Journal Pre-proof

Immune literacy: reading, writing and editing adaptive immunity

Lucia Csepregi, Roy Ehling, Bastian Wagner, Sai T. Reddy

PII: S2589-0042(20)30711-2

DOI: https://doi.org/10.1016/j.isci.2020.101519

Reference: ISCI 101519

To appear in: *ISCIENCE*

Please cite this article as: Csepregi, L., Ehling, R., Wagner, B., Reddy, S.T, Immune literacy: reading, writing and editing adaptive immunity, *ISCIENCE* (2020), doi: https://doi.org/10.1016/j.isci.2020.101519.

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2020

Journal Pre-proof

Immune literacy: reading, writing and editing adaptive immunity

Lucia Csepregi, Roy Ehling, Bastian Wagner and Sai T Reddy*

Department of Biosystems Science and Engineering, ETH Zurich, 4058, Basel, Switzerland
Correspondence: *sai.reddy@ethz.ch

ABSTRACT

Advances in reading, writing and editing DNA are providing unprecedented insights into the complexity of immunological systems. This combination of systems and synthetic biology methods is enabling the quantitative and precise understanding of molecular recognition in adaptive immunity, thus providing a framework for reprogramming immune responses for translational medicine. In this review, we will highlight state-of-the-art methods such as immune repertoire sequencing, immunoinformatics and immunogenomic engineering and their application towards adaptive immunity. We showcase novel and interdisciplinary approaches that have the promise of transforming the design and breadth of molecular and cellular immunotherapies.

INTRODUCTION

Lymphocyte B and T cells play central roles in adaptive immunity by mounting molecularly-targeted responses against foreign pathogens, providing both short-term and long-term immunological protection. Molecular recognition and specificity against pathogenic antigens is provided through adaptive immune receptors: B cell receptors (BCRs, secreted version of antibodies) and T cell receptors (TCRs), which are both dimeric protein complexes consisting of heavy (HC) and light chains (LC) for BCR and alpha (⍰) and beta chains (β) (or γ and δ) for TCR. The diversity of both BCR and TCR needed for recognizing foreign antigens is generated through combinatorial rearrangements of genomic germline variable-, (diversity-), joining-segments (V(D)J recombination) during lymphocyte development (Tonegawa 1983). Imprecise joining of these segments during V(D)J recombination leads to template-independent addition or deletion of nucleotides, which together with combinatorial pairing of receptor chains further increases diversity. Due to these diversification processes, each organism contains a highly unique B and T cell repertoire, in which their composition, dynamics and evolution are constantly shaped by extrinsic (i.e., pathogens, vaccination) and intrinsic forces (i.e., tumors, autoimmunity, aging). This leaves each individual with a unique immunological fingerprint that provides information about the current and past states of adaptive immunity (Greiff, Bhat et al. 2015).

Rapid progress in immune repertoire sequencing enables high-throughput interrogation of the genetic diversity of adaptive immune receptor repertoires at both the individual and population-wide level, thus providing unprecedented molecular insight on adaptive immunity in the context of infection, vaccination, aging, autoimmunity and cancer (Emerson et al. 2017; Galson et al. 2015; Egorov et al. 2018; Bashford-Rogers et al. 2019; 2016). Immune receptor sequencing as well as recent progress in single-cell approaches excel at capturing the diversity of immune cell populations and their complex transcriptional profiles (Papalexi and Satija 2018). Increasing efforts are also being put into immunoinformatic tools to decode repertoire sequence information, with a primary goal of identifying specificity to antigens (Kidd et

Journal Pre-proof

al. 2014; Shugay et al. 2018; Brown et al. 2019). Furthermore, immunogenomic engineering platforms are being developed to interrogate immune receptors at high-throughput, which are providing a means to unravel the functional properties of BCRs and TCRs and tune their specificity and selectivity (Pogson et al. 2016; Eyquem et al. 2017). Recently, advances in genome editing technology are making it possible to reprogram the specificity and function of adaptive immune cells, which leads to the promise of generating "living drugs" that can sense and respond dynamically to disease targets (June et al. 2018; Li et al. 2019). In this review, we will highlight methods from both systems and synthetic immunology which are being utilized in a highly interdisciplinary manner to address fundamental questions in adaptive immunity and develop molecular and cellular immunotherapies (Figure 1).

Reading adaptive immunity by immune repertoire sequencing

Immune repertoire sequencing of adaptive immune receptor repertoires (BCRs and TCRs) has become an indispensable tool for the interrogation of the highly complex adaptive immune response. Being able to elucidate immune repertoire sequences and their quantitative composition on a nucleotide-level within and across individuals facilitates the identification of specific immune cells driving a response upon antigen recognition. When choosing from the variety of sequencing platforms, there are many aspects that need to be taken into consideration such as sequencing depth, read length and sequencing accuracy. Hereinafter, we provide a brief overview on the technological progress of sequencing platforms in the context of immune repertoires.

Platforms for immune repertoire sequencing

The emergence of deep sequencing platforms has enabled immune repertoire sequencing and analysis at a high-throughput level (Weinstein et al. 2009; Robins et al. 2009; Freeman et al. 2009; Wang et al. 2010; Robins et al. 2010; Warren et al. 2011; Venturi et al. 2011). Massively parallel sequencing was made possible due to the miniaturization and simultaneous operation of sequencing reactions on millions of DNA templates (Margulies et al. 2005). The most widely used technology from Illumina relies on DNA template amplification before sequencing in order to obtain sufficient signal for detection (Metzker 2010). A variety of different instruments are provided by Illumina (e.g. MiSeq, HiSeq, NextSeq and NovaSeq) with outputs ranging from 15 Gb and 25 million reads (MiSeq) to 6000 Gb and 20 billion reads (NovaSeq) with read lengths up to 300 bp (MiSeq). The overall error rate of all Illumina instruments is relatively low and lies around 0.1% with the most common error being single nucleotide substitutions (Pfeiffer et al. 2018). Sequencing errors in immune repertoires lead to artificial increase of clonal diversity as erroneous clones cannot be distinguished from naturally occurring clonal variants, leading to potentially flawed immunological interpretations. As Illumina platforms perform bridge amplification, it is possible to read both ends of the DNA fragments by paired-end sequencing and achieve longer read lengths (e.g., 2 x 300 bp). This is advantageous as merging two overlapping reads can cover the entire variable regions of BCRs and TCRs and partially correct for errors in clonal identifier regions such as complementarity-determining region 3 (CDR3) (Schirmer et al. 2015; Friedensohn, Khan, and Reddy 2017). To date, well-established Illumina devices clearly dominate the field of immune repertoire sequencing as they offer high quality short-sequencing reads at unexcelled throughput. Deep resolution of the vast immune repertoire sequence space enables better insights into the diversity of unique BCR and TCR sequences (>10^{13} in

Journal Pre-proof

humans) and allows detection of shared clones across individuals, so called 'public clones' (Soto et al. 2019; Briney et al. 2019; Greiff, Menzel, et al. 2017; DeWitt et al. 2018; Emerson et al. 2017).

Paired-end sequencing on Illumina allows coverage of one variable region of immune receptors, however for approaches that rely on complete immune receptor sequence coverage and hence require longer read lengths, an alternative platform is offered by Pacific Biosciences (PacBio) (Eid et al. 2009). PacBio's sequencing technology is based on single-molecule sequencing. Each DNA template molecule gets isolated in a microwell, together with a single polymerase and fluorescently labeled nucleotides. Importantly, the DNA templates are circularized before sequencing, allowing for repeated readout by the polymerase. This allows for consensus reading, achieving read accuracies of over 99% with high-fidelity read lengths up to 15 kb (Wenger et al. 2019). With the recent introduction of the PacBio Sequel II system, sequencing depths of 4 million reads per run can be achieved. Hence, PacBio can be utilized for sequencing immune receptor repertoires where variable regions [e.g., variable heavy (VH) and variable light (VL)] are physically linked in the context of single-cell sequencing (DeKosky et al. 2013; McDaniel et al. 2016). Moreover, long-read sequencing is in particular valuable for in vitro screening platforms (e.g., phage display) which for instance utilize recombinant antibody fragment libraries known as single-chain variable fragments (scFvs). Sequencing of full-length scFvs is important for covering both VH and VL diversity which is critical in assessing quality of such libraries and their utility for drug discovery (Hemadou et al. 2017; Giudicelli et al. 2017; Han et al. 2018).

Finally, Oxford Nanopore offers another single-molecule sequencing platform with promising future implications for immune repertoire sequencing due to the capacity of producing ultra-long reads (up to 2 Mb) (Clarke et al. 2009; Payne et al. 2019). This can be leveraged for the detection of additional diversification mechanisms in immune receptors, such as chromosomal integrations into variable regions (Tan et al. 2016), as well as for studies combining transcriptome and V(D)J sequence analysis (Cole et al. 2020; Byrne et al. 2017).

Single cell profiling of immune repertoires

Studies performing immune repertoire sequencing on bulk B or T cell populations mostly focus on one of the receptor transcripts, as for both BCR and TCR the corresponding VH/VL or Vα/Vβ chains are expressed as separate transcripts from different chromosomes, respectively. However, there is enormous value in obtaining single-cell resolution as it provides essential information on natural pairing of variable chains, which enables deeper insights into the repertoire diversity and allows for validation of specificity with recombinant expression systems. In recent years, many innovative single-cell sequencing technologies based on droplet microfluidics, microwells or combinatorial indexing have emerged with varying levels of throughput (DeKosky et al. 2015; Stoeckius et al. 2017; Howie et al. 2015; Rosenberg et al. 2018).

The commercial 10X Genomics Chromium technology uses automated droplet-based microfluidics for a variety of genomic approaches most notably single-cell transcriptome sequencing (Zheng et al. 2017). Since the recent introduction of specific protocols and kits for targeted sequencing of BCR and TCR

Journal Pre-proof

variable regions (V(D)J libraries), it has become the most widely used platform for single-cell immune repertoire sequencing. Its key features are gel beads in emulsion (GEMs) that incorporate a gel bead together with a single cell, enzymes and poly(dT) primer. The GEM approach does not result in physical linkage of receptor chain transcripts, instead it relies on barcoding of cell derived templates, thus providing a digital linkage of sequences. 5' prime capture of full-length BCR/TCR transcripts ranging from 5'UTR to the constant regions are possible, thus allowing retrieval of variable region and isotype information of 100-10.000 cells per sample. Within one run, libraries of multiple samples can be prepared simultaneously with capture efficiencies of ~50-65%. Due to the fragmentation of amplicons, resulting V(D)J libraries can be short-read sequenced by Illumina platforms (e.g., 26 x 91 bp).

Using the 10X GEM system, Goldstein and colleagues performed single-cell sequencing from over 10^5 B cells from different organisms. They used antigen-baiting to label B cells from immunized animals, followed by flow cytometry enrichment and then 10X single-cell V(D)J sequencing. Based on the obtained information they recombinantly expressed a subset of the full-length VH-VL pairs and confirmed antigen-specific binding (Goldstein et al. 2019). This demonstrates the value of obtaining paired receptor data by high-throughput single-cell sequencing experiments in order to rapidly identify multiple antigen-binding antibodies. Along this line, Setliff et al. developed the method called 'LIBRA-seq' which uses DNA-barcoded antigens to match BCR sequences with antigen specificity by a sequencing-based readout. This approach led to the discovery of several HIV-1- and influenza-specific antibodies (Setliff et al. 2019). In addition to rapid antibody discovery, LIBRA-seq holds promise for vaccine and immunotherapy development as it allows screening of multiple antigens or epitopes at once. However, this strategy works only for B cells that have surface expression of BCRs, for cells that lack membrane-bound BCR (secreted antibodies) such as plasma cells, a recent droplet-based microfluidic platform 'CelliGO' was developed, which allows antigen-specific sorting and single-cell sequencing within one workflow (Gérard et al. 2020). Similar as for B cells, 10X Genomics can also be utilized for antigen-specific screening of T cells by the use of DNA-barcoded peptide-MHC multimers (Overall et al. 2020).

In addition to profiling immune repertoires from single cells, there is substantial efforts to also integrate single-cell transcriptome data, as this provides deeper insight into the dependencies between immune cell clonal diversity and specificity with phenotype and function (Saikia et al. 2019; Volden and Vollmers 2020; Stubbington et al. 2016; Croote et al. 2018; Afik et al. 2017). For example, Singh et al. combined targeted capture and long-read sequencing (Oxford Nanopore) of both BCR and TCR mRNA transcripts with short-read transcriptome sequencing (Illumina), all at a single-cell resolution (Singh et al. 2019). This method was used on tumor-infiltrating lymphocytes from a human breast cancer patient, revealing the repertoire and gene expression profile of clonally expanded lymphocytes. Horns et al. combined 10X immune repertoire and transcriptome sequencing with bulk VH repertoire sequencing to profile human B cell responses to influenza vaccination (Horns, Dekker, and Quake 2020). The impact and benefit of performing both single-cell immune repertoire and transcriptome analysis becomes in particular evident during the outbreak of SARS-CoV2. Several studies performed single-cell sequencing on PBMCs from Covid-19 patients and were able to both phenotypically characterize immune cells as well as extract

Journal Pre-proof

immune receptor sequences that correlated with infection (Cao et al. 2020; Liao et al. 2020; Wen et al. 2020).

Immunoinformatics: Decoding of immune receptor repertoires

In the era of 'big data', the use of specialized computational tools is crucial, in particular when interrogating the complexity of adaptive immunity. Immunoinformatics combines biological, computational and statistical approaches. Its applications make use of genomic, proteomic as well as structural data and range from the study of immune-related genes, receptor epitope predictions to in silico vaccination, among others (Zvyagin et al. 2020; Brown et al. 2019). While advances in antigen-specific single cell sorting enable identification of receptor sequence and specificity, it is still restricted in throughput when compared to the depth of data generated by bulk immune repertoire sequencing. In this section we will focus on immunoinformatic tools for decoding immune receptor repertoires obtained from immune repertoire sequencing studies (Figure 2). In particular, we describe computational methods that are used in the discovery and characterization of (potential) antigen-specific immune receptors providing valuable information for the design of immune therapies.

Data pre-processing and V(D)J assignment

Before performing in-depth immune repertoire sequencing analysis, the obtained raw data needs to be pre-processed in order to generate error-corrected sequences for proper analysis and interpretation. In short, pre-processing consists of removal of low quality sequences and sequence length trimming (e.g. primer sequence removal). For paired-end sequencing, reads will be merged by consensus building. Advanced library preparation protocols that use unique molecular identifiers (UMIs) can be applied for error correction (Shugay et al. 2014; Egorov et al. 2015; Turchaninova et al. 2016; Khan et al. 2016; Friedensohn et al. 2018; Ma et al. 2018; Ahmed et al. 2020), which is highly advantageous for obtaining accurate information on clonal diversity.

As each receptor sequence is a product of somatic recombination with possible N/ P nucleotide editing and somatic hypermutation (for B cells), each sequence needs to be assigned to its corresponding unmutated V(D)J germline segments. However, the high variability present in repertoire sequences can make assignments to germline a complicated task. Additionally, the number of allelic variants is large and still not fully known, hence, reference germline databases are only partially complete (Corcoran et al. 2016; Ohlin et al. 2019; Safonova and Pevzner 2019; Watson and Breden 2012). Proper V(D)J assignments strongly depend on the completeness of those databases as well as on the choice of annotation software packages to infer utilized V(D)J segments (Smakaj et al. 2019; López-Santibáñez-Jácome, Avendaño-Vázquez, and Flores-Jasso 2019). There is a variety of software tools that can be utilized for data pre-processing and/or V(D)J alignment: e.g., pRESTO and Change-O (Vander Heiden et al. 2014; Gupta et al. 2015), MiXCR (Bolotin et al. 2015), IMGT (Giudicelli, Chaume, and Lefranc 2004; Alamyar et al. 2012), IgBlast (Ye et al. 2013); for new germline allele identification: e.g. TIgGER (Gadala-Maria et al. 2015; 2019) and IgDiscover (Corcoran et al. 2016)).

Clonal analysis

Journal Pre-proof

Clonal selection and expansion are foundational principles of adaptive immunity, where the biological definition of clones is a subset of cells that originated from a common ancestor (e.g., naive B cell or T cell). Therefore, the bioinformatic definition of clones in immune repertoire sequencing data is crucial for downstream analysis. The clonal grouping or clustering approach, also referred to as clonotyping, is usually based on the more diverse VH for BCR and Vβ chain for TCR. Special attention is given to the junction of V(D)J recombination: the CDR3, which is the most variable region and decisive for antigen recognition and binding, thus the VH/Vβ CDR3 are preferentially used to identify BCR and TCR clones. For T cells, all identical TCR sequences resemble a unique clone, for B cells, however, inferring clones and their clonal lineage is non-trivial as clonal members do not comprise identical V(D)J sequences due to somatic hypermutation (Hershberg and Luning Prak 2015). The most common clonotyping strategy for the identification of clonally related B cells utilizes a distance-based approach via hierarchical clustering (Yaari and Kleinstein 2015; Greiff, Miho, et al. 2015). Briefly, BCR sequences with the same annotated V- and J-gene germline segments and same junction length are grouped together and those sequences that possess a junction similarity above a set threshold (e.g. 80-90%) are defined as originating from the same clone (same can be done for the LC as well). However, setting a fixed distance cut-off for clonal definition is far from optimal as it cannot account for different levels of clonal diversification within a repertoire. Therefore, much effort is put into more advanced and alternative strategies that make use of different grouping criteria or thresholds to infer clones, including ones based on probabilistic models or spectral clustering with adaptive thresholds (Ralph and Matsen IV 2016; Nouri and Kleinstein 2018; 2020). Just recently, Lindenbaum and colleagues introduced an alignment-free method bypassing initial V- and J-gene assignments (Lindenbaum et al. 2020).

Diversity profiles

Upon antigen recognition and activation, certain subsets of B and T cells undergo extensive proliferation (and affinity maturation in the case for B cells). This induces temporal changes in immune repertoire dynamics and composition which can be captured by diversity profiles. Such quantitative analysis on a global repertoire level is of interest when investigating vaccination response or discriminating between health and disease. Diversity measurements for immune repertoires are adapted from mathematical ecology which are used to assess the diversities of ecosystems. Commonly used diversity measures include species richness, Shannon's entropy and Simpson's index, among others, which differ in their weighting of rare clones (Chaudhary and Wesemann 2018). However, they are sensitive to sampling depth and results can vary depending on which diversity index is used. In order to visualize full clone size distribution, Hill-based diversity profiles which include a continuum of single diversity indices (including the above-mentioned ones) can be used (Greiff, Bhat, et al. 2015; Rosenfeld, Meng, Luning Prak, et al. 2018).

Clonal Evolution

Since B cells can undergo somatic hypermutation upon antigen recognition, a B cell lineage is generated which consists of BCR sequence variants that share the same specificity but with different affinities. Phylogenetic trees are often used to reconstruct both ancestral and intermediate relationships between B cell clonal sequences, thereby enabling tracking of clonal evolution, which can be in particular indicative for antigen-specificity (Yermanos et al. 2018; Hoehn et al. 2016; 2019) (note: lineage tree construction

Journal Pre-proof

might be integrated into the clonotyping step). Commonly used phylogenetic methods range from maximum parsimony and maximum likelihood to Bayesian analysis. Using maximum parsimony, phylogenetic trees are reconstructed by assuming the least amount of mutations, whereas the other methodologies include substitution models or probabilistic priors which can incorporate certain features thought to impact sequence evolution such as mutation rates (Yermanos et al. 2017; Ralph and Matsen IV 2016; Hoehn, Lunter, and Pybus 2017). Under the assumption that B cell affinity maturation drives improved affinity, study of the tree topology may reveal clonal selection and allow the identification of clonal sequences or features correlating with increased affinities towards the antigen of interest. Reconstruction of the evolutionary path of antigen-reactive B cells is of particular use in the field of rational vaccine design e.g. for HIV research, in which increased somatic hypermutation levels are known to be important features of broadly neutralizing antibodies (bNAbs). Extensive studies led to the discovery of a panel of HIV Env-specific bNAbs, these B cells evolved in different patients after years of chronic infection but exhibited common sequence elements depending on the epitope being recognized (Wu et al. 2015; Zhu et al. 2013). Thus, efforts are undertaken to rationally design immunogens which specifically target germline-precursor B cells to induce their maturation towards bNAb secreting cells - a promising approach referred to as ‘germline targeting’ (Sok et al. 2016; Briney et al. 2016; Jardine et al. 2016; Steichen et al. 2019).

