

Adaptive immune receptor repertoire analysis

In the format provided by the
authors and unedited

Supplementary Table 1 summarizes major tools for AIRR-sequencing data pre-processing , germline gene annotation and novel gene/allele inference and discovery.

Sr. No	Tools/software	Summary	Conceived for bulk or single-cell (sc) analysis
Pre-processing of sequencing data			
1.	VDJServer ¹	A scalable web-based resource for immune repertoire pre-processing (based on pRESTO and IgBlast), basic repertoire analysis and data sharing.	bulk
2.	immuneDB ²	A Python package and web-based user interface for storing and analyzing large quantities of AIRR-seq data. Interoperable with multiple tools and format, with tools for sequence pre-processing.	bulk
3.	Change-O ³ & pRESTO ⁴ (Immcantation suite)	Python packages for AIRR-seq data processing, including quality control, UMI correction, alignment, and clonal assignment. Offers tools for both pre- and post-processing functionalities.	bulk/sc
4.	MiXCR ⁵	A Java package for AIRR-seq data processing, UMI correction, alignment, and clonal assignment. Also supports scAIRR-seq processing and reconstruction of AIR sequences from RNA-seq data.	bulk/sc
5.	Cell Ranger ⁶	A command line tool to pre-process 10X Genomics single-cell sequencing data. Contains all the functions for sequence alignment, demultiplexing, integration with scRNA-seq data, and supports cell surface feature barcoding.	sc
6.	Vidjil ⁷	An open-source web application for automated preprocessing of AIRR-seq. Specialized in repertoire analysis in hemopathies, integrated with patient databases.	bulk
7.	LymAnalyzer ⁸	A Java package for pre-processing of AIRR-seq data. Performs gene alignment, PCR error correction, novel allele inference,	bulk

		SHM analysis, but no UMI correction or clonotyping.	
8.	Abalign ⁹	A Python package with graphical user interface containing a sequence alignment algorithm built specifically for BCRs. Takes input from sequencing fasta files, but does not support UMI correction.	bulk
9.	dandelion ¹⁰	A Python package to preprocess 10X scAIRR-seq data. Main functionality includes re-annotation of Cell Ranger output, reorganizing non-productive sequences, and improved annotation of $\gamma\delta$ TCR.	sc
10.	sciReptor ¹¹	A Perl & R package to pre-process single-cell BCR sequencing data. Takes input from fastq files and performs quality controls, consensus building, gene annotations, and with the option to integrate flow cytometry data.	sc
11.	IMperm ¹²	An immune paired-end reads merger for sequencing data written in C, capable of handling all types of paired-end reads, eliminating adapter contamination and merging low-quality and minor/non-overlapping reads.	bulk
12.	IMonitor ¹³	A Perl and R package for pre-processing of AIRR-seq data. Performs multiple round of alignment to ensure accurate allele identification, and possesses functionalities for PCR error correction and bias minimization.	bulk
13.	iROAR ¹⁴	A Python package to correct for over amplification bias in multiplex-based AIRR-seq library. Accepts clonotype tables in VDJTools format as input, calculates Over Amplification Rate, and output adjusted clonal count and frequency.	bulk
Germline gene annotation			
14.	TRUST4 ¹⁵	A C package for the extraction, alignment, and annotation of AIR sequences from RNA-seq libraries. Output reconstructed contigs with annotations. Also supports single-cell libraries with cell barcodes.	bulk/sc
15.	SONAR ¹⁶	A Python package for ontogenic analysis of the antibody repertoire. Specialized in longitudinal tracking of antibody sequences.	bulk

16.	IMGT ^{17,18}	A web-based interface for gene annotation of AIR sequences using V-QUEST (high throughput annotation using high V-QUEST).	bulk
17.	IMSEQ ¹⁹	A software in C++ for germline gene assignment from AIRR-seq fastq files. Also supports UMI error correction.	bulk
18.	IgBLAST ²⁰	A web-based tool for alignment of TCR and BCR sequences. Simple to use, supports a variety of germline databases and custom options.	bulk
19.	TraCeR ²¹ & BraCer ²²	Extracts and reconstructs rearranged TCRs and BCRs from short read RNA-seq data. Lacks support for AIRR data representation.	sc
20.	ANARCI ²³	A web-based tool to annotate amino acid sequences of BCRs and TCRs. Supports a wide variety of numbering schemes, but cannot take nucleotide sequences as input, and lacks clonal assignment functionalities.	bulk
Novel gene/allele inference and discovery			
21.	Partis ^{24–27}	A Hidden Markov Model-based framework for BCR and TCR sequence annotation, simulation, clonal family and germline inference, and affinity prediction. Does not support UMI correction and clonotyping.	bulk/sc
22.	TIgGER ²⁸ (Immcanatation suite)	An R package for novel allele inference from AIRR-seq sequences using a predefined germline sequence database.	bulk
23.	IgDiscover ²⁹	A Python package for annotating and discovering new V,D,J genes from AIRR-seq sequencing data using an existing VDJ database as a starting point. Output candidate novel V genes.	bulk
24.	IgDetective ³⁰	A Python package to predict V,D,J genes from genome assembly. Takes input from a genome assembly, detect the RSS motif, then predict the novel genes either based on similarity with existing genes (Iterative), or without any assumptions (Blind). Outputs a table of predicted genes.	bulk

