UKAS PROFICIENCY TESTING
4367

EMQN CIC Office
Unit 4, Enterprise House
Pencroft Way
Manchester Science Park
Manchester, M15 6SE, UK
Tel: +44 (0)161 757 1591
Email: office@emqn.org

EMQN CIC Policy on Variant Nomenclature

(v13, updated 2024)

BACKGROUND

Over the years, EMQN CIC has observed extensive disparities in the nomenclature used to describe DNA sequence variants. Variant nomenclature inconsistencies are unacceptable, potentially leading to misinterpretation both in laboratory and clinical settings (Tack *et al*., 2016). To avoid confusion and clinical errors it is important that diagnostic laboratories lead the way in adopting a standardised approach to describing genomic variation.

Nomenclature guidelines are available from the Human Genome Variation Society (HGVS), the international body for defining gene variation nomenclature, under the umbrella of the Human Genome Organization (HUGO) and the International Federation of Human Genetics Societies (IFHGS). EMQN CIC recommends that all labs follow HGVS guidelines when reporting genomic variation (website: http://varnomen.hgvs.org/). However, we recognise the guidance provided by HGVS for certain types of variation is difficult to interpret. Please see below for further guidance.

POLICY

We have produced the following notes to assist in the interpretation of HGVS nomenclature guidelines.

1. **The EMQN CIC expects laboratories to report their results using HGVS nomenclature** (http://varnomen.hgvs.org/) **recommending that laboratories implement the latest version (currently 20.05). EMQN CIC will not usually deduct marks for minor HGVS errors and will allow a reasonable timescale for laboratories to adapt their reporting policies when new versions of HGVS are published (usually a complete annual cycle of EQA schemes). During this time period comments will be made but no points will be deducted for failure to implement new guidance. EMQN CIC does not expect laboratories to amend their nomenclature for guidance which has not yet been accepted e.g. proposed changes still the subject of community consultation.**
2. **Reference sequences:**
    a. A sequence variant should always be described in the context of a reference sequence, referred to by means of a unique sequence identifier or accession number. Reference sequences define the numbering system and default state of a sequence and for this reason, reporting variation requires inclusion of the reference sequence for accurate interpretation of the nomenclature supplied. HGVS recommends that reference sequences come from data sources that provide stable and permanent identifiers i.e. RefSeq (NCBI) and Ensembl (EBI).
    b. In recognition of the discontinuation of the LRG reference sequence project, EMQN CIC:
        i. Encourages community engagement with the MANE initiative, a joint project between NCBI and EMBL-EBI, which aims to create a universal standard for transcript annotation in the genomics era.
        ii. EMQN CIC supports the use of MANE Select and MANE Plus Clinical as denoted by the MANE initiative, for the standardization of variant annotation, interpretation and reporting
        iii. Recognizes that the MANE initiative is still in development and will not penalise participating laboratories for using the correct LRG reference in the 2024 scheme rounds.
        iv. EMQN CIC encourages laboratories to review their use of LRG reference sequences and substitute with the appropriate MANE Select or MANE Plus Clinical reference where practicable. Where this is not practicable, please include the reason why in the "Generic Data collection form" submitted with reports in each EMQN scheme in 2024, including describing the impediments to using the MANE Select or MANE Plus Clinical. This information will be summarised in each EMQN 2024 scheme report.
        v. For more information on MANE, please see our webinar here. information on the initiative can be found at the NCBI website or EBI website.
    c. It is recommended to include reference sequence accessions for all genes analysed during testing, even when reporting that no pathogenic sequence variants have been identified. Reporting the reference sequence accessions enables the reader to understand, which transcript/exons have been included within the scope of testing/analysis; this information is particularly pertinent when the cause of a suspected genetic disorder has not been identified.

