

がんゲノム検査 標準化フォーマット

英名 : CATS
(Cancer Genomic Test Standardized)
Format

解説資料

By C-CAT ゲノムデータ管理室

v1.1.0

2021/07/02

目次

I.	はじめに	4
I-1.	目的	4
I-2.	用語	4
I-3.	必須条件について	4
I-4.	ファイル情報.....	5
II.	C-CAT に特有の事項.....	6
II-1.	ファイルの送付形態.....	6
II-2.	入力の範囲	6
II-3.	お願い.....	6
II-4.	注	6
III.	metaData タグ.....	7
III-1.	schemaVersion キー	7
III-2.	referenceGenome タグ.....	7
III-2-1.	referenceGenome タグ内の解説.....	7
III-2-2.	referenceGenome タグ記載例.....	7
III-3.	configOptions タグ	8
III-3-1.	configOptions タグ内の解説.....	8
III-3-2.	configOptions タグ記載例	10
III-4.	comments タグ	10
III-4-1.	comments タグ内の解説	11
III-4-2.	comments タグ記載例.....	11
IV.	testInfo タグ.....	13
IV-1.	testInfo タグ内の解説	13
IV-2.	testInfo タグ記載例.....	13
V.	variants タグ.....	15
V-1.	shortVariants タグ.....	15
V-1-1.	shortVariants タグ内の解説.....	15
V-1-2.	shortVariants タグ記載例.....	18
V-2.	copyNumberAlterations タグ.....	21
V-2-1.	copyNumberAlterations タグ内の解説.....	21
V-2-2.	copyNumberAlterations タグ記載例.....	23
V-3.	rearrangements タグ	24
V-3-1.	rearrangements タグ内の解説	24
V-3-2.	rearrangements タグ記載例	28
VI.	otherBiomarkers タグ	31
VI-1.	otherBiomarkers タグ内の解説.....	31

VI-2. otherBiomarkers タグ記載例	32
VII. compositeBiomarkers タグ	34
VII-1. compositeBiomarkers タグ内の解説.....	34
VII-2. compositeBiomarkers タグ記載例.....	34
VIII. sequencingSamples タグ	36
VIII-1. sequencingSamples タグ内の解説.....	36
VIII-2. sequencingSamples タグ記載例.....	37
IX. 注意事項.....	38
IX-1. itemId.....	38
IX-1-1. itemId の記載例	38
IX-2. matePieceLocation.....	38
IX-2-1. matePieceLocation の記載例	39
X. お問い合わせ先	41

I. はじめに

I-1. 目的

包括的がんゲノムプロファイリング検査の結果得られる遺伝子異常データのフォーマットは、現在検査会社ごとに違いがある。この差異のため、統一的な枠組みの下同一のソフトウェアによって、遺伝子異常データに、候補となる適応薬剤や臨床試験を注釈付けることは、第三者にとって困難となっている。がんゲノム検査結果の解釈の統一化や均てん化を効率的に促進するためには、包括的がんゲノムプロファイリング検査における遺伝子異常データの統一的なフォーマットを定めることが必要である。

本資料は、包括的がんゲノムプロファイリング検査によって得られる遺伝子異常データの統一的なフォーマット、CATS（がんゲノム検査標準化、cancer genomic test standardized）フォーマットの仕様を解説した資料である。CATS フォーマットのデータ・スキーマは、JSON 定義ファイル "schema.json" によって定義されており、本資料ではそれを解説する。

利用の想定は以下である。包括的がんゲノムプロファイリング検査を行う検査会社等は、CATS フォーマットにて、C-CAT のような検査結果注釈機関へ遺伝子異常データを送付する。検査結果注釈機関は、CATS フォーマットで表現された遺伝子異常データに対し、臨床データも勘案して、候補となる適応薬剤や臨床試験の注釈付けをする。本フォーマットでは遺伝子異常データを対象とする。臨床データは、病院の電子カルテに保管されていて検査会社は知ることが出来ない可能性が高いため、対象としない。

I-2. 用語

- 検査結果注釈機関：検査会社等から包括的がんゲノムプロファイリング検査の遺伝子異常データを、また、病院等から臨床データを受領し、がん知識ベースを利用して、個々の遺伝子異常に候補となる適応薬剤および臨床試験情報を注釈付けする機関。例として、C-CAT。
- 検査結果注釈書：包括的がんゲノムプロファイリング検査の個々の遺伝子異常データに、候補となる適応薬剤や臨床試験情報が紐付けられた書類。例として、C-CAT 調査結果。検査結果注釈書とは異なり、検査結果報告書は、検査会社が発行する検査結果の報告書。
- がん知識ベース：がんの遺伝子異常に、候補となる適応薬剤や臨床試験を関連付けたデータベース。例として、C-CAT CKDB (cancer knowledge database)や OncoKB (Chakravarty et al, 2017, JCO Precision Oncology)。

I-3. 必須条件について

- 必須：JSON の親タグが存在するとき、必須の項目
- 任意：検査結果注釈書において薬剤や臨床試験情報に紐付けられたり、検査会社の検査結果報告書の内容がより反映されて検査結果注釈書の精度を高める可能性があったりする、入力推奨の項目

I-4. ファイル情報

- 文字コード：UTF8
- タイプ：JSON
- 拡張子：json

II. C-CAT に特有の事項

II-1. ファイルの送付形態

包括的がんゲノムプロファイリング検査による遺伝子異常データは、検査会社等から C-CAT へ、CATS (cancer genomic test standardized) フォーマットにて送付する必要があります。

II-2. 入力範囲

データ品質が確保された遺伝子異常 (shortVariant、copyNumberAlteration、rearrangement、otherBiomarker) が、CATS フォーマットへの入力の対象です。偽陽性と考えられる遺伝子異常は入力しないで下さい。入力された遺伝子異常に対し、C-CAT 調査結果に出力するかしないかは、後述するタグ ("reported") を指定することで選択できます。

必ず、承認範囲にある遺伝子異常の情報は入力し、かつ、C-CAT 調査結果へ出力するようにして下さい。

II-3. お願い

- 任意項目は入力推奨の項目であり、できるだけ入力してください。より多く入力していただくと、C-CAT 調査結果においてより多くの薬剤や臨床試験情報が付加されたり、検査会社の検査結果報告書をより反映して C-CAT 調査結果の精度がより高まったりする可能性があります。また、現在の版の C-CAT 調査結果ではたとえ使用されてなくても、将来の版で使用され、より多くの情報が付与される可能性があります。
- データ品質が確保された全ての遺伝子異常を（後述する "reported": false の遺伝子異常も含めて）、できるだけ記載して下さい。そうでないと、検査会社の検査結果報告書のフォーマットや仕様が変わった場合、検査会社に問い合わせをしたり、場合によっては C-CAT 調査結果の作成が滞ったりする可能性があります。

