established in HGVS (Human Genome Variation Society, http://www.hgvs.org/) nomenclature. Amino acid changes can be included to give more information.

EXAMPLE c.76A > T_ p.Asn26Tyr.

7.8.3.2 HGVS nomenclature

The detailed explanation on description of sequence changes in DNA level can be found at http://varnomen.hgvs.org/recommendations/DNA/.

EXAMPLE Diverse sequence variants can be presented as follows:

Substitutions: in HGVS, ">" indicates a substitution at DNA level (i.e. c.76A > T).

Deletions: in HGVS, "del" indicates a deletion (i.e. c.76delA).

Duplications: in HGVS, "dup" indicates a duplication (i.e. c.76dupA).

Insertion: in HGVS, “ins” indicates an insertion (i.e. c.76_77insG).

Insertion/deletion: in HGVS, “delins” indicates an indel (c.112_117delinsTG). Indels are described as a deletion followed by an insertion.

Inversions: in HGVS, ‘inv’ indicates an inversion (i.e. c.76_83inv).

Conversions: in HGVS, ‘con’ indicates a conversion. The example “g.123_678conNG_012232.1:g.9456_10011” describes a gene conversion replacing nucleotides 123 to 678 of the reference genomic sequence with nucleotides 9 456 to 10 011 from the sequence as present in GenBank file NG_012232.1.

Translocations: in HGVS, translocations are described at the molecular level using the format “t(X;4)(p21.2;q35)”, followed by the usual numbering, indicating the position translocation breakpoint. t(X;4)(p21.2;q35)(c.857+101_857+102) denotes a translocation breakpoint in the intron between coding DNA nucleotides 857+101 and 857+102, joining chromosome bands Xp21.2 and 4q34.

7.8.3.3 Effects of variants

Type of variants can be explained using free texts.

EXAMPLE 1 Substitution, Deletion, Duplication, Insertion, Inversion and Conversion.

Additionally, the effects of a variant on protein function such as missense, nonsense and silent can be given in parentheses.

EXAMPLE 2 Substitution (missense).

7.8.3.4 Sequence variant ID

All known variants can be reported using other unique IDs with the name and version of the database to which the IDs refer.

If the sequence variant has dbSNP ID[4], the variant can be represented by dbSNP ID. The prefix “rs” represents dbSNP ID (i.e. rs10000). dbSNP is maintained by NCBI.

ClinVar[5] is a database for the relationships among human variations and phenotypes, with supporting evidence, maintained by NCBI. If ClinVar ID is used, the identifier should be prefixed with “ClinVar”: (i.e. ClinVar:17661) as in HL7 FHIR Standard Profile for Genetics.

If a variant is a somatic variant, COSMIC (catalogue of somatic mutations in cancer)[6] identifier can be used with other identifiers. COSMIC id should be prefixed with “COSM” (i.e. COSM12979) as in HL7 FHIR Standard Profile for Genetics. COSMIC is a database to store and display somatic variant information and related details and contains information relating to human cancers. COSMIC is maintained by the Wellcome Trust Sanger Institute.

7.9 Classification of variants

7.9.1 General

Currently, there are two different approaches to classify genetic variants: one focuses on the pathogeny and the other focuses on clinical relevance based on the existence of possible medication. These two classifications have different meanings, the institution can temporally choose which one will be used or make a new classification by combining two existing classifications.

4) http://www.ncbi.nlm.nih.gov/snp/

5) http://www.ncbi.nlm.nih.gov/clinvar/

6) https://cancer.sanger.ac.uk/cosmic/

7.9.2 Classification of variants based on the pathogeny

This classification follows the ACMG recommendations for standards for interpretation and reporting of sequence variations: Revision 2015,[28] IARC 5-class system,[29] or College of American Pathologists guideline[30]: pathogenic, likely pathogenic, unknown significance, likely benign, and benign.

7.9.3 Classification of variants based on clinical relevance

In a clinical setting, alternative classification is necessary to provide a proper treatment for the patient.[31] In this document, we revised the classification: identified, likely identified, uncertain, and not identified.

