

The Adaptive Immune Receptors in a Big Data World

Brandon J. DeKosky^{*, †, ‡, §}

^{*} The Ragon Institute of Mass General, MIT, and Harvard, Cambridge, MA, USA

[†] Department of Chemical Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA

[‡] The Koch Institute for Integrative Cancer Research, Massachusetts Institute of Technology, Cambridge, MA, USA

[§]Correspondence:

Correspondence addressed to dekosky@mit.edu

Running Title

Adaptive Immune Receptors in a Big Data World

Abstract

Antibodies and T cell receptors are the molecular basis of immune memory, and also have become the foundation for generations of clinically successful biologics. Decades of research have established a broad range of systems to observe, identify, discover, and optimize the adaptive immune receptors, from both natural and synthetic sources. Recent advances in high-throughput display, next-generation sequencing analysis, and machine learning data mining are accelerating our capabilities to characterize immune receptor repertoires, including their functional properties. This commentary discusses the unique aspects that guide modern immune receptor studies, with an eye toward the near-term horizons in large functional dataset collection and analysis.

Key words

Antibody, T cell receptor, single-cell analysis, immune repertoire, bioinformatics

The Initial Frameworks

Analysis of single adaptive immune receptors was launched in the 1970's when Köhler and Milstein reported the hybridoma method, which for the first time enabled scientists to capture and harness monoclonal antibody and T cell receptor (TCR) variants (1). The isolation of monoclonal antibodies and monoclonal TCRs led to an explosion of scientific knowledge along with translational uses for adaptive immune receptors. Hybridoma technology, and later single-cell RT-PCR (2), enabled the first molecular glimpses of the vast sets of unique specificities that comprise our immune repertoires. Studies explored immunogenetics like V-gene usage, heavy:light and alpha:beta gene pairing, and somatic hypermutation. Statistics on the scale of hundreds of clones revealed gene associations with particular immune response features (3, 4).

Hybridoma and single-cell sequencing provided a limited view of large immune repertoires, while the technologies of PCR, cloning, and especially protein display joined forces to rapidly accelerate progress in antibody and TCR affinity engineering against to target proteins (5). However, it was not until the 2000's when next-generation sequencing technologies were released that more comprehensive data could be obtained for adaptive immune repertoires (6). By using next-generation sequencing, we began for the first time to discern individual adaptive immune receptors as a part of large adaptive immune response networks with precise molecular clarity. These technical advances of recent decades have now brought our community into the current world of big data for adaptive immune receptors.

Parallels between Antibodies and TCRs

B cells and T cells work synergistically *in vivo*, and many common features are shared amongst their immune receptors. Chief among these are the features of central tolerance, clonal selection and expansion *in vivo*, and the need to balance sufficient antigen specificity while minimizing cross-reactivity. A key difference between antibodies and TCRs are their mechanisms of action. Antibody molecules are secreted by B cells, and both the soluble protein and the cells are the effector agents. In contrast, T cell receptor functions are intricately linked to the T cell that encodes them, and the T

cell is the actual effector agent of TCR-driven responses. Affinity and multivalency dynamics of these two classes of immune receptors are also profoundly different. Antibodies are expressed as various molecular classes that present different numbers of binding arms: membrane-bound B cell receptors (BCRs) are multivalent, soluble IgG and IgE are bivalent, soluble IgA is bi- or tetravalent, and soluble IgM is pentavalent. Somatic hypermutation plays an important role for antibodies to establish sufficient affinities to function in soluble, especially bivalent, formats. In contrast, T cells can express thousands of membrane-bound TCRs on their surface, and TCR recognition often (but not always) relies on multivalency. These low-affinity, multi-avid interactions of TCRs are ideally suited to recognize the subtle differences between distinct peptides presented by the same MHC molecule, whereas high-affinity antibody interactions are well-suited to the soluble nature of antibodies that bind tightly to foreign antigens and traffic them to appropriate effector cells and proteins within the body. Soluble antibodies have some activation differences based on glycosylation states, but still the antibody protein is sufficient to effectuate an immune response. In contrast, the response cascade triggered by a surface-expressed TCR is highly dependent on cellular state. T cell control mechanisms include numerous regulatory checkpoints and exhaustion pathways that help to mitigate the high degree of cross-reactivity that most TCRs can display to related peptides, including self peptides. These biological, functional, and structural differences between antibodies and TCRs are summarized in **Figure 1**. These differences lead to divergence in the experimental platforms and technologies needed to understand, evaluate, and manipulate adaptive immunity via the humoral vs. cellular immune compartments.

The Two-Chain Nature of Antibodies and TCRs

The common two-chain nature of adaptive immune receptors (heavy and light chain for antibodies; alpha and beta chain for T cell receptors) provides a diversity generation mechanism that allows immune receptors to recognize varied binding targets. Because the heavy:light and alpha:beta chains are encoded on separate mRNAs and derived from distinct chromosomes inside the cell, apart from being present inside the same cell membranes there is no physical connection between

complementary chain pairs until quaternary structures are formed post-translation. Due to the difficulty of sequencing multiple genes on separate chromosomes together, the earliest next-generation sequencing technologies analyzed heavy, light, alpha, and beta chains separately (6). However, given the importance of two-chain immune receptors in mammals, repertoire-scale technologies need to collect information on both chains together to fully analyze and understand immune repertoire properties (7–9). In developing antibody repertoires, it appears so far that most heavy chain V-genes can successfully pair with most light chain V-genes (4, 10). Then, somatic hypermutation enhances antibody affinity both through antigen interactions, and also often by refining the heavy:light interfaces for improved recognition (11, 12). Recent evidence also demonstrates that antibodies across multiple individuals often utilize the same heavy:light pairings when recognizing common antigens, suggesting a functional and structural convergence of recognition modes (13). Similar features can be observed in T cell receptors that target the same pMHC (14). Because T cells do not somatically hypermutate, the specific alpha:beta gene combinations likely provide an even greater role in antigen recognition than has currently been observed for B cells (15). The development of technologies to efficiently analyze paired heavy:light or paired alpha:beta chains are discussed in detail in the following section.

Determining the targets of antibodies and TCRs helps to reveal the workings of adaptive immunity, and also can identify immune receptors or drug candidates that efficiently target a disease. Immune receptors have been discovered from individuals that can efficiently prevent viral infections or cancers, or that exacerbate diseases like autoimmunity. New vaccines or molecular therapeutics can then be devised to treat or cure based if immune receptor analyses reveal new molecular mechanisms of a disease. In some cases, a functional study of immune receptors helps identify target antigens, which is becoming especially common for T cell receptors in cancer (16). Antibody and TCR discovery is also commonly performed to identify new drug candidates. While native chain pairing information is very helpful for antibody drug discovery applications, it is not strictly required, and many antibodies have been discovered through phage display of randomly paired heavy and light chains. Still, native heavy:light chain pairing is often superior to random pairings, including both for performance and stability features (17). For TCRs, native alpha:beta pairings are extremely important

to reduce the risks of off-target therapeutic toxicity, which was observed in some early clinical studies of TCR-based T cell (TCR-T) therapy (18). Using natively paired alpha:beta TCRs as immune drugs can reduce toxicity risks because those TCRs have gone through central tolerance and are unlikely to present as high toxicity risks as artificial alpha:beta pairs, especially for autologous TCR-T applications.