II-4. 注

各項目内で C-CAT に特有の注意事項は、「※」にて記載しています。

III. metaData タグ

メタ情報を記載する。

schemaVersion キー、referenceGenome タグ、configOptions タグ、comments タグが含まれる。

キー	必須条件	データ型	説明
metaData	必須	オブジェクト	メタデータ集約タグ

III-1. schemaVersion キー

キー	必須条件	データ型	説明
schemaVersion	必須	文字列 正規表現： ^[0-9\\.]+\$	JSON ファイルのスキーマ定義バージョン

III-2. referenceGenome タグ

リファレンスゲノム配列に関する情報を記載する。

キー	必須条件	データ型	説明
referenceGenome	必須	オブジェクト	リファレンスゲノム配列に関する情報

III-2-1. referenceGenome タグ内の解説

キー	必須条件	データ型	説明
name	任意	文字列 正規表現： ^\.+\$	検査会社にて解析で使したリファレンスゲノム配列名を記載する。
grcRelease	必須	文字列 正規表現： ^GRC\.+\$	リファレンスゲノム配列の GRC (ゲノム参照コンソーシアム) リリース ID を記載する。
descriptions	任意	配列(長さ 0~N、 文字列 正規表現： ^\.+\$)	name タグに記載したリファレンスゲノム配列の説明。 使用言語および改行については、comments タグ内の contents タグを参照。

III-2-2. referenceGenome タグ記載例

```
(例 1. NCBI の記載例)
"referenceGenome": {
  "name": "GRCh38.p13",
```

```

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

```

(例 2. UCSC の記載例)

```

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

```

(例 3. GDC の記載例)

```

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

```

III-3. configOptions タグ

C-CAT CKDB のようながん知識ベースとのマッチングや、C-CAT 調査結果のような検査結果注釈書への掲載を制御するタグ。

キー	必須条件	データ型	説明
configOptions	任意	オブジェクト	がん知識ベースとのマッチングや、検査結果注釈書への掲載に関する制御情報集約タグ。

III-3-1. configOptions タグ内の解説

キー	必須条件	データ型	説明
typeLabelsInterpretedAsKbAmplification	任意	配列(長さ1~4、文字列) [選択式]	がん知識ベースにおける"amplification" (コピー数増幅) として解釈される、検査会社使用の遺伝子異常ラベル。 次の選択肢から選択 (配列内で重複不可)。 ・ "copyNumberAlterationType: amplification" ・ "copyNumberAlterationType: gain"

			<ul style="list-style-type: none"> • "copyNumberAlterationType: duplication" • "rearrangementType: duplication" (デフォルト: "copyNumberAlterationType: amplification", "copyNumberAlterationType: gain", "copyNumberAlterationType: duplication")
typeLabelsInterpretedAsKbLoss	任意	配列(長さ1~3、文字列) [選択式]	がん知識ベースにおける"loss" (コピー数減少) として解釈される、検査会社使用の遺伝子異常ラベル。 次の選択肢から選択 (配列内で重複不可)。 • "copyNumberAlterationType: loss" • "copyNumberAlterationType: deletion" • "copyNumberAlterationType: homozygous deletion" • "rearrangementType: deletion" (デフォルト: "copyNumberAlterationType: loss", "copyNumberAlterationType: deletion", "copyNumberAlterationType: homozygous deletion")
typeLabelsInterpretedAsKbGeneFusion	任意	配列(長さ1~8、文字列) [選択式]	がん知識ベースにおける"geneFusion" (遺伝子融合) として解釈される、検査会社使用の遺伝子異常ラベル。 次の選択肢から選択 (配列内で重複不可)。 • "rearrangementType: gene fusion" • "rearrangementType: gene fusion and frameshift variant" • "rearrangementType: bidirectional gene fusion" • "rearrangementType: duplication" • "rearrangementType: tandem duplication" • "rearrangementType: deletion" • "rearrangementType: inversion" • "rearrangementType: truncation" • "rearrangementType: other" (デフォルト: "rearrangementType: gene fusion", "rearrangementType: gene fusion and frameshift variant", "rearrangementType: bidirectional gene fusion")
hideAlleleFrequency	任意	ブール型	変異型アレル頻度数値の検査結果注釈書への掲載を行わない(true)指定が可能

			※承認範囲の情報に関しては、キー自体を記載せずデフォルトを使用するか、false を記載する。 (デフォルト: false)
hideCnaValue	任意	ブール型	コピー数異常数値に関し、上記 hideAlleleFrequency タグと同じ。
hideMsiValue	任意	ブール型	マイクロサテライト不安定性 (MSI) 数値に関し、上記 hideAlleleFrequency タグと同じ。
hideTmbValue	任意	ブール型	腫瘍変異負荷 (TMB) 数値に関し、上記 hideAlleleFrequency タグと同じ。
hideLohValue	任意	ブール型	ヘテロ接合性の消失 (LOH) 数値に関し、上記 hideAlleleFrequency タグと同じ。

III-3-2. configOptions タグ記載例

(例)

```
"configOptions": {
  "hideTmbValue": true,
  "hideLohValue": true,
  "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"
  ]
}
```

III-4. comments タグ

遺伝子異常(variants)、バイオマーカー(otherBiomarkers)、sequencing sample 情報 (sequencingSamples)に対するコメントを記載する。itemIds キーとコメントを入れる contents キーが含まれる。

キー	必須条件	データ型	説明
----	------	------	----

comments	任意	配列(長さ 0 ~N、オブジェクト)	コメント情報集約タグ。 配列の各オブジェクトは一意であること。
----------	----	--------------------	------------------------------------

III-4-1. comments タグ内の解説

キー	必須条件	データ型	説明
itemIds	必須	配列(長さ 0~N、文字列 正規表現： <code>^.+ \$</code>)	遺伝子異常(variants)、バイオマーカー(otherBiomarkers)、sequencing sample 情報(sequencingSamples)の itemId (複数も可、重複は不可)を記載。 症例全体のコメントを記載する場合、itemIds の長さは 0 とする。 ※itemId を記載した場合、C-CAT 調査結果へは掲載されない
contents	必須	配列(長さ 1~N、文字列 正規表現： <code>^.+ \$</code>)	itemId に対するコメントの内容。 説明文は、英語または日本語。説明文中の改行コードは無視され、説明文に改行を含める場合は配列の要素で分けて記載。

III-4-2. comments タグ記載例

(例)

