Minutes of IARC meeting 112, Jan 9th, 2023

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

1. **Approval of minutes of meeting 111**
   Approved

2. **Submissions of non-functional alleles to GenBank**
   It has been noted (MC) that submissions of alleles with stop codons are not readily accepted by GenBank and not available to common search functions. This must be kept in mind in dealing with particular analysis algorithms.

3. **Primers and literature reference of S00036 to be updated**
   In process (GY, WL will finalize the information).

4. **Submission of data for IGLV2-14*i02 (IGLV2-14*04) and of data for additional inferred IGHV P1 sequences in VDJbase that have not yet been affirmed (e.g. IGHV3-30*04 C201T G317A (IGHV3-30*18 G113C C114T) (P1_I70), IGHV3−13*01_G290A_T300C (P1_10), IGHV4-61*01 A41G (P1_I23)) to ENA/GenBank - update**
   Data defining alleles IGHV3-30*04_C201T_G317A, IGHV4-61*01_A41G, and IGHV3−13*01_G290A_T300C are now present in GenBank as OX384049-51 and ENA as ERR10444291-3. AP and WL will proceed with submission to OGRDB.

5. **IARC - Notes on development of human IGH germline sets that are optimized for use in AIRR-seq analysis.**
   Discussion. A text defining the basis for inclusion of alleles is being drafted. WL, AC, and CW will review it with a focus on inclusion criteria of alleles and circulate it to the entire IARC. MC et al. are about to publish a larger number of inferred alleles (incl. genomic confirmation of many of these genes) and information will be made available upon publication by the journal.

6. **Process to assess novel germline alleles using gAIRR process**
   Submissions are in process by data generators.

7. **Assessment of inference TRBV12-4*01_C87T in P4_I24_S1 (S00036)**
   TRBV12-4*01_C87T has been inferred in nine genotypes in the VDJbase P4 data set, including in VDJbase P4_I24_S1, a haplotypable data set (based on heterozygocity in TRBJ1-6). The genotype is also implied to carry TRBV12-4*01. No other gene apart from TRBV12-3 in the IMGT database is closely related to these alleles of TRBV12-4. The novel allele is the most expressed allele in the repertoire (72% allelic frequency; 1.47% of the total error-free population). It is represented by 499 error-free sequences and 467 unique CDR3s in the error-free set. Haplotyping based on allelic diversity in TRBJ1-6 demonstrates perfect separation from TRBV12-4*01.

   The data defining the 3'-end of sequences is negatively affected by trimming during the rearrangement process. IARC affirms the sequence (as TRBV12-4*i01) at Level 1 up to and including base 323 based on inference alone. It is acknowledged that the allele most likely carries 3 additional bases, typically AGC, at base positions 324-326. Analysis of sequences that have not been trimmed upstream of each position strongly suggest that the 3'-end of the sequence is indeed AGC. For many applications, IARC recommends the use of germline sequences representing their most likely full length base sequence. There is, however, insufficient evidence to make an affirmation of the full length sequence based on inference alone. In this case genomic data confirming the full length sequence is available from two other subjects (Rodrigues et al. Cell Genomics 2, 12, 100228 (https://doi.org/10.1016/j.xgen.2022.100228 (Supplementary Table S4)). Hence the full length sequence is approved based on inference in combination with genomic data.

   The locus that carries human TRBV 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 in the Notes section.

```
>TRBV12-4*01_C87T (TRBV12-4*i01)
```

