Minutes of IARC meeting 115, Feb 9th, 2023

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

1. **Approval of minutes of meeting 114**
   Approved

2. **Assessment of inference TRBV5−6*01_T284G in P4_I12_S1 (S00036)**
   TRBV5−6*01_T284G (substitution: L95W) has been inferred in four genotypes in the VDJbase P4 data set, including in VDJbase P4_I12_S1, a haplotypable data set (based on heterozygocity in TRBJ1-6). The genotype is also implied to carry TRBV5-6*01. No other gene in the IMGT database is highly similar to these alleles of TRBV5-6. The novel allele is the lesser expressed allele in the repertoire (35% allelic frequency; 0.43% of the total error-free population). It is represented by 122 error-free sequences and 118 unique CDR3s in the error-free set. Haplotyping based on allelic diversity in TRBJ1-6 demonstrates perfect separation from TRBV5-6*01. IARC affirms the sequence as TRBV5-6*i01 at Level 1 up to and including base 324. It is acknowledged that the allele most likely carries 1 additional base, typically G, at base position 325. 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.

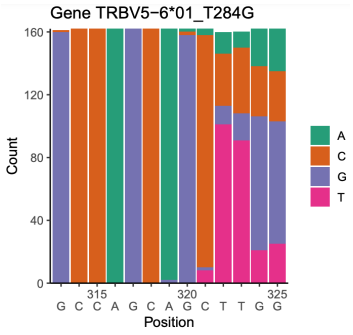
```
>TRBV5-6*i01 (TRBV5-6*01_T284G)
GACGCTGGAGTCACCCAAAGTCCCACACACCTGATCAAAACGAGAGGACA
GCAAGTGACTCTGAGATGCTCTCCTAAGTCTGGGCATGACACTGTGTCCT
GGTACCAACAGGCCCTGGGTCAGGGGCCCCAGTTTATCTTTCAGTATTAT
GAGGAGGAAGAGAGACAGAGAGGCAACTTCCCTGATCGATTCTCAGGTCA
CCAGTTCCCTAACTATAGCTCTGAGCTGAATGTGAACGCCTTGTGGCTGG
GGGACTCGGCCCTCTATCTCTGTGCCAGCAGCTTG.
```

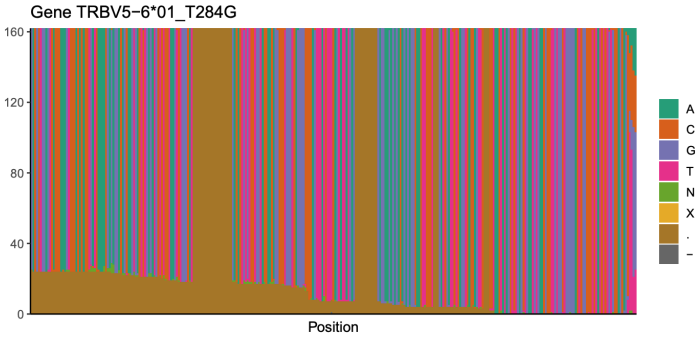
Result summary: TRBV5-6*01_T284G	No rearrangement found		
V-GENE and allele	Homsap TRBV5-6*01 F	score = 1356	identity = 99.63% (272/273 nt)
FR-IMGT lengths, CDR-IMGT lengths	[5.6.X]		

1. Alignment for V-GENE and allele identification

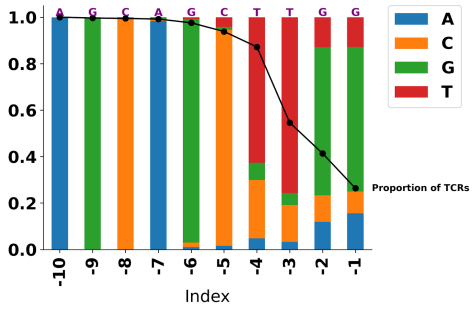
Closest V-REGIONs (evaluated from the V-REGION first nucleotide to the 2nd-CYS codon)

		Score	Identity
L36092	Homsap TRBV5-6*01 F	1356	99.63% (272/273 nt)
L36092	Homsap TRBV5-5*01 F	1212	93.77% (256/273 nt)
L36092	Homsap TRBV5-7*01 ORF	1212	93.77% (256/273 nt)
X57611	Homsap TRBV5-5*02 (F)	1203	93.41% (255/273 nt)
X58801	Homsap TRBV5-5*03 (F)	1203	93.41% (255/273 nt)

