TECHNICAL SPECIFICATION

ISO/TS 4425

First edition
2023-04

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

Informatique génomique — Éléments de données et leurs métadonnées pour décrire les informations relatives à l'instabilité des microsatellites (MSI) du séquençage massif parallèle d'ADN

Reference number
ISO/TS 4425:2023(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 ..... **v**
Introduction ..... **vi**
1 Scope **1**
2 Normative references **1**
3 Terms and definitions **1**
4 Abbreviated terms **5**
5 Microsatellite instability (MSI) **5**
6 Composition of elements for describing MSI status on clinical DNA NGS report **6**
6.1 General 6
6.2 Summary part 7
6.3 Detail part 7
7 Fields and their nomenclature of required data **7**
7.1 General 7
7.2 Clinical sequencing order 8
7.2.1 General 8
7.2.2 Clinical sequencing order code 8
7.2.3 Date and time 8
7.3 Information on subject of care 9
7.3.1 General 9
7.3.2 Subject of care identifier 9
7.3.3 Subject of care name 9
7.3.4 Subject of care birth date 9
7.3.5 Subject of care sex 9
7.3.6 Subject of care ancestry 9
7.3.7 Referring diagnosis 10
7.4 Information on legally authorized person ordering clinical sequencing 10
7.4.1 General 10
7.5 Performing laboratory 10
7.5.1 General 10
7.5.2 Basic information on performing laboratory 10
7.5.3 Information on report generator 10
7.5.4 Information of legally confirmed person on sequencing report 10
7.6 Biospecimen information 10
7.6.1 General 10
7.6.2 Type of specimen 10
7.7 MSI status result information 11
7.7.1 General 11
7.7.2 MSI status 11
7.8 Recommended treatment 11
7.8.1 General 11
7.8.2 Medication 11
7.8.3 Clinical trial information 11
7.8.4 Other recommendations 11
7.8.5 Supporting information 11
8 Fields and their nomenclature of optional data **12**
8.1 General 12
8.2 Reference genome information 12
8.3 MSI information 13
8.3.1 General 13
8.3.2 Criteria of MSI status 13
8.3.3 Genomic position for determining MSI status 13

8.3.4 Genomic position against markers of alternative method 13
8.3.5 Clinical implication of MSI status 13
8.4 Sequencing information 13
8.4.1 Clinical sequencing date 13
8.4.2 Sequencing type 13
8.4.3 Quality control metrics 14
8.4.4 Sequencing platform information 14
8.4.5 Analysis platform information 14
Annex A (informative) **Example structure of MSI status report** **16**
Bibliography **19**

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. 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 a specific molecular character of their tumor mass, such as PD-L1 or CTLA4 surface protein expression[2]. 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 immuno-therapy 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 standard for describing MSI status, value, and metadata. ISO/TS 20428 focuses on only DNA variations compared with the reference genome. According to some research results, 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 — Data elements and their metadata for describing the microsatellite instability (MSI) information of clinical massive parallel DNA sequencing

1 Scope

This 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

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

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

3.4
deoxyribonucleic acid
DNA
molecule that encodes genetic information in the nucleus of cells

[SOURCE: ISO 25720:2009, 4.7]

3.5
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]

3.6
exome
part of the genome formed by exons

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

3.7
gene
basic unit of hereditary material that encodes and controls the expression of a protein or protein subunit

3.8
indel
insertion (3.9) or/and *deletion* (3.3)

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

3.9
insertion
contiguous addition of one or more bases into a genomic sequence

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

3.10
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)

3.11
microsatellite instability
MSI
condition of genetic hypermutability (predisposition to mutation) that results from impaired DNA mismatch repair (MMR)

3.12
DNA mismatch repair
MMR
system for recognizing and repairing erroneous insertion, deletion, and misincorporation of bases that can rise during DNA replication and recombination, as well as repairing some forms of DNA damage

3.13
nucleotide
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]

3.14
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 16577:2022, 3.6.47]

3.15
quality 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: $Q = -10\log10(e)$, 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, 3.30]

3.16
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]

3.17
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 23092-1:2020, 3.22]

3.18
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]

3.19
variation
sequence variation
DNA sequence variation
differences of DNA sequence among individuals in a population

[SOURCE: ISO 25720:2009, 4.8]

3.20
small indel
insertion (3.9) or *deletion* (3.3) of 2 nucleotides to 100 nucleotides

[SOURCE: ISO/TS 20428:2017, 3.32]

3.21
subject of care
person who uses, or is a potential user of, a healthcare service

[SOURCE: ISO/TS 22220:2011, 3.2, modified — Note to entry and second preferred term deleted.]

