

**CATS**  
**(Cancer Genomic Test Standardized)**  
**Format**

**Document for details**

**By Section of Genomic Data Management,**  
**C-CAT**

**v1.4.0**

**2024/03/19**

# Contents

I. Introduction .....	4
I-1. Objectives .....	4
I-2. Terms .....	4
I-3. About the condition field .....	5
I-4. Format information .....	5
II. Matters specific to C-CAT .....	6
II-1. Sending format of files .....	6
II-2. Scope of inputs .....	6
II-3. Requests .....	6
II-4. Notes .....	6
III. metaData tag .....	8
III-1. schemaVersion key .....	8
III-2. referenceGenome tag .....	8
III-2-1. Tags within referenceGenome tag .....	8
III-2-2. Examples of referenceGenome tag .....	9
III-3. configOptions tag .....	9
III-3-1. Tags within configOptions tag .....	10
III-3-2. Example of configOptions tag .....	13
III-4. comments tag .....	15
III-4-1. Tags within comments tag .....	15
III-4-2. Example of comments tag .....	15
IV. testInfo tag .....	17
IV-1. Tags within testInfo tag .....	17
IV-2. Example of testInfo tag .....	17
V. sequencingSamples tag .....	19
V-1. Tags within sequencingSamples tag .....	19
V-2. Example of sequencingSamples tag .....	21
VI. variants tag .....	23
VI-1. shortVariants tag .....	23
VI-1-1. Tags within shortVariants tag .....	23
VI-1-2. Example of shortVariants tag .....	31
VI-2. copyNumberAlterations tag .....	34
VI-2-1. Tags within copyNumberAlterations tag .....	35
VI-2-2. Example of copyNumberAlterations tag .....	38
VI-3. rearrangements tag .....	39
VI-3-1. Tags within rearrangements tag .....	39

VI-3-2. Example of rearrangements tag.....	46
VII. otherBiomarkers tag.....	53
VII-1. Tags within otherBiomarkers tag.....	53
VII-2. Example of otherBiomarkers tag.....	54
VIII. expressions tag.....	56
VIII-1. Tags within expressions tag.....	56
VIII-2. Example of expressions tag.....	58
IX. armLevelChanges tag .....	59
IX-1. Tags within armLevelChanges tag .....	59
IX-2. Example of armLevelChanges tag .....	60
X. nonHumanContents tag .....	62
X-1. Tags within nonHumanContents tag .....	62
X-2. Example of nonHumanContents tag .....	63
XI. compositeBiomarkers tag .....	64
XI-1. Tags within compositeBiomarkers tag .....	64
XI-2. Example of compositeBiomarkers tag .....	65
XII. druggability tag.....	66
XII-1. Tags within druggability tag.....	66
XII-2. Example of druggability tag.....	67
XIII. Other notes .....	69
XIII-1. itemId.....	69
XIII-1-1. Example of itemId description.....	69
XIII-2. matePieceLocation.....	70
XIII-2-1. Example of matePieceLocation description.....	70
XIII-3. sampleMetrics.....	73
XIV. For inquires .....	73

## I. Introduction

### I-1. Objectives

Differences in the formats for results in comprehensive genomic profiling tests of cancer makes it difficult for a third party to annotate alterations with candidate drugs and clinical trials using the same software under a simple framework. To efficiently promote uniformity and homogeneity in the interpretation of cancer genomic test results, it is necessary to define a standardized format to present alteration data in cancer comprehensive genome profiling tests.

This document describes the CATS (cancer genomic test sandardized) format, a standardized format for presenting alteration data in cancer comprehensive genomic profiling tests. The data schema of the CATS format is defined by the JSON definition file "schema.json", the specifications of which are explained in this document.

This format is used as follows: A laboratory that performs cancer comprehensive genome profiling tests sends alteration data in the CATS format to a testing annotation organization such as C-CAT. The organization annotates alteration data in the CATS format with candidate drugs and clinical trials, taking into account the clinical data of patients. This format only applies to alteration data, not clinical data, because clinical data are stored in electronic medical records in hospitals and are unlikely to be accessible to testing companies.

### I-2. Terms

- *Testing annotation organization*: An organization that receives alteration data in cancer comprehensive genome profiling tests from testing companies and clinical data from hospitals and uses the cancer knowledge base to annotate each alteration with candidate drugs and clinical trials. For example, C-CAT.
- *Testing annotation document*: A document that associates individual alterations in a cancer comprehensive genomic profiling test with candidate drugs and clinical trials for each patient. An example is the C-CAT Findings document. In contrast to a testing annotation document, a testing report document is a report of test results issued by a testing company.
- *Cancer knowledge base*: A database that associates alterations in cancer with candidate drugs and clinical trials. Examples include C-CAT CKDB (cancer knowledge database) and OncoKB (Chakravarty et al, 2017, JCO Precision Oncology).

### I-3. About the condition field

- Required: Required field when the JSON parent tag exists. When the parent tag is optional, the required field is described as "[required]" because it is assumed that the parent tag exists.
- Optional: Recommended field that may be associated with drug and clinical trial information in testing annotation documents or that increases the accuracy of testing annotation documents according to testing reports issued by testing companies. If it is strongly recommended, it is described as "optional (recommended)".

### I-4. Format information

- Character code: UTF8
- Type: JSON
- Extension: json

## II. Matters specific to C-CAT

### II-1. Sending format of files

Alteration data in cancer comprehensive genomic profiling tests should be sent to C-CAT from the testing company in CATS (cancer genomic test standardized) format.

### II-2. Scope of inputs

Quality-assured data on alterations are in the scope of inputs in CATS format. Please do not input alterations that are determined as to be false positives. You can choose whether or not to show alterations and their associated knowledge base annotations in C-CAT Findings by the tag ("reported" and "grade") described later.

Note: "Alteration" indicates shortVariants, copyNumberAlterations, rearrangements, otherBiomarkers, expressions, armLevelChanges, nonHumanContents and compositeBiomarkers, and does not indicate each of the detailed items at lower levels. "Variant" indicates shortVariants, copyNumberAlterations and rearrangements.

With these above attentions, please be sure to input data on alterations in the testing company's test result report or equivalent report. Please be sure to output data on alterations that are approved by the concerned authorities in C-CAT Findings.

### II-3. Requests

- It is recommended to provide inputs for as many optional fields as possible. As a result, more information on drugs or clinical trials may be added to C-CAT Findings, and C-CAT Findings based on laboratory testing reports will be more accurate. Additionally, more information linked to these fields may be added in future versions, even if such information does not appear in the current version of C-CAT Findings. C-CAT may separately contact the testing company for inputs in optional fields as necessary.
- Please try your best to input quality-assured data on all alterations (including those with the tag of "reported": false, as explained below) in CATS format. Otherwise, in case of any changes in the format or specifications of laboratory testing reports, C-CAT may send inquiries to the testing company and the production of C-CAT Findings may be delayed.

### II-4. Notes

Detailed precautions specific to C-CAT are indicated by ‘\*’ in the Description below.

### III. metaData tag

This tag is used to define the metadata.

It contains 4 keys: schemaVersion, referenceGenome, configOptions, and comments.

Key	Condition	Data type	Description
metaData	required	object	Aggregation tag for metadata.

#### III-1. schemaVersion key

Key	Condition	Data type	Description
schemaVersion	required	string regex: ^[0-9\.\-]+	Schema version of this format.

#### III-2. referenceGenome tag

Key	Condition	Data type	Description
referenceGenome	required	object	For information on a reference genome sequence.

##### III-2-1. Tags within referenceGenome tag

Key	Condition	Data type	Description
name	optional	string regex: ^.+	Name of a reference genome sequence used in your test.
grcRelease	required	string regex: ^GRC.+	GRC (Genome Reference Consortium) release ID of a reference genome sequence.
descriptions	optional	array (length: 0-N, string regex: ^.+)	Description of the reference genome sequence in the name tag. See the contents tag within the comments tag for usable languages and new lines.



### III-2-2. Examples of referenceGenome tag

(Example1. for NCBI)

```
"referenceGenome": {  
  "name": "GRCh38.p13",  
  "grcRelease": "GRCh38.p13",  
  "descriptions": [  
    "Homo sapiens (human) genome assembly GRCh37 (hg19) from the Genome Reference Consortium."  
  ]  
}
```

(Example2. for UCSC)

```
"referenceGenome": {  
  "name": "hg38Patch11",  
  "grcRelease": "GRCh38.p11",  
  "descriptions": [  
    "GRCh38 Genome Reference Consortium Human Reference 38 (GCA_000001405.22))"  
  ]  
}
```

(Example3. for GDC)

```
"referenceGenome": {  
  "name": "GRCh38.d1.vd1",  
  "grcRelease": "GRCh38",  
  "descriptions": [  
    "Homo sapiens (human) genome assembly GRCh38 (hg38) from GDC, GRCh38.d1.vd1"  
  ]  
}
```

### III-3. configOptions tag

This tag controls matching to cancer knowledge bases, such as C-CAT CKDB.

Key	Condition	Data type	Description
configOptions	optional	object	Aggregation tag that controls matching to cancer knowledge bases.

### III-3-1. Tags within configOptions tag

Key	Condition	Data type	Description
typeLabelsInterpretedAsKbAmplification	optional	array (length: 1-4, string) [choice]	<p>The testing company's labels for variants, which are interpreted as copyNumberAlteration's(or rearrangement's) "amplification" (copy number amplification) in cancer knowledge bases.</p> <p>Choose from the following options (must not be duplicated in an array).</p> <ul style="list-style-type: none"> <li>• "copyNumberAlterationType: amplification"</li> <li>• "copyNumberAlterationType: gain"</li> <li>• "copyNumberAlterationType: duplication"</li> <li>• "rearrangementType: duplication"</li> </ul> <p>(Default: "copyNumberAlterationType: amplification", "copyNumberAlterationType: gain", "copyNumberAlterationType: duplication")</p>
typeLabelsInterpretedAsKbLoss	optional	array (length: 1-5, string) [choice]	<p>The testing company's labels for variants, which are interpreted as copyNumberAlteration's(or rearrangement's) "loss" (copy number loss) in cancer knowledge bases.</p> <p>Choose from the following options (must not be duplicated in an array).</p> <ul style="list-style-type: none"> <li>• "copyNumberAlterationType: loss"</li> <li>• "copyNumberAlterationType: deletion"</li> <li>• "copyNumberAlterationType: homozygous deletion"</li> <li>• "rearrangementType: deletion"</li> <li>• "rearrangementType: exon skipping"</li> </ul> <p>(Default: "copyNumberAlterationType: loss", "copyNumberAlterationType: deletion", "copyNumberAlterationType: homozygous deletion")</p>