Network analysis

Network-based analysis offers another possibility to infer clonal relationships and their organization within immune repertoires based on sequence-similarity. In an immune repertoire network, nodes represent unique clones (CDR3 or full V(D)J sequences) which are connected by edges depending on their sequence similarity. Sequence distance matrix calculations are usually based on Levenshtein or hamming distance (e.g. with a Levenshtein distance cut-off of 1, only those nodes are connected that differ at most by one nucleotide or amino acid). This approach allows visualization and detection of clonal clusters that build up a unique repertoire architecture. For quantitative analysis, metrics describing each repertoire architecture on a global as well as on a local level can be deployed. For instance, measuring the connectivity degree distribution, meaning quantifying the number of connections each node exhibits, provides insights into the clonal expansion profile of the repertoire in question. Dissecting clonal clusters that exhibit high connectivity (intraclonal diversity) as well as integrating clonal frequency information, may thus offer hints for antigen-induced expansion of specific clones (Bashford-Rogers et al. 2013; Miho et al. 2018). Of note, network-based analysis allowed the characterization of public clones showing that they tend to be highly connected and associated with maintaining the repertoire architecture. Madi and others performed a network-based analysis of both murine and human immune repertoires and observed that public clones were found to be among the most connected nodes, possibly providing a common basis for functional immunity (Madi et al. 2017; Miho et al. 2019). Albeit the emergence and functionality of public clones are not yet fully understood, there is increasing evidence for their role in neutralization of common pathogenic antigens, autoimmune diseases and cross-reactivity (Madi et al. 2014; Greiff, Weber, et al. 2017; Zhao et al. 2016; Hershberg and Luning Prak 2015; Khosravi-Maharlooei et al. 2019). Recently, public T cell clones have been shown to recognize peptides on multiple HLAs which makes them a

Journal Pre-proof

promising target for immunotherapy (Galperin et al. 2018). Furthermore, the identification of public clones is of particular interest for population-wide targeting in the context of vaccine design (Setliff et al. 2018).

Clonal convergence and identification of antigen-specific patterns

The occurrence of identical or similar shared immune receptor sequences within or across different individuals in response to an antigenic challenge hints towards the concept of antigen-associated clonal convergence (Parameswaran et al. 2013; Trück et al. 2015; Emerson et al. 2017; Pogorelyy et al. 2018; Ehrhardt et al. 2019; Akbar et al. 2019). One straightforward approach to identify clones that specifically arose upon an antigenic challenge is to perform overlap analysis, which can be visualized with Venn diagrams. Frequency information can also be included for pairwise quantitative comparisons by applying metrics such as the Morisita-Horn overlap index or the cosine-similarity metric (Rosenfeld et al. 2018; Meng et al. 2017). However, the enormous immune receptor diversity makes antigen-driven public clones - based on exact immune receptor sequence - a statistically unlikely event. Thus, increasing efforts are being undertaken to develop advanced clustering tools to detect sequence patterns which hint towards antigen-driven repertoire convergence (De Neuter et al. 2018; Pogorelyy et al. 2019; Meysman et al. 2019; Zhang et al. 2020).

Two seminal approaches performed by Dash et al. and Glanville et al. utilized immune repertoire sequencing of antigen-specific single T cells for the detection of sequence motif patterns that were associated with a specific antigen (Dash 2017, Glanville 2017). Dash and colleagues utilized a CDR-weighted distance metric to measure similarity between any TCR sequences (TCRdist) (Dash et al. 2017). Glanville and colleagues developed an algorithm called GLIPH for identifying TCR specificity groups, where all TCRs in a group recognize identical or highly similar pMHC ligands. Also here, conserved motifs as well as global similarity between CDRs (in particular CDR3s) are used in the clustering approach. GLIPH can be used directly on TCR repertoires without prior knowledge of antigen specificities, to identify TCRs with shared specificity as well as their HLA restriction (Glanville et al. 2017; Huang et al. 2020). Using the insights on antigen-specific TCR sequence motifs, both groups developed a classification system with which they could predict antigen-specificity with high accuracy.

Along this line, increasing efforts were put into the development of novel bioinformatic, statistical and machine learning models; these include hidden Markov models (HMMs), support vector machines (SVMs), decision trees, or artificial neural networks (ANNs) (Mason et al. 2019; Greiff, Weber, et al. 2017; Widrich et al. 2020; Konishi et al. 2019; Sidhom et al. 2018; Fischer et al. 2019). For example, immune receptor sequences identified through antigen-specific cell sorting can be grouped into antigen-specific binders and non-binders. All sequences then get encoded as numerical representations by methods such as one-hot encoding, k-mer decomposition or vector embedding. Utilizing classical machine learning or more advanced deep learning models, one can then train and fine-tune the models using the embedded sequences for which specificity is already known in such a way that the models are later able to classify unseen input sequences into binder and non-binder. Of note, due to the increased usage of these novel machine- and deep learning strategies, models for in silico simulation of immune repertoire are highly valuable for a better understanding, training and evaluation of such complex computational tools (Weber

Journal Pre-proof

et al. 2020; Yermanos et al. 2017; Marcou, Mora, and Walczak 2018; Sethna et al. 2019; Safonova, Lapidus, and Lill 2015).

Just recently, Friedensohn and colleagues developed an unsupervised deep learning approach for the identification of convergent antibody sequence clusters from bulk BCR repertoires (not antigen-sorted) (Friedensohn et al. 2020). Briefly, they performed immune repertoire sequencing of antibody VH chains from four different cohorts of immunized mice. The resulting sequences were used as input and training of a variational autoencoder (VAE) combined with a Gaussian mixture model (GMM), which clustered the encoded sequences in a lower dimensional latent space. Cluster membership of novel, unseen sequences could then be used to predict their antigen specificity. Moreover, cluster-derived sequences as well as in-silico generated sequences were recombinantly expressed and shown to be specific for the corresponding antigen.

Progress in the field of deep learning approaches has promising implications for rational guided immune receptor design. As demonstrated in the study of Glanville et al. and Friedensohn et al., the identification of antigen-specific motifs or clusters can be leveraged for in-silico generation of de novo antigen-specific variants (Glanville et al. 2017; Friedensohn et al. 2020; Davidsen et al. 2019; Amimeur et al. 2020). Thus, different deep learning models can be trained (e.g. with sequences derived from large combinatorial mutagenesis libraries) and used to identify antigen-specific sequences with improved features such as optimized affinity or developability parameters out of both natural or in silico generated immune receptor sequences. This has promising implications for time and cost-effective immune receptor engineering and optimization approaches (Mason et al. 2019; Chen et al. 2020; Liu et al. 2020).

Immune receptor engineering with synthetic display platforms

One of the most widely utilized synthetic platforms is phage display, which relies on genetic encoding and expression of the variable regions of an immune receptor (as a fusion to viral capsid protein), thus providing a phenotype-genotype linkage (Smith 1985). The displayed variable domains when derived from a BCR/antibody can either be in an scFV format in which the VL and VH domains are connected via a peptide linker or in a fragment antigen-binding (Fab) format, which consists of the variable and the first constant domain of each light and heavy chain. In the context of a TCR, direct display is challenging due to low stability and expression of a soluble TCR, thus the most common format uses disulfide-bond linked TCRs (dsTCRs) in which additional disulfide bonds increase the stability (Boulter et al. 2003) or the scTCR format in which the alpha and beta chain are connected via a peptide linker. One of the main advantages of phage display over other synthetic platforms is the ability to create very large recombinant libraries of up to 1×10^{11} variants. These libraries are constructed either from natural immune repertoires (e.g., infected patient or immunized animals) or synthetic repertoires via different mutagenesis and DNA synthesis approaches. Phage libraries are screened via biopanning which involves the addition of the library to immobilized antigen, washing out unbound phage variants and eluting the bound phage. The bound phages are then amplified and subjected to multiple rounds of panning to enrich for antigen-specific phage. The introduction of immune repertoire sequencing in the phage display screening process gives valuable insights, guides the selection process and improves library design (Barreto et al. 2019; Dias-Neto

Journal Pre-proof

et al. 2009; Glanville et al. 2015; Ravn et al. 2010). Deep sequencing of panning rounds has uncovered that many high affinity binding clones become depleted due to fitness cost and that enriched clones identified can exhibit higher antigen selectivity (Ravn et al. 2010; Saggy et al. 2012; Wang et al. 2010). To recover specific clones, immune repertoire sequencing data can be used to design primers that bind to the CDRH3 region of a specific clone and amplify it via an inverse PCR approach (Spiliotopoulos et al. 2015).

Another powerful platform is yeast display, where variable regions of immune receptors are fused to yeast surface protein (Boder and Wittrup 1997). Similar to phage display, an antibody or TCR can be displayed on yeast in scFv, Fab or scTCR formats (Sivelle et al. 2018; Kieke et al. 1999; Holler, Chlewicki, and Kranz 2003; Smith, Harris, and Kranz 2015). Yeast display libraries can be screened with a significant advantage of utilizing fluorescence-associated cell sorting (FACS), which provides real-time assessment and quantitative characterization of the binding properties of receptor variants. Compared to phage the maximum library size of yeast is smaller with $1 \times 10^{9\text{-}10}$ variants (Benatuil et al. 2010), but due to efficient homologous recombination in yeast there is no need for cloning in vitro which simplifies the construction of libraries (Ma et al. 1987). Another advantage of yeast is the presence of protein folding control, posttranslational modifications and glycosylation resulting in immune receptors with more stable properties, which is particularly important for antibody drug development (Boder, Raeeszadeh-Sarmazdeh, and Price 2012).

Immune receptor engineering with synthetic immune cell platforms

Although phage and yeast display are able to express immune receptor fragments, they both still lack important features present in the naive expression of immune receptors by B and T cells. For example, certain receptor frameworks are difficult to express in bacteria or yeast cells and are therefore lost. Furthermore, expression as in the truncated scFv or Fab format does not always successfully translate into the IgG format, thus slowing or preventing the development of therapeutic antibodies. Furthermore, scTCR binding does not necessarily point to eventual TCR signaling. While it is fast and inexpensive to recombinantly express proteins in bacteria or yeast using standard DNA plasmid systems, mammalian cells do not replicate plasmids. In order to use repeated cycles of mutagenesis and selection rounds akin to yeast and phage, advanced genome editing tools are necessary. With the advent and rapid advancement of CRISPR-based methods, mammalian display has entered an exciting new stage with opportunities for engineering immune receptors. In the following sections, we will briefly explain CRISPR systems and elaborate how these can be applied to mammalian display platforms to engineer specificity, selectivity and functionality of immune receptors.

CRISPR-based genome editing

Since its adaptation to mammalian cells in 2013 (Cong et al. 2013), the CRISPR/Cas9 system has dominated the field of genome editing, due to its versatility, simplicity and efficiency. CRISPR/Cas9 was discovered in bacteria as a form of adaptive immunity against phage viruses. As a genome editing tool, the system consists of only two main parts: the Cas9 nuclease enzyme and the guide RNA (gRNA). Beyond Cas9, many related CRISPR systems have been discovered, repurposed and utilized for various applications, including immune cell manipulations of both RNA and DNA. While natural DNA repair mechanisms are leveraged by designer nucleases, novel approaches rely on targeting other effectors to

Journal Pre-proof

specific regions of the genome, such as nucleotide-converting enzymes (Base editors), or reverse transcriptases (Prime editing) to directly synthesize a desired edit into a DNA nick. Despite their utility in other fields (Webber et al. 2019), these are less suited to cause profound changes in immune cells, since they can only be used to modify a small region within the genome (Marzec, Brąszewska-Zalewska, and Hensel 2020).

CRISPR systems are still in their infancy in regards to medical translatability, not yet reaching the more extensively used, classical genome modification tools (e.g., retroviruses, lentiviruses, adeno-associated viruses and zinc-finger nucleases). In a clinical context, only Cas9 has been tested so far and concluded to be safe for use in treating cancer patients with CRISPR-modified T cells expressing chimeric antigen receptors (CARs) for cancer treatment, which showed promising results related to efficacy and safety (Stadtmauer et al. 2020).

Antibody screening with mammalian display

One of the first mammalian display approaches was based on HEK-293T cells transiently transfected with plasmids encoding an scFv library. Since the scFv was fused to the transmembrane domain of human platelet-derived growth factor receptor (PDGFR) it was displayed on the cell surface (Ho, Nagata, and Pastan 2006). Similar to yeast display, the mammalian cell library was screened via flow-cytometry. One drawback of this transient transfection approach is that the number of plasmids per cell is variable and therefore results in the expression of multiple scFv variants per single cell as well as expression level fluctuations due to different plasmid copy numbers. Furthermore, the expression is temporary which prevents multiple rounds of enrichment or downstream assays. One approach to overcome transient expression consists of using the sindbis virus to stably integrate scFv in the genome of BAK cells, which enabled the long-term display of scFv on the surface of mammalian cells (Beerli et al. 2008). However, since the sindbis virus randomly integrates in the genome, the problem of multiple variants and copies per cell was not solved. To address this challenge, a genome integration system was developed based on the Flp recombinase (Flp-In cell line developed by Invitrogen), which enabled site-specific single copy integration and through fusion with the transmembrane domain of PDGFR resulted in the display of full IgG on the surface of CHO cells (Zhou et al. 2010). Since therapeutic antibodies are administered in the secreted format and many functional and developability assays require soluble antibodies, several approaches have been developed to enable display but also secretion of antibodies in the same platform. One approach used Cre recombination sites flanking the transmembrane domain, which facilitated Cre-mediated deletion of the transmembrane domain and therefore switching from antibody display to antibody secretion (Tomimatsu et al. 2013). Another approach used the RIRR sequence that can be recognized and partially cleaved by the furin enzyme, leading to simultaneous secretion and display of antibodies (Zhou et al. 2013). Yet another approach is based on alternative splicing, thus allowing titration of the secretion to display ratio (Aebischer-Gumy et al. 2020; Horlick et al. 2013). More recently, several novel mammalian display platforms have been developed by taking advantage of CRISPR systems (Figure 3). For example, Pogson and Parola et al. used Cas9 and homology-directed repair (HDR) to exchange antibody sequences at the endogenous immunoglobulin locus of a hybridoma cell line, thus enabling endogenous promoter driven expression, display as well as secretion of full-length IgG (Pogson et al. 2016; Parola et al. 2019). Furthermore Mason et al. established the constitutive expression of Cas9 in this

Journal Pre-proof

cell line and demonstrated that Cas9 HDR with degenerate oligonucleotide templates could be used to generate mutagenesis libraries into antibody variable regions with high efficiency (Mason et al. 2018).

T cell platforms for functional TCR engineering

Functional screening platforms for interrogation of engineered TCRs outside of the primary cell context are useful tools to approach challenges related to safety, specificity, developability and efficacy of T cell therapies. Despite a plethora of reports on the utilization of CRISPR in immortalized Jurkat T cell lines(Chi, Weiss, and Wang 2016; Simeonov et al. 2017; Bray et al. 2018; Borowicz et al. 2020; Chan et al. 2020), functional engineering of TCR in this context has received little attention.

The first efforts in the field of testing TCRs in Jurkats was accomplished by equipping TCR-negative Jurkat T cells with a tumor specific TCR through retroviral integration. Activation of the TCR was measured through an NFAT-Luciferase reporter (Aarnoudse et al. 2002), however the TCRs either did not activate or activated only weakly. CD8 was identified as an important driver of T cell activation (Dembić et al. 1987) and the lack thereof may explain why Jurkats (which are CD4+) were inadequate to simulate T cell activation for cancer recognition. Aarnoudse and colleagues consequently generated CD8+ Jurkats through retroviral transduction. Non-endogenous behavior meant that modification of Jurkats was necessary to transform them into a viable platform for TCR engineering. Because of this, activation in a cellular context has been subpar in evaluating TCR functionality for TCR engineering. Fast and inexpensive functional screening of transiently expressed TCRs in CD8+ and CD4+ Jurkats has been established, but the transient nature prevented further engineering approaches and selections. Current research in the field has been overwhelmingly focused on affinity-based selections of TCRs (Kessels et al. 2001; Chervin et al. 2008; Malecek et al. 2013; Schmitt et al. 2017; Wagner et al. 2019; Spindler et al. 2020). As highlighted in previous sections, affinity maturation of scTCR-pMHC pairs has been done by phage and yeast display screening. However, these display technologies exhibit rather poor scTCR expression and stability and most importantly they are unable to interrogate the natural physiological context of TCR functional specificity: antigen-induced activation of TCR signaling pathways. Modifications to the TCR structure have been made recently, enabling improved expression and folding (Gunnarsen et al. 2018; Froning et al. 2020).

Current approaches for TCR engineering heavily rely on affinity maturation to improve engineered T cell proliferation, cytokine secretion and cytotoxicity. This is based on the notion that the affinity and function of T-cell mediated responses are correlated (Zehn, Lee, and Bevan 2009; Schmid et al. 2010). However, despite this common assumption, several reports on modeling and screening TCR-pMHC interactions suggest that more research is necessary (Gálvez, Gálvez, and García-Peñarrubia 2019; Birnbaum et al. 2014; Sibener et al., 2018). Affinity maturation, however, seems necessary to engineer soluble TCRs or cell therapies since naturally-occurring TCRs are often too low affinity to be effective.

Unfortunately, TCRs engineered for higher affinity binding to peptide-MHC, as well as TCRs derived from transgenic mice have been linked to patient deaths in cell therapy clinical trials. Currently on-going clinical trials have raised further safety concerns (NCT04044768, NCT02592577). These outcomes are related to

Journal Pre-proof

the fact that TCRs can possess a broad, cross-reactive specificity profile, in some cases recognizing thousands of different peptide off-targets (Bentzen et al. 2018). In one clinical trial, a TCR derived from an immunized transgenic mouse targeting the tumor-associated antigen MAGE-A3 peptide was discovered to be cross-reactive towards a similar peptide, (MAGE-A12), displayed in the brain, leading to severe neurotoxicity including patient death (Morgan et al. 2010). Another TCR targeting MAGE-A3, this time derived from an immunized patient and affinity matured through phage display, was used in a clinical trial, also leading to patient death. Retrospectively it was discovered that this engineered TCR possessed cross-reactivity to titin-peptide expressed on cardiomyocytes and was the source of the observed severe cardiotoxicity. Engineering by phage display actually introduced mutations that led to cross-reactivity with titin (Cameron et al. 2013).

A number of reports demonstrate the use of mammalian cells for TCR screening and engineering, most of which involved viral transformations and affinity-based selections (Kessels et al. 2001; Chervin et al. 2008; Malecek et al. 2013; Schmitt et al. 2017; Wagner et al. 2019; Spindler et al. 2020) In order to overcome the challenges related to TCR affinity and cross reactivity, it is necessary to engineer TCR specificity not only on the basis of binding and function (Rosskopf et al. 2018), but also while detecting cross-reactivity. To that end, the recently established TCR-accepting T cell (TnT) platform was developed through several CRISPR-editing steps in the Jurkat T cell line (Vazquez-Lombardi et al. 2020). Combining CRISPR-targeting, functional selections and immune repertoire sequencing, it was possible to profile mutational landscapes of tumor specific TCRs, as well as accurately predicting off-targets that led to patient deaths in engineered TCR clinical trials. Additionally, the authors showed that TnT cells could be used to engineer synthetic TCR variants with enhanced affinity to the MAGE-A3 cancer antigen with no detectable cross-reactivity, representing promising and safe clinical candidates for TCR cell therapy.

Engineering chimeric antigen receptors (CARs)

CARs and their antigen-binding domains are often discovered or engineered using phage and yeast display, resulting in high affinity scFvs. Similar to high affinity TCRs, these high-affinity CAR T cells may exhibit significant toxicities due to on-target off-tumor activity. Therefore, understanding the relationship between affinity, selectivity and CAR activation is crucial to creating safer and more efficacious treatments. In a recent report, reducing the affinity and increasing the valency/avidity of an scFv enhanced the ability of CAR T cells to distinguish between normal and tumor cells (Slaga et al. 2018). It is also possible to quickly evaluate new CAR constructs in a mammalian reporter cell line by relying on an NFAT and NFkB dual-fluorescent reporter system upon stimulation (Rydzek et al. 2019). This cell line was generated using viral transductions and transposon-mediated integration of an scFv library targeting the tumor-associated antigen ROR1, which was screened to identify scFv variants that showed improved signalling. However, due to difficulties with multiple, transposase-mediated genomic insertions, discerning the effect of single variants was challenging. Recently, a CRISPR-mediated CAR T cell reporter cell line was generated (Di Roberto et al. 2020). This platform was used to identify CAR variants that were highly active and selective for high antigen expression levels. Site-saturation mutagenesis libraries of the scFv were screened, followed by immune repertoire sequencing. Selection for both signalling and binding guided the isolation of specific variants with enhanced antigen level discrimination. Interestingly, the

Journal Pre-proof

authors report a striking divergence between affinity and CAR T cell responsiveness profiles, indicating again the importance of functional screening.

Engineering synthetic immunity

Active and passive immunizations (e.g., antibody serum therapy) have been clinically used for over a century. These established therapeutics have eradicated a large number of pathogenic diseases and continue to be relevant for novel diseases (e.g., Covid-19). However, some pathogens and afflictions continue to elude natural immunity, such as HIV and cancer. In some cases, it may be necessary to move past active and passive immunization and combine both into a novel concept of synthetic immunity, in which adaptive immune responses are engineered on a cellular level. Recent success in treating hematological cancers due to the approval of CAR T cell therapies are testament to this. In the following sections, we will discuss the theoretical basis and progress to date in respect to engineering adaptive immune responses in T cells and B cells by genome editing (Figure 4).

