Title

The genotype list string code syntax for exchanging nomenclature-level genotyping results in clinical and research data management and analysis systems.

Journal

HLA: Immune Response Genetics, 102(4)

Authors

Mack, Steven
Sauter, Jürgen
Robinson, James
et al.

Publication Date

2023-10-01

DOI

10.1111/tan.15145

Peer reviewed

HHS Public Access

Author manuscript

HLA. Author manuscript; available in PMC 2024 October 01.

Published in final edited form as:

HLA. 2023 October ; 102(4): 501–507. doi:10.1111/tan.15145.

The Genotype List String Code Syntax for Exchanging Nomenclature-Level Genotyping Results in Clinical and Research Data Management and Analysis Systems

Steven J. Mack[1,*], **Jürgen Sauter**[2], **James Robinson**[3,4], **Kazutoyo Osoegawa**[5], **Lloyd McKenzie**[6], **Joel Schneider**[7], **Martin Maiers**[7], **Robert P. Milius**[7]

[1]Department of Pediatrics, University of California, San Francisco, Oakland, CA, USA

[2]DKMS, Tübingen, Germany

[3]Anthony Nolan Research Institute, Royal Free Campus, London, UK

[4]UCL Cancer Institute, Royal Free Campus, London, UK

[5]Histocompatibility & Immunogenetics Laboratory, Stanford Blood Center, Stanford Health Care, Palo Alto, CA, USA

[6]Dogwood Health Consulting, North Saanich, BC, Canada

[7]National Marrow Donor Program, Minneapolis, MN, USA

Abstract

The nomenclatures used to describe HLA and Killer-cell Immunoglobulin-like Receptor (KIR) alleles distinguish unique nucleotide and peptide sequences and patterns of expression, but are insufficient for describing genotyping results, as description of ambiguities and relations across loci require terminology beyond allele names. The Genotype List (GL) String grammar describes genotyping results for genetic systems with defined nomenclatures, like HLA and KIR, documenting what is known and unknown about a given genotyping result. However, the accuracy of a GL String is dependent on the reference database version under which it was generated. Here, we describe the GL String Code (GLSC) system, which associates each GL String with meta-data describing the specific reference context in which the GL String was created, and in which it should be interpreted. GLSC is a defined syntax for exchanging GL Strings in the context of a specific gene-family namespace, allele-name code-system, and pertinent reference database version. GLSC allows HLA and KIR genotyping data to be transmitted, parsed and interpreted in the appropriate context, in an unambiguous manner, on modern data-systems, including Health

[*]Corresponding Author: Steven J. Mack, steven.mack@ucsf.edu, 5700 Martin Luther King Jr. Way, Oakland, CA 94609-1673, USA.

AUTHOR CONTRIBUTIONS

All authors have read and approved the final manuscript. Steven Mack, Martin Maiers, Robert Milius, Kazutoyo Osoegawa, James Robinson and Jürgen Sauter participated in DaSH Hackathon discussions regarding GL String extensions and improvements. Kazutoyo Osoegawa and Jürgen Sauter provided model data. Lloyd McKenzie, Robert Milius and Joel Schneider provided input and expertise on HL7 FHIR systems. Steven Mack drafted the paper and supplementary material. All authors made contributions to the final version of the paper. No artificial intelligence systems were applied in the writing of the paper or for the work described.

CONFLICT OF INTEREST STATEMENT

The authors have no conflicts of interest to declare.

Level 7 Fast Healthcare Interoperability Resource systems. Technical specification for GLSC can be found at https://glstring.org.

Keywords

genotype list string code; killer-cell immunoglobulin-like receptor; HLA; HL7 FHIR; electronic healthcare record systems; codeable concept

