Agenda of IARC meeting 91, Jan 10th, 2022

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

1. **Approval of minutes of meeting 90**
   Approved

2. **Update on the mouse germline F1 manuscript**
   AC updated on the status of the manuscript and the path forward. WL will assess if identical alleles exist that may result in common multiple assignments that might compromise the analysis pipeline.

3. **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**
   AP described the most recent version of Functional groups reference book (https://ayeletperes.github.io/reference_book2/).

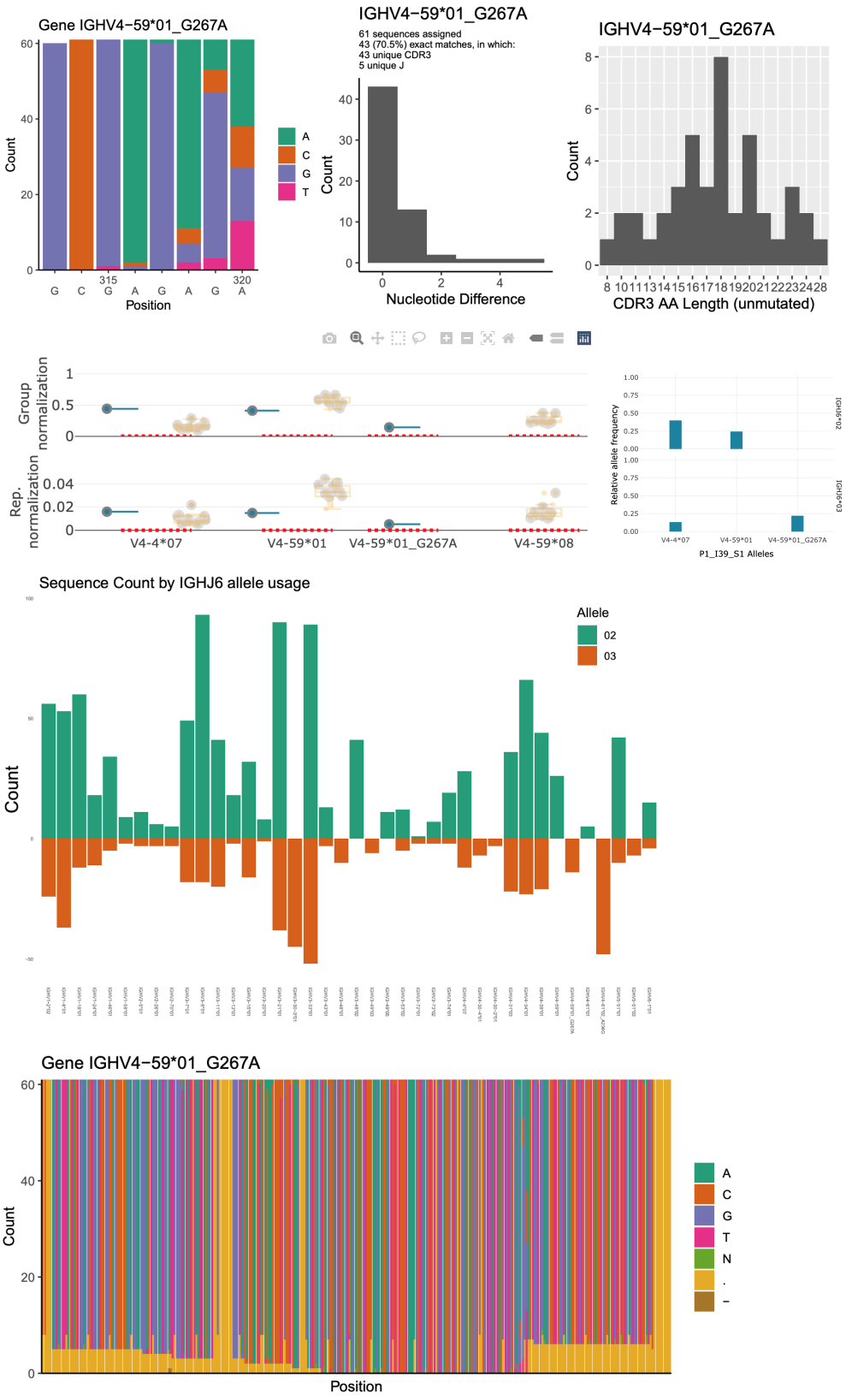
4. **Update on assessment of difficult to identify SNPs towards the 3'-end of alleles (see §4b of Meeting 90)**
   Difficulties to identify SNP's close to the 3'-end of some inferences (like IGHV3-15*01_a313t_c317g) may relate to read calls made by IgBLAST. AP will retrieve reads associated to the inference of IGHV3-15*01_a313t to assess how the supporting data supports the inference or its variant (IGHV3-15*01_a313t and IGHV3-15*01_a313t_c317g).

