HGVS Recommendations for the Description of Sequence Variants: 2016 Update

Johan T. den Dunnen,[1]* Raymond Dalgleish,[2] Donna R. Maglott,[3] Reece K. Hart,[4] Marc S. Greenblatt,[5] Jean McGowan-Jordan,[6] Anne-Francoise Roux,[7] Timothy Smith,[8] Stylianos E. Antonarakis,[9] and Peter E.M. Taschner[10] on behalf of the Human Genome Variation Society (HGVS), the Human Variome Project (HVP), and the Human Genome Organisation (HUGO)

[1]Human Genetics & Clinical Genetics, Leiden University Medical Center, Leiden, Nederland; [2]Department of Genetics, University of Leicester, Leicester, United Kingdom; [3]National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, Maryland; [4]Invitae, Inc., San Francisco, California; [5]University of Vermont College of Medicine, Burlington, Vermont; [6]Children's Hospital of Eastern Ontario and University of Ottawa, Ottawa, Ontario, Canada; [7]Laboratoire de Génétique Moléculaire, CHRU Montpellier, Montpellier, France; [8]Human Variome Project International Coordinating Office, Melbourne, Australia; [9]Department of Genetic Medicine, University of Geneva Medical School, Geneva, Switzerland; [10]Generade Centre of Expertise Genomics and University of Applied Sciences Leiden, Leiden, The Netherlands

ABSTRACT: The consistent and unambiguous description of sequence variants is essential to report and exchange information on the analysis of a genome. In particular, DNA diagnostics critically depends on accurate and standardized description and sharing of the variants detected. The sequence variant nomenclature system proposed in 2000 by the Human Genome Variation Society has been widely adopted and has developed into an internationally accepted standard. The recommendations are currently commissioned through a Sequence Variant Description Working Group (SVD-WG) operating under the auspices of three international organizations: the Human Genome Variation Society (HGVS), the Human Variome Project (HVP), and the Human Genome Organization (HUGO). Requests for modifications and extensions go through the SVD-WG following a standard procedure including a community consultation step. Version numbers are assigned to the nomenclature system to allow users to specify the version used in their variant descriptions. Here, we present the current recommendations, HGVS version 15.11, and briefly summarize the changes that were made since the 2000 publication. Most focus has been on removing inconsistencies and tightening definitions allowing automatic data processing. An extensive version of the recommendations is available online, at http://www.HGVS.org/varnomen.
Hum Mutat 37:564–569, 2016. © 2016 Wiley Periodicals, Inc.

KEY WORDS: standards; sequence variation; mutation; database; nomenclature

Introduction

Sequence variant nomenclature needs to be accurate, unambiguous, and stable, but sufficiently flexible to allow description of all known classes of sequence variation. After the publication of some initial guidelines [Ad Hoc Committee on Mutation Nomenclature, 1996; Antonarakis, 1998], the HGVS proposed a more comprehensive set of recommendations [den Dunnen and Antonarakis, 2000], now known as the HGVS recommendations/nomenclature (http://www.HGVS.org/varnomen). These guidelines have gradually acquired world-wide acceptance and are currently acknowledged as the standard nomenclature in molecular diagnostics [Gulley et al., 2007; Richards et al., 2015]. As the recommendations were used, it was recognized that the initial proposal had a few errors, contained some inconsistencies, and did not cover all types of sequence variants (e.g., complex changes). This paper will summarize the current recommendations, the result of the evolution of the original recommendations [den Dunnen and Antonarakis, 2000] applied in practice: HGVS recommendations version 15.11.

The HGVS Recommendations

The HGVS recommendations are designed to be stable, meaningful, memorable, and unequivocal. Still, modifications may be necessary to remove inconsistencies, clarify confusing conventions, and/or to extend the recommendation to represent cases that were not previously covered. To allow users to specify up to what time-point they have followed the HGVS recommendations, a version number is now included. Any change in the recommendations will result in a new dated version number and all changes introduced will be specified in the version list. The recommendations described here represent the HGVS recommendations for the description of sequence variants version 15.11 (i.e., November 2015).