CRISPR-mediated T cell engineering

Adoptive T cell transfers to cure cancer are not a recent idea. Tumor infiltrating lymphocytes (TILs) were extracted, expanded and used in the 1980s, achieving durable remissions in some patients. A plethora of tumor evasion mechanisms such as downregulation of MHC or the secretion of immune modulators, as well as the complex manufacturing negatively impacted the success of adoptively transferred TILs. In the late 1980s, new tools were developed that facilitated the manipulation of T cells, driven by advances in DNA reading, writing and editing, thus, enabling the delivery of exogenous DNA and its detection in immune cells. Despite recent clinical successes in treating hematological tumors with engineered CAR T cells, considerable limitations and safety concerns persist. In the following sections, we will highlight the major developments in engineering T cells as potent immunotherapies.

As already highlighted, CRISPR is still at a very early stage regarding clinical translation, as there are currently no approved CRISPR-based cell therapies approved by the FDA or EMA but there are a number of ongoing clinical trials (clinicaltrials.gov, NCT03752541, NCT04026100, NCT03166878, NCT03229876). CAR T cell responses against solid malignancies sometimes suffer from poor durability and persistence. However, this may be related to genome editing methods that rely on viral vectors for the random integration of transgenes, which can result in oncogenic transformation, unstable expression levels and silencing (Maude et al. 2014; Turtle et al., 2016). Recent advances in targeted genome editing into endogenous genomic loci offer the potential to alleviate some of these shortcomings. For example, it was demonstrated that targeted genomic integration of CAR genes into the endogenous TCR alpha chain constant region (TRAC) locus resulted in more uniform expression and increased potency of CAR T cells, when compared to random integration by retrovirus (Eyquem et al. 2017). Furthermore, various CAR constructs were tested in respect to their influence on T cell phenotypes, including differentiation potential, exhaustion, activation and memory. Disrupting the endogenous TCR through CRISPR-mediated knockout was reported to impair long-term persistence in vivo compared to CAR T cells with endogenous TCR, despite the advantage of avoiding potential allo-reactivity of engineered T cells towards the host (Stenger et al. 2020). It is yet unclear if targeting CARs to TCR loci suffers from similar limitations. It is conceivable,

Journal Pre-proof
however, that the lack of TCR expression itself results in this reduced long-term persistence. Lentiviral knock-in of engineered TCRs with CRISPR-mediated knockout of endogenous TCRα and β genes and the PDCD1 gene has recently been reported in a first in human phase I clinical trial, with results suggesting safety and durability of treatment (Stadtmauer et al. 2020)

Adeno-associated viruses (AAVs) are exceptional at transducing haematopoietic cells and are able to deliver large quantities of DNA, thus providing a source for HDR templates for CRISPR-mediated transgene integration. In a recent study, a full TCRα and β variable domain was integrated into the TRAC locus using nucleofected PCR-amplified double-stranded DNA (dsDNA) and Cas9-Ribonucleoprotein (Roth et al. 2018) as well as parallel knockout of endogenous TCRα and β genes to avoid mispairing (Schober et al. 2019). In addition to targeted integration, there are a number of clinical trials investigating concurrent knockout of immunomodulatory genes, such as PDCD1 (NCT03399448, NCR03545815). In order to achieve efficient on-target integration as well as multiplexed knockout of such genes, novel CRISPR systems have been used. In a single step, anti-CD22/19 CAR T cells with PDCD1 knockout were generated using the CRISPR-Cas12a system, all encoded by a singular knock-in knock-out (KIKO) vector, containing the crRNAs as well as the DNA template. Double knock-in of anti-CD19 and anti-CD22 CARs into primary human T cells was also achieved at high rates of up to ~75%, while Cas9 edits were unable to achieve the same result (up to ~5%) (Dai et al. 2019). Double knock-in of large DNA constructs as well as multiplexed knockout of specific genes may be beneficial for treating complex cancers (e.g. in solid tumors with inhibitory tumor microenvironments). An alternative to gene ablation by concomitant knockout is reported in another study by Roth and colleagues, in which a pool of gene constructs that potentially enhance engineered T cell function is transfected alongside a TCR of defined specificity. Aforementioned advances in non-viral DNA delivery enable the integration of large DNA fragments, and thus, multiple genes. Here, the authors delivered a barcoded 36-member therapeutic knock-in library into the TRAC locus. T cell behavior in terms of proliferation and other functional parameters were assessed including experiments in which enriched T cells were extracted in a solid human tumor xenograft mouse model (Roth et al. 2020).

CRISPR-mediated B cell engineering

Engineering B cells to express precisely defined monoclonal antibodies against pathogens and disease would represent a powerful way to provide synthetic immunity. In contrast to T cell genome editing, B cell genome editing has only very recently started to emerge as a possible immunotherapy strategy. This is largely because most cellular therapy efforts have been focused on cancer indications, where T cells play a dominant role for therapeutic intervention. Additionally, genomic modification of B cells has been much more difficult to achieve.

One of the first reports of editing primary human B cells attempted to knock-in targeted nucleotides with co-delivery of single-stranded DNA (ssDNA) oligonucleotides and electroporation of Cas9 RNP types (Wu et al. 2018). Despite achieving efficiencies of 10% or more HDR, integration was limited to a few base-pairs. The authors also reported that insertions or deletions occurred much more frequently than what has been observed in other cells, pointing towards differences in the behavior of the DNA repair machinery in B cells compared to T cells. Generating larger knock-in seemingly required other tools and vectors. Extending the ex vivo culture as well as using AAV6 to deliver an HDR template, site-specific integration

Journal Pre-proof
rates of up to 25% (Johnson et al. 2018) and 40% (Hung et al. 2018) were achieved in primary human B cells. Johnson and colleagues also found that Cas9 mRNA delivery instead of RNPs may be detrimental to activation and growth regulation, possibly due to cytosolic RNA sensors (Johnson et al. 2018). With this groundwork completed, many research groups attempted to achieve the same goal in both murine and human B cells: introducing a transgene coding for a BCR variable region of defined specificity into the endogenous VH locus. In contrast to T cell therapies that were tested in immunodeficient mice, B cell therapies require a functioning immune system to assess homing to secondary lymphoid organs and additional stimulation to mature, differentiate, undergo somatic hypermutation and class-switch recombination. The first study on integrating predefined BCRs into polyclonal murine B cells using Cas9 RNP and long ssDNA into the endogenous locus relied on targeting the intron downstream of the recombined VDJ_H to be able to use it partially as a 5' homology arm, thus resulting in integration of bNAbs against HIV (Hartweger et al. 2019). Concomitant CRISPR-mediated knock-out of endogenous murine Cκ prevented mispairing of the introduced antibody. Despite exciting and promising data, the study did not conclusively show immunization dependent recruitment of engineered B cells, as well as antigen-dependent clonal expansion, the primary benchmark for creating an adoptive immunization. Further studies confirmed the inability of stimulated engineered cells to form a functional B cell memory response, despite the presence of memory phenotype markers which is likely an artifact from ex vivo culture. A study in primary human B cells equally attempted to integrate HIV bnAbs into the endogenous locus, using a different strategy (Voss et al. 2019). Here, the authors deleted approximately 1 megabase from the V gene locus while introducing an HDR template with homology arms 5' and 3' of the deletion in order to be able to target a polyclonal mixture of B cells. While integration was achieved, editing efficiencies for this strategy remained extremely low at <0.1%. While not directly integrating antibodies, another study attempted to engineer murine B cells to express chimeric BCRs to sense antigen and in response secrete a protein of interest (Pesch et al. 2019). Similar to other studies, it was found that HDR editing rates using the conventional B cell ex vivo culture protocol and dsDNA transfection was very low at less than 1%. By optimizing murine B cell culture and transfection, Moffett and colleagues achieved higher editing rates of ~4% using dsDNA and ~16% with AAV6, as well as co-expression of separate antibodies from both antibody loci, representing an important milestone in B cell editing (Moffett et al. 2019). Engineered B cells were transferred to immunocompromised $RAG1^{-/-}$ mice and while they persisted in vivo for an extended period and formed protective serum titers, the engineered cells did not actively respond to infection in a memory or plasma cell like manner. This is probably due to the B cell phenotype derived from ex vivo culture as well as the lack of T cell help in $RAG^{-/-}$ mice. In a recent study, Nahmad and colleagues achieved high HDR rates using AAV-DJ/8 to deliver the HDR templates and engineered HIV-specific CD45.1 B cells, which were then transferred to immunocompetent CD45.2 mice (Nahmad et al. 2020). These engineered B cells were traced in regards to persistence, indicating successful clonal expansion in response to antigen. A synonymously re-coded anti-HIV bnAb containing AID-hotspots was introduced to assess somatic hypermutation. While some mutations may be attributed to AAV preparation and B cell engineering, patterns in enrichment of mutations that follow transfer and immunization were observed, implying an in vivo selection. Shortly after, another similar study used LPS to expand primary murine B cells, reproducing the results observed by Nahmad and colleagues, i.e. demonstrating expansion of engineered memory B cells and long-lived plasma cells, as well as somatic hypermutation (Huang et al.

Journal Pre-proof

2020). This exciting progress is paving the way for B cell genome editing as a feasible translational strategy for synthetic immunity.

CONCLUSION

Manipulation of the immune system in order to protect from infectious diseases has been performed for several centuries, with one of the first historical examples being the smallpox vaccine invented by Edward Jenner in 1796, which is considered the official founding of the field of immunology. Since then, the means used to manipulate immunity have changed profoundly catalyzed by a greater understanding of the molecular and cellular intricacies that govern immune function. As extensively reviewed here, the field of immunology is increasingly relying on novel methods in immune repertoire sequencing, immunoinformatics and immunogenomic engineering, the result of this is the emergence of the transformative new field of systems and synthetic immunology, which offers the possibility to answer long-standing fundamental questions and to reprogram and engineer immunity.

References

Aarnoudse, Corlien A., Margreet Krüse, Renate Konopitzky, Nathalie Brouwenstijn, and Peter I. Schrier. 2002. “TCR Reconstitution in Jurkat Reporter Cells Facilitates the Identification of Novel Tumor Antigens by CDNA Expression Cloning.” *International Journal of Cancer* 99 (1): 7–13. https://doi.org/10.1002/ijc.10317.

Aebischer-Gumy, Christel, Pierre Moretti, Romain Ollier, Christelle Ries Fecourt, François Rousseau, and Martin Bertschinger. 2020. “SPLICELECT^{TM}: An Adaptable Cell Surface Display Technology Based on Alternative Splicing Allowing the Qualitative and Quantitative Prediction of Secreted Product at a Single-Cell Level.” *MAbs* 12 (1): 1709333. https://doi.org/10.1080/19420862.2019.1709333.

Afik, Shaked, Kathleen B. Yates, Kevin Bi, Samuel Darko, Jernej Godec, Ulrike Gerdemann, Leo Swadling, et al. 2017. “Targeted Reconstruction of T Cell Receptor Sequence from Single Cell RNA-Seq Links CDR3 Length to T Cell Differentiation State.” *Nucleic Acids Research* 45 (16): e148–e148. https://doi.org/10.1093/nar/gkx615.

Ahmed, Ibrahim, Felicia A. Tucci, Aure Aflalo, Kenneth G. C. Smith, and Rachael J. M. Bashford-Rogers. 2020. “Ultrasensitive Amplicon Barcoding for Next-Generation Sequencing Facilitating Sequence Error and Amplification-Bias Correction.” *Scientific Reports* 10 (1): 1–10. https://doi.org/10.1038/s41598-020-67290-1.

Akbar, Rahmad, Philippe A. Robert, Milena Pavlović, Jeliazko R. Jeliazkov, Igor Snapkov, Andrei Slabodkin, Cédric R. Weber, et al. 2019. “A Compact Vocabulary of Paratope-Epitope Interactions Enables Predictability of Antibody-Antigen Binding.” bioRxiv. https://doi.org/10.1101/759498.

Alamyar, Eltaf, Patrice Duroux, Marie-Paule Lefranc, and Véronique Giudicelli. 2012. “IMGT® Tools for the Nucleotide Analysis of Immunoglobulin (IG) and T Cell Receptor (TR) V-(D)-J Repertoires, Polymorphisms, and IG Mutations: IMGT/V-QUEST and IMGT/HighV-QUEST for NGS.” In *Immunogenetics: Methods and Applications in Clinical Practice*, edited by Frank T. Christiansen and Brian D. Tait, 569–604. Methods in Molecular Biology. Totowa, NJ: Humana Press. https://doi.org/10.1007/978-1-61779-842-9_32.

Amimeur, Tileli, Jeremy M. Shaver, Randal R. Ketchem, J. Alex Taylor, Rutilio H. Clark, Josh Smith, Danielle Van Citters, et al. 2020. “Designing Feature-Controlled Humanoid Antibody Discovery Libraries Using Generative Adversarial Networks.” bioRxiv. https://doi.org/10.1101/2020.04.12.024844.

Barreto, Kris, Bharathikumar V. Maruthachalam, Wayne Hill, Daniel Hogan, Ashley R. Sutherland, Anthony Kusalik, Humphrey Fonge, John F. DeCoteau, and C. Ronald Geyer. 2019. “Next-Generation Sequencing-Guided Identification and Reconstruction of Antibody CDR Combinations from Phage Selection Outputs.” *Nucleic Acids Research* 47 (9): e50–e50. https://doi.org/10.1093/nar/gkz131.

Bashford-Rogers, R. J. M., L. Bergamaschi, E. F. McKinney, D. C. Pombal, F. Mescia, J. C. Lee, D. C. Thomas, et al. 2019. “Analysis of the B Cell Receptor Repertoire in Six Immune-Mediated Diseases.” *Nature* 574 (7776): 122–26. https://doi.org/10.1038/s41586-019-1595-3.

Bashford-Rogers, R. J. M., K. A. Nicolaou, J. Bartram, N. J. Goulden, L. Loizou, L. Koumas, J. Chi, et al. 2016. “Eye on the B-ALL: B-Cell Receptor Repertoires Reveal Persistence of Numerous B-Lymphoblastic Leukemia Subclones from Diagnosis to Relapse.” *Leukemia* 30 (12): 2312–21. https://doi.org/10.1038/leu.2016.142.

Bashford-Rogers, Rachael J.M., Anne L. Palser, Brian J. Huntly, Richard Rance, George S. Vassiliou, George A. Follows, and Paul Kellam. 2013. “Network Properties Derived from Deep Sequencing of Human B-Cell Receptor Repertoires Delineate B-Cell Populations.” *Genome Research* 23 (11): 1874–84. https://doi.org/10.1101/gr.154815.113.

Beerli, Roger R., Monika Bauer, Regula B. Buser, Myriam Gwerder, Simone Muntwiler, Patrik Maurer, Philippe Saudan, and Martin F. Bachmann. 2008. “Isolation of Human Monoclonal Antibodies by Mammalian Cell Display.” *Proceedings of the National Academy of Sciences* 105 (38): 14336–41. https://doi.org/10.1073/pnas.0805942105.

Benatuil, Lorenzo, Jennifer M. Perez, Jonathan Belk, and Chung-Ming Hsieh. 2010. “An Improved Yeast

Journal Pre-proof

Transformation Method for the Generation of Very Large Human Antibody Libraries." *Protein Engineering, Design and Selection* 23 (4): 155–59. https://doi.org/10.1093/protein/gzq002.

Bentzen, Amalie K., Lina Such, Kamilla K. Jensen, Andrea M. Marquard, Leon E. Jessen, Natalie J. Miller, Candice D. Church, et al. 2018. "T Cell Receptor Fingerprinting Enables In-Depth Characterization of the Interactions Governing Recognition of Peptide–MHC Complexes." *Nature Biotechnology* 36 (12): 1191–96. https://doi.org/10.1038/nbt.4303.

Birnbaum, Michael E., Juan L. Mendoza, Dhruv K. Sethi, Shen Dong, Jacob Glanville, Jessica Dobbins, Engin Özkan, Mark M. Davis, Kai W. Wucherpfennig, and K. Christopher Garcia. 2014. "Deconstructing the Peptide-MHC Specificity of T Cell Recognition." Cell 157 (5): 1073–87. https://doi.org/10.1016/j.cell.2014.03.047.

Boder, Eric T., Maryam Raeeszadeh-Sarmazdeh, and J. Vincent Price. 2012. "Engineering Antibodies by Yeast Display." *Archives of Biochemistry and Biophysics*, Antibody Engineering, 526 (2): 99–106. https://doi.org/10.1016/j.abb.2012.03.009.

Boder, Eric T., and K. Dane Wittrup. 1997. "Yeast Surface Display for Screening Combinatorial Polypeptide Libraries." *Nature Biotechnology* 15 (6): 553–57. https://doi.org/10.1038/nbt0697-553.

Borowicz, Paweł, Hanna Chan, Daniel Medina, Simon Gumpelmair, Hanna Kjelstrup, and Anne Spurkland. 2020. "A Simple and Efficient Workflow for Generation of Knock-in Mutations in Jurkat T Cells Using CRISPR/Cas9." *Scandinavian Journal of Immunology* 91 (4): e12862. https://doi.org/10.1111/sji.12862.

Boulter, Jonathan M., Meir Glick, Penio T. Todorov, Emma Baston, Malkit Sami, Pierre Rizkallah, and Bent K. Jakobsen. 2003. "Stable, Soluble T-Cell Receptor Molecules for Crystallization and Therapeutics." *Protein Engineering* 16 (9): 707–11. https://doi.org/10.1093/protein/gzg087.

Bray, Cara, David Wright, Sonja Haupt, Sharyn Thomas, Hans Stauss, and Rose Zamoyska. 2018. "Crispr/Cas Mediated Deletion of PTPN22 in Jurkat T Cells Enhances TCR Signaling and Production of IL-2." *Frontiers in Immunology* 9. https://doi.org/10.3389/fimmu.2018.02595.

Briney, Bryan, Anne Inderbitzin, Collin Joyce, and Dennis R. Burton. 2019. "Commonality despite Exceptional Diversity in the Baseline Human Antibody Repertoire." *Nature* 566 (7744): 393–97. https://doi.org/10.1038/s41586-019-0879-y.

Briney, Bryan, Devin Sok, Joseph G. Jardine, Daniel W. Kulp, Patrick Skog, Sergey Menis, Ronald Jacak, et al. 2016. "Tailored Immunogens Direct Affinity Maturation toward HIV Neutralizing Antibodies." *Cell* 166 (6): 1459-1470.e11. https://doi.org/10.1016/j.cell.2016.08.005.

Brown, Alex J., Igor Snapkov, Rahmad Akbar, Milena Pavlović, Enkelejda Miho, Geir K. Sandve, and Victor Greiff. 2019. "Augmenting Adaptive Immunity: Progress and Challenges in the Quantitative Engineering and Analysis of Adaptive Immune Receptor Repertoires." *Molecular Systems Design & Engineering* 4 (4): 701–36. https://doi.org/10.1039/C9ME00071B.

Byrne, Ashley, Anna E. Beaudin, Hugh E. Olsen, Miten Jain, Charles Cole, Theron Palmer, Rebecca M. DuBois, E. Camilla Forsberg, Mark Akeson, and Christopher Vollmers. 2017. "Nanopore Long-Read RNAseq Reveals Widespread Transcriptional Variation among the Surface Receptors of Individual B Cells." *Nature Communications* 8 (1): 16027. https://doi.org/10.1038/ncomms16027.

Cameron, Brian J., Andrew B. Gerry, Joseph Dukes, Jane V. Harper, Vivekanandan Kannan, Frayne C. Bianchi, Francis Grand, et al. 2013. "Identification of a Titin-Derived HLA-A1–Presented Peptide as a Cross-Reactive Target for Engineered MAGE A3–Directed T Cells." *Science Translational Medicine* 5 (197): 197ra103. https://doi.org/10.1126/scitranslmed.3006034.

Cao, Yunlong, Bin Su, Xianghua Guo, Wenjie Sun, Yongqiang Deng, Linlin Bao, Qinyu Zhu, et al. 2020. "Potent Neutralizing Antibodies against SARS-CoV-2 Identified by High-Throughput Single-Cell Sequencing of Convalescent Patients' B Cells." *Cell*, May. https://doi.org/10.1016/j.cell.2020.05.025.

Chan, Waipan, Rachel A. Gottschalk, Yikun Yao, Joel L. Pomerantz, and Ronald N. Germain. 2020. "Efficient Immune Cell Genome Engineering with Improved CRISPR Editing Tools." bioRxiv. https://doi.org/10.1101/2020.02.13.947002.

Chaudhary, Neha, and Duane R. Wesemann. 2018. "Analyzing Immunoglobulin Repertoires." *Frontiers in Immunology* 9 (March): 462. https://doi.org/10.3389/fimmu.2018.00462.

