DOI: 10.1111/imr.12480

INVITED REVIEW

WILEY Immunological Reviews

Antibodyomics: bioinformatics technologies for understanding B-cell immunity to HIV-1

Peter D. Kwong[1,2] | Gwo-Yu Chuang[1] | Brandon J. DeKosky[1] | Tatyana Gindin[2] | Ivelin S. Georgiev[3] | Thomas Lemmin[4] | Chaim A. Schramm[1,2,5] | Zizhang Sheng[2,5] | Cinque Soto[1] | An-Suei Yang[6] | John R. Mascola[1] | Lawrence Shapiro[1,2,5]

[1]Vaccine Research Center, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD, USA

[2]Department of Biochemistry & Molecular Biophysics, Columbia University, New York, NY, USA

[3]Vanderbilt Vaccine Center and Department of Pathology, Microbiology, and Immunology, Vanderbilt University Medical Center, Nashville, TN, USA

[4]Department of Pharmaceutical Chemistry, University of California San Francisco, San Francisco, CA, USA

[5]Department of Systems Biology, Columbia University, New York, NY, USA

[6]Genomics Research Center, Academia Sinica, Taipei, Taiwan

Correspondence
Lawrence Shapiro, Columbia University, New York, NY, USA
Email: shapiro@convex.hhmi.columbia.edu

Summary

Numerous antibodies have been identified from HIV-1-infected donors that neutralize diverse strains of HIV-1. These antibodies may provide the basis for a B cell-mediated HIV-1 vaccine. However, it has been unclear how to elicit similar antibodies by vaccination. To address this issue, we have undertaken an informatics-based approach to understand the genetic and immunologic processes controlling the development of HIV-1-neutralizing antibodies. As DNA sequencing comprises the fastest growing database of biological information, we focused on incorporating next-generation sequencing of B-cell transcripts to determine the origin, maturation pathway, and prevalence of broadly neutralizing antibody lineages (Antibodyomics1, 2, 4, and 6). We also incorporated large-scale robotic analyses of serum neutralization to identify and quantify neutralizing antibodies in donor cohorts (Antibodyomics3). Statistical analyses furnish another layer of insight (Antibodyomics5), with physical characteristics of antibodies and their targets through molecular dynamics simulations (Antibodyomics7) and free energy perturbation analyses (Antibodyomics8) providing information-rich output. Functional interrogation of individual antibodies (Antibodyomics9) and synthetic antibody libraries (Antibodyomics10) also yields multi-dimensional data by which to understand and improve antibodies. Antibodyomics, described here, thus comprise resolution-enhancing tools, which collectively embody an information-driven discovery engine aimed toward the development of effective B cell-based vaccines.

KEYWORDS

bioinformatics, B-cell ontogeny, broadly neutralizing antibodies, HIV vaccine, information technology, massively parallel sequencing

1 | INTRODUCTION

Despite the proven ability of the human immune system to generate broadly neutralizing antibodies in response to HIV infection,[1–3] to date no broadly neutralizing antibody against HIV-1 has been induced by vaccination. Vaccine induction of effective HIV-1-neutralizing antibodies thus represents a major unfulfilled goal in modern immunology.

Many approaches are being used to achieve this goal, several of which are described in this volume on antibodies and immunity to HIV-1. One approach based on decision, information, and game theory, proposes resolution-enhancing multi-dimensional analyses

This article is part of a series of reviews covering B cells and Immunity to HIV appearing in Volume 275 of *Immunological Reviews*.

as a search strategy.[4] Here, we describe a collection of technologies (Antibodyomics1–10) that provide information-rich analyses of the genetic and immunological processes controlling the development of effective HIV-1-neutralizing antibodies (Fig. 1). Our description is not intended to represent a comprehensive summary of the field; rather, we have been asked by the editors of this volume to synthesize key aspects and to focus on our specific contributions and ideas. The majority of the Antibodyomics technologies incorporate data generated by massively parallel next-generation sequencing (NGS) of B-cell transcripts to enhance the resolution of the genetic information underlying antibody development.[5–9] We have also focused on technologies that allowed increased resolution of function relative to physical characteristics of antibodies and their targets. Together, these multi-dimensional technologies may afford sufficient resolution of antibodies and relevant immunological processes so they can be optimized for passive delivery or recapitulated by vaccination.

2 | ANTIBODYOMICS1

NGS of B-cell transcripts provides a means to interrogate the human antibody repertoire and to understand the B cell-based response to infection. We developed a software suite, Antibodyomics1, which can be used to organize cross-sectional NGS data containing millions of B-cell transcripts into distinct lineages and to map their development (Fig. 2). At its core, Antibodyomics1 assigns immunoglobulin germline genes to each NGS transcript, parses framework and CDR regions from each transcript, and uses sequence metrics to identify related antibodies in unrelated donors. Antibodyomics1 was originally developed to analyze 454 pyrosequencing data from HIV-1 infected donors [10] and can now accept sequencing data from the Illumina platform.[5,11] Here, we highlight applications of the Antibodyomics1 pipeline.

The earliest application of Antibodyomics1 involved the discovery and characterization of antibodies from HIV-1–infected individuals.[9] Specifically, heavy chain sequences from antibodies of the VRC01 class [a multidonor class of antibodies directed at the CD4-binding site (12)] were used to analyze B-cell transcripts from an HIV-1 infected donor (donor 74). This analysis yielded hundreds of new VRC01-class antibody sequences, with lineage analysis allowing the tracing back to earlier developmental intermediates. This initial application was aided by the sequence of VRC-PG04, a VRC01-class antibody identified by probe sorting of donor 74 B cells.[9] The results appeared promising enough that we undertook the de novo identification of VRC01-class antibodies in a new donor, C38, from which no VRC01-class antibodies were known. By using only template VRC01-class antibodies from non-C38 donors, we were able to identify VRC01-class heavy and light chains in donor C38. Reconstitution of identified heavy and light chain from donor C38 yielded antibodies capable of neutralizing diverse strains of HIV-1.[13]

One of the pitfalls of NGS involving antibody repertoires is the loss of heavy and light chain pairing information. A solution to this pitfall implemented in Antibodyomics1 involves the phylogeny-based pairing of heavy and light chain transcripts.[14] Notably, we observed somatic variants positioned similarly in heavy and light chain phylogenetic trees, when paired and reconstituted, had reduced autoreactivity when compared to mismatched pairings [see Fig. 3C in (14)]. These results indicate that the phylogenetically determined pairings approximate those in vivo, providing a partial solution to the heavy/light chain pairing pitfall.

We also used the Antibodyomics1 pipeline to define the developmental pathway of the 10E8 antibody from donor N152.[15] Heavy and light chain NGS data were processed bioinformatically to obtain somatic variants, and a combination of sequence sieving and CDR3-sequence filters were used to construct maximum-likelihood phylogenetic (ML) trees [see Figs 3 and 4 in (15)]. Alternative ML trees were ranked by parsimony of mutation, and pairing of the inferred sequences was accomplished using the topologies of the heavy and light chain trees. The inferred 10E8 developmental pathway revealed unfavorable interactions with antigen by the unmutated common ancestor (UCA) to be resolved early in development [see Fig. 2C in (15)]. Moreover, we found that early intermediates could bind to MPER peptide, but could not recognize peptide in the context of membrane [see Fig. 6B and C in (15)], a critical attribute for 10E8 neutralization only acquired by latter intermediates. These findings are now being investigated in vaccine studies aimed at eliciting antibodies similar to 10E8—and demonstrate the vaccine impact of Antibodyomics1.

The applicability of the Antibodyomics1 technology continues to expand, and the current Antibodyomics1.1 pipeline is available upon request (PDK, CS, LS). We recently used Antibodyomics1 to identify the sequences of vaccine-induced antibodies capable of neutralizing diverse group 1 and 2 influenza A viruses.[16] In addition to antibodies against other pathogens, Antibodyomics1 can in principle be applied to investigate diverse areas such as autoimmunity, inflammation, and B-cell disorders including lymphomas or IgA nephropathy. In principle, the B-cell record may contain information on the exposure to pathogens throughout the lifetime of an organism. How comprehensively one can sequence the B-cell compartment and how diverse B-cell transcripts are across donor populations remain areas of active research.

3 | ANTIBODYOMICS2

To add a temporal dimension to the analysis of B-cell repertoires, we reimagined the Antibodyomics1 pipeline.[13] Currently, the Software for Ontogenic aNalysis of Antibody Repertoires (SONAR; Antibodyomics2.1) provides a basis of Antibodyomics2. SONAR comprises a suite of tools for processing NGS data from B-cell receptor libraries, including germline V(D)J gene annotation, lineage identification, phylogenetic analysis, ancestor sequence inference, and longitudinal maturation rate calculation [(17) manuscript submitted]. The SONAR Antibodyomics2.1 software is available at https://github.com/scharch/SONAR and a Docker image can be obtained from https://hub.docker.com/r/scharch/sonar/.

A

Antibodyomics	Application	Data	
		Input	Output
1	Antibody identification and lineage analysis	NGS	New antibodies, lineage relationship
2	Longitudinal lineage development	NGS	Temporal relationships in lineage
3	Delineation of neutralization into structurally defined epitopes	Serum neutralization	Donor antibody specificity
4	Heavy-/light-chain paired NGS for antibody identification and lineage analysis	Paired NGS	New antibodies, lineage relationship
5	Antibody-class quantification and class-guided design	Epitopes, NGS, neutralization	Vaccine process quantification
6	Likelihood of gene SHM	NGS	Mutation probabilities
7	Antibody and antigen dynamics	Structure	Antibody interactions
8	Prediction of antibody binding energies	Structure	Binding energies
9	Antibody improvement by surface-matrix screening	Structure	Improved antibodies
10	Comparison of synthetic and natural antibody libraries for vaccine development	Structure, NGS, and natural libraries	Well-configured antibody repertoires

B

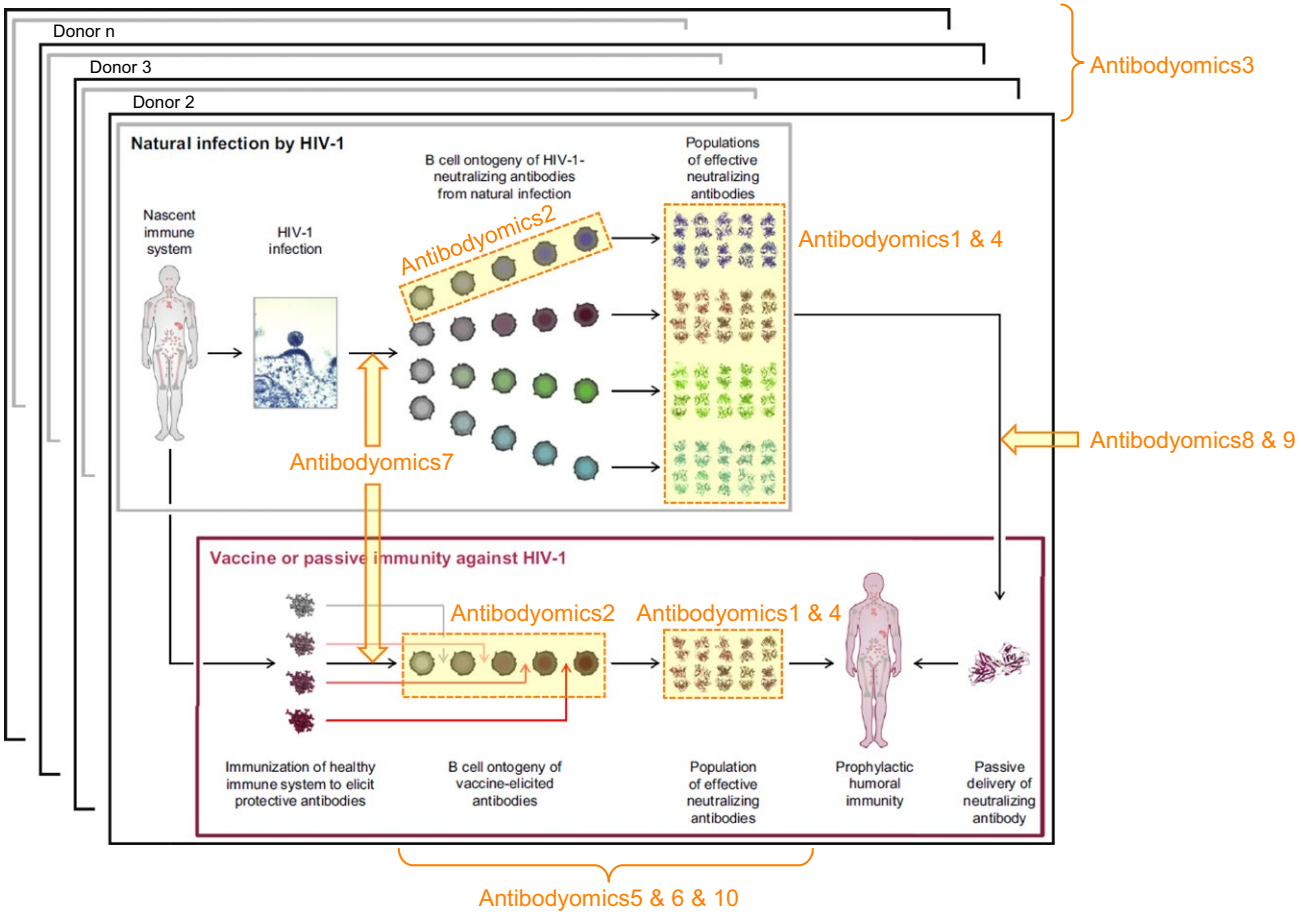

FIGURE 1 Antibodyomics comprise resolution-enhancing tools to understand and to manipulate broadly neutralizing immunity to HIV-1. (A) Applications and data for Antibodyomics1–10. (B) Schematic depiction of antibody lineage development in natural infection (top panel) and active or passive immunization (bottom panel). Areas of resolution-enhancing data generation by the Antibodyomics technologies delineated in this review are highlighted in red. (Adapted from Kwong and Mascola, Immunity 2012)

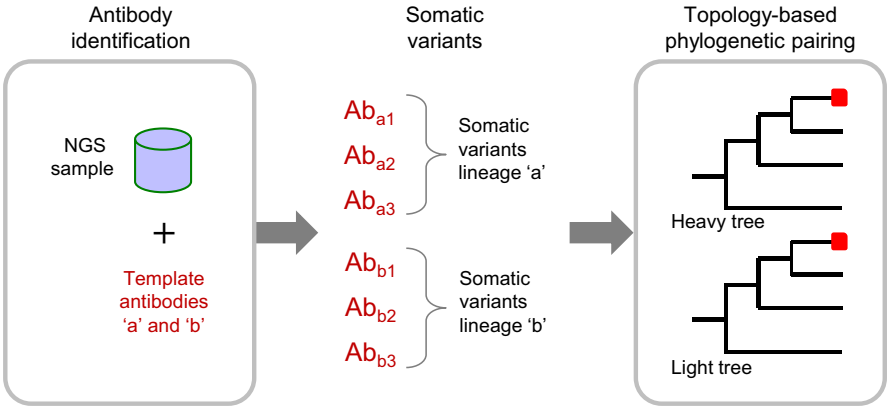

FIGURE 2 Antibodyomics1: antibody lineage identification and developmental analyses via next-generation sequencing (NGS) of B-cell transcripts. Antibodies can be identified based upon similarity to template antibodies (left), these antibodies can be grouped into lineages (middle), and their heavy and light chains can be paired phylogenetically (right)

Longitudinal sampling of the antibodyome provides temporal resolution of antibody repertoire dynamics and of B-cell ontogenies[6,7,18] (Fig. 3A–C). The temporal dimension can also be leveraged to better understand key interactions with virus that helped drive antibody evolution[5–7,19] (Fig. 3D), as well as intrinsic properties of antibody maturation[18,20] (Fig. 3E). Altogether, these analyses provide developmental insights with implications for vaccine design (Fig. 3F) and to enhance the functionality of natural antibodies for clinical use (Fig. 3G).

