

Minutes of IARC meeting 109, Nov 7th, 2022

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

1. Approval of minutes of meeting 108

Approved

2. Functional groups reference book - update

Manuscript is in process.

3. 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

Submission of IGHV approved by ENA and submission to OGRDB is pending.

4. IARC - Notes on development of human IGH germline sets that are optimized for use in AIRR-seq analysis.

Discussion. WL will create a table defining the different gene set types. CW and AC will summarize data from OR's publications. AC will focus on details of valid sequences of Matsuda germline genes. MO will upload VBASE data in Drive for review.

CW raised the question on where to host the sets, possibly OGRDB.

Discussion of the role of AIRR-C in the process, whether AIRR-C endorses the process and IARC is responsible for the sets.

5. 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 heterozygosity 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.

Further discussion of the future process of affirmation of challenging 3'-end of heavily trimmed alleles based on transcriptomic data. It was suggested to affirm what is immediately obvious from the data but also to explicitly recommend use of likely full length sequences for many applications where short sequences pose an analytical problem.

The inference will be further discussed at the next meeting.

>TRBV12-4*01_C87T

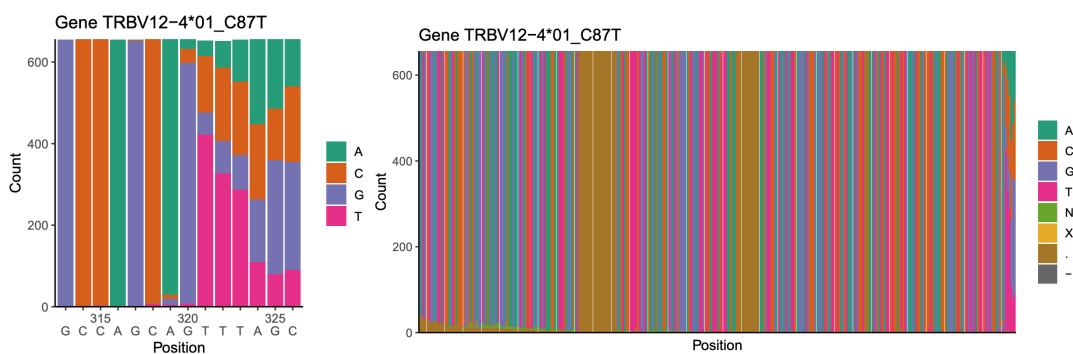
```
GATGCTGGAGTTATCCAGTCACCCCGGCACGAGGTGACAGAGATGGGACAAGAAGTG
ACTCTGAGATGTAAACCAATTTTCAGGACATGACTACCTTTTCTGGTACAGACAGACC
ATGATGCGGGGACTGGAGTTGCTCATTACTTTAACAACAACGTTCCGATAGATGAT
TCAGGGATGCCCCGAGGATCGATTCTCAGCTAAGATGCCTAATGCATCATTCTCCACT
CTGAAGATCCAGCCCTCAGAACCCAGGGACTCAGCTGTGTACTTCTGTGCCAGCAGT
. . . . .
```

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

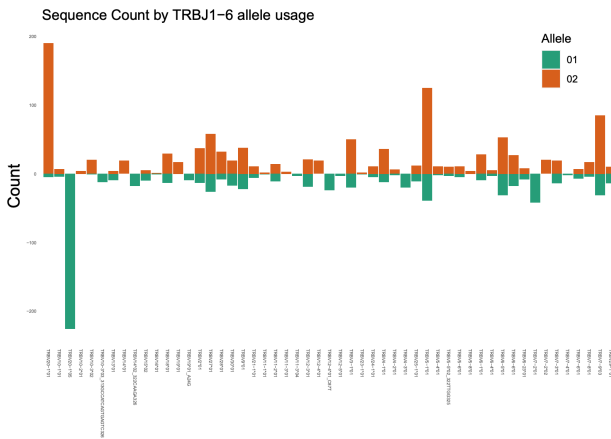
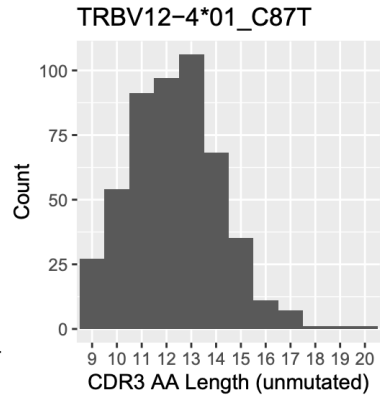
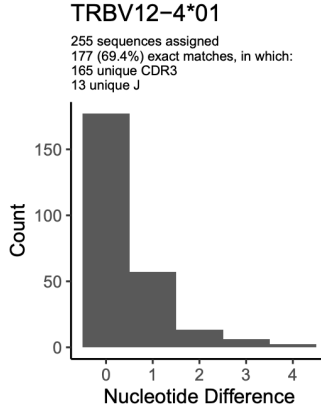
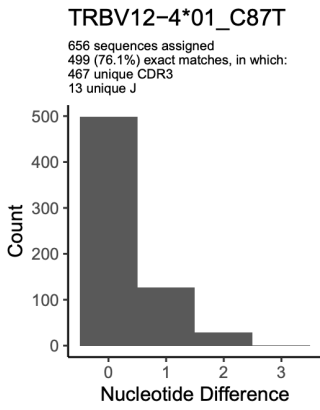
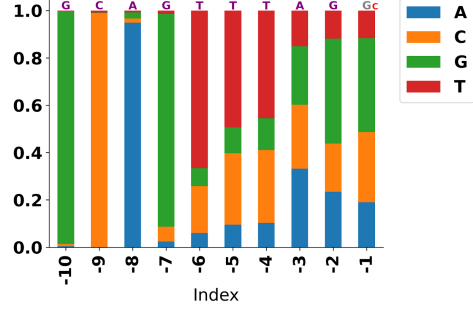
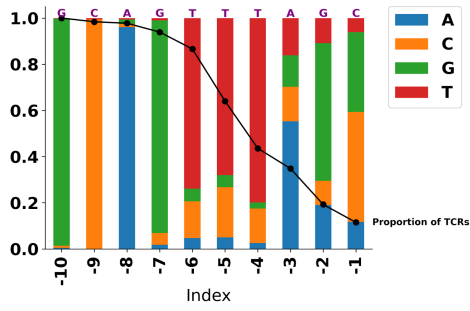
Closest V-REGIONS (evaluated from the V-REGION first nucleotide to the 2nd-CYS codon)

		Score	Identity
X02546	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:



6. Next meeting
 Monday Nov 21st at 11.00 UTC.