25.	IMPre ³¹	A Perl and C package for predicting V/J germline gene from AIRR-seq data. Takes input as fasta files for AIR sequences and outputs a list of inferred germline genes and annotations.	bulk
26.	SEARCH-D ³²	A Python package for <i>de novo</i> identification of IGHD genes in humans. Takes input from a fasta file containing the IGHD locus in the genome assembly, and outputs de novo predicted IGHD genes.	bulk
27.	RABHIT ³³	An R package for inference of haplotypes and detection of chromosomal deletions from AIRR-seq data. Takes input from clonotype tables in AIRR format, and outputs a table of possible haplotypes/deletions.	bulk

Supplementary Table 2 summarizes major tools for repertoire analysis by category: summary repertoire statistics, single-cell analysis, repertoire similarity analysis, clonal network analysis, SHM and phylogenetic analysis, clustering methods, simulation tools, and machine learning methods. Notably, the list of software is non-exhaustive.

Sr. No.	Tools/software	Summary	Compatibility
Summary repertoire statistics			
1.	Alakazam ^{3,34} (Immcountation suite)	An R package to calculate summary repertoire metrics from outputs of pRESTO, Change-O, and AIRR-format data. Optimized to work with other Immcountation packages.	BCR/TCR
2.	Immunarch ³⁵	An R package to analyze annotated clonotype data from both bulk and single-cell AIRR-seq. Supports reading clonotype output directly from other pre-processing packages (MiXCR, Cell Ranger, vdjil, imseq, AIRR format).	BCR/TCR
3.	VDJtools ³⁶	A Java package to analyze AIRR-seq data. One of the earlier packages that bundles multiple repertoire analysis functions in one package. Input needs to be converted to VDJTools format before use.	BCR/TCR
4.	Vegan ³⁷	An R package with multiple analysis tools for ecology, in which diversity and dissimilarity analyses are the most relevant to AIRR-seq.	BCR/TCR
5.	Sumrep ³⁴	An R package for the assessment and comparison of general repertoire properties.	BCR/TCR
6.	Abalign ⁹	A Python with graphical user interface to calculate summary	BCR

		repertoire statistics, SHM profiling, B-cell lineage tree construction, and humanness assessment by aligning with BCR sequences from the Observed Antibody Space database.	
7.	Recon ³⁸	A Python package to estimate the true diversity of the repertoire from sequencing data. Takes input from a table of clonotype ID and observed size of each clonotype, and outputs a table of Hill numbers for the distribution.	BCR/TCR
8.	AIRRscape ³⁹	An R Shiny web application for interactive exploration of AIRR-seq data. Takes AIRR-format data as input, and output visualizations of gene usage, CDR3 length distribution, and sequence motifs comparison.	BCR/TCR
9.	TCR_nextflow ⁴⁰	A Nextflow pipeline for pre- and post-processing of TCR repertoire data recovered from bulk RNA-seq data. Performs automated sequence alignment, clonotyping with MiXCR, and summary repertoire statistics with Immunarch.	TCR
Single-cell analysis			
10.	CoNGa ⁴¹	A Python package to detect correlation between gene expression and AIR sequences in scAIRR-seq/scRNA-seq data using graph-versus-graph analysis. Takes input from the output gene expression and AIR contigs tables from 10X Genomics Cell Ranger.	BCR/TCR
11.	Tessa ⁴² & Benisse ⁴³	A Python and R package to integrate TCR (Tessa) or BCR (Benisse) sequences with its scRNA-seq data to investigate the functional landscape of these sequences. Takes the AIR sequence matrix and its RNA-seq count matrix as inputs, and outputs a numerical embedding of the AIR sequences for further analysis.	BCR/TCR
12.	Scirpy ⁴⁴	A Python package to analyze scAIRR-seq data. Directly integrates with the Scanpy package (scRNA-seq) and output from many established pipelines (TraCer & BraCer, Cell Ranger, BDRhapsody, AIRR format).	BCR/TCR
13.	scRepertoire ⁴⁵	An R package to process 10X scAIRR-seq data, with built-in integration support with the Seurat package (scRNA-seq) and AIRR-seq sequence autoencoding (Trex and Ibex). Performs clonotyping from contigs and support flexible clonotype definitions.	BCR/TCR
14.	dandelion ¹⁰	A Python package for 10X scAIRR-seq repertoire analysis. Provides functions for spectratyping, calculating gene usage, diversity, mutations, clonal overlap, and network analysis.	BCR/TCR
15.	Platypus & ePlatypus ^{46,47}	An R package for the data analysis of scAIRR-seq data. Functionalities include summary repertoire statistics, sequence similarity networks, integration with gene expression data, sequence encoding and feature extraction of	BCR/TCR