5. **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). These include

   a. IGHV4-59*01_g267a (P1_I39)

      This allele is associated to a diverse set of reads (including diverse CDR3 and CDR3 lengths) as recorded in VDJbase. It shows lower expression as compared to IGHV4-59*01. It haplotypes well in relation to IGHV4-59*01. The coverage of the gene in particular of the 5'-end

may need further evaluation. This allelic variant is not found at NCBI using BLAST with L2-sequence+inferred allele sequence as search sequence. The novel allele is inferable up to and including base 319.



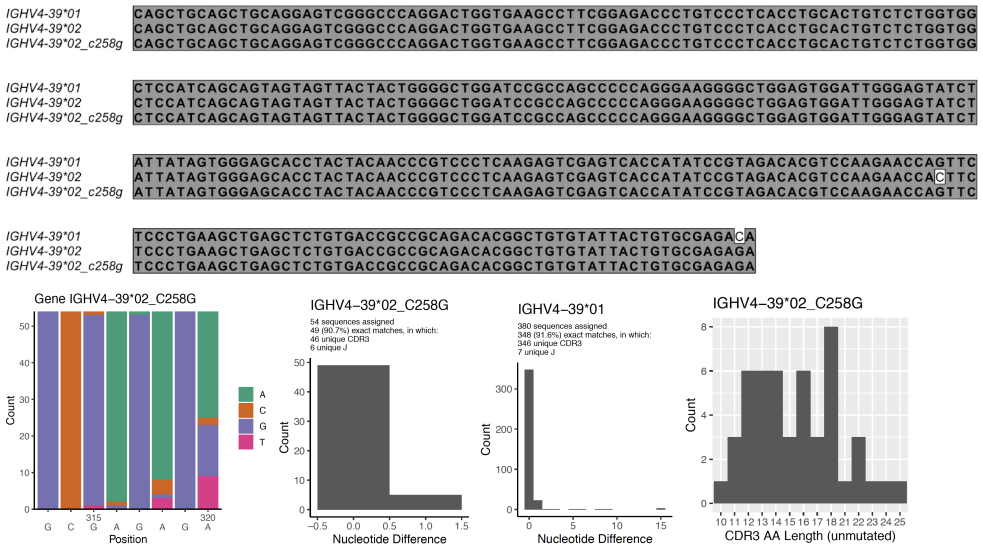
AP will extract the raw reads associated to the inference in question, to allow for an assessment of the lack of identification of some of the bases close to the 5’ end.

Decision: IGHV4-59*01_g267a is considered a valid inference that should, pending assessment of the allele’s 5’-end, move forward and be submitted to OGRDB to allow final assessment and possible affirmation as a Level 0 or level1 sequence.

b. IGHV4-39*02_c258g (IGHV4-39*01_c319g) (previously in VDJbase in haplotypable samples P1_I52, P1_I91; now in P1_I52 and in non-haplotypable sample P1_I65)

This allele is represented by a substantial number of reads, several unique CDR3s, CDR3 lengths, and IGHJs. However it is represented by much fewer sequences than the highly similar allele IGHV4-39*01 that is present in the genotype and from which it differs by only a single base at the penultimate base (position 319). “Functional groups reference book” does not differentiate between these alleles. Haplotyping (VDJbase) suggests that this variant is present in the same haplotypes as IGHV4-39*01. This allelic variant might just be inferred based on rearrangements that have been trimmed at the 3’-end?