Chen, Xingyao, Thomas Dougherty, Chan Hong, Rachel Schibler, Yi Cong Zhao, Reza Sadeghi, Naim Matasci, Yi-Chieh Wu, and Ian Kerman. 2020. "Predicting Antibody Developability from Sequence Using Machine Learning." *bioRxiv*, June, 2020.06.18.159798. https://doi.org/10.1101/2020.06.18.159798.

Chervin, Adam S., David H. Aggen, John M. Raseman, and David M. Kranz. 2008. "Engineering Higher Affinity T Cell Receptors Using a T Cell Display System." *Journal of Immunological Methods* 339 (2): 175–84. https://doi.org/10.1016/j.jim.2008.09.016.

Chi, Shen, Arthur Weiss, and Haopeng Wang. 2016. "A CRISPR-Based Toolbox for Studying T Cell Signal Transduction." *BioMed Research International* 2016: 1–10. https://doi.org/10.1155/2016/5052369.

Clarke, James, Hai-Chen Wu, Lakmal Jayasinghe, Alpesh Patel, Stuart Reid, and Hagan Bayley. 2009. "Continuous Base Identification for Single-Molecule Nanopore DNA Sequencing." *Nature Nanotechnology* 4 (4): 265–70. https://doi.org/10.1038/nnano.2009.12.

Cole, Charles, Ashley Byrne, Matthew Adams, Roger Volden, and Christopher Vollmers. 2020. "Complete Characterization of the Human Immune Cell Transcriptome Using Accurate Full-Length CDNA Sequencing." *Genome Research*, April. https://doi.org/10.1101/gr.257188.119.

Cong, Le, F. Ann Ran, David Cox, Shuailiang Lin, Robert Barretto, Naomi Habib, Patrick D. Hsu, et al. 2013. "Multiplex Genome Engineering Using CRISPR/Cas Systems." *Science (New York, N.Y.)* 339 (6121): 819–23. https://doi.org/10.1126/science.1231143.

Corcoran, Martin M., Ganesh E. Phad, Néstor Vázquez Bernat, Christiane Stahl-Hennig, Noriyuki Sumida, Mats A. A. Persson, Marcel Martin, and Gunilla B. Karlsson Hedestam. 2016. "Production of Individualized V Gene Databases Reveals High Levels of Immunoglobulin Genetic Diversity." *Nature Communications* 7 (1): 1–14.

Journal Pre-proof
https://doi.org/10.1038/ncomms13642.
Croote, Derek, Spyros Darmanis, Kari C. Nadeau, and Stephen R. Quake. 2018. “High-Affinity Allergen-Specific Human Antibodies Cloned from Single IgE B Cell Transcriptomes.” *Science* 362 (6420): 1306–9. https://doi.org/10.1126/science.aau2599.
Dai, Xiaoyun, Jonathan J. Park, Yaying Du, Hyunu R. Kim, Guangchuan Wang, Youssef Errami, and Sidi Chen. 2019. “One-Step Generation of Modular CAR-T Cells with AAV–Cpf1.” *Nature Methods* 16 (3): 247–54. https://doi.org/10.1038/s41592-019-0329-7.
Davidsen, Kristian, Branden J Olson, William S DeWitt III, Jean Feng, Elias Harkins, Philip Bradley, and Frederick A Matsen IV. 2019. “Deep Generative Models for T Cell Receptor Protein Sequences.” Edited by Detlef Weigel, Arup K Chakraborty, Eric Huseby, and Curtis Callan. *ELife* 8 (September): e46935. https://doi.org/10.7554/eLife.46935.
De Neuter, Nicolas, Wout Bittremieux, Charlie Beirnaert, Bart Cuypers, Aida Mrzic, Pieter Moris, Arvid Suls, et al. 2018. “On the Feasibility of Mining CD8+ T Cell Receptor Patterns Underlying Immunogenic Peptide Recognition.” *Immunogenetics* 70 (3): 159–68. https://doi.org/10.1007/s00251-017-1023-5.
DeKosky, Brandon J., Gregory C. Ippolito, Ryan P. Deschner, Jason J. Lavinder, Yariv Wine, Brandon M. Rawlings, Navin Varadarajan, et al. 2013. “High-Throughput Sequencing of the Paired Human Immunoglobulin Heavy and Light Chain Repertoire.” *Nature Biotechnology* 31 (2): 166–69. https://doi.org/10.1038/nbt.2492.
DeKosky, Brandon J., Takaaki Kojima, Alexa Rodin, Wissam Charab, Gregory C. Ippolito, Andrew D. Ellington, and George Georgiou. 2015. “In-Depth Determination and Analysis of the Human Paired Heavy- and Light-Chain Antibody Repertoire.” *Nature Medicine* 21 (1): 86–91. https://doi.org/10.1038/nm.3743.
Dembić, Zlatko, Werner Haas, Rose Zamoyska, Jane Parnes, Michael Steinmetz, and Harald von Boehmer. 1987. “Transfection of the CD8 Gene Enhances T-Cell Recognition.” *Nature* 326 (6112): 510–11. https://doi.org/10.1038/326510a0.
DeWitt, William S, III, Anajane Smith, Gary Schoch, John A Hansen, Frederick A Matsen IV, and Philip Bradley. 2018. “Human T Cell Receptor Occurrence Patterns Encode Immune History, Genetic Background, and Receptor Specificity.” Edited by Aleksandra M Walczak, Arup K Chakraborty, Yuval Elhanati, and Bram Gerritsen. *ELife* 7 (August): e38358. https://doi.org/10.7554/eLife.38358.
Di Roberto, Raphaël B., Rocío Castellanos Rueda, Samara Frey, David Egli, Rodrigo Vazquez-Lombardi, and Sai T. Reddy. 2020. “A Functional Screening Strategy for Engineering Chimeric Antigen Receptors with Reduced On-Target, Off-Tumor Activation.” Mol. Ther. https://doi.org/10.1016/j.ymthe.2020.08.003.
Dias-Neto, Emmanuel, Diana N. Nunes, Ricardo J. Giordano, Jessica Sun, Gregory H. Botz, Kuan Yang, João C. Setubal, Renata Pasqualini, and Wadih Arap. 2009. “Next-Generation Phage Display: Integrating and Comparing Available Molecular Tools to Enable Cost-Effective High-Throughput Analysis.” *PLoS ONE* 4 (12). https://doi.org/10.1371/journal.pone.0008338.
Egorov, Evgeny S., Sofya A. Kasatskaya, Vasiliy N. Zubov, Mark Izraelson, Tatiana O. Nakonechnaya, Dmitriy B. Staroverov, Andrea Angius, et al. 2018. “The Changing Landscape of Naive T Cell Receptor Repertoire With Human Aging.” *Frontiers in Immunology* 9. https://doi.org/10.3389/fimmu.2018.01618.
Egorov, Evgeny S., Ekaterina M. Merzlyak, Andrew A. Shelenkov, Olga V. Britanova, George V. Sharonov, Dmitriy B. Staroverov, Dmitriy A. Bolotin, et al. 2015. “Quantitative Profiling of Immune Repertoires for Minor Lymphocyte Counts Using Unique Molecular Identifiers.” *The Journal of Immunology* 194 (12): 6155–63. https://doi.org/10.4049/jimmunol.1500215.
Ehrhardt, Stefanie A., Matthias Zehner, Verena Krähling, Hadas Cohen-Dvashi, Christoph Kreer, Nadav Elad, Henning Gruell, et al. 2019. “Polyclonal and Convergent Antibody Response to Ebola Virus Vaccine RVSV-ZEBOV.” *Nature Medicine* 25 (10): 1589–1600. https://doi.org/10.1038/s41591-019-0602-4.
Eid, John, Adrian Fehr, Jeremy Gray, Khai Luong, John Lyle, Geoff Otto, Paul Peluso, et al. 2009. “Real-Time DNA Sequencing from Single Polymerase Molecules.” *Science* 323 (5910): 133–38. https://doi.org/10.1126/science.1162986.
Emerson, Ryan O., William S. DeWitt, Marissa Vignali, Jenna Gravley, Joyce K. Hu, Edward J. Osborne, Cindy Desmarais, et al. 2017. “Immunosequencing Identifies Signatures of Cytomegalovirus Exposure History and HLA-Mediated Effects on the T Cell Repertoire.” *Nature Genetics* 49 (5): 659–65. https://doi.org/10.1038/ng.3822.
Eyquem, Justin, Jorge Mansilla-Soto, Theodoros Giavridis, Sjoukje J. C. van der Stegen, Mohamad Hamieh, Kristen M. Cunanan, Ashlesha Odak, Mithat Gönen, and Michel Sadelain. 2017. “Targeting a CAR to the TRAC Locus with CRISPR/Cas9 Enhances Tumour Rejection.” *Nature* 543 (7643): 113–17. https://doi.org/10.1038/nature21405.
Fischer, David S., Yihan Wu, Benjamin Schubert, and Fabian J. Theis. 2019. “Predicting Antigen-Specificity of Single T-Cells Based on TCR CDR3 Regions.” bioRxiv. https://doi.org/10.1101/734053.
Freeman, J. Douglas, René L. Warren, John R. Webb, Brad H. Nelson, and Robert A. Holt. 2009. “Profiling the T-Cell Receptor Beta-Chain Repertoire by Massively Parallel Sequencing.” *Genome Research* 19 (10): 1817–24. https://doi.org/10.1101/gr.092924.109.
Friedensohn, Simon, Tarik A. Khan, and Sai T. Reddy. 2017. “Advanced Methodologies in High-Throughput Sequencing of Immune Repertoires.” *Trends in Biotechnology* 35 (3): 203–14. https://doi.org/10.1016/j.tibtech.2016.09.010.
Friedensohn, Simon, John M. Lindner, Vanessa Cornacchione, Mariavittoria Iazeolla, Enkelejda Miho, Andreas Zingg, Simon Meng, Elisabetta Traggiai, and Sai T. Reddy. 2018. “Synthetic Standards Combined With Error and Bias Correction Improve the Accuracy and Quantitative Resolution of Antibody Repertoire Sequencing in Human Naïve and Memory B Cells.” *Frontiers in Immunology* 9. https://doi.org/10.3389/fimmu.2018.01401.
Friedensohn, Simon, Daniel Neumeier, Tarik A. Khan, Lucia Csepregi, Cristina Parola, Arthur R. Gorter de Vries, Lena Erlach, Derek M. Mason, and Sai T. Reddy. 2020. “Convergent Selection in Antibody Repertoires Is

Journal Pre-proof

Revealed by Deep Learning." *bioRxiv*, February, 2020.02.25.965673. https://doi.org/10.1101/2020.02.25.965673.

Froning, Karen, Jack Maguire, Arlene Sereno, Flora Huang, Shawn Chang, Kenneth Weichert, Anton J. Frommelt, et al. 2020. "Computational Stabilization of T Cell Receptors Allows Pairing with Antibodies to Form Bispecifics." *Nature Communications* 11 (1): 2330. https://doi.org/10.1038/s41467-020-16231-7.

Gadala-Maria, Daniel, Moriah Gidoni, Susanna Marquez, Jason A. Vander Heiden, Justin T. Kos, Corey T. Watson, Kevin C. O'Connor, Gur Yaari, and Steven H. Kleinstein. 2019. "Identification of Subject-Specific Immunoglobulin Alleles From Expressed Repertoire Sequencing Data." *Frontiers in Immunology* 10. https://doi.org/10.3389/fimmu.2019.00129.

Gadala-Maria, Daniel, Gur Yaari, Mohamed Uduman, and Steven H. Kleinstein. 2015. "Automated Analysis of High-Throughput B-Cell Sequencing Data Reveals a High Frequency of Novel Immunoglobulin V Gene Segment Alleles." *Proceedings of the National Academy of Sciences* 112 (8): E862–70. https://doi.org/10.1073/pnas.1417683112.

Galperin, Moran, Carine Farenc, Madhura Mukhopadhyay, Dhilshan Jayasinghe, Amandine Decroos, Daniela Benati, Li Lynn Tan, et al. 2018. "CD4+ T Cell–Mediated HLA Class II Cross-Restriction in HIV Controllers." *Science Immunology* 3 (24). https://doi.org/10.1126/sciimmunol.aat0687.

Galson, Jacob D., Johannes Trück, Anna Fowler, Elizabeth A. Clutterbuck, Márton Münz, Vincenzo Cerundolo, Claudia Reinhard, et al. 2015. "Analysis of B Cell Repertoire Dynamics Following Hepatitis B Vaccination in Humans, and Enrichment of Vaccine-Specific Antibody Sequences." *EBioMedicine* 2 (12): 2070–79. https://doi.org/10.1016/j.ebiom.2015.11.034.

Gálvez, Jesús, Juan J. Gálvez, and Pilar García-Peñarrubia. 2019. "Is TCR/PMHC Affinity a Good Estimate of the T-Cell Response? An Answer Based on Predictions From 12 Phenotypic Models." *Frontiers in Immunology* 10 (March). https://doi.org/10.3389/fimmu.2019.00349.

Gérard, Annabelle, Adam Woolfe, Guillaume Mottet, Marcel Reichen, Carlos Castrillon, Vera Menrath, Sami Ellouze, et al. 2020. "High-Throughput Single-Cell Activity-Based Screening and Sequencing of Antibodies Using Droplet Microfluidics." *Nature Biotechnology*, March, 1–7. https://doi.org/10.1038/s41587-020-0466-7.

Giudicelli, Véronique, Denys Chaume, and Marie-Paule Lefranc. 2004. "IMGT/V-QUEST, an Integrated Software Program for Immunoglobulin and T Cell Receptor V–J and V–D–J Rearrangement Analysis." *Nucleic Acids Research* 32 (suppl_2): W435–40. https://doi.org/10.1093/nar/gkh412.

Giudicelli, Véronique, Patrice Duroux, Sofia Kossida, and Marie-Paule Lefranc. 2017. "IG and TR Single Chain Fragment Variable (ScFv) Sequence Analysis: A New Advanced Functionality of IMGT/V-QUEST and IMGT/HighV-QUEST." *BMC Immunology* 18 (1): 35. https://doi.org/10.1186/s12865-017-0218-8.

Glanville, J, S D'Angelo, TA Khan, ST Reddy, L Naranjo, F Ferrara, and ARM Bradbury. 2015. "Deep Sequencing in Library Selection Projects: What Insight Does It Bring?" *Current Opinion in Structural Biology* 33 (August): 146–60. https://doi.org/10.1016/j.sbi.2015.09.001.

Glanville, Jacob, Huang Huang, Allison Nau, Olivia Hatton, Lisa E. Wagar, Florian Rubelt, Xuhuai Ji, et al. 2017. "Identifying Specificity Groups in the T Cell Receptor Repertoire." *Nature* 547 (7661): 94–98. https://doi.org/10.1038/nature22976.

Goldstein, Leonard D., Ying-Jiun J. Chen, Jia Wu, Subhra Chaudhuri, Yi-Chun Hsiao, Kellen Schneider, Kam Hon Hoi, et al. 2019. "Massively Parallel Single-Cell B-Cell Receptor Sequencing Enables Rapid Discovery of Diverse Antigen-Reactive Antibodies." *Communications Biology* 2 (1): 1–10. https://doi.org/10.1038/s42003-019-0551-y.

Greiff, Victor, Pooja Bhat, Skylar C. Cook, Ulrike Menzel, Wenjing Kang, and Sai T. Reddy. 2015. "A Bioinformatic Framework for Immune Repertoire Diversity Profiling Enables Detection of Immunological Status." *Genome Medicine* 7 (1): 49. https://doi.org/10.1186/s13073-015-0169-8.

Greiff, Victor, Ulrike Menzel, Enkelejda Miho, Cédric Weber, René Riedel, Skylar Cook, Atijeh Valai, et al. 2017. "Systems Analysis Reveals High Genetic and Antigen-Driven Predetermination of Antibody Repertoires throughout B Cell Development." *Cell Reports* 19 (7): 1467–78. https://doi.org/10.1016/j.celrep.2017.04.054.

Greiff, Victor, Enkelejda Miho, Ulrike Menzel, and Sai T. Reddy. 2015. "Bioinformatic and Statistical Analysis of Adaptive Immune Repertoires." *Trends in Immunology* 36 (11): 738–49. https://doi.org/10.1016/j.it.2015.09.006.

Greiff, Victor, Cédric R. Weber, Johannes Palme, Ulrich Bodenhofer, Enkelejda Miho, Ulrike Menzel, and Sai T. Reddy. 2017. "Learning the High-Dimensional Immunogenomic Features That Predict Public and Private Antibody Repertoires." *The Journal of Immunology* 199 (8): 2985–97. https://doi.org/10.4049/jimmunol.1700594.

Gunnarsen, Kristin Støen, Lene Støkken Høydahl, Ralf Stefan Neumann, Kaare Bjerregaard-Andersen, Nicolay Rustad Nilssen, Ludvig Magne Sollid, Inger Sandlie, and Geir Åge Løset. 2018. "Soluble T-Cell Receptor Design Influences Functional Yield in an E. Coli Chaperone-Assisted Expression System." *PLOS ONE* 13 (4): e0195868. https://doi.org/10.1371/journal.pone.0195868.

Gupta, Namita T., Jason A. Vander Heiden, Mohamed Uduman, Daniel Gadala-Maria, Gur Yaari, and Steven H. Kleinstein. 2015. "Change-O: A Toolkit for Analyzing Large-Scale B Cell Immunoglobulin Repertoire Sequencing Data." *Bioinformatics* 31 (20): 3356–58. https://doi.org/10.1093/bioinformatics/btv359.

Han, Seung Yub, Alesia Antoine, David Howard, Bryant Chang, Woo Sung Chang, Matthew Slein, Gintaras Deikus, et al. 2018. "Coupling of Single Molecule, Long Read Sequencing with IMGT/HighV-QUEST Analysis Expedites Identification of SIV Gp140-Specific Antibodies from ScFv Phage Display Libraries." *Frontiers in Immunology* 9. https://doi.org/10.3389/fimmu.2018.00329.

Hartweger, Harald, Andrew T. McGuire, Marcel Horning, Justin J. Taylor, Pia Dosenovic, Daniel Yost, Anna Gazumyan, et al. 2019. "HIV-Specific Humoral Immune Responses by CRISPR/Cas9-Edited B Cells." *The Journal of Experimental Medicine* 216 (6): 1301–10. https://doi.org/10.1084/jem.20190287.

Journal Pre-proof

Hemadou, Audrey, Véronique Giudicelli, Melissa Laird Smith, Marie-Paule Lefranc, Patrice Duroux, Sofia Kossida, Cheryl Heiner, et al. 2017. “Pacific Biosciences Sequencing and IMGT/HighV-QUEST Analysis of Full-Length Single Chain Fragment Variable from an In Vivo Selected Phage-Display Combinatorial Library.” *Frontiers in Immunology* 8. https://doi.org/10.3389/fimmu.2017.01796.

Hershberg, Uri, and Eline T. Luning Prak. 2015. “The Analysis of Clonal Expansions in Normal and Autoimmune B Cell Repertoires.” *Philosophical Transactions of the Royal Society B: Biological Sciences* 370 (1676): 20140239. https://doi.org/10.1098/rstb.2014.0239.

Ho, Mitchell, Satoshi Nagata, and Ira Pastan. 2006. “Isolation of Anti-CD22 Fv with High Affinity by Fv Display on Human Cells.” *Proceedings of the National Academy of Sciences* 103 (25): 9637–42. https://doi.org/10.1073/pnas.0603653103.

Hoehn, Kenneth B., Anna Fowler, Gerton Lunter, and Oliver G. Pybus. 2016. “The Diversity and Molecular Evolution of B-Cell Receptors during Infection.” *Molecular Biology and Evolution* 33 (5): 1147–57. https://doi.org/10.1093/molbev/msw015.

Hoehn, Kenneth B., Jason A. Vander Heiden, Julian Q. Zhou, Gerton Lunter, Oliver G. Pybus, and Steven H. Kleinstein. 2019. “Repertoire-Wide Phylogenetic Models of B Cell Molecular Evolution Reveal Evolutionary Signatures of Aging and Vaccination.” *Proceedings of the National Academy of Sciences* 116 (45): 22664–72. https://doi.org/10.1073/pnas.1906020116.

Hoehn, Kenneth B., Gerton Lunter, and Oliver G. Pybus. 2017. “A Phylogenetic Codon Substitution Model for Antibody Lineages.” *Genetics* 206 (1): 417–27. https://doi.org/10.1534/genetics.116.196303.

Holler, Phillip D., Lukasz K. Chlewicki, and David M. Kranz. 2003. “TCRs with High Affinity for Foreign PMHC Show Self-Reactivity.” Nature Immunology 4 (1): 55–62. https://doi.org/10.1038/ni863.

Horlick, Robert A., John L. Macomber, Peter M. Bowers, Tamlyn Y. Neben, Geoffery L. Tomlinson, Irina P. Krapf, Jennifer L. Dalton, Petra Verdino, and David J. King. 2013. “Simultaneous Surface Display and Secretion of Proteins from Mammalian Cells Facilitate Efficient in Vitro Selection and Maturation of Antibodies.” *The Journal of Biological Chemistry* 288 (27): 19861–69. https://doi.org/10.1074/jbc.M113.452482.