		machine learning tasks. Takes input from the output of 10X Genomics Cell Ranger software. ePlatypus is an expansion of Platypus with a database component to store and access AIRR-seq data.	
16.	sciReptor ¹¹	An R & Python package for the data analysis of scBCR-seq data. In addition to summary repertoire statistics, the package also supports the integration of non-sequencing data such as flow cytometry surface marker expression.	BCR
Repertoire similarity analysis			
17.	immuneREF ⁴⁸	An R package for calculating a similarity score between repertoires based on multiple features such as diversity, germline gene usage, overlap, network architecture. Takes input from AIRR-format clonotype tables, and outputs a set of similarity score for each feature and a global similarity score between repertoires.	BCR/TCR
18.	Repertoire Dissimilarity Index (RDI) ⁴⁹ (Immcanation suite)	An R package to calculate the Repertoire Dissimilarity Index, a non-parametric method to quantify differences in germline V,D,J gene usage with adjustment for sampling depth and simulated data for calibration of the RDI index calculation.	BCR/TCR
19.	CompAIRR ⁵⁰	A command line tool to compare repertoires and compute their overlap at ultra-high throughput.	BCR/TCR
Clonal network analysis			
20.	igraph ⁵¹	A collection of network analysis tools available in R, python, C/C++ and mathematica.	BCR/TCR
21.	imNET ⁵²	A Python software package for generation and analysis of large-scale immunological and biological sequence networks.	BCR/TCR
22.	cytoscape ⁵³	A generalized open source software platform for visualizing complex networks and integrating these with any type of attribute data. Not suitable for large networks.	BCR/TCR
23.	Gephi ⁵⁴	A generalized visualization and exploration software for all kinds of graphs and networks.	BCR/TCR
24.	networkx	A Python package for the creation, manipulation, and study of the structure, dynamics, and functions of complex networks.	BCR/TCR
25.	ALICE ⁵⁵	An R package for the detection of TCRs involved in immune response from AIRR-seq data based on recombination frequency and sequence similarity.	TCR
26.	TraCeR ²¹ & BraCer ²²	A sub-command to create network graph of reconstructed BCR/TCR clonotypes from RNA-seq data.	BCR/TCR

SHM and phylogenetic analysis (BCRs)			
27.	MiXCR ⁵ (export command)	A sub-command of MiXCR to construct clonal lineage from MiXCR-defined clonotypes.	BCR
28.	SHazaM ³ & dowser ⁵⁶ (Immcanation suite)	R packages for analysis of SHM and lineage reconstruction in BCR repertoire data.	BCR
29.	IgIDivA ⁵⁷	An R Shiny application for analysis of SHM. Takes input from annotated BCR sequences by IMGT/HighV-QUEST and clonotyping by tripr.	BCR
30.	ClonalTree ⁵⁸	A Python package to construct BCR lineage trees using maximum parsimony. Take raw BCR fasta sequences as input, and output lineage trees.	BCR
31.	BraCeR ²²	A sub-command of BraCeR to infer clonal lineage from constructed BCR sequences based on PHYLIP.	BCR
32.	Ape ⁵⁹	An R package for reading, constructing, plotting, and manipulating phylogenetic trees.	BCR
33.	IgPhyML ^{60,61}	A python tool for building phylogenetic trees from BCRs and for testing the evolutionary hypotheses of B cell affinity maturation (integrated with the Immcanation framework).	BCR
34.	gctree ⁶²	A command line utility and Python package for using sequence abundance information in phylogenetic inference.	BCR
35.	PHYLIP ⁶³	A collection of programs for inferring phylogenies (evolutionary trees).	BCR
36.	AncesTree ⁶⁴	A Java package graphical interface to visualize phylogenetic trees. Takes input from the output of IgPhyML and PHYLIP.	BCR
37.	ViCloD ⁶⁵	A Python package (and web interface) for visualization of clonal lineage. Takes input from AIRR-format clonotype table.	BCR
Clustering methods			
38.	SCOPer ⁶⁶ (Immcanation suite)	An R package to cluster clonotypes based on predefined threshold of sequence similarity. Also supports cell barcodes to cluster on single-cell level.	BCR/TCR
39.	ClusTCR ⁶⁷	A Python package for efficient clustering of large sets of TCR sequences, independent of antigen specificity.	TCR
40.	GIANA ⁶⁸	A Python package for TCR sequence clustering.	TCR
41.	TCRdist ⁶⁹	A Python package for clustering of TCR sequences by	TCR

