Cell Systems

CellPress

Voices

What are the current driving questions in immune repertoire research?



Victor Greiff
University of Oslo and Oslo University Hospital

Mapping adaptive immune specificity

Adaptive immune receptors (AIRs: B cell receptors [BCRs] and/or T cell receptors [TCRs]) can recognize nearly any antigen by incredibly diverse AIR repertoires (AIRRs). AIRs are crucial to fighting infection but may inflict autoimmunity. Although current public AIR(R) sequence and structural databases are expanding rapidly in size, the majority of the stored data are antigen *un*labeled. This leaves our knowledge of antigen specificity, one of the hallmark features of AIR mechanisms of action, scarce and hinders the precision manipulation and prediction of adaptive immunity.

Fundamental questions require investigation: how do antigen-specific AIRRs differ across individuals and immune states (health, infection, disease)? And how common are antigen-specific precursor AIR lineages for different antigens and in different populations, and are these present at sufficient frequency to initiate an immune response?

Antigen-specific high-throughput AIRR-sequencing and structural biology technologies as well as growing interdisciplinary expertise in systems immunology, statistics, and machine learning are now beginning to offer the opportunity to resolve these questions.

Specifically, large-scale antigen-specific AIRR data may soon enable the development of computational and machine learning methods that predict antigen specificity from the AIR sequence or structure, unlocking the so far inaccessible antigen-specific information in publicly available AIRR data.

AIRR-focused competitions, similar to those that facilitated the development of Alpha-Fold, may be needed to sharpen our currently insufficient tools for predicting antigen-specific adaptive immunity.



Eline T. Luning Prak
University of Pennsylvania

All for one and one for all

The immune repertoire can be thought of as a collection of B and T cell clones that embodies an individual's lifetime history of immune exposures and responses. The repertoire can be accessed through next-generation sequencing of antibody and TCR gene rearrangements, but translating these data to an informed view about one's state of health or level of immune protection is still in its infancy.

We need more detailed datasets on antigen-enriched and multireactive B and T cell populations. We need more data on immune responses in tissues and how those responses compare to the blood. We need data on individuals who are young, old, healthy, and sick. We need data on diverse individuals who have different antibody and TCR genes and who live in different parts of the world where different pathogens and antigens are encountered. We need to know how many clones to sample and how deeply to sample them.

The COVID-19 pandemic has shown us that sharing and standardization of immune data, including immune repertoire data, can facilitate large-scale studies in which public clones and shared motifs associated with immune responses to the virus and to vaccines are identified. Ongoing efforts to standardize and annotate datasets by groups such as the AIRR Community promote data sharing and facilitate discovery on a global scale. With each new repertoire that is contributed to public data, our ability to decipher the clonal content and landscape improves: all for one and one for all.



Tim O'Donnell
Icahn School of Medicine at Mount Sinai

Optimism for antigen prediction

Sequencing the immune receptors expressed in a population of T cells or B cells is now routine.

However, identifying the antigens they recognize still requires expensive or low-throughput experiments. Prediction of antigen specificity given a TCR or antibody sequence would be a breakthrough with wide utility across immunology and clinical diagnostics.

The past two years have seen a revolution in structural biology, with AlphaFold2 from DeepMind able to predict protein structure often to experimental accuracy. Methods like AlphaFold2, however, rely on evolutionary information from homologous proteins to guide the prediction, a technique that performs poorly for the somatically generated variable regions that drive antigen recognition by TCRs and antibodies.

Models inspired by AlphaFold2 have recently been introduced that achieve similar accuracy while dispensing with the need for homologous sequences. Early results suggest these may be better able to predict the structure of antibody variable regions. However, serious challenges remain, including scoring receptor/epitope interactions and sampling multiple conformations of the often-flexible variable loops. The accuracy required for these steps is high because the ideal use case is searching many receptor sequences against many potential antigens. While much work remains, I am optimistic that progress on antigen prediction will be possible by making use of the expanding toolbox of ideas from general protein structure prediction.



Francesca Finotello
University of Innsbruck

Deciphering T cell responses in space and time

Thanks to their TCRs, T cells are able to recognize a wealth of different antigens, constituting a major line of defense against infection and cancer. The outbreak of the COVID-19 pandemic has prompted the development of new methodologies and resources that can advance our understanding of TCR-antigen specificity and motivated a worldwide community effort for the curation of thousands of TCR sequences recognizing SARS-CoV-2 epitopes.

Given the vast diversity of all possible TCR-epitope combinations, this rapidly growing body of experimentally determined pairs represents a rich dataset for the training of algorithms for TCR specificity prediction. In parallel, deep learning approaches inspired by recent breakthroughs in protein structure prediction could provide new means for predicting TCR specificity for unseen epitopes playing a role in infection, autoimmunity, and cancer.

