information services
gwasanaethau gwybodaeth

VERIFYING NOMENCLATURE OF DNA VARIANTS AND SHARING VARIANT DATA

Jan Higgins[1], Raymond Dalgleish[2], Johan den Dunnen[3], Greg Barsh[4], Peter Causey-Freeman[3], David N. Cooper[5], Sara Cullinan[6], Kay Davies[7], Huw Dorkins[8,9], James Evans[10], Li Gong[11], Issei Imoto[12,13], Teri Klein[11], Bruce Korf[14], Adya Misra[15], Sarah Ratzel[16], Juergen Reichardt[17], Heidi Rehm[18], Katsushi Tokunaga[19], Karen Weck[10,20], Garry R. Cutting[21]

1   American College of Medical Genetics and Genomics, Bethesda, Maryland, USA
2   Department of Genetics and Genome Biology, University of Leicester, Leicester, UK
3   Human Genetics and Clinical Genetics, Leiden University Medical Center, Leiden, The Netherlands
4   HudsonAlpha Institute for Biotechnology, Huntsville, Alabama; Department of Genetics, Stanford University, Stanford, California, USA
5   Institute of Medical Genetics, School of Medicine, Cardiff University, Heath Park, Cardiff CF14 4XN, UK
6   Deputy Editor, American Journal of Human Genetics
7   MDUK Oxford Neuromuscular Centre, Department of Physiology, Anatomy and Genetics, University of Oxford, Oxford, UK
8   St Peter's College, University of Oxford, Oxford, UK
9   Department of Clinical Genetics, University Hospitals of Leicester NHS Trust, Leicester, UK
10   Department of Genetics, School of Medicine, University of North Carolina, Chapel Hill, NC, USA
11   Department of Biomedical Data Science, Stanford University, Stanford, CA, USA
12   The Social, Ethical and Legal Issues Committee, Japanese Society for Genetic Counseling, Tokyo, Japan
13   Department of Molecular Genetics, Aichi Cancer Center Research Institute, Nagoya, Japan
14   Editor-in-chief, American Journal of Human Genetics
15   Senior Editor, PLOS Medicine
16   Science Editor, American Journal of Human Genetics
17   Australian Institute of Tropical Health and Medicine, James Cook University, Smithfield, QLD 4878, Australia
18   Medical & Population Genetics Program and Genomics Platform, Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA; Department of Pathology, Harvard Medical School, Boston, MA 02114, USA; Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA 02142, USA
19   Genome Medical Science Project, National Center for Global Health and Medicine, Tokyo, Japan
20   Department of Pathology and Laboratory Medicine, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA
21   McKusick-Nathans Department of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD, USA

Communicating author: Garry R. Cutting
                Miller Research Building Rm 559
                733 N. Broadway
                Baltimore MD 21205
                gcutting@jhmi.edu

Journal Requirements

Conflict of Interest Statement

Authors will be asked to provide a conflict of interest statement during the submission process. For details on what to include in this section, see the Conflict of Interest section in the Editorial Policies and Ethical Considerations section below. Submitting authors should ensure they liaise with all co-authors to confirm agreement with the final statement.

Keywords

HGVS, DNA variants, ClinVar, LOVD

Documenting the variation in our genomes is important both for research and clinical care. Accuracy in the annotation of DNA variants is therefore essential (Tack, Deans, Wolstenholme, Patton, & Dequeker, 2016);(Yen et al., 2017). To address this issue, the Human Variome Project convened a committee to evaluate the feasibility of requiring authors to verify that all variants submitted for publication complied with a widely accepted standard for annotation. Members of the committee represented expertise in journal publication, variant annotation and data sharing. Following a series of monthly committee discussions and consultations with external experts over a two-year period, a consensus was reached that in principle all published variants should conform to criteria developed by the Human Genome Variation Society (HGVS). Furthermore, committee members agreed that authors were responsible for the accuracy of variant notation and should ideally be required to confirm in writing that all variants reported in their manuscripts conformed to HGVS criteria. In addition, the committee recommended that editors request evidence of verification in the form of output files from online publicly available software tools or records documenting submission to online databases that verified nomenclature [e.g. ClinVar (Landrum et al., 2020), Leiden Open Variation Database (LOVD) (Fokkema et al., 2011)]. Provision of such evidence would assist authors and editors should database curators or others question variant annotation, or in the event that one or more variants could not be annotated according to HGVS criteria.

