SYSTEMATIC ANALYSIS OF B-CELL REPERTOIRES

By

Elaine Chang Chen

Dissertation

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

in

Pathology, Microbiology, and Immunology

May 13th, 2022

Nashville, Tennessee

Approved by:

Ivelin S. Georgiev, Ph.D.

Kristen Ogden, Ph.D.

YuWei Zhu, Ph.D.

Jeffrey Rathmell, Ph.D.

James E. Crowe, Jr., M.D

To my parents, Arrina Chang and Tim Chen, thank you for your everlasting support.

ACKNOWLEDGMENTS

First and foremost, I am incredibly grateful for my thesis mentor Dr. James Crowe. His passion for science and tireless work ethic inspires me to become a better scientist every day. Having the opportunity to train in the Crowe lab has been one of the best experiences of my professional life. Thank you for taking a chance on me and setting me up for success through constant encouragement and support. You gave me the freedom, resources, and opportunities to explore and dive into my curiosities with the sky as the limit while teaching me to believe and have confidence in myself as a scientist. There is no other place I could imagine being more perfect for my scientific training, and I will be forever grateful for your mentorship.

I would like to thank the Department of Pathology, Microbiology, and Immunology at Vanderbilt University and the Vanderbilt Program in Molecular Medicine for providing me with the opportunities to train and learn throughout the last 3.5 years. I would like to acknowledge the support and mentorship from my committee members: Drs. Ivelin S. Georgiev, Kristen M Ogden, Yuwei Zhu, and Jeffrey C. Rathmell. I would also like to thank the mentorship I received from my clinical mentor: Dr. Clarence Buddy Creech. I am grateful for the support and guidance you have all given me to become a better scientist.

This dissertation would not have been possible without support from the following funding sources: the Immunological mechanisms of disease training program (T32-AI138932-01A1), Therapeutic human monoclonal antibody treatments for filoviruses (U19AI142785), Structure-based design of trimer interface epitope focused universal influenza vaccines (U01AI50739), Identification of novel human antibodies and vaccines against influenza (U19AI117905), Human antibody-based countermeasures against the Wuhan Coronavirus SARS-CoV-2 (R01AI157155), Human Vaccines Project, Dolly Parton COVID-19 Research Fund, and Defense Advanced Research Projects agency.

This body of work took a village and could not have been accomplished without the help of many collaborators. I would like to thank Dr. Alexander Bukreyev's group at the University of Texas Medical Branch, more specifically Philipp Illinykh and Kai Huang, for their help on authentic filovirus neutralization assays and animal studies to determine antibody protection *in vivo* of all my Ebola public clonotype antibodies. Emma Winkler from Dr. Michael Diamond's group at Washington University for help on authentic virus neutralization assays and animal studies for the SARS-CoV-2 work. Carley Cabel from Dr. Curtis Throne's group and Shuaizhi Li from Dr. Samuel Campos' group at the University of Arizona for authentic virus neutralization assays using double-stranded RNA staining of SARS-CoV-2. Edgar Davidson and Dr. Benjamin Doranz from Integral Molecular for their help on epitope mapping using alanine-scanning mutagenesis. Stefano Bonissone and Dr. Natalie Castellana from Abterra Biosciences help identify antibodies present in sera. And Erica Saphire's group at Scripps for providing Ebola glycoproteins used for memory B-cell sorting.

I would also like to thank all the past and present members of the Crowe lab. Thank you to Dr. Rob Carnahan for giving me numerous opportunities to learn, lead, and contribute to projects; Rachel Nargi for training me on microscale expression and purification as well as helping me purify many rounds of antibody; Dr. Elad Binshtein for always being willing to help me on EM experiments despite all the other structures you have to solve; Dr. Naveen Suryadevara for reagents and expertise on SARS-CoV-2; Andrew Trivette and Joseph Reidy for assistance in ordering large panels of antibodies; Rachel Sutton for helping keep things running and shipping out antibodies to collaborators while I was away at my internship; and Jessica Rodriguez, Erica Armstrong, and Chris Gainza for all the help on protein expression and purifications. A special thanks to Luke Myers and Sam Day, who have been instrumental in helping me with all things bioinformatics. From teaching me scripting to fixing my code, thank you for letting me spend hours upon hours sitting looking over your shoulders in tech core and

allowing me to learn from you both. I would also like to thank my fellow trainees, both past and present, for their advice, friendship, and support.

I am also highly grateful to Dr. Cinque Soto for taking me under your wing when I first joined the lab. Thank you for dedicating time and patience to training me from someone with no prior computational experience to being computationally literate. I also wanted to thank you for constantly challenging and pushing me to become an independent scientist. You taught me to be confident in my scientific knowledge and not shy away from speaking up and taking chances when opportunities are presented.

I would also like to express my gratitude to Drs. Seth Zost and Pavlo Gilchuk for helping me become the scientist I am today. Thank you both for your guidance, expertise, and support. No matter how busy you are, you somehow always find the time to take a meeting, help, or train me on new techniques or discuss the best ways to approach papers and projects. Thank you both for the opportunity to learn from and work with both of you in the last couple of years.

On a more personal note, thank you to my friends who I have met along this crazy ride of graduate school: Esther Kang and Jonathan Davies for our weekly game nights that kept me sane through the last few years; Sydni Smith and Sirena Tran for our little support group cheering each other on through every milestone; Sam Lisy for always pumping me up and giving me that boost; Lindsey Guerin for all the wine and cheese nights; and Ivette Perez for our impulse coffee dates. I am grateful for all of your friendships and memories that we have shared here in Nashville and look forward to future adventures. Thank you to my friends outside of graduate school who have continually cheered me on from the sidelines, from encouraging me during graduate school applications to writing this dissertation: Sara Yasrebi, Vicky Ho, Christine Madamba, Jasmine Wang, Kristina Coleman, and Dave Molina. I am so grateful for each and every one of you and could not have made it this far into graduate school without your support.

I also thank my family. First, my parents, Arrina Chang and Tim Chen, who have always strived to put our education first growing up, encouraged us to pursue higher education and be great role models to look up to growing up. Thank you for constantly pushing me to be curious and continually learn, shaping me into the person I am today. I am also grateful for my sisters and best friends: Eunise and Ellie Chen, who has always been just a phone call away for laughter and tears. You both inspire me every day to work hard and dream big. I can't wait for the day we celebrate all 3 Dr. E.C. Chens! And lastly, thank you to my partner, Jackie Mak, for the continual encouragement to chase my dreams. From driving me between jobs and classes and picking me up at midnight from the lab back in undergrad to moving to Nashville to support me during the ups and downs of graduate school - thank you for your unwavering love and support. I cannot wait for our next adventure.

Table of Contents

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS/NOMENCLATURE

AID	Activation induced cytidine deaminase
AL amyloidosis	Light chain amyloidosis
APC	Antigen presenting cells
cDNA	Complementary DNA
CDR	Complementarity determining regions
D gene	Diversity gene
EBOV	Ebola virus
EBV	Epstein-Barr virus
ELISA	Enzyme linked immunoassay
FACS	Fluorescence activated sorting
FR	Framework region
gDNA	Genomic DNA
GP	Glycoprotein
HA	Hemagglutinin
HIV	Human Immunodeficiency Virus
IAV	Influenza A
IBV	Influenza B
IGH	Immunoglobulin heavy
IGK	Immunoglobulin kappa
IGL	Immunoglobulin lambda
J	Joining gene
MHC	Major histocompatibility complex
mRNA	Messenger RNA
NA	Neuraminidase
NGS	Next generation sequencing
NTD	N-terminal domain
PBMC	Peripheral blood mononuclear cells
PCR	Polymerase chain reaction
RBD	Receptor binding domain
RTCA	Real time cell analysis assay
S protein	Spike protein
SHM	Somatic hyper mutation
TCR	T-cell receptor
TdT	Terminal deoxynucleotidyl transferase
T_{FH}	T folliular helper cells
T_{reg}	Regulatory T cells
V	Variable gene

CHAPTER I

INTRODUCTION

Thesis overview

This body of work is the culmination of my dissertation on the interrogation of antibody repertoires utilizing B-cell receptor sequencing. My dissertation is divided into seven chapters. In chapter I, I provide an overview on antibodies, utilization of sequence analysis in antibody repertoire studies, and background on each of the disease states discussed in this dissertation: light chain amyloidosis, Influenza, Ebola virus, and SARS-CoV-2 virus.

In chapter II, I describe the development of a sequence analysis toolbox for repertoire-wide sequencing studies. First, I describe an antibody sequencing study performed on bone marrow aspirates from seven patients diagnosed with light chain (AL) amyloidosis, which revealed a dominating light chain in each patient's repertoire, the contributing cause of disease. The data also revealed differing patterns of overall antibody repertoire disruption in different patients. Analysis pipelines built during this study were then applied to understanding antibodies isolated from one of the first SARS-CoV-2 infected individuals in North America. Additional methods were then developed to facilitate matching of exact complementarity determining regions amino acid sequences to identify public T-cell receptor clonotypes in NCBI's GenBank repository.

In chapter III, I describe the applications of utilizing clustering of antibody sequences in conjunction with high-throughput characterization of antibodies to influenza related antibody discovery projects. Utilizing clustering of antibody sequences in conjunction with functional datasets permitted antigen-specific assignments to sequences that would otherwise have unknown antigen reactivities. Additionally, this method was applied to multiple antibody

discovery efforts where there were large sets of antibody sequences, to systematically down select for antibody synthesis and characterization. Such techniques were applied to characterize large panels of antibodies to influenza A and B as well as antibodies identified within the overlap of antibody sequences present in the plasmablast and bone marrow long-lived plasma cells of donors vaccinated with the annual influenza vaccine. The ability to cluster similar sequences to identify clonal families led to further expanding the sequence analysis tool box to building phylogenetic trees of sequences from the same lineage. This allowed for the inference of unmutated common ancestors of interesting antibody lineages, which was executed on a class of influenza A hemagglutinin trimer interface-directed antibodies.

In chapter IV, I describe work done on understanding the private antibody response to the Ebola glycoprotein (GP) from a survivor of the 2014 Ebola virus outbreak. Here, I detail the largest collection of paired antibody sequences originating from a single donor to a viral antigen. The depth of paired sequencing achieved in this study allowed for investigation on the genetic and functional diversity of the B-cell repertoire within an individual to a viral antigen, and identification of a population of antibodies present in both the serum and memory B cell repertoire. Additionally I identified several clonally expanded families of antibodies that arose from germline-encoded binding and neutralizing properties.

In chapter V, I describe the public antibody response to the Ebola glycoprotein within convalescent and vaccinated individuals. The large set of paired antibody sequences achieved from the experiments described in chapter IV served as an optimal platform to mine for public clonotypes. This study reveals the high prevalence and systematic characterization of public clonotypes to the EBOV GP. Neutralizing public clonotypes targeted diverse epitopes on the GP, a subset of which possess cross-reactive activity. These neutralizing public clonotypes also inferred protection in mice *in vivo* to lethal challenge. Several public clonotypes retained neutralization capacity once reverted to germline sequence, indicating that naïve B cells may

encode for EBOV GP reactivity, leading to the elicitation of protective public clonotypes in multiple donors.

In chapter VI, I describe the public antibody responses to the SARS-CoV-2 spike protein in convalescent and vaccinated individuals. Here, public clonotypes shared between convalescent donors were first identified, with an emphasis of three novel public clonotypes previously undescribed, deemed as groups 1, 2, and 3. Groups 1 and 2 public clonotypes are the first public antibodies to be described to the S2 domain of the spike protein. The group 3 public clonotype targets a cross-reactive epitope to SARS-COV and SARS-CoV-2, and confers *in vivo* protection in mice after viral challenge. After deep single cell sequencing of individual's response to the Pfizer-BioNTech vaccine, additional public clonotypes were identified for future studies.

In chapter VII, I summarize the body of work I completed during my dissertation. I also discuss the future directions implicated for the work described in this dissertation.

Part I: Antibodies

The adaptive immune system and memory retention

In the event of exposure to a foreign pathogen, the innate arm of the immune system is activated. Cytokines will attract antigen presenting cells (APCs) such as macrophages and dendritic cells, which will phagocytose the pathogen, and express the antigens on their major histocompatibility complex (MHC) (Kenneth, 2008).

These APCs then migrate to the lymph nodes to activate the adaptive arm of the immune system. The adaptive immune system encompasses T and B cells, adaptive immunity's cellular and humoral components. Antigens present on the MHCs of APCs activate naïve T-cells to differentiate. CD8+ T-cells differentiate into memory T-cells and cytotoxic T-cells. CD4+ T-cells

can differentiate T_H1, T_H2, T_H17, regulatory T-cells (T_{reg}), and T follicular helper cells (T_{FH}). Here in the lymph nodes, naïve B cells will be activated by $T_{H1/2}$ and T_{FH} cells, allowing for naïve B cells to start clonal expansion and affinity maturation, allowing for the development of B cells with an increased affinity to the target antigen. These "new and improved" B cells are then activated for class switching and differentiation into plasma cells, memory B cells, and plasma-blasts (Kenneth, 2008). Plasma cells migrate to the bone marrow, where they reside and secrete antibodies into the circulation. Memory B cells will either enter circulation or stay in the lymph nodes. And plasmablasts will enter circulation and secrete antibodies as the immediate response to the antigen with a life span ranging from 3-5 days (for short-lived plasma cells) to several months (for long-lived plasma cells) (Khodadadi et al., 2019) (**Figure I-1**).

If the pathogen is reintroduced, the adaptive immune system is now poised for a quick and timely response. Memory B cells specific to the antigen are positioned to be quickly activated for expansion and differentiation once re-exposure occurs. Plasma cells will continually secrete high-affinity antibodies in circulation, surveying the body for antigens they recognize. Having antibodies patrol in circulation allows the immune system to quickly respond to pathogens as antibodies are responsible for various effector functions, including neutralization, opsonization, phagocytosis, antibody-dependent cellular cytotoxicity, and complement activation (Palm and Henry, 2019; Quast and Tarlinton, 2021).

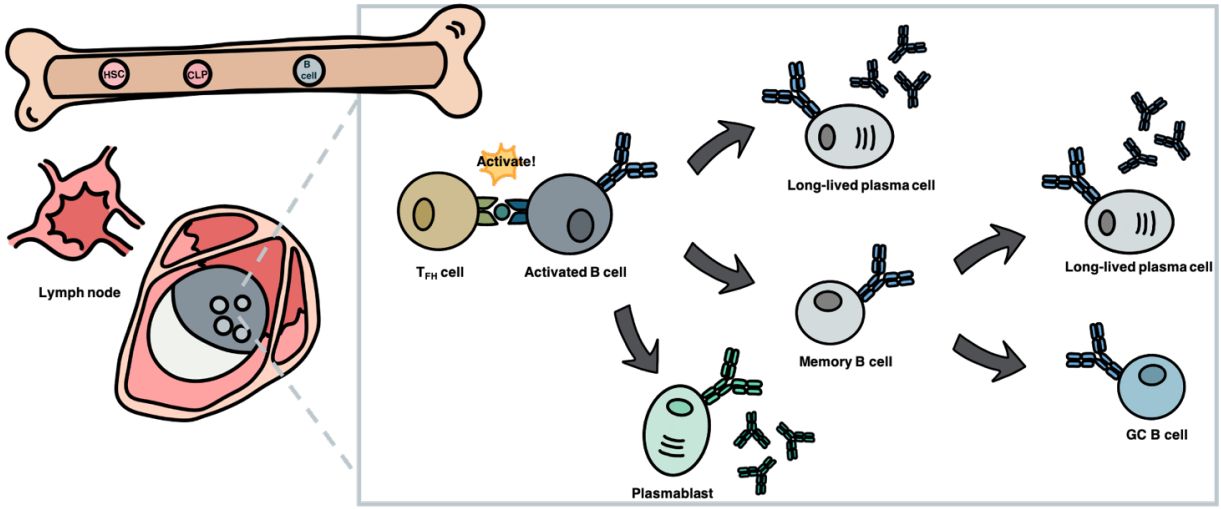


Figure I-1: Adaptive immune system.

A schematic illustrating the link between T and B cells allowing for B-cell activation. Naïve B-cells migrate from the bone marrow to lymph nodes allowing for activation by T cells in response to a foreign antigen. Activated B cells then differentiate into multiple B-cell subsets: Long-lived plasma cells, memory B cells, or plasmablasts.

<u>Understanding antibody repertoires informs vaccines and therapeutics</u>

Memory B cells and antibodies in circulation can be thought of as a logbook, intricately capturing all foreign antigens an individual has responded to in their lifetime. As an individual survives different infections or is exposed to varying antigens throughout their lifetime, the humoral immune system stores each exposure in the memory in the form of antigen-specific antibodies/ B cells. And such antibodies were likely partially responsible for helping the individual recover from said infections. Therefore the B-cell repertoire within an individual is a library of antibodies to all the pathogens that an individual had been exposed to in the past.

Prior to the development of abilities to isolate monoclonal and polyclonal antibodies, utilizing plasma as therapy for infectious diseases was exercised for severe infections, revealing

that antibodies mount protection and treatment against pathogens. Since then, polyclonal antibodies from survivors of infection have been used for decades to prevent viral diseases (Stokes, 1951), further showing the utility of antibodies in treating and preventing diseases. On the monoclonal level, antibodies identified in survivors of different infections have since then been studied and developed to combat diseases as either prophylaxis or treatment. Kohler and Milstein produced the first hybridoma in 1975, where human B cells are fused with multiple myeloma cells after immortalization of B cells using Epstein-Barr virus (EBV), allowing for screening of binding and neutralization to viral targets (Kohler and Milstein, 1975). This technology has since been applied to many targets in infectious and non-infectious disease indications such as Palivizumab as a prophylactic to respiratory syncytial virus (RSV) and anti-CD3 antibodies for cancer.

Understanding the human antibody response to different pathogens can be beneficial as it can lead to the discovery of clinical antibody candidates. On the flip side, it can also inform the development of vaccines as understanding the antibody response on a molecular level allows for detailing of neutralizing and non-neutralizing epitopes on different antigens, which can be utilized for rational vaccine design as well as benchmarking vaccine candidates (Pedrioli and Oxenius, 2021).

The anatomy of an antibody

B cells display a membrane-bound immunoglobulin known as the B-cell receptor (BCR). An antibody is the secreted form of a BCR. The structure is identical to those of an antibody except for a small portion of the carboxy terminus of the heavy chain constant region (Schatz et al., 1989).

Antibodies are composed of two heavy chains and two light chains. These polypeptides are held together by a disulfide bond to form two functional regions of an antibody: Fab (fragment antibody binding) and Fc (fragment crystallizable). The variable region is the top half of the fab.

The heavy chain variable region comprises a V, D, and J gene, and the light chain variable region is composed of a V and J gene. The tip of the variable region encompasses three complementarity determining region (CDR) loops, also known as hypervariable regions, CDR1, CDR2, and CDR3. There are three loops on the heavy chain and three on the light chain, which clasps onto the antigen, accounting for the antibody's specificity (Schramm and Douek, 2018). Mutations for antibodies are usually concentrated in these regions to allow for specificity to antigen. Usually, the CDR3 region is known to be the most mutated out of the three loops and therefore also is known to make up the majority of antigen contact residues. The constant region is the bottom part of the Y, and determines the effector functions for the antibody (Kenneth, 2008) (**Figure I-2**).

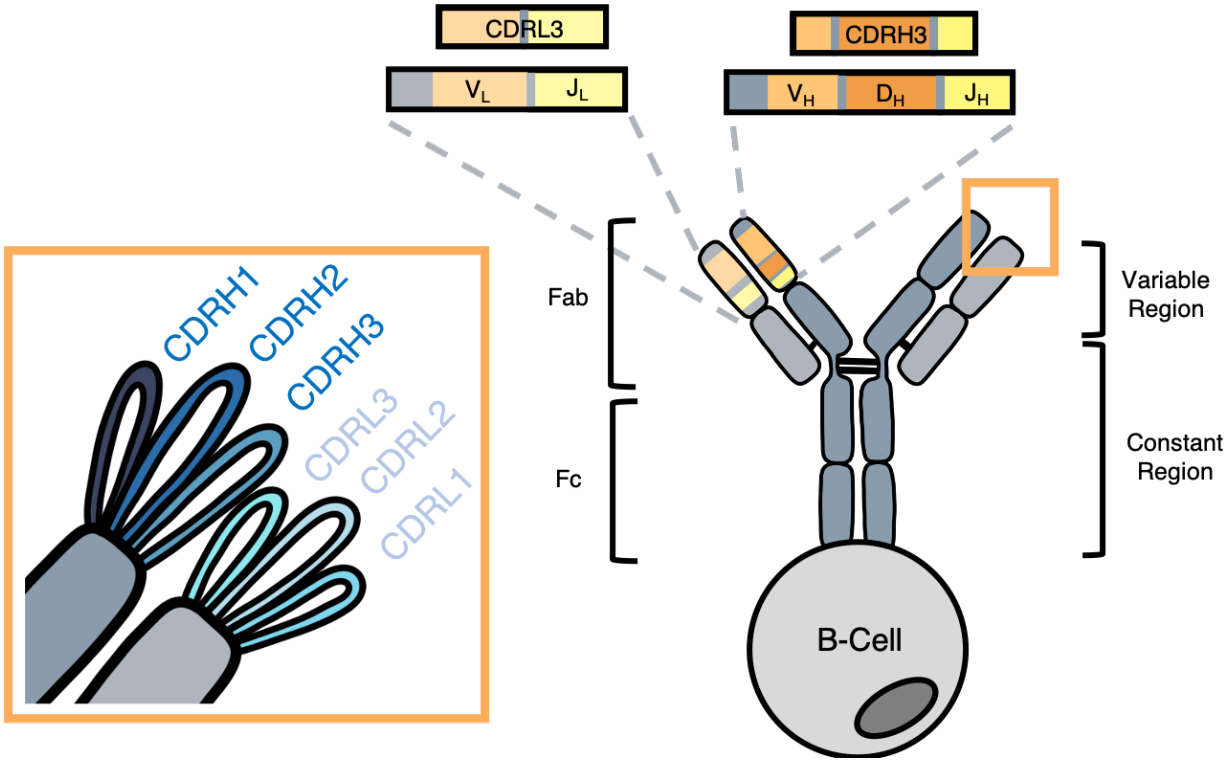


Figure I-2: The anatomy of an antibody

Schematic of a B-cell receptor with the heavy chain in dark grey and light chain in light grey. Variable and constant regions are identified. The variable region of the antibody is composed of a V, D, and J gene on the heavy chain and a V and J gene on the light chain. The tips of the variable region encompasses three loops on the heavy and light chain, known as the complementarity determining region (CDRs). Based on Janeway Immunobiology, 9th edition.

VDJ recombination leads to high diversity of antibody sequences

B cells start development in the bone marrow. A B cell will start its fate as a hematopoietic stem cell (HSC), following a common lymphoid progenitor(CLP) cell, where it commits to either becoming a T-cell or B-cell. Once the cell differentiates into a Pro B cell, it seals its fate as a B cell and undergoes VDJ recombination to become a naïve B cell.

During VDJ recombination, various variable (V), diversity (D), and joining (J) genes for the heavy chain, and V and J genes for the light chain, are recombined randomly using recombination

activating genes, forming B-cell receptors (**Figure I-3**). With approximately 40 IGHV gene segments, 27 IGHD gene segments, and 6 IGHJ gene segments, the recombination of one V, D, and J segment per antibody approximates roughly 10^4 variations of possible heavy chains. This diversity is further amplified as the same process is applied to the light chains V and J genes yielding approximately 320 light chain recombinations. The pairing of different heavy and light chains to complete an antibody is, therefore, further diversified for roughly 3.5 x 10^6 different antibodies. In addition to combinatorial diversity, an additional layer of diversity is added at the junctions. Junctional diversity is added at the junctions between different gene segments during the recombination stage. Recombination-activating genes (RAG) enzymes will add p-nucleotides, and exonucleases may trim nucleotides encoded by the germline gene segments, leading to additional codon variations in the junctions between each gene segment. Terminal deoxynucleotidyl transferase (TdT) may also add non-templated nucleotides within junctions, adding to the junctional diversity. The junctional diversity contributes to roughly 10^{11} different antibodies (Kenneth, 2008; Soto et al., 2019)(**Figure I-3**).

Once the B cell has undergone VDJ recombination, it is deemed a naïve B cell where it undergoes testing against central tolerance using stromal cells in the bone marrow. If the naïve B cells are non-autoreactive, they are positively selected and migrate to the lymph nodes that house the germinal centers. In the germinal centers, naïve B cells can be activated by binding to antigens present on APCs or T_{FH} cells. Once activated, these B cells undergo clonal expansion and affinity maturation through somatic hypermutation (SHM) (Victora et al., 2010). During this process, activated B cells will proliferate and hypermutate, creating many clones with either an increase or decrease in affinity to the antigen that stimulated the B cell in the first place.

Random mutations are introduced into the heavy and light chains by activation-induced cytidine deaminase (AID). AID acts on cytidine and deaminates the residue, allowing for

introduction of mutations. Many of these mutations are located in the complementarity determining regions (CDRs) (Wei et al., 2015). If antibodies increase their affinity to antigen after introducing mutations, these antibodies will continue undergoing the next round of SHM, leading into a positive feedback loop, where these antibodies will then undergo the next round of mutations. Consequently, if antibodies decrease their affinity to antigen, these B cells will undergo apoptosis. Through this process, antibodies are affinity matured and gain the ability to bind with a higher affinity to the antigen it is selected for, therefore introducing an additional layer of diversity.

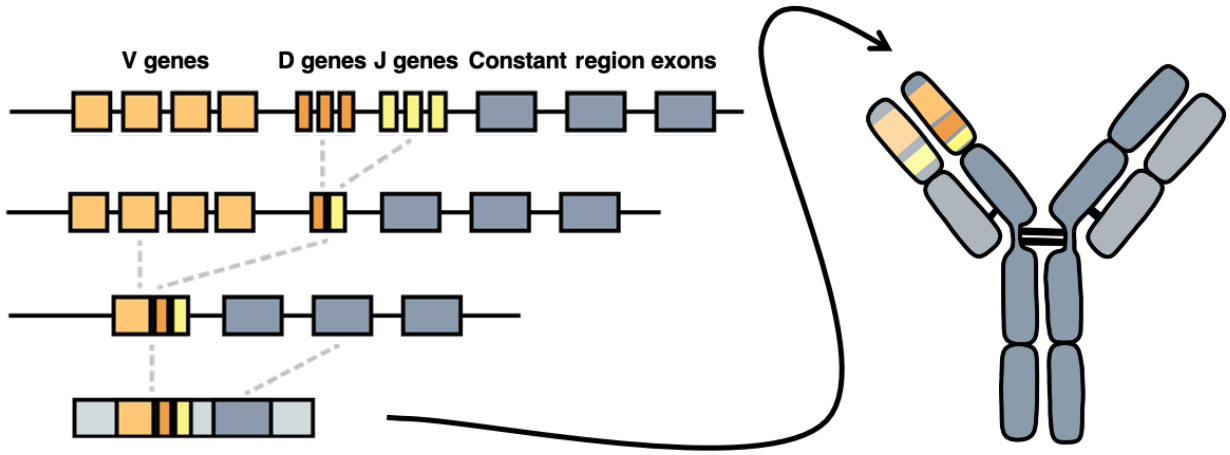


Figure I-3: VDJ recombination

V genes (orange), D genes (dark orange), and J genes (yellow) recombine to make naïve B-cells in the bone marrow. These gene segments are then recombined with constant regions (grey) to determine the isotype of the antibody.

Methods of antibody discovery and sequencing

Since the beginning of antibody discovery originating from the first hybridoma made in 1975 (Kohler and Milstein, 1975), many antibody discovery technologies have been developed to deploy high through-put antibody discovery. The multitude of antibody discovery methods

available can be split into two types: front-end sequencing-based methods and front-end screening-based methods.

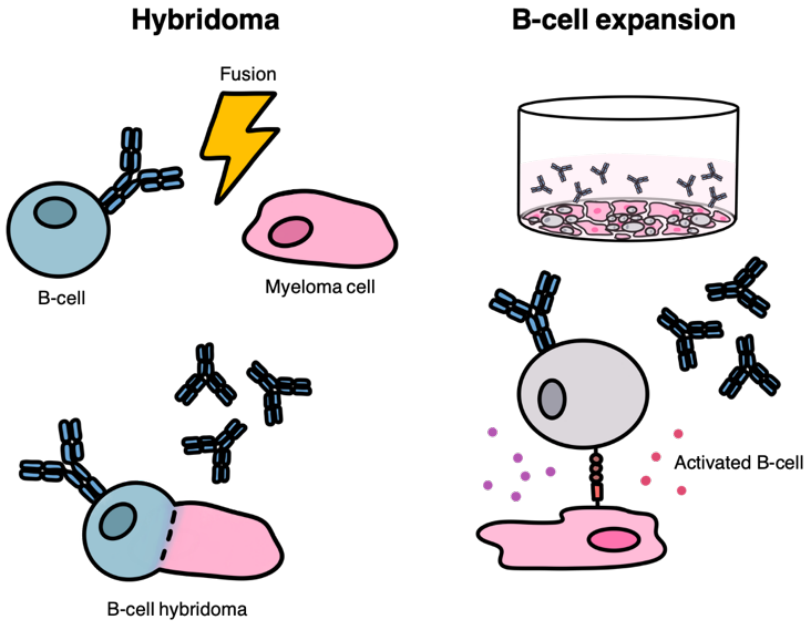


Figure I-4: Hybridoma and B-cell expansion technologies

Left panel illustrates the hybridoma process. B cells are fused with myeloma cells to generate B cell hybridomas, an immortalized cell line for that B cell which will continually secrete antibody into the supernatant. Right panel illustrates the B cell expansion process using 3T3 cells. 3T3 cells with CD40 ligand activates B cells along with cytokines in the media, causing B cells to continually secrete antibody into the supernatant.

With the hybridoma methodology, B cells are fused with myeloma cells after immortalization of B cells using Epstein-Barr virus. These hybridomas grow in culture, secreting antibodies in the culture supernatant. Secreted antibodies can then be screened for binding or neutralization to the target antigen or pathogen (Smith and Crowe, 2015). Similar to the hybridoma methodology, there have also been methods developed for B cell expansion without undergoing the fusion process. B cells of interest are counted and diluted to limiting dilutions and co-cultured with a mouse embryonic fibroblast cell line (3T3) along with anti-CD40 ligand and IL-4 to stimulate B-cell growth (Nojima et al., 2011).

Similarly, these stimulated B cells secrete antibodies in the culture supernatant can be used for screening assays. Hybridomas or stimulated B cells yielding positive hits from the screen are then sequenced individually (**Figure I-4**). With both methods, screening assays are usually done on the front end to down select B cells by their functional phenotype before individual B-cell colonies are sequenced. As a result, panels of antibodies are created with functional characteristics in mind.

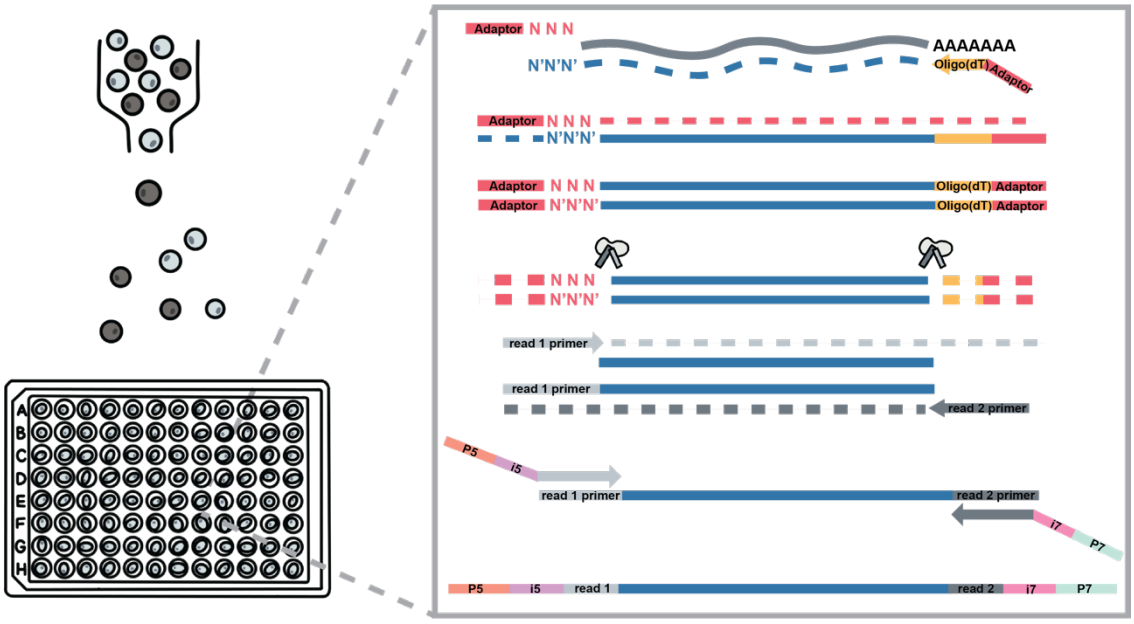


Figure I-5: Single cell sequencing workflow

Single cell sequencing workflow is depicted. B cells of interest are sorted into lysis buffer contained in a 96-well plate. Following, cDNA is generated. If sequencing is done via sanger sequencing, the amplified cDNA can be sequenced. To proceed with next generation sequencing, adaptors are cleaved off and Illumina primers and indexes are added on for flow cell sequencing compatibility.

Conversely, a more agnostic approach utilizes antigen-specific B cell sorting into 96-well plates. Generally, antigens are used to label memory B cells, and cells are sorted using fluorescence-activated cell sorting (FACS). These B cells are sorted into single cells into lysis

buffer where the mRNA is reverse transcribed to make cDNA for each cell. The heavy and light chains are amplified by polymerase chain reactions (PCR) separately using variable gene and constant region-specific primers. Alternatively, one can skip the cDNA generation step and directly go from cell in lysis buffer to PCR reactions (Smith et al., 2009). In both methods, PCR products are then sequenced accordingly to achieve paired heavy and light chain sequencing using either Sanger sequencing or indexed for Illumina-based sequencing (**Figure I-5**). Although with this method, the ability to screen antibodies after sorting into plates using multiple functional assays is lost. Therefore, screening to antigen reactivity is often done on the front end via gating using flow cytometry. Therefore, all antibodies for which you get productive sequencing are ideally specific to your target.

Each B cell is theoretically contained in a well with all the methods described above. Therefore, when obtaining the heavy and light chain sequences, the native pairing of each antibody is retained with the well acting as an index linking the heavy and light chain. These natively paired heavy and light chain sequences can then be used to produce recombinant versions of identified antibodies to further characterize the functional phenotypes of discovered antibodies. However, there is a limit on the number of plates an individual can process at a single time for all of these methods, therefore capping the throughput of these methods at hundreds of antibodies in one round of discovery (Smith et al., 2009).

With significant decreases in the cost of sequencing, many have circumvented the bottleneck of scale with these methods and increased the sample size of B cells sequenced using bulk B-cell sequencing. mRNA from a B cell sample is isolated in bulk and reverse transcribed using antibody variable gene and constant region-specific primers to create cDNA libraries. These libraries are then sequenced using next-generation sequencing (NGS) and bioinformatically filtered (Robinson, 2015). Sequences derived using NGS methods range between millions of reads for each prep depending on the sample (**Figure I-6**). Due to the

exponential increase in antibody sequences derived from bulk sequencing, repertoire wide studies could be performed to have an understanding of what the antibody repertoire looks like on a sequence level at a scale that was unachievable by the hybridoma, B-cell expansion, or single-cell sequencing methods (Burkholder et al., 2017).

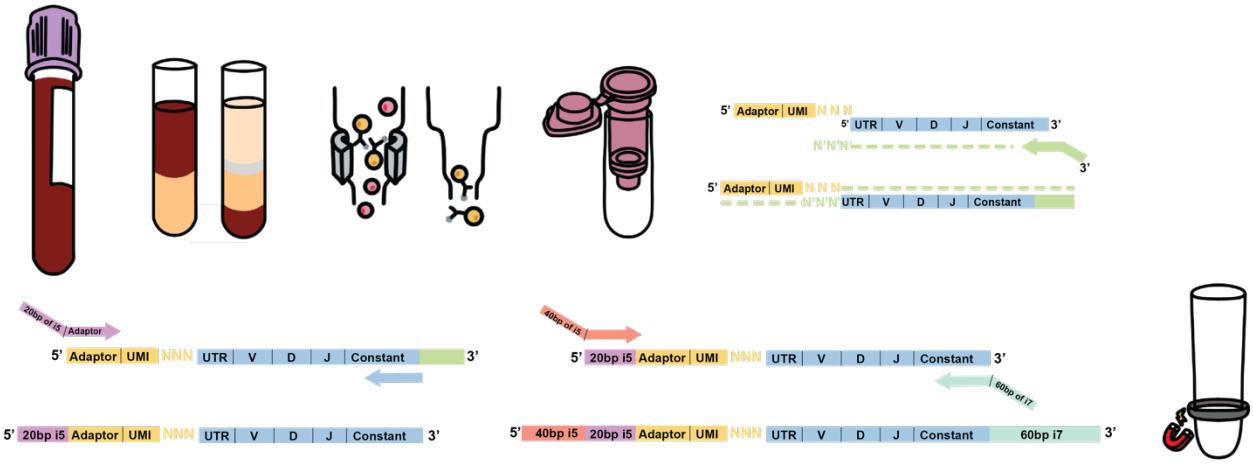


Figure I-6: Next generation sequencing

Schematic of next generation sequencing library prep. PBMCs are isolated from blood, and total B cells are enriched. From the B-cell pool, bulk mRNA is isolated for the following library prep. cDNA is generated and amplified with gene specific primers. Illumina indexes are appended at the end and a magnetic bead cleanup is done before loading onto a sequencing cell.

