Agenda of IARC meeting 83, October 18th, 2021

In attendance: Ayelet Peres, Gur Yaari, Andrew Collins, Martin Corcoran, Corey Watson, William Lees, Mats Ohlin

1. **Approval of minutes of meeting 82**
   Approved

2. **Inference of IGHV3-33*01_G75C**
   IGHV3-33*01_G75C was inferred in a genotype (VDJbase P1_I46). This inference has previously been pre-assessed at IARC meeting 62 (https://www.antibodysociety.org/wordpress/wp-content/uploads/2021/04/Meeting-62-3_11_20-minutes.pdf). Among alleles of the IGHV3-30/IGHV3-30-3/IGHV3-33 set the genotype in addition only carried IGHV3-33*01 as defined by OGRDB. The two alleles were supported by similar numbers of sequences (allelic frequency 46%). IGHV3-33*01_G75C was associated to multiple (388) unique CDR3s. The genotype of this subject as defined by VDJbase also features IGHV3-30*18/IGHV3-30-5*01. IgDiscover-based analysis has also identified IGHV3-30*18 in this genotype. This lack of definition of this allele in the haplotyped data in VDJbase is being addressed. Haplotyping based on alleles of IGHJ6 supported its presence (ratio: 99:1 (IGHV3-33*01) and 2:98 (IGHV3-33*01_G75C), respectively). IGHV3-33*01_G75C is identical to IGHV3-33*08 that entered into the IMGT-DB on March 5th, 2021 based on its presence in GenBank entry accession number AC279998. IARC infers, based on information in the transcriptome of this subject, at level 1 the sequence 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. The allele is given the name IGHV3-33*i01. We recognize that this allele might have been located at IGHV3-30, IGHV3-30-3, IGHV3-30-5, and/or IGHV3-33 in this subject and IARC gene naming does not reflect a position on this matter.

   >IGHV3-33*i01
   CAGGTGCAGCTGGTGGAGTCTGGGGGAGGCGTGGTCCAGCCTGGG
   AGGTCCCTGAGACTCTCCTGTGCAGCCTCTGGATTCACCTTCAGTAG
   CTATGGCATGCACTGGGTCCGCCAGGCTCCAGGCAAGGGGCTGGAG
   TGGGTGGCAGTTATATGGTATGATGGAAGTAATAAATACTATGCAGACT
   CCGTGAAGGGCCGATTCACCATCTCCAGAGACAATTCCAAGAACACG
   CTGTATCTGCAAATGAACAGCCTGAGAGCCGAGGACACGGCTGTGTA
   TTACTGTGCGAGAG.

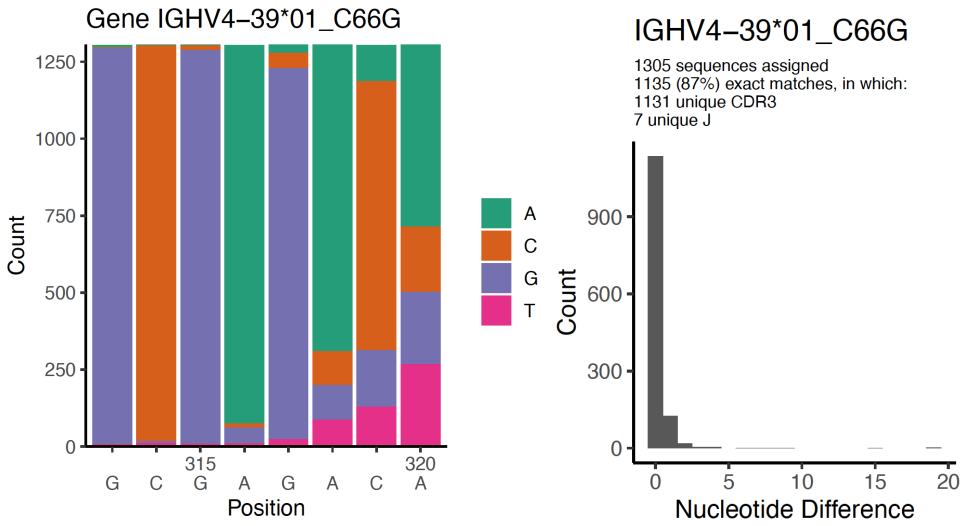
3. **Inference of IGHV3-9*01_T307C**
   IGHV3-9*01_T307C was inferred in subject S82 (P1_I86_S1; ERR2567259), a genotype that also expressed IGHV3-9*01. This inference has previously been pre-assessed at IARC meeting 62 (https://www.antibodysociety.org/wordpress/wp-content/uploads/2021/04/Meeting-62-3_11_20-minutes.pdf). IGHV3-9*01_T307C was well expressed, represented by 453 sequences, allelic frequency 36%, represented 1.6% of the unmutated population, and had 373 unique CDR3s in the unmutated population. It showed a haplotype ratio (based on association with alleles of IGHJ6) of 100:0 while IGHV3-9*01 showed a ratio of 1:99. IARC infers the sequence at level 1 up to and including base 319. It is acknowledged that the allele most likely carries 3 additional bases, typically ATA at base position 320-322. The distribution of bases at each of these positions is such that inference of them cannot be achieved.The allele is given the name IGHV3-9*i01.

