Minutes of IARC meeting 120, April 24, 2023

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

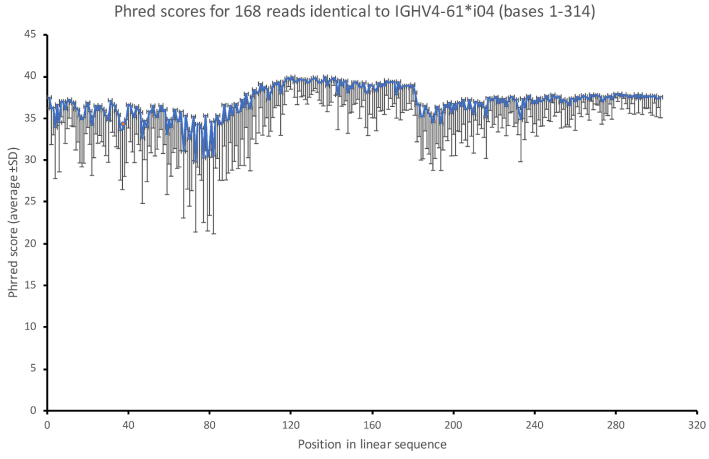
1. **Approval of minutes of meeting 119**
   Approved

2. **Next meeting**
   To be determined

3. **Quality of reads supporting an inference**
   Read quality was originally considered as a part of the affirmation process as local read problems have been shown to contribute to diversity (doi: 10.3389/fimmu.2017.01433, 10.1016/j.dib.2017.06.031). As a case study PEAR-assembled FASTQ data (ERR2567200; VDJbase P1_I23) for reads (100% sequence identity from base 1-314) supporting inference of IGHV4-61*01 A41G were studied. This particular inference has been made both using TIgGER (VDJbase) and IgDiscover (doi: 10.3389/fimmu.2021.730105) inference technology.

   Reads of IGHV4-61-similar genes in the genotype of the subject as defined by OGRDB-submission S00038:

IGHV4-4*02	413	363	87.89	61.00	1.26	403	355	IGHJ6	0:100
IGHV4-4*07	262	225	85.88	39.00	0.78	261	224	IGHJ6	100:0
IGHV4-59*01	761	670	88.04	100.00	2.32	743	657	IGHJ6	48:52
IGHV4-61*01	131	115	87.79	57.00	0.40	128	115	IGHJ6	7:93
IGHV4-61*01_A41G	100	91	91.00	43.00	0.32	98	91	IGHJ6	100:0



(base 38, corresponding to IMGT base 41, is highlighted by an orange dot)

In all, no evidence of local poor read quality in bases supporting base 41 was found. This information be used to get additional confidence in challenging inferences. In addition, current inference tool settings (e.g.IgDiscover shared CDR3 settings) may prevent inference based on this type of error.

4. **Germline reference set manuscript**
   Discussion of the draft version 6. It is important to note that the reference sets have been derived without direct reference to IMGT reference gene sets. Discussion on how to include pseudogenes and poorly expressed genes and genes that may differ substantially in level of expression between individuals. Discussion on how to manage sequences that are present in more than one gene or in genes that are not the same as defined by its current name. Comments to be added to the manuscript by May 1st, 2023.

5. **Assessment of inference TRBV7-7*01_C315T in P4_I9_S1 (S00036)**
   TRBV7-7*01_C315T has been inferred in seventeen genotypes in the VDJbase P4 data set, including in VDJbase P4_I9_S1, a haplotypable data set (based on heterozygocity in TRBJ1-6). The genotype is also implied to carry TRBV7-7*01. No other gene apart from IGHV7-6 (alleles of which also carry C315) in the IMGT database is highly similar to these alleles of TRBV7-7. The novel allele is the most expressed allele in the repertoire (58% allelic frequency; 0.16% of the total error-free population). It is represented by 37 error-free sequences and 33 unique CDR3s in the error-free set. Haplotyping based on allelic diversity in TRBJ1-6 demonstrates association of TRBV7-7*01_C315T with only one of the haplotypes (only few recorded cases; TRBV7-7*01 was not associated with any allele of TRBJ1-6).

   The allele has also been identified as TRBV7-7*01_S0326 and Sanger validated (GenBank MZ339373) (Corcoran et al. (2023) Immunity 56, 635-652.E6 (DOI: 10.1016/j.immuni.2023.01.026)). MC reported that the gene (as published in Immunity (DOI: 10.1016/j.immuni.2023.01.026) has been seen in six Sanger sequenced genomic clones derived from two subjects.