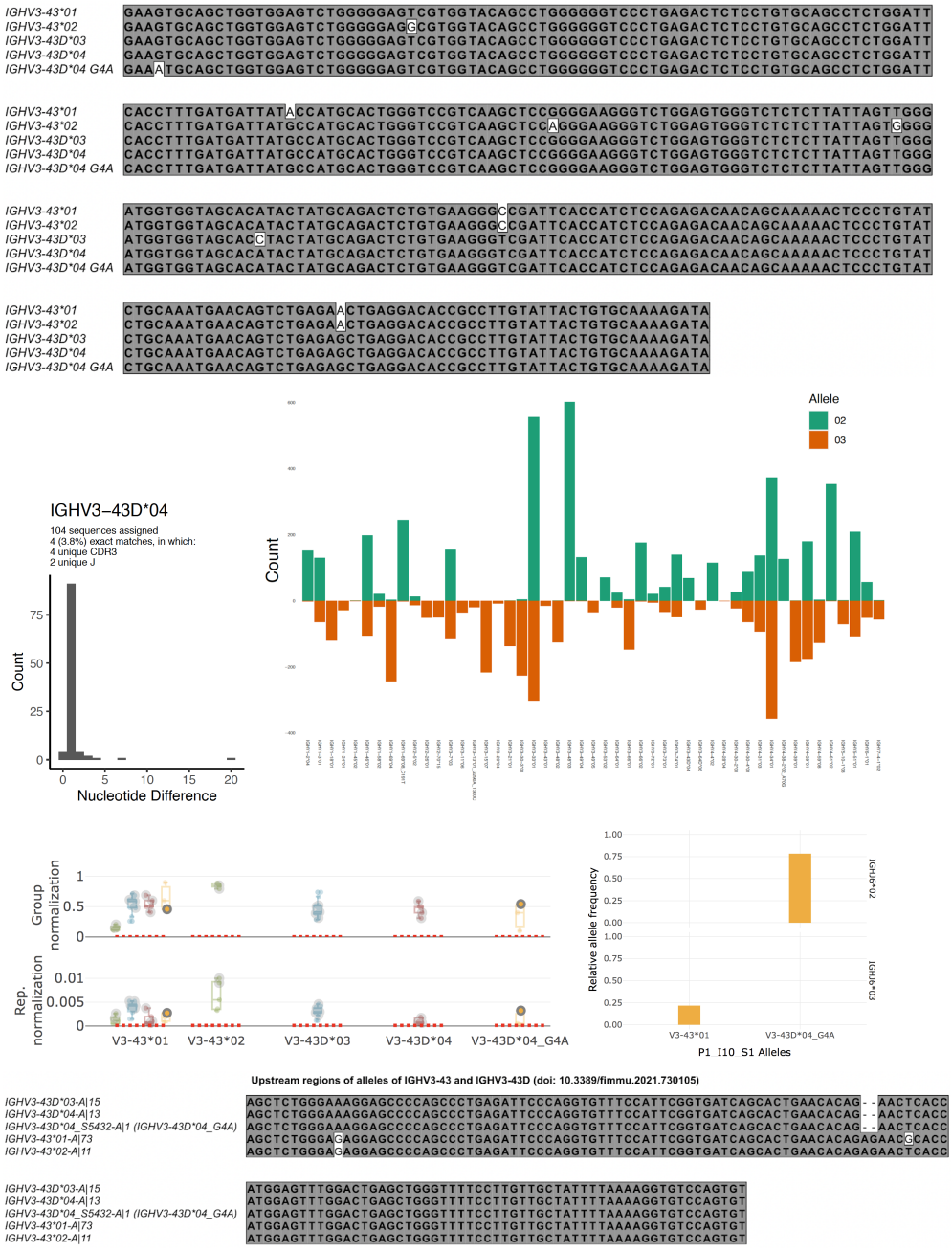
Decision: The validity of inferred allele IGHV4-39*02_c258g was questioned and it should not move forward and be submitted to OGRDB.



c. IGHV3-43D*04 G4A (P1_10)

This allele has not been inferred in sample P1_I10 as outlined in VDJbase (December 21, 2021). Rather, IGHV3-43D*04 was inferred in this sample but with almost all instances referring to reads that carry 1 base difference from IGHV3-43D*04. “Functional groups reference book” however identifies this allele and haplotyping indicates that it is present on the haplotype that does not carry the most similar allele IGHV3-43*01. This haplotype also lacks the adjacent gene IGHV4-39. The novel allele has also been independently inferred in this sample in a past study (doi: 10.3389/fimmu.2021.730105). It was shown to harbour an upstream region typical of other alleles of IGHV3-43D but not of IGHV3-43 (doi: 10.3389/fimmu.2021.730105) and it is present on a haplotype that also carries IGHV4-38-2, a gene that commonly accompanies IGHV3-43D. Altogether it is more likely that this allele is a variant of IGHV3-43D that of IGHV3-43. This is also a haplotype that carries a large deletion of genes ranging from IGHV1-8 to IGHV3-30. This allelic variant is not found at NCBI using BLAST with L2-sequence+inferred allele sequence as search sequence.