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

↑ itemIds配列の長さが0の場合は検査全体へのコメントを表す。

↑ 説明文に改行を含める場合は、配列の要素として分けて記載する。

↑ 特定の変異検出結果に対するコメントを記載する場合は、対象のitemIdを記載する。

```

"contents": [
  "TSC1 functions independently of TSC2 and mTORC1."
],
{
  "itemIds": [
    "variant-1",
    "variant-5"
    ↑ 複数の変異検出結果に対するコメントを記載する場合は、複数のitemIdを記載する。
  ],
  "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 タグ

検査情報を記載する。

キー	必須条件	データ型	説明
testInfo	必須	オブジェクト	検査情報集約タグ

IV-1. testInfo タグ内の解説

キー	必須条件	データ型	説明
testId	必須	文字列 正規表現： ^\.+\$	検査会社で使用する任意の ID
testType	必須	文字列 [選択式]	検査で使用する検体の組み合わせ。 ・ "tumor-only": 腫瘍検体のみで解析を行っている場合 ・ "tumor and matched-normal": 腫瘍検体と正常検体の検査で使用している場合 ・ "tumor-only (cell-free)": セルフリー検体のみで解析を行っている場合 ・ "tumor (cell-free) and matched-normal": セルフリー検体と正常検体の検査で使用している場合
softwareName	任意	文字列 正規表現： ^\.+\$	使用した遺伝子解析ソフトウェア名
softwareVersion	任意	文字列 正規表現： ^\.+\$	softwareName に記載したソフトウェアのバージョン
panelName	必須	文字列 正規表現： ^\.+\$	使用したパネル検査名。 ※保険収載されていないパネル検査をご希望の方は、事前に C-CAT まで要相談のこと。
panelVersion	必須	文字列 正規表現： ^\.+\$	panelName に記載したパネル検査のバージョン

IV-2. testInfo タグ記載例

```
(例)
"testInfo": {
  "testId": "12345678901231900001",
  "testType": "tumor and matched-normal",
```

```
"softwareName": "variant caller A",  
"softwareVersion": "ver.1.2",  
"panelName": "Multi-gene Panel A",  
"panelVersion": "ver.1.03-00"  
}
```

V. variants タグ

検出された遺伝子異常を記載する。shortVariants タグ、copyNumberAlterations タグ、rearrangements タグが含まれる。

キー	必須条件	データ型	説明
variants	任意	オブジェクト	遺伝子異常情報集約タグ

V-1. shortVariants タグ

1 塩基変異、塩基配列の挿入や欠失、欠失と挿入などの結果を記載する。

キー	必須条件	データ型	説明
shortVariants	任意	配列(長さ 1～N、オブジェクト)	1 塩基変異、塩基配列の挿入や欠失、欠失と挿入などの結果集約タグ。 配列の各オブジェクトは一意であること。

V-1-1. shortVariants タグ内の解説

キー	必須条件	データ型	説明
itemId	必須	文字列 正規表現： ^\.+\$	遺伝子異常に割り当てる ID。 一症例内で一意の文字列であること。
chromosome	必須	文字列 正規表現： ^[a-zA-Z0-9_\-]+\$	染色体番号
position	必須	整数	染色体上の物理位置。座標の表現方法には 1-based 座標系を使用し、VCF v4.3 に従って記載する。（記載例として VCF v4.3 の page 13 にあるように、リファレンス塩基が atCga、変異塩基が at-ga、リファレンス塩基の C の位置が 3 の時、"position": 2, "referenceAllele": "TC", "alternateAllele": "T" と表記する。後述の通り、"referenceAllele" はリファレンス塩基、"alternateAllele" は変異塩基を表す。）
referenceAllele	必須	文字列	リファレンス塩基。

		正規表現： ^[ACGTN]+\$	VCF v4.3 に従って記載する。記載例は 上記 position の説明を参照。
alternateAllele	必須	文字列 正規表現： ^[ACGTN*]+\$	変異塩基。 VCF v4.3 に従って記載する。記載例は 上記 position の説明を参照。 トライ・アレル以上のマルチ・アレル は、別の shortVariants タグの要素とし て分けて記載する。その際、comments タグに対象の itemId とマルチ・アレル である旨を記載する。
alternateAlleleFrequency	必須	数値	変異型アレル頻度(0～1 の範囲)
totalReadDepth	任意	整数	合計読取深度(最小値 1)
alternateAlleleReadDepth	任意	整数	変異型塩基の読取深度(最小値 1)
variantType	任意	文字列 [選択式]	検査会社の報告書に記載されている遺 伝子異常のタイプ。 次の値から選択して記載。 ・ "SNV" ・ "insertion" ・ "deletion" ・ "delins" ・ "indel" ・ "MNV" ※上記以外のタイプを使用する場合 は、C-CAT に要相談
transcripts	必須	配列(長さ: 1 ～N、オブジ ェクト)	代表転写産物の情報。 配列の各オブジェクトは一意であるこ と。 (一つの遺伝子に対し、一つの代表転 写産物。)
transcriptId	必須	文字列 正規表現： ^[^\\s]+\$	転写産物 ID を記載(例: NM_000368. <u>4</u>)。 正確性のため、sub-number (上の赤字 下線部) を含めることを強く推奨す る。インタージェニック領域の変異の ため転写産物 ID が存在しない場合は、 "transcriptId": null と記載することも 可。
transcriptDatabaseName	必須	文字列 [選択式]	転写産物 ID に対するデータベース名。 次の 2 つから選択する。 ・ "RefSeq"

			<ul style="list-style-type: none"> ・ "Ensembl" transcriptId が null の場合は、 "transcriptDatabaseName": null と記載することも可。
transcriptDatabaseVersion	任意	文字列 正規表現： $^{\wedge} \cdot + \$$	上記データベースのバージョンを記載。 ※記載が無い場合は、C-CAT が決めたバージョンが使用される。
geneSymbol	必須	文字列 正規表現： $^{\wedge} [^{\wedge} \backslash \backslash s] + \$$	検査会社の報告書に記載されている遺伝子名(遺伝子記号)を記載。 遺伝子が存在しない場合は "geneSymbol": null と記載。 「TERT promoter」のように、関連付けられる遺伝子があれば入力する。
strand	任意	文字列 [選択式]	転写の向き。 参照ゲノム配列と同じ向きの場合は "+", 逆向きの場合は "-". transcriptId が null の場合は、 "strand": null と記載することも可。
cdsChange	必須	文字列 正規表現： $^{\wedge} \cdot + \$$	検査会社の報告書に記載されている DNA レベルでの変化を記載。 HGVS に準拠した表記が推奨される。 インタージェニック領域等で RNA がコードされない場合は、(n.* のようなノンコーディング領域の表記法以外に) null とすることも可能。
aminoAcidsChange	必須	文字列 正規表現： $^{\wedge} \cdot + \$$	検査会社の報告書に記載されているタンパク質レベルでの変化を記載。 HGVS に準拠した表記が推奨される。 非翻訳領域等でアミノ酸変化がない場合は、null とすることも可能。
calculatedEffects	任意	配列(長さ: 0 ~ N、文字列 正規表現： $^{\wedge} \cdot + \$$)	"splicing_variant" など、転写産物上の変異(異常)の効果は Sequence Ontology で記載する。 snpEff ツールの「Effect (Sequence Ontology)」や、annovar ツールの VCF 「Func.refGene」に相当する項目。annovar では、User Guide, Gene-based Annotation の Output file 1

			(refSeq gene annotation) に解説があり、当該項目の用語を Sequence Ontology に変換できる。 配列の 1 要素につき 1 用語を割り当てる。 配列の各文字列は一意であること。
testMethod	必須	文字列 [選択式]	得られた検出結果が DNA 検体由来か RNA 検体由来かを記載。 ・ "DNA-seq": DNA 検体由来 ・ "RNA-seq": RNA 検体由来
variantOrigin	任意	文字列 [選択式]	体細胞由来か生殖細胞系列由来かを記載。 ※C-CAT では、体細胞由来か生殖細胞系列由来かで、検索するがん知識ベースの対象を変えている。入力がない場合は、体細胞変異（異常）の知識ベースが使用される。 ・ "somatic": 体細胞由来 ・ "germline": 生殖細胞系列由来 ・ "likely somatic": 典型的には腫瘍のみの検査において、体細胞由来と予想される場合 ※体細胞変異（異常）の知識ベースが使用される。 ・ "likely germline": 典型的には腫瘍のみの検査において、生殖細胞系列由来と予想される場合 ※生殖細胞系列変異（異常）の知識ベースが使用される。
reported	必須	ブール型	検査会社の検査結果報告書、またはそれに準ずる報告書に記載された変異であるか否か。True の場合、がん知識ベースを使った注釈付けを行う。

V-1-2. shortVariants タグ記載例

(例 1. SNV の記載例)