typeLabelsInterpretedAsKbGeneFusion	optional	array (length: 1-4, string) [choice]	<p>The testing company's labels for variants, which are interpreted as rearrangement's "geneFusion" (gene fusion) in cancer knowledge bases.</p> <p>Choose from the following options (must not be duplicated in an array).</p> <ul style="list-style-type: none"> <li>• "rearrangementType: gene fusion"</li> <li>• "rearrangementType: gene fusion and frameshift variant"</li> <li>• "rearrangementType: bidirectional gene fusion"</li> <li>• "rearrangementType: other"</li> </ul> <p>(Default: "rearrangementType: gene fusion", "rearrangementType: gene fusion and frameshift variant", "rearrangementType: bidirectional gene fusion")</p>
typeLabelsInterpretedAsKbInversion	optional	array (length: 1-3, string) [choice]	<p>The testing company's labels for variants, which are interpreted as rearrangement's "inversion" (gene inversion) in cancer knowledge bases.</p> <p>Choose from the following options (must not be duplicated in an array).</p> <ul style="list-style-type: none"> <li>• "rearrangementType: inversion"</li> <li>• "rearrangementType: truncation"</li> <li>• "rearrangementType: other"</li> </ul> <p>(Default: "rearrangementType: inversion")</p>
typeLabelsInterpretedAsKbDeletion	optional	array (length: 1-8, string) [choice]	<p>The testing company's labels for variants, which are interpreted as rearrangement's(or copyNumberAlteration's) "deletion" (gene deletion) in cancer knowledge bases.</p> <p>Choose from the following options (must not be duplicated in an array).</p> <ul style="list-style-type: none"> <li>• "rearrangementType: deletion"</li> <li>• "rearrangementType: truncation"</li> <li>• "rearrangementType: splice variant"</li> <li>• "rearrangementType: exon skipping"</li> <li>• "rearrangementType: other"</li> <li>• "copyNumberAlterationType: deletion"</li> <li>• "copyNumberAlterationType: homozygous deletion"</li> <li>• "copyNumberAlterationType: loss"</li> </ul> <p>(Default: "rearrangementType: deletion")</p>

typeLabelsInterpretedAsKbDuplication	optional	array (length: 1-6, string) [choice]	<p>The testing company's labels for variants, which are interpreted as rearrangement's(or copyNumberAlteration's) "duplication" (gene duplication) in cancer knowledge bases.</p> <p>Choose from the following options (must not be duplicated in an array).</p> <ul style="list-style-type: none"> <li>• "rearrangementType: duplication"</li> <li>• "rearrangementType: tandem duplication"</li> <li>• "rearrangementType: other"</li> <li>• "copyNumberAlterationType: duplication"</li> <li>• "copyNumberAlterationType: amplification"</li> <li>• "copyNumberAlterationType: gain"</li> </ul> <p>(Default: "rearrangementType: duplication", "rearrangementType: tandem duplication")</p>
typeLabelsInterpretedAsKbTruncation	optional	array (length: 1-7, string) [choice]	<p>The testing company's labels for variants, which are interpreted as rearrangement's(or copyNumberAlteration's) "truncation" (gene truncation) in cancer knowledge bases.</p> <p>Choose from the following options (must not be duplicated in an array).</p> <ul style="list-style-type: none"> <li>• "rearrangementType: truncation"</li> <li>• "rearrangementType: deletion"</li> <li>• "rearrangementType: inversion"</li> <li>• "rearrangementType: other"</li> <li>• "copyNumberAlterationType: loss"</li> <li>• "copyNumberAlterationType: deletion"</li> <li>• "copyNumberAlterationType: homozygous deletion"</li> </ul> <p>(Default: "rearrangementType: truncation")</p>
typeLabelsInterpretedAsKbExonSkipping	optional	array (length: 1-4, string) [choice]	<p>The testing company's labels for variants, which are interpreted as rearrangement's "exon skipping" (exon skipping) in cancer knowledge bases.</p> <p>Choose from the following options (must not be duplicated in an array).</p> <ul style="list-style-type: none"> <li>• "rearrangementType: exon skipping"</li> <li>• "rearrangementType: splice variant"</li> <li>• "rearrangementType: deletion"</li> <li>• "rearrangementType: other"</li> </ul> <p>(Default: "rearrangementType: exon skipping")</p>

typeLabelsInterpretedAsTranslocation	optional	array (length: 1-2, string) [choice]	<p>The testing company's labels for variants, which are interpreted as rearrangement's "translocation" (translocation) in cancer knowledge bases.</p> <p>Choose from the following options (must not be duplicated in an array).</p> <ul style="list-style-type: none"> <li>• "rearrangementType: translocation"</li> <li>• "rearrangementType: other"</li> </ul> <p>(Default: "rearrangementType: translocation")</p>
typeLabelsInterpretedAsKbRearrangement	optional	array (length: 1-18, string) [choice]	<p>The testing company's labels for variants, which are interpreted as rearrangement's(or copyNumberAlteration's) "rearrangement" (gene rearrangement) in cancer knowledge bases.</p> <p>Choose from the following options (must not be duplicated in an array).</p> <ul style="list-style-type: none"> <li>• "rearrangementType: gene fusion"</li> <li>• "rearrangementType: gene fusion and frameshift variant"</li> <li>• "rearrangementType: bidirectional gene fusion"</li> <li>• "rearrangementType: duplication"</li> <li>• "rearrangementType: tandem duplication"</li> <li>• "rearrangementType: deletion"</li> <li>• "rearrangementType: inversion"</li> <li>• "rearrangementType: truncation"</li> <li>• "rearrangementType: splice variant"</li> <li>• "rearrangementType: exon skipping"</li> <li>• "rearrangementType: translocation"</li> <li>• "rearrangementType: other"</li> <li>• "copyNumberAlterationType: amplification"</li> <li>• "copyNumberAlterationType: gain"</li> <li>• "copyNumberAlterationType: duplication"</li> <li>• "copyNumberAlterationType: loss"</li> <li>• "copyNumberAlterationType: deletion"</li> <li>• "copyNumberAlterationType: homozygous deletion"</li> </ul> <p>(Default: "rearrangementType: other")</p>

### III-3-2. Example of configOptions tag

(Example)

```

"configOptions": {
  "typeLabelsInterpretedAsKbAmplification": [
    "copyNumberAlterationType: amplification",
    "copyNumberAlterationType: gain",
    "copyNumberAlterationType: duplication"
  ],
  "typeLabelsInterpretedAsKbLoss": [
    "copyNumberAlterationType: loss",
    "copyNumberAlterationType: deletion",
    "rearrangementType: deletion"
  ],
  "typeLabelsInterpretedAsKbGeneFusion": [
    "rearrangementType: gene fusion",
    "rearrangementType: gene fusion and frameshift variant",
    "rearrangementType: bidirectional gene fusion"
  ],
  "typeLabelsInterpretedAsKbInversion": [
    "rearrangementType: inversion"
  ],
  "typeLabelsInterpretedAsKbDeletion": [
    "rearrangementType: deletion"
  ],
  "typeLabelsInterpretedAsKbDuplicationDeletion": [
    "rearrangementType: duplication",
    "rearrangementType: tandem duplication"
  ],
  "typeLabelsInterpretedAsKbTruncation": [
    "rearrangementType: truncation"
  ]
}

```

### III-4. comments tag

You can comment on gene alterations (variants), biomarkers (otherBiomarkers), and information on sequencing samples (sequencingSamples). This tag contains the itemIds and contents tags.

Key	Condition	Data type	Description
comments	optional	array (length: 1-N, object)	Aggregation tag for comment information. Each object in the array must be unique.

#### III-4-1. Tags within comments tag

Key	Condition	Data type	Description
itemIds	required	array (length: 0-N, string regex: ^.+)\$)	The itemIds (multiple unique itemIds possible) to indicate alterations and information on sequencing samples (sequencingSamples). Possible to comment on a test overall if you set the length of this key to be zero. * The content will not be shown in C-CAT Findings, if itemId is specified.
contents	required	array (length: 1-N, string regex: ^.+)\$)	The content of the comment for itemId. The description can be in English or Japanese. Please use array elements if you make new lines, because we ignore line feed codes in this tag. * The total number of characters for each element is within 4,000 characters including the line feed code.

#### III-4-2. Example of comments tag

(Example)

```
"comments": [  
  {
```

```
    "itemIds": [],
```

**Note: If the length of the itemIds array is zero, it represents a comment on this test overall.**

```
    "contents": [  
      "Amplification of the FGFR1 gene is observed in 5 to 20% of squamous cell
```

```
      carcinomas, and it has been reported that FGFR1 is sensitive to FGFR inhibitors in  
      vitro.",
```

```
      "FGFR2 and FGFR3 gene activating mutations and FGFR3 gene fusions have been  
      reported one after another, and their frequency is low at around 3%, but therapeutic  
      effects with FGFR inhibitors are expected."
```

Note: If you want to make new lines in testing annotation documents, sentences or phrases should be separated as elements of an array.

```
]
},
{
  "itemIds": [
    "variant-1"
```

Note: The itemId of the mutation should be included to comment on a specific mutation.

```
],
  "contents": [
    "TSC1 functions independently of TSC2 and mTORC1."
  ]
},
{
  "itemIds": [
    "variant-1",
    "variant-5"
```

Note: You can comment on multiple mutations altogether by listing multiple itemIds.

```
],
  "contents": [
    "Although CD4 T cell percentage in Tsc1-/- mice was not strongly affected by Bim deficiency in vivo, TCR-mediated apoptosis of Tsc1-/- Bcl2l11-/- double knockout CD4 T cells was less pronounced compared with that of Tsc1-/- cells. (Kai Yang et al.)"
  ]
}
]
```



#### IV. testInfo tag

This tag is used to provide test information.

Key	Condition	Data type	Description
testInfo	required	object	Aggregation tag for test information.

##### IV-1. Tags within testInfo tag

Key	Condition	Data type	Description
testId	required	string regex: ^.+	Any ID used by the testing company.
testType	required	string [choice]	The combination of specimens used in the test. <ul style="list-style-type: none"><li>• "tumor-only": test using tumor samples only</li><li>• "tumor and matched-normal": test using tumor and matched normal samples</li><li>• "tumor-only (cell-free)": test using cell-free tumor samples only</li><li>• "tumor (cell-free) and matched-normal": test using cell-free tumor samples and normal samples</li></ul>
targetRegion Version	optional	string regex: ^.+	Target area version defined in the bed file.
softwareName	optional	string regex: ^.+	Name of gene analysis software.
softwareVersion	optional	string regex: ^.+	Version of gene analysis software.
panelName	required	string regex: ^.+	Name of your genomic profiling (gene panel) test. * Please inform C-CAT beforehand if you want to use a genomic test NOT approved to use under National Health Insurance.
panelVersion	required	string regex: ^.+	Version of your test.

##### IV-2. Example of testInfo tag

(Example)

```
"testInfo": {  
  "testId": "12345678901231900001",  
  "testType": "tumor and matched-normal",  
  "targetRegionVersion": "target region A",  
  "softwareName": "variant caller A",  
  "softwareVersion": "ver.1.2",  
  "panelName": "Multi-gene Panel A",  
  "panelVersion": "ver.1.03-00"  
}
```

## V. sequencingSamples tag

This tag is used to provide the information on sequencing results in the NGS run.

