Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Systems Biology

The adaptive immune receptor repertoire community as a model for FAIR stewardship of big immunology data

Jamie K. Scott[1,2] and Felix Breden[3]

Abstract

Systems biology involves network-oriented, computational approaches to modeling biological systems through analysis of big biological data. To contribute maximally to scientific progress, big biological data should be FAIR: findable, accessible, interoperable, and reusable. Here, we describe high-throughput sequencing data that characterize the vast diversity of B- and T-cell clones comprising the adaptive immune receptor repertoire (AIRR-seq data) and its contribution to our understanding of COVID-19 (coronavirus disease 19). We describe the accomplishments of the AIRR community, a grass-roots network of interdisciplinary laboratory scientists, bioinformaticians, and policy wonks, in creating and publishing standards, software and repositories for AIRR-seq data based on the FAIR principles.

Addresses

[1] Department of Molecular Biology and Biochemistry, Simon Fraser University, Burnaby, BC, V5A 1S6, Canada

[2] Faculty of Health Sciences, Simon Fraser University, Burnaby, BC, V5A 1S6, Canada

[3] Department of Biological Sciences, Simon Fraser University, Burnaby, BC, V5A 1S6, Canada

Corresponding author: Breden, Felix (breden@sfu.ca)

Current Opinion in Systems Biology 2020, **24**:71–77

This review comes from a themed issue on **Systems immunology & host-pathogen interaction (2020)**

Edited by **Ramit Mehr**

For complete overview of the section, please refer the article collection - Systems immunology & host-pathogen interaction (2020)

Available online 10 October 2020

https://doi.org/10.1016/j.coisb.2020.10.001

Keywords

Adaptive immunity, B-cell and T-cell receptor repertoires, FAIR Principles, Open source, Adaptive immune receptor repertoire (AIRR) community.

Introduction

Systems immunology encompasses computational approaches for modeling the immune system and its dynamic responses through the use of high-dimensional and 'big' biological data (*i.e.*, digital data having features of (i) large volume, (ii) collection and processing at high velocity, (iii) variable sources of origin, and thus, (iv) requiring a means of verifying data quality) [1]. High-throughput sequencing of the diverse receptors of adaptive immunity: [B-cell receptors (BcRs) of B lymphocytes and the antibodies (Abs) secreted by their plasma cell progeny, and the T-cell receptors (TcRs) of T lymphocytes], has allowed characterization of adaptive immune receptor repertoires (AIRRs) at the level of single clones (see Box 1 for Glossary). AIRR sequence (AIRR-seq) data portray the frequencies of B- and T-cell clones in a lymphocyte population, and, when coupled with phenotypic data, can reflect the dynamics of clonal differentiation during an immune response. AIRR-seq data can marry well with other types of big biological data that characterize immune responses (*i.e.*, flow cytometric/CYTOF data, RNA-sequencing (RNA-seq) data, metabolomic, proteomic, and microbiome data and digitized microscopic data); together these data types are contributing to systems-level analyses of immune responses. However, for such analyses to be practical, and for data from different sources to be comparable, big data and their metadata should be standardized; the software used to analyze them should complement those standards; and repositories used to store such data should be federated into a virtual commons with gateway functions to allow simultaneous querying of the entire commons for defined data and/or associated metadata. The universal adoption of such standards would go a long way toward enabling large-scale, multidimensional analyses.

The AIRR community (AIRR-C) was established to bring these features to AIRR-seq data [2]. The AIRR-C comprises a network of volunteer bioinformaticians, laboratory scientists, and policy experts from academia and industry who support open science [3] insofar as possible, and are working together to establish standards for reporting [4], and storing and sharing [5] AIRR-seq data following FAIR practices [6]. Its vision is to promote a community of AIRR-seq data generators and users who share its core values of collaboration, inclusivity, transparency, and data and materials sharing. To achieve these goals, the AIRR-C communicates the resources it develops through publications, meetings, and workshops, including standard formats for AIRR-seq data and metadata storage and analysis, computational

Box 1. Glossary of terms relevant to the AIRR community.