The first antibody lineage to be studied with Antibodyomics2 was the CH103 lineage, a CD4-binding site-targeting antibody,[7] which was also analyzed in detail with software developed by Kepler and colleagues.[21] Transcripts of this lineage were isolated from four time points spanning 130 weeks, and used to infer the UCA and other intermediates. Analysis of antibody-virus co-evolution revealed that viral diversification around the CH103 epitope was important for the development of breadth and potency of neutralization by the CH103 lineage. A second antibody lineage, CH235, was later identified from the same donor[5,22]; longitudinal studies of this lineage revealed similarities to other CD4-binding site-directed lineages, which may help guide efforts to elicit this class of antibodies via vaccination.[5] Together, the longitudinal information from these three studies provided an in-depth look at the antibody response to HIV in a single donor, including how multiple antibody lineages target similar epitopes and can interact to drive the emergence of neutralization breadth.

We have also used Antibodyomics2 to investigate the origins of the CAP256-VRC26 lineage, a V1V2-targeting broadly neutralizing antibody.[6,19] We were able to identify autologous viruses that likely helped trigger lineage development and showed that toleration of early autologous escape mutants correlates strongly with the acquisition of breadth.[19] We also showed that the CAP256-VRC26 lineage shared developmental features with other V1V2-targeting antibodies, including the creation of a protruding loop by recombination and a propensity for early intermediates to interact with a specific subset of viral strains.[11] In this case, the addition of longitudinal information allowed for the developmental tracking of multiple branches of a broadly neutralizing lineage and for an understanding of why select evolutionary pathways lead to the development of neutralization breadth.[19]

A recent study reported similar longitudinal analyses of the development of a V3-targeting antibody lineage,[8] providing insights into the diversification and selection of B cells responding to HIV as the virus escapes from immune pressure. Antibodyomics2 has been used to investigate the antibodyome in animal models,[23] comparing B-cell responses to vaccination and infection. We have also conducted Antibodyomics2 analyses of antibody lineages for which the available data start only long after initial infection.[18,24] Although we were unable to infer the UCA for these lineages, the longitudinal data provided important insights into the continued evolution of a lineage even after many years and showed how an antibody lineage can split into multiple independently evolving clades.[18,24]

The Antibodyomics2 technology can be utilized for monitoring B-cell responses to vaccination. In non-human primates, for instance, we showed that while the choice of adjuvant could impact the ELISA titer of the vaccine response, it did not increase the levels of somatic hypermutation (SHM) in vaccine-elicited lineages, which remained far below those seen in models of infection.[23] Antibodyomics2 was used to monitor the elicitation and expansion of influenza broadly neutralizing antibodies in response to immunization with an H5N1 strain [(16) and manuscript in preparation]. We have also used longitudinal information to analyze the rate of antibody evolution. HIV-1 evolves at an extraordinarily rapid rate; analysis of three broadly neutralizing antibody lineages showed that antibodies can match this rate of evolution even after years of maturation, while new lineages can evolve at rates up to 15-fold faster.[20] The decrease in evolutionary rate over time appears to be partly due to Darwinian selection effects and the depletion of AID hotspot motifs in the antibody sequence. Thus, Antibodyomics2 can provide insights into basic immunological functions.

Antibodyomics2 has revealed important commonalities in the development of HIV-1 broadly neutralizing antibody classes that target the CD4-binding site, the V1V2 region, and glycan-V3 epitopes. The ontogenic pathways for each of these classes are distinct.[25] Ongoing work aims to use the Antibodyomics2 technology to elucidate the pathways taken by broadly neutralizing antibodies targeting other sites of viral vulnerability. We are continuing to use Antibodyomics2 to assess novel vaccine designs and investigate the mechanisms underlying evolutionary rate changes. We also plan to further develop the capability of Antibodyomics2 to mine the rich data contained in longitudinal samples.

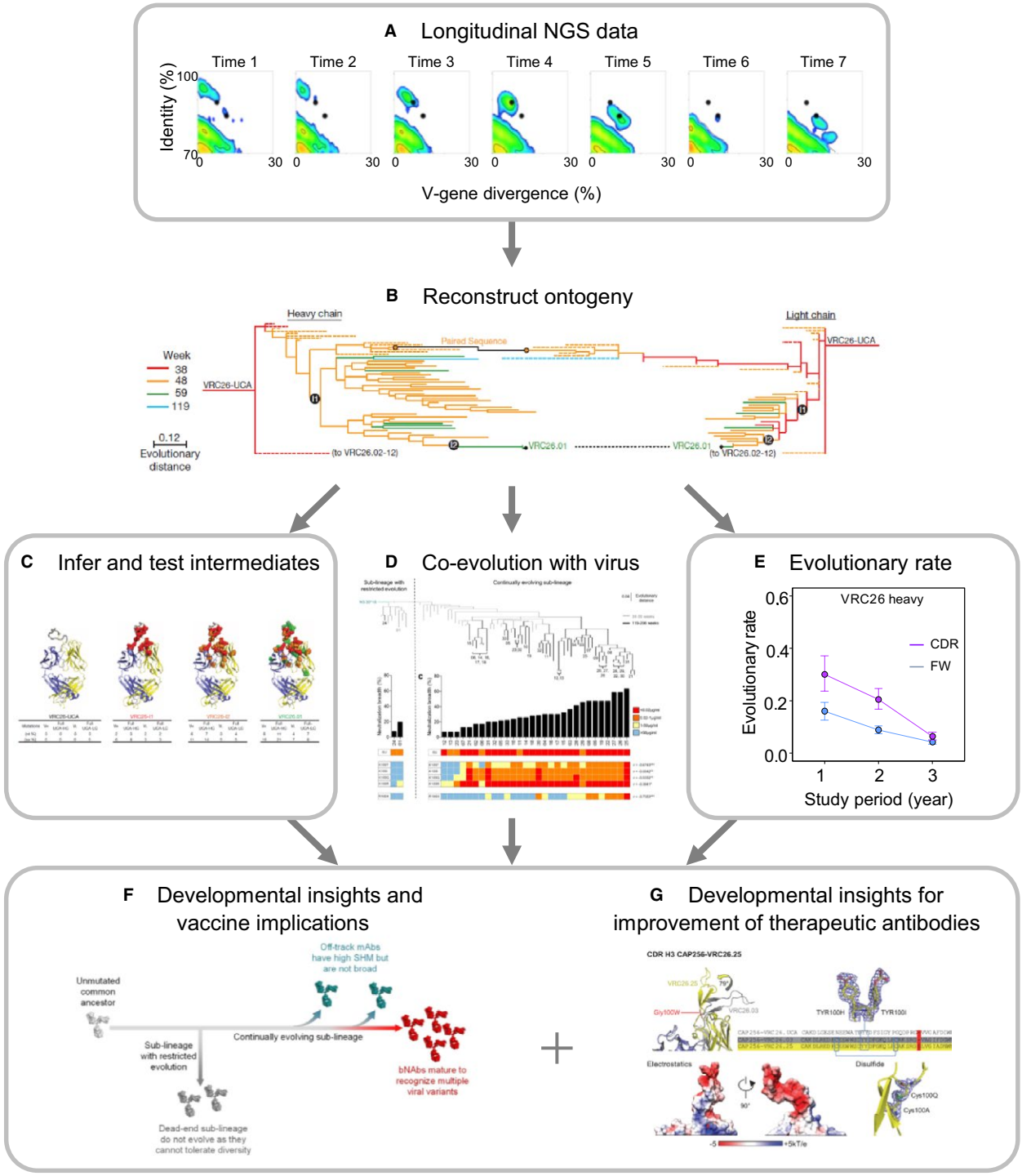

FIGURE 3 Antibodyomics2: Translating longitudinal samples into developmental pathways and maturation rates. Antibodyomics2 focuses on the longitudinal development of neutralizing antibody lineages, using (A) multiple deep-sequencing experiments across time (B) to reconstruct detailed ontogenies. These ontogenies can be used (C) to infer UCA and early intermediates, providing details on the emergence of desired features, such as heterologous neutralization; (D) to study co-evolution between an antibody lineage or lineages and the eliciting pathogen, revealing critical interactions that elicited the desired features; and (E) to observe intrinsic properties of antibody maturation, such as the evolutionary rate and how it changes over time. All of these analyses provide insights into antibody development which can be leveraged (F) to design better vaccines and (G) to improve naturally occurring antibodies for use as therapeutics. Panels A, D, and F are adapted from Bhiman et al., Nature Med 2015. Panels B and C are adapted from Doria-Rose et al. Nature 2014. Panel E is adapted from Sheng et al. PLoS Comput Biol 2016. Panel G is adapted from Doria-Rose et al. J. Virol 2016

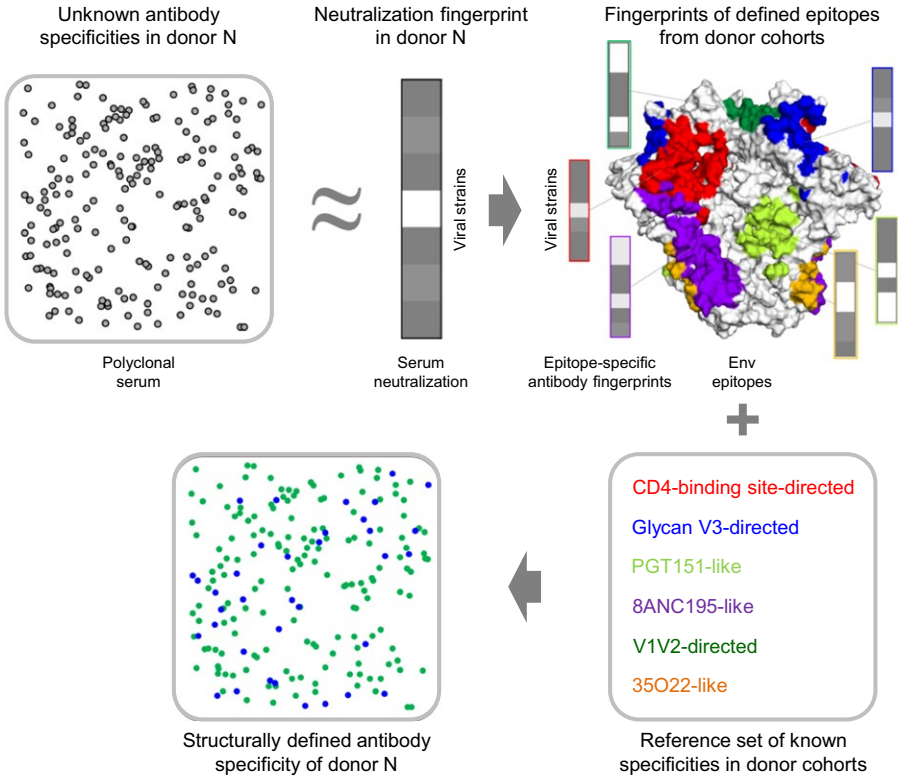

FIGURE 4 Antibodyomics3: Delineation of neutralization specificities by deconvolution of neutralization fingerprint into structurally defined epitopes. (top left) Polyclonal serum analyzed by neutralization (top middle, heatmap for a diverse set of HIV-1 strains shown as white to dark gray for increasing neutralization potency), compared to epitope-specific monoclonal antibody neutralization fingerprints over the same set of strains (top right, with reference set of known fingerprints and corresponding epitopes indicated on a model of the HIV-1 trimer spike). (bottom right) Deconvolution to reference set of known antibody specificities, defines the antibody specificities in donor N (bottom left)

4 | ANTIBODYOMICS3

Antibody responses to HIV-1 infection are typically complex and polyclonal, targeting a variety of epitopes on the Env glycoprotein.[26] However, there is limited evidence that a large number of diverse antibody lineages can develop neutralization breadth within a given individual.[6,18,27,28] Current studies indicate that each individual typically develops one or a small number of broadly neutralizing antibody lineages targeting a limited set of Env epitopes. Indeed, while the general HIV-specific antibody diversity is extremely high, only a small set of sites of vulnerability on Env have been discovered to date.[6,9,27–38] Antibodies against these sites have been found in multiple donors, indicating that such broadly neutralizing epitopes represent a general mode of effective Env recognition by antibodies.[39]

Therefore, for a large fraction of HIV-infected individuals with broadly neutralizing polyclonal sera, the primary contributor to serum neutralization is a combination of known antibody specificities. Most experimental techniques for characterizing the epitope specificities targeted by polyclonal sera involve knowledge about previously identified antibodies and epitopes: some of the standard techniques include binding competition with monoclonal antibodies, differential binding to wildtype vs epitope-knockout antigen, and binding to epitope-peptides.[40–48] These techniques are typically time-consuming and may not always possess a sufficient level of specificity to correctly identify the component antibodies in a polyclonal response.[49]

As an alternative to purely experimental techniques, we developed a data analysis-based computational method for serologic analysis of HIV-infected individuals.[49] The method works as follows: we define an antibody neutralization fingerprint as the potency-dependent pattern with which a set of diverse HIV-1 strains is neutralized by a given monoclonal antibody. As we have shown that neutralization fingerprints are epitope-specific[49] (i.e. two antibodies targeting the same epitope have similar fingerprints, whereas two antibodies targeting different epitopes have different fingerprints), we can use mathematical techniques to deconvolute the neutralization of a polyclonal serum as a combination of the fingerprints for the known broadly neutralizing monoclonal antibodies (Fig. 4). Using this approach, we can predict the relative contribution to polyclonal neutralization by each of the known major broadly neutralizing monoclonal antibody specificities. As serum neutralization data are obtained as one of the first steps in the characterization of any given serum, the neutralization fingerprinting approach does not place an additional experimental burden, and can therefore be extremely efficient compared to standard experimental serum mapping techniques.

