TECHNICAL SPECIFICATION

ISO/TS 4424

First edition
2023-12

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

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

3.17
target capture
method to capture genomic regions of interest from a DNA sample prior to sequencing

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

3.18
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]

3.19
whole exome sequencing
WES
technique for sequencing the exomes of the protein-coding genes in a genome

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

3.20
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 immunogenic neoantigens is not clear, TMB represents a quantifiable measure of the number of tumor mutations that can inform treatment selection.

TMB is mostly measured by NGS – whole genome sequencing (WGS), whole exome sequencing (WES), and targeted gene sequencing[8]. In 2020, commercial targeted DNA sequencing for TMB was approved as the companion diagnostics to inhibit the immune checkpoint system (PD-L1) for patients with solid tumors in the United States. In addition, various targeted gene panel assays are being developed for multi-purposes, including TMB assessment, and are used in various clinical trials. So, it could be expected that more methods would be approved. Therefore, in order to increase the utility of the TMB results from multi agencies or methods, the essential factors such as the way of describing TMB results and metadata derived from the estimate process of TMB on the clinical report should be defined and standardized.

6 Composition of elements for describing TMB on the clinical DNA NGS report

6.1 General

The clinical DNA NGS report, including TMB results, may mainly consist of two parts, the same as ISO/TS 20428: the summary and detailed parts. Because the result of TMB and related data could be calculated from the result of DNA sequencing, it is general to supply the result of TMB with the result of clinical DNA NGS in the report. The summary part includes the subset of required fields to report the main result of DNA sequencing, including TMB status, concisely for clinicians. The detailed section should contain all required fields (see Clause 7) and the selected optional fields (see Clause 8).

6.2 Summary part

The summary section should report the overall interpretation of a test based on clinical implications. It should contain the calculated result of TMB status. The additional clinical significance of the relevant therapies can be included. In addition, the summary section may include the subset of the detailed section, as necessary.

6.3 Detail part

The detailed section should contain all the required and selected optional fields. The required fields mainly focus on helping clinicians by providing the necessary TMB status result and information, interpretation results, and related treatments. The information that can be only described is included in the required fields to minimize the length of the 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 consent from the subject of care.

Annex A demonstrates the composition of the clinical DNA sequencing report, including the TMB result.

7 Fields and their nomenclature of required data

7.1 General

In the field of required data, data elements, their metadata, and cardinality (see Table 1) were chosen mainly for clinical practice using TMB reports with clinical DNA sequencing based on NGS, including targeted DNA sequencing, whole exome sequencing, or whole genome 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
		Order received date		One
		Specimen collection date		One
		Report date		One
		Addendum creation date		Many

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

Table 1 *(continued)*

Data elements		Metadata (Primary)	Cardinality
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
	Referring diagnosis	ICD [11]	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 [12]	One
TMB result information	TMB value	NUMERIC (mut/MB)	One
	TMB status	ENUM (“TMB-high”, “TMB-intermediate”, “TMB-low”) [a]	One
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 the 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 measure TMB status, a detailed description of the order can be given as free text.

7.2.3 Date and time

7.2.3.1 General

All dates and times in the report shall be represented in accordance with the ISO 8601 series. ISO 8601 remove unambiguity on day-date conventions.

EXAMPLE Date represented as ‘YYYY-MM-DD’.

7.2.3.2 Order date

The order date is when a clinician ordered the necessary sequencing test to measure TMB status.

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. The order date (see 7.2.3.2) and order received date could be different.

7.2.3.4 Specimen collection date

The specimen collection date is when a specimen is taken from the patient. Since the specimen can be collected in diverse ways, such as surgery, biopsy, or blood collection on 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 TMB status report. It might be divided into report generation date and confirmation date 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, such as the update of the analysis pipeline or change of result by reviewers. 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. ISO/TS 22220 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 healthcare organizations.

7.3.2 Subject of care identifier

The subject of care identifier indicates the unique identifier of the patient or recipient of care.

7.3.3 Subject of care name

The subject of care name indicates the name of the patient or recipient of care. Some parts of the name could be removed or blanked by the indication of the institute for protecting personal information.

7.3.4 Subject of care birth date

The birth date shall be represented in accordance with the ISO 8601 series 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 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 persons who ordered or prescribed the TMB status test shall be represented in accordance with ISO/TS 27527. ISO/TS 27527 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 regarding the laboratory that performs the sequencing.

7.5.2 Basic information on performing laboratory

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

7.5.3 Information on report generator

Information on the 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 physicians shall be provided 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 the subject of care.

7.6.2 Type of specimen

Type of specimen can be represented in accordance with the Standard PREanalytical Code (SPREC) of the International Society for Biological and Environmental Repositories. Currently, SPREC Version 3.0 is the up-to-date version.

