

© ISO ##### - All rights reserved

ISO ~~/AWI~~ TS 4425:2022(E)

ISO TC 215/SC 1/WG 1

Secretariat: KATS

~~Date: 2022-11-17~~

Genomics informatics — Data elements and their metadata for describing the microsatellite instability (MSI) information of clinical massive parallel DNA sequencing

~~WD stage~~

Warning for WDs and CDs

This document is not an ISO International Standard. It is distributed for review and comment. It is subject to change without notice and may not be referred to as an International Standard.

Recipients of this draft are invited to submit, with their comments, notification of any relevant patent rights of which they are aware and to provide supporting documentation.

<https://standards.iteh.ai/catalog/standards/sist/44bd29a0-3c72-4a82-8f94-a7745a71c545/iso-dts-4425>

To help you, this guide on writing standards was produced by the ISO/TMB and is available at

A model manuscript of a draft International Standard (known as "The Rice Model") is available at

© ISO #### - All rights reserved

© ISO 20XX

iTeh STANDARD PREVIEW (standards.iteh.ai)

ISO/DTS 4425

<https://standards.iteh.ai/catalog/standards/sist/44bd29a0-3a72-4a82-8f94-a7745a71c545/iso-dts-4425>

© ISO 2023

All rights reserved. Unless otherwise specified, or required in the context of its implementation, no part of this publication may be reproduced or utilized otherwise in any form or by any means, electronic or mechanical, including photocopying, or posting on the internet or an intranet, without prior written permission. Permission can be requested from either ISO at the address below or ISO's member body in the country of the requester.

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 ~~www.iso.org~~

Published in Switzerland

iTeh STANDARD PREVIEW
(standards.iteh.ai)

ISO/DTS 4425

<https://standards.iteh.ai/catalog/standards/sist/44bd29a0-3a72-4a82-8f94-a7745a71c545/iso-dts-4425>

Contents

Foreword.....	7
Introduction	8
1 Scope.....	9
2 Normative references	9
3 Terms and definitions.....	9
4 Abbreviated terms	13
5 Microsatellite instability (MSI)	15
6 Composition of elements for describing MSI status on clinical DNA NGS report	16
6.1 General.....	16
6.2 Summary part.....	16
6.3 Detail part.....	16
7 Fields and their nomenclature of required data	16
7.1 General.....	16
Table 1 — Data elements, their metadata, and cardinality for required fields.....	17
7.2 Clinical sequencing order	18
7.2.1 General.....	18
7.2.2 Clinical sequencing order code	18
7.2.3 Date and time.....	18
7.3 Information on subject of care	19
7.3.1 General.....	19
7.3.2 Subject of care identifier	19
7.3.3 Subject of care name.....	19
7.3.4 Subject of care birth date	19
7.3.5 Subject of care sex	19
7.3.6 Subject of care ancestry.....	20
7.3.7 Referring diagnosis.....	20
7.4 Information on legally authorized person ordering clinical sequencing.....	20
7.4.1 General.....	20
7.5 Performing laboratory.....	20
7.5.1 General.....	20
7.5.2 Basic information on performing laboratory	20
7.5.3 Information on report generator	20
7.5.4 Information of legally confirmed person on sequencing report.....	20
7.6 Biospecimen information	20
7.6.1 General.....	20
7.6.2 Type of specimen.....	20
7.7 MSI status result information.....	21
7.7.1 General.....	21
7.7.2 MSI status	21
7.8 Recommended treatment.....	21
7.8.1 General.....	21
7.8.2 Medication	21
7.8.3 Clinical trial information	21
7.8.4 Other recommendations	21
7.8.5 Supporting information.....	21

8	Fields and their nomenclature of optional data	22
8.1	General.....	22
Table 2 — Data elements and their metadata for optional fields		
8.2	Reference genome information	23
8.3	MSI information	23
8.3.1	General.....	23
8.3.2	Criteria of MSI status	23
8.3.3	Genomic position for determining MSI status.....	23
8.3.4	Genomic position against markers of alternative method.....	23
8.3.5	Clinical implication of MSI status	23
8.4	Sequencing information.....	23
8.4.1	Clinical sequencing date.....	23
8.4.2	Sequencing type	24
8.4.3	Quality control metrics	24
8.4.4	Sequencing platform information	24
8.4.5	Analysis platform information	25
Annex A (informative)	Example structure of MSI status report.....	26
Bibliography.....		30

iTeh STANDARD PREVIEW
(standards.iteh.ai)

ISO/DTS 4425

<https://standards.iteh.ai/catalog/standards/sist/44bd29a0-3a72-4a82-8f94-a7745a71c545/iso-dts-4425>

Foreword

ISO (the International Organization for Standardization) is a worldwide federation of national standards bodies (ISO member bodies). The work of preparing International Standards is normally carried out through ISO technical committees. Each member body interested in a subject for which a technical committee has been established has the right to be represented on that committee. International organizations, governmental and non-governmental, in liaison with ISO, also take part in the work. ISO collaborates closely with the International Electrotechnical Commission (IEC) on all matters of electrotechnical standardization.

The procedures used to develop this document and those intended for its further maintenance are described in the ISO/IEC Directives, Part 1. In particular, the different approval criteria needed for the different types of ISO documents should be noted. This document was drafted in accordance with the editorial rules of the ISO/IEC Directives, Part 2 (see www.iso.org/directives).

Attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights. ISO shall not be held responsible for identifying any or all such patent rights. Details of any patent rights identified during the development of the document will be in the Introduction and/or on the ISO list of patent declarations received (see www.iso.org/patents).

Any trade name used in this document is information given for the convenience of users and does not constitute an endorsement.