Recommended guidelines:

1) Variant notation ~~must~~ should adhere to current recommendations of the Human Genome Variation Society (HGVS)(den Dunnen et al., 2016). Variants in the text and tables should be presented at the DNA level, with the option of including (predicted) protein-based designations and RNA-based designations (when RNA is analyzed; see additional considerations below). Ideally, variants will be defined with respect to a specific genomic DNA reference sequence with ~~a~~ the reference genome version clearly appended. The latest nomenclature updates, examples of acceptable nomenclature and guidance concerning reference sequences can be found at https:/varnomen.hgvs.org/.

2) Compliance with HGVS nomenclature ~~must~~ should be verified using tools such as the Mutalyzer program (den Dunnen, 2016) or VariantValidator (Freeman, Hart, Gretton, Brookes, & Dalgleish, 2018) each of which ~~each~~ offers a batch mode to facilitate rapid checking of multiple variant descriptions. Alternatively, variants previously submitted to existing online databases that ensure nomenclature compliance (e.g. ClinVar, LOVD), will have been assigned unique IDs by the database in question. Authors should include those unique IDs with the variants in their submitted manuscripts.

3) Authors ~~must~~ should provide files generated by verification tools or database submission IDs to confirm that all variants compl~~ied~~y with HGVS nomenclature.

4) Editorial staff will review files provided by authors as a compliance check.

5) Authors will be responsible for the accuracy of variant annotation in their manuscripts, just as they are responsible for the accuracy of figures, tables and data files.

Important considerations include:

- Reference sequences defined in the HGVS nomenclature guidelines ~~must~~ should be used ~~for to~~ report~~ing~~ sequence variants. Authors should always include the Accession and Version Number of the relevant reference sequence(s) (*e.g.*, RefSeq NM_003002.3, LRG_9t1 or GenBank NC_000011.10) in the Materials and Methods section and as a footnote in any tables listing variants.

- If alternative nomenclature schemes are commonly found in the literature, they may also be used *in addition to approved nomenclature*, but they must be defined clearly and unambiguously (e.g., *F5* p.Arg534Gln and factor V Leiden).

- Standard HGVS nomenclature using g. annotation and identifying the genome build must be used for non-coding variants, including those variants identified in genome-wide association studies (GWAS) (e.g., NC_000017.11:g.50201450C>T). Variants may also be described using dbSNP genomic location identifiers, *in addition to approved nomenclature,* if the specific nucleotide change is also included.

To assess the burden that these requirements are likely to place~~d~~ on authors and editors, staff at two journals (*Human Mutation* and *Genetics in Medicine*) performed a pilot project. We documented compliance with notation requirements and recurrent issues over a period of 4 months. Concurrently, staff who developed and implemented validation tools (Mutalyzer and VariantValidator) agreed to provide assistance to authors and to document unresolved issues.

Among 425 initial submissions to *Human Mutation* over a 6-month period from April 2019, 349 had variant data and over half (208) included the validation file. During the same time frame, 187 manuscripts under revision reported variants, of which 162 included a validation file, while the remaining 25 indicated that alternative validation software had been used or that variants had been previously reported. *Genetics in Medicine* only checked revised manuscripts for variant data. Over a 3-month period, from April 2019, *GIM* received 193 revised manuscripts. Thirty-five of these contained variant data and all included a variant file. Four had to be sent back to the authors ~~to~~ for errors to be correct~~ed errors~~, but all revised manuscripts complied with the requirement to provide a file confirming that validation had been performed before the article was processed as a revision.

