# MIRING Elements and Formats

| Number | | Element | Components | Messaging Instructions and Notes |
|---|---|---|---|---|
| 1 | | Message Annotation | | |
| | 1.1 | | Unique MIRING Message Identifier | Identifies the MIRING message generator (e.g., an International Organization for Standardization (ISO)/International Electrotechnical Commission (IEC) standard 6523 organization identifier (OID)[1, 2]) and the specific MIRING message. |
| | 1.2 | | Message Generator Contact Information | Email, mailing address, website, phone number, etc. |
| | 1.3 | | Platform Documentation (MIRING element 6) Reference | e.g., citation of a peer-reviewed publication or an entry in the National Center for Biotechnology Information (NCBI) Genetic Testing Registry (GTR) (ncbi.nlm.nih.gov/gtr/) |
| | 1.4 | | Read Processing Documentation (MIRING element 7) Reference | A reference to the location of a structured report documenting the use of programs/scripts (including parameters and order of use) to process the primary read data in order to make allele calls. |
| | 1.5 | | Primary Data Availability | A: Public, and available as defined in MIRING element 1.6<br>B: Private, and potentially available by contacting the message generator as defined in MIRING element 1.2 |
| | 1.6 | | Primary Data (MIRING element 8) Reference | Provided when permitted.<br>e.g., referenced to data in the NCBI Sequence Read Archive (SRA) (ncbi.nlm.nih.gov/sra/) |

| Number | | Element | Components | Messaging Instructions and Notes |
|---|---|---|---|---|
| 2 | | Reference Context | | |
| | 2.1 | | Reference Sequence Database Version for Allele Calling | Identify for each locus included in the message |
| | 2.2 | | Individual Reference Sequences Applied | Identify the source database and accession number of each individual sequence applied in the message. |
| | 2.2.1 | | Reference Sequence Identifier | A unique identifier ranging from 0 to n-1, where n is the number of reference sequences in MIRING element 2.2. |
| | 2.3 | | Reference Sequence Source Type | Specified for MIRING elements 2.1 and 2.2<br>A: Public and curated<br>B: Public and uncurated<br>C: Not public<br>D: No reference |
| 3 | | Full Genotype | | Defined in MIRING elements 3.1 and 3.2 |
| | 3.1 | | Pertinent Locus/Loci | Genetic loci as defined in an International Nucleotide Sequence Database Collaboration (INSDC) resource[3]. All of the loci tested as part of the work reported in the MIRING message should be included. |
| | 3.2 | | Formatted Genotype | If a genotype is detected for a given locus, report that genotype in Genotype List (GL) String format[4], or an equivalent format. If a locus is identified in MIRING element 3.1, but no sequence data are generated for that locus, report that genotype as 'Absent' in the GL String. |
| | 3.3 | | Genotype Uniform Resource Identifier (URI) | e.g., derived from the GL Service (gl.nmdp.org) |

| Number | | Element | Components | Messaging Instructions and Notes |
|---|---|---|---|---|
| 4 | | Consensus Sequence | | The use of MIRING elements 4.1 and 4.2 to describe the consensus sequence allows the sequences to be reported in FASTA format[5-7]. FASTA format is not a required component of a MIRING message, but a FASTA formatted component of a MIRING-derived genotype report should use the pipe-delimited header format described in 4.2. |
| | 4.1 | | Consensus Sequence Block (CSB) | A contiguous nucleotide sequence organized in the 5' to 3' direction and written using IUPAC and IUBMB nucleotide base symbols[8]. Multiple sequence blocks may be included in a MIRING message. |
| | 4.2 | | Consensus Sequence Descriptor | For FASTA representations of consensus sequence, assign each CSB a pipe-delimited descriptor comprised by MIRING elements 4.2.1 – 4.2.7. |
| | 4.2.1 | | Consensus Sequence Block Identifier | Ranges from 0 to n-1, where n is the number of CSBs included in the message. CSB identifier numbers must increase in the 5' to 3' order of CSBs. |
| | 4.2.2 | | Reference Sequence Identifier | MIRING Element 2.2.1 pertinent to each CSB. If the reference sequence is identified as being of type D (no Reference; MIRING element 2.3) the entire CSB is considered to be a novel polymorphism, but does not need to be independently documented as part of MIRING element 5. |
| | 4.2.3 | | Reference Sequence Coordinate | The position in the reference sequence (MIRING element 2.2.1) (indexed from 0) corresponding to the 1st position of the CSB. |

