Agenda of IARC meeting 82, October 11th, 2021

In attendance: Ayelet Peres, Gur Yaari, Andrew Collins, Martin Corcoran, Mats Ohlin, James Heather (Guest)

1. **Approval of minutes of meeting 81**
   Approved

2. **Reflection on learning of IARC's allele approval process**
   JH reflected on the process.

3. **Further elaboration on details of inference of IGHV1-69*04_C184T**
   During our pre-assessment of IGHV1-69*04_C184T (IARC meeting 60) we assessed this inference in sample VDJbase P1_I77_S1 and not in P1_I50_S1, the sample that was used for assessment of the inference of IGHV1-69*i03 at IARC meeting 81. The current version of VDJbase (Version: Pipeline 25/9/20) does not infer this allele in P1_I77_S1. Instead it infers IGHV1-69*04 (as well as IGHV1-69*02). All these alleles are highly similar.

However, all reads associated to IGHV1-69*04 in P1_I77_S1 are mutated, in line with the presence of a variant of IGHV1-69*04 in this subject.

Current OGRDBplots data defining IGHV1-69*04 in VDJbase (Version: Pipeline 25/9/20) P1_I77_S1
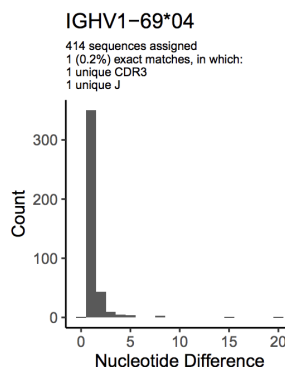
Separate inference by IgDiscover 0.12.2 had allowed for identification of IGHV1-69*04_C184T in P1_I50_S1 (ERR2567224) as well as in P1_I77_S1 (ERR2567250) (DOI: 10.3389/fimmu.2021.730105). Altogether this demonstrates that the original pre-assessment of inference IGHV1-69*01_C184T in sample P1_I77_S1 (assessed during IARC meeting 60) was valid.

4. **Inference of IGHV3-30*19_T189C**

IGHV3-30*19_T189C was inferred in one genotype (VDJbase P1_I49). 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). Among alleles of the IGHV3-30/IGHV3-30-3/IGHV3-33 set the genotype in addition also carried IGHV3-30-3*01 and IGHV3-33*01 as defined in OGRDB. The three alleles are supported by similar numbers of sequences (357, 141, and 540), multiple (116, 59, 200) unique CDR3s among unmutated reads. Haplotyping based on alleles of IGHJ6 supported its presence (IGHV3-30*19_T189C ratio: 100:0; IGHV3-30-3*01: 0:100; IGHV3-33*01: 56:44). The sequences are quite similar and theoretically IGHV3-30*19_T189C could have been generated as a chimeric PCR products of rearranged sequences derived from the 5'-part IGHV3-33*01 and the 3'-part of IGHV3-30-3*01. However, haplotyping based on IGHJ6 associates IGHV3-30*19_G189C and IGHV3-30-3*01 to different alleles of IGHJ6. Furthermore, a separate study of this data set (DOI: 10.3389/fimmu.2021.730105) defined that the upstream sequences of IGHV3-30*19_T189C and IGHV3-33*01 are substantially different. IGHV3-30*18 have also been found associated to both haplotypes of this genotype but it will currently not feature in OGRDB as two alleles have the same sequence, preventing their inclusion in the genotype. In any event, this allele cannot easily have contributed (by being involved in PCR chimeras) to the inference of IGHV3-30*19_T189C as it differs from the novel allele both in the allele's 5'- and 3' end. In addition, its upstream sequence is different from that of IGHV3-30*19_T189C. Altogether there is no evidence that IGHV3-30*19_T189C has been generated as a PCR chimera. IARC infers the sequence at Level 1 up to and including base 319 in agreement with past practice. It is acknowledged that the allele most likely carries one additional base, typically A at base position 320. A trailing "." 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 allele is given the name IGHV3-30*i01. We recognize that this allele might be located at IGHV3-30, IGHV3-30-3, IGHV3-30-5, and/or IGHV3-33 and IARC gene naming does not reflect a position on this matter.

>IGHV3-30*19_T189C
CAGGTGCAGCTGGTGGAGTCTGGGGGAGGCGTGGTCCAGCCTGGGAG
GTCCCTGAGACTCTCCTGTGCAGCGTCTGGATTCACCTTCAGTAGCTATG
GCATGCACTGGGTCCGCCAGGCTCCAGGCAAGGGGCTGGAGTGGGTG
GCAGTTATATCATATGATGGAAGCAATAAATACTACGCAGACTCCGTGAAG
GGCCGATTCACCATCTCCAGAGACAATTCCAAGAACACGCTGTATCTGCA
AATGAACAGCCTGAGAGCTGAGGACACGGCTGTGTATTACTGTGCGAGA
G.