SVD-WG

To support overall acceptance of the HGVS recommendations, three organizations (HGVS, HVP, and HUGO) joined forces to establish the Sequence Variant Description Working Group (SVD-WG). The SVD-WG operates following a standard procedure (for details see http://www.variome.org/svd-wg.html), discusses incoming requests to modify or extend the recommendations and, where necessary, makes a proposal for changes. When finalized, the proposal is published on the nomenclature Website and, for a 2-month

*Correspondence to: Johan T. den Dunnen, Human Genetics (S04-030), Leiden University Medical Center, P.O. Box 9600, Leiden 2300RC, The Netherlands. E-mail: varnomen@HGVS.org

Table 1. Nomenclature Definitions with Example Variant Descriptions

Substitution (>)	g.1318G>T	A change where one nucleotide is replaced by one other nucleotide
Deletion| (del)	g.3661_3706del	A change where one or more nucleotides are not present (deleted)
Inversion (inv)	g.495_499inv	A change where more than one nucleotide replaces the original sequence and is the reverse-complement of the original sequence (e.g., CTCGA to TCGAG)
Duplication (dup)	g.3661_3706dup	A change where a copy of one or more nucleotides are inserted directly 3' of the original copy of that sequence
Insertion (ins)	g.7339_7340insTAGG	A change where one or more nucleotides are inserted in a sequence and where the insertion is not a copy of a sequence immediately 5'
Conversion (con)	g.333_590con1844_2101	A specific type of deletion-insertion where a range of nucleotides replacing the original sequence are a copy of a sequence from another site in the genome
Deletion-insertion (delins/indel)	g.112_117delinsTG	A change where one or more nucleotides are replaced by one or more other nucleotides and which is not a substitution, inversion, or conversion

Read "*a change where*" as "*a change where in a specific sequence compared to the reference sequence . . .* "

period, opened for "*Community Consultation.*" People interested in the work of the SVD-WG can register for a mailing list, ensuring that they will receive a copy of all proposals and decisions made. After 2 months, all comments are collected and evaluated by the SVD-WG. When no major concerns are received the proposal is accepted, published on the nomenclature Website and a new version number is assigned to the recommendations. Proposals receiving many comments are either modified or considered rejected. The entire procedure described was completed for the first time in October 2015 and resulted in the acceptance of two new proposals (see *Community Consultation* below).

Terminology

In contrast to the original recommendations, the terms "polymorphism" and "mutation" are no longer used because both terms have assumed imprecise meanings in colloquial use. Polymorphism is confusing because in some disciplines it refers to a sequence variation that is not disease causing, whereas in other disciplines it refers to a variant found at a frequency of 1% or higher in a population. Similarly, mutation is confusing since it is used both to indicate a "change" and a "disease-causing change." In addition, "mutation" has developed a negative connotation [Condit et al., 2002; Cotton, 2002], whereas the term "variant" has a positive value in discussions between medical doctors and patients by dedramatizing the implication of the many, often largely uncharacterized, changes detected. Therefore, following recommendations of the Human Genome Variation Society (HGVS) and American College of Medical Genetics (ACMG) [Richards et al., 2015], we only use neutral terms such as "variant", "alteration," and "change."

Definitions

To enhance clarity as well as to facilitate computational analysis and description of sequence variants, the basic types of variants had to be defined more strictly. In addition, descriptions have been prioritized, meaning that when a description is possible according to several classes, for example, as a duplication or an insertion, one specific class is preferred. The priority assigned is (1) deletion, (2) inversion, (3) duplication, (4) conversion, and (5) insertion. These changes made it possible to generate a formalized description of the HGVS standard in Extended Backus-Naur Form [Laros et al., 2011] and to develop software tools that can check and/or generate HGVS descriptions [Wildeman et al., 2008; Hart et al., 2014].

