Minutes of IARC meeting 116, Feb 27th, 2023

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

1. **Approval of minutes of meeting 115**
   Approved

2. **Next meeting**
   March 6th 2023 at 11.00 UTC

3. **Assessment of inference IGHV3-30*04 C201T G317A/IGHV3-30*18 G113C C114T** (P1_I70_S1)
   IGHV3-30*04 C201T has been inferred in one genotype (P1_I70) in VDJbase P1 data set. The genotype also carries related alleles like IGHV3-30-3*01 and IGHV3-30*18. It is represented by 394 error-free sequences and 374 unique CDR3s of different lengths in the error-free set. Haplotyping based on allelic diversity in IGHJ6 demonstrates association of the haplotype defined by IGHJ6*03 (99:1 ratio). IGHV3-30-3*01 and IGHV3-30*18 are more abundant but also present on both haplotypes. On the haplotype in question IGHV3-30-3*01, IGHV3-30*18, and inferred IGHV3-30*04 C201T are present at approximately a 2:2:1 ratio.

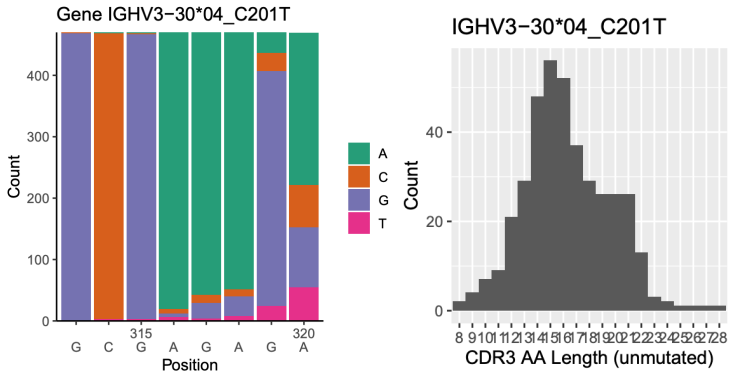
   Inspection of the sequences associated with the inference demonstrates that base 317 is A, not G as implied by the outcome of the inference, demonstrating that IGHV3-30*04 C201T G317A (a sequence identical to IGHV3-30*18 G113C C114T) is the allele in question. A past study has, using IgDiscover v.0.12, inferred IGHV3-30*04 C201T G317A/IGHV3-30*18 G113C C114T in this data set (Huang et al. Front Immunol 12:730105; DOI: 10.3389/fimmu.2021.730105). Upstream regions of these alleles were also inferred in that study. The upstream region of IGHV3-30*04 C201T G317A was identical to that of IGHV3-30*18 but differed in four positions from that of IGHV3-30-3*01.