Coding regions of genes identified in sample P1__I49:

```
                      10        20        30        40        50        60        70
IGHV3-30*19_T189C  CAGGTGCAGCTGGTGGAGTCTGGGGGAGGCGTGGTCCAGCCTGGGAGGTCCCTGAGACTCTCCTGTGCAG
IGHV3-33*01        CAGGTGCAGCTGGTGGAGTCTGGGGGAGGCGTGGTCCAGCCTGGGAGGTCCCTGAGACTCTCCTGTGCAG
IGHV3-30-3*01      CAGGTGCAGCTGGTGGAGTCTGGGGGAGGCGTGGTCCAGCCTGGGAGGTCCCTGAGACTCTCCTGTGCAG
IGHV3-30*18        CAGGTGCAGCTGGTGGAGTCTGGGGGAGGCGTGGTCCAGCCTGGGAGGTCCCTGAGACTCTCCTGTGCAG

                      80        90       100       110       120       130       140
IGHV3-30*19_T189C  CGTCTGGATTCACCTTCAGTAGCTATGGCATGCACTGGGTCCGCCAGGCTCCAGGCAAGGGGCTGGAGTG
IGHV3-33*01        CGTCTGGATTCACCTTCAGTAGCTATGGCATGCACTGGGTCCGCCAGGCTCCAGGCAAGGGGCTGGAGTG
IGHV3-30-3*01      CCTCTGGATTCACCTTCAGTAGCTATGCTATGCACTGGGTCCGCCAGGCTCCAGGCAAGGGGCTGGAGTG
IGHV3-30*18        CCTCTGGATTCACCTTCAGTAGCTATGGCATGCACTGGGTCCGCCAGGCTCCAGGCAAGGGGCTGGAGTG

                     150       160       170       180       190       200       210
IGHV3-30*19_T189C  GGTGGCAGTTATATCATATGATGGAAGCAATAAATACTACGCAGACTCCGTGAAGGGCCGATTCACCATC
IGHV3-33*01        GGTGGCAGTTATATGGTATGATGGAAGTAATAAATACTATGCAGACTCCGTGAAGGGCCGATTCACCATC
IGHV3-30-3*01      GGTGGCAGTTATATCATATGATGGAAGCAATAAATACTACGCAGACTCCGTGAAGGGCCGATTCACCATC
IGHV3-30*18        GGTGGCAGTTATATCATATGATGGAAGTAATAAATACTATGCAGACTCCGTGAAGGGCCGATTCACCATC

                     220       230       240       250       260       270       280
IGHV3-30*19_T189C  TCCAGAGACAATTCCAAGAACACGCTGTATCTGCAAATGAACAGCCTGAGAGCTGAGGACACGGCTGTGT
IGHV3-33*01        TCCAGAGACAATTCCAAGAACACGCTGTATCTGCAAATGAACAGCCTGAGAGCCGAGGACACGGCTGTGT
IGHV3-30-3*01      TCCAGAGACAATTCCAAGAACACGCTGTATCTGCAAATGAACAGCCTGAGAGCTGAGGACACGGCTGTGT
IGHV3-30*18        TCCAGAGACAATTCCAAGAACACGCTGTATCTGCAAATGAACAGCCTGAGAGCTGAGGACACGGCTGTGT

                     290
IGHV3-30*19_T189C  ATTACTGTGCGAGAGA
IGHV3-33*01        ATTACTGTGCGAGAGA
IGHV3-30-3*01      ATTACTGTGCGAGAGA
IGHV3-30*18        ATTACTGTGCGAAAGA
```