Key	Condition	Data type	Description
sequencingSamples	required	array (length: 1-N, object)	Aggregation tag for information on sequencing results. Each object in the array must be unique.

The maximum length of this array is 4, since the tumorOrNormal tag and the nucleicAcid tag can each take two values "tumor" or "normal" and "DNA" or "RNA", respectively.

### V-1. Tags within sequencingSamples tag

Key	Condition	Data type	Description
itemId	required	string regex: ^.+	The ID of the sample. It must be a unique string of characters within a single case.
tumorOrNormal	required	string [choice]	Whether the sequencing results are of tumor specimen or normal specimen. • "tumor" • "normal"
nucleicAcid	required	string [choice]	Whether the sequencing results are derived from DNA or RNA samples. • "DNA": DNA sample • "RNA": RNA sample
resultQuality	optional	object	The information of the sequencing results determined by the testing company.
state	[required]	string [choice]	The quality of the sequencing results. • "pass": Pass • "reference": Reference purpose only • "fail": Fail
reason	optional	string regex: ^.+	The reason and details for the state. e.g.) Possibility of contamination. Low read depth.
specimenType	optional	string regex: ^.+	The name of the specimen typewritten in the testing company's report. e.g.) FFPE Intact

allReadCount	optional	number	The number of all reads including unmapped reads.
sampleMetrics	optional (recommended)	array (length: 1-N, object)	Measured read information. Each object in the array must be unique.
value	[required]	number	Measured value such as read count and read depth. e.g.) 91.52
unit	[required]	string regex: ^.+ \$	Unit. Units may vary depending on test types. e.g.) % Input "null" for values without unit.
targetType	[required]	string [choice]	Type of target. Select one from: • "total": target and off target • "target": target * Inform C-CAT for other options.
duplicateType	[required]	string [choice]	Type of duplicate. Select one from: • "unique": unique read • "duplicate": duplicate read • "unique+duplicate": all read * Inform C-CAT beforehand if you want other options.
unitType	[required]	string [choice]	Type of value. Select one from: • "count": read count • "percentage": read percentage • "medianDepth": median depth • "meanDepth": mean depth * See "XIII-3. sampleMetrics" below for supplemental explanation when "percentage" is selected. * Inform C-CAT for other options.
tumorContent	optional	array (length: 1-N, object)	The value of the tumor content written in the testing company's report.
value	[required]	number	Tumor content.

			The unit is %.
type	[required]	string [choice]	Type of tumor content. <ul style="list-style-type: none"> <li>• "pathological": pathologically measured.</li> <li>• "estimate": estimated from sequence data.</li> <li>• "unclear": unclear.</li> </ul>

## V-2. Example of sequencingSamples tag

(Example)

```
"sequencingSamples": [
{
  "itemId": "sequence-1-tumor-dna",
  "tumorOrNormal": "tumor",
  "nucleicAcid": "DNA",
  "resultQuality": {
    "state": "pass"
  },
},
"specimenType": "FFPE",
"sampleMetrics": [
{
  "value": 91.52,
  "unit": "%",
  "targetType": "total",
  "duplicateType": "duplicate",
  "unitType": "percentage"
},
{
  "value": 87.31,
  "unit": "%",
  "targetType": "total",
  "duplicateType": "unique+duplicate",
  "unitType": "percentage"
},
{
  "value": 247,
  "unit": null,
  "targetType": "target",
  "duplicateType": "unique",
  "unitType": "medianDepth"
},
{
  "value": 238,
  "unit": null,
  "targetType": "target",
  "duplicateType": "unique",
  "unitType": "meanDepth"
}
]
```

```

    ]
  },
  {
    "itemId": "sequence-2-tumor-rna",
    "tumorOrNormal": "tumor",
    "nucleicAcid": "RNA",
    "resultQuality": {
      "state": "reference",
      "reason": "degraded"
    },
    "allReadCount": 11928428,
    "sampleMetrics": [
      {
        "value": 9542742,
        "unit": null,
        "targetType": "total",
        "duplicateType": "unique+duplicate",
        "unitType": "count"
      }
    ],
    "tumorContent": [
      {
        "value": 50,
        "type": "pathological"
      }
    ]
  },
  {
    "itemId": "sequence-3-normal-dna",
    "tumorOrNormal": "normal",
    "nucleicAcid": "DNA"
  }
]

```

## VI. variants tag

This tag is used to provide information on detected variants. It contains shortVariants tag, copyNumberAlterations tag, and rearrangements tag.

Key	Condition	Data type	Description
variants	optional	object	Aggregation tag for information on variants.

### VI-1. shortVariants tag

This tag is used to define information on SNV (single nucleotide variation), insertion, deletion, delins (simultaneous insertion and deletion), indel (insertion and deletion), and MNV (multiple-nucleotide variation).

Key	Condition	Data type	Description
shortVariants	optional	array (length: 1-N, object)	Aggregation tag for information on single nucleotide variation (SNV), insertion (insertion), deletion (deletion), deletion and insertion (delins), insertion and deletion (indel), and multiple-nucleotide variant (MNV) of nucleotides. Each object in the array must be unique.

#### VI-1-1. Tags within shortVariants tag

key	Condition	Data type	Description
itemId	required	string regex: ^.+	An ID assigned to an variant. It must be a unique string of characters within a single case.
chromosome	required	string regex: ^[a-zA-Z0-9_#-]+\$	Chromosome number.
position	required	integer	Physical position in a chromosome. Please use the 1-based coordinate system and describe according to VCF v4.3 (e.g., as stated on page 13 of VCF v4.3, when the

			reference base is atCga and the variant base is at-ga, and C position in the reference base is 3, denote the physical position as "position": 2, "referenceAllele": "TC", and "alternateAllele": "T". As described below, "referenceAllele" represents a reference base and "alternateAllele" represents a variant base.)
cytoband	optional	string regex: ^.+ \$	Cytoband where the variant exists.
referenceAllele	required	string regex: ^[ACGTN]+ \$	Reference base. Please describe according to VCF v4.3 (see the example above).
alternateAllele	required	string regex: ^[ACGTN¥*]+ \$	Variant base. Please describe according to VCF v4.3 (see the example above). Multi-alleles – tri-alleles and more – should be listed as different elements in the shortVariants tag. In that case, indicate the itemIds and that the variants are multi-alleles in the comments tag.
alternateAlleleFrequency	required	number	Variant allele frequency (ranging from 0 to 1).
totalReadDepth	optional (recommend)	integer	Total read depth (minimum value 1).
alternateAlleleReadDepth	optional (recommend)	integer	Variant allele read depth (minimum value 0).



variantType	optional	string [choice]	<p>Short variant type written in the report by the testing company. Select one from:</p> <ul style="list-style-type: none"> <li>• "SNV"</li> <li>• "insertion"</li> <li>• "deletion"</li> <li>• "delins"</li> <li>• "indel"</li> <li>• "MNV"</li> </ul> <p>* Please inform C-CAT beforehand if you use other types.</p>
transcripts	required	array (length: 1-N, object)	<p>Information on the representative transcript. Each object in the array must be unique.</p> <p>* When creating C-CAT Findings, the sequence is scanned from the beginning and the first transcript information for each gene is used.</p>
transcriptId	[required]	string regex: $^{\wedge}[\wedge\text{¥s}]+\text{\$}$	<p>Transcription product ID (e.g., NM_000368<sup><u>4</u></sup>).</p> <p>It is strongly recommended to include sub-numbers (underlined in red above) for accurate computation. If no transcriptId is present due to a mutation in an intergenic region, then you can define "transcriptId": null.</p>
transcriptDatabaseName	[required]	string [choice]	<p>Database of the transcript ID.</p> <p>Select from the following two options.</p> <ul style="list-style-type: none"> <li>• "RefSeq"</li> <li>• "Ensembl"</li> </ul>

			If the transcriptId is null, then you can define "transcriptDatabaseName": null.
transcriptDatabaseVersion	optional	string regex: ^.+	Version of the database. * If this is not defined, the version determined by C-CAT will be used.
regionStructure	optional	array (length: 1-N, object)	Information of the variant's structure.
type	[required]	string [choice]	Additional information on gene position, as written in the testing company's report. Select one from: <ul style="list-style-type: none"> <li>• "promoter"</li> <li>• "upstream"</li> <li>• "downstream"</li> <li>• "intergenic"</li> <li>• "5'UTR"</li> <li>• "3'UTR"</li> <li>• "exon"</li> <li>• "intron"</li> </ul> * Please consult C-CAT if you want to use other types.
number	optional	integer	If "exon" or "intron" is selected for type, input the exon number or intron number.
totalNumber	optional	integer	If "exon" or "intron" is selected for type, input the total exon number or total intron number.
geneSymbol	[required]	string regex: ^[^s]+	Name of the gene (gene symbol), as written in the testing report document by the testing company. Enter "geneSymbol": null if a gene does not exist.

			Enter an associated gene such as "TERT promoter", "TERT" should be listed. The "promoter" is described in the "regionType" tag.
regionName	optional	string regex: ^.+	Name of specific region where the alteration exists, which is written in the testing company's report. e.g. D4Z4 repeat
strand	optional	string [choice]	Direction of transcription. If the orientation is the same as that of the reference genome sequence, it is "+"; if the orientation is opposite, it is "-". If the transcriptId is null, then you can define "strand": null.
cdsChange	[required]	string regex: ^.+	Enter changes at the DNA level, as written in the testing company's report. In principle, the cdsChange should conform to HGVS. When an RNA is not transcribed as in an intergenic region, you can input null (in addition to the notation of non-coding regions such as n.*).
aminoAcidsChange	[required]	string regex: ^.+	Enter changes at the protein level, as written in the testing company's report. Amino acids should be written with a single letter. In principle, other amino acids should conform to HGVS.

			If no change is observed in amino acids as in an untranslated region, you can input null.
calculatedEffects	optional	array (length: 1-N, string regex: ^.+)\$	<p>Note the effects of alterations on transcripts, such as "splicing_variant", using Sequence Ontology terms.</p> <p>This field corresponds to "Effect (Sequence Ontology)" of the snpEff tool and VCF "Func.refGene" of the annovar tool. For annovar, this is explained at Output file 1 (refSeq gene annotation) on Gene-based Annotation in User Guide, whereby terms provided by annovar can be converted to Sequence Ontology terms. One term should be assigned to each element of an array.</p> <p>Each string in the array must be unique.</p>
skippedExonRanges	optional (recommend)	array (length: 1-N, object)	Exon with skipping potential.
transcriptId	[required]	string regex: ^[^¥s]+\$	Refer to the description in the transcripts tag.
transcriptDatabaseName	[required]	string [choice]	Refer to the description in the transcripts tag.
transcriptDatabaseVersion	optional	string regex: ^.+)\$	Refer to the description in the transcripts tag.
geneSymbol	[required]	string regex:	Refer to the description in the transcripts tag.