3. **Numbering: As dictated by HGVS guidelines, nucleotide numbering starts with "c.1" at the A of the ATG translation initiation (start) codon (it is also recommended to clearly state that nucleotide numbering starts with the A of the ATG translation initiation site – this is particularly important in genes where legacy numbering does not start from the ATG initiation codon e.g. *BRCA1* and *BRCA2*).**
    a. Legacy numbering is seen in some genes where variants are named from base 1 of a cDNA reference sequence. This can lead to confusion and inappropriate testing e.g. *BRCA2* gene variants in the BIC database using accession number U43746.1 where the A of the ATG codon is at position 229.
    b. **Legacy numbering is only allowed in addition to approved numbering**. Any different numbering system to that used in the reference sequence (e.g. constituent in legacy numbering), must be noted on your report (for example; state that the numbering system starts with nucleotide 1 being the beginning of the transcript).
4. **Legacy nomenclature**: It is recognised that some recurrent variants come with historical names which do not comply at all with the HGVS recommendations (e.g. 'DeltaF508' in the CFTR gene or 'Factor V Leiden variant' in the FV gene). It appears unrealistic that the use of such legacy names in the clinical context will ever be replaced by systematic nomenclature. Therefore, it is recommended that HGVS nomenclature is used in addition to the legacy name when appropriate.
    a. If application of the HGVS guidelines results in a different nomenclature from that used in the past, then you should quote both old and new nomenclatures to avoid confusion with previous reports of the same variant.
5. In a disease such as cystic fibrosis, laboratory reports frequently list the panel of variants tested. To avoid a lengthy list of both old and HGVS compliant nomenclature, only variants identified in the patient or family need to be quoted using HGVS nomenclature.
6. Where the information is available, it is recommended that genetic variants are described at both the DNA and the protein level. **Due to redundancy in the genetic code, it is not sufficient to describe variants at the protein level only.** The protein description is designated as the unprocessed AA chain. If the protein is a theoretical prediction, this should be indicated using parentheses e.g. p.(Arg21Val).
7. If the testing method used does not unambiguously identify a specific base change, the nomenclature used should reflect this. New nomenclature are now available for incompletely specified bases (http://varnomen.hgvs.org/bg-material/standards/#aacode) which use IUAP-IUB nucleotide/amino acid codes e.g. B for A,G or T or K for G or T or N for A,G,C or T. Similarly, the symbol Xaa may be used for an unspecified or unknown AA in the protein description. Finally, the symbol '^' can also be used to indicate 'or'.
8. In the full disease-specific and molecular pathology EQAs, laboratories are required to use HGVS nomenclature to describe variants, but are not required to provide full HGVS genotypes describing both alleles using HGVS. For example, it is acceptable to state: "...is heterozygous for the *CFTR* c.1521_1523delCTT p.(Phe508del) variant", rather than stating the full genotype c.[1521_1523delCTT];[=] p.[(Phe508del)];[(=)]. If a laboratory chooses to use HGVS that describes alleles it must be written in accordance with HGVS rules concerning description of the normal allele; c.= indicating that the whole reference sequence has been screened and no variation from the reference sequence was detected. See also point 15 and 16
9. **In the technical EQA scheme for Sanger Sequencing** it is no longer required that variants are reported using full HGVS genotype nomenclature e.g. it is acceptable to state: "...is heterozygous for the *CFTR* c.1521_1523delCTT p.(Phe508del) variant", rather than stating the full genotype c.[1521_1523delCTT];[=] p.[(Phe508del)];[(=)]. Remember c.= now means that you have tested the whole cDNA transcript and found no differences from the reference sequence (see point 15 and 16 below). Additionally, it is acceptable to describe an allele with no clinically relevant variants in words rather than using HGVS e.g. no pathogenic variant identified (homozygous).
10. Measurement of zygosity in tumour DNA is problematic: samples may have normal cell contamination, multiple cell lines, or amplifications and it is therefore recommended to omit zygosity for tumour samples.
11. It is recommended that laboratories follow HGVS guidance on the naming of mosaic variants:
    a. c.85C=/>T - The sample is a mix of cells containing c.85C= and c.85C>T
    b. It is currently recommended to restrict the use of this nomenclature to the naming of germline mosaics
12. **Exon numbering:** HGVS nomenclature does not provide recommendations for exon numbering, and states that exon numbers are not required for unambiguous variant descriptions, nucleotide positions are sufficient. However, EMQN recognizes the clinical utility of describing exonic copy number variations (CNVs) in words for clinical reporting and offers the following guidance on reporting exon numbers:
    a. ENSEMBL reference sequences include exon numbering, but NCBI RefSeq reference sequences do not. As Legacy / Custom exon numbering exists historically for some genes, the system used for exon numbering (if used) should be unambiguous for NCBI RefSeq reference sequences. EMQN CIC will therefore assume that whenever exon numbers are provided, systematic exon numbering (numbering exons from start to end including non-translated exons) has been used for NCBI RefSeq reference sequences, unless otherwise stated in the clinical report.