INTRODUCTION

The need for standardized nomenclature systems to describe genes and gene products has been recognized for almost 70 years[1, 2]. The World Health Organization Nomenclature Committee (WHO Nomenclature Committee) for Factors of the HLA system[3] and its Killer-cell Immunoglobulin-like Receptor (KIR) gene sub-committee[4] maintain allele-name nomenclature systems for the HLA and KIR genes, which encode cell-surface molecules that facilitate the interaction of immune cells as part of the mechanisms of adaptive and innate immunity[5–9]. Though they differ significantly in key specifics, both gene systems' allele name nomenclatures apply a set of fields to describe the unique peptide, coding-nucleotide and non-coding nucleotide sequences for each allele[10], with an additional field for HLA alleles describing the (often predicted) antigenicity of each unique protein sequence[11], as illustrated in Figure 1. In both cases, the HLA and KIR nomenclatures describe what is known about a given gene's nucleotide and peptide sequences (and antigenicity for HLA) for specific reference specimens. Both gene families display high levels of polymorphism[12], with 36,263 unique HLA nucleotide sequences, encoding 21,013 unique HLA proteins, known as of April of 2023, and 1617 unique KIR nucleotide sequences, encoding 703 unique proteins, known as December of 2022.

In HLA and KIR genotyping experiments, it is often the case that the complete nucleotide sequence of an HLA or KIR gene is not determined[13]. Given the large number of known HLA and KIR alleles, accurate reporting of an experiment usually requires that a list of multiple possible alleles sharing the interrogated sequence be returned as an experimental finding. In cases where un-phased sequences are assessed for a gene, multiple possible combinations of these sequences must be considered, often resulting in multiple possible genotypes[14]. The Genotype List (GL) String grammar was developed to standardize the reporting of these and other types of ambiguity that can result from HLA and KIR genotyping experiments[15].

As shown in Table 1, GL Strings apply a set of six hierarchically parsed delimiters (?,^,|,+,~, and /) to accurately describe what is known and unknown about an HLA or KIR genotype for a given genotyping experiment[16]. Combined, these delimiters accurately and comprehensively represent what is known and unknown about an HLA or KIR genotype. For example, HLA-A*23:01:01:01+HLA-A*24:02:01:01/HLA-A*24:02:01:96|HLA-A*23:17:01:01+HLA-A*24:462 describes two alternative HLA-A genotypes, delimited by the "|", one of which includes two alleles that are identical in the sequenced region and are delimited by the "/". The alleles encoded by individual copies the *HLA-A* gene are delimited by the "+".

GL Strings can further describe what is known and unknown regarding multiple loci. HLA-A*02:01/HLA-A*02:02+HLA-A*03:01|HLA-A*02:07+HLA-A*03:06^HLA-B*08:01+HLA-B*44:02/HLA-B*44:03^HLA-DRB1*03:01~HLA-DRB3*01:01+HLA-DRB1*03:01~HLA-DRB3*01:01 describes genotypes for the *HLA-A*, *HLA-B*, *HLA-DRB1*, and *HLA-DRB3* genes. In addition to the "|", "/", and "+" operators described above, genotypes for loci that are not in phase are delimited by the "+", while alleles of loci for which phase has been experimentally determined are delimited with the "~".

Finally, GL Strings can describe what is known and unknown regarding paralogous loci. KIR2DL5A*00105+KIR2DL5A*00105?KIR2DL5A*00105+KIR2DL5B*00802+?KIR2DL5B*00802+KIR2DL5B*00802 describes three possible genotypes for the KIR2DL5A and KIR2DL5B genes. The possible genotypes are delimited by the "?" operator. In this case, nucleotide sequences correspond to alleles at both the KIR2DL5A and KIR2DL5B genes, and copy number information confirms two copies of these sequences.

However, the HLA nomenclature format has changed over time, and the number of known HLA and KIR alleles has grown with continuous allele discovery. HLA allele name and sequence data are housed in the IPD-IMGT/HLA Database[17], which has had 97 releases since December of 1998. KIR allele name and sequence data are housed in the IPD-KIR Database[18], which has had 18 releases since July of 2003. Because each database release includes new allele sequences and names, a GL String that accurately represents genotyping performed in the context of one database release is likely to be inaccurate for other releases. For instance, a newly discovered allele that shares a stretch of nucleotides with previously identified alleles should be included in a respective ambiguous typing result, but cannot be included if the original genotyping had been performed before that allele was identified. If the GL String for that genotyping is interpreted in the wrong temporal context, it might appear this this new allele had been excluded as a potential typing result, when it had not been. Therefore, to accurately represent a genotyping, a GL String must be considered in the reference database context under which it was generated.