Big biological data: Are digital data having features of (i) large volume, (ii) collection at high velocity *via* high-throughput technologies, (iii) data originating from variable sources, thus, (iv) requiring a means of verifying data quality.
FAIR data principles: Making big biological data findable, accessible, interoperable, and reusable.
Systems immunology: Uses systems biology approaches to model dynamic networks characterizing the immune system and its responses from multidimensional and big biological data.
Antigen: A molecule or complex that generates a specific immune response by binding to an antibody, B-cell receptor, or T-cell receptor.
Adaptive immune responses: Immune responses that are initiated by antigen-mediated selection of B- and/or T-cell clones and that mediate antigen-specific effector functions and memory.
Immunological memory: A state of the immune system that produces a rapid response on re-exposure to antigen (*e.g.*, a virus), after it has been cleared from the organism.
T cell: A type of lymphocyte that mediates adaptive immunity by killing cells harboring foreign antigens, or by orchestrating immune responses *via* cell−cell interactions and/or secreted cytokines.
B cell: A type of lymphocyte that mediates adaptive immunity, mainly through the secretion of antibodies.
TcR: T-cell receptor; the antigen-binding receptor of a T cell
BcR B-cell receptor; the antigen-binding receptor of a B cell
Ab: Antibody (AKA immunoglobulin); the secreted form of a BcR
Variable region: The N-terminal region of the two chains of an Ab, BcR, or TcR that fold together to form their antigen-binding domain (AKA the variable domain). They are encoded by rearranged V, (D), and J gene segments.
Germline gene segments: Unrearranged V, D, and J gene segments encoded in the 'germline' genome. In developing B and T cells, they recombine to form VDJ open reading frames that encode the variable domains of Abs, BcRs, and TcRs.
CDR3: The third complementarity-determining region of the variable regions of BcR, Ab, and TcR. The CDR3s comprise the most diverse regions of a variable domain and the main contact site with antigen.
AIRR-C: The adaptive immune receptor repertoire (AIRR) community.
WG: Working Group of the AIRR-C; currently, there are 8 WGs.
COVID-19 coronavirus disease 19.
SARS-CoV-2 severe acute respiratory syndrome coronavirus 2.

tools, data repositories, and websites. As one example, the AIRR Data Commons (ADCs) houses more than 3 billion AIRR-seqs and their associated metadata in a standardized format. It provides these data to the public in a standardized format that is easily queryable and makes available open-source software for one to establish their own repository within the ADC [7]. Recent studies comprising ~700 million COVID-19 (coronavirus disease 19)−related AIRR-seqs are easily accessible in a standard format through the ADC as implemented by the iReceptor Gateway [8]. We describe in the following context the biological and genetic underpinnings of AIRR-seq data and provide examples of their contributions to our understanding of the immune responses elicited by COVID-19.

The immunogenetics of AIRR diversity

The adaptive immune system comprises the T cells and B cells, which use highly diverse receptors to recognize and respond to foreign and self antigens. T cells recognize fragments of protein antigens (peptides) in complex with major histocompatibility (MHC) molecules on the surface of cells. A given T cell can directly kill a cell that displays foreign antigen-MHC molecules and/or initiate cytokine production that will regulate the immune response. The T and B cells in vertebrates are highly diverse and complex due to the diversity of their heterodimeric TcRs (comprising α/β or γ/δ chains) and BcRs (comprising heavy and light chains); each chain is encoded at a different genetic locus. B cells that are activated by antigen differentiate to become plasma cells that secrete their BcRs as soluble Abs.

During its development, each B and T cell independently recombines gene segments in the loci for each of its chains to encode variable domains that fold together to produce the unique, heterodimeric antigen-binding domain of a TcR or BcR. TcR and BcR/Ab diversity is generated from combining one variable (V) gene segment, one diversity (D) gene segment (only for the V domains of the heavy chains of BcRs/Abs, and the β or δ chains of TcRs), and one joining (J) gene segment. Thus, each genetic locus that encodes a TcR or BcR chain contains clusters of multiple, different 'germline' V, (D), and J gene segments from which a single V, (D), and J gene segment is used to produce one of the two chains of a heterodimeric BcR/Ab or TcR. Besides the combinatorial diversity provided from recombination of the V, (D), and J gene clusters, and from combining the two chains of the heterodimer, diversity is also produced at the junctions between VD and DJ and VJ joins, in a complex process called 'imprecise joining' that can even add 'nontemplated' nucleotides at random to these junctions. The two junctions spanning a V−D−J recombination and the single junction of a V−J recombination encode the most diverse region of the variable domain in each chain of a BcR/Ab or TcR, its third complementarity-determining region (CDR3). The CDR3 of each chain of a BcR/Ab or TcR is also responsible for making the dominant contacts with antigen; thus, these regions, which encode the greatest sequence diversity, are also positioned in the receptor to make the greatest contribution to its specificity and affinity for antigen.