Several issues surfaced during the pilot project. Prime among these was the question of whether journal editors or authors should be responsible for correcting nomenclature errors. Two factors contributed to the committee's response to this issue: 1) authors have primary responsibility for the accuracy of data presented in their publications and 2) editorial staff do not have sufficient expertise or work force capacity to perform these corrections. Therefore, the committee decided that authors should have final responsibility for correcting the nomenclature of any variants not conforming to HGVS guidelines. In addition, because a substantial fraction of published variants ha~~ve~~s documented nomenclature errors (e.g. (Deans, Fairley, den Dunnen, & Clark, 2016)), the committee decided that all variants presented in a manuscript, whether new, or previously reported, should be verified. This approach will progressively correct extant nomenclature errors. Mutalyzer and VariantValidator autocorrect variant descriptions and create warnings in their results file~~s~~, except on rare occasions where there are syntax errors that cannot be interpreted/resolved. The HGVS nomenclature committee can assist when autocorrection does not occur.

Upon reviewing results of the pilot project, the committee agreed that requiring authors to verify that variants complied with HGVS nomenclature is a reasonable step toward standardizing the worldwide inventory of human variation. Editorial policy can require compliance with the recommendations to describe sequence variants adequately before manuscripts are accepted and

published. Ideally, submission of variants and phenotypes to a public database, with ~~the intrinsic~~its implicit quality checks ~~of the database~~, prior to initial manuscript submission would streamline the ~~review of~~ manuscript~~s~~ review process (den Dunnen, 2019). Some databases (e.g., ClinVar, LOVD) allow an embargo period after submission to allow for publication time. Finally, the process should not place an undue burden on editorial staff, as it only requires documentary evidence that authors have used the validation tools or have submitted their variants to a database that has validated the descriptions.

Issues that remain for editorial staff to consider for their own journals include the extent to which authors are required to correct nomenclature errors, guidelines for describing variants that cannot be defined using current HGVS description criteria and manuscript types that may be exempt from annotation requirements. The committee recognized that a minor fraction of variants has ambiguous notation. Consequently, asking authors to provide documentation that errors cannot be resolved after two attempts may be a reasonable compromise. Likewise, rejecting a manuscript due to annotation error may depend on the lack of effort on the authors' part to correct remediable errors.

Table 1 ~~contains examples of~~lists the journals that currently have editorial policies requiring validation of variant nomenclature, illustrating the feasibility of broadly implementing such policies.

References:

Deans, Z. C., Fairley, J. A., den Dunnen, J. T., & Clark, C. (2016). HGVS Nomenclature in p~~P~~ractice: An e~~E~~xample from the United Kingdom National External Quality Assessment Scheme. *Human Mutation*, *37*(6), 576–578. https://doi.org/10.1002/humu.22978

den Dunnen, J. T. (2016). Sequence Variant Descriptions: HGVS Nomenclature and Mutalyzer. *Current Protocols in Human Genetics*, *90*(1), 7.13.1-7.13.19. https://doi.org/10.1002/cphg.2

den Dunnen, J. T. (2019). Efficient variant data preparation for Human Mutation manuscripts: Variants and phenotypes. *Human Mutation*, *40*(8), 1009. https://doi.org/10.1002/humu.23830

den Dunnen, J. T., Dalgleish, R., Maglott, D. R., Hart, R. K., Greenblatt, M. S., Mcgowan-Jordan, J., … Taschner, P. E. M. (2016). HGVS Recommendations for the d~~D~~escription of s~~S~~equence v~~V~~ariants: 2016 Update. *Human Mutation*, *37*(6), 564–569. https://doi.org/10.1002/humu.22981

Fokkema, I. F. A. C., Taschner, P. E. M., Schaafsma, G. C. P., Celli, J., Laros, J. F. J., & den Dunnen, J. T. (2011). LOVD v.2.0: The next generation in gene variant databases. *Human Mutation*, *32*(5), 557–563. https://doi.org/10.1002/humu.21438

Freeman, P. J., Hart, R. K., Gretton, L. J., Brookes, A. J., & Dalgleish, R. (2018). VariantValidator: Accurate validation, mapping, and formatting of sequence variation descriptions. *Human Mutation*, *39*(1), 61–68. https://doi.org/10.1002/humu.23348

Landrum, M. J., Chitipiralla, S., Brown, G. R., Chen, C., Gu, B., Hart, J., … Kattman, B. L. (2020). ClinVar: Improvements to accessing data. *Nucleic Acids Research*, *48*(D1), D835–D844. https://doi.org/10.1093/nar/gkz972