The development of a standard means of representing the relationship between a genotyping result and its reference context has proven challenging. GL Strings can be very long, requiring modifications of software and databases to accommodate them. In turn, these accommodations necessitate the concerted efforts of informatic professionals to implement changes on multiple systems that share data internationally. Attempts to include GL Strings in clinical reports written using healthcare standards have focused on Health Level 7 International (HL7) Fast Healthcare Interoperability Resource [FHIR], an international clinical data interoperability standard developed by the HL7 organization to exchange health data using the FHIR specification[19]. Since its development in 2012, the FHIR standard has become widely used in healthcare communication systems, Electronic Health Record (EHR) systems, and mobile healthcare applications.

An HL7 FHIR *codeable concept* data element associates a *code* with the *code system* that defines and controls that code (see https://hl7.org/fhir/R4/datatypes.html#CodeableConcept, https://hl7.org/fhir/R4/terminologies.html, and http://hl7.org/fhir/R4/terminologies-systems.html for details). For example, a codeable concept

that includes the code “57290-9” and code system “http://loinc.org” defines the Logical Observation Identifiers Names and Codes (LOINC) code for the laboratory test “HLA-A [Type] by High resolution”[1] (https://loinc.org/57290-9). A result of this laboratory test could be a single HLA-A allele, which could similarly be recorded as a codeable concept by providing e.g., “A*01:01:01:01” as the code, and providing the Uniform Resource Locator (URL) for the IPD-IMGT/HLA Database, “https://www.ebi.ac.uk/ipd/imgt/hla”, as the code system. An HL7 FHIR codeable concept can optionally include a *version* of the code system. A compact JavaScript Object Notation (JSON) example of an HLA-A allele included in an HL7 FHIR *observation.valueCodeableConcept* is illustrated in Figure 2.

This approach is feasible for individual alleles, but a GL String cannot serve as a code because the IPD-IMGT/HLA Database cannot be specified as the code system; it does not define the GL String grammar. A new system is needed to identify the gene namespace and reference context together with the GL String. Here, we introduce the GL String Code (GLSC) system, which encapsulates GL Strings with meta-data specifying the gene-namespace (which includes the code system under which the allele names should be parsed) and the pertinent reference database version or date of genotyping in a single text string. Unlike GL Strings, GLSCs can be included in HL7 FHIR *valueCodeableConecpts*. Use of GLSC in medical data transmission systems enables the most effective use of genotyping results for medical applications.

METHODS and RESULTS

The GLSC system was developed as part of the Data Standards Hackathons for Next Generation Sequencing (DaSH for NGS) (https://github.com/nmdp-bioinformatics/dash/wiki), which have been developing tools, systems and services for standardized management of immunogenomic data for the last decade[23, 24]. An early DaSH for NGS product was the GL Service[25], a RESTful web service envisioned as a Digital Object Identifier system for HLA and KIR genotypes. The GL Service generated a short, unique Uniform Resource Identifier (URI) that corresponded to a submitted GL String, and retrieved that GL String when the URI was entered in an internet browser. The GL Service allowed URIs of uniform length (less than 100 characters) to be exchanged in lieu of potentially very-long (hundreds to thousands of characters) GL Strings. To account for the growth and changes in allele names with each reference database release, each URI included the specific gene database and release-version under which its GL String had been registered, and GL Strings for each release-version were recorded in separate instances.

GL Service URIs were compact and easily transmitted, but dereferencing them required internet access, as well as the maintenance of GL Service instances for each HLA or KIR reference database release. These requirements limited the utility of the GL Service, and it has been decommissioned. The GLSC system was developed to address these limitations, by transmitting a full GL String in the context under which it was created.

[1]Nunes et al. defined a “high resolution” HLA genotype as identifying the amino acid sequence encoded by class I exons 2 and 3, and class I exon 2, resulting in two-field allele names [20]. With the advent of NGS typing methods, the “high resolution” term has been applied to three-field allele names that may display allelic ambiguity due to unsequenced exons [21].

