FINAL
DRAFT

TECHNICAL SPECIFICATION

ISO/DTS 4424

ISO/TC **215**/SC **1**

Secretariat: **KATS**

Voting begins on:
2023-09-15

Voting terminates on:
2023-11-10

Genomics informatics — Data elements and their metadata for describing the tumor mutation burden (TMB) information of clinical massive parallel DNA sequencing

Informatique génomique — Éléments de données et leurs métadonnées pour décrire les informations relatives à la charge tumorale mutationnelle (TMB) du séquençage massif parallèle d'ADN

Reference number
ISO/DTS 4424:2023(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 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*, 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 (NGS) technologies, clinical sequencing has been applied to realize personalized and precision medicine. ISO/TS 20428[1] was published to standardize the clinical sequencing reports in electronic health records. After introducing NGS panel sequencings (whole genome, whole exome, targeted gene sequencing), they are widely used in the clinical field.

In the field of cancer treatment, various treatment strategies were tried differently from traditional anti-cancer chemotherapies. Recently, drugs related to the immune system were developed and more efficient for patients with specific tumor molecular characteristics. It is the immune checkpoint blockade drug such as the first approved drug – Ipilimumab, an anti-cytotoxic T-lymphocyte antigen (CTLA4) for non-small cell lung cancer[2]. Tumors can use these checkpoints to protect themselves from immune system attacks. Currently approved checkpoint therapies block inhibitory checkpoint receptors. Blockade of negative feedback signaling to immune cells thus results in a continued immune response against tumors. It was reported that the status of Programmed Death-Ligand 1 (PD-L1) expression or the status of TMB (Tumor Mutation Burden) could be used as the predictive marker for the efficacy of the immune checkpoint blockade because TMB is considered an indirect measurement of how many tumor cell-specific peptide fragments are presenting and the increase of antigen-presenting leads more immune reaction[3].

The status of TMB can be calculated and reported from detected genomic variants by NGS DNA sequencing. According to national regulatory agencies, including US-FDA, several NGS sequencing products are being approved for companion diagnostics[4]. Some NGS sequencing products provide TMB status and value on their NGS sequencing report. CLIA-certified labs or equivalent-level agencies in countries also serve the TMB value from their own methods. It is forecasted that more clinical NGS sequencing will be approved to report TMB.[5]

However, there is no international standard for describing TMB status, value, and metadata. The previous ISO/TS 20428 focused on only DNA variations compared with the reference genome. Some research results said that TMB values and how to describe them are different even if using the same sequencing data. The absence of a standard for TMB representation makes it difficult for clinicians and researchers not only to use TMB results for clinical decision support but also for secondary analysing purposes when receiving from more than one sequencing lab. Related metadata should be essential to expand the usage of TMB values.

In this document, the data elements and their standardized metadata for TMB in electronic health records will be described. The clinical report for TMB will provide proper information on bioinformatics analysis to help clinical decisions.

Genomics informatics — Data elements and their metadata for describing the tumor mutation burden (TMB) information of clinical massive parallel DNA sequencing

1 Scope

This document identifies data elements and metadata to represent the information about tumor mutation burden (TMB) when reporting the value for the biomarker using clinical massive parallel DNA sequencing.

This document covers the TMB status and related metadata such as mutation type, sequencing types, and target areas of sequencing from human samples for clinical practice and research.

This document is not intended

— to define experimental protocols or methods for calculating the value of tumor mutation burden,

— for the other biological species, and

— for the Sanger sequencing methods.

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

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

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
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]

3.2
clinical sequencing
next generation sequencing or later sequencing technologies with human samples for clinical practice and clinical trials

[SOURCE: ISO/TS 20428:2017, 3.5]

3.3
deoxyribonucleic acid
DNA
molecule that encodes genetic information in the nucleus of cells

[SOURCE: ISO 25720:2009, 4.7]

3.4
DNA sequencing
determining the order of nucleotide bases (adenine, guanine, cytosine and thymine) in a molecule of DNA

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

[SOURCE: ISO 17822:2020, 3.19]

3.5
exome
part of the genome formed by exons

[SOURCE: ISO/TS 20428:2017, 3.13]

3.6
gene
basic unit of hereditary material that encodes and controls the expression of a protein or protein subunit

[SOURCE: ISO 11238:2018, 3.29]

3.7
gene panel
technique for sequencing the targeted genes in a genome

[SOURCE: ISO/TS 20428:2017, 3.15]

3.8
germline
series of germ cells each descended or developed from earlier cells in the series, regarded as continuing through successive generations of an organism

[SOURCE: ISO/TS 20428:2017, 3.17]

3.9
nucleotide
base
base pair
monomer of a nucleic acid polymer such as DNA or RNA

Note 1 to entry: Nucleotides are denoted as letters ('A' for adenine; 'C' for cytosine; 'G' for guanine; 'T' for thymine which only occurs in DNA; and 'U' for uracil which only occurs in RNA). The chemical formula for a specific DNA or RNA molecule is given by the sequence of its nucleotides, which can be represented as a string over the alphabet ('A', 'C', 'G', 'T') in the case of DNA, and a string over the alphabet ('A', 'C', 'G', 'U') in the case of RNA. Bases with unknown molecular composition are denoted with 'N'.