Tack, V., Deans, Z. C., Wolstenholme, N., Patton, S., & Dequeker, E. M. C. (2016). What's in a Name? A c~~C~~oordinated a~~A~~pproach toward the c~~C~~orrect u~~U~~se of a u~~U~~niform n~~N~~omenclature to i~~I~~mprove p~~P~~atient r~~R~~eports and d~~D~~atabases. *Human Mutation*, *37*(6), 570–575. https://doi.org/10.1002/humu.22975

Yen, J. L., Garcia, S., Montana, A., Harris, J., Chervitz, S., Morra, M., … Church, D. M. (2017). A variant by any name: Quantifying annotation discordance across tools and clinical databases.

Table 1: ~~List of examples j~~Journals that have editorial policies requiring validation of variant nomenclature

Human Mutation	https://onlinelibrary.wiley.com/page/journal/10981004/homepage/forauthors.html
Genetics in Medicine	https://www.nature.com/gim/authors-and-referees
European Journal of Human Genetics	https://www.nature.com/ejhg/authors-and-referees/policies
American Journal of Human Genetics	https://www.cell.com/ajhg/authors
European Journal of Medical Genetics	https://www.elsevier.com/journals/european-journal-of-medical-genetics/1769-7212/guide-for-authors
Journal of Medical Genetics	https://jmg.bmj.com/pages/authors/
Human Genome Variation	https://www.nature.com/documents/hgv-gta.pdf
Cold Spring Harbor Molecular Case Studies	http://molecularcasestudies.cshlp.org/site/misc/ifora_details.xhtml
Neuromuscular Disorders	https://www.nmd-journal.com/content/authorinfo
The Journal of Immunology	https://www.jimmunol.org/info/authors
Journal of Microbiology, Immunology and Infection	https:// www.elsevier.com/journals/journal-of-microbiology-immunology-and-infection/1684-1182/guide-for-authors
Journal of the Chinese Medical Association	https://journals.lww.com/jcma/Pages/instructionstoauthors.aspx
Taiwan Journal of Ophthalmology	http://www.e-tjo.org/contributors.asp

List of URLS

- Human Variome Project https://www.humanvariomeproject.org/
- Human Genome Variation Society https://varnomen.hgvs.org/
- ClinVar https://www.ncbi.nlm.nih.gov/clinvar/
- LOVD https://www.lovd.nl/
- Mutalyzer https://mutalyzer.nl/
- HGVS nomenclature guidelines http://varnomen.hgvs.org/bg-material/refseq/
- *Human Mutation* https://onlinelibrary.wiley.com/journal/10981004
- *Genetics in Medicine* https://www.nature.com/gim/
- VariantValidatorhttps://variantvalidator.org/
- *Human Mutation*
  https://onlinelibrary.wiley.com/page/journal/10981004/homepage/forauthors.html
- *Genetics in Medicine* https://www.nature.com/gim/authors-and-referees
- *European Journal of Human Genetics* https://www.nature.com/ejhg/authors-and-referees/policies
- *American Journal of Human Genetics*   https://www.cell.com/ajhg/authors
- *European Journal of Medical Genetics*   https://www.elsevier.com/journals/european-journal-of-medical-genetics/1769-7212/guide-for-authors
- *Journal of Medical Genetics* https://jmg.bmj.com/pages/authors/
- *Human Genome Variation*   https://www.nature.com/documents/hgv-gta.pdf
- *Cold Spring Harbor Molecular Case Studies*
  http://molecularcasestudies.cshlp.org/site/misc/ifora_details.xhtml
- *Neuromuscular Disorders*   https://www.nmd-journal.com/content/authorinfo
- *The Journal of Immunology*   https://www.jimmunol.org/info/authors
- *Journal of Microbiology, Immunology and Infection*   https://www.elsevier.com/journals/journal-of-microbiology-immunology-and-infection/1684-1182/guide-for-authors
- *Journal of the Chinese Medical Association*
  https://journals.lww.com/jcma/Pages/instructionstoauthors.aspx
- *Taiwan Journal of Ophthalmology*   http://www.e-tjo.org/contributors.asp