Decision: IGHV3-43D*04 G4A is considered a valid inference that should, pending appropriate annotation in VDJbase, move forward and be submitted to OGRDB to allow final assessment and possible affirmation as a level 1 sequence.

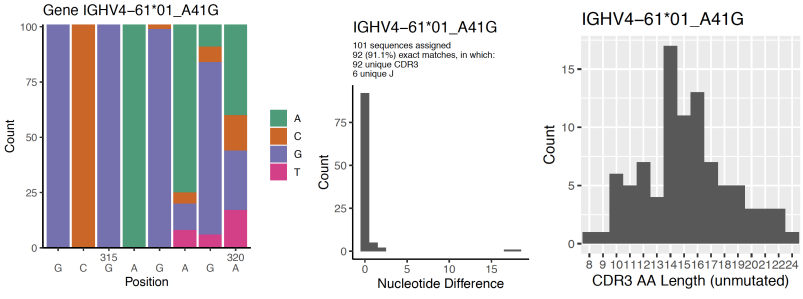


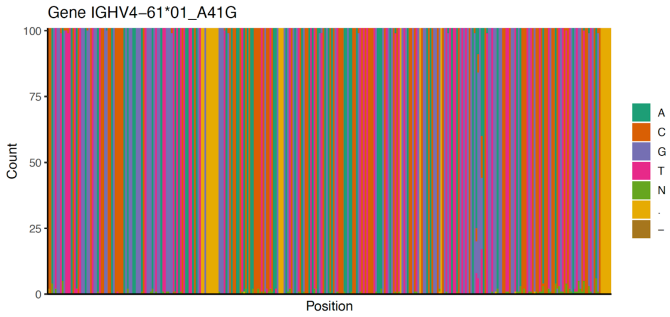
d. IGHV4-61*01_a41g (level 0 when assessed at meeting 63) (P1_I23)

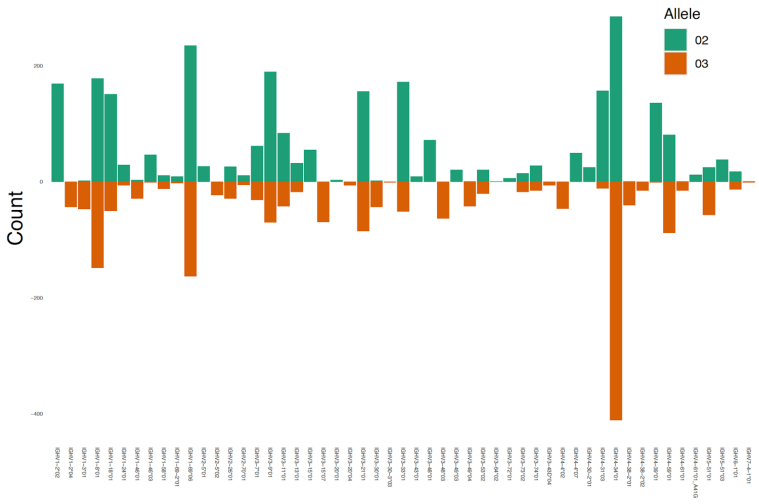
This allele is expressed at a level similar as somewhat poorly expressed allele IGHV4-61*01 (group normalized frequency 43% for IGHV4-61*01 A41G in “Functional groups reference book”), an allele

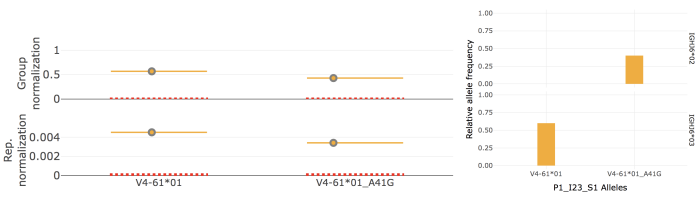
that is also present in the genotype. Multiple reads as exact matches, multiple CDR3 lengths, and appropriate haplotyping relates to this inference. The novel allele has also been inferred in this sample in a past study (doi: 10.3389/fimmu.2021.730105). This allelic variant is not found at NCBI using BLAST with L2-sequence+inferred allele sequence as search sequence. The novel allele is inferable up to and including base 319.