		distance. Takes input from a table containing the TCR α/β nucleotide sequences and the corresponding epitope, and output the TCR sequence clusters.	
42.	GLIPH ⁷⁰	A command line tool for grouping TCR sequences by their CDR3 motifs. Take inputs from a table of TCRs containing CDR3 amino acid sequences, V and J genes, count and metadata. Output a list of motifs, TCR convergence groups, and predicted HLA binding.	TCR
43.	TCRanno ⁷¹	A Python package to predict and quantify antigen specificity from TCR sequences based on databases of experimentally validated TCR-epitope amino acid sequence pairs.	TCR
44.	TCRMatch ⁷²	A Python package to predict antigen specificity based on sequence similarity to validated antigen-specific sequences.	TCR
Simulation tools			
45.	immuneSIM ⁷³	An R package to generate human and mouse AIR sequences with many user-defined properties, including: germline gene usage, clone count distribution, insertion and deletion likelihood, SHM likelihood, and motif implantation.	BCR/TCR
46.	IGoR ⁷⁴	A C++ software to infer V(D)J recombination from sequencing data. Functionalities include calculating generation probability of a sequence, hypermutation models, generating sequences following a recombination model.	BCR/TCR
46.	OLGA ⁷⁵	A Python software to calculate the generation probability of CDR3 sequences from a recombination model. Can also generate sequences from a recombination model.	BCR/TCR
47.	SoNNia ⁷⁶	A Python software to infer selection pressure on the generation of V(D)J sequences. Supports paired-chain sequences.	BCR/TCR
48.	Absolut! ⁷⁷	A C++ package to generate synthetic antibody-antigen binding pairs at a large scale. Take inputs from: an antigen PDB entry and discretize the structure into a 3D lattice, an antibody CDR3 amino acid sequence and compute their binding to the lattice antigen. Outputs antibody-antigen binding features to be used for machine learning tasks.	BCR
49.	Echidna ⁷⁸	An R package to simulate scAIRR-seq and RNA-seq data with the ability to define repertoire features such as germline gene usage, SHM, gene expression levels and also spatial locations. Compatible with the package Platypus.	BCR/TCR
50.	DIMSIM ⁷⁹	A C++ tool for the simulation of B cell clonal lineage evolution in order to benchmark clonal lineage reconstruction methods.	BCR
51.	IMPIAntS ⁸⁰	A Python package to simulate BCR heavy chain sequences for benchmarking of repertoire analysis tools. Simulated	BCR

		data can have specific characteristics defined including germline gene usage, junctional indels, SHM by position and frequency, and also structure of NGS sequencing data.	
Machine learning methods			
52.	immuneML ⁸¹	A Python platform for machine learning-based analysis and classification of AIRRs. Provides broad support for a variety of AIRR-seq data format, training models, tools for benchmarking ML methods, and simulated data for model evaluation.	BCR/TCR
53.	Immune2vec ⁸²	A Python package based on Natural Language Processing for repertoire classification. The model uses the word2vec model for embedding the cdr3 AA sequences into vectors with fixed size.	BCR/TCR
54.	TCRex ⁸³	A web tool for the prediction of epitope binding based on the human TCR β CDR3 sequence. The model was trained based on manually curated data on McPAS-TCR, VDJdb, and ImmuneCODE TM .	TCR
55.	DeepTCR ⁸⁴	A Python package containing various deep learning methods applicable to TCR data. Functionalities include classifying antigen-specific TCRs, identifying antigen-specific TCRs from scRNA-seq data, and TCR repertoire classification.	TCR
56.	NetTCR ⁸⁵	A web tool to predict the probability of binding between TCR CDR sequences and MHC-I peptides.	TCR
57.	NoisET ⁸⁶	A Python package for learning experimental noise related to sequence depth from sequencing data replicates in order to discover true clonal expansion following stimulus. Takes input from clonotype tables and outputs a list of clonotypes responding to a stimulus based on the inferred noise model.	TCR
58.	AIRIVA ⁸⁷	A generative model that learns a low-dimensional, interpretable, and compositional representation of TCR repertoires devoid of population genetics and sequencing depth effects. It can learn from unlabeled samples; generate <i>in silico</i> TCR repertoires by intervening on the latent factors, and identify disease-associated TCRs.	TCR
59.	DeepRC ⁸⁸	A Python package utilizing deep learning for the classification of immune repertoires. Takes input from either experimental or simulated AIRR-seq data.	BCR/TCR
60.	MAL-ID ⁸⁹	An interpretive machine learning framework that can screen for multiple illnesses simultaneously. It learns decision rules based on repertoire composition, CDR3 clustering, and language embedding.	BCR/TCR

Supplementary Table 3: AIRR-seq sequence databases

Tools/software	Summary	Compatibility
iReceptor ⁹⁰	A web-based user interface for the discovery, analysis and download of data from the AIRR Data Commons.	bulk/sc
VDJServer Community Data Portal ¹	A web portal to access community sequencing data stored on VDJServer and to data in other repositories in the AIRR Data Commons.	bulk/sc
McPAS-TCR ⁹¹	A manually curated web database of TCR sequences that are associated with pathology in human and mice.	bulk/sc
VDJdb ⁹²	A web portal for a curated database of TCR sequences with known antigen specificities in specific MHC contexts.	bulk/sc
CoV-AbDab ⁹³	A web database of coronavirus binding antibody sequences with associated protein structures.	bulk/sc
OGRDB ⁹⁴	A web database of germline AIR sequences for human and mouse. Specialized in curating novel alleles and germline gene sets specific for each mouse strains.	bulk/sc
VDJbase ⁹⁵	A web database of published human and mouse AIR sequences with accompanied genotype and haplotype data inferred across populations.	bulk/sc