		$^{\wedge}[\text{^}\text{¥s}]+\text{\$}$	
strand	optional	string [choice]	Refer to the description in the transcripts tag.
exonRange	[required]	array (length: 2, integer)	Range of skipped exon numbers. Start and end exon numbers are included in the range.
totalExonNumber	optional	integer	Total exon number.
sampleItemId	required	string regex: $\text{^}.\text{+}\text{\$}$	Described the itemId of sequencing sample information (sequencing Samples). Represents sample information in which the mutation was detected.
variantOrigin	optional (recommend)	string [choice]	<p>Somatic or germline origin.</p> <p>* C-CAT uses a different cancer knowledge base, according to whether alterations are somatic or germline. If no input is provided, the knowledge base for somatic alterations is used.</p> <ul style="list-style-type: none"> <li>• "somatic": derived from somatic cells</li> <li>• "germline": derived from germline cells</li> <li>• "likely somatic": typically in a tumor-only test, the alteration is likely to be somatic, and the knowledge base for somatic alterations is used.</li> <li>• "likely germline": typically in a tumor-only test, the alteration is likely to be</li> </ul>

			germline, and the knowledge base for germline alterations is used.
matched	optional	boolean	Whether or not the variant is identified using both tumor and matched-normal samples.
approved	optional (recommend)	boolean	Whether the variant is approved (true) or not (false) in the testing company's test result report or equivalent report. (Default: true)
grade	optional	string [choice]	The variant grade treated in the testing company's report or the other related report. Select one from: <ul style="list-style-type: none"> <li>• "clear": unambiguous detection</li> <li>• "equivocal": equivocal detection, such as low confidence</li> <li>• "suppl": supplementary information related to variants regardless of the accuracy of detection.</li> </ul> (Default: "clear") * In the case of clear or equivocal, the variant is annotated with cancer knowledge base. Please consult C-CAT beforehand if you use other types.
reported	required	boolean	Whether or not the alteration is reported in testing company's report or similar documents.

			* In case of true, the variant is outputted to C-CAT finding document.
--	--	--	--

#### VI-1-2. Example of shortVariants tag

(Example1. for SNV)

```
{
  "itemId": "variant-1",
  Note: The itemId is a character string at the discretion of the testing company for the detected variant.
  "chromosome": "9",
  "position": 135781005,
  "cytoband": "q34.13",
  "referenceAllele": "C",
  "alternateAllele": "G",
  Note: Describe "position" and "referenceAllele", "alternateAllele" according to the rules of VCF v4.3.
  "alternateAlleleFrequency": 0.54,
  "alternateAlleleReadDepth": 108,
  "totalReadDepth": 200,
  "variantType": "SNV",
  "transcripts": [
    {
      "transcriptId": "NM_000368.4",
      "transcriptDatabaseName": "RefSeq",
      "transcriptDatabaseVersion": "Release 99",
      "regionStructure": [
        {
          "type": "exon",
          "number": 15,
          "totalNumber": 23
        }
      ],
      "geneSymbol": "TSC1",
      "cdsChange": "c.1960C>G",
      "aminoAcidsChange": "p.Q654E",
      "calculatedEffects": [
        "missense_variant"
      ]
    }
  ],
  "sampleItemId": "sequence-1-tumor-dna",
  "variantOrigin": "somatic",
  "matched": true,
  "reported": true
}
```

(Example2. for insertion)

```
{
  "itemId": "variant-2",
  "chromosome": "8",
  "position": 37553560,
  "cytoband": "p11.23",
  "referenceAllele": "A",
  "alternateAllele": "AAGCGGC",
  "alternateAlleleFrequency": 0.4953,
  "alternateAlleleReadDepth": 368,
  "totalReadDepth": 743,
  "variantType": "insertion",
  "transcripts": [
    {
      "transcriptId": "NM_025069.2",
      "transcriptDatabaseName": "RefSeq",
      "transcriptDatabaseVersion": "Release 99",
      "regionStructure": [
        {
          "type": "exon",
          "number": 1,
          "totalNumber": 2
        }
      ],
      "geneSymbol": "ZNF703",
      "cdsChange": "c.63_64insAGCGGC",
      "aminoAcidsChange": "G21_G22insSG"
    }
  ],
  "sampleItemId": "sequence-1-tumor-dna",
  "variantOrigin": "somatic",
  "matched": true,
  "reported": true
}
```

(Example3. for deletion)

```
{
  "itemId": "variant-3",
  "chromosome": "1",
  "position": 27097751,
  "cytoband": "p.36.11",
  "referenceAllele": "TC",
  "alternateAllele": "T",
  "alternateAlleleFrequency": 0.1203,
  "alternateAlleleReadDepth": 32,
  "totalReadDepth": 266,
  "variantType": "deletion",
  "transcripts": [
    {
      "transcriptId": "ENST00000324856.13",

```



```

"transcriptDatabaseName": "Ensembl",
"transcriptDatabaseVersion": "v99",
"regionStructure": [
  {
    "type": "exon",
    "number": 12,
    "totalNumber": 20
  }
],
"geneSymbol": "ARID1A",
"cdsChange": "c.3340delC",
"aminoAcidsChange": "p.P1115fs*46",
"calculatedEffects": [
  "frameshift_variant"
]
},
"sampleItemId": "sequence-3-normal-dna",
"variantOrigin": "somatic",
"matched": true,
"reported": true
}

```

(Example4. for delins)

```

{
  "itemId": "variant-4",
  "chromosome": "1",
  "position": 26696982,
  "cytoband": "p36.11",
  "referenceAllele": "GC",
  "alternateAllele": "TT",
  "alternateAlleleFrequency": 0.0993,
  "alternateAlleleReadDepth": 52,
  "totalReadDepth": 524,
  "variantType": "delins",
  "transcripts": [
    {
      "transcriptId": "NM_006015.6",
      "transcriptDatabaseName": "RefSeq",
      "transcriptDatabaseVersion": "Release 99",
      "regionStructure": [
        {
          "type": "exon",
          "number": 1,
          "totalNumber": 20
        }
      ],
      "geneSymbol": "ARID1A",
      "cdsChange": "c.579_580delinsTT",
      "aminoAcidsChange": "p.E193_P194delinsDS"
    }
  ]
}

```

```

],
"sampleItemId": "sequence-1-tumor-dna",
"variantOrigin": "somatic",
"matched": true,
"reported": true
}

```

(Example5. for "TERT promoter")

```

{
  "itemId": "variant-5",
  "chromosome": "5",
  "position": 1295113,
  "cytoband": "p15.33",
  "referenceAllele": "G",
  "alternateAllele": "A",
  "alternateAlleleFrequency": 0.163,
  "alternateAlleleReadDepth": 15,
  "totalReadDepth": 92,
  "variantType": "SNV",
  "transcripts": [
    {
      "transcriptId": "ENST00000310581.9",
      "transcriptDatabaseName": "Ensembl",
      "transcriptDatabaseVersion": "Release 99",
      "regionStructure": [
        {
          "type": "promoter"
        }
      ],
      "geneSymbol": "TERT",
      "regionName": "TERT promoter",
      "cdsChange": "n.1295113C>T",
      "aminoAcidsChange": null,
      "calculatedEffects": [
        "TF_binding_site_variant"
      ]
    }
  ],
  "sampleItemId": "sequence-1-tumor-dna",
  "variantOrigin": "somatic",
  "matched": true,
  "reported": true
}

```

## VI-2. copyNumberAlterations tag

This tag is used to provide information on copy number alterations (CNAs).

Key	Condition	Data type	Description
-----	-----------	-----------	-------------

copyNumberAlterations	optional	array (length: 1-N, object)	Aggregation tag for information on copy number alterations (CNAs). Each object in the array must be unique.
-----------------------	----------	--------------------------------	--

#### VI-2-1. Tags within copyNumberAlterations tag

Key	Condition	Data type	Description
itemId	required	string regex: ^.+	An ID assigned to an variant. It must be a unique string of characters within a single case.
chromosome	optional (recommend)	string regex: ^[a-zA-Z0-9_#-]+	Chromosome number.
startPosition	optional (recommend)	integer	Physical starting position in a chromosome. Please use the 1-based coordinate system.
startCytoband	optional	string regex: ^.+	Cytoband at the starting position.
endPosition	optional (recommend)	integer	Physical ending position in a chromosome. Please use the 1-based coordinate system.
endCytoband	optional	string regex: ^.+	Cytoband at the ending position.
copyNumberMetrics	optional (recommend)	array (length: 1-N, object)	Measurements and units of the copy number alteration. Array of objects composed of two keys, value and unit. If there are two or more values in different units, register them as an array with the length of 2 or more. Each object in the array must be unique.
value	[required]	number	Aberrated copy number measurement.

unit	[required]	string [choice]	Unit for measured value. Select one from: <ul style="list-style-type: none"> <li>• "copy number": Copy number</li> <li>• "fold-change": Ratio of (standardized) reading depth of tumor samples to normal samples</li> <li>• "log2 fold-change": log2 transformation of "fold-change"</li> <li>• "fraction-of-gene": fraction of a CNA region to the gene region of interest</li> </ul> * Please consult C-CAT if you want to use other units.
copyNumberAlterationType	required	string [choice]	CNA type written in the report by the testing company. Select one from: <ul style="list-style-type: none"> <li>• "amplification"</li> <li>• "gain"</li> <li>• "duplication"</li> <li>• "loss"</li> <li>• "deletion"</li> <li>• "homozygous deletion"</li> <li>• "neutral"</li> </ul> * Please consult C-CAT if you want to use other types.
transcripts	required	array (length: 1-N, object)	Refer to the description in the shortVariants tag.
transcriptId	optional	string regex: ^[^¥s]+\$	Refer to the description in the shortVariants tag.
transcriptDatabaseName	optional	string [choice]	Refer to the description in the shortVariants tag. If the transcriptId is entered, this key is recommended to input as well.
transcriptDatabaseVersion	optional	string regex: ^\.+\$	Refer to the description in the shortVariants tag.
regionStructure	optional	array (length: 1-N, object)	Refer to the description in the shortVariants tag.

type	[required]	string [choice]	Refer to the description in the shortVariants tag.
number	optional	integer	Refer to the description in the shortVariants tag.
totalNumber	optional	integer	Refer to the description in the shortVariants tag.
geneSymbol	[required]	string regex: ^[^\$]+	Refer to the description in the shortVariants tag.
regionName	optional	string regex: ^.+	Refer to the description in the shortVariants tag.
strand	optional	string [choice]	Refer to the description in the shortVariants tag.
cdsChange	optional	string regex: ^.+	Refer to the description in the shortVariants tag.
aminoAcidsChange	optional	string regex: ^.+	Refer to the description in the shortVariants tag.
calculatedEffects	optional	array (length: 0-N, string regex: ^.+)	Refer to the description in the shortVariants tag.
skippedExonRanges	optional (recommend)	array (length: 1-N, object)	Refer to the description in the shortVariants tag.
transcriptId	[required]	string regex: ^[^\$]+	Refer to the description in the shortVariants tag.
transcriptDatabaseName	[required]	string [choice]	Refer to the description in the shortVariants tag.
transcriptDatabaseVersion	optional	string regex: ^.+	Refer to the description in the shortVariants tag.

geneSymbol	[required]	string regex: ^[^¥s]+\$	Refer to the description in the shortVariants tag.
strand	optional	string [choice]	Refer to the description in the shortVariants tag.
exonRange	[required]	array (length: 2, integer)	Refer to the description in the shortVariants tag.
totalExonNumber	optional	integer	Refer to the description in the shortVariants tag.
sampleItemId	required	string regex: ^.+ \$	Refer to the description in the shortVariants tag.
variantOrigin	optional (recommend)	string [choice]	Refer to the description in the shortVariants tag.
matched	optional	boolean	Refer to the description in the shortVariants tag.
approved	optional (recommend)	boolean	Refer to the description in the shortVariants tag.
grade	optional	string [choice]	Refer to the description in the shortVariants tag.
reported	required	boolean	Refer to the description in the shortVariants tag.