Decision: IGHV4-61*01_a41g is considered a valid inference that should move forward and be submitted to OGRDB to allow final assessment and possible affirmation as a level 1 sequence.





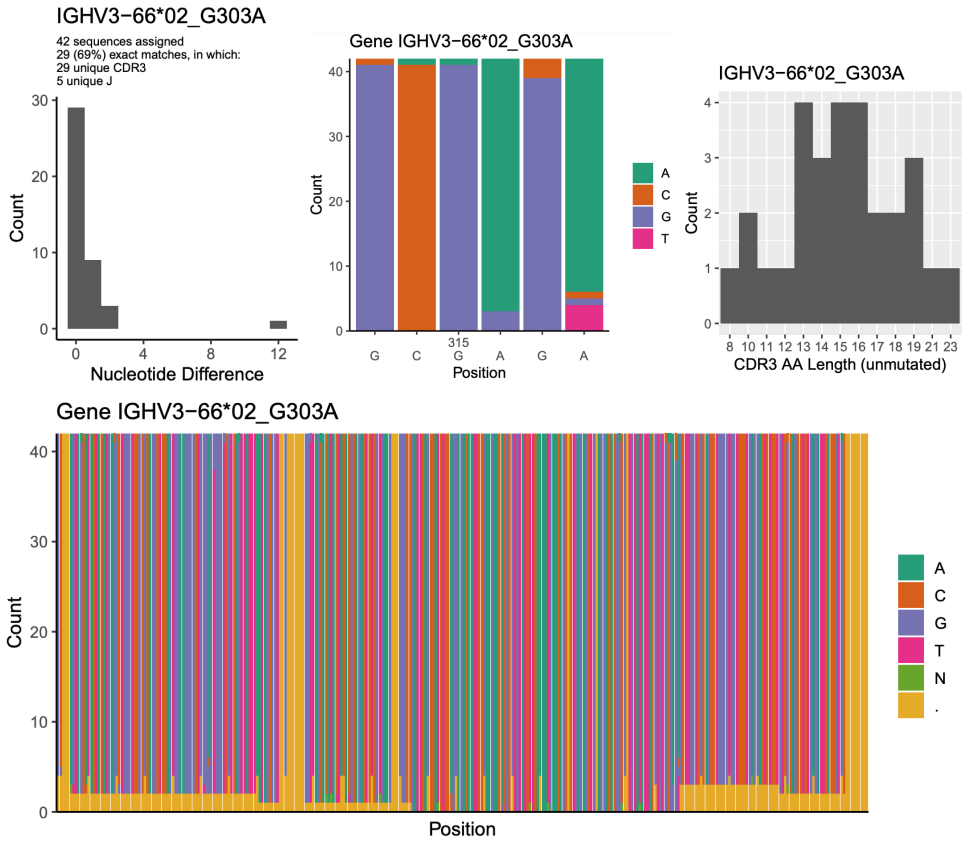


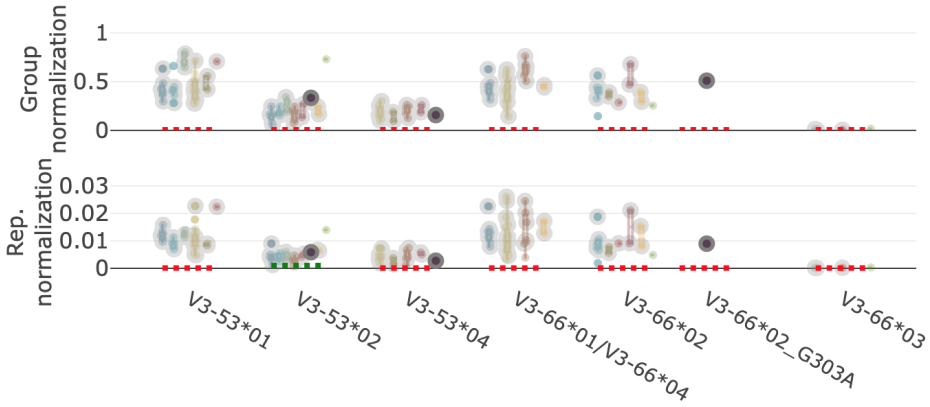


e. IGHV3-66*02_g303a (P1_I28)