Recent technologies have dramatically expanded the scope of paired alpha:beta and heavy:light receptors that can be sequenced, cloned, and analyzed. Still, it is important to note that we still face severe biological and technical shortcomings for immune repertoire analysis. Biological sampling challenges remain severe, and it is impossible to collect full organism-level tissue samples from human clinical studies. The most advanced large-scale reported technologies for paired immune receptor sequencing accommodate a maximum of $\sim 10^7$ input B or T cells. Even though we can sequence more paired immune receptors than ever before, technical platforms can analyze the $\sim 10^7$ B cells but not the $\sim 10^8$ T cells from a single human blood draw. Still, these numbers remain far from the $\sim 2 \times 10^{11}$ B cells and $\sim 4 \times 10^{11}$ T cells in a single individual (19). Even if we could recover all those cells, and pair all of their immune receptors, the current 2x300 bp sequencing platforms max out around $\sim 4 \times 10^8$ reads per run, and long-read sequencing technologies analyze only around 2.5×10^7 reads per run. Thus, despite tremendous advances in single-cell processing and sequencing platforms, we can still only observe a small fraction of human immune receptors directly. Given the limited observation windows that we have for antibody and TCR repertoires, it is important to collect the most complete information possible from the limited cell types samples we do obtain.

Connecting Sequence to Function

Several approaches have been devised to explore the functional properties of immune receptors, beginning with the hybridoma techniques that are still commonly used today. Single-cell RT-PCR after flow cytometry assisted sorting (FACS) into well plates also matured over decades, and is now a refined protocol (2). These single-cell RT-PCR approaches were transformed by advances in antigen design and development to enable precise identification of desired antigen-specific cells,

including designed recombinant protein antigens to identify specific B cells (20) and multivalent pMHC constructs to identify TCRs (21). More recently, antigen barcoding methods have been established to more easily detect immune receptor sequences and antigen binding profiles concurrently (22–24), and these techniques also can analyze full transcriptomes and link B- or T-cell state with antigen specificity. One limitation of these primary immune cell-based approaches is that the cells are non-renewable, and thus each antigen specificity measurement is subject to single-cell measurement error.

An important alternative to analyzing primary immune cells is to use a display platform to study immune receptor function. Display platforms connect an RNA or DNA sequence to the function of the protein that it encodes, with many technologies established over the years that are useful for immune repertoire screening applications. For antibodies, a broad variety of display platforms have been developed including phage display, mRNA display, bacterial display, and yeast display (25). Several display formats have also been established for antibody display, including single-chain fragment variable (scFv), fragment antigen binding (Fab), and full IgG formats; each of these are successfully used across multiple display platforms. Each platform offers unique advantages and disadvantages. Phage display has proven to be a workhorse technology for the biomedical community, and phage display supports remarkable library sizes with extremely high diversity and screening throughput. However, phage display can be limited by substantial replication bias across rounds, the limitations of protein expression in bacterial systems, is not compatible with flow cytometry, and can only be used for affinity analyses (not functional analyses). Yeast display has been growing in use due to its robust nature, eukaryotic expression systems, low cost and low replication bias across rounds, use in flow cytometry, and ample library sizes, although yeast are generally limited to proteins that do not require mammalian post-translational modifications and are mostly used for affinity-based screens. Most recently, mammalian display has been emerging as a popular platform. Technologies to improve library construction in mammalian cells have accelerated growth in mammalian display, which has the capacity to express mammalian proteins in their native formats while also being compatible with functional screening systems (26). These major advantages outweigh the higher cost and complexity

of mammalian cell screening for many groups, and thus mammalian display screening platforms are anticipated to become even more widely used in the coming years.

T cell receptors can also be incorporated into display platforms and screened in varied formats, although with somewhat more difficulty than for antibodies. Antibodies evolved as secreted single-molecule, soluble formats, and the structural configuration for soluble expression which supports their efficient use across many systems and fusion protein formats. In contrast, T cell receptors are always membrane-bound in nature, and they are a little more difficult to express than antibodies but still achieve successful display in phage, yeast, and mammalian cells (27). Because mammalian gene manipulation have advanced quickly over the past few years, it can be anticipated that the use of mammalian platforms for both antibody and TCR display will continue to expand.

Our group has established renewable display library screening techniques to circumvent the limitations of functionally mapping immune repertoires using primary cells. Across several studies, we generate paired chain gene libraries and clone them into display platforms to enable repeated screening experiments (8, 10, 28–33). An advantage of renewable libraries is that they provide the ability to collect accurate functional data by analyzing repertoires with statistical coverage (e.g., 10-fold more tests than the known library size), which cannot be achieved with primary cells. These display-based assays can bin immune repertoires based on performance metrics like affinity, which can require testing different concentrations of antigen to evaluate across an entire affinity range, and likewise cannot be performed efficiently with primary cells. Renewable screening platforms can also evaluate cross-reactivity to antigen panels (30, 34–36), which is more difficult with primary cells because the multiple antigens compete with each other for cell surface binding. A renewable library screening approach also enables interrogation of cell populations that do not express high levels of immune receptors. Important examples are plasma cells from spleen and bone marrow, which secrete the majority of antibodies in circulation but do not express surface B cell receptors and therefore are not compatible with standard antigen-bait approaches (32). Finally, a critical advantage of renewable formats is that the immune receptors can also be screened for functional activity (e.g., pathogen neutralization, activation against a cancer cell, etc.), which cannot be achieved at a library scale with

primary cells and antigen-bait approaches (37). Collecting paired immune receptor sequence libraries and cloning into display libraries thus enables direct analysis of the functional performance of immune receptors *en masse* – bringing us potentially one step closer to a more complete, large-data understanding of immune receptor repertoires.

Challenges and Opportunities in Functional Antibody Repertoire Analysis

While antigen affinity is important and is often used as a screening parameter, an antibody's affinity is rarely the functional drug property. For example in antiviral antibodies, virus neutralization *in vitro* is often considered the most important functional property for passive antibody prophylaxis and to evaluate vaccine responses. Other antibody effector functions like antibody-dependent cellular cytotoxicity (ADCC), antibody-dependent cellular phagocytosis (ADCP), and complement deposition can also be important to protect against infections, especially for viruses where breadth and potency trade-offs exist and non-neutralizing antibodies can contribute substantially to immune protection (e.g., influenza (38)). Autoimmune antibodies can recruit complement one aspect of disease pathogenesis (39), or even have receptor agonist or antagonist properties that primarily cause disease (39). Receptor agonism or antagonism is often much more important than affinity for drug discovery targets, and in some cases, the desired function is anti-correlated with affinity (40). The trade-offs between screening for affinity and screening functional performance have become even more important for multispecific and fusion proteins that use antibody domains, including T-cell engagers and affinity-tuned constructs. In multispecifics, each binding partner can have many possible variants with precisely known affinities, and yet still the ultimate combination of binding arms, linkers, and formats that will obtain a molecule with functional activity is very difficult to predict (41). Thus because affinity is only rarely the most important feature of an antibody, large-scale data collection for antibodies needs to focus on functional data collection to accurately explore antibody repertoire performance and obtain the necessary data for rigorous AI/ML model training.

