



Universiteit
Leiden

The Netherlands

Knowledge multiplies when shared — when calling things by their right name: improving the validation and exchange of genetic data in research and diagnostics

Fokkema, I.F.A.C.

Citation

Fokkema, I. F. A. C. (2025, December 9). *Knowledge multiplies when shared — when calling things by their right name: improving the validation and exchange of genetic data in research and diagnostics*. Retrieved from <https://hdl.handle.net/1887/4285050>

Version: Publisher's Version

License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/4285050>

Note: To cite this publication please use the final published version (if applicable).

HGVS Nomenclature 2024: improvements to community engagement, usability, and computability

Reece K Hart¹, **Ivo F.A.C. Fokkema**², Marina DiStefano³, Ros J. Hastings^{4,5},
Jeroen F.J. Laros^{2,6}, Rachel L. Taylor^{7,8}, Alex Wagner^{9,10}, Johan T. den Dunnen^{2,11}

1 - MyOme, Inc., San Francisco, USA.

2 - Department of Human Genetics, Leiden University Medical Center, Leiden, The Netherlands.

3 - Medical and Population Genetics Group, The Broad Institute of MIT and Harvard, Cambridge, USA.

4 - GenQA, Oxford University Hospitals NHS Trust, Oxford, UK.

5 - GenQA, NHS Lothian, Edinburgh, UK.

6 - Department of Bioinformatics and Computational Services, National Institute for Public Health and the Environment, Bilthoven, The Netherlands.

7 - EMQN CIC, Manchester, UK.

8 - Manchester University NHS Foundation Trust, Manchester, UK.

9 - Institute for Genomic Medicine, Nation-wide Children's Hospital, Columbus, USA.

10 - Department of Pediatrics, the Ohio State University College of Medicine, Columbus, USA.

11 - Department of Clinical Genetics, Leiden University Medical Center, Leiden, The Netherlands.

Published in: *Genome Medicine* 2024; 16:149.

DOI: 10.1186/s13073-024-01421-5.

5.1 Abstract

Background The Human Genome Variation Society (HGVS) Nomenclature is the global standard for describing and communicating variants in DNA, RNA, and protein sequences in clinical and research genomics. This manuscript details recent updates to the HGVS Nomenclature, highlighting improvements in governance, community engagement, website functionality, and underlying implementation of the standard.

Methods The HGVS Variant Nomenclature Committee (HVNC) now operates under the Human Genome Organization (HUGO), facilitating broader community feedback and collaboration with related standards organizations. The website has been redesigned using modern documentation tools and practices. The specification was updated to include guidance for transcript selection and to align with recent cross-consortia recommendations for the representation of gene fusions. A formal computational grammar was introduced to improve the precision and consistency of variant descriptions.

Results Major improvements in HGVS Nomenclature v.21.1 include a redesigned website with enhanced navigation, search functionality, and mobile responsiveness; a new versioning policy aligned with software management practices; formal mechanisms for community feedback and change proposals; and adoption of Extended Backus-Naur Form (EBNF) for defining syntax. The specification now recommends MANE Select transcripts where appropriate and includes updated guidance for representing adjoined transcripts and gene fusions. All content is freely available under permissive licenses at hgvs-nomenclature.org.

Conclusions These advancements establish a more sustainable foundation for maintaining and evolving the HGVS Nomenclature while improving its accessibility and utility. The introduction of formal computational grammar marks a crucial step toward unambiguous variant descriptions that can be reliably processed by both humans and machines. Combined with enhanced community engagement mechanisms and improved guidance, these changes position the HGVS Nomenclature to better serve the evolving needs of clinical and research genomics while maintaining the stability that users require.

5.2 Background

Accurate interpretations of sequence variants in clinical genomic testing depend on variants being described, communicated, and compared using conventions that are understood uniformly by researchers and clinicians. The HGVS Nomenclature^{1,2} is the de facto international standard for fostering consistency in the presentation and description of sequence variation, including clinical reports, scientific publications, variant databases, and proficiency testing/external quality assessment (PT/EQA). Although originally intended for the description of variants in humans, the HGVS Nomenclature is widely used to describe variants in other species.

Standards for representing sequence variants are crucial for ensuring consistency, accuracy, and interoperability in sequence variant data. They enable researchers, clinicians, and laboratories to communicate findings effectively, minimizing misunderstandings and errors. By adhering to established standards, sequence variants can be uniformly interpreted, compared, and integrated across various platforms and studies. This uniformity is essential for advancing research, improving diagnostic accuracy (including the efficient identification of patients eligible for therapeutic intervention), and facilitating the development of new therapies. Standards also support regulatory compliance and PT/EQA, further underscoring their critical role in the field of genetics and genomics. Additionally, in the context of software development, standards provide a clear framework for building tools and applications that can process, analyze, and share variation data seamlessly. They ensure that software developed by different groups can work together, enhancing collaboration and innovation within the scientific community.

There are several standards for representing variation within biological sequences (Table 5.1). While it may initially seem redundant to have multiple standards, each implements a distinct set of trade-offs made in the context of the history of a rapidly growing field. As standards evolve to support the representation of new kinds of information, they must do so in a way that preserves the interpretability of existing data and with an awareness of adjacent standards. The International System for Cytogenomic Nomenclature (ISCN) guidelines were developed in 1971 to standardize the representation of normal and abnormal cytogenomic results as inferred from karyotyping;³ ISCN later evolved to handle new technologies such as fluorescent in-situ hybridization (FISH), chromosomal microarrays, region-specific assays, optical genome mapping (OGM) and, in collaboration with HGVS, sequencing. In 1996, the HGVS created recommendations to represent variants identified by sequencing,^{1,2,4,5} which was still relatively low throughput and manual. Analysis requirements for next-generation sequencing (NGS) data led to the development of the Variant Call Format (VCF) in 2008,⁶ which supports only genomic sequences. In 2015, the Variation Modeling Consortium

Table 5.1: **Comparison of sequence variant representation and presentation standards.** *Coordinate systems* describes the distinct ways in which position may be defined on a reference sequence. *Normalization* describes the convention by which deletion/insertion variants are normalized (also referred to as aligned, shuffled, or shifted) in the context of sequence repeats. *Variant primitive operations* describes the distinct primitive operations that are supported by the standard; “deletion/insertion” in this column refers to the replacement of a span of sequence with another sequence, either of which might be zero-length. *Primary uses, strengths, and limitations* provides a cursory and incomplete summary of typical usage; nuances are omitted for brevity. The phrase “deletion/insertion” is used to describe deletion, insertion, and combined deletion and insertion variant types.