To date, the neutralization fingerprinting approach has shown utility in a number of areas. (i) First, it can be used to efficiently screen donor sera for target antibody specificities. For example, in (49) we analyzed a large set of sera for VRC01-like antibodies, predicted a strong VRC01-like signal for a previously uncharacterized donor, 127/C, and confirmed the predictions through antibody isolation and characterization studies; (ii) Furthermore, the fingerprinting approach can be used to determine the relative contribution to overall serum neutralization by a particular antibody lineage. For

example, for donor CAP256, we observed that the CAP256-VRC26 lineage was the primary contributor to serum neutralization, likely indicating the lack of other major broadly neutralizing specificities.[6] Alternatively, an identified antibody lineage may not contribute substantially to serum neutralization, which may be dominated by other, yet to be discovered, specificities. The fingerprinting analysis can therefore lead to a better understanding of the complexity of polyclonal sera and the effects different component antibodies can have on polyclonal neutralization; (iii) While the sera from many donors appear to contain antibodies that are similar to already known specificities, new epitopes continue to be discovered.[27,30,33] When an antibody targeting a new epitope is discovered, the neutralization fingerprinting approach can be used to screen HIV-infected sera for neutralization signals matching the new epitope, therefore both guiding efforts for identifying additional antibodies targeting this epitope, as well as providing an estimate of the overall prevalence of such antibodies in infected donors. The latter can be an important factor when evaluating targets for epitope-specific vaccine design, with the idea that epitopes that are more prevalent in the population may be better vaccine candidates;[27] and (iv) In addition to the analysis of polyclonal sera, the neutralization fingerprinting approach has been useful for the characterization of the epitope specificities of monoclonal antibodies. For example, such analysis can reveal connections between antibody genotype, structural mode of recognition, and neutralization fingerprint similarity.[39] Furthermore, the fingerprinting analysis can help trace the neutralization evolution of different members of the same antibody lineage.[5,18]

Overall, the neutralization fingerprinting method for serologic analysis of antibody responses to HIV-1 infection has proven to be a useful tool in the armamentarium of techniques available to researchers.[5,6,18,27,30,32,39,49–51] The source code for Antibodyomics3 is provided as a supplement to Georgiev et al.[49], and updated versions can be requested from the authors. Additional improvements in the computational methodology should further enhance the usefulness of the fingerprinting approach by improving the accuracy and reliability of the predictions. As neutralization data for large HIV-infected cohorts become available, the fingerprinting technology can play an important role in deciphering the elicitation of broadly neutralizing antibody specificities at the population level, and can therefore help guide efforts for epitope-specific vaccine design. More generally, as indicated by initial analysis of influenza antibody fingerprints,[16] the fingerprinting technology should be extendable to other viruses with high levels of antigen sequence diversity, such as influenza A and hepatitis C viruses.

5 | ANTIBODYOMICS4

Antibodyomics4 applies NGS to the analysis of paired antibody heavy and light chain variable regions (V_H and V_L, respectively). We utilize a technology that links V_H and V_L sequences prior to high-throughput sequencing.[52] The ability to analyze in high throughput both heavy and light chain sequences permits deconvolution of the ontogeny of HIV broadly neutralizing antibodies and rapid discovery of new antibody variants, both in the context of HIV/SHIV infections and in response to vaccination.

Conventional short-read NGS technologies fail to pair V_H and V_L because antibody heavy and light chain genes are on separate chromosomes and expressed on distinct mRNA strands.[53] In the early years of high-throughput antibody sequencing, this pairing problem remained fully unresolved and researchers relied on single-cell RT-PCR and Sanger sequencing for antibody discovery. To address this issue, a new technology was developed that used poly(dT) magnetic beads to capture and purify mRNA from a single B cell (Fig. 5A). This method links V_H and V_L sequences by using overlap extension RT-PCR prior to NGS, providing the capability to analyze up to 10 million B cells in a single analysis.[52,54,55]

The first use of high-throughput paired V_H:V_L sequencing with HIV-1-directed broadly neutralizing antibodies helped elucidate the early ontogeny of the CAP256-VRC26 antibody, which targets the V1/V2 region of the fully assembled HIV trimer.[6] In this study, longitudinal V_H:V_L data were used to align the distinct V_H and V_L phylogenetic trees and identify the earliest known paired CAP256-VRC26 V_H:V_L sequence (48 weeks postinfection, Fig. 5D). V_H:V_L sequencing was later expanded for the analysis of other donors that expressed V1/V2-targeting broadly neutralizing antibodies, and in particular it provided additional information on the diversity and ontogeny of two different bNAb lineages in Donor 0219 from a single sample analysis.[11,28] This work helped reveal the structural similarities between different antibody solutions to broad V1/V2-based HIV neutralization and outlined new strategies for V1/V2-targeting vaccine designs. Paired V_H:V_L sequencing has also been applied to address several problems beyond HIV.[56,57]

Recent applications of paired V_H:V_L sequencing have included an analysis of the elicitation of VRC01-class antibody lineages in transgenic mouse models.[61] VRC01-class antibodies have strict requirements of both heavy and light chain genes and V_H-only or V_L-only sequencing is insufficient for high-confidence VRC01-class antibody identification.[12] This work revealed that particular immunization regimens with the eOD-GT8 germline-targeting minimal antigen[59,60] were able to increase the fraction of VRC01-like antibodies in class-switched splenic B cells by approximately 200-fold.[61] Interestingly, the finding that only the class-switched sequences among B-cell splenocytes were enriched for VRC01-like characteristics (which has also been reported in low-throughput analyses of a different mouse model[60]) suggested that class-switched B cells may be the optimal biological compartment to analyze for HIV bNAb precursor enrichment in transgenic murine vaccine models.

Other ongoing efforts have focused on targeted gene priming based on heavy and light chain genes to enrich sequence data for variants of a known antibody lineage. This new technology has resulted in the isolation of over 2000 complete V_H:V_L sequences of the VRC34 antibody class, which targets the fusion peptide as a new vulnerable epitope on the HIV trimer.[30] V_H:V_L genetic lineage targeting is

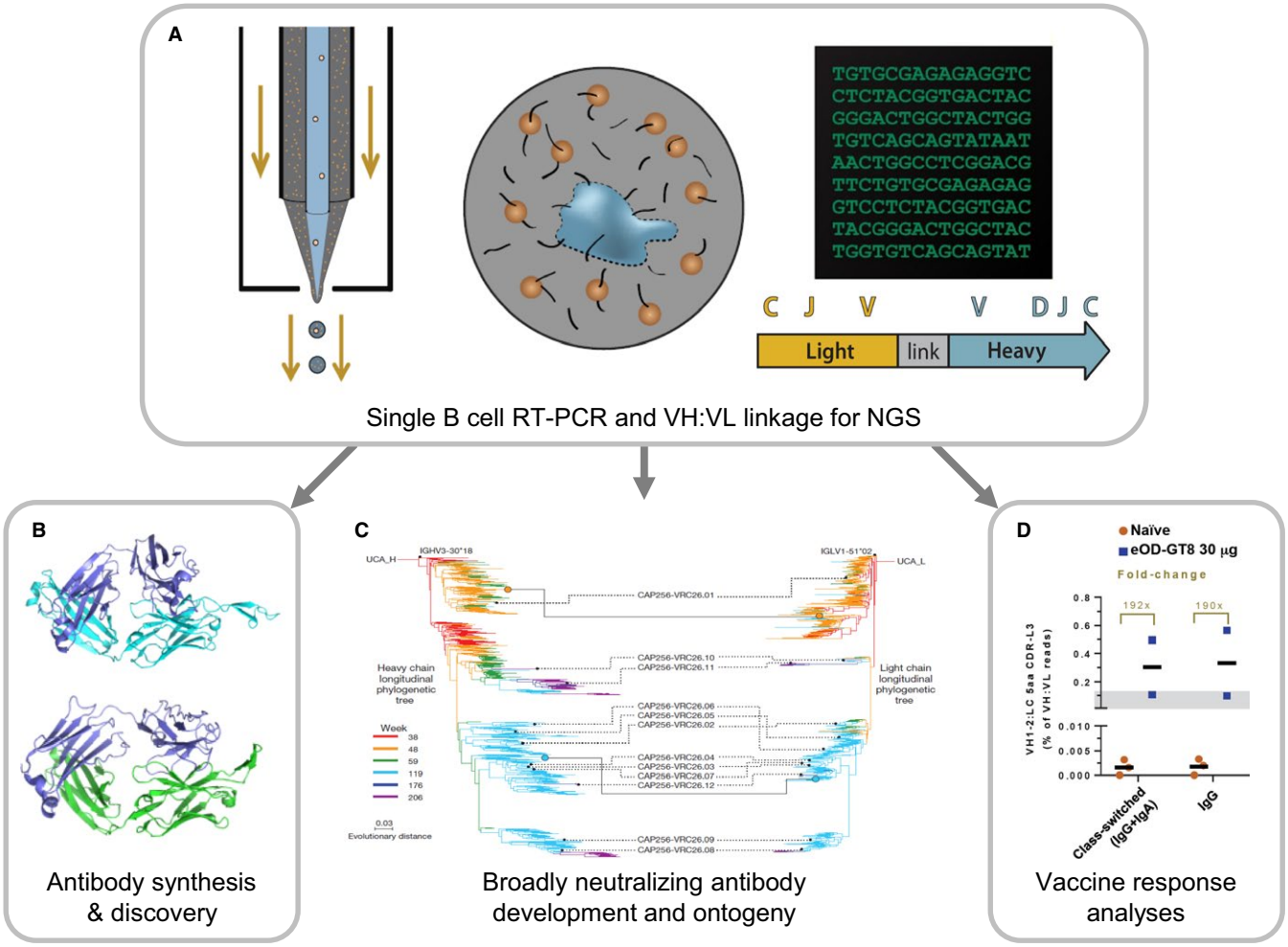

FIGURE 5 Antibodyomics4: Paired sequencing of antibody heavy and light chains for discovery of new antibodies, quantitative analysis of vaccine responses, and elucidation of B-cell ontogenies. (A) B cells are isolated using a flow-focusing nozzle (*left*) to encapsulate single B cell inside emulsion droplets along with lysis buffer and poly(dT) magnetic beads to capture paired heavy and light chain mRNA (center). After linkage overlap extension RT-PCR, an approximately 850 bp linked VH:VL construct is sequenced using any of several NGS technologies, including Illumina MiSeq 2 × 250 or 2 × 300 and the Pacific Biosciences long-read platforms. (B) Antibodyomics4 has been applied to discover new antibody families, including the CH04 and VRC26 broadly neutralizing families shown here. (C) Paired VH:VL sequencing has been used to quantify the enrichment of VRC01-class antibody precursors using various immunization protocols in transgenic mice, as a testable model of vaccine development. Importantly, the VRC01-class antibodies in particular has known requirements of both heavy and light chains, and therefore high-throughput analyses must incorporate paired heavy:light information for robust VRC01-precursor identification. (D) Antibodyomics4 is being used to elucidate the ontogeny of HIV-directed broadly neutralizing antibodies and to help in the design of vaccine antigens. Panel A adapted from DeKosky et al., Nat Medicine 2015; panel B adapted from Goramn et al., NSMB 2016; panel C adapted from Doria-Rose et al., Nature 2014; panel D adapted from Tian et al., Cell 2016

particularly promising for the identification of early antibody variants from longitudinal samples (both human and primate), because early antibodies may not bind to B-cell probes with sufficient affinity for single-cell sorting, and also may be weakly or narrowly neutralizing and therefore refractory to identification via single-cell microneutralization. Sequencing of a comprehensive set of antibody variants without the biases introduced by B-cell probe binding or microneutralization is helping generate a more complete genetic understanding of early antibody development to accelerate the rational design of HIV vaccines.[62]

The ability to glean new insights regarding HIV-directed broadly neutralizing antibody development using high-throughput paired heavy and light chain analysis has opened up several exciting possibilities. High-throughput paired heavy:light sequences have permitted massive computational structure analysis to interrogate antibodies in the sequenced repertoire with genetic characteristics of known broadly neutralizing antibodies.[63] Paired heavy and light chain sequences permit antibody discovery via sequence-based identification and testing of new variants based on known genetic characteristics. These ongoing efforts in the high-throughput interrogation of paired heavy and light chain antibody sequences are broadening our understanding of antibody diversity and development, both in natural infection and in response to experimental vaccination.

6 | ANTIBODYOMICS5

Multidonor class antibodies (sometimes called stereotypic or public clonotypes) arise in multiple donors and share common features of antigen recognition and B-cell ontogeny.[12] Such antibodies may be more amenable targets for immunogen design because they are, by definition, reproducible.[12] A growing number of multidonor class broadly neutralizing antibodies have been identified for HIV-1,[5,11,12] and for other diseases [e.g. influenza[16,64,65] and dengue virus[66]]. In addition to providing vaccine templates, quantification of the prevalence of these multidonor class broadly neutralizing antibodies in sera could potentially allow their use as biomarkers in vaccine development (Fig. 6).

In Antibodyomics5, we delineate the frequency of multidonor class antibodies in target populations,[12] derive sequence signatures for the multidonor class,[12,39] and use these signatures to quantify the prevalence of the multidonor antibody class in the B-cell compartment.[12,39] This information can then be used to design or to select immunogens that maximize the elicitation frequency of target antibodies.[59–61,67]

In the first application of Antibodyomics5, we used both frequentist and Bayesian approaches to estimate the frequency of recombination events suitable for the development of VRC01-class antibodies.[12] These estimates indicated several recombinations to occur each day, with the roadblock to broad neutralization being the extraordinary SHM required for function.[40] Additional multidonor classes have now been identified including CD4-binding-site antibodies from the VH1-46 germline[5,39] and V1V2-directed antibodies characterized by a protruding anionic loop formed by recombination.[6,11] More recently, mouse models incorporating knock-in germline genes has demonstrated the priming of multidonor VRC01 class antibodies in response to germline-targeting immunogens.[58,60,67] We have also applied Antibodyomics5 to the quantification of multidonor class antibodies that neutralize both group 1 and group 2 influenza A viruses.[16] We delineated antibody sequence signatures and quantified antibodies that satisfy these sequence signatures in NGS datasets from multiple donors with different vaccination histories.[52,68] Notably, one vaccination regimen from the VRC 310 trial involving immunization with a diverse H5 hemagglutinin increased the frequency of two multidonor classes by over 100-fold,[16] despite titers too low to detect in sera. This demonstrates the exquisite sensitivity of B-cell quantification, where one transcript in about one million can be detected—a more sensitive quantification than allowed by assessment of serum neutralization.