While functional screening efforts have been pursued for many years, substantial progress is still needed. Most platforms are insufficiently flexible, have limited throughput, or require complex

instrumentation that raises experimental costs to the point that repertoire-scale analysis is not feasible for most groups or for collecting large-scale data (ideally, 10^5 or more data points). For example, the Beacon single-cell platform can achieve function-first screening, but it requires multiple days of on-chip culture and the sequencing throughput is limited to a few thousand antibody-secreting cells. Functional yeast reporter systems have also been established for purpose-built systems, but are limited to yeast-based reporter mechanisms (42). Fluorescence-activated droplet sorting (FADS) enables multi-cell capture of antibody-secreting cells and reporter cells, with commercial instrumentation recently becoming available (43–45). However, FADS is still limited by the throughput of the droplet sorter and by the Poisson distribution statistical limitations of obtaining a single antibody-secreting cell and a single reporter cell in the same droplet, restricting the scale of repertoire information that can be collected. To address these challenges, our group has recently reported an approach to combine antibody secretion and functional reporter systems into the same cell, allowing efficient screening for functional read-outs using drug-ready molecules secreted by mammalian cells(37). Because the reporter cell and the secreting cell are one and the same, it becomes simple to break the droplets and sort secreting cells directly based on the reporter gene expression. This approach easily enables the functional analysis of >20 million single antibody-secreting cells within a single afternoon. We anticipate that these advances in large-scale functional screening will accelerate progress in antibody repertoire analysis and directed evolution for drug discovery.

Challenges and Opportunities in Functional TCR Repertoire Analysis

TCRs present numerous complications for high-throughput display screening. TCRs are optimized for low-affinity, low binding energy interactions, and cross-reactivity occurs from common MHC protein contacts and closely related bound peptides (21). The combination of low affinity and substantial cross-reactivity makes TCR discovery presents unique challenges compared with antibodies because TCRs are prone to rapid dissociation as monomers, and the difference between on-target and off-target affinities is lower. Rather than *in vivo* selection for high binding affinity like antibodies, TCRs are instead chosen by the immune system to sensitively distinguish between

different degrees of low-affinity contacts. T cells also modulate surface TCR expression *in vivo*, in addition to co-receptor proteins like CD3 and CD8 (46), dynamically tuning TCR activation sensitivity according to innate cell and extrinsic immune-based signals. The dynamic tuning helps to avoid damaging autoimmune responses while preserving T cell potency for inflammatory threats. The combination of avidity-based TCR selection and cell-based modulation of TCR signal amplifiers make T cells impressively capable to distinguish between self and non-self *in vivo*. How to collect the necessary data to understand molecular immune dynamics for TCR-pMHC interactions, and also to predict the behavior of T cells that encode those TCRs after a certain stimulus, remains an important but unsolved problem.

To add to the complexity, not all binding TCRs will actively transduce the appropriate signals and activate T cells. TCR activation often occurs from low-affinity interactions, and appears to be mediated by energy vs. distance dynamics of the TCR-pMHC bond, rather than by overall affinity (47, 48). It is important to express multivalent TCRs (e.g. on mammalian cells) since low-affinity interactions often require suitable avidity for binding and/or activation. Multimers such as streptavidin tetramers or dextramers are often used to enhance avidity and prolong dissociation times for cell-based sorting, although these affinity-based screens do not directly collect information in functional TCR signaling (21). Mammalian cell lines such as Jurkat and SKW3 have been established to both analyze TCR affinity and also serve as activation reporters, often through native activation markers like CD69 expression or transcription factor expression like NF- κ B. These cell lines normally express the necessary co-receptor protein CD3, and also often express CD8 or CD4.

A common technique to identify antigen-specific TCRs uses T cell proliferation or cytokine expression after pulsing PBMC with antigen, and analyzing the activated T cells after culture (49). These co-culture assays identify certain populations of antigen-specific T cells, but are limited by the responsive capacity of primary T cells initially placed in culture. Some primary T cells are less sensitive and in an exhaustion state, whereas others are prone to sympathetically expand in a non-specific manner in a stimulatory environment. In many cases, T cells relevant to specific disease states cannot be sufficiently detected with peptide pulse expansion assays. T cell exhaustion is especially prevalent

for cancer-specific T cells, for which solutions that do not rely on antigen-specific primary cell proliferation are urgently needed. Another important challenge is to identify the pMHC antigens that are presented in disease-relevant contexts – for example, certain autoimmune peptides, or cancer antigens, which can include neoantigens, tumor-associated antigens, splice variants and/or peptides containing post-translational modifications. Improved methods for antigen-specific TCR analysis and for disease-relevant peptide identification are needed to improve upon the limitations of current TCR-pMHC identification assays.

To help address these challenges, our group has established a series of assays for renewable screening of TCR repertoires (31, 33). These platforms support unbiased functional screening against a set of desired pMHC tetramers or peptide-pulsed antigen-presenting cells (APCs). We first capture paired alpha:beta T cell receptor chains from T cell repertoire, and clone them *en masse* into mammalian display and reporter cell lines. By transferring the TCR libraries for expression in an activation-capable cell line, we ensure that every T cell in the library is on “equal footing” with respect to activation state. We can screen libraries by both affinity (31) and activation (33), and we can analyze TCR reactivities against cancer cell line and autoimmune model cell lines as well. A long-term goal is to use this platform to better understand the molecular characteristics of activating TCRs against pMHC antigens that are most important for disease, including against autoimmunity and cancers for which antigen-specific TCRs and the pMHCs that they recognize are both insufficiently described.

Catalyzing Progress with Big Data

Machine learning (ML) and artificial intelligence (AI) are transforming the way we understand, analyze, and interact with our world. A few exciting areas that AI/ML could assist researchers and clinicians include: 1) enhance basic sequence and structural understanding of adaptive immune interaction landscapes, 2) accelerate the design and discovery of new antibody- and TCR-based therapeutics, and 3) establish clinical biomarkers from immune receptor repertoire analyses to evaluate health, perform disease diagnosis & prognosis, and predict future disease risks. Successful AI/ML implementations nearly always occur in data-rich environments. Thus the key question for the

above applications is: do we have enough data already? And if not, how do we collect the data that these algorithms need to be successful? The most rapid progress for ML's impact appears to have been made in antibody developability. Several studies have compiled large datasets on various features that can influence developability features, which supported model development with a reasonable degree of accuracy for developability prediction (50). Mining immunized animal models to identify potentially optimal antibodies and mutations can also be a rich data source (51, 52), and library:library interaction data can also provide ample data points for effective discovery (53).

De novo immune receptor design against a desired target is still low-yield (a small number of designed antibodies are actual binders) but is advancing rapidly. The best current antibody design technologies still require experimental screening due to low hit rates, and often also require follow-up optimization. Most immune receptor design models are built around structural and protein:protein docking predictions. These structural prediction efforts are generally trained in part on crystallography data that cannot fully account for protein dynamics. The high diversity and untemplated nature of the antibody CDR3 regions also pose a challenge for structure prediction and docking algorithms, although with important recent progress (54). The high binding energies of antibody:antigen interactions provide a tailwind for affinity-based antibody computational design. Still, the greatest challenges woare to design functional antibodies (agonists, allosteric modulators, and neutralizers), which can require dynamics that are not neatly captured in most structural data. Functional antibodies may also require binding to very precise regions of an antigen that are difficult to target specifically by immunization or library panning. Multispecifics present another major design challenge, and also rely heavily on dynamics and yet have many higher degrees of freedom compared to monoclonal antibodies. We need to collect more richly annotated datasets to address these antibody design & evaluation challenges in a satisfactory manner using AI, with a special focus on functional data as opposed to affinity data.

