HL7 FHIR Proposal for Copy Number Range and Decimal

Version 2

What's new in this version?

We propose removing the invariant currently on the **copy-number** component because it is not possible to apply constraints to only one of multiple data types. We propose adding guidance to the component with the same content as two of the constraints. A modification to the definition has also been added.

Background

Copy number variants are often described with a single copy number value. However, due to the nature of assays, such as qPCR[1] and microarray[2], it can be difficult to assign an exact copy number. Therefore, labs sometimes report a range instead of an exact copy number. They may alternatively choose to report decimal numbers, even if whole numbers are biologically correct.

In general, it can be difficult to estimate a copy number from measurements taken on a heterogenous population of cells, such as a tumor. As indicated in the ISCN 2020 specifications (see next section), it is appropriate to express the measurements using a range. Furthermore, regardless of cell population, many assays use fluorescent probes to detect copy number. These testing methodologies do not always provide an exact, discrete value for this data point.

According to Ma et al.[1], qPCR assays can produce "results [that] fall between integers (e.g., copy number measuring 1.3), making interpretation more difficult." These assays amplify a locus with a known copy number as a control, and a locus whose copy number is under investigation, which can be compared against the control. Even when the PCR run is performed with the utmost care and under the same conditions between the control and target samples, it is still possible to end up with an inexact copy number measurement due to the nature of molecular kinetics. Currently, labs may report this output case as a decimal or a range.

Our customers have requested that we implement copy number range into our variant discrete data model so that they can view this information properly, as well as to allow decimal copy numbers to be stored. In turn, we are bringing this feedback to the HL7 Clinical Genomics Work Group.

Existing conventions

ISCN

According to ISCN 2020[3], 14.2.5 (Mixed Cell Populations and Uncertain Copy Number) indicates the use of ranges to represent copy number if an estimation is not available.

> To indicate a mixed cell population, the proportion of the sample with the abnormality can be estimated and included in brackets following the copy number. If the proportion of abnormal DNA cannot be estimated, the copy number range should be given using a tilde (~) or **mos** may be used.

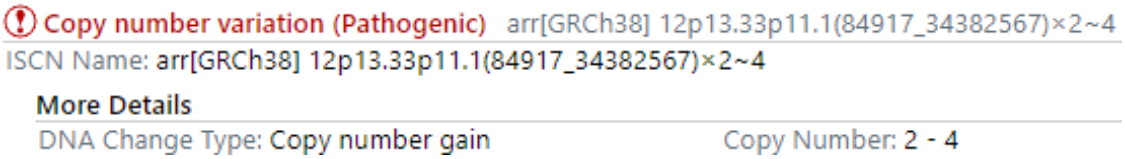
It also provides an example:

arr[GRCh38] 12p13.33p11.1(84917_34382567)×2~4

Microarray analysis shows a two copy gain of the short arm of chromosome 12, resulting in tetrasomy 12p. Although this result likely indicates an isochromosome of 12p, such as those found in Pallister Killian syndrome, FISH or chromosome analysis is required to confirm. The approximate sign is used to indicate that the number of copies of this region varies from 2 to 4.

Representation in Epic

In Epic, copy number ranges can be displayed in variant details. The following screenshot represents the ISCN example, with a copy number range of 2 to 4.



Proposed Changes to FHIR IG

We propose changing the data type of the **copy-number** component on the R4 **Variant** resource profile from **Quantity** to **Quantity | Range**.

Example

```
.component[1].code = 82155-3
.component[1].value.low = 2
.component[1].value.high = 4
...
```

We also considered allowing just the **Range** data type and inserting guidance that "'High' can be omitted for single [values]", like **exact-start-end**. However, there is a valid use case for representing copy number ranges with a lower bound, but not an upper bound, which would not be representable with this guidance in place.

Constraints

As discussed in the Background section, we recommend allowing decimal copy numbers. In version 1 of this proposal, we suggested removing the constraint "If present, the value SHALL be a whole number" from the existing instance of invariant cnt-3. However, while it may be possible to create a new invariant with the remaining two invariants (below), to our understanding, it does not appear to be possible to apply an invariant to only one data type, so we propose to remove the invariant entirely. Nevertheless, we also propose to add guidance for the two remaining (to be removed) constraints:

1. There SHALL be a code with a value of '1' if there is a value.

2. If system is present, it SHALL be UCUM.

This guidance will help prevent the use of nonsensical units (e.g., 3 mL) and unrecognized systems.

Changes to definition

The current definition of the **copy-number** component is:

> The copy number of the large variant. In HGVS, this is the numeric value following the “X”. It is a unit-less value. Note that a copy number of 1 can imply a deletion.

We propose changing it to:

> The copy number of the large variant. In HGVS, this is the numeric value following the “X”. It is a unit-less value. If the value determined by the assay is not a whole number, a range or decimal number can be specified. Note that a copy number of 1 can imply a deletion.

References

1 Ma L, Chung WK. Quantitative analysis of copy number variants based on real-time LightCycler PCR. Current protocols in human genetics. 2014 Jan;80(1):7-21.

2 Nagl S, Schaeferling M, Wolfbeis OS. Fluorescence analysis in microarray technology. Microchimica Acta. 2005 Sep 1;151(1-2):1-21.

3 McGowan-Jordan J, Hastings RJ, Moore S, editors. ISCN 2020: An International System for Human Cytogenomic Nomenclature (2020). Reprint Of: Cytogenetic and Genome Research 2020, Vol. 160, No. 7-8. Karger, S; 2020.