The definitions are shown in Table 1. The consequences of these definitions are, for example, that an A>T substitution should not be described as an inversion. Similarly, a change where one nucleotide

Table 2. Reference Sequences

Numbering scheme	Prefix	Position numbering in relation to
Genomic DNA[a]	g.	First nucleotide of the genomic reference sequence
Coding DNA[a]	c.	First nucleotide of the translation start codon of the coding DNA reference sequence
Noncoding DNA[a,b]	n.	First nucleotide of the noncoding DNA reference sequence
Mitochondrial DNA	m.	First nucleotide of the mitochondrial DNA reference sequence
RNA	r.	First nucleotide of the translation start codon of the RNA reference sequence or first nucleotide of the noncoding RNA reference sequence
Protein	p.	First amino acid of the protein sequence

[a]For diagnostic applications, it is strongly recommended to use Locus Reference Genomic sequence (LRG) [Dalgleish et al., 2010; MacArthur et al., 2014]. When no LRG is available, one should be requested; "pending" LRGs should not be used. If there is no LRG, a RefSeq sequence [O'Leary et al., 2016] is recommended.
[b]Reference sequence recently added after community consultation by the HGVS/HVP/HUGO sequence variant description working group (SVD-WG, http://www.hgvs.org/mutnomen/accepted002.html).

is replaced by more than one other nucleotide is not a substitution but a deletion-insertion (delins/indel). A duplication is defined as a tandem copy of an upstream sequence. The position of a duplication is represented by defining the position of the nucleotide, or the range of nucleotides, that is duplicated (see *Variability in Repeated Sequences* below). When the "duplicated" sequence is not directly 3' to the original copy, it should be described as an insertion. Insertions are in general short and de novo, that is, not a copy of an existing sequence from elsewhere in the genome. For larger duplicating insertions, that is, having a copy elsewhere in the genome, one should define the original source sequence with respect to a reference sequence and a nucleotide range.

Reference Sequences

Discussions regarding the preferred use of reference sequences (Table 2) remain lively. The most fundamental recommendation is that one has to describe what was observed, not what has been inferred, that is, report original observations using an appropriate reference sequence. When genomic DNA is sequenced, a genomic reference sequence should be the preferred choice. However, especially in diagnostics, reporting based on a coding DNA reference sequence is far more popular. The reason is simple, from the description one immediately gets some information regarding the location of the variant, namely, exonic or intronic, 5' of the ATG or 3' of the stop codon, and, by dividing the nucleotide number by 3, the number of the amino acid residue that is affected (see Fig. 1; Tables 3–6). For
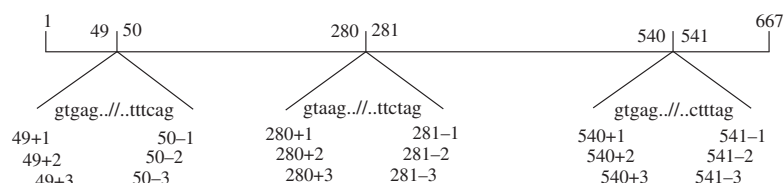
Coding DNA reference
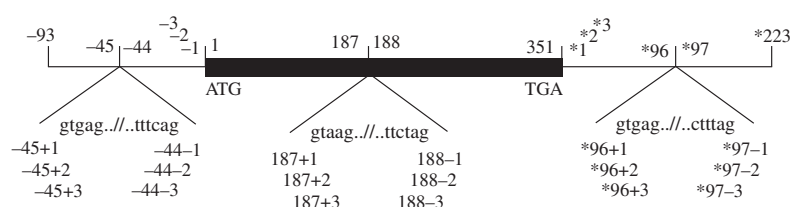
Noncoding DNA reference