#### VI-2-2. Example of copyNumberAlterations tag

(Example)

```
{
  "itemId": "variant-9",
  "chromosome": "1",
  "startPosition": 8921059,
  "startCytoband": "p36.23",
  "endPosition": 8939151,
  "endCytoband": "p36.23",
  "copyNumberMetrics": [
    {
      "value": 0.2309,
      "unit": "fold-change"
    },
    {
      "value": -2.1147,
      "unit": "log2 fold-change"
    }
  ],
}
```

```

"copyNumberAlterationType": "loss",
"transcripts": [
{
  "transcriptDatabaseName": "RefSeq",
  "transcriptDatabaseVersion": "Release 99",
  "geneSymbol": "EN01"
}
],
"sampleItemId": "sequence-1-tumor-dna",
"variantOrigin": "somatic",
"matched": true,
"reported": false
}

```

### VI-3. rearrangements tag

This tag is used to provide information on rearrangements such as fusions, duplications, large deletions, and inversions.

Key	Condition	Data type	Description
rearrangements	optional	array (length: 1-N, object)	Aggregation tag for information on rearrangements, such as fusions, duplications, large deletions, and inversions. Each object in the array must be unique.

#### VI-3-1. Tags within rearrangements tag

Key	Condition	Data type	Description
itemId	required	string regex: ^.+	An ID assigned to a variant. It must be a unique string of characters within a single case.
breakends	required	array (length: 2, object)	Two breakends of the rearrangement. Each object in the array must be unique.
chromosome	required	string regex: ^[a-zA-Z0-9_#-]+\$	Chromosome number.

startPosition	required	integer	Physical starting position in a chromosome. Please use the 1-based coordinate system.
startCytoband	optional	string regex: ^.+ \$	Cytoband at the starting position.
endPosition	required	integer	Physical ending position in a chromosome. Please use the 1-based coordinate system.
endCytoband	optional	string regex: ^.+ \$	Cytoband at the ending position.
referenceAllele	optional (recommend)	string regex: ^[ACGTN]+\$	Reference base. Please describe according to VCF v4.3. See "XIII-2. matePieceLocation" below for a detailed explanation.
alternateAllele	optional (recommend)	string regex: ^[ACGT¥[¥]:0-9]+\$	Variant base. Please describe according to VCF v4.3. See "XIII-2. matePieceLocation" below for a detailed explanation.
matePieceLocation	optional (recommend)	string [choice]	If the sequence of interest is bound to another sequence on the upstream (or downstream) of this sequence of interest along the reference genome sequence, it is input as "upstream" (or "downstream"). It is strongly recommended to input this key for accurately identifying the genomic changes. See "XIII-2. matePieceLocation" below for a detailed explanation.
supportingReadCount	optional	integer	The number of reads that support the breakends, or the numerator of alternateAlleleFrequency.
totalReadCount	optional	integer	The total number of reads that support and do not support breakends, or the



			denominator of alternateAlleleFrequency.
alternateAlleleFrequency	optional (recommend)	number	Variant allele frequency of a breakend (ranging from 0 to 1). Please input a variant allele frequency calculated for each breakend.
transcripts	required	array (length: 1-N, object)	Refer to the description in the shortVariants tag.
transcriptId	optional	string regex: ^[^¥s]+\$	Refer to the description in the shortVariants tag.
transcriptDatabaseName	optional	string [choice]	Refer to the description in the shortVariants tag. If the transcriptId is entered, this key is recommended to input as well.
transcriptDatabaseVersion	optional	string regex: ^.+ \$	Refer to the description in the shortVariants tag.
regionStructure	optional (recommend)	array (length: 1-N, object)	Refer to the description in the shortVariants tag.
type	[required]	string [choice]	Refer to the description in the shortVariants tag.
number	optional (recommend)	integer	Refer to the description in the shortVariants tag.
totalNumber	optional	integer	Refer to the description in the shortVariants tag.
geneSymbol	[required]	string regex: ^[^¥s]+\$	Refer to the description in the shortVariants tag.

regionName	optional (recommend)	string regex: ^.+ \$	Refer to the description in the shortVariants tag.
strand	optional	string [choice]	Refer to the description in the shortVariants tag.
cdsChange	optional	string regex: ^.+ \$	Refer to the description in the shortVariants tag.
aminoAcidsChange	optional	string regex: ^.+ \$	Refer to the description in the shortVariants tag.
calculatedEffects	optional	array (length: 1-N, string regex: ^.+ \$)	Refer to the description in the shortVariants tag.
orderedGenePairs	optional (recommend)	array (length: 1-N, array)	<p>An array of an ordered pair(s) of gene names that traverses a breakpoint of a rearrangement, where the pair(s) is written in the transcriptional order.</p> <p>This array has, as a child element(s), an array(s) in length 2 that stores a pair of geneSymbols of transcripts. The geneSymbols in each pair should be aligned in the transcriptionally forward direction and should be unique. For example, for geneSymbols A and B, if A is in the upstream of transcription and B is the downstream, it is represented as</p> <p>"orderedGenePairs": [ [ "A", "B" ] ].</p> <p>* This directional information is taken into account in the cancer knowledge base search. For example, in the case of</p>

			<p>[ ["A", "B"] ], ["A", "B"] alone is searched. In the case of [ ["A", "B"], ["B", "A"] ], both ["A", "B"] and ["B", "A"] are respectively searched. You can control the search in more detail, such as by [ ["A", "B"], ["B", "A"], ["C", "B"] ].</p> <p><u>* Please specify such an appropriate input in this tag if you want to control how gene pairs are searched.</u></p> <p>* If this tag is omitted, the pair of geneSymbols under the breakends tag is treated as a pair without information on the transcriptional direction, and both direction of the pair will be respectively searched in the knowledge base.</p>
(Child element of orderedGenePairs)	required	array (length: 2, string regex: ^.+)\$	An array of strings of length 2 with the geneSymbol described in transcripts aligned in the forward direction of transcription. e.g.) ["EML4", "ALK"]
affectedGenes	optional (recommend)	array (length: 1-N, string regex: ^.+)\$	Genes affected by gene rearrangement. If multiple genes are deleted due to rearrangement, this key is recommended to input.
skippedExonRanges	optional (recommend)	array (length: 1-N, object)	Exon information skipped by rearrangement. If “exon skipping” is selected for rearrangementType, this key is recommended to input as well.
transcriptId	[required]	string regex: ^[^\s]+\$	Refer to the description in the shortVariants tag.
transcriptDatabaseName	[required]	string [choice]	Refer to the description in the shortVariants tag.

			If the transcriptId is entered, this key is recommended to input as well.
transcriptDatabaseVersion	optional	string regex: ^.+	Refer to the description in the shortVariants tag.
geneSymbol	[required]	string regex: ^[^\$]+	Refer to the description in the shortVariants tag.
strand	optional	string [choice]	Refer to the description in the shortVariants tag.
exonRange	[required]	array (length: 2, integer)	Range of skipped exon numbers. Start and end exon numbers are included in the range.
totalExonNumber	optional	integer	Refer to the description in the shortVariants tag.
insertedSequence	optional	string regex: ^[ACGTN]+\$	Sequence inserted between the two breakends of the genome sequence. If an inserted sequence does not exist, input null.
supportingReadCount	optional	integer	The number of reads that support rearrangement, or the numerator of alternateAlleleFrequency.
totalReadCount	optional	integer	The total number of reads that support and do not support rearrangement, or the denominator of alternateAlleleFrequency.
alternateAlleleFrequency	optional (recommend)	number	Variant allele frequency of a rearrangement (ranging from 0 to 1). Please input a value calculated for a rearrangement.
expressionLevelMetrics	optional	array (length: 1-N, object)	Information on expression levels in RNA-seq. Each object in the array must be unique.
value	[required]	number	Value of an expression level
unit	[required]	string [choice]	Unit of an expression level. Select one from the following options:

			<ul style="list-style-type: none"> <li>• "TPM"</li> <li>• "FPKM"</li> <li>• "FPM"</li> <li>• "RPKM"</li> <li>• "RPM"</li> </ul> <p>* Please consult C-CAT if you want to use other units.</p>
type	optional	string [choice]	<p>Type of summary statistics. Select one from the following options:</p> <ul style="list-style-type: none"> <li>• "mean"</li> <li>• "median"</li> <li>• "standard deviation"</li> </ul> <p>(Default: null)</p> <p>If the expression level represents a measured value (not a summary statistic), you can input null.</p> <p>* Please consult C-CAT if you want to use other choices.</p>
sampleSize	optional	integer	<p>Sample size when calculating summary statistics.</p> <p>* If a non-null value is inputted in type, it is recommended to input this item as well.</p> <p>(Default: null)</p>
isControl	optional	boolean	<p>Whether the expression level was taken from a control (normal) sample (true) or from a tumor sample (false).</p> <p>(Default: false)</p>
rearrangementNames	optional (recommend)	array (length: 1-N, string regex: ^.+)\$	<p>Individual rearrangement name given by the testing company. For example, "EML4-ALK fusion"</p> <p>Each string in the array must be unique.</p>
rearrangementType	required	string [choice]	<p>Rearrangement type written in the report by the testing company.</p> <p>Select one from:</p> <ul style="list-style-type: none"> <li>• "gene fusion"</li> <li>• "gene fusion and frameshift variant"</li> </ul>

			<ul style="list-style-type: none"> <li>• "bidirectional gene fusion"</li> <li>• "duplication"</li> <li>• "tandem duplication"</li> <li>• "deletion"</li> <li>• "inversion"</li> <li>• "truncation"</li> <li>• "splice variant"</li> <li>• "exon skipping"</li> <li>• "translocation"</li> <li>• "other"</li> </ul> <p>* Please consult C-CAT if you want to use other types.</p>
sampleItemId	required	string regex: ^.+	Refer to the description in the shortVariants tag.
variantOrigin	optional (recommend)	string [choice]	Refer to the description in the shortVariants tag.
matched	optional	boolean	Refer to the description in the shortVariants tag.
approved	optional (recommend)	boolean	Refer to the description in the shortVariants tag.
grade	optional	string [choice]	Refer to the description in the shortVariants tag.
reported	required	boolean	Refer to the description in the shortVariants tag.