Taken together, the *potential diversity* of the BcR/Ab and TcR repertoires is vast, far larger than the number of T and B cells in one's body. This enormous receptor diversity allows the adaptive immune system to respond

with incredible specificity to a limitless variety of antigens. With the right costimulation, foreign antigens will 'select' 'naïve' B- and T-cell clones from their respective repertoires by virtue of binding directly to their BcRs, or as peptide fragments in complex with MHC to their TcRs, respectively. These selected B- and T-cell clones then divide and differentiate into effector cells (*e.g.*, Ab-secreting plasma cells or cytotoxic, helper or regulatory T cells) to carry out the effector functions of adaptive immunity. Some descendants of these activated clones will also become long-lived memory B and T cells that await future encounters with the same antigen; they form the basis of immunological memory. Thus, adaptive immune responses are (*i*) initiated *via* selection by antigen of B- and T-cell clones in a repertoire whose BcRs or TcRs bind antigen, and (*ii*) mediated by these cells' expansion and differentiation into effector and memory cells. Note that a final level of diversification occurs during B-cell responses: with T-cell help, the recombined V(D)J genes of BcRs (but not TcRs) accumulate somatic mutations; antigen then selects 'stronger-binding BcRs' out of the pool of mutants of a selected B-cell clone in a process called 'affinity maturation', which mediates the development of high-affinity Abs.

With the advent of high-throughput cDNA sequencing came its modification to characterize adaptive immune receptor (*i.e*, BcR/Ab and/or TcR) repertoires (*i.e.*, AIRRs). AIRR-seq data can be generated from high-throughput sequencing of bulk or single B cells, Ab-secreting plasma cells, or T cells by (i) amplification and bulk sequencing of the V(D)J rearrangements from a *single chain* of a BcR/Ab or TcR from cDNA or genomic DNA prepared from whole-cell populations that are often isolated based on cell phenotype (*e.g.*, cell-surface markers and size) [9,10], or (ii) amplifying and sequencing the 'paired' V(D)J rearrangements from cDNAs encoding *both chains of a BcR/Ab or TcR* from single cells [11–13]. Thus, clonal expansions can be identified from AIRR-seqs appearing at high frequency, a hallmark of 'immunodominant' clones in an antigen-specific response, and when coupled with cell phenotype data (from flow cytometry and/or RNA-sequencing) can provide insights into the effector and memory functions of those clones. The lineages of somatically mutated B cells arising from a single clone can also be deduced from AIRR-seq data. And finally, the sequences of an individual's 'germline' V, D, and J gene segments can be inferred from the AIRR-seq repertoires of unmutated BcRs and TcRs, providing a bird's-eye view of one's germline gene haplotype [14–16].

AIRR-seq data inform COVID-19 immune responses

Recently the AIRR-C called for increased data sharing to help overcome the COVID-19 pandemic (https://www.antibodysociety.org/airr-community/covid-19-demands-increased-public-sharing-of-biomedical-research-data/). Many COVID-19 researchers have responded, with some providing data from studies during the preprint stage. As of November 1 2020, the ADCs included >700 M receptor sequences from 17 studies of patients affected with COVID-19 all curated in accordance with the AIRR-C Standards (Table 1).

