Summary of the Adaptive Immune Receptor Repertoire (AIRR)
2016 Community Meeting
June 27-30 at NIH Fishers Lane Facility in Rockville, MD

The Adaptive Immune Receptor Repertoire (AIRR) Community has been organizing since a planning meeting of about 20 people in 2014, which set the agenda for a full Community meeting of about 70 people in Vancouver in 2015 (airr-community.org). Three Working Groups (**WG**) were established at the 2015 meeting: Common Repository for sharing AIRR Data; Minimal Standards for publishing and depositing AIRR data in a common repository; and Tools and Resources for producing and analyzing AIRR data. The WGs were tasked with researching their topics, consulting with the Community, and bringing recommendations to the 2016 AIRR Community Meeting.

The Organizing Committee sent invitiations to an email list of over 500 researchers, mostly involved in producing and using AIRR data, but also experts in consent, legal, intellectual property (**IP**), and computer security issues. The 4-day meeting had 116 regular registrants, and 20-45 webinar registrants for each of its six sessions. Each of the three WGs presented its preliminary recommendations on the first day; these were initially discussed, then revisited in each corresponding session during the second and third days. Finally, on the fourth day, revised recommendations were presented by each WG, discussed and voted upon. This summary provides a short description of the recommendations, as well as plans for each WG's activities over the coming year. The Community also discussed and voted to become an independent society, or join an existing society, and to reconvene in 2017.

Summaries of the recommendations from the three WGs, along with plans for the future of the Community Initiative developed by the Organizing Committee, are listed in point form below, and follow the full meeting agenda, which can be found at:
https://palladianpartners.cvent.com/AIRR2016

Common Repository Working Group:
(Co-leaders: Lindsay Cowell & Corey Watson)

The goal of the AIRR Common Repository Working Group (**CRWG**) was to develop a set of recommendations that promote the deposition, sharing, and use of B-cell and T-cell receptor sequence data in public data repositories. Over the course of 2015/2016, the CRWG met bi-weekly to debate issues and ideas, culminating in a draft of recommended guidelines, which were presented and discussed in detail at the AIRR 2016 Community Meeting. These have since been refined by the CRWG to reflect the concerns and issues raised by the Community. The guidelines comprise a total of 14 recommendations intended to:
      (1) State general principles for sharing of AIRR data;
(2) Define the characteristics of AIRR compliant repositories for data deposition, storage and access; and
      (3) Describe a distributed model for compliant repositories for AIRR data, linked by a   central registry.

Recommendations
The following includes a summary of recommendations for which the CRWG reports there was general consensus amongst community members. In many instances the details have not been finalized, and will be the subject of future efforts of the CRWG over the coming year.

• The CRWG recommends that, as a requisite for publication, data should be deposited in a public open-access database with no/few restrictions over deposition, access, storage, curation, and use. Depositors of data and repositories should have no right to interfere with access to and use of the data by others, including through the assertion of any IP rights. Exceptions to open-data sharing should only be considered in rare circumstances.

o Specifically, in conjunction with the Minimal Standards Working Group, the CRWG will explore the possibility of SRA/GenBank serving as the main public database for data deposition. As part of this, the groups will finalize details regarding what metadata should be included with data submission.

• The CRWG recommends that data sharing be facilitated by deposition of open AIRR data into an AIRR-compliant common repository, following a distributed model, in which compliant repositories are supported by a central registry.

o Compliance will be assessed based on each repository meeting a core set of criteria, the details of which will be developed by the CRWG over the coming year, and voted upon by electronic ballot or at a future AIRR meeting. Generally, these will require that repositories ensure data openness (as outlined above) while protecting personal health information, and adhere to the use of defined operational and data frameworks that will allow interoperability, performance, maintainability, and evolution.

Future Work
• The CRWG will work collaboratively with the other working groups to finalize recommendations to be approved by the AIRR community.

o Specifically, collaborative efforts will focus on the development of:
(1) customized metadata for AIRR sequencing data in the SRA/GenBank;
(2) standardized data elements with computable specifications, queries to be supported by the distributed repositories, and a standardized data-submission process and associated submission formats; and
(3) more detailed specifications for recommended technologies for implementation.

• The CRWG and AIRR Community will work with repositories to establish an accreditation system for compliance. This should include an appropriate system for the citation/attribution of the repository and/or the data.

- The CRWG and AIRR Community will explore the idea of seeking funding for the development and maintenance of a central registry of compliant repositories.

- The CRWG and AIRR Community will work with funders and journals to establish mechanisms for compliance with these recommendations.

- The CRWG will participate in preparing the AIRR Community white paper document.

- The CRWG and AIRR Community will work to develop consistent consent documents that are compliant with best practices for the broad sharing of AIRR data.

Minimal Standards Working Group:
(Co-leaders: Nina Luning Prak & Steven Kleinstein)

The goal of the Minimal Standards WG (**MSWG**) over the past year was to define a set of metadata and data elements that should be made available when publishing AIRR sequencing studies, or when submitting AIRR sequencing data to a data repository. The MSWG developed a set of six high-level principles along with a detailed set of data elements for implementing these principles.

Recommendations
- The MSWG proposes a framework of agreed-upon metadata and data for publication and submission to data repositories, to facilitate sharing of information from different experiments. The standard will span the six principles outlined by the MSWG comprising:
	(1) study, (2) sample, (3) processing, (4) raw data files (fastq output from the sequencer), (5) data processing (software) and (6) processed data files (including V(D)J assignment).
- The MSWG recommends that the proposed data elements are required for publication.
- The MSWG recommends that metadata and data covering principles 1-6 be deposited in a central repository, such as SRA/GenBank. Storing analysis results (such as clonal determination and diversity analysis) are beyond the scope of this standard, and would use more specialized databases.
- The MSWG recommends that funding mechanisms should be pursued for infrastructure support to enable investigators to more efficiently perform the time-consuming process of data entry.