This allele has been inferred in VDJbase only in one of the samples of P1. There are relatively few sequences assigned but these are diverse. The clone represents 51% of the G25 (IGHV3-53/IGHV3-66) group in the “Functional groups reference book”. The novel allele has also been inferred in this sample in a past study (doi: 10.3389/fimmu.2021.730105). The sequence is inferred up to and including base 318 as the IMGT reference sequence only extends up to and including base 318. This allelic variant is not found at NCBI using BLAST with the inferred allele sequence as search sequence. As in the case of IGHV4-59*01_g267a, further assessment of the 5’end is required as some bases are missing in the relevant plot.

Decision: IGHV3-66*02_g303a is considered a valid inference that should, pending assessment of the allele’s 5’-end, move forward and be submitted to OGRDB to allow final assessment and possible affirmation as a level 0 or level 1 sequence.

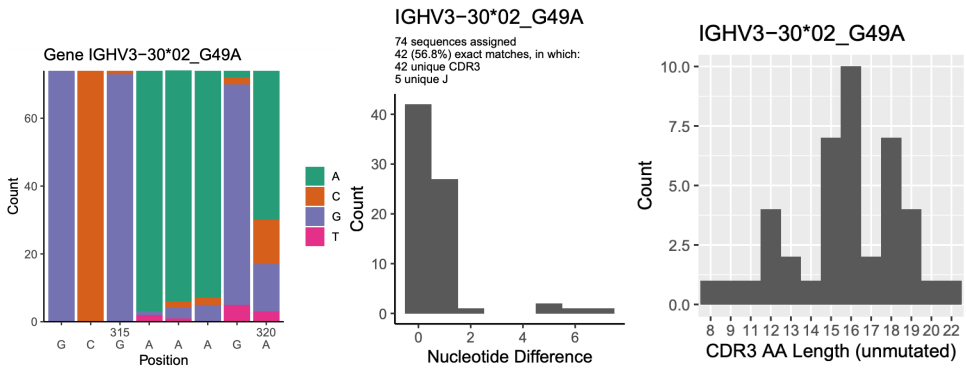


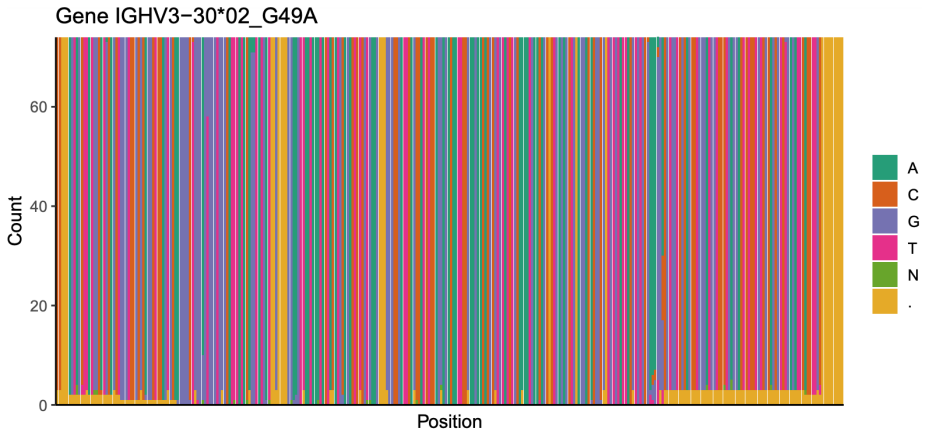