```
{
  "itemId": "variant-1",
  ↑ itemIdには検出された遺伝子異常に対し、検査会社随意の文字列を振る。
  "chromosome": "9",
  "position": 135781005,
```

```
"referenceAllele": "C",
"alternateAllele": "G",
```

↑ "position"と"referenceAllele"、"alternateAllele"はVCF v4.3のルールに従って記載する。

```
"alternateAlleleFrequency": 0.54,
"alternateAlleleReadDepth": 108,
"totalReadDepth": 200,
"variantType": "SNV",
"transcripts": [
  {
    "transcriptId": "NM_000368.4",
    "transcriptDatabaseName": "RefSeq",
    "transcriptDatabaseVersion": "Release 99",
    "geneSymbol": "TSC1",
    "cdsChange": "c.1960C>G",
    "aminoAcidsChange": "p.Q654E",
    "calculatedEffects": [
      "missense_variant"
    ]
  }
],
"testMethod": "DNA-seq",
"variantOrigin": "somatic",
"reported": true
}
```

(例 2. insertion の記載例)

```
{
  "itemId": "variant-2",
  "chromosome": "8",
  "position": 37553560,
  "referenceAllele": "A",
  "alternateAllele": "AAGCGGC",
  "alternateAlleleFrequency": 0.4953,
  "alternateAlleleReadDepth": 368,
  "totalReadDepth": 743,
  "variantType": "insertion",
  "transcripts": [
    {
      "transcriptId": "NM_025069.2",
      "transcriptDatabaseName": "RefSeq",
      "transcriptDatabaseVersion": "Release 99",
      "geneSymbol": "ZNF703",
      "cdsChange": "c.63_64insAGCGGC",
      "aminoAcidsChange": "G21_G22insSG"
    }
  ],
  "testMethod": "DNA-seq",
  "variantOrigin": "somatic",
  "reported": true
}
```

(例 3. deletion の記載例)

```
{
  "itemId": "variant-3",
  "chromosome": "1",
  "position": 27097751,
  "referenceAllele": "TC",
  "alternateAllele": "T",
  "alternateAlleleFrequency": 0.12,
  "alternateAlleleReadDepth": 32,
  "totalReadDepth": 266,
  "variantType": "deletion",
  "transcripts": {
    "transcriptId": "ENST00000324856.13",
    "transcriptDatabaseName": "Ensembl",
    "transcriptDatabaseVersion": "v99",
    "geneSymbol": "ARID1A",
    "cdsChange": "c.3340delC",
    "aminoAcidsChange": "p.P1115fs*46",
    "calculatedEffects": [
      "frameshift_variant"
    ]
  },
  "testMethod": "DNA-seq",
  "variantOrigin": "somatic",
  "reported": true
}
```

(例 4. delins の記載例)

```
{
  "itemId": "variant-4",
  "chromosome": "1",
  "position": 26696982,
  "referenceAllele": "GC",
  "alternateAllele": "TT",
  "alternateAlleleFrequency": 0.25,
  "alternateAlleleReadDepth": 52,
  "totalReadDepth": 524,
  "variantType": "delins",
  "transcripts": {
    "transcriptId": "NM_007294.4",
    "transcriptDatabaseName": "RefSeq",
    "transcriptDatabaseVersion": "Release 99",
    "geneSymbol": "BRCA1",
    "cdsChange": "c.579_580delinsTT",
    "aminoAcidsChange": "p.E193_P194delinsDS"
  },
  "testMethod": "DNA-seq",
  "variantOrigin": "somatic",
  "reported": true
}
```

```
}
```

(例 5. TERT promoter の記載例)