Despite the large scale of antibody sequences derived from these studies, there was no method of taking the further step of the investigation into understanding what these antibodies functionally do. As the native pairing of each antibody's heavy and light chain sequences is lost at the bulk mRNA isolation step of the protocol, there is no method of producing the antibody recombinantly to test in functional assays. The only way to assign a function to the antibodies which were bulk sequenced was to see if antibodies with a similar sequence to the one in question were previously discovered using the hybridoma, B-cell expansion, or single-cell

sequencing methods. There is often a tradeoff between getting large sets of antibody sequences and having the ability to functionally characterize antibodies (**Figure I-7**).

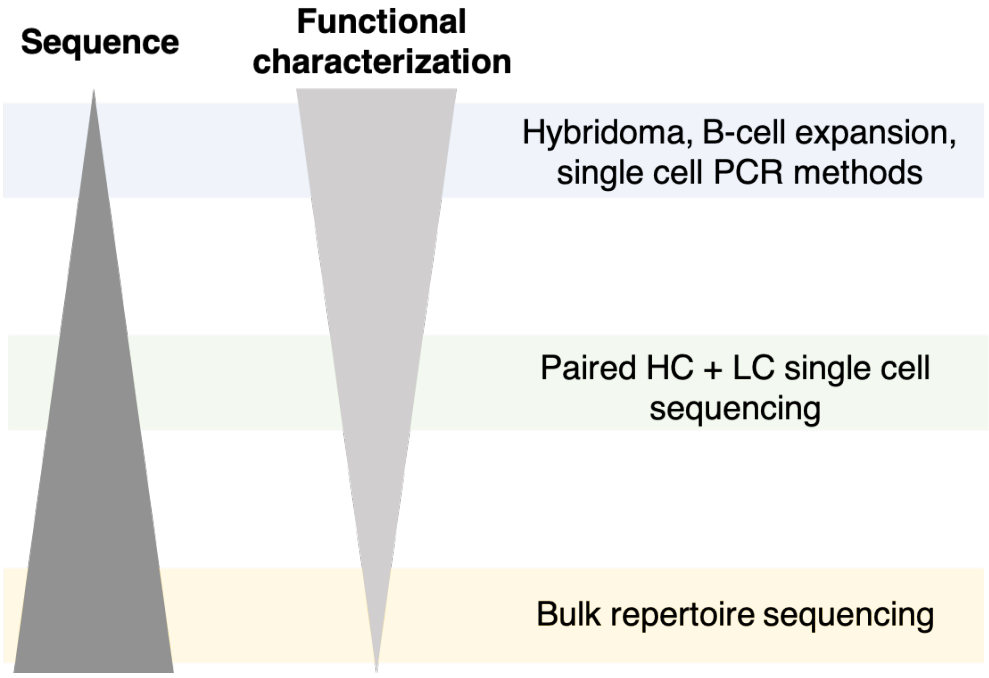


Figure I-7: The trade-off between antibody sequencing and functional characterization

Number of antibodies is indicated in the thickness of each triangle. Abundance in sequence (dark grey) is shown on the left, and abundance in functional characterization (light grey) is shown on the right. Each antibody discovery method is indicated in the boxes in relationship to the amount of sequencing or functional characterization for that method.

Using paired chain single cell sequencing in antibody discovery

Until recent technology was developed, it was impossible to perform high-throughput sequencing of paired heavy and light chain sequences. The only method to achieve natively paired heavy and light chain sequences was through either hybridoma, B-cell expansions, or single-cell PCR (Pedrioli and Oxenius, 2021). This was a bottleneck for being able to sample a more significant portion of the B-cell repertoire while retaining the ability to functionally validate

antibodies. To overcome the scarcity of paired heavy and light chain sequencing, Georgiou and colleagues developed using single-cell encapsulation of B cells with emulsion overlap extension PCR linking the heavy and light chain transcripts (DeKosky et al., 2013). Additionally, 10X genomics released a commercially available kit where single B cells are encapsulated in oil droplet emulsions with barcoded beads, allowing for indexing of both heavy and light chain sequences from the same cell during reverse transcription and PCR. As unique barcodes index each cell's heavy and light chain at the cDNA level, the library prep can be performed subsequently and sequenced using NGS (Goldstein et al., 2019). Output sequences using this methodology are then filtered bioinformatically and the naively paired heavy and light chains are identified from the barcoded index (**Figure I-8**).

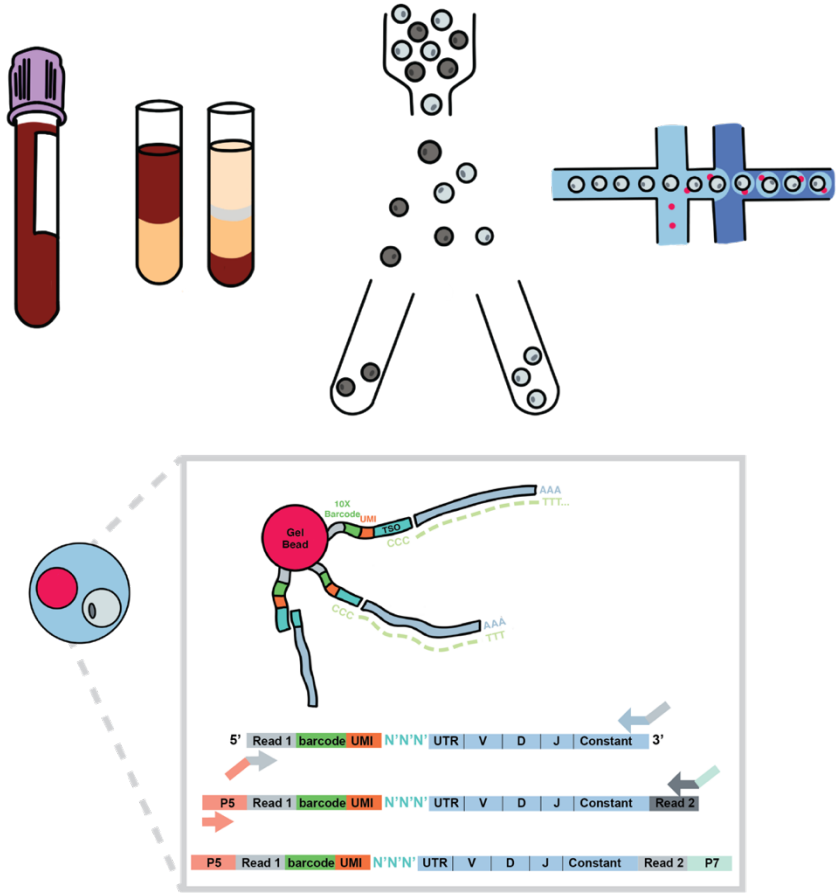


Figure I-8: Paired single cell sequencing

Schematic of paired single cell sequencing. Starting from a blood draw, followed by PBMC isolation, and sorting for the cell population of interest. Sorted cells are flowed through a microfluidic device and encapsulated in an oil emulsion droplet in a barcoded bead. Such barcoding technology allows for retention of native heavy and light chain pairing.

The scale of paired antibody sequences from a single sample of 7,000 input cells can be in the range of couple of thousands, thereby dramatically increasing the sample pool of paired antibody sequence generation by more than ten fold from traditional discovery methods. As each antibody's native heavy and light pairing is retained in this method, sequences derived from such experiments may be expressed and functionally characterized. These technologies radically transformed the scale at which antibody discovery could be performed.

Microscale expression, purification, and validation of antibodies

In conjunction with the advent of paired single-cell sequencing, methods have been developed to express and purify antibodies at a small scale, enabling high-throughput screening and characterization of antibodies (Gilchuk et al., 2020a). This yields large amounts of functional data in combination with paired full-length sequencing allowing for sequence analysis on the repertoire scale with the ability to relate it back to the functional characteristics of antibodies.

Sequences of the heavy and light chains belonging to each antibody are synthesized in 96-well plates. They are then transfected and purified in 96-well format, usually yielding enough antibody to screen for binding and neutralization to the targets of interest (**Figure I-9**). Having antibodies in a 96-well format is an optimal layout for downstream assays. For example, antibodies could then be easily screened against four antigens on a single 384-well plate for an enzyme linked immunoassay (ELISA) based screen. Additionally, neutralization assays may be optimized to operate in a 96-well format as well; antibodies can be tested for neutralization at a single concentration. With the use of the real-time cell analysis assay (RTCA), the neutralization of several viruses can be measured by cellular impedance in a 96 or 384 well format. Utilizing

the 384-well format allows testing of each antibody at four different concentrations within the same experiment, which will help rank panels of antibodies by potency (Zost et al., 2020b).

These methods enable the ability to characterize hundreds of antibodies in multiple functional assays. As antibody discovery is often analogized to finding a needle in the haystack, broadening the starting pool of antibodies with such methods will help increase the chances of finding "the one". Additionally, these methods allow for high-throughput characterization of hundreds of antibodies a time, allowing for characterization of antibodies on a repertoire-wide scale.

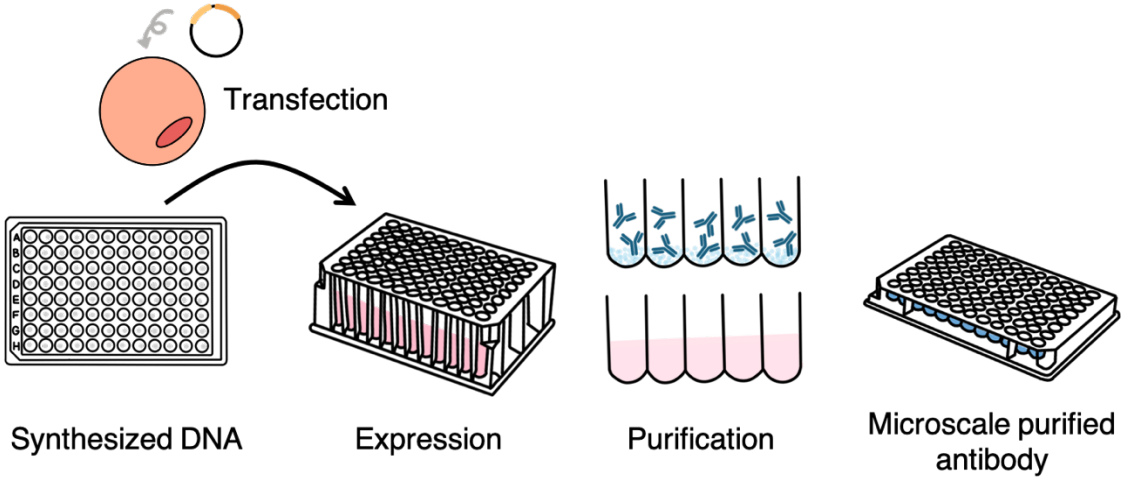


Figure I-9: Schematic of microscale expression and purification process

Plasmids are first synthesized by Twist Bioscience as previously described (Chng et al., 2015), and transfected into ExpiCHO cells. These cells are expressed for 4 days, upon which they are harvested per manufacture's directions. Antibodies are purified by using protein A resin loaded into columns, resulting in purified antibodies in 96-well format.

Berkeley Lights Beacon

Another recent advancement in antibody discovery is the Berkeley Lights Beacon. This optofluidic instrument allows for thousands of B cells to be loaded into their "pens," enabling for

measurements of antigen reactivity for thousands of cells simultaneously. As activated B cells are loaded into each pen and continue to secrete antibody, beads coated with the desired antigen is then floated over each of the pens to measure antigen reactivity. If the B cell in a pen secretes antibodies specific to the said antigen, there would be a “bloom” of color. B cells of interest can then be exported from the instrument for sequencing (**Figure I-10**). Using the Beacon as a screening method allows for thousands of antibodies to be screened simultaneously, which can help down select which B cells are of interest for further studies (Zost et al., 2020b).

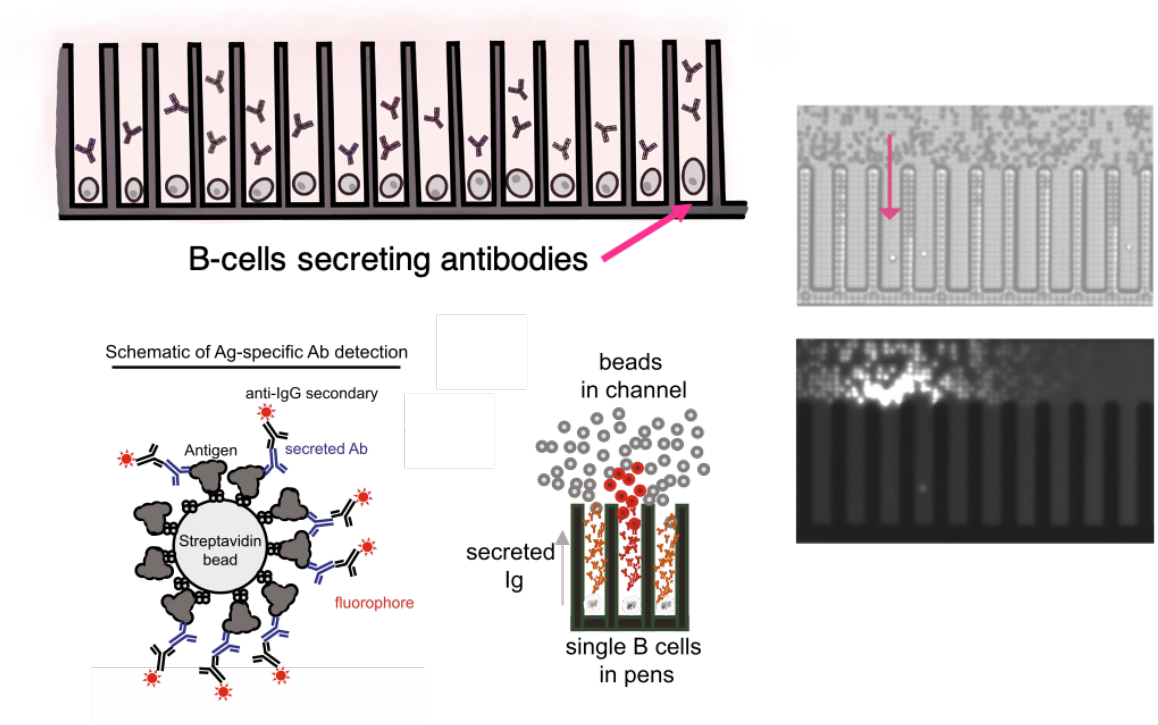


Figure I-10: Schematic of the Berkeley Lights Beacon “bloom” assay

The top panel shows a cartoon rendition of how the B cells are loaded into each pen. The bottom panel is a schematic of how the beads are coated with antigen, and how antibody binding in each pen is detected. If the antibodies produced by the B cell in said pen are antigen reactive, there will be fluoresce detected at the top of the pen. On the right side are pictures taken from the Berkeley Lights Beacon showing the assay on the instrument. Figure taken from Zost *et al.*, Nature Medicine 2021 and Chen *et al.*, Cell Reports 2021.

PART II: Antibody repertoire sequence analysis

Gene usages and relationships to functional properties

As previously mentioned, an antibody's heavy chain is encoded by a V, D, and J gene, and a V and J gene encodes the light chain. The combinations of these genes in conjunction with somatic mutations acquired during the maturation determine the antibody's specificity to antigen and thereby functional characteristics such as viral neutralization. Therefore, several features are commonly identified when approaching an antibody sequence data set (Chaudhary and Wesemann, 2018).

Antibody variable genes extend from framework (FR) 1 to 4, with the three CDRs scattered in the middle. The V gene encodes for most of the antibody variable region covering from FR1 to the CDR3, and the J gene encodes the latter half of the variable region from CDR3 to FR4. During VDJ recombination on the heavy chain, the D gene is first recombined with the J gene creating a D-J segment. This D-J segment is then recombined again with a V gene, creating a V-D-J transcript (**Figure I-11**). During each step of recombination, the ends of genes are commonly trimmed or mutated, leaving the D gene on the heavy chain, often unrecognizable. Therefore, the V gene and J gene identification from antibody sequences are often one of the initial steps when analyzing sequences (Yaari and Kleinstein, 2015).

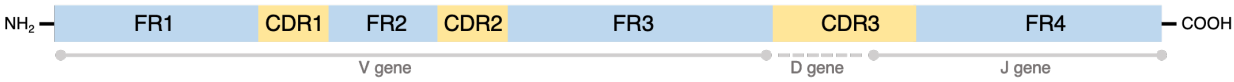


Figure I-11: Antibody heavy chain variable sequence.

Schematic illustrating the antibody variable region. The framework regions of the antibody sequence are in blue. The complementarity determining regions are in yellow. Each of the V, D, or J genes encoding for that region are identified at the bottom in grey.

As illustrated by the central dogma, DNA sequences are translated into amino acid sequences, allowing for protein folding. At the sequence level, amino acids encompass different biochemical properties such as polarity, acidity, hydrophobicity, and charge, contributing to the antibody's functional activity once the protein is folded. Antibody-antigen interactions involve a variety of forces such as electrostatic forces, hydrogen bonds, van der waals forces, and hydrophobic forces. With the use of high-resolution structural biology, critical residues contributing to antibody-antigen interaction are identified, leading to the identification of sequence motifs with favorable properties as antiviral antibodies. As the V and J gene encodes the majority of the variable region, several sequence motifs are associated with specific V and J gene usages, as many of these motifs are germline-encoded. "Germline-encoded" indicates specific residues and properties of an antibody are encoded by the gene, without the introduction of somatic mutations (Kenneth, 2008).

Therefore, identifying gene usages can correlate different functional phenotypes of antibodies to various pathogens. For example, *IGHV1-69* have been described in a variety of viral broadly neutralizing antibodies (Chen et al., 2019) including but not limited to HIV (Huang et al., 2004), Hepatitis C (Bailey et al., 2017a), influenza (Joyce et al., 2016), and Ebola (Murin et al., 2021). *IGHV1-69* encodes two hydrophobic residues at the tip of the CDRH2 loop, which provides the structural basis for epitope recognition. Many viral envelope glycoproteins have conserved hydrophobic regions that contribute to fusion machinery during infection. The hydrophobic loop encoded by *IGHV1-69* targets these conserved regions, setting the basis for broadly neutralizing antibodies. Another example is *IGHV3-53/3-66* antibodies to SARS-CoV-2. Residues encoded by *IGHV3-53/3-66* on the CDRH1 and CDRH2 form an extensive hydrogen bond network with the receptor-binding domain (RBD) of the SARS-CoV-2 spike protein, leading to virus neutralization (Yuan et al., 2020a). As many co-structures of antibodies have been

identified to various antigens, more essential sequence motifs are being uncovered, and analysis of antibody sequencing may allow for predictions of antibody-antigen interactions.

The CDR3 and somatic mutations

CDRs within an antibody are responsible for direct contact with antigens, and the CDR3 region contains the highest sequence diversity. The expansive diversity of the CDRH3 contributes to its ability to adapt and recognize multiple antigens, therefore commonly ranking it as a significant contributor to antigen recognition. The length of CDR3s may contribute to different antigen specificities and functional properties. As an example, one of the common characteristics of broadly neutralizing HIV antibodies is a long heavy chain CDR3, allowing for the antibody to penetrate the glycan shield of the HIV Env trimer or reach less accessible epitopes such as gp41 (Pejchal et al., 2011; Yu and Guan, 2014).

Integrating CDR3 length usage is, although imperfect, an incredibly useful method in the identification of clonal families. As previously mentioned, the process of VDJ recombination assembles naïve B cells in the bone marrow. As different V, D, and J genes are being stitched together; nucleotides are being added and subtracted at junctions between gene segments, thereby altering the length of the antibody sequence. On the heavy chain, the D gene segment undergoes the recombination process twice, thereby altering the length of the sequence in this region multiple times. This region encodes for the CDR3 sequence, therefore contributing to the high diversity of CDR3s within the antibody repertoire. Once a productive antibody is formed, the naïve B cell migrates to the lymph nodes for clonal expansion and somatic hypermutation. During the process of somatic hypermutation, point mutations are introduced, therefore, in most cases, not affecting the length of the CDR3 sequence. There are exceptions in which the

antibody sequence length varies during SHM (Kenneth, 2008). But a simple guideline that may be utilized is: Antibodies evolved from the same naïve B cell or antibodies from the same clonal family should retain the same CDR3 length as mutations introduced during SHM should not alter the CDR3 length of the antibody (**Figure I-12**). Therefore, variation in sequence length is retained at the level of VDJ recombination. This definition becomes incredibly useful when clustering large datasets for clonal family identification as pre-binning sequences by CDR3 length dramatically decreases computation time and power.

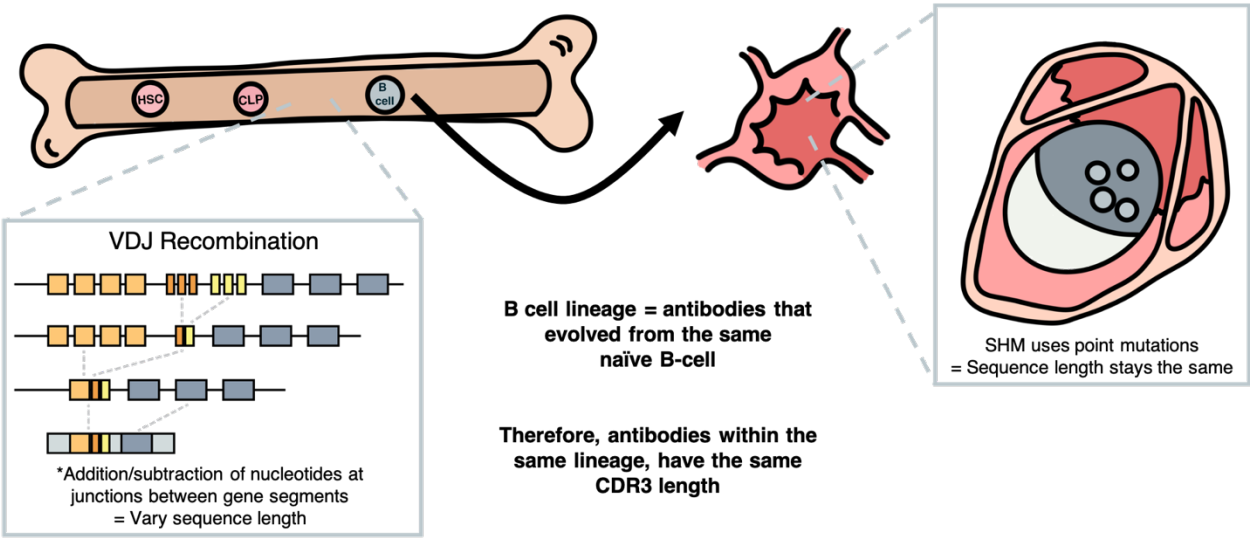


Figure I-12: Relationship between CDR3 lengths and clonal families

A schematic illustrating the relationship between CDR3 lengths and clonal families. The left side of the figure illustrates VDJ recombination, after which makes a naïve B-cell. As sequence length typically varies during VDJ recombination, and therefore prior to the generation of a naïve B-cell, B-cells within the same clonal family should encode the same CDR3 length.

Another common characteristic of broadly neutralizing HIV antibodies is that they are highly somatically mutated (Klein et al., 2013; Wiehe et al., 2018). This is likely due to the longitudinal nature of the disease, therefore prolonging antibody maturation by undergoing

multiple rounds of somatic hypermutation. Within the germinal center, B cells compete for an array of signals delivered in an affinity-dependent manner. Therefore B cells with higher affinity BCRs can out-compete lower affinity B cells, leading to a rapid expansion of higher affinity SHM derived variants of these high-affinity B cells and generating antibodies with significant accumulations of somatic mutations (Mesin et al., 2016) (**Figure I-13**). Therefore, identification of the amount of somatic mutations antibodies can also be insightful for predicting an antibody's binding affinity and, in turn functional characteristics.

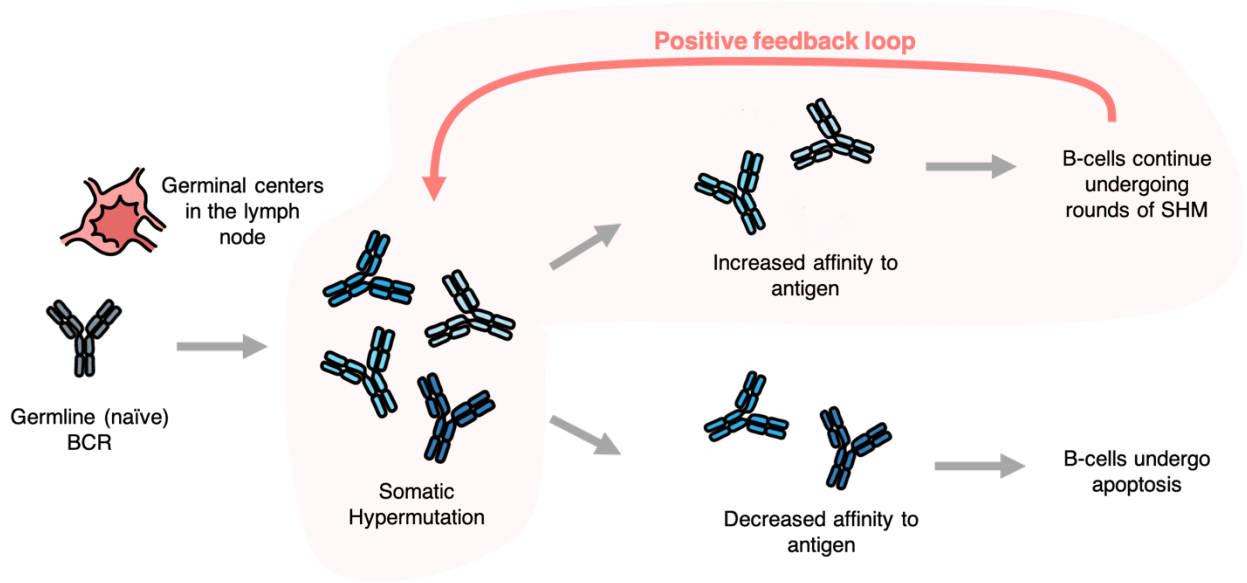


Figure I-13: Somatic hypermutation

Simplified schematic of somatic hypermutation. Process starts as the naïve B-cell enters the germinal centers and is activated for clonal expansion and somatic hypermutation. Point mutations are introduced and mutations that lead to an increased affinity to selected antigen are led to subsequent rounds of somatic hypermutation. B-cells that acquire mutations leading to the decrease in antigen affinity undergo apoptosis.

Data processing of antibody sequences

IgBlast and IMGT/HighV-Quest are two of the most popular programs for processing immune repertoire sequencing data (Brochet et al., 2008; Ye et al., 2013). Both tools use

template-based nucleotide alignments between a database of germline genes and query sequence to infer the most plausible germline genes used for an antibody and identify an antibody's FR and CDR regions. PyIR is an IgBlast wrapper using Python3, assembled to process large sets of immunoglobulin sequence data (Soto et al., 2020). Taking a FASTA formatted file as an input, PyIR runs all sequences through IgBlast and returns a zipped JSON file. This output file contains baseline sequence information such as inferred V gene, D gene, J gene, CDR sequences, CDR lengths, number of somatic mutations, and more. This output file is the basis for many sequence analyses described in the chapters of this dissertation.

Leveraging both bulk and paired sequencing

As previously described, bulk NGS sequencing allows for extensive in-depth sampling of the antibody repertoire. However, antibody sequences identified in bulk sequencing cannot be functionally validated due to lack of data on the natively paired heavy or light chain sequence (Georgiou et al., 2014). Utilizing paired single-cell sequencing retains the native pairing information, permitting downstream functional characterization of sequenced antibodies. Despite the ability to accommodate more sampling depth compared to the hybridoma, B-cell expansion, or single-cell sequencing methods, the depth of sequencing is still a magnitude less than bulk sequencing-based methods. Depth of sequencing is essential for understanding antibody evolution, clonal families, and sample diversity (Bashford-Rogers et al., 2013; Robinson, 2015).

In the optimal scenario, a representative antibody of each clonal family is captured in paired single-cell sequencing and characterized. Bulk sequencing can be leveraged by clustering similar antibody sequences to the one characterized for investigation on clonal families and antibody evolution, therefore merging the best of both worlds, my maintaining both

functional characterization and depth of sequencing (Zost et al., 2021a) (**Figure I-14**). Methods as such can also be utilized for indexing the human antibody repertoire to investigate not only the diversity of antibody sequences but also the diversity of antibody functional characteristics present in a given repertoire.

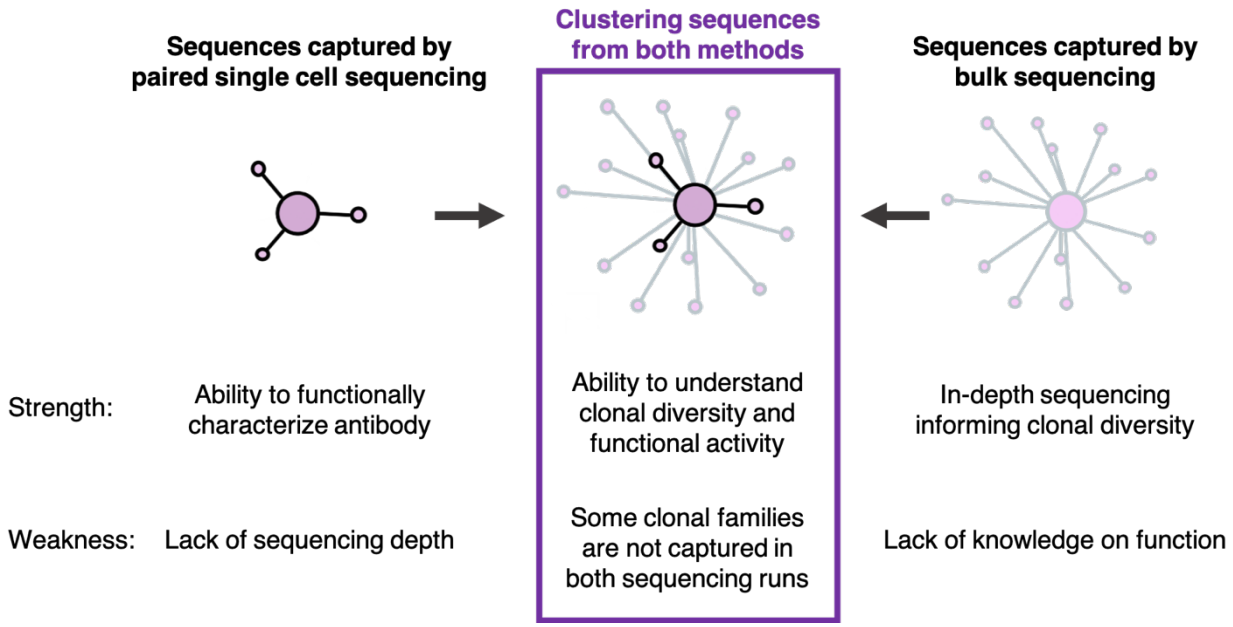


Figure I-14: Leveraging both bulk and paired sequencing

Schematic demonstrating the strengths and weaknesses of using paired single cell sequencing (left) as well as bulk sequencing (right). Leveraging the powers of both methods through clustering sequences from both datasets allows for the ability to understand clonal diversity and evolution along with functional characterization.

Sequence clustering

Sequence clustering is extremely powerful when undergoing repertoire-level studies. There are many types of clustering used in data science. Here, I will focus on hierarchical clustering with respect to clustering antibody sequences. There are two types of hierarchical clustering: top-down and bottom-up (Rodriguez et al., 2019). The top-down treats all sequences as the same cluster and continually splits each cluster until all sequences are in an individual

cluster. The bottom-up method treats each sequence as an individual until successfully merged. The hierarchy of clusters in both methods is represented as a dendrogram (**Figure I-15**). The bottom-up method, also known as agglomerative hierarchical clustering, is the most commonly used for analyzing antibody sequences.

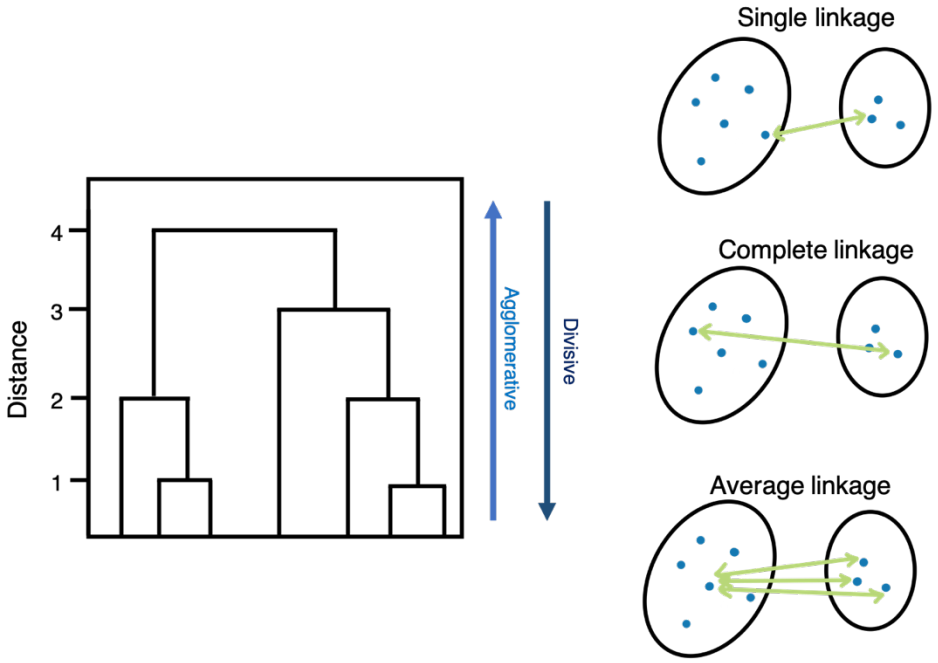


Figure I-15: Hierarchical clustering

Hierarchy of clusters are represented as a dendrogram as shown here in the left panel. Each node on the dendrogram is representative of a sequence cluster. Right panel is a representation of a example dendrogram. Divisive clustering reads the dendrogram from top down. And agglomerative clustering reads the dendrogram from bottom up. The right panel shows the differences between single-linkage, complete-linkage and average-linkage clustering. Each blue dot represents an antibody sequence and circles represent clusters. The green line is the distance calculated between the two clusters.

There are three types of hierarchical clustering: single, complete, and average linkage. In single-linkage clustering, the distance between clusters is defined as the shortest distance between two points in each cluster. In complete-linkage clustering, the distance between clusters is defined as the longest distance between two points in each cluster. And in average linkage clustering, the distance between two clusters is defined as the average distance between each point in one cluster to every point in the other cluster. Complete-linkage clustering is the most conservative of the three, as it entails that every sequence within a cluster is at most a defined similarity to one another. On the flip side, single-linkage clustering entails that every sequence within a cluster is at least a defined similarity to one another.

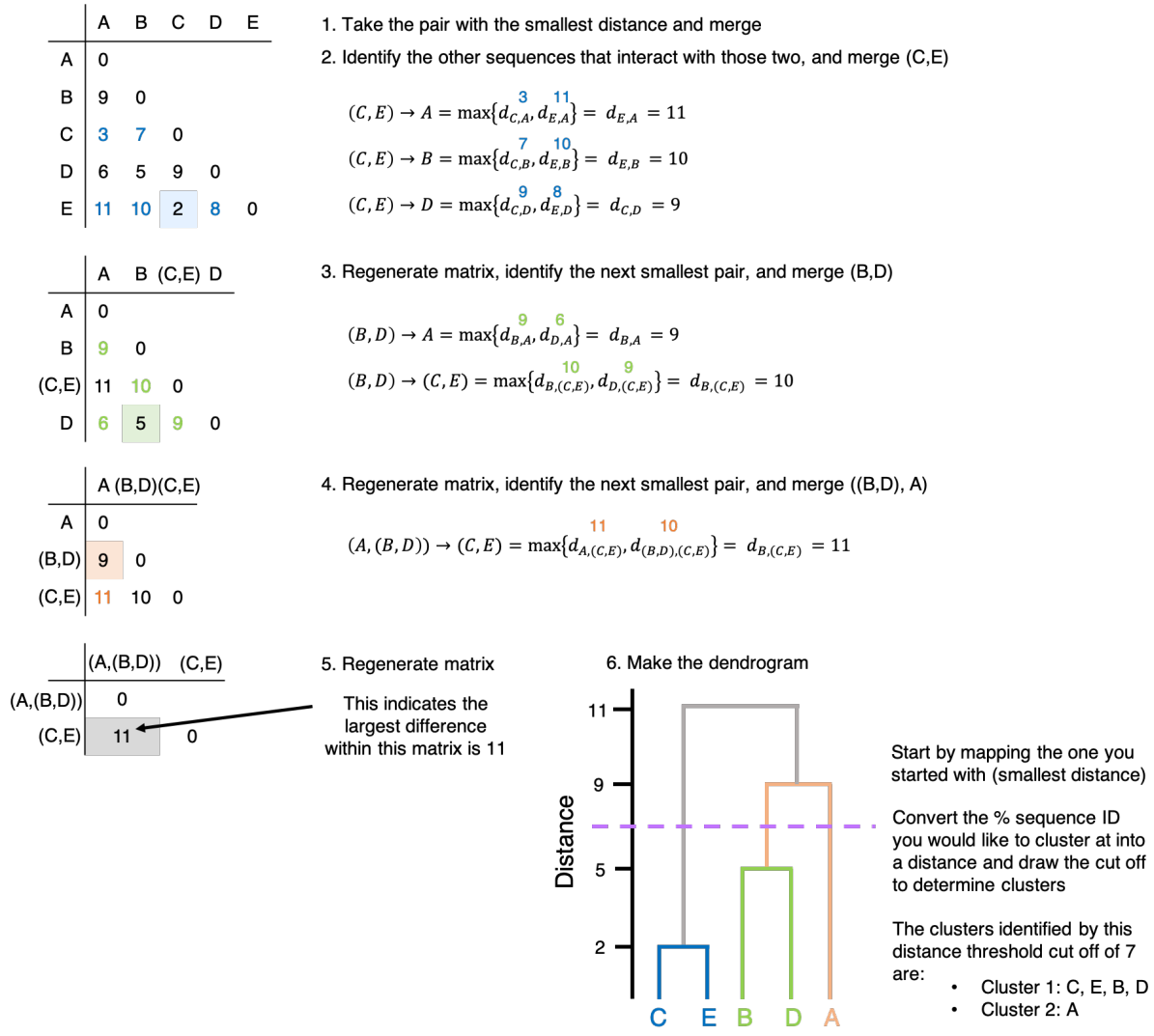


Figure I-16: Mathematics behind hierarchical clustering

Each letter (A, B, C, D, E) indicates an antibody sequence. Numbers within the matrix represent the distance (difference) between the two sequences. As each pair of sequences are selected to merge, the maximum distance between the selected sequences and other sequences in the matrix are identified. Colors of the distances correlate between the matrix and dendrogram.