For an explanation of the voluntary nature of standards, the meaning of ISO specific terms and expressions related to conformity assessment, as well as information about ISO's adherence to the World Trade Organization (WTO) principles in the Technical Barriers to Trade (TBT), see www.iso.org/iso/foreword.html.

This document was prepared by Technical Committee ISO/TC 215, *Health informatics*, Subcommittee SC 1, *Genomics informatics*.

Any feedback or questions on this document should be directed to the ~~user's~~ national standards body. A complete listing of these bodies can be found at www.iso.org/members.html.

Introduction

Massively parallel sequencing is a high-throughput analytical approach to nucleic acid sequencing that allows whole genomes, transcriptomes, and specific nucleic acid targets. These advanced technologies have been used in the clinical field, and clinical sequencing has been applied to realize personalized medicine and precision medicine. For clinical usage, ISO/TS 20428^[1] Health informatics — Data elements and their metadata for describing structured clinical genomic sequence information in electronic health records has been published. ISO/TS 20428^[1] has been developed for clinical usage.

In the field of cancer treatment, various treatment strategies were performed differently from traditional anti-cancer chemotherapies. One of those strategies is the control of human immune system that maintains the action to extract cancer cells. Recent outcomes of clinical trials show that this immune therapy is efficient for some patients who have the specific molecular character of their tumor mass, such as PD-L1 or CTLA4 surface protein expression.^[2] So, As a result, these molecular characters are used as biomarkers for selecting patients. In colon cancer, according to several clinical trials, it is reported that the status of MSI (microsatellite instability) is regarded as a biomarker that drugs based on immunotherapy are more efficient for the patient with MSI-H (high).^[3]

The status of MSI can be calculated and reported by small nucleotide deletion on a specific region of human genome reference with NGS sequencing.^[4] According to US FDA, four NGS sequencing products were approved for companion diagnostics. Among these products, three NGS sequencing provide MSI status and value on their NGS sequencing report. CLIA-certified labs or equivalent level agencies in countries also are servicing the MSI status from their methods.^[5] It is forecasted that more clinical NGS sequencing will be approved to report MSI.

However, there is no international standard for describing MSI status, value, and metadata. The previous ISO/TS 20428 focused on only DNA variations compared with the reference genome. According to some research results said that, MSI status and the way to describe it are different even if using the same sequencing data. This makes it difficult for clinicians and researchers not only to use MSI status results for clinical decisions but also for secondary analyzing purposes when receiving from more than one sequencing lab. Related metadata should be essential to expand the usage of MSI status results.

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

Genomics informatics — General guidelines for describing the microsatellite instability (MSI) information of clinical massive parallel DNA sequencing

1 Scope

This ~~Technical Specification document~~ identifies data elements and metadata to represent the information about microsatellite instability (MSI) for reporting the value of the biomarker using clinical massive parallel DNA sequencing.

This document covers information about the MSI test result and related data, such as used resources, data generation condition, and data processing information which are helpful to clinical diagnosis and research.

This document is not intended

- ~~to define~~for defining experimental protocols or methods for calculating the value of microsatellite instability (MSI).
- for the other biological species than human resource, or
- 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 27527:2010, *Health informatics — Provider identification*

3 Terms and definitions

For the purposes of this document, the following terms and definitions apply.

ISO and IEC maintain ~~terminological~~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

ISO/~~AWI-TS~~DTS 4425:2023(E)

[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]

3.2.1 deletion
contiguous removal of one or more bases from a genomic sequence

[SOURCE: ISO/IEC 23092-2:~~2019~~2020, 3.4]

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/TS 17822-~~1~~:~~2014~~:2020, 3.~~2019~~]

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 11238:2012, 2.1.16]~~

3.7 indel
insertion (3.8) or/and *deletion* (3.2)

[SOURCE: ISO/TS 20428:2017, 3.18]

3.8 insertion
contiguous addition of one or more bases into a genomic sequence

[SOURCE: ISO/IEC 23092-2:~~2019~~2020, 3.18]

3.9

10

microsatellite

repetitive DNA elements, also known as simple sequence repeats (SSR), consisting of short in tandem repeat motifs of one to a few nucleotides that tend to occur in non-coding DNA of eukaryotic genomes and that are sometimes referred to as variable number of tandem repeats (VNTRs)

[SOURCE: ISO 16577:2016(en), 3.111]

3.10**microsatellite instability****MSI**

condition of genetic hypermutability (predisposition to mutation) that results from impaired DNA mismatch repair (MMR)

3.11**DNA mismatch repair****MMR**

system for recognizing and repairing erroneous insertion, deletion, and misincorporation of bases that can arise during DNA replication and recombination, as well as repairing some forms of DNA damage

3.12**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 23092-2:20192020, 3.20]

3.13**polymerase chain reaction****PCR**

in vitro enzymatic technique to increase the number of copies of a specific DNA fragment by several orders of magnitude

[SOURCE: ISO 16577:20162022, 3.1496.47]

3.14**quality score****Q score****Phred quality score****Q score**

sequencing quality score of a given nucleotide base

Note 1 to entry: Q is defined by the following equation: $Q = -10\log_{10}(e)$, 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 %.

ISO/~~AWI-TSDTS~~ 4425:2023(E)

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/~~DIS~~ 20397-2:2021, 3.30]

3.15 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 and then starts another round of reading from the opposite end.

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

3.16 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 increases within the string from left to right.

[SOURCE: ISO/IEC 23092-1:2019/2020, 3.22]

3.17 ~~read~~ sequence read

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

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

3.18 ~~variation~~ sequence variation DNA sequence variation

differences of DNA sequence among individuals in a population

[SOURCE: ISO 25720:2009, 4.8]

3.19 ~~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]