```
{
  "itemId": "variant-5",
  "chromosome": "5",
  "position": 1295113,
  "referenceAllele": "G",
  "alternateAllele": "A",
  "alternateAlleleFrequency": 0.163,
  "alternateAlleleReadDepth": 15.9,
  "totalReadDepth": 92,
  "variantType": "SNV",
  "transcripts": {
    "transcriptId": "ENST00000310581.9",
    "transcriptDatabaseName": "Ensembl",
    "transcriptDatabaseVersion": "Release 99",
    "geneSymbol": "TERT",
    "cdsChange": "n.1295113C>T",
    "aminoAcidsChange": null,
    "calculatedEffects": [
      "TF_binding_site_variant"
    ]
  },
  "testMethod": "DNA-seq",
  "variantOrigin": "somatic",
  "reported": true
}
```

V-2. copyNumberAlterations タグ

検出されたコピー数異常に関して記載するタグである。

キー	必須条件	データ型	説明
copyNumberAlterations	任意	配列(長さ 1 ～N、オブ ジェクト)	コピー数異常集約タグ。 配列の各オブジェクトは一意であること。

V-2-1. copyNumberAlterations タグ内の解説

キー	必須条件	データ型	説明
itemId	必須	文字列 正規表現： ^.+	遺伝子異常に割り当てる ID。 一症例内で一意の文字列であること。
chromosome	任意	文字列 正規表現：	染色体番号

		^[a-zA-Z0-9_\\-]+\$	
startPosition	任意	整数	染色体上の開始位置(1-based による記載)
endPosition	任意	整数	染色体上の終了位置(1-based による記載)
copyNumberMetrics	任意	配列(長さ: 0~N、オブジェクト)	コピー数異常の測定値と単位。 value、unit の 2 つのキーから成るオブジェクトの配列。unit が異なる値が 2 つ以上ある場合は、長さ 2 以上の配列で登録。 配列の各オブジェクトは一意であること。
value	必須	数値	コピー数異常の測定値
unit	必須	文字列 [選択式]	測定値 value の単位。 単位は次から選択可能。 <ul style="list-style-type: none"> ・ "absolute copy number": 絶対コピー数 ・ "fold-change": 正常検体に対する腫瘍検体の（標準化された）読取深度の比 ・ "log2 fold-change": “fold-change”の log2 変換 ・ "fraction-of-gene": 測定された遺伝子領域のうちコピー数変化した領域の割合。 ※その他の単位を使用する場合には、事前に C-CAT に要相談。
copyNumberAlterationType	必須	文字列 [選択式]	検査会社が報告書に記載したコピー数異常のタイプ。 次から選択して記載。 <ul style="list-style-type: none"> ・ "amplification" ・ "gain" ・ "duplication" ・ "loss" ・ "deletion" ・ "homozygous deletion" ・ "neutral" ※上記以外のタイプを使用する場合は、事前に C-CAT に要連絡。
transcripts	必須	配列(長さ: 1~N、オブジェクト)	shortVariants タグ内の解説を参照

transcriptId	任意	文字列 正規表現： ^[^\s]+\$	shortVariants タグ内の解説を参照
transcriptDatabaseName	任意	文字列 [選択式]	shortVariants タグ内の解説を参照。 transcriptId を入力した場合、本項目も入力を推奨する。
transcriptDatabaseVersion	任意	文字列 正規表現： ^\.+\$	shortVariants タグ内の解説を参照
geneSymbol	必須	文字列 正規表現： ^[^\s]+\$	shortVariants タグ内の解説を参照
strand	任意	文字列 [選択式]	shortVariants タグ内の解説を参照
cdsChange	任意	文字列 正規表現： ^\.+\$	shortVariants タグ内の解説を参照
aminoAcidsChange	任意	文字列 正規表現： ^\.+\$	shortVariants タグ内の解説を参照
calculatedEffects	任意	配列(長さ：0～N、文字列 正規表現： ^\.+\$)	shortVariants タグ内の解説を参照
testMethod	必須	文字列 [選択式]	shortVariants タグ内の解説を参照
variantOrigin	任意	文字列 [選択式]	shortVariants タグ内の解説を参照
reported	必須	ブール型	shortVariants タグ内の解説を参照

V-2-2. copyNumberAlterations タグ記載例

```
(例)
{
  "itemId": "variant-9",
  "chromosome": "1",
  "startPosition": 8921059,
  "endPosition": 8939151,
  "copyNumberMetrics": [
```