All of the preceding recommendations were voted upon and approved by a majority in attendance on June 30, 2016, at the AIRR 2016 Community Meeting.

Future Work

• Publish the minimal standards in a high-impact journal, with input/authorship from the AIRR community; include in this publication the standards in a checklist format for easy adoption.

• In collaboration with the Common Repository group, develop a more detailed specification and controlled vocabulary for the proposed set of data elements to facilitate interoperability of repositories.

• Work with SRA to develop a prototype implementation of the AIRR sequencing data standard. This will include: (1) adapting the current SRA data submission template to capture the AIRR data elements, and (2) extending the SRA data validator to implement the controlled vocabularies and ontology linkages in the standard.

• Continue to refine the standards based on implementation experience, and new sequencing technologies.

• Discuss/update the standards during future AIRR Community Meetings.

• Participate in preparing the AIRR Community white paper document.

Tools and Resources WG:
(Co-leaders: Chaim Schramm & Erick Matsen)

The goal of the Tools and Resources Working Group (**TRWG**) is to promote standards of rigor and reproducibility for large-scale AIRR research. Much of our discussion has and continues to take place on the http://B-T.CR online forum. We encourage all members of the AIRR Community to join and participate in forum discussions. More generally, we believe that the AIRR Community should work towards reducing entry barriers, and identify new ways to explain and visualize the outstanding importance of AIRR research for understanding the immune system in health and disease.

Recommendations
Principles for community organization & development
The AIRR Community should:

● Promote transparent sharing of tools, methods, and information to enable review, contribution, and continued development.

● Provide a platform for sharing current experiences with methods, their limitations, and emerging technical issues.

● Discuss and communicate current challenges to ensure that tool/resource developers are focusing on the needs of Community members.

Sharing software tools
Software tools should:

● Use a license that is Open Source Initiative (OSI) approved, or a Creative Commons licence at least as liberal as Attribution-NonCommercial.

● Be hosted on publicly available repositories with versioning.

● Use a community-approved data format.

● Strive for modularity and interoperability with other tools.

● Include example data and checks for expected output.

- Provide information about run parameters as part of output.
- Clearly list hardware, operating system, and software dependencies, and/or provide a script to build a virtual machine.
- Clarify whether the software is still under active development, what level of support users can expect, and how to obtain support.

Sharing protocols and biological resources
- Protocols should be made publicly available in full detail, such that there is sufficient information to reproduce the entire experiment.
- Change history should be clearly shown as part of the protocol: Date of change, what was changed, and where it was changed.
- Protocols themselves should be kept in a versioned DOI-granting repository and updated as necessary.
- Biological materials (*i.e.*, plasmids, cell lines) should be made available to interested researchers *via* commercial repositories (*e.g.*, Addgene for vectors, ATCC for cell lines).

All of the preceding recommendations were voted upon and approved by a majority in attendance on June 30, 2016 at the AIRR 2016 Community meeting

Future work
The TRWG identified five major areas of focus for future work. In order to give proper attention to each of these needs, we created five new working groups that will work together to achieve the aims of the Tools and Resources WG.

Protocols and Best Practices Working Group (led by Encarnita Mariotti-Ferrandiz)
This working group will be responsible for developing recommended best practices and validation methods for new protocols.

Biological Standards Working Group (led by Melissa Smith)
This working group will be responsible for coordinating the development of reference samples that can be used as controls. The working group will reach out to established organizations such as NIST and Genome in a Bottle, as well as companies like ATCC and Novartis to help encourage ease of use and broad adoption.

Software Standards Working Group (led by Erick Matsen)
This working group will be responsible for developing a list of standard datasets with which software tools can be tested and compared. This will include both real and simulated data with a variety of characteristics matched to potential applications.

Germline Database Working Group (led by Corey Watson and Andrew Collins)
This working group will be responsible for developing appropriate metadata fields for the documenting of novel germline alleles and for establishing standards for versioned, inclusive databases. This will involve interfacing with established germline repositories, such as IMGT, IgPdb, and VBase to get the widest possible buy-in.

File Formats Working Group (led by Uri Laserson)
This working group will be responsible for developing standardized names for data fields which can be understood and interpreted by all software tools, allowing interoperability between pipelines from different developers. Close collaboration with the Minimal Standards Working Group is expected.

Meeting Organizing Committee:
(Tom Kepler, Jamie Scott, Danny Douek, & Felix Breden, plus the WG co-leaders listed above)

Approved
•   Become a formal society by joining an established society or starting an independent society.
•   Organize a Community Meeting in the Spring or Summer of 2017.
•   For virtual voting between Community Meetings, eligible voters will consist of participants in the 2015 Community Meeting, 2016 Community Meeting (face-to-face and webinar participants), and researchers and other stakeholders in the AIRR Community identified and agreed to by the eligible voters.

Future work
❖ Investigate joining an established society (*e.g.*, The Antibody Society, International Immunology Congress, *etc.*)
  ▪ Bring possibilities to the Community for a virtual discussion and vote.
❖ Determine a location for the next meeting and apply for grant funding to support it.
  ▪ Bring possibilities to the Community for a virtual discussion and vote on venue.
❖ Recruit members to participate in Working Groups and other functions of the Community, such as fundraising, communicating through social media, shaping the white paper document, etc.

Please contact Felix Breden (breden@sfu.ca) for further information.