

GUIDELINES FOR SUBMITTING VARIATION DATA BASED ON HGVS NOMENCLATURE

The LRG collaboration has developed a template file to support submission of information about sequence variants that have identified to differ from a reference standard. Although submission based on the LRG sequence as the reference standard is highly recommended, submissions based on a public sequence, specified by accession and version, will be accepted.

The following documents the steps to follow:

1. OBTAIN THE TEMPLATE FILE

The template for submitting the information about observed variations can be obtained from either of the central databases:

EBI: ftp://ftp.ebi.ac.uk/pub/databases/lrgex/docs/submission_template.xls

NCBI:

ftp://ftp.ncbi.nih.gov/refseq/H_sapiens/RefSeqGene/submission_template.xls

2. SELECT THE PUBLIC SEQUENCE FOR YOUR REPORTING STANDARD

We strongly recommend use of an LRG or a RefSeqGene (with the version) as the sequence standard in the [HGVS name](#). We will process any submission based on a sequence in a public database (*e.g.* [INSDC](#)), but there may be delays if the alignment of that sequence to an LRG/RefSeqGene is not perfect. For all sequences that are versioned (INSDC, RefSeq, RefSeqGene), you must supply the version in the HGVS expression.

Examples:

- a. Single nucleotide change based on a genomic sequence (based on [rs72659347](#))
 - i. LRG_5:g.5505C>A
 - ii. NG_008123.1:g.5505C>A
- b. Single nucleotide change based on a coding DNA sequence (based on [rs72659347](#))

- i. LRG_5t1:c.392C>A
- ii. NM_022356.3:c.392C>A
- iii. LRG_5t2: c.392C>A
- iv. NM_001146289.1:c.392

3. IDENTIFY THE HGVS NOMENCLATURE TO REPORT EACH VARIANT

The Human Genome Variation Society ([HGVS](http://www.hgvs.org), www.hgvs.org) has established standards for reporting variation based on a reference sequence. This site

<http://www.hgvs.org/mutnomen/>

provides complete documentation for the method, which involves reporting the position and the sequence change observed. The position is unambiguously established on the sequence named as the numbering standard. HGVS provides recommendations for genomic, cDNA and protein standards, and these rules will not be repeated here.

We strongly recommend that you submit your data with nucleotide sequence coordinates (as exemplified above), rather than protein. As you understand, some protein changes could be caused by more than one nucleotide change, and it is much more useful to know the exact nucleotide variation that has been observed, if that is what your assay or report has detected.

4. COMPLETE THE FORM

The submission form contains some instructions for each field, indicating whether input is required or optional. You will note that lines 10-14 request information about the submission and line 19 onwards request information about the variants themselves. Please note that in the variants section, you can use one line to submit information about each observation; or summarize data according to a variant/ethnicity/gender combination.

- a. The summary data about the submission.
 - i. Lines 10/11 are used to provide a public identifier for the submitter. This identifier must be unique (e.g. a URL), and support web-based lookup of the submitters identity.
 - ii. Line 12 is used to identify the submission unit or batch. This must be unique in combination with the submitter's identification. If you think you may ever want to update your submission, you should choose your batch identifiers carefully and remember to include them in your update. The central databases will compare your current submission to any previous submission from you with the same batch name. If a previous

submission is identified, the databases will use the difference between the submissions to insert, add, or update data. In other words, if you submit a batch named BBBBBB which contains 10 records, and then re-submit a batch named BBBBBB with only 1 record (which was not in the previous 10), then previous 10 will be treated as 'withdrawn' batch BBBBBB from you will now contain only one current record. If a previous batch is not identified, all data submitted will be treated as new.

- iii. Lines 13 and 14 provide information about the person to contact if there are any questions about a submission.
- b. Variation-specific information
 - i. Not all columns need to be completed. Those which are absolutely required are:
 - 1. A. The HGVS name of the variant
 - 2. J. The number of independent observations (if not submitted, the value will be converted to 1)
 - 3. L. Whether the variant is germline, somatic, or unknown in the sample as described
 - ii. We strongly encourage submission of data in the rest of the columns. In particular, submission of columns F and K will allow the central databases to provide URLs to your internet resource.

5. SUBMIT THE FORM

Once your form is completed, you can e-mail it to submit-variants@lrg-sequence.org. Alternatively, if the submission is large (> 4 Mb), upload it to our FTP server <ftp-private.ebi.ac.uk>. To obtain a login and password, please send an e-mail with subject line '[LRG] FTP details' to submit-variants@lrg-sequence.org.

6. DOWNSTREAM PROCESSING

EBI/NCBI will process your submission, converting the data to an accessioned record in dbSNP or [dbVar/DGV](#) as appropriate. Any uncertainty in determining the location on the reference nucleotide sequence that resulted in the observed change in amino acid sequence will delay complete processing of your submission. If we are unable to resolve the ambiguity, the submission will be returned to you for clarification. Within two weeks, that submission will be compared to other submissions, clustered on the LRG/RefSeqGene, and assigned a reference identifier (e.g. rs#) . You will be given an opportunity to review your submission before it is made public.

7. REPORTS

When data are completely processed and ready to be made public, a report will be generated merging the data you supplied with any other information that may be available

about that variant. These data will be integrated into the updatable layer of the LRG, as well as provided for download from the central databases.