```

{
  "value": 0.2309,
  "unit": "fold-change"
},
{
  "value": -2.1147,
  "unit": "log2 fold-change"
}
],
"copyNumberAlterationType": "loss",
"transcripts": [
{
  "transcriptDatabaseName": "RefSeq",
  "transcriptDatabaseVersion": "Release 99",
  "geneSymbol": "EN01"
}
],
"testMethod": "DNA-seq",
"variantOrigin": "somatic",
"reported": false
}

```

V-3. rearrangements タグ

遺伝子融合、重複、長い欠失、逆位などの遺伝子再構成の情報を記載する。

キー	必須条件	データ型	説明
rearrangements	任意	配列(長さ 1～N、オブジェクト)	遺伝子再構成情報集約タグ。 配列の各オブジェクトは一意であること。

V-3-1. rearrangements タグ内の解説

キー	必須条件	データ型	説明
itemId	必須	文字列 正規表現： <code>^[.]+\$</code>	遺伝子異常に割り当てる ID。 一症例内で一意の文字列であること。
breakends	必須	配列(長さ: 2、オブジェクト)	遺伝子再構成の 2 つのブレイクエンド。 配列の各オブジェクトは一意であること。
chromosome	必須	文字列 正規表現： <code>^[a-zA-Z0-9_\\-]+\$</code>	染色体番号

startPosition	必須	整数	染色体上の開始位置(1-based による記載)
endPosition	必須	整数	染色体上の終了位置(1-based による記載)
matePieceLocation	任意	文字列 [選択式]	参照配列上の結合点から見たときに、物理位置上流で別の配列と結合している場合は"upstream"、物理位置下流で結合している場合は"downstream"。遺伝子融合やその他の再構成("rearrangementType": "other") の場合は、正確なゲノム変化の把握のため、記載することを強く推奨する。詳細な説明は後述「VII-2. matePieceLocation」を参照。
totalReadCount	任意	整数	ブレイクエンドとしての合計リード数。
alternateAlleleFrequency	任意	数値	ブレイクエンドとしての変異型アレル頻度(0~1 の範囲) ブレイクエンドごとにアレル頻度を計算した値を記載する。
transcripts	必須	配列(長さ: 1~N、オブジェクト)	shortVariants タグ内の解説を参照
transcriptId	任意	文字列 正規表現: <code>^[^\s]+\$</code>	shortVariants タグ内の解説を参照
transcriptDatabaseName	任意	文字列 [選択式]	shortVariants タグ内の解説を参照。 transcriptId を入力した場合、本項目も入力を推奨する。
transcriptDatabaseVersion	任意	文字列 正規表現: <code>^[.]+\$</code>	shortVariants タグ内の解説を参照
geneSymbol	必須	文字列 正規表現: <code>^[^\s]+\$</code>	shortVariants タグ内の解説を参照

strand	任意	文字列 [選択式]	shortVariants タグ内の解説を参照
cdsChange	任意	文字列 正規表現： ^.+ \$	shortVariants タグ内の解説を参照
aminoAcidsChange	任意	文字列 正規表現： ^.+ \$	shortVariants タグ内の解説を参照
calculatedEffects	任意	配列(長さ：0～N、文字列 正規表現： ^.+ \$)	shortVariants タグ内の解説を参照
genePairs	任意	配列(長さ：0～N、文字列 正規表現： ^.+ \$)	<p>遺伝子再構成においてブレークエンドをまたぐ遺伝子名の転写方向順序ありペア。</p> <p>transcripts に記載した geneSymbol を、転写の順方向に合わせて "-" でつなぐ。たとえば、geneSymbol A, B に対し、A が転写上流、B が転写下流にある場合、"A-B" と表す。</p> <p>※がん知識ベース検索の際、この方向情報が加味される。本タグの記載が無い場合は、geneSymbol のペアは、転写方向の情報が無いものとして扱われ、遺伝子名の順序なしペアで知識ベースが検索される。</p> <p>※本タグを省略する（かつ、matePieceLocation の記載がない）場合は、geneSymbol のあらゆるペアが順序なしで検索されますが、検索を細かく制御したい場合は、["A-B", "B-A", "C-B"] のように記載下さい。</p> <p>配列の各文字列は一意であること。</p>
insertedSequence	任意	文字列 正規表現： ^[ACGTN]+\$	<p>ゲノム配列の2つのブレークエンドの間に挿入された配列。</p> <p>挿入された配列が存在しない場合は、null を記載する。</p>

supportingReadCount	任意	整数	サポートリード数
totalReadCount	任意	整数	合計リード数
alternateAlleleFrequency	任意	数値	遺伝子再構成としての変異型アレル頻度(0~1 の範囲) 1 つの遺伝子再構成として計算した値を記載する。
expressionLevelMetrics	任意	配列(長さ: 0~N, オブジェクト)	RNA-seq から得られた発現量の情報。 配列の各オブジェクトは一意であること。
value	必須	数値	発現量の値
unit	必須	文字列 [選択式]	発現量の単位。 単位は次から選択。 <ul style="list-style-type: none"> • "TPM" • "FPKM" • "FPM" • "RPKM" • "RPM" ※上記以外の単位を使用する場合は、事前に C-CAT に要連絡。
rearrangementNames	任意	配列(長さ: 0~N、文字列 正規表現: ^\.[+)\$	検査会社が付けた個々の遺伝子再構成の名前。例えば、"EML4-ALK fusion"。 配列の各文字列は一意であること。
rearrangementType	必須	文字列 [選択式]	検査会社が報告書に記載した遺伝子再構成のタイプ。 次の選択肢から選択。 <ul style="list-style-type: none"> • "gene fusion" • "gene fusion and frameshift variant" • "bidirectional gene fusion" • "duplication" • "deletion" • "inversion" • "truncation" • "splice variant" • "tandem duplication" • "other"

			※上記以外のタイプを使用する場合は、事前に C-CAT に要連絡。
testMethod	必須	文字列 [選択式]	shortVariants タグ内の解説を参照
variantOrigin	任意	文字列 [選択式]	shortVariants タグ内の解説を参照
reported	必須	ブール型	shortVariants タグ内の解説を参照

V-3-2. rearrangements タグ記載例

(例 1. genePairs が存在する場合)

```
{
  "itemId": "variant-13",
  "breakends": [
    {
      "chromosome": "2",
      "startPosition": 42510050,
      "endPosition": 42510050,
      "matePieceLocation": "downstream",
      "transcripts": [
        {
          "transcriptDatabaseName": "RefSeq",
          "transcriptDatabaseVersion": "Release 99",
          "geneSymbol": "EML4"
        }
      ]
    },
    {
      "chromosome": "2",
      "startPosition": 29445240,
      "endPosition": 29445240,
      "matePieceLocation": "upstream",
      "transcripts": [
        {
          "transcriptDatabaseName": "RefSeq",
          "transcriptDatabaseVersion": "Release 99",
          "geneSymbol": "ALK"
        }
      ]
    }
  ],
  "genePairs": [
    "EML4-ALK"
  ],
  "supportingReadCount": 30,
  "totalReadCount": 430,
  "alternateAlleleFrequency": 0.07,
```

```
"rearrangementType": "other",
"variantOrigin": "somatic",
"testMethod": "DNA-seq",
"reported": false
}
```

(例 2. ブレークエンドごとに合計リード数、アレル頻度を記載する場合)

```
{
  "itemId": "variant-13",
  "breakends": [
    {
      "chromosome": "2",
      "startPosition": 42510050,
      "endPosition": 42510050,
      "matePieceLocation": "downstream",
      "totalReadCount": 330,
      "alternateAlleleFrequency": 0.06,
      "transcripts": [
        {
          "transcriptDatabaseName": "RefSeq",
          "transcriptDatabaseVersion": "Release 99",
          "geneSymbol": "EML4"
        }
      ]
    },
    {
      "chromosome": "2",
      "startPosition": 29445240,
      "endPosition": 29445240,
      "matePieceLocation": "upstream",
      "totalReadCount": 570,
      "alternateAlleleFrequency": 0.07,
      "transcripts": [
        {
          "transcriptDatabaseName": "RefSeq",
          "transcriptDatabaseVersion": "Release 99",
          "geneSymbol": "ALK"
        }
      ]
    }
  ],
  "genePairs": [
    "EML4-ALK"
  ],
  "supportingReadCount": 30,
  "rearrangementType": "other",
  "variantOrigin": "somatic",
  "testMethod": "DNA-seq",
  "reported": false
}
```

(例 3. insertedSequence が存在する場合)