Figure 1. Nucleotide numbering for a coding (top) and noncoding (bottom) DNA reference sequence; black box = the protein coding sequence. In the coding DNA reference sequence, nucleotide numbering starts with c.1 at the A of the ATG translation initiation codon. Numbering proceeds until the last nucleotide of the translation termination codon (TG**A**, TA**A**, or TA**G**). Nucleotides 5′ of the ATG are numbered c.-1, c.-2, and so on, nucleotides 3′ of the stop codon c.*1, c.*2, and so on. Intronic nucleotides are numbered based on the closest flanking exon nucleotide, on the 5′ side going into the intron such as c.187+1, c.187+2, and so on, on the 3′ side going in to the intron such as c.188-1, c.188-2, and so on. When introns have an uneven number of nucleotides, the central nucleotide (N) is linked to the upstream exon, like c.187+N. Nucleotide numbering for a noncoding DNA reference sequence starts with nucleotide c. and ends at the end of the reference sequence. Intronic nucleotides are numbered as for a coding DNA reference sequence.

Table 3. Nucleotide Numbering

| | Genomic reference sequence | | Coding DNA reference sequence LRG_199t1, NM_004006.2 |
Location nucleotide	NC_000023.10	LRG_199 (DMD)	
5′ transcription start	g.33231774	g.130953	c.-2345
In 5′ UTR	g.33229552	g.133175	c.-123
(In intron in 5′ UTR)[a]			(c.-55+23 / c.-54-23)[b]
A of the ATG start codon	g.33229429	g.133298	c.1
In coding DNA	g.32862930	g.499797	c.234
In intron, 5′ side	g.32380903	g.981883	c.5325+2[a]
In intron, 3′ side	g.32366647	g.996080	c.5326-2[a]
G of TAG stop codon	g.31140036	g.2222791	c.11058
In 3′ UTR	g.31139691	g.2222836	c.*345
(In intron in 3′ UTR)[a]			(c..*54+23 / c.*55-23)[b]
3′ transcription end	g.31136580	g.2226247	c.*3456

Nucleotide numbering using a genomic reference sequence (NC_000023.10 [genome build GRCh37/hg19], and LRG_199) and a coding DNA reference sequence (DMD gene, LRG_199t1, or NM_004006.2). Nucleotide numbering starts at 1; there is no nucleotide 0.
[a]Coding DNA reference sequence NM_004006.2 does not contain intron sequences; LRG_199 is required for this description.
[b]Hypothetical example, the DMD gene does not contain an intron in the 5′ or 3′ UTR.

diagnostic applications, a new reference was recently introduced, the Locus Reference Genomic sequence (LRG) [Dalgleish et al., 2010; MacArthur et al. 2014]. The HGVS recommendations strongly advise the use of an LRG. When for a gene of interest no LRG is available, one should be requested as soon as possible. "Pending" LRGs should not be used (they might change before being approved). If there is no current LRG, the use of a RefSeq sequence [O'Leary et al., 2016], with its version (RefSeqGene or transcript) is recommended.

Existing Standards

To become compliant with prior standards and conventions, a few changes to the original recommendations were required. The most prominent of these concerned termination variants at the protein level, where "X" had previously been used to describe a stop codon, to be changed to "Ter" or "*," for example, p.Trp123Ter or p.W123* (previously p.Trp123X, see Table 6). With this change, the recommendations now follow the IUPAC-IUB nomenclature and its use of symbols for amino acids and peptides published in 1984 [IUAPC-IUB Joint Commission on Biochemical Nomenclature (JCBN), 1984] where "X" is defined to indicate "any amino acid." "Ter" (termination codon) is used in three-letter amino acid code, and "*" can be used in both one- and three-letter code.

It should be noted that HGVS recommendations follows the full IUPAC-IUBMB "Nomenclature for Incompletely Specified Bases in Nucleic Acid Sequences" (Tables 2 and 6). This facilitates, for example, the description at DNA level of the uncertainty that remains