For T cell receptors, ML has been used extensively to mine immunoproteomic peptide elution data and refine peptide display predictions. Several recent reports also explored AI/ML for TCR-pMHC interaction prediction (55), and prediction accuracy is improving quickly. TCR:pMHC interaction

prediction may be more tractable than many antibody:antigen predictions due to the common binding epitope and restricted angles of approach for TCR docking, and the lack of somatic hypermutation makes TCR sequence space much smaller than antibody sequence space. However, structure prediction of TCR:pMHC presents three major complexities that hinder progress: a) TCR:pMHC complexes have low affinity, which is more difficult to detect in energy-based docking models, b) cross-reactivity is common, and c) TCR activation strength is only affinity-driven, but is also a function of bond energy vs. distance. Due to these lower interaction energies and the importance of the energy vs. distance dynamics, we may need extensive functional TCR training data to design antigen-specific activating TCRs with minimal cross-reactivity. The known TCR structures are also biased towards higher-affinity TCRs, and additional structures of low-affinity, activating TCRs is likely to provide important data to accelerate progress. These unique challenges for TCR structural prediction are counterbalanced by the relatively lower diversity of the TCR and the common recognition modes (and thus more limited degrees of freedom for TCR:pMHC interactions) that can help support sufficient dataset collection and accelerate progress in TCR:pMHC prediction approaches.

More high-quality data is needed to address all these challenges. One relatively rapid way to collect large-scale affinity data is the use of library:library technologies (56), which promise to fill out large sets of affinity information and are very useful for AI/ML training. The ability to obtain large-scale functional information on antibody or TCR interactions, especially when combined with large-scale affinity data in parallel, will help to more finely discriminate between affinity/avidity interactions versus functional interactions. In pursuit of these aims, our group is collecting large-scale data on immune receptor performance for both affinity (28, 30, 34–36) and functional activity (33, 37). We hope that these rich datasets will catalyze new progress in AI/ML performance by combining functional data layers along with structure prediction. Pairing large-scale functional data to structural knowledge will unite advances in both experimental and computational techniques to unlock future opportunities in functional immune receptor design and prediction.

The Future of Immune Receptor Technologies

The study and use of immune receptors will be dramatically transformed in coming decades as scientific tools to control immune responses and design new drugs continue to expand, while the costs of drug distribution & production continue to fall. For antibody therapies, active product dose manufacturing costs will fall from a combination of higher-potency drugs from refined discovery pipelines, longer half-life Fc modifications, and high-concentration formulations to allow infrequent and at-home subcutaneous delivery. Future improvements in nucleic acid drug delivery may also transform the protein drug treatment landscape by simplifying manufacturing & distribution. Extrapolating from the ~200 currently approved monoclonal antibodies, the large pipeline of drugs in clinical trials, and current growth from annual FDA antibody approvals, it is conceivable that we could have close to 1,000 or more antibodies approved by 2040. At some point in the coming decades, reductions in active product dose costs and formulation/distribution advances will shift antibodies from a specialty drug class into more of a commodity, priced perhaps like the upper tier of today's over-the-counter drugs. AI/ML will eventually become broadly useful to design and predict the performance of antibody drugs. Experiments will not be replaced or eliminated, but instead will focus on: 1) large-scale data collection and for model training that will enhance AI/ML accuracy and partner with computational tools on focused tasks, and 2) efficient large-scale validation of computationally designed molecules, and 3) engineering or optimization of initial candidates. Growth in immune repertoire analyses to measure broad clinical biomarkers will continue more slowly than drug discovery due to the tighter restrictions on clinical data collection, analysis, and interpretation compared to bench research. Still, immune repertoire analysis for diagnostics, prognosis, and treatment will build on its currently established base in minimal residual disease detection, and will develop fastest in therapeutic settings with strong patient-to-patient presentation differences and difficult-to-predict disease courses that also have high patient burdens, such as cancer immunotherapy and severe autoimmune diseases.

TCRs are ideally suited for personalized medicines, and are already successfully used in personalized T cell and TCR-T cancer therapies. With sufficient data, improved design & selection capabilities will improve TCR-T treatment timelines and reduce costs for broader, worldwide

378 therapeutic growth. TCR-T are somewhat scalable on an individualized nature since the TCR
379 molecule can be packaged as manufacturing-friendly nucleic acids and the resulting active drug
380 product (the TCR-T cell) can multiply within the patient, which reduces the need for ongoing drug
381 administration. Thus if the TCR-pMHC matching problem can be conclusively resolved and with
382 sufficient reliability for medical settings, cancer therapy and autoimmune disease therapy will be
383 radically transformed by personalized, designer TCR therapeutics. At some point in the foreseeable
384 future, TCR drugs will be computationally designed based on immune repertoire and disease markers
385 (e.g., based on tumor biopsy data), and a nucleic acid to deliver TCRs could be available for a quick
386 subcutaneous injection at a local pharmacy. With robust and reliable TCR-pMHC matching, TCR
387 immune repertoire sequencing will become a biomarker for autoimmune monitoring, cancer
388 suppression or progression, viral disease susceptibility, and as a general marker of health state.

389 Continued scientific progress will further blend the antibody and TCR modalities for synthetic
390 and translational purposes. For example, soluble TCRs will be used like antibodies as the components
391 of multispecifics against pMHC antigens (e.g., T cell engagers). Antibodies will also be used like
392 TCRs, for example in pMHC-specific antibodies, and of course in various formats of CAR-T. With
393 sufficient functional data and safety/developability prediction data, personalized antibody therapies
394 could become a reality. An important first application for personalized antibody design would be
395 pMHC-specific antibodies, which may be a tractable problem due to the restricted degrees of freedom
396 and common MHC binding contacts for antibody-pMHC interactions. Such antibodies could have a
397 common framework for each MHC protein and vary only in the specific peptide contacts, and would
398 have numerous high-value applications to support their initial development in autoimmunity and
399 cancer treatment. In some ways, pMHC-specific antibodies could be simpler to design and quality
400 control than TCR-T because antibodies are transiently dosed as protein rather than as cell-based
401 TCR-T. Progress is already occurring quickly in *de novo* designs for proteins that bind pMHC (57),
402 and designed antibodies and TCRs to precisely target pMHCs seem right around the corner.

403 As treatment costs continue to fall we may see expanded prophylactic uses for both antibodies
404 and TCRs. Studies of antiviral immune protection have revealed extremely potent and broad anti-viral

405 antibody responses in a subset of patients over the past 15 years; research into other diseases like
406 cancers may reveal similar outstanding protective adaptive immune features in certain individuals. As
407 costs for antibody and TCR drug delivery become competitive with traditional vaccines, we may begin
408 augmenting our immune systems with exceptional adaptive immune proteins or immune constructs,
409 rather than vaccinate and rely on our own immune responses against infectious diseases. For
410 example, highly potent anti-flu antibodies with extended half-life modifications may be given in small
411 annual doses before flu season, or even genetically programmed to be produced by our cells
412 throughout the year. Anti-cancer antibody cocktails could be given to patients with a suitable risk
413 profile on a prophylactic basis. These expanded drug uses of immune receptors in healthy individuals
414 will only be possible by first addressing the safety barriers for interventions in healthy individuals, for
415 which precedent exists for the several monoclonal antibodies already used for infectious disease
416 prevention in approved therapeutics.