The three elements of a GLSC (Gene Family [GF] Namespace, GF Nomenclature Version or GL String creation Date, and GL String) are described in Table 2. Multiple nomenclatures may be in use within a given GF Namespace. Within the *hla* namespace, the GLSC system supports the following identifiers: allele names, G groups and P groups defined by the WHO Nomenclature Committee, multiple allele code (MAC) designations defined by the National Marrow Donor Program (NMDP) (e.g., the HLA-A*24:AMG MAC represents the HLA-A*24:02/HLA-A*24:09N ambiguity [https://bioinformatics.bethematchclinical.org/hla-resources/allele-codes/allele-code-lists/]), and the additional codes (XXXX, NNNN, UUUU, and NEW) and allele family XX codes defined by the World Marrow Donor Association (WMDA) [26]. Within the *kir* namespace, only allele names defined by the WHO Nomenclature Committee's KIR subcommittee are supported.

For HLA alleles, G and P groups and KIR alleles, the GF Nomenclature Version should identify the corresponding IPD-IMGT/HLA or IPD-KIR Database version under which the genotyping results were generated. For NMDP MACs and WMDA additional and allele family XX codes, when a specific database version is unavailable, the date on which the GL String was constructed can be provided instead, under the assumption that the date will identify the potential range of reference database releases under which the genotype data were generated. When dates are provided, they must adhere to the HL7 FHIR date type (https://hl7.org/fhir/R4/datatypes.html#date), and follow the yyyy, yyyy-mm, or yyyy-mm-dd format. The elements of a GLSC are delimited with pound signs (#). An example GLSC for an HLA genotype generated under IPD-IMGT/HLA Database release version 3.25.0 is illustrated in Figure 3. Full technical specifications for GLSC, along with additional examples, can be found online at https://glstring.org/syntax-1.1.html.

Because each GLSC includes a GL String, GLSCs are compatible with any current application of a GL String. The addition of GF Namespace and Nomenclature Version or Date information in a GLSC will foster the standardization of allele identifiers generated in different reference database epochs, allowing the GL String data in a GLSC to be updated with respect to allele identifier changes, or nucleotide sequence extensions for a given allele, over time. Efforts to add support for GLSC and to implement these types of standardizations for extant data-analysis tools and data-storage systems are underway.

Exchanging GLSCs on HL7 FHIR Data Systems

GLSCs are easily read by both humans and machine systems. Use of GLSC facilitates the transmission of HLA and KIR genotype data through modern FHIR systems, by including GSLCs in HL7 FHIR systems as part of a codeable concept. HL7 FHIR Observation resources that report laboratory results use the codeable concept in the *observation.valueCodeableConcept* element as the result. In these cases, the GLSC grammar is defined by the "http://glstring.org" code system. In addition to the GLSC itself, HLR FHIR systems using a *valueCodeableConcept* may include additional details of the code system and its version. Using the grammar defined for the GL String Code, an example codeable concept is illustrated in Figure 4.

GLSC was used in the development of the HL7 Genomics Reporting FHIR Implementation Guide and in particular the section on Hiscompatibility Reporting (http://hl7.org/fhir/uv/genomics-reporting/). This Implementation Guide was further constrained in an HLA Reporting Implementation Guide (https://fhir.nmdp.org/ig/hla-reporting/)[27].

DISCUSSION AND CONCLUSION

Immunogenetic genotyping efforts have been underway for almost 40 years. Over these last four decades, innovations in nucleotide sequencing technologies, and the concomitant development of computing and informatics technologies, have driven the rapid growth in the discovery of HLA and KIR polymorphism, and its application for basic science, clinical solutions and human health. It is likely that the number of new HLA and KIR alleles identified each year will continue to grow, as estimates suggest that the Human population harbors millions of HLA alleles at each locus[12, 28]. The need for sophisticated informatic systems to manage and exchange these data in an automated fashion will only increase as new therapeutic applications requiring HLA and KIR genotyping are innovated. A key DaSH goal has been the development of seamless, easy-to-use systems for the exchange and application of HLA and KIR data, with the aim of enabling widespread, effective application of immunogenetic health-care data.