```
{
  "itemId": "variant-14",
  "breakends": [
    {
      "chromosome": "14",
      "startPosition": 234567,
      "endPosition": 234567,
      "matePieceLocation": "downstream",
      "transcripts": [
        {
          "transcriptDatabaseName": "RefSeq",
          "transcriptDatabaseVersion": "Release 99",
          "geneSymbol": null
        }
      ]
    },
    {
      "chromosome": "2",
      "startPosition": 321672,
      "endPosition": 321672,
      "matePieceLocation": "upstream",
      "transcripts": [
        {
          "transcriptDatabaseName": "RefSeq",
          "transcriptDatabaseVersion": "Release 99",
          "geneSymbol": "LINC01865"
        }
      ]
    }
  ],
  "insertedSequence": "GTNNNNNCAT",
  "supportingReadCount": 30,
  "alternateAlleleFrequency": 0.07,
  "rearrangementType": "other",
  "variantOrigin": "somatic",
  "testMethod": "DNA-seq",
  "reported": false
}
```

VI. otherBiomarkers タグ

variants タグにある遺伝子異常以外のバイオマーカーに関する情報を記載する。現在、マイクロサテライト不安定性(MSI)、腫瘍変異負荷(TMB)、ヘテロ接合性の消失(LOH)の3種類のバイオマーカーに対応している。

キー	必須条件	データ型	説明
otherBiomarkers	任意	配列(長さ 0～N、オブジェクト)	variants タグにある遺伝子異常以外のバイオマーカーに関する情報集約タグ。 配列の各オブジェクトは一意であること。

VI-1. otherBiomarkers タグ内の解説

キー	必須条件	データ型	説明
itemId	必須	文字列 正規表現： ^\.+\$	バイオマーカーに割り当てる ID。 一症例内で一意の文字列であること。
biomarkerType	必須	文字列 [選択式]	記載するバイオマーカーのタイプ。 タイプは次から選択。 ・ "TMB": 腫瘍変異負荷(Tumor Mutation Burden) ・ "MSI": マイクロサテライト不安定性(Micro-Satellite Instability) ・ "LOH": ヘテロ接合性の消失(Loss Of Heterozygosity) ※上記以外のタイプを使用する場合は、事前に C-CAT に要連絡。
biomarkerMetrics	任意	配列(長さ 0～N、オブジェクト)	検査数値と単位。 配列の各オブジェクトは一意であること。
value	必須	数値	検査数値 (例) 5.15
unit	必須	文字列 正規表現： ^\.+\$	検査数値の単位。 単位は検査内容によって異なる。 (例) %
state	任意	文字列 [選択式]	バイオマーカーの状態を次の選択肢から選択。 ・ "high" ・ "low"

			<ul style="list-style-type: none"> • "intermediate" • "stable" • "cannot be determined" <p>検査を実施したが、結果が上記以外の場合は、null を記載。</p> <p>※上記以外の選択肢を使用する場合は、事前に C-CAT に要連絡。</p>
descriptions	任意	配列(長さ 0～N、文字列正規表現： ^.+ \$)	<p>検査値の求め方や意味などの説明文。</p> <p>使用言語および改行については、comments タグ内の contents タグを参照。</p>
biomarkerOrigin	任意	文字列 [選択式]	shortVariants タグ内の解説を参照
reported	必須	ブール型	shortVariants タグ内の解説を参照

VI-2. otherBiomarkers タグ記載例

(例)

```
"otherBiomarkers": [
{
  "itemId": "biomarker-1",
  "biomarkerType": "MSI",
  "biomarkerMetrics": [
    {
      "value": 5.15,
      "unit": "%"
    },
    {
      "value": 2,
      "unit": "MSIsensor score"
    }
  ],
},
],
```

↑ 同じ検査項目で異なる単位の値が複数存在する場合は、"biomarkerMetrics" タグ内に配列表記する。

```
  "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"
  ],
  "reported": true
},
{
```



```
"itemId": "biomarker-2",
"biomarkerType": "TMB",
"biomarkerMetrics": [
  {
    "value": 34.5680122,
    "unit": "Muts/Mb"
  }
],
"state": "high",
"reported": true
},
{
  "itemId": "biomarker-3",
  "biomarkerType": "LOH",
  "biomarkerMetrics": [
    {
      "value": 24.14,
      "unit": "%"
    }
  ],
  "state": "neutral",
  "reported": true
}
]
```

VII. compositeBiomarkers タグ

shortVariants タグや copyNumberAlterations タグ、rearrangements タグの要素を組み合わせて、複合マーカー（遺伝子異常の組み合わせ、3 遺伝子の融合など）の情報を記載する。

キー	必須条件	データ型	説明
compositeBiomarkers	任意	配列(長さ 0～N、オブジェクト)	複合マーカーの情報集約タグ。 配列の各オブジェクトは一意であること。

VII-1. compositeBiomarkers タグ内の解説

キー	必須条件	データ型	説明
itemId	必須	文字列 正規表現： ^\.+\$	複合マーカーに割り当てる ID。 一症例内で一意の文字列であること。
componentItemIds	必須	配列(長さ 2～N、文字列 正規表現： ^\.+\$)	対象となる遺伝子異常(itemId)の配列。 配列の各文字列は一意であること。
biomarkerNames	必須	配列(長さ 1～N、文字列 正規表現： ^\.+\$)	検査会社の報告書に記載されている複合マーカー名を記載。 配列の各文字列は一意であること。
descriptions	任意	配列(長さ 0～N、文字列 正規表現： ^\.+\$)	複合マーカーの説明文。 使用言語および改行については、comments タグ内の contents タグを参照。
reported	必須	ブール型	shortVariants タグ内の解説を参照

VII-2. compositeBiomarkers タグ記載例

(例)

```
"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  
},  
]
```

VIII. sequencingSamples タグ

次世代シーケンサーによる sequencing sample 情報を記載する。

キー	必須条件	データ型	説明
sequencingSamples	任意	配列(長さ 1~N、オブジェクト)	sequencing sample 情報集約タグ。 配列の各オブジェクトは一意であること。

sequencingSamples タグの配列の長さは、tumorOrNormal の値が腫瘍検体(tumor)または正常検体(normal)、testMethod の値が DNA 検体(DNA-seq)または RNA 検体(RNA-seq)とすると、最大で 4 である。

VIII-1. sequencingSamples タグ内の解説

キー	必須条件	データ型	説明
itemId	必須	文字列 正規表現： <code>^[.]+\$</code>	項目の ID。 一症例内で一意の文字列であること。
tumorOrNormal	必須	文字列 [選択式]	sequencing sample 情報が腫瘍検体(tumor)か正常検体(normal)かを記載。 ・ "tumor" ・ "normal"
testMethod	必須	文字列 [選択式]	得られた検出結果が DNA 検体由来か RNA 検体由来かを記載。 ・ "DNA-seq": DNA 検体由来 ・ "RNA-seq": RNA 検体由来
duplicateReadsPercentage	任意	数値	重複率(%)
mappedReadsPercentage	任意	数値	マッピング率(%)
meanReadDepth	任意	数値	平均読取深度
medianReadDepth	任意	数値	読取深度の中央値
suspectedSampleStates	任意	配列(長さ 0~N、文字列) [選択式]	「コンタミ」のような疑わしい DNA または RNA サンプルの状態を選択（配列内で重複不可）。 ・ "contaminated": コンタミの可能性

			<ul style="list-style-type: none"> • "deaminated": FFPE DNA におけるシトシンの脱アミノ化が顕著な可能性 • "fragmentated": (FFPE DNA) 断片化が顕著な可能性 • "degraded": (RNA) 分解が顕著な可能性
--	--	--	--

VIII-2. sequencingSamples タグ記載例

(例)