Standard and supporting organization	Sequence types, coordinate systems, and normalization	Variant primitive operations	Primary uses, strengths, limitations
Human Genome Variation Society (HGVS) Nomenclature Human Genome Organization (HUGO)	<ul style="list-style-type: none"> DNA, RNA, Protein Contiguous linear, Discontiguous (exonic/intronic), Circular Right-justified normalization 	<ul style="list-style-type: none"> Substitution Deletion/insertion Adjoined transcripts Duplication Extension Frameshift Inversion Repeat 	Predominant standard in representing variation in databases, publications, and clinical reports. Supports most sequence and variant types, as well as phased alleles and HGVS/ISCN. Optionality desired by user groups necessarily creates some ambiguity.
International System for Cytogenomic Nomenclature (ISCN)	<ul style="list-style-type: none"> DNA Chromosomal bands Contiguous linear normalization 	<ul style="list-style-type: none"> Aneuploidy Derived chromosomes Amplification Deletion Insertion Duplications Gene fusions Inversions Translocations ISCN/HGVS 	Long-standing standard for reporting genomic variation from karyotypes, FISH, microarray, genome mapping, and region-specific assays in constitutional and neoplasia cytogenomics. Represents variants not easily represented by other standards.
Sequence Position Deletion Insertion (SPDI) National Center for Biotechnology Information (NCBI)	<ul style="list-style-type: none"> DNA, RNA, Protein Contiguous linear Fully justified normalization 	<ul style="list-style-type: none"> Deletion/insertion (as a generalization of substitutions, duplications, inversions, etc.) 	Primarily used within NCBI for the precise and unambiguous representation of variants that can be represented as a combination of deletion and insertion events. The fully justified normalization using the SPDI algorithm expresses position ambiguity of variants in repeat regions better than left- or right-justified/normalized/shuffled variants.
Variant Call Format (VCF) Global Alliance for Genomics and Health (GA4GH)	<ul style="list-style-type: none"> DNA Contiguous linear Left-justified normalization, with anchor 	<ul style="list-style-type: none"> Deletion/insertion Deletion (DEL) Breakend (BND) Copy number (CNV) Insertion (INS) Inversion (INV) Translocation (TRA) 	The de facto standard file format for storing variants from NGS. Provides support for annotations, quality scores, and multiple samples within a single file. Supports only genomic sequences and file-based storage.
Variation Representation Specification (VRS) Global Alliance for Genomics and Health (GA4GH)	<ul style="list-style-type: none"> DNA, RNA, Protein Contiguous linear Fully justified normalization 	<ul style="list-style-type: none"> Deletion/insertion Adjoined sequences Copy number Genotype Haplotype Repeat 	Extensible standard for representing variants on all sequence types and sharing variants between systems and in complex data structures. Primarily focused on precise representation and the minimization of ambiguity, but also supports affirmative imprecision of some concepts. Semi-structured representation limits human readability.

(VMC) was formed to draft the Variation Representation Specification (VRS),⁷ with the goal of facilitating data sharing and communication among researchers, diagnostic labs, and clinical environments through Health Level Seven (HL7) and Fast Healthcare Interoperability Resources 1 (FHIR) standards; the VMC was later expanded into the Variation Representation group within the Global Alliance for Genomics and Health (GA4GH),⁸ which continues to develop the specification.⁹ In 2018, the National Center for Biotechnology Information (NCBI)¹⁰ developed the Sequence Position Deletion Insertion (SPDI) format and the Variant Overprecision Correction Algorithm (VOCA), which introduced normalization (also known as shuffling or shifting) to a fully justified representation as an alternative to right-normalized variants used by the HGVS Nomenclature or left-normalized variants used by VCF.¹¹ Taken together, SPDI and VOCA represent the position ambiguity of deletion-insertion variants in repeat regions more accurately than left or right normalization. VRS later adopted fully justified normalization.

This manuscript summarizes recent and major updates in the HGVS Nomenclature. These changes include new governance under the auspices of the HUGO, new community discussion and feedback mechanisms, and a new website that is easier to maintain and use. This release of the HGVS Nomenclature also initiates the adoption of a formal computational grammar for sequence variation and provides new guidance regarding the use of versioning to facilitate the compatibility of software. The intention of these changes is to enable the smooth and controlled evolution of the HGVS Nomenclature in order to better serve the needs of the scientific and clinical communities who rely on it.

5.3 Methods

5.3.1 Oversight and governance: incorporation into HUGO

The HGVS Variant Nomenclature Committee (HVNC),¹² a subcommittee of HUGO, maintains the HGVS Nomenclature and coordinates community feedback. The HVNC was formed from the Sequence Variant Description Working Group, a committee within the Human Variome Project (HVP), when HUGO, the HVP, and HGVS merged into one organization. “HGVS Nomenclature” was retained as the name of the standard because of its familiarity in the community and is now the preferred name for the project. Previous names — Mutnomen, VarNomen, and HGVS Recommendations — are discouraged.

The HVNC is currently composed of eleven representatives from academia, commercial, and national institutions with an interest in research and/or clinical genomics. Representatives serve 4-year terms and meet approximately six times per year. Recognizing the interconnectedness of standards efforts and the importance of consistency among them,

the HVNC coordinates with the Human Genome Nomenclature Committee (HGNC), the International System for Cytogenomic Nomenclature, the Global Alliance for Genomics and Health (GA4GH),⁸ and the Variant Interpretation for Cancer Consortium (VICC).¹³ Membership openings will be announced on the hgvs-announcements mailing list (discussed below) and the website.

5.3.2 User communities and community engagement

The HGVS Nomenclature and the HVNC serve a diverse audience of diagnostic genetic testing laboratories, clinicians, researchers, database developers, and publishers throughout the life sciences. The HVNC seeks representation from these groups on the committee and through several communications channels.

