IARC Meeting 40: August 7th 2019: minutes

The meeting commenced at 20:00 AEDT. AC, MC, CS, MO and WL were in attendance. Corey Watson and Davide Bagnara joined the meeting as observers.

1. The minutes for Meeting 39 were accepted.
2. WL outlined developments with VDJbase and OGRDB. Two changes of OGRDB are waiting for deployment: using SRA/ENA accession numbers to gather metadata for inclusion in OGRDB; and the reporting to IARC members if sequences that are similar or identical to a submitted inference have already been published. MC made the point that we would need to agree on acceptable processes for the calling of genotypes, if this was to be implemented. The many recent developments of VDJbase were highlighted. Improved reports are now up and running, though bugs are still likely to be in the system and will hopefully be resolved quickly. The problem of aberrant inferences was mentioned, and AC reported that he had recently been in discussion with Gur Yaari on ways to resolve this issue.
3. The status of the OGRDB manuscript (for Nucleic Acids Research database edition) was discussed. In particular, endorsement by the AIRR community was discussed. It was determined that emails have been sent to the AIRR leadership, but not to the membership, seeking support for the endorsement. With submission required by August 15th, AC agreed to write to Nina, suggesting that we will proceed with submission, and we will add AIRR community endorsement (should that occur) during the review and revision period.
4. CS highlighted the fact that OGRDB records many submissions as still being 'under review'. It was agreed that priority would be given to clarifying which submissions were still under review, and which could be finalised. This will be done at the next IARC meeting.
5. DB reported that sequences that were evaluated at early meetings of the IARC, and which have not been entered into OGRDB, will soon be released by Genbank. This will allow the sequences to enter the OGRDB.
6. MC answered a series of questions regarding Submission S00028 before leaving the meeting at 20:45 AEST. In particular, MC confirmed that primers used for amplification of heavy chain rearrangements did not overlap with the sequence that encodes the mature protein. The rules of the IARC required him to leave before the evaluation of this submission, because of his involvement with the studies that generated the sequences.
7. The meeting considered the sequence **IGHV3-7*02_A318G**, which was previously considered by the IARC as part of submission S00003. The committee noted a rearrangement frequency of 0.7%, with 10188 alignments, though only 2443 perfect matches to the inferred allele. Haplotyping based on IGHJ6 alleles confirmed the segregation of reads associated with these alleles to the different haplotypes Alignments were also seen to three other IGHV3-7 sequences, though only one of these sequences was abundant (IGHV3-7*03: 20569 alignments, 4062 unmutated alignments). Alignments to the *01 allele were at a trivial level (<0.01% of 'unmutated' alignments), though alignments to the *02 allele were seen in 0.09% of all unmutated sequences. They likely originated from an introduction of A318 as part of the V-DJ rearrangement process. This represented 3% of

all alignments to the IGHV3-7 gene. There were 933 different CDR3s associated with the inference, including 233 different CDR3s associated with unmutated alignments. All members of the committee who were present agreed that the sequence should be accepted as a valid inference, which would raise this previously recognized Level 0 sequence (IGHV3-7*i01) to Level 2. This will happen when the Genbank record is made publicly available. Before publication of the inference through OGRDB, the committee will seek comment from the submitter regarding the relatively low percentage of unmutated sequences seen in alignments to the IGHV3-7 gene compared to other genes in the database. In line with previous policy, the submitted sequence was recognized up to and including nucleotide 319. A trailing "." in the affirmed sequence as shown in the IARC record will indicate IARC's opinion that the sequence is likely to contain an additional 3' nucleotide for which there is insufficient evidence to make an affirmation. The sequence has been submitted to GenBank and will be available with accession number MN244239 and the Select Set is available in SRA/ENA as SRR9867872. **(NOTE: The analysis reported here was subsequently modified. The sequence was reconsidered at IARC Meeting 43, where the affirmation of the sequence as a Level 2 sequence was reaffirmed. The numbers of alignments shown here differ from the numbers agreed upon at Meeting 43.)**