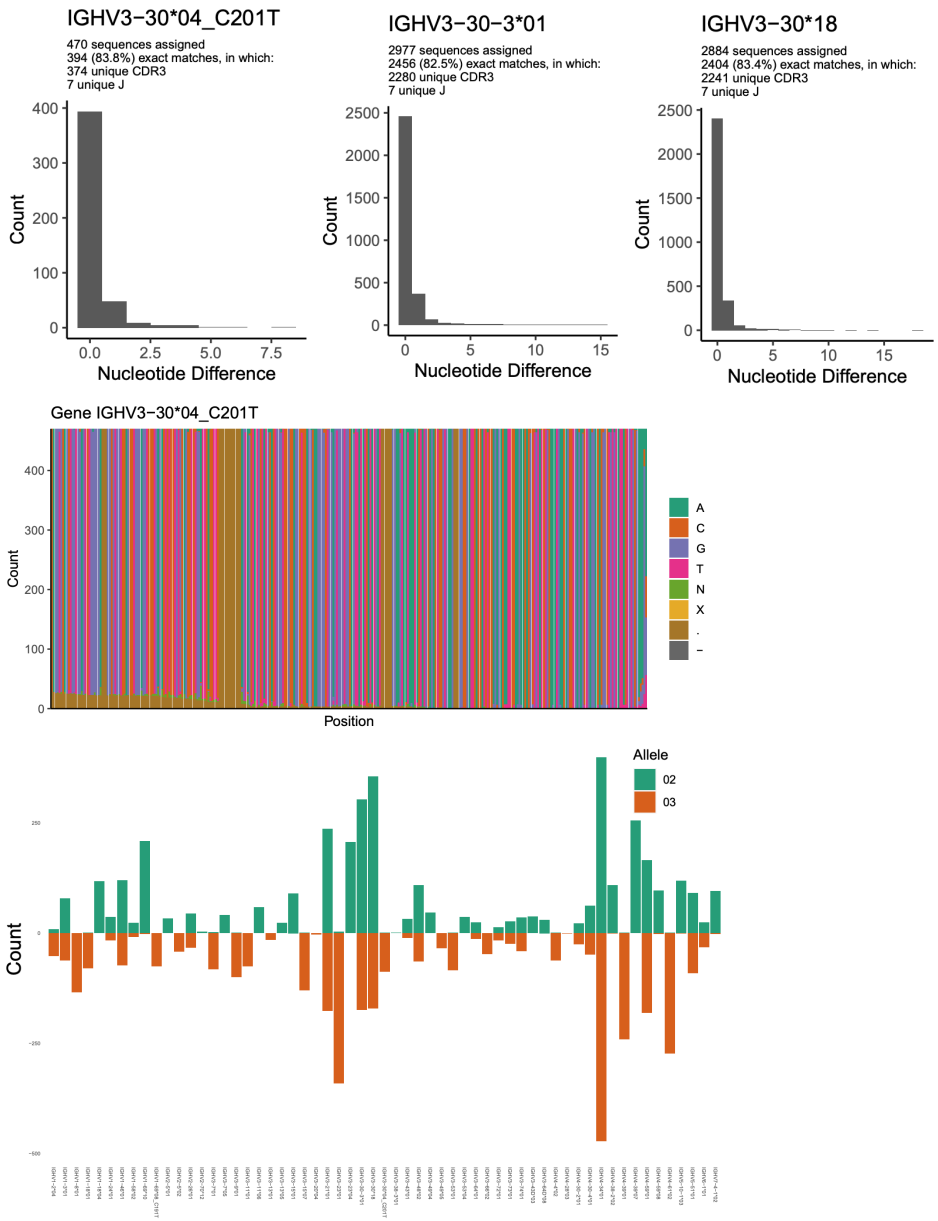
   IARC affirmed the sequence at Level 0 at Meeting 115, to possibly be upgraded to level 1 pending further discussion. IARC now affirms the sequence at level 1 up to and including base 319. It is acknowledged that the allele most likely carries 1 additional base, typically A, at base positions 320. The trailing “.” indicates IARC’s opinion that the sequence is likely to contain

additional 3’-nucleotides for which there is insufficient evidence to make an affirmation. The GenBank record (OX384049) is publicly available and reports the sequence of IGHV3-30*04 C201T G317A, up to and including base 320. For use in a reference germline gene set, IARC recommends the use of the expected full length sequence.

```
>IGHV3-30*i02
CAGGTGCAGCTGGTGGAGTCTGGGGGAGGCGTGGTCCAGCCTGGGAGGTCCCTGAGA
CTCTCCTGTGCAGCCTCTGGATTCACCTTCAGTAGCTATGCTATGCACTGGGTCCGC
CAGGCTCCAGGCAAGGGGCTGGAGTGGGTGGCAGTTATATCATATGATGGAAGTAAT
AAATACTATGCAGACTCCGTGAAGGGCCGATTCACCATCTCCAGAGACAATTCCAAG
AACACGCTGTATCTGCAAATGAACAGCCTGAGAGCTGAGGACACGGCTGTGTATTAC
TGTGCGAAAG.
```

