Agenda of IARC meeting 92, Jan 24th, 2022

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

1. **Approval of minutes of meeting 91**
   Postponed until next meeting

2. **Update on the mouse germline F1 manuscript**
   AC presented process

3. **Update on the Review VDJbase Inferences feature of OGRDB**
   WL presented novel OGRDB features relevant for facilitated access to information of novel alleles in VDJbase

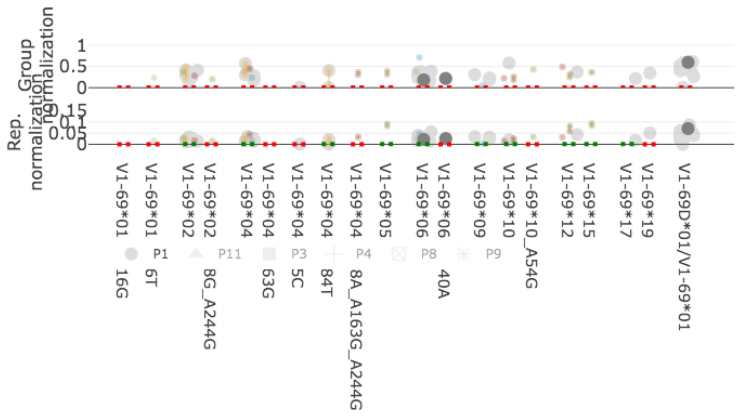
4. **AP's web interface for germline gene repertoire assessment of AIRR-seq data (Functional groups reference book) and integration of information with output of OGRDB**
   (https://ayeletperes.github.io/reference_book2/)
   AP presented updates. Discussion.

5. **Update on assessment of difficult to identify SNPs towards the 3'-end of alleles (see §4b of Meeting 90)**
   AP is updating on on-going work on this matter

6. **Additional novel alleles in VDJbase, study P1 (contd.)**
   VDJbase P1 study carries a number of other possible novel alleles that have not been affirmed in the past in addition to those discussed during Meeting 90 (§4) and Meeting 91 (§5). These include

   a. IGHV1-69*06_g240a (level 1 when assessed at meeting 60) (P1_I48), also assessed at meeting 91

      This allele now appears in a recent update of Functional groups reference book 2, in which case it represents 401 assignments and a 22% group normalized frequency within IGHV1-69/IGHV1-69D (G8). Overall frequency of this inference: 2.6%



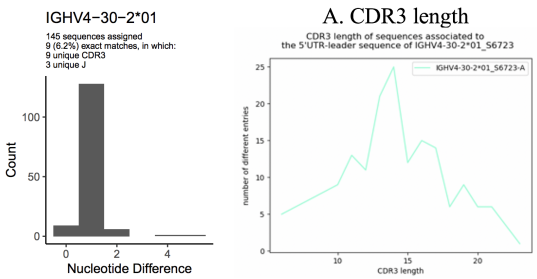
b. IGHV3-30*04_c201t_g317a (P1_I70)
Annotated as IGHV3-30*18_g113c_c114t in VDJbase. Not found in the original version of “Functional groups reference book” but it is now found in version 2. It has also been independently inferred in a past study (doi: 10.3389/fimmu.2021.730105). Found in many rearrangements with a wide diversity of CDR3 lengths and a high frequency of unmutated reads. A separate study has identified this allele associated to IGHJ6*03 in a genotype that also carries IGHV3-30-3*01 and IGHV3-30*18 associated to both haplotypes (doi: 10.3389/fimmu.2021.730105). (IGHV3-30*18 is not yet well annotated by VDJbase or Functional Groups Reference Book as it is identical to IGHV3-30-5*01.) The inferred upstream region of IGHV3-30*04_C201T_G317A was shown to be identical to that of IGHV3-30*18 but entirely different from that of IGHV3-30-3*01. This allelic variant is not found at NCBI using BLAST with the inferred allele sequence as search sequence. The novel allele is inferable up to and including base 319.

Decision: Missing bases at the 5’-end as annotated in OGRDBstats plots must be addressed. The analysis of this complicated part of the IGHV locus will be enhanced as IGHV3-30*18/IGHV3-30-5*01 is annotated in haplotype analysis. When resolved, it is anticipated that this allele can be submitted to OGRDB for final affirmation as a Level 1 sequence.

