IARC Meeting 9: March 28th 2018: minutes

The meeting commenced at 22:00 AEST. AC, MC and MO were in attendance. CS joined the meeting at 22:40 AEST.

1. An exchange of emails between AC and Marie-Paule Lefranc was made available to the committee. The emails concerned the IMGT position regarding inferences that could allow a truncated sequence, within the accepted repertoire, to be extended. The committee will further explore the issue, in subsequent weeks, by submitting information to IMGT based upon IARC consideration of data from Davide Bagnara. IGHV4-4*01_S5769 (IgDiscover nomenclature) from B16 was highlighted as a candidate sequence for this purpose.
2. The committee was informed of an email from AC to Steven Kleinstein, discussing the fact that TIgGER by default applies its algorithm to nucleotides up to 312. A response from SK has not yet been received.
3. The committee considered
   >IGHV1-69*14_S3451 (G163A) from B12
   CAGGTCCAGCTGGTGCAGTCTGGGGCTGAGGTGAAGAAGCCTGGGTCCTCGGTGAAG
   GTCTCCTGCAAGGCTTCTGGAGGCACCTTCAGCAGCTATGCTATCAGCTGGGTGCGAC
   AGGCCCCTGGACAAGGGCTTGAGTGGATGGGAAGGATCATCCCTATCTTTGGTACAGC
   AAACTACGCACAGAAGTTCCAGGGCAGAGTCACGATTACCGCGGACAAATCCACGAGC
   ACAGCCTACATGGAGCTGAGCAGCCTGAGATCTGAGGACACGGCCGTGTATTACTGTG
   CGAGAG
   The allele was present at high frequency (2.3% of exact matches) in the IgDiscover analysis, and defined as the fourth most abundantly expressed allele of IGHV1-69/69D. The gene was also inferred by TIgGER but not by partis. CDR3-based cross-over analysis suggested caution was required, and haplotype analysis indicated 3 alleles on the haplotype carrying IGHV1-69*14_S3451. There was strong support for the recognition of the sequence, but given the complexity of the haplotype of B12, with an apparent large gene duplication, the committee decided the sequence should be moved to Level 0. It was also noted that this sequence is identical to the partial sequence IGHV1-69*p29 in IgPdb. IGHV1-69*p29 is missing the first 28 nucleotides, including a critical nucleotide that distinguishes IGHV1-69*06 from IGHV1-69*14.
4. The committee considered
   >IGHV2-70*01_S4660 (A124G) from B12
   CAGGTCACCTTGAGGGAGTCTGGTCCTGCGCTGGTGAAACCCACACAGACCCTCACAC
   TGACCTGCACCTTCTCTGGGTTCTCACTCAGCACTAGTGGAATGTGTGTGAGCTGGGT
   CCGTCAGCCCCCAGGGAAGGCCCTGGAGTGGCTTGCACTCATTGATTGGGATGATGAT
   AAATACTACAGCACATCTCTGAAGACCAGGCTCACCATCTCCAAGGACACCTCCAAAAA
   CCAGGTGGTCCTTACAATGACCAACATGGACCCTGTGGACACAGCCACGTATTACTGT
   GCACGGA
   The sequence is present at very low frequency (0.04%) in the IgDiscover analysis, and the committee initially thought the sequence should be rejected considering the provisional (but arbitrary) cut-off defined for allowed inferences. TIgGER did not infer any new allele of this

gene while partis did not infer this allele but inferred a different new IGHV2-70 allele. Haplotype analysis is however supportive of this inference, and CDR3-based cross-over analysis is not strongly supportive, but not immediately prohibitive given the status of this sample in general. It was noted that this sequence is identical to IGHV2-70*p16 in IgPdb. The sequence was therefore moved to Level 0.

5. The committee considered
   >IGHV1-58*01_S8523 (G61T) from B16
   CAAATGCAGCTGGTGCAGTCTGGGCCTGAGGTGAAGAAGCCTGGGACCTCAGTGAAG
   TTCTCCTGCAAGGCTTCTGGATTCACCTTTACTAGCTCTGCTATGCAGTGGGTGCGACA
   GGCTCGTGGACAACGCCTTGAGTGGATAGGATGGATCGTCGTTGGCAGTGGTAACACA
   AACTACGCACAGAAGTTCCAGGAAAGAGTCACCATTACCAGGGACATGTCCACAAGCA
   CAGCCTACATGGAGCTGAGCAGCCTGAGATCCGAGGACACGGCCGTGTATTACTGTG
   CGGCAG
   Like IGHV2-70*01_S4660 from B12, this sequence is present at very low frequency (0.02%) in the IgDiscover analysis. It is also represented by few exact matches (exact: n=71; CDR3s_exact: n=22). Very few reads can be used for haplotype analysis, which consequently is inconclusive. CDR3-based crossover analysis shows poor values but similar to those of many genes of this particular sample. The sequence was not inferred by TIgGER or partis. The sequence is identical to IGHV1-58*p03 in IgPdb. It was moved to Level 0.
6. It was suggested that a separate list of sequences might be developed for sequences that are present at very low frequencies, but that otherwise meet all the requirements for acceptance of inferences.
7. The committee discussed whether or not consideration of a submitted sequence could be informed by additional evidence that is independent of the submission. It was suggested the such evidence might be used to ‘lower’ flags that were raised to suggest some concern with a sequence. It was pointed out that this other evidence could also be subject to similar systematic errors as inferences. As an obvious source of additional evidence is IgPdb, it was decided that the next meeting should focus on whether or not IgPdb evidence can be used by IARC. It was pointed out that if other evidence, including IgPdb submissions, is to be used in support of an inference, such information must be quality controlled just as inferences themselves must be assessed. This will be done by continuing deliberations regarding the three sequences that were designated as Level 0, at this meeting
8. The next meeting will be on Friday April 6th at 21:00 AET.