Horns, Felix, Cornelia L. Dekker, and Stephen R. Quake. 2020. “Memory B Cell Activation, Broad Anti-Influenza Antibodies, and Bystander Activation Revealed by Single-Cell Transcriptomics.” *Cell Reports* 30 (3): 905-913.e6. https://doi.org/10.1016/j.celrep.2019.12.063.

Howie, Bryan, Anna M. Sherwood, Ashley D. Berkebile, Jan Berka, Ryan O. Emerson, David W. Williamson, Ilan Kirsch, et al. 2015. “High-Throughput Pairing of T Cell Receptor α and β Sequences.” *Science Translational Medicine* 7 (301): 301ra131-301ra131. https://doi.org/10.1126/scitranslmed.aac5624.

Huang, Deli, Jenny Tuyet Tran, Alex Olson, Thomas Vollbrecht, Mariia V. Guryleva, Mary Tenuta, Roberta P. Fuller, et al. 2020. “Vaccine Elicitation of HIV Broadly Neutralizing Antibodies from Engineered B Cells.” *bioRxiv*, March, 2020.03.17.989699. https://doi.org/10.1101/2020.03.17.989699.

Hung, King L., Iana Meitlis, Malika Hale, Chun-Yu Chen, Swati Singh, Shaun W. Jackson, Carol H. Miao, Iram F. Khan, David J. Rawlings, and Richard G. James. 2018. “Engineering Protein-Secreting Plasma Cells by Homology-Directed Repair in Primary Human B Cells.” *Molecular Therapy* 26 (2): 456–67. https://doi.org/10.1016/j.ymthe.2017.11.012.

Jardine, J. G., D. W. Kulp, C. Havenar-Daughton, A. Sarkar, B. Briney, D. Sok, F. Sesterhenn, et al. 2016. “HIV-1 Broadly Neutralizing Antibody Precursor B Cells Revealed by Germline-Targeting Immunogen.” *Science* 351 (6280): 1458–63. https://doi.org/10.1126/science.aad9195.

Johnson, Matthew J., Kanut Laoharawee, Walker S. Lahr, Beau R. Webber, and Branden S. Moriarity. 2018. “Engineering of Primary Human B Cells with CRISPR/Cas9 Targeted Nuclease.” *Scientific Reports* 8 (1): 12144. https://doi.org/10.1038/s41598-018-30358-0.

June, Carl H., Roddy S. O’Connor, Omkar U. Kawalekar, Saba Ghassemi, and Michael C. Milone. 2018. “CAR T Cell Immunotherapy for Human Cancer.” *Science* 359 (6382): 1361–65. https://doi.org/10.1126/science.aar6711.

Kessels, Helmut W. H. G., Monika C. Wolkers, Marly D. van den Boom, Martin A. van den Valk, and Ton N. M. Schumacher. 2001. “Immunotherapy through TCR Gene Transfer.” *Nature Immunology* 2 (10): 957–61. https://doi.org/10.1038/ni1001-957.

Khan, Tarik A., Simon Friedensohn, Arthur R. Gorter de Vries, Jakub Straszewski, Hans-Joachim Ruscheweyh, and Sai T. Reddy. 2016. “Accurate and Predictive Antibody Repertoire Profiling by Molecular Amplification Fingerprinting.” *Science Advances* 2 (3): e1501371. https://doi.org/10.1126/sciadv.1501371.

Khosravi-Maharlooei, Mohsen, Aleksandar Obradovic, Aditya Misra, Keshav Motwani, Markus Holzl, Howard R. Seay, Susan DeWolf, et al. 2019. “Cross-Reactive Public TCR Sequences Undergo Positive Selection in the Human Thymic Repertoire.” *The Journal of Clinical Investigation* 129 (6): 2446–62. https://doi.org/10.1172/JCI124358.

Kidd, Brian A., Lauren A. Peters, Eric E. Schadt, and Joel T. Dudley. 2014. “Unifying Immunology with Informatics and Multiscale Biology.” *Nature Immunology* 15 (2): 118–27. https://doi.org/10.1038/ni.2787.

Kieke, M. C., E. V. Shusta, E. T. Boder, L. Teyton, K. D. Wittrup, and D. M. Kranz. 1999. “Selection of Functional T Cell Receptor Mutants from a Yeast Surface-Display Library.” Proceedings of the National Academy of Sciences 96 (10): 5651–56. https://doi.org/10.1073/pnas.96.10.5651.

Konishi, Hiroki, Daisuke Komura, Hiroto Katoh, Shinichiro Atsumi, Hirotomo Koda, Asami Yamamoto, Yasuyuki Seto, et al. 2019. “Capturing the Differences between Humoral Immunity in the Normal and Tumor Environments from Repertoire-Seq of B-Cell Receptors Using Supervised Machine Learning.” *BMC Bioinformatics* 20 (1): 267. https://doi.org/10.1186/s12859-019-2853-y.

Li, Dan, Xue Li, Wei-Lin Zhou, Yong Huang, Xiao Liang, Lin Jiang, Xiao Yang, et al. 2019. “Genetically Engineered T Cells for Cancer Immunotherapy.” *Signal Transduction and Targeted Therapy* 4 (1): 1–17. https://doi.org/10.1038/s41392-019-0070-9.

Liao, Mingfeng, Yang Liu, Jing Yuan, Yanling Wen, Gang Xu, Juanjuan Zhao, Lin Cheng, et al. 2020. “Single-Cell

Journal Pre-proof

Landscape of Bronchoalveolar Immune Cells in Patients with COVID-19." *Nature Medicine*, May, 1–3. https://doi.org/10.1038/s41591-020-0901-9.

Lindenbaum, Ofir, Nima Nouri, Yuval Kluger, and Steven Kleinstein. 2020. "Alignment Free Identification of Clones in B Cell Receptor Repertoires." *bioRxiv*, March, 2020.03.30.017384. https://doi.org/10.1101/2020.03.30.017384.

Liu, Ge, Haoyang Zeng, Jonas Mueller, Brandon Carter, Ziheng Wang, Jonas Schilz, Geraldine Horny, Michael E. Birnbaum, Stefan Ewert, and David K. Gifford. 2020. "Antibody Complementarity Determining Region Design Using High-Capacity Machine Learning." *Bioinformatics* 36 (7): 2126–33. https://doi.org/10.1093/bioinformatics/btz895.

López-Santibáñez-Jácome, Laura, S. Eréndira Avendaño-Vázquez, and Carlos Fabián Flores-Jasso. 2019. "The Pipeline Repertoire for Ig-Seq Analysis." *Frontiers in Immunology* 10 (April): 899. https://doi.org/10.3389/fimmu.2019.00899.

Ma, Hong, Sam Kunes, Peter J. Schatz, and David Botstein. 1987. "Plasmid Construction by Homologous Recombination in Yeast." *Gene* 58 (2): 201–16. https://doi.org/10.1016/0378-1119(87)90376-3.

Ma, Ke-Yue, Chenfeng He, Ben S. Wendel, Chad M. Williams, Jun Xiao, Hui Yang, and Ning Jiang. 2018. "Immune Repertoire Sequencing Using Molecular Identifiers Enables Accurate Clonality Discovery and Clone Size Quantification." *Frontiers in Immunology* 9 (February). https://doi.org/10.3389/fimmu.2018.00033.

Madi, Asaf, Asaf Poran, Eric Shifrut, Shlomit Reich-Zeliger, Erez Greenstein, Irena Zaretsky, Tomer Arnon, et al. 2017. "T Cell Receptor Repertoires of Mice and Humans Are Clustered in Similarity Networks around Conserved Public CDR3 Sequences." Edited by Arup K Chakraborty. *ELife* 6 (July): e22057. https://doi.org/10.7554/eLife.22057.

Madi, Asaf, Eric Shifrut, Shlomit Reich-Zeliger, Hilah Gal, Katharine Best, Wilfred Ndifon, Benjamin Chain, Irun R. Cohen, and Nir Friedman. 2014. "T-Cell Receptor Repertoires Share a Restricted Set of Public and Abundant CDR3 Sequences That Are Associated with Self-Related Immunity." *Genome Research* 24 (10): 1603–12. https://doi.org/10.1101/gr.170753.113.

Malecek, Karolina, Shi Zhong, Katelyn McGary, Connie Yu, Kevin Huang, Laura A. Johnson, Steven A. Rosenberg, and Michelle Krogsgaard. 2013. "Engineering Improved T Cell Receptors Using an Alanine-Scan Guided T Cell Display Selection System." *Journal of Immunological Methods* 392 (1–2): 1–11. https://doi.org/10.1016/j.jim.2013.02.018.

Marcou, Quentin, Thierry Mora, and Aleksandra M. Walczak. 2018. "High-Throughput Immune Repertoire Analysis with IGoR." *Nature Communications* 9 (1): 561. https://doi.org/10.1038/s41467-018-02832-w.

Margulies, Marcel, Michael Egholm, William E. Altman, Said Attiya, Joel S. Bader, Lisa A. Bemben, Jan Berka, et al. 2005. "Genome Sequencing in Microfabricated High-Density Picolitre Reactors." *Nature* 437 (7057): 376–80. https://doi.org/10.1038/nature03959.

Marzec, Marek, Agnieszka Brąszewska-Zalewska, and Goetz Hensel. 2020. "Prime Editing: A New Way for Genome Editing." *Trends in Cell Biology* 30 (4): 257–59. https://doi.org/10.1016/j.tcb.2020.01.004.

Mason, Derek M, Simon Friedensohn, Cédric R Weber, Christian Jordi, Bastian Wagner, Simon Meng, Pablo Gainza, Bruno E Correia, and Sai T Reddy. 2019. "Deep Learning Enables Therapeutic Antibody Optimization in Mammalian Cells by Deciphering High-Dimensional Protein Sequence Space." bioRxiv. https://doi.org/10.1101/617860.

Mason, Derek M., Cédric R. Weber, Cristina Parola, Simon M. Meng, Victor Greiff, William J. Kelton, and Sai T. Reddy. 2018. "High-Throughput Antibody Engineering in Mammalian Cells by CRISPR/Cas9-Mediated Homology-Directed Mutagenesis." *Nucleic Acids Research* 46 (14): 7436–49. https://doi.org/10.1093/nar/gky550.

Maude, Shannon L., Noelle Frey, Pamela A. Shaw, Richard Aplenc, David M. Barrett, Nancy J. Bunin, Anne Chew, et al. 2014. "Chimeric Antigen Receptor T Cells for Sustained Remissions in Leukemia." *New England Journal of Medicine* 371 (16): 1507–17. https://doi.org/10.1056/NEJMoa1407222.

McDaniel, Jonathan R., Brandon J. DeKosky, Hidetaka Tanno, Andrew D. Ellington, and George Georgiou. 2016. "Ultra-High-Throughput Sequencing of the Immune Receptor Repertoire from Millions of Lymphocytes." *Nature Protocols* 11 (3): 429–42. https://doi.org/10.1038/nprot.2016.024.

Meng, Wenzhao, Bochao Zhang, Gregory W. Schwartz, Aaron M. Rosenfeld, Daqiu Ren, Joseph J. C. Thome, Dustin J. Carpenter, et al. 2017. "An Atlas of B-Cell Clonal Distribution in the Human Body." *Nature Biotechnology* 35 (9): 879–84. https://doi.org/10.1038/nbt.3942.

Metzker, Michael L. 2010. "Sequencing Technologies — the next Generation." *Nature Reviews Genetics* 11 (1): 31–46. https://doi.org/10.1038/nrg2626.

Meysman, Pieter, Nicolas De Neuter, Sofie Gielis, Danh Bui Thi, Benson Ogunjimi, and Kris Laukens. 2019. "On the Viability of Unsupervised T-Cell Receptor Sequence Clustering for Epitope Preference." *Bioinformatics* 35 (9): 1461–68. https://doi.org/10.1093/bioinformatics/bty821.

Miho, Enkelejda, Rok Roškar, Victor Greiff, and Sai T. Reddy. 2019. "Large-Scale Network Analysis Reveals the Sequence Space Architecture of Antibody Repertoires." *Nature Communications* 10 (1): 1–11. https://doi.org/10.1038/s41467-019-09278-8.

Miho, Enkelejda, Alexander Yermanos, Cédric R. Weber, Christoph T. Berger, Sai T. Reddy, and Victor Greiff. 2018. "Computational Strategies for Dissecting the High-Dimensional Complexity of Adaptive Immune Repertoires." *Frontiers in Immunology* 9. https://doi.org/10.3389/fimmu.2018.00224.

Moffett, Howell F., Carson K. Harms, Kristin S. Fitzpatrick, Marti R. Tooley, Jim Boonyaratanakornkit, and Justin J. Taylor. 2019. "B Cells Engineered to Express Pathogen-Specific Antibodies Protect against Infection." *Science Immunology* 4 (35). https://doi.org/10.1126/sciimmunol.aax0644.

Morgan, Richard A, James C Yang, Mio Kitano, Mark E Dudley, Carolyn M Laurencot, and Steven A Rosenberg. 2010. "Case Report of a Serious Adverse Event Following the Administration of T Cells Transduced With a

Journal Pre-proof

Chimeric Antigen Receptor Recognizing ERBB2." *Molecular Therapy* 18 (4): 843–51. https://doi.org/10.1038/mt.2010.24.

Nahmad, Alessio D., Yuval Raviv, Miriam Horovitz-Fried, Ilan Sofer, Tal Akriv, Daniel Nataf, Iris Dotan, et al. 2020. "B Cells Engineered to Express an Anti-HIV Antibody Allow Memory Retention, Class Switch Recombination and Clonal Selection in Mice." *bioRxiv*, March, 2020.02.28.970822. https://doi.org/10.1101/2020.02.28.970822.

Nouri, Nima, and Steven H. Kleinstein. 2018. "Optimized Threshold Inference for Partitioning of Clones From High-Throughput B Cell Repertoire Sequencing Data." *Frontiers in Immunology* 9 (July). https://doi.org/10.3389/fimmu.2018.01687.

Nouri, Nima, and Steven H. KLeinstein. 2020. "Somatic Hypermutation Analysis for Improved Identification of B Cell Clonal Families from Next-Generation Sequencing Data." *PLOS Computational Biology* 16 (6): e1007977. https://doi.org/10.1371/journal.pcbi.1007977.

Ohlin, Mats, Cathrine Scheepers, Martin Corcoran, William D. Lees, Christian E. Busse, Davide Bagnara, Linnea Thörnqvist, et al. 2019. "Inferred Allelic Variants of Immunoglobulin Receptor Genes: A System for Their Evaluation, Documentation, and Naming." *Frontiers in Immunology* 10. https://doi.org/10.3389/fimmu.2019.00435.

Overall, Sarah A., Jugmohit S. Toor, Stephanie Hao, Mark Yarmarkovich, Sara M. O'Rourke, Giora I. Morozov, Son Nguyen, et al. 2020. "High Throughput PMHC-I Tetramer Library Production Using Chaperone-Mediated Peptide Exchange." *Nature Communications* 11 (1): 1909. https://doi.org/10.1038/s41467-020-15710-1.

Papalexi, Efthymia, and Rahul Satija. 2018. "Single-Cell RNA Sequencing to Explore Immune Cell Heterogeneity." *Nature Reviews Immunology* 18 (1): 35–45. https://doi.org/10.1038/nri.2017.76.

Parameswaran, Poornima, Yi Liu, Krishna M. Roskin, Katherine K.L. Jackson, Vaishali P. Dixit, Ji-Yeun Lee, Karen L. Artiles, et al. 2013. "Convergent Antibody Signatures in Human Dengue." *Cell Host & Microbe* 13 (6): 691–700. https://doi.org/10.1016/j.chom.2013.05.008.

Parola, Cristina, Daniel Neumeier, Simon Friedensohn, Lucia Csepregi, Mariangela Di Tacchio, Derek M. Mason, and Sai T. Reddy. 2019. "Antibody Discovery and Engineering by Enhanced CRISPR-Cas9 Integration of Variable Gene Cassette Libraries in Mammalian Cells." *MAbs* 11 (8): 1367–80. https://doi.org/10.1080/19420862.2019.1662691.

Payne, Alexander, Nadine Holmes, Vardhman Rakyan, and Matthew Loose. 2019. "BulkVis: A Graphical Viewer for Oxford Nanopore Bulk FAST5 Files." *Bioinformatics* 35 (13): 2193–98. https://doi.org/10.1093/bioinformatics/bty841.

Pesch, Theresa, Lucia Bonati, William Kelton, Cristina Parola, Roy A. Ehling, Lucia Csepregi, Daisuke Kitamura, and Sai T. Reddy. 2019. "Molecular Design, Optimization, and Genomic Integration of Chimeric B Cell Receptors in Murine B Cells." Frontiers in Immunology 10. https://doi.org/10.3389/fimmu.2019.02630.

Pfeiffer, Franziska, Carsten Gröber, Michael Blank, Kristian Händler, Marc Beyer, Joachim L. Schultze, and Günter Mayer. 2018. "Systematic Evaluation of Error Rates and Causes in Short Samples in Next-Generation Sequencing." *Scientific Reports* 8 (1): 1–14. https://doi.org/10.1038/s41598-018-29325-6.

Pogorelyy, Mikhail V., Anastasia A. Minervina, Mikhail Shugay, Dmitriy M. Chudakov, Yuri B. Lebedev, Thierry Mora, and Aleksandra M. Walczak. 2019. "Detecting T Cell Receptors Involved in Immune Responses from Single Repertoire Snapshots." *PLOS Biology* 17 (6): e3000314. https://doi.org/10.1371/journal.pbio.3000314.

Pogorelyy, Mikhail V., Anastasia A. Minervina, Maximilian Puelma Touzel, Anastasiia L. Sycheva, Ekaterina A. Komech, Elena I. Kovalenko, Galina G. Karganova, et al. 2018. "Precise Tracking of Vaccine-Responding T Cell Clones Reveals Convergent and Personalized Response in Identical Twins." *Proceedings of the National Academy of Sciences* 115 (50): 12704–9. https://doi.org/10.1073/pnas.1809642115.

Pogson, Mark, Cristina Parola, William J. Kelton, Paul Heuberger, and Sai T. Reddy. 2016. "Immunogenomic Engineering of a Plug-and-(Dis)Play Hybridoma Platform." *Nature Communications* 7 (1): 12535. https://doi.org/10.1038/ncomms12535.

Ralph, Duncan K., and Frederick A. Matsen IV. 2016. "Likelihood-Based Inference of B Cell Clonal Families." *PLOS Computational Biology* 12 (10): e1005086. https://doi.org/10.1371/journal.pcbi.1005086.

Ralph, Duncan K., and Frederick A. Matsen IV. 2016. "Consistency of VDJ Rearrangement and Substitution Parameters Enables Accurate B Cell Receptor Sequence Annotation." *PLOS Computational Biology* 12 (1): e1004409. https://doi.org/10.1371/journal.pcbi.1004409.

Ravn, U., F. Gueneau, L. Baerlocher, M. Osteras, M. Desmurs, P. Malinge, G. Magistrelli, L. Farinelli, M. H. Kosco-Vilbois, and N. Fischer. 2010. "By-Passing in Vitro Screening—next Generation Sequencing Technologies Applied to Antibody Display and in Silico Candidate Selection." *Nucleic Acids Research* 38 (21): e193–e193. https://doi.org/10.1093/nar/gkq789.

Robins, Harlan S., Paulo V. Campregher, Santosh K. Srivastava, Abigail Wacher, Cameron J. Turtle, Orsalem Kahsai, Stanley R. Riddell, Edus H. Warren, and Christopher S. Carlson. 2009. "Comprehensive Assessment of T-Cell Receptor β-Chain Diversity in Aβ T Cells." *Blood* 114 (19): 4099–4107. https://doi.org/10.1182/blood-2009-04-217604.

Robins, Harlan S., Santosh K. Srivastava, Paulo V. Campregher, Cameron J. Turtle, Jessica Andriesen, Stanley R. Riddell, Christopher S. Carlson, and Edus H. Warren. 2010. "Overlap and Effective Size of the Human CD8+ T Cell Receptor Repertoire." *Science Translational Medicine* 2 (47): 47ra64-47ra64. https://doi.org/10.1126/scitranslmed.3001442.

Rosenberg, Alexander B., Charles M. Roco, Richard A. Muscat, Anna Kuchina, Paul Sample, Zizhen Yao, Lucas T. Graybuck, et al. 2018. "Single-Cell Profiling of the Developing Mouse Brain and Spinal Cord with Split-Pool Barcoding." *Science* 360 (6385): 176–82. https://doi.org/10.1126/science.aam8999.

Rosenfeld, Aaron M., Wenzhao Meng, Dora Y. Chen, Bochao Zhang, Tomer Granot, Donna L. Farber, Uri Hershberg, and Eline T. Luning Prak. 2018. "Computational Evaluation of B-Cell Clone Sizes in Bulk Populations."