EXAMPLE TIS (Solid tissue), CTC (Circulating tumor cell), PL1 (Plasma, single spun).

7.7 TMB status result information

7.7.1 General

This field describes the TMB result of the subject of care from clinical NGS DNA sequencing and related data for clinical usage.

7.7.2 TMB value

TMB value shall be reported as the number of effective variants per megabase of the genomic sequence by identifying variants from the result of DNA sequencing. TMB value should be expressed numerically in the unit of mut/Mb.

7.7.3 TMB status

The TMB state should be described by being based on the cutoff using the calculated TMB value. According to the standards of the manufacturer or analysis agency, the TMB status should be categorized into two or three stages such as TMB-high, TMB-low or TMB-high, TMB-moderate, and TMB-low.

7.8 Recommended treatment

7.8.1 General

The recommended treatment, such as medication, clinical guidelines, or clinical trials, can be reported to help clinicians. According to the result of TMB status, the relevant clinical decision shall be provided with supporting information, including clinical guidelines, clinical trials, or academic articles.

7.8.2 Medication

As regards the diagnosis and the result of TMB status about the subject of care, the recommended clinical care guide shall be described, including not only the drug name but also the condition of 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 or the drug information portal maintained by the domestic drug information regulatory institute (e.g. Drug information portal of National Institutes of Health (NIH)) can be used to represent the classification of active ingredients of drugs. International Nonproprietary Names (INN) can be alternatively used as a drug name since INN facilitates the identification of pharmaceutical substances or active pharmaceutical ingredients. Indicators of the national insurance system or medical practices used by each organization can also be used.

7.8.3 Clinical trial information

The clinical trial information can be provided to help clinicians. It should be represented in accordance with Universal Trial Number (UTN) used in the World Health Organization (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.

7.8.4 Other recommendation

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

7.8.5 Supporting information

The supporting information on the provided medication or clinical care can be described using free text.

EXAMPLE Clinical care decision guideline (ex. National Comprehensive Cancer Network (NCCN)) including version, national registry identifier or certified clinical information database.

8 Fields and their nomenclature of optional data

8.1 General

The data elements for optional fields may be used to understand the process of calculating TMB value and the condition of DNA sequencing. These fields can also be applied to clinical trials and translational research. Other fields might be added based on the institutional decision even if they are not included in this list.

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
Reference genome version			Genome Reference Consortium Human Genome release ID	One
TMB information	Criteria of TMB status		TEXT	One
	Approach of filtering germline variants		TEXT	One
	Variant types used for calculating TMB value		TEXT	One
	TMB region covered		NUMERIC (Mb or Mbase)	One
	Calibrated TMB value		NUMERIC (mut/Mb)	One
Sequencing information	Clinical sequencing date		ISO 8601 (all parts)	One
	Sequencing type		TEXT	One
	Quality control metrics	Percentage data quality > Q30	NUMERIC (%)	One
		Depth of target	NUMERIC (X)	One
	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
		Variant caller		Many
		Annotation tools		Many
References			TEXT	Many

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

8.2 Reference genome version

The version of the reference genome used for processing the generated reads or identifying genomic variants to calculate TMB value could be supplied. The reference sequence should be represented in accordance with the 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 European Bioinformatics Institute (EBI), can be used as well.

EXAMPLE LRG_1

8.3 TMB information

8.3.1 General

In this field, data elements or related data can be supplied, which can explain the condition of calculating TMB value using NGS DNA sequencing results or would be useful for the translational research or merge of TMB data generated by multiple sites.

8.3.2 Criteria of TMB status

The criteria for determining the stage of TMB status should be described in this field. The described criteria should be able to explain all the levels in the status that can be provided.

EXAMPLE 1 TMB-high: ≥ 10 mut/Mb, TMB-intermediate: 10 < and > 2 mut/Mb, TMB-low: ≤ 2 mut/Mb

EXAMPLE 2 TMB-high: ≥ 10 mut/Mb, TMB-low: < 10 mut/Mb

8.3.3 Approach of filtering germline variants

The method for filtering germline variants should be described in this field. If only tumor specimen was used, the database and version used for filtering germline variants can be specified.

EXAMPLE paired normal tissue used, tumor-only, tumor-only with dbSNP build 155

8.3.4 Variant types used for calculating TMB value

Variant types identified when selecting essential variants for calculating TMB value should be described.

EXAMPLE synonymous and non-synonymous, non-synonymous only

8.3.5 TMB region covered

The genomic region used for calculating TMB value within the target region should be indicated.