Before any clustering is done, a proximity matrix between each data point (in this case sequence) is calculated using a distance function. The distance function measures how different each antibody sequence is from the other. Once a distance matrix is created, the pair of sequences with the smallest distance is identified. This pair is then merged by taking the largest

distance between each merge point. Once the maximum distance is identified for each merge point, a new matrix is generated. This process is continually done until there is only one distance number left. The last distance left in the matrix indicates the largest difference within the given matrix. If clustering using single-linkage or average linkage instead of complete linkage, the merging would occur by taking the minimum distance between each merge point (for single linkage) or the average distance (for average linkage). Once a distance threshold is established for the level of similarity clusters would share, a line is drawn across the dendrogram, thereby defining each cluster as the sequences linked together by the dendrogram under the threshold line (**Figure I-16**).

Hierarchical clustering is a very computationally expensive method of clustering. Calculations required for clustering exponentially increase as more sequences are added into the formula (Gupta et al., 2017). The distance between the new sequence and all existing sequences has to be calculated to fill out the matrix for every sequence added. Therefore, it is beneficial to pre-group sequences together before clustering, generating smaller groups to apply clustering to.

Clustering is often utilized to identify sequences similar to one another, with implications for identifying additional antibodies within a clonal family as well as similar antibodies within multiple datasets. Therefore, one method of pre-grouping sequences together prior to clustering is by binning sequences with the same V gene and J gene. When clustering to identify clonal families, binning by the CDR3 length is also helpful. As mentioned previously, antibodies with the same clonal family ideally have the same CDR3 lengths. All sequences grouped are now encoded by the same V and J genes. Within an antibody sequence, the V gene encodes for the majority of the sequence real estate. Therefore, pre-grouping antibody sequences neglects the need to cluster the antibody sequence on the full-length sequence. As the CDR3 sequence is known to maintain the most diversity, clustering on the CDR3 sequence alone is sufficient in

differentiating between sequence groups or clonal families. With the length of the CDR3 sequence being substantially shorter than the full-length antibody sequence, computation time and power are also conserved in comparison to clustering on the full antibody sequence (Chen et al., 2010).

Sensitivity vs specificity in clustering

There are two factors to assess when clustering: sensitivity and specificity. Specificity is defined as the fraction of unrelated sequences inferred as a part of a clonal group. Sensitivity is defined as the fraction of clonal relationships successfully identified (**Figure I-17**). Experiments comparing different clustering methods are often done on simulated sequence sets, where naïve antibody sequences simulate clonal lineages by introducing mutations that match experimentally observed mutations. Therefore, creating a sample set with defined clonal lineages to compare against is useful for such experiments (Gupta et al., 2017; Zhou and Kleinstein, 2019).

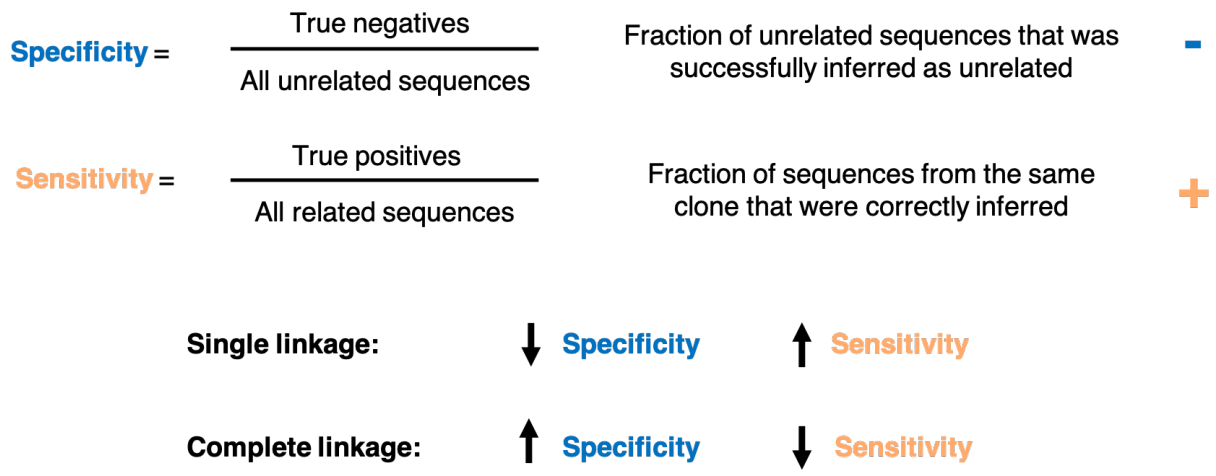


Figure I-17: Relationship between sensitivity and specificity

There is a relationship between sensitivity and specificity when identifying the type of clustering one wants done on a dataset. Specificity is important when inferring antigenic reactivity and sensitivity is important when understanding clonal lineages.

Sensitivity and specificity are important aspects to consider when applying different clustering methods. Complete-linkage clustering offers high specificity but at the cost of significant loss in sensitivity. This may be crucial in studies that are highly dependent on the accurate calling of sequences to the same clone, such as assigning antigen specificity to clonal relatives. Additionally, clustering on the CDR3 can be performed on either the nucleotide or amino acid sequence. Although clustering on nucleotide sequence is more computationally taxing, it has been shown to deliver higher sensitivity in identifying clusters than utilizing the amino acid sequence (Gupta et al., 2017).

Distance metrics and identity thresholds used in sequence clustering

Two main distance metrics are used when identifying the difference between antibody sequences: Hamming distance and Levenshtein distance. Hamming distance is a metric for comparing two strings of equal length, measuring the number of different positions. Every position that differs between the two strings is calculated as a distance of one. Levenshtein distance accommodates for strings of varying lengths. It allows for the deletion and addition of letters within a string, counting the number of steps it takes to change the first string into the second, therefore identifying the differences between two strings (**Figure I-18**). Between the two distance metrics, Hamming distance is the computationally cheaper one of the two.

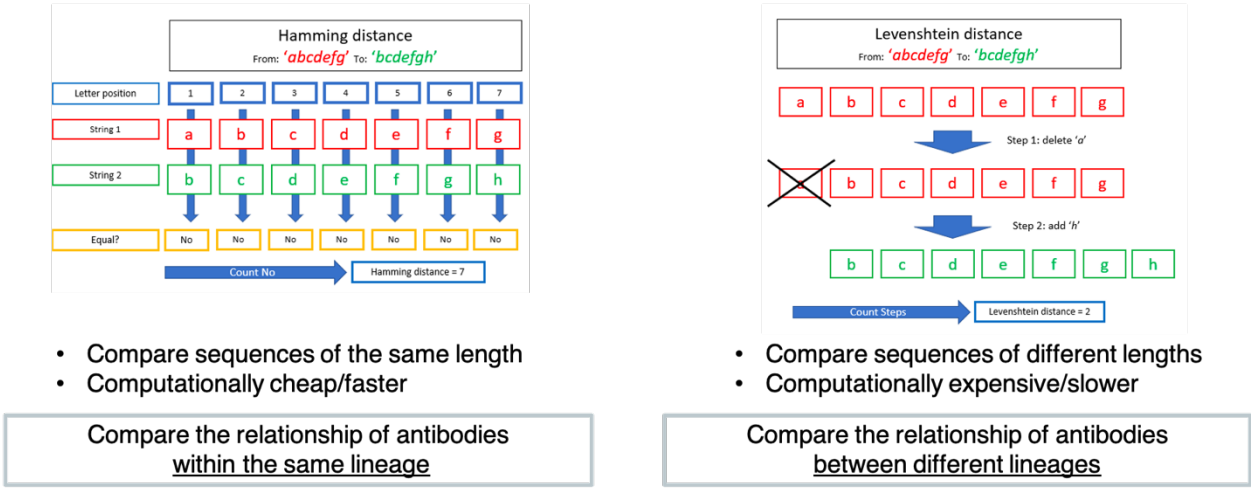


Figure I-18: Hamming vs levenshtein distances

Schematic illustrating the difference between hamming (left) vs levenshtein (right) distances. Hamming distance is calculated between two strings of the same length. Levenshtein distance can accommodate for strings with different lengths by identifying the shortest path to align the two strings.

As previously mentioned, antibodies belonging to the same clonal family ideally have the same CDR3 sequence. Therefore if sequences were pre-grouped by V gene, J gene, and CDR3 length, the hamming distance could be used as the distance metric for clustering. However, if sequences of varying CDR3 lengths were being computed, Levenshtein distance

would have to be used instead to accommodate varying string lengths. These distances are the numbers used to create the distance matrix before clustering.

Both distance metrics are often normalized to 1, allowing a fair comparison between sequences with varying lengths. For example, a five-letter string with one letter difference cannot be compared to a ten-letter string with a one-letter difference. As for the first string, it would be a 20% identity difference in comparison to the second string with a 10% identity difference. Therefore, identity thresholds are often converted to a normalized distance. An identity threshold of 80% similarity, where all clones within the same cluster are at 80% alike, would allow for a 20% difference between sequences. Therefore the normalized distance threshold would be 0.2.

Many different thresholds are used depending on the type of clustering as well as the data set. Studies utilizing clustering to identify clonal families have varied from 80-95% identity thresholds on the CDR3 (**Table I-1**). Additional methods of identifying clonal families via clustering include clustering on junctions. When considering thresholds, it is important to consider where the antibody sequences originated from. For example, B cells that have been previously sorted by antigen specificity can likely be clustered at a lower threshold in comparison to B cells that had not been sorted and just bulk sequenced from B-cell isolation. Reactivity to antigen adds a layer of “clustering” within itself, down selecting unrelated B cells by antigen specificity.

Additionally, the ability to cluster sequences on both heavy and light chains is a luxury only afforded by paired sequence data sets. Even though clustering based on the heavy chain alone can determine clonal relationships with >80% confidence, incorporating the light chain is beneficial in increasing sensitivity. Although light chain clustering alone does not support clonal clustering with high enough granularity to determine clonal families, antibodies within the same clonal family also tended to carry light chains with junction sequences that were at least 80%

similar to each other. Using of the light chain as a “subcluster” allows for the accurate identification of misclustered clones at the heavy chain level (Zhou and Kleinstein, 2019).

Threshold	Bulk sequencing studies	Paired single cell sequencing studies
90	Horns et al., *PNAS 2019* 90% on single-linkage clustering	
85	Rosenfeld et al., *Front. Immunol. 2018* 85% on CDRH3	
80		Goldstein et al., *Nat. Comm. 2019* 80% on CDRH3, and match on VL

Table I-1: Identity thresholds on CDR3 sequences for clustering
This table lists a couple examples of the different CDR3 sequence identities used for clustering to identify clonal families.

Antibody phylogenies

One of the benefits of having deep sequencing datasets is identifying multiple genetically similar antibodies representing a clonal family. An antibody sibling is a member of a clonal family. Identifying multiple siblings within a clonal family allows for the construction of phylogenies to detail the evolutionary pathway of antibodies within the clonal family (Yermanos et al., 2018). Having an accurate overview of the antibody maturation process in response to an antigen drives new vaccination strategies aimed at priming the B-cell precursors effectively expressing germline-encoded antibodies and walking the lineage of B cells through utilizing designed immunogens to initiate curated rounds of somatic hypermutation. Understanding the nature of the elicitation of a protective antibody response is a fundamental step to building tools for designing next-generation vaccines (Rappuoli et al., 2016).

Epitopes targeted by broadly neutralizing antibodies have been integrated into the design of recombinant proteins utilized to induce affinity maturation of B-cell lineages of these broadly neutralizing antibodies. This fine-tuned priming of the B-cell precursors at each step of

the maturation pathway requires prior knowledge of the evolutionary pathway of the targeted antibody. Once a clonal family is identified, there are multiple methods by which a phylogeny can be constructed. There are two types of tree structures: rooted and unrooted. A rooted tree assumes there is a common ancestor of all sequences. Therefore, the path from the root to the node (sequence) would define the evolutionary pathway. An unrooted tree is mostly used to specify relationships between nodes (in our case, antibody sequences) but is not used to specify an evolutionary path.

Prior to reconstructing a phylogenetic tree, a multiple sequence alignment is created using all the sequences within the clonal family. There are multiple programs available for multiple sequence alignments. Some of the most commonly used ones are ClustalW (Thompson et al., 1994), ClustalO (Sievers et al., 2011), MUSCLE (Edgar, 2004), or MAFFT (Katoh et al., 2002). Once a multiple sequence alignment is made, a phylogenetic tree may be constructed.

There are two types of construction methods for phylogenetic trees: Distance-based and character-based methods (Yermanos et al., 2018). Distance-based methods infer relationships based on sequence similarity and are computationally fast to build. However, many assumptions are made with distance-based methods, including assuming rates of evolution are the same among different lineages and may be sensitive to unbalanced evolutionary trees. Two common distance-based tree construction methods are UPGMA and neighbor-joining. Character-based/ sequence-based trees are often preferred over distance-based trees as they allow for different models of sequence evolution. Character-based tree construction methods include maximum parsimony and maximum likelihood. Maximum parsimony trees assume there should be no more evolutionary steps than necessary, therefore constructing a tree that minimizes the number of changes between ancestors and descendants. Maximum likelihood trees evaluate the likelihood of every substitution of every mutation in a tree based on

evolutionary models. Although the maximum likelihood model is computationally intensive, it often outperforms other methods. Through this dissertation, the maximum likelihood method is used in all tree reconstructions.

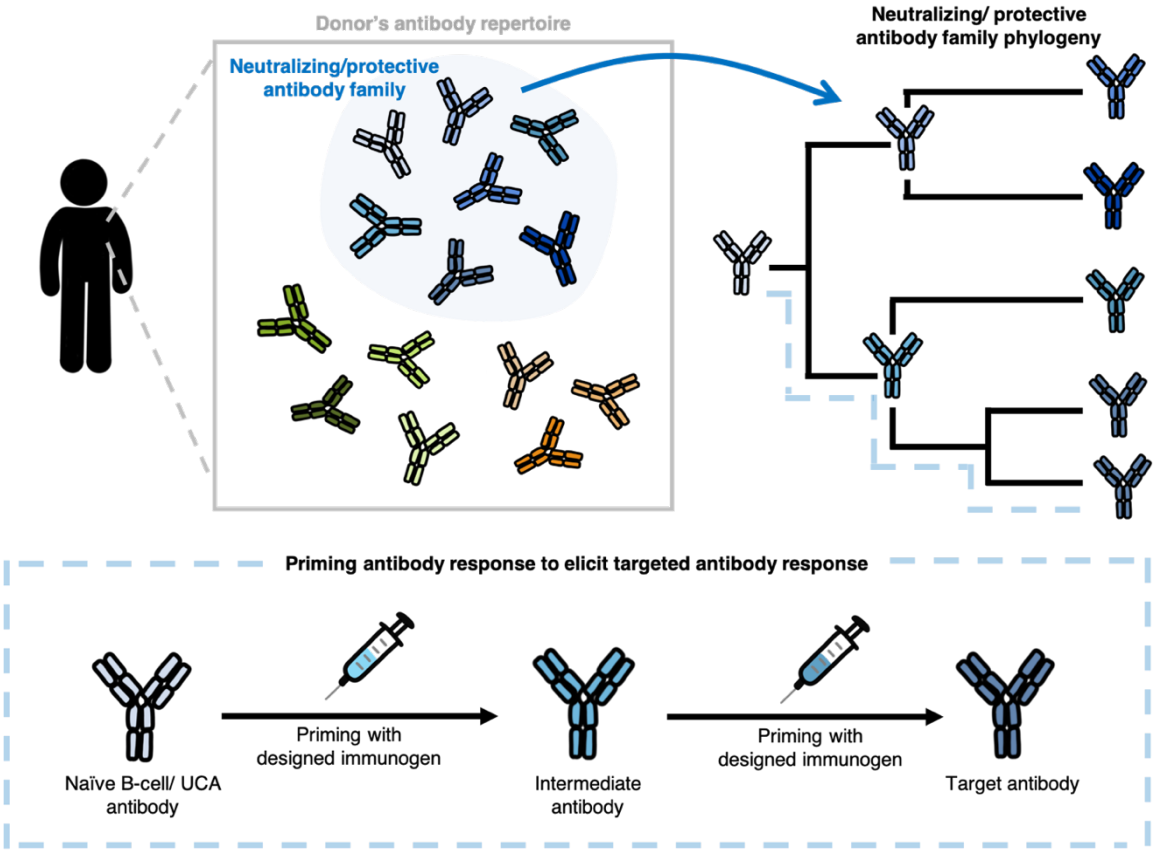


Figure I-19: Reverse vaccinology

Schematic representation of reverse vaccinology. The antibody repertoire to a specific target is specified and characterized to identify a target clonal family. This is likely a family of antibodies with favorable anti-viral properties. Once the clonal family's sequences are identified, a the phylogeny of the antibody lineage can be built.

Unmutated common ancestor and germline revertant antibodies

Germline-encoded reactivity refers to an antibody's functional characteristics that are likely encoded by the germline genes prior to the addition of somatic hypermutations. Therefore, the reactivity a naïve B cell would have as it exits the bone marrow. There are two methods to identify the sequence of these naïve B cells. The first is inferring an unmutated common ancestor, and the second is by reverting an antibody sequence back to the germline.

An unmutated common ancestor (UCA) is the inferred antibody sequence of the naïve B cell within a lineage. During phylogenetic tree reconstruction, intermediates at all nodes with the tree are inferred if not identified within the sequence set. The UCA is the node at the beginning of the tree. This sequence usually contains zero to few mutations when aligned to the inferred V, D, and J genes. However, for accurate phylogenetic tree reconstruction, and by proxy an accurate inferred UCA, there must be enough sequences within a clonal family to build the tree in the first place (Bonsignori et al., 2018).

If there are not enough sequences within a clonal family to reconstruct a tree, the second method of achieving the sequence of the theoretical naïve B cell is through reverting the antibody sequence back to germline. This is often done by aligning the antibody sequence to the inferred V, D, and J genes and reverting every somatically mutated residue back to the germline sequence(Dai et al., 2016).

Both the UCA or the germline revertant (GR) antibody can then be functionally characterized, such as binding or neutralization, to identify germline-encoded reactivity to the target antigen, revealing if there are germline precursors of the targeted antibodies. Understanding germline-encoded reactivity is important in reverse vaccinology as if the naïve B cell already has some pathogen-specific activity. It is likely easier to stimulate antibodies originating from that lineage, increasing affinity to said antigen. Therefore targeting germline precursors of potent and neutralizing antibodies can be used in reverse vaccinology (**Figure I-19**).

PART III: Public clonotypes

What are public clonotypes?

As previously described, B cell development and maturation process yields a large diversity of B cells within every individual. Therefore, this large diversity as well as the different B cells that are activated in an individual due to the different types of infections they acquire over the years of life, it was believed that B-cell repertoires remained private to each individual as it was inconceivable that with the sheer number of variations of possible B cells, that individuals would make the same antibodies. However, in recent years due to the decrease in the price of sequencing leading to the increase in BCR sequencing being done, it has been shown that unrelated individuals can share the same or similar antibody responses. These shared, or convergent antibody responses are called public clonotypes.

Public clonotypes have been identified in human antibody repertoires in response to a variety of viral pathogens, including influenza virus (Joyce et al., 2016; Pappas et al., 2014; Zost et al., 2021b), respiratory syncytial virus (Mukhamedova et al., 2021), hepatitis C virus (Bailey et al., 2017a), HIV (Setliff et al., 2018; Zhou et al., 2015), SARS-CoV-2 (Chen et al., 2021a; Sakharkar et al., 2021; Schmitz et al., 2021; Tan et al., 2021), as well as healthy individuals (Briney et al., 2019; Soto et al., 2019) revealing selection of genetically similar B cell receptors in memory cells in the circulation of diverse immune individuals. Public clonotypes are of immense interest as understanding viral epitopes that commonly induce antibodies in diverse individuals has implications for predicting common responses to vaccines and viral proteins in large populations. Additionally, if diverse individuals make the same antibody response to an antigen, there could be a constant and collective selective pressure on that epitope, resulting in a high

potential for escape variants at that site which could lead to the prediction of escape variants to several viruses.

It is likely that before the generation of large sets of sequencing data, public clonotypes were not being identified as frequently due to the shallow pool of B-cell sequences available to mine, hinting that identification of public clonotypes may be a numbers game. Therefore, the more sequences available to mine, the more public clonotypes one can identify. Due to the availability of large sets of antibody sequences, many of the first public clonotypes were identified using bulk NGS sequencing. However, due to lack of paired sequence information, the functional characteristics of identified public clonotypes remained unknown unless similar antibodies had previously been found and characterized. The inability to functionally validate the characteristics of these public clonotypes left the following question unanswered: what are the implications of public clonotypes on humoral and population immunity?

Mining datasets for public clonotypes requires very large numbers of sequences and validation of identified public clonotypes by testing recombinant immunoglobulins for specificity, and antiviral function require antibody gene datasets containing authentically-paired heavy and light chain genes from single B cells. With the generation of paired-chain antibody sequence sets exponentially increasing in recent years, and the spike in paired antibody sequences generated to SARS-CoV-2 due to the current pandemic (Kreer et al., 2020; Robbiani et al., 2020; Seydoux et al., 2020; Zost et al., 2020b), sets a unique position to mine and characterize public clonotypes on an unprecedented scale.

Definition of public clonotypes

Public clonotypes are understood as similar antibody sequences shared between unrelated individuals. However, there is no strict definition as there are many definitions for what

constitutes a public clonotype. In some literature, it is defined as the heavy chain sequence only (Nielsen et al., 2020; Setliff et al., 2018), or the light chain sequence only (Cohen-Dvashi et al., 2020; Zost et al., 2021b), in others it may be a sequence motif in combination with a gene usage (Dong et al., 2021), or it could be a definition that includes both the heavy and light chain sequences (Chen et al., 2021a). Additionally, there are multiple clustering methods and thresholds of similarity in which sequences are clustered to identify public clonotypes.

Many of the first public clonotypes were primarily identified using heavy chain bulk sequencing. Interesting antibodies were identified using traditional discovery approaches, and the sequence of such antibodies would be the query to look for similar antibodies sequences present in other individuals. This scenario implies that antibodies sharing a similar heavy chain sequence would yield a similar function. However, these public clonotypes were not often tested as the light chain sequence of mined antibodies often remained unknown, therefore limiting the capability to compare the functional characteristics of such public clonotypes. Often these public clonotypes would be mined by matching on the same V gene, J gene, CDRH3 length, and clustering of a threshold on the CDR3 sequence. Thresholds are usually maintained relatively high, around 80-90% identity as antigen specificity is implied in such analysis but not often tested due to lack of light chain pairing; therefore, a more conservative clustering approach is taken.

However, a disadvantage to clustering purely on the heavy chain sequences for public clonotypes is that light chain-driven public clonotypes would be left unidentified. Several cases are shown in influenza (Zost et al., 2021b) and Ebola (Cohen-Dvashi et al., 2020) have demonstrated the light chain sequence can drive public clonotypes. In these cases, public clonotypes are antibodies with the same phenotypic function and share a similar light chain sequence despite a different heavy chain. These antibodies are usually identified as they share similar intriguing phenotypic properties such as mode of neutralization or broad reactivity, of which sequences of all antibodies with the same properties are compared. These antibodies encoded

by a light chain variable gene segment usually incorporate shared somatic mutations and have motifs in the heavy chain despite diverse heavy chain variable gene usage.

Another approach to identifying public clonotypes is clustering sets of sequences in an unbiased manner using both the heavy and light chains. Commonly these antibodies are identified through binning on the same V_H, J_H, CDRH3 length, and V_L, J_L, and clustered to a threshold of percent identity on the CDRH3. Studies demonstrating clonal relationships within repertoires reported that clustering on the heavy chain is sufficient. However, the incorporation of subclustering on the light chain could be crucial in contexts when evaluating antigen binding and specificity (Zhou and Kleinstein, 2019). Large sets of sequences allow for systematic identification of public clonotypes. The threshold of sequence identity on the CDRH3 can be adjusted to a lower percent identity as the incorporation of sub clustering on the light chain gene usages will “catch” any of the clones that may not belong in the public clonotype. Thresholds may also be adjusted based on the input source of B cells of which sequences are being analyzed (ex: if they have been previously antigen sorted). Additionally, through the use of high-throughput antibody expression, purification, and characterization using microscale methods, all public clonotypes identified can be tested to ensure their functional phenotype bins are similar in nature (**Figure I-20**).

Depending on the use case of public clonotypes identified, several clustering methods and thresholds can be used. Groups have investigated that clustering using the amino acid sequence is more efficient, but nucleotide-based methods have higher sensitivity in defining groups of clonotypes. Similarly, single-linkage clustering is more efficient and sensitive but has low specificity and can be used for most studies. However, if attempting to link antigen-specific sequences or identification of clonal relatives, these studies may benefit from the high confidence in specificity in each clonal connection provided by complete-linkage clustering (Gupta et al., 2017).

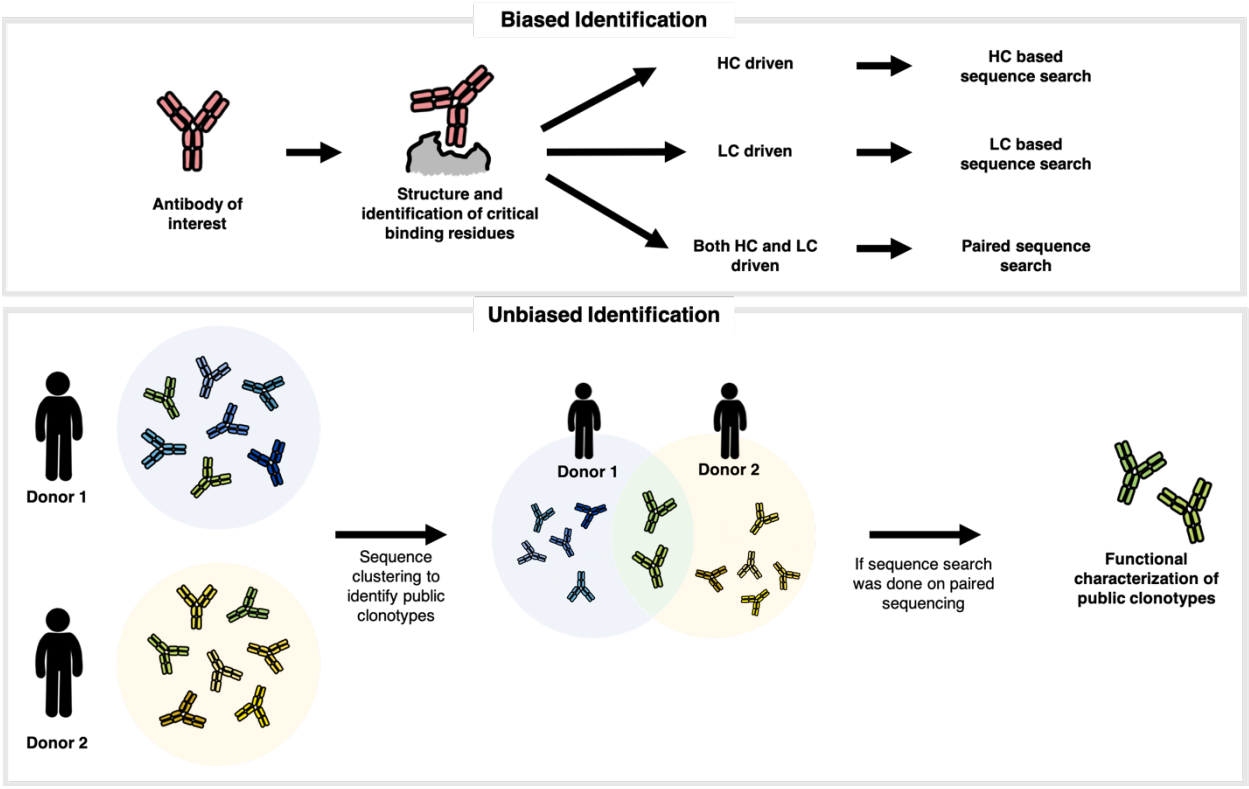


Figure I-20: Identification of public clonotypes

There are two categories to how to identify public clonotypes: biased and unbiased. With a biased method, the antibody is already functionally characterized. Based on where the critical residues map, public clonotypes can be identified using the motifs identified as a basis for the sequence search. With an unbiased method, pools of antibody sequences are mined for identification public clonotypes. The identified public clonotypes are then functionally characterized.

Greater implications of public clonotypes

One of the key challenges to reverse vaccinology is that naïve B cells that recognize the correct epitope might be rare and buried deep in the antibody repertoire (Havenar-Daughton et al., 2018). However, mining for and understanding the properties of public clonotypes reveals immunodominant B cell responses within immune populations, which may be of benefit for rational design of vaccines that may exhibit immunogenicity in a broader segment of the population. Knowing the public clonotype profile following natural infection also can enhance experimental

vaccine testing since the immunogenicity for desirable public antibodies recognizing cross-reactive sites of vulnerability for potent neutralization can be recognized at the cDNA sequence level.

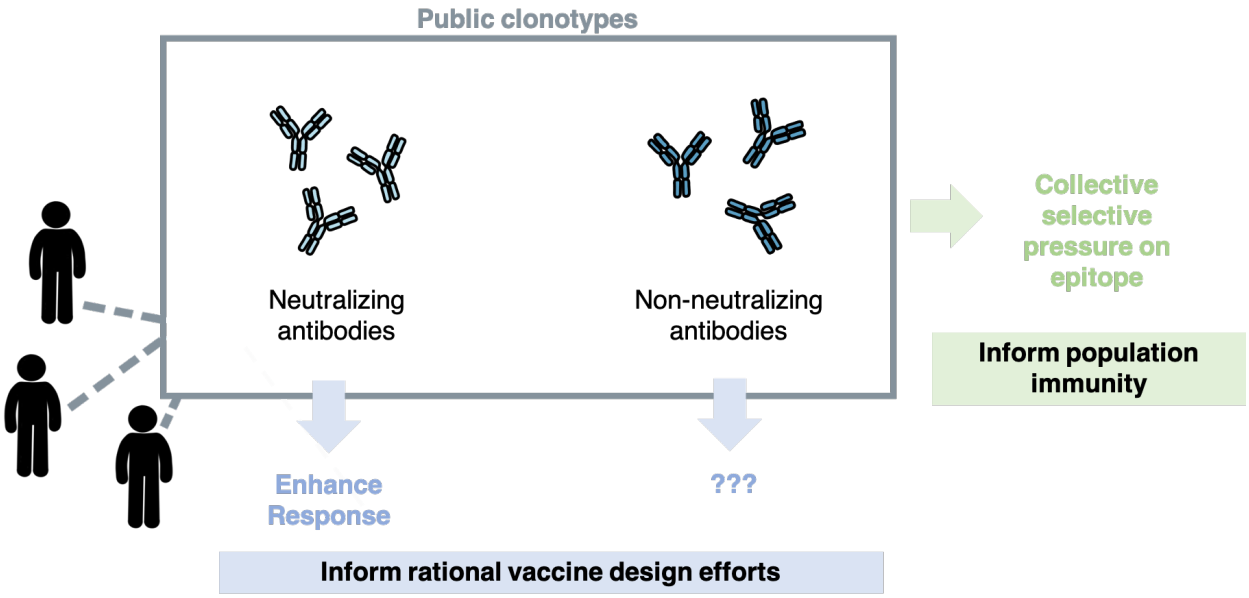


Figure I-21: Greater implications for public clonotypes

Implications for understanding the characteristics of public clonotypes include informing rational vaccine design (blue) as well as population immunity (green).

On the flip side, the broad induction of public antibody clonotypes recognizing the protective antigen of an RNA virus can lead to constant and collective pressure on certain epitopes to viruses leading to the rapid selection of escape mutant variant viruses, as evident in the evolution of SARS-CoV-2. Therefore, identifying public clonotypes to different pathogens can be utilized as a method of identifying potential escape variants which will impact public health (**Figure I-21**). Furthermore, public clonotypes have demonstrated success in predicting antigen specificity using machine learning. Mining and characterizing a larger database of public clonotypes may help further understand different sequence motifs encoded in antibodies and their relationship to structure and function.

Part IV: Immunoglobulin light chain amyloidosis

Introduction to light chain amyloidosis

Amyloidosis is a collective term for the extracellular deposition of abnormal proteins either in a single organ (localized) or throughout the body (systemic). These amyloid proteins come from abnormal folding caused by a proteolytic event in the amino acid sequence making the protein thermodynamically unstable and therefore prone to self aggregation. These unstable proteins form protofilaments, which aggregate becoming amyloid fibrils, which makes up the amyloid buildup (Blancas-Mejia and Ramirez-Alvarado, 2013). There are different subtypes of amyloidosis characterized according the origin of the deposited proteins.

Light chain (AL) amyloidosis is the most common form of systemic amyloidosis, estimated to affect 5-12 million people a year. AL amyloidosis is caused by a B cell, usually a plasma cell, which secretes unstable immunoglobulin light chains. The clonal plasma cell produce the amyloid in the form of immunoglobulin light chains. These amyloidogenic free light chains then accumulate and deposit aggregate in peripheral organs causing organ dysfunction, and subsequently organ failure. Common organs involved in disease prognosis include the heart and kidneys, with cardiac involvement being the main driver in disease prognosis and mortality (Palladini et al., 2015). Patients with advanced disease, especially when there is cardiac involvement, are at high risk of death within a few months. The amyloidogenic plasma cell clone is characteristically small and indolent. However, when the clone accounts for 10-15% of the bone marrow cells, the patient prognosis is poor, similar to that of patients with multiple myeloma (Desport et al., 2012). There is a strong correlation for patients diagnosed with AL amyloidosis and multiple myeloma with every 10 patients diagnosed with multiple myeloma, 1-2 patients are diagnosed with AL amyloidosis (Bahlis and Lazarus, 2006).

Diagnosis of AL amyloidosis

AL amyloidosis has a diverse clinical presentation profile, as symptoms are dependent on the organ involvement and tropism of the amyloid fibrils. Due to the wide variety of symptoms, AL amyloidosis patients are frequently led to delays in diagnosis such that organ dysfunction has already advanced by the time treatment is initiated. Diagnosis of AL amyloidosis requires (1) the demonstration of amyloid in the tissue and (2) demonstration of plasma cell dyscrasia. For proper diagnosis, biopsy of affected tissues are performed and stained with a Congo red stain. Tissue amyloid deposits will look apple green under polarized microscopy (Rysava, 2019). This requirement of doing an invasive procedure like a biopsy to obtain tissue is another factor towards the delay in diagnosis. Plasma cell dyscrasia is done after tissue diagnosis.

The long term objective for AL amyloidosis treatment regimens is to improve organ function and prolong survival. Therefore, elimination of the plasma cells secreting the amyloidogenic light chains are targeted with chemotherapy and immunotherapy. However, the current standard of care is to treat patients symptomatically with supportive therapy rather than targeting the problem, as there is currently no FDA approved drug for AL amyloidosis. Often times, depending on the prognosis of disease, patients are separated into transplant eligible and ineligible to replace the diseased organs (Elsayed et al., 2020). Therefore, early diagnosis is vital for the progression of AL amyloidosis.