The locus on chromosome 14 that carries human IGHV genes is highly complex. Genes may be duplicated or deleted, and identical sequences may be found in more than one gene. The IGHV3-30, IGHV3-30-3, IGHV3-30-5, and IGHV3-33 harbors very similar alleles, some of which are identical. For instance IGHV3-30*18 is found at both IGHV3-30 and IGHV3-30-5 (in that case: IGHV3-30-5*01) and IGHV3-30*04 is also found at IGHV3-30-3 (as IGHV3-30-3*03). Inference does not provide proof of the gene of the inferred allele. The gene of the inferred allele IGHV3-30*04 C201T G317A cannot be defined. Any name (with an “i” allele designation) of an inferred allele does not imply that its precise genetic location is known. It just, according to past practice of IARC, relates to the most similar allele presently found in the IMGT database, or to the gene with the lowest alphanumeric value, should alleles of multiple genes be equally matched to the novel allele in question. Other similar genes have been mentioned above.





Upstream regions:

```
IGHV3-30-3*01-A|69      AGCTCTGGGAGACGAGCCCAGCACTGGAAGTCGCCGGTGTTTCCATTCGGTGATCATCACTGAACACAGA
IGHV3-30*04 C201T G317A-A|1 AGCTCTGGGAGAGGAGCCCAGCACTAGAAGTCGGCGGTGTTTCCATTCGGTGATCAGCACTGAACACAGA
IGHV3-30*18-A|92         AGCTCTGGGAGAGGAGCCCAGCACTAGAAGTCGGCGGTGTTTCCATTCGGTGATCAGCACTGAACACAGA

IGHV3-30-3*01-A|69      GGACTCACCATGGAGTTTGGGCTGAGCTGGGTTTTCCTCGTTGCTCTTTTAAGAGGTGTCCAGTGT
IGHV3-30*04 C201T G317A-A|1 GGACTCACCATGGAGTTTGGGCTGAGCTGGGTTTTCCTCGTTGCTCTTTTAAGAGGTGTCCAGTGT
IGHV3-30*18-A|92         GGACTCACCATGGAGTTTGGGCTGAGCTGGGTTTTCCTCGTTGCTCTTTTAAGAGGTGTCCAGTGT
```

4. **Assessment of inference IGHV3−13*01_G290A_T300C in P1_I10_S1 (S00038)**

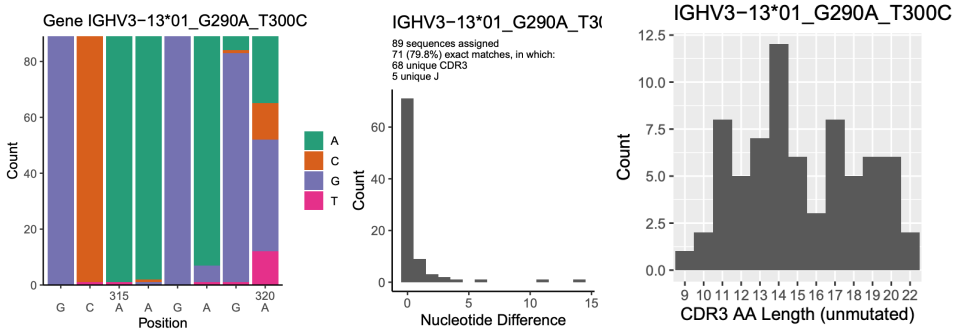
   IGHV3−13*01_G290A_T300C has been inferred in one genotype (P1_I10) in VDJbase P1 data set. The genotype does not carry a related allele and the opposite haplotype carries a large deletion that involves IGHV3-13 (Gidoni et al. (2019) Nat Commun 10, 628. DOI: 10.1038/s41467-019-08489-3). It