Overall, the quantification afforded by Antibodyomics5 of multidonor antibodies and their elicitation by vaccination enables a quantitative approach to B-cell ontogeny-based vaccine design. Currently, these quantifications are made on an ad hoc basis, generally using frequentist approaches and incorporating other antibodyomics engines, such as Antibodyomics1 or Antibodyomics2 to select transcripts or to refine sequence signatures. Such quantifications can also be made with animal models incorporating knock-ins of human immunoglobulin genes and their developmental intermediates, providing a powerful tool to monitor and optimize the development of ontogeny-based vaccines.

7 | ANTIBODYOMICS6

Mature antibodies in the human antibody repertoire usually contain approximately 5–10% SHM at the nucleotide level and approximately 10–20% SHM at the amino acid level; in contrast, anti-HIV-1 broadly neutralizing antibodies can have up to approximately 20–35% nucleotide SHMs.[6,7,18,22] To date, the maturation pathways of several broadly neutralizing antibody lineages have been characterized.[6–8,18,22] Extensive efforts are being invested in the development of vaccines to mimic the natural antibody-virus co-evolution process, with the goal of eliciting similar broadly neutralizing antibodies.[69] To quantify expectations, we developed Antibodyomics6 to investigate factors that modulate the probabilities of antibody SHM and to provide models by which to evaluate the likelihood of a particular pattern of antibody mutation. This information can then be used to predict the likelihood of specific mutations in response to vaccination and infection, allowing for quantitative evaluation of vaccine strategies through comparison with simulations (Fig. 7).

Antibody SHM is mainly initiated by activation-induced cytosine deaminase (AID),[70,71] which mutates the antibody variable region with a rate of approximately 10^{-3} mutations per site per B-cell generation.[72] Studies have shown that AID preferentially mutates a position at or around hotspot motifs,[73–75] which are more frequently observed in the CDR regions than the framework regions.[74,76,77] Moreover, targeted nucleotide positions have increased chances to generate transition mutations (A↔G, C↔T), which are generated more frequently than transversion mutations, such as A↔C, A↔T, C↔G, or G↔T.[74] Thus, antibody-gene mutability (the non-uniform mutation frequency and nucleotide substitution bias within the antibody variable region) leads to a distribution of substitution probabilities at the nucleotide level. As a consequence, patterns of amino acid substitutions are also affected.[78] The use of NGS to sample antibody repertoires provides a sufficiently large basis set of antibody transcripts from which a mutation spectrum for each position of each germline V gene can be compiled (Fig. 7, left and middle). By applying this method, we found that antibodies originating from the same germline V gene share a substantial percentage of amino acid mutations in the V region, irrespective of antigen specificity.[5] This suggests that intrinsic mutability modulates the positions and identities of substitutions in the antibody variable region. This also suggests that functional selection is not the sole (or even the dominant) factor regulating antibody maturation process. These conclusions are supported by our recent observations in an HIV Env vaccination study that high-frequency somatic mutations observed in elicited antibodies were also high-frequency mutations in non-HIV neutralizing antibodies.[61] One implication of these findings is that mutations with very low occurrence frequency, but are functionally critical, may constitute dominant roadblocks to elicitation.[79]

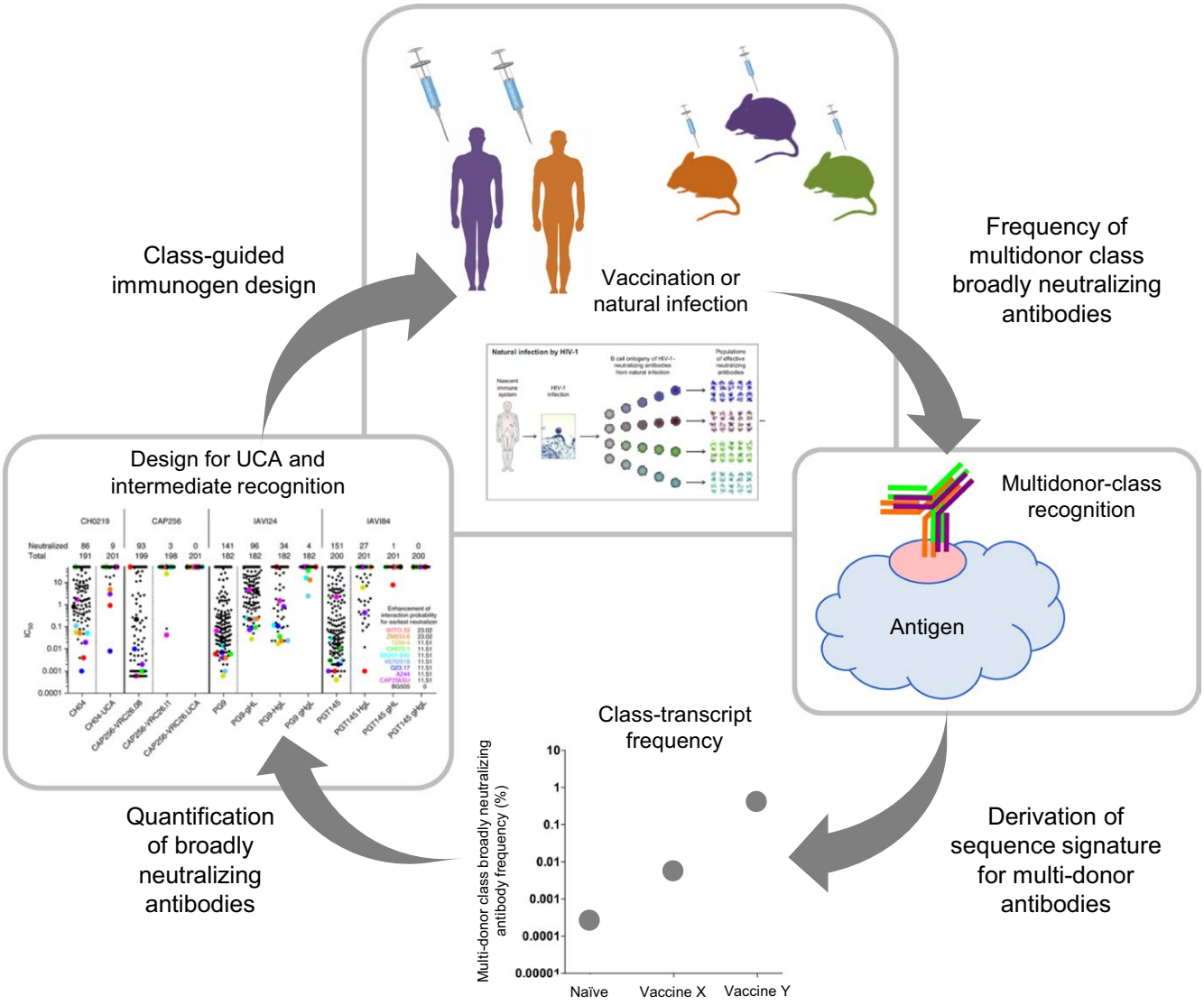

FIGURE 6 Antibodyomics5: Quantification of multidonor antibody responses and class-guided design. Broadly neutralizing antibodies that arise in multiple donors and share common features of antigen recognition and B-cell ontogeny may have utility as vaccine templates due to the potential for similar antibodies to be elicited by a common immunogen in the general population. In Antibodyomics5, we delineate the frequency of multidonor class antibodies, derive sequence signatures, and quantify the prevalence of a given multidonor antibody class in the B-cell compartment. This information can be used to guide the design or the selection of vaccine candidate aimed at "re"-eliciting the target broadly neutralizing antibody classes. Panels adapted from Kwong and Mascola, Immunity 2012 and Gorman et al., NSMB 2016

Ongoing developments include construction of new in silico models of the mutation space to achieve improved simulation of the antibody maturation process. Currently, the construction of mutational profiles for Antibodyomics6 is implemented as an extension of Antibodyomics2.1, through scripts selecting appropriate antibody repertoires to build gene-specific mutational profiles.[5] Antibodyomics6 may shed light on mechanisms modulating antibody maturation, and future versions would be expected to provide quantitative estimates of the likelihood of eliciting by vaccination mutational patterns similar to those observed in broadly neutralizing antibody lineages developed in response to natural infection.

8 | ANTIBODYOMICS7

Molecular dynamics (MD) simulations provide a means to extend atomic-level structural information, which can be helpful in multiple cases including (i) the docking of an antibody to an antigen when the high-resolution complex structure is unavailable; (ii) the conformation of the antigen, especially for mobile regions where crystallographic information provides only a snapshot; and (iii) the accessibility of antibody to protein surface, especially with glycopeptide antigens. In Antibodyomics7, we apply MD to the analyses of antigen conformation and of antibody-antigen interactions (Fig. 8).

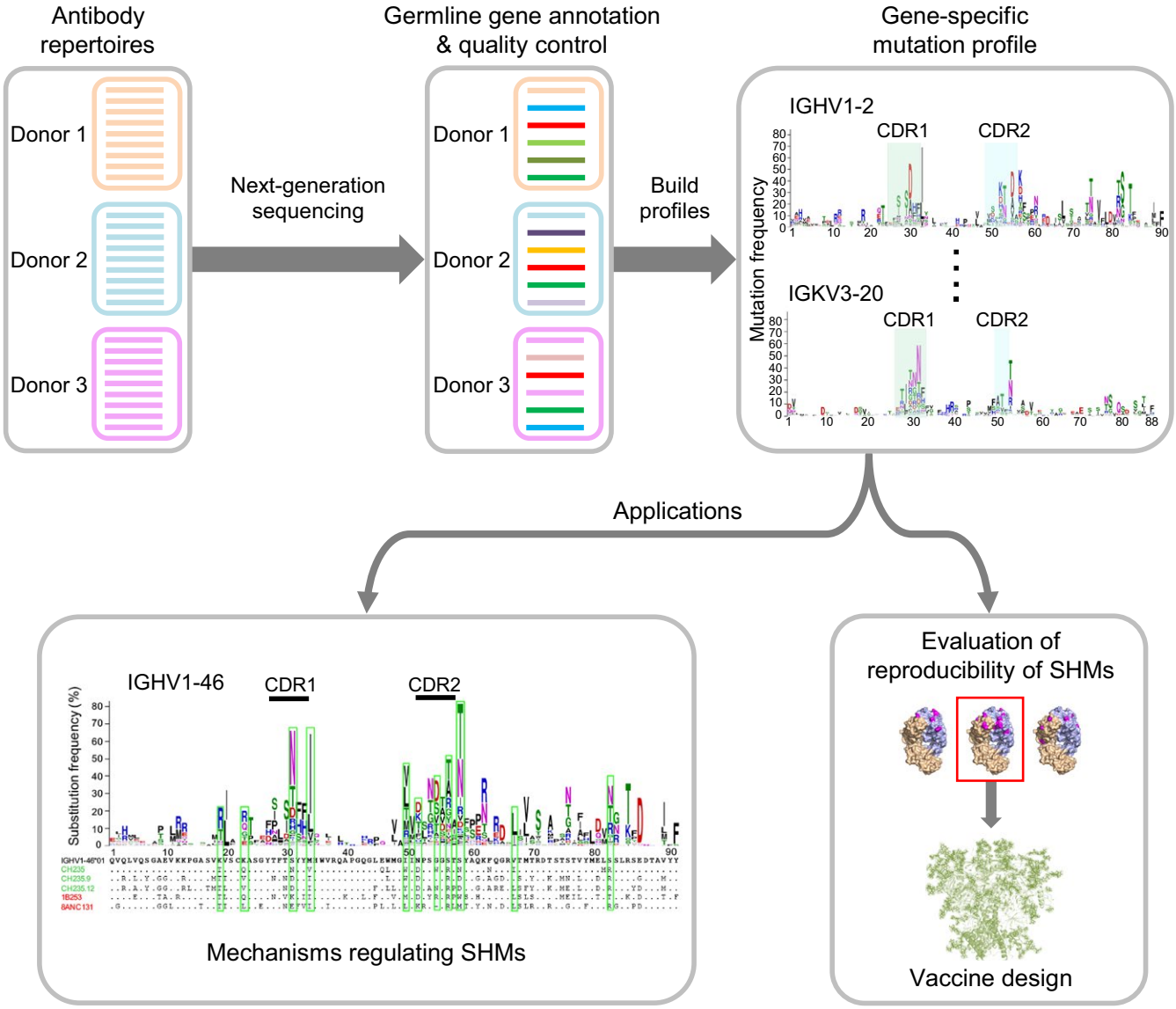

FIGURE 7 Antibodyomics6: Impact of gene mutability and substitution bias on the development of broadly neutralizing antibodies. In Antibodyomic6, we focused on the investigation of position-specific mutation probability and substitution bias of the antibody variable region (Top row), the knowledge derived from which will be used to address a series of immunological questions. Briefly, we sequence antibody repertoires from multiple donors with NGS and process transcripts to identify antibody clones (Top row left and middle). We choose a representative transcript from each clone and construct a mutation profile for each germline V gene using all clones identified (Top row right). We then use the mutation profiles and in silico models to understand the mutation patterns observed in bnAbs and to evaluate the likelihood of re-eliciting similar mutations by vaccination (Bottom row)

Experimentally determined atomic-level structures can be used to guide the design of candidate vaccine antigens. However, obtaining high-resolution structures of antibody-antigen complexes may in some cases be challenging, and computational techniques can be used to integrate data from different experimental methods to derive atomic-level biologically relevant models. In particular, molecular dynamics flexible fitting (MDFF)[80] has proven to be an effective method for fitting atomic models into low-resolution cryo-EM density maps.[81] To create an atomistic model of the glycosylated HIV-1 Env trimer in complex with antibody CAP256-VRC269, we used an X-ray crystal structure of the BG505 SOSIP.664 Env trimer (PDB ID: 4TVP) and of the broadly neutralizing antibody CAP256-VRC263 (PDB ID: 4OD1) with a negative stain 3D reconstruction of the complex[82] along with a combination of homology modeling and MDFF to obtain an atomic-level model of the complex. This model revealed specific interactions between the trimer and antibody, which we validated experimentally using mutagenesis and hydrogen-deuterium exchange mass spectrometry [see Fig. 4 and Fig. S3 of (11)].