Taken together, the initial analyses from these studies indicate convergence of AIRR-sequences among patients affected with COVID-19, evidence of restricted V-gene usage, and characteristic patterns of TcR and BcR oligoclonal expansion vs. polyclonal activation among AIRR-seq repertoires from patients experiencing different degrees of disease severity (Table 1). The ability to validate such patterns by comparing results across studies and institutions should be greatly facilitated by integrated searches across the ADCs through the iReceptor Gateway. For example, Meysman et al. [17] compared longitudinal data from multiple studies to reinforce the observation that SARS-CoV-2 (severe acute respiratory syndrome coronavirus 2) reactive T-cell diversity increases over the course of disease progression. For more information on obtaining or sharing COVID-19 data contact support@ireceptor.org.

Much of the work of the AIRR-C is performed by its Working Groups (**WGs**), which consist of AIRR-C members and other interested persons who are organized around a common goal and meet virtually (Box 2). A central function of several WGs is to develop and publish standards for the annotation, storage, and sharing of AIRR-seq data. The AIRR-C's first set of standards, the miAIRR standards, were published by the Minimal Standards WG (https://www.antibodysociety.org/the-airr-community/airr-working-groups/minimal_standards/); they consist of a minimal set of metadata for describing an AIRR-seq data set when these are published or deposited in an AIRR-compliant repository [4]. These recommendations focus on metadata that would support standard quality control and reproducibility of AIRR-seq data. The Data Representation WG complements the Minimal Standards WG by focusing on the receptor sequences, or rearrangement-level of AIRR-seq data. It is primarily responsible for developing standardized file formats, schemas and data field names to represent annotated AIRR-seqs, and downstream data representations [5] such as the AIRR rearrangement-compatible TSV file format (AIRR.tsv) that allows interoperability among all AIRR-seq data analysis tools and miAIRR-compliant databases.

The Common Repository WG works closely with these two data standards groups to provide an ADC for the scientific community. The ADC is envisioned as a federated set of MiARR-compliant AIRR-seq data stored at several institutions world-wide. At present, two major repositories, VDJServer (Univ. of Texas Southwestern Medical Center) and the iReceptor Public Archive (IPA;

Table 1

COVID-19 AIRR-seq studies available in AIRR-C Data Commons.

Study	Publication status	Ab/BCR; TCR; or Both	Cohort	Source material	No. of sequences (M)	Notes/conclusions
Galson et al. [22]	Preprint	Ab/BCR	19 COVID-19 patients	Peripheral blood samples	3.7	Convergent clonotypes were identified among patients but not healthy controls.
Kim et al. [23]	Preprint	Ab/BCR	Severe COVID-19 patients	Peripheral blood samples	3.4	VH clonotypes encoded by specific V and J germline gene segments (either IGHV3-53 or IGHV3-66 and IGHJ6) were identified in different Ab isotypes (IgG1, IgA1 and IgA2) with minimal mutations, and paired with diverse light chains; resulting Abs bound to the RBD.
Kuri-Cervantes et al. [24]	Published	Ab/BCR	Moderate, severe, and recovered COVID-19 patients	Peripheral Blood samples	11.9	Selective clustering of severe COVID-19 cases through unbiased analysis of the aggregated immunological phenotypes (B-cell repertoire clonality; SARS-CoV-2-specific Abs; etc.).
Montague et al. [25]	Preprint	Ab/BCR	19 COVID-19 patients of varying disease severities	Peripheral blood samples	18.9	Expanded rare clonal lineages shared among patients; V-gene specific responses are highly individualized; longer CDRH3 lengths in Abs from patients with moderate and severe symptoms.
Nielsen et al. [26]	Published	Ab/BCR	Admitted patients with symptoms of COVID-19 and confirmed SARS-CoV-2 infection by RT-qPCR	Peripheral blood samples	8.9	B cells utilized a limited subset of V genes, and extensive activation of IgG and IgA B cells without significant somatic mutation; expansion of B-cell clones with highly similar sequences shared among patients.
Liao et al. [27]	Published	TCR	12 COVID-19 patients of varying disease severities	Bronchoalveolar lavage fluid	66.9	Greater expansion of T cell clones in moderate disease than severe disease.
Minervina et al. [28]	Preprint	TCR	2 COVID-19 mild cases	Peripheral blood samples	87.1	SARS-CoV-2-responding CD4+ and CD8+ T cell clones were detected in the memory fraction of both preinfection and postinfection samples; TCR sequences from these clones exhibited characteristic motifs
Nolan et al. [29]	Preprint	TCR	Multiple cohorts from 7 studies	Peripheral Blood samples	308	Sequences from 7 projects by Adaptive Biotechnologies and Microsoft available; no preliminary conclusions.
Shomuradova et al. [30]	Preprint	TCR	31 COVID-19 convalescent patients	Peripheral Blood samples	3.9	Possible T-cell cross reactivity in some patients; sequences of some T-cell clones responding to S protein were shared across multiple patients.
Schultheiβ et al. [31]	Published	Both	Cohort 1: recovered individuals, mostly after mild disease Cohort 2: severe disease, requiring hospitalization	Peripheral blood samples	13.3	Convergent B-cell responses to SARS-CoV-2 are mainly from naïve cells. T-cell clusters emerge over the disease course that are shared among recovering patients.