Data	Germline gene/allele	Haplotype	
		IGHJ6*02	IGHJ6*03
ERR2567243	IGHV3-30*04_S7005	0	63
	IGHV3-30*18	253	128
	IGHV3-30-3*01	195	111

c. IGHV4-30-2*01_g70a (P1_I74)
Not inferred in the current version of VDJbase or represented in “Functional groups reference book” (neither in version 1 or version 2). OGRDBstats plots indicate the presence of IGHV4-30-2*01 in this sample but most sequences have 1 mutation. The novel allele has been independently inferred in a past study (doi: 10.3389/fimmu.2021.730105), a study that demonstrated a high diversity of CDR3 lengths. It was also inferred in the original study by Mikocziova et al. There is no other allele of IGHV4-30-2 in the genotype. The sample cannot be haplotyped.

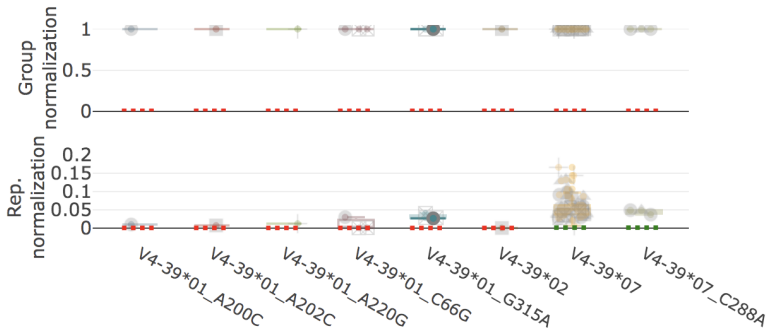
Decision: If this allele is to be assessed as a valid allele it must be identified and annotated in VDJbase and access to full OGRDBstats plots will be required for final affirmation. The absence of another similar allele in the genotype will facilitate its affirmation. The number of sequences associated to this inference is relatively small and it is anticipated that a final affirmation might assign it to either level 0 or level 1.



d. IGHV4-39*01_g315a (P1_I60)
This allele is currently not seen in VDJbase or featured in “Functional groups reference book” version 1 but it features as the sole allele of functional group G41 in this sample in “Functional groups reference book” version 2. As of Jan 13, 2022, IGHV4-39*01 is not featured in Functional Groups Reference Book (either version). It was not inferred by IgDiscover but the frequency of perfect matches to IGHV4-39*01 is lower than that of other subgroup 4 alleles, suggesting the presence of

two alleles of this gene in the genotype. Extensive scrutiny of reads associated to IGHV4-39*01 in the underlying data indicates the occurrence of two similarly sized populations, one with 315G and one with 315A, distinctly different from other alleles of IGHV4 subgroup alleles.

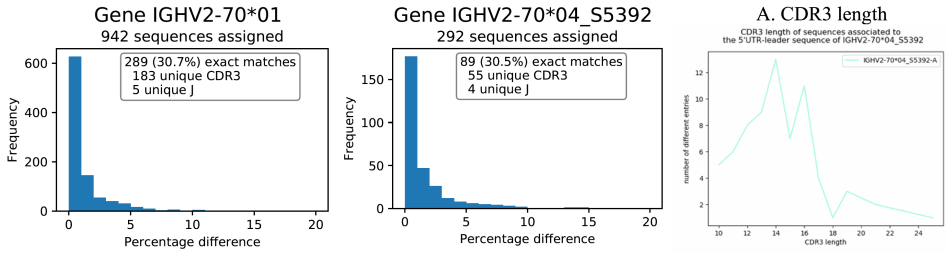
Decision: If this allele is to be assessed as a valid allele it must be identified and annotated in VDJbase. OGRDBstats plots for instance defining the alleles 3'-end will be required for its affirmation. It is anticipated that a final affirmation might assign it to level 1.