In terms of antigen conformation, dynamic properties may have significant effects, which may not be apparent in static X-ray structures. In particular, highly mobile regions and molecules exposed to

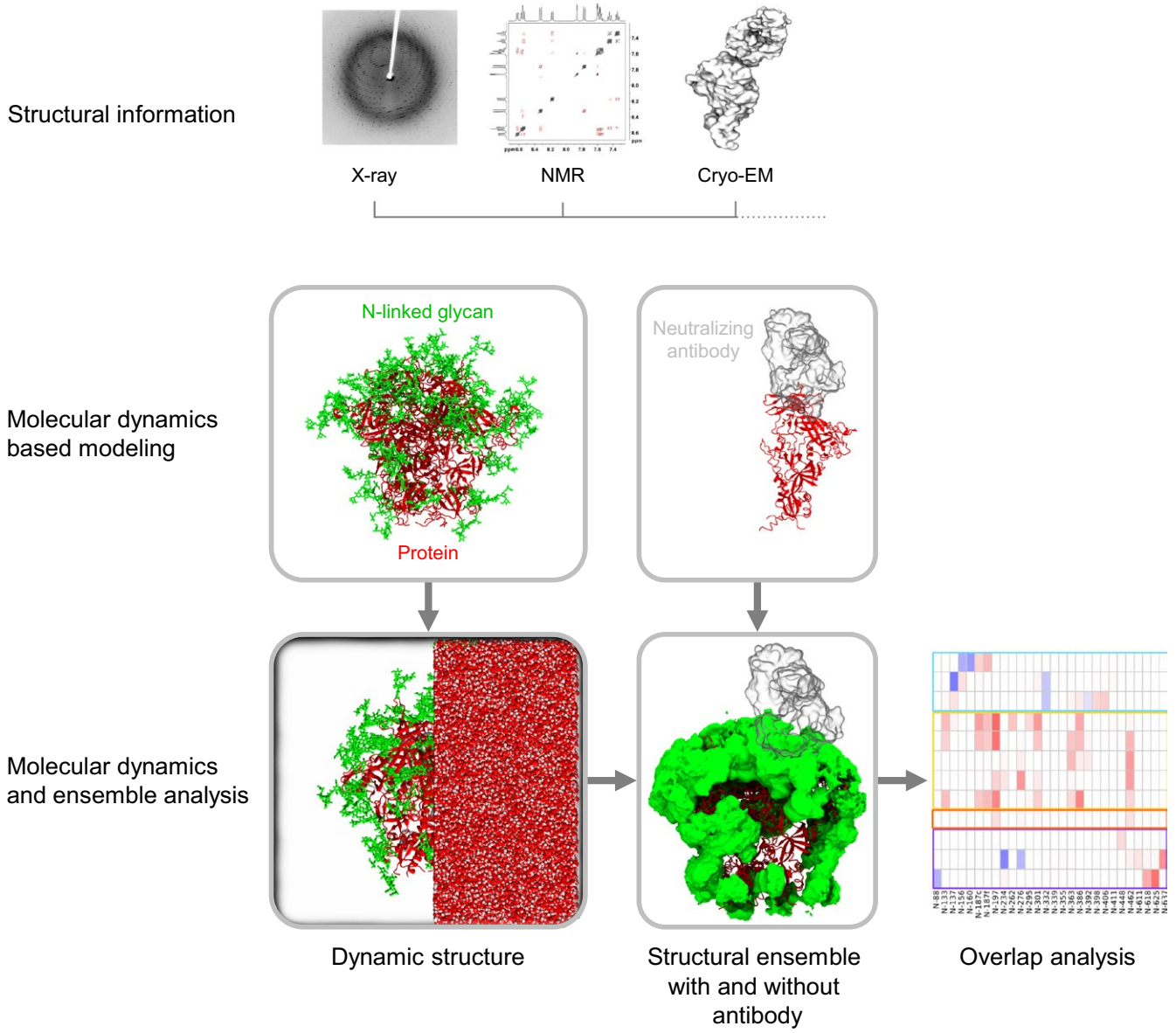

FIGURE 8 Antibodyomics7: Molecular dynamics in antibody docking and overlap analysis. Molecular dynamics simulations provide a means to further extend atomic-level structural information by integrating available experimental data (top row). The generated models (middle row) and structural ensembles on the microsecond time scales can be analyzed to provide insight into antibody interactions (bottom row). Overlap analysis panel adapted from Stewart-Jones et al., Cell 2016

solvent are rarely defined structurally, and classical MD has proven to be a valuable tool to probe dynamics of such molecules[81,83–85]. For example, the eight amino acids at the N-terminal of the fusion peptide of HIV-1 Env in its closed prefusion state are not visualized in most crystal structures.[32,86,87] However, we observed antibody VRC34 to recognize these eight residues as the primary determinant of binding.[30] MD of the prefusion closed Env trimer revealed this eight amino acid region to be highly mobile with average solvent accessibilities in the MD simulations that correlated with VRC34 recognition [see Fig. 4B and C (30)].

MD has been particularly useful in assessing the impact of the HIV-1 glycan shield.[87] *N*-linked glycosylation is important in post-translational modifications of the HIV Env trimer and account for nearly 50% of its mass.[88–91] We used MD to characterize the properties of the glycan shield and to identify specific glycans that obstruct antibody access to known sites of vulnerability.[87]

Altogether, Antibodyomics7 utilizes molecular modeling, molecular simulations, and their interplay with available experimental data to advance our understanding of antibody interactions with antigen in active and passive immunity. What is the relationship between the accessibility of an exposed antigen surface and its immunogenicity? To what extent do glycan clashes alter antibody elicitation? What is the relative immunogenicity of exposed protein surface and exposed *N*-linked glycan? These and related questions may be addressed with

Antibodyomics7 as (i) site-specific immunogenicity is determined, (ii) additional crystal structures are determined, and (iii) additional MD simulations are analyzed.

9 | ANTIBODYOMICS8

One of the fastest growing areas of biotherapeutics is the use of monoclonal antibodies, especially in the treatment of cancer, autoimmune disorders, and infectious diseases.[92] One roadblock to achieving effective therapies is the long and costly process of antibody optimization, which may require years of effort. Computational approaches may aid in this process, but their effective implementation has been difficult because current tools fail to yield predictions of sufficient accuracy. Antibodiomics8 aims to develop and apply computational methods that provide accurate biophysical characterization and improved functional efficacy of antibodies directed against HIV-1.

As discussed in the prior section, MD simulations are widely used to define the conformational space of macromolecules,[83] and free energy perturbation (FEP) methodology constitutes one of the most rigorous MD-based approaches to determine the binding free energy for complex biological systems.[93] FEP exploits the fact that free energy is a state function, so that the difference in free energy between two states does not depend on the path between the two states. As such, FEP uses alchemical transformations to mutate one ligand into another in the presence and absence of binding partners, and then uses the thermodynamic cycle to compute the free energy difference between the physical states of the system (Fig. 9). With improvements in force fields and sampling algorithms,[94–96] FEP methods have been shown to produce robust, accurate predictions of changes in binding affinities for modifications of small molecule ligands binding to protein receptors.[97–100] The goal of Antibodyomics8 is to achieve similar accuracy for predicting changes in binding affinity for antibody-antigen complexes upon mutation of one or more antibody residues—a considerable challenge given the complexity of antibody-antigen interfaces.

Initial method benchmarking and optimization studies performed using retrospective alanine scanning data of antibody-paratope residues yielded encouraging results (A. Clark, unpublished data). The effects of antibody interface mutants for three VRC01-class antibodies [VRC01, VRC03 (40), and VRC-PG04 (9)] on HIV-1 gp120 binding were quantified by surface-plasmon resonance (SPR) measurements. By using FEP simulations, we were able to reproduce the experimental changes in binding affinities for alanine mutants to gp120 (average error less than 1 kcal/mole)—the level of accuracy sufficient to provide meaningful prospective predictions of binding free energies for antibody improvement.

Currently, we are extending the FEP approach to characterize other classes of broadly neutralizing antibodies directed against different epitopes on the HIV-1 viral spike. Antibodyomics8 is expected to provide detailed information on the per-residue level of 'hot spots'[101] responsible for HIV-1 gp120 glycoprotein binding and characteristics of broadly neutralizing antibodies and, thus, guide the design of improved antibodies for clinical use. Antibodyomics8 is also expected to expand our current understanding of antibody diversity and accelerate rational vaccine design through the use of quantitative computational tools.

10 | ANTIBODYOMICS9

As an effective HIV-1 vaccine may be elusive, passive immunization with broadly neutralizing antibodies might be an alternative strategy for generating protective immunity to HIV-1. Advancement of a candidate antibody into manufacturing and clinical trials often requires improvement of multiple molecular properties. Example of such properties include potency, optimization of which can lead to reduce dosage, thus reducing the cost of the therapy; solubility, improvement of which prevents antibody aggregation and leads to better manufacturability; and autoreactivity, reduction of which can increase antibody half-life and prevent interaction with self-antigens.

Previously we showed that arginine scanning of antibody surfaces could be used to identify paratope residues that interact with antigen.[11,102] In Antibodyomics9, we extend this methodology to a surface-matrix approach in which a collection of individual surface mutations are generated and their properties are assessed (Fig. 10). Selected mutations include hydrophobic changes (e.g. phenylalanine/tryptophan), hydrophilic protrusions (e.g. arginine), large steric protrusions (*N*-linked glycan), and side chain removal (e.g. poly-glycine). As hydrophobic interactions are often a major contributor to binding affinity, hydrophobic mutations at different surface locations can identify variants with enhanced potency.[103,104] Similarly, hydrophilic mutations and *N*-linked glycan can identify variants with enhanced solubility.[105,106] Moreover, *N*-linked glycan can modulate autoreactivity and also inform antibody proximity to antigen.[107] Additionally, mutation of stretches of surface residues to glycine can remove clashes with the antigen, thereby improving potency.[108] In the specific case where the antibody epitope is proximal to or partially comprised of membrane, hydrophobic and/or arginine mutations may enhance potency through interaction with lipid membrane. Mutants are evaluated with various functional assays, including binding, neutralization, solubility, and autoreactivity. Mutations that yield improved properties can be further interrogated, including the assessment of mutations with all 20 amino acid types, or at proximal residue positions. Mutations can be combined, and the surface-matrix screening approach can be applied iteratively to the leading antibody candidate. Of note, individual mutations may provide optimal assessment of structure/activity relationships.

We are currently in the process of applying Antibodyomics9 to antibody 10E8, which targets the membrane-proximal external region (MPER) and is capable of neutralizing approximately 98% of HIV-1 isolates.[109] By using a surface-matrix scan of phenylalanine/tryptophan, arginine, *N*-linked glycan, and poly-glycine mutations, we were able to identify mutations that improved the neutralization potency of 10E8. Incorporation of these mutations to 10E8v4, a soluble version of 10E8,[110] yielded variants with acceptable solubility,

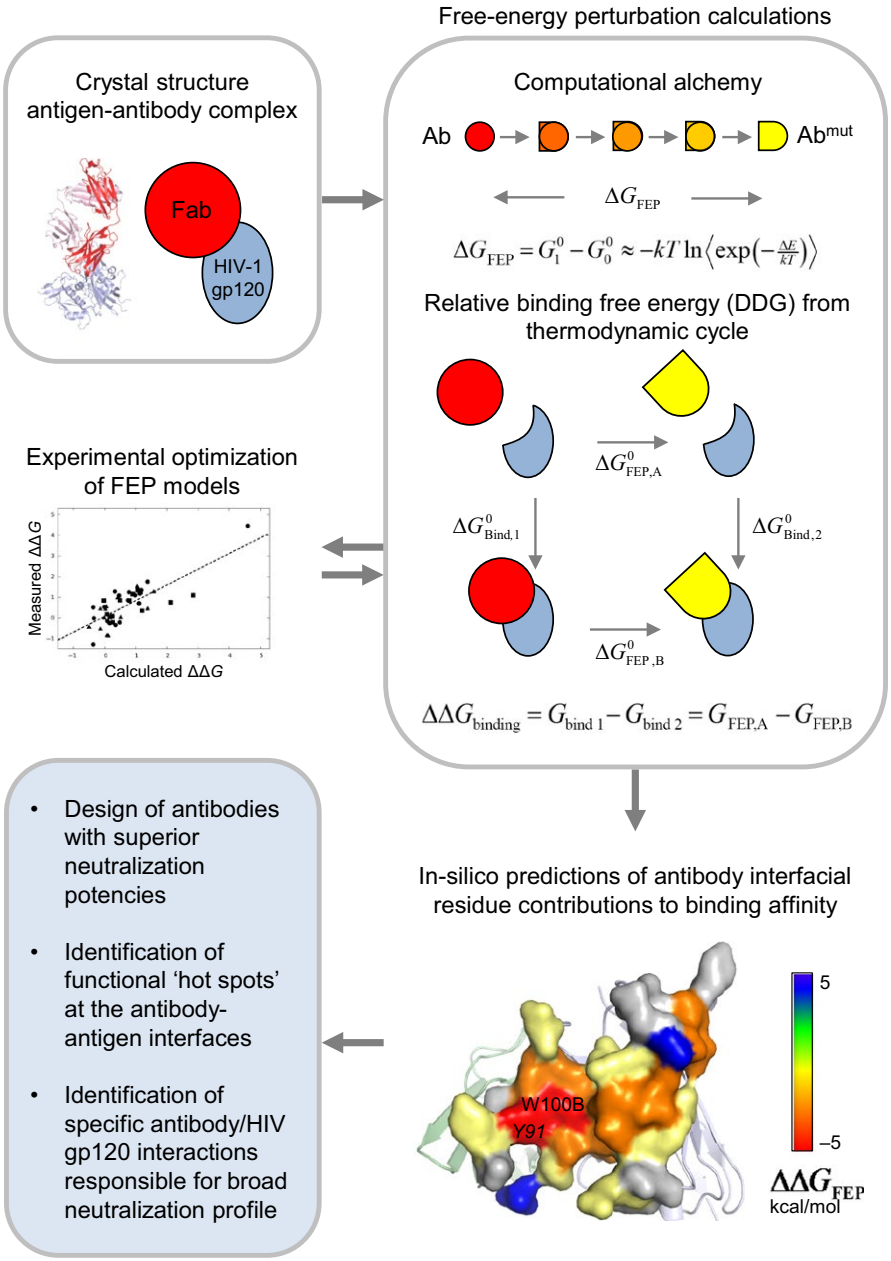

FIGURE 9 Antibodyomics8: Free energy perturbation analysis of antibody-antigen interactions. With crystallographic structure of the antibody-antigen complex as a starting point, we use thermodynamic cycle to compute the relative free energies of binding (ΔΔ*G*) between antibody (red) and a mutant (yellow). The vertical legs of the cycle correspond to the physical binding process, and horizontal legs to 'alchemical' transformation of antibody into mutant in solvent (top) and in complex with HIV-1 gp120 glycoprotein (bottom). FEP is used to compute the free energy associated with these alchemical transformations, and ΔΔ*G* is calculated as the difference between the free energy required to transform the antibodies in complex with gp120 and in solution. By iterating over all antibody-gp120 interfacial residues, a heatmap of residue contributions to binding affinity can be constructed

low poly-reactivity, and improved IC_{80} potencies of >10-fold better than the parent 10E8. We plan to apply Antibodyomics9 to improve other naturally broadly neutralizing HIV-1 antibodies, including those that target the CD4-binding site and the V1V2 apex of HIV.[39,82] The improvement of manufacturing characteristics and therapeutic properties for promising anti-HIV-1 antibodies may accelerate their clinical development and lead to their use in the prevention and treatment of HIV-1.