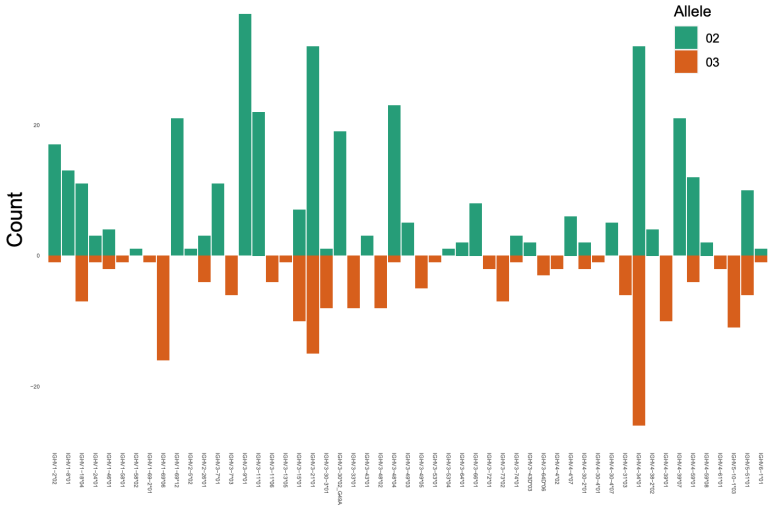
f. IGHV3-30*02_g49a (P1_I47)

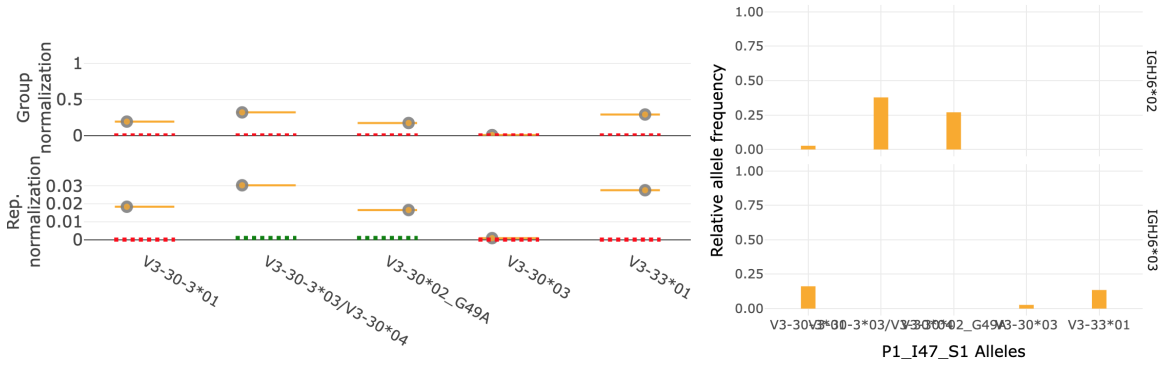
This allele is reported in VDJbase and “Functional groups reference book” and it has been independently inferred in this sample in a past study (doi: 10.3389/fimmu.2021.730105). It is represented by relatively few sequences but these are diverse. Haplotyping in VDJbase/OGRDBstats is complicated by the fact that IGHV3-30*18 is not featured in the current analysis by RabHit. A past study (doi: 10.3389/fimmu.2021.730105) identified a reasonable separation of alleles by haplotyping. As in the case of IGHV4-59*01_g267a, further assessment of the 5’end is required as some bases are missing in the relevant plot.

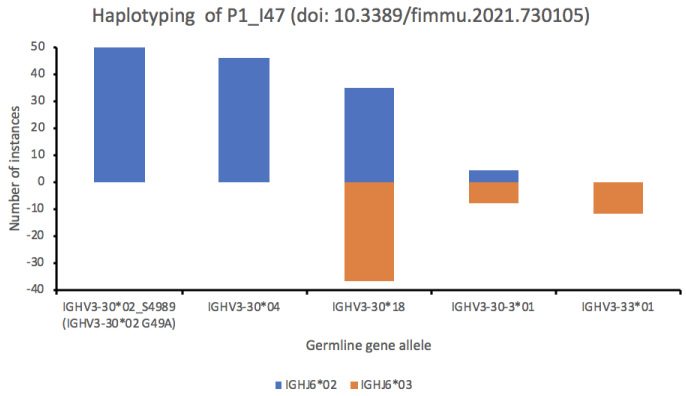
Decision: IGHV3-30*02_g49a is considered a valid inference that should, pending assessment of the allele’s 5’-end, move forward and be submitted to OGRDB to allow final assessment and possible affirmation as a level 0 or level 1 sequence.




Gene IGHV3−30*02_G49A
Count
Position
A
C
G
T
N
.