417 As we continue to expand our abilities to manipulate immune protection, we will transition from
418 blunt techniques like immunosuppressants and live-attenuated or inactivated vaccines toward fine-
419 grained, molecular control of immunity using molecular analysis and optimization of specific adaptive
420 immune receptors. Personalized, large-data tools will guide patients and their doctors to help maintain
421 optimal immune health. Immune systems will be read and analyzed *en masse*, biomarkers will be
422 tracked for diagnosis and prognosis, and new drugs will be designed, all in partnership with guided
423 computational assistance. The tight union of wet lab data, clinical data, and computational analyses
424 represents the most important theme of the coming decades as we accelerate our explorations of
425 adaptive immune receptors in science and medicine.

426 **Acknowledgements**

427 Many thanks to Camila Franca for assistance with figure preparation.

428

429 **Author contributions (with CRediT details):**

430 B.J.D.: Conceptualization, Funding acquisition, Visualization, Writing – original draft, and
431 Writing – review & editing.

432

433 **Funding:**

434 This work was delivered as part of the MATCHMAKERS team, supported by the Cancer
435 Grand Challenges partnership financed by CRUK (CGCATF-2023/100009), the National
436 Cancer Institute (OT2CA297514), and The Mark Foundation for Cancer Research. This work
437 was also supported by NIH grants 1R01AI181684, 1R01AI192975, 1U01AI169587,
438 R21AI166396, R21AI144408, R21AI143407, R21AI178021, and DP5OD023118, and by The
439 Mark and Lisa Schwartz AI/ML Initiative, The American Cancer Society, The Gates
440 Foundation, the Koch Institute for Cancer Research, the MIT Research Support Committee,
441 the MIT Department of Chemical Engineering, and The Ragon Institute.

442

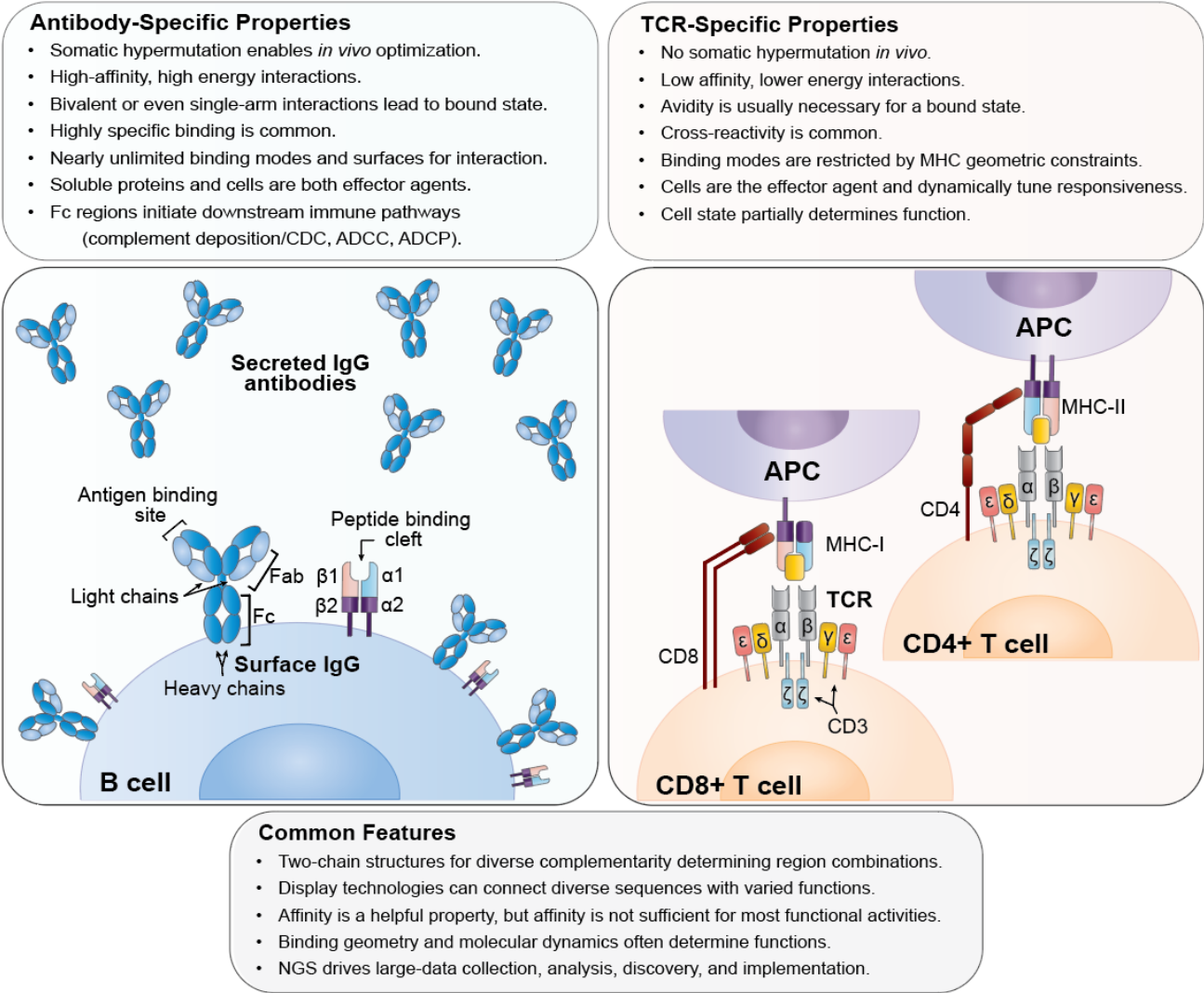
443 **Conflicts of interest:**

444 B.J.D. is a co-inventor of patents related to immune receptor discovery technologies currently
445 assigned to The University of Texas at Austin, The University of Kansas, Massachusetts
446 General Hospital, and Massachusetts Institute of Technology.

447

448 **Data availability:**

449 No new data is reported.



451

452 **Figure 1. Unique and common features of the adaptive immune receptors.** The listed
453 properties constitute important factors in the design of molecular sequencing, functional
454 screening, discovery, and engineering studies for antibodies and TCRs. CDC-complement
455 dependent cytotoxicity; ADCC-antibody dependent cell cytotoxicity; ADCP-antibody
456 dependent cell phagocytosis; APC-antigen presenting cell.