EXAMPLE 35,2 Mb

8.3.6 Calibrated TMB value

According to the definition of TMB, whole exome sequencing (WES) can be considered the standard method for generating data to calculate TMB value. However, the TMB values of each analysis site are varied by the variability of conditions for generating NGS data and calculating TMB – the type of

sequencing, panel size, and criteria of filtering variants. To facilitate the comparability across the sites, calibrated TMB value can be supplied through the process of calibration to estimate the TMB value of WES by internal or open-source calibration methods.

8.4 Sequencing information

8.4.1 Clinical sequencing date

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

8.4.2 Sequencing type

The sequencing type which was used for generating sequencing data should be described. The sequencing type may be determined according to the sequenced part of the whole genome and can be classified into WGS, WES, or targeted sequencing.

EXAMPLE WES or targeted sequencing

8.4.3 Quality control metrics

8.4.3.1 Percentage data quality > Q30

The Q30 rate is the proportion of the Phred quality score over 30 for each base in the sequencing read of the sequencing instrument. A Phred quality score is a measure of the quality of the identification of the nucleobases generated by automated DNA sequencing. If Phred assigns a quality score of 30 to a base, the chances that this base is called incorrectly are 1 in 1 000.

8.4.3.2 Depth of target

The depth of target is the mean number of sequencing depths aligned at each base of the target area. With hybridization-based capture, whole-genome sequencing or UMI (unique molecular identifier) tagging method, PCR duplicates should be removed.

8.4.4 Sequencing platform information

8.4.4.1 General

In this field, the condition which was used for generating sequencing data should be given, such as type of sequencer, library preparation method, target capture strategy if used, and information of read characters.

8.4.4.2 Type of sequencers

The specific sequencer that performs the sequencing may be given in the free text.

EXAMPLE Illumina®[1] MiSeq®[2], NextSeq®[3] 550, NovaSeq®[4] 6000, X10, THERMO SCIENTIFIC®[5] Ion PGM®[6], PacBio®[7] Sequel

8.4.4.3 Library preparation methods

The sequencing library preparation methods may be given in the free text.

EXAMPLE Sequencer compatible library kit (Illumina® TruSeq®[8] RNA library kit, THERMO SCIENTIFIC® Ion Total RNA-Seq Kit, ArcherDx®[9] fusion flex series, or any kind of used RNA seq library prep kits)

8.4.4.4 Target capture methods

The exome or targeted region capture methods should be notified.

EXAMPLE Amplicon, probe capture, or none.

8.4.4.5 Read type

Sequencing read type might be given using free text.

EXAMPLE Single-end, Paired-end

8.4.4.6 Read length

The sequencing read length information might be given using free text.

EXAMPLE 101 bp, 35-250 bp, <1 kb

8.4.5 Analysis platform information

8.4.5.1 General

The information on the analysis procedure with generated data might be included in the free text. This information might be used for 2nd analysis with TMB data from multiple sites.

---

1) Illumina® is the registered trademark of Illumina, Inc. This information is given for the convenience of users of this document and does not constitute an endorsement by ISO.

2) HiSeq® is the registered trademark of Illumina, Inc. This information is given for the convenience of users of this document and does not constitute an endorsement by ISO.

3) NextSeq® is the registered trademark of Illumina, Inc. This information is given for the convenience of users of this document and does not constitute an endorsement by ISO.

4) NovaSeq® is the registered trademark of Illumina, Inc. This information is given for the convenience of users of this document and does not constitute an endorsement by ISO.

5) THERMO SCIENTIFIC® is the registered trademark of Thermo Fisher Scientific Inc. This information is given for the convenience of users of this document and does not constitute an endorsement by ISO.

6) Ion PGM® is the trademark of Thermo Fisher Scientific Inc. This information is given for the convenience of users of this document and does not constitute an endorsement by ISO.

7) PacBio® is the trademark of Pacific Biosciences of California, Inc. This information is given for the convenience of users of this document and does not constitute an endorsement by ISO.

8) TruSeq® is the registered trademark of Illumina, Inc. This information is given for the convenience of users of this document and does not constitute an endorsement by ISO.

9) ArcherDx® is the registered trademark of ArcherDX, Inc. This information is given for the convenience of users of this document and does not constitute an endorsement by ISO.

8.4.5.2 Alignment tools

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

EXAMPLE BWA-MEM 0.7.17, Bowtie 2.4.4

8.4.5.3 Variants caller

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

EXAMPLE Mutact2 4.1.4.1, GATK 4.1.4.1

8.4.5.4 Annotation tool

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

EXAMPLE VEP release-104, ANNOVAR

Annex A
(informative)

Example structure of TMB report

Tables A.1 and A.2 show the composition of the TBM report with clinical DNA sequencing using example values.