Supplementary references

1. Christley, S. *et al.* VDJServer: A Cloud-Based Analysis Portal and Data Commons for Immune Repertoire Sequences and Rearrangements. *Front. Immunol.* **9**, 976 (2018).
2. Rosenfeld, A. M., Meng, W., Luning Prak, E. T. & Hershberg, U. ImmuneDB, a Novel Tool for the Analysis, Storage, and Dissemination of Immune Repertoire Sequencing Data. *Front. Immunol.* **9**, 2107 (2018).
3. Gupta, N. T. *et al.* Change-O: a toolkit for analyzing large-scale B cell immunoglobulin repertoire sequencing data. *Bioinformatics* **31**, 3356–3358 (2015).
4. Vander Heiden, J. A. *et al.* pRESTO: a toolkit for processing high-throughput sequencing raw reads of lymphocyte receptor repertoires. *Bioinformatics* **30**, 1930–1932 (2014).
5. Bolotin, D. A. *et al.* MiXCR: software for comprehensive adaptive immunity profiling. *Nat. Methods* **12**, 380–381 (2015).
6. Zheng, G. X. Y. *et al.* Massively parallel digital transcriptional profiling of single cells. *Nat. Commun.* **8**, 14049–14049 (2017).
7. Giraud, M. *et al.* Fast multiclonal clusterization of V(D)J recombinations from high-throughput sequencing. *BMC Genomics* **15**, 409 (2014).
8. Yu, Y., Ceredig, R. & Seoighe, C. LymAnalyzer: a tool for comprehensive analysis of next generation

- sequencing data of T cell receptors and immunoglobulins. *Nucleic Acids Res.* **44**, e31 (2016).
9. Zong, F. *et al.* Abalign: a comprehensive multiple sequence alignment platform for B-cell receptor immune repertoires. *Nucleic Acids Res.* (2023) doi:10.1093/nar/gkad400.
 10. Suo, C. *et al.* Dandelion utilizes single cell adaptive immune receptor repertoire to explore lymphocyte developmental origins. *bioRxiv* 2022.11.18.517068 (2023) doi:10.1101/2022.11.18.517068.
 11. Imkeller, K., Arndt, P. F., Wardemann, H. & Busse, C. E. sciReptor: analysis of single-cell level immunoglobulin repertoires. *BMC Bioinformatics* **17**, 67 (2016).
 12. Zhang, W. *et al.* IMperm: a fast and comprehensive IMMune Paired-End Reads Merger for sequencing data. *Brief. Bioinform.* **24**, (2023).
 13. Zhang, W. *et al.* IMonitor: A Robust Pipeline for TCR and BCR Repertoire Analysis. *Genetics* **201**, 459–472 (2015).
 14. Smirnova, A. O. *et al.* The use of non-functional clonotypes as a natural calibrator for quantitative bias correction in adaptive immune receptor repertoire profiling. *Elife* **12**, (2023).
 15. Song, L. *et al.* TRUST4: immune repertoire reconstruction from bulk and single-cell RNA-seq data. *Nat. Methods* **18**, 627–630 (2021).
 16. Schramm, C. A. *et al.* SONAR: A High-Throughput Pipeline for Inferring Antibody Ontogenies from Longitudinal Sequencing of B Cell Transcripts. *Front. Immunol.* **7**, 372 (2016).
 17. Brochet, X., Lefranc, M.-P. & Giudicelli, V. IMGT/V-QUEST: the highly customized and integrated system for IG and TR standardized VJ and VDJ sequence analysis. *Nucleic Acids Res.* **36**, W503–W508 (2008).