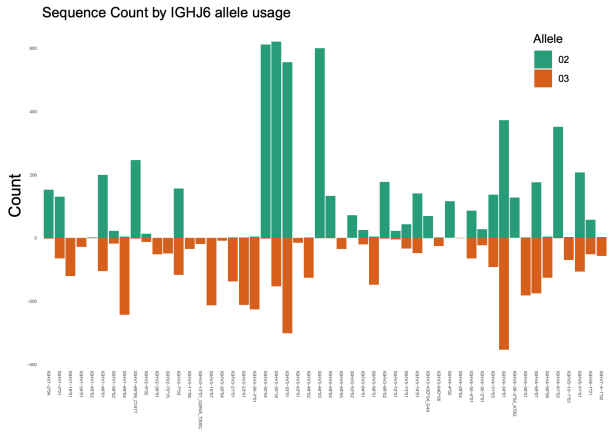
represents 0.23% of the total unmutated population, it is represented by 71 unmutated error-free sequences and 68 unique CDR3s of different lengths in the error-free set. Haplotyping based on allelic diversity in IGHJ6 demonstrates association of the haplotype defined by IGHJ6*03 (100:0 ratio). This allele is also inferred in multiple other data sets, two of which can be haplotyped. In both cases the inferred allele separates appropriately from IGHV3-13*05 (P1_I69) and IGHV3-13*04 (P1_I93) as determined by haplotyping.

IARC affirms the sequence as IGHV3-13*i01 at Level 1 up to and including base 319. It is acknowledged that the allele most likely carries 1 additional base, typically A, at base positions 320. Trailing "." indicates IARC's opinion that the sequence is likely to contain additional 3'-nucleotides for which there is insufficient evidence to make an affirmation. For use in a reference germline gene set, IARC recommends the use of the expected full length sequence.

```
>IGHV3-13*i01
GAGGTGCAGCTGGTGGAGTCTGGGGGAGGCTTGGTACAGCCTGGGGGGTC
CCTGAGACTCTCCTGTGCAGCCTCTGGATTCACCTTCAGTAGCTACGACA
TGCACTGGGTCCGCCAAGCTACAGGAAAAGGTCTGGAGTGGGTCTCAGCT
ATTGGTACTGCTGGTGACACATACTATCCAGGCTCCGTGAAGGGCCGATT
CACCATCTCCAGAGAAAATGCCAAGAACTCCTTGTATCTTCAAATGAACA
GCCTGAGAGCCGAGGACACGGCCGTGTATTACTGTGCAAGAG.
```

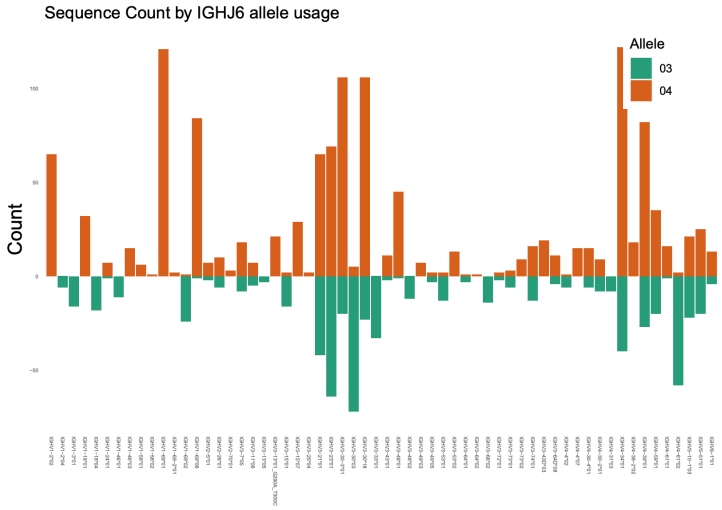
The locus on chromosome 14 that carries human IGHV genes is highly complex. Genes may be duplicated or deleted, and identical sequences may be found in more than one gene. The name (with an "i" allele designation) of an inferred allele does not imply that its precise genetic location is known. It just relates to the most similar allele presently found in the IMGT database, or to the gene with the lowest alphanumeric value, should alleles of multiple genes be equally matched to the novel allele in question. No other highly similar genes have been described.