The development of GLSC and its integration into HL7 FHIR systems is a key step in meeting those goals. A GLSC in a HL7 FHIR *valueCodeableConcept* describes both the genotype and the information necessary to interpret it in its proper context in a lossless fashion, which can be translated into more accurate, granular electronic health data for clinicians and patients. Looking beyond the application of GLSC for HLA and KIR genotypes, future versions of the GLSC system could easily extend its use to additional genetic nomenclature systems. For example, sequencing results for gene systems that lack a formal nomenclature (e.g., the *ABO* and *Leukocyte Immunoglobulin-Like Receptor* genes) could be named using a Gene Feature Enumeration approach[29], and transmitted via HL7 FHIR.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGEMENTS

We thank the participants of the Data Standards Hackathons for NGS for their continued participation and for helpful discussion on the topics of data-representation and data-sharing. The work described here was supported by National Institutes of Health (NIH) National Institute of Allergy and Infectious Disease (NIAID) grant R01AI128775 (SM, MM), US Office of Naval Research (ONR) grant N00014-20-1-2832 (MM, RM), and Office of the National Coordinator for Health Information Technology (ONC) support for Sync for Genes Phases 1, 2 and 3 (MM, RM). The content is solely the responsibility of the authors and does not necessarily reflect the official views of the NIAID, NIH, ONC, or United States Government.

Abbreviations:

DaSH	Data Standards Hackathon

DOI	Digital Object Identifier
EHR	Electronic Health Record
FHIR	Fast Healthcare Interoperability Resource
GF	Gene Family
GL	Genotype List
GLSC	Genotype List String Code
HL7	Health Level 7
JSON	JavaScript Object Notation
KIR	Killer-cell Immunoglobulin-like Receptor
LOINC	Logical Observation Identifiers Names and Codes
MAC	Multiple Allele Code
NGS	Next Generation Sequencing
NMDP	National Marrow Donor Program
URL	Uniform Resource Locator
URI	Uniform Resource Identifier
WHO	World Health Organization
WMDA	World Marrow Donor Association

REFERENCES

1. Tanaka Y, Report of the international committee on genetic symbols and nomenclature. International Union of Biological Sciences B, 1957. 30: p. 1–6.
2. Human gene mapping 5. Edinburgh Conference (1979). Birth Defects Orig Artic Ser, 1980. 15(11): p. 1–236. [PubMed: 6938249]
3. Committee, W.N., Nomenclature for factors of the HL-a system. Bull World Health Organ, 1968. 39(3): p. 483–6. [PubMed: 5303912]
4. Trowsdale J, Genetic and functional relationships between MHC and NK receptor genes. Immunity, 2001. 15(3): p. 363–74. [PubMed: 11567627]
5. Parham P, et al. , Nature of polymorphism in HLA-A, -B, and -C molecules. Proceedings of the National Academy of Sciences of the United States of America, 1988. 85(11): p. 4005–4009. [PubMed: 3375250]
6. Lechler R and Warrens A, eds. HLA in health and disease 2 ed. 2000, Academc press: San Diego. 472.
7. Selvakumar A, et al. , Genomic organization and allelic polymorphism of the human killer cell inhibitory receptor gene KIR103. Tissue Antigens, 1997. 49(6): p. 564–73. [PubMed: 9234477]
8. Vales-Gomez M, et al. , Kinetics of interaction of HLA-C ligands with natural killer cell inhibitory receptors. Immunity, 1998. 9(3): p. 337–44. [PubMed: 9768753]
9. Moretta A, et al. , P58 molecules as putative receptors for major histocompatibility complex (MHC) class I molecules in human natural killer (NK) cells. Anti-p58 antibodies reconstitute lysis of MHC

class I-protected cells in NK clones displaying different specificities. J Exp Med, 1993. 178(2): p. 597–604. [PubMed: 8340759]