Upstream sequence of alleles of IGHV3-30/3-30-3/3-33 in samples of data set P1 of VDJbase (DOI: 10.3389/fimmu.2021.730105) (IGHV3-30*19_T189C is named IGHV3-30*19_S5956 in this panel):

```
                          10        20        30        40        50        60        70
IGHV3-30*19_S5956-A|1  AGCTCTGGGAGACGAGCCCAGCACTGGAAGTCGCCGGTGTTTCCATTCGGTGATCATCACTGAACACAGA
IGHV3-30-3*01-A|69     AGCTCTGGGAGACGAGCCCAGCACTGGAAGTCGCCGGTGTTTCCATTCGGTGATCATCACTGAACACAGA
IGHV3-30*04-A|6        AGCTCTGGGAGACGAGCCCAGCACTGGAAGTCGCCGGTGTTTCCATTCGGTGATCATCACTGAACACAGA
IGHV3-30*04-B|12       AGCTCTGGGAGACGAGCCCAGCACTGGAAGTCGCCGGTGTTTCCATTCGGTGATCAGCACTGAACACAGA
IGHV3-30*01-A|11       AGCTCTGGGAGAGGAGCCCAGCACTGGAAGTCGCCGGTGTTTCCATTCGGTGATCAGCACTGAACACAGA
IGHV3-30*02-A|3        AGCTCTGGGAGAGGAGCCCAGCACTGGAAGTCGCCGGTGTTTCCATTCGGTGATCATCACTGAACACAGA
IGHV3-30*02-B|6        AGCTCTGGGAGAGGAGCCCAGCACTAGAAGTCGGCGGTGTTTCCATTCGGTGATCAGCACTGAACACAGA
IGHV3-30*02_S4989-A|1  AGCTCTGGGAGAGGAGCCCAGCACTAGAAGTCGGCGGTGTTTCCATTCGGTGATCAGCACTGAACACAGA
IGHV3-30*03-A|5        AGCTCTGGGAGAGGAGCCCAGCACTAGAAGTCGGCGGTGTTTCCATTCGGTGATCAGCACTGAACACAGA
IGHV3-30*04_S7005-A|1  AGCTCTGGGAGAGGAGCCCAGCACTAGAAGTCGGCGGTGTTTCCATTCGGTGATCAGCACTGAACACAGA
IGHV3-30*18-A|92       AGCTCTGGGAGAGGAGCCCAGCACTAGAAGTCGGCGGTGTTTCCATTCGGTGATCAGCACTGAACACAGA
IGHV3-33*01-A|96       AGCTCTGGGAGAGGAGCCCAGCACTAGAAGTCGGCGGTGTTTCCATTCGGTGATCAGCACTGAACACAGA
IGHV3-33*01_S3418-A|1  AGCTCTGGGAGAGGAGCCCAGCACTAGAAGTCGGCGGTGTTTCCATTCGGTGATCAGCACTGAACACAGA

                          80        90       100       110       120       130
IGHV3-30*19_S5956-A|1  GGACTCACCATGGAGTTTGGGCTGAGCTGGGTTTTCCTCGTTGCTCTTTTAAGAGGTGTCCAGTGT
IGHV3-30-3*01-A|69     GGACTCACCATGGAGTTTGGGCTGAGCTGGGTTTTCCTCGTTGCTCTTTTAAGAGGTGTCCAGTGT
IGHV3-30*04-A|6        GGACTCACCATGGAGTTTGGGCTGAGCTGGGTTTTCCTCGTTGCTCTTTTAAGAGGTGTCCAGTGT
IGHV3-30*04-B|12       GGACTCACCATGGAGTTTGGGCTGAGCTGGGTTTTCCTCGTTGCTCTTTTAAGAGGTGTCCAGTGT
IGHV3-30*01-A|11       GGACTCACCATGGAGTTTGGGCTGAGCTGGGTTTTCCTCGTTGCTCTTTTAAGAGGTGTCCAGTGT
IGHV3-30*02-A|3        GGACTCACCATGGAGTTTGGGCTGAGCTGGGTTTTCCTCGTTGCTCTTTTAAGAGGTGTCCAGTGT
IGHV3-30*02-B|6        GGACTCACCATGGAGTTTGGGCTGAGCTGGGTTTTCCTCGTTGCTCTTTTAAGAGGTGTCCAGTGT
IGHV3-30*02_S4989-A|1  GGACTCACCATGGAGTTTGGGCTGAGCTGGGTTTTCCTCGTTGCTCTTTTAAGAGGTGTCCAGTGT
IGHV3-30*03-A|5        GGACTCACCATGGAGTTTGGGCTGAGCTGGGTTTTCCTCGTTGCTCTTTTAAGAGGTGTCCAGTGT
IGHV3-30*04_S7005-A|1  GGACTCACCATGGAGTTTGGGCTGAGCTGGGTTTTCCTCGTTGCTCTTTTAAGAGGTGTCCAGTGT
IGHV3-30*18-A|92       GGACTCACCATGGAGTTTGGGCTGAGCTGGGTTTTCCTCGTTGCTCTTTTAAGAGGTGTCCAGTGT
IGHV3-33*01-A|96       GGACTCACCATGGAGTTTGGGCTGAGCTGGGTTTTCCTCGTTGCTCTTTTAAGAGGTGTCCAGTGT
IGHV3-33*01_S3418-A|1  GGACTCACCATGGAGTTTGGGCTGAGCTGGGTTTTCCTCGTTGCTCTTTTAAGAGGTGTCCAGTGT
```

OGRDBplots illustration of bases in rearrangements
associated to IGHV3-30*19_T189C:

Gene IGHV3−30*19_T189C
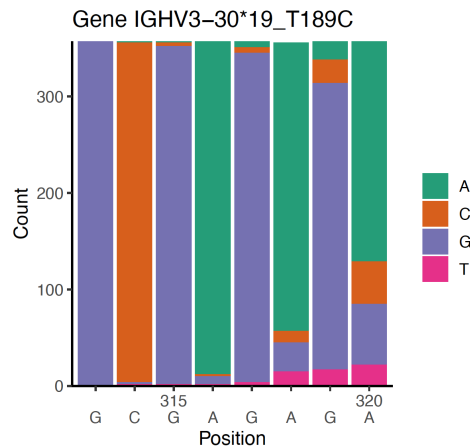
5. **Followup: The problem of identical sequences associated to
   different/duplicated genes in germline gene databases and the impact of
   multiple assignments in sequence analysis and VDJbase's
   representation of genotypes/haplotypes**
   Presentation of strategies and discussion of ways forward.

6. **Next meeting**
   October 18, 2021 at 10.00 UTC