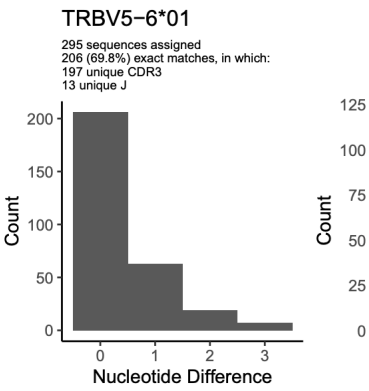


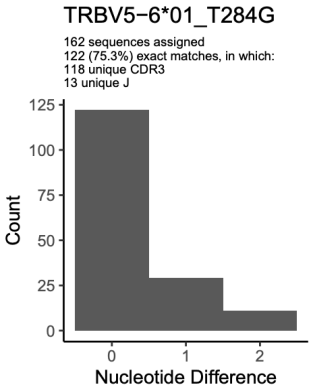


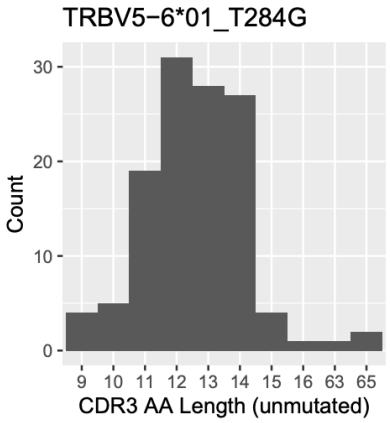
Consensus plot of 3'-end:

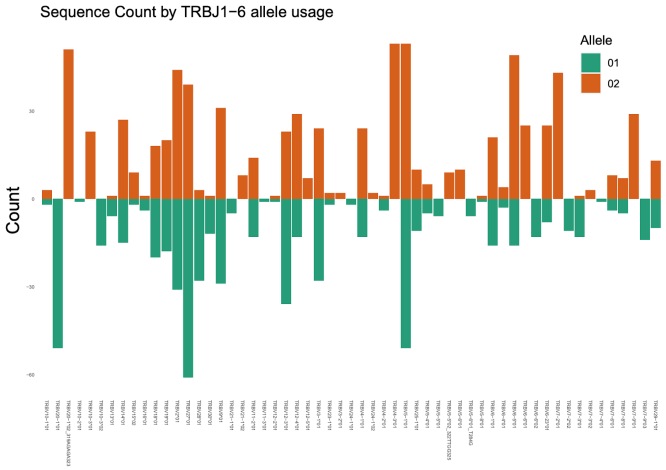


All data:



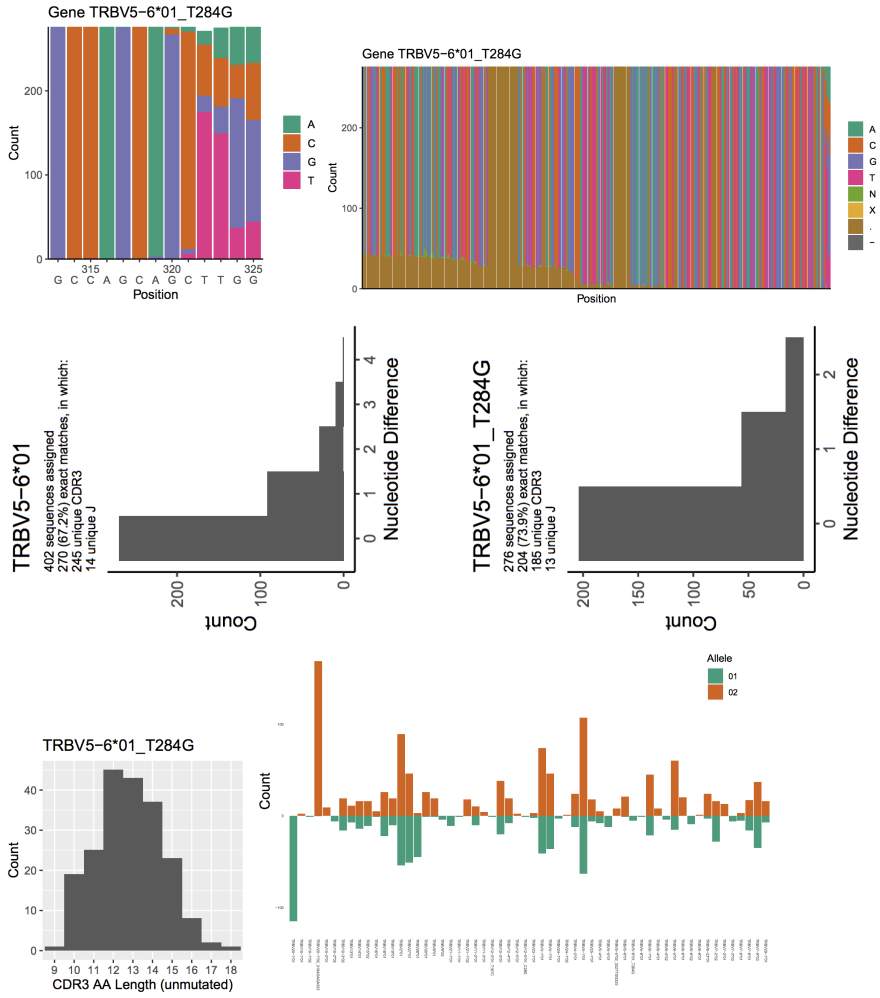


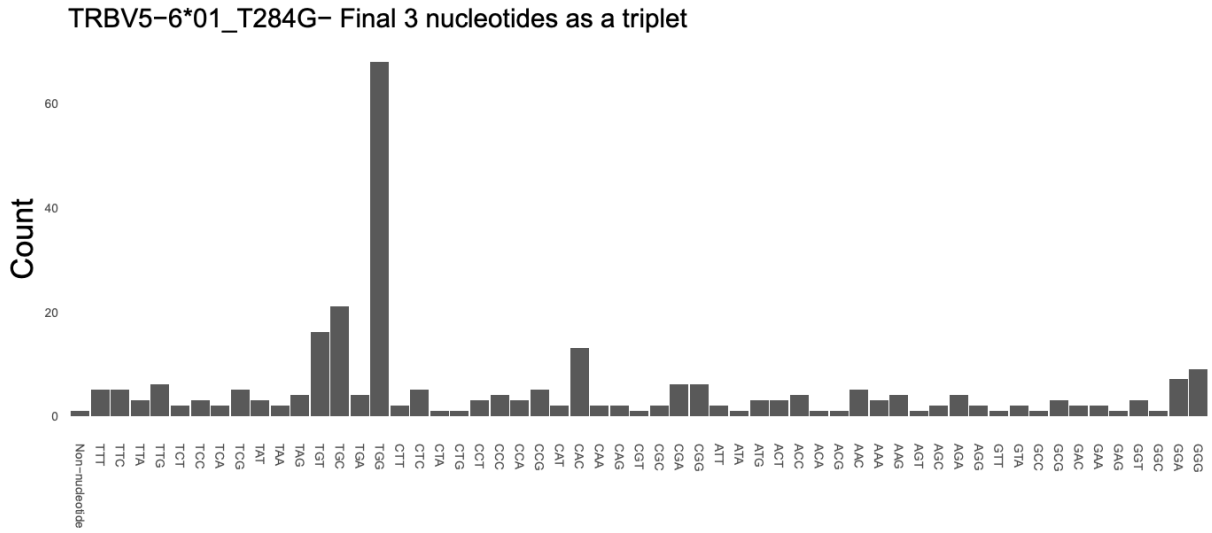




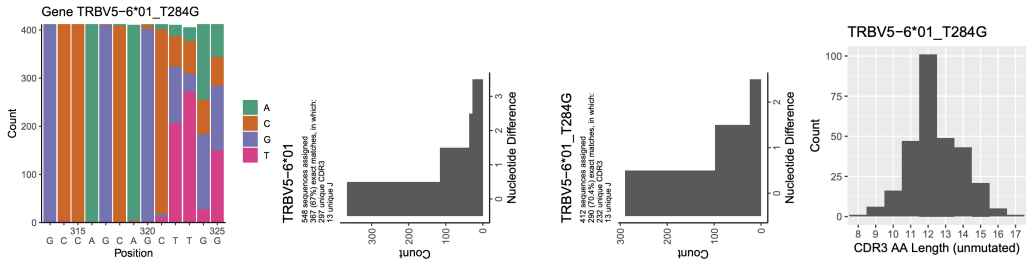
Other inferences:

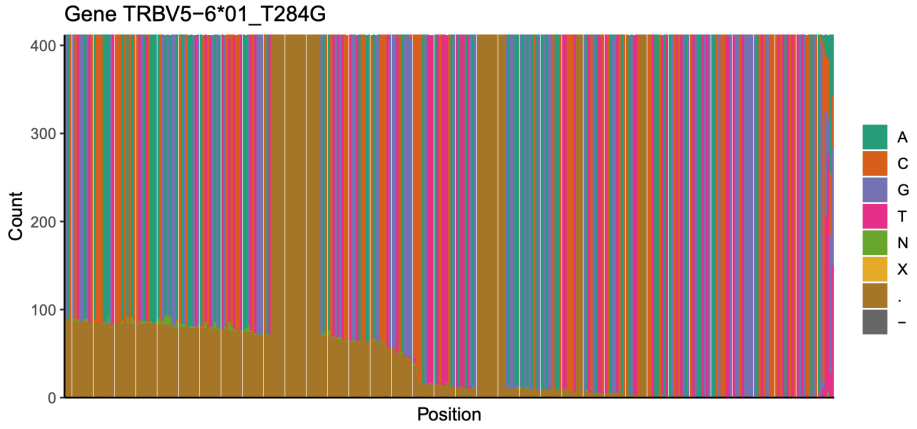
P1_I16

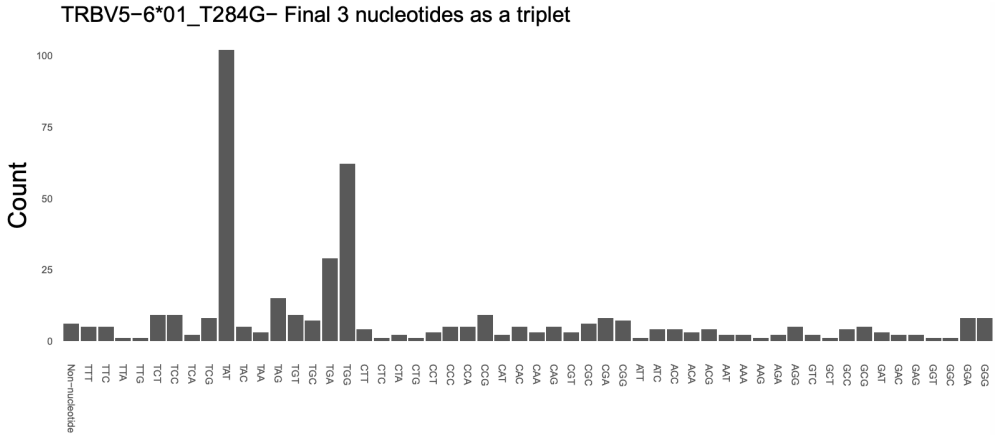



TRBV5−6*01_T284G− Final 3 nucleotides as a triplet
Count
Non−nucleotide

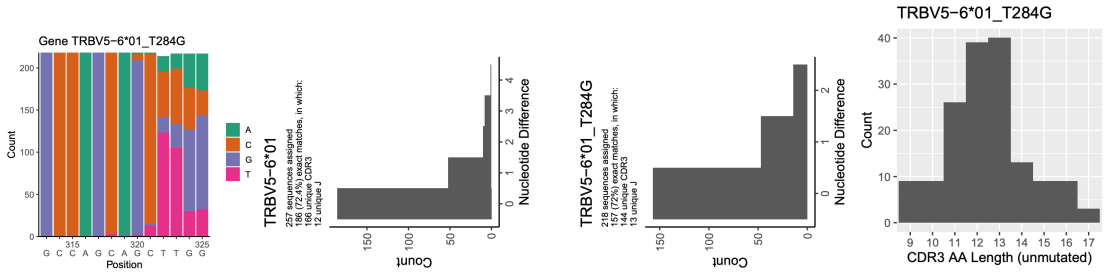
P4_I17:

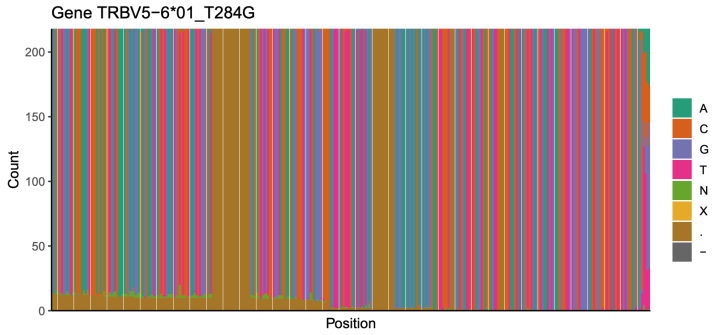

Gene TRBV5−6*01_T284G
Count
Position
A
C
G
T
TRBV5−6*01
Nucleotide Difference
TRBV5−6*01_T284G
TRBV5−6*01_T284G
CDR3 AA Length (unmutated)

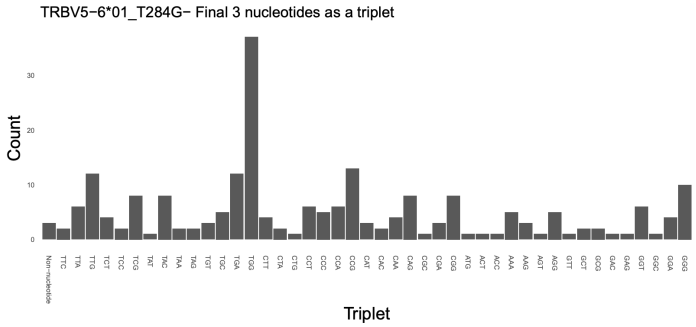

Gene TRBV5−6*01_T284G
Count
Position
A
C
G
T
N
X
.
−


TRBV5−6*01_T284G− Final 3 nucleotides as a triplet
Count
Non−nucleotide

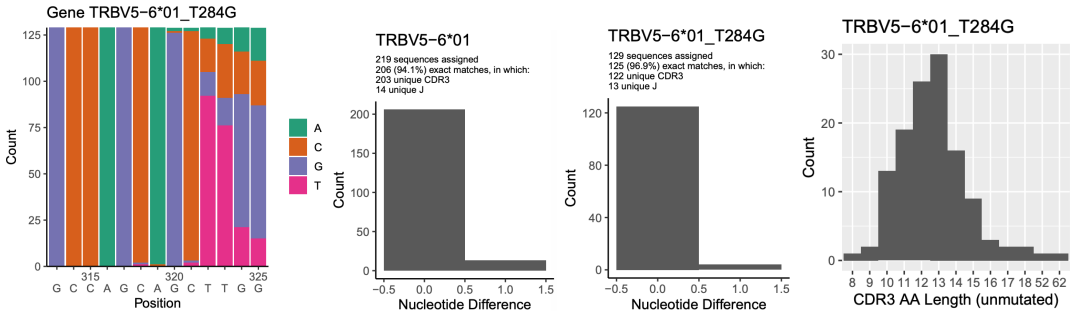
P4_I28:

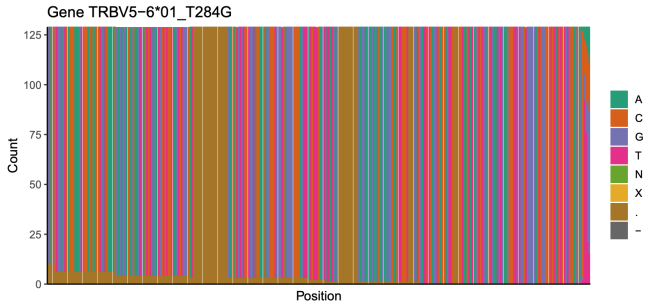

Gene TRBV5−6*01_T284G
Count
Position
A
C
G
T
TRBV5−6*01
257 sequences assigned
186 (72.4%) exact matches, in which:
166 unique CDR3
12 unique J
Nucleotide Difference
TRBV5−6*01_T284G
218 sequences assigned
157 (72%) exact matches, in which:
144 unique CDR3
13 unique J
TRBV5−6*01_T284G
CDR3 AA Length (unmutated)