As the free antibody light chain secreted by a population of plasma cells generally thought to be clonal (Gertz, 2016), the current best practices for determining patient hematologic disease status involve measuring the absolute quantity of free light chain proteins in serum (Comenzo et al., 2012; Kumar et al., 2012). Free light chain ratio is determined by measuring serum free light chains in patients and identifying the kappa-to-lambda light chain ratio. A complete hematologic remission is defined in part by normalization of the free light chain

ratio. However, many patients with complete hematologic responses do not experience an organ response, or they even experience organ progression (Comenzo et al., 2012; Manwani et al., 2019; Palladini and Merlini, 2019). Furthermore, some patients present with relatively low levels of serum free light chain protein and discordantly advanced organ involvement (Milani et al., 2017). One plausible mechanism for this finding is the persistence of small numbers of cells from plasma cell clones that continue to produce amyloidogenic light chain protein, which contributes to progressive organ dysfunction at low concentrations. As many as 60% of AL amyloidosis patients in complete hematologic response may have residual clonal amyloidogenic plasma cell populations, as measured by next generation flow cytometric analysis of circulating white blood cells (Kastritis et al., 2018; Sidana et al., 2020). More sensitive and specific methods for determining hematologic disease status are needed.

Part V: Influenza Virus

Influenza virus

Influenza is a respiratory disease caused by four types: A, B, C, and D. Influenza A infects a wide variety of species, including humans, birds, and pigs, contributing to its zoonotic potential to recombine multiple strains. Influenza B is primarily found in humans. Influenza C is found in humans but has also been shown to cross over to pigs and dogs. And Influenza D is mainly found in cattle (CDC, 2022b).

Influenza A (IAV) and B (IBV) are the two types associated with the seasonal epidemics every year, with Influenza A infections accounting for 75% of cases and Influenza B infection accounting for 25% of cases annually (Taubenberger and Morens, 2008). Although flu is often characterized by annual seasonal epidemics, there have also been sporadic and unpredictable global pandemic outbreaks. A pandemic flu outbreak occurs every 10-50 years and is

characterized by the introduction of new strains that are antigenically distinct from previously circulating strains, often rising from recombination events with zoonotic strains, resulting in an antigenic shift. This antigenic shift causes pandemics due to the lack of pre-existing immunity within humans to these antigenically shifted strains, contributing to the severity of infection (Kim et al., 2018).

Symptoms associated with influenza virus infection can vary from mild respiratory diseases such as sore throat, cough, runny nose, and muscle fatigue to severe, with some cases leading to pneumonia and subsequently death. It is estimated that annually, influenza epidemics result in ~1 billion infections, of which 3-5 million develop severe illnesses, leading to 300,000 – 500,000 deaths (WHO).

Influenza A and B viruses

Influenza viruses are a member of the Orthomyxoviridae family. They are negative-sense, single-stranded RNA viruses with a segmented genome. Both Influenza A and B viruses encode 8 RNA segments which encode for the viral glycoproteins, RNA polymerase subunits, nucleoprotein, matrix protein, membrane protein, nonstructural protein, and nuclear export protein (Krammer et al., 2018).

Influenza has two viral glycoproteins that stud the surface of the virion (**Figure I-22**). Hemagglutinin (HA), which is responsible for viral entry, and neuraminidase (NA), which is responsible for viral release. HA attaches virions to cells by binding to terminal sialic acid residues, initiating the viral life cycle, and NA cleaves the terminal sialic acids from N-linked glycans, therefore releasing the virions completing the viral life cycle. The HA trimer is divided into the head domain and the stalk domain. The HA head domain holds the receptor-binding site and has high plasticity, constantly undergoing mutational change, and the stalk domain is the more conserved region, housing the machinery for membrane fusion. The NA tetramer is also

divided into a head and stalk. The head of the NA contains the NA active site, responsible for the sialic acid cleavage event (Gamblin and Skehel, 2010).

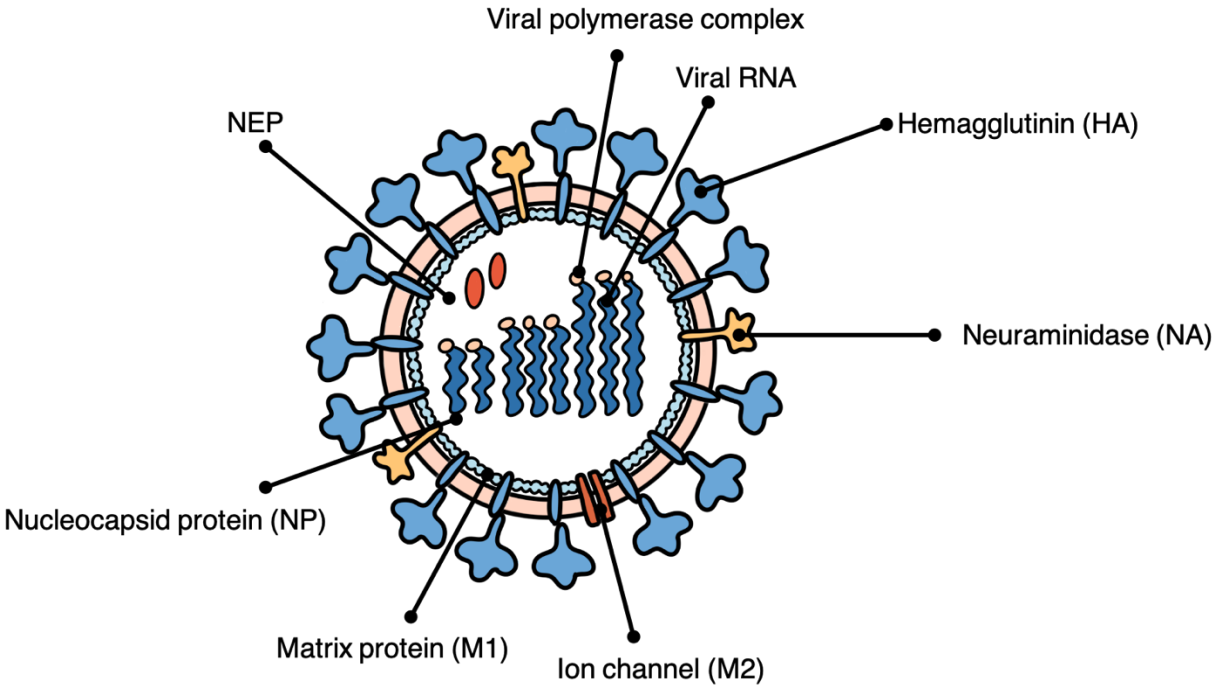


Figure I-22: Influenza virion

Cartoon with a representative virion of Influenza A. Each virion comprises 8 single stranded RNA segments encoding for the different parts of the virion. The two main surface proteins are Hemagglutinin (blue) and Neuraminidase (orange). The ion channel, M2 (dark orange) is also on the surface.

As the HA and NA are the most antigenically variable between strains, they are utilized to classify subtypes. Categorized by the HA, influenza A is split into groups 1 and 2, and within the two groups, they are further divided into subtypes, classified by similarity in stalk structure. There are a total of 16 antigenically different HA and 9 antigenically different NA subtypes (18 HA and 11 NA if including bat influenza-A viruses) (Tong et al., 2012) (**Figure I-23**). Within influenza B, there are two lineages: Victoria and Yamagata. There are four subtypes of

influenza attributed to the seasonal influenza epidemics – two from influenza A (H1N1 and H3N2) and two from influenza B (Victoria and Yamagata) (Rota et al., 1990).

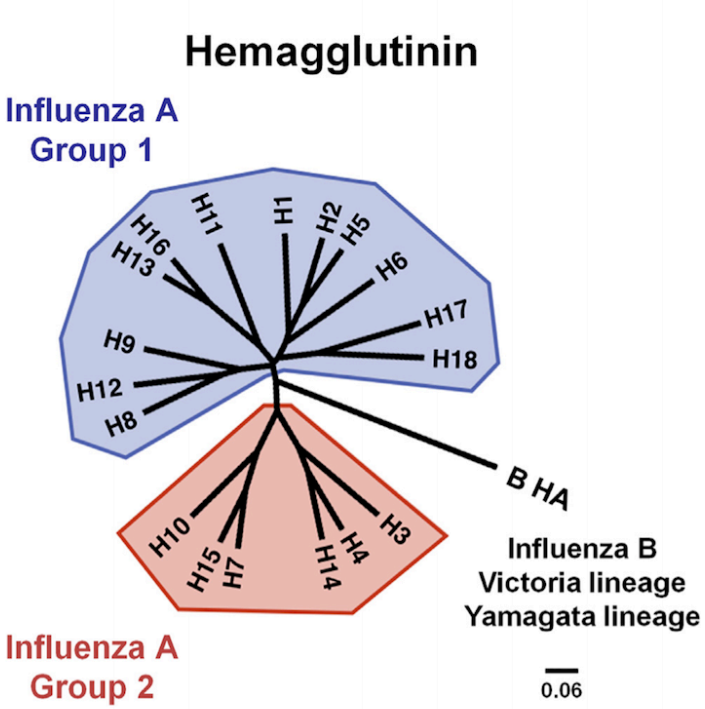


Figure I-23: Influenza A subtypes

Influenza A subtype grouping. Group 1 is shown in blue and group 2 is shown in red. Scale bar indicates 6% change at the amino acid level. Figure reproduced with permission from Coughlan et al., Cell Host and Microbe, 2018.

Antigenic drift and shift

Due to the zoonotic characteristics of influenza A viruses, viruses from different animal reservoirs may cross-species barriers through recombination events. The segmented nature of the virus enables genome reassortment upon dual-infection, where the genomic RNA segments of two subtypes recombine to form new subtypes. This process results in the inception of a new viral subtype, therefore, termed antigenic shift and the causal source of pandemics. An example case of this was the 1918 pandemic where the majority of the population was naïve to H1 HA and N1 NA. A similar case was seen in 1957, where H2N2 emerged, and again in 1968 when H3N2 emerged. At each point, there was minimal population immunity to not only the HA and NA subtypes alone but also combined (Krammer, 2019). When population immunity is naïve to emerging strains, these pandemic viruses spread quickly through the human population (**Figure**

I-24). However, as the majority of the population becomes infected, population immunity is developed targeting the emerged virus, and these viruses taper becoming seasonal strains, causing milder symptoms and lower fatalities.

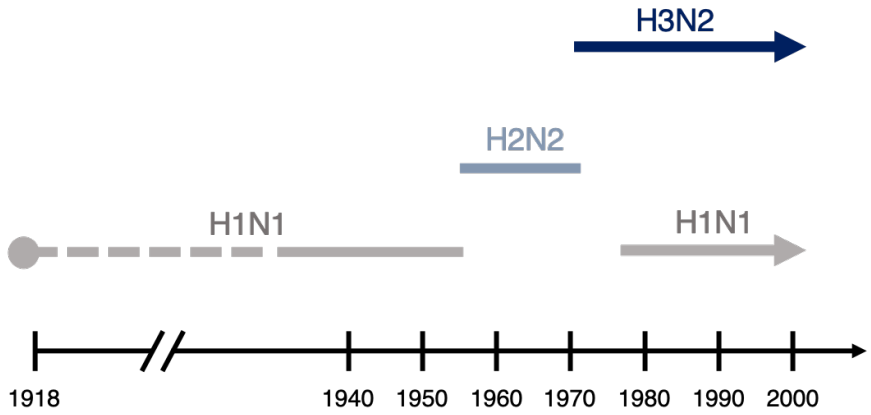


Figure I-24: Pandemic influenza A timeline

Timeline of pandemic influenza A outbreaks caused by antigenic shift.

As the population mounts immunity to the emerged virus, there is an evolutionary pressure exerted by the population's antibody response. To increase viral fitness, the virus is forced to evolve, evading preexisting population immunity by the introduction of mutations, termed antigenic drift. The constant evolutionary interplay between influenza and the humoral immune system alludes to the constant mutable behavior of the virus (Bahadoran et al., 2016). Therefore, the reason why our influenza vaccines have to be updated on an annual basis. Two times a year, circulating strains are compiled to make recommendations on the composition for the annual vaccine. One H1N1 strain, an H3N2 strain, and two strains of influenza B are selected for the quadrivalent vaccines distributed globally to maintain population immunity to circulating viruses.

Humoral responses to influenza

Although supported by surveillance data, predictions for vaccine strains are not perfect. Therefore, the antibody response to influenza virus infection and vaccination is studied to help better design vaccines and therapeutics. All proteins from the influenza virion can be targeted by the antibody response, but the two proteins that are most commonly studied are antibodies to HA and NA. Neutralizing and protective antibodies specific to both HA and NA have proven importance for immune protection. However, due to the antigenic diversity of both proteins, efforts have been zoomed in on understanding cross-reactive antibodies and their epitopes (Zost et al., 2019).

Antibodies to different viral proteins have different antiviral properties. Antibodies to the head of the HA typically block receptor binding through either steric hindrance or direct binding in the receptor binding site, therefore preventing the binding of sialic acid, constituting neutralization. Antibodies targeting the stalk of the HA can lock it in a pre-fusion formation, thereby preventing the fusion of the viral particle with the host cell (Tan et al., 2014; Wu and Wilson, 2020). Stalk-directed antibodies can also block the protease cleavage site, preventing the virus from producing mature HA on the viral surface. Some HA-directed antibodies can also prevent viral egress (Krammer and Palese, 2013). Antibodies targeting the NA can inhibit enzymatic activity via either sterically blocking or directly binding to the active site, therefore preventing viral egress (Stadlbauer et al., 2019). Antibodies to HA, NA, as well as M2 have also been shown to have indirect antiviral effects through Fc mediated effector functions (Vanderven and Kent, 2020).

The receptor for HA is sialic acid (Taubenberger and Kash, 2010), which is present on multiple cell types, therefore causing non-specific binding. Blocking the receptor binding site to overcome the “stickiness” can help overcome this issue as HA would not be reactive to sialic acid anymore. However, blocking the receptor binding site would prevent the capture of B cells

reactive to the receptor binding site, a large fraction of the neutralizing antibody response. As HA is one of the primary targets for antibody responses to influenza, it removes the ability to antigenically sort B cells for sequencing and subsequently antibody discovery. One of the methods used to circumvent this is by isolating and sequencing the plasmablast response. Doing so infers that the plasmablast response is reacting to recent antigen exposure; therefore, confirmation of infection or vaccination is required for such studies. Sequencing of the plasmablast response to infection captures B cells targeting not only HA and NA but also different viral proteins as well (Jackson et al., 2014).

Part VI: Ebola Virus

Introduction to the Ebola virus

Ebola virus is in the Mononegavirales order, Filoviridae family, under the Ebolavirus genus. It is an enveloped single-stranded, non-segmented, negative-sense RNA virus with genomes about 19kb long containing seven arranged genes: VP30, VP40, glycoprotein (GP), VP30, VL24, and L (**Figure I-25**). These genes create seven structural proteins and one non-structural protein, the soluble GP (sGP). The Ebola virion is filamentous in shape, with the GP studding the surface (Noda et al., 2002).

Phylogenetic relationships within the Filoviridae family are determined by the coding region of the glycoprotein gene (**Figure I-26**). Within the Ebolavirus genus, Zaire (EBOV), Bundibugyio (BDBV), and Sudan (SUDV) are the three strains that have caused outbreaks throughout history and are highly lethal in humans. Taï forest (TAFV) has caused a single case of severe but non-lethal disease. And Reston (RESTV) has caused disease in non-human primates but has seemed to be asymptomatic in humans. Marburg virus (MARV), a member of

the Filoviridae family but not within the Ebolavirus genus, has also caused outbreaks in humans(CDC, 2022a).

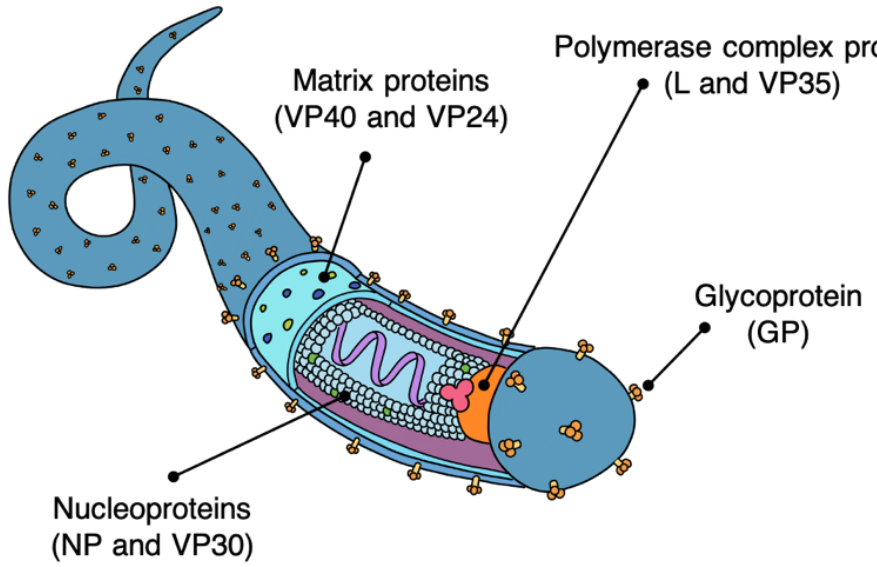


Figure I-25: Ebola virus virion

Cartoon rendition of the Ebola virus virion. The glycoprotein (orange) studs the surface of the virion.

Filoviruses cause severe disease in humans with devastatingly high mortality rates. Abrupt onset of flu-like symptoms such as fever, chills, and myalgia, follows an incubation period of 8-12 days post-exposure, usually from contact with infected individuals or animals. Subsequent symptoms include systemic, respiratory, vascular, gastrointestinal, and neurological manifestations. Hemorrhagic fever symptoms will often develop at the peak of illness. One of the telling signs of EBOV infection is a maculopapular rash by days 5-7 post-exposure. Non-fatal cases usually present with fever at 5-9 days with improvement by day 11 due to the development of humoral antibodies. Fatal cases will often develop symptoms early in infection (6-16 days) and likely progress into hemorrhagic shock. EBOV has the highest mortality rate ranging from 60-90%. SUDV, BDBV, and MARV follows with mortality rates at 50-60%, 40%, and 70-85% (CDC, 2022a; Malvy et al., 2019).

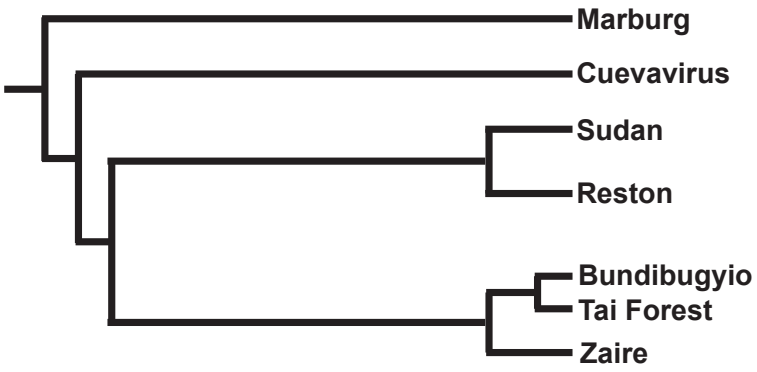


Figure I-26: Filoviridae family tree

Filoviridae family tree. Adapted from Chippaux et al., J Venom Anim Toxins Incl Trop Dis, 2014 (Open Access)

EBOV was first reported in 1976 when it appeared in the Democratic Republic of Congo (DRC) and Sudan. Ever since its inception, there have been ongoing outbreaks, with the 2014-2016 outbreak as the largest Ebola outbreak in history leading to 11,000 deaths (**Figure I-27**) (WHO, 2022a). These periodic outbreaks are global health concerns, and the epidemic potential for not only EBOV but other viruses within the family accelerates the need for the development of therapeutics and vaccines.

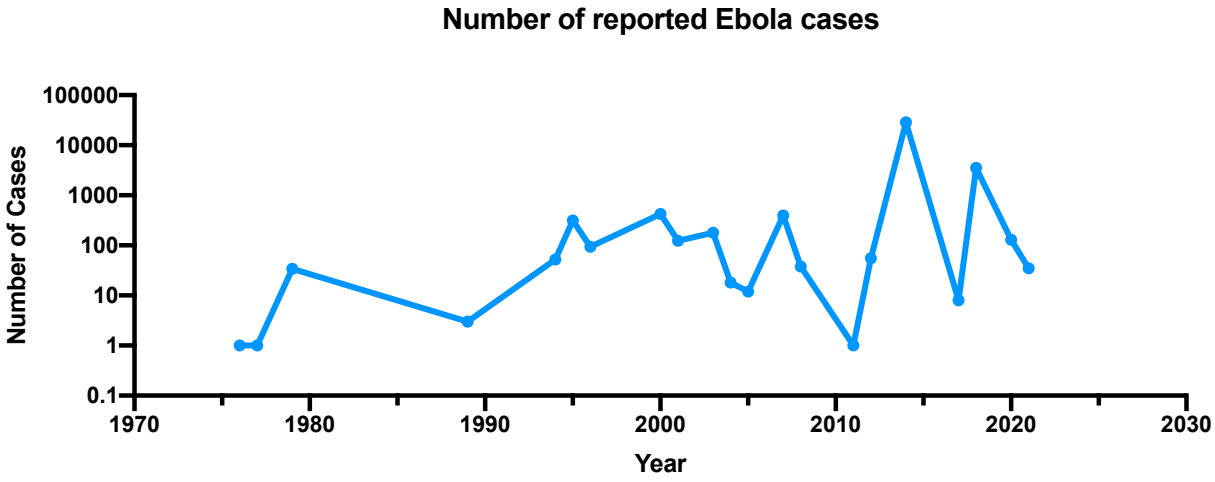


Figure I-27: Periodic outbreaks caused by EBOV

Data collected from the WHO and plotted by number of cases to each year shows the periodic outbreaks constantly reoccurring. Plotted using Graphpad Prism 9.

Ebola virus viral life cycle

During infection, the virus first binds target cells through two relatively nonspecific receptors: first C-type lectins, which interact with the glycans on the EBOV GP, and second the phosphatidylserine receptors that interact with the viral envelope. The virion is then internalized by micropinocytosis (Nanbo et al., 2010). Once internalized into endosomes, the low pH allows cathepsins to remove the heavily glycosylated mucin-like domain (MLD) and glycan cap, exposing the receptor-binding domain (RBD). This allows for NPC1 to bind to the RBD. NPC1 engages the hydrophobic cavity at the GP head and triggers a conformational change (Wang et al., 2016). The hydrophobic loop of the GP2 will then insert itself into the host membrane, forming a transmembrane 6-helix bundle fusion pore. This allows for the release of the ribonucleoprotein complex into the cytoplasm of the host cell. The viral RNA is following replicated and translated. All components of the viral particle are then transported to the cell surface, and virion assembly is then directed by VP40 to start the budding process. The GP then antagonizes tetherin, which helps the budded viral particle to be released by the cell (Hoenen et al., 2019) (**Figure I-28**).

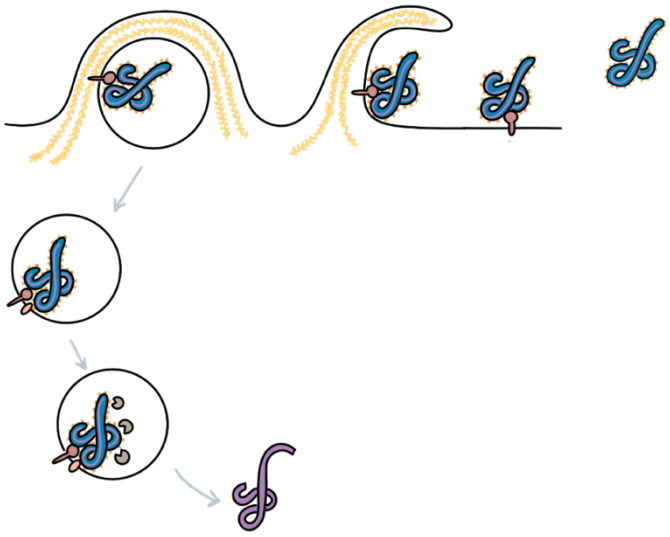


Figure I-28: Ebola virus entry

The Ebola virus entry pathway. Starting with non specific interactions leading to micropinocytosis of the virus into the host cell. Once virus is internalized into endosomes, the glycoprotein is cleaved and the virus is following fused with the endosomal membrane and RNA is released into the host cell cytoplasm.

The Ebola virus glycoprotein

The membrane-anchored EBOV GP is essential in both viral entry and egress and is the sole known target for protective antibodies (Lee and Saphire, 2009), therefore making it the primary target for antiviral vaccines and therapies. The mature EBOV GP is composed of two subunits, GP1 and GP2, that heterodimerize through disulfide bonds and associate to form trimers (**Figure I-29**). The GP1 includes the glycan cap, MLD, and core. When trimerized, the receptor-binding domain (RBD) lays in the middle of the chalice-like shape. The GP2 includes the fusion loop (EBOV GP is a class I fusion protein) and transmembrane region of the GP. A disulfide linkage between GP1 (140kDa) and GP2 (26kDa) helps the assembly of the 450kDa trimer (Beniac and Booth, 2017; Lee et al., 2008; Moller-Tank and Maury, 2015).

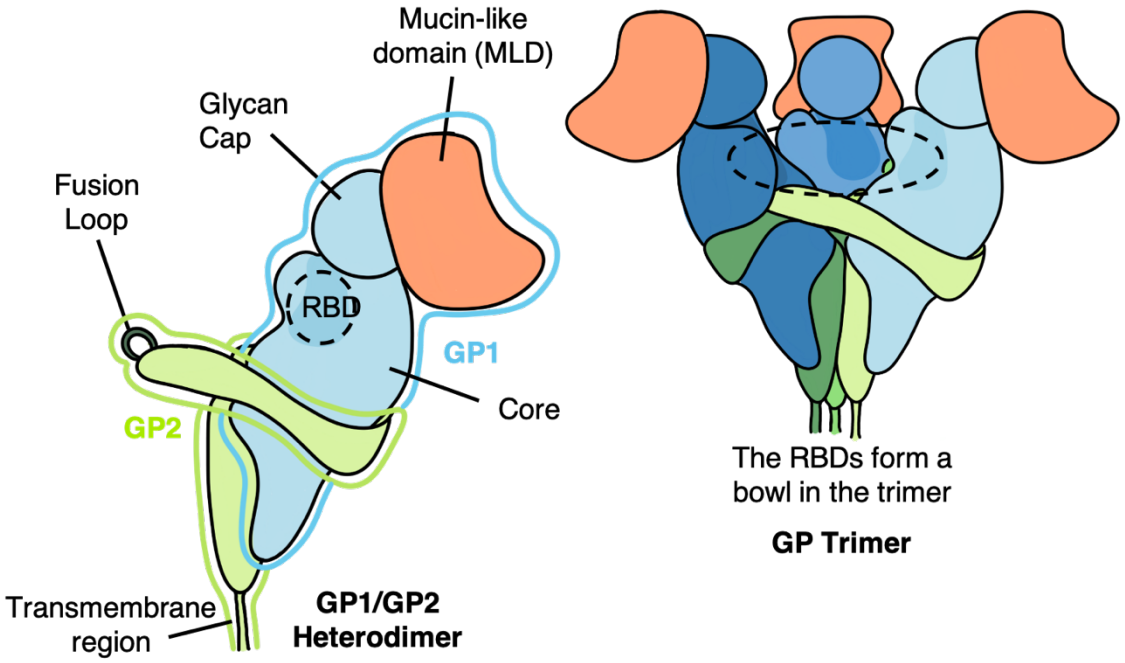


Figure I-29: Ebola glycoprotein

Cartoon of the Ebola glycoprotein. The GP is composed of GP1 (blue) and GP2 (green), which heterodimerize to form a trimer. Adapted from Moller-Tank et al., PLOS Pathogens 2015 (Open Access).

Antibodies to Ebola virus

The 2014-2016 Ebola epidemic led to a unanimous effort for antibody discovery against the EBOV GP in hopes of identifying therapeutic candidates. Multiple groups since then have isolated hundreds of antibodies from convalescent donors as well as donors vaccinated with the Ervbo vaccine (Bornholdt et al., 2016; Ehrhardt et al., 2019; Flyak et al., 2016; Williamson et al., 2019). Scattered throughout the GP, there are many classes of antibodies elicited to the EBOV GP broadly binned into four sites: glycan cap, GP1 core, receptor binding domain, and the base. Within the base-specific antibodies, they can be further binned into three types: antibodies targeting the base, the tip of the IFL, and the IFL-cathepsin loop. This gives a total of 6 antibody classes to the EBOV GP (Saphire and Aman, 2016). Combinations of antibodies targeting several of the listed epitopes have been proven protective and are used as antibody therapeutics to EBOV (Davidson et al., 2015; Gilchuk et al., 2020b; Pascal et al., 2018). There are three main mechanisms for neutralization for EBOV antibodies: inhibition of cathepsin cleavage, interference with structural rearrangements of the GP2, and blocking receptor binding.

KZ52 is one of the first identified antibodies from a convalescent donor (Lee et al., 2008), binding to the interface of GP1 and GP2 at the base of the trimer, interfering with membrane fusion. Despite potent neutralization and protection in rodents, the antibody failed to protect non-human primates in later studies. However, these antibodies targeting this epitope had later shown to be useful in antibody cocktails, leading to the development of ZMapp. The ZMapp antibody cocktail contains 2G7 and 4G7 targeting the base and 13C6 targeting the glycan cap and has shown 100% protection in NHPs to lethal challenge (Qiu et al., 2014). However, the ZMapp cocktail antibodies only possess reactivity to EBOV and do not recognize other strains such as BDBV and SUDV.

Base-directed and glycan cap-directed antibodies have both been shown to have potent neutralization properties. Many antibodies targeting both the glycan cap and base have also since been identified with cross-reactivity to multiple ebolaviruses as well. A broad ebolavirus antibody cocktail containing rEBOV-548 and rEBOV520 directs one antibody to the glycan cap and one to the base and has been shown to infer protection to all three lethal strains of ebolavirus: EBOV, BDBV, and SUDV (Gilchuk et al., 2020b).

Nestled under the glycan cap is the receptor-binding domain, the directed epitope for receptor NPC1. Several neutralizing antibodies have been identified to this epitope, with mAb114 being one of the first characterized (Corti et al., 2016), is currently approved for compassionate use for EVD treatment. These antibodies directly block NPC1 binding, therefore resulting in neutralization. The RBS is often a site of interest as regions within the RBD are widely conserved across the filovirus family, thereby expanding reactivity outside of the ebolavirus genus. Paradoxically from what has been seen in EBOV convalescent individuals, antibodies isolated from MARV convalescent individuals have been shown to primarily target the NPC1 RBD. Previous studies have shown that the RBD on EBOV GP is obscured by the MLD and glycan cap, therefore only becoming exposed after cathepsin cleavage. The MARV GP RBD is more accessible due to the lack of the MLD, therefore allowing the identification of larger pools of antibodies reactive to the RBD (Flyak et al., 2015).

Antibodies targeting internal fusion loop (IFL) epitopes are of interest due to their potential for broadly neutralizing activity as IFL is conserved across the ebolavirus genus. Such antibodies can be separated into two categories: ones that target the tip of the IFL and ones that target the IFL-cathepsin loop. 6D6 and FVM02 are examples of antibodies that had been identified to the tip of the IFL (Furuyama et al., 2016; Keck et al., 2016). Although antibodies targeting this tip have varying neutralization potencies, they have shown broad reactivity. The second category of antibodies targets the IFL-cathepsin loop. mAb100, an antibody in this

group, binds a quaternary epitope covering residues by the IFL as well as the GP1, thereby locking the GP in its pre-fusion conformation and altering cleavage (Misasi et al., 2016).

It has been suggested that the early antibody response to the EBOV GP targets the glycan cap or the sGP, with a low frequency of neutralizing antibodies (Williamson et al., 2019). However, as EBOV persists in the central nervous system, the eye, and testes after acute infection, there is a continued increase in antibody SHM, avidity, and neutralization titers (Davis et al., 2019). Additionally, Wec and colleagues (Wec et al., 2019) have optimized antibodies through affinity maturation, improving protective efficacy against EBOV, BDBV, and SUDV (Bornholdt et al., 2019). Taken together, this suggests that EBOV-specific antibodies need to be affinity matured to achieve potent neutralization.

Vaccines and therapeutic monoclonal antibodies targeting Ebola virus disease

Since the 2014 epidemic, there has been lots of progress in achieving EBOV-specific vaccines and therapeutics. Two antibody-based drugs, Inmazeb™ and Ebanga™, received approval from the U.S. Food and Drug Administration (FDA) in 2020 (Bornholdt et al., 2019; Corti et al., 2016; Gilchuk et al., 2020b; Levine, 2019; Pascal et al., 2018). And ERVEBO, a vaccine for EBOV, was also approved by the FDA in 2020. ERVEBO, manufactured by Merk, is produced by using recombinant vesicular stomatitis virus with the EBOV GP grafted on in place of the VSV G protein (Wong et al., 2014).

However, many of the approved therapeutics are monospecific to EBOV only. The sporadic nature of these epidemics leaves a need for pan-ebolavirus and pan-filovirus therapeutics and vaccines. Understanding the human antibody repertoire induced to ebolavirus GPs will help efforts to identify the next generation of pan-ebolavirus antibody therapeutics and aid the design and development of broadly protective vaccines.

Part VII: SARS-CoV-2 Virus

Introduction to the SARS-CoV-2 virus

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) emerged in December 2019 in Wuhan, China causing an outbreak of viral pneumonia. This highly transmissible nature of the virus quickly spread worldwide, causing the coronavirus disease 2019 (COVID-19) pandemic. Since then, there have been 350 million cases and 5.6 million deaths worldwide. SARS-CoV-2 belongs to the Coronaviridae family, a beta-coronavirus genus with positive sense, unsegmented genomes spanning 30kb long encoding for roughly 29 proteins. 16 non-structural, nine accessory factors, and four structural proteins(Gordon et al., 2020). The four structural proteins encoded include the spike (S), membrane, envelope, and nucleocapsid. While the membrane, envelope, and nucleocapsid proteins promote virion assembly during viral replication, the S protein is the single surface protein decorating the surface of the virion.

SARS-CoV-2 is the 7th known coronavirus to infect humans. Within the six additional coronaviruses, two are known to be highly pathogenic: Severe Acute Respiratory Syndrome coronavirus (SARS) and Middle East Respiratory Syndrome coronavirus (MERS). SARS emerged in 2002 as the first pandemic coronavirus, infecting 8,096 people in 29 countries leading to 774 deaths(CDC, 2016). MERS emerged in 2013, despite having a lower caseload with roughly 2,500 cases, possesses a higher mortality rate at 35%(WHO, 2022b). The following four are known to be endemic and are some of the known causes of the common cold. These viruses include two from the alpha coronavirus genus: OC43, HKU1, and two from the beta coronavirus genus: NL63, and 229E (**Figure I-30**).

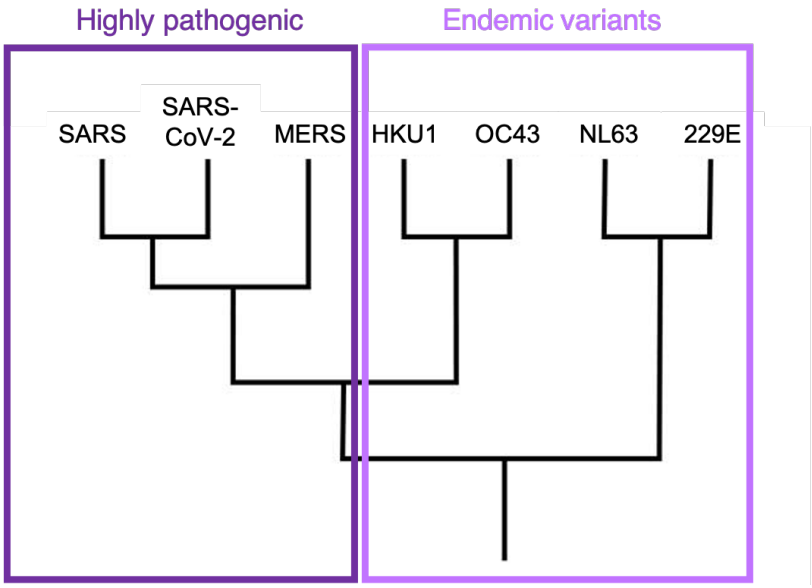


Figure I-30: Coronavirus phylogenetic relationship

Phylogenetic tree indicating the highly pathogenic and endemic strains of coronavirus.

Upon infection, the receptor-binding domain of the SARS-CoV-2 spike engages with angiotensin-converting enzyme 2 (ACE2) (Lan et al., 2020) (**Figure I-31**). Besides receptor binding, cell surface serine protease TMPRSS2 is required for proteolytic cleavage of the S protein is also required by host-cell derived proteases to permit fusion. This dual interaction with ACE2 and proteolytic cleavage promotes viral uptake and fusion, thereby releasing the RNA for replication and translation (Zhang et al., 2021). Together the replicated RNA and translated structural proteins result in budding into the lumen of secretory vesicular compartments. The virions are then secreted via exocytosis from the infected cell.

Humoral response to SARS-CoV-2

The spike (S) protein is the principal antigen recognized in the protective antibody response to SARS-CoV-2 and is; therefore, the target for many monoclonal antibody discovery efforts focused on developing medical countermeasures against COVID-19. The S protein is composed of two subunits: S1 and S2. S1 houses the receptor-binding domain (RBD) and N-

terminal domain (NTD), and S2 houses the fusion peptide and heptad repeat regions, therefore mediating the fusion between virus and host cell membrane (Bosch et al., 2003; Tortorici and Veesler, 2019).