Table 4. DNA Level Descriptions

Variant type	g. description	c. description[a]	Remarks
Substitution	g.32662262G>A	c.1318G>A	
Deletion	g.32466684_32466698del	c.3661_3706del	Specification of deleted nucleotides(s) optional
Duplication	g.32466684_32466698dup	c.3661_3706dup	Specification duplicated nucleotide(s) optional
Insertion	g.31792279_31792280insTAGG	c.7339_7340insTAGG	Specification of inserted nucleotides mandatory
Inversions	g.32481638_32481654inv	c.3334_3350inv	Minimum size: 2 nucleotides
Deletion-insertion (indel)	g.32867914_32867919delinsTG	c.112_117delinsTG	Specification of inserted nucleotides mandatory
Translocation			No recommendation yet[b]
Repetitive DNA stretch	g.31836932T[22]	c.7309+1160T[22]	Describe first nucleotide and repeat unit or range
	g.33170306TAA[9] or	c.31+59093TAA[9] or	of first repeat unit with number of repeat units
	g.33170306_33170308[9]	c.31+59093_31+59095[9]	between brackets
Two variants on one chromosome (in cis)	g.[32841486C>T;33038273G>C]	c.[76C>T; 283G>C]	
Two variants on two different chromosomes (in trans)		c.[76C>T];[283G>C]	
	g.[32841486C>T];[33038273G>C]		
Two variants, phase unknown (on one or two chromosomes)		c.[76C>T(;)283G>C]	
	g.[32841486C>T(;)33038273G>C]		

Variants are described in relation to hypothetical genomic and coding DNA reference sequences. A more extensive collection of examples is available from the HGVS nomenclature Website.
[a]For another location of the nucleotide relative to a coding DNA reference sequence, see Table 3.
[b]Subject of proposal SVD-WG004.

Table 5. RNA Level Descriptions

Variant type	r. description	Remarks
Substitution	r.1318g>u	
Deletion	r.3661_3706del	No specification of deleted nucleotide(s)
Duplication	r.3661_3706dup	No specification of duplicated nucleotide(s)
Insertion	r.7339_7340instagg	Mandatory specification of inserted nucleotide(s)
	r.456_457ins456+87_456+121	
Inversions	r.3334_3350inv	Minimum size 2 nucleotides
Deletion-insertion (indel)	r.112_117delinsug	Mandatory specification of inserted nucleotides
Two variants on 1 chromosome (in cis)	r.[76c>u;283g>c]	
Two variants on different chromosomes (in trans)	r.[76c>u];[283g>c]	
Two variants, phase unknown	r.[76c>u(;)283g>c]	
One DNA change yielding two transcripts	r.[76a>c,70_77del]	
Predictions	*r.spl r.?*	Affects splicing unknown consequences

Variants are described in relation to a hypothetical RNA reference sequence. Compared with DNA descriptions, lower case nucleotides are used and "u" instead of "t." A more extensive collection of examples is available from the HGVS nomenclature Website.

when publications list variants only at protein level, for example, p.(Ile321Leu) on DNA level as c.961A>Y (Y being C or T).

Modifications to Original Recommendations

The 2000 recommendations [Den Dunnen and Antonarakis, 2000] contained some minor inconsistencies that demanded modification. A leading feature, to reduce confusion, was to use specific symbols for one purpose only. In the 2000 recommendations, the "+" character was used both in nucleotide numbering (for intronic nucleotides) and to list combined variants or alleles. The recommendation is now to "+" only in nucleotide numbering and to use ";" to list combined variants or alleles: c.[76C>T; 283G>C] for two variants known to be on one molecule (in cis), c.[76C>T];[283G>C] for the same two variants known to be on two different molecules (in trans), and c.[76C>T(;)283G>C] for two variants where the phase is unknown (see Table 4). Similarly, "," was introduced to replace the "+" to describe several RNA transcripts deriving from one variant at DNA level. In the example from Den Dunnen and Antonarakis (2000), r.[76a>c,70_77del] now replaces r.[76a>c+70_77del] (see Table 5).

The alternative use of c.IVS# and c.EX# in the description of nucleotide positions has been retracted. Such descriptions are indirect and cause confusion when different exon/intron numbering schemes are used for a gene. They also compromise the development of software tools. Some tools, for example, the Mutalyzer Name Checker, accept the c.IVS format, supporting conversion to the recommended format [Wildeman et al., 2008]. Nucleotide positions should be specific and include numbers only (Fig. 1; Table 3).