11 | ANTIBODYOMICS10

As random searching for functional antibodies to counter the near-infinite diversity of potential immunogens may be problematic, well-configured antibody repertoires may be key to protective humoral immunity.[111] The human immune system, for example, typically responds to a novel antigen by generating antigen-specific antibodies within just a few weeks.[112]

Unfortunately, our understanding of the principles underlying well-configured antibody repertoires has been hindered by two technical barriers: first, the antibody repertoire (approximately 10^{11} in humans and approximately 10^9 in mice) is sufficiently large as to impede characterization of all component antibodies; and second, experimental systems to test antibody repertoires may encounter difficulties in synthesizing large numbers of component antibodies. Continued technical advances in NGS and in antibody library synthesis have, however, begun to make headway in addressing these technical barriers [reviewed in (53, 113)], and a number of fundamental

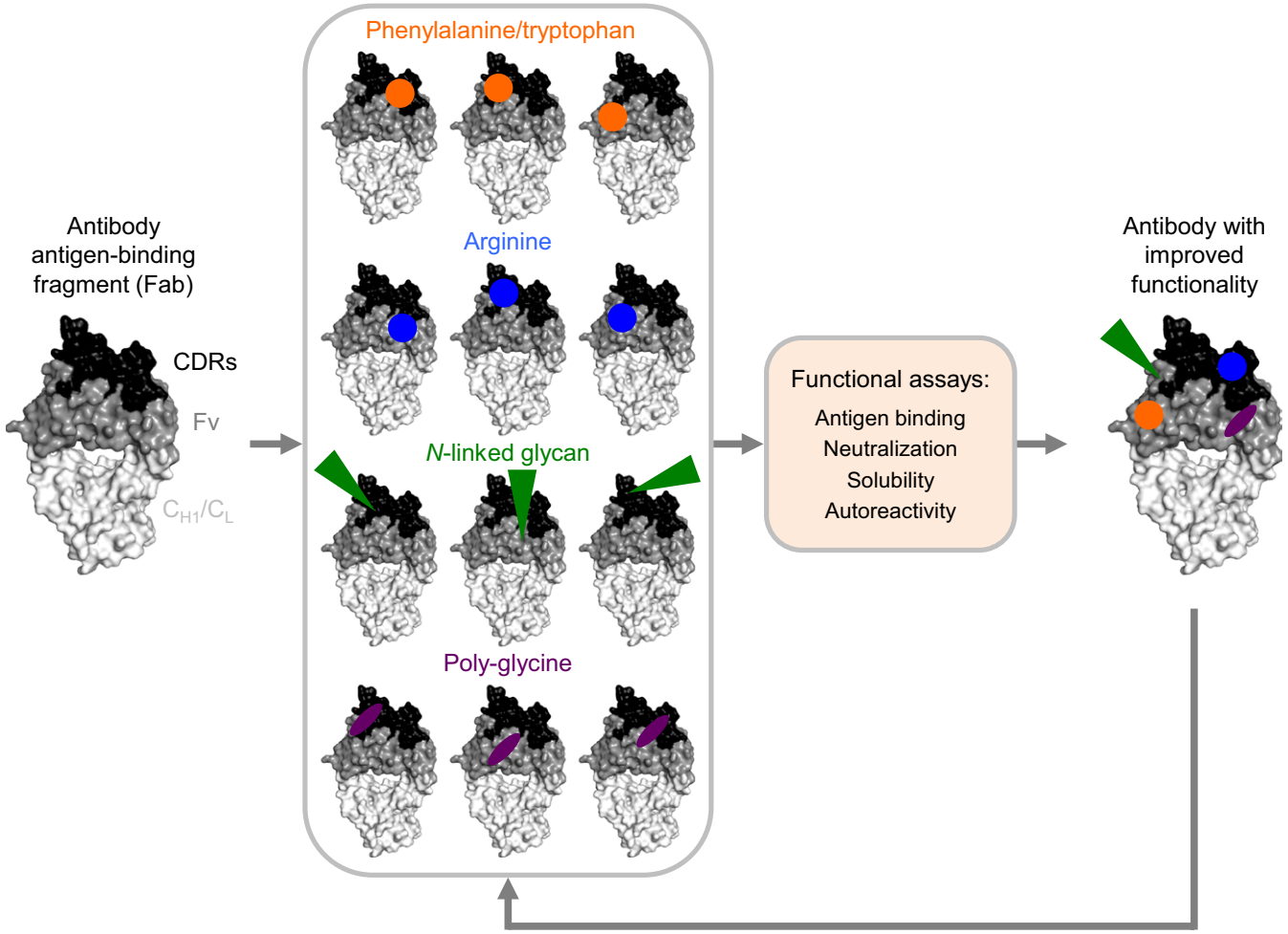

FIGURE 10 Antibodyomics9: Antibody improvement by surface-matrix screening. To optimize an antibody, we introduce individual mutations on Fv-surface residues to create a matrix of information of surface-chemical modifications. Such mutations include phenylalanine/tryptophan, arginine, *N*-linked glycan, and poly-glycine. All mutants are then evaluated for the desired properties (e.g. antigen binding, neutralization, solubility, and autoreactivity). Mutations with improved properties can then be combined and the surface-matrix screening can be repeated

insights into antibody-antigen affinity and specificity are now apparent: (i) functional antibodies recognize protein-antigen surfaces mainly through aromatic side chains in their complementarity-determining regions (CDRs); (ii) the major portion of the antibody-protein binding energy is contributed by interactions of aromatic side chains from antibody with backbone atoms and side chain-carbon atoms on protein antigens, which comprise three-quarters of protein surfaces on average[114]; (iii) direct hydrogen bonds across the antibody-antigen interface involve mostly polar backbone groups and small hydrophilic residues on the antibody.[114] One implication of these insights is that a synthetic antibody library bearing paratopes with diverse structural contours—enriched with aromatic residues among short chain hydrophilic residues—should be capable of recognizing the surfaces of most protein antigens through binding to common physicochemical features: carbon atoms and polar backbone groups.[114]

We have analyzed the antigen recognition capabilities of mouse antibody repertoires by examining the functionality of corresponding synthetic phage-displayed antibody libraries—including libraries specifically enriched with aromatic residues among short chain hydrophilic residues (Fig. 11).[114–117] NGS was employed to sample the antibody repertoire, and we used the extracellular domain (ECD) of human epidermal growth factor receptor 2 (HER2) as a model antigen.[111] Computational analysis of the NGS data indicated that the antibody repertoires from naïve mice and from vaccinated mice immunized with various immunization protocols were largely skewed toward a CDR-canonical structure. A synthetic antibody library was constructed based on the structural and sequence characteristics of the mouse antibody repertoires and tested for HER2-specific recognition. The synthetic antibody library was indeed able to generate antibodies with equal affinity and specificity as the affinity-matured antibodies from the immunized mice. Moreover, the synthetic antibody library was also able to generate antibodies that recognized diverse protein surfaces in addition to the extracellular domain of HER2. These results suggest that the predominant CDR-canonical structure in mouse antibody repertoires is a robust framework capable of encoding highly functional antibodies against diverse protein antigens. The results confirm the utility of phage-displayed synthetic

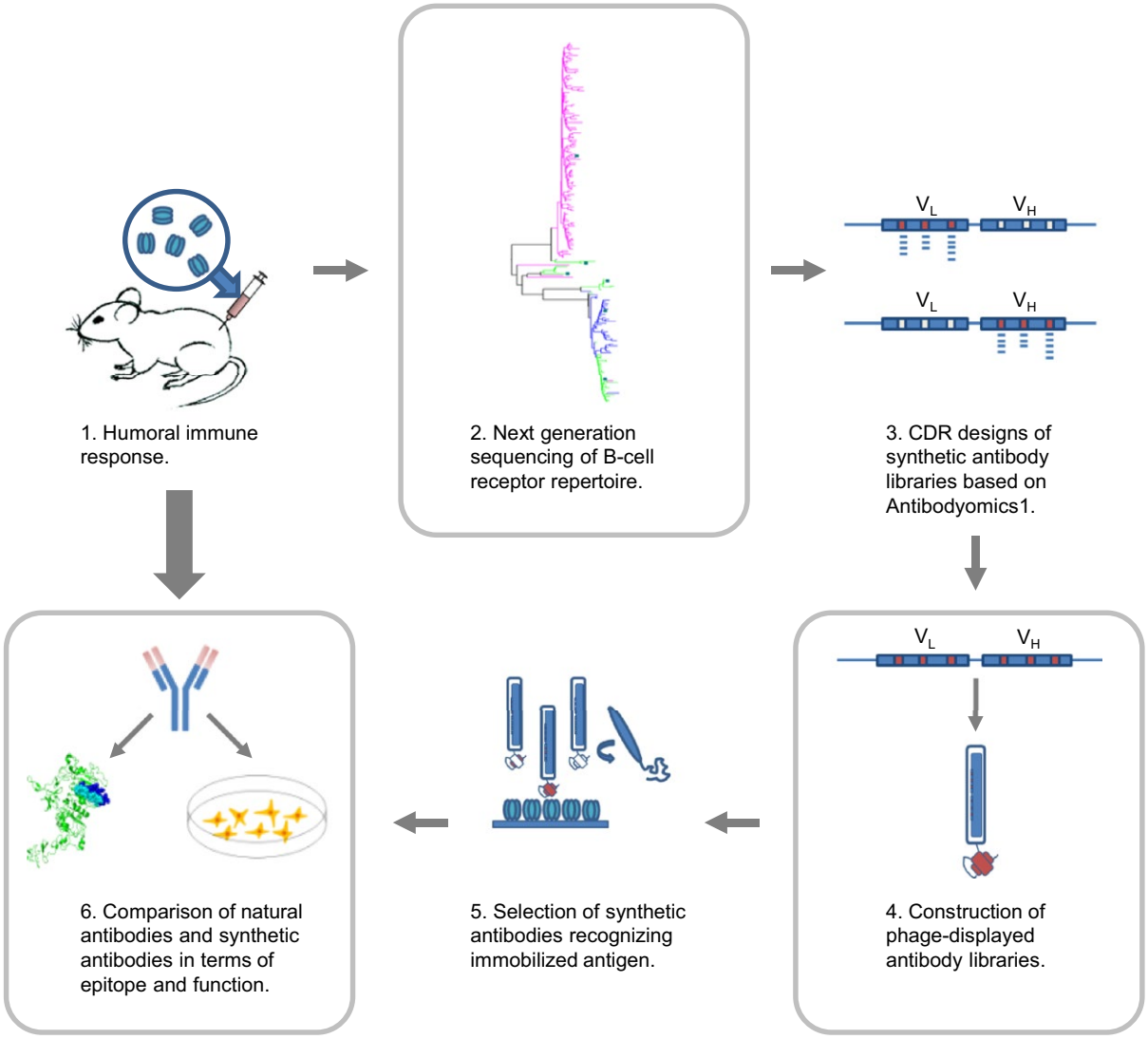

FIGURE 11 Antibodyomics10: Synthetic antibody libraries in vaccine development. Comparison of synthetic antibody libraries, such as those derived from phage display, with natural antibody repertoires, allows for insight into antibody features required for recognition without confounding factors such as autoreactivity, T-cell help, and restrictions arising from surface display of B-cell receptors

antibody libraries in testing hypotheses concerning natural antibody repertoires and in developing artificial antibody repertoires to discover functional antibodies targeting diverse epitopes.[111,114–117]

In addition to providing insight into natural antibody repertoires, we have explored synthetic antibody libraries skewed to include unusual features observed in broadly neutralizing antibodies that target HIV-1, such as long protruding loops and tyrosine sulfation.[11,102,118,119] We observed antibodies from libraries enriched in long anionic loops, similar to those observed in antibodies PG9, CH01, and CAP256-VRC26,[6,11,28,37,102] to have poor solubility. Moreover, we are beginning to understand the features that allow for recognition of highly glycosylated antigens. Synthetic libraries encoding unusual features may thus provide tools to validate or to challenge sequence signatures thought to be important in recognizing the heavily glycosylated HIV-1 viral spike.

12 | CONCLUSION

The resolution-enhancing multi-dimensional technologies described here yield information-rich datasets that can synergize with one another to provide in-depth understanding of the development of broadly neutralizing immunity to HIV-1, both in the context of natural infection (Fig. 1B, upper panels) and in active or passive immunization (Fig. 1B, lower panels). In particular, Fig. 1B delineates the initiation of B-cell lineages that develop through cycles of affinity maturation into populations of neutralizing antibodies, shows the use of select antibodies for passive delivery, and depicts the incorporation of lineage information into the design of ontogeny-based immunogens. Overlaid are the resolution-enhancing Antibodyomics technologies, highlighting appropriate points of application and the immunological processes that they inform.

TABLE 1 Antibodyomics embodiments, software, and applications

			References	
Antibodyomics	**Application**	**Embodiment**	**Software**	**Applications**
1	Antibody identification and lineage analysis	Perl-scripted pipeline	1.0: (13) 1.1: (16)	(9–16, 39, 108, 110, 120)
2	Longitudinal lineage development	Integrated suite including modules in Python, Perl, and R	2.0: (6) 2.1: (17)	(5, 6, 11, 17–20, 23, 24, 82)
3	Delineation of neutralization into structurally defined epitopes	Mathematica scripts	3.0: (49)	(5, 6, 16, 27, 30, 32, 39, 49–51, 110, 121–123)
4	Heavy/light chain paired NGS for antibody identification and lineage analysis	Bash-scripted pipeline and Python-scripted pipeline	Bash 2.0: (52) Python 2.0: (55)	(6, 11, 56, 57, 61, 63)
5	Antibody class quantification and class-guided design	Ad hoc frequentist, Bayesian, and other statistical analyses	None	(11, 12, 16, 124)
6	Likelihood of gene SHM	Perl- and Python-scripted pipeline	6.0 (Z. Sheng, unpublished data)	(5, 61, Z. Sheng, unpublished data)
7	Antibody and antigen dynamics	Tcl/Tk analysis scripts	7.0: (87)	(11, 30, 87)
8	Prediction of antibody binding energies	C++ utilizing CUDA with FEP module in C++ and python	Desmond molecular dynamics program (125)	(97, 126–128)
9	Antibody improvement by surface-matrix screening	Perl-scripted pipeline	None	(107, 109)
10	Comparison of synthetic and natural antibody libraries for vaccine development	Manual comparison	None	(111)

In addition to ontogeny-based strategies of vaccine design, the Antibodyomics technologies described here can be applied to other vaccine strategies. For example, Antibodyomics1, 2, and 4 can be applied to any vaccine strategy for which lineage information is crucial. Similarly, Antibodyomics5 and 6 generate statistical information on elicited lineages and their SHM requirements, and these technologies should inform lineage-based strategies. Meanwhile, Antibodyomics3 provides information on the specificity of serum, with general applicability to many vaccine-related problems. Other Antibodyomics technologies may have more specific applications, such as Antibodyomics8 and 9, which involve antibody recognition and improvement, and Antibodyomis10, which is designed to provide specific insight into confounding factors such as autoreactivity, T-cell help, and commonality of protruding loops. Table 1 summarizes the Antibodyomics technologies and their software embodiments.