Previously, communication with the HVNC (and its predecessor, the Sequence Variant Description Working Group) required posting questions to a Facebook group or sending an email to a dedicated email address that was accessible only to the committee's chair. The Facebook group is now obsolete and the single-user email address has been replaced by two new email groups. The hgvs-nomenclature-announcements@googlegroups.com group is a low-volume mailing list for official announcements, such as new releases.¹⁴ The hgvs-nomenclature@googlegroups.com is intended for community discussion; in 2024, over one hundred new members have joined and approximately three questions have been asked and answered per week.¹⁵ Conversations for both groups are archived and publicly available without registration.

The HVNC has formalized the process by which the community may propose changes to the HGVS Nomenclature. As described on the new HGVS Nomenclature website, anyone may submit a proposal for changes to the standard. Proposals will be created in GitHub for public discussion and reviewed periodically by the HVNC. Proposals that are accepted will be drafted as changes to the HGVS Nomenclature text and refined with community input.

5.4 Results

5.4.1 Improvements to the management and presentation of the HGVS Nomenclature

5.4.1.1 Website implementation

The previous version of the HGVS Nomenclature website was written using a mix of Markdown¹⁶ and the HyperText Markup Language (HTML), which made it difficult to update and format consistently. Although changes were listed on a specific Versions page, the

previous site presented only one version of the HGVS Nomenclature, making it difficult to investigate the evolution of the recommendations. The new site is written nearly entirely using Markdown for consistency, rendered using MkDocs,¹⁷ and automatically deployed to Read the Docs.¹⁸

Many of the most significant new features of the HGVS Nomenclature website were created based on feedback from users. Examples include the collapsible navigation that makes it easy to understand the site layout, the version number shown in the top panel on all pages (Figure 5.1), and a new versions page that itemizes changes between versions (Figure 5.2). The new website enables users to easily switch between versions of the HGVS Nomenclature. Other major changes to the website include corrections of numerous small errors and inconsistencies, integrated searching with previews, the ability to navigate between different versions of the HGVS Nomenclature, and consistent styling that provides for distinguishing valid examples and invalid counterexamples (in red). The new website is responsive and remains functional across a wide range of devices, including phones, tablets, and laptops. An archived, prior version of the HGVS Nomenclature website (version 20.05) remains available.¹⁹

5.4.1.2 Change processes, versioning, and release processes

The processes for updating the HGVS Nomenclature now follow conventional source code management practices. The GitHub repository,²⁰ which stores the official documentation and code of the HVNC, is managed by a group of maintainers. Significant changes are discussed prior to implementation, and all changes are submitted as “pull requests” for review by maintainers prior to incorporation. As appropriate, the maintainers will release the aggregated changes with a new version tag and release notes. The site will be built and deployed automatically, typically within a few minutes.

The HVNC has developed a new versioning policy based on “semantic versioning”²¹ that is intended to help users and adopters manage compatibility between providers and consumers of tools. Each release of the HGVS Nomenclature will be tagged with a version that has the format X.Y.Z, often called the major, minor, and patch numbers, respectively. When backward incompatible changes are made, the major version (X) will be incremented; such changes are expected to be very rare. When new features or capabilities are added, the minor version (Y) will be incremented. For trivial changes, such as typographic fixes or clarifications, the patch version (Z) will be incremented. Because the patch level carries no change in the intent of the HGVS Nomenclature, two versions may be compared by the major and minor versions alone.

The HGVS Nomenclature now recommends that data providers — that is, tools or databases that present HGVS variant descriptions — should advertise the version of the HGVS Nomenclature.

a)

Home

21.0
2024-11-16

Search

Home

HGVS Nomenclature

Recommendations

General

DNA

RNA

Protein

Uncertain

HGVS Syntax Summary

Style Conventions

Grammar Elements

Checklist

Background

Basics

HGVS simple

Standards

Educational Material

History

Numbering

Reference Sequences

Glossary

Consultation

HGVS Variant Nomenclature Committee (HVNC)

Software

Versions

21.1 Series

21.0 Series

Home

21.0
2024-11-16

Search

The **HGVS Nomenclature** is an internationally-recognized standard for the description of DNA, RNA, and protein sequence variants. It is used to convey variants in clinical reports and to share variants in publications and databases.

The HGVS Nomenclature is administered by the **HGVS Variant Nomenclature Committee (HVNC)** under the auspices of the **Human Genome Organization (HUGO)**.

Get Connected with the Community

Sign up for Announcements

Engage with the Community

Contribute Ideas and proposals

Join the HGVS Nomenclature Announcements group and mailing list to receive official administrative announcements from the HVNC such as new releases and membership opportunities. This is a very low-volume list.

Join the HGVS Nomenclature group and mailing list. This is the preferred forum for discussing questions and issues regarding the HVNC. Messages sent to hgvs-nomenclature@googlegroups.com will be distributed to members by email. You do not need to be a member to send a message to the group. **However, you will not see a reply unless you join the group or manually check the HGVS Nomenclature conversation archive.**

When members of the community reach out to the HGVS nomenclature committee with questions via the google group email, we aim to respond within 2 business days of email receipt. If you haven't received a response within four business days, please post a reminder under your question in the google group. For a more expedited response in relation to a clinical inquiry, please email the committee directly. We cannot guarantee that even clinical inquiries will be answered within 2 business days or on weekends.

Read the Docs

stable 21.1.0 21.0.4 21.0.3 21.0.2 21.0.1 21.0.0 main

On Read the Docs

Project Home Builds

Search

Search docs

Citations

HGVS Nomenclature 2024: Improvements to community engagement, usability, and computability.

Hart RK, Fokkema IPAC, DiStefano M, Hastings R, Laros JFJ, Taylor R, Wagner A, den Dunnen JT

b)

Home

21.0
2024-11-16

Search

The **HGVS Nomenclature** is an internationally-recognized standard for the description of DNA, RNA, and protein sequence variants. It is used to convey variants in clinical reports and to share variants in publications and databases.

The HGVS Nomenclature is administered by the **HGVS Variant Nomenclature Committee (HVNC)** under the auspices of the **Human Genome Organization (HUGO)**.

Get Connected with the Community

Sign up for Announcements

Engage with the Community

Contribute Ideas and proposals

Join the HGVS Nomenclature Announcements group and mailing list to receive official administrative announcements from the HVNC such as new releases and membership opportunities. This is a very low-volume list.

