Minutes of IARC meeting 124, August 9th, 2023

In attendance: Ayelet Peres, Andrew Collins, Martin Corcoran, William Lees, Corey Watson, Mats Ohlin

1. **Approval of minutes of meeting 123**
   Approved

2. **Next meeting**
   August 21st, 2023 at 10.00 UTC

3. **Process for affirmation of germline genes defined by long read sequencing**
   (see Rodriguez et al. (2023) https://doi.org/10.1038/s41467-023-40070-x)

   Discussion of matters of automated gene annotation and the need for manual curation for instance in the case of highly similar (*e.g.* duplicated) genes.

   It was decided that sequences seen more than once in this data (as annotated in VDJbase) will be used to confirm sequences also reported in prior key studies defined in the manuscript. Sequences associated to well-known duplicated genes (like IGHV1-69/69D, IGHV2-70/70D, and the IGHV3-30 -> IGHV3-33 region) will not be assessed in this way with respect to allocation to a particular gene location.

4. **Germline reference set manuscript**

   The paper will be circulated to the AIRR-C when refinement of the sets and text has been finalized.

   Specific matters related to Source and Reference Sets were addressed:

   1. It was decided that when we have a candidate sequence (seen only in one key study) that was not supported by other evidence and thus failed to enter the Reference Set by the standard criteria, it will be accepted if we have confirmation from Rodriguez et al. (2023) (https://doi.org/10.1038/s41467-023-40070-x).
   2. It was decided that when we have a sequence that was not seen in a key study (or Gidoni-VDJbase data), but which has been reported elsewhere

and is an officially-named sequence, it could now enter into the Source and Reference Sets if the sequence was reported from 2 or more individuals in Rodriguez et al. (2023) (https://doi.org/10.1038/s41467-023-40070-x). For instance, Rodriguez et al. reports multiple individuals with IGHD3-10*03, IGHD3-16*03 and IGHD5-18*02. These are recent additions to the IMGT Reference Directory. With this new publication, these sequences could be considered candidates. As they are seen in multiple subjects, it was decided that they will be added into the Source and Reference Sets.

3. It was decided that IGKV3-15*01_A117G and IGKV3-15_G107A, important for one of the manuscript's figures, will be submitted by AP to ENA and assessed by the IARC for inclusion into the Source and Reference Sets.

5. OGRDB

The presentation of germline genes will be updated so as to allow easier visualization of alleles affirmed through the regular IARC process.

6. Sustainability of database resources

Groups at the University of Texas, Simon Fraser University, University of Louisville, Cincinnati Childrens' Hospital, La Jolla Institute and Bar-Ilan University have won a 5-year award from the US NIAID to develop an "Integrated AIRR Knowledge Commons" by integrating and extending related databases. The groups include two run by IARC members: Corey and Gur. The funding will enable continued support and development of VDJbase and OGRDB as well as allowing us to work on extending their reach and utility.