### VI-3-2. Example of rearrangements tag

(Example1. In cases where orderedGenePairs is present)

```
{
  "itemId": "variant-13",
  "breakends": [
    {
      "chromosome": "2",
      "startPosition": 42510050,
      "startCytoband": "p21",
      "endPosition": 42510050,
      "endCytoband": "p21",
      "referenceAllele": "A",
      "alternateAllele": "A[1:29445240[",
      "matePieceLocation": "downstream",
      "transcripts": [
```

```

{
  "transcriptId": "NM_001145076.3",
  "transcriptDatabaseName": "RefSeq",
  "transcriptDatabaseVersion": "Release 99",
  "regionStructure": [
    {
      "type": "exon",
      "number": 7,
      "totalNumber": 22
    }
  ],
  "geneSymbol": "EML4"
}
],
{
  "chromosome": "2",
  "startPosition": 29445240,
  "startCytoband": "p23.2",
  "endPosition": 29445240,
  "endCytoband": "p23.2",
  "referenceAllele": "A",
  "alternateAllele": "]2:42510050]A",
  "matePieceLocation": "upstream",
  "transcripts": [
    {
      "transcriptId": "NM_001353765.2",
      "transcriptDatabaseName": "RefSeq",
      "transcriptDatabaseVersion": "Release 99",
      "regionStructure": [
        {
          "type": "exon",
          "number": 3,
          "totalNumber": 10
        }
      ],
      "geneSymbol": "ALK"
    }
  ]
}
],
"orderedGenePairs": [
  ["EML4", "ALK"]
],
"supportingReadCount": 30,
"totalReadCount": 430,
"alternateAlleleFrequency": 0.07,
"rearrangementType": "gene fusion",
"sampleItemId": "sequence-1-tumor-dna",
"variantOrigin": "somatic",
"matched": true,
"reported": false

```

```
}
```

(Example2. When describing the totalReadCount and alternateAlleleFrequency for each breakend)

```
{
  "itemId": "variant-14",
  "breakends": [
    {
      "chromosome": "2",
      "startPosition": 42510050,
      "startCytoband": "p21",
      "endPosition": 42510050,
      "endCytoband": "p21",
      "matePieceLocation": "downstream",
      "totalReadCount": 330,
      "alternateAlleleFrequency": 0.06,
      "transcripts": [
        {
          "transcriptId": "NM_001145076.3",
          "transcriptDatabaseName": "RefSeq",
          "transcriptDatabaseVersion": "Release 99",
          "regionStructure": [
            {
              "type": "exon",
              "number": 7,
              "totalNumber": 22
            }
          ],
          "geneSymbol": "EML4"
        }
      ]
    },
    {
      "chromosome": "2",
      "startPosition": 29445240,
      "startCytoband": "p23.2",
      "endPosition": 29445240,
      "endCytoband": "p23.2",
      "matePieceLocation": "upstream",
      "totalReadCount": 570,
      "alternateAlleleFrequency": 0.07,
      "transcripts": [
        {
          "transcriptId": "NM_001353765.2",
          "transcriptDatabaseName": "RefSeq",
          "transcriptDatabaseVersion": "Release 99",
          "regionStructure": [
            {
              "type": "exon",
              "number": 3,
```



```

        "totalNumber": 10
      }
    ],
    "geneSymbol": "ALK"
  }
]
}
],
"orderedGenePairs": [
  ["EML4", "ALK"]
],
"supportingReadCount": 30,
"rearrangementType": "gene fusion",
"sampleItemId": "sequence-1-tumor-dna",
"variantOrigin": "somatic",
"matched": true,
"reported": false
}

```

(Example3. In cases where insertedSequence is present)

```

{
  "itemId": "variant-15",
  "breakends": [
    {
      "chromosome": "14",
      "startPosition": 234567,
      "startCytoband": "p13",
      "endPosition": 234567,
      "endCytoband": "p13",
      "matePieceLocation": "downstream",
      "transcripts": [
        {
          "transcriptId": null,
          "transcriptDatabaseName": "RefSeq",
          "transcriptDatabaseVersion": "Release 99",
          "geneSymbol": null
        }
      ]
    }
  ],
  {
    "chromosome": "2",
    "startPosition": 321672,
    "startCytoband": "p25.3",
    "endPosition": 321672,
    "endCytoband": "p25.3",
    "matePieceLocation": "upstream",
    "transcripts": [
      {
        "transcriptId": null,
        "transcriptDatabaseName": "RefSeq",
        "transcriptDatabaseVersion": "Release 99",

```

```

    "geneSymbol": "LINC01865"
  }
]
},
"insertedSequence": "GTNNNNNCAT",
"supportingReadCount": 30,
"alternateAlleleFrequency": 0.07,
"rearrangementType": "other",
"sampleItemId": "sequence-1-tumor-dna",
"variantOrigin": "somatic",
"matched": true,
"reported": false
}

```

(Example4. In cases of exon skipping)

```

{
  "itemId": "variant-16",
  "breakends": [
    {
      "chromosome": "7",
      "startPosition": 116771654,
      "startCytoband": "p22.1",
      "endPosition": 116771654,
      "endCytoband": "p22.1",
      "transcripts": [
        {
          "transcriptId": "NM_000245",
          "transcriptDatabaseName": "RefSeq",
          "transcriptDatabaseVersion": "Release 99",
          "regionStructure": [
            {
              "type": "exon",
              "number": 13,
              "totalNumber": 21
            }
          ],
          "geneSymbol": "MET"
        }
      ]
    },
    {
      "chromosome": "7",
      "startPosition": 116774881,
      "startCytoband": "p22.1",
      "endPosition": 116774881,
      "endCytoband": "p22.1",
      "transcripts": [
        {
          "transcriptId": "NM_000245",
          "transcriptDatabaseName": "RefSeq",

```

```

    "transcriptDatabaseVersion": "Release 99",
    "regionStructure": [
      {
        "type": "exon",
        "number": 15,
        "totalNumber": 21
      }
    ],
    "geneSymbol": "MET"
  }
],
"skippedExonRanges": [
  {
    "transcriptId": "NM_000245",
    "transcriptDatabaseName": "RefSeq",
    "transcriptDatabaseVersion": "Release 99",
    "geneSymbol": "MET",
    "exonRange": [14, 14],
    "totalExonNumber": 21
  }
],
"supportingReadCount": 30,
"alternateAlleleFrequency": 0.07,
"rearrangementType": "exon skipping",
"sampleItemId": "sequence-2-tumor-rna",
"variantOrigin": "somatic",
"matched": false,
"reported": true
}

```

(Example5. In cases of translocation)

```

{
  "itemId": "variant-17",
  "breakends": [
    {
      "chromosome": "14",
      "startPosition": 105857793,
      "startCytoband": "p11.2",
      "endPosition": 105857793,
      "endCytoband": "p11.2",
      "matePieceLocation": "downstream",
      "transcripts": [
        {
          "transcriptId": null,
          "transcriptDatabaseName": null,
          "geneSymbol": "IGH",
          "regionName": "IGH"
        }
      ]
    }
  ]
}

```

```

},
{
  "chromosome": "8",
  "startPosition": 127735729,
  "startCytoband": "p11.23",
  "endPosition": 127735729,
  "endCytoband": "p11.23",
  "matePieceLocation": "upstream",
  "transcripts": [
    {
      "transcriptId": null,
      "transcriptDatabaseName": null,
      "regionStructure": [
        {
          "type": "upstream"
        }
      ],
      "geneSymbol": "MYC"
    }
  ]
},
"supportingReadCount": 120,
"totalReadCount": 435,
"alternateAlleleFrequency": 0.31,
"rearrangementType": "translocation",
"sampleItemId": "sequence-1-tumor-dna",
"variantOrigin": "somatic",
"matched": false,
"reported": true
}

```

## VII. otherBiomarkers tag

This tag is used to provide information on biomarkers other than variants defined in the variants tag. Currently, Micro-Satellite Instability (MSI), Tumor Mutation Burden (TMB), and Loss Of Heterozygosity (LOH) can be noted.

Key	Condition	Data type	Description
otherBiomarkers	optional	array (length: 1-N, object)	Aggregation tag for information on biomarkers. Each object in the array must be unique.

### VII-1. Tags within otherBiomarkers tag

Key	Condition	Data type	Description
itemId	required	string regex: ^.+	An ID assigned to a biomarker. It must be a unique string of characters within a single case.
biomarkerType	required	string [choice]	Type of biomarkers. Select one from: • "TMB": Tumor Mutation Burden • "MSI": Micro-Satellite Instability • "LOH": Loss Of Heterozygosity * Inform C-CAT for other biomarkers.
biomarkerMetrics	optional (recommend)	array (length: 1-N, object)	Inspection value information. Each object in the array must be unique. * When creating C-CAT Findings, the first object in the array is used first.
value	[required]	number	Measured value. e.g.) 5.15
unit	[required]	string regex: ^.+	Unit for the measured value. Units may vary depending on test types. e.g.) % If the inspection value is a scalar, you can input null.
type	[required]	string regex: ^.+	Type of inspection value. e.g.) Mutations per megabase MSIsensor score percentage of MSI sites
state	optional (recommend)	string [choice]	Select the biomarker status from the following options. • "high"

			<ul style="list-style-type: none"> <li>• "low"</li> <li>• "intermediate"</li> <li>• "stable"</li> <li>• "cannot be determined"</li> </ul> <p>If the test was performed but the result is not listed above, describe null.</p> <p>* Inform C-CAT for other options.</p>
descriptions	optional	array (length: 1-N, string regex: ^.+)\$	Descriptions such as how to obtain the testing value and the meaning of the value. Refer to the contents tag within the comments tag for usable languages and new lines.
biomarkerOrigin	optional (recommend)	string [choice]	Refer to the description in the variantOrigin tag in shortVariants tag.
sampleItemId	required	string regex: ^.+)\$	Refer to the description in the shortVariants tag.
matched	optional	boolean	Refer to the description in the shortVariants tag.
approved	optional (recommend)	boolean	Refer to the description in the shortVariants tag.
grade	optional	string [choice]	Refer to the description in the shortVariants tag.
reported	required	boolean	Refer to the description in the shortVariants tag.