Join the HGVS Nomenclature group and mailing list. This is the preferred forum for discussing questions and issues regarding the HVNC. Messages sent to hgvs-nomenclature@googlegroups.com will be distributed to members by email. You do not need to be a member to send a message to the group. **However, you will not see a reply unless you join the group or manually check the HGVS Nomenclature conversation archive.**

When members of the community reach out to the HGVS nomenclature committee with questions via the google group email, we aim to respond within 2 business days of email receipt. If you haven't received a response within four business days, please post a reminder under your question in the google group. For a more expedited response in relation to a clinical inquiry, please email the committee directly. We cannot guarantee that even clinical inquiries will be answered within 2 business days or on weekends.

Read the Docs

stable 21.1.0 21.0.4 21.0.3 21.0.2 21.0.1 21.0.0 main

On Read the Docs

Project Home Builds

Search

Search docs

Citations

HGVS Nomenclature 2024: Improvements to community engagement, usability, and computability.

Hart RK, Fokkema IPAC, DiStefano M, Hastings R, Laros JFJ, Taylor R, Wagner A, den Dunnen JT

Figure 5.1: **The landing page for hgvs-nomenclature.org.** a) The landing page for hgvs-nomenclature.org shows the “foldable” navigation menu on the left, a search bar with a clear indication of the selected version, and the Read the Docs version selector to navigate to other versions. The version selector is minimized and unobtrusive by default but expanded here to show functionality. b) Mobile version of the landing page. The navigation menu, search features, and selected version are hidden in the “hamburger menu” denoted by the ≡ symbol.

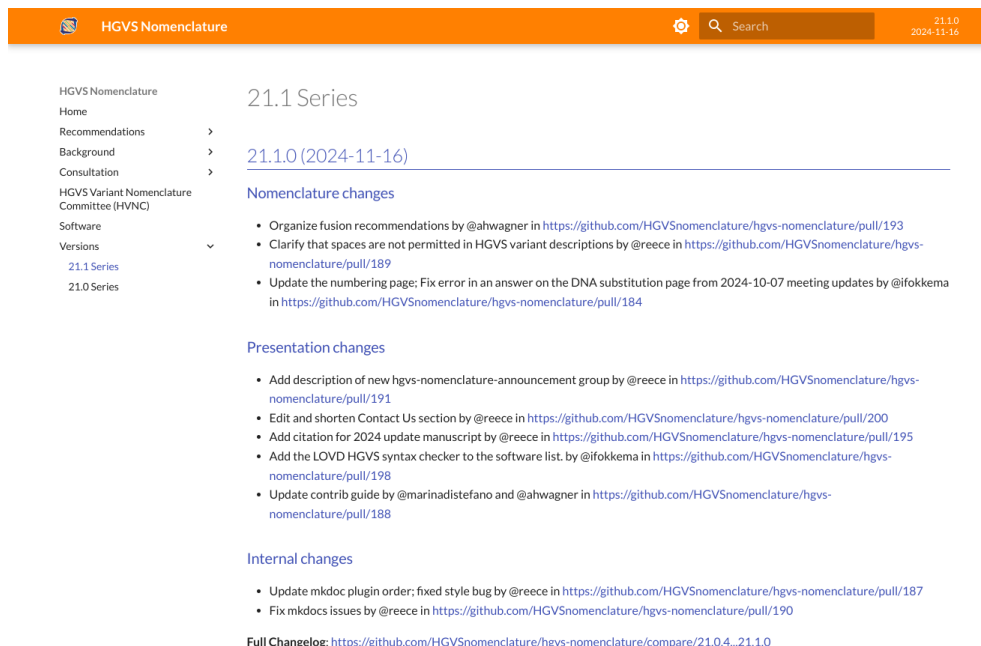


Figure 5.2: The release notes page provides details of every change to the website since the previous version.

ture that they use. Similarly, data consumers — that is, websites or software tools that accept variant descriptions as input — should advertise the versions of the HGVS Nomenclature that they accept. Data consumers should use the same major version as the provider and a minor version that is greater than or equal to that of the provider; other combinations risk backward incompatibility (when the consumer’s major version is less than the provider’s major version) or lack of a feature required to parse a variant (when the consumer’s minor version is less than the provider’s minor version). Details are given on the HGVS Nomenclature Versions page.²²

5.4.1.3 Presentation of HGVS Nomenclature grammar

Historically, the HGVS Nomenclature was presented as a set of guidelines and an informal grammar primarily aimed at human readers, providing recommendations for describing sequence variants. This approach was accessible to readers but required human interpretation and preference that is no longer suitable when building interoperable and scalable systems for the analysis of sequence variation. To enhance the precision, consistency, and interoperability of sequence variant descriptions, the HGVS Nomenclature is transitioning to the use of Extended Backus-Naur Form (EBNF) for defining its syntax.²³ EBNF is a formal notation that uses symbols and patterns to rigorously define the structure of a language, making it both transparent to human readers and precise enough for automated processing by software.

a)

DNA Recommendations

Substitution Variant

Definitions

Substitution

a sequence change where, compared to a reference sequence, one nucleotide is replaced by one other nucleotide.

Description

Format: "prefix""position_substituted""reference_nucleotide"">"new_nucleotide", e.g. g.1234>G

"prefix" = reference sequence used = g.

"position_substituted" = position nucleotide substituted = 123

"reference_nucleotide" = nucleotide at reference position = A

">" = type of change is a substitution =>

"new_nucleotide" = substituted nucleotide = G

Note

- prefix reference sequences accepted are g, m, c, and n (genomic, mitochondrial, coding DNA and non-coding DNA).

b)

Substitution

Substitution: a sequence change where, compared to a reference sequence, one nucleotide is replaced by one other nucleotide.

Syntax

Simple sequence substitution

Syntax

sequence_identifier ":" coordinate_type "." position reference_sequence ">"

alternate_sequence

Examples

NC_008803.18:g.33888255C>A

Genome reference with coordinates from aligned transcript

Syntax

sequence_identifier "(" transcript_identifier ":" c." transcript_position reference_sequence ">" alternate_sequence

Examples

NG_012232.1(NM_004886.2):c.93+16>T

Explanation of Symbols

- alternate_sequence: The new sequence (nucleic or amino acid) that will occur at the specified position in the given sequence_identifier

Figure 5.3: **Syntax for a DNA substitution.** a) Syntax for a DNA substitution from the previous website.²⁴ b) Syntax for a DNA substitution from the updated website.²⁵ The new presentation clarifies that there are two distinct forms of a DNA substitution variant and includes an explanation of the syntactic components of each form.

