Minutes of IARC meeting 132, February 12th, 2024

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

1. **Approval of minutes of meeting 131**
   Approved

2. **Next meeting**
   February 26, 2024 at 11.00 UTC

3. **The final decision to allow affirmation of novel alleles without GenBank accession numbers**
   Sequences can now be affirmed in the absence of GenBank accession numbers as long as data supporting the allele is available in a repository.

   > Currently, when considering a novel allele inferred from AIRR-seq data, IARC requires GenBank deposition of the submitted sequence. As implemented at NCBI, this deposition usually requires a set of supporting reads from the original repertoire sequencing. Both requirements can cause significant delay in the approval process, and limit the breadth of datasets from which inferences can be drawn.
   > Going forward, and in considering novel inferences derived from genomic source data as well as those derived from AIRR-seq data, IARC will require the original base data from which the inference was derived (for example the repertoire sequencing in the case of AIRR-seq data) to be deposited in a publicly-accessible repository with long-term funding, such as an INSDC repository, but will not usually require additional records to be created. The base data, together with the submission record and notes on OGRDB, will provide an adequate basis for reproducibility. It should be noted that OGRDB is regularly backed up to Zenodo, providing long-term archival in the event that OGRDB itself is discontinued. IARC retains the option to request additional data to be deposited prior to approval, and the discretion to determine which repositories are suitable.

   WL has proceeded with a change in OGRDB that will allow for submission of sequences in the absence of a GenBank accession number.

4. **AIRR-C Exec meeting in February (tentative periodic report)**
   Inferred Allele Review Committee

- **Current general objectives of IARC**
  - Affirmation of novel human IG AND TR alleles
  - Define human IGV germline allele reference sets
  - Aid definition of macaque IG germline set by providing expertise to other WG
  - Consolidation of the use of permanent identifiers
  - Further integration of genomic and transcriptomic data
  - Processes for allele submission and publication of germline gene reference sets
  - Manuscript(s) on novel alleles and IARC processes
  - Guidelines for the generation and growth of germline sets to establish formally recognized gene sets
  - Funding mechanisms moving forward
  - Consolidation of the human TCR IARC process
  - IARC sustainability.
- **Number of participants in IARC: 9 (W/M : 2/7)**
- **Main highlights**
  - **Germline gene sets**
    - An AIRR-C-endorsed paper describing the AIRR-C germline reference set has been published in Frontiers in Immunology (https://doi.org/10.3389/fimmu.2023.1330153).
    - The AIRR-C germline reference set at OGRDB can now be selected in the IgBLAST online tool

  - **Affirmation without GenBank accession number.** A process to allow affirmation of novel alleles without GenBank accession numbers has been approved as long as data supporting the allele is available in a sustainable repository.

- **Process for germline sets based on genomic data.** Roadmap for building Ig germline sets based on genomic data is being considered; evidence, assessment, and naming.
- **Publications**
  - Collins *et al.* (2024) **Front Immunol** 14, 1330153 (https://doi.org/10.3389/fimmu.2023.1330153)
  - Poster on the Germline Gene Set at PEGS Europe In Lisbon, Portugal.
- **Strength and weakness of the group**
  - \+ Dedicated team
  - \- Long-term sustainability and recruitment, incl. process for expansion to other species
- **Future plans**
  - Collaboration with IUIS TR-IG
  - Roadmap for use of data generated by novel tools as well as genomic data in affirmation processes
  - Provide expertise for generation of macaque germline reference set
  - Enhancement of germline gene sets
  - Affirmation of novel alleles

5. **TR-IG process**
   IARC co-lead has been invited to a future discussion with IUIS TR-IG. Presentation of future process to be taken by TR-IG and collaborating partners.

6. **AIRR-C-meeting in Porto 2024**
   a. report - to be prepared: Discussion
   b. IARC co-chairs 2024-2025: Discussion of election of two IARC co-chairs for next period. To be continued.

7. **Inference of IGKV3-15*01_G107A**
   https://ogrdb.airr-community.org/submission/S00044
   The inference was originally reported in Mikocziova et al. iScience 2021, 24, 103192 (https://doi.org/10.1016/j.isci.2021.103192).
   IGKV3-15*01_G107A was inferred in VDJbase sample P1_I12_S1. This sample's genotype that also carry IGKV3-15*01 and a substantially lower number of sequences of IGKV3D-15*01 (about 9-fold). The relatively high