8. The meeting considered the sequence ***IGHV3-20*01_S2478***. A note will be made in OGRDB noting that at the time of the generation of the S00028 data, the sequence was not in the IMGT reference directory. At the time of evaluation of the submission by the IARC, the sequence has been accepted into the IMGT reference directory as IGHV3-20*04. No further action will therefore be taken by the IARC.
9. The meeting considered the sequence ***IGHV3-21*01_S5868***. The sequence represents IGHV3-21*01 C159T C164A in the IMGT numbering system. The committee noted a rearrangement frequency of 1.5%, with 9064 alignments including 5167 perfect matches to the inferred allele. Alignments were also seen to four other IGHV3-21 sequences, though only one of these sequences was abundant (IGHV3-21*01: 12914 alignments, 8386 unmutated alignments). Haplotyping based on IGHJ6 alleles confirmed the segregation of reads associated with these alleles to the different haplotypes Alignments to the three other alleles shown in the genotype were at trivial levels (<0.01% of 'unmutated' alignments, in each case). There were 1394 different CDR3s associated with the inference, including 614 different CDR3s associated with unmutated alignments. All members of the committee who were present agreed that the sequence should be accepted as a Level 1 sequence as IGHV3-21*i01. In line with previous policy, the submitted sequence was recognized up to and including nucleotide 319. A trailing "." in the affirmed sequence as shown in the IARC record indicates IARC's opinion that the sequence is likely to contain an additional 3' nucleotide for which there is insufficient evidence to make an affirmation. The sequence has been submitted to GenBank and is available with accession number MK308859 and the Select Set is available in SRA/ENA as SRR9842443. **(NOTE: The analysis reported here was subsequently modified. The sequence was reconsidered at IARC Meeting 43, where the affirmation of the sequence as a Level 1 sequence was reaffirmed. The**

numbers of alignments shown here differ from the numbers agreed upon at Meeting 43.)

10. The meeting considered the sequence ***IGHV3-53*02_S0744.*** The sequence represents IGHV3-53*02 C288T in the IMGT numbering system. The committee noted a rearrangement frequency of 0.7%, with 5149 alignments including 2426 perfect matches to the inferred allele. Alignments were also seen to three other IGHV3-53 sequences, though only one of these sequences was abundant (IGHV3-53*04: 3448 alignments, 2149 unmutated alignments). Haplotyping based on IGHJ6 alleles confirmed the segregation of reads associated with these alleles to the different haplotypes Alignments to the two other alleles shown in the genotype were at trivial levels (<0.01% of 'unmutated' alignments, in each case). There were 742 different CDR3s associated with the inference, including 292 different CDR3s associated with unmutated alignments. All members of the committee who were present agreed that the sequence should be accepted as a Level 1 sequence as IGHV3-53*i01. In line with previous policy, the submitted sequence was recognized up to and including nucleotide 319. A trailing "." in the affirmed sequence as shown in the IARC record indicates IARC's opinion that the sequence is likely to contain an additional 3' nucleotide for which there is insufficient evidence to make an affirmation. The sequence has been submitted to GenBank and is available with accession number MK308860 and the Select Set is available in SRA/ENA as SRR9842442. **(NOTE: The analysis reported here was subsequently modified. The sequence was reconsidered at IARC Meeting 43, where the affirmation of the sequence as a Level 1 sequence was reaffirmed. The numbers of alignments shown here differ from the numbers agreed upon at Meeting 43.)**

11. MO pointed out that there has still been no resolution of issues regarding the nucleotide 320 in sequences accepted into the IMGT reference directory, nor of the problem regarding the shortening of the sequence IGHV1-3*i01 (IGHV1-3*04 as defined by IMGT), caused by an error in the original GenBank record associated with this inference, an issue that was resolved by GenBank as of May 6, 2019. AC will re-visit these issues with IMGT when he notifies them of the latest IGHV gene affirmations.
12. The next meeting (Meeting 41) will be on Wednesday August 21st at 20:00 AEST. CS will send out electronic meeting invites.

The meeting ended at 21:20 AEST.