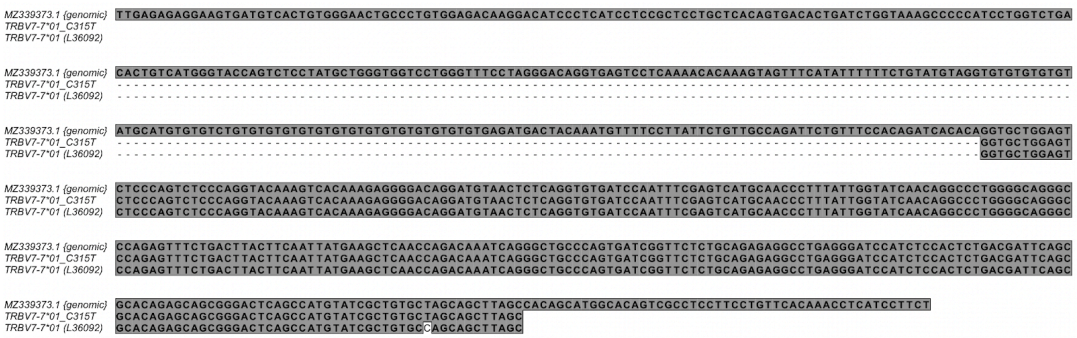
>MZ339373
TTGAGAGAGGAAGTGATGTCACTGTGGGAACTGCCCTGTGGAGACAAGGACATC
CCTCATCCTCCGCTCCTGCTCACAGTGACACTGATCTGGTAAAGCCCCCATCCT
GGTCTGACACTGTC**ATGGGTACCAGTCTCCTATGCTGGGTGGTCCTGGGTTTCC**
TAGGGACAGGTGAGTCCTCAAAACACAAAGTAGTTTCATATTTTTTCTGTATGT

AGGTGTGTGTGTGTATGCATGTGTGTCTGTGTGTGTGTGTGTGTGTGTGTGTGT
GTGTGAGATGACTACAAATGTTTTCCTTATTCTGTTGCCAGATTCTGTTTCCAC
AG**ATCACACAGGTGCTGGAGTCTCCCAGTCTCCCAGGTACAAAGTCACAAAGAG**
GGGACAGGATGTAACTCTCAGGTGTGATCCAATTTCGAGTCATGCAACCCTTTA
TTGGTATCAACAGGCCCTGGGGCAGGGCCCAGAGTTTCTGACTTACTTCAATTA
TGAAGCTCAACCAGACAAATCAGGGCTGCCCAGTGATCGGTTCTCTGCAGAGAG
GCCTGAGGGATCCATCTCCACTCTGACGATTCAGCGCACAGAGCAGCGGGACTC
AGCCATGTATCGCTGTGCTAGCAGCTTAGCCACAGCATGGCACAGTCGCCTCCT
TCCTGTTCACAAACCTCATCCTTCT
heptamer 23-mer nonamer

It has also been identified in a BAC clone with accession number AC229888

The process of affirmation of the allele and incorporation of not only transcriptomic but also genomic data will be brought up for further discussion during next meeting.

>TRBV7-7*i01 (TRBV7-7*01_C315T)
GGTGCTGGAGTCTCCCAGTCTCCCAGGTACAAAGTCACAAAGAGGGGACAGGAT
GTAACTCTCAGGTGTGATCCAATTTCGAGTCATGCAACCCTTTATTGGTATCAA
CAGGCCCTGGGGCAGGGCCCAGAGTTTCTGACTTACTTCAATTATGAAGCTCAA
CCAGACAAATCAGGGCTGCCCAGTGATCGGTTCTCTGCAGAGAGGCCTGAGGGA
TCCATCTCCACTCTGACGATTCAGCGCACAGAGCAGCGGGACTCAGCCATGTAT
CGCTGTGCTAGCAGCTTAGC



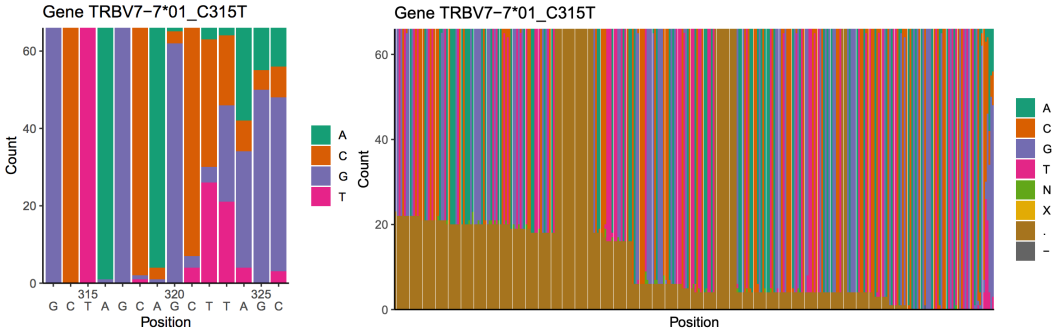
Result summary: TRBV7-7*01_C315T	No rearrangement found		
V-GENE and allele	Homsap TRBV7-7*01 F	score = 1380	identity = 100.00% (276/276 nt)
FR-IMGT lengths, CDR-IMGT lengths	[5.6.X]		

