

Agenda of IARC meeting 80, September 27th, 2021

In attendance: Corey Watson, Ayelet Peres, Gur Yaari, Andrew Collins, Martin Corcoran, William Lees, James Heather (guest), Mats Ohlin

1. Approval of minutes of meeting 79

Postponed until next meeting

2. Update on manuscript submitted to Genes and Immunity

The manuscript has been provisionally accepted for publication.

3. TCR inference as an IARC activity and status of gene submission

The main focus is on participating in and learning from the current IARC's processes and assessment discussions.

4. Assessment of inference IGHV3-21*01_A184G_T190A_A191C

IGHV3-21*01_A184G_T190A_A191C was inferred in the genotype of subject S76 (VDJbase: P1_I80). This inference has previously been pre-assessed at IARC meeting 61

(https://www.antibodysociety.org/wordpress/wp-content/uploads/2020/12/Meeting-61-13_10_20-minutes.pdf). The genotype also carried IGHV3-21*01. Both alleles were represented in high numbers (2.9% and 2.5% of the total unmutated population). IGHV3-21*01_A184G_T190A_A191 was represented by 1306 sequences, 1092 unmutated sequences, an allelic frequency of 53%, and 1069 unique CDR3s in the unmutated sequence set. Haplotyping based on differences in IGHJ6 alleles was not possible for this dataset.

>IGHV3-21*01_A184G_T190A_A191C

```
GAGGTGCAGCTGGTGGAGTCTGGGGGAGGCCTGGTCAAGCCTGGGGG
GTCCCTGAGACTCTCTGTGCAGCCTCTGGATTCACCTTCAGTAGCTATA
GCATGAACTGGGTCCGCCAGGCTCCAGGGAAGGGGCTGGAGTGGGTCT
CATCCATTAGTAGTAGTGGTAGTACCATATACTACGCAGACTCAGTGAAGG
GCCGATTCACCATCTCCAGAGACAACGCCAAGAACTCACTGTATCTGCAA
ATGAACAGCCTGAGAGCCGAGGACACGGCTGTGTATTACTGTGCGAGAG
A
```

The sequence is inferred at Level 1. IARC infers the sequence up to and including base 320 based on the convincing evidence of the sequence end as presented in the OGRDBplot.

XX, xx, as data generators/submitters, abstained from participating in the decision.

5. The 3'end problem in inference.

The final decision on inference of IGHV3-21*01_A184G_T190A_A191C was preceded by an extensive discussion on handling of 3'-ends (see DOI: 10.3389/fimmu.2019.00435: "If a sequence is reported up to and including base 320, the final base will only be affirmed by IARC if supporting analysis is provided.") and the advantages and disadvantages of inferring a full sequence. It was considered that the convincing data presented in OGRDBplots for IGHV3-21*01_A184G_T190A_A191C and the precedent for A as base 320 in alleles of IGHV3-21 provided such supporting information.

6. The problem of identical sequences associated to different/duplicated genes in germline gene databases and the impact of multiple assignments in sequence analysis and VDJbase's representation of genotypes/haplotypes

AP presented a tool for assessment of, and actual analysis of identical sequences found in duplicated IGHV genes and highly similar genes. It was decided that analysis should be carried out using a starting database where such alleles had been merged into single database entries.

7. Next meeting

October 4th 2021 at 10.00 UTC