3.22
target capture
method to capture genomic regions of interest from a DNA sample prior to sequencing

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

3.23
targeted sequencing
technique used for sequencing only selected/targeted genomic regions of interest from a DNA sample

[SOURCE: ISO/TS 22692:2020, 3.22, modified — Note to entry and second preferred term deleted.]

3.24
whole exome sequencing
WES
technique for sequencing the exomes of the protein-coding genes in a genome

3.25
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

ATC	Anatomical Therapeutic Chemical
CTLA4	Cytotoxic T-Lymphocyte Associated Protein 4
EBI	European Bioinformatics Institute
FHIR	Fast Healthcare Interoperability Resources
HL7®	Health Level Seven
IDMP	Identification of Medicinal Product
IMPID	Investigational MPID
INN	International Nonproprietary Names
MPID	Medicinal Product Identifier
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
WHO	World Health Organization
UTN	Universal Trial Number

5 Microsatellite instability (MSI)

The DNA mismatch repair (MMR) pathway plays an important role in the cell cycle process to recognize and repair mismatches during DNA replication. The major components are four key enzymes coded for by the following genes: MLH1, MSH2, MSH6, PMS2, and EPCAM. MMR function doesn't work when mutational inactivation in the five genes or epigenetic inactivation occurs. It is called Deficiency of mismatch repair (dMMR). One of the most related diseases is Lynch syndrome[6]. Lynch syndrome, also known as hereditary non-polyposis colorectal cancer (HNPCC), is the most common cause of hereditary colorectal cancer. People with Lynch syndrome are more likely to get colorectal cancer and other cancers at a younger age (under 50). Patients develop dMMR tumors following the inactivation of the second wild-type allele through somatic mutation, loss of heterozygosity, or epigenetic silencing. These alterations – mutation or epigenetic inactivation is related to not only Lynch syndrome but also revealed differences in the case of cancer type. However, both lead to the accumulation of short sequences of DNA repeated throughout the genome-specific location and an increased risk of malignant transformation in certain tissues. These tumors have a higher frequency of somatic mutations compared with non-dMMR cancers and are assumed to have a large range of tumor neoantigens (high tumor mutation burden) and a highly immunogenic signature, including a high proportion of tumor-infiltrating lymphocytes. Defective mismatch repair results in a high tumor mutation burden and abundant neo-antigen formation, which can be recognized by the host immune system. Microsatellite instability (MSI) is found in 1,5 % to 3,5 % of all human cancers, such as colorectal, endometrial, ovarian, and cancers of the stomach, small intestine, pancreas, biliary tract, and ureter. The human genome contains more than 19 million microsatellites, short tandem repeats of motifs of 1 nucleotide to 6 nucleotides, typically spanning 10 nucleotides to 60 nucleotides in total length[7]. However, if the MMR function doesn't work well, nucleotide error accumulates, especially in the human genomic position, including microsatellites.

Sometimes, certain polymorphic microsatellites can serve as an individual's molecular barcode, which can be used in forensic identification.

The status of MSI is used as a biomarker for predicting the prognosis of colorectal cancer or selecting patients who be more effective with immune therapy. Recent clinical studies have shown that MSI status predicts clinical benefits from immune checkpoint blockade (ICB) with PD-1/PD-L1 interaction inhibitors. In May 2017, a new drug-pembrolizumab (KEYTRUDA®[1)]) was approved, which is a humanized antibody against the programmed death receptor-1 (PD-1), for the treatment of patients with any advanced solid cancer harboring a high tumor mutation burden as measured by the presence of microsatellite instability (MSI-H) in the US[8]. In addition, combined therapy, including nivolumab and ipilimumab, was approved for MSI-H metastatic colorectal cancer. This diagnostic and treatment strategy was recommended through NCCN guidelines for many types of cancers. MSI status also has prognostic significance, most notably in colorectal cancer, where testing is recommended in clinical practice guidelines for all patients.