 18. Alamyar, E., Giudicelli, V., Li, S., Duroux, P. & Lefranc, M.-P. IMGT/HighV-QUEST: the IMGT® web portal for immunoglobulin (IG) or antibody and T cell receptor (TR) analysis from NGS high throughput and deep sequencing. *Immunome Res.* **8**, 26 (2012).
 19. Kuchenbecker, L. *et al.* IMSEQ—a fast and error aware approach to immunogenetic sequence analysis. *Bioinformatics* **31**, 2963–2971 (2015).
 20. Ye, J., Ma, N., Madden, T. L. & Ostell, J. M. IgBLAST: an immunoglobulin variable domain sequence analysis tool. *Nucleic Acids Res.* **41**, W34–40 (2013).
 21. Stubbington, M. J. T. *et al.* T cell fate and clonality inference from single-cell transcriptomes. *Nat. Methods* **13**, 329–332 (2016).
 22. Lindeman, I. *et al.* BraCeR: B-cell-receptor reconstruction and clonality inference from single-cell RNA-seq. *Nat. Methods* **15**, 563–565 (2018).
 23. Dunbar, J. & Deane, C. M. ANARCI: antigen receptor numbering and receptor classification. *Bioinformatics* **32**, 298–300 (2016).
 24. Ralph, D. K. & Matsen, F. A., 4th. Consistency of VDJ Rearrangement and Substitution Parameters Enables Accurate B Cell Receptor Sequence Annotation. *PLoS Comput. Biol.* **12**, e1004409 (2016).
 25. Ralph, D. K. & Matsen, F. A., 4th. Per-sample immunoglobulin germline inference from B cell receptor deep sequencing data. *PLoS Comput. Biol.* **15**, e1007133 (2019).
 26. Ralph, D. K. & Matsen, F. A., 4th. Inference of B cell clonal families using heavy/light chain pairing information. *PLoS Comput. Biol.* **18**, e1010723 (2022).
 27. Ralph, D. K. & Matsen, F. A., 4th. Using B cell receptor lineage structures to predict affinity. *PLoS Comput. Biol.* **16**, e1008391 (2020).
 28. Gadala-Maria, D. *et al.* Identification of Subject-Specific Immunoglobulin Alleles From Expressed Repertoire Sequencing Data. *Front. Immunol.* **10**, 129 (2019).
 29. Corcoran, M. M. *et al.* Production of individualized V gene databases reveals high levels of immunoglobulin genetic diversity. *Nat. Commun.* **7**, 13642 (2016).
 30. Sirupurapu, V., Safonova, Y. & Pevzner, P. A. Gene prediction in the immunoglobulin loci. *Genome Res.* **32**, 1152–1169 (2022).
 31. Zhang, W. *et al.* IMPre: An Accurate and Efficient Software for Prediction of T- and B-Cell Receptor Germline Genes and Alleles from Rearranged Repertoire Data. *Front. Immunol.* **7**, 457 (2016).
 32. Safonova, Y. & Pevzner, P. A. V(DD)J recombination is an important and evolutionarily conserved mechanism for generating antibodies with unusually long CDR3s. *Genome Res.* **30**, 1547–1558 (2020).
 33. Peres, A., Gidoni, M., Polak, P. & Yaari, G. RABHIT: R Antibody Haplotype Inference Tool. *Bioinformatics* **35**, 4840–4842 (2019).
 34. Olson, B. J. *et al.* sumrep: A Summary Statistic Framework for Immune Receptor Repertoire Comparison and Model Validation. *Front. Immunol.* **10**, 2533 (2019).
 35. ImmunoMind Team. immunarch: An R Package for Painless Bioinformatics Analysis of T-Cell and B-Cell Immune Repertoires. Preprint at <https://doi.org/10.5281/zenodo.3367200> (2019).
 36. Shugay, M. *et al.* VDJtools: Unifying Post-analysis of T Cell Receptor Repertoires. *PLoS Comput. Biol.* **11**,