Journal Pre-proof

Frontiers in Immunology 9. https://doi.org/10.3389/fimmu.2018.01472.

Rosenfeld, Aaron M., Wenzhao Meng, Eline T. Luning Prak, and Uri Hershberg. 2018. “ImmuneDB, a Novel Tool for the Analysis, Storage, and Dissemination of Immune Repertoire Sequencing Data.” *Frontiers in Immunology* 9. https://doi.org/10.3389/fimmu.2018.02107.

Rosskopf, Sandra, Judith Leitner, Wolfgang Paster, Laura T. Morton, Renate S. Hagedoorn, Peter Steinberger, and Mirjam H. M. Heemskerk. 2018. “A Jurkat 76 Based Triple Parameter Reporter System to Evaluate TCR Functions and Adoptive T Cell Strategies.” *Oncotarget* 9 (25): 17608–19. https://doi.org/10.18632/oncotarget.24807.

Roth, Theodore L., P. Jonathan Li, Franziska Blaeschke, Jasper F. Nies, Ryan Apathy, Cody Mowery, Ruby Yu, et al. 2020. “Pooled Knockin Targeting for Genome Engineering of Cellular Immunotherapies.” *Cell* 181 (3): 728-744.e21. https://doi.org/10.1016/j.cell.2020.03.039.

Roth, Theodore L., Cristina Puig-Saus, Ruby Yu, Eric Shifrut, Julia Carnevale, P. Jonathan Li, Joseph Hiatt, et al. 2018. “Reprogramming Human T Cell Function and Specificity with Non-Viral Genome Targeting.” *Nature* 559 (7714): 405–9. https://doi.org/10.1038/s41586-018-0326-5.

Rydzek, Julian, Thomas Nerreter, Haiyong Peng, Sabrina Jutz, Judith Leitner, Peter Steinberger, Hermann Einsele, Christoph Rader, and Michael Hudecek. 2019. “Chimeric Antigen Receptor Library Screening Using a Novel NF-KB/NFAT Reporter Cell Platform.” *Molecular Therapy* 27 (2): 287–99. https://doi.org/10.1016/j.ymthe.2018.11.015.

Safonova, Yana, Alla Lapidus, and Jennie Lill. 2015. “IgSimulator: A Versatile Immunosequencing Simulator.” *Bioinformatics* 31 (19): 3213–15. https://doi.org/10.1093/bioinformatics/btv326.

Safonova, Yana, and Pavel A. Pevzner. 2019. “De Novo Inference of Diversity Genes and Analysis of Non-Canonical V(DD)J Recombination in Immunoglobulins.” *Frontiers in Immunology* 10. https://doi.org/10.3389/fimmu.2019.00987.

Saggy, Ido, Yariv Wine, Leeron Shefet-Carasso, Limor Nahary, George Georgiou, and Itai Benhar. 2012. “Antibody Isolation from Immunized Animals: Comparison of Phage Display and Antibody Discovery via V Gene Repertoire Mining.” *Protein Engineering, Design and Selection* 25 (10): 539–49. https://doi.org/10.1093/protein/gzs060.

Saikia, Mridusmita, Philip Burnham, Sara H. Keshavjee, Michael F. Z. Wang, Michael Heyang, Pablo Moral-Lopez, Meleana M. Hinchman, Charles G. Danko, John S. L. Parker, and Iwijn De Vlaminck. 2019. “Simultaneous Multiplexed Amplicon Sequencing and Transcriptome Profiling in Single Cells.” *Nature Methods* 16 (1): 59–62. https://doi.org/10.1038/s41592-018-0259-9.

Schirmer, Melanie, Umer Z. Ijaz, Rosalinda D’Amore, Neil Hall, William T. Sloan, and Christopher Quince. 2015. “Insight into Biases and Sequencing Errors for Amplicon Sequencing with the Illumina MiSeq Platform.” *Nucleic Acids Research* 43 (6): e37–e37. https://doi.org/10.1093/nar/gku1341.

Schmid, Daphné A., Melita B. Irving, Vilmos Posevitz, Michael Hebeisen, Anita Posevitz-Fejfar, J.-C. Floyd Sarria, Raquel Gomez-Eerland, et al. 2010. “Evidence for a TCR Affinity Threshold Delimiting Maximal CD8 T Cell Function.” *The Journal of Immunology* 184 (9): 4936–46. https://doi.org/10.4049/jimmunol.1000173.

Schmitt, Thomas M., David H. Aggen, Kumiko Ishida-Tsubota, Sebastian Ochsenreither, David M. Kranz, and Philip D. Greenberg. 2017. “Generation of TCRs of Higher Affinity by Antigen-Driven Differentiation of Progenitor T Cells in Vitro.” *Nature Biotechnology* 35 (12): 1188–95. https://doi.org/10.1038/nbt.4004.

Schober, Kilian, Thomas R. Müller, Füsun Gökmen, Simon Grassmann, Manuel Effenberger, Mateusz Poltorak, Christian Stemberger, et al. 2019. “Orthotopic Replacement of T-Cell Receptor α- and β-Chains with Preservation of near-Physiological T-Cell Function.” *Nature Biomedical Engineering* 3 (12): 974–84. https://doi.org/10.1038/s41551-019-0409-0.

Sethna, Zachary, Yuval Elhanati, Curtis G. Callan, Aleksandra M. Walczak, and Thierry Mora. 2019. “OLGA: Fast Computation of Generation Probabilities of B- and T-Cell Receptor Amino Acid Sequences and Motifs.” *Bioinformatics* 35 (17): 2974–81. https://doi.org/10.1093/bioinformatics/btz035.

Setliff, Ian, Wyatt J. McDonnell, Nagarajan Raju, Robin G. Bombardi, Amyn A. Murji, Cathrine Scheepers, Rutendo Ziki, et al. 2018. “Multi-Donor Longitudinal Antibody Repertoire Sequencing Reveals the Existence of Public Antibody Clonotypes in HIV-1 Infection.” *Cell Host & Microbe* 23 (6): 845-854.e6. https://doi.org/10.1016/j.chom.2018.05.001.

Setliff, Ian, Andrea R. Shiakolas, Kelsey A. Pilewski, Amyn A. Murji, Rutendo E. Mapengo, Katarzyna Janowska, Simone Richardson, et al. 2019. “High-Throughput Mapping of B Cell Receptor Sequences to Antigen Specificity.” *Cell* 179 (7): 1636-1646.e15. https://doi.org/10.1016/j.cell.2019.11.003.

Shugay, Mikhail, Dmitriy V. Bagaev, Ivan V. Zvyagin, Renske M. Vroomans, Jeremy Chase Crawford, Garry Dolton, Ekaterina A. Komech, et al. 2018. “VDJdb: A Curated Database of T-Cell Receptor Sequences with Known Antigen Specificity.” *Nucleic Acids Research* 46 (D1): D419–27. https://doi.org/10.1093/nar/gkx760.

Shugay, Mikhail, Olga V. Britanova, Ekaterina M. Merzlyak, Maria A. Turchaninova, Ilgar Z. Mamedov, Timur R. Tuganbaev, Dmitriy A. Bolotin, et al. 2014. “Towards Error-Free Profiling of Immune Repertoires.” *Nature Methods* 11 (6): 653–55. https://doi.org/10.1038/nmeth.2960.

Sibener, Leah V., Ricardo A. Fernandes, Elizabeth M. Kolawole, Catherine B. Carbone, Fan Liu, Darren McAffee, Michael E. Birnbaum, et al. 2018. “Isolation and Visualization of a Structural Trigger That Uncouples TCR Signaling from PMHC Binding.” Cell 174 (3): 672. https://doi.org/10.1016/j.cell.2018.06.017.

Sidhom, John-William, H. Benjamin Larman, Petra Ross-MacDonald, Megan Wind-Rotolo, Drew M. Pardoll, and Alexander S. Baras. 2018. “DeepTCR: A Deep Learning Framework for Understanding T-Cell Receptor Sequence Signatures within Complex T-Cell Repertoires.” bioRxiv. https://doi.org/10.1101/464107.

Simeonov, Dimitre R., Benjamin G. Gowen, Mandy Boontanrart, Theodore L. Roth, John D. Gagnon, Maxwell R. Mumbach, Ansuman T. Satpathy, et al. 2017. “Discovery of Stimulation-Responsive Immune Enhancers with CRISPR Activation.” *Nature* 549 (7670): 111–15. https://doi.org/10.1038/nature23875.

Journal Pre-proof

Singh, Mandeep, Ghamdan Al-Eryani, Shaun Carswell, James M. Ferguson, James Blackburn, Kirston Barton, Daniel Roden, et al. 2019. “High-Throughput Targeted Long-Read Single Cell Sequencing Reveals the Clonal and Transcriptional Landscape of Lymphocytes.” *Nature Communications* 10 (1): 1–13. https://doi.org/10.1038/s41467-019-11049-4.

Sivelle, Coline, Raphaël Sierocki, Kelly Ferreira-Pinto, Stéphanie Simon, Bernard Maillere, and Hervé Nozach. 2018. “Fab Is the Most Efficient Format to Express Functional Antibodies by Yeast Surface Display.” *MAbs* 10 (5): 720–29. https://doi.org/10.1080/19420862.2018.1468952.

Slaga, Dionysos, Diego Ellerman, T. Noelle Lombana, Rajesh Vij, Ji Li, Maria Hristopoulos, Robyn Clark, et al. 2018. “Avidity-Based Binding to HER2 Results in Selective Killing of HER2-Overexpressing Cells by Anti-HER2/CD3.” *Science Translational Medicine* 10 (463). https://doi.org/10.1126/scitranslmed.aat5775.

Smakaj, Erand, Lmar Babrak, Mats Ohlin, Mikhail Shugay, Bryan Briney, Deniz Tosoni, Christopher Galli, et al. 2019. “Benchmarking Immunoinformatic Tools for the Analysis of Antibody Repertoire Sequences.” Edited by Inanc Birols. *Bioinformatics*, December, btz845. https://doi.org/10.1093/bioinformatics/btz845.

Smith, G. P. 1985. “Filamentous Fusion Phage: Novel Expression Vectors That Display Cloned Antigens on the Virion Surface.” *Science (New York, N.Y.)* 228 (4705): 1315–17. https://doi.org/10.1126/science.4001944.

Smith, Sheena N., Daniel T. Harris, and David M. Kranz. 2015. “T Cell Receptor Engineering and Analysis Using the Yeast Display Platform.” *Methods in Molecular Biology (Clifton, N.J.)* 1319: 95–141. https://doi.org/10.1007/978-1-4939-2748-7_6.

Sok, D., B. Briney, J. G. Jardine, D. W. Kulp, S. Menis, M. Pauthner, A. Wood, et al. 2016. “Priming HIV-1 Broadly Neutralizing Antibody Precursors in Human Ig Loci Transgenic Mice.” *Science* 353 (6307): 1557–60. https://doi.org/10.1126/science.aah3945.

Soto, Cinque, Robin G. Bombardi, Andre Branchizio, Nurgun Kose, Pranathi Matta, Alexander M. Sevy, Robert S. Sinkovits, Pavlo Gilchuk, Jessica A. Finn, and James E. Crowe. 2019. “High Frequency of Shared Clonotypes in Human B Cell Receptor Repertoires.” *Nature*, February. https://doi.org/10.1038/s41586-019-0934-8.

Spiliotopoulos, Anastasios, Jonathan. P. Owen, Ben. C. Maddison, Ingrid Dreveny, Helen. C. Rees, and Kevin. C. Gough. 2015. “Sensitive Recovery of Recombinant Antibody Clones after Their in Silico Identification within NGS Datasets.” *Journal of Immunological Methods* 420 (May): 50–55. https://doi.org/10.1016/j.jim.2015.03.005.

Spindler, Matthew J., Ayla L. Nelson, Ellen K. Wagner, Natasha Oppermans, John S. Bridgeman, James M. Heather, Adam S. Adler, et al. 2020. “Massively Parallel Interrogation and Mining of Natively Paired Human TCRαβ Repertoires.” *Nature Biotechnology* 38 (5): 609–19. https://doi.org/10.1038/s41587-020-0438-y.

Stadtmauer, Edward A., Joseph A. Fraietta, Megan M. Davis, Adam D. Cohen, Kristy L. Weber, Eric Lancaster, Patricia A. Mangan, et al. 2020. “CRISPR-Engineered T Cells in Patients with Refractory Cancer.” *Science* 367 (6481). https://doi.org/10.1126/science.aba7365.

Steichen, Jon M., Ying-Cing Lin, Colin Havenar-Daughton, Simone Pecetta, Gabriel Ozorowski, Jordan R. Willis, Laura Toy, et al. 2019. “A Generalized HIV Vaccine Design Strategy for Priming of Broadly Neutralizing Antibody Responses.” *Science* 366 (6470): eaax4380. https://doi.org/10.1126/science.aax4380.

Stenger, Dana, Tanja Andrea Stief, Theresa Käuferle, Semjon Willier, Felicitas Rataj, Kilian Schober, Binje Vick, et al. n.d. “Endogenous TCR Promotes in Vivo Persistence of CD19-CAR-T Cells Compared to a CRISPR/Cas9-Mediated TCR Knockout CAR.” *Blood*. Accessed June 8, 2020. https://doi.org/10.1182/blood.2020005185.

Stoeckius, Marlon, Christoph Hafemeister, William Stephenson, Brian Houck-Loomis, Pratip K. Chattopadhyay, Harold Swerdlow, Rahul Satija, and Peter Smibert. 2017. “Simultaneous Epitope and Transcriptome Measurement in Single Cells.” *Nature Methods* 14 (9): 865–68. https://doi.org/10.1038/nmeth.4380.

Stubbington, Michael J. T., Tapio Lönnberg, Valentina Proserpio, Simon Clare, Anneliese O. Speak, Gordon Dougan, and Sarah A. Teichmann. 2016. “T Cell Fate and Clonality Inference from Single-Cell Transcriptomes.” *Nature Methods* 13 (4): 329–32. https://doi.org/10.1038/nmeth.3800.

Tan, Joshua, Kathrin Pieper, Luca Piccoli, Abdirahman Abdi, Mathilde Foglierini, Roger Geiger, Claire Maria Tully, et al. 2016. “A LAIR1 Insertion Generates Broadly Reactive Antibodies against Malaria Variant Antigens.” *Nature* 529 (7584): 105–9. https://doi.org/10.1038/nature16450.

Tomimatsu, Kosuke, Shin-ei Matsumoto, Hayato Tanaka, Makiko Yamashita, Hidekazu Nakanishi, Kiichiro Teruya, Saiko Kazuno, et al. 2013. “A Rapid Screening and Production Method Using a Novel Mammalian Cell Display to Isolate Human Monoclonal Antibodies.” *Biochemical and Biophysical Research Communications* 441 (1): 59–64. https://doi.org/10.1016/j.bbrc.2013.10.007.

Tonegawa, S. 1983. “Somatic Generation of Antibody Diversity.” *Nature* 302 (5909): 575–81. https://doi.org/10.1038/302575a0.

Trück, Johannes, Maheshi N. Ramasamy, Jacob D. Galson, Richard Rance, Julian Parkhill, Gerton Lunter, Andrew J. Pollard, and Dominic F. Kelly. 2015. “Identification of Antigen-Specific B Cell Receptor Sequences Using Public Repertoire Analysis.” *The Journal of Immunology* 194 (1): 252–61. https://doi.org/10.4049/jimmunol.1401405.

Turchaninova, M. A., A. Davydov, O. V. Britanova, M. Shugay, V. Bikos, E. S. Egorov, V. I. Kirgizova, et al. 2016. “High-Quality Full-Length Immunoglobulin Profiling with Unique Molecular Barcoding.” *Nature Protocols* 11 (9): 1599–1616. https://doi.org/10.1038/nprot.2016.093.

Turtle, Cameron J., Laïla-Aïcha Hanafi, Carolina Berger, Theodore A. Gooley, Sindhu Cherian, Michael Hudecek, Daniel Sommermeyer, et al. 2016. “CD19 CAR–T Cells of Defined CD4+:CD8+ Composition in Adult B Cell ALL Patients.” *The Journal of Clinical Investigation* 126 (6): 2123–38. https://doi.org/10.1172/JCI85309.

Vander Heiden, Jason A., Gur Yaari, Mohamed Uduman, Joel N. H. Stern, Kevin C. O’Connor, David A. Hafler, Francois Vigneault, and Steven H. Kleinstein. 2014. “PRESTO: A Toolkit for Processing High-Throughput Sequencing Raw Reads of Lymphocyte Receptor Repertoires.” *Bioinformatics* 30 (13): 1930–32.

Journal Pre-proof

https://doi.org/10.1093/bioinformatics/btu138.

Vazquez-Lombardi, Rodrigo, Johanna S. Jung, Florian Bieberich, Edo Kapetanovic, Erik Aznauryan, Cédric R. Weber, and Sai T. Reddy. 2020. “CRISPR-Targeted Display of Functional T Cell Receptors Enables Engineering of Enhanced Specificity and Prediction of Cross-Reactivity.” bioRxiv. https://doi.org/10.1101/2020.06.23.166363.

Venturi, Vanessa, Máire F. Quigley, Hui Yee Greenaway, Pauline C. Ng, Zachary S. Ende, Tina McIntosh, Tedi E. Asher, et al. 2011. “A Mechanism for TCR Sharing between T Cell Subsets and Individuals Revealed by Pyrosequencing.” *The Journal of Immunology* 186 (7): 4285–94. https://doi.org/10.4049/jimmunol.1003898.

Volden, Roger, and Christopher Vollmers. 2020. “Highly Multiplexed Single-Cell Full-Length CDNA Sequencing of Human Immune Cells with 10X Genomics and R2C2.” *bioRxiv*, January, 2020.01.10.902361. https://doi.org/10.1101/2020.01.10.902361.

Voss, James E, Alicia Gonzalez-Martin, Raiees Andrabi, Roberta P Fuller, Ben Murrell, Laura E McCoy, Katelyn Porter, et al. 2019. “Reprogramming the Antigen Specificity of B Cells Using Genome-Editing Technologies.” Edited by Tomohiro Kurosaki, Gisela Storz, and Tomoharu Yasuda. *ELife* 8 (January): e42995. https://doi.org/10.7554/eLife.42995.

Wagner, Ellen K., Ahlam N. Qerqez, Christopher A. Stevens, Annalee W. Nguyen, George Delidakis, and Jennifer A. Maynard. 2019. “Human Cytomegalovirus-Specific T Cell Receptor Engineered for High Affinity and Soluble Expression Using Mammalian Cell Display.” *Journal of Biological Chemistry*, February, jbc.RA118.007187. https://doi.org/10.1074/jbc.RA118.007187.

Wang, Chunlin, Catherine M. Sanders, Qunying Yang, Harry W. Schroeder, Elijah Wang, Farbod Babrzadeh, Baback Gharizadeh, et al. 2010. “High Throughput Sequencing Reveals a Complex Pattern of Dynamic Interrelationships among Human T Cell Subsets.” *Proceedings of the National Academy of Sciences* 107 (4): 1518–23. https://doi.org/10.1073/pnas.0913939107.

Warren, René L., J. Douglas Freeman, Thomas Zeng, Gina Choe, Sarah Munro, Richard Moore, John R. Webb, and Robert A. Holt. 2011. “Exhaustive T-Cell Repertoire Sequencing of Human Peripheral Blood Samples Reveals Signatures of Antigen Selection and a Directly Measured Repertoire Size of at Least 1 Million Clonotypes.” *Genome Research* 21 (5): 790–97. https://doi.org/10.1101/gr.115428.110.

Watson, C. T., and F. Breden. 2012. “The Immunoglobulin Heavy Chain Locus: Genetic Variation, Missing Data, and Implications for Human Disease.” *Genes & Immunity* 13 (5): 363–73. https://doi.org/10.1038/gene.2012.12.

Webber, Beau R., Cara-lin Lonetree, Mitchell G. Kluesner, Matthew J. Johnson, Emily J. Pomeroy, Miechaleen D. Diers, Walker S. Lahr, et al. 2019. “Highly Efficient Multiplex Human T Cell Engineering without Double-Strand Breaks Using Cas9 Base Editors.” *Nature Communications* 10 (1): 5222. https://doi.org/10.1038/s41467-019-13007-6.

Weber, Cédric R., Rahmad Akbar, Alexander Yermanos, Milena Pavlović, Igor Snapkov, Geir K. Sandve, Sai T. Reddy, and Victor Greiff. 2020. “ImmuneSIM: Tunable Multi-Feature Simulation of B- and T-Cell Receptor Repertoires for Immunoinformatics Benchmarking.” *Bioinformatics* 36 (11): 3594–96. https://doi.org/10.1093/bioinformatics/btaa158.

Weinstein, Joshua A., Ning Jiang, Richard A. White, Daniel S. Fisher, and Stephen R. Quake. 2009. “High-Throughput Sequencing of the Zebrafish Antibody Repertoire.” *Science* 324 (5928): 807–10. https://doi.org/10.1126/science.1170020.