References

1. Kohler, G., and C. Milstein. 1975. CONTINUOUS CULTURES OF FUSED CELLS SECRETING ANTIBODY OF PREDEFINED SPECIFICITY. *Nature* 256: 495–497.
2. Tiller, T. 2011. Single B cell antibody technologies. *New Biotechnology* 28: 453–457.
3. Schoettler, N., D. Ni, and M. Weigert. 2012. B cell receptor light chain repertoires show signs of selection with differences between groups of healthy individuals and SLE patients. *Molecular Immunology* 51: 273–282.
4. Brezinschek, H.-P., S. J. Foster, T. Dörner, R. I. Brezinschek, and P. E. Lipsky. 1998. Pairing of variable heavy and variable κ chains in individual naive and memory B cells. *The Journal of Immunology* 160: 4762–4767.
5. 2018. Nobel work that galvanized an industry. *Nat Biotechnol* 36: 1023–1023.
6. Mhanna, V., H. Bashour, K. Lê Quý, P. Barennes, P. Rawat, V. Greiff, and E. Mariotti-Ferrandiz. 2024. Adaptive immune receptor repertoire analysis. *Nat Rev Methods Primers* 4: 6.
7. DeKosky, B. J., G. C. Ippolito, R. P. Deschner, J. J. Lavinder, Y. Wine, B. M. Rawlings, N. Varadarajan, C. Giesecke, T. Dorner, S. F. Andrews, P. C. Wilson, S. P. Hunicke-Smith, C. G. Willson, A. D. Ellington, and G. Georgiou. 2013. High-throughput sequencing of the paired human immunoglobulin heavy and light chain repertoire. *Nat Biotech* 31: 166–169.
8. DeKosky, B. J., T. Kojima, A. Rodin, W. Charab, G. C. Ippolito, A. D. Ellington, and G. Georgiou. 2015. In-depth determination and analysis of the human paired heavy- and light-chain antibody repertoire. *Nat. Med.* 21: 86–91.
9. Pai, J. A., and A. T. Satpathy. 2021. High-throughput and single-cell T cell receptor sequencing technologies. *Nat Methods* 18: 881–892.
10. DeKosky, B. J., O. I. Lungu, D. Park, E. L. Johnson, W. Charab, C. Chrysostomou, D. Kuroda, A. D. Ellington, G. C. Ippolito, J. J. Gray, and G. Georgiou. 2016. Large-scale sequence and structural comparisons of human naive and antigen-experienced antibody repertoires. *PNAS* 113: E2636–E2645.
11. Rhodes, E. R., J. G. Faris, B. M. Petersen, and K. G. Sprenger. 2023. Common framework mutations impact antibody interfacial dynamics and flexibility. *Front. Immunol.* 14: 1120582.
12. Burkovitz, A., I. Sela-Culang, and Y. Ofran. 2014. Large-scale analysis of somatic hypermutations in antibodies reveals which structural regions, positions and amino acids are modified to improve affinity. *The FEBS Journal* 281: 306–319.
13. Jaffe, D. B., P. Shahi, B. A. Adams, A. M. Chrisman, P. M. Finnegan, N. Raman, A. E. Royall, F. Tsai, T. Vollbrecht, D. S. Reyes, N. L. Hepler, and W. J. McDonnell. 2022. Functional antibodies exhibit light chain coherence. *Nature* 1–6.

14. Venturi, V., K. Kedzierska, D. A. Price, P. C. Doherty, D. C. Douek, S. J. Turner, and M. P. Davenport. 2006. Sharing of T cell receptors in antigen-specific responses is driven by convergent recombination. *Proc. Natl. Acad. Sci. U.S.A.* 103: 18691–18696.
15. Quigley, M. F., H. Y. Greenaway, V. Venturi, R. Lindsay, K. M. Quinn, R. A. Seder, D. C. Douek, M. P. Davenport, and D. A. Price. 2010. Convergent recombination shapes the clonotypic landscape of the naive T-cell repertoire. *Proc. Natl. Acad. Sci. U.S.A.* 107: 19414–19419.
16. Klebanoff, C. A., S. S. Chandran, B. M. Baker, S. A. Quezada, and A. Ribas. 2023. T cell receptor therapeutics: immunological targeting of the intracellular cancer proteome. *Nat Rev Drug Discov* 22: 996–1017.
17. Adler, A. S., D. Bedinger, M. S. Adams, M. A. Asensio, R. C. Edgar, R. Leong, J. Leong, R. A. Mizrahi, M. J. Spindler, S. R. Bandi, H. Huang, P. Tawde, P. Brams, and D. S. Johnson. 2018. A natively paired antibody library yields drug leads with higher sensitivity and specificity than a randomly paired antibody library. *mAbs* 0: 1–13.
18. Linette, G. P., E. A. Stadtmauer, M. V. Maus, A. P. Rapoport, B. L. Levine, L. Emery, L. Litzky, A. Bagg, B. M. Carreno, P. J. Cimino, G. K. Binder-Scholl, D. P. Smethurst, A. B. Gerry, N. J. Pumphrey, A. D. Bennett, J. E. Brewer, J. Dukes, J. Harper, H. K. Tayton-Martin, B. K. Jakobsen, N. J. Hassan, M. Kalos, and C. H. June. 2013. Cardiovascular toxicity and titin cross-reactivity of affinity-enhanced T cells in myeloma and melanoma. *Blood* 122: 863–871.
19. Sender, R., Y. Weiss, Y. Navon, I. Milo, N. Azulay, L. Keren, S. Fuchs, D. Ben-Zvi, E. Noor, and R. Milo. 2023. The total mass, number, and distribution of immune cells in the human body. *Proc. Natl. Acad. Sci. U.S.A.* 120: e2308511120.
20. Wu, X., Z.-Y. Yang, Y. Li, C.-M. Hogerkorp, W. R. Schief, M. S. Seaman, T. Zhou, S. D. Schmidt, L. Wu, L. Xu, N. S. Longo, K. McKee, S. O'Dell, M. K. Louder, D. L. Wycuff, Y. Feng, M. Nason, N. Doria-Rose, M. Connors, P. D. Kwong, M. Roederer, R. T. Wyatt, G. J. Nabel, and J. R. Mascola. 2010. Rational Design of Envelope Identifies Broadly Neutralizing Human Monoclonal Antibodies to HIV-1. *Science* 329: 856–861.
21. Christophersen, A. 2020. Peptide-MHC class I and class II tetramers: From flow to mass cytometry. *HLA* 95: 169–178.
22. Bentzen, A. K., A. M. Marquard, R. Lyngaa, S. K. Saini, S. Ramskov, M. Donia, L. Such, A. J. S. Furness, N. McGranahan, R. Rosenthal, P. T. Straten, Z. Szallasi, I. M. Svane, C. Swanton, S. A. Quezada, S. N. Jakobsen, A. C. Eklund, and S. R. Hadrup. 2016. Large-scale detection of antigen-specific T cells using peptide-MHC-I multimers labeled with DNA barcodes. *Nat Biotechnol* 34: 1037–1045.
23. Zhang, S.-Q., K.-Y. Ma, A. A. Schonnesen, M. Zhang, C. He, E. Sun, C. M. Williams, W. Jia, and N. Jiang. 2018. High-throughput determination of the antigen specificities of T cell receptors in single cells. *Nature Biotechnology* 36: 1156–1159.

