

## Minutes of IARC meeting 100, June 8th, 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 99

Approved

### 2. Recap of AIRR-C VI meeting

Discussion on the role of IARC in AIRR-C and in relation to IUIS process. Discussion on legal aspects of processes to generate germline gene sets including use of existing germline sequence sets as part of that process. Discussion on how we could integrate genomic and transcriptomic data to generate unique datasets that are not in conflict with copywrite or existing intellectual database intellectual property claims.

### 3. The future combined role of genomic and inference data in IARC processes - IARC's role vs. that of GLDB

IARC should proceed with efforts to develop human germline sets in collaboration with GLDB. Discussion on how IARC and GLDB through outreach could recruit data and support to develop a sustainable position of their efforts.

### 4. Functional groups reference book - update

Testing and comparison with TIgGER in process. Discussion of how data provided in the reference book can impact intellectual discussion of repertoire studies including at the *WG2 - Evolutionary footprints in immunoglobulin V genes* working group meeting.

### 5. Use of the novel temporary label system in the case of novel inferred human alleles

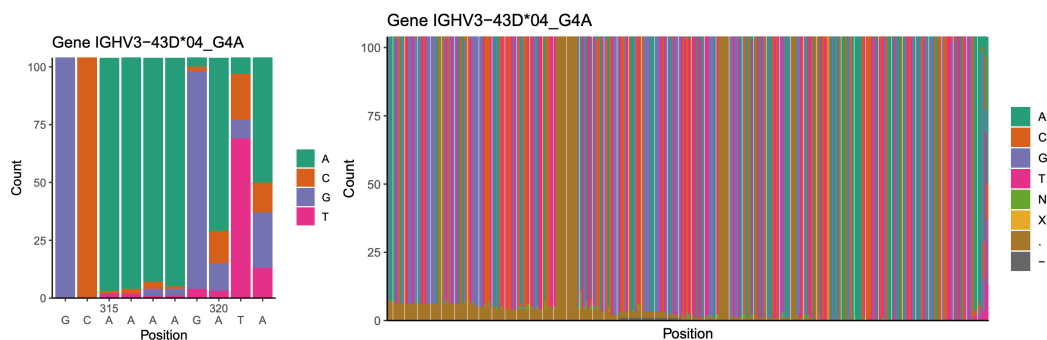
This is an option for further development but the system is not yet established for human sequences.