## VII-2. Example of otherBiomarkers tag

```
(Example)
"otherBiomarkers": [
{
  "itemId": "biomarker-1",
  "biomarkerType": "MSI",
  "biomarkerMetrics": [
    {
      "value": 5.15,
      "unit": "%",
      "type": "percentage of MSI sites"
    },
    {
      "value": 2,
      "unit": null,
```

```
    "type": "MSIsensor score"
  }
],
```

Note: If one inspection item has values in multiple units, use array notation in the "biomarkerMetrics" tag.

```
    "state": "stable",
    "descriptions": [
      "MSI sensor score 10 points or more was MSI-H, 3 points or more and less than 10 points was indeterminate (MSI-I), and less than 3 points was microsatellite stable (MSS).",
      "https://www.gi-cancer.net/gi/ronbun/archives/201901-01.html"
    ],
    "sampleItemId": "sequence-1-tumor-dna",
    "matched": true,
    "reported": true
  },
  {
    "itemId": "biomarker-2",
    "biomarkerType": "TMB",
    "biomarkerMetrics": [
      {
        "value": 34.5680122,
        "unit": "Muts/Mb",
        "type": "Mutations per megabase"
      }
    ],
    "state": "high",
    "sampleItemId": "sequence-1-tumor-dna",
    "reported": true
  },
  {
    "itemId": "biomarker-3",
    "biomarkerType": "LOH",
    "biomarkerMetrics": [
      {
        "value": 24.14,
        "unit": "%",
        "type": "LOH score"
      }
    ],
    "state": "intermediate",
    "sampleItemId": "sequence-1-tumor-dna",
    "reported": true
  }
]
```

## VIII. expressions tag

This tag is used to provide the information on gene expression.

Key	Condition	Data type	Description
expressions	optional	array (length: 1-N, object)	Aggregation tag for information on gene expressions. Each object in the array must be unique.

### VIII-1. Tags within expressions tag

Key	Condition	Data type	Description
itemId	required	string regex: ^.+ \$	An ID assigned to an expression. It must be a unique string of characters within a single case.
readCount	optional (recommend)	integer	Number of reads.
transcripts	required	array (length: 1-N, object)	Refer to the description in the shortVariants tag.
transcriptId	optional	string regex: ^[^\$]+ \$	Refer to the description in the shortVariants tag.
transcriptDatabaseName	optional	string [choice]	Refer to the description in the shortVariants tag. If the transcriptId is entered, this key is recommended to input as well.
transcriptDatabaseVersion	optional	string regex: ^.+ \$	Refer to the description in the shortVariants tag.
regionStructure	optional	array (length: 1-N, object)	Refer to the description in the shortVariants tag.
type	[required]	string [choice]	Refer to the description in the shortVariants tag.
number	optional	integer	Refer to the description in the shortVariants tag.



totalNumber	optional	integer	Refer to the description in the shortVariants tag.
geneSymbol	[required]	string regex: ^[^¥s]+\$	Refer to the description in the shortVariants tag.
regionName	optional	string regex: ^.\$	Refer to the description in the shortVariants tag.
strand	optional	string [choice]	Refer to the description in the shortVariants tag.
cdsChange	optional	string regex: ^.\$	Refer to the description in the shortVariants tag.
aminoAcidsChange	optional	string regex: ^.\$	Refer to the description in the shortVariants tag.
calculatedEffects	optional	array (length: 1-N, string regex: ^.\$)	Refer to the description in the shortVariants tag.
expressionLevelMetrics	required	array (length: 1-N, object)	Refer to the description in the rearrangements tag.
value	[required]	number	Refer to the description in the rearrangements tag.
unit	[required]	string [choice]	Refer to the description in the rearrangements tag.
type	optional (recommend)	string [choice]	Refer to the description in the rearrangements tag.
sampleSize	optional	integer	Refer to the description in the rearrangements tag.
isControl	optional	boolean	Refer to the description in the rearrangements tag.
sampleItemId	required	string regex: ^.\$	Refer to the description in the shortVariants tag.

approved	optional (recommend)	boolean	Refer to the description in the shortVariants tag.
grade	optional	string [choice]	Refer to the description in the shortVariants tag.
reported	required	boolean	Refer to the description in the shortVariants tag.

## VIII-2. Example of expressions tag

(Example)

```
"expressions": [
{
  "itemId": "expression-1",
  "readCount": 9928,
  "transcripts": [
    {
      "transcriptDatabaseName": "RefSeq",
      "transcriptDatabaseVersion": "Release 99",
      "geneSymbol": "MET"
    }
  ],
  "expressionLevelMetrics": [
    {
      "value": 28.24,
      "unit": "TPM",
      "isControl": false
    },
    {
      "value": 142.7,
      "unit": "TPM",
      "type": "mean",
      "sampleSize": 20,
      "isControl": true
    },
    {
      "value": 134.2,
      "unit": "TPM",
      "type": "standard deviation",
      "sampleSize": 20,
      "isControl": true
    }
  ],
  "sampleItemId": "sequence-4-normal-rna",
  "reported": true
}
]
```

## IX. armLevelChanges tag

This tag is used to provide the information about chromosome – chromosome arm change.

Key	Condition	Data type	Description
armLevelChanges	optional	array (length: 1-N, object)	Aggregation tag for armLevelChange information. Each object in the array must be unique.

### IX-1. Tags within armLevelChanges tag

Key	Condition	Data type	Description
itemId	required	string regex: ^.+\$\$	An ID assigned to chromosome – chromosome arm change. It must be a unique string of characters within a single case.
armLevelChangeType	required	string [choice]	Type of the change for chromosome or chromosome arm level change. Select one from: For more information, see International System for Human Cytogenomic Nomenclature (ISCN) 2020 for the details of these abbreviations. <ul style="list-style-type: none"> <li>• "amp": amplification</li> <li>• "del": deletion</li> <li>• "t": translocation</li> <li>• "inv": inversion</li> <li>• "dup": duplication</li> <li>• "cx": complex reconstruction</li> </ul> * Please consult C-CAT if you want to use other types.
cytobands	required	array (length: 1-N, object)	Cytoband information.
chromosome	[required]	string regex: ^[a-zA-Z0-9_#-]+\$	Chromosome number.
arm	optional (recommend)	string [choice]	Short or long arm. Select from the following two options. <ul style="list-style-type: none"> <li>• "p": short arm</li> </ul>

			• "q": long arm
region	optional (recommend)	string regex: ^.+ \$	Cytoband region
armLevelChangeNames	optional (recommend)	array (length: 1- N, string regex: ^.+ \$)	Name of chromosome – chromosome arm change in the report by the testing company. Each string in the array must be unique.
sampleItemId	required	string regex: ^.+ \$	Refer to the description in the shortVariants tag.
approved	optional (recommend)	boolean	Refer to the description in the shortVariants tag.
grade	optional	string [choice]	Refer to the description in the shortVariants tag.
reported	required	boolean	Refer to the description in the shortVariants tag.

#### IX-2. Example of armLevelChanges tag

(Example)

```

"armLevelChanges": [
{
  "itemId": "arm-1",
  "armLevelChangeType": "amp",
  "cytobands": [
    {
      "chromosome": "7",
      "arm": "q"
    }
  ],
  "sampleItemId": "sequence-1-tumor-dna",
  "reported": true
},
{
  "itemId": "arm-2",
  "armLevelChangeType": "del",
  "cytobands": [
    {
      "chromosome": "17",
      "arm": "p"
    }
  ]
}

```

```

    }
  ],
  "sampleItemId": "sequence-1-tumor-dna",
  "reported": true
},
{
  "itemId": "arm-3",
  "armLevelChangeType": "t",
  "cytobands": [
    {
      "chromosome": "X",
      "arm": "q",
      "region": "28"
    },
    {
      "chromosome": "14",
      "arm": "q",
      "region": "11.2"
    }
  ],
  "armLevelChangeNames": [
    "t(X;14)(q28;q11.2)"
  ],
  "sampleItemId": "sequence-1-tumor-dna",
  "reported": true
},
{
  "itemId": "arm-4",
  "armLevelChangeType": "amp",
  "cytobands": [
    {
      "chromosome": "21"
    }
  ],
  "armLevelChangeNames": [
    "iAMP21"
  ],
  "sampleItemId": "sequence-1-tumor-dna",
  "reported": true
}
]

```

## X. nonHumanContents tag

This tag is used to provide the information on non-human content.

Key	Condition	Data type	Description
nonHumanContents	optional	array (length: 1-N, object)	Aggregation tag for information on non-human contents. Each object in the array must be unique.

### X-1. Tags within nonHumanContents tag

Key	Condition	Data type	Description
itemId	required	string regex: ^.+ \$	The ID of the item. It must be a unique string of characters within a single case.
organism	required	string regex: ^.+ \$	The name of the species, as listed in the testing company's report.
contentMetrics	optional	array (length: 1-N, object)	Refer to the description in the biomarkerMetrics tag in otherBiomarkers tag.
value	[required]	number	Refer to the description in the biomarkerMetrics tag in otherBiomarkers tag.
unit	[required]	string regex: ^.+ \$	Refer to the description in the biomarkerMetrics tag in otherBiomarkers tag.
type	[required]	string regex: ^.+ \$	Refer to the description in the biomarkerMetrics tag in otherBiomarkers tag.
descriptions	optional	array (length: 1-N, string regex: ^.+ \$)	Description on nonHumanContent. Refer to the contents tag within the comments tag for usable languages and new lines.
sampleItemId	required	string regex: ^.+ \$	Refer to the description in the shortVariants tag.
approved	optional	boolean	Refer to the description in the shortVariants tag.

grade	optional	string [choice]	Refer to the description in the shortVariants tag.
reported	required	boolean	Refer to the description in the shortVariants tag.

## X-2. Example of nonHumanContents tag

(Example)

```
"nonHumanContents": [
  {
    "itemId": "nonHuman-1",
    "organism": "HBV",
    "contentMetrics": [
      {
        "value": 65,
        "unit": "reads-per-million",
        "type": "virus-derived read sequences"
      }
    ],
    "descriptions": [
      "Hepatitis B virus."
    ],
    "sampleItemId": "sequence-1-tumor-dna",
    "reported": false
  }
]
```

## XI. compositeBiomarkers tag

This tag provides information on composite markers (e.g., combination of variants; fusion composed of three genes) that are represented by the combinations of elements in the shortVariants, copyNumberAlterations, and rearrangements tags.

Key	Condition	Data type	Description
compositeBiomarkers	optional	array (length: 1-N, object)	Aggregation tag for information on composite markers. Each object in the array must be unique.

### XI-1. Tags within compositeBiomarkers tag

Key	Condition	Data type	Description
itemId	required	string regex: ^.+ \$	An ID assigned to a composite marker. It must be a unique string of characters within a single case.
componentItemIds	required	array (length: 2-N, string regex: ^.+ \$)	Array of component variants (itemIds). Each string in the array must be unique.
biomarkerNames	required	array (length: 1-N, string regex: ^.+ \$)	The name of the composite marker, as listed in the testing company's report. Each string in the array must be unique.
descriptions	optional	array (length: 0-N, string regex: ^.+ \$)	Description on composite markers Refer to the contents tag within the comments tag for usable languages and new lines.
approved	optional	boolean	Refer to the description in the shortVariants tag.
grade	optional	string [choice]	Refer to the description in the shortVariants tag.
reported	required	boolean	Refer to the description in the shortVariants tag.



## XI-2. Example of compositeBiomarkers tag

(Example)

```
"compositeBiomarkers": [  
  {  
    "itemId": "composite-1",  
    "componentItemIds": [  
      "variant-14",  
      "variant-15"  
    ],  
    "biomarkerNames": [  
      "BRAF-NRG1-ALK fusion"  
    ],  
    "descriptions": [  
      "Three genes are fused together to produce the fusion gene BRAF-NRG1-ALK."  
    ],  
    "reported": true  
  }  
]
```

## XII. druggability tag

This tag provides information on druggability.