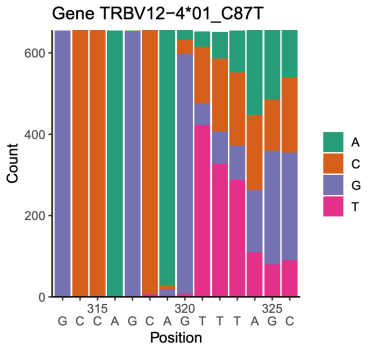
GATGCTGGAGTTATCCAGTCACCCCGGCACGAGGTGACAGAGATGGGACAAGAAGTG
ACTCTGAGATGTAAACCAATTTCAGGACATGACTACCTTTTCTGGTACAGACAGACC
ATGATGCGGGGACTGGAGTTGCTCATTTACTTTAACAACAACGTTCCGATAGATGAT
TCAGGGATGCCCGAGGATCGATTCTCAGCTAAGATGCCTAATGCATCATTCTCCACT
CTGAAGATCCAGCCCTCAGAACCCAGGGACTCAGCTGTGTACTTCTGTGCCAGCAGT
TTAGC

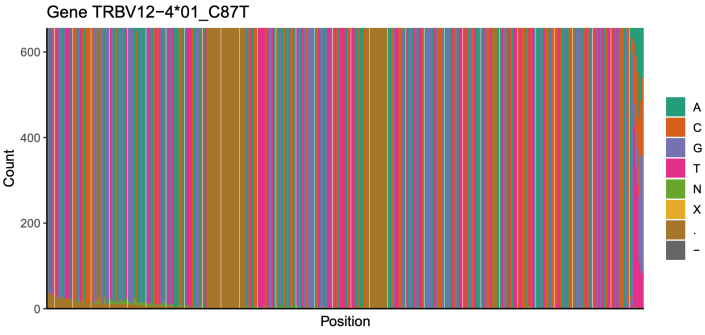
Result summary: TRBV12-4*01_C87T	**No rearrangement found**		
V-GENE and allele	Homsap TRBV12-4*01 F, or Homsap TRBV12-4*02 (F)	score = 1371	identity = **99.64%** (275/276 nt)
FR-IMGT lengths, CDR-IMGT lengths	[5.6.X]		

1. Alignment for V-GENE and allele identification

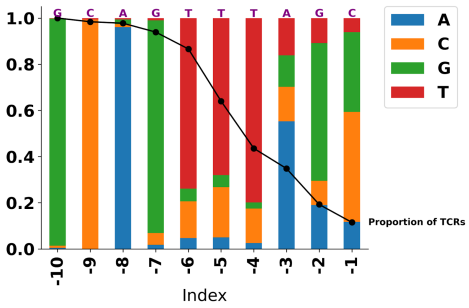
Closest V-REGIONs (evaluated from the V-REGION first nucleotide to the 2nd-CYS codon)

		Score	Identity
K02546	Homsap TRBV12-4*01 F	1371	99.64% (275/276 nt)
M14264	Homsap TRBV12-4*02 (F)	1371	99.64% (275/276 nt)
X07192	Homsap TRBV12-3*01 F	1326	97.83% (270/276 nt)
X07223	Homsap TRBV12-5*01 F	957	82.97% (229/276 nt)
X07224	Homsap TRBV12-1*01 P	867	79.35% (219/276 nt)