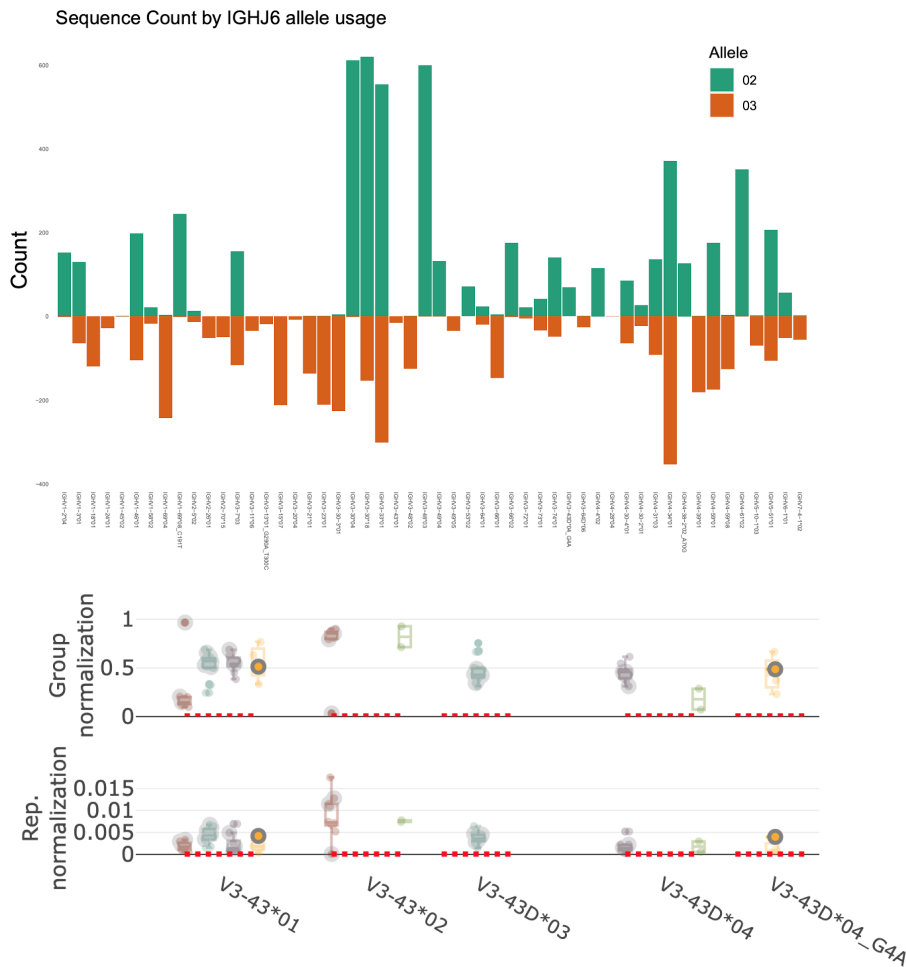
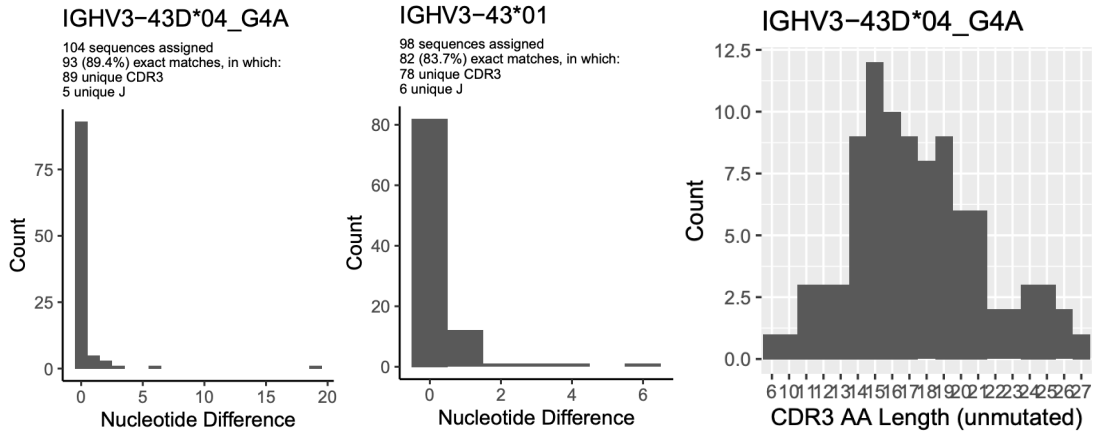
### 6. Assessment of inference IGHV3-43D\*04\_G4A in P1\_I10\_S1 (S00037)

IGHV3-43D\*04\_G4A was inferred in subject S10 (VDJ-base: P1\_I10\_S1). The genotype also carried IGHV3-43\*01 but no other allele of IGHV3-43D. The inference was supported by a relatively small number of sequences (104) and

unmutated sequences (93), a low overall frequency in the unmutated population (0.3%) and a small but diverse set of unique CDR3s (89) in the unmutated sequence set. Its allelic ratio was 100%. Haplotyping based on allelic diversity in IGJH6 was possible and the alleles distributed to one haplotype, a haplotype different from that associated with the allele IGHV3-43 (IGHV3-43D\*04\_G4A: 100:0; IGHV3-43\*01: 0:100). Reads of IGHV3-43D\*04\_G4A were associated to the same haplotype as those of IGHV4-38-2, a common duplication combination. Reads of IGHV3-43D\*04\_G4A were associated to an upstream region similar to that of other alleles of IGHV3-43D (but not similar to alleles of IGHV3-43) (doi: 10.3389/fimmu.2021.730105). Altogether, the association of this allele to IGHV3-43D is reasonable. IARC affirms the sequence at level 1 up to and including base 321 in agreement with past practice. It is acknowledged that the allele most likely carries one additional base, typically A at base position 322. A trailing "." indicates IARC's opinion that the sequence is likely to contain one additional 3'-nucleotides for which there is insufficient evidence to make an affirmation. The allele is given the name IGHV3-43D\*i02.

```
>IGHV3-43D*i02 (IGHV3-43D*04_G4A)
GAAATGCAGCTGGTGGAGTCTGGGGGAGTTCGTGGTACAGCCTGGGGGGTCCCTGAGA
CTCTCCTGTGCAGCCTCTGGATTCACCTTTGATGATTATGCCATGCACTGGGTCCGT
CAAGCTCCGGGAAGGGTCTGGAGTGGGTCTCTCTTATTAGTTGGGATGGTGGTAGC
ACATACTATGCAGACTCTGTGAAGGGTTCGATTCACCATCTCCAGAGACAACAGCAAA
AACTCCCTGTATCTGCAAATGAACAGTCTGAGAGCTGAGGACACCGCCTTGTATTAC
TGTGCAAAAAGAT.
```