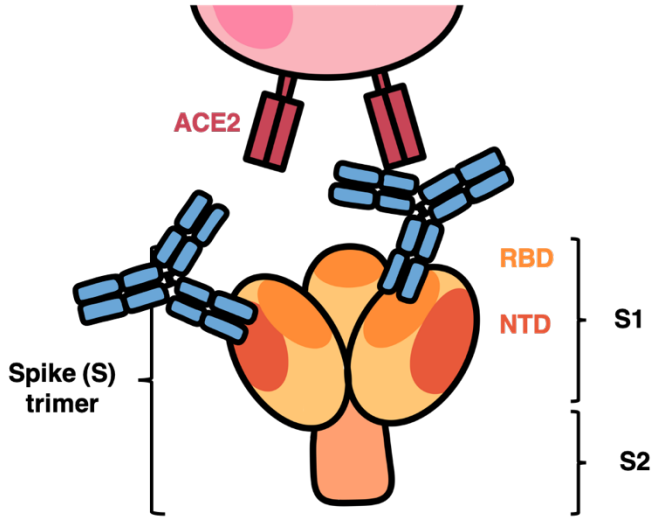


Figure I-31: SARS-CoV-2 spike protein

Cartoon rendition of the SARS-CoV-2 S protein. The S protein is composed of the S1 domain that houses the RBD and NTD, as well as the S2 domain which houses the fusion machinery. Most antibody discovery efforts have been focused on the RBD region of S1.

As the SARS-CoV-2 pandemic started to unravel, antibodies were being discovered at an unprecedented scale and speed that had not previously been possible for other pathogens. In the last year, there had been more antibody sequences published to SARS-CoV-2 than to any other pathogen revealing characteristics of the human antibody response to SARS-CoV-2. The RBD of the S is the target for most neutralizing antibodies (**Figure I-31**). Within the antibodies targeting the RBD, there are four classes of antibodies deemed classes 1, 2, 3, and 4. Class 1 and 2 antibodies target epitopes overlapping with the ACE2 binding site. Class 3 antibodies are potently neutralizing but do not directly bind the ACE2 binding site. And class 4 antibodies target a cryptic epitope outside the RBD and are generally less potent (Barnes et al., 2020). NTD specific antibodies have also been found to be protective *in vivo* but require Fc effector functions for optimal protection (Cerutti et al., 2021; Suryadevara et al., 2021). Although

the majority of discovery efforts have been focused on the S1 domain, antibodies to S2 have also been described. A class of antibodies directed at the S2 domain have demonstrated neutralization of SARS-CoV-2 by inhibition of membrane fusion. S2P6, an antibody belonging to this class, had shown to broadly neutralize sarbecoviruses, merbecoviruses, and embecoviruses through the same mechanism, illustrating this may be a site conserved across coronaviruses (Pinto et al., 2021).

Vaccines and antibodies to SARS-CoV-2

Multiple SARS-CoV-2 vaccines have been developed in the last two years, aimed at producing neutralizing antibody responses to the S protein. Multiple platforms have been used to deliver S protein as vaccine candidates including but not limited to mRNA (by Moderna and Pfizer/BioNTech), adenoviral vectors (Astrazeneca/Oxford, Johnson and Johnson), inactivated virus (Sinovac), protein subunits (Novavax), and virus-like particles (Glaxo Smith Kine).

Several antibody therapeutics have undergone clinical development for SARS-CoV-2. However, with the majority of the neutralizing antibody response targeting the RBD of the S protein, there is focused evolutionary pressure at the RBD. Various variants of SARS-CoV-2 have since emerged with multiple mutations in the RBD region of the S protein, which have led to the loss of neutralization for many clinically developed antibodies to SARS-CoV-2. Bamlanivimab or LY-CoV555, developed by Eli Lilly, was the first antibody that received emergency use authorization (EUA) in 2020. However, the emergence of the E484K mutation in beta, gamma, delta, epsilon, and kappa variants rendered the antibody non-neutralizing. This antibody was later accompanied by Etesevimab (LY-CoV016), creating an antibody cocktail. LY-CoV016's neutralization was later abrogated by the K417N/T mutation, therefore leading to FDA's retrieval of EUA for the cocktail. Regeneron's antibody cocktail is made up of two antibodies: Casirivimab (REGN10933) and Imdevimab (REGEN10987). REGEN10933 has

reduced neutralization against beta and gamma variants which contain E417N and E484 mutations but do not completely lose binding. And REGEN10987 was negatively affected by the N440K mutation, which recently emerged in the omicron variant (Starr et al., 2021). Adintrevimab (ADG20), developed by Adagio therapeutics, had shown a reduction in potency to the omicron variant but had held up against previous variants (Planas et al., 2021). Evusheld, a cocktail developed by AstraZeneca, contains AZD1061 (COV2-2130) and AZD8895 (COV2-2196). AZD8895 exhibited reduced neutralization to the E484 mutations but still maintained relatively stable neutralization until the introduction of Q493R, S477N, and E484A mutations which arose in the omicron variant, and AZD1061 is sensitive to the L452R mutation (Greaney et al., 2021; VanBlargan et al., 2022). Sotrovimab (S309), developed by Glaxo Smith Kine and Vir Biotechnology has so far retained neutralizing activity against all variants.

Part VIII: Gaps of knowledge

Soto and colleges as well as Briney and colleauges have described that the experimental diversity of the circulating antibody repertoire is substantially smaller than expected. The theoretical diversity had been calculated as 10^{11} (Trepel, 1974), in comparison to the experimental diversity described in these studies calculated as 11^{6} (Briney et al., 2019; Soto et al., 2019). However, there have not been previous studies demonstrating the what fraction of the B-cell repertoire is specific for a viral antigen. In addition to understanding the number of clonal families or antibodies there are to a specific antigen, the diversity of functional phenotypes for antibodies to specific antigens is also an important avenue of study to better understand antibody epitopes and classes of antibodies for discovery efforts.

Along with this, high frequency of public clonotypes have been described with large bulk sequencing datasets, revealing that large sequence datasets are required for the mining of

public clonotypes (Briney et al., 2019; Setliff et al., 2018; Soto et al., 2019). However, most public clonotypes have previously been identified in bulk sequencing studies, there has been a gap in knowledge in understanding what exactly do these public clonotypes do functionally, and how their functions contribute to humoral immunity.

Throughout this dissertation, we discuss the both the functional and genetic diversity of antibodies to AL amyloidosis, Influenza, Ebola virus, and SARS-CoV-2. Additionally, we also discuss the prevalence and characteristics of public clonotypes to Ebola virus and SARS-COV-2.

CHAPTER II

DEVELOPMENT OF A SEQUENCE ANALYSIS TOOLBOX FOR REPERTOIRE-WIDE SEQUENCING STUDIES

This chapter is an adaptation of the following published manuscripts:

Chen E.C.*, Rubinstein S.*, Soto C., Bombardi R.G., Day S.B., Myers L., Zaytsev A., Majedi M., Cornell F., Crowe J.E. Diverse patterns of antibody variable gene repertoire disruption in patients with amyloid light chain (AL) amyloidosis. *PLOS ONE*. 2020; 15(7): e0235713. doi: 10.1371/journal.pone.0235713 (*contributed equally)

Soto, C.S., Bombardi, R.G., Kozhevnikov, M., Sinkovits, R.S., **Chen E.C.**, Branchizio, A., Kose, N., Day, S.B., Pikinton, M., Gujral, M., Mallal, S., and Crowe, J.E.. High Frequency of Shared Clonotypes in Human T Cell Receptor Repertoires. *Cell Reports* 2020; 32(2): e107882. doi: 10.1016/j.celrep.2020.107882

Zost, S.J., Gilchuk, P., Chen, R.E. Case, J.B., Riedy J.X., Trivette, A., Nargi, R.S., Sutton, R.E., Suryadevara N., **Chen E.C.**, Binshtein E., Shrihari S., Chu, H.Y., Didier, J.E., MacRenaris, K.W., Jones, T., Day, S., Myers, L., Lee., F.E., Nguyen, D.C., Sanz, I., Martinez, D.R., Rothlauf, P.W., Bloyet, L.M., Whelan, S.P.J., Baric, R.S., Thackray, L.B., Diamond, M.S., Carnahan, R.H., and Crowe, J.E. Rapid isolation and profiling of a diverse panel of human monoclonal antibodies targeting the SARS-CoV-2 spike protein. *Nature Medicine.* 2020*;* 26**,** 1422–1427 (2020). doi: 10.1038/s41591-020-0998-x

CHAPTER OVERVIEW

In the last few years, there has been a substantial increase in the amount of B-cell receptor and T-cell receptor sequencing utilizing next-generation sequencing technologies. During my initial studies since joining the Crowe laboratory, I built a toolbox of sequence analysis tools enabling the visualization and profiling of immunoglobulin repertoire sequencing. In this chapter, I will focus on the sequence analysis I have done regarding (1) profiling the B-cell repertoire to AL amyloidosis patients, (2) sequence matching and profiling of T-cell receptors with publicly available sequences, and (3) understanding the characteristics of SARS-COV-2 specific antibodies.

The studies described in this chapter would not have been possible without many colleagues of the Crowe lab and their contributions in performing the library preps and assistance in sequence analysis. I would like to thank Robin Bombardi for executing the majority of the sequencing library preparations for both the B-cell receptor(BCR) and T-cell receptor studies(TCR). Patient sample identification was done by Dr. Samuel Rubeinstein. I would also like to thank Dr. Cinque Soto for his expertise and guidance on the TCR studies. Lastly, I would like to thank Drs. Seth Zost, Pavlo Gilchuk, Dr. Naveenchandra Suryadevara, Elad Binshtein, Robert Carnahan, and Rachel Nargi, Rachel Sutton, Joseph Reidy, Andrew Trivette, and many others for isolating and sequencing SARS-CoV-2 specific B cells.

INTRODUCTION

Amyloidoses are systemic illnesses caused by the extracellular deposition into tissue of amyloid proteins, which are generally subunits of normal serum proteins consisting largely of beta-pleated sheet regions. The most common amyloidosis in the United States is light chain (AL) amyloidosis, in which the amyloidogenic protein typically is free antibody light chain secreted by a population of plasma cells generally thought to be clonal (Gertz, 2016). The current best practices for determining patient hematologic disease status involve measuring the absolute quantity of free light chain proteins in serum (Comenzo et al., 2012; Kumar et al., 2012). Free light chain ratio is determined by measuring serum free light chains in patients and identifying the kappa-to-lambda light chain ratio. A complete hematologic remission is defined in part by normalization of the free light chain ratio. However, many patients with complete hematologic responses do not experience an organ response, or they even experience organ progression (Comenzo et al., 2012; Manwani et al., 2019; Palladini and Merlini, 2019). Furthermore, some patients present with relatively low levels of serum free light chain protein and discordantly advanced organ involvement(Milani et al., 2017). One plausible mechanism for

this finding is the persistence of small numbers of cells from plasma cell clones that continue to produce amyloidogenic light chain protein, which contributes to progressive organ dysfunction at low concentrations. As many as 60% of AL amyloidosis patients in complete hematologic response may have residual clonal amyloidogenic plasma cell populations, as measured by next generation flow cytometric analysis of circulating white blood cells(Kastritis et al., 2018; Sidana et al., 2020). More sensitive and specific methods for determining hematologic disease status are needed.

Early in B cell development, immunoglobulin germline genes rearrange to encode the B cell receptor expressed on naïve B cells, which then is modified and diversified by somatic hypermutation after B cells respond to antigen in germinal centers. This process allows B cells and plasma cells in a lineage derived from a single cell (a clone) to persist with one distinct heavy chain and one light chain sequence. Emerging high-throughput sequencing (HTS) technologies have made it possible to sequence the heavy and light chain variable regions of millions or billions of B and T cell receptors from single samples, making it possible to identify and track clonal lymphocyte populations. The ability to track clonal populations of B or T cells has implications for monitoring disease progression over time in settings where one might track the proliferation of an aberrant B cell population involved in cancer progression. The clonoSEQ assay by Adaptive Biotechnologies(Robins, 2013) was approved by the FDA for the detection and monitoring of minimal residual disease (MRD) in bone marrow samples from multiple myeloma or B-cell acute lymphoblastic leukemia (ALL) patients(Robins, 2013). The Adaptive Biotechnologies assay uses genomic DNA as input and has been shown to provide sufficient sensitivity for clinical utility(Robins, 2013). While the clonoSEQ assay is the only FDA-approved assay to date using immune repertoire sequencing as a clinical diagnostic test, the technology is relatively expensive and does not provide full-length sequencing of the variable region, which is of interest for following B cell clones that can expand in lineages due to somatic hypermutation. We tested the

performance of an alternative reverse transcription-based methodology that uses mRNA-based sequencing (mRNA) to determine if we could identify the dominant clonal populations and features of immune repertoire disruption for patients with AL amyloidosis. The mRNA-based method provides full-length sequencing of the antibody variable region, which is pertinent to following the occurrence of somatic mutations outside of the junctional region targeted in the more size-limited amplicons in current genomic DNA-based methods. We identified the dominant clonotype in each subject studied, we also tracked the clonal population of the dysplastic B cell in one subject over time. With this approach, we identified a subset of patients with dysplastic B cells that also expressed high levels of a particular heavy chain protein. Interestingly, when we examined the top ten most frequently represented mRNA clonotypes present in the antibody gene transcripts for each subject, we identified somatic variants of several clones, suggesting that AL amyloidosis disease might be driven in part by antigen stimulation. This proof-of-concept study demonstrates the applicability of mRNA-based sequencing for the detection of over-represented clonal populations and important features of the overall immune repertoire in B cells from patients with AL amyloidosis. This will help identify the clonal plasma cells in patients with hematologically active disease.

RESULTS

Bone marrow aspirates from AL amyloidosis patients are sequenced

Four patients had detailed clinical information available for review. The remaining patients consented for tissue bank participation but not linkage with the electronic health record. The four patients with detailed clinical information had active hematologic disease at the time of sample collection. Three had newly diagnosed, untreated AL amyloidosis, and the fourth had relapsed after initial therapy. Organ involvement was biopsy proven in three patients (AM1, AM2, AM3) and was clinically suspected in a fourth (AM5) on the basis of proven fat pad

involvement as well as a suggestive cardiac presentation and echocardiogram. Organ involvement was as follows: 1) cardiac, renal, and duodenal involvement, 2) cardiac and renal involvement, 3) duodenal involvement alone, and 4) cardiac involvement alone. All patients had active AL amyloidosis disease at the time the samples were acquired (**Table II-1**).

The medical records of the patients showed that clinical testing revealed the patient bone marrow aspirate mononuclear cell suspensions had a median of 14% plasma cells (range: 6 to 90), and the serum had a median free light chain concentration of 4.11 mg/dL (range, 2.7 to 118) at the time points studied. Three patients had lambda light chain disease, and one had kappa light chain disease, in all cases, immunofixation was free light chain only. Three had Mayo stage III disease, and one had Mayo stage II disease.

The viability of the previously cryopreserved mononuclear cells in the bone marrow aspirate suspension after thawing ranged from 30 to 42%. The estimated number of B cells per donor used in the sequencing reactions ranged between 1.7×10^4 to 2.3×10^4 B cells (**Table II-1**). We sequenced the heavy and light chain immune repertoires of seven donors using Illumina's paired-end (PE) Mi seq platform. We then used the PyIR bioinformatics pipeline to process and correct errors in the reads. After processing and error correction, we obtained an average of 1,963 unique and productive heavy chain variable gene reads per donor and an average of 2,803 unique and productive light chain variable gene reads per donor (**Table II-1**). We adopted the V3J clonotype definition from Soto *et al.*(Soto et al., 2019) to group together somatic variants belonging to the same lineage. A V3J clonotype was defined as comprising clones with sequences using the identical CDR3 amino acid sequence and the same V and J germline gene assignments (ignoring allelic distinctions). Thus, any somatic variants sharing these three properties were considered to belong to the same V3J clonotype and thus the same B cell clonal lineage. After grouping somatic variants using the V3J clonotype definition, we obtained an average of 1,393 unique V3J clonotypes per donor for heavy chains and an

average of 1,557 unique V3J clonotypes per donor for light chains (**Table II-1**). For downstream analysis of the dominant V3J clonotype, we pooled together all replicate samples from the same subject.

Patient	Clinical tissue status [cardiac (C), renal (R), duodenal (D) or not available (na)]	Number of viable plasma cells* in the sample (x 10^5)	Light chain variable gene sequences obtained after de-duplication of biological replicates**		Genetic features of the dominant light chain variable gene***		
			Total unique reads	Total unique clonotypes	V genes	J genes	CDR3 (amino acids)
AM1	C, R, D	1.2	3,921	1,683	*IGLV3-21*	*IGLJ3*	QVWDRSSDRPV
AM2 *Time-point 1*	na	5.1	8,601	4,937	***IGLV3-25***	***IGLJ2***	**QSADSSGTYEVI**
AM2 *Time-poin 2*	na	0.7	2,331	1,165	***IGLV3-25***	***IGLJ2***	**QSADSSGTYEVI**
AM2 *Time-point 3*	na	6.8	1,702	630	***IGLV3-25***	***IGLJ2***	**QSADSSGTYEVI**
AM3	C	4.0	4,524	2,885	*IGLV1-47*	*IGLJ1*	AAWDGSLSGYV
AM4	na	0.4	1,788	1,307	***IGLV2-14***	***IGLJ1***	**SS**FT**S**SSSY**V**
AM5	C, R	1.0	577	266	***IGLV2-14***	***IGLJ1***	**SS**YT**S**ITNTL**V**
AM6	na	1.1	1,316	919	*IGKV3-20*	*IGKJ4*	QQYGTSPLT
AM7	na	1.5	475	223	*IGLV6-57*	*IGLJ3*	QSYQGSSGV

* The total number of viable mononuclear cells in the aliquot of cryopreserved bone marrow aspirate sample was multiplied by plasma cell percentage to achieve number of viable plasma cells in the sample.
** Three replicates for: AM1, AM2 Time-point 1, AM2 Time-point 2, AM3, and AM4. Two replicates for: AM2 Time-point 3. One replicate for:AM5, AM6, and AM7.
*** **Bold** or **bold/underlined** entries highlight two different sets of samples with common genetic features.

Table II-1: Results of antibody gene repertoire sequence analysis experiments for bone marrow aspirate specimens from seven patients with AL amyloidosis.

Each subject's number of sequence reads, unique clonotype, V and J gene of the dominant clone as well as CDR3 are listed in the table. Information is listed for both heavy and light chain repertoires. Technical replicates were analyzed to ensure the dominant clone remained the dominant clone in all replicates, indicated by the dashes.

Dominating light chain clone is identified in every patient

One of the hallmarks of AL amyloidosis is the presence of a single clone that dominates the light chain repertoires. We sought to ask if there is diversity in the light chain variable genes that were over-represented in the next generation sequencing (NGS) of these subjects. One metric for measuring diversity often used in immune repertoire sequencing is Shannon entropy

(Soto et al., 2019). We computed the Shannon entropy value for the light chain repertoires using the V3J clonotype definition and found values for the light chain repertoire for each subject: AM1 at 1.66, AM2 at 5.02, 1.14, 0.10, AM3 at 4.5, AM4 at 6.03, AM5 at 0.34, AM6 at 1.19, and AM7 at 0.51. For comparison, we analyzed the Shannon entropy in three healthy subjects (designated HIP1, 2, or 3) using large repertoire data sets from Soto *et al*., (Soto et al., 2019). The Shannon entropy values for each of the healthy subjects were as follows: HIP1 at 9.80, HIP2 at 9.47, and HIP3 at 8.76. Thus, the repertoire of amyloidogenic patients clearly exhibited a general profound lack of diversity, consistent with the presence of a dominant clonotype.

To determine the fraction of heavy and light chain repertoires accounted for by each V3J clonotype, we divided the total number of unique somatic variants associated with each clonotype by the total number of unique somatic variants for that subject (**Figure II-1**). Since AL amyloidosis is a disease affecting the immunoglobulin light chain, we reasoned that the V3J clonotype with the largest number of somatic variants likely corresponded to the variable gene of the aberrant light chain sequence causing the disease. For three of the seven subjects considered here (AM2, AM5 and AM7) a single V3J clonotype accounted for approximately 50% of the light chain repertoire (**Figure II-1A**). If we relaxed the cutoff to just 30% of the total number of unique reads, four of seven repertoires tested (AM1, AM2, AM5 and AM7) had a single dominant V3J clonotype for the light chain repertoire. A comparison with light chain sequencing from three healthy subjects revealed that in those healthy subjects the most prevalent single V3J clonotype accounted for less than 1% of the light chain repertoire (**Figure II-1A**, see subjects HIP1, HIP2 or HIP3). While none of the subjects with AL amyloidosis shared the exact same dominant V3J clonotype in their light chain repertoire, two out of the seven subjects shared the identical V-J gene combination and the identical CDR3 length for their dominant V3J clonotype (**Table II-1, Figure II-1A** and **Tables II-2**). The dominant V3J clonotype from the heavy chain sequencing from these seven subjects was much less pronounced (**Figure II-1B**). In fact, four out of seven subjects (AM1, AM5,

AM6, AM7) had a similar profile for clonotype dominance as seen in heavy chain sequencing from three healthy donors HIP1, HIP2 or HIP3. However, we identified three subjects with what appeared to be a corresponding heavy chain that was highly over-expressed (**Figure II-1**, subjects AM2, AM3, and AM4). This finding suggests that the dysplastic cell clone in these patients coordinately over-expressed both an antibody light chain and a heavy chain. In contrast, the repertoires of the other subjects studied showed over-expression of an antibody light chain only.

We examined the reproducibility of the findings in the cases where we generated two or three technical replicates on the same RNA extract from a single aspirate sample. In each of the cases (separate replicates in total for light and heavy chains), the technical replicates provided identical results within the replicate group (**Table II-2**).

In one subject, we tracked the single aberrant V3J clonotype over three years to simulate the use of this technique in a clinical monitoring scenario, and found the clone was present with varying levels over time (see subject AM2 data in **Table II-1**, **Figure II-1A** and **Table II-2**). The samples were collected with 18 months between the first and second time point, and 4 months between the second and third time point. These data suggest that this method could be used for tracking the persistence of a dominant pathogenic clone in a patient over time.

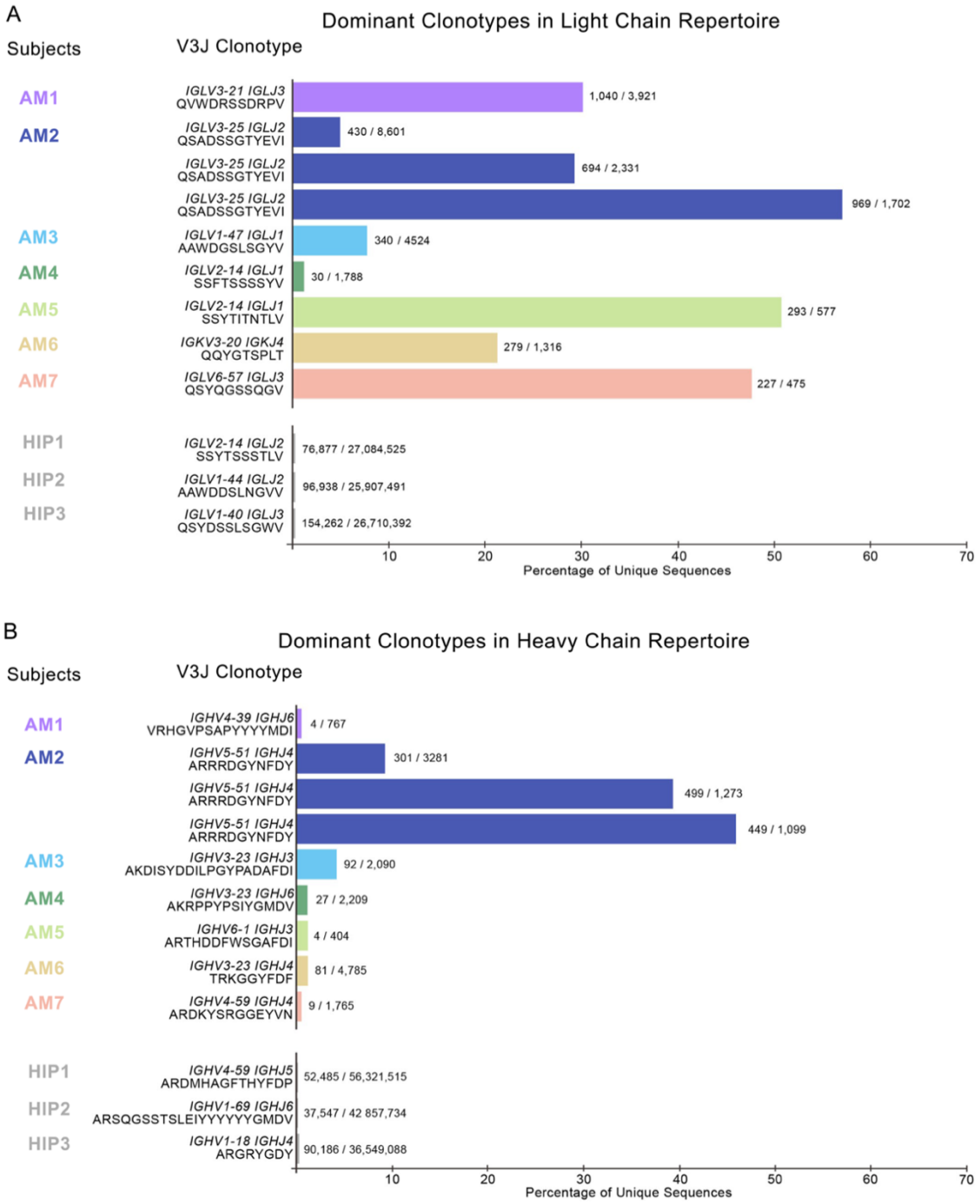


Figure II-1: Dominant clonotypes in the light and heavy chain immune repertoires of subjects affected with light chain amyloidosis

(A) Most abundant light chain V3J clonotypes in each subject **(B)** Most abundant heavy V3J clonotypes in each subject. Percentages were obtained by dividing the total number of unique sequences containing the V3J clonotype in each subject by the total number of unique sequences for the entire repertoire of each subject. The somatic variant count for the most prevalent V3J clonotype appears at the end of each bar graph. For comparison, we also included sequencing data from 3 healthy subjects denoted as HIP1, HIP2 or HIP3.

Patient	Light chain					Heavy chain				
	Sequences obtained		Genetic features of the dominant sequence			Sequences obtained		Genetic features of the dominant sequence		
	Sequence reads for each of 1-3 technical replicates	Unique clonotypes for each of 1-3 technical replicates	$V_{K \text{ or } L}$ gene*	$J_{K \text{ or } L}$ gene*	LCDR3 (amino acids)*	Sequence reads for each of 3 technical replicates	Unique clonotypes for each of 3 technical replicates	V_H gene*	J_H gene*	HCDR3 (amino acids)*
AM1	1,963 2,296 1,953	1,006 1,139 1,029	*IGLV3-21* – –	*IGLJ3* – –	QVWDRSSDRPV ---------- ----------	351 391 357	340 366 334	*IGHV4-39* – –	*IGHJ6* – –	VRHGVPSAPYYYYMDI ---------------- ----------------
AM2 *Time-point 1*	5,754 4,755	3,575 2,946	***IGLV3-25*** –	***IGLJ2*** –	**QSADSSGTYEVI** ----------	1,225 1,372 1,260	1,071 1,185 1,082	***IGHV5-51*** – –	***IGHJ4*** – –	**ARRRDGYNFDY** ---------- ----------
Time-point 2	1,590 843 1,018	857 476 591	– – –	– – –	---------- ---------- ----------	470 601 640	288 322 344	– – –	– – –	---------- ---------- ----------
Time-point 3	82 958 977	176 309 299	– – –	– – –	---------- ---------- ----------	353 456 497	185 241 240	– – –	– – –	---------- ---------- ----------
AM3	2,020 2,457 2,250	1,489 1,727 1,647	*IGLV1-47* – –	*IGLJ1* – –	AAWDGSLSGYV ---------- ----------	288 1,060 1,069	226 931 979	*IGHV3-23* – –	*IGHJ3* – –	AKDISYDDILPGYPADAFDI -------------------- --------------------
AM4	994 788 989	817 659 807	*IGLV2-14* – –	*IGLJ1* – –	SSFTSSSSYV ---------- ----------	614 837 819	269 300 291	*IGHV3-23* – –	*IGHJ6* – –	AKRPPYSIYGMDV ------------- -------------
AM5	577	266	*IGLV2-14*	*IGLJ1*	SSYTITNTLV	154 191 113	151 187 110	*IGHV6-1* – –	*IGHJ3* – –	ARTHDDFWSGAFDI -------------- --------------
AM6	1,316	919	*IGKV3-20*	*IGKJ4*	QQYGTSPLT	2,491 2,609 2,770	1,960 2,072 2,199	*IGHV3-23* – –	*IGHJ4* – –	TRKGGYFDF --------- ---------
AM7	475	223	*IGLV6-57*	*IGLJ3*	QSYQGSSGV	828 791 843	735 702 755	*IGHV4-59* – –	*IGHJ4* – –	ARDKYSRGGEYVN ------------- -------------

Table II-2: Detailed results of antibody gene repertoire sequence analysis experiments for bone marrow aspirate specimens from seven patients with AL amyloidosis.

* The – symbol indicates the genetic feature is identical to that of the sequence above. Bolded features show consistency of the dominant clone within one individual over a time of 3 years.

We constructed a multiple sequence alignment using full length sequences from the dominant V3J clonotype within each donor, and for the three timepoints belonging to subject AM2. The V and J germline gene assignments of these light chain sequences were inferred from IgBlast, and aligned with the sequences from each donor. Within each donor, there are consistent somatic mutations throughout the length of the sequence persisting within most of the somatic variants of the dominant V3J clone. In AM1, these mutations include P8S, R19T, T21A in FR1,

K30T in CDR1, K38T, V44L in FR2, F61L in FR3, and S92R, H96R in CDR3.In AM2, throughout all three time points, these mutations include: A28D, K31T and Q32K in CDR1, Y50H IN FR2, K51R and S53T in CDR2, P60S, V72A and Y89H in FR3, V100E, an valine insertion at 101, and V102I in the CDR3. There does not seem to be an ongoing hypermutation process during the period of three years when the samples were obtained from AM2. In AM3, these mutations include S25A in FR1, S31I in CR1, Q39H, L40F, T43A, K46N, and Y50H in FR2, G65A and S73F in FR3, D94G in CDR3, and T108I in FR4. In AM4, these mutations include S27G and G31T in CDR2, L48V in FR2, E52A in CDR2, Y93F and T98S in CDR3, and T103P in FR4. In AM5, these mutations include Y32D in CDR1, Y34H and G43D in FR2, S55T in CDR2, G60W, S62P, N63D, G78D, A90T, and Y97F in FR3, S103I, S104T, S105N, Y108L in CDR3, and T102S, K105R in FR4. In AM6, these mutations include R18G, T20N in CDR1, S28T, V29I and S32N in CDR1, A35T and P41F in FR2, A52Sin CDR2, S64T, S87N, V86M in FR3, and S94T in CDR3. In AM7, these mutations include T19I in FR1, S31T in CDR1, N52D in CDR2, S69R in FR3, D95Q, S96G, N98S, and W100G in CDR3. Additionally, every subject's dominant clone CDR3 was mutated by at least one amino acid residue.

Gene usages of every dominating clone

Heat maps were constructed showing the different V-J gene combinations used in the kappa (**Figure II-2A**) or lambda repertoire (**Figure II-2B**) for each subject. The number of sequences observed with the same V-J gene combinations are depicted in each heat map as a Z-score, normalized for each subject. The dominant clone described previously (**Table II-1** and **Figure II-1**), could be recognized easily on the heat maps in **Figure II-2** (highlighted in the figure by black boxes), with the number of sequences containing the V-J combination denoted above each box. Also, we tracked the V-J combination containing the dominant clone in subject AM2

over the different time points in which biopsies were drawn (designated AM2.1, AM2.2 or AM2.3 in **Figure II-3**). The pathogenic clone was easily distinguished at all time points.

Every sample contains its respective highly used V and J genes, including the healthy sample. However, despite the healthy control sample having some more highly used IGLV/IGLJ combinations, the healthy control samples do not have any light chain expanded clonotypes (V-J-CDR3). The highest representation of a clonotype within the healthy repertoires is at 0.6% of the repertoire, present in HIP3 (**Figure II-4**). Additionally, out of all the amyloid patient samples, the dominant clone with the lowest representation is AM4, at about 1.6%, three-fold higher than what was seen in HIP3 (**Figure II-4**). This shows that although there are some light chain V-J combinations that are more highly used than others, there is a need to go one level deeper, and look at the V-J-CDR3 clonotype to determine the dysplastic clone present in each patient. The abundance of the top clonotypes within each donor will reveal the genotype of the dysplastic clones.

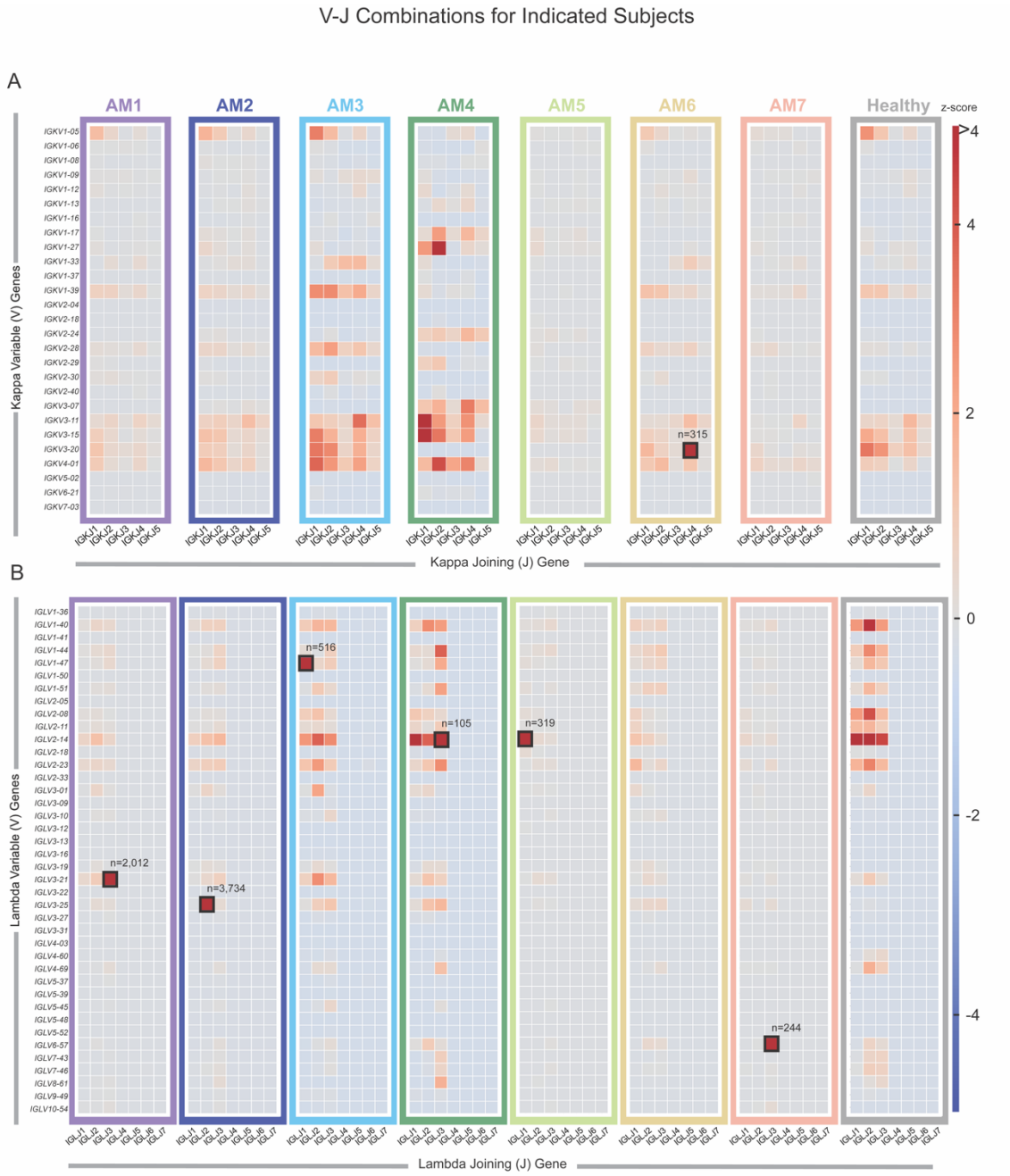


Figure II-2: Light chain inferred V and J germline gene usage from repertoires belonging to subjects with amyloidosis.

(A) Heatmap based on Vκ/Jκ germline gene usage. **(B)** Heatmap based on Vλ/Jλ germline gene usage. The frequency counts were derived from the total number of unique V3J clonotypes from each repertoire. The V/J frequency counts were transformed into a Z-score by first subtracting away the average frequency and then normalizing by the standard deviation of each subject. The colored box around each individual heatmap denotes repertoire data from each individual donor. The number of unique somatic variants for each dominant clonotype is indicated by a black box. For comparison we also included sequencing data from 3 healthy subjects from the Human Immunome Project (designated HIP1, HIP2 or HIP3). These data sets were combined and appear as the “healthy” dataset on the plot.