[SOURCE: ISO/IEC 23092-2:2020, 3.20]

3.10
quality score
Q score
Phred quality score
sequencing quality score of a given nucleotide base

Note 1 to entry: Q is defined by the following equation: Q = -10log10(e), where e is the estimated probability of the base call being wrong.

Note 2 to entry: A quality score of 20 represents an error rate of 1 in 100, with a corresponding call accuracy of 99 %.

Note 3 to entry: Higher quality scores indicate a smaller probability of error. Lower quality scores can result in a significant portion of the reads being unusable. Low quality scores may also indicate false-positive variant calls, resulting in inaccurate conclusions.

[SOURCE: ISO 20397-2:2021, 3.32]

3.11
read type
type of run in the sequencing instrument

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

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

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

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

3.12
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 standard 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]

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

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

3.14
sequence variation
DNA sequence variation
variation
differences of DNA sequence among individuals in a population

[SOURCE: ISO 25720:2009, 4.8]

3.15
single nucleotide variant
SNV
DNA sequence variation that occurs when a single nucleotide, A, T, C, or G, in the genome (or other target sequence) differs between templates

[SOURCE: ISO 20395:2019, 3.35]

3.16
specimen
biospecimen
biological 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]

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

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

3.18
target capture
method to capture genomic regions of interest from a DNA sample prior to sequencing

[SOURCE: ISO/TS 20428:2017, 3.36]

3.19
targeted sequencing
disease-targeted gene panels
the technique used for sequencing only selected/targeted genomic regions of interest from a DNA sample

[SOURCE: ISO/TS 22692:2020, 3.30]

3.20
whole exome sequencing
WES
technique for sequencing the exomes of the protein-coding genes in a genome

[SOURCE: ISO/TS 20428:2017, 3.38]

3.21
whole genome sequencing
WGS
technique that determines the complete DNA sequence of an organism's genome at a single time

[SOURCE: ISO/TS 20428:2017, 3.39]

4 Abbreviated terms

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

ATC	Anatomical Therapeutic Chemical
CTLA4	Cytotoxic T-Lymphocyte Associated Protein 4
EBI	European Bioinformatics Institute
IDMP	Identification of Medicinal Product
IMPID	Investigational MPID
INN	International Nonproprietary Names
MHC	Major Histocompatibility Complex
MPID	Medicinal Product Identifier
NCBI	National Center for Biotechnology Information
NCCN	National Comprehensive Cancer Network
NGS	Next Generation Sequencing
NIH	National Institutes of Health
PD-L1	Programmed Death-Ligand 1
PD-1	Programmed cell Death protein 1
SPREC	Standard PREanalytical Code
TMB	Tumor Mutation Burden
WES	Whole Exome Sequencing
WGS	Whole Genome Sequencing
WHO	World Health Organization
UTN	Universal Trial Number
UMI	Unique Molecular Identifier

5 Tumor mutation burden (TMB)

Molecular characterization of tumors utilizing next-generation sequencing methods (NGS) is currently in the focused area of personalized medicine. Tumor mutation burden (TMB) is considered one of the molecular markers in the field of tumor diagnostics. Recently, clinical trials showed that immune checkpoint inhibitors are more effective with patients with high TMB regarding response rates and survival benefits. The simple mean of TMB is how many mutations occurred in tumor cells[7]. Neoantigens are novel tumor cell surface peptides presenting by Major Histocompatibility Complex (MHC), some of which can be recognized as foreign to the body by the immune system, resulting in increased T-cell reactivity and thereby leading to an antitumor immune response. To prevent the excessive immune reaction, immune checkpoint proteins (ex, PD-L1, PD-1, CTLA4) were expressed on the surface of tumor cells or T-cells. As binding these proteins between tumor cells and T-cells, the immune reaction is decreased.

Immune checkpoint inhibitors enhance antitumor T-cell activity by inhibiting immune checkpoint molecules. So, the status of neo-antigen or TMB can be a suitable clinical biomarker to guide treatment decisions for immune checkpoint inhibitors. Although selecting which mutations are likely to induce