>5'UTR-leader sequence of IGHV3-43D\*04\_G4A (doi: 10.3389/fimmu.2021.730105)

AGCTCTGGGAAAGGAGCCCCAGCCCTGAGATTCCCAGGTGTTTCCATTTCGGTGATCAGCACTGAACACA  
GAACTCACCATGGAGTTTGGACTGAGCTGGGTTTTCTTGTGCTATTTTAAAAGGTGTCAGTGT

Upstream region sequences of alleles of IGHV3-43/IGHV3-43D in the  
VDJbase P1 data (doi: 10.3389/fimmu.2021.730105)

	10	20	30	40	50	60	70
IGHV3-43D*03-A 15	AGCTCTGGGAAAGGAGCCCCAGCCCTGAGATTCCCAGGTGTTTCCATTCCGGTGATCAGCACTGAACACAG						
IGHV3-43D*04-A 13	AGCTCTGGGAAAGGAGCCCCAGCCCTGAGATTCCCAGGTGTTTCCATTCCGGTGATCAGCACTGAACACAG						
IGHV3-43D*04 G4A (S5432)-A 1	AGCTCTGGGAAAGGAGCCCCAGCCCTGAGATTCCCAGGTGTTTCCATTCCGGTGATCAGCACTGAACACAG						
IGHV3-43*01-A 73	AGCTCTGGGAAAGGAGCCCCAGCCCTGAGATTCCCAGGTGTTTCCATTCCGGTGATCAGCACTGAACACAG						
IGHV3-43*02-A 11	AGCTCTGGGAAAGGAGCCCCAGCCCTGAGATTCCCAGGTGTTTCCATTCCGGTGATCAGCACTGAACACAG						