IMGT, the international ImMunoGeneTics information system.

Box 2. AIRR Community Working Groups and Published Standards.

Name	Working Group (WG) or Subcommittee (SC)	Yea Established	Description and purpose	Publication or url for standards
Minimal standards	WG	2015	Develops miAIRR standards for the submission of AIRR-seq data sets. Supports reproducibility, standard quality control, and data deposition in a common repository.	Rubelt et al. [4]
Data representation	WG	2016	Develops standardized file formats, schemas and data field names to represent MiAIRR metadata, annotated BcR/Ab and TcR sequences, and downstream data representations.	Vander Heiden et al. [5]
Common repository	WG	2015	Promotes and facilitates deposition, access, and sharing/reuse of AIRR-seq data sets through the creation of common repositories	Christley et al. [7]
Germline Database	WG	2016	Promotes the development of complete and accurate sets of reference germ line IG and TCR genes for multiple species, and promotes the accurate analysis and reporting of the germ line genes that can be identified in repertoire studies (works closely with the IARC SC*)	Lees et al. [20]
Biological resources	WG	2016	Coordinates the development of reference samples that can be used as controls.	NA
Software	WG	2016	Encourages practices that enable AIRR-compliant software tools to work, and to work with one another.	https://docs.airr-community.org/en/stable/swtools/airr_swtools_standard.html
Legal and ethics	WG	2019	Provides a forum for focused discussion and development of legal and/or ethical standards for the AIRR-C.	NA
Diagnostics	WG	2019	Explores AIRR-seq data as diagnostic tools for clinical applications.	NA
Executive	SC	2015	Provides leadership, oversees manuscript endorsement, receives reports from WGs and other SCs, manages votes and elections, draft proposals regarding governance and other services to the AIRR-C.	Breden et al. [2]
Meetings	SC	2016	Plans and runs AIRR-C meetings and events.	NA
Communications	SC	2017	Communicates activities of the AIRR-C to the AIRR-C and the scientific community.	NA
*Inferred Alleles Review Committee (IARC)	SC	2017	Judges the validity of germline immunoglobulin and TCR genes that are inferred from AIRR-seq data. IARC advises IMGT and the IUIS nomenclature IG/TCR/MHC subcommittee of their findings. It also reviews inferred sequences and evidence in support of their existence and makes these data available to through the OGRDB database.	Ohlin et al. [19]

Simon Fraser University, Canada) curate over 4-billion V(D)J rearrangements from 5938 repertoires from 49 studies, including the majority of AIRR-seq studies from patients affected with COVID-19. The ADC can be queried *en masse* through the iReceptor Gateway (www.ireceptor.org) and the results downloaded, allowing the user to rapidly and seamlessly search through billions of AIRR sequences and to download only what is wanted. The code for all of these resources is open source, and data availability follow the FAIR principles.

One very active area of development for the AIRR-C is identification and cataloging of the germline V, D, and J gene segments that recombine to produce AIRRs. The loci encoding these gene segments are highly complex and repetitive, making them difficult to sequence. Thus, little is known about their full sequences nor if polymorphisms within them map to disease susceptibility [18]. The AIRR-seq data set from an individual can be used to infer one's germline gene haplotype; knowledge of the immunoglobulin germline gene haplotype for an individual allows stronger conclusions to be drawn when determining somatic mutations from one's BcR/Ab sequence data. The AIRR-C's Inferred Allele Review Committee developed a set of requirements for recognizing and validating new V, D, and J alleles from AIRR-seq data [19]; OGRDB, established by the Germline Database WG, curates these inferred alleles [20], in collaboration with the standard database for these genes, the Immunogenetics information system (http://www.imgt.org/) [21].