Identified: the clinically relevant (actionable) variants are detected.

Likely identified: the variants with likely clinical relevance are detected.

Uncertain: the variants with uncertain clinical relevance are detected.

Not identified: no clinically relevant variants are detected.

7.10 Recommended treatment

7.10.1 General

The recommended treatment such as medication or clinical trials can be reported to help clinicians.

7.10.2 Classification of variants based on clinical relevance

The associated medication can be represented by MPID (Medicinal Product IDentifier) or IMPID (Investigational MPID) of ISO 11615. The Anatomical Therapeutic Chemical (ATC) classification maintained by the WHO Collaborating Centre for Drug Statistics Methodology, can be alternatively used to represent the classification of active ingredients of drugs. International Nonproprietary Names (INN) facilitates the identification of pharmaceutical substances or active pharmaceutical ingredients. Each INN is a unique name that is globally recognized and is public property. If INN exists, INN can be alternatively used as drug name.

However, it will be most directly beneficial to individual patients for the recommended target of therapy to be represented using the same coding scheme that has already been used within the particular country or EHR system. This way, the treatment recommendation and the actual prescription or course of chemotherapy are semantically aligned within the record of a single patient, who is likely to be treated in a local environment rather than globally. Therefore, if there is a national or local standard, it can be used temporarily. The examples are RxNorm, which is maintained by the US. National Library of Medicine, or KD Code (Korea Drug Code), which is maintained by the Korea Health Insurance Review and Assessment Service.

7.10.3 Clinical trial information

The information from clinical trials that test the identified variants-targeted drugs can help clinicians. The clinical trial information should be represented following guidance established in ClinicalTrials.gov ID or EudraCT trial number. Another domestic registry ID can be alternatively used.

ClinicalTrials.gov is a registry of federally and privately supported clinical trials conducted in the United States and around the world. If there are the related clinical trials to detect variants, the ClinicalTrials.gov ID can be given. In addition, the target condition, intervention, phase, and study results can be attached.

EXAMPLE NCT00844506.

EudraCT[7)] is a database of all clinical trials which commenced in Europe. EudraCT number can also be used to represent the related clinical trials.

7) https://eudract.ema.europa.eu/

7.10.4 Known protocols related to a variant

If the found variants are classified as clinical actionable variants, the established clinical guidelines such as CPIC (Clinical Pharmacogenetics Implementation Consortium) guideline[8)][32] or NCCN (National Comprehensive Cancer Network) guideline[9)] should be included.

7.10.5 Other recommendation

Other recommendations can be included, for example other laboratory tests including another sequencing order or testing relatives.

7.11 Addendum

If there is an addendum, it should be described in the required field. The reason for creating an addendum can be included.

8 Fields and their nomenclature of optional data

8.1 General

The data elements for optional fields in the structured clinical sequencing report focus mainly on the support of clinical decision by giving more detailed information (see Table 3). These fields can also be applied to clinical trials and translational research. Cardinality indicates that this data element shall appear once in the report (1..1), one to many (1..N), or zero to many (0..N).

The fields that are not listed in this document can be also used based on institutional decision.

Table 3 — Data elements and their metadata for optional fields

Data elements			Metadata (primary)	Cardinality
Medical history			ICD code with the version used	1..1
Family history/Pedigree information			HL7 v3 IG: Family history/ Pedigree interoperability	0..N
Reference genome version			Genome Reference Consortium Human Genome release ID	0..N
Racial genome information			TEXT	1..1
Karyotypic sex			TEXT	1..1
Genetic variation	Gene symbols and names		HGNC	0..N
	Sequence variation information	Notation	HGVS	0..N
		Effects of variants	TEXT	
		Sequence variant ID	Database unique ID	
		Zygosity	Text	
	HGVS version		HGVS version number	1..1

[a] ENUM indicates the contents should be chosen among the given category.