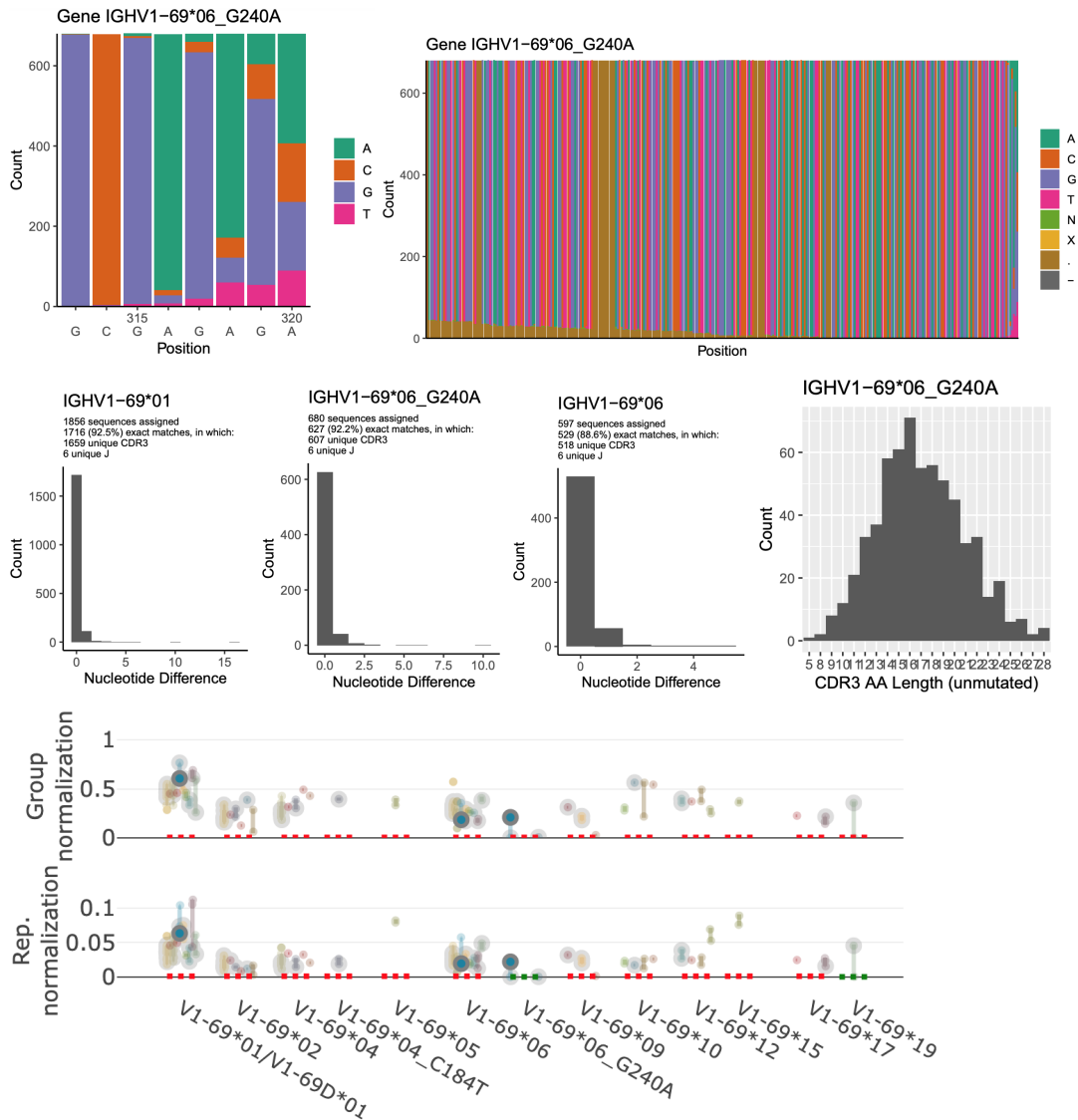
  

	80	90	100	110	120	130
IGHV3-43D*03-A 15	A - - ACTCACCATGGAGTTTGGACTGAGCTGGGTTTTCCCTTGTTCATTTTAAAAGGTGTCCAGTGT					
IGHV3-43D*04-A 13	A - - ACTCACCATGGAGTTTGGACTGAGCTGGGTTTTCCCTTGTTCATTTTAAAAGGTGTCCAGTGT					
IGHV3-43D*04 G4A (S5432)-A 1	A - - ACTCACCATGGAGTTTGGACTGAGCTGGGTTTTCCCTTGTTCATTTTAAAAGGTGTCCAGTGT					
IGHV3-43*01-A 73	AGAACTCACCATGGAGTTTGGACTGAGCTGGGTTTTCCCTTGTTCATTTTAAAAGGTGTCCAGTGT					
IGHV3-43*02-A 11	AGAACTCACCATGGAGTTTGGACTGAGCTGGGTTTTCCCTTGTTCATTTTAAAAGGTGTCCAGTGT					

## 7. Assessment of inference IGHV1-69\*06\_G240A in P1\_I48\_S1 (S00037)

IGHV1-69\*06\_G240A was inferred in subject S45 (VDJ-base: P1\_I48\_S1). The genotype also carried IGHV1-69\*01 and IGHV1-69\*06. Given the common duplication of IGHV1-69 in the human IGHV locus, this is not an unexpected finding. The inference was supported by a relative large number of sequences (680) and unmutated sequences (627), a high overall frequency in the unmutated population (2.4%) and a large and diverse set of unique CDR3s (607) in the unmutated sequence set. Its allelic ratio was 22%. Haplotyping based on allelic diversity in IGHJ6 was not possible. IARC affirms the sequence at level 1 up to and including base 319 in agreement with past practice. It is acknowledged that the allele most likely carries one additional base, typically A at base position 320. A trailing "." indicates IARC's opinion that the sequence is likely to contain one additional 3'-nucleotides for which there is insufficient evidence to make an affirmation. The allele is given the name IGHV1-69(D)\*i04. This name highlights the uncertainty of which of the two duplicated genes the allele is most likely associated with.

```
>IGHV1-69(D)*i04 (IGHV1-69*06_G240A)
CAGGTGCAGCTGGTGCAGTCTGGGGCTGAGGTGAAGAAGCCTGGGTCCCGGTGAAG
GTCTCCTGCAAGGCTTCTGGAGGCACCTTCAGCAGCTATGCTATCAGCTGGGTGCGA
CAGGCCCTTGACAAGGGCTTGGAGTGGATGGGAGGGATCATCCCTATCTTTGGTACA
GCAAACACTACGCACAGAAGTTCCAGGGCAGAGTCACGATTACCGCAGACAAATCCAGG
AGCACAGCCTACATGGAGCTGAGCAGCCTGAGATCTGAGGACACGGCCGTGTATTAC
TGTGCGAGAG.
```