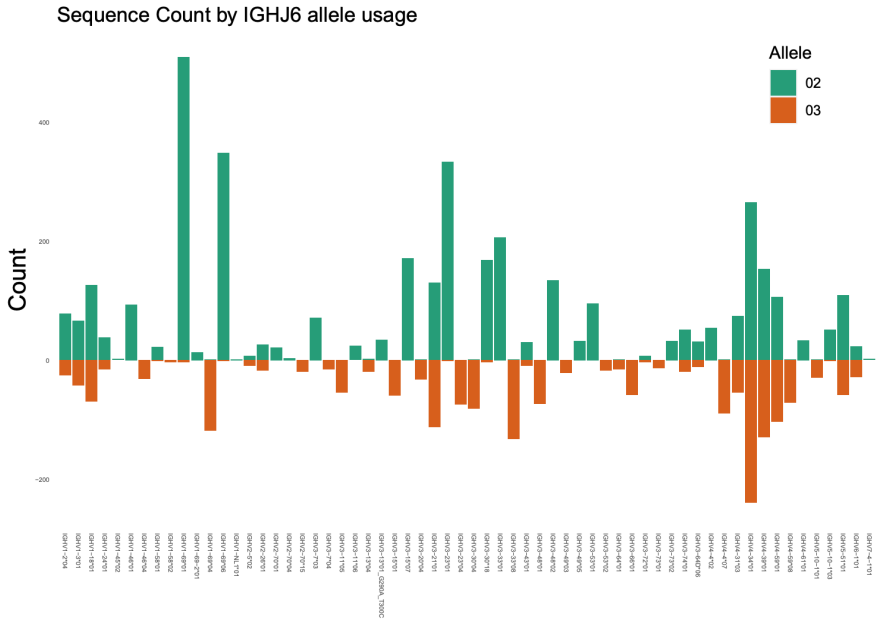


Haplotyping of other inferences carrying IGHV3-13*01_G290A_T300C

Inference in P1_I69_S1



Inference in P1_I93_S1:



5. **Assessment of inference IGHV4-61*01_A41G in P1_I23_S1 (S00038)**
   IGHV4-61*01_A41G has been inferred in one genotype (P1_I23) in VDJbase P1 data set. The genotype also carries IGHV4-61*01. IGHV4-61*01_A41G represents 0.32% of the total unmutated population. It is represented by 91 unmutated error-free sequences and 91 unique CDR3s of different lengths in the error-free set. Haplotyping based on allelic diversity in IGHJ6 demonstrates association of the haplotype defined by IGHJ6*02 (100:0 ratio) (IGHV4-61*01 shows a haplotype ratio of 7:93).

   Other genes carry alleles defined by IMGT, alleles that are highly similar to IGHV4-61*01 A41G. These genes include IGHV4-59 and IGHV4-4 (>97% sequence identity). In no case do these alleles carry the A41G SNP.

   IARC affirms the sequence as IGHV4-61*i04 at Level 1 up to and including base 319. It is acknowledged that the allele most likely carries 1 additional base, typically A, at base positions 320. Trailing “.” indicates IARC’s opinion that the sequence is likely to contain additional 3’-nucleotides for which there is insufficient evidence to make an affirmation. For use in a reference germline gene set, IARC recommends the use of the expected full length sequence.

   >IGHV4-61*i04
   CAGGTGCAGCTGCAGGAGTCGGGCCCAGGACTGGTGAGGCCTTCGGAGACCCTGTCC
   CTCACCTGCACTGTCTCTGGTGGCTCCGTCAGCAGTGGTAGTTACTACTGGAGCTGG
   ATCCGGCAGCCCCCAGGGAAGGGACTGGAGTGGATTGGGTATATCTATTACAGTGGG
   AGCACCAACTACAACCCCTCCCTCAAGAGTCGAGTCACCATATCAGTAGACACGTCC
   AAGAACCAGTTCTCCCTGAAGCTGAGCTCTGTGACCGCTGCGGACACGGCCGTGTAT
   TACTGTGCGAGAG.

The locus on chromosome 14 that carries human IGHV genes is highly complex. Genes may be duplicated or deleted, and identical sequences may be found in more than one gene. The name (with an “i” allele designation) of an inferred allele does not imply that its precise genetic location is known. It just relates to the most similar allele presently found in the IMGT database, or to the gene with the lowest alphanumeric value, should alleles of multiple genes be equally matched to the novel allele in question. Other similar genes have been mentioned above.