- e1004503 (2015).
37. Oksanen, J. *et al.* Package ‘vegan’. *Community ecology package, version 2*, 1–295 (2013).
 38. Kaplinsky, J. & Arnaout, R. Robust estimates of overall immune-repertoire diversity from high-throughput measurements on samples. *Nat. Commun.* **7**, 11881 (2016).
 39. Waltari, E., Nafees, S., McCutcheon, K. M., Wong, J. & Pak, J. E. AIRRscape: An interactive tool for exploring B-cell receptor repertoires and antibody responses. *PLoS Comput. Biol.* **18**, e1010052 (2022).
 40. Rubio, T. *et al.* A Nextflow pipeline for T-cell receptor repertoire reconstruction and analysis from RNA sequencing data. *Immunoinformatics* **6**, 100012 (2022).
 41. Schattgen, S. A. *et al.* Integrating T cell receptor sequences and transcriptional profiles by clonotype neighbor graph analysis (CoNGA). *Nat. Biotechnol.* **40**, 54–63 (2022).
 42. Zhang, Z., Xiong, D., Wang, X., Liu, H. & Wang, T. Mapping the functional landscape of T cell receptor repertoires by single-T cell transcriptomics. *Nat. Methods* **18**, 92–99 (2021).
 43. Zhang, Z. *et al.* Interpreting the B-cell receptor repertoire with single-cell gene expression using Benisse. *Nature Machine Intelligence* **4**, 596–604 (2022).
 44. Sturm, G. *et al.* Scirpy: a Scanpy extension for analyzing single-cell T-cell receptor-sequencing data. *Bioinformatics* **36**, 4817–4818 (2020).
 45. Borcherding, N., Bormann, N. L. & Kraus, G. scRepertoire: An R-based toolkit for single-cell immune receptor analysis. *F1000Res*. **9**, 47 (2020).
 46. Yermanos, A. *et al.* Platypus: an open-access software for integrating lymphocyte single-cell immune repertoires with transcriptomes. *NAR Genom Bioinform* **3**, lqab023 (2021).
 47. Kreiner, V. *et al.* ePlatypus: an ecosystem for computational analysis of immunogenomics data. *bioRxiv* 2022.09.28.509709 (2022) doi:10.1101/2022.09.28.509709.
 48. Weber, C. R. *et al.* Reference-based comparison of adaptive immune receptor repertoires. *Cell Rep Methods* **2**, 100269 (2022).
 49. Bolen, C. R., Rubelt, F., Vander Heiden, J. A. & Davis, M. M. The Repertoire Dissimilarity Index as a method to compare lymphocyte receptor repertoires. *BMC Bioinformatics* **18**, 155 (2017).
 50. Rognes, T., Scheffer, L., Greiff, V. & Sandve, G. K. CompAIRR: ultra-fast comparison of adaptive immune receptor repertoires by exact and approximate sequence matching. *Bioinformatics* **38**, 4230–4232 (2022).
 51. Csardi, G. & Nepusz, T. The igraph software package for complex network research. *InterJournal* vol. Complex Systems 1695 Preprint at <https://igraph.org> (2006).
 52. Miho, E., Roškar, R., Greiff, V. & Reddy, S. T. Large-scale network analysis reveals the sequence space architecture of antibody repertoires. *Nat. Commun.* **10**, 1321–11 (2019).
 53. Shannon, P. *et al.* Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504 (2003).
 54. Bastian, M., Heymann, S. & Jacomy, M. Gephi: An Open Source Software for Exploring and Manipulating Networks. *ICWSM* **3**, 361–362 (2009).
 55. Pogorelyy, M. V. *et al.* Detecting T cell receptors involved in immune responses from single repertoire snapshots. *PLoS Biol.* **17**, e3000314 (2019).
 56. Hoehn, K. B., Pybus, O. G. & Kleinstei, S. H. Phylogenetic analysis of migration, differentiation, and class switching in B cells. *PLoS Comput. Biol.* **18**, e1009885 (2022).
 57. Zaragoza-Infante, L. *et al.* IglDivA: immunoglobulin intraclonal diversification analysis. *Brief. Bioinform.* **23**, (2022).
 58. Abdollahi, N., Jeusset, L., de Septenville, A., Davi, F. & Bernardes, J. S. Reconstructing B cell lineage trees with minimum spanning tree and genotype abundances. *BMC Bioinformatics* **24**, 70 (2023).
 59. Paradis, E. & Schliep, K. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* **35**, 526–528 (2019).
 60. Hoehn, K. B., Lunter, G. & Pybus, O. G. A Phylogenetic Codon Substitution Model for Antibody Lineages. *Genetics* **206**, 417–427 (2017).
 61. Hoehn, K. B. *et al.* Repertoire-wide phylogenetic models of B cell molecular evolution reveal evolutionary signatures of aging and vaccination. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 22664–22672 (2019).
 62. DeWitt, W. S., 3rd, Mesin, L., Vitoria, G. D., Minin, V. N. & Matsen, F. A., 4th. Using Genotype Abundance to Improve Phylogenetic Inference. *Mol. Biol. Evol.* **35**, 1253–1265 (2018).
 63. Felsenstein, J. PHYLIP (phylogeny inference package) version 3.695. *Distributed by the*.
 64. Foglierini, M., Pappas, L., Lanzavecchia, A., Corti, D. & Perez, L. AncesTree: An interactive immunoglobulin lineage tree visualizer. *PLoS Comput. Biol.* **16**, e1007731 (2020).
 65. Jeusset, L. *et al.* ViCloD, an interactive web tool for visualizing B cell repertoires and analyzing intraclonal diversities: application to human B-cell tumors. *bioRxiv* 2022.11.28.518196 (2023)