10. Marsh SGE, et al. , Killer-cell immunoglobulin-like receptor (KIR) nomenclature report, 2002. Human Immunology, 2003. 64(1): p. 648–654. [PubMed: 12770798]
11. Marsh SG, et al. , Nomenclature for factors of the HLA system, 2010. Tissue Antigens, 2010. 75(4): p. 291–455. [PubMed: 20356336]
12. Robinson J, et al. , Distinguishing functional polymorphism from random variation in the sequences of >10,000 HLA-A, -B and -C alleles. PLoS Genet, 2017. 13(6): p. e1006862. [PubMed: 28650991]
13. Schöfl G, et al. , 2.7 million samples genotyped for HLA by next generation sequencing: lessons learned. BMC Genomics, 2017. 18(1): p. 161. [PubMed: 28196473]
14. Hollenbach JA, et al. , A community standard for immunogenomic data reporting and analysis: proposal for a STrengthening the REporting of Immunogenomic Studies statement. Tissue Antigens, 2011. 78(5): p. 333–44. [PubMed: 21988720]
15. Milius RP, et al. , Genotype List String: a grammar for describing HLA and KIR genotyping results in a text string. Tissue Antigens, 2013. 82(2): p. 106–12. [PubMed: 23849068]
16. Mack SJ, et al. , GL String 1.1: Extending the Genotype List String Grammar for Describing HLA and KIR Genotypes. HLA 2023 Aug;102(2):206–212. [PubMed: 37286192]
17. Barker DJ, et al. , The IPD-IMGT/HLA Database. Nucleic Acids Res, 2023. 51(D1): p. D1053–d1060. [PubMed: 36350643]
18. Robinson J, et al. , IPD--the Immuno Polymorphism Database. Nucleic Acids Res, 2013. 41(Database issue): p. D1234–40. [PubMed: 23180793]
19. Chronaki C and Ploeg F, Towards mHealth Assessment Guidelines for interoperability: HL7 FHIR. Stud Health Technol Inform, 2016. 224: p. 164–9. [PubMed: 27225573]
20. Nunes E, et al. , Definitions of histocompatibility typing terms: Harmonization of Histocompatibility Typing Terms Working Group. Hum Immunol, 2011. 72(12): p. 1214–1216. [PubMed: 21723898]
21. Do MD, et al. , High-Resolution HLA Typing of HLA-A, -B, -C, -DRB1, and -DQB1 in Kinh Vietnamese by Using Next-Generation Sequencing. Front Genet, 2020. 11: p. 383. [PubMed: 32425978]
22. Balgansuren G, et al. , HLA-B leader genotypes in a clinical population. HLA, 2023 Jul;102(1):44–51. [PubMed: 36929133]
23. Matern BM, et al. , Standard reference sequences for submission of HLA genotyping for the 18th International HLA and Immunogenetics Workshop. HLA, 2021. 97(6): p. 512–519. [PubMed: 33719220]
24. Osoegawa K, et al. , Challenges for the standardized reporting of NGS HLA genotyping: Surveying gaps between clinical and research laboratories. Hum Immunol, 2021. 82(11): p. 820–828. [PubMed: 34479742]
25. Milius RP, et al. , The GL service: Web service to exchange GL string encoded HLA & KIR genotypes with complete and accurate allele and genotype ambiguity. Hum Immunol 2016 Mar;77(3):249–256. [PubMed: 26621609]
26. Bochtler W, et al. , World Marrow Donor Association guidelines for use of HLA nomenclature and its validation in the data exchange among hematopoietic stem cell donor registries and cord blood banks. Bone Marrow Transplant, 2007. 39(12): p. 737–41. [PubMed: 17438587]
27. Freimuth RR, et al., 6 - Clinical genomic data on FHIR®: Case studies in the development and adoption of the Genomics Reporting Implementation Guide, in Genomic Data Sharing, McCormick J and Pathak J, Editors. 2023, Academic Press. p. 91–110.
28. Klitz W, Hedrick P, and Louis EJ, New reservoirs of HLA alleles: pools of rare variants enhance immune defense. Trends Genet, 2012. 28(10): p. 480–6. [PubMed: 22867968]
29. Mack SJ, A gene feature enumeration approach for describing HLA allele polymorphism. Hum Immunol, 2015. 76(12): p. 975–81. [PubMed: 26416087]
30. Osoegawa K, et al. , Tools for building, analyzing and evaluating HLA haplotypes from families. Hum Immunol, 2019. 80(9): p. 633–643. [PubMed: 30735756]