Conclusions

In summary, AIRR-seq data have made significant contributions to our understanding of the immune system, and they promise to lend an important dimension to high-dimensional systems approaches in immunology research. In future, progress in systems immunology would be enhanced greatly by the collective efforts of researchers, institutions, national research funding agencies, journals, scientific societies, and industry to implement FAIR data practices, and open-source practices where possible, for all types of big biological data. The COVID-19 pandemic has created an urgent need for critical diagnostics, vaccines, and therapeutics, and the scientific community has stepped up in sharing data freely to meet that need in a timely fashion. However, given the current lack of standards for sharing many types of big biological data, impediments exist to their general usability. The AIRR-C serves as a model for implementing such a goal, through its grass-roots approach to developing standards for a specific type of immunological big data based upon FAIR data and open-source practices.

Conflict of interest statement

Nothing declared.

Acknowledgements

This work was supported by The Canada Research Chairs Program (J.K.S.) and the Canadian Institutes for Health Research, European Union's Horizon 2020 Research and Innovation Programme under grant agreement No. 825821 (iReceptor Plus), and CANARIE (F.B.).

References

Papers of particular interest, published within the period of review, have been highlighted as:

- • of special interest

1. Li Y, Chen L: **Big biological data: challenges and opportunities**. *Dev Reprod Biol* 2014, **12**:187–189.

2. • Breden F, Luning Prak ET, Peters B, Rubelt F, Schramm CA, Busse CE, Vander Heiden JA, Christley S, Bukhari SAC, Thorogood A, *et al.*: **Reproducibility and reuse of adaptive immune receptor repertoire data**. *Front Immunol* 2017, **8**:1418.
Introduces the goals of the AIRR Community Initiative and urges researchers to join the Community's efforts

3. Besançon L, Peiffer-Smadja N, Segalas C, Jiang H, Masuzzo P, Smout C, Deforet M, Leyrat C: **Open Science saves lives: lessons from the COVID-19 pandemic**. *BioRxiv* 2020, https://doi.org/10.1101/2020.08.13.249847. preprint doi.

4. • Rubelt F, Busse CE, Bukhari SAC, Bürckert JP, Mariotti-Ferrandiz E, Cowell LG, Watson CT, Marthandan N, Faison WJ, Hershberg U, *et al.*: **Adaptive Immune Receptor Repertoire Community recommendations for sharing immune-repertoire sequencing data**. *Nat Immunol* 2017, **18**:1274–1278.
Reports the AIRR Community's recommendations for a minimal set of metadata that should be used to describe an AIRR-seq data set when published or when deposited in an AIRR-compliant repository

5. • Vander Heiden JA, Marquez S, Marthandan N, Bukhari SAC, Busse CE, Corrie B, Hershberg U, Kleinstein SH, Matsen Iv FA, Ralph DK, *et al.*: **AIRR Community standardized representations for annotated immune repertoires**. *Front Immunol* 2018, **9**:2206.
Describes the standardized data representations for storing and sharing annotated antibody/B-cell and T-cell receptor repertoire (AIRR-seq) rearrangement data

6. Wilkinson MD, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A, Blomberg N, Boiten J-W, da Silva Santos LB, Bourne PE, *et al.*: **The FAIR Guiding Principles for scientific data management and stewardship**. *Sci. Data* 2016, **3**, 160018.