We note that applications of Antibodyomics technologies are not unique to the investigations of immunity to HIV-1. Indeed, we recently applied Antibodyomics1 to understand the induction of antibodies capable of neutralizing diverse group 1 and group 2 subtypes of influenza A virus, and also used Antibodyomics3 to classify reconstituted neutralizing functionalities.[16]

More generally, the purpose of the Antibodyomics technologies is to afford a data-rich view of the genetic and immunological processes controlling B cell-lineage development. The information-rich output may also provide insights into the evolution and specific immunological events allowing for lineage development. When combined with atomic-level structural information on antibodies, antigens, and their complexes, a similarly data-rich view of the chemistry of antibody/antigen interaction can be obtained. Overall, the Antibodyomics technologies described here comprise an information-rich discovery platform, which has been developed specifically for HIV-1, but should be broadly applicable to other systems that require an understanding of antibody lineages.

ACKNOWLEDGEMENTS

We thank G. Georgiou and A. Ellington for developing the paired sequencing methodology used in Antibodyomics4 and R. Friesner for developing the FEP methodology used in Antibodyomics8. We thank T. Bylund, S. Patel, C.-H. Shen, and Z. Zhang for crucial input to development and applications, B. Honig for bioinformatics guidance, J. Stuckey for assistance with figures, and members of the Structural Biology Section and Structural Bioinformatics Core Section of the Vaccine Research Center for helpful comments. Support for this work was provided by a Swiss National Foundation of Science Fellowship (164691) to T. L., by NIH grant P01-AI104722 to L. S., by Academia Sinica grant MOST104-0210-01-09-02 and Taiwan Protein Project grant MOST105-0210-01-12-01 to A. S. Y., and by the Intramural Research Program of the Vaccine Research Center, National Institute of Allergy and Infectious Diseases, National Institutes of Health (NIH).

COMPETING FINANCIAL INTERESTS

BJD declares a competing financial interest in the form of a patent held by the University of Texas at Austin.

AUTHOR CONTRIBUTIONS

The Antibodyomics program was initiated by P. D. K., J. R. M., and L. S. with sections on Antibodyomics1 compiled by C. S., Antibodyomics2 by C. A. S. and Z. S., Antibodyomics3 by I. S. G., Antibodyomics4 by B. J. D., Antibodyomics5 by G. Y. C., P. D. K., and L. S., Antibodyomics6 by Z. S. and C. A. S., Antibodyomics7 by T. L. and C. S., Antibodyomics8 by T. G., Antibodyomics9 by G. Y. C., and Antibodyomics10 by A.-S. Y.

REFERENCES

1. Burton DR, Mascola JR. Antibody responses to envelope glycoproteins in HIV-1 infection. *Nat Immunol*. 2015;16:571–576.
2. Burton DR, Poignard P, Stanfield RL, Wilson IA. Broadly neutralizing antibodies present new prospects to counter highly antigenically diverse viruses. *Science*. 2012;337:183–186.
3. Hraber P, Seaman MS, Bailer RT, Mascola JR, Montefiori DC, Korber BT. Prevalence of broadly neutralizing antibody responses during chronic HIV-1 infection. *AIDS*. 2014;28:163–169.
4. Nabel GJ. Philosophy of science. The coordinates of truth. *Science*. 2009;326:53–54.
5. Bonsignori M, Zhou T, Sheng Z, et al. Maturation pathway from germline to broad HIV-1 neutralizer of a CD4-mimic antibody. *Cell*. 2016;165:449–463.
6. Doria-Rose NA, Schramm CA, Gorman J, et al. Developmental pathway for potent V1V2-directed HIV-neutralizing antibodies. *Nature*. 2014;509:55–62.
7. Liao HX, Lynch R, Zhou T, et al. Co-evolution of a broadly neutralizing HIV-1 antibody and founder virus. *Nature*. 2013;496:469–476.
8. MacLeod DT, Choi NM, Briney B, et al. Early antibody lineage diversification and independent limb maturation lead to broad HIV-1 neutralization targeting the Env high-mannose patch. *Immunity*. 2016;44:1215–1226.
9. Wu X, Zhou T, Zhu J, et al. Focused evolution of HIV-1 neutralizing antibodies revealed by structures and deep sequencing. *Science*. 2011;333:1593–1602.
10. Zhu J, O'Dell S, Ofek G, et al. Somatic populations of PGT135-137 HIV-1-neutralizing antibodies identified by 454 pyrosequencing and bioinformatics. *Front Microbiol*. 2012;3:315.
11. Gorman J, Soto C, Yang MM, et al. Structures of HIV-1 Env V1V2 with broadly neutralizing antibodies reveal commonalities that enable vaccine design. *Nat Struct Mol Biol*. 2016;23:81–90.
12. Zhou T, Zhu J, Wu X, et al. Multidonor analysis reveals structural elements, genetic determinants, and maturation pathway for HIV-1 neutralization by VRC01-class antibodies. *Immunity*. 2013;39:245–258.
13. Zhu J, Wu X, Zhang B, et al. De novo identification of VRC01 class HIV-1-neutralizing antibodies by next-generation sequencing of B-cell transcripts. *Proc Natl Acad Sci USA*. 2013;110:E4088–E4097.
14. Zhu J, Ofek G, Yang Y, et al. Mining the antibodyome for HIV-1-neutralizing antibodies with next-generation sequencing and phylogenetic pairing of heavy/light chains. *Proc Natl Acad Sci USA*. 2013;110:6470–6475.
15. Soto C, Ofek G, Joyce MG, et al. Developmental pathway of the MPER-directed HIV-1-neutralizing antibody 10E8. *PLoS ONE*. 2016;11:e0157409.
16. Joyce MG, Wheatley AK, Thomas PV, et al. Vaccine-induced antibodies that neutralize group 1 and group 2 influenza A viruses. *Cell*. 2016;166:609–623.
17. Schramm CA, Sheng Z, Zhang Z, Mascola JR, Kwong PD, Shapiro L. SONAR: a high-throughput pipeline for inferring antibody ontogenies from longitudinal sequencing of B cell transcripts. *Front. Immunol*. 2016;7:372. doi: 10.3389/fimmu.2016.00372.
18. Wu X, Zhang Z, Schramm CA, et al. Maturation and diversity of the VRC01-antibody lineage over 15 years of chronic HIV-1 infection. *Cell*. 2015;161:470–485.
19. Bhiman JN, Anthony C, Doria-Rose NA, et al. Viral variants that initiate and drive maturation of V1V2-directed HIV-1 broadly neutralizing antibodies. *Nat Med*. 2015;21:1332–1336.
20. Sheng Z, Schramm CA, Connors M, et al. Effects of Darwinian selection and mutability on rate of broadly neutralizing antibody evolution during HIV-1 infection. *PLoS Comput Biol*. 2016;12:e1004940.
21. Kepler TB. Reconstructing a B-cell clonal lineage. I. Statistical inference of unobserved ancestors. *F1000Res*. 2013;2:103.
22. Gao F, Bonsignori M, Liao HX, et al. Cooperation of B cell lineages in induction of HIV-1-broadly neutralizing antibodies. *Cell*. 2014;158:481–491.
23. Francica JR, Sheng Z, Zhang Z, et al. Analysis of immunoglobulin transcripts and hypermutation following SHIV(AD8) infection and protein-plus-adjuvant immunization. *Nat Commun*. 2015;6:6565.
24. Huang J, Kang BH, Ishida E, et al. A CD4-binding site antibody to HIV that evolves extraordinary neutralization breadth. *Immunity*. In Press.
25. Derdeyn CA, Moore PL, Morris L. Development of broadly neutralizing antibodies from autologous neutralizing antibody responses in HIV infection. *Curr Opin HIV AIDS*. 2014;9:210–216.
26. Scheid JF, Mouquet H, Feldhahn N, et al. Broad diversity of neutralizing antibodies isolated from memory B cells in HIV-infected individuals. *Nature*. 2009;458:636–640.
27. Huang J, Kang BH, Pancera M, et al. Broad and potent HIV-1 neutralization by a human antibody that binds the gp41-gp120 interface. *Nature*. 2014;515:138–142.
28. Bonsignori M, Montefiori DC, Wu X, et al. Two distinct broadly neutralizing antibody specificities of different clonal lineages in a single HIV-1-infected donor: implications for vaccine design. *J Virol*. 2012;86:4688–4692.
29. Landais E, Huang X, Havenar-Daughton C, et al. Broadly neutralizing antibody responses in a large longitudinal sub-Saharan HIV Primary Infection Cohort. *PLoS Pathog*. 2016;12:e1005369.
30. Kong R, Xu K, Zhou T, et al. Fusion peptide of HIV-1 as a site of vulnerability to neutralizing antibody. *Science*. 2016;352:828–833.
31. Scharf L, Scheid JF, Lee JH, et al. Antibody 8ANC195 reveals a site of broad vulnerability on the HIV-1 envelope spike. *Cell Rep*. 2014;7:785–795.
32. Pancera M, Zhou T, Druz A, et al. Structure and immune recognition of trimeric pre-fusion HIV-1 Env. *Nature*. 2014;514:455–461.
33. Falkowska E, Le KM, Ramos A, et al. Broadly neutralizing HIV antibodies define a glycan-dependent epitope on the prefusion conformation of gp41 on cleaved envelope trimers. *Immunity*. 2014;40:657–668.
34. Walker LM, Huber M, Doores KJ, et al. Broad neutralization coverage of HIV by multiple highly potent antibodies. *Nature*. 2011;477:466–470.
35. Scheid JF, Mouquet H, Ueberheide B, et al. Sequence and structural convergence of broad and potent HIV antibodies that mimic CD4 binding. *Science*. 2011;333:1633–1637.
36. Corti D, Langedijk JP, Hinz A, et al. Analysis of memory B cell responses and isolation of novel monoclonal antibodies with neutralizing breadth from HIV-1-infected individuals. *PLoS ONE*. 2010;5:e8805.