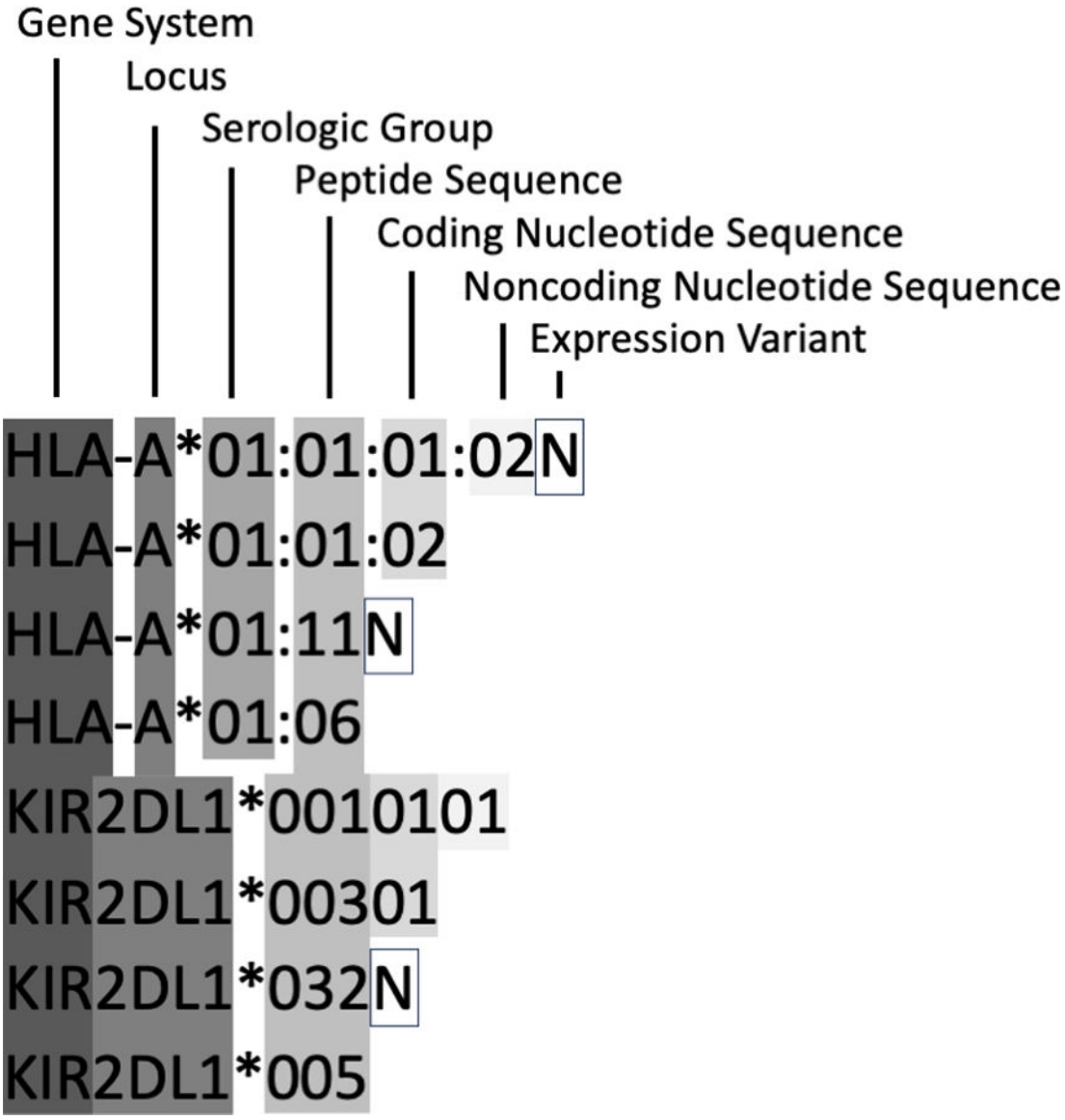

Figure 1.

Structural Elements of HLA and KIR Allele Names

Four HLA and four KIR alleles with names of different lengths are shown. HLA alleles include colon (:) delimiters between fields. KIR allele names do not include a serologic group. Delimiters do not precede expression variants in HLA alleles. This figure is derived from Marsh et al. 2010[11] and Marsh et al. 2003[10].

```
"valueCodeableConcept": {
        "coding": [ {
            "system": "http://www.ebi.ac.uk/ipd/imgt/hla",
            "version": "3.25.0",
            "code": "HLA-A*01:01:01:01"
         } ]
}
```

Figure 2.