Allele
02
03
Count


Group normalization
Rep. normalization
V3-30-3*01
V3-30-3*03/V3-30*04
V3-30*02_G49A
V3-30*03
V3-33*01
Relative allele frequency
IGHJ6*02
IGHJ6*03
P1_I47_S1 Alleles


Haplotyping of P1_I47 (doi: 10.3389/fimmu.2021.730105)
Number of instances
IGHV3-30*02_S4989 (IGHV3-30*02 G49A)
IGHV3-30*04
IGHV3-30*18
IGHV3-30-3*01
IGHV3-33*01
Germline gene allele
IGHJ6*02
IGHJ6*03

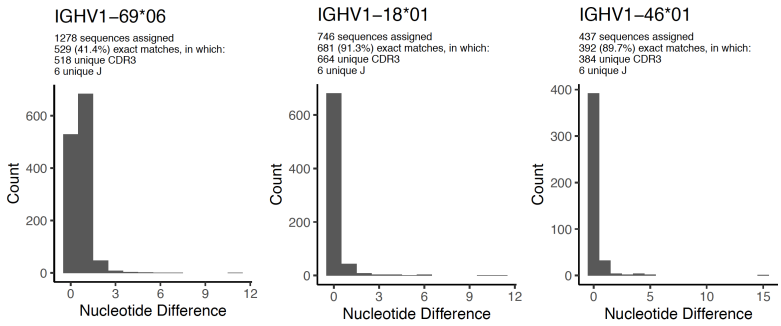
g. IGHV6-1*01_t91c (P1_I2, P1_I3 P1_I4)

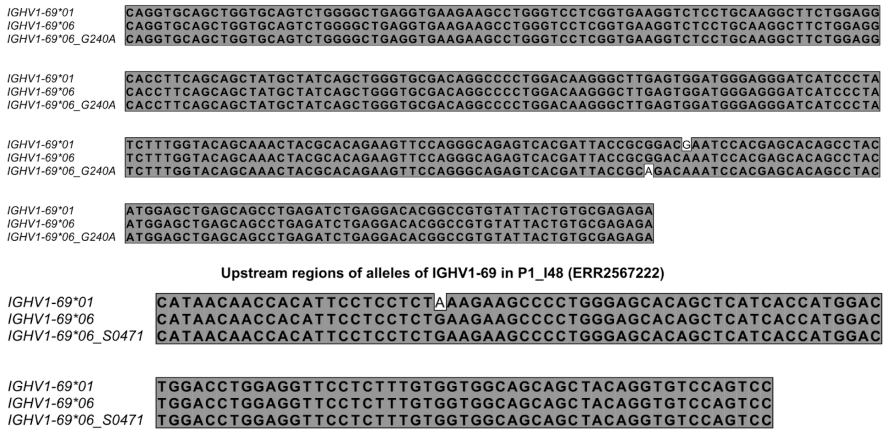
This allele is now absent in VDJbase (as of Dec 21, 2021). Four occurrences in “Functional groups reference book” (samples P1_I1, P1_I2, P1_I3, P1_I4) all with very few (1-9) assignments (maximum mutations allowed=0).

Decision: IGHV6-1*01_t91c should not move forward and be submitted to OGRDB.

h. IGHV1-69*06_g240a (level 1 when assessed at meeting 60) (P1_I48)

This allelic variant is currently not featured in VDJbase (but was defined in a past version of VDJbase) or “Functional groups reference book”. It has also been independently inferred in this sample in a past study (doi: 10.3389/fimmu.2021.730105). VDJbase indicates a genotype carrying IGHV1-69*01 and IGHV1-69*06 but with an unusually (compare with the profile of other alleles iof IGHV1) high frequency of sequences with one mutation in the case of IGHV1-69*06 (OGRDBstats plots for IGHV1-69*01 is currently not made available as the presence of the identical IGHV1-69D*01 prevents the generation of such data.) The past IgDiscover-based study (doi: 10.3389/fimmu.2021.730105) had previously identified IGHV1-69*01 and IGHV1-69*06 G240A in this sample but detailed analysis of the raw data suggests that the genotype carries IGHV1-69*01, IGHV1-69*06, and IGHV1-69*06 G240A of which the former has a separate upstream region. This allelic variant is not found at NCBI using BLAST with L2-sequence+inferred allele sequence as search sequence.





Decision: Further assessment of this inference is required before submission to OGRDB.

6. Next meeting

Jan 24th, 2022 at 11.00 UTC