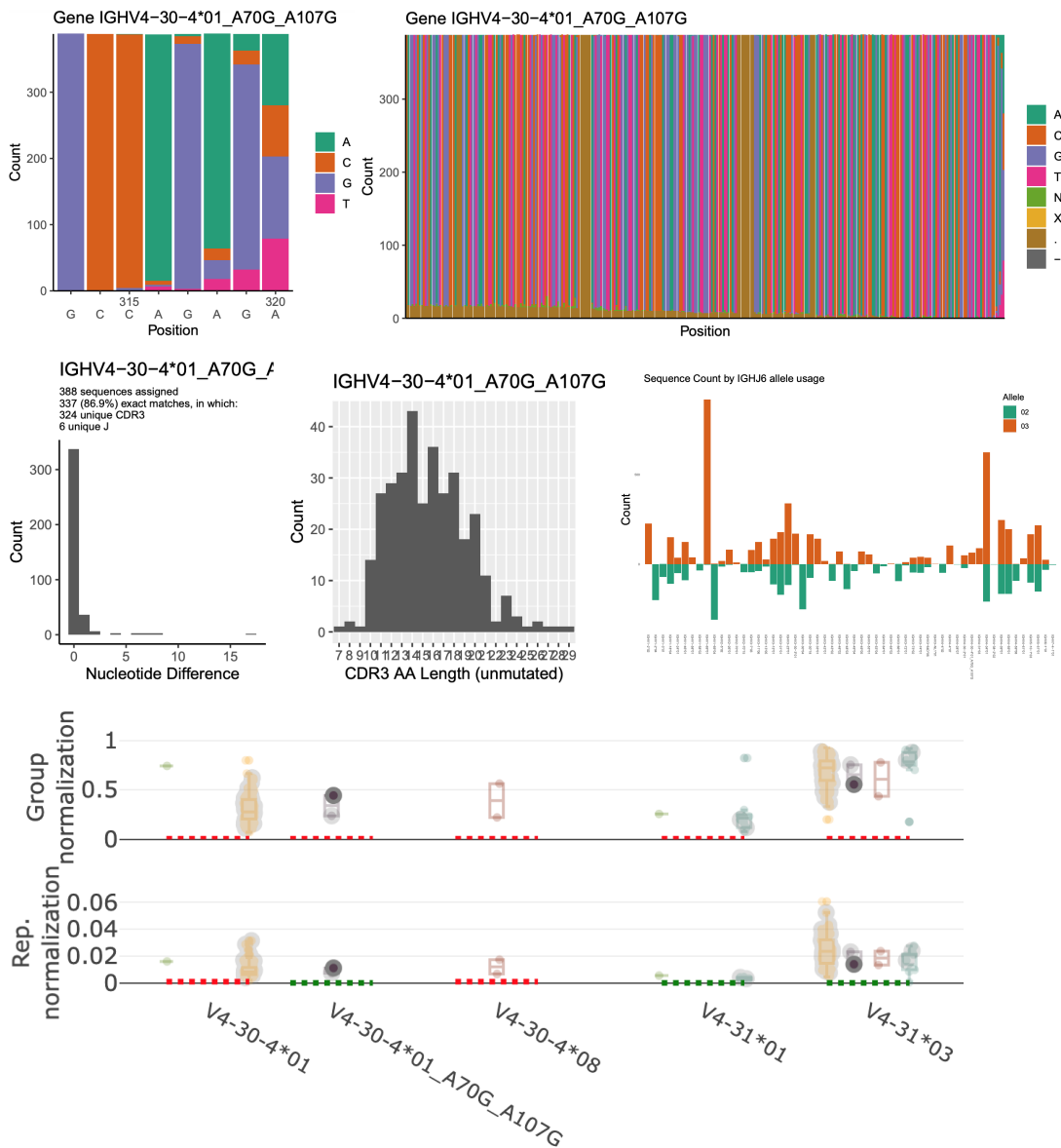


## 8. Assessment of inference IGHV4-30-4\*01\_A70G\_A107G in P1\_I41\_S1 (S00037)

IGHV4-30-4\*01\_A70G\_A107G was inferred in subject S38 (VDJ-base: P1\_I41\_S1). The genotype carried no other allele of IGHV4-30-4. The inference was supported by a relative large number of sequences (388) and unmutated sequences (337), a high overall frequency in the unmutated population (0.95%) and a large and diverse set of unique CDR3s (324) in the unmutated sequence set. Its allelic ratio was 100%. Haplotyping based on allelic diversity in IGHJ6 was possible. All reads were associated to one haplotype, a haplotype that also carried IGHV4-30-2, IGHV3-30-3, and IGHV4-31. IARC affirms the sequence at level 1 up to and including base 