This transition is essential for improving the reliability of sequence variant descriptions in computational systems, as it enables the development of tools that can automatically validate, parse, and interpret these descriptions with minimal ambiguity. By adopting EBNF, the HGVS Nomenclature is better positioned to support scalable genetic analyses, ensuring that descriptions are not only consistent across different platforms but also interoperable with other genomic standards and tools.

HGVS Nomenclature grammar is stored in a single, computer-readable file structure indexed by molecule type (DNA, RNA, protein) and variant type (substitution, deletion, etc.). All presentations of HGVS Nomenclature syntax rules are generated from this single file. Because the current HGVS Nomenclature is not easily represented as a strictly valid grammar, this release adopts the syntax of EBNF for variant descriptions without attempting to reconcile missing or ambiguous elements of the previous release. Nonetheless, the new presentation often clarifies rules that were previously difficult to interpret unambiguously. For example, the HGVS Nomenclature allows two distinct forms of DNA substitutions: a “simple” substitution on a reference sequence and a more complex form that imposes a transcript structure on an underlying genomic sequence. The existence of two forms is not obvious on the previous site but is clearly presented in the new website (Figure 5.3).

The HGVS Nomenclature has historically been grouped by molecule type and then classified by variant type. This arrangement makes it difficult for users to identify commonalities of a variant type across molecule types. For example, it is difficult to understand how DNA, RNA, and protein forms of substitutions might appear together and users often needed to navigate at least three pages to identify the rules for a single variant type, leading to frequent requests for reorganization. Storing grammar rules in a single file enables reuse and reformatting to

HGVS Syntax Summary

variant type	molecule type	forms, syntax, and examples
sub	dna	Simple sequence substitution sequence_identifier ":" coordinate_type "." position reference_sequence ">" alternate_sequence NC_000023.10:g.33030255C>A
		Genome reference with coordinates from aligned transcript sequence_identifier "(" transcript_identifier ":" transcript_position reference_sequence ">" alternate_sequence NG_012232.1(NM_004006.2):c.93+1G>T
	rna	sequence_identifier ":" coordinate_type "." position reference_nucleotide ">" new_nucleotide NM_004006.3:r.123c>g
	aa	Experimentally ascertained protein consequence sequence_identifier ":" coordinate_type "." aa_position alternate_base NP_003997.1:p.Trp24Cys NP_003997.1:p.Trp24Ter NP_003997.1:p.W24*
		Predicted protein consequence sequence_identifier ":" coordinate_type "." "(" aa_position alternate_base ")" NP_003997.1:p.(Trp24Cys)

Figure 5.4: **An excerpt of the new HGVS Syntax Summary pages.**²⁶ The HGVS Syntax Summary shows all syntax rules in HGVS, organized primarily by variant type and secondarily by molecule type. This organization complements that of the website and facilitates readers to compare syntax rules. The syntax rules in the summary and molecule-specific page (Figure 5.3b) are drawn from the same file, thereby guaranteeing consistency between these two views.

summarize the syntax rules by variant type, thereby facilitating comparisons across molecule types (Figure 5.4).

5.4.1.4 List of available software

As the HGVS Nomenclature migrates towards the full adoption of a computational grammar, software will become essential to using the standard reliably and effectively. The new website includes a list of software that has been submitted by the authors and deemed to meet reasonably objective criteria.²⁷ Software must be a library, web API, or web user interface. Only openly available software will be listed, and well-recognized licenses by the Open Source Initiative are preferred. Currently, software may claim one or more of a set of functionality categories in order to help users find tools appropriate for their uses. Tools must also be primarily intended to manipulate HGVS expressions; tools that merely use HGVS expressions for input or output are not included. Tools that have tests and/or were described in a peer-reviewed publication are also preferred. To submit a tool to be included in the software list, developers can submit a pull request to the HGVS Nomenclature repository.²⁰

5.4.2 New and updated features in the HGVS Nomenclature specification

While most of the recent changes to the HGVS Nomenclature focused on how the specification is maintained and communicated, two substantive changes were made in response to timely

requests from the community. A new syntax for gene fusions enables the HGVS Nomenclature to be used for adjoined transcripts. In addition, the HGVS Nomenclature 21.0 updates guidance for the selection of transcripts.

5.4.2.1 Aligning conventions for adjoined transcripts and gene fusions

As defined by VICC, a gene fusion is “the joining of two or more genes resulting in a chimeric transcript and/or a novel interaction between a rearranged regulatory element with the expression of a partner gene transcript”.^{28,29} An adjoined transcript results from the transcription of a gene fusion and is represented in HGVS using a double colon, e.g., **NM_002354.2:r.-358_555::NM_000251.2:r.212_*279**. This definition, alongside the definitions of chimeric transcript fusions and regulatory fusions, has been added to the HGVS Nomenclature glossary. Prior definitions of the terms “RNA Fusion” and “Fusion Transcript” (from community proposal SVD-WG007) have been deprecated in favor of the community-aligned term “adjoined transcript”, focusing on the precise semantics of representing two adjoined transcript sequences. A new page describing the HGVS syntax for adjoined transcripts is available under the RNA Recommendations section of the HGVS Nomenclature.

The VICC fusion nomenclature is compatible with the HGVS Nomenclature recommendations for intervening nucleic acid sequences (described as linker sequences) and adjoined transcripts. It also provides recommendations for an exon-based nomenclature that shares the structure of adjoined transcripts as used in the HGVS Nomenclature. In recognizing that the human genomics community has a need for representing regulatory fusions, exon-based chimeric transcript fusions, and gene-level fusions,²⁹ the HGVS Nomenclature and ISCN now endorse the use of the VICC nomenclature for the representation of gene fusion events, in addition to their prior recommendation to use the HGNC naming conventions³⁰ for fusion genes.