8) https://www.pharmgkb.org/guidelineAnnotations

9) https://www.nccn.org/guidelines/category_1

Table 3 *(continued)*

Data elements			Metadata (primary)	Cardinality
Detailed sequencing information	Clinical sequencing date		ISO 8601	1..1
	Quality control metrics		NUMERIC	1..N
	Base calling information	Read depth	NUMERIC	1..1
		Reference allelic depth		
		Alternative allelic depth		
		Allele frequency		
		Genotype		
	Sequencing platform information	Type of sequencers	TEXT	1..N
		Library preparation methods		
		Target capture methods		
		Read type	ENUM (“single-end”, “paired-end”)[a]	1..1
		Read length	TEXT	1..1
	Analysis platform information	Alignment tools	TEXT	1..N
		Variant calling tools		
		Other tools		
		Chromosome coordination system	ENUM (“zero-based”, “one-based”, “zero-based, half-open”)	1..1
		Annotation tools and databases	TEXT	1..N
References			TEXT	1..N

[a] ENUM indicates the contents should be chosen among the given category.

8.2 Subject of care population

The population of the subject of care should be reported to represent his or her genetic origin. The population information should be represented following guidance established in HL7 v3 Code System Race[10]. Alternatively, if there are national standards, those coding systems can be used, for example, US FDA guidance for Industry–Collection of Race and Ethnicity Data in Clinical Trials.[26]

8.3 Medical history

The medical history of the subject of care can be reported. When reporting medical history, the relevant standard terminology should be used, for example, ICD code with the version used for disease names or IDMP for medication.

8.4 Family history/Pedigree information

All family history or pedigree information should be represented following guidance established in HL7 v3 Implementation Guide: Family History/Pedigree Interoperability, Release 1.[34]

10) https://terminology.hl7.org/3.1.0/CodeSystem-v3-Ethnicity.html

8.5 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 following guidance established in Genome Reference Consortium Human Genome release ID[11] or Locus Reference Genomic ID.

If there is an update, the patch number should be appended.

EXAMPLE GRCh38 or GRCh37.p13 (GRCh37 Patch Release 13).

8.6 Populational genomic information

When populational genomic information is reported, the reference data set for populational information should be notified using 1000 Genomes[12], Genome Aggregation Database (gnomAD)[13], or other sequencing database which can give populational information.

8.7 Karyotopic sex

The karyotypic sex of the subject of care should be given.

8.8 Genetic variation

The remaining variants, which are not reported in the required part, can be listed in the optional part. The variants should be reported following the instructions in 7.8. However, in optional fields, HGVS version and zygosity of the variant should be notified, i.e. HGVS Version 15.11.

8.9 Classification of variants

8.9.1 General

The classification of the variants which are reported in the optional part should be reported following the instructions in 7.9.

8.9.2 Classification of variants as secondary finding

Secondary findings may be enumerated as a result of the incidental or accidental discovery of genomic variants in genes that are included in a predefined list during the active search conducted in accordance with the specific objectives of the analysis.

8.10 Detailed sequencing information

8.10.1 General

Until now, there is no standard for this purpose. The notation used in conventional bioinformatics fields can be used alternatively.

8.10.2 Clinical sequencing date

Clinical sequencing date is the date on which the performing laboratory generates the sequencing results using the received specimen. Usually, after receiving samples, the laboratory gathers enough samples to run the sequencer to optimize the efficiency of sequencer. In addition, the sequencing procedure takes a day and more to generate the analysis results. The date should be represented following guidance established in ISO 8601 as other date information.

11) https://www.ncbi.nlm.nih.gov/grc/human

12) https://www.internationalgenome.org/

13) https://gnomad.broadinstitute.org/

8.10.3 Quality control metrics

The relevant quality control (QC) metrics for sequencing and analysis can be given. The report can include the overall QC metrics for biospecimen, the QC metrics for all variants, or the QC metrics for specific variants based on the report generator's decision. QC metrics can be given based on ISO/TS 22692:2020.