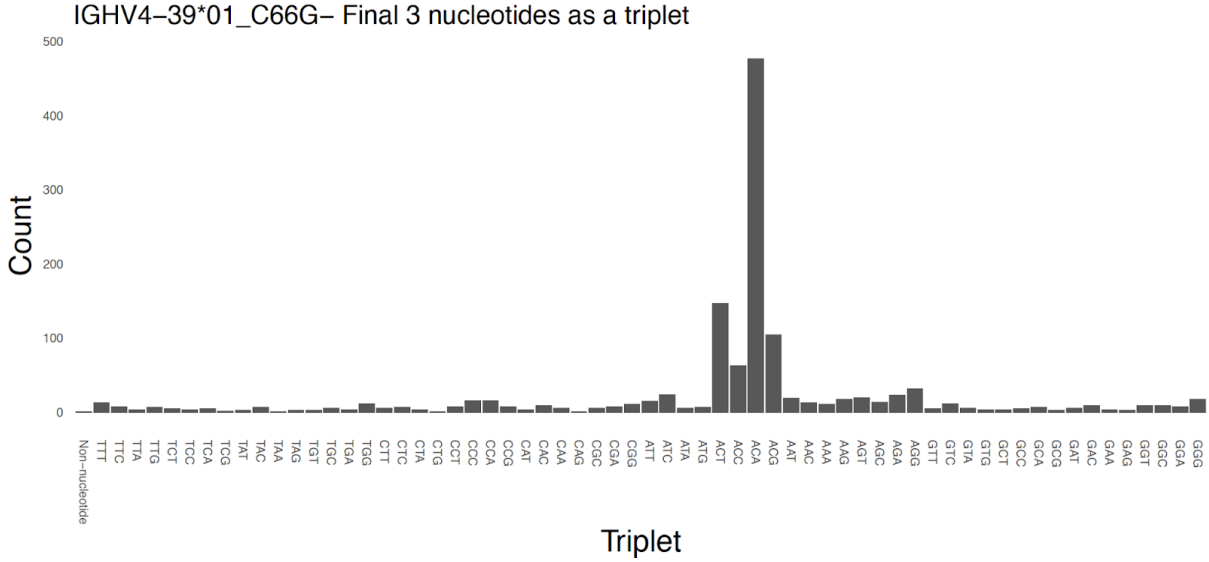
   >IGHV3-9*i01
   GAAGTGCAGCTGGTGGAGTCTGGGGGAGGCTTGGTACAGCCTGGC
   AGGTCCCTGAGACTCTCCTGTGCAGCCTCTGGATTCACCTTTGATGA
   TTATGCCATGCACTGGGTCCGGCAAGCTCCAGGGAAGGGCCTGGAG
   TGGGTCTCAGGTATTAGTTGGAATAGTGGTAGCATAGGCTATGCGGAC
   TCTGTGAAGGGCCGATTCACCATCTCCAGAGACAACGCCAAGAACTC
   CCTGTATCTGCAAATGAACAGTCTGAGAGCTGAGGACACGGCCTTGT
   ATCACTGTGCAAAAG...