7. • Christley S, Aguiar A, Blanck G, Breden F, Bukhari SAC, Busse CE, Jaglale J, Harikrishnan SL, Laserson U, Peters B, *et al.*: **The ADC API: a web API for the programmatic query of the AIRR Data Commons**. *Front. Big Data* 2020, **3**:22.
Outlines the API (Application Programming Interface) for curating and sharing data in the AIRR Data Commons (ADC)

8. • Corrie BD, Marthandan N, Zimonja B, Jaglale J, Zhou Y, Barr E, Knoetze N, Breden FMW, Christley S, Scott JK, *et al.*: **iReceptor: a platform for querying and analyzing antibody/B-cell and T-cell receptor repertoire data across federated repositories**. *Immunol Rev* 2018, **284**:24–41.
Describes the iReceptor Science Gateway (ireceptor.org) which queries the AIRR Data Commons based on sequences or metadata

9. Rubelt F, Bolen CR, McGuire HM, Vander Heiden JA, Gadala-Maria D, Levin M, Euskirchen GM, Mamedov MR, Swan GE, Dekker CL, *et al.*: **Individual heritable differences result in unique cell lymphocyte receptor repertoires of naïve and antigen-experienced cells**. *Nat Commun* 2016, **7**:11112.

10. Lin SG, Ba Z, Du Z, Zhang Y, Hu J, Alt FW: **Highly sensitive and unbiased approach for elucidating antibody repertoires**. *Proc Natl Acad Sci USA* 2016, **113**:7846–7851.

11. Tan YC, Kongpachith S, Blum LK, Ju CH, Lahey LJ, Lu DR, Cai X, Wagner CA, Lindstrom TM, Sokolove J, *et al.*: **Barcode-enabled sequencing of plasmablast antibody repertoires in rheumatoid arthritis**. *Arthritis Rheum* 2014, **66**:2706–2715.

12. DeKosky BJ, Kojima T, Rodin A, Charab W, Ippolito GC, Ellington AD, Georgiou G: **In-depth determination and analysis**

of the human paired heavy- and light-chain antibody repertoire. *Nat Med* 2015, **21**:86–91.

13. Yost KE, Chang HY, Satpathy AT: **Tracking the immune response with single-cell genomics**. *Vaccine* 2020, **38**: 4487–4490.

14. Corcoran MM, Phad GE, Vázquez Bernat N, Stahl-Hennig C, Sumida N, Persson MA, Martin M, Karlsson Hedestam GB: **Production of individualized V gene databases reveals high levels of immunoglobulin genetic diversity**. *Nat Commun* 2016, **7**:13642.

15. Gadala-Maria D, Gidoni M, Marquez S, Vander Heiden JA, Kos JT, Watson CT, O'Connor KC, Yaari G, Kleinstein SH: **Identification of subject-specific immunoglobulin alleles from expressed repertoire sequencing data**. *Front Immunol* 2019, **10**:129.

16. Omer A, Shemesh O, Peres A, Polak P, Shepherd AJ, Watson CT, Boyd SD, Collins AM, Lees W, Yaari G: **VDJbase: an adaptive immune receptor genotype and haplotype database**. *Nucleic Acids Res* 2020, **48**:D1051–D1056.

17. Meysman P, Postovskaya A, De Neuter N, Ogunjimi B, Laukens K: **Tracking SARS-CoV-2 T cells with epitope-T-cell receptor recognition models**. *BioRxiv* 2020, https://doi.org/10.1101/2020.09.09.289355. preprint doi.

18. Watson CT, Breden F: **The immunoglobulin heavy chain locus: genetic variation, missing data, and implications for human disease**. *Gene Immun* 2012, **13**:363–373.

19. • Ohlin M, Scheepers C, Corcoran M, Lees WD, Busse CE, Bagnara D, Thörnqvist L, Bürckert JP, Jackson KJL, Ralph D, *et al.*: **Inferred allelic variants of immunoglobulin receptor genes: a system for their evaluation, documentation, and naming**. *Front Immunol* 2019, **10**:435.
Outlines the criteria for evaluating the confidence in alleles of the immunoglobulin and T-cell receptor genes, as inferred from NGS sequencing of antibody/B-cell and T-cell receptor repertoires (AIRR-seq) data

20. • Lees W, Busse CE, Corcoran M, Ohlin M, Scheepers C, Matsen FA, Yaari G, Watson CT, Community AIRR, Collins A, *et al.*: **OGRDB: a reference database of inferred immune receptor genes**. *Nucleic Acids Res* 2020, **48**:D964–D970.
Describes a data base for curating alleles of the immunoglobulin and T-cell receptor genes, as inferred from NGS sequencing of antibody/B-cell and T-cell receptor repertoires (AIRR-seq) data

21. Lefranc M-P, Giudicelli V, Duroux P, Jabado-Michaloud J, Folch G, Aouinti S, Carillon E, Duvergey H, Houles A, Paysan-Lafosse T, *et al.*: **IMGT®, the international ImMunoGeneTics information system® 25 years on**. *Nucleic Acids Res* 2015, **43**: D413–D422.