8.10.4 Base calling information

8.10.4.1 General

Information on base calling that is generated by base calling software for identifying a nucleotide sequence can be notified. The detailed fields will be as detailed in 8.10.4. The other fields can be added based on the report generator's decision.

8.10.4.2 Read depth

The average number of nucleotides contributing to a portion of an assembly can be reported as used in conventional bioinformatics fields.

EXAMPLE 100x.

8.10.4.3 Reference allelic depth

Allelic depth for the reference allele can be reported as used in conventional bioinformatics fields.

EXAMPLE 50x.

8.10.4.4 Alternative allelic depth

Allelic depth for the alternative allele can be reported.

EXAMPLE 50x.

8.10.4.5 Variant Allele frequency

The frequency of alternative allele at each locus for each individual can be reported.

EXAMPLE 0,3, 30 %.

8.10.4.6 Genotype

The pair of alleles present at a single locus can be reported.

EXAMPLE AA, AC, AG, AT, CC, CG, CT, GG, GT, TT.

8.10.5 Sequencing platform information

8.10.5.1 General

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

8.10.5.2 Type of sequencers

The specific sequencer that performs the sequencing should be given, for example Illumina Hiseq 2500, Thermo Fisher Ion Torrent, or Illumina MiSeq.

8.10.5.3 Library preparation methods

The sequencing library preparation methods should be given, i.e. SureSelectXT.

8.10.5.4 Target capture methods

The exome or targeted region capture methods should be notified, i.e. Amplicon, probe capture.

8.10.5.5 Read type

Sequencing read type such as single-end or paired-end should be given.

8.10.5.6 Read length

The sequencing read length information should be given, i.e. 101 bp or 35-250 bp.

8.10.6 Analysis platform information

8.10.6.1 General

The primary, secondary (if applicable tertiary) analysis pipelines should be mentioned. The parameter setting for pipelines such as GATK 3.5, CASAVA 1.7, Complete Genomics v2.2, or Torrent Suite 5.0.2 should be also mentioned to confirm the reliability of variant call.

8.10.6.2 Alignment tools

The name of alignment tool and its version should be given, i.e. BWA-MEM 0.7.12.

8.10.6.3 Variant calling tools

The name of variant calling tool and its version should be given, i.e. GATK 3.5 or SAMTools 1.3.1.

8.10.6.4 Other tools

The name of other tools (e.g. for manipulating NGS data and formats) can be given.

EXAMPLE PICARD 1.9.3.

8.10.6.5 Chromosome coordinate system

The chromosome number can be started at 0 or 1. Therefore, they should be given as the 0-based, 0-based half open, or 1-based coordinated system.

8.10.6.6 Annotation tools and databases

The name of annotation tools and source of databases that is publicly available or private should be reported, i.e. ANNOVAR (2016Feb01), SnpEff, 4.3, Ensembl v74.

8.11 References

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

Annex A
(informative)

Example structure of clinical sequencing report

Tables A.1 to A.3 demonstrates the content of clinical sequencing report with the informative values.

Table A.1 — Example of the summary part of a clinical sequencing report

Fields[a]			Value[b]	Representation[c]
Clinical sequencing order	Clinical sequencing order code	Order code	14-RM-0000056	14-RM-0000056
	Date and time	Order date	2014–04–18	April 18, 2014
		Report date	2014–04–25	April 25, 2014
	Specimen information		13-S-048435_A1	13-S-048435_A1
Information on subject of care	Identifiers		12345678	12345678
	Name		Gildong Hong	Gildong Hong
Biomaterial information	Type of sample		CEN	Non-blood tissue
Genetic variations	Gene symbols and names		HGNC:1097, BRAF	BRAF
Recommended treatment	Medication		Vemurafenib, L01XE15 (ATC code)	Vemurafenib
Overall interpretation			Identified	Identified

[a] The selected fields for the Summary part.

[b] Computer-processable code or value for each field.

[c] Human-readable representation in each field of the sequencing report.