5.4.2.2 Updated guidance for selection of transcripts

Historically, the HGVS Nomenclature has recommended the use of the longest transcript of a gene that represents the known biological and clinical significance of that locus. However, due to transcript sequence differences between RefSeq and Ensembl, and differences between transcripts and the primary assembly, the choice of transcript was practically challenging.

The Matched Annotation from the NCBI and EMBL-EBI Select (MANE) transcript set defines representative transcripts and corresponding proteins for human protein-coding genes.³¹ MANE transcripts share identical exon structure and sequence between NCBI and Ensembl, and therefore, identifiers from those databases may be used interchangeably. Furthermore, MANE transcript sequences match the GRCh38 reference assembly identically, which avoids

complications that arise with variation that overlaps alignment mismatches and gaps between transcripts and GRCh37. The MANE Select subset defines one transcript at each protein-coding locus that represents the known biology at that locus. The HGVS Nomenclature now recommends the use of MANE Select and MANE Plus Clinical transcripts where possible and appropriate. In addition, the previous recommendation to use Locus Reference Genomic sequences was withdrawn.

5.5 Discussion

The recent updates to the HGVS Nomenclature represent a fundamental shift in how the standard is maintained and presented to the community. Taken together, these changes significantly advance the accessibility and usability of the HGVS Nomenclature.

The reorganization under the auspices of HUGO and the establishment of the HVNC play a pivotal role in maintaining and updating the nomenclature. The HVNC's efforts to coordinate with other standards organizations, such as GA4GH, ISCN, and VICC, ensure that the HGVS Nomenclature evolves in harmony with related standards. Despite the diversity of goals and views, the individuals and organizations behind the standards continually seek to improve interoperability. The introduction of new mechanisms for community feedback, including a formal process for proposing changes, highlights the importance of community involvement in the standard's evolution. This inclusive approach facilitates the adoption of changes that reflect the needs and insights of diverse invested users.

The redesigned HGVS Nomenclature website addresses previous usability challenges by incorporating features such as integrated search, version tracking, and a responsive design. These enhancements improve the overall user experience, making it easier for users to access, navigate, and interpret the information they need. The inclusion of a collapsible navigation menu, an intuitive search bar, and a version selector contributes to a more streamlined and accessible platform. In addition, numerous errors and inconsistencies from previous versions were resolved with this new release. These changes ensure that the website remains functional across a wide range of devices, enhancing accessibility for users worldwide.

The primary mission of the HGVS Nomenclature is to facilitate reliable communication of sequence variants, which requires that the HGVS Nomenclature is stable and consistent. Nonetheless, modifications will be required from time to time in order to address new scientific needs, resolve inconsistencies, or clarify conventions. A key goal of the HVNC is to thoughtfully balance the need for the HGVS Nomenclature to evolve while also ensuring that existing data sharing interactions are not disrupted and communicating the changes to the community clearly. The HVNC has introduced a versioning policy that balances the need

for stability with the necessity of evolution. This policy allows for controlled updates to the nomenclature, ensuring that changes are communicated clearly to the community and that existing data-sharing interactions are not disrupted.

The adoption of EBNF for the HGVS Nomenclature marks a crucial advancement. This formal computational grammar provides a clear and unambiguous method for specifying the syntax of sequence variant descriptions. By ensuring that the nomenclature is both human- and machine-readable, EBNF enhances the accuracy and consistency of variant descriptions and validations. This dual readability is particularly important for enabling the development of interoperable and scalable systems that validate, manipulate, and analyze sequence variation.

Recent updates to the HGVS Nomenclature include a new syntax for gene fusion events and updated guidance for transcript selection. The new syntax for gene fusions enables the nomenclature to describe adjoined transcripts accurately, reflecting a critical need in current genomic research and clinical practice. The updated guidance for transcript selection, which now recommends the use of the MANE Select transcripts, addresses practical challenges associated with transcript sequence differences and alignment mismatches. These changes are essential for ensuring that the nomenclature remains relevant and useful in the context of advancing scientific knowledge and clinical requirements.

Transitioning from human-readable descriptions of sequence variants to a technical specification based on a formal grammar presents several challenges. One major issue is the inherent ambiguity and variability in natural language, which can lead to inconsistencies and misinterpretations when describing sequence variation. Human descriptions often rely on context and implied knowledge, which are difficult to capture in a formal grammar, leaving guidance open to inconsistent interpretation and incorrect usage. Additionally, creating a comprehensive and unambiguous formal specification requires significant effort to define and standardize all possible variant types and their representations, ensuring that they are both human- and machine-readable. This process also involves extensive community engagement and consensus-building to adopt the new specification as well as the development of tools and resources to support its implementation and use. The shift necessitates retraining and adaptation for users who are accustomed to the previous, less formalized system, which can be a substantial barrier to smooth adoption. Moreover, ensuring interoperability with existing data and systems that were designed around the old nomenclature adds another layer of complexity to this transition. The benefits of this transition, some of which are already evident in this release, outweigh the challenges associated with migrating to a formal grammar for the HGVS Nomenclature.

Professional and Compliance organizations, such as the American College of Medical Genetics

(ACMG), College of American Pathologists (CAP), European Molecular Genetics Quality Network Community Interest Company (EMQN CIC), and Genomic Quality Assessment (GenQA), recommend the use of the HGVS Nomenclature (and ISCN for cytogenomic descriptions) during proficiency testing (PT) and external quality assessment (EQA) to standardize the description of sequence variants. These organizations adopt an educational approach to work globally, ensuring that genomic diagnostic testing facilities, research laboratories, and test kit manufacturers accurately and consistently communicate their findings to avoid misrepresentation of test results, which can lead to diagnostic errors and patient mismanagement. However, since some variants have multiple valid representations — for example, protein variants may use either one- or three-letter amino acid codes — exact matching of variant descriptions is often challenging. The HGVS Variant Nomenclature Committee seeks to collaborate with compliance organizations to gradually reduce optionality in the specification by providing clearer recommendations, thereby minimizing unnecessary ambiguity. Minimizing optionality will benefit the entire community, not just compliance organizations.

Conformance tests are crucial for ensuring consistent and accurate application of the HGVS Nomenclature across platforms and tools. These tests, planned as future work, will validate adherence to HGVS standards, reducing errors in sequence variant descriptions and enhancing interoperability between systems. Ultimately, these tests will uphold the integrity of the HGVS Nomenclature, ensuring it remains a trusted standard in genomics.