1. Alignment for V-GENE and allele identification

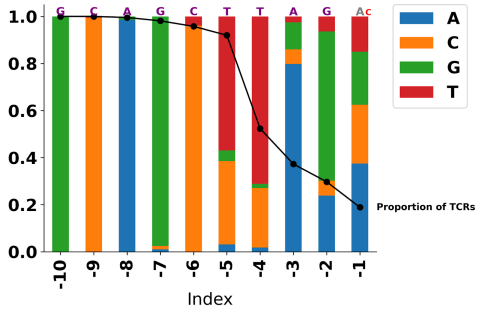
Closest V-REGIONs (evaluated from the V-REGION first nucleotide to the 2nd-CYS codon)

		Score	Identity
L36092	Homsap TRBV7-7*01 F	1380	100.00% (276/276 nt)
X57607	Homsap TRBV7-7*02 (F)	1371	99.64% (275/276 nt)
L36092	Homsap TRBV7-6*01 F	1281	96.01% (265/276 nt)
X58806	Homsap TRBV7-6*02 (F)	1272	95.65% (264/276 nt)
M11953	Homsap TRBV7-8*01 F	1119	89.49% (247/276 nt)

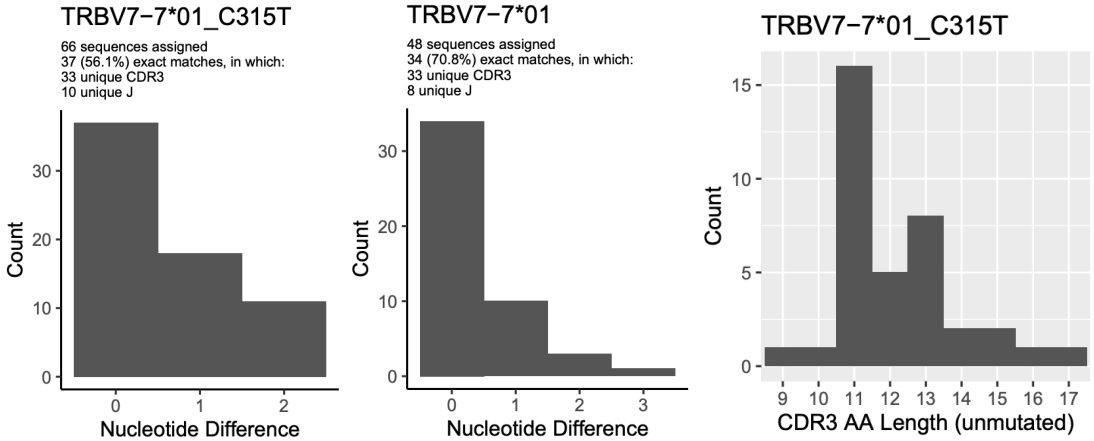
(Note: the difference between TRBV7-7*01_C315T and TRBV7-7*01 is outside of the range assessed by IMGT/V-QUEST)

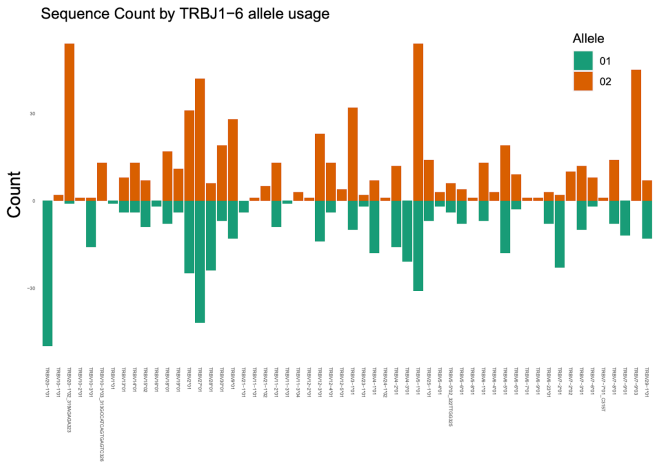


Consensus plot of 3'-end:



All data:




Sequence Count by TRBJ1−6 allele usage
Allele
01
02
Count
30
0
−30