Key	Condition	Data type	Description
druggability	optional	array (length: 1-2, object)	Aggregation tag for information on druggability. Each object in the array must be unique.

### XII-1. Tags within druggability tag

Key	Condition	Data type	Description
companionDiagnostics	optional (recommend)	array (length: 1-N, object)	Information on companion diagnostics.
itemId	[required]	string regex: ^.+ \$	An ID assigned to a druggability. It must be a unique string of characters within in this case report.
alterationItemId	[required]	string regex: ^.+ \$	Target alteration(itemId).
drugNames	[required]	array (length: 1-N, string regex: ^.+ \$)	Array of drugs.
descriptions	optional	array (length: 1-N, string regex: ^.+ \$)	Description on druggability. Refer to the contents tag within the comments tag for usable languages and new lines.
approved	optional (recommend)	boolean	Refer to the description in the shortVariants tag.
reported	[required]	boolean	Refer to the description in the shortVariants tag.
comprehensiveGenomicProfiling	optional (recommend)	array (length: 1-N, object)	Information on comprehensive genomic profiling.

itemId	[required]	string regex: ^.+ \$	An ID assigned to a druggability. It must be a unique string of characters within a single case.
alterationItemId	[required]	string regex: ^.+ \$	Target alteration(itemId).
drugNames	[required]	array (length: 1- N, string regex: ^.+ \$)	Array of drugs.
knowledgeBase	optional	array (length: 1- N, string regex: ^.+ \$)	The name of the knowledge base.
evidenceLevel	optional (recommend)	string regex: ^.+ \$	Evidence level.
cancerCode	optional	string regex: ^.+ \$	Cancer code.
cancerName	optional	string regex: ^.+ \$	Cancer name.
descriptions	optional	array (length: 1- N, string regex: ^.+ \$)	Description on druggability. Refer to the contents tag within the comments tag for usable languages and new lines.
approved	optional (recommend)	boolean	Refer to the description in the shortVariants tag.
reported	[required]	boolean	Refer to the description in the shortVariants tag.

## XII-2. Example of druggability tag

(Example)

```
"druggability": {
  "companionDiagnostics": [
```

```

{
  "itemId": "druggability-CDx-1",
  "componentItemid": "variant-1",
  "drugNames": [
    "afatinib",
    "erlotinib"
  ],
  "approved": true,
  "reported": true
},
"comprehensiveGenomicProfiling": [
  {
    "itemId": "druggability-CGP-1",
    "componentItemid": "variant-2",
    "drugNames": [
      "alectinib",
      "crizotinib"
    ],
    "knowledgeBase": [
      "OncoKB"
    ],
    "evidenceLevel": "A",
    "cancerCode": "NSCLC",
    "cancerName": "Non-Small Cell Lung Cancer",
    "approved": true,
    "reported": true
  }
]
}

```

### XIII. Other notes

Precautions.

#### XIII-1. itemId

The value of itemId must be unique within the file. The value can be any string.

##### XIII-1-1. Example of itemId description

Examples of itemId values for each tag are as follows.

- For variants tag

"itemId": "variant-1"
-----------------------

"itemId": "variant-2"
-----------------------

"itemId": "variant-3"
-----------------------

- For otherBiomarkers tag

"itemId": "biomarker-1"
-------------------------

"itemId": "biomarker-2"
-------------------------

"itemId": "biomarker-3"
-------------------------

- For sequencingSamples tag

"itemId": "sequence-1-tumor-dna"
----------------------------------

"itemId": "sequence-2-tumor-rna"
----------------------------------

"itemId": "sequence-3-normal-dna"
-----------------------------------

- For compositeBiomarkers tag

"itemId": "composite-1"
-------------------------

"itemId": "composite-2"
-------------------------

"itemId": "composite-3"
-------------------------

### XIII-2. matePieceLocation

Here, we provide an explanation for matePieceLocation in the breakends tag. We assume that "upstream" and "downstream" correspond to the direction of decreasing and increasing positional coordinates along a chromosome in a reference genome sequence, respectively.

Note: These "upstream" and "downstream" are different from those in the transcriptional direction.

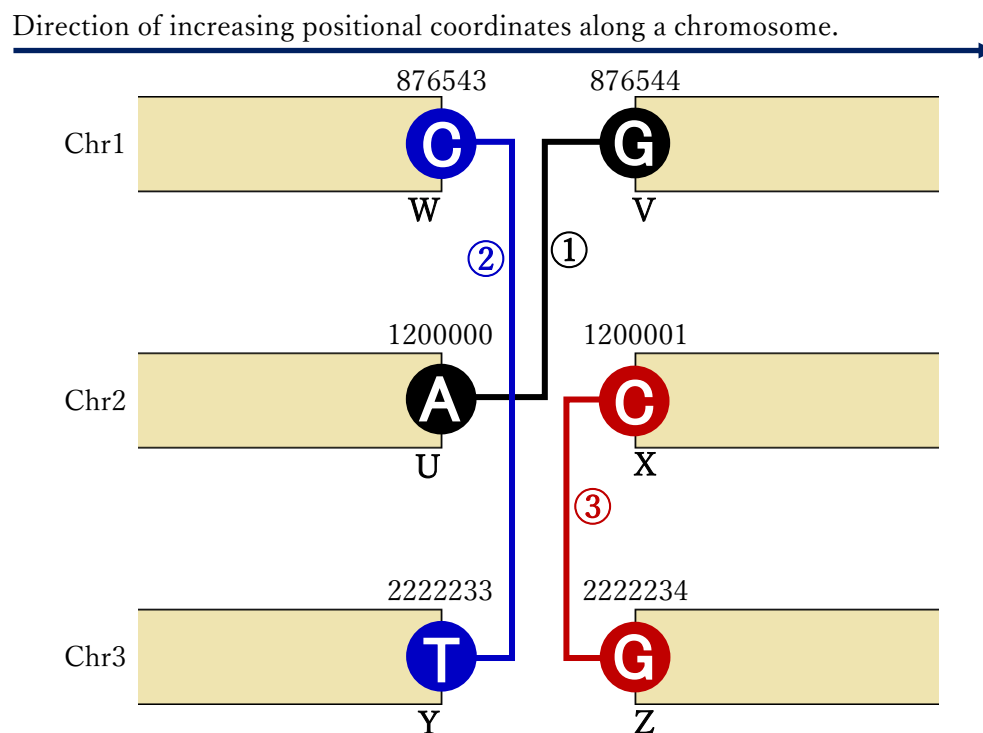
#### XIII-2-1. Example of matePieceLocation description

We illustrate how to input matePieceLocation and present representations in VCF format (v4.3). The examples below are modified from the figure in the following documents:

The Variant Call Format Specification VCF v4.3 and BCF v2.2

<https://samtools.github.io/hts-specs/VCFv4.3.pdf>

Example of rearrangements:



Descriptions in the standardized format:

- Example①

For the chromosome 2 junction point, the downstream sequence of the junction point is replaced with another sequence; therefore, the value of matePieceLocation is "downstream".

In contrast, for the chromosome 1 junction point, the upstream sequence of the junction point is replaced with another sequence; therefore, the value of matePieceLocation is "upstream".

(Example①)

```
"breakends": [  
  {  
    "chromosome": "2",  
    "startPosition": 1200000,  
    "endPosition": 1200000,  
    "matePieceLocation": "downstream"  
  },  
  {  
    "chromosome": "1",  
    "startPosition": 876654,  
    "endPosition": 876654,  
    "matePieceLocation": "upstream"  
  }  
]
```

• Example②

For the chromosome 1 junction point, the downstream sequence of the junction point is replaced with another sequence; therefore, the value of matePieceLocation is "downstream".  
The same is true for chromosome 3; thus, the value of matePieceLocation is "downstream".

(Example②)

```
"breakends": [  
  {  
    "chromosome": "1",  
    "startPosition": 876543,  
    "endPosition": 876543,  
    "matePieceLocation": "downstream"  
  },  
  {  
    "chromosome": "3",  
    "startPosition": 2222233,  
    "endPosition": 2222233,  
    "matePieceLocation": "downstream"  
  }  
]
```

• Example③

For the chromosome 2 junction point, the upstream sequence of the junction point is replaced with another sequence; therefore, the value of matePieceLocation is "upstream".  
The same is true for the chromosome 3 junction point; thus, the value of matePieceLocation is "upstream".

(Example③)

```
"breakends": [  
  {  
    "chromosome": "2",  
    "startPosition": 1200001,  
    "endPosition": 1200001,  
    "matePieceLocation": "upstream"  
  },  
  {  
    "chromosome": "3",  
    "startPosition": 2222234,  
    "endPosition": 2222234,  
    "matePieceLocation": "upstream"  
  }  
]
```

Description in VCF format:

The following representation for the above figure is based on VCF v4.3: The numbers in the leftmost column in the table below correspond to the numbers in the figure above.

	#CHROM	POS	ID	REF	ALT	QUAL	FILTER	INFO
①	1	876544	bnd_V	G	]2:1200000]G	6	PASS	SVTYPE=BN
	2	1200000	bnd_U	A	A[1:876544[	6	PASS	SVTYPE=BN
②	1	876543	bnd_W	C	C]3:2222233]	6	PASS	SVTYPE=BN
	3	2222233	bnd_Y	T	T]1:876543]	6	PASS	SVTYPE=BN
③	2	1200001	bnd_X	C	[3:2222234[C	6	PASS	SVTYPE=BN
	3	2222234	bnd_Z	G	[2:1200001[G	6	PASS	SVTYPE=BN



### XIII-3. sampleMetrics

For the sampleMetrics tag in the sequencingSamples tag, the following table provides supplemental explanation on the numerator and denominator when selecting percentages for the unitType tag. Column A in the table below shows example values for the targetType and duplicateType tags when selecting percentages for the unitType tag. Column B shows the values of the targetType, duplicateType, and unitType tags corresponding to the numerator and denominator of the percentage, respectively. The denominator may correspond to the value of the allReadCount tag rather than the values of those tags (e.g., row #1 in the table below).

#	A (unitType = 'percentage')		B			
	targetType	duplicateType		targetType	duplicate Type	unitType
1	total	unique+duplicate	numerator	total	unique+	count
			denominator	allReadCount		
2	total	unique	numerator	total	unique	count
			denominator	total	unique+	count
3	total	duplicate	numerator	total	duplicate	count
			denominator	total	unique+	count
4	target	unique+duplicate	numerator	target	unique+	count
			denominator	allReadCount		
5	target	unique	numerator	target	unique	count
			denominator	total	unique	count
6	target	duplicate	numerator	target	duplicate	count
			denominator	total	duplicate	count

### XIV. For inquiries

Please contact the C-CAT Help Desk.

E-Mail: [helpdesk\\_c-cat@ml.res.ncc.go.jp](mailto:helpdesk_c-cat@ml.res.ncc.go.jp)