Recommendations regarding the description of variants that have not been fully characterized, for example, deletion or duplication breakpoints, have been clarified. In such cases, the description should indicate the region of uncertainty, using the format (5′border_3′border). The suggestion to describe exon deletions and duplications using the format c.77-?_923+?del (or dup) was therefore retracted. Descriptions of genomic deletions therefore have formats such as g.2345_6789del (break point sequenced), g.(1234_3456)_(5678_7890)del (breakpoints defined but not sequenced), g.(?_3456)_6789del (5′ break point undefined, 3′break point sequenced).

Reporting one variant using different descriptions creates confusion and underrepresentation of its frequency. Preventing this is especially important in stretches of repeated DNA sequences. Although not stated prominently, the 2000 guidelines specified normalization (shuffling) to the 3′ end. This so-called 3′ rule states that for variants in stretches of repeated DNA sequences the most 3′ position possible is arbitrarily deemed to have been changed. Consequently, the change of TTT to TT is described as g.3del (not g.1del or g.2del). A corollary of the 3′ rule is that predicted duplications

Table 6. Protein Level Descriptions

Variant type	p. description	Remarks
	p.(Arg490Ser)	The protein change is predicted (no experimental proof)
Substitution	p.Arg490Ser/p.R490S p.Trp87Ter / p.Trp78∗/p.W87∗	Both three- (preferred) and one-letter amino acid code may be used; ∗ accepted for one- and three-letter code
Deletion	p.Asp388_Gln393del	No specification of deleted amino acid(s)
Duplication	p.Asp388_Gln393dup	No specification of duplicated amino acid(s)
Insertion	p.Ala228_Val229insTrpPro p.Ala228_Val229insLys∗	Mandatory specification of inserted amino acids
Inversions		Not possible
Deletion-insertion (indel)	p.L7_H8delinsWQQFRTG	Mandatory specification of inserted amino acids
Frame shift	p.(Arg97fs) p.(Arg97Profs∗23)	Short and long form accepted; long form contains "fsTer" or "fs∗"
Extension	p.Met1ValextMet-12 p.Ter110GlnextTer17	Short and long form accepted; long form contains "fsTer" or "fs∗"
Repetitive amino acid stretch	p.Gln34[22] p.Ser7_Ala9[6]	Describe first amino acid repeat unit
Two protein coding variants on one chromosome (in cis)	p.[Trp78∗;Arg490Ser]	
Two protein coding variants on two different chromosomes (in trans)	p.[Trp78∗];[Arg490Ser]	
Two protein coding variants, phase unknown	p.[Trp78∗(;)Arg490Ser]	
Predictions	p.?	Unknown consequences

Variants are described in relation to a hypothetical protein reference sequence. A more extensive collection of examples is available from the HGVS nomenclature Website.

and deletions of amino acids are similarly normalized to the most C-terminal position.

The description of variability in repeated sequences has been slightly modified and specified more precisely. Such changes are described by defining the first nucleotide of the repeat unit, or the range of the first repeat unit, with the number of repeat units specified between square brackets, for example, g.123_124[4]. For short/simple repeats, it is allowable to include the content of the repeated unit, using the format "position-first-nucleotide-repeat-content[number]" such as g.123TG[4]. When the size of the repeat unit is uncertain, this should be specified using parentheses; g.-128GGC[(600_800)].

For descriptions at the protein level, it is strongly recommended to use the three-letter amino acid code. The three-letter code retains better compatibility with IUPAC and leads to fewer errors when used, especially for those amino acids where the first letter differs from their one-letter code (e.g., aspartic acid [D], asparagine [N], and arginine [R]). The use of the one-letter code should be restricted to the description of long sequences.