319 in agreement with past practice. It is acknowledged that the allele most likely carries one additional base, typically A at base position 320. A trailing “.” indicates IARC’s opinion that the sequence is likely to contain one additional

3'-nucleotides for which there is insufficient evidence to make an affirmation.  
The allele is given the name IGHV4-30-4\*i02.

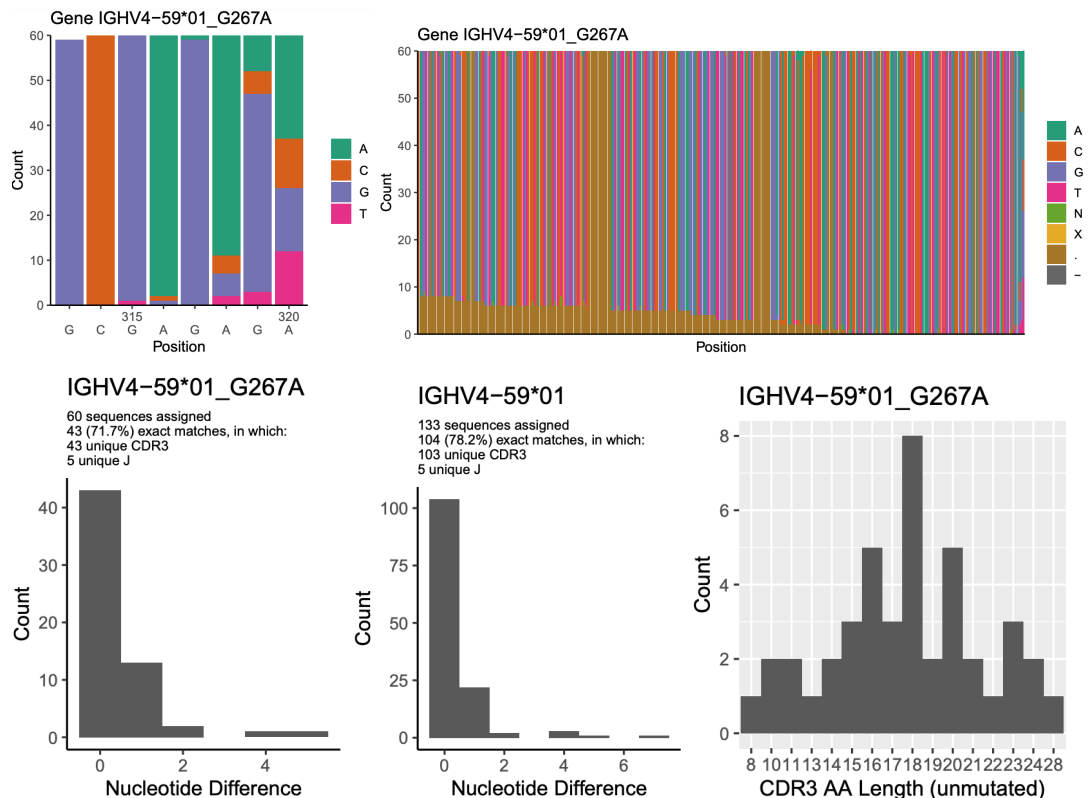
```
>IGHV4-30-4*i02 (IGHV4-30-4*01_A70G_A107G)
CAGGTGCAGCTGCAGGAGTCGGGCCAGGACTGGTGAAGCCTTCACAGAC
CCTGTCCCTCACCTGCGCTGTCTCTGGTGGCTCCATCAGCAGTGGTGGTT
ACTACTGGAGTTGGATCCGCCAGCCCCAGGGAAGGGCCTGGAGTGGATT
GGGTACATCTATTACAGTGGGAGCACCTACTACAACCCGTCCCTCAAGAG
TCGAGTTACCATATCAGTAGACACGTCCAAGAACCAGTTCTCCCTGAAGC
TGAGCTCTGTGACTGCCGCAGACACGGCCGTGTATTACTGTGCCAGAG.
```