In cancer, the characterization of T cell responses is particularly challenging. This is not only because of the sheer diversity of tumor antigens, but also because of the multifactorial mechanisms governing T cell expansion and polarization in space and time. In this light, the integrative analysis of transcriptomic and TCR data from single-cell and spatial transcriptomics can help us answer a fundamental question in cancer immunology: what are the intrinsic and extrinsic factors that determine the expansion of tumor-reactive T cells, ultimately impacting the efficacy of cancer immunotherapy?



Sai T. Reddy
ETH Zurich

Convergence in immune repertoires

Immune repertoire research is focused on the targeted and deep sequencing of lymphocyte BCRs (and their secreted version antibodies) and TCRs, which recognize molecular structures (antigens or epitopes) on foreign pathogens and diseased cells and thus serve as the basis of adaptive immune recognition.

A central question of adaptive immunity is the extent to which antigen specificity results in convergent selection of BCR or TCR sequences in different individuals. Recent progress in the field has revealed that despite the exceptionally diverse repertoire of human antibody sequences, there is evidence of certain convergent (or shared) sequences present in different individuals; convergence can either be independent of antigen or associated with a common antigen specificity. In the very recent and intensely studied area of neutralizing antibodies targeting the SARS-CoV-2 receptor binding domain, there have been a number of convergent antibody sequence patterns identified in vaccinated or convalescent individuals. Similarly, TCR sequence patterns have been detected that are associated with pathogen or tumor-associated antigens (peptide-MHC). This massive increase of immune repertoire sequencing data is also providing fertile ground to apply methods in machine and deep learning. For example, deep unsupervised clustering has been used to identify complex patterns associated with antigen-associated convergence. As machine and deep learning approaches continue to advance, they may facilitate detection of convergence at a much larger scale and eventually support the development of immune diagnostics, the discovery of therapeutic antibodies, and the design of vaccine immunogens.



Aleksandra Walczak and Thierry Mora
LPENS and CNRS

Novel data analysis approaches

Advances in different high-throughput techniques and new detailed experiments have given us access to new information on receptor repertoires that makes up for the diversity of adaptive immune responses. However, while we have more and more sequences from an increasing number of conditions and species, we still do not really know how to read this information. We do not fully understand how the cells of the immune repertoire work together to protect us from many evolving pathogenic challenges. The challenges are at many levels, and once again we are learning that cataloging is important but does not give answers about how a complex biological system works. In short, we need to derive new analysis techniques that combine information from different experiments.

One of the challenges is linked to the specificity of lymphocyte receptors. While we know that lymphocytes are cross-reactive, meaning they can recognize more than one epitope, we do not know which epitopes are recognized by which B or T cell. We also do not know what the range of cross-reactivity is. Recent work has shown that T cells with similar sequences can recognize the same antigen, but we do not know how many and what kind of mutations set the cross-reactivity scale. We know that lymphocyte receptor sequences that are apparently dissimilar in sequence can also recognize similar antigens. Are there some hidden sequence features that can help us identify which lymphocyte receptors can recognize the same or similar antigens?

Another challenge is linked to self-recognition and lymphocyte maturation. How is the repertoire formed to avoid harboring too many TCRs recognizing the organism's own proteins, which would otherwise lead to autoimmune diseases? How is avoiding self-proteins achieved in BCRs, if at all? It now seems the process of T cell thymic selection is leakier than originally believed.

Yet another challenge concerns somatic repertoire evolution. We are starting to identify responding lymphocytes from whole repertoire data. How does the repertoire evolve in the long term, and how is homeostasis maintained? Can we understand what differentiates a successful response from a failed one and use it to help guide our immune system?

From another perspective, there is the global evolutionary question of how repertoires of different species compare and how they evolved. A combination of theoretical, experimental, and detailed statistical analysis combined with publicly available datasets means these questions are starting to be within our reach.

DECLARATION OF INTERESTS

V.G. declares advisory board positions in aiNET GmbH, Enpicom B.V., Specifica Inc., Adaptyv Biosystems, EVQLV, and Omniscope. V.G. is a consultant for Roche/Genentech and immunai. E.T.L.P. has no financial interests to declare; however, she does serve on the board of directors of the Antibody Society (TAbS) and on the executive sub-committee of the Adaptive Immune Receptor Repertoire (AIRR) Community. S.T.R. is a co-founder of deepCDR Biologics (acquired by Alloy Therapeutics) and co-founder of Engimmune Therapeutics. S.T.R. is on the Scientific Advisory Board of Alloy Therapeutics and Engimmune Therapeutics. S.T.R. holds shares in Alloy Therapeutics and Engimmune Therapeutics.