```
"sequencingSamples": [
  {
    "itemId": "sequence-1",
    "tumorOrNormal": "tumor",
    "testMethod": "DNA-seq",
    "duplicateReadsPercentage": 91.52,
    "mappedReadsPercentage": 87.31,
    "meanReadDepth": 247.8,
    "medianReadDepth": 238
  },
  {
    "itemId": "sequence-2",
    "tumorOrNormal": "tumor",
    "testMethod": "RNA-seq",
    "suspectedSampleStates": [
      "degraded"
    ]
  },
  {
    "itemId": "sequence-3",
    "tumorOrNormal": "normal",
    "testMethod": "DNA-seq"
  }
]
```

IX. 注意事項

注意事項について記述する。

IX-1. itemId

itemId の値は、ファイル内で一意であること。値は、任意の文字列。

IX-1-1. itemId の記載例

itemId が登場する各タグ内での itemId の値の記載例について示す。

- variants タグ内

"itemId": "variant-1"

"itemId": "variant-2"

"itemId": "variant-3"

- otherBiomarkers タグ内

"itemId": "biomarker-1"

"itemId": "biomarker-2"

"itemId": "biomarker-3"

- sequencingSamples タグ内

"itemId": "sequence-1"

"itemId": "sequence-2"

"itemId": "sequence-3"

- compositeBiomarkers タグ内

"itemId": "composite-1"

"itemId": "composite-2"

"itemId": "composite-3"

IX-2. matePieceLocation

breakends タグ内に存在する matePieceLocation について解説する。前提として、染色体の参照配列上で物理位置の番地が少ない場所を上流 (upstream)、番地が多い場所を下流 (downstream) とする。

※ 転写産物配列の上流・下流とは異なる

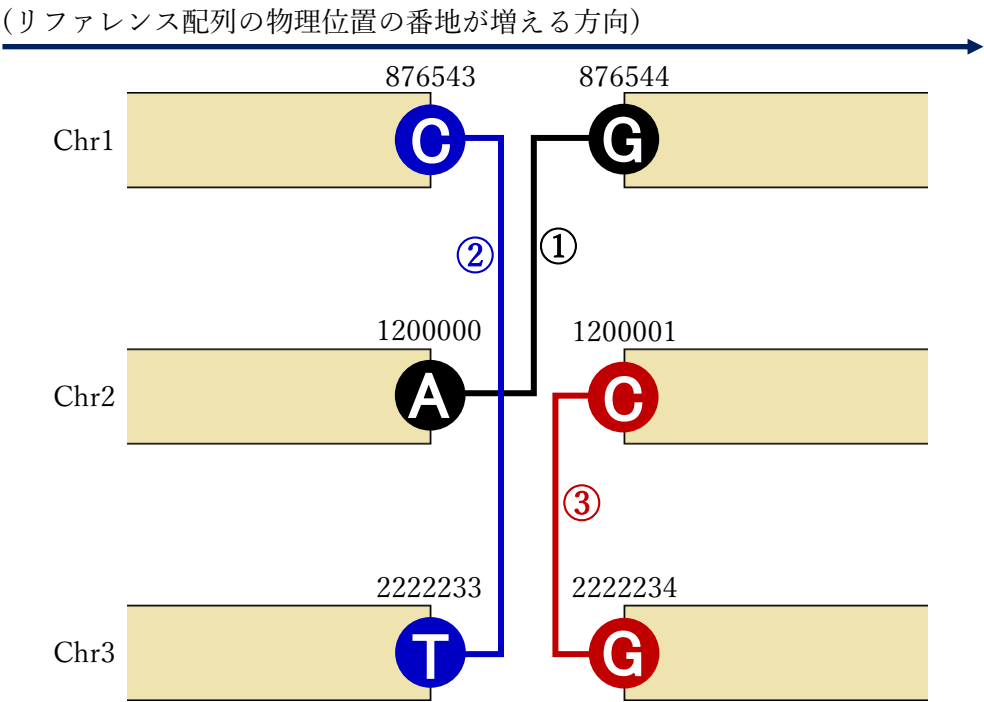
IX-2-1. matePieceLocation の記載例

再構成の例を挙げて VCF 形式 (v4.3) での表現を提示しながら matePieceLocation の記載方法を示す。

なお再構成と VCF の例は以下のドキュメントの図表を改変したものである:

The Variant Call Format Specification VCF v4.3 and BCF v2.2
<https://samtools.github.io/hts-specs/VCFv4.3.pdf>

再構成の例:



VCF 形式での記載:

上図のそれぞれの遺伝子再構成の VCF v4.3 に従った記載は次のようになる。以下例の左にある番号が上図の番号に対応する。

	#CHROM	POS	ID	REF	ALT	QUAL	FILTER	INFO
①	1	876544	bnd_V	G]2:1200000]G	6	PASS	SVTYPE=BND
	2	1200000	bnd_U	A	A[1:876544[6	PASS	SVTYPE=BND
	1	876543	bnd_W	C	C]3:2222233]	6	PASS	SVTYPE=BND

②	3	2222233	bnd_Y	T	T]1:876543]	6	PASS	SVTYPE=BNĐ
③	2	1200001	bnd_X	C	[3:2222234[C	6	PASS	SVTYPE=BNĐ
	3	2222234	bnd_Z	G	[2:1200001[G	6	PASS	SVTYPE=BNĐ

本フォーマットでの記載:

・ 例①

染色体番号 2 側の結合点から見ると、参照配列下流側が別の配列に置き換わっている
で、ブレイクエンドのオブジェクトの `matePieceLocation` の値は"downstream"となる。
一方、染色体番号 1 側の結合点から見ると、上流側が別の配列に置き換わっている
ので、`matePieceLocation` の値は"upstream"となる。

(例①)

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

・ 例②

染色体番号 1 側の結合点から見ると、参照配列下流側が別の配列に置き換わっている
で、オブジェクトの `matePieceLocation` の値は "downstream" となる。
染色体番号 3 側の結合点から見た場合も同様なので、`matePieceLocation` の値は
"downstream" となる。

(例②)

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



```
"chromosome": "3",  
"startPosition": 2222233,  
"endPosition": 2222233,  
"matePieceLocation": "downstream"  
}  
]
```

・ 例③

染色体番号 2 側の結合点から見ると、参照配列上流側が別の配列に置き換わっているの
で、オブジェクトの matePieceLocation の値は "upstream" となる。

染色体番号 3 側の結合点から見た場合も同様なので、matePieceLocation の値は
"upstream" となる。

(例③)

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

X. お問い合わせ先

C-CAT ヘルプデスク

E-Mail: helpdesk_c-cat@ml.res.ncc.go.jp

以上