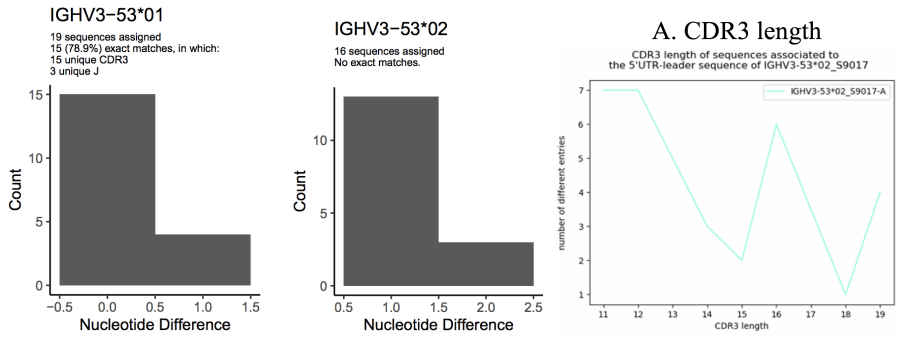
e. IGHV2-70*04 A14G (P1_I5)
Inferred in a separate study (DOI: 10.3389/fimmu.2021.730105) by IgDiscover in ERR2567182 (P1_I5) but it is not seen in VDJbase or reported in "Functional groups reference book". Reads associated with this inferred germline gene allele are diverse in CDR3 length and have a high level of unmutated reads, similar to the total frequency of exact matches in the same subject. Known alleles of IGHV2-70 carry either an A or a G at base position 14. No OGRDBStats is available for IGHV2-70*04 or variants thereof in VDJbase. This genotype also carries IGHV2-70*01, an allele that features three differences from IGHV2-70*04_S5392. These alleles are expressed at a 3.2:1 ratio. The inferred upstream regions of these two alleles are identical. The IMGT database also reports an incomplete allele (IGHV2-70*05) that in its defined sequence (that does not cover base 14) differs only in one position (base 116) from IGHV2-70*04_S5392. Due to its incompleteness, IGHV2-70*05 does not perform well during the sequence alignment procedure that is part of the inference process but there is no evidence from the processed data that IGHV2-70*05 should be the source of a major part of the transcriptome related to IGHV2-70 of this subject.

Decision: If this allele is to be assessed as a valid allele it must be identified and annotated in VDJbase. OGRDBstats plots will be

required for its affirmation. It is anticipated that a final affirmation might assign it to level 0 or level 1.



f. IGHV3-53*02 C259T (P1_I22)
This allele was inferred in a separate study (https://doi.org/10.3389/fimmu.2021.730105) by IgDiscover in ERR2567198 (P1_I21) but it is not seen in VDJbase or reported in “Functional groups reference book”. This sample in general carries few clonotypes of each gene. Nevertheless, the CDR3 length of this inferred allele is diverse. OGRDBstats plots indicate the presence of IGHV3-53*01 and IGHV3-53*02 in the genotype but all of the reads of the latter carry at least one mutation. The frequency of reads associated to each allele is similar.



Decision: If this allele is to be assessed as a valid allele it must be identified and annotated in VDJbase and access to full OGRDBstats plots will be required for final affirmation. The number of sequences associated to this inference is small and it is anticipated that a final affirmation might assign it at most to level 0.

7. **Upcoming AIRR conference - status and IARC strategy**
Discussion on the goals set at the 5th AIRR Community Meeting

Goals for 2021 and progress since 2020 AIRR meeting

- **Continue to evaluate human IGH, IGK, and IGL inferences** - ongoing; implementation of tools in VDJbase and Functional Groups Reference Book;
- **Establish a human TCR IARC** - initiated in collaboration with James Heather; it is challenging to engage the TCR community at large;
- **Establish IARC(s) for non-human species** - work initiated within the current IARC; first manuscript in preparation; process for non-positional naming of alleles in development;
- **Develop tools and guidelines for the evaluation of the 'quality' of submitted datasets** - ongoing; process mainly within the Software Working Group;
- **Streamline procedures for the submission and assessment of human IGHV inferences** - challenging to engage wider community in submission process; process established for submission to ENA - guidelines in place; features in VDJbase/OGRDB facilitates submission;
- **Develop strategies to make the work of the IARC(s) sustainable** - application for funding still under review; engagement by specialized teams (TCR; IG of other species etc.) will be required to create a sustainable organization moving forward.

8. **Next meeting**
   February 7th, 2022 at 11.00 UTC