37. Walker LM, Phogat SK, Chan-Hui PY, et al. Broad and potent neutralizing antibodies from an African donor reveal a new HIV-1 vaccine target. *Science*. 2009;326:285–289.
38. Huang J, Ofek G, Laub L, et al. Broad and potent neutralization of HIV-1 by a gp41-specific human antibody. *Nature*. 2012;491:406–412.
39. Zhou T, Lynch RM, Chen L, et al. Structural repertoire of HIV-1-neutralizing antibodies targeting the CD4 supersite in 14 donors. *Cell*. 2015;161:1280–1292.
40. Wu X, Yang ZY, Li Y, et al. Rational design of envelope identifies broadly neutralizing human monoclonal antibodies to HIV-1. *Science*. 2010;329:856–861.
41. Gray ES, Taylor N, Wycuff D, et al. Antibody specificities associated with neutralization breadth in plasma from human immunodeficiency virus type 1 subtype C-infected blood donors. *J Virol*. 2009;83:8925–8937.
42. Walker LM, Simek MD, Priddy F, et al. A limited number of antibody specificities mediate broad and potent serum neutralization in selected HIV-1 infected individuals. *PLoS Pathog*. 2010;6:e1001028.
43. Li Y, Svehla K, Louder MK, et al. Analysis of neutralization specificities in polyclonal sera derived from human immunodeficiency virus type 1-infected individuals. *J Virol*. 2009;83:1045–1059.
44. Stamatatos L, Morris L, Burton DR, Mascola JR. Neutralizing antibodies generated during natural HIV-1 infection: good news for an HIV-1 vaccine? *Nat Med*. 2009;15:866–870.
45. Binley JM, Lybarger EA, Crooks ET, et al. Profiling the specificity of neutralizing antibodies in a large panel of plasmas from patients chronically infected with human immunodeficiency virus type 1 subtypes B and C. *J Virol*. 2008;82:11651–11668.
46. Dhillon AK, Donners H, Pantophlet R, et al. Dissecting the neutralizing antibody specificities of broadly neutralizing sera from human immunodeficiency virus type 1-infected donors. *J Virol*. 2007;81:6548–6562.
47. Doria-Rose NA, Klein RM, Manion MM, et al. Frequency and phenotype of human immunodeficiency virus envelope-specific B cells from patients with broadly cross-neutralizing antibodies. *J Virol*. 2009;83:188–199.
48. Li Y, Migueles SA, Welcher B, et al. Broad HIV-1 neutralization mediated by CD4-binding site antibodies. *Nat Med*. 2007;13:1032–1034.
49. Georgiev IS, Doria-Rose NA, Zhou T, et al. Delineating antibody recognition in polyclonal sera from patterns of HIV-1 isolate neutralization. *Science*. 2013;340:751–756.
50. Crooks ET, Tong T, Chakrabarti B, et al. Vaccine-elicited tier 2 HIV-1 neutralizing antibodies bind to quaternary epitopes involving glycan-deficient patches proximal to the CD4 binding site. *PLoS Pathog*. 2015;11:e1004932.
51. Cortez V, Wang B, Dingens A, et al. The broad neutralizing antibody responses after HIV-1 superinfection are not dominated by antibodies directed to epitopes common in single infection. *PLoS Pathog*. 2015;11:e1004973.
52. DeKosky BJ, Kojima T, Rodin A, et al. In-depth determination and analysis of the human paired heavy- and light-chain antibody repertoire. *Nat Med*. 2015;21:86–91.
53. Georgiou G, Ippolito GC, Beausang J, Busse CE, Wardemann H, Quake SR. The promise and challenge of high-throughput sequencing of the antibody repertoire. *Nat Biotechnol*. 2014;32:158–168.
54. DeKosky BJ, Ippolito GC, Deschner RP, et al. High-throughput sequencing of the paired human immunoglobulin heavy and light chain repertoire. *Nat Biotechnol*. 2013;31:166–169.
55. McDaniel JR, DeKosky BJ, Tanno H, Ellington AD, Georgiou G. Ultra-high-throughput sequencing of the immune receptor repertoire from millions of lymphocytes. *Nat Protoc*. 2016;11:429–442.
56. Lavinder JJ, Wine Y, Giesecke C, et al. Identification and characterization of the constituent human serum antibodies elicited by vaccination. *Proc Natl Acad Sci USA*. 2014;111:2259–2264.
57. Wang B, Kluwe CA, Lungu OI, et al. Facile discovery of a diverse panel of anti-Ebola virus antibodies by immune repertoire mining. *Sci Rep*. 2015;5:13926.
58. Dosenovic P, von Boehmer L, Escolano A, et al. Immunization for HIV-1 broadly neutralizing antibodies in human Ig knockin mice. *Cell*. 2015;161:1505–1515.
59. Jardine J, Julien JP, Menis S, et al. Rational HIV immunogen design to target specific germline B cell receptors. *Science*. 2013;340:711–716.
60. Jardine JG, Ota T, Sok D, et al. Priming a broadly neutralizing antibody response to HIV-1 using a germline-targeting immunogen. *Science*. 2015;349:156–161.
61. Tian M, Cheng C, Chen X, et al. Induction of HIV-1 neutralizing antibodies in mice with diverse precursor repertoires. *Cell*. 2016;166:1–14.
62. Kwong PD, Mascola JR. Human antibodies that neutralize HIV-1: identification, structures, and B cell ontogenies. *Immunity*. 2012;37:412–425.
63. DeKosky BJ, Lungu OI, Park D, et al. Large-scale sequence and structural comparisons of human naive and antigen-experienced antibody repertoires. *Proc Natl Acad Sci USA*. 2016;113:E2636–E2645.
64. Ekiert DC, Bhabha G, Elsliger MA, et al. Antibody recognition of a highly conserved influenza virus epitope. *Science*. 2009;324:246–251.
65. Sui J, Hwang WC, Perez S, et al. Structural and functional bases for broad-spectrum neutralization of avian and human influenza A viruses. *Nat Struct Mol Biol*. 2009;16:265–273.
66. Parameswaran P, Liu Y, Roskin KM, et al. Convergent antibody signatures in human dengue. *Cell Host Microbe*. 2013;13:691–700.
67. Jardine JG, Kulp DW, Havenar-Daughton C, et al. HIV-1 broadly neutralizing antibody precursor B cells revealed by germline-targeting immunogen. *Science*. 2016;351:1458–1463.
68. Jiang N, He J, Weinstein JA, et al. Lineage structure of the human antibody repertoire in response to influenza vaccination. *Sci Transl Med*. 2013;5:171ra119.
69. Haynes BF, Kelsoe G, Harrison SC, Kepler TB. B-cell-lineage immunogen design in vaccine development with HIV-1 as a case study. *Nat Biotechnol*. 2012;30:423–433.
70. Muramatsu M, Kinoshita K, Fagarasan S, Yamada S, Shinkai Y, Honjo T. Class switch recombination and hypermutation require activation-induced cytidine deaminase (AID), a potential RNA editing enzyme. *Cell*. 2000;102:553–563.
71. Wood GS. The immunohistology of lymph nodes in HIV infection: a review. *Prog AIDS Pathol*. 1990;2:25–32.
72. McKean D, Huppi K, Bell M, Staudt L, Gerhard W, Weigert M. Generation of antibody diversity in the immune response of BALB/c mice to influenza virus hemagglutinin. *Proc Natl Acad Sci USA*. 1984;81:3180–3184.
73. Cowell LG, Kepler TB. The nucleotide-replacement spectrum under somatic hypermutation exhibits microsequence dependence that is strand-symmetric and distinct from that under germline mutation. *J Immunol*. 2000;164:1971–1976.
74. Betz AG, Rada C, Pannell R, Milstein C, Neuberger MS. Passenger transgenes reveal intrinsic specificity of the antibody hypermutation mechanism: clustering, polarity, and specific hot spots. *Proc Natl Acad Sci USA*. 1993;90:2385–2388.
75. Goyenechea B, Milstein C. Modifying the sequence of an immunoglobulin V-gene alters the resulting pattern of hypermutation. *Proc Natl Acad Sci USA*. 1996;93:13979–13984.
76. Wei L, Chahwan R, Wang S, et al. Overlapping hotspots in CDRs are critical sites for V region diversification. *Proc Natl Acad Sci USA*. 2015;112:E728–E737.
77. Zheng NY, Wilson K, Jared M, Wilson PC. Intricate targeting of immunoglobulin somatic hypermutation maximizes the efficiency of affinity maturation. *J Exp Med*. 2005;201:1467–1478.

78. Mirsky A, Kazandjian L, Anisimova M. Antibody-specific model of amino acid substitution for immunological inferences from alignments of antibody sequences. *Mol Biol Evol*. 2015;32:806–819.
79. Klein F, Diskin R, Scheid JF, et al. Somatic mutations of the immunoglobulin framework are generally required for broad and potent HIV-1 neutralization. *Cell*. 2013;153:126–138.
80. Trabuco LG, Villa E, Schreiner E, Harrison CB, Schulten K. Molecular dynamics flexible fitting: a practical guide to combine cryo-electron microscopy and X-ray crystallography. *Methods*. 2009;49:174–180.
81. Zhao G, Perilla JR, Yufenyuy EL, et al. Mature HIV-1 capsid structure by cryo-electron microscopy and all-atom molecular dynamics. *Nature*. 2013;497:643–646.
82. Doria-Rose NA, Bhiman JN, Roark RS, et al. New member of the V1V2-directed CAP256-VRC26 lineage that shows increased breadth and exceptional potency. *J Virol*. 2016;90:76–91.
83. Adcock SA, McCammon JA. Molecular dynamics: survey of methods for simulating the activity of proteins. *Chem Rev*. 2006;106:1589–1615.
84. Karplus M, Kuriyan J. Molecular dynamics and protein function. *Proc Natl Acad Sci USA*. 2005;102:6679–6685.
85. Karplus M, McCammon JA. Molecular dynamics simulations of biomolecules. *Nat Struct Biol*. 2002;9:646–652.
86. Julien JP, Cupo A, Sok D, et al. Crystal structure of a soluble cleaved HIV-1 envelope trimer. *Science*. 2013;342:1477–1483.
87. Stewart-Jones GB, Soto C, Lemmin T, et al. Trimeric HIV-1-Env structures define glycan shields from clades A, B, and G. *Cell*. 2016;165:813–826.
88. Pritchard LK, Spencer DI, Royle L, et al. Glycan clustering stabilizes the mannose patch of HIV-1 and preserves vulnerability to broadly neutralizing antibodies. *Nat Commun*. 2015;6:7479.
89. Doores KJ, Bonomelli C, Harvey DJ, et al. Envelope glycans of immunodeficiency virions are almost entirely oligomannose antigens. *Proc Natl Acad Sci USA*. 2010;107:13800–13805.
90. Behrens AJ, Vasiljevic S, Pritchard LK, et al. Composition and antigenic effects of individual glycan sites of a trimeric HIV-1 envelope glycoprotein. *Cell Rep*. 2016;14:2695–2706.
91. Lasky LA, Groopman JE, Fennie CW, et al. Neutralization of the AIDS retrovirus by antibodies to a recombinant envelope glycoprotein. *Science*. 1986;233:209–212.
92. Mahajan PB. Recent advances in application of pharmacogenomics for biotherapeutics. *AAPS J*. 2016;18:605–611.
93. De Vivo M, Masetti M, Bottegoni G, Cavalli A. Role of molecular dynamics and related methods in drug discovery. *J Med Chem*. 2016;59:4035–4061.
94. Swails JM, York DM, Roitberg AE. Constant pH replica exchange molecular dynamics in explicit solvent using discrete protonation states: implementation, testing, and validation. *J Chem Theory Comput*. 2014;10:1341–1352.
95. Abrams C, Bussi G. Enhanced sampling in molecular dynamics using metadynamics, replica-exchange, and temperature-acceleration. *Entropy-Switz*. 2014;16:163–199.
96. Rocchia W, Masetti M, Cavalli A. Enhanced sampling methods in drug design. *RSC Drug Discov*. 2012;23:273–301.
97. Wang L, Berne BJ, Friesner RA. On achieving high accuracy and reliability in the calculation of relative protein-ligand binding affinities. *Proc Natl Acad Sci USA*. 2012;109:1937–1942.
98. Huang X, Hagen M, Kim B, Friesner RA, Zhou R, Berne BJ. Replica exchange with solute tempering: efficiency in large scale systems. *J Phys Chem B*. 2007;111:5405–5410.
99. Wang L, Friesner RA, Berne BJ. Replica exchange with solute scaling: a more efficient version of replica exchange with solute tempering (REST2). *J Phys Chem B*. 2011;115:9431–9438.
100. Aubert B, Barate R, Boutigny D, et al. Measurement of time-dependent CP asymmetries in B0–>D(*)+/-D+/- decays. *Phys Rev Lett*. 2005;95:131802.
101. Clackson T, Wells JA. A hotspot of binding energy in a hormone-receptor interface. *Science*. 1995;267:383–386.
102. McLellan JS, Pancera M, Carrico C, et al. Structure of HIV-1 gp120 V1/V2 domain with broadly neutralizing antibody PG9. *Nature*. 2011;480:336–343.
103. Diskin R, Scheid JF, Marcovecchio PM, et al. Increasing the potency and breadth of an HIV antibody by using structure-based rational design. *Science*. 2011;334:1289–1293.
104. Young L, Jernigan RL, Covell DG. A role for surface hydrophobicity in protein-protein recognition. *Protein Sci*. 1994;3:717–729.
105. Hebert DN, Lamriben L, Powers ET, Kelly JW. The intrinsic and extrinsic effects of N-linked glycans on glycoproteostasis. *Nat Chem Biol*. 2014;10:902–910.
106. Trevino SR, Scholtz JM, Pace CN. Amino acid contribution to protein solubility: Asp, Glu, and Ser contribute more favorably than the other hydrophilic amino acids in RNase Sa. *J Mol Biol*. 2007;366:449–460.
107. Chuang GY, Zhang B, McKee K, et al. Eliminating antibody polyreactivity through addition of N-linked glycosylation. *Protein Sci*. 2015;24:1019–1030.
108. Rudicell RS, Kwon YD, Ko SY, et al. Enhanced potency of a broadly neutralizing HIV-1 antibody in vitro improves protection against lentiviral infection in vivo. *J Virol*. 2014;88:12669–12682.
109. Kwong PD, Kwon YD, Chuang GY, et al. Informatics-based improvement of HIV-1 neutralizing antibody 10E8. Conference on Retroviruses and Opportunistic Infections 2016;Abstract Number: 310LB.
110. Kwon YD, Georgiev IS, Ofek G, et al. Optimization of the solubility of HIV-1-neutralizing antibody 10E8 through somatic variation and structure-based design. *J Virol*. 2016;90:5899–5914.
111. Chen HS, Hou SC, Jian JW, et al. Predominant structural configuration of natural antibody repertoires enables potent antibody responses against protein antigens. *Sci Rep*. 2015;5:12411.
112. Cohn M, Mitchison NA, Paul WE, Silverstein AM, Talmage DW, Weigert M. Reflections on the clonal-selection theory. *Nat Rev Immunol*. 2007;7:823–830.
113. Knappik A, Ge L, Honegger A, et al. Fully synthetic human combinatorial antibody libraries (HuCAL) based on modular consensus frameworks and CDRs randomized with trinucleotides. *J Mol Biol*. 2000;296:57–86.
114. Peng HP, Lee KH, Jian JW, Yang AS. Origins of specificity and affinity in antibody-protein interactions. *Proc Natl Acad Sci USA*. 2014;111:E2656–E2665.
115. Hsu HJ, Lee KH, Jian JW, et al. Antibody variable domain interface and framework sequence requirements for stability and function by high-throughput experiments. *Structure*. 2014;22:22–34.
116. Chang HJ, Jian JW, Hsu HJ, et al. Loop-sequence features and stability determinants in antibody variable domains by high-throughput experiments. *Structure*. 2014;22:9–21.
117. Yu CM, Peng HP, Chen IC, et al. Rationalization and design of the complementarity determining region sequences in an antibody-antigen recognition interface. *PLoS ONE*. 2012;7:e33340.
118. Andrabi R, Voss JE, Liang CH, et al. Identification of common features in prototype broadly neutralizing antibodies to HIV envelope V2 apex to facilitate vaccine design. *Immunity*. 2015;43:959–973.
119. Huang CC, Venturi M, Majeed S, et al. Structural basis of tyrosine sulfation and V-H-gene usage in antibodies that recognize the HIV type 1 coreceptor-binding site on gp120. *Proc Natl Acad Sci USA*. 2004;101:2706–2711.
120. Sundling C, Zhang Z, Phad GE, et al. Single-cell and deep sequencing of IgG-switched macaque B cells reveal a diverse Ig repertoire following immunization. *J Immunol*. 2014;192:3637–3644.

121. Chuang GY, Acharya P, Schmidt SD, et al. Residue-level prediction of HIV-1 antibody epitopes based on neutralization of diverse viral strains. *J Virol*. 2013;87:10047–10058.
122. Chuang GY, Liou D, Kwong PD, Georgiev IS. NEP: web server for epitope prediction based on antibody neutralization of viral strains with diverse sequences. *Nucleic Acids Res*. 2014;42:W64–W71.
123. Ofek G, Zirkle B, Yang Y, et al. Structural basis for HIV-1 neutralization by 2F5-like antibodies m66 and m66.6. *J Virol*. 2014;88:2426–2441.
124. Joyce MG, Zhang B, Ou L, et al. Iterative structure-based improvement of a fusion-glycoprotein vaccine against RSV. *Nat Struct Mol Biol*. 2016. In press.
125. Bowers KC, Chow E, Xu H, et al. , ed. Scalable algorithms for molecular dynamics simulations on commodity clusters, 2006 November 11–17; Tampa, Florida.
126. Shivakumar D, Williams J, Wu YJ, Damm W, Shelley J, Sherman W. Prediction of absolute solvation free energies using molecular dynamics free energy perturbation and the OPLS force field. *J Chem Theory Comput*. 2010;6:1509–1519.
127. Wang L, Deng Y, Knight JL, et al. Modeling local structural rearrangements using FEP/REST: application to relative binding affinity predictions of CDK2 inhibitors. *J Chem Theory Comput*. 2013;9:1282–1293.
128. Wang L, Wu YJ, Deng YQ, et al. Accurate and reliable prediction of relative ligand binding potency in prospective drug discovery by way of a modern free-energy calculation protocol and force field. *J Am Chem Soc*. 2015;137:2695–2703.