expression of IGKV3-15*01_G107A suggests that the allele was found in the proximal and to the typically less expressed distal IGKV locus. It was supported by 811 unmutated sequences and 655 unique CDR3s in the unmutated sequence set. The allelic ratio for IGKV3-15*01_G107A / IGKV3-15*01 was 56/44. Haplotyping associated IGKV3-15*01_G107A to IGKJ2*01 while IGKV3-15*01 associated to IGKJ2*04. IGKV3D-15*01 associated to both haplotypes as defined by alleles of IGKJ2.
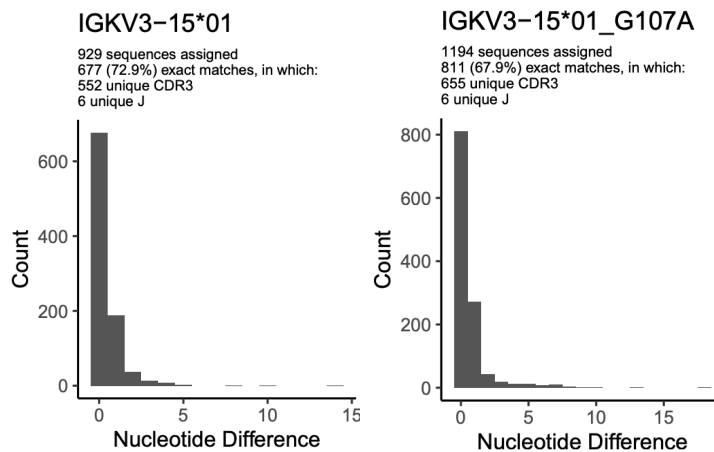
The sequence is also inferred in VDJbase P1_I43_S1 and P1_I91_S1 but no haplotyping data are available for those samples.
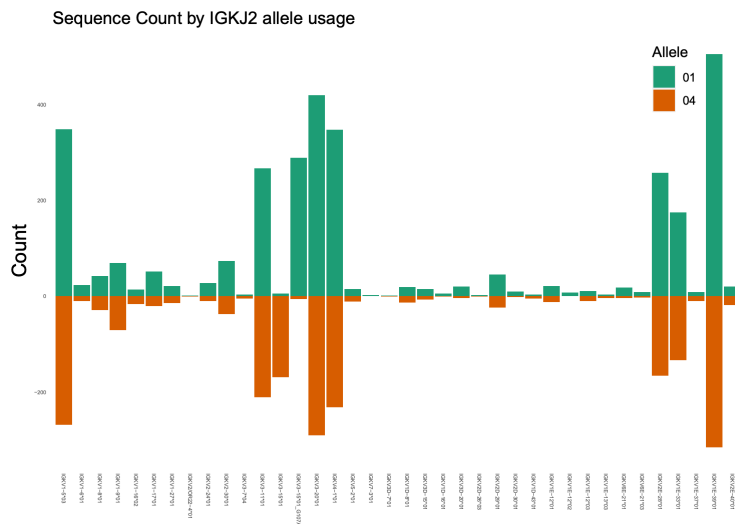
There is no entry in GenBank or among genomic entries in VDJbase that is identical to the full sequence of IGKV3-15*01_G107A.

>IGKV3-15*01_G107A
GAAATAGTGATGACGCAGTCTCCAGCCACCCTGTCTGTGTCTCCAGGGG
AAAGAGCCACCCTCTCCTGCAGGGCCAGTCAGAGTGTTAACAGCAACTT
AGCCTGGTACCAGCAGAAACCTGGCCAGGCTCCCAGGCTCCTCATCTAT
GGTGCATCCACCAGGGCCACTGGTATCCCAGCCAGGTTCAGTGGCAGTG
GGTCTGGGACAGAGTTCACTCTCACCATCAGCAGCCTGCAGTCTGAAGA
TTTTGCAGTTTATTACTGTCAGCAGTATAATAACTGGCCTCC

Data for P1_I12_S1:

IGKV3−15*01

929 sequences assigned
677 (72.9%) exact matches, in which:
552 unique CDR3
6 unique J

IGKV3−15*01_G107A

1194 sequences assigned
811 (67.9%) exact matches, in which:
655 unique CDR3
6 unique J

Sequence Count by IGKJ2 allele usage

A new VDJbase analysis is in progress, an analysis that will include an analysis of the 3'-end. CW will investigate the possible presence of this allele in unpublished genomic data that would support the inference and in particular the definition of the 3'-end. The inference is to be further discussed at the next IARC meeting.

