Human Immunology xxx (2015) xxx–xxx

Contents lists available at ScienceDirect

journal homepage: www.elsevier.com/locate/humimm

Minimum information for reporting next generation sequence genotyping (MIRING): Guidelines for reporting HLA and KIR genotyping via next generation sequencing

Steven J. Mack [a,*], Robert P. Milius [b], Benjamin D. Gifford [c], Jürgen Sauter [d], Jan Hofmann [d], Kazutoyo Osoegawa [e], James Robinson [f,g], Mathijs Groeneweg [h], Gregory S. Turenchalk [i], Alex Adai [i], Cherie Holcomb [j], Erik H. Rozemuller [k], Maarten T. Penning [k], Michael L. Heuer [b], Chunlin Wang [l], Marc L. Salit [m,n], Alexander H. Schmidt [d], Peter R. Parham [o], Carlheinz Müller [p], Tim Hague [q], Gottfried Fischer [r], Marcelo Fernandez-Viña [e], Jill A. Hollenbach [s], Paul J. Norman [o], Martin Maiers [b]

[a] Children's Hospital Oakland Research Institute, Oakland, CA, USA
[b] National Marrow Donor Program, Minneapolis, MN, USA
[c] One Lambda, Thermo Fisher Scientific, Brown Deer, WI, USA
[d] DKMS German Bone Marrow Donor Center, Tübingen, Germany
[e] Department of Pathology, Stanford University, Stanford, CA, USA
[f] Anthony Nolan Research Institute, Royal Free Hospital, London, UK
[g] University College London Cancer Institute, University College London, London, UK
[h] Maastricht University Medical Centre, Maastricht, The Netherlands
[i] Bioinformatics, Roche Sequencing, Pleasanton, CA, USA
[j] Roche Molecular Systems, Pleasanton, CA, USA
[k] GenDX, Utrecht, The Netherlands
[l] Stanford Genome Technology Center, Stanford University, Stanford, CA, USA
[m] National Institute of Standards and Technology, Stanford, CA, USA
[n] Department of Bioengineering, Stanford University, Stanford, CA, USA
[o] Department of Structural Biology, Stanford University, Stanford, CA, USA
[p] Zentrales Knochenmarkspender-Register Deutschland, Ulm, Germany
[q] Omixon, Budapest, Hungary
[r] Medical University of Vienna, Vienna, Austria
[s] Department of Neurology, University of California, San Francisco, CA, USA

ARTICLE INFO

Article history:
Received 5 January 2015
Revised 30 August 2015
Accepted 22 September 2015
Available online xxxx

Keywords:
NGS
HLA
KIR

ABSTRACT

The development of next-generation sequencing (NGS) technologies for *HLA* and *KIR* genotyping is rapidly advancing knowledge of genetic variation of these highly polymorphic loci. NGS genotyping is poised to replace older methods for clinical use, but standard methods for reporting and exchanging these new, high quality genotype data are needed. The Immunogenomic NGS Consortium, a broad collaboration of histocompatibility and immunogenetics clinicians, researchers, instrument manufacturers and software developers, has developed the Minimum Information for Reporting Immunogenomic NGS Genotyping (MIRING) reporting guidelines. MIRING is a checklist that specifies the content of NGS genotyping results as well as a set of messaging guidelines for reporting the results. A MIRING message includes five categories of structured information – message annotation, reference context, full genotype,

Abbreviations: BAM, binary alignment/map; CSB, consensus sequence block; dbGAP, Genotype and Phenotype database; EMBL, European Molecular Biology Laboratory; ENA, European Nucleotide Archive; GL, Genotype List; GRC, Genome Reference Consortium; GTR, Genetic Testing Registry; HLA, human leukocyte antigen; HIEDFS, HLA Information Exchange Data Format Standards; HIPAA, Health Insurance Portability and Accountability Act; IEC, International Electrotechnical Commission; IDAWG, Immunogenomic Data Analysis Working Group; IHIW, International HLA and Immunogenetics Workshop; IMGT, immunogenetics; INGSDC, Immunogenomic Next Generation Sequencing Data Consortium; INSDC, International Nucleotide Sequence Database Collaboration; IPD, immuno polymorphism database; ISO, International Organization for Standardization; IUBMB, International Union of Biochemistry and Molecular Biology; IUPAC, International Union of Pure and Applied Chemistry; KIR, killer-cell immunoglobulin-like receptor; MIBBI, Minimum Information for Biological and Biomedical Investigations; MIRING, Minimum Information for Reporting Immunogenomic NGS Genotyping; NCBI, National Center for Biotechnology Information; NGS, next generation sequencing; OID, organization identifier; PIPEDA, Personal Information Protection and Electronic Documents Act; SBT, Sanger sequencing based typing; SFF, standard flowgram format; SRA, Sequence Read Archive; SSOP, sequence-specific oligonucleotide probe; SSP, sequence-specific priming; URI, uniform resource identifier; VCF, variant call format.