- doi:10.1101/2022.11.28.518196.
66. Nouri, N. & Kleinstein, S. H. A spectral clustering-based method for identifying clones from high-throughput B cell repertoire sequencing data. *Bioinformatics* **34**, i341–i349 (2018).
 67. Valkiers, S., Van Houcke, M., Laukens, K. & Meysman, P. ClusTCR: a Python interface for rapid clustering of large sets of CDR3 sequences with unknown antigen specificity. *Bioinformatics* (2021) doi:10.1093/bioinformatics/btab446.
 68. Zhang, H., Zhan, X. & Li, B. GIANA allows computationally-efficient TCR clustering and multi-disease repertoire classification by isometric transformation. *Nat. Commun.* **12**, 4699 (2021).
 69. Dash, P. *et al.* Quantifiable predictive features define epitope-specific T cell receptor repertoires. *Nature* **547**, 89–93 (2017).
 70. Glanville, J. *et al.* Identifying specificity groups in the T cell receptor repertoire. *Nature* **547**, 94–98 (2017).
 71. Luo, J. *et al.* Quantitative Annotations of T-Cell Repertoire Specificity. *bioRxiv* 2023.01.29.526145 (2023) doi:10.1101/2023.01.29.526145.
 72. Chronister, W. D. *et al.* TCRMatch: Predicting T-Cell Receptor Specificity Based on Sequence Similarity to Previously Characterized Receptors. *Front. Immunol.* **12**, 640725 (2021).
 73. Weber, C. R. *et al.* immuneSIM: tunable multi-feature simulation of B- and T-cell receptor repertoires for immunoinformatics benchmarking. *Bioinformatics* **36**, 3594–3596 (2020).
 74. Marcou, Q., Mora, T. & Walczak, A. M. High-throughput immune repertoire analysis with IGoR. *Nat. Commun.* **9**, 561 (2018).
 75. Sethna, Z., Elhanati, Y., Callan, C. G., Walczak, A. M. & Mora, T. OLGA: fast computation of generation probabilities of B- and T-cell receptor amino acid sequences and motifs. *Bioinformatics* **35**, 2974–2981 (2019).
 76. Isacchini, G., Walczak, A. M., Mora, T. & Nourmohammad, A. Deep generative selection models of T and B cell receptor repertoires with soNNia. *Proc. Natl. Acad. Sci. U. S. A.* **118**, (2021).
 77. Robert, P. A. *et al.* Unconstrained generation of synthetic antibody–antigen structures to guide machine learning methodology for antibody specificity prediction. *Nature Computational Science* **2**, 845–865 (2022).
 78. Han, J. *et al.* Echidna: integrated simulations of single-cell immune receptor repertoires and transcriptomes. *Bioinform Adv* **2**, vbac062 (2022).
 79. Zhang, C., Bzikadze, A. V., Safonova, Y. & Mirarab, S. A scalable model for simulating multi-round antibody evolution and benchmarking of clonal tree reconstruction methods. *Front. Immunol.* **13**, 1014439 (2022).
 80. Yang, X. *et al.* Novel Allele Detection Tool Benchmark and Application With Antibody Repertoire Sequencing Dataset. *Front. Immunol.* **12**, 739179 (2021).
 81. Pavlović, M. *et al.* The immuneML ecosystem for machine learning analysis of adaptive immune receptor repertoires. *Nature Machine Intelligence* **3**, 936–944 (2021).
 82. Ostrovsky-Berman, M., Frankel, B., Polak, P. & Yaari, G. Immune2vec: Embedding B/T Cell Receptor Sequences in $\mathbb{R}^N$ Using Natural Language Processing. *Front. Immunol.* **12**, 680687 (2021).
 83. Gielis, S. *et al.* Detection of Enriched T Cell Epitope Specificity in Full T Cell Receptor Sequence Repertoires. *Front. Immunol.* **10**, 2820 (2019).
 84. Sidhom, J.-W., Larman, H. B., Pardoll, D. M. & Baras, A. S. DeepTCR is a deep learning framework for revealing sequence concepts within T-cell repertoires. *Nat. Commun.* **12**, 1605 (2021).
 85. Montemurro, A., Jessen, L. E. & Nielsen, M. NetTCR-2.1: Lessons and guidance on how to develop models for TCR specificity predictions. *Front. Immunol.* **13**, 1055151 (2022).
 86. Koraichi, M. B., Touzel, M. P., Mazzolini, A., Mora, T. & Walczak, A. M. NoiseET: Noise Learning and Expansion Detection of T-Cell Receptors. *J. Phys. Chem. A* **126**, 7407–7414 (2022).
 87. Pradier, M. F. *et al.* AIRIVA: A Deep Generative Model of Adaptive Immune Repertoires. *arXiv [q-bio.QM]* (2023).
 88. Widrich, M. *et al.* Modern hopfield networks and attention for immune repertoire classification. *Adv. Neural Inf. Process. Syst.* **33**, 18832–18845 (2020).
 89. Zaslavsky, M. E. *et al.* Disease diagnostics using machine learning of immune receptors. *bioRxiv* (2023) doi:10.1101/2022.04.26.489314.
 90. Corrie, B. D. *et al.* iReceptor: A platform for querying and analyzing antibody/B-cell and T-cell receptor repertoire data across federated repositories. *Immunol. Rev.* **284**, 24–41 (2018).
 91. Tickotsky, N., Sagiv, T., Prilusky, J., Shifrut, E. & Friedman, N. McPAS-TCR: a manually curated catalogue of pathology-associated T cell receptor sequences. *Bioinformatics* (2017) doi:10.1093/bioinformatics/btx286.

92. Shugay, M. *et al.* VDJdb: a curated database of T-cell receptor sequences with known antigen specificity. *Nucleic Acids Res.* (2018) doi:10.1093/nar/gkx760.
93. Raybould, M. I. J., Kovaltsuk, A., Marks, C. & Deane, C. M. CoV-AbDab: the coronavirus antibody database. *Bioinformatics* **37**, 734–735 (2021).
94. Lees, W. *et al.* OGRDB: a reference database of inferred immune receptor genes. *Nucleic Acids Res.* **48**, D964–D970 (2020).
95. Omer, A. *et al.* VDJbase: an adaptive immune receptor genotype and haplotype database. *Nucleic Acids Res.* **48**, D1051–D1056 (2020).