Wen, Wen, Wenru Su, Hao Tang, Wenqing Le, Xiaopeng Zhang, Yingfeng Zheng, Xiuxing Liu, et al. 2020. “Immune Cell Profiling of COVID-19 Patients in the Recovery Stage by Single-Cell Sequencing.” *Cell Discovery* 6 (1): 1–18. https://doi.org/10.1038/s41421-020-0168-9.

Wenger, Aaron M., Paul Peluso, William J. Rowell, Pi-Chuan Chang, Richard J. Hall, Gregory T. Concepcion, Jana Ebler, et al. 2019. “Accurate Circular Consensus Long-Read Sequencing Improves Variant Detection and Assembly of a Human Genome.” *Nature Biotechnology* 37 (10): 1155–62. https://doi.org/10.1038/s41587-019-0217-9.

Widrich, Michael, Bernhard Schäfl, Milena Pavlović, Geir Kjetil Sandve, Sepp Hochreiter, Victor Greiff, and Günter Klambauer. 2020. “DeepRC: Immune Repertoire Classification with Attention-Based Deep Massive Multiple Instance Learning.” *bioRxiv*, April, 2020.04.12.038158. https://doi.org/10.1101/2020.04.12.038158.

Wu, Chung-An M., Theodore L. Roth, Yuriy Baglaenko, Dario M. Ferri, Patrick Brauer, Juan Carlos Zuniga-Pflucker, Kristina W. Rosbe, Joan E. Wither, Alexander Marson, and Christopher D. C. Allen. 2018. “Genetic Engineering in Primary Human B Cells with CRISPR-Cas9 Ribonucleoproteins.” *Journal of Immunological Methods* 457: 33–40. https://doi.org/10.1016/j.jim.2018.03.009.

Wu, Xueling, Zhenhai Zhang, Chaim A. Schramm, M. Gordon Joyce, Young Do Kwon, Tongqing Zhou, Zizhang Sheng, et al. 2015. “Maturation and Diversity of the VRC01-Antibody Lineage over 15 Years of Chronic HIV-1 Infection.” *Cell* 161 (3): 470–85. https://doi.org/10.1016/j.cell.2015.03.004.

Yaari, Gur, and Steven H. Kleinstein. 2015. “Practical Guidelines for B-Cell Receptor Repertoire Sequencing Analysis.” *Genome Medicine* 7 (1): 121. https://doi.org/10.1186/s13073-015-0243-2.

Ye, Jian, Ning Ma, Thomas L. Madden, and James M. Ostell. 2013. “IgBLAST: An Immunoglobulin Variable Domain Sequence Analysis Tool.” *Nucleic Acids Research* 41 (W1): W34–40. https://doi.org/10.1093/nar/gkt382.

Yermanos, Alexander Dimitri, Andreas Kevin Dounas, Tanja Stadler, Annette Oxenius, and Sai T. Reddy. 2018. “Tracing Antibody Repertoire Evolution by Systems Phylogeny.” *Frontiers in Immunology* 9. https://doi.org/10.3389/fimmu.2018.02149.

Yermanos, Alexander, Victor Greiff, Nike Julia Krautler, Ulrike Menzel, Andreas Dounas, Enkelejda Miho, Annette Oxenius, Tanja Stadler, and Sai T. Reddy. 2017. “Comparison of Methods for Phylogenetic B-Cell Lineage Inference Using Time-Resolved Antibody Repertoire Simulations (AbSim).” *Bioinformatics* 33 (24): 3938–46. https://doi.org/10.1093/bioinformatics/btx533.

Zehn, Dietmar, Sarah Y. Lee, and Michael J. Bevan. 2009. “Complete but Curtailed T Cell Response to Very Low

Journal Pre-proof

Affinity Antigen." *Nature* 458 (7235): 211–14. https://doi.org/10.1038/nature07657.
Zhang, Hongyi, Longchao Liu, Jian Zhang, Jiahui Chen, Jianfeng Ye, Sachet Shukla, Jian Qiao, et al. 2020. "Investigation of Antigen-Specific T-Cell Receptor Clusters in Human Cancers." *Clinical Cancer Research* 26 (6): 1359–71. https://doi.org/10.1158/1078-0432.CCR-19-3249.
Zhao, Yunqian, Phuong Nguyen, Peter Vogel, Bofeng Li, Lindsay L. Jones, and Terrence L. Geiger. 2016. "Autoimmune Susceptibility Imposed by Public TCRβ Chains." *Scientific Reports* 6 (1): 37543. https://doi.org/10.1038/srep37543.
Zheng, Grace X. Y., Jessica M. Terry, Phillip Belgrader, Paul Ryvkin, Zachary W. Bent, Ryan Wilson, Solongo B. Ziraldo, et al. 2017. "Massively Parallel Digital Transcriptional Profiling of Single Cells." *Nature Communications* 8 (1): 14049. https://doi.org/10.1038/ncomms14049.
Zhou, Chen, Frederick W. Jacobsen, Ling Cai, Qing Chen, and David Shen. 2010. "Development of a Novel Mammalian Cell Surface Antibody Display Platform." *MAbs* 2 (5): 508–18. https://doi.org/10.4161/mabs.2.5.12970.
Zhou, Yuanping, Junjie Wang, Ivan Zhou, Haibo Lou, Chang-Zheng Li, Zhen-Rui Chen, Zhe-Huan Zhang, et al. 2013. "Simultaneous Expression of Displayed and Secreted Antibodies for Antibody Screen." *PLOS ONE* 8 (11): e80005. https://doi.org/10.1371/journal.pone.0080005.
Zhu, Jiang, Gilad Ofek, Yongping Yang, Baoshan Zhang, Mark K. Louder, Gabriel Lu, Krisha McKee, et al. 2013. "Mining the Antibodyome for HIV-1–Neutralizing Antibodies with next-Generation Sequencing and Phylogenetic Pairing of Heavy/Light Chains." *Proceedings of the National Academy of Sciences* 110 (16): 6470–75. https://doi.org/10.1073/pnas.1219320110.
Zvyagin, Ivan V., Vasily O. Tsvetkov, Dmitry M. Chudakov, and Mikhail Shugay. 2020. "An Overview of Immunoinformatics Approaches and Databases Linking T Cell Receptor Repertoires to Their Antigen Specificity." *Immunogenetics* 72 (1): 77–84. https://doi.org/10.1007/s00251-019-01139-4.

Figures 1-4 were created with BioRender.com.

Journal Pre-proof

Journal Pre-proof

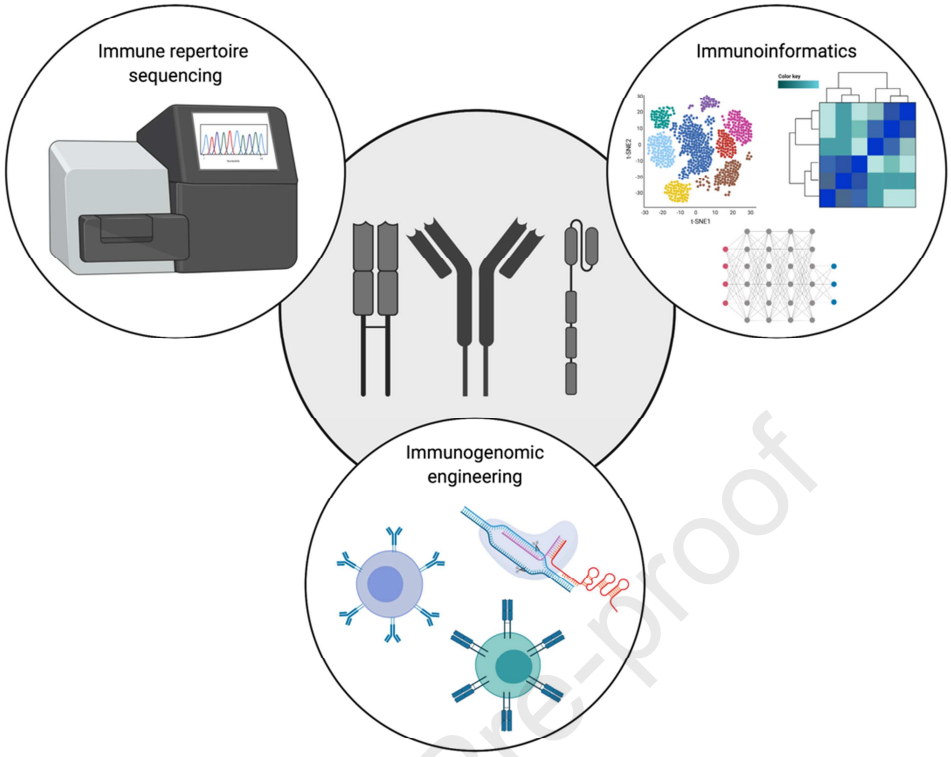


Figure 1: Reading, writing and editing adaptive immunity with state-of-the-art methods such as immune repertoire sequencing, immunoinformatics and immunogenomic engineering. The combination of these systems and synthetic immunology tools offers the possibility to answer long-standing immunological questions and to reprogram and engineer adaptive immunity.

Journal Pre-proof

Journal Pre-proof

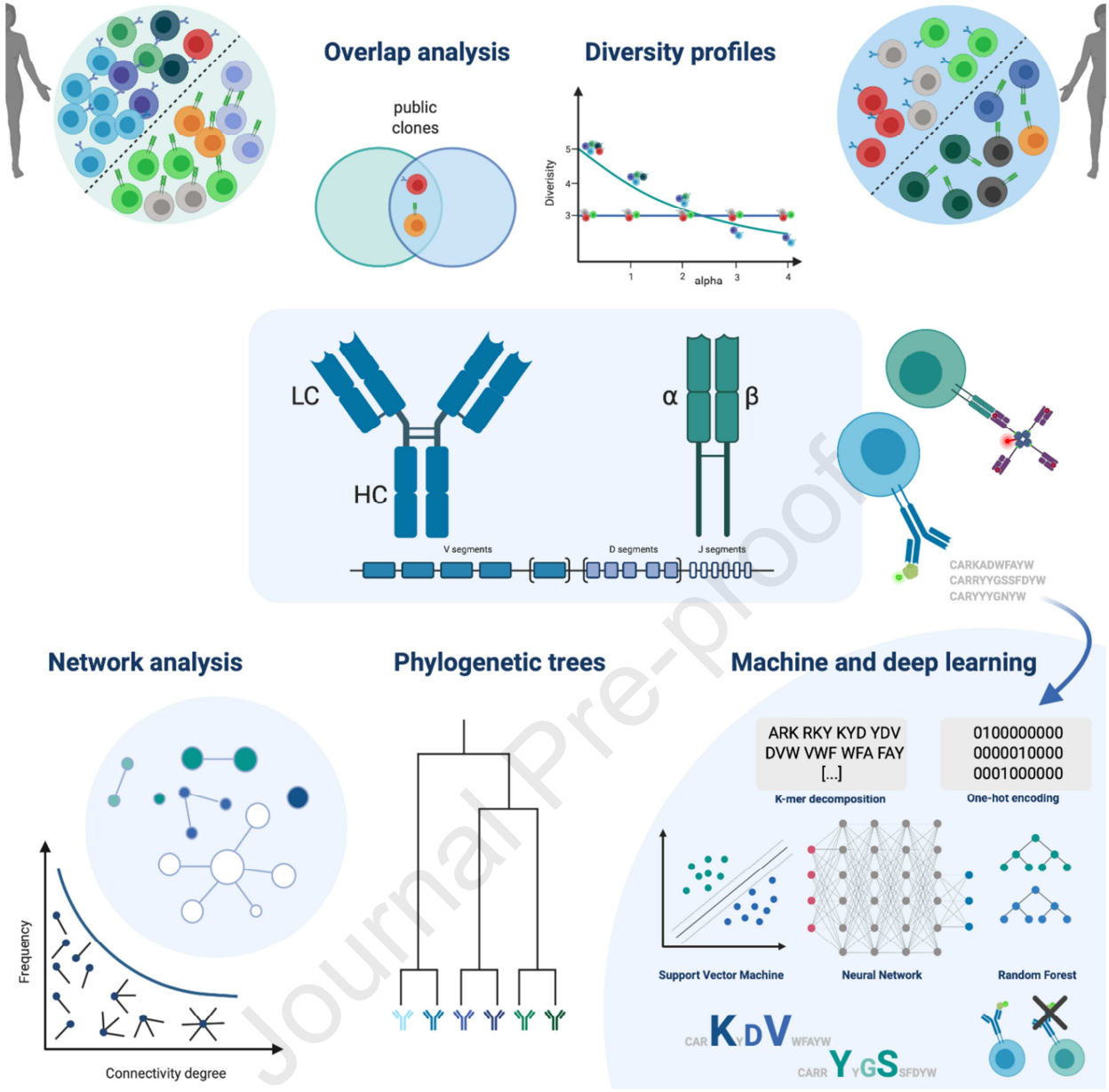


Figure 2: Immunoinformatics analysis and visualization of immune repertoires. Each organism contains a highly unique B and T cell repertoire, generated by V(D)J recombination, whose composition, dynamics and evolution gets shaped by extrinsic (i.e., pathogens, vaccination) and intrinsic forces (i.e., tumors, autoimmunity, aging). Diversity and clonal expansion within immune repertoires can be captured by e.g. Hill-based diversity profiles which include a continuum of diversity indices (clones get weighted by the parameter alpha – the higher alpha, the more weight is put on high-frequency clones). Clonal relationships and organization within immune repertoires can be inferred by network-based analysis with nodes representing unique clones, which are connected by edges depending on their sequence similarity. Connectivity degree distributions quantify the number of connections each node exhibits, thus, providing insights into clonal expansion profiles. Clonal evolution analysis is in particular useful for B cells which undergo somatic hypermutation upon antigen recognition. Phylogenetic trees allow the reconstruction of both ancestral and intermediate relationships between B cell clonal sequences, thus, facilitating the identification of affinity maturated clonal variants. Increasing efforts are put into detection of convergent immune responses upon antigenic challenge, e.g. via clonal overlap analysis for the discovery of shared clones across individuals (e.g. via Venn diagrams) or by the identification of shared sequence patterns. Machine- and deep learning models (e.g. support vector machine, random forest, or neural networks) can be used for the classification of immune receptor sequences into antigen-binder and non-binder. For instance, these models can be trained with immune receptor sequences of known antigen-specificity which get encoded as numerical representations by methods such as k-mer decomposition or one-hot encoding.

Journal Pre-proof

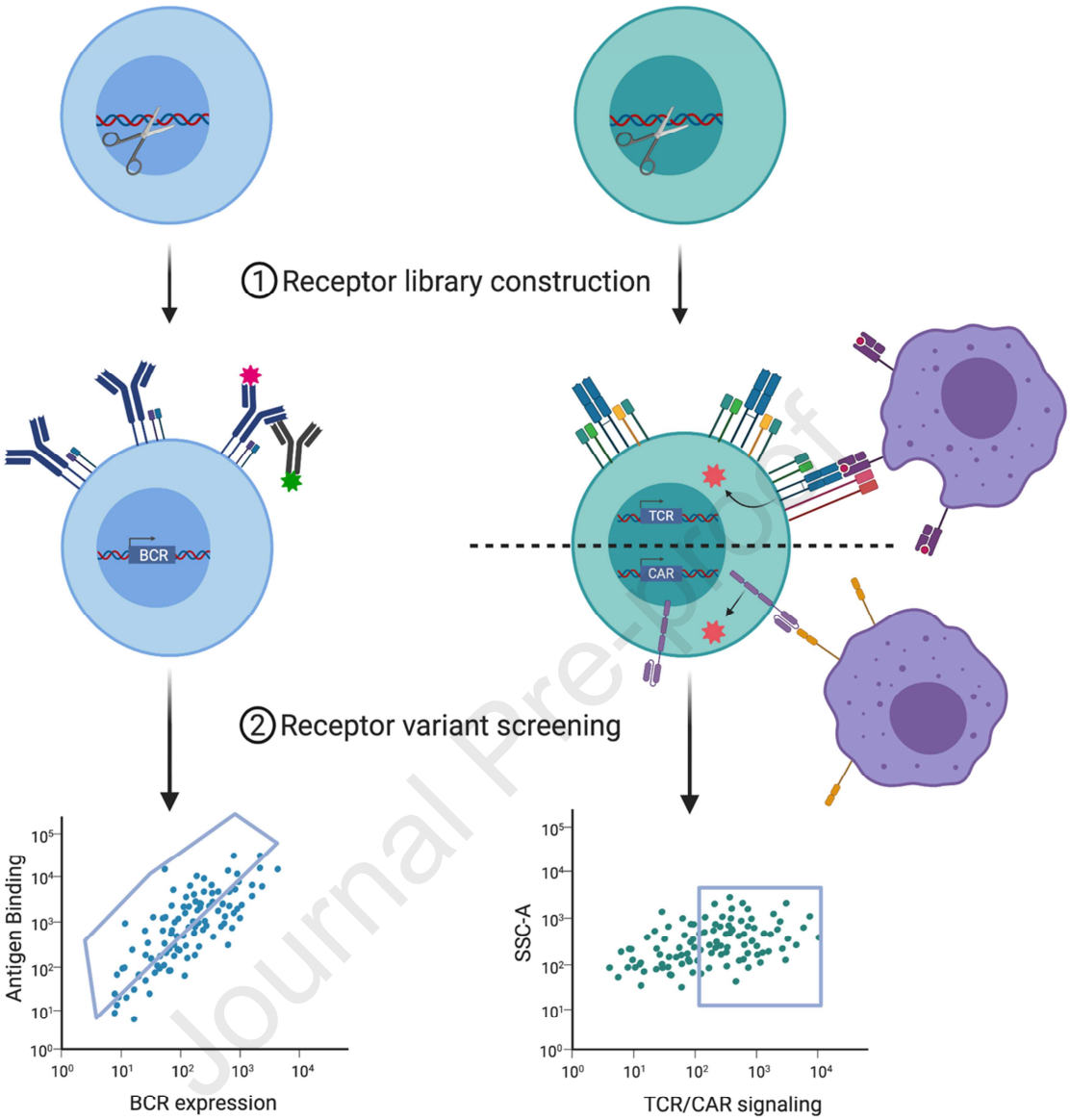


Figure 3: Synthetic mammalian display platforms for the discovery and engineering of immune receptors. The construction of large mammalian receptor libraries is becoming possible due to novel genome editing tools like CRISPR/Cas9 that can be used to introduce diversity in the immune receptor locus of B or T cell lines. These cell libraries can be screened using high-throughput methods to assess affinity, specificity and functional activity. For the selection of immune receptors with a desired affinity and specificity, fluorescence-activated cell sorting (FACS) and magnetic-activated cell sorting (MACS) can be utilized. Functional receptor activity on the other hand can be assessed via co-culture systems that include a target presenting cell line that activates the receptor signaling pathway of the library cell line followed by subsequent screening based on signaling activity via FACS.