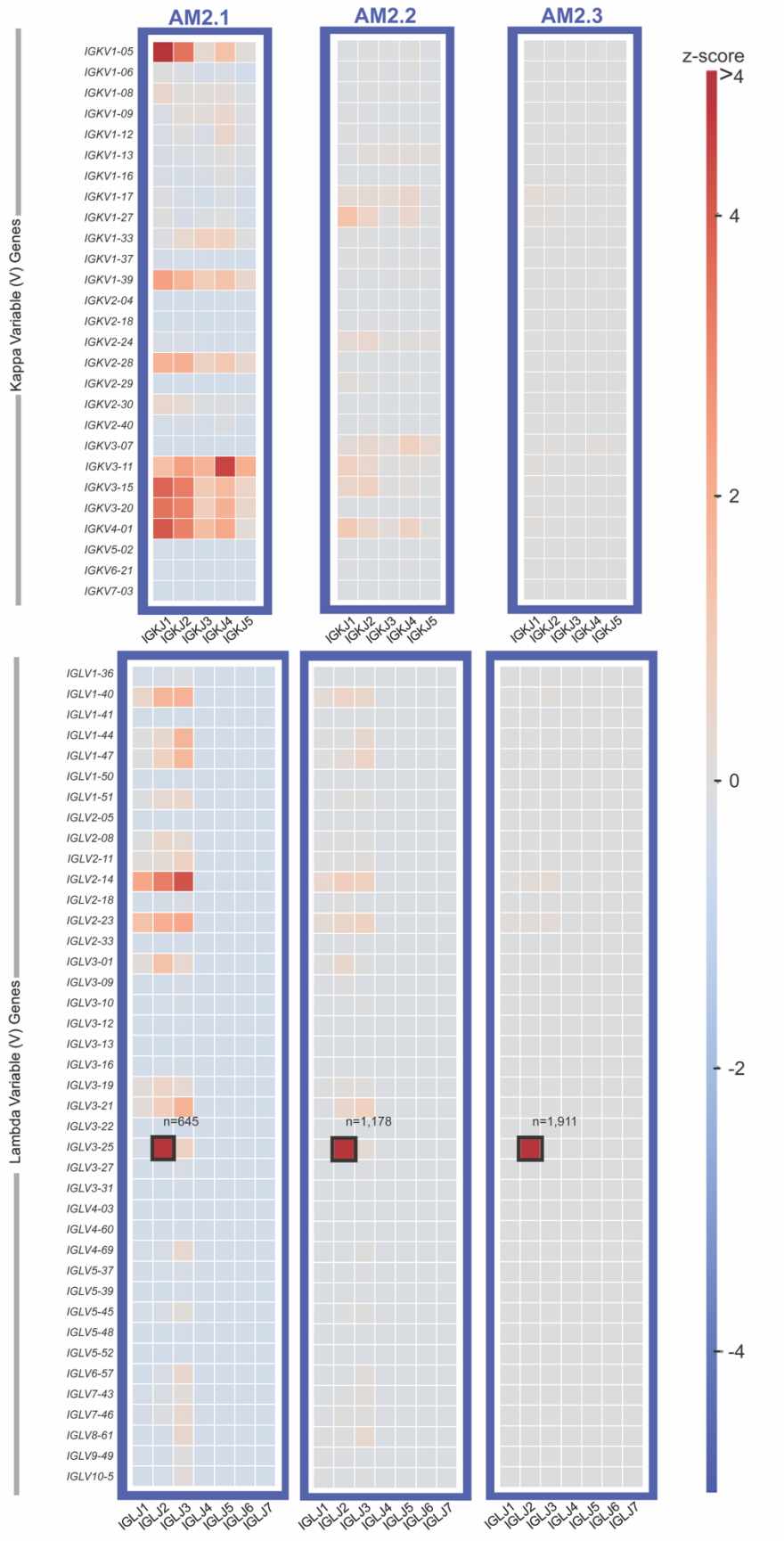


Figure II-3: Light chain inferred V and J germline gene usage in the repertoire of subject AM2 over time.

Heatmaps were generated from each of three time points at which a biopsy was taken. The heatmaps are based on Vκ/Jκ germline gene usage (upper panel) or Vλ/Jλ germline gene usage (lower panel). The V/J frequency counts were transformed into a Z-score by first subtracting away the average frequency and then normalizing by the standard deviation of each subject. The colored box around each individual heatmap denotes repertoire data from each individual donor. The number of unique somatic variants for each dominant clonotype is indicated by a black box.

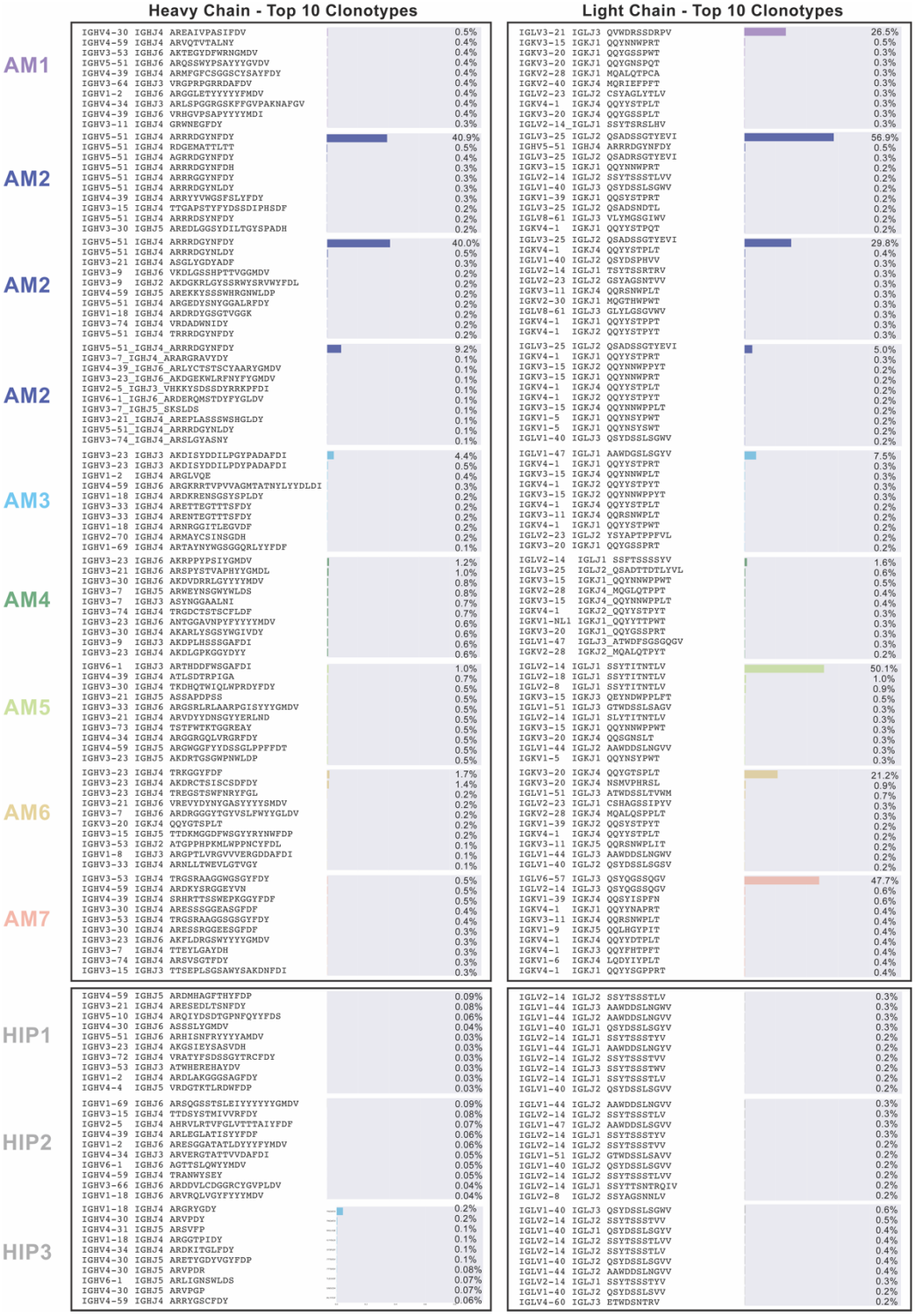


Figure II-4: The ten most abundant V3J clonotypes ordered by the number of unique somatic variants.

(A) Light chain V3J clonotypes **(B)** Heavy chain V3J clonotypes. Percentages were obtained by dividing the total number of somatic variants containing the V3J clonotype with the total number of somatic variants for the entire repertoire for each donor. The somatic variant count for the most prevalent V3J clonotype in each donor is shown at the top of each graph.

Searching for TCR V3J Clonotypes in Genbank

In addition to studies on identifying sequence features to AL amyloidosis, I assisted in a project identifying the diversity of T-cell receptors (TCR) in the lab. The collection of TCRs generated by somatic recombination is large but unknown. Therefore, we generated large TCR repertoire datasets as a resource to facilitate detailed studies of the role of TCR clonotypes and repertoires in health and disease. We estimated the size of individual human recombined and expressed TCRs by sequence analysis to contain between 5 and 21 million TCR clonotypes. We had sought to determine the biological properties of the TCR clonotypes sequenced. As the sequencing on the TCR was done using bulk sequencing, we were unable to functionally test sequenced identified. Therefore, I used data from public repositories containing functionally characterized paired TCR sequences for comparison. Previous studies have shown that mining next-generation sequencing of B cell repertoires can lead to discovery of sequences with a high degree of similarity to functionally known antibodies(Kovaltsuk et al., 2018; Krawczyk et al., 2019). Here, I searched GenBank (Clark et al., 2016) for TCRb V3J clonotypes that matched to those in our HIP subjects using exact matching criterion that required the V germline gene, J germline gene, and CDR3 amino acid sequences to be identical between clonotypes from GenBank and those sequenced from our HIP subjects. I found 557 TCRb clonotype matches in GenBank, with 134 matches associated with patented sequences, 49 matches associated with recognition of epitopes from viral pathogens, 8 matches associated with bacterial pathogens, 249 matches associated with autoimmune disorders, 66 matches associated with cancer, and 51 matches associated with other diseases (Figure II-5A). I then selected full-length sequences from a handful of these matches and generated alignments between the sequences from GenBank and those obtained from our HIP subjects (Figure II-5B). The sequences derived from gDNA sequencing were not of full length, and for clarity, used amino acids from the closest-matching germline gene to fill in the missing framework regions.

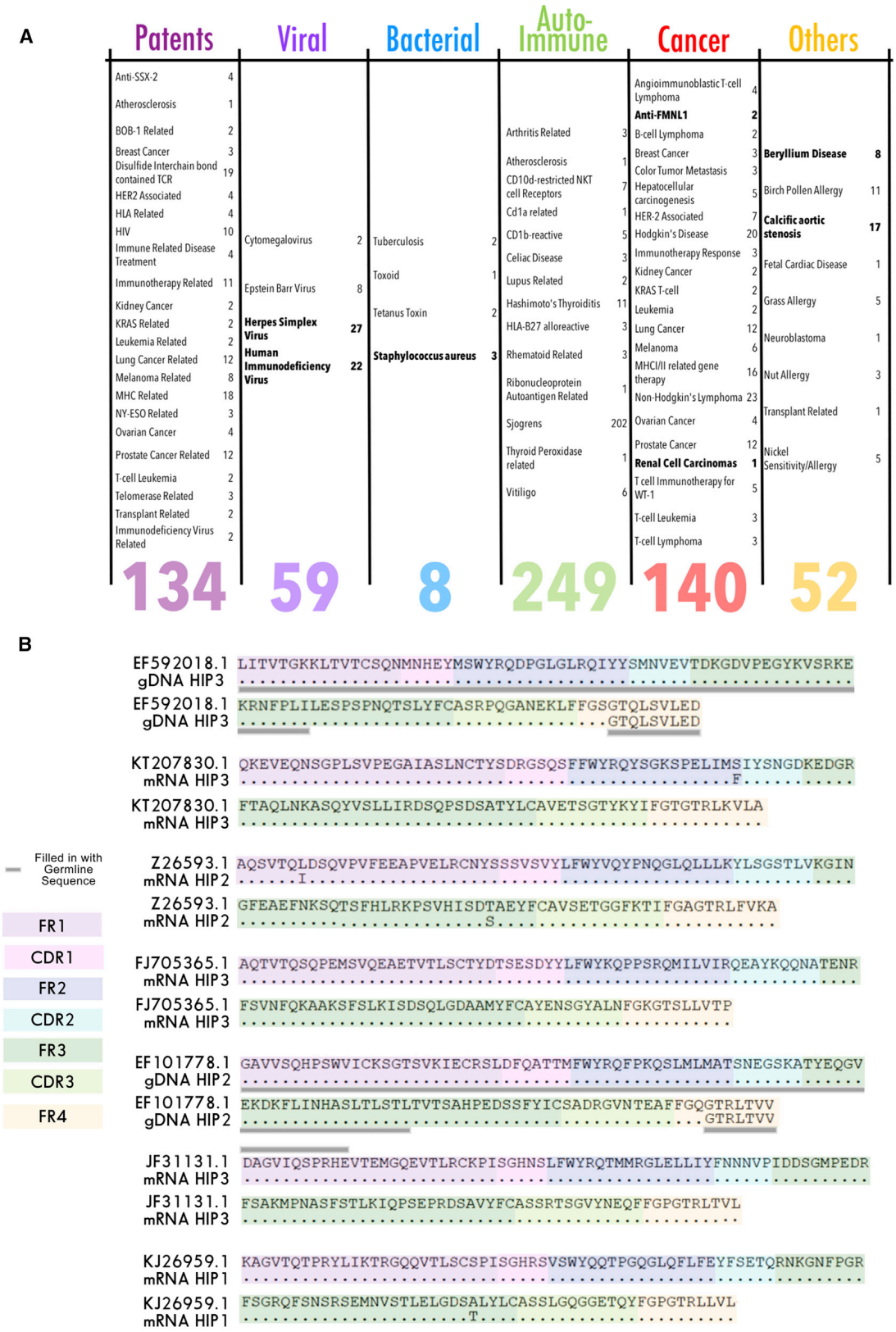


Figure II-5: TCRb Clonotypes from HIP Subjects Appearing in GenBank
(Legend on next page)

Figure II-5 legend: TCRb V3J clonotypes from HIP subjects were used to search against the entire GenBank database for possible matches. (A) Exact matches were grouped either as patented sequences or into one of five categories that focused on the target: viral target, bacterial target, autoimmune target, cancer target, and other. (B) Representative amino acid sequence alignments between the V region from GenBank and the V region from the HIP subject. In cases in which the sequence was missing in the framework region or regions, the closest-matching germline sequence was used to fill in the missing region. The filled-in portion of the sequence is highlighted by the thick gray line of the alignment.

The alignments show TCRb chains that bind to epitopes specific to viral antigens, such as herpes simplex virus 2, HIV gag protein, and influenza hemagglutinin (Table II-4). Thus, it appears possible to use the GenBank repository to infer possible specificities for several TCRb clonotypes derived from the HIP subjects.

Genbank Number	HIP Matches	Target[a]
EF592018.1	- HIP3 gDNA - HIP4 gDNA	HSV-2 VP22 49-57 epitope RPRGEVRFL
KT207830.1	- HIP3 mRNA	Specific for the HLA*02-restricted HIV-1 Gag SL9 epitope (SLYNTVATL; HIV_{HXB2} amino acid position 77–85)
Z26593.1	- HIP2 mRNA	CL-1 is a human T cell clone that is specific for a peptide derived from influenza hemagglutinin (HA 307-31 9) presented in the context of HLA-DR1.
FJ795365.1	- HIP3 mRNA	HLA-A2 specific and binds to peptide derived from FMNL1
EF101778.1	- HIP2 gDNA - HIP4 gDNA - HIP3 gDNA - HIP5 gDNA	Renal tumor antigens
JF731131.1	- HIP3 mRNA - HIP1 gDNA - HIP2 gDNA - HIP4 gDNA	The peptide bound is from antigens expressed on the tricuspid valve leaflet and should be on a cytoplasmic origin
KJ026959.1	- HIP1 mRNA, and gDNA - HIP2 mRNA, and gDNA - HIP3 mRNA, and gDNA - HIP4 gDNA - HIP5 gNDA - HIP6 gDNA	Be-loaded HLA-DP2–mimotope-2 (FWIDLFETIG) tetramer

[a]Citations for publications can be found in the Genbank information

Table II-3: Possible functional characterization for selected HIP TCR V3J clonotypes. TCR clonotypes identified in Genbank with previous literature validated

Sequence features of SARS-CoV-2 specific antibodies

During the beginning of the COVID-19 pandemic, 389 recombinant SARS-CoV-2-reactive human mAbs were isolated from a pooled sample of convalescent donors. B cells were antigenically tagged with SARS-CoV-2 S2P and RBD fused with mouse Fc. Reactive antibodies were then expanded, then split into two workflows. Half of the cells were sequenced using the 10X chromium single cell encapsulation automated system, and the other half was loaded onto a Berkeley Lights Beacon optofluidic instrument to test for antigen reactivity before exporting antigen reactive B cells from the instrument for sequencing.

I examined the sequences for the 389 antibodies to assess the diversity of antigen-specific B cell clonotypes discovered. The analysis showed that among the 389 mAbs, 321 unique amino acid sequences were present and 313 unique V_H–J_H–CDRH3–V_L–J_L– CDRL3 clonotypes were represented, with diverse usage of antibody variable genes (**Figure II-6A**). The length distributions of CDR3 amino acids in the heavy and light chains were typical of human repertoires (**Figure II-6B**) (Soto et al., 2019). The high relatedness of sequences to the inferred germline variable genes observed for this panel of antibodies (**Figure II-6C**) contrasts with the much higher levels of somatic mutation seen in B cell recall responses against common human pathogens such as influenza (Wrammert et al., 2011). These data suggest that the SARS-CoV-2 antibodies were likely induced during the primary response to SARS-CoV-2 infection and not by a recall response to a distantly related seasonal coronavirus.

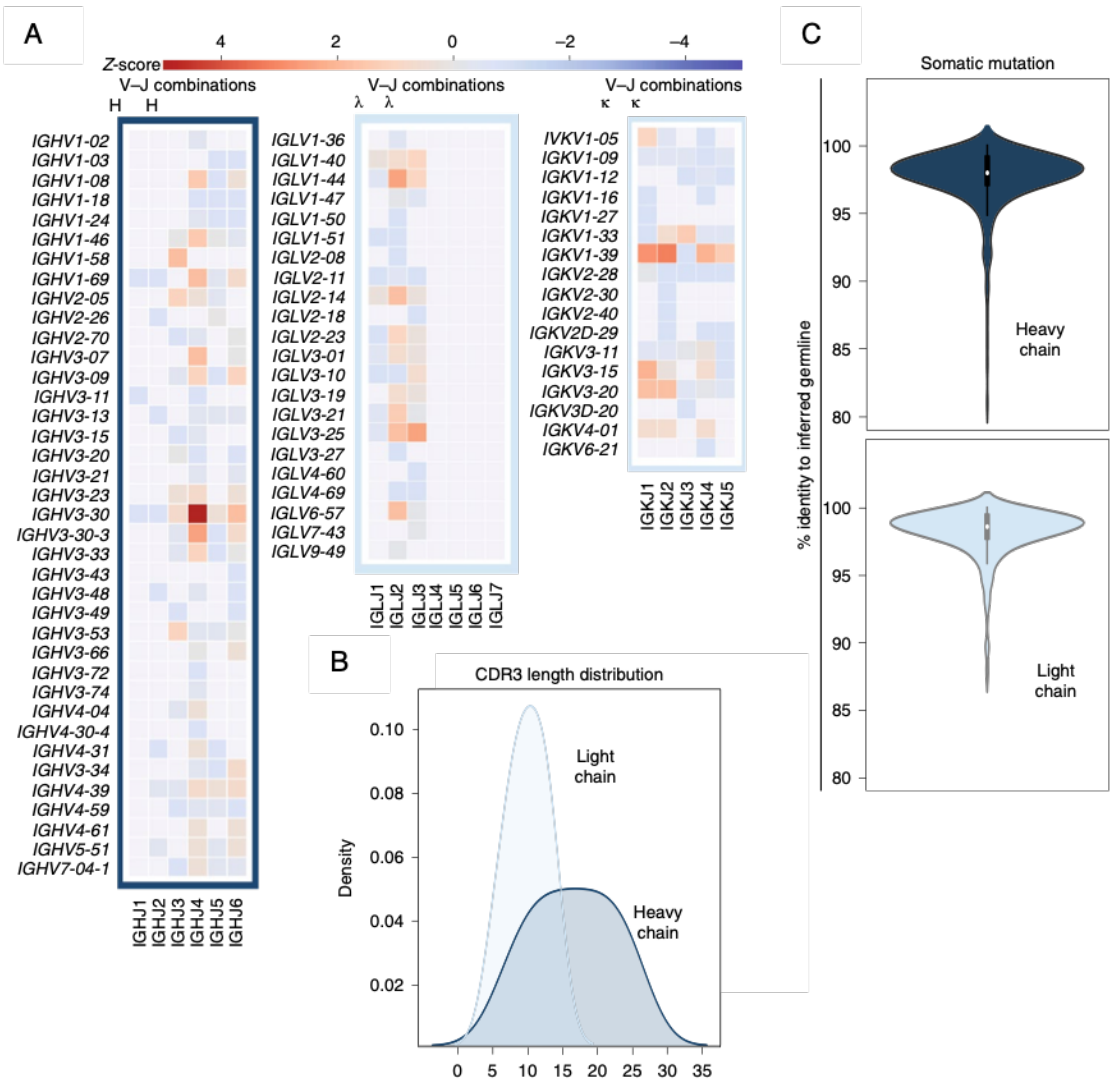


Figure II-6: Sequence reactivity of 389 human antibodies to SARS-CoV-2

(A) Heatmap showing usage of antibody variable-gene segments for variable (*V*) and joining (*J*) genes. Of the 389 antibodies tested in **b** and **c**, 324 were found to have unique sequences, and those unique sequences were analyzed for genetic features. The frequency counts are derived from the total number of unique sequences with the corresponding *V* and *J* genes. The *V*/*J* frequency counts then were transformed into a *Z*-score by first subtracting the average frequency, then normalizing by the s.d. of the set of antigen-reactive mAbs. red denotes more common gene usage, and blue denotes less common gene usage. **(B)** CDr3 amino acid length distribution. The CDr3 of each sequence was determined using PyIR software. The amino acid length of each CDr3 was counted. The distribution of CDr3 amino acid lengths for heavy or light chains then was plotted as a histogram and fitted using kernel density estimation for the curves. **(C)** Divergence from inferred germline gene sequences. The number of mutations of each mAb relative to the inferred germline variable gene was counted for each clone. These numbers then were transformed into percent values and plotted as violin plots. For the heavy chain, values range from 81 to 100, with a median of 98, a 25th quartile of 97.3 and a 75th quartile of 99. For the light chain, values range from 87.5 to 100, with a median of 98.6, a 25th quartile of 97.9 and a 75th quartile of 99.3.

DISCUSSION

This study on identifying genetic signatures of dysplastic clones in AL amyloidosis patients was one of the initial studies I worked on in the Crowe lab. Prior to joining the lab, I had no experience in coding in the python language and scripting for data analysis. Therefore, the project allowed for training and familiarization with analyzing antibody sequences. Gaining this expertise allowed me to apply these methods to analyze the antibody isolated to the SARS-CoV-2 spike protein during the beginning of the SARS-CoV-2 pandemic. Additionally, analysis of sequence features were also used for the identification of V3J clonotypes shared between TCR repertoire dataset generated in the lab with GenBank entries. These queries allowed for identification of public clonotypes, and the functionality of each clonotype.

Together, the data within the main project in this chapter, show that the dominant V3J clonotype in AL amyloidosis patients can be identified using relatively shallow mRNA-based NGS repertoire sequencing, and the mRNA for the dominant clone can be tracked easily in patients over time. Since we used both heavy and light chain antibody gene sequencing, we also could distinguish between patients who exhibited only a dominant light chain mRNA and those with coordinated over-expression of a single heavy and light chain mRNA. It will be of interest in the future to determine if these diverse genetic features correlate with differences in clinical outcomes or response to therapy.

However, despite these encouraging results, this study does have several limitations. The sample size is relatively small, limiting our ability to draw sweeping conclusions. Many patients had a relatively high burden of hematologic disease for a population of patients with AL amyloidosis. All samples used in this study were obtained at time points during which the patients had active disease. Therefore, additional studies will be needed to determine the sensitivity and specificity of this method for patients with a lower burden of disease. By acquiring bone marrow aspirates of patients ranging from high to low burden of disease, sensitivity can be assayed by

doing an end-point dilution until the assay can no longer detect the dysplastic clone after ranking the top ten clonotypes present in each patient. Specificity can be achieved by comparing results from using the mRNA sequencing method discussed in this paper in parallel with Adaptive Technology's clonoSeq test. The clonoSeq test is currently FDA approved for monitoring measurable residual disease (MRD) in both B-cell acute lymphoblastic leukemia as well as multiple myeloma, and therefore would be a good way to measure the specificity of the assay in detecting dysplastic clones. It would also be of interest to use this assay on samples from patients in apparent complete clinical remission to determine if this method can identify the subclinical persistence of residual clones in the bone marrow. Previously, we examined the expressed antibody variable gene repertoires from 10 different human tissues using RNA samples derived from a large number of individuals (see Briney *et al.*, (Briney et al., 2014)). Mao *et al.* also looked at B-cell responses in a variety of tissues (Briney et al., 2014). Based on our early work and the work of others, it should be possible to detect differences in clonal B-cell populations within different tissues from individuals with AL. Moreover, we should be able to detect differences in specific features of the repertoire between diseased and non-diseased tissues (*e.g.*, CDR3 length or somatic mutation). Whether or not these differences are statistically significant when compared to sequencing from whole blood needs to be studied in more detail. Assuming we could obtain tissue samples from the organs of patients with and without AL, we see no limitation in applying our method to much larger sample size. With potential access to a large bank of amyloidogenic tissues, this technology can be applied to investigate sequence features of light chains that are prominent for aggregating in certain tissues. Studies like this will be informative in patients that have amyloid deposits in less high-risk organs, therefore subjecting them to close monitoring instead of toxic treatments. Despite these limitations of the current study, the data demonstrate the straightforward use of shallow mRNA-based antibody variable gene sequence analysis on small samples to identify the recombined genes distinguishing a clonal population of plasma cells

for each of the patients under study. The data also offer intriguing snapshots of diversity in genetic features of the clones and repertoires in AL amyloidosis patients that were not previously appreciated. There are many methods being developed to measure minimal residual disease, especially for multiple myeloma. Both flow cytometry techniques and next-generation sequencing techniques have proven to have great specificity. However, the main differentiator between the two methods is that the flow cytometry assay requires fresh samples and immediate processing after biopsy. Additionally, flow cytometric studies are often operator-dependent. Therefore, this study is valuable in establishing another assay option for assessing minimal disease, especially for light chain amyloidosis. This proof-of-concept study suggests it will be of value to apply this method in a larger population of patients with a lower burden of disease to define the role of somatic variation and coordinated heavy and light chain over-expression in pathogenesis and response to therapy.

The main limitation that comes with bulk next-generation sequencing is the inability to functionally validate clones. Therefore, we searched through the Genbank repository to find previously reported clonotypes that are identical to those in our experimental repertoires. Many clonotypes for which we found matches have been validated functionally. Matching clonotypes appear in diverse patents and many published academic studies, especially cancer-related studies. Comparing these data with those from studies of single T cell transcriptomics with linked TCRa and TCRb-chain repertoires will be of interest for achieving a deeper understanding of human immune responses. Additionally, this method may be useful when identifying if certain biologics have been patented in the past before pursuing a drug discovery effort.

From identifying genetic features of antibodies isolated to SARS-CoV-2 spike protein, we found that there was a high relatedness of sequences to inferred germline variable genes suggesting that the responses in the antibodies isolated were primary responses to SARS-CoV-2 and not through a recall response to distantly related coronaviruses. Additionally, the high

identity to germline hints that it does not take multiple rounds of somatic hypermutation to achieve recognition to the SARS-CoV-2 spike, therefore inferring that a productive antibody response may be mounted quickly after initial antigen exposure. Future work on tracking how these lineages evolve can give insight into rational vaccine design.

METHODS

Bone marrow aspirate B-cell sequencing. Bone marrow samples were obtained from patients through the Vanderbilt University Medical Center (VUMC) hematologic malignancy tissue bank after written informed consent was obtained. The study was approved by the Vanderbilt University Medical Center Institutional Review Board. Mononuclear cells were isolated from the bone marrow aspirate samples from seven subjects (designated AM1, AM2, AM3, AM4, AM5, AM6, or AM7) with AL amyloidosis with or without multiple myeloma; samples collected at three separate timepoints for subject AM2 were included. Approximately 2×10^6 mononuclear cells per donor sample that had been cryopreserved previously were thawed, washed and counted with a viability dye. The estimated number of B cells in the sample was determined by multiplying the total viable mononuclear cell count in the sample by the percent of plasma cells in each subject. The cells were pelleted in a tabletop centrifuge prior to isolation of total cell RNA from the resulting pellet using the RNeasy Mini Kit (Qiagen), according to the manufacturer's recommendations. Each resulting pool of purified RNA was split equally by volume to create aliquots for up to three technical replicates for subsequent target enrichment using a previously described 5'RACE enrichment approach (Turchaninova et al., 2016). Briefly, a cDNA synthesis primer mix (10 µM each) was combined with template RNA and incubated at 70°C for 2 min, and then the incubation temperature was decreased to 42°C to anneal the synthesis primers (5 min). Primer annealed template RNA then was mixed with 5X First-strand buffer (Clontech), DTT (20 mM from the SMARTScribe synthesis kit; Clontech), 5' template switch oligo (TSO) adapter containing a Unique Molecular Identifier (10 µM), dNTP mix (10 mM each), 20 units RNAse inhibitor (RNAsin, Promega) and 10X SMARTscribe Reverse Transcriptase (Clontech) prior to incubation at 42°C for 60 min. Immediately following reverse transcription, first-strand cDNA was purified using AMPure SPRIselect beads (Beckman Coulter) at a ratio of 1X per volume RT reaction. Purified cDNA was PCR amplified by mixing

Table S2. Primers used for next generation sequencing		
Primer	**Application**	**Sequence**
First-strand cDNA synthesis		
SmartNNNext	5′ – template-switch oligo with sequencing illumina adapter. U = dU, rG-riboG	AGATGUGTAUAAGAGACAGNNNNUNNNNUNNNNUCTT(rG)$_4$
Human IGH cDNA synthesis primer mix		
hIGG_r1	Primer for cDNA synthesis, human IgG heavy-chain mRNA	GAAGTAGTCCTTGACCAGGCA
hIGM_r1	Primer for cDNA synthesis, human IgM heavy-chain mRNA	GTGATGGAGTCGGGAAGGAAG
hIGA_r1	Primer for cDNA synthesis, human IgA heavy-chain mRNA	GCGACGACCACGTTCCCATCT
hIGD_r1	Primer for cDNA synthesis, human IgD heavy-chain mRNA	GGACCACAGGGCTGTTATC
hIGE_r1	Primer for cDNA synthesis, human IgE heavy-chain mRNA	AGTCACGGAGGTGGCATTG
Human IGL cDNA synthesis primer mix		
hIGLC_r1	Primer for cDNA synthesis, human IgL light-chain mRNA	GCTCCCGGGTAGAAGT
hIGKC_r1	Primer for cDNA synthesis, human IgK light-chain mRNA	GCGTTATCCACCTTCC
First PCR amplification		
Common primer	Step-out primer, anneals on the switch adaptor	AGATGTGTATAAGAGACAG
Human IGH reverse primer mix		
Common-hIGGE_r2	Nested primer with sequencing illumina adaptor, human IgG/IgE heavy-chain cDNA	AGATGTGTATAAGAGACAGARGGGGAAGACSGATG
Common-hIGA_r2	Nested primer with sequencing illumina adaptor, human IgA heavy-chain cDNA	AGATGTGTATAAGAGACAGCAGCGGGAAGACCTTG
Common-hIGM_r2	Nested primer with sequencing illumina adaptor, human IgM heavy-chain cDNA	AGATGTGTATAAGAGACAGAGGGGGAAAAGGGTTG
Common-hIGD_r2	Nested primer with sequencing illumina adaptor, human IgD heavy-chain cDNA	AGATGTGTATAAGAGACAGATATGATGGGGAACAC
Human IGL reverse primer mix		
Common-hIGL_r2	Nested primer with sequencing illumina adaptor, human IgL heavy-chain cDNA	AGATGTGTATAAGAGACAGGYGGGAACAGAGTGAC
Common-hIGK_r2	Nested primer with sequencing illumina adaptor, human IgK heavy-chain cDNA	AGATGTGTATAAGAGACAGGATGGTGCAGCCACAG
Second PCR amplification		
F-common	Step-out primer with sequencing and P7 illumina adapters	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG
R- common	Step-out primer with sequencing and P5 illumina adapters	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG
Third PCR amplification		
Fc_i7	Step-out primer with index 1 illumina adapter	CAAGCAGAAGACGGCATACGAGAT[i7]GTCTCGTGGGCTCGG
Fc_i5	Step-out primer with index 2 illumina adapter	AATGATACGGCGACCACCGAGATCTACAC[i5]TCGTCGGCAGCGTC

Table II-4: Primers used for next generation sequencing.

Primers used for each step of library prep and sequencing are listed. All primers are a modified version that was provided by Chdakov at the time of sequencing to incorerate Illumina Nextera adapters instead of Truseq adapters [10].

with 2X Q5 Master Mix (NEB), dNTP solution (10 mM each), universal forward primer extending off the TSO (10 µM), gene-specific first PCR reverse primer mix (10 µM). PCR 1 cycling was performed as follows: 95°C for 1 min 30s followed by 18 cycles of 95°C for 10s, 60°C for 20s and 72°C for 40s, followed by a final extension of 72°C for 4 min. First-round PCR products were purified using AMPure SPRIselect beads at a ratio of 0.6X per volume PCR 1 reaction (Beckman Coulter). PCR 2 cycling was performed identically to PCR 1 except using extension primers with sample indexes and performing 14 cycles for PCR 2. Final PCR products were purified using AMPure SPRIselect beads at a ratio of 0.6X per volume RT reaction (Beckman Coulter). Concentrations were determined by fluorometric quantitation (Qubit) and pooled by

combining equimolar portions of each individual sample. Pooled PCR products were used to generate a sequencing library using the NEBNext Ultra Library Prep Kit for Illumina sequencing by applying standard protocol according to the manufacturer's recommendations. Libraries were sequenced on an Illumina MiSeq platform using MiSeq Reagent Kit V3 (600 cycles) and symmetric PE-300 sequencing. All primers used in this protocol were modified from a previously described approach(Turchaninova et al., 2016) and are listed (**Table II-4**).

Bioinformatic sequence processing. The bioinformatics processing of all NGS data was performed using the PyIR sequence processing pipeline that uses IgBLAST(Ye et al., 2013). Briefly, we merged all paired-end (PE) reads to generate full-length contigs using the program USEARCH v9.18(Edgar, 2010). The overlap region (*-fastq_minovlen*) was set to 15 nucleotides, and the maximum number of differences in the overlap region (*-fastq_maxdiffpct*) was set to 10. All merged reads were filtered in the following order using our MongoDB database: (1) Removal of any read that had an average Phred score of less than 30; (2) Removal of any read that had an E-value larger than 10^{-6} for *IGHV*/*IGHJ* germline assignments; (3) Removal of any read that did not have a defined CDR3; (4) Removal of any read containing a stop codon; (5) Removal of any read that was out of frame at the junction region; and (6) Removal of reads for which the nucleotide length from framework 1 through framework 3 was less than 250 nucleotides. All remaining reads were labeled as productive reads (but not error-corrected). To correct for any sequencing errors, we binned all raw reads associated with a productive read, using a Universal Molecular Identifier (UMI), CDR3 length (in amino acids) and variable (V) and joining (J) gene assignment (ignoring allele). The grouped raw reads were re-oriented from 5' to 3' and then used to generate a consensus sequence based on a the most frequent nucleotide at each position. The consensus reads from each grouping were then processed again using our PyIR pipeline and subjected to the same set of filters already

described above. All remaining error-corrected reads were labeled productive and used for all analysis. Authors E.C.C. and C.S. analyzed the data, and all authors had access to primary data. **Data sharing statement.** The dataset(s) used in this article are available in the Sequence Read Archive (SRA, https://www.ncbi.nlm.nih.gov/sra) under Bioproject number PRJNA637633.

