

## IARC Meeting 49: February 4th 2020: minutes

The meeting commenced at 10:30 UTC. AC, CS, MO, and WL were in attendance. MC attended the start of the meeting, and then excused himself as the meeting was called to assess inferences with which he has been associated.

1. The minutes of meetings 47 and 48 were approved.
2. The challenges of light chain inferences were discussed. The consequences of ambiguities at the 3' end of IGKV and IGLV were discussed, and it was proposed that the Light Chain sub-group of the GLDB WG and the Software WG be asked to give the issue some attention. The implications of the existence of the proximal and distal IGKV gene sets were discussed. It was agreed that the naming of affirmed IGKV by the IARC will be made with reference to the most similar previously-reported sequence, with the acknowledgement that in the absence of mapping studies, the location of inferred sequences in either the proximal or distal cluster will be uncertain. It was noted that haplotyping of kappa and lambda genes will rarely be possible. It was also noted that light chain datasets are likely to include relatively high levels of somatic point mutations, unless sorted naive cells have been used for library preparation. Finally it was noted that instructions for the preparation of select sets may require modification for light chain genes, as the current recommendations (for IGHV) suggests that bases 1-312 should be conserved to >99.6%. In the case of light chain, a longer range will need to be stated.
3. The committee considered IGKV1D-8\*01\_S8780 (C59T) of Genotype A007 VK, Submission S00028. The submitted IGKV1D-8\*01\_S8780 sequence is as follows:  
GTCATCTGGATGACCCAGTCTCCATCCTTACTCTCTGCATCTACAGGAGACAGAGTCAT  
CATCAGTTGTCGGATGAGTCAGGGCATTAGCAGTTATTTAGCCTGGTATCAGCAAAAAC  
CAGGGAAAGCCCCTGAGCTCCTGATCTATGCTGCATCCACTTTGCAAAGTGGGGTCCC  
ATCAAGGTTTCAGTGGCAGTGGATCTGGGACAGATTTCACTCTCACCATCAGTTGCCTG  
CAGTCTGAAGATTTTGC AACTTATTACTGTCAACAGTATTATAGTTTTCCCTCC  
The sequence was seen in 0.3% of all unmutated rearrangements, with 747 sequences including 538 perfect matches to the inferred allele. There was abundant variation in the CDR3 regions of the aligned sequences, though relatively large numbers of sequences were clonally related (42 "Unique CDR3s with unmutated"). The IGKV1D-8\*01 allele was also present in the genotype, at a similar frequency (0.4% of all unmutated sequences; 976 sequences; 758 unmutated sequences). It was noted that the select set available in SRA/ENA is very limited in size (essentially only representing 2 different rearrangements, either involving IGKJ1 or IGKJ2) as only full-length sequences (covering all bases of the V gene up to and including base 335) had been included in the submitted set. IGKV1-8\*01 was also present at high frequency (7743 sequences), but this sequence has 12 nucleotide differences to the inference under consideration. Haplotype data is unavailable. Plots of the final 3' nucleotides showed a high level of variability, making it impossible to determine the final three nucleotides with certainty. The sequence, up to and including nucleotide 332, was

affirmed as a Level 1 sequence. Uncertainty regarding nucleotides 333-335 will be indicated in IARC and OGRDB publications by three dots at the end of the affirmed sequence.

4. The committee made a preliminary assessment of IGLV inferences in submission S00028. Discussion particularly focused upon 'Nucleotide difference' plots (available through Submission Notes). It was noted that many plots, of inferred sequences and previously-reported sequences, include prominent peaks for 'highly mutated' sequences and there was evidence that some inferences rather represented two different reported alleles. As the plots are not bell-curves but sharp peaks, they can not be easily explained by clonal expansions. AC will seek an explanation from the submitters, and will write to Martin Corcoran to ask whether the general lack of variability in light chain CDR3 regions could complicate IgDiscover analysis.
5. The next meeting (Meeting 50) will again be devoted to the consideration of light chain inferences, and will be held on February 24th at 10:00 UTC.

The meeting ended at 11:30 UTC.