Journal Pre-proof

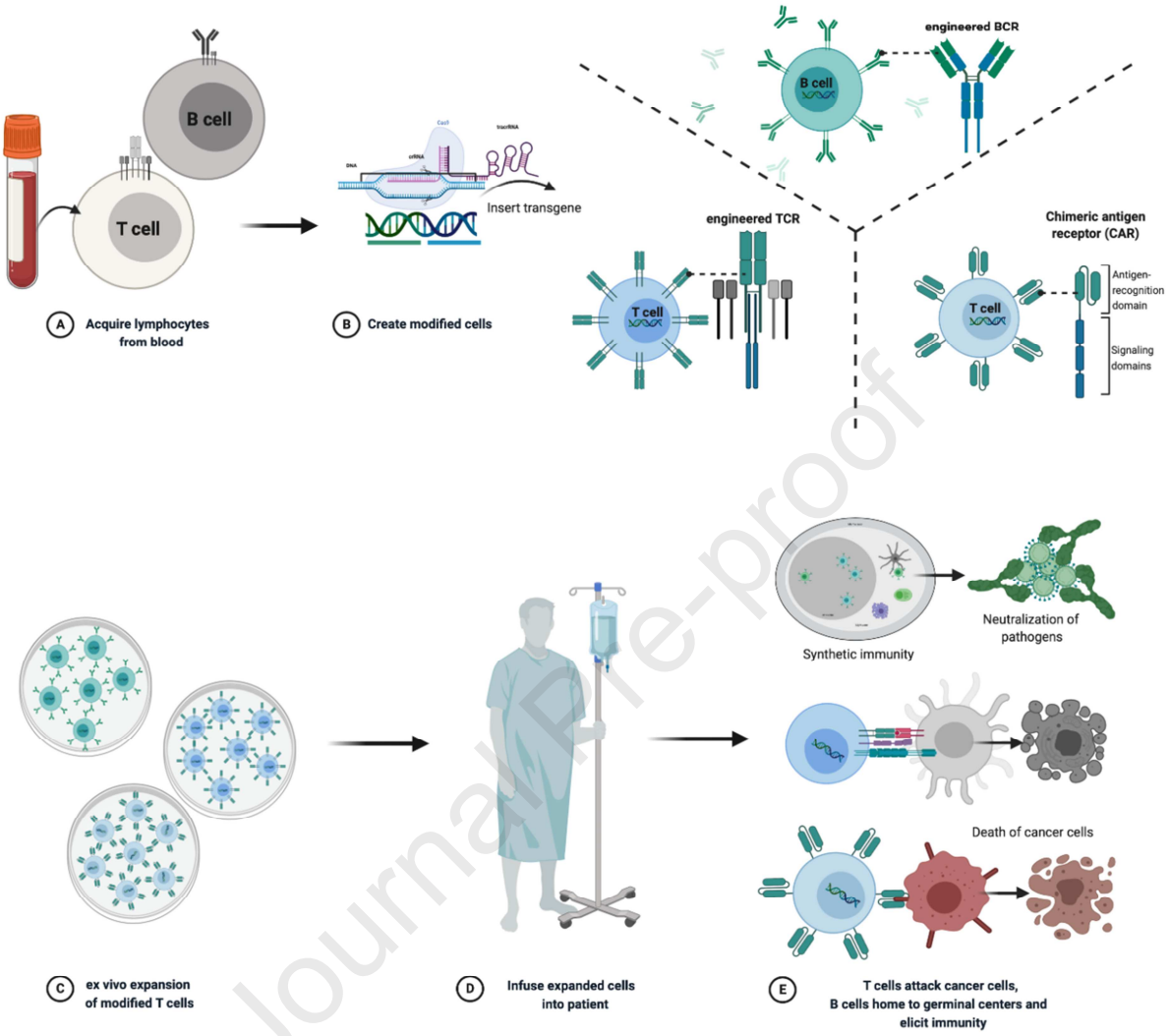
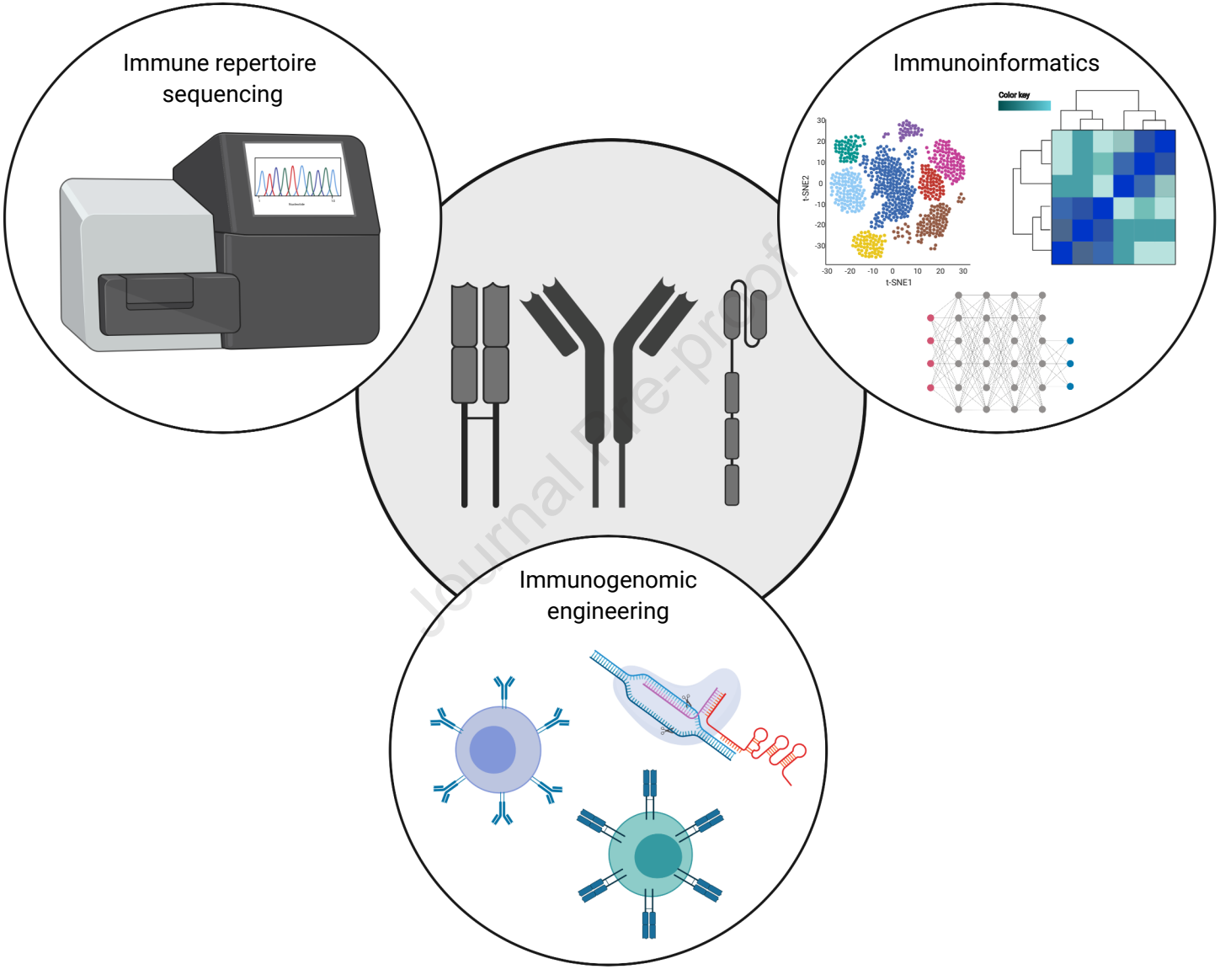
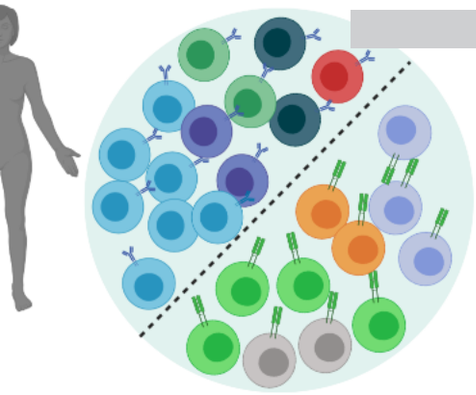


Figure 4: Workflow of adoptive cellular therapy. (A) First, B or T cells are isolated from whole blood or PBMCs. Pure lymphocyte populations are (B) re-engineered using CRISPR systems. (C) After ex vivo expansion, (D) autologous cells are infused into the patient, (E) where they exert their effector functions either by attacking cancer cells (modified T cells) or by homing to lymphoid tissue to differentiate and secrete antibodies that confer a synthetic immunity towards pathogens.

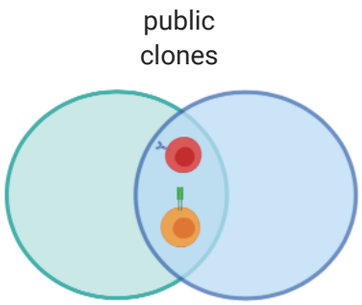
Journal Pre-proof



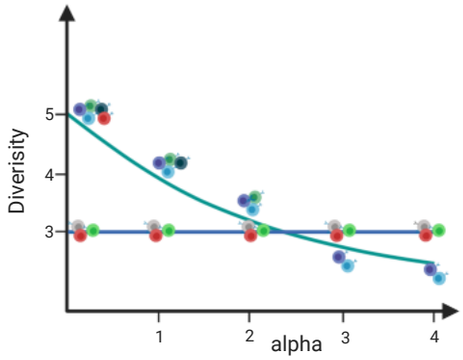
Journal Pre-proof

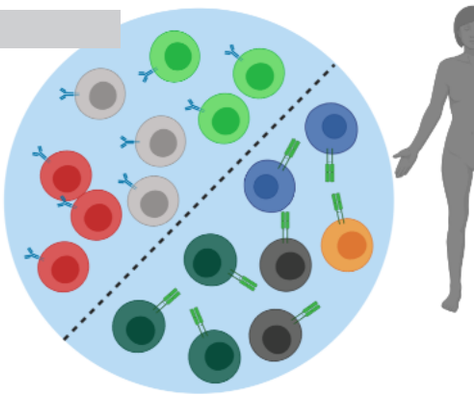


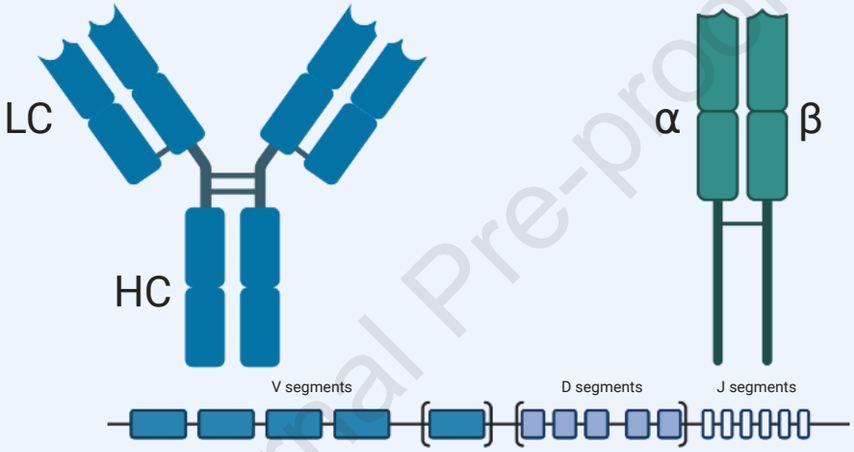
Overlap analysis

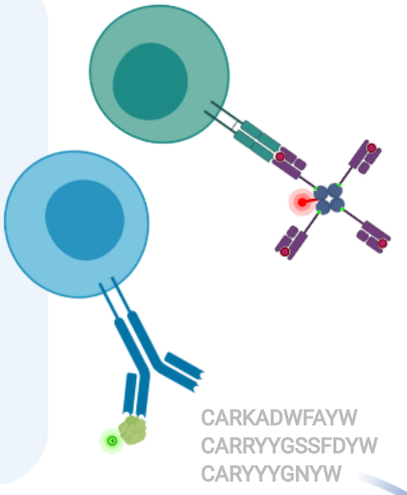


Diversity profiles

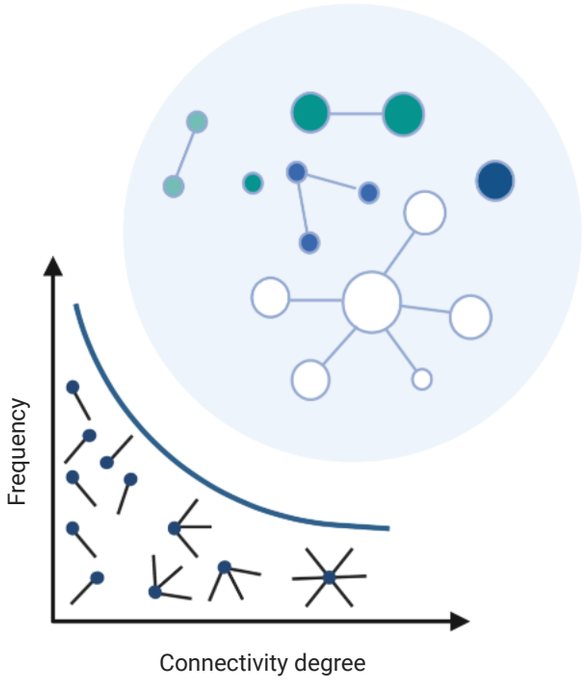




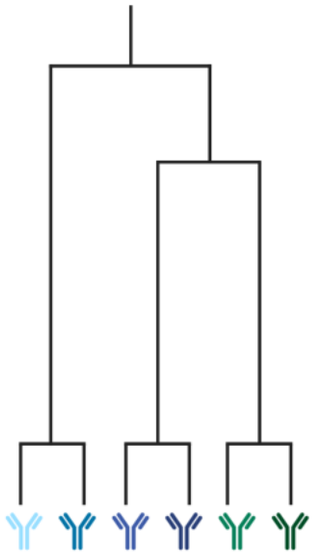




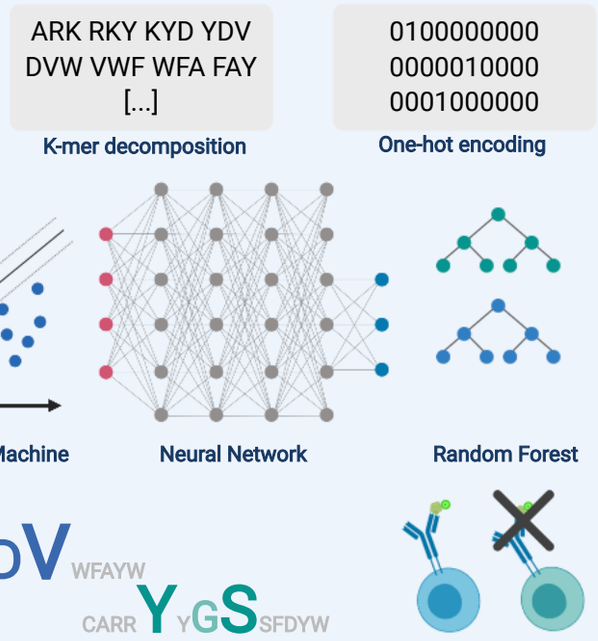
Network analysis



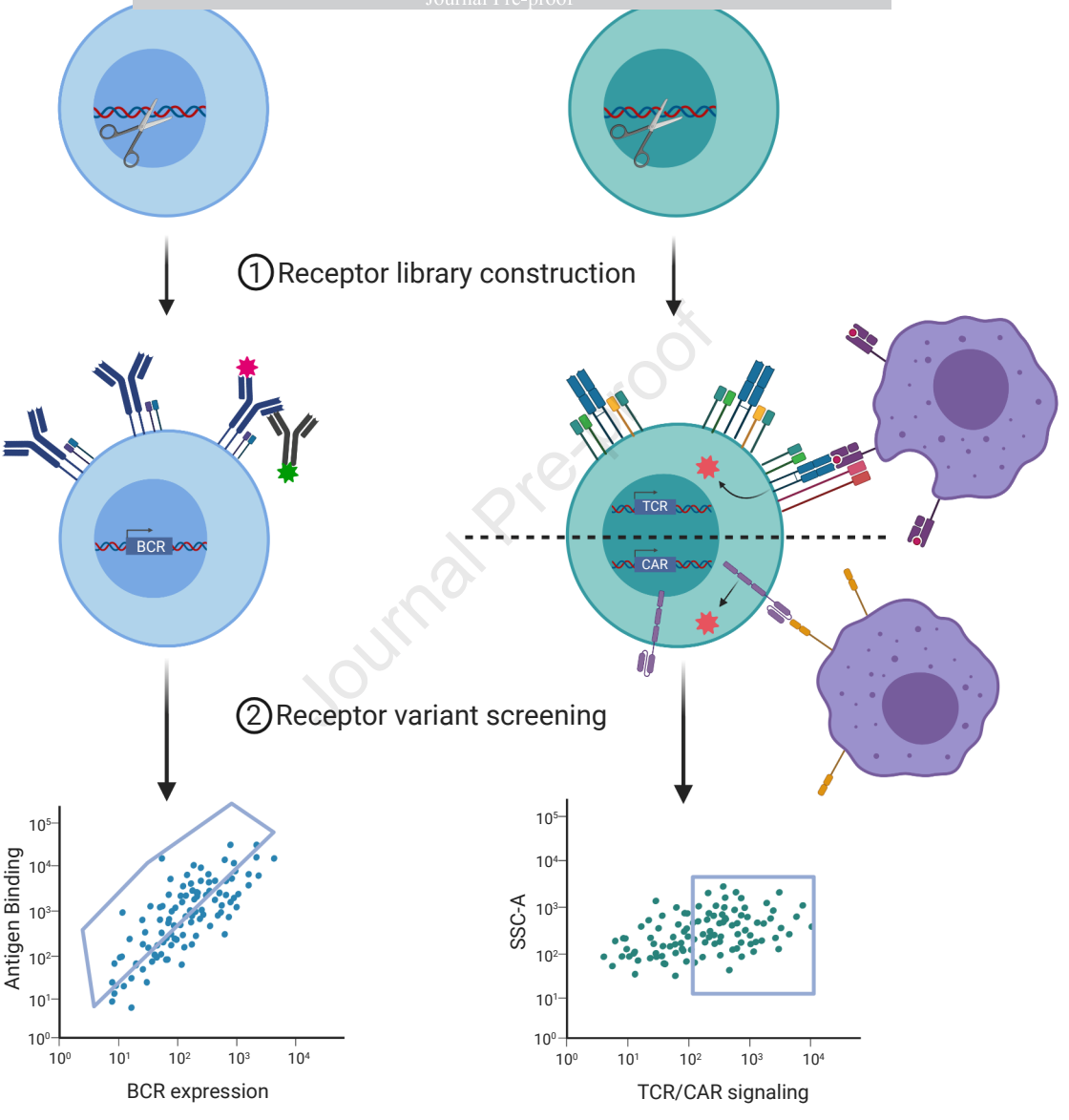
Phylogenetic trees



Machine and deep learning

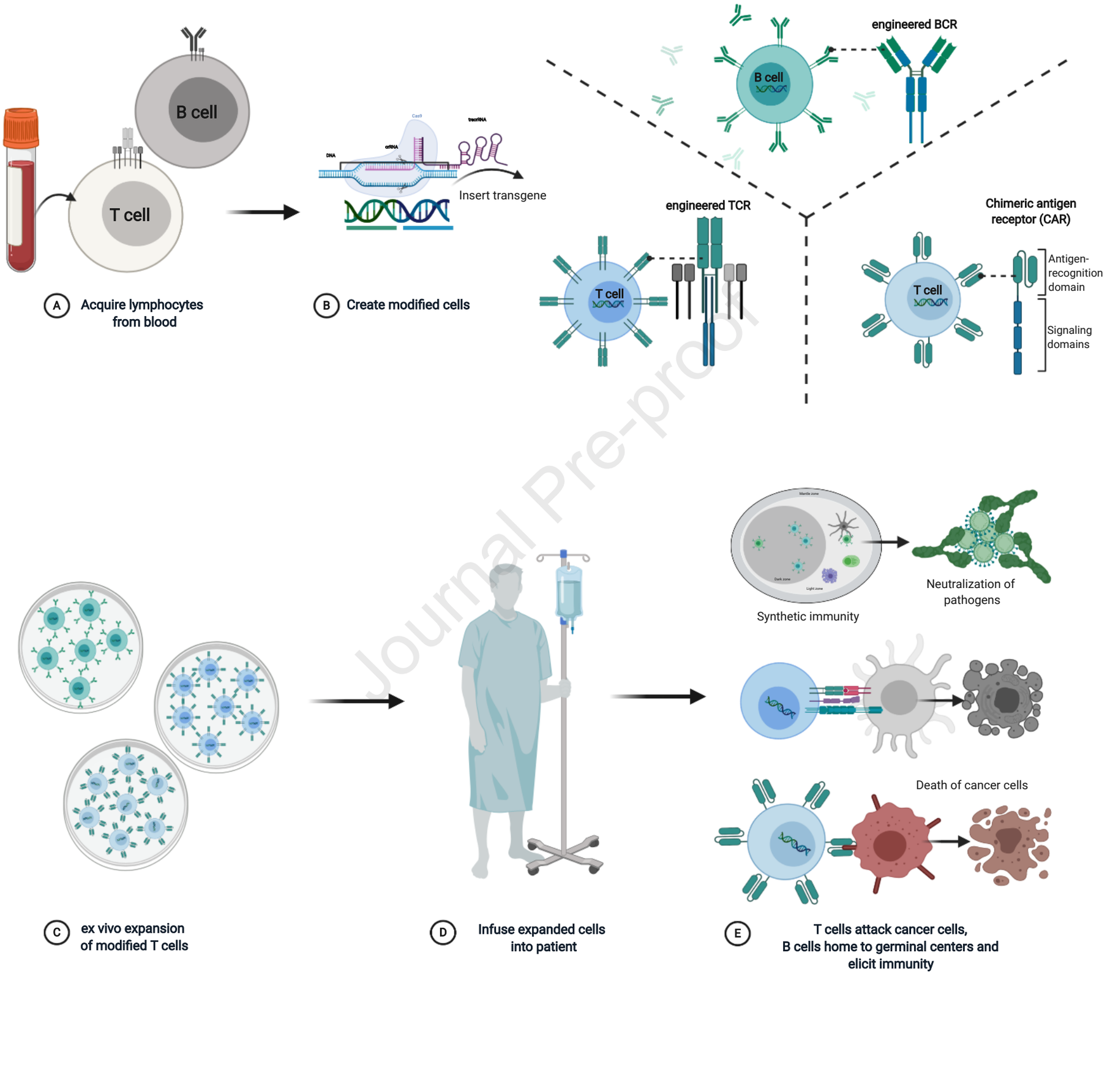


Journal Pre-proof



Journal Pre-proof

Journal Pre-proof



Journal Pre-proof