At the protein level, besides changing the description of a translational stop codon variant from X to Ter/∗, the description of frame shifts has been specified and additions have been made to describe variants affecting the translation start and stop codon. In addition, the recommendation has been made that descriptions at the protein level should make clear whether experimental proof was available or not. When not, one should list the predicted consequences in parentheses. Variants that are predicted to shift the translational reading frame should be described using either a short or a long form; p.(Arg97fs) and p.(Arg97Profs∗23), respectively. For "fsTer#"/"fs∗#," it is specified that "#" indicates at which codon number the new reading frame ends with a stop codon. The number of the stop in the new reading frame is calculated starting at the first amino acid that is changed by the frame shift, ending at the stop codon (∗#).

A newly added recommendation is to describe changes directly affecting the start or stop codon as an N- or C-terminal extension. Amino acids upstream of the original start site are numbered using a minus sign. For example, p.Met1ValextMet-12 describes the observed N-terminal extension of 12 amino acids (Met-12 to Thr-1) of the protein as the consequence of a variant (DNA c.1A>G) that changes amino acid Met1 to Val. Similarly, p.Ter110GlnextTer17 describes the observed extension of the C-terminus of the protein with 17 new amino acids as a consequence of a variant (DNA c.331T>C) that changes Ter110 to Gln.

Community Consultation

The Sequence Variant Description Working Group (SVD-WG) currently commissioning the variant description recommendations under the auspices of HGVS, HVP, and HUGO recently completed the first round of proposals. Proposal SVD-WG001 was accepted allowing the description g.123G= to indicate that a variant screen was performed but no change detected. Similarly, when a variant g.456G>A has no predicted consequences at the protein level, this can be described as p.(Arg152=). Proposal SVD-WG002 was accepted allowing the use of a noncoding DNA reference sequence using the prefix "n." (n.963G>C; Table 2). Proposal SVD-WG003 to further specify the description of exon deletions/duplications detected using MLPA is still under discussion. Discussions are ongoing to achieve a harmonization of the nomenclature systems of the HGVS and the International Standing Committee on Human Cytogenetic Nomenclature [ISCN, 2013], but these have not yet been finalized. The HGVS nomenclature as presented in this paper is version 15.11.

Dissemination

The HGVS nomenclature pages list an email address for questions (VarNomen @ HGVS.org). The chair of the SVD-WG collects all requests for information and/or clarification of the recommendations. In most cases, questions can be answered easily; in rare cases, the SVD-WG is consulted first before an answer is sent. To promote the recommendations, a Facebook page (http://www.facebook.com/HGVSmutnomen) has been started where, on a regular basis, topics of interest are discussed. These include simplified Q&As, meetings where the recommendations are discussed and the release of new proposals for Community consultation. Current action points for the SVD-WG are the development of educational material and a restructuring of the HGVS nomenclature pages.

Although the guidelines have gradually acquired world-wide acceptance, there is still room for improvement. Initiated by *Human Mutation*, the first journal to demand their use, strong support is coming from many journals making the use of HGVS nomenclature for sequence variant descriptions mandatory. All major variant databases support HGVS nomenclature and professional

organizations have started to demand its use in clinical diagnostic reporting [Gulley et al., 2007; Richards et al., 2015]. However, in specific disciplines, HGVS nomenclature is used less frequently and alternative nomenclature systems survive [e.g., Berwouts et al., 2011; Kalman et al., 2016]. EQA providers have detected the problem [Tack et al., 2016] and started to more actively promote HGVS nomenclature by asking participants to follow the recommendations and subtracting marks when laboratories fail to do so correctly. In this respect, it should be noted that excellent open source support tools have been developed that make it very easy to check whether correct HGVS descriptions are reported [Wildeman et al., 2008; Hart et al., 2014]. The latest version of the Mutalyzer suite includes a Variant Description Extractor that, based on a reference and sample sequence, will report all changes present in HGVS nomenclature (https://www.mutalyzer.nl/description-extractor). Finally, the HGVS nomenclature Website is currently being completely reconstructed to increase usability and a start is being made to add educational and training material.

Acknowledgments

References

Ad Hoc Committee on Mutation Nomenclature. 1996. Update on nomenclature for human gene mutations. Hum Mutat 8:197–202.