Traditionally, MSI status testing is most performed via PCR and/or IHC analysis of tumor tissue specimens. The IHC is a screening method for detecting four abnormal proteins translated with MMR pathway genes by staining. One main disadvantage of the IHC method is the inability to detect MSI caused by point mutations or small indel mutations in MMR proteins which are not functioned but can still produce a positive IHC result. The rate of false positives is higher because of this reason. Another method is PCR based method. In 1997, a study[22] profiled microsatellite loci, creating the first standardized PCR-based MSI panel using five microsatellites that accurately segregated 200 colorectal carcinomas with and without MSI. The following year, a National Cancer Institute working group adopted a modified microsatellite panel (termed the Bethesda panel), which included two poly-A homopolymers (BAT-25, BAT-26) and three dinucleotide repeats (D5S346, D2S123, D17S250)[9]. Microsatellite panels composed only of mononucleotide loci (poly-A homopolymers BAT-25, BAT-26, NR21, NR-24, NR-27/MONO-27) are now preferred because mononucleotides offer greater sensitivity than dinucleotides for detecting MSI-high (MSI-H) tumors. Currently, several MSI loci (5 of 8) PCR testing is commercialized and popular. However, it is known that the PCR method is limited to the number of target loci. To overcome the limitations of traditional methods of IHC and PCR, next-generation sequencing (NGS) assays and novel computational methods have been developed to detect MSI status. NGS technologies have been used as the diagnostic platform in the 2010s, and several targeted sequencing methods were approved as comprehensive tumor analysis or companion diagnostics. In addition, NGS can give the result of detected not only somatic alterations but also the status of MSI. The flexibility of selecting a target for MSI and easily increase in the number of targets make the NGS method for MSI to be more accurate. According to the market report, it is forecasted that the NGS method will be used more in the clinical diagnostic field. Therefore, the essential factors as the way of describing MSI results and metadata derived from the calculating process of MSI should be defined and standardized for clinical DNA sequencing reports.

6 Composition of elements for describing MSI status on clinical DNA NGS report

6.1 General

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

1) KEYTRUDA® is a registered trademark of Merck Sharp & Dohme Corp., a subsidiary of Merck & Co., Inc. This information is given for the convenience of users of this document and does not constitute an endorsement by ISO of the product named.

6.2 Summary part

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

6.3 Detail part

The detailed section should contain all the required fields and the selected optional fields. The required fields mainly focus on helping clinicians by providing the necessary MSI status result and information, interpretation results, and related treatments. They include all fields required for clinical practice. 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 MSI status result.

7 Fields and their nomenclature of required data

7.1 General

In the field of required data, data elements, their metadata, and cardinality were chosen for mainly clinical practice using MSI status reports with clinical DNA sequencing using NGS, including targeted DNA sequencing or 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). The summary of data elements, their metadata, and cardinality are shown in Table 1. Table A.1 provides 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
Information on subject of care	Identifier		ISO/TS 22220:2011	One
	Name			One
	Birth date		ISO 8601 (all parts)	One
	Sex		ISO/TS 22220:2011	One
	Ancestry		HL7® v3 Code System Ethnicity	One
	Referring diagnosis		ICD [10]	One
Information of legally authorized person ordering clinical sequencing			ISO/TS 27527:2010	One

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

Table 1 *(continued)*

Data elements		Metadata (Primary)	Cardinality
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 [11]	One
MSI result information	MSI status	ENUM ("MSI-H", "MSI-intermediate", "MSS")	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 MSI status, the 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. The ISO 8601 series remove unambiguity on day-date conventions.

EXAMPLE Date represented as YYYY-MM-DD.

7.2.3.2 Order date

The order date is the date when a clinician ordered the necessary sequencing test to measure MSI 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 can be different.

7.2.3.4 Specimen collection date

The specimen collection date is the date 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 MSI 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.

7.3.3 Subject of care name

The subject of care name indicates the name of the patient. Some parts of name can 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.

7.3.5 Subject of care sex

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

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

7.3.6 Subject of care ancestry

The ancestry of the subject of care shall be notified to represent their genetic origin. The ancestry information should be represented in accordance with the HL7®[2]v3 Code System Race[12]. Alternatively,

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

if there are national standards, those coding systems can be used, for example, HL7® FHIR US Core Implementation Guide, US Core Race Extension[13].