An HLA-A allele in a compact JSON HL7 FHIR observation.valueCodeableConcept

The FHIR specification describes three data types for transmitting coded data: **code**, **Coding** and **CodeableConcept**. The **Coding** datatype encapsulates the code system, version, code and display associated with a coded value. The **code** data type contains only the code, and may be used when the code system is indicated by the definition of the data element in which it appears. The **CodableConcept** data type may contain multiple **Coding** representations of the same concept. In the valueCodeableConcept shown, the code system is the URL for the IPD-IMGT/HLA Database, the version is the IPD-IMGT/HLA Database release version under which this HLA-A allele was genotyped (shown in bold), and the code is the name of the allele.

hla#3.25.0#HLA-A*02:01:01:01/HLA-A*02:01:01:02L+HLA-A*24:02:01:01^HLA-B*18:01:01:02/HLA-B*18:01:01:04+HLA-B*35:01:01:01/HLA-B*35:01:01:02^HLA-C*04:01:01:06+HLA-C*07:01:01:01^HLA-DRB1*04:02:01+HLA-DRB1*11:04:01^HLA-DRB3*02:02:01:02+HLA-DRB3*02:02:01:02^HLA-DRB4*01:03:01:01/HLA-DRB4*01:03:01:03+HLA-DRB4*01:03:01:01/HLA-DRB4*01:03:01:03^HLA-DQB1*03:01:01:03+HLA-DQB1*03:02:01^HLA-DQA1*03:01:01+HLA-DQA1*05:05:01:02/HLA-DQA1*05:05:01:04/HLA-DQA1*05:05:01:01^HLA-DPB1*04:02:01:01+HLA-DPB1*107:01/HLA-DPB1*13:01:01^HLA-DPA1*01:03:01:05+HLA-DPA1*02:01:01:01

Figure 3.

An Example Genotype List String Code

The Genotype List String Code # delimiters are shown in bold. This Genotype List String was generated under IPD-IMGT/HLA Database release version 3.25.0. These HLA data were previously published[30].

```
"valueCodeableConcept": {
        "coding": [ {
                "system": "http://glstring.org",
                "version": "1.1",
                "code":"hla#3.25.0#HLA-A*01:01:01:01/HLA-A*01:02+HLA-A*24:02:01:01"
        } ]
}
```

Figure 4.

Genotype List String Code embedded in a compact JSON HL7 FHIR *valueCodeableConcept*

A GL String Code is shown in bold. The inclusion of the GLSC in an HL7 FHIR message requires that it is enclosed within in *valueCodeableConcept* and *coding* tags, which include definition of the code system, in this case, https://glstring.org, and the code system version, in this case 1.1.

Table 1.

Genotype List String Delimiters

Precedence[1]	Delimiter	Identifies	Example
0	?	Possible Loci	*KIR2DL5A*00104?KIR2DL5B*00804*
1	^	Unphased Genes	*HLA-DRB1*11:02:01+HLA-DRB1*11:04:01^HLA-DRB3*02:02:01:01+HLA-DRB3*02:02:01:02*
2	|	Possible Genotypes	*HLA-DPB1*03:01:01+HLA-DPB1*04:01:01:01|HLA-DPB1*124:01+HLA-DPB1*350:01*
3	+	Copies of Genes	*HLA-DPB1*03:01:01+HLA-DPB1*04:01:01:01*
4	~	Phased Genes	*HLA-DRB1*13:01:01:01~HLA-DRB3*02:02:01:02*
5	/	Possible Alleles	*HLA-A*24:02:01:01/HLA-A*24:02:01:96*

1: GL String delimiters are evaluated in the numerical ascending precedence order shown.

Table 2:

Elements of a Genotype List String Code

Element	Description	Examples
Gene Family Namespace	The set of code-systems specific to a particular gene	HLA, KIR
Gene Family Nomenclature Version	The base version of the nomenclature system used by the described gene family namespace; when a nomenclature version is not available, the date on which the GL String was constructed can be used	For HLA 3.25.0, 3.27.1, 3.51.0
		For KIR 2.21.0, 2.7.0
		For dates 2018, 2017-10-15, 2015-01
GL String[1]	The reported GL String	HLA-A*02:02+HLA-A*03:01^ HLA-DRB1*03:01~HLA-DRB3*01:01

1: As recommended in the original description of the GL String format[15], the ‘HLA’ prefix must be included for each HLA allele in a GL String. A discussion of this requirement is included as Supplementary Material.