| Number | | Element | Components | Messaging Instructions and Notes |
|---|---|---|---|---|
| | 4.2.4 | | Phase Set | When phase information is available, identify the lowest numbered CSB (using MIRING element 4.2.1) sharing phase with a given CSB; assign the lowest numbered CSB in a phase set its own CSB identifier. If no phase information is available for a given CSB, assign that CSB its own CSB identifier |
| | 4.2.5 | | Copy Number | 1 to n, where n is the number of distinct sequences represented by the CSB (e.g., haploid = 1, diploid = 2, etc.). |
| | 4.2.6 | | Reference Sequence Match | 1: CSB exactly matches the sequence range (MIRING element 4.2.3) of the reference sequence (MIRING Element 2.2.1).<br>0: CSB does not exactly match the sequence range of the reference sequence.<br>When reference sequence match = 0, a description of novel polymorphisms (MIRING element 5) is expected unless value for MIRING element 2.3 = D. |
| | 4.2.7 | | Sequence Continuity | 1: no sequence gaps occur between a given CSB and the preceding CSB in the same phase set.<br>0: there is no phase information for this consensus sequence block, or there is a sequence gap between this CSB and the preceding CSB. |

| Number | | Element | Components | Messaging Instructions and Notes |
|---|---|---|---|---|
| 5 | | Novel Polymorphisms | | Define novel polymorphisms (identified in MIRING element 4.2.6) using MIRING elements 5.1 – 5.8. These elements allow novel polymorphisms to be reported using variant call format (VCF)[9] or an equivalent. VCF is not a required component of a MIRING message, but a VCF component of a MIRING-derived genotype report should use the format described in elements 5.1-5.8. |
| | 5.1 | | Reference | MIRING element 2.2.1 |
| | 5.2 | | Position | The position in the reference sequence (MIRING element 2.2.1) (indexed from 1) corresponding to the position of the reported variant sequence. |
| | 5.3 | | Variant Identifier | A composite value comprised by the CSB identifier including the variant, and a number ranging from 0 to n-1, where n is the number of sequence variants reported, separated by a pipe (e.g. 0\|12). |
| | 5.4 | | Reference Sequence | The sequence in the reference (MIRING element 2.2.1) at the position (MIRING element 5.2). |
| | 5.5 | | Variant Sequence | The variant sequence identified at the position (MIRING element 5.2). This is the equivalent of the VCF ALT column[9]. |
| | 5.6 | | Quality Score | Quality score for the sequence variant reported in MIRING element 5.5. This is the equivalent of the VCF QUAL column[9]. |
| | 5.7 | | Quality Filter Status | PASS or FAIL value for MIRING element 5.6 |
| | 5.8 | | INSDC Accession Number | When possible, provide a GenBank or EMBL-ENA accession number for the novel sequence |

| Number | | Element | Components | Messaging Instructions and Notes |
|---|---|---|---|---|
| 6 | | Platform Documentation | | A peer-reviewed publication, or the identifier of a record deposited in the NCBI GTR or an equivalent resource, documenting the specific details of the methodology and pertinent versions of the platform and instrument-dependent analysis software applied to obtain the unmapped reads and quality scores (MIRING element 8).<br>Relevant platform-dependent information must include: instrument version, instrument-dependent software version identifier(s), reagent versions and lot number, sequence read lengths, expected amplicon/insert length, reference sequences applied, and sequence feature/region targeted.<br>Inclusion of primer target locations is optional. |
| 7 | | Read Processing Documentation | | The specific details of the instrument-independent processing of the primary data (MIRING element 8), documented using the SRA Analysis XSD XML Schema[10] or an equivalent; e.g., instrument-independent analysis software version identifier(s), analysis software parameters used, details of the cutoff values and reference sequences (defined in MIRING element 2) used to filter the data for read quality and/or mapping quality, along with the final read depth obtained and a confidence score of the zygosity for the SNPs used to infer the final genotype. This information is not included in the MIRING message, but must be associated with the primary data (MIRING element 8) by the message generator, and can be accessed using MIRING element 1. |

| Number | | Element | Components | Messaging Instructions and Notes |
|---|---|---|---|---|
| 8 | | Primary Data | | When permitted, unmapped reads with quality scores (e.g., Sanger FASTQ[11] or standard flowgram format (SFF)[12] formatted files), as generated by the instrument (defined in MIRING element 6), must be retained and should be made available as the primary NGS data. Adapter sequences may be excluded from the primary data. These primary data are not included in the MIRING message, but are accessed using MIRING element 1. |

## Legend

MIRING is both a checklist of eight elements that constitute a NGS HLA or KIR genotyping result, and a set of messaging guidelines for transmitting that NGS HLA or KIR genotyping result using five of those elements. Genotyping reports can be generated from a MIRING message. The MIRING guidelines include semantic definitions for a MIRING message, but are not intended to impose syntactic constraints on the message; they are principles that must be met, regardless of the structure of the message.