8. Inference of IGKV3-15*01_A117G
https://ogrdb.airr-community.org/submission/S00043
The inference was originally reported in Mikocziova et al. iScience 2021, 24, 103192 (https://doi.org/10.1016/j.isci.2021.103192).
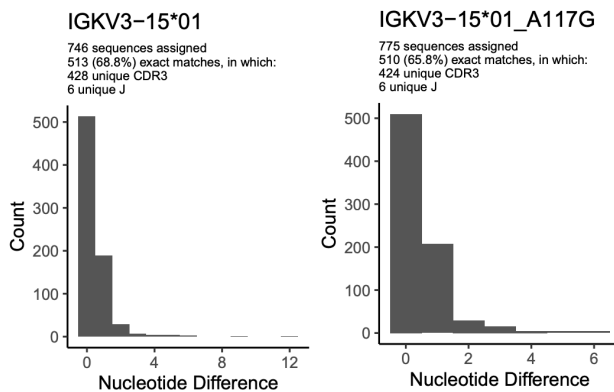
IGKV3-15*01_A117G was inferred in VDJbase sample P1_I24_S1. This sample's genotype that also carry IGKV3-15*01 and a substantially lower number of sequences of IGKV3D-15*01 (about 8-fold). The relatively high expression of IGKV3-15*01_A117G suggests that the allele was found in the proximal and to the typically less expressed distal IGKV locus. It was supported by 510 unmutated sequences and 424 unique CDR3s in the unmutated sequence set. The allelic ratio for IGKV3-15*01_A117G / IGKV3-15*01 was 51/49. Haplotyping associated IGKV3-15*01_A117G to IGKJ2*01 while IGKV3-15*01 associated to IGKJ2*03. IGKV3D-15*01 associated to both haplotypes as defined by alleles of IGKJ2.

The sequence is also inferred in VDJbase P1_I33_S1, P1_I43_S1, and P1_I76_S1but no haplotyping data are available for those samples.
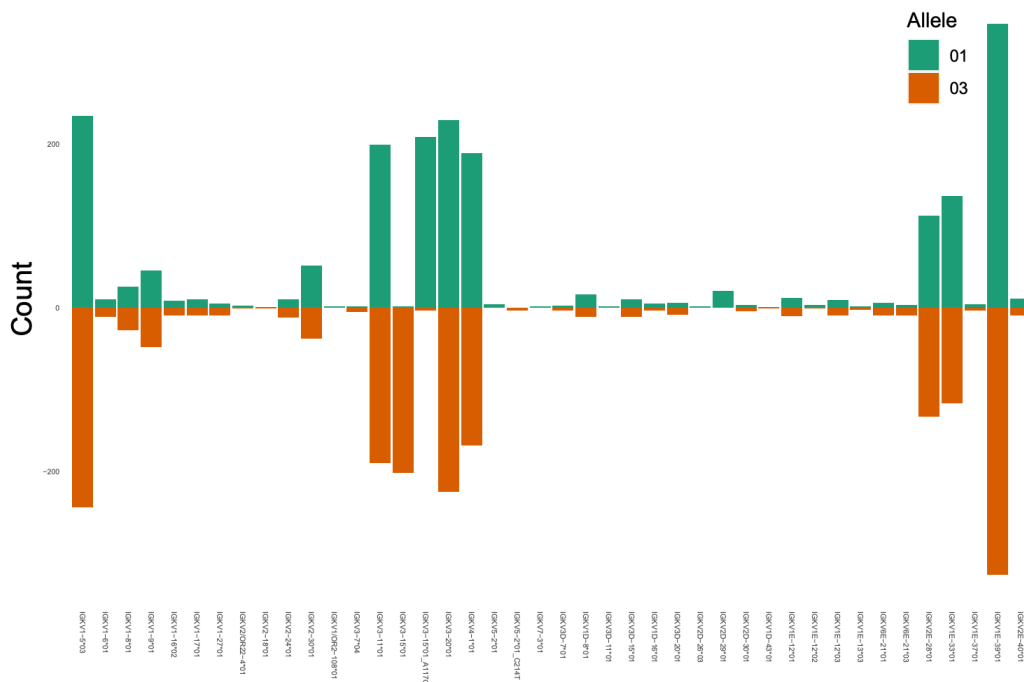
There is no entry in GenBank (apart from to rearranged sequences: MW718549 and MW718534) or among genomic entries in VDJbase that is identical to the full sequence of IGKV3-15*01_G107A.

>IGKV3-15*01_A117G

GAAATAGTGATGACGCAGTCTCCAGCCACCCTGTCTGTGTCTCCAGGGG
AAAGAGCCACCCTCTCCTGCAGGGCCAGTCAGAGTGTTAGCAGCAACTT
GGCCTGGTACCAGCAGAAACCTGGCCAGGCTCCCAGGCTCCTCATCTAT
GGTGCATCCACCAGGGCCACTGGTATCCCAGCCAGGTTCAGTGGCAGTG
GGTCTGGGACAGAGTTCACTCTCACCATCAGCAGCCTGCAGTCTGAAGA
TTTTGCAGTTTATTACTGTCAGCAGTATAATAACTGGCCTCC

Sequence Count by IGKJ2 allele usage

A new VDJbase analysis is in progress, an analysis that will include an
analysis of the 3'-end. CW will investigate the possible presence of this allele
in unpublished genomic data that would support the inference and in
particular the definition of the 3'-end. The inference is to be further discussed
at the next IARC meeting.