22. Galson JD, Schaetzle S, Bashford-Rogers RJM, Raybould MIJ, Kovaltsuk A, Kilpatrick GJ, Minter R, Finch DK, Dias J, James L, *et al.*: **Deep sequencing of B cell receptor repertoires from COVID-19 patients reveals strong convergent immune signatures**. *BioRxiv* 2020, https://doi.org/10.1101/2020.05.20.106294.

23. Kim SI, Noh J, Kim S, Choi Y, Yoo DK, Lee Y, Lee H, Jung J, Kang CK, Song K-H, *et al.*: **Stereotypic neutralizing V_H clonotypes against SARS-CoV-2 RBD in COVID-19 patients and the healthy population**. *BioRxiv* 2020, https://doi.org/10.1101/2020.06.26.174557. preprint doi.

24. Kuri-Cervantes L, Pampena MB, Meng W, Rosenfeld AM, Ittner CAG, Weisman AR, Agyekum RS, Mathew D, Baxter AE, Vella LA, *et al.*: **Comprehensive mapping of immune perturbations associated with severe COVID-19**. *Sci Immunol* 2020, **5**, eabd7114.

25. Montague Z, Lv H, Otwinowski J, DeWitt WS, Isacchini G, Yip GK, Ng WW, Tsang OT, Yuan M, Liu H, Wilson IA, *et al.*: **Dynamics of B-cell repertoires and emergence of cross-reactive responses in COVID-19 patients with different disease severity**. *BioRxiv* 2020, https://doi.org/10.1101/2020.07.13.20153114. preprint.

26. Nielsen SCA, Yang F, Jackson KJL, Hoh RA, Röltgen K, Jean GH, Stevens BA, Lee JY, Rustagi A, Rogers AJ, *et al.*: **Human B cell clonal expansion and convergent antibody responses to SARS-CoV-2**. *Cell Host Microbe* 2020, **S1931–3128**: 30503–30505.

27. Liao M, Liu Y, Yuan J, Wen Y, Xu G, Zhao J, Cheng L, Li J, Wang X, Wang F, *et al.*: **Single-cell landscape of bronchoalveolar immune cells in patients with COVID-19**. *Nat Med* 2020, **26**:842–844.

28. Minervina AA, Komech EA, Titov A, Koraichi MB, Rosati E, Mamedov IZ, Franke A, Efimov GA, Chudakov DM, Mora T, *et al.*: **Longitudinal high-throughput TCR repertoire profiling reveals the dynamics of T cell memory formation after mild COVID-19 infection**. *BioRxiv* 2020, https://doi.org/10.1101/2020.05.18.100545. preprint doi.

29. Nolan S, Vignali M, Klinger M, Dines JN, Kaplan IM, Svejnoha E, Craft T, Boland K, Pesesky M, Gittelman RM, *et al.*: **A large-scale database of T-cell receptor beta (TCRβ) sequences and binding associations from natural and synthetic exposure to SARS-CoV-2**. *BioRxiv* 2020, https://doi.org/10.21203/rs.3.rs-51964/v1. preprint.

30. Shomuradova AS, Vagida MS, Sheetikov SA, Zornikova KV, Kiryukhin D, Titov A, Peshkova IO, Khmelevskaya A, Dianov DV, *et al.*: **SARS-CoV-2 epitopes are recognized by a public and diverse repertoire of human T-cell receptors**. *BioRxiv* 2020, https://doi.org/10.1101/2020.05.20.20107813. preprint doi.

31. Schultheiß C, Paschold L, Simnica D, Mohme M, Willscher E, von Wenserski L, Scholz R, Wieters I, Dahlke C, Tolosa E, *et al.*: **Next-generation sequencing of T and B cell receptor repertoires from COVID-19 patients showed signatures associated with severity of disease**. *Immunity* 2020, **53**:442–455.e4.