Elements 1-5 constitute the MIRING message, suitable for reporting a genotyping result. Elements 6-8 constitute the contextual resource for MIRING messages, but are not included in MIRING messages; instead, these elements are referenced in MIRING element 1[13].

Where possible, MIRING elements are consistent with established formats for describing genetic and genomic data (e.g., FASTA[5-7], FASTQ[11], Variant Call Format (VCF)[9] and Genotype List (GL) String formats[4]). In addition, these elements leverage existing genetic and genomic data-resources (e.g., the IMGT/HLA and IPD-KIR Databases[14], the NCBI Genetic Testing Registry (GTR)[15] and International Nucleotide Sequence Database Collaboration (INSDC)[16]).

## Literature Cited


1. 32, I.I.J.S. Information technology — Structure for the identification of organizations and organization parts — Part 1: Identification of organization identification schemes. 1998 [cited 2014 October 26]; Available from: https://www.iso.org/obp/ui/#iso:std:iso-iec:6523:-1:ed-1:v1:en.
2. 32, I.I.J.S. Information technology — Structure for the identification of organizations and organization parts — Part 2: Registration of organization identification schemes. 1998 [cited 2014 October 26]; Available from: https://www.iso.org/obp/ui/#iso:std:iso-iec:6523:-2:ed-1:v1:en.
3. Nakamura, Y., G. Cochrane, and I. Karsch-Mizrachi, The International Nucleotide Sequence Database Collaboration. Nucleic Acids Res, 2013. 41(Database issue): p. D21-4.
4. Milius, R.P., 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.
5. Lipman, D.J. and W.R. Pearson, Rapid and sensitive protein similarity searches. *Science*, 1985. 227(4693): p. 1435-41.
6. Pearson, W.R. and D.J. Lipman, Improved tools for biological sequence comparison. *Proc Natl Acad Sci USA*, 1988. 85(8): p. 2444-8.
7. Neuwald, A.F., et al., Extracting protein alignment models from the sequence database. *Nucleic Acids Research*, 1997. 25: p. 1665-1677.
8. Cornish-Bowden, A., Nomenclature for incompletely specified bases in nucleic acid sequences: recommendations 1984. Nucleic Acids Res, 1985. 13(9): p. 3021-30.
9. Danecek, P., et al., The variant call format and VCFtools. *Bioinformatics*, 2011. 27(15): p. 2156-8.
10. SRA Handbook [Internet], in Analysis Submission Guide, N.C.f.B.I. (US), Editor 2010-: Bethesda (MD).
11. Cock, P.J., et al., The Sanger FASTQ file format for sequences with quality scores, and the Solexa/Illumina FASTQ variants. *Nucleic Acids Res*, 2010. 38(6): p. 1767-71.

12. Malde, K., Flower: extracting information from pyrosequencing data. Bioinformatics, 2011. 27(7): p. 1041-2.
13. Mack S.J., et al., Minimum information for reporting next generation sequence genotyping (MIRING): Guidelines for reporting HLA and KIR genotyping via next generation sequencing. *Hum Immunol*. 2015 Sep 23. pii: S0198-8859(15)00448-6. doi: 10.1016/j.humimm.2015.09.011.
14. Robinson, J., J.A. Halliwell, and S.G.E. Marsh, IMGT/HLA and the Immuno Polymorphism Database. *Methods Mol Biol*, 2014. 1184: p. 109-21.
15. Rubinstein, W.S., et al., The NIH genetic testing registry: a new, centralized database of genetic tests to enable access to comprehensive information and improve transparency. *Nucleic Acids Res*, 2013. 41(Database issue): p. D925-35.
16. Nakamura, Y., G. Cochrane, and I. Karsch-Mizrachi, The International Nucleotide Sequence Database Collaboration. *Nucleic Acids Res*, 2013. 41(Database issue): p. D21-4.