**9. Assessment of inference IGHV4-59\*01\_G267A in P1\_I39 (S00037)**

IGHV4-59\*01\_G267A was inferred in subject S36 (VDJ-base: P1\_I39\_S1). The genotype carried one other allele of IGHV4-59. This sample carries few IGHV reads. The inference was consequently supported by a relative small number of sequences (60) and unmutated sequences (43), a relatively low overall frequency in the unmutated population (0.58%) and a small but diverse set of unique CDR3s (43) in the unmutated sequence set. Its allelic ratio was 31%. Haplotyping based on allelic diversity in IGHJ6 was possible with excellent separation from reads associated to IGHV4-59\*01 (IGHV4-59\*01\_G267A: 0:100; IGHV4-59\*01: 100:0). IARC postpones affirmation of the allele to a later meeting.

```
>IGHV4-59*01_G267A
CAGGTGCAGCTGCAGGAGTCGGGCCAGGACTGGTGAAGCCTTCGGAGACCCTGTCC
CTCACCTGCACTGTCTCTGGTGGCTCCATCAGTAGTTACTACTGGAGCTGGATCCGG
CAGCCCCAGGGAAGGGACTGGAGTGGATTGGGTATATCTATTACAGTGGGAGCACC
AACTACAACCCCTCCCTCAAGAGTCGAGTCACCATATCAGTAGACACGTCCAAGAAC
CAGTTCTCCCTAAAGCTGAGCTCTGTGACCGCTGCGGACACGGCCGTGTATTACTGT
GCGAGAG.
```





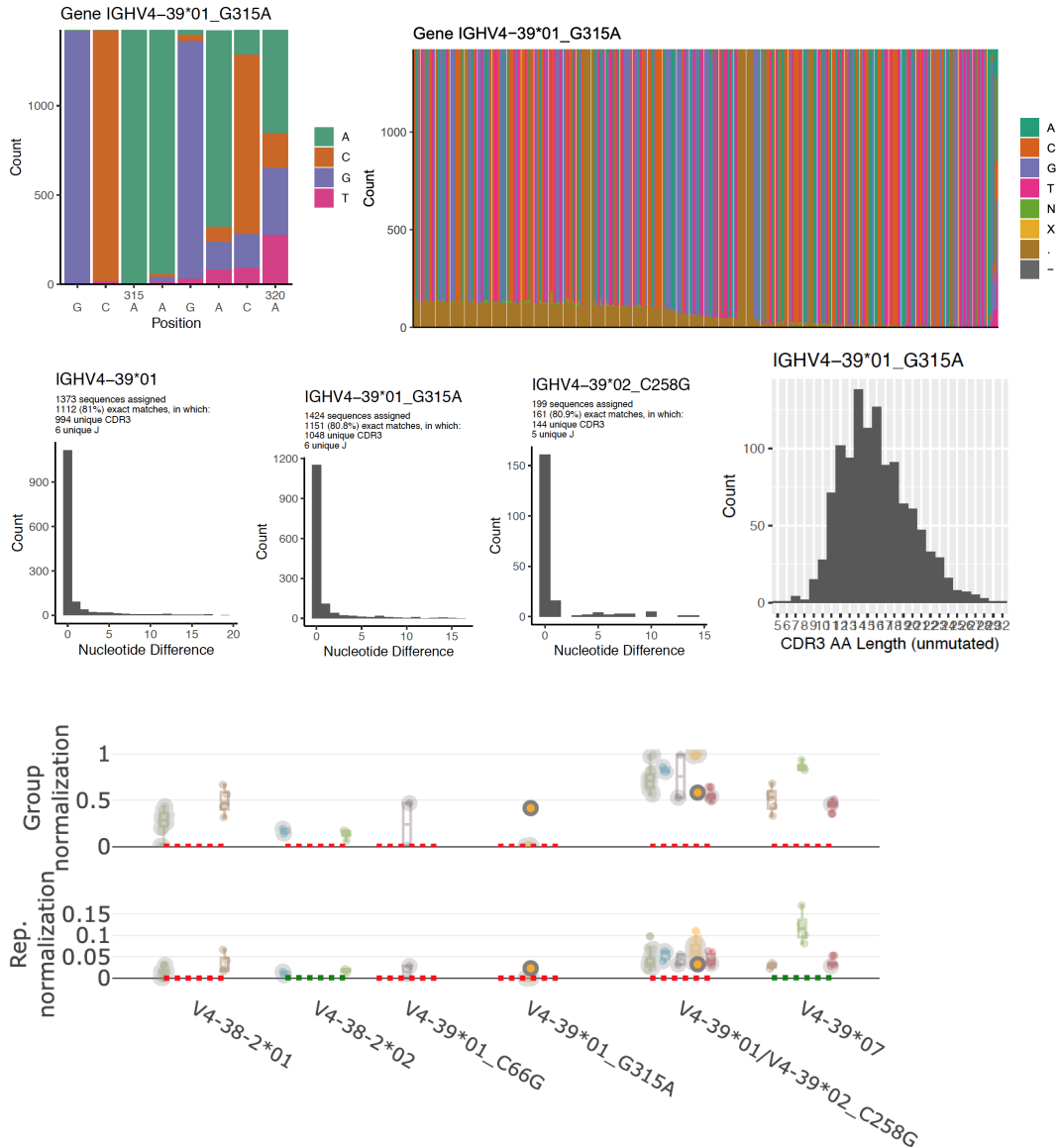
## 10. Assessment of inference IGHV4-39\*01 G315A in P1\_I60 (S00037)