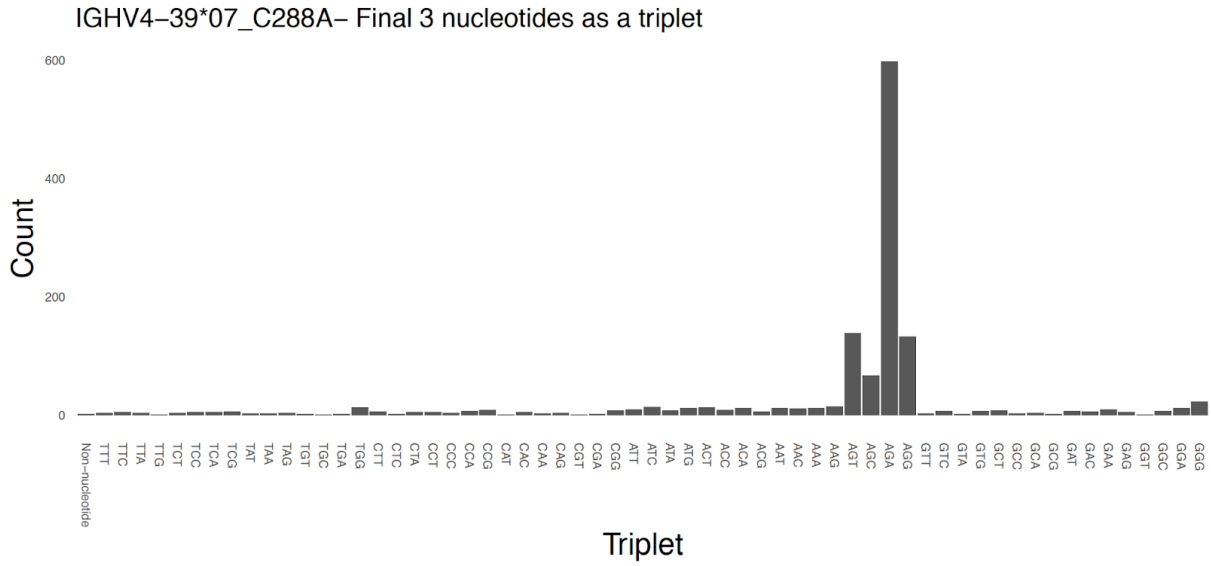
4. **Inference of IGHV4-39*01_C66G**
   IGHV4-39*01_C66G was inferred in subject S86 (VDJbase: P1_I90_S1), a genotype that also carries IGHV4-39*01. This inference has previously been pre-assessed at IARC meeting 62 (https://www.antibodysociety.org/wordpress/wp-content/uploads/2021/04/Meeting-62-3_11_20-minutes.pdf). IGHV4-39*01_C66G was well expressed, represented by 1305 sequences, 1135 unmutated sequences, 1131 unique CDR3s in the unmutated sequence set, showed an allelic frequency of 48%, and represented 3.3% of all unmutated reads of the genotype. Haplotyping based on allelic diversity in IGHJ6 was possible and the alleles distributed well between haplotypes (IGHV4-39*01_C66G: 2:98; IGHV4-39*01: 100:0). 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. It is known that the second last base of alleles of IGHV4-39 may be either C or G. The data describing the

final bases of this inferred allele strongly supports that this base (319) is a C. The inference of IGHV4-39*07_C288A in data set VDJbase: P1_I29_S1 in contrast suggests that this allele features a G as the penultimate residue, illustrating that the inference process can pick up a difference at the second last base of alleles. The allele representing IGHV4-39*01_C66G is given the name IGHV4-39*i01.

>IGHV4-39*i01
CAGCTGCAGCTGCAGGAGTCGGGCCCAGGACTGGTGAAGCCTTCG
GAGACCCTGTCCCTCACGTGCACTGTCTCTGGTGGCTCCATCAGCA
GTAGTAGTTACTACTGGGGCTGGATCCGCCAGCCCCCAGGGAAGGG
GCTGGAGTGGATTGGGAGTATCTATTATAGTGGGAGCACCTACTACAA
CCCGTCCCTCAAGAGTCGAGTCACCATATCCGTAGACACGTCCAAGA
ACCAGTTCTCCCTGAAGCTGAGCTCTGTGACCGCCGCAGACACGGC
TGTGTATTACTGTGCGAGAC.







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**
   Discussion. Specifically, the assignment of human genes/alleles to defined subgroups were scrutinised and cases of potential outliers were identified. Problematic separation of some alleles (e.g. those of IGHV4-4 and IGHV4-59) was highlighted.

6. **Next meeting**
   October 25th, 2021 at 10.00 UTC