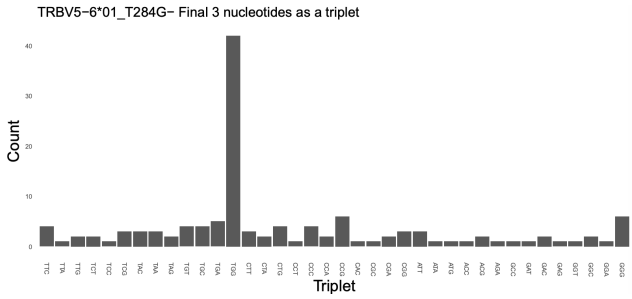

Gene TRBV5−6*01_T284G
Count
Position
A
C
G
T
N
X
.
-

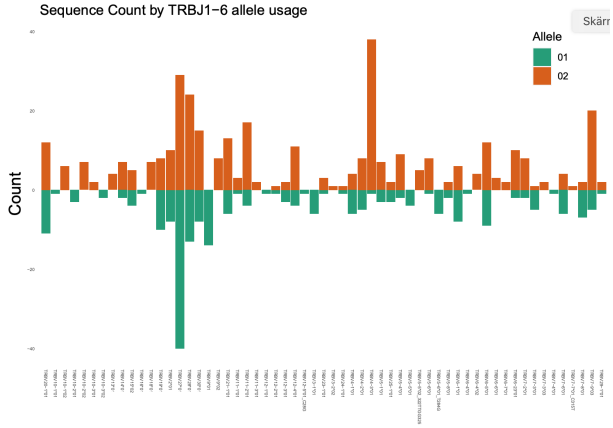

TRBV5−6*01_T284G− Final 3 nucleotides as a triplet
Count
Triplet

P16_I4


Gene TRBV5−6*01_T284G
Count
Position
A
C
G
T
TRBV5−6*01
219 sequences assigned
206 (94.1%) exact matches, in which:
203 unique CDR3
14 unique J
Nucleotide Difference
TRBV5−6*01_T284G
129 sequences assigned
125 (96.9%) exact matches, in which:
122 unique CDR3
13 unique J
TRBV5−6*01_T284G
CDR3 AA Length (unmutated)


Gene TRBV5−6*01_T284G
Count
Position
A
C
G
T
N
X
.
-





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.

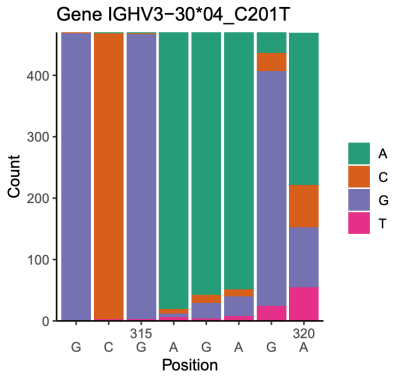
   Inspection of the sequences associated with the inference demonstrates that base 317 is A, not G as implied by the outcome of the inference, implying that IGHV3-30*04 C201T G317A (a sequence identical to IGHV3-30*18 G113C C114T) is the allele in question. A past study has, using IgDiscover v.0.12, inferred 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

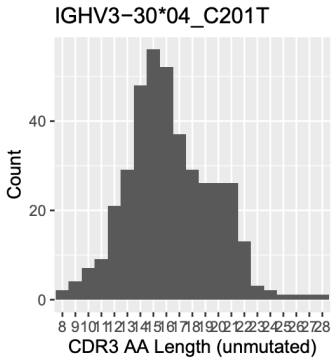
was identical to that of IGHV3-30*18 but differed in four positions from that of IGHV3-30-3*01.

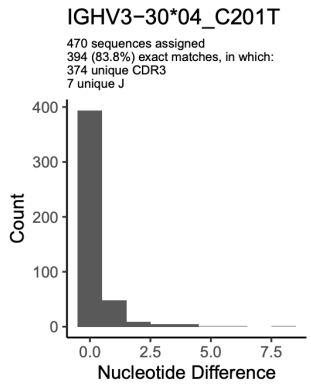
IARC affirms the sequence at Level 0, to possibly be upgraded to level 1 pending further discussion and when an updated GenBank record (AP) is publicly available, up to and including base 319 . It is acknowledged that the allele most likely carries 1 additional base, typically A, at base positions 320. 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. Considering the uncertainty of the alleles gene location, we currently do not assign a name to the sequence.

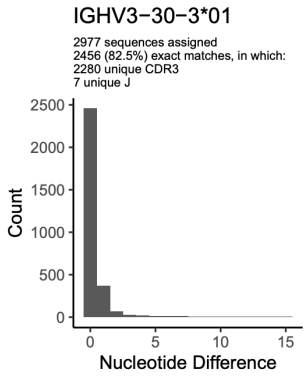
>
CAGGTGCAGCTGGTGGAGTCTGGGGGAGGCGTGGTCCAGCCTGGGAGGTCCCTG
AGACTCTCCTGTGCAGCCTCTGGATTCACCTTCAGTAGCTATGCTATGCACTGG
GTCCGCCAGGCTCCAGGCAAGGGGCTGGAGTGGGTGGCAGTTATATCATATGAT
GGAAGTAATAAATACTATGCAGACTCCGTGAAGGGCCGATTCACCATCTCCAGA
GACAATTCCAAGAACACGCTGTATCTGCAAATGAACAGCCTGAGAGCTGAGGAC
ACGGCTGTGTATTACTGTGCGAAAG.

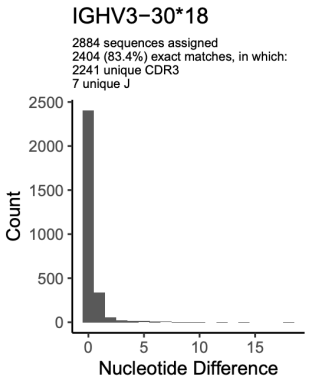
The locus on chromosome 14 that carries human IGHV genes is highly complex. Genes may be duplicated or deleted, and identical sequences may be found in more than one gene. The IGHV3-30, IGHV3-30-3, IGHV3-30-5, and IGHV3-33 harbors very similar alleles, some of which are identical. For instance IGHV3-30*18 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. A 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.

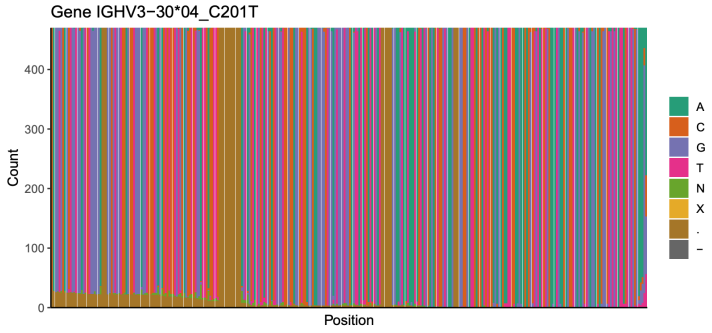


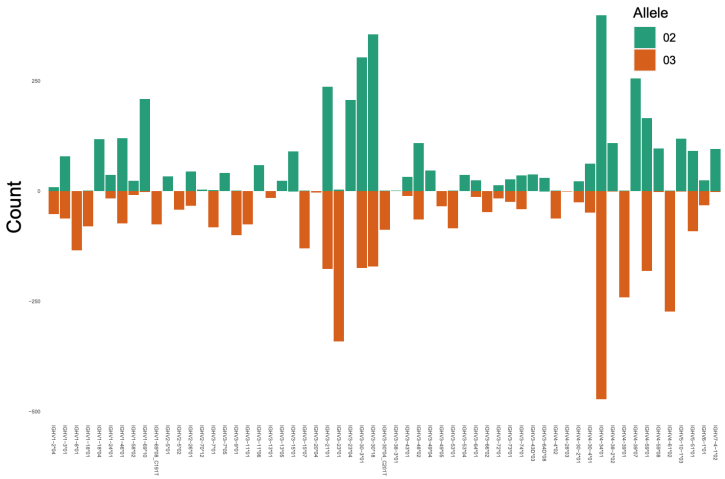












Upstream regions:

*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 IGHV1-13*05 (P1_I69) and IGHV1-13*04 (P1_I93) as determined by haplotyping.

   Discussion

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).

   Discussion.

6. **Next meeting**
   February 27th 2023 at 11.00 UTC