Agenda and Minutes of IARC Meeting 68, April 26, 2021 at 10.00 UTC via Zoom video system.

Present at meeting: Andrew Collins, Martin Corcoran, Corey Watson, William Lees, Ayelet Peres (guest), Mats Ohlin

1. Approval of minutes of meeting 67

   Minutes of meeting 67 were approved

2. Consideration of inference IGHV3-21*01_G278C (VDJbase P4_I27_S1) (continued) that is not inferred in the current version of VDJbase (release 03-2021).

   AP reports that TIgGER modifications resulted in lower noise but also removed some inferences that depend on background data that has little noise. As the data in question are based on peripheral blood B cell transcripts, likely rich in unmutated IgM/IgD-encoding transcripts. They thus carry little mutational noise, likely preventing this inference. The TIgGER process is being updated to balance sensitivity and specificity also when using this type of data.
   MO reports that this allele is well inferred in the present data set using IgDiscover 0.11 and 0.12.
   Decision: The inferences that are currently not featured in VDJbase are put on hold until the sensitivity/specificity balance of TIgGER has been updated. This decision currently affects inference of IGHV3-21*01_G278C, and inferences of IGHV4-30-4*01_A70G_A107G and IGHV4-61*01_A41G (both of which were provisionally assessed in the past (Meetings 62 and 63, respectively)).

3. Final consideration of inference IGHV1-18*01_A196G previously discussed at Meeting 59.

   At the time of meeting 59, focus was to put the processes for approval of inferences from data sets other than P1 as presented in VDJbase in place, rather than on complete documentation of the inferences. Consequently, the inference was not completely assessed at that time although IARC viewed the inference very positively.
   Decision: This inference will be completely reconsidered as the allele has been submitted via the OGRDB submission system.

4. Observation of specific features of inferences in sample P4_I7_S1.
5. Reconsideration of the strategy to name inferred sequences

(discussed together)

   It has been observed (see minutes of Meeting 67 §5) that inference calls in VDJbase may not always be consistent with the principle of using the lowest alphanumeric name for

an inference. For instance, IGHV1-69*05 G54A may be called as IGHV1-69*13 G238A. This may be a consequence of short entries in the starting database used for inference. MO reports that IgDiscover inference results in the same outcome in terms of naming. Similarly, under some conditions IGHV4-4*09 may be called either as IGHV4-4*08_S8301 (a two base extension (G319-A320) of the allele IGHV4-4*09, an entry in the IMGT database that most likely lacks the final two bases of its sequence) and/or IGHV4-59*08_S0184 (a two base extension (in this case: C319-A320) of the allele IGHV4-4*09) if IGHV4-4*09 itself is not present in the database itself. This case highlights both issues with inference of the final bases of a gene and the difficulty associated with proper assignment of names to alleles of duplicated or otherwise highly similar genes (IGHV1-69(D), IGHV2-70(D), IGHV3-30/3-30-3/3-30-5/33, IGHV3-23(D), IGHV4-4/59/61 etc. This matter was further discussed. CW brought up issues with recent official names given to alleles of for instance IGHV2-70(D) as a timely example.

6 Update on the submission process of recent, provisionally affirmed sequences based on VDJbase data

AP informed that sets required for submission to GenBank/SRA are being generated and submission will be performed to OGRDB in the immediate future. It was recognized that it is important to make sure that submission moves through that preferred process and that the more demanding TPA process is avoided.

7. Possible elevation of level 0 sequences that enter into the IMGT data base by routes other than through an IARC review process.

It was agreed that Level 0 sequences assessed by IARC in the past that subsequently enter into the IMGT germline gene database by other means should be elevated to Level 1 based on this fact to publicly highlight the inference in question. At this moment this is the case for IGHV2-70*i01. Discussion on the kind of background data that supports such inferences by IMGT. Such decisions will have to be made after consideration of the data type used to approve the gene/allele. MC pointed out that genomic sequences derived from rearranged lymphoblastoid cell lines may be mutated even though the gene under investigation has not been part of the rearrangement process itself.

8. It was agreed that next meeting's major task will be to assess ideas for an IARC manuscript.

9. Next meeting date: May 18th at 10 AM UTC