Antonarakis SE. 1998. Recommendations for a nomenclature system for human gene mutations. Hum Mutat 11:1–3.

Berwouts S, Morris MA, Girodon E, Schwarz M, Stuhrmann M, Dequeker E. 2011. Mutation nomenclature in practice: findings and recommendations from the cystic fibrosis external quality assessment scheme. Hum Mutat 32:1197–1203.

Condit CM, Achter PJ, Lauer I, Sefcovic E. 2002. The changing meanings of "mutation:" a contextualized study of public discourse. Hum Mutat 19:69–75.

Cotton RGH. 2002. Communicating "mutation:" modern meanings and connotations. Hum Mutat 19:2–3.

Dalgleish R, Flicek P, Cunningham F, Astashyn A, Tully RE, Proctor G, Chen Y, McLaren WM, Larsson P, Vaughan BW, Béroud C, Dobson G, et al. 2010. Locus Reference Genomic sequences: an improved basis for describing human DNA variants. Genome Med 2:24.1–24.7.

den Dunnen JT, Antonarakis SE. 2000. Mutation nomenclature extensions and suggestions to describe complex mutations: a discussion. Hum Mutat 15:7–12.

Gulley ML, Braziel RM, Halling KC, Hsi ED, Kant JA, Nikiforova MN, Nowak JA, Ogino S, Oliveira A, Polesky HF, Silverman L, Tubbs RR, et al. 2007. Clinical laboratory reports in molecular pathology. Arch Pathol Lab Med 131:852–863.

Hart RK, Rico R, Hare E, Garcia J, Westbrook J, Fusaro VA. 2014. A Python package for parsing, validating, mapping and formatting sequence variants using HGVS nomenclature. Bioinformatics 31:268–270.

IUPAC-IUB Joint Commission on Biochemical Nomenclature (JCBN). 1984. Nomenclature and symbolism for amino acids and peptides. Recommendations 1983. Eur J Biochem 138:9–37.

Kalman LV, Agúndez J, Appell ML, Black JL, Bell GC, Boukouvala S, Bruckner C, Bruford E, Caudle K, Coulthard SA, Daly AK, Tredici AD, et al. 2016. Pharmacogenetic allele nomenclature: International workgroup recommendations for test result reporting. Clin Pharmacol Ther 99:172–185.

Laros JF, Blavier A, den Dunnen JT, Taschner PEM. 2011. A formalized description of the standard human variant nomenclature in Extended Backus-Naur Form. BMC Bioinformatics 12 Suppl 4:S5.

MacArthur JAL, Morales J, Tully RE, Astashyn A, Gil L, Bruford EA, Larsson P, Flicek P, Dalgleish R, Maglott DR, Cunningham F. 2014. Locus Reference Genomic: reference sequences for the reporting of clinically relevant sequence variants. Nucleic Acids Res 42:D873–D878.

O'Leary NA, Wright MW, Brister JR, Ciufo S, Haddad D, McVeigh R, Rajput B, Robbertse B, Smith-White B, Ako-Adjei D, Astashyn A, Badretdin A, et al. 2016. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. Nucleic Acids Res 44:D733–D745.

Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, Grody WW, Hegde M, Lyon E, Spector E, Voelkerding K, Rehm HL, ACMG Laboratory Quality Assurance Committee. 2015. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet Med 17:405–424.

ISCN 2013: *An International System for Human Cytogenetic Nomenclature*, Shaffer LG, McGowan-Jordan J, Schmid M (eds); S. Karger, Basel 2013.

Tack V, Deans ZC, Wolstenholme N, Patton S, Dequeker EMC. 2016. What's in a name? A coordinated approach towards the correct use of a uniform nomenclature to improve patient reports and databases. Hum Mutat 37:570–575.

Wildeman M, van Ophuizen E, den Dunnen JT, Taschner PEM. 2008. Improving sequence variant descriptions in mutation databases and literature using the Mutalyzer sequence variation nomenclature checker. Hum Mutat 29:6–13.