13. Laboratories reporting the presence of exon deletions and duplications will not be penalised for not using HGVS nomenclature (but HGVS may also be used). A clear statement of which exons are deleted/duplicated with an appropriate reference sequence is acceptable. As with other tests, the methodology must be stated in the report and if a commercial kit used then the kit name and version number must be provided.
    a. Please note changes **HGVS MLPA nomenclature to include the format** c.(649+1_650-1)_(1331+1_1332-1)del has not been adopted (SVD-WG003: status new proposal to be made) and therefore this nomenclature is not currently recommended by EMQN CIC.
14. Laboratories reporting the presence of triplet repeat variants, or other large scale genomic changes (for example, results of methylation analysis, uniparental disomy, copy number analysis, uniparental disomy analysis) will not be penalised for not using HGVS nomenclature. However, where a specific gene (rather than genomic region) is being tested then, to properly identify the gene/transcript tested, the reference sequence and the HGNC gene symbol for that gene should be included in all reports.
15. Please note if laboratories report full genotypes, they must be given using correct HGVS nomenclature. From 2019, if minor errors regarding the placement of brackets, semi-colons etc. are observed then no marks will be deducted but feedback comments will be given. If major errors are observed such as incorrect nucleotide or amino acid then appropriate marks will be deducted in the genotyping category.
16. **Describing the 'normal allele:** Recent changes to HGVS nomenclature have led to a variety of different descriptions of the normal allele depending on the situation e.g. a predictive test, sequencing of the whole cDNA, sequencing of a single exon etc. To avoid unnecessarily penalising laboratories EMQN CIC suggests a written description e.g. no pathogenic variant identified, heterozygous for…. or an appropriate HGVS annotation e.g. c.456C= (the C at position 456 is unchanged from the reference sequence) or c.400_600= (there are no differences between the tested sequence and the reference sequence between bases 400 and 600 of the cDNA).

FEEDBACK

Please let us know of any difficulties in applying these guidelines. Feedback from you will help us establish consensus recommendations.

REFERENCES

- Tack, V., Deans, Z. C., Wolstenholme, N., Patton, S., & Dequeker, E. M. C. (2016). What's in a Name? A Coordinated Approach toward the Correct Use of a Uniform Nomenclature to Improve Patient Reports and Databases. *Human Mutation*. https://doi.org/10.1002/humu.22975

SUMMARY OF CHANGES TO DOCUMENT

Version no.	Change(s)
1	New document
2	Updated email address
3	Updated address
4	Major rewrite - Statement added that labs should use v2 nomenclature.
5	Updated to include new sections on use of full HGVS, duplication/deletions and other large scale genomic changes. Also clarified background and policy and added a section for feedback. Indicated version and update date on title so that labs see when reviewed.
6	Updated policy with recommendations on tumour DNA - not using full HGVS and how to describe mosaicism if you do not know if the change is somatic or germline
7	Title changed to "Variant" Nomenclature from "Mutation" nomenclature
8	Major rewrite to address HGVS version 15.11
9	Alternatives for naming the normal allele following introduction of HGVS v15.11
10	Policy 9: Sanger scheme no longer requires the use of 'full' HGVS to specify zygosity, Policy 14 updated to position on minor HGVS errors and Document title HGVS version updated to 19.01
11	HGVS version updated to 20.05, Policy 2 clarified best practice to include a reference sequence even when no pathogenic variant has been identified. Clarified use of full HGVS and naming of 'normal' alleles in the Sanger scheme policy 9. Policy 10 clarified use of mosaic nomenclature is only recommended in the germline context. Updated address, logo and EMQN CIC.
12	Policy 9: Removed recommendation for use of LRG reference sequences and added EMQN position statement for MANE transcripts. Policy 12: Added a separate policy for exon numbering. Updated address, logo and EMQN CIC.
13	Updated the MANE transcript advice. Clarified use of allele nomenclature (points 8, 9, 16). Added new EMQN logo.