In conclusion, the recent updates to the HGVS Nomenclature signify a major advancement in the field of sequence variant descriptions. These changes, driven by community engagement and technological advancements, position the nomenclature as a precise, consistent, and interoperable standard that can effectively support the needs of the scientific and clinical communities.

5.6 Conclusions

The modernization of the HGVS Nomenclature through improved governance, community engagement, and technical infrastructure represents a significant advancement for the genomics community. The transition to HUGO oversight, implementation of formal feedback mechanisms, and adoption of software development best practices have created a more sustainable and responsive foundation for maintaining this critical standard. The redesigned website, with its enhanced search capabilities, consistent styling, and clearer presentation of syntax rules, makes the nomenclature more accessible and easier to understand for both new and experienced users.

Perhaps most importantly, the introduction of a formal computational grammar marks a

crucial step toward unambiguous variant descriptions that can be reliably processed by both humans and machines. Combined with the new versioning policy and guidance on transcript selection, these changes position the HGVS Nomenclature to better serve the evolving needs of clinical genomics and research communities while maintaining the stability that users require. As genomic testing becomes increasingly central to healthcare and research, these improvements to the HGVS Nomenclature will help ensure accurate and consistent communication of sequence variants across the global genomics ecosystem.

5.7 Acknowledgements

The HVNC acknowledges and is grateful for the many suggestions and corrections from innumerable users over the years. The HVNC appreciates and celebrates the many years of guidance from past members of the committee: Oleg Agofonov, Stylianos Antonarakis, Raymond Dalglish, Els Dequeker, Peter Freeman, Marc Greenblatt, Donna Maglott, Amy McAllister, Jean McGowan-Jordan, Terence Murphy, Anne-Francoise Roux, Tim Smith, and Peter Taschner. The HVNC is grateful to HUGO and Claire Lee for organizational support.

5.8 Authors' contributions

RKH: Conceptualization, Methodology, Software, Validation, Resources, Writing (Original Draft, Review & Editing), Supervision, Project administration, Funding acquisition. IFACF: Conceptualization, Methodology, Software, Validation, Resources, Writing (Review & Editing), Project administration. MDS: Conceptualization, Methodology, Writing (Review & Editing). RH: Conceptualization, Methodology, Writing (Review & Editing). JFJL: Conceptualization, Methodology, Validation, Writing (Review & Editing). RT: Conceptualization, Methodology, Writing (Review & Editing). AW: Conceptualization, Methodology, Software, Validation, Resources, Writing (Original Draft, Review & Editing), Supervision, Project administration, Funding acquisition. JTdD: Conceptualization, Methodology, Validation, Resources, Writing (Review & Editing), Supervision.

5.9 Funding

AW was supported by the National Human Genome Research Institute award R35HG011949.

5.10 Competing interests

RKH is an employee and shareholder of MyOme, Inc. RH is an employee of GenQA. RT is an employee of EMQN CIC.

5.11 References

- [1] Johan T. den Dunnen, Raymond Dalglish, Donna R. Maglott, Reece K. Hart, Marc S. Greenblatt, Jean McGowan-Jordan, Anne-Francoise Roux, Timothy Smith, Stylianos E. Antonarakis, and Peter E.M. Taschner. *HGVS Recommendations for the Description of Sequence Variants: 2016 Update*. Human Mutation 2016; 37 (6) 564–569.
- [2] J.T. den Dunnen and E. Antonarakis. *Nomenclature for the description of human sequence variations*. Human Genetics 2001; 109 (1) 121–124.
- [3] J. McGowan-Jordan, S. Moore, and R. Hastings, eds. *ISCN 2020: an international system for human cytogenomic nomenclature*. Karger, 2020.
- [4] Ad Hoc Committee on Mutation Nomenclature. *Update on nomenclature for human gene mutations*. Human Mutation 1996; 8 (3) 197–202.
- [5] Stylianos E. Antonarakis and Nomenclature Working Group. *Recommendations for a Nomenclature System for Human Gene Mutations*. Human Mutation 1998; 11 1–3.
- [6] Petr Danecek, Adam Auton, Goncalo Abecasis, Cornelis A. Albers, Eric Banks, Mark A. DePristo, Robert E. Handsaker, Gerton Lunter, Gabor T. Marth, Stephen T. Sherry, Gilean McVean, and Richard Durbin. *The variant call format and VCFtools*. Bioinformatics 2011; 27 (15) 2156–2158.
- [7] Alex H. Wagner, Lawrence Babb, Gil Alterovitz, Michael Baudis, Matthew Brush, Daniel L. Cameron, Melissa Cline, Malachi Griffith, Obi L. Griffith, Sarah E. Hunt, David Kreda, Jennifer M. Lee, Stephanie Li, Javier Lopez, Eric Moyer, Tristan Nelson, Ronak Y. Patel, Kevin Riehle, Peter N. Robinson, Shawn Rynearson, Helen Schuilenburg, Kirill Tsukanov, Brian Walsh, Melissa Konopko, Heidi L. Rehm, et al. *The GA4GH Variation Representation Specification: A computational framework for variation representation and federated identification*. Cell Genomics 2021; 1 (2).
- [8] Heidi L. Rehm, Angela J.H. Page, Lindsay Smith, Jeremy B. Adams, Gil Alterovitz, Lawrence J. Babb, Maxmillian P. Barkley, Michael Baudis, Michael J.S. Beauvais, Tim Beck, Jacques S. Beckmann, Sergi Beltran, David Bernick, Alexander Bernier, James K. Bonfield, Tiffany F. Boughtwood, Guillaume Bourque, Sarion R. Bowers, Anthony J. Brookes, Michael Brudno, Matthew H. Brush, David Bujold, Tony Burdett, Orion J. Buske, Moran N. Cabili, et al. *GA4GH: International policies and standards for data sharing across genomic research and healthcare*. Cell Genomics 2021; 1 (2).
- [9] Wesley Goar, Lawrence Babb, Srikar Chamala, Melissa Cline, Robert R. Freimuth, Reece K. Hart, Kori Kuzma, Jennifer Lee, Tristan Nelson, Andreas Pilić, Kevin Riehle, Anastasia Smith, Kathryn Stahl, Andrew D. Yates, Heidi L. Rehm, and Alex H. Wagner. *Development and application of a computable genotype model in the GA4GH Variation Representation Specification*. Pacific Symposium on Biocomputing 2023; (2023) 383–394.
- [10] National Center for Biotechnology Information.
URL: <https://www.ncbi.nlm.nih.gov/> (visited on August 12, 2024).
- [11] J. Bradley Holmes, Eric Moyer, Lon Phan, Donna Maglott, and Brandi Kattman. *SPDI: data model for variants and applications at NCBI*. Bioinformatics 2020; 36 (6) 1902–1907.