24. Setliff, I., A. R. Shiakolas, K. A. Pilewski, A. A. Murji, R. E. Mapengo, K. Janowska, S. Richardson, C. Oosthuysen, N. Raju, L. Ronsard, M. Kanekiyo, J. S. Qin, K. J. Kramer, A. R. Greenplate, W. J. McDonnell, B. S. Graham, M. Connors, D. Lingwood, P. Acharya, L. Morris, and I. S. Georgiev. 2019. High-Throughput Mapping of B Cell Receptor Sequences to Antigen Specificity. *Cell* 179: 1636-1646.e15.
25. Valldorf, B., S. C. Hinz, G. Russo, L. Pekar, L. Mohr, J. Klemm, A. Doerner, S. Krah, M. Hust, and S. Zielonka. 2022. Antibody display technologies: selecting the cream of the crop. *Biological Chemistry* 403: 455–477.
26. Slavny, P., M. Hegde, A. Doerner, K. Parthiban, J. McCafferty, S. Zielonka, and R. Hoet. 2024. Advancements in mammalian display technology for therapeutic antibody development and beyond: current landscape, challenges, and future prospects. *Front. Immunol.* 15.
27. Rosenberg, A. M., and B. M. Baker. 2022. Engineering the T cell receptor for fun and profit: Uncovering complex biology, interrogating the immune system, and targeting disease. *Current Opinion in Structural Biology* 74: 102358.
28. Wang, B., B. J. DeKosky, M. R. Timm, J. Lee, E. Normandin, J. Misasi, R. Kong, J. R. McDaniel, G. Delidakis, K. E. Leigh, T. Niezold, C. W. Choi, E. G. Viox, A. Fahad, A. Cagigi, A. Ploquin, K. Leung, E. S. Yang, W.-P. Kong, W. N. Voss, A. G. Schmidt, M. A. Moody, D. R. Ambrozak, A. R. Henry, F. Laboune, J. E. Ledgerwood, B. S. Graham, M. Connors, D. C. Douek, N. J. Sullivan, A. D. Ellington, J. R. Mascola, and G. Georgiou. 2018. Functional interrogation and mining of natively paired human V_H:V_L antibody repertoires. *Nature Biotechnology* 36: 152–155.
29. Banach, B. B., G. Cerutti, A. S. Fahad, C.-H. Shen, M. O. de Souza, P. S. Katsamba, Y. Tsybovsky, P. Wang, M. S. Nair, Y. Huang, I. M. F. Urdaniz, P. J. Steiner, M. Gutierrez-Gonzalez, L. Liu, S. N. L. Acevedo, A. Nazzari, J. R. Wolfe, Y. Luo, A. S. Olia, I.-T. Teng, J. Yu, T. Zhou, E. R. Reddem, J. Bimela, X. Pan, B. Madan, A. D. Laflin, R. Nimrania, K.-T. Yuen, T. A. Whitehead, D. D. Ho, P. D. Kwong, L. Shapiro, and B. J. DeKosky. 2021. Paired heavy and light chain signatures contribute to potent SARS-CoV-2 neutralization in public antibody responses. *Cell Reports* 2020.12.31.424987.
30. Fahad, A. S., M. R. Timm, B. Madan, K. E. Burgomaster, K. A. Dowd, E. Normandin, M. F. Gutiérrez-González, J. M. Pennington, M. O. De Souza, A. R. Henry, F. Laboune, L. Wang, D. R. Ambrozak, I. J. Gordon, D. C. Douek, J. E. Ledgerwood, B. S. Graham, L. R. Castilho, T. C. Pierson, J. R. Mascola, and B. J. DeKosky. 2021. Functional Profiling of Antibody Immune Repertoires in Convalescent Zika Virus Disease Patients. *Front. Immunol.* 12.
31. Fahad, A. S., C.-Y. Chung, S. N. Lopez Acevedo, N. Boyle, B. Madan, M. F. Gutiérrez-González, R. Matus-Nicodemos, A. D. Laflin, R. R. Ladi, J. Zhou, J. Wolfe, S. Llewellyn-Lacey, R. A. Koup, D. C. Douek, H. H. Balfour Jr, D. A. Price, and B. J. DeKosky. 2022. Immortalization

- and functional screening of natively paired human T cell receptor repertoires. *Protein Engineering, Design and Selection* 35: gzab034.
32. Pan, X., S. N. López Acevedo, C. Cuziol, E. De Tavernier, A. S. Fahad, P. S. Longjam, S. P. Rao, D. Aguilera-Rodríguez, M. Rezé, C. A. Bricault, M. F. Gutiérrez-González, M. O. de Souza, J. M. DiNapoli, E. Vigne, M. A. Shahsavarian, and B. J. DeKosky. 2023. Large-scale antibody immune response mapping of splenic B cells and bone marrow plasma cells in a transgenic mouse model. *Frontiers in Immunology* 14.
 33. Fahad, A. S., C. Y. Chung, S. N. López Acevedo, N. Boyle, B. Madan, M. F. Gutiérrez-González, R. Matus-Nicodemos, A. D. Laflin, R. R. Ladi, J. Zhou, J. Wolfe, S. Llewellyn-Lacey, R. A. Koup, D. C. Douek, H. H. Balfour, D. A. Price, and B. J. DeKosky. 2023. Cell activation-based screening of natively paired human T cell receptor repertoires. *Sci Rep* 13: 8011.
 34. Madan, B., B. Zhang, K. Xu, C. W. Chao, S. O'Dell, J. R. Wolfe, G.-Y. Chuang, A. S. Fahad, H. Geng, R. Kong, M. K. Louder, T. D. Nguyen, R. Rawi, A. Schön, Z. Sheng, R. Nimrania, Y. Wang, T. Zhou, B. C. Lin, N. A. Doria-Rose, L. Shapiro, P. D. Kwong, and B. J. DeKosky. 2021. Mutational fitness landscapes reveal genetic and structural improvement pathways for a vaccine-elicited HIV-1 broadly neutralizing antibody. *Proc Natl Acad Sci USA* 118: e2011653118.
 35. Banach, B. B., P. Tripathi, L. D. S. Pereira, J. Gorman, T. D. Nguyen, M. Dillon, A. S. Fahad, P. K. Kiyuka, B. Madan, J. R. Wolfe, B. Bonilla, B. Flynn, J. R. Francica, N. K. Hurlburt, N. Kisalu, T. Liu, L. Ou, R. Rawi, A. Schon, C.-H. Shen, I.-T. Teng, B. Zhang, M. Pancera, A. H. Idris, R. A. Seder, P. D. Kwong, and B. J. DeKosky. 2022. Highly protective antimalarial antibodies via precision library generation and yeast display screening. *Journal of Experimental Medicine* 219.
 36. Banach, B. B., S. Pletnev, A. S. Olia, K. Xu, B. Zhang, R. Rawi, T. Bylund, N. A. Doria-Rose, T. D. Nguyen, A. S. Fahad, M. Lee, B. C. Lin, T. Liu, M. K. Louder, B. Madan, K. McKee, S. O'Dell, M. Sastry, A. Schön, N. Bui, C.-H. Shen, J. R. Wolfe, G.-Y. Chuang, J. R. Mascola, P. D. Kwong, and B. J. DeKosky. 2023. Antibody-directed evolution reveals a mechanism for enhanced neutralization at the HIV-1 fusion peptide site. *Nat Commun* 14: 7593.
 37. Gutierrez-Gonzalez, M., A. S. Fahad, C. L. Delley, C.-Y. Chung, S. Jin, N. Boyle, M. O. de Souza, A. Pirhanov, N. M. S. Galvez, C. T. Franca, S. Marglous, E. Bhagat, D. Vincent, D. Neumeier, Y. Cao, N. Doria-Rose, S. T. Reddy, A. G. Schmidt, A. B. Balazs, A. R. Abate, and B. J. DeKosky. 2025. High-Throughput Antibody Neutralization Screening in Massively Parallel Droplet Arrays. 2025.01.31.635965.
 38. Chandler, T. L., A. Yang, C. E. Otero, S. R. Permar, and S. L. Caddy. 2023. Protective mechanisms of nonneutralizing antiviral antibodies. *PLoS Pathog* 19: e1011670.
 39. Pisetsky, D. S. 2023. Pathogenesis of autoimmune disease. *Nat Rev Nephrol* 19: 509–524.