Research participants for SARS-CoV-2 study. We studied four patients in North America with recent laboratory-confirmed symptomatic SARS-CoV-2 infections that were acquired in China. The studies were approved by the Institutional Review Board of Vanderbilt University Medical Center, and subsite studies were approved by the Institutional Review Board of the University of Washington or the Research Ethics Board of the University of Toronto. Samples were obtained after written informed consent. Patient 1 (35-year-old male) was the earliest reported case of SARS-CoV-2 infection in the United States, who presented with disease in Seattle, Washington, on 19 January 2020, a blood sample was obtained for study on 19 February 2020. Patient 2 (52-year-old female) was infected following close exposure in Beijing, China, to an infected person from Wuhan, China, during the period between 23 January 2020 to 29 January 2020. She presented with mild respiratory disease symptoms from 1 February 2020 to 4 February 2020 that occurred after travel to Madison, Wisconsin. She obtained a diagnosis of infection by testing at the US Centers for Disease Control on 5 February 2020. Blood samples were obtained for study on 7 March 2020 and 8 March 2020. Patient 3 (a 56-year-old male) and patient 4 (a 56-year-old female) are a married couple and residents of Wuhan, China, who traveled to Toronto, Canada, on 22 January 2020. Patient 3 first developed a cough without fever on 20 January 2020 in the city of Wuhan, where he had a normal chest X-ray on that day. He few to Canada with persisting cough and arrived in Canada on 22 January 2020, where he became febrile. He presented to a hospital in Toronto, 23 January 2020 complaining of fever, cough and shortness of breath; a nasopharyngeal swab was positive by PCR testing for SARS-CoV-2. His chest X-ray at that time was abnormal, and he was admitted

for non-intensive-care-unit impatient care. He improved gradually with supportive care, was discharged 30 January 2020 and rapidly became asymptomatic except for a residual dry cough that persisted for a month. He had a negative nasopharyngeal swab PCR test on 19 February 2020. Patient 4 is the wife of patient 3 who traveled with her husband from Wuhan. She was never symptomatic with respiratory symptoms or fever but was tested because of her exposure. Her nasopharyngeal swab was positive for SARS-CoV-2 by PCR, on 24 January 2020; repeat testing in follow-up on 21 February 2020 was negative. PBMCs were obtained by leukapheresis from patients 3 and 4 on 10 March 2020, which was 50 d since symptom onset of patient 3. Samples were transferred to Vanderbilt University Medical Center in Nashville, Tennessee, on 14 March 2020.

Patient selection and target-specific memory B cells isolation for SARS-CoV-2 study. B cell responses to SARS-CoV-2 in PBMCs from a cohort of four patients with documented previous infection with the virus were analyzed for antigen specificity, and PBMCs were used for SARS-CoV-2-specific B cell enrichment. The frequency of SARS-CoV-2 S protein-specific B cells was identified by antigen-specific staining with either biotinylated S2Pecto or RBD–mFc protein. Briefly, B cells were purified magnetically (STEMCELL Technologies) and stained with anti-CD19–APC (BioLegend clone HIB19 cat. no. 982406, lot B270238, 1:10 dilution), anti-IgD– FITC (BioLegend clone IA6–2, cat. no. 348206, lot B258195, 1:20 dilution), and anti-IgM–FITC (BioLegend clone MHM-88, cat. no. 314506, lot B218736, 1:20 dilution) phenotyping antibodies and biotinylated antigen. A DAPI stain was used as a viability dye to distinguish dead cells. Antigen-labeled class-switched memory B cell–antigen complexes (CD19+IgM– IgD–Ag+DAPI–) were detected with a R-PE-labeled streptavidin conjugate (Thermo Fisher Scientific, S866, 1:500 dilution). After identification of the two patients with the highest B cell response against SARS-CoV-2 (patients 3 and 4), target-specific memory B cells were isolated by flow-cytometric sorting using an SH800 cell sorter (Sony) from pooled

PBMCs of these two patients, after labeling of B cells with either biotinylated S2Pecto or RBD–mFc proteins. Flow-cytometric data were analyzed with the SH800 software and FlowJo version 10 (Tree Star). Overall, from >4 ×108 PBMCs, 2,126 RBD–mFc-reactive and 5,544 S2Pecto-reactive B cells were sorted and subjected to further analysis. Several methods were implemented for the preparation of sorted B cells for sequencing. Approximately 4,500 sorted cells were subjected to direct sequencing immediately after flow cytometric sorting. The remaining cells were expanded in culture for 8 d in the presence of irradiated 3T3 feeder cells that were engineered to express human CD40L, IL-21 and BAFF, as described previously22. The expanded lymphoblastoid cell lines (LCLs) secreted high levels of S-protein-specific antibodies, as confirmed by ELISA, to detect antigen-specific human antibodies in culture supernatants. Approximately 40,000 expanded LCLs were sequenced using the Chromium sequencing method (10x Genomics).

Microfluidic device selection of single antigen-specific B cells for SARS-CoV-2 study. Activated memory B cells were screened using Berkeley Lights' Beacon optofluidic system. Purified B cell samples were imported automatically onto OptoSelect 11k chips in a novel plasmablast survival medium that promotes antibody secretion and preserves cell viability44. Single-cell penning was then performed using OEP technology, in which light is used to transfer B cells into individual nanoliter-volume chambers (NanoPens). Using this light-based manipulation, thousands of LCLs were transferred into pens across multiple chips in each workflow. We performed an on-chip, fluorescence-based assay to identify antibodies that bound SARS-CoV-2 S2Pecto or RBD–mFc protein. We prepared 6- to 8-μm and 10- to 14-μm RBD–mFc-conjugated beads by coupling biotinylated RBD–mFc protein to streptavidin-coated polystyrene particles (Spherotech). We prepared 6- to 8-μm S2Pecto protein-conjugated beads by coupling full-length S2Pecto protein to streptavidin-coated polystyrene particles. Assays consisted of mixing beads conjugated with the RBD–mFc or S2Pecto proteins with Alexa Fluor

(AF)-labeled anti-human secondary antibodies (Thermo Fisher Scientific cat. no. A-21090) at a 1:100 dilution and importing this assay mixture into OptoSelect 11k chips. Antigen-specific antibodies bound the antigen-conjugated beads, which then sequestered the fluorescent secondary antibody. Cells secreting antigen-specific antibodies were identified by locating the NanoPens immediately adjacent to the fluorescent beads. We also performed an on-chip assay to select antibodies that blocked the interaction of hACE2 and the RBD of SARS-CoV-2. The blocking assay was performed by first co-incubating LCLs and 10- to 14-μm RBD–mFc-conjugated streptavidin-coated beads (Spherotech) in the NanoPen chambers to allow for secreted antibodies to saturate the antigen. Then, a mixture of recombinant hACE2 with a FLAG tag (Sigma-Aldrich cat. no. SAE0064), a rat anti-FLAG AF 647 antibody at a 1:50 dilution (BioLegend clone L5, cat. no. 637315, Lot B265929) and an anti-human IgG AF 568 antibody at a 1:100 dilution (Thermo Fisher Scientific cat. no. A-21090) was perfused through the OptoSelect 11k chip and allowed to diffuse into the NanoPen chambers. RBD-binding antibodies were identified by locating pens with RBD–mFc-conjugated beads that were fluorescent when imaged using the Beacon TRED filter cube. Simultaneously, hACE2 binding to the RBD-coated beads was detected using a Cy5 filter cube. NanoPen chambers containing RBD–mFc-conjugated beads with fluorescence in both filter cubes were classified as containing B cells secreting anti-SARS-CoV-2 antibodies that bound RBD but that did not demonstrate hACE2-blocking activity. NanoPen chambers that contained RBD–mFc-conjugated beads that were fluorescent in the TRED channel but non-fluorescent in the Cy5 channel contained secreted antibodies that had both bound RBD and blocked hACE2–RBD interaction. Antigen-specific cells of interest were exported from specific NanoPen chambers to individual wells of 96-well reverse transcription–PCR plates containing lysis buffer.

Sequencing and cloning of single antigen-specific B cells for SARS-CoV-2 study.

After export from the Beacon instrument, antibody heavy- and light-chain sequences for B cells secreting RBD–mFc- or S2Pecto-binding antibodies were amplified and recovered using components of the Opto Plasma B Discovery cDNA Synthesis Kit (Berkeley Lights). Antibody heavy- and light-chain sequences were amplified through a 5′ RACE approach using the kit's included 'BCR Primer 2' forward primer and isotype-specific reverse primers. The 5′-RACE-amplified cDNA was sequenced using the Pacific Biosciences Sequel platform using the SMRTbell Barcoded Adapt Complete Prep-96 kit (Pacific Biosciences) and a 6-h movie time. In a redundant sequencing approach, heavy- and light-chain sequences were amplified using a cocktail of custom V- and J-gene-specific primers (similar to previously described human Ig gene-specific primers45) from the original 5′-RACE-amplified cDNA while the products of the gene-specific amplification were sent for Sanger sequencing (GENEWIZ). The sequences generated by these two approaches were analyzed using our Python-based antibody variable-gene analysis tool (PyIR; https://github.com/crowelab/PyIR)46 to identify which V and J genes most closely matched the nucleotide sequence. Heavy- and light-chain sequences were then amplified from the original cDNA using cherry-picked V- and J-gene-specific primers most closely corresponding to the V and J gene calls made by PyIR. These primers include adapter sequences which allow Gibson-based cloning into a monocistronic IgG1 expression vector (pMCis_G1). Similar to a vector described below, this vector contains an enhanced 2A sequence and GSG linker that allows simultaneous expression of mAb heavy and light chain genes from a single construct upon transfection47. The pMCis_G1 vector was digested using the New England BioLabs restriction enzyme FspI, and the amplified paired heavy- and light-chain sequences were cloned through Gibson assembly using NEBuilder HiFi DNA Assembly Master Mix. After recovered sequences were cloned into pMCis_ G1 expression constructs, recombinant antibodies were expressed in CHO cells and were purified by affinity

chromatography as detailed below. Antigen-binding activity was confirmed using plate-based ELISA. Generation of antibody variable-gene libraries from single B cells. As an alternative approach, we also used a second major approach for isolation of SARS-CoV-2-reactive antibodies. In some experiments, the Chromium Single Cell V(D)J workflow with B-cell-only enrichment option was used for generating linked heavy-chain and light-chain antibody profiling libraries. Approximately 2,866 directly sorted S2Pecto or 1,626 RBD–mFc protein-specific B cells were split evenly into 2 replicates each and separately added to 50µl of RT Reagent Mix, 5.9µl of Poly-dt RT Primer, 2.4µl of Additive A and 10µl of RT Enzyme Mix B to complete the Reaction Mix as per the vendor's protocol, which then was loaded directly onto a Chromium chip (10x Genomics). Similarly, for the remaining sorted cells that were expanded in bulk, approximately 40,000 cells from 2 separate sorting approaches were split evenly across 4 reactions and processed separately as described above, before loading onto a Chromium chip. The libraries were prepared following the User Guide for Chromium Single Cell V(D)J Reagents kits (CG000086_REV K).

Generation of antibody variable-gene libraries from single B cells for SARS-CoV-2 study. As an alternative approach, we also used a second major approach for isolation of SARS-CoV-2-reactive antibodies. In some experiments, the Chromium Single Cell V(D)J workflow with B-cell-only enrichment option was used for generating linked heavy-chain and light-chain antibody profiling libraries. Approximately 2,866 directly sorted S2Pecto or 1,626 RBD–mFc protein-specific B cells were split evenly into 2 replicates each and separately added to 50µl of RT Reagent Mix, 5.9µl of Poly-dt RT Primer, 2.4µl of Additive A and 10µl of RT Enzyme Mix B to complete the Reaction Mix as per the vendor's protocol, which then was loaded directly onto a Chromium chip (10x Genomics). Similarly, for the remaining sorted cells that were expanded in bulk, approximately 40,000 cells from 2 separate sorting approaches were split evenly across 4 reactions and processed separately as described above, before

loading onto a Chromium chip. The libraries were prepared following the User Guide for Chromium Single Cell V(D)J Reagents kits (CG000086_REV K). Next-generation DNA sequence analysis of antibody variable genes. Chromium Single Cell V(D)J B-Cell-enriched libraries were quantified, normalized and sequenced according to the User Guide for Chromium Single Cell V(D)J Reagents kits (CG000086_REV C). The two enriched libraries from direct flow cytometric cell sorting were sequenced on a NovaSeq sequencer (Illumina) with a NovaSeq 6000 S1 Reagent Kit (300 cycles) (Illumina). The four enriched libraries from bulk expansion were sequenced on a NovaSeq sequencer with a NovaSeq 6000 S4 Reagent Kit (300 cycles) (Illumina). All enriched V(D)J libraries were targeted for sequencing depth of at least 5,000 raw read pairs per cell.

Bioinformatics analysis of single-cell sequencing data for SARS-CoV-2 study.

Following sequencing, all samples were demultiplexed and processed through the 10x Genomics Cell Ranger software (version 2.1.1). The down-selection to identify lead candidates for expression was carried out in two phases. In the first phase, all paired antibody heavy- and light-chain variable-gene cDNA nucleotide sequences obtained that contained a single heavy- and light-chain sequence were processed using PyIR. We considered heavy- and light-chain-encoding gene pairs productive and retained them for additional downstream processing if they met the following criteria: (1) did not contain a stop codon, (2) encoded an intact CDR3 and (3) contained an in-frame junctional region. The second phase of processing eliminated redundant sequences (those with identical amino acid sequences). Any antibody variant that was designated as an IgM isotype (based on the sequence and assignment using the 10x Genomics Cell Ranger V(D)J software, version 2.1.1) was removed from consideration (while IgG and IgA isotype antibodies were retained). The identities of antibody variable-gene segments, CDRs and mutations from inferred germline gene segments were determined by using PyIR.

Donors for the T-cell study. We leukapheresed five healthy HIV-negative adult subjects with no recently reported acute infections or vaccinations. Adult samples were obtained after informed consent from the Vanderbilt Clinical Trials Center. The study was approved by the Institutional Review Board of Vanderbilt University Medical Center (VUMC). The subjects consisted of two adult females (designated as subject HIP1 or HIP5) and three adult males (designated as subjects HIP2, HIP3 or HIP4) (Table 1). HLA typing (Mack et al., 2013; Robinson et al., 2003) was carried out for all HIP subjects (Table S2). Leukopaks containing large numbers of PBMCs obtained by leukapheresis were collected from all subjects at VUMC. Following leukapheresis, peripheral blood mononuclear cells (PBMCs) were isolated with Ficoll-Histopaque by density gradient centrifugation and cryopreserved in multiple aliquots containing 1 3 107 , 2 3 107 , 5 3 107 , 1 3 108 or 2 3 108 cells in each cryovial in a one mL volume. The cells were cryopreserved in the vapor phase of liquid nitrogen until use.

Cell Sorting for the T-cell study. For human subjects HIP2, HIP3, HIP4 and HIP5 subsets of naive and memory CD4+ and CD8+ T cells were obtained by magnetic activated cell sorting (MACS) prior to gDNA extraction and immunosequencing using the Adaptive Biotechnolgies Immunosequencing TCRb kit. CD4+ were first bead isolated by negative selection (Miltenyi) then the CD45RA Naive CD4s were isolated by positive selection with the negative fraction including memory cells. CD8 subset include a large fraction of CD8+ effector memory cells that are CD45RA+ (TemRA) which are highly clonal and if included as ‘‘naı¨ve’’ would skew the diversity in the naive fraction. Therefore, after negatively selecting CD8+ T cells, we needed two steps to isolate the naive from the memory cells. We first positively selected the CD45RO+ fraction. The negative fraction included CD8+ TemRA cells and CD8+naive cells. We then used CCR7-PE antibody (Miltenyi) and an anti-PE magnetic bead conjugated antibody (Miltenyi) to positively select the naive cells from the CCR7negative TemRA fraction. The TemRA cells were combined with the CD45RO fraction for CD8 memory TCR sequencing and

the CD45RA+CCR7+ was naive. All cell sorted populations were assessed for purity and quantity using analytical flow cytometry. A summary of resulting enriched T cell fractions.

mRNA sequencing. The mRNA sequencing method, performed by AbHelix LLC (http://www.abhelix.com/, South Plainfield, NJ, USA), was only utilized for human subjects HIP1, HIP2, and HIP3. For each of these human subjects, approximately 9 3 108 PBMCs were counted by hemacytometer and aliquoted into 5 or 6 biological replicates prior to total RNA extraction using the RNAeasy Maxi kit (QIAGEN). All extractions were performed on separate days and care was taken ensure no cross-contamination between replicate samples or human subject samples. Purified total RNA was shipped and processed at AbHelix, LLC. The AbHelix assay is designed to sequence 5 chains targeting B cell receptors (IgG, IgM, IgA, IgK and IgL) and 2 chains targeting T cell receptors TCRa and TCRb. The total RNA was divided evenly per B or T cell receptor chain type, so only 2/7 of the total RNA provided was utilized for TCRa and TCRb sequencing. The data from the B cell sequencing of HIP1, HIP2 and HIP3 at AbHelix was used in a separate but similar study. For the T cells assays used here, total RNA samples were reversed transcribed using the oligo d(T)18 in 3-5 ug per 20 ul reaction (SuperScript IV Reverse Transcriptase, ThermoFisher, CA). Multiple reactions of reverse transcription were combined per biological replicate and purified using magnetic beads. The purified RT products were divided evenly for the first round of PCR amplification specific to human TCRb and TCRa. The 50 multiplex PCR primers are designed within the leader sequences of each productive V-gene and the 30 primers within the constant regions but in close approximation to the J-C junctions. The resulting 1st PCR products were purified with magnetic beads and subject to the second round of PCR amplification to add Illumina index and adaptor sequences. The resulting PCR products were purified with magnetic beads and pooled for sequencing with PE 2x250 on an Illumina HiSeq 2500. Phusion High-Fidelity DNA Polymerase (ThermoFisher, CA) was used in all PCR amplification reactions and care was taken to minimize the number of cycles to achieve

adequate amplification. Purified libraries were quantitated using the Qubit 3.0 fluorometer (Thermo Fisher Scientific) prior to size determination using a Bioanalyzer 2100 (Agilent). Final libraries were quantified using the KAPA Library Quantification kit (Roche) before sequencing on an Illumina HiSeq 2500. A single PE-250 flow cell was dedicated to each biological replicate from each human subject (a total of 5-6 biological replicates and corresponding flow cells per human subject).

TCRb V3J clonotypes in GenBank for the T-cell study. The entire GenBank (Clark et al., 2016) database (release 231) was downloaded from GenBank: ftp://ftp.ncbi.nih.gov/genbank/ and processed using PyIR. Only those sequences from GenBank with Vb and Jb matches and that passed our quality filtering were considered. We used all V3J clonotypes from HIP1, HIP2, HIP3, HIP4 or HIP5 to search through the processed set of GenBank V3J clonotypes to find exact matches.

CHAPTER III

USING CLUSTERING OF ANTIBODY SEQUENCES TO INVESTIGATE INFLUENZA VIRUS-SPECIFIC ANTIBODY REPERTOIRES

Parts of this chapter is an adaption of the following manuscript as well as several works in progress:

Zost S.J., Dong J., Gilchuk I.M., Gilchuk P., Thornburg N.J., Bangaru S., Kose N., Finn J.A., Bombardi R., Soto C., **Chen E.C.**, Nargi R.S., Sutton R.E., Irving R.P., Suryadevara N., Westover J.B., Carnahan R.H., Turner H.L., Li S., Ward A.B., Crowe J.E. Canonical features of human antibodies recognizing the influenza hemagglutinin trimer interface. *The Journal of Clinical Investigation*. 2021; 131(15):e146791. doi: 10.1172/JCI146791

CHAPTER OVERVIEW

There are two analysis pipelines that were constructed at the beginning stages of my time in the Crowe lab. The first was a pipeline to cluster paired and bulk sequencing to help identify clusters/clonal families of antibodies that would have the same functional profile. This was built as there was a large set of 1,191 antibodies that had been isolated and characterized by Dr. Pavlo Gilchuk and Rachel Nargi from a donor acutely infected with H3N2. This effort was executed to identify therapeutic antibodies to influenza A. From the same donor, a series of bulk sequencing was performed on the donor's antibody response to the yearly influenza vaccine. Therefore a clustering method was implemented along with guidance from Dr. Cinque Soto and Dr. Seth Zost to bin and identify the functional characteristics of influenza-specific antibody sequences that had otherwise not been tested. Successful implementation of clustering allowed for the assessment of the evolutionary pathway of several antibody lineages with interesting functional phenotypes. Therefore the second pipeline I had worked on is the construction of phylogenetic trees. These two pipelines were implemented to many sets of influenza-specific antibody sequences. Additionally, the overlap of the bone marrow resident long-lived plasma

cells and plasmablast antibody sequences were identified and characterized in response to vaccination.

The work in this chapter would not have been possible without the help of Dr. Cinque Soto, Sam Day, and Luke Myers, who helped with the construction of both pipelines. Dr. Seth Zost, Dr. Pavlo Gilchuk, and Rachel Nargi performed the cell sorting of the long-lived plasma cells and plasmablasts in response to vaccination and infection. They also contributed to the characterization of several panels of antibodies described in this chapter.

INTRODUCTION

It is estimated that annually, influenza epidemics result in ~1 billion infections, of which 3-5 million develop severe illnesses, leading to 300,000 – 500,000 deaths (WHO), therefore remaining a serious health threat to the world. Due to the plasticity of the influenza genome, new viral strains continuously emerge annually, necessitating the annual reformulation of vaccines. Due to natural infections as well as vaccinations, humans are repeatedly exposed to the influenza virus and viral proteins. When stimulated by a new exposure, memory B cells can undergo rounds of somatic hypermutation and selection (Viant et al., 2020). Therefore, there is an ongoing selection in humoral immunity alongside the ongoing evolution of the virus. Antibody lineages may be re-activated every year during annual antigen exposure and can either gain or lose breadth across different strains (Schmidt et al., 2015). Examination of the evolutionary history of different B-cell lineages specific to influenza with beneficial properties such as protection and neutralization may have utility for vaccine design. Therefore, characterizing the dynamics and evolution of the B-cell repertoire using both bulk and paired single-cell sequencing allows for a detailed understanding of clonal dynamics and antigen specificity to antibodies specific to influenza, helping paint an integrated picture of the molecular detail of the repertoire to influenza.

In addition to aiding the design of next-generation influenza vaccines, monoclonal antibodies targeting conserved epitopes on influenza antigens are of therapeutic interest for passive immunotherapy. The two targets most discovery efforts are targeted to are HA or NA. Protective HA-directed antibodies usually mediates the binding of the virus to sialic acid receptors, therefore disrupting viral entry by preventing fusion of viral and host cell membranes. Protective NA-directed antibodies often target viral egress as NA is essential for cleaving terminal sialic acid residues to release newly formed virions (Matrosovich et al., 2004). The third protein present on the surface of the virus is the M2 protein (M2e), which is an ion channel. Antibodies targeting all three targets have demonstrated neutralization of virus as well as protection *in vivo* (Bangaru et al., 2019; Beerli et al., 2009; Stadlbauer et al., 2019).

HA-directed antibodies have been considered mediators of infection for years and therefore is, the antibody population measured when determining vaccine efficacy. However, the antigenic drift of HA remains faster than NA (Abed et al., 2002). Therefore inhibition of NA activity has been the basis of many antiviral therapeutics to influenza, such as zanamivir and oseltamivir. Despite this, little is known about the human antibody response to NA, and many vaccine development efforts are focused on targeting the HA. Previous work by Chen and colleagues identified that, unlike vaccination, natural influenza infection induces a high proportion of NA-reactive B cells (Chen et al., 2018). Therefore, recent work demonstrated the isolation of antibodies from previously infected individuals revealed cross-reactive NA-specific antibodies, which have therapeutic potential (Madsen et al., 2020).

This chapter contains several projects I contributed to in respect to antibody repertoires and responses to influenza: (1) Flu time series in a donor profiled by H3N2 natural infection, (2) a lineage of hemagglutinin trimer antibodies, (3) A lineage of M1 and M2 cross-reactive antibodies elicited by H3N2 natural infection, (4) The clonally expanded antibody repertoire in

response to H1N1 natural infection, (5) The clonally expanded antibody repertoire to influenza B natural infection B, and (6) kinetics of immune response to vaccination.

RESULTS

Influenza Time Series in a donor profiled by H3N2 natural infection

A donor was identified with an active H3N2 infection in the fall of 2017. On day 7 of infection, blood was drawn, and the plasmablast response to infection was sorted. Sorted plasmablasts were following sequenced using paired single-cell sequencing. A total of 4,037 sequences were derived, of which 1,191 antibodies were synthesized, microscale expressed, purified, and characterized for binding to a variety of influenza viral proteins. This same donor had blood drawn for a month at days 0, 3, 4, 5, 6, 7, 10, 13/14, and 27 after vaccination in 2013, 2014, 2015, 2016, and 2017 (**Figure III-1**). B cells from these blood draws were isolated and sequenced using bulk sequencing. This deep longitudinal dataset of sequencing within a single donor serves as a great platform to understand antibody evolution and dynamics to influenza exposure.

Different thresholds of clustering were attempted to identify a realistic clustering threshold. As different clustering thresholds were applied, the functional “bins” of each cluster were used to identify if there were any antibodies that had clustered with certain families that did not seem to share a similar functional phenotype. Binning antibodies by function served as a form of “validating” each threshold sequence were being clustered at. We had identified that within the panel of 1,191 antibodies that had been isolated from a plasmablast response and were validated, binning sequences by the same V_H gene, J_H gene, and CDRH3 length, followed by clustering at 60% sequence identity on the CDRH3 and binning at the back end with the V_L and J_L genes had binned sequences with similar functions together. From comparison with

literature and identification of functional bins, the following clustering scheme was applied. Clonal families were first determined using the paired sequence set. Antibodies were clustered by binning by V_H gene, J_H gene, and CDRH3 length, followed by complete-linkage clustering at 80% identity. These clusters were further subclustered using the V_L and J_L genes.

Out of the 4,037 sequences, there were 1,343 clonal families with a panel of 1,191 antibodies that had been functionally validated coming from 671 clonal families. The 671 clonal families were then assigned a function based on the reactivity all antibodies within that clonal family exhibited to the different antigens that they were tested against (NA (N1, N4, N8. N9), NP, M1, HA (H3, H1, H5, H7), and M2). From this, 215/671 clonal families were successfully assigned a functional phenotype. 54 clonal families were HA reactive, 106 were NA reactive, 18 were NP reactive, 15 were M1 reactive, and 22 were M2 reactive (**Figure III-1**).

These 1,343 clonal families were then used as a basis to pull in bulk sequences from each year of vaccination sequencing. By using a “chaining” method, paired sequence antibodies that had clustered together at a CDR3 similarity of 80% identity would continually stay together in the same clonal family. As sequences from each year of bulk sequencing were pulled in, the original cluster of antibody sequences must not be broken apart. Any sequence that falls within the 80% sequence identity of the already existing cluster is then pulled in as an additional member of the clonal family. This allowed us to index each clonal family by its functional characteristics – thereby allowing us to visualize the dynamics of the antibody response to each antigen.

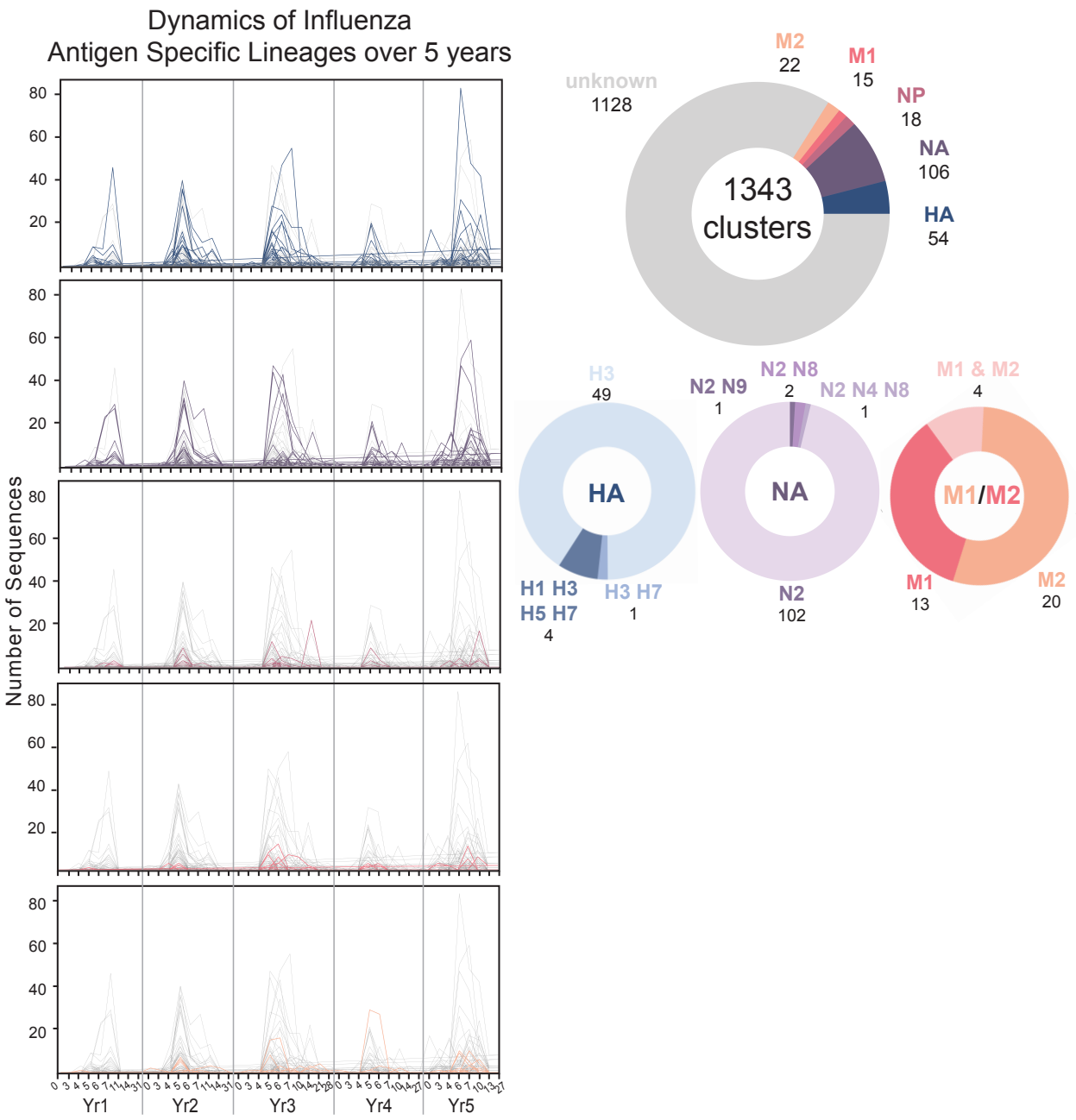


Figure III-1: Dynamics of influenza antigen-specific lineages over 5 years

Clustering using chaining based off the of paired sequence data is exhibited. In the left panel, each line is a lineage, the colored lines in each row represent the antigen specificity. HA lineages are in blue, NA lineages are in purple, NP lineages are in violet, M1 lineages are in pink, and M2 lineages are in orange.

A lineage of hemagglutinin trimer antibodies

Broadly reactive antibodies targeting the influenza A virus hemagglutinin (HA) head domain are thought to be rare and to require extensive somatic mutations or unusual structural features to achieve breadth against divergent HA subtypes. Within the large set of sequences described in the previous section, a lineage of antibodies was identified specific to the trimer interface of the influenza A HA head (**Figure III-2**). This lineage of antibodies was not only found in sequencing from every year of vaccination; they also are identified as a shared antibody response (public clonotype) encoded by a light chain variable gene segment (*IGKV1-39*) incorporating a shared somatic mutation.

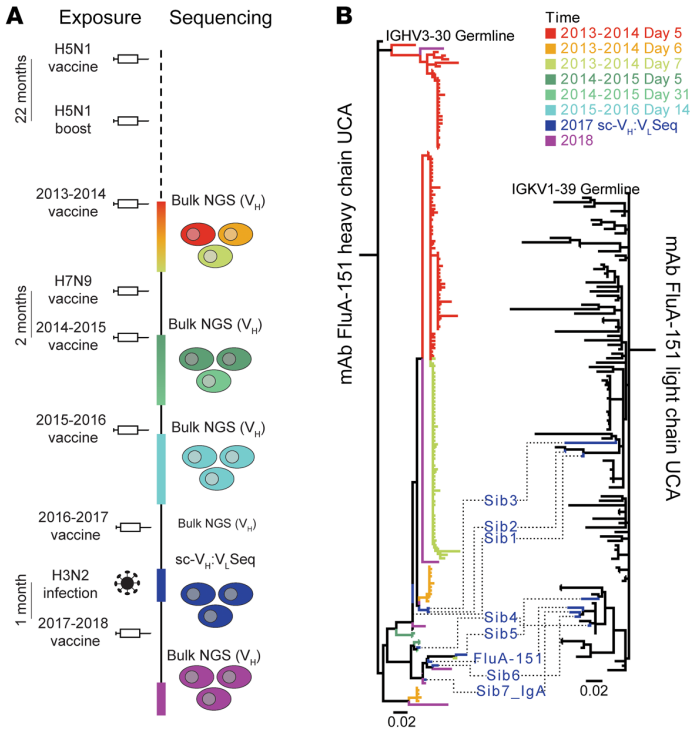


Figure III-2: Trimer interface antibody lineage recalled in the response to vaccination and natural infection.

(A) Timeline showing the vaccination history of the research subject, with exposures and repertoire sequencing indicated. (**B**) Phylogenetic trees showing the FluA-151 lineage over 4 years of vaccination and infection, with branch colors corresponding to sequencing time point. At left, the heavy chain phylogeny for FluA-151 is color-coded based on year of vaccination and days postvaccination. At right, the light chain phylogeny for FluA-151 is shown. Paired heavy-light sequences identified by single-cell sequencing are shown with dashes connecting the heavy chain and light chain trees.

Lineage of M1 and M2 cross-reactive antibodies elicited by H3N2 natural infection

From the same data set described in the last two sections, we also identified an M1 M2 cross-reactive lineage of antibodies utilizing the genes: *IGHV3-53/IGKJ4* and *IGKV1-39/IGKJ5*. The same lineage of antibodies re-appear in each year of vaccination sequencing as well (**Figure III-1**).

Features	Panel	10X	Year 1	Year 2	Year 3	Year 4	Year 5
IGHV3-53 *IGHJ4* *IGKV1-39* *IGKJ5* CDRH3: 14 AA	12	27	34	47	45	17	18

Table III-1: Number of sequences from the M1 M2 cross-reactive lineage.

This table indicates the number of sequences pulled identified to the M1 M2 cross-reactive lineage in each year of vaccination time series sequencing as well as in the plasmablast sequencing from natural infection (10X column).

As this class of antibodies was identified from a microscale expressed and purified panel, the twelve antibodies from the panel were then scaled up for follow on studies. EC_{50} binding curves were evaluated via ELISA, with M1 coated on the plate (**Figure III-4**). Binding to

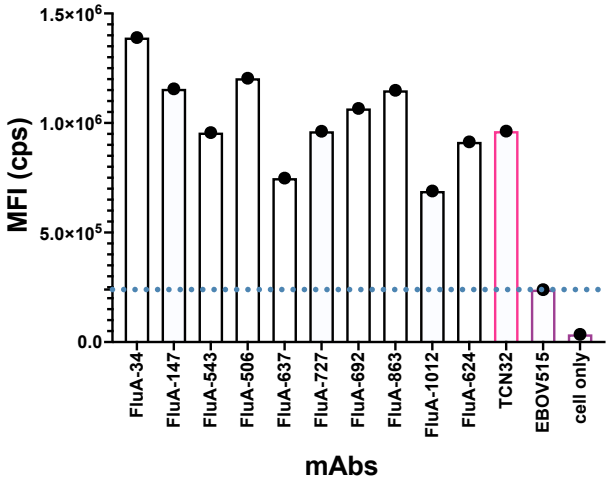


Figure III-3: M1 M2 cross-reactive lineage of antibodies to M2 on cell surface display

Cell surface display of M2 on 293F cells, with the binding of each antibody within the panel. Positive control is TCN32 (pink), and negative control is EBOV-515 (purple).

M2 was also assessed via cell surface display (**Figure III-3**). M2 construct was transfected into HEK 293F cells and expressed for 72 hours. Binding of the 12 antibodies were then assessed via flow cytometry.

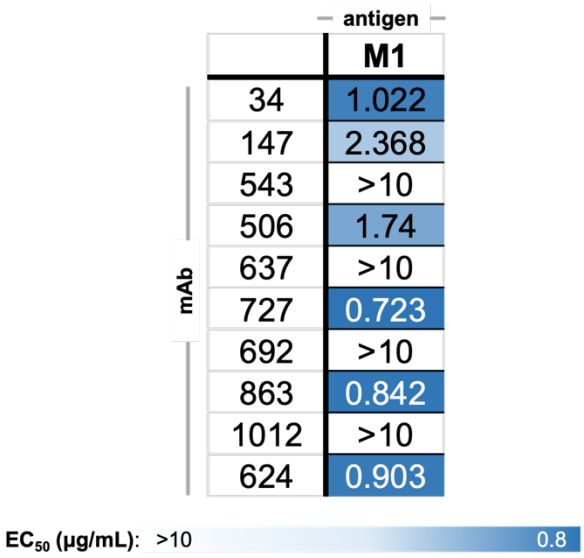


Figure III-4: M1 binding of antibodies

EC_{50} values of each of the antibodies to M1 measured via ELISA.

When aligning the M1 and M2 sequences, there is a conserved peptide sequence: MSLLTEVET. As this lineage of antibodies is cross-reactive to both M1 and M2, this is the hypothesized binding site. This peptide resides in the ectodomain of the M2 ion channel, and antibodies targeting this region have shown to be protective against infection(Grandea et al., 2010) and therefore is possible that it contributes to humoral immunity as this lineage is boosted every year from vaccination.

Antibody repertoire to H1N1 natural infection

A donor presented with an H1N1 acute infection with the circulating strain in December 2019 in Taiwan. On day 7, post-onset of symptoms, blood was drawn, and the plasmablast response was sorted and sequenced using paired single-cell sequencing. 14,504 paired heavy and light chain sequences were obtained. Sequences were then clustered by V_H gene, J_H gene,

and CDRH3 length, followed by complete-linkage clustering at 80% identity. These clusters were further subclustered using the V_L and J_L genes. The set of 14,504 sequences clustered into 7,723 clonal families. With the most clonally expanded family encapsulating 236 sequences represented by IGHV4-31/IGHJ6 and IGKV3-20/IGKJ1. A representative antibody of 94 of the most clonally expanded families was then synthesized, microscale expressed, purified, and characterized for binding to different influenza viral proteins. To determine a "hit rate," all expressed antibodies were also tested for binding to the 2019-2020 Fluzone vaccine (**Figure III-5**).