- [12] HGVS Variant Nomenclature Committee. *HGVS Nomenclature: HGVS Variant Nomenclature Committee (HVNC)*.
URL: <https://hgvs-nomenclature.org/hvnc/> (visited on August 16, 2024).
- [13] Alex H. Wagner, Brian Walsh, Georgia Mayfield, David Tamborero, Dmitriy Sonkin, Kilannin Krysiak, Jordi Deu-Pons, Ryan P. Duren, Jianjiong Gao, Julie McMurry, Sara Patterson, Catherine del Vecchio Fitz, Beth A. Pitel, Ozman U. Sezerman, Kyle Ellrott, Jeremy L. Warner, Damian T. Rieke, Tero Aittokallio, Ethan Cerami, Deborah I. Ritter, Lynn M. Schriml, Robert R. Freimuth, Melissa Haendel, Gordana Raca, Subha Madhavan, et al. *A harmonized meta-knowledgebase of clinical interpretations of somatic genomic variants in cancer*. *Nature Genetics* 2020; 52 (4) 448–457.
- [14] *HGVS Nomenclature Announcements - Google Groups*.
URL: <https://groups.google.com/g/hgvs-nomenclature-announcements> (visited on November 12, 2024).
- [15] *HGVS Nomenclature - Google Groups*.
URL: <https://groups.google.com/g/hgvs-nomenclature> (visited on August 12, 2024).
- [16] *Markdown Guide*.
URL: <https://www.markdownguide.org/> (visited on August 12, 2024).
- [17] *MkDocs*.
URL: <https://www.mkdocs.org/> (visited on August 12, 2024).
- [18] *Read the Docs*.
URL: <https://about.readthedocs.com/> (visited on August 12, 2024).
- [19] HGVS Variant Nomenclature Committee. *HGVS Nomenclature Archive*.
URL: <https://archive.hgvs-nomenclature.org/> (visited on August 12, 2024).
- [20] Reece K. Hart, Ivo F.A.C. Fokkema, Marina DiStefano, Ros Hastings, Jeroen F.J. Laros, Rachel Taylor, Alex Wagner, and Johan T. den Dunnen. *HGVSnomenclature/hgvs-nomenclature: HGVS Nomenclature GitHub*.
URL: <https://github.com/HGVSnomenclature/hgvs-nomenclature/> (visited on August 12, 2024).
- [21] Tom Preston-Werner. *Semantic Versioning*.
URL: <https://semver.org/> (visited on August 12, 2024).
- [22] HGVS Variant Nomenclature Committee. *HGVS Nomenclature: Versions*.
URL: <https://hgvs-nomenclature.org/versions/> (visited on August 12, 2024).
- [23] Jeroen F.J. Laros, André Blavier, Johan T. den Dunnen, and Peter E.M. Taschner. *A formalized description of the standard human variant nomenclature in Extended Backus-Naur Form*. *BMC Bioinformatics* 2011; 12 Suppl 4 (Suppl 4).
- [24] HGVS Variant Nomenclature Committee. *HGVS Nomenclature Archive: DNA Substitution*.
URL: <https://archive.hgvs-nomenclature.org/recommendations/DNA/variant/substitution/> (visited on August 12, 2024).
- [25] HGVS Variant Nomenclature Committee. *HGVS Nomenclature: DNA Substitution*.
URL: <https://hgvs-nomenclature.org/recommendations/DNA/substitution/> (visited on August 12, 2024).

- [26] HGVS Variant Nomenclature Committee. *HGVS Nomenclature: HGVS Syntax Summary*. URL: <https://hgvs-nomenclature.org/recommendations/summary/> (visited on August 12, 2024).
- [27] HGVS Variant Nomenclature Committee. *HGVS Nomenclature: Software*. URL: <https://hgvs-nomenclature.org/software/> (visited on August 12, 2024).
- [28] *VICC Gene Fusion Specification*. URL: <https://fusions.cancervariants.org/en/latest/terminology.html> (visited on August 12, 2024).
- [29] Alex H. Wagner, Yasmine Akkari, Marilyn Li, Angshumoy Roy, Karen Tsuchiya, and Gordana Raca. *Recommendations for future extensions to the HGNC gene fusion nomenclature*. *Leukemia* 2021; 35 (12) 3611–3612.
- [30] Elspeth A. Bruford, Cristina R. Antonescu, Andrew J. Carroll, Arul Chinnaiyan, Ian A. Cree, Nicholas C.P. Cross, Raymond Dalgleish, Robert Peter Gale, Christine J. Harrison, Rosalind J. Hastings, Jean Loup Huret, Bertil Johansson, Michelle Le Beau, Cristina Mecucci, Fredrik Mertens, Roel Verhaak, and Felix Mitelman. *HUGO Gene Nomenclature Committee (HGNC) recommendations for the designation of gene fusions*. *Leukemia* 2021; 35 (11) 3040–3043.
- [31] Joannella Morales, Shashikant Pujar, Jane E. Loveland, Alex Astashyn, Ruth Bennett, Andrew Berry, Eric Cox, Claire Davidson, Olga Ermolaeva, Catherine M. Farrell, Reham Fatima, Laurent Gil, Tamara Goldfarb, Jose M. Gonzalez, Diana Haddad, Matthew Hardy, Toby Hunt, John Jackson, Vinita S. Joardar, Michael Kay, Vamsi K. Kodali, Kelly M. McGarvey, Aoife McMahon, Jonathan M. Mudge, Daniel N. Murphy, et al. *A joint NCBI and EMBL-EBI transcript set for clinical genomics and research*. *Nature* 2022; 604 (7905) 310–315.