7.3.7 Referring diagnosis

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

7.4 Information on legally authorized person ordering clinical sequencing

7.4.1 General

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

7.5.4 Information of legally confirmed person on sequencing report

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

7.6 Biospecimen information

7.6.1 General

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

7.6.2 Type of specimen

The type of specimen can be represented in accordance with the 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 MSI status result information

7.7.1 General

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

7.7.2 MSI status

MSI status shall be reported as the classification of words by the judgment criteria. MSI status can be categorized into two or three stages, such as MSI-H, MSS or MSI-H, MSI-intermediate, MSS, or other words expressing the same meaning.

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 MSI status, the relevant clinical decision shall be provided with supporting information, including clinical guidelines, clinical trials, or academic articles.

7.8.2 Medication

Regarding the diagnosis and the result of MSI status concerning the subject of care, the recommended clinical care guide shall be described with the drug name. 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, NIH) can be used to represent the classification of active ingredients of drugs. International Nonproprietary Names (INN) it can be alternatively used as a drug name since INN facilitates the identification of pharmaceutical substances or active pharmaceutical ingredients. 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 using MSI status results. It should be represented in accordance with UTN (Universal Trial Number) used in the WHO International Clinical Trials Registry Platform. Alternatively, US ClinicalTrials.gov ID, EU EudraCT number or another domestic registry ID can be used.

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

7.8.4 Other recommendations

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 (e.g. 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 can be used to understand the process of calculating MSI status 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
MSI information	Criteria of MSI status		TEXT	One
	Genomic position for determining MSI status		NUMERIC	One
	Genomic position against markers of alternative method		TEXT	One
	Clinical implication of MSI status		TEXT	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 information

The version of the reference genome used for processing the generated reads or identifying genomic variants to calculate MSI status can be supplied. The reference sequence should be represented in accordance with Genome Reference Consortium Human Genome release ID or Locus Reference Genomic ID.

Reference sequences should be represented in accordance with Genome Reference Consortium Human Genome release ID[13]. If there is an update, the patch number should be appended.

EXAMPLE GRCh38.p13 (GRCh38 Patch Release 13).

LRG (Locus Reference Genomic) ID,[14] which is maintained by EBI, can be used as well.

EXAMPLE LRG_1.

8.3 MSI information

8.3.1 General

In this field, data elements or related data can be supplied, which can explain the condition of determining MSI status using NGS DNA sequencing result or would be useful for the translational research or merge of MSI status data generated by multiple sites.

8.3.2 Criteria of MSI status

The criteria for determining the status of MSI can be described in this field. If the criteria is not available, the evidence or confidence level of determining the MSI status can be described with an explanation.

EXAMPLE MSI-H: ≥90 % of positive in tested genomics position (region), MSS: <90 % of positive in tested genomics position (region)

8.3.3 Genomic position for determining MSI status

The number of genomic positions (region) for determining MSI status can be supplied in this field. The list of test positions would be supplied with the result of each position.

8.3.4 Genomic position against markers of alternative method

The field includes the genomic location (BAT25, BAT26, D2S123, D5S346, and D17S250) where alternative popular methods (qPCR) could be supplied.

8.3.5 Clinical implication of MSI status

In this field, the clinical implication of MSI status can be supplied not only about the tumor type of the prescription case but all tumor types in which the clinical translational research or usages have been reported. In addition, the result of MSI status about the prescription case can be explained for clinicians to get more information about supplied recommended treatment in the required data section and the characteristics of molecular status for prognosis with references.

8.4 Sequencing information

8.4.1 Clinical sequencing date

The clinical sequencing date is the date when the performing laboratory generates the MSI status results using the received specimen. The date should be represented in accordance with the ISO 8601 series 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 can 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. If Phred assigns a quality score of 30 to a base, the chances that this base is called incorrectly are 1 in 1000.

8.4.3.2 Depth of target

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 can be given in free text.

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

8.4.4.3 Library preparation methods

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

EXAMPLE Sequencer compatible library kit (Illumina TruSeq RNA library kit, Ion Total RNA-Seq Kit, ArcherDx fusion flex series, or any 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 can be given using free text.

EXAMPLE Single-end, Paired-end.

8.4.4.6 Read length

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

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

8.4.5 Analysis platform information

8.4.5.1 General

The information of analysis procedure with generated data can be included in free test. This information can be used for 2nd analysis of MSI status data from multiple sites.

8.4.5.2 Alignment tools

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

EXAMPLE BWA-MEM 0.7.17, Bowtie 2.4.4

8.4.5.3 Variant caller

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

EXAMPLE VEP release-104, ANNOVAR

Annex A
(informative)

Example structure of MSI status report

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