* Corresponding author at: Children's Hospital Oakland Research Institute, 5700 Martin Luther King Jr. Way, Oakland, CA 94609, USA.
E-mail address: SJMACK@CHORI.ORG (S.J. Mack).

http://dx.doi.org/10.1016/j.humimm.2015.09.011

MIRING
Genotyping
Data standards

consensus sequence and novel polymorphism – and references to three categories of accessory information – NGS platform documentation, read processing documentation and primary data. These eight categories of information ensure the long-term portability and broad application of this NGS data for all current histocompatibility and immunogenetics use cases. In addition, MIRING can be extended to allow the reporting of genotype data generated using pre-NGS technologies. Because genotyping results reported using MIRING are easily updated in accordance with reference and nomenclature databases, MIRING represents a bold departure from previous methods of reporting *HLA* and *KIR* genotyping results, which have provided static and less-portable data. More information about MIRING can be found online at miring.immunogenomics.org.

1. Introduction

Next-generation sequencing (NGS) offers high-throughput generation of phased sequences for the highly polymorphic human leucocyte antigen (*HLA*) and killer-cell immunoglobulin-like receptor (*KIR*) genes, allowing their rapid, high-resolution genotyping. NGS methods may be more generally described as single-molecule sequencing methods [1]. In some cases, these methods offer full-gene sequence results [1–4]. In general, all NGS methods offer higher resolution and lower ambiguity genotypes than standard methods such as "Sanger" sequencing based typing (SBT), and sequence-specific oligonucleotide probe (SSOP) or primer (SSP) methods [3–6], and do not require the use of secondary genotyping methods to resolve ambiguities.

Any method for genotyping *HLA* and *KIR* using genomic DNA requires at least three components: the *genotyping instrument*, *reference sequences*, and *analysis software*. The *genotyping instrument* generates primary sequence data, which is interpreted by the *analysis software*, using the *reference sequences* to identify the subject's genotype. A wide variety of instruments, reference sequence resources, and data analysis programs are available for both NGS and pre-NGS genotyping approaches, and are used in different combinations.

In some cases, the different methods may not generate the same results for a given subject. Such discrepancies may derive from the instrumentation, reference sequences, software, or a combination of these components. However, as Hollenbach et al. [7] have described, there is no standard format for reporting a genotyping result or for documenting the components that were applied to generate that result. In the absence of such documentation, the source of discrepancies in genotyping results is rarely identifiable. In addition, it becomes impossible to directly relate the *HLA* and *KIR* genotypes of subjects genotyped using different methods, as genetic differences between individuals may not be distinguishable from methodological differences between genotyping approaches. This lack of clarity has important implications for meta-analytical approaches to population or disease association studies that seek to combine and compare data across different studies. In general, ambiguity regarding the source of genotyping discrepancies impedes technical advances and optimization, and frustrates reproducible research.

Guidelines for reporting and documenting genotyping results are essential for evaluating *HLA* and *KIR* genotypes generated using different instruments, reference sequences or data-analysis programs. The active and ongoing development of NGS methods requires the adoption of a single extensible and adaptable standard for reporting and documenting NGS genotyping results.

Here we describe the Minimum Information for Reporting NGS Genotyping (MIRING) checklist, a set of Minimum Information for Biological and Biomedical Investigations (MIBBI) [8,9] reporting guidelines developed by a consortium of immunogenomic researchers and clinicians, NGS instrument manufacturers and software developers, *HLA* and *KIR* sequence database developers and administrators, bone marrow donor registries and donor centers.

2. Description of MIRING

2.1. MIRING development

The standard reporting of *HLA* and *KIR* genotypes is a long unmet need of the histocompatibility and immunogenetics community [7,10–14]. The specific need for NGS genotype reporting guidelines emerged from a survey of immunogenomic data management and analysis practices [15], carried out by the Immunogenomic Data Analysis Working Group (IDAWG) as part of the 16th International HLA and Immunogenetics Workshop (IHIW) [16]. The survey uncovered a lack of consistency between laboratories and the resulting impact on downstream analytical results. The development of MIRING began with the formation of the Immunogenomic Next Generation Sequencing Data Consortium (INGSDC) (ngs.immunogenomics.org) by the IDAWG and the HLA Information Exchange Data Format Standards (HIEDFS) group. The INGSDC met several times between 2012 and 2014, and identified the minimum information needed to accurately report NGS genotyping results for the *HLA* and *KIR* genes for clinical and research applications. Further MIRING development took place as part of the BeTheMatch Foundation's Data Standards 'Hackathons' for NGS-based typing held in September of 2014 and February of 2015 (dash.immunogenomics.org). Implementations of MIRING are being evaluated as part of a 17th IHIW Informatics Component (ihiws.org/informatics-of-genomic-data/) project; bioinformatic tools for generating, exchanging and consuming MIRING messages are being developed as part of this project as well. The participation of interested investigators in this IHIWS project is welcome.

2.2. MIRING goals

To meet the current needs of the histocompatibility and immunogenetics community for reporting and exchanging NGS genotype data, the elements of a MIRING message were designed with the following goals:

1. To facilitate downstream analyses and data management for current research and clinical use cases for molecular genotyping data in the histocompatibility and immunogenetics field.
2. To permit the re-analysis of NGS *HLA* or *KIR* genotyping results in the context of past, present and (foreseeable) future molecular nomenclatures and methods of describing HLA and KIR allele diversity.
3. To permit the comparison and evaluation of genotyping performance between different NGS platforms and analysis methods.

Please cite this article in press as: Mack SJ 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), http://dx.doi.org/10.1016/j.humimm.2015.09.011