596 40. Yu, X., C. M. Orr, H. T. C. Chan, S. James, C. A. Penfold, J. Kim, T. Inzhelevskaya, C. I.
597 Mockridge, K. L. Cox, J. W. Essex, I. Tews, M. J. Glennie, and M. S. Cragg. 2023. Reducing
598 affinity as a strategy to boost immunomodulatory antibody agonism. *Nature* 614: 539–547.

599 41. Keri, D., M. Walker, I. Singh, K. Nishikawa, and F. Garces. 2024. Next generation of
600 multispecific antibody engineering. *Antibody Therapeutics* 7: 37–52.

601 42. Yaginuma, K., W. Aoki, N. Miura, Y. Ohtani, S. Aburaya, M. Kogawa, Y. Nishikawa, M.
602 Hosokawa, H. Takeyama, and M. Ueda. 2019. High-throughput identification of peptide agonists
603 against GPCRs by co-culture of mammalian reporter cells and peptide-secreting yeast cells
604 using droplet microfluidics. *Sci Rep* 9: 10920.

605 43. Gérard, A., A. Woolfe, G. Mottet, M. Reichen, C. Castrillon, V. Menrath, S. Ellouze, A. Poitou, R.
606 Doineau, L. Briseno-Roa, P. Canales-Herrerias, P. Mary, G. Rose, C. Ortega, M. Delincé, S.
607 Essono, B. Jia, B. Iannascoli, O. R.-L. Goff, R. Kumar, S. N. Stewart, Y. Pousse, B. Shen, K.
608 Grosselin, B. Saudemont, A. Sautel-Caillé, A. Godina, S. McNamara, K. Eyer, G. A. Millot, J.
609 Baudry, P. England, C. Nizak, A. Jensen, A. D. Griffiths, P. Bruhns, and C. Brenan. 2020. High-
610 throughput single-cell activity-based screening and sequencing of antibodies using droplet
611 microfluidics. *Nature Biotechnology* 1–7.

612 44. Wang, Y., R. Jin, B. Shen, N. Li, H. Zhou, W. Wang, Y. Zhao, M. Huang, P. Fang, S. Wang, P.
613 Mary, R. Wang, P. Ma, R. Li, Y. Tian, Y. Cao, F. Li, L. Schweizer, and H. Zhang. 2021. High-
614 throughput functional screening for next-generation cancer immunotherapy using droplet-based
615 microfluidics. *Sci. Adv.* 7: eabe3839.

616 45. Segaliny, A. I., J. Jayaraman, X. Chen, J. Chong, R. Luxon, A. Fung, Q. Fu, X. Jiang, R. Rivera,
617 X. Ma, C. Ren, J. Zimak, P. N. Hedde, Y. Shang, G. Wu, and W. Zhao. 2023. A high throughput
618 bispecific antibody discovery pipeline. *Commun Biol* 6: 1–14.

619 46. Xiao, Z., M. F. Mescher, and S. C. Jameson. 2007. Detuning CD8 T cells: down-regulation of
620 CD8 expression, tetramer binding, and response during CTL activation. *The Journal of*
621 *Experimental Medicine* 204: 2667–2677.

622 47. Sibener, L. V., R. A. Fernandes, E. M. Kolawole, C. B. Carbone, F. Liu, D. McAfee, M. E.
623 Birnbaum, X. Yang, L. F. Su, W. Yu, S. Dong, M. H. Gee, K. M. Jude, M. M. Davis, J. T. Groves,
624 W. A. Goddard, J. R. Heath, B. D. Evavold, R. D. Vale, and K. C. Garcia. 2018. Isolation of a
625 Structural Mechanism for Uncoupling T Cell Receptor Signaling from Peptide-MHC Binding. *Cell*
626 174: 672-687.e27.

627 48. Zhao, X., E. M. Kolawole, W. Chan, Y. Feng, X. Yang, M. H. Gee, K. M. Jude, L. V. Sibener, P.
628 M. Fordyce, R. N. Germain, B. D. Evavold, and K. C. Garcia. 2022. Tuning T cell receptor
629 sensitivity through catch bond engineering. *Science* 376: eabl5282.

49. Tippalagama, R., L. Y. Chihab, K. Kearns, S. Lewis, S. Panda, L. Willemsen, J. G. Burel, and C. S. Lindestam Arlehamn. 2023. Antigen-specificity measurements are the key to understanding T cell responses. *Front. Immunol.* 14: 1127470.
50. Fernández-Quintero, M. L., A. Ljungars, F. Waibl, V. Greiff, J. T. Andersen, T. T. Gjølberg, T. P. Jenkins, B. G. Voldborg, L. M. Grav, S. Kumar, G. Georges, H. Kettenberger, K. R. Liedl, P. M. Tessier, J. McCafferty, and A. H. Laustsen. 2023. Assessing developability early in the discovery process for novel biologics. *mAbs* 15: 2171248.
51. Mason, D. M., S. Friedensohn, C. R. Weber, C. Jordi, B. Wagner, S. M. Meng, R. A. Ehling, L. Bonati, J. Dahinden, P. Gainza, B. E. Correia, and S. T. Reddy. 2021. Optimization of therapeutic antibodies by predicting antigen specificity from antibody sequence via deep learning. *Nat Biomed Eng* 5: 600–612.
52. Erlach, L., S. Friedensohn, D. Neumeier, D. M. Mason, and S. T. Reddy. 2025. Antibody affinity engineering using antibody repertoire data and machine learning. .
53. Agarwal, A. A., J. Harrang, D. Noble, K. L. McGowan, A. W. Lange, E. Engelhart, M. C. Lahman, J. Adamo, X. Yu, O. Serang, K. J. Minch, K. Y. Wellman, D. A. Younger, R. M. Lopez, and R. O. Emerson. 2025. AlphaBind, a domain-specific model to predict and optimize antibody–antigen binding affinity. *mAbs* 17: 2534626.
54. Hitawala, F. N., and J. J. Gray. 2024. What does AlphaFold3 learn about antigen and nanobody docking, and what remains unsolved? .
55. Deng, L., C. Ly, S. Abdollahi, Y. Zhao, I. Prinz, and S. Bonn. 2023. Performance comparison of TCR-pMHC prediction tools reveals a strong data dependency. *Front. Immunol.* 14: 1128326.
56. Johnson, S., and B. J. DeKosky. 2024. Expanding the landscape of antibody discovery. *Cell Reports Methods* 4: 100936.
57. Hickok, G. H., and I. M. Stromnes. Beyond the native repertoire. .