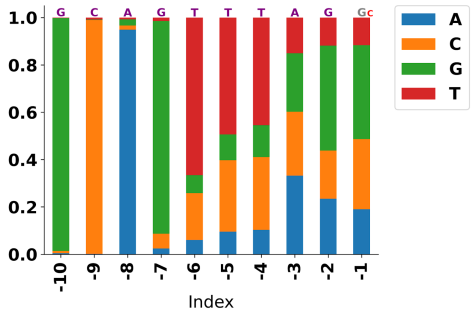


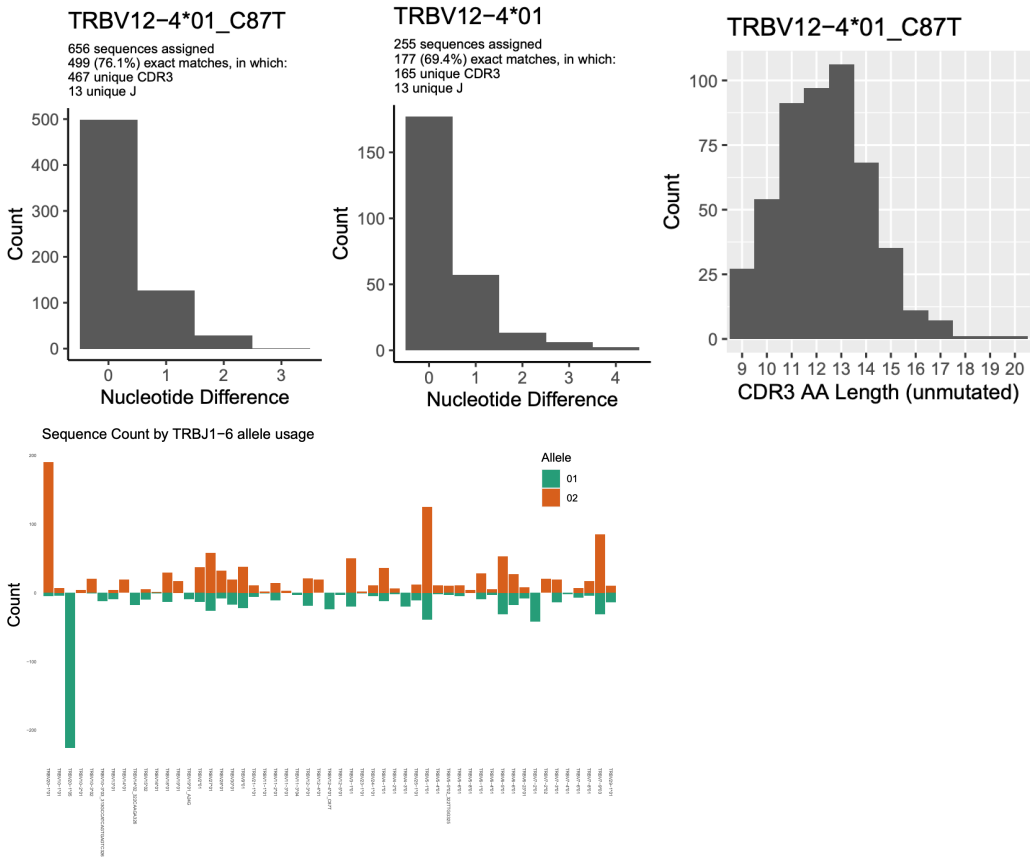


Consensus plot of 3’-end:



All data:





8. **Initial discussion of inference TRBV5−6*01_T284G in P4_I12_S1 (S00036)**

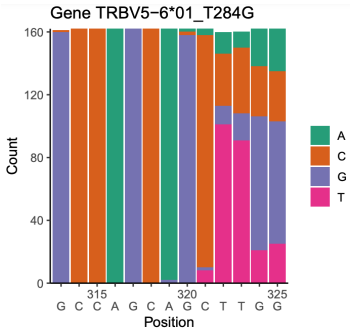
   TRBV5−6*01_T284G (substitution: L95W) has been inferred in four genotypes in the VDJbase P4 data set, including in VDJbase P4_I12_S1, a haplotypable data set (based on heterozygocity in TRBJ1-6). The genotype is also implied to carry TRBV5-6*01. No other gene in the IMGT database is highly similar to 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. No supplementary genomic confirmation seems to be available at this time. Further discussion will be required at a future meeting.

