

Appendix C: AIRR-C Sub-committee (SC) and Working Group (WG) Reporting Template

Current

Date of this report: April 20, 2022

SC/WG Name: Inferred Allele Review Committee (IARC)

SC/WG Co-leaders: Mats Ohlin

SC/WG Active Members (list): Ayelet Peres, Gur Yaari, Martin Corcoran, Andrew Collins, Corey Watson, William Lees, Mats Ohlin, James Heather (co-opted)

Purpose: to review human immunoglobulin germline gene inferences from AIRR-Seq datasets, and to communicate affirmed sequences to the IG gene naming authorities.

Goals:

- Continue to evaluate human IGH, IGK, and IGL inferences
- Establish a human TCR IARC
- Establish IARC(s) for non-human species
- Develop tools and guidelines for the evaluation of the 'quality' of submitted datasets
- Streamline procedures for the submission and assessment of human IGHV inferences
- Develop strategies to make the work of the IARC(s) sustainable

Long-term vision and how WG products integrate with the AIRR-C mission:

IARC provides resources and knowhow that enable correct annotation and analysis of IG and TR transcriptomic data, in agreement with the AIRR Community's mission to standardize, analyze and share AIRR-seq data.

Products (if any):

- 9 inferred immunoglobulin alleles submitted to and approved by IUIS
- A published commentary on the use of short read sequencing for AIRR allele discovery
- A submitted manuscript on Balb/c germline alleles identified by inference (co-authored with other WGs)
- Germline gene sets for Balb/c and C57BL/6 mouse strains published on OGRDB

Resources (if any):

- OGRDB with enhanced functionalities
 - Gene set feature being added
 - Better support for submission
 - Improved OGRDBStats plot functionality.
- A Functional Groups Reference Book app to support IARC's assessment of inferred germline genes

Progress report on current purpose, goals, products and resources:

Since AIRR Community Meeting V, IARC has contributed the following activities/outputs.

- IARC has, as of April 19, 2022, met 36 times since AIRR Community Meeting V.
- Continue to evaluate human IGH IGK, and IGL inferences
 - 9 inferences of IGHV alleles have been affirmed, reported to IUIS and included in the IMGT reference set
 - A Functional Groups Reference Book App is being developed (Ayelet Peres and Gur Yaari) to aid future assessment of novel, inferred alleles by IARC.
 - We have initiated the integration of inferred and genomic data for enhanced understanding of AIRRs.
- Establish a human TCR IARC
 - The process to establish a TCR IARC has been initiated. James Heather has been co-opted to IARC with the intention to bring additional TCR expertise into the team and to transfer expertise of inference, affirmation, and process
 - We have initiated the assessment of novel inferred alleles of the TRBV locus.
- Establish IARC(s) for non-human species
 - Alleles of germline genes of mouse and macaque are currently mostly dealt with within the existing IARC as several members are experts on these germline gene repertoires.
 - Members of IARC co-authored a manuscript on a mouse Balb/c IGHV gene set
 - Balb/c and C57BL/6 expressed germline gene sets have been published on OGRDB.
- Develop tools and guidelines for the evaluation of the 'quality' of submitted datasets
 - In process in collaboration with Software WG.
 - New functionalities within OGRDB (enhanced plots, germline sets etc.) and Functional Groups Reference Book will aid submission and assessment of novel alleles.
 - IARC authored a commentary on the use of short read sequence data for AIRR allele discovery (doi: 10.1038/s41435-021-00152-6).
 - Members of IARC co-authored manuscripts on sequencing and annotation.
 - Members of IARC co-developed a temporary allele label system (manuscript in progress).
- Streamline procedures for the submission and assessment of human IGHV inferences
 - OGRDB has been integrated with VDJbase to simplify submission of VDJbase inferred alleles.
 - Tooling to create record sets for NIH/ENA submission has been improved, and documentation explaining the submission process has been rewritten.
 - A submission process with ENA has been established.
- Develop strategies to make the work of the IARC(s) sustainable
 - An application for funding of infrastructure has been submitted (decision pending).

Proposed plans for the coming interval:

Purpose: To generate resources for and advocate for better practice in and understanding of IG and TR annotation and analysis in the community interested in studies of AIRR.

Goals:

- To affirm additional novel human IGHV, IGKV, IGLV, and TRV alleles, and extend partial genes.
- To in collaboration with other WG identify expressed AIRR alleles and publish human IGV germline allele sets
- To, by providing expertise to other WGs/IARCs, consolidate a macaque IG germline set.
- To, in collaboration with other WGs, consolidate the use of temporary labels in instances where formal gene assignment of alleles cannot be made with confidence.
- To further integrate genomic and transcriptomic data of genes and alleles of IG loci.
- To, in collaboration with other WGs, establish processes for allele submission and germline gene sets of other species to OGRDB.
- To submit manuscript(s) on novel alleles and IARC processes.
- To publish guidelines for generation and growth of germline sets to establish formally recognized gene sets.
- To establish funding mechanisms moving forward.
- To consolidate human TCR IARC process.
- To develop IARC sustainability.

Products (if any):

- Affirmed alleles of AIRR germline genes, submitted to IUIS for final approval.
- Published manuscripts of relevance to the IARC's responsibilities.
- Germline gene sets (human and macaque).

Long-term vision and how WG or SC products integrate with the AIRR-C mission:

IARC provides resources and knowhow that enable correct annotation and analysis of IG and TR transcriptomic data, in agreement with the AIRR Community's mission to standardize, analyze and share AIRR-seq data

Proposed SC/WG Co-leaders: Mats Ohlin