Draft Recommendations from the Tools and Resources Working Group

Ramy Arnaout, Davide Bagnara, Christian E. Busse, Martin Corcoran, Rion Dooley, Danny Douek, Simon Frost, Victor Greiff, William Lees, Mats Ohlin, Sai Reddy, Adrian Shepherd, Corey Watson, Frederick A. Matsen IV, and Chaim A. Schramm

Overview. The adaptive immune-receptor repertoire (**AIRR**) Community will benefit greatly from cooperation among groups developing tools and resources for AIRR research. The goal of the Tools and Resources Working Group is to promote standards for AIRR tools and resources in order to enable rigorous and reproducible immune repertoire research at the largest scale possible. To achieve this goal, we have compiled a list of recommendations. Background information and discussions for all of these recommendations may be found in the online forums at http://B-T.CR; to facilitate viewpoint diversity, 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.

I. **Principles for community organization and development**
   A. The AIRR Community should promote transparent sharing of tools, methods, and information to enable review, contribution, and continued development.
   B. The AIRR Community should provide a platform for sharing current experiences with methods and their limitations.
   C. The AIRR Community should discuss and communicate current challenges to ensure that tool/resource developers are focusing on the needs of users.
II. **Guidelines for developing and sharing software tools**
   A. Software tools should use OSI-approved open-source licenses and be hosted on publicly available repositories with a clear versioning system.
   B. Software tools should use community-curated standard file formats and strive for modularity and interoperability with other tools (Please see discussion at http://b-t.cr/t/-/25).
   C. Software tools should include example data and checks for expected output, and provide information about run parameters as part of output.
   D. Software tools should clearly list dependencies and/or provide scripts to build a virtual machine.
   E. The AIRR Community should agree on basic data sets that can be used for software testing (Please see discussions at http://b-t.cr/-/40 and http://b-t.cr/t/-/129).
   F. Publications describing software tools should make it clear whether the software is still under active development, what level of support users can expect, and how to obtain support.
III. **Guidelines for sharing protocols**
   A. The AIRR Community should choose a central, searchable, location to list available protocols and reagents.
   B. Protocols should be made publicly available in full detail, and the AIRR Community should consider developing a template to assist in this process.

    C. Protocols should be kept in a versioned repository and updated as necessary. https://figshare.com/ is an established DOI-granting service that offers versioning.

IV.   **Guidelines for sharing biological resources**
    A. Novel reagents (plasmids, cell lines, antibodies, *etc*.) should be made available to all qualified researchers at cost. We urge researchers to do this through repositories (*e.g*., Addgene for vectors) to ensure long-term availability.
    B. The AIRR Community should work to develop standardized benchmark biological resources to test library preparation methods.
    C. In addition to sharing primer sequences, additional information such as concentration, cycling conditions, and type of polymerase should be made available as part of a shared protocol.

V.   **Key resource: An open and dynamic germline database**
    A. Sequences should be made available with a license that enables further sharing and extension. (Please see discussion at http://b-t.cr/t/-/102)
    B. The level of evidence for each germline gene should be made clear, by a metric determined by the AIRR Community.
    C. There should be a clear and objective system by which new germline genes are added to the database. This should include a means of submitting germline genes inferred from high-throughput sequencing data.
    D. The community should decide if it would be best to extend the IMGT naming convention, or start afresh. (Please see discussion at http://b-t.cr/t/-/43)
    E. The community should work towards the definition and creation of metadata to enable germline databases to be conveniently used with junction analysis tools.

VI.   **Key resource: A place to share many of the resources discussed above**
    A. The AIRR Community should establish a publicly editable website or wiki (*e.g*., *via* github.com or Wikibooks) that will allow us to maintain lists of available datasets, protocols, and other resources.

Requested feedback from AIRR attendees. There remain many open challenges to developing and sharing AIRR tools and resources that the Working Group would like to address. We are specifically interested in feedback on the following questions:

1. What would be the critical features of a germline-gene database, and how can they be implemented? How can we incorporate various types of information, levels of confidence, and need to interconnect with more established resources (Please see discussions at the above links.)
2. What types of standardized resources are desirable for testing new protocols and benchmarking analysis tools? How can these be collaboratively developed by the AIRR Community?
3. What are the outstanding computational challenges in AIRR analysis? As a community, how can we best address these needs? (Please see discussion at http://b-t.cr/t/-/127)
4. What are critical features for shared file formats? (Please see discussion at http://b-t.cr/t/-/25.) What about annotating clonal families (http://b-t.cr/t/-/109) or known TCR specificities?

5. How can we encourage sharing of resources and the adoption of best practices for tool development?
6. How can explain and visualize the outstanding importance of AIRR research for understanding the immune system in health and disease? (Please see discussion at http://b-t.cr/t/-/103/)