45/94 (47%) antibodies were determined as a positive hit using reactivity to the 2014-2015 Fluzone vaccine. Out of the 45 positive hits, two antibodies, H1N1-92, and H1N1-93, showed cross-reactive HA activity to H1, H3, H5, and H7 (**Figure III-5**). Subsequent testing revealed that H1N1-92 targets the head domain of the HA and is encoded by IGHV4-59/IGHJ6 and IGKV1-39/IGKJ2. Additional analysis into this antibody showed that this antibody is a member of the trimer interface public clonotype described in the above section. Interestingly, 15/45 antibodies exhibited cross-reactivity against H1 and H5 HAs, all of which (through subsequent testing) targeted the stem region of HA, demonstrating the conservation of HA stem across different subtypes.

TWIST name	A/Taiwan/D-1921/2020 rHA (H1)	A/California/04/2009 rHA (H1)	A/Texas/50/2012 rHA (H3)	A/Vietnam/1203/2004 rHA (H5)	A/Hong Kong/125/2017 rHA (H7)	A/Taiwan/D-1921/2020 rN1	A/Brisbane/07/2007 rN2	2014-2015 rN2 (+glyc)	N8 + N9	NP	M1	NS1	2019-2020 Fluzone vaccine
H1N1-85_mCisK	0.115	0.281	0.106	0.123	0.106	0.117	0.108	0.118	0.121	0.426	0.108	0.137	1.551
H1N1-73_mCisK	2.846	2.545	3.501	0.106	0.111	0.128	0.135	0.114	0.16	0.112	0.113	0.114	1.555
H1N1-38_mCisK	0.102	0.157	0.117	0.14	0.13	0.121	0.131	0.119	0.143	0.119	0.112	0.142	1.989
H1N1-93_mCisK	3.309	3.408	3.572	3.572	3.582	0.102	0.162	0.116	0.126	0.114	0.11	0.127	2.015
H1N1-92_mCisK	3.554	3.488	3.613	3.591	3.102	0.123	0.126	0.109	0.116	0.121	0.117	0.131	2.022
H1N1-78_mCisK	0.102	0.464	0.117	0.107	0.115	0.115	0.115	0.101	0.121	2.444	0.119	0.124	2.181
H1N1-91_mCisL	0.125	0.499	0.125	0.106	0.106	0.14	0.138	0.117	0.113	0.144	0.11	0.127	2.319
H1N1-8_mCisK	0.219	0.14	0.356	0.186	0.175	0.168	0.25	0.19	0.182	2.539	0.193	0.143	2.377
H1N1-63_mCisL	0.136	0.125	0.115	0.153	0.14	0.137	0.132	0.153	0.138	0.209	0.138	0.192	2.413
H1N1-45_mCisK	0.114	0.445	0.127	0.142	0.106	0.126	0.126	0.153	0.123	0.112	0.115	0.159	2.63
H1N1-32_mCisK	0.106	0.396	0.114	0.106	0.115	0.137	0.115	0.108	0.118	0.114	0.114	0.137	2.639
H1N1-48_mCisK	0.116	0.281	0.132	0.124	0.173	0.119	0.134	0.115	0.143	0.14	0.136	0.153	2.649
H1N1-35_mCisK	0.102	0.21	0.103	0.103	0.105	0.109	0.132	0.111	0.118	0.106	0.126	0.132	2.675
H1N1-94_mCisK	0.097	0.362	0.108	0.109	0.105	0.11	0.114	0.109	0.129	0.121	0.116	0.135	2.762
H1N1-66_mCisK	0.114	0.675	0.103	0.128	0.112	0.122	0.102	0.123	0.104	0.112	0.114	0.134	2.8
H1N1-22_mCisL	0.104	0.142	0.123	0.118	0.131	0.11	0.117	0.108	0.127	0.118	0.161	0.137	2.835
H1N1-77_mCisK	0.111	0.148	0.124	0.122	0.124	0.125	0.126	0.117	0.147	0.112	0.13	0.129	2.841
H1N1-3_mCisL	0.116	0.186	0.104	0.15	0.128	0.157	0.111	0.144	0.146	0.115	0.113	0.14	2.85
H1N1-72_mCisK	0.128	0.143	0.12	0.144	0.137	0.123	0.115	0.142	0.157	0.129	0.126	0.18	2.999
H1N1-41_mCisL	0.755	1.629	0.69	2.561	0.559	0.111	0.136	0.113	0.131	0.121	0.151	0.158	3.384
H1N1-54_mCisK	3.502	3.477	0.158	0.109	0.112	0.107	0.128	0.113	0.135	0.105	0.13	0.125	3.429
H1N1-4_mCisK	1.611	2.656	0.162	2.971	0.252	0.179	0.125	0.207	0.204	0.124	0.17	0.173	3.445
H1N1-69_mCisK	3.51	3.371	0.116	0.109	0.122	0.117	0.122	0.112	0.133	0.108	0.128	0.126	3.465
H1N1-37_mCisK	0.169	0.25	1.685	0.105	0.113	1.707	0.249	0.773	0.272	0.107	0.124	0.13	3.467
H1N1-90_mCisL	3.474	3.353	0.109	2.389	0.148	0.131	0.118	0.129	0.136	0.127	0.118	0.155	3.472
H1N1-95_mCisL_hG1	3.399	3.438	0.141	0.193	0.139	0.124	0.129	0.152	0.125	0.246	0.164	0.246	3.476
H1N1-13_mCisK	3.482	3.429	0.112	3.375	0.157	0.142	0.129	0.118	0.154	0.108	0.204	0.135	3.477
H1N1-104_mCisL_hG1	0.986	3.409	0.27	0.21	0.338	0.173	0.524	0.187	0.29	0.222	0.277	0.194	3.481
H1N1-64_mCisK	3.385	3.456	0.11	3.508	0.103	0.129	0.13	0.132	0.108	0.129	0.116	0.134	3.486
H1N1-27_mCisK	3.612	3.562	0.11	0.123	0.127	0.122	0.113	0.112	0.12	0.114	0.122	0.142	3.502
H1N1-43_mCisK	3.543	3.371	0.134	3.585	0.189	0.134	0.152	0.173	0.188	0.155	0.133	0.174	3.508
H1N1-56_mCisL	3.39	3.466	0.124	3.531	0.106	0.131	0.143	0.141	0.12	0.115	0.125	0.134	3.519
H1N1-23_mCisK	3.507	3.436	0.115	1.949	0.109	0.114	0.11	0.112	3.081	0.112	0.115	0.128	3.522
H1N1-97_mCisL_hG1	2.815	3.483	0.121	0.18	0.131	1.18	0.119	0.127	0.146	0.153	0.131	0.18	3.525
H1N1-89_mCisK	3.556	3.357	0.105	0.121	0.105	0.111	0.11	0.111	0.112	0.11	0.117	0.133	3.528
H1N1-58_mCisK	3.447	3.431	0.1	2.862	0.104	0.111	0.102	0.106	0.108	0.11	0.108	0.135	3.528
H1N1-7_mCisK	3.553	3.37	0.264	3.55	0.296	0.221	0.292	0.175	0.291	0.124	0.168	0.175	3.53
H1N1-60_mCisK	3.429	3.402	0.111	3.402	0.122	0.13	0.117	0.12	0.126	0.13	0.121	0.148	3.531
H1N1-71_mCisK	3.497	3.266	0.137	0.492	0.111	0.12	0.111	0.12	0.119	0.138	0.119	0.149	3.532
H1N1-29_mCisK	1.434	3.363	0.106	0.123	0.097	0.102	0.1	0.103	0.106	0.106	0.106	0.12	3.533
H1N1-26_mCisL	2.872	3.371	0.151	0.12	0.131	0.108	0.16	0.109	0.174	0.174	0.321	0.17	3.541
H1N1-6_mCisK	3.482	3.42	0.317	3.538	0.301	0.212	0.444	0.185	0.232	0.134	0.191	0.193	3.557
H1N1-53_mCisK	3.581	3.441	0.125	0.138	0.166	0.125	0.217	0.122	0.239	0.237	0.132	0.186	3.558
H1N1-98_mCisL_hG1	3.265	3.394	0.187	0.19	0.185	0.151	0.185	0.17	0.236	0.157	0.192	0.265	3.564
H1N1-86_mCisK	3.57	3.288	0.105	3.605	0.11	0.126	0.122	0.126	0.116	0.123	0.12	0.164	3.565

Figure III-5: H1N1 donor derived antibody panel

Antibodies isolated from the H1N1 naturally infected donor with positive hits to the 2019-2020 Fluzone vaccine are shown. Experimental data shown is the OD 450 nm reading from ELISA with the antigen coated on the plate at 1μg/mL listed across the top of the table and the antibody names across the y-axis. Darker blue indicates higher OD reading.

Antibody repertoire to influenza B (Victoria) natural infection

A donor acutely infected with Influenza B presented January 2020, likely infected with the Victoria strain, which was the circulating strain in North America at the time. On day 7, post-onset of symptoms, blood was drawn, and the plasmablast response was sorted and sequenced using paired single-cell sequencing. 7,201 paired heavy and light chain sequences were obtained. Sequences were then clustered by V_H gene, J_H gene, and CDRH3 length, followed by complete-linkage clustering at 80% identity. These clusters were further subclustered using the V_L and J_L genes. The set of 7,201 sequences clustered into 4,753 clonal families. The most clonally expanded family had 60 members, encoded by *IGHV3-30/IGHJ4* and *IGKV4-1/IGKJ1*. Out of the 181 clonally expanded families (determined by if there were five or more sequences encapsulated in the clonal family), a representative of each clonal family along with 113 other antibodies chosen from non-clonally expanded families were synthesized, microscale expressed, purified, and characterized (**Figure III-6** and **Figure III-7**).

Out of the 312 antibodies tested, there was an overall 60% antigen reactivity determined by reactivity to the 2014-2015 Fluzone vaccine. Out of the 256 clones that were antigen reactive, 87 were HA reactive (**Figure III-7**), and 42 were NA reactive (**Figure III-6**). Besides assessing binding, antibodies were also tested for neutralizing using a real-time cell analysis system that reads cellular impedance as a measure for neutralization against B/Colorado/6/2017.

	ELISA binding (OD_{450})								HAI		Neut	NA-Fluor	
mAb	B/Singapore/2016 rHA (Yamagata)	B/Iowa/2017 rHA (Victoria)	B/Darwin/2019 rHA (Victoria)	B/Singapore/2016 rNA (Yamagata)	B/Colorado/2017/rNA (Victoria)	A/Brisbane/2007/rN2	A/Texas /2012/rHA	2019-2020 Fluzone vaccine	B/Colorado/6/2017 (Victoria)	B/Indiana/17/2017 (Yamagata)	RTCA % Neut (B/Colorado/6/2017, Victoria	B/Colorado/06/2017 (Victoria) % signal	B/Indiana/17/2017 (Yamagata) % signal
1G05	0.068	0.29	0.067	1.62	3.202	0.066	0.073	2.626	0	0	4	28	0
2D10	0.07	0.154	0.069	1.224	2.693	0.067	0.076	1.781	0	0	82	90	117
2E01	0.07	0.204	0.09	1.41	3.414	0.082	0.082	2.048	0	0	110	5	0
FluB-104	0.075	0.133	0.074	2.135	3.754	0.075	0.083	2.361	0	0	69	76	86
FluB-129	0.075	0.26	0.075	3.273	3.738	0.093	0.085	2.73	0	0	3	153	89
FluB-130	0.08	0.273	0.079	3.516	3.794	0.078	0.086	2.719	0	0	13	155	86
FluB-133	0.075	0.072	0.071	0.074	2.342	0.072	0.078	1.056	0	0	32	97	90
FluB-15	0.075	0.066	0.071	0.077	2.158	0.076	0.072	1.236	0	0	22	93	83
FluB-155	0.069	0.078	0.069	1.798	3.754	0.069	0.076	2.404	0	0	28	82	86
FluB-157	0.072	0.068	0.07	0.076	3.231	0.074	0.079	0.682	0	0	16	100	96
FluB-159	0.081	0.377	0.079	3.569	3.745	0.071	0.081	3.531	0	0	23	121	72
FluB-163	0.075	0.073	0.069	0.085	3.677	0.069	0.072	1.479	0	0	79	130	96
FluB-174	0.1	0.092	0.076	0.237	0.766	0.081	0.102	1.922	0	0	20	110	90
FluB-18	0.071	0.09	0.07	1.358	3.522	0.07	0.078	1.841	0	0	68	128	66
FluB-2	0.223	0.212	0.168	0.606	3.754	0.118	0.095	2.786	0	0	15		
FluB-205	0.079	0.074	0.075	0.08	3.78	0.078	0.085	1.661	0	0	36	75	82
FluB-208	0.087	0.114	0.111	0.081	3.344	0.092	0.118	1.992	0	0	22	107	96
FluB-209	0.075	0.073	0.075	0.075	0.764	0.093	0.071	0.616	0	0	60	89	99
FluB-210	0.076	0.076	0.077	0.133	2.385	0.076	0.077	1.11	0	0	48	91	85
FluB-217	0.074	0.215	0.218	0.856	3.739	0.077	0.075	1.77	0	0	26	122	77
FluB-220	0.08	0.125	0.079	1.728	1.816	0.078	0.078	2.158	0	0	78	28	19
FluB-222	0.091	0.088	0.101	0.116	3.513	0.097	0.09	0.6	0	0	73	131	102
FluB-242	0.08	0.079	0.079	0.082	1.299	0.081	0.085	1.263	0	0	25	102	94
FluB-243	0.076	0.072	0.077	0.075	3.705	0.072	0.077	1.936	0	0	31	381	89
FluB-259	0.079	0.19	0.077	3.345	3.734	0.081	0.093	3.198	0	0	93	67	93
FluB-26	0.066	0.188	0.066	1.61	3.423	0.066	0.075	2.618	0	0	26	155	74
FluB-291	0.086	0.316	0.124	3.479	3.617	0.087	0.084	2.959	0	0	9	126	82
FluB-302	0.084	0.161	0.112	2.545	3.764	0.12	0.113	2.421	0	0	6	98	81
FluB-315	0.082	0.09	0.08	2.188	3.735	0.078	0.08	2.354	0	0	24	171	82
FluB-339	0.079	0.078	0.078	0.085	0.58	0.078	0.078	0.197	0	0	22	119	97
FluB-357	0.084	0.092	0.081	1.296	3.647	0.079	0.085	1.747	0	0	79	71	80
FluB-362	0.096	0.082	0.081	0.962	3.703	0.085	0.085	1.517	0	0	69	73	83
FluB-39	0.068	0.118	0.066	0.696	3.044	0.068	0.079	2.312	0	0	75	133	86
FluB-47	0.07	0.067	0.069	0.532	0.839	0.069	0.071	0.077	0	0	13	108	85
FluB-48	0.068	0.068	0.069	0.188	3.571	0.068	0.072	1.874	0	0	22	414	62
FluB-49	0.123	0.147	0.148	2.15	3.77	0.143	0.105	3.017	0	0	2	99	92
FluB-50	0.07	0.066	0.07	0.077	3.195	0.065	0.071	1.46	0	0	61	131	94
FluB-53	0.074	0.088	0.083	0.073	3.553	0.085	0.086	2.081	0	0	34		
FluB-55	0.067	0.071	0.073	0.068	3.066	0.071	0.08	1.492	0	0	58	203	92
FluB-66	0.073	0.106	0.07	2.196	3.467	0.069	0.079	2.522	0	0	8	128	85
FluB-84	0.07	0.111	0.067	1.45	3.253	0.064	0.076	1.823	0	0	16	96	79
FluB-87	0.07	0.096	0.067	1.213	3.279	0.068	0.079	2.042	0	0	64	58	84
FluB-94	0.094	0.071	0.07	0.131	2.757	0.076	0.08	1.53	0	0	65	192	69
FluB-96	0.082	0.087	0.096	0.608	3.823	0.089	0.083	2.033	0	0	92	114	68
FluB-99	0.076	0.075	0.075	0.47	3.35	0.079	0.07	1.665	0	0	25	109	76

Figure III-6: Influenza B donor derived NA specific antibodies

NA reactive antibodies from the Flu B antibody panel. Starting from the left shows ELISA binding at OD450 nm read. Neutralization is identified in green normalized to cells only wells. Blue cells on the right indicates NAI activity detected by NA-Fluor.

55/87 HA reactive clones showed neutralizing activity, and 15/42 NA reactive clones showed neutralizing activity. 2/42 NA clones also exhibited neuraminidase inhibition activity detected by NA-Fluor. Hemagglutinin inhibition was also tested (HAI activity) for all antibodies that exhibited reactivity to HA.

	ELISA binding (OD_{450})								HAI		Neut	NA-Fluor	
mAb	B/Singapore/2016 rHA (Yamagata)	B/Iowa/2017 rHA (Victoria)	B/Darwin/2019 rHA (Victoria)	B/Singapore/2016 rNA (Yamagata)	B/Colorado/2017/rNA (Victoria)	A/Brisbane/2007/rN2	A/Texas /2012/rHA	2019-2020 Fluzone vaccine	B/Colorado/6/2017 (Victoria)	B/Indiana/17/2017 (Yamagata)	RTCA % Neut (B/Colorado/6/2017, Victoria	B/Colorado/06/2017 (Victoria) % signal	B/Indiana/17/2017 (Yamagata) % signal
FluB-1	0.692	1.247	1.15	0.073	0.079	0.072	0.074	2.485	0	0	51		
FluB-102	1.434	3.35	3.442	0.088	0.086	0.111	0.096	3.42	0	0	40		
FluB-107	0.066	3.661	3.754	0.068	0.08	0.076	0.071	3.657	0	0	22		
FluB-112	0.074	3.766	3.807	0.08	0.084	0.081	0.084	3.703	0	0	27		
FluB-115	2.32	3.773	3.634	0.075	0.08	0.088	0.079	3.736	+	+	86	103	91
FluB-116	2.143	3.506	3.739	0.08	0.092	0.092	0.082	3.544	0	0	77	98	89
FluB-117	3.485	3.712	3.741	0.079	0.083	0.104	0.081	3.682	0	0	77	114	95
FluB-12	0.066	3.163	3.571	0.079	0.08	0.083	0.079	3.471	0	0	47		
FluB-121	0.075	3.66	3.842	0.07	0.086	0.099	0.086	3.771	0	0	72		
FluB-125	1.924	3.493	2.68	0.077	0.079	0.082	0.078	3.359	0	+	85		
FluB-136	2.05	3.802	3.81	0.085	0.142	0.089	0.083	3.697	0	0	48		
FluB-139	0.069	3.665	3.749	0.072	0.086	0.086	0.08	3.682	0	0	38		
FluB-142	2.826	2.6	3.035	0.089	0.095	0.093	0.091	3.669	0	0	-1		
FluB-144	0.099	3.637	3.65	0.078	0.086	0.088	0.087	3.596	0	0	72		
FluB-151	0.345	2.417	3.453	0.087	0.087	0.084	0.082	3.457	0	0	68		
FluB-154	0.407	1.764	2.618	0.078	0.077	0.082	0.081	2.207	0	0	29		
FluB-16	0.08	2.234	3.1	0.071	0.092	0.075	0.072	3.46	0	0	60		
FluB-160	0.08	3.828	3.801	0.079	0.088	0.088	0.088	3.708	0	0	75		
FluB-162	0.073	3.68	3.782	0.083	0.09	0.09	0.086	3.608	0	0	31		
FluB-165	0.069	2.306	3.24	0.076	0.085	0.08	0.081	2.813	0	0	34		
FluB-168	0.076	3.599	3.678	0.075	0.093	0.082	0.086	3.607	0	0	73		
FluB-17	0.223	3.741	3.657	0.105	0.102	0.103	0.101	3.779	0	0	72		
FluB-170	0.544	3.667	3.669	0.076	0.095	0.078	0.085	3.72	+	0	80	112	95
FluB-171	0.069	3.676	3.67	0.071	0.086	0.079	0.076	3.667	+	0	81		
FluB-180	0.071	3.765	3.818	0.071	0.081	0.083	0.083	3.736	0	0	28		
FluB-182	0.074	3.697	3.701	0.079	0.106	0.093	0.099	3.636	0	0	34		
FluB-183	0.111	3.668	3.75	0.095	0.128	0.084	0.096	3.635	0	0	52		
FluB-184	0.077	3.67	3.777	0.081	0.092	0.088	0.092	3.686	0	0	52		
FluB-185	0.785	0.881	1.132	0.076	0.084	0.076	0.082	0.914	0	0	89		
FluB-194	0.085	3.789	3.768	0.095	0.087	0.096	0.084	3.789	0	0	47		
FluB-195	0.082	1.885	2.119	0.083	0.083	0.08	0.078	1.777	0	0	40		
FluB-199	0.09	3.661	3.76	0.096	0.095	0.104	0.092	3.754	+	0	81		
FluB-20	0.358	1.423	1.896	0.082	0.075	0.075	0.077	2.938	0	0	92	113	86
FluB-203	0.071	3.793	3.793	0.073	0.127	0.092	0.102	3.778	+	0	89		
FluB-204	1.608	3.531	3.733	0.114	0.089	0.08	0.082	3.361	0	0	56		
FluB-214	0.082	3.499	3.648	0.082	0.09	0.098	0.099	3.533	+	0	81		
FluB-215	0.078	3.665	3.846	0.093	0.091	0.104	0.084	3.782	0	0	50		
FluB-218	0.081	3.778	3.85	0.084	0.107	0.122	0.096	3.824	0	0	83		
FluB-229	0.086	3.726	3.8	0.075	0.089	0.078	0.08	3.72	+	0	76		
FluB-232	0.088	3.822	3.816	0.104	0.109	0.092	0.103	3.731	+	0	91		
FluB-236	0.078	3.634	3.726	0.087	0.114	0.127	0.102	3.589	0	0	83		
FluB-237	0.077	3.693	3.703	0.075	0.109	0.098	0.106	3.666	0	0	80		
FluB-24	0.07	2.751	3.233	0.071	0.076	0.089	0.092	3.513	0	0	80		
FluB-240	0.082	3.768	3.72	0.076	0.106	0.101	0.095	3.696	+	0	82		
FluB-246	0.084	3.755	3.69	0.081	0.091	0.099	0.103	3.661	0	0	32		
FluB-247	3.639	3.47	3.607	0.129	0.142	0.12	0.114	3.751	0	0	26		
FluB-248	3.611	3.825	3.802	0.125	0.125	0.11	0.11	3.738	0	0	37		
FluB-249	2.061	3.736	3.761	0.088	0.106	0.112	0.104	3.75	+	+	81		
FluB-254	0.079	3.793	3.656	0.104	0.128	0.134	0.124	3.643	+	0	83		
FluB-256	3.476	3.359	3.547	0.094	0.106	0.106	0.107	3.583	0	0	75		
FluB-263	0.094	3.706	3.76	0.123	0.14	0.103	0.109	3.746	+	0	76		
FluB-264	3.742	3.776	3.794	0.107	0.086	0.11	0.083	3.808	0	0	8		
FluB-267	0.072	3.753	3.754	0.076	0.112	0.096	0.118	3.711	0	0	27		
FluB-279	0.106	0.551	0.582	0.09	0.112	0.079	0.098	2.351	0	0	6		
FluB-28	0.067	1.965	3.234	0.078	0.083	0.087	0.093	3.23	0	0	8		
FluB-284	0.145	3.654	3.732	0.123	0.1	0.138	0.092	3.665	0	0	58		
FluB-285	0.088	3.663	3.696	0.089	0.11	0.123	0.141	3.625	0	0	86		
FluB-286	0.078	2.411	2.826	0.081	0.091	0.085	0.132	2.973	0	0	12		
FluB-295	0.117	2.573	0.703	0.117	0.11	0.085	0.092	3.478	+	0	93		
FluB-31	0.214	3.793	1.221	0.096	0.095	0.092	0.09	3.832	+	0	55		
FluB-323	0.077	3.758	3.792	0.08	0.447	0.099	0.092	3.637	0	0	64		
FluB-328	0.1	3.867	3.842	0.097	0.106	0.1	0.107	3.783	0	0	74		
FluB-33	1.998	3.079	3.251	0.078	0.084	0.084	0.081	3.475	0	0	3		
FluB-342	0.085	3.689	3.662	0.081	0.097	0.109	0.12	3.663	0	0	52		
FluB-349	0.082	3.667	3.5	0.087	0.104	0.123	0.108	3.654	0	0	73		
FluB-350	0.107	3.737	3.574	0.102	0.119	0.139	0.125	3.646	0	0	56		
FluB-366	3.609	3.642	3.62	0.143	0.133	0.099	0.103	3.636	0	0	6		
FluB-374	0.083	3.615	3.626	0.088	0.133	0.099	0.101	3.57	0	0	85		
FluB-380	0.119	0.566	0.655	0.119	0.111	0.119	0.101	0.508	0	0	16		
FluB-42	0.814	0.596	0.781	0.083	0.07	0.07	0.119	3.536	0	0	8		
FluB-43	0.174	3.75	3.278	0.105	0.09	0.096	0.094	3.792	+	+	100		
FluB-51	0.062	2.692	2.823	0.065	0.074	0.08	0.085	3.369	0	0	67		
FluB-57	0.068	2.894	3.086	0.068	0.079	0.072	0.08	3.567	+	+	86		
FluB-58	0.07	2.638	3.266	0.072	0.081	0.072	0.083	3.51	0	0	38		
FluB-60	0.845	1.993	2.316	0.074	0.076	0.079	0.082	3.109	0	0	83		
FluB-62	0.069	3.148	3.294	0.069	0.084	0.081	0.081	3.503	0	0	70		
FluB-69	0.065	3.676	0.074	0.063	0.078	0.063	0.064	0.217	0	0	8		
FluB-72	3.564	3.621	3.698	0.117	0.1	0.087	0.096	3.795	0	0	92		
FluB-73	0.069	2.924	3.572	0.07	0.075	0.074	0.075	3.503	0	0	39		
FluB-74	0.069	2.637	3.162	0.07	0.074	0.073	0.088	3.405	+	0	86		
FluB-76	0.071	1.988	1.489	0.068	0.081	0.075	0.091	3.432	+	0	89		
FluB-82	3.663	3.61	3.559	0.085	0.09	0.11	0.084	3.565	0	0	17		
FluB-83	0.071	3.009	2.525	0.074	0.08	0.08	0.085	3.606	+	0	84		
FluB-88	0.071	3.16	3.273	0.069	0.083	0.085	0.084	3.565	0	0	70		
FluB-91	0.067	3.054	1.692	0.069	0.077	0.069	0.08	3.581	+	0	87		
FluB-92	0.069	1.519	3.008	0.071	0.076	0.073	0.083	3.157	0	0	25		
FluB-95	3.53	3.676	3.734	0.105	0.146	0.095	0.11	3.769	0	0	15		

Figure III-7: Influenza B donor derived HA specific antibodies

Flu B antibodies exhibiting HA reactivity is shown here. Starting from ELISA binding to viral antigens read at OD450nm. HAI was tested for all HA reactive antibodies with HAI activity shown in green. Neutralization activity is also shown as percent neutralization in green, normalized to cells only wells on each plate.

Several of the HA-specific antibodies that exhibited cross-reactive binding to both Yamagata and Victoria influenza B strains HAs were selected for additional studies (**Figure III-8**). Out of the 7 antibodies selected, they bin into two groups based on competition binding via ELISA (**Figure III-9**).

mAb	HA antigens: B/Singapore/INFTT16-0610/2016 (Yamagata)	B/Darwin/07/2019 (Victoria)	B/Iowa/06/2017 (Victoria)
FluB-43	-	0.45	0.02
FluB-249	1.8	0.069	0.012
FluB-115	0.601	0.389	0.052
FluB-125	0.485	1.265	0.204
FluB-72	0.014	0.023	0.019
FluB-117	0.012	0.024	0.011
FluB-256	0.033	0.114	0.021

Figure III-8: Influenza B HA cross-reactive HA antibodies binding

Influenza B HA reactive antibodies were tested to determine EC_{50} values to each of the HA antigens. One Yamagata lineage HA was tested and two Victoria lineage HAs were tested. Binding potency is indicated by the shade of blue.

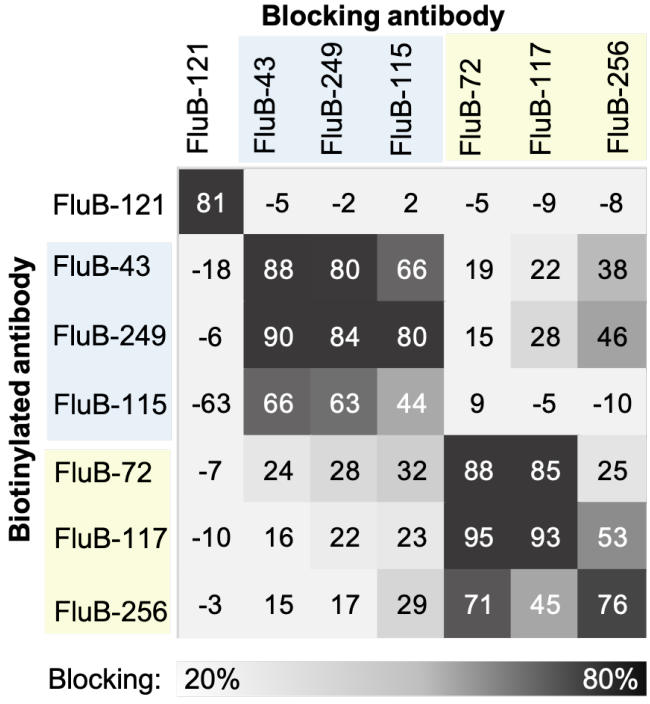


Figure III-9: Influenza B HA specific antibodies competition binding by BLI

Cross-reactive Flu B HA antibodies bin into two main groups shown in the blue and yellow. Loaded antigen was B/Iowa/06/2017.

The heavy chain gene usage of the influenza B specific repertoire was also assessed. *IGHV4-39* and *IGHV3-30* were the most commonly used genes within the repertoire. A breakdown of the functionally validated antibodies using these gene usages is identified, showing a large variety of functional phenotypes (**Figure III-10**).

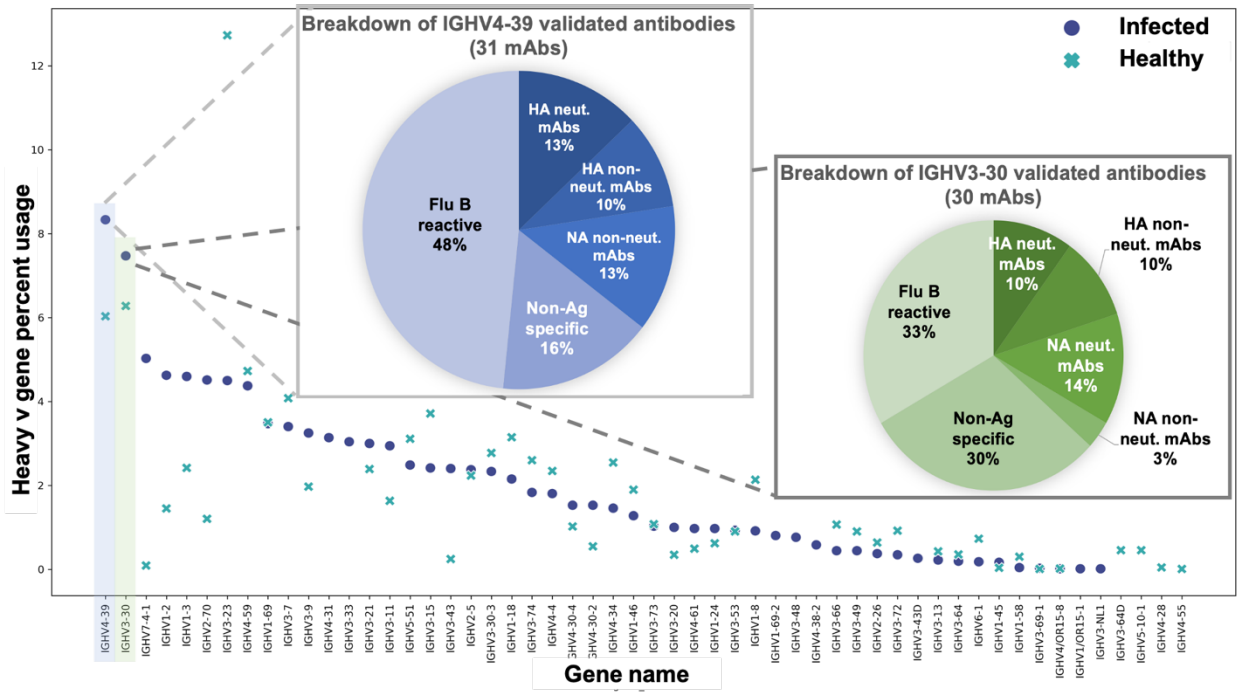


Figure III-10: Heavy chain gene usages of Influenza B specific repertoire

Heavy chain gene assignments of each antibody sequenced within the Influenza B plasmablast response are shown with the dark blue. A healthy data set of memory B-cells is used as a comparison shown in the turquoise crosses.

<u>Kinetics of Immune response to vaccination project</u>

Four donors designated as KIRV7, 8, 9, and 10 were enrolled in a study where they received the 2018-2019 season influenza vaccine. 7 days post-vaccine, whole blood was drawn to sort for their plasmablast response. On either day 29 or 57 post-vaccine, bone marrow was

drawn from each of the donors to sort for their plasma cell response. All sequencing was done using paired single-cell sequencing.

From each of the donors, we identified the shared clonotypes in plasmablast and bone marrow plasma cell repertoires. A clonotype within this study is identified as an exact match on the V gene, J gene, and CDR3 amino acid sequence. Within KIRV7, we identified 55 clonotypes present in both the bone marrow plasma cell repertoire as well as the plasmablast repertoire. Within KIRV8, 9, and 10, we identified 53, 183, and 64 clonotypes, respectively (**Figure III-11**).

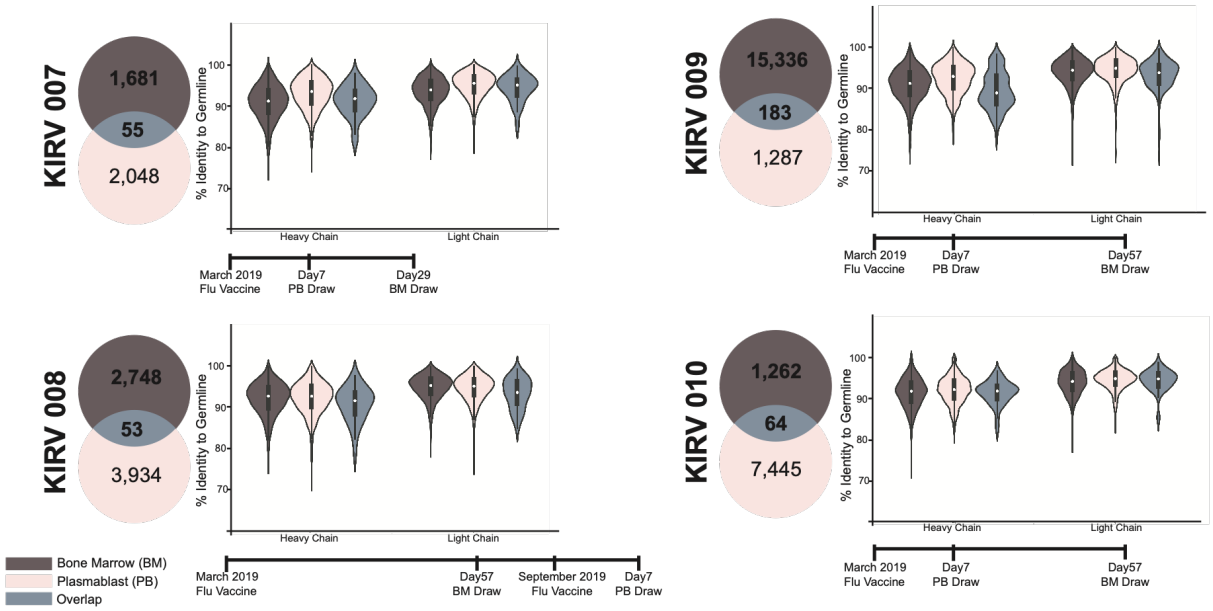


Figure III-11: Plasmablast and bone marrow repertoire overlaps
Number of plasmablast sequences in each donor is indicated in pink and bone marrow sequences in brown. The overlap repertoire is indicated in grey. Graphs indicate the % identity to germline with all the sequenced identified within each group. The time line under each graph indicates the sampling each donor had.

The clones identified in the overlap between the bone marrow and plasmablast repertoire were synthesized, and microscale expressed, purified, and tested for binding against several influenza viral proteins. While most clones did not show antigen reactivity, one lineage

of clones derived from KIRV10 was specific to NA. Testing of all clones within this lineage showed cross-reactivity to both Darwin and Yamagata influenza B NAs. Additionally, the lineage derived from the antibody KIRV10_OL19 is a possible public clonotype with a previously identified antibody by Madsen and colleagues, 2D10, as they use similar gene usages (**Figure III-14** and **Figure III-12**).