IGHV4-39\*01\_G315A was inferred in subject S57 (VDJ-base: P1\_I60\_S1). The genotype carried one other allele of IGHV4-39 (IGHV4-31\*01, and a variant allele with a low number of associated reads associated that likely represents 3'-trimmed reads of IGHV4-39\*01). The inference was supported by a large number of sequences (1424) and unmutated sequences (1151), a high overall frequency in the unmutated population (2.47%) and a large set of unique diverse CDR3s (1048) in the unmutated sequence set. It represented 48% of sequences associated with the gene. Haplotyping based on allelic diversity in IGJ6 was not possible. IARC affirms the sequence at level 1 up to and including base 319 in agreement with past practice. It is acknowledged that the allele most likely carries one additional base, typically A at base position 320. A trailing "." indicates IARC's opinion that the sequence is likely to contain one additional 3'-nucleotides for which there is insufficient evidence to make an affirmation. The allele is given the name IGHV4-39\*i03.

```
>IGHV4-39*01_G315A
```



CAGCTGCAGCTGCAGGAGTCGGGCCAGGACTGGTGAAGCCTTCGGAGACCCTGTCC  
 CTCACCTGCACTGTCTCTGGTGGCTCCATCAGCAGTAGTAGTTACTACTGGGGCTGG  
 ATCCGCCAGCCCCAGGGAAGGGGCTGGAGTGGATTGGGAGTATCTATTATAGTGGG  
 AGCACCTACTACAACCCGTCCCTCAAGAGTCGAGTCACCATATCCGTAGACACGTCC  
 AAGAACCAGTTCTCCCTGAAGCTGAGCTCTGTGACCGCCGCAGACACGGCTGTGTAT  
 TACTGTGCAAGAC.



(the few reads associated to IGHV4-39\*02\_C258G likely represent mostly unmutated reads of IGHV4-39\*01 that have trimmed away two bases at the 3'-end and added a G at position 319).

## 11. Tentative submission of additional inferences of VDJbase P1

We should look into submission of the following inferred sequences

- a. IGHV3-30\*04 C201T G317A/IGHV3-30\*18 G113C C114T (P1\_I70)
- b. IGHV3-13\*01\_G290A\_T300C (P1\_10)
- c. IGHV4-61\*01 A41G (P1\_I23)

## **12. Next meeting**

To be decided