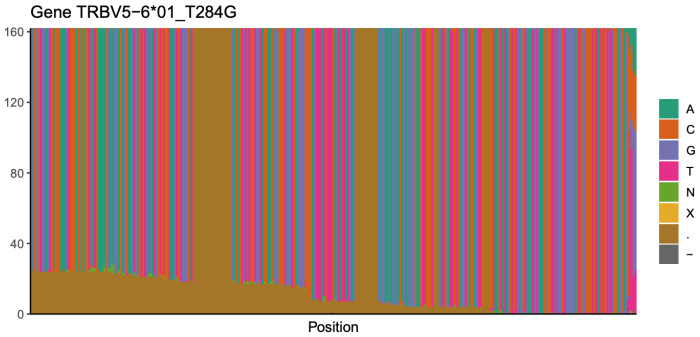
Result summary: TRBV5-6*01_T284G	No rearrangement found		
V-GENE and allele	Homsap TRBV5-6*01 F	score = 1356	identity = 99.63% (272/273 nt)
FR-IMGT lengths, CDR-IMGT lengths	[5.6.X]		

1. Alignment for V-GENE and allele identification

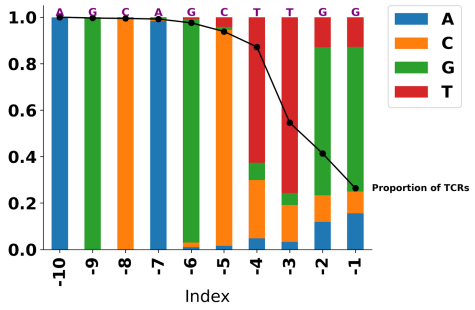
Closest V-REGIONs (evaluated from the V-REGION first nucleotide to the 2nd-CYS codon)

		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)

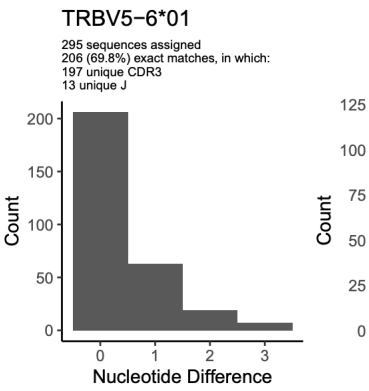


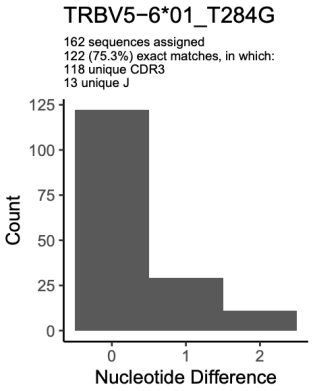


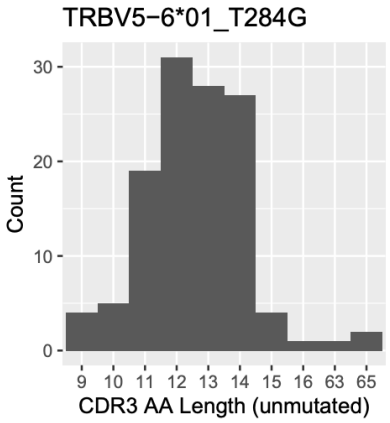
Consensus plot of 3'-end:

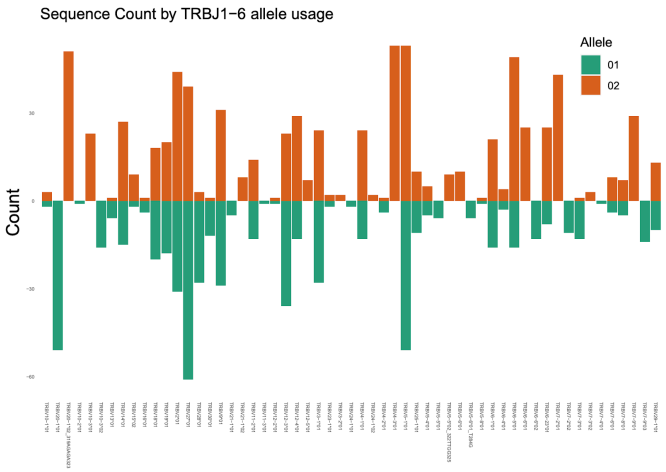


All data:









9. Next meeting

January 23rd, 2023 at 11.00 UTC