Human Immunology 76 (2015) 393–394

Contents lists available at ScienceDirect

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

Introducing a new manuscript format: Enabling access to immunogenomic population data with short population reports

Steven J. Mack *, Derek Middleton

The capacity to replicate the findings of a study is key to the advancement of research, and access to the data on which a study is predicated is required for true replication. Population genetic studies have long been a focus of the immunogenetic research community, but access to the primary genotype data underlying these studies has historically been limited. With the notable exception of the International HLA and Immunogenetics Workshops, it is primarily allele and haplotype frequency data that are made available upon publication for most immunogenetic population studies. However, such frequency data are the products of a prior analysis, and may reflect unreported methodological biases. Access to the primary genotype data allows independent validation and true replication of the study, and knowledge of the ambiguities associated with those genotype data maximizes the utility of population data for replication and meta-analyses.

With this issue of Human Immunology, we are introducing a new manuscript format. Structured descriptions of reference populations, populations of anthropological interest and control populations for disease studies, along with genetic data and minimal analyses, can now be submitted as Short Population Reports. Short Population Reports will be peer-reviewed, and will follow a standard format (described below), with the aim of fostering the availability and archiving of the genetic data underlying immunogenetic population studies.

Human Immunology is partnering with the Allele Frequencies Net Database (AFND; http://www.allelefrequencies.net) to archive and make primary ambiguous genotype data, allele frequency data and haplotype frequency data for the HLA, KIR, cytokine and MIC genes publically accessible, along with demographic data for each population. Details of the AFND data-submission process can be found at http://www.allelefrequencies.net/submit.

While public access to immunogenomic allele and haplotype frequency data has always been part of AFND's mission, the availability of primary ambiguous genotype data is new. These data will be made available on AFND in one of three ways, as determined by the data submitter in accordance with their data-sharing permissions; the primary data will be (i) publically available for download, or (ii) maintained privately on AFND, and made available for analysis by anyone using software that will be provided by AFND or (iii) maintained privately on AFND and will only be analyzed by AFND for data validation purposes. We encourage all authors to seek data-sharing permissions that allow maximal sharing of their primary data.

Genotype, allele and haplotype data deposited in the AFND can be referenced in the corresponding Short Population Report with a unique identifier provided by the AFND.

As described below, Short Population Reports should include only three analyses – evaluation of Hardy–Weinberg equilibrium proportions, for validation of the genotyping data, calculation of allele frequencies, and estimation (or calculation) or haplotype frequencies. Population studies that include additional analyses should be submitted as Research Articles. However, authors are encouraged to cite specific Short Population Reports in Research Articles that present more detailed analyses; this approach fosters more in-depth presentation and discussion of methods and analyses in population studies published as Research Articles.

An example Short Population Report authored by Williams and Middleton is included in this issue.

1. The Short Population Report format

The title of a Short Population Reports should include the name of the population and its geographic region of origin in no more than 150 characters. The body of a Short Population Report should include the following in no more than 1000 words:

(1) An abstract of up to 100 words describing the population and associated genetic data.
(2) Statements regarding the status of informed consent for the collection and use of specimens, ethical use of human subjects in research, and the public availability of the associated data.
(3) A description of the geographic origin of the population, indicating the general region where the samples were collected, and the region to which the population is indigenous if these locations differ.
(4) A brief anthropological and demographic overview of each population's history, including information regarding potential ancestral populations, the history of migrations and any changes in the historical range of the population, and the degree and extent of contact with neighbors or other populations.

* Corresponding author.

http://dx.doi.org/10.1016/j.humimm.2015.03.014

(5) A summary of the languages spoken by the members of the population, along with and any pertinent historical linguistic information. Authors are encouraged to use the Ethnologue language codes (ethnologue.com/codes) to describe linguistic information in a standard fashion.
(6) A summary of any relevant cultural or ethnographic information for the population (e.g., ethnic distinctions, marriage patterns, caste structures).
(7) A description of the methodology employed in obtaining samples, including:
   (a) the rationale for collecting the population sample,
   (b) the rationale for selecting the sites from which the samples were obtained,
   (c) information regarding the degree of relatedness among individuals, and
   (d) information on whether or not data was collected as controls in a disease association study.
(8) A summary of the typing methods used to generate the genotype data for this population, including:
   (a) genotyping system(s) manufacturer and version,
   (b) pertinent reference sequence database version (e.g. the IMGT/HLA or IPD-KIR Database release containing the alleles that the typing methods used can detect),
   (c) gene features (exons, introns and UTRs) interrogated, and
   (d) the rules used to resolve genotyping ambiguity and obtain "allele calls".
(9) The following three types of analyses of the genotype data, the associated methodological documentation, and the unique AFND identifier for data (which is issued to the author by AFND after the data have been checked e.g., AFND000123):
   (a) evaluation of deviation from Hardy–Weinberg expectations for each locus,
   (b) calculation of allele frequencies for each locus, and
   (c) when multi-locus data are presented, estimation of haplotype frequencies (or calculation of haplotype frequencies if phase is known).

   Allele and haplotype frequency tables should be included as Supplementary data.
(10) Up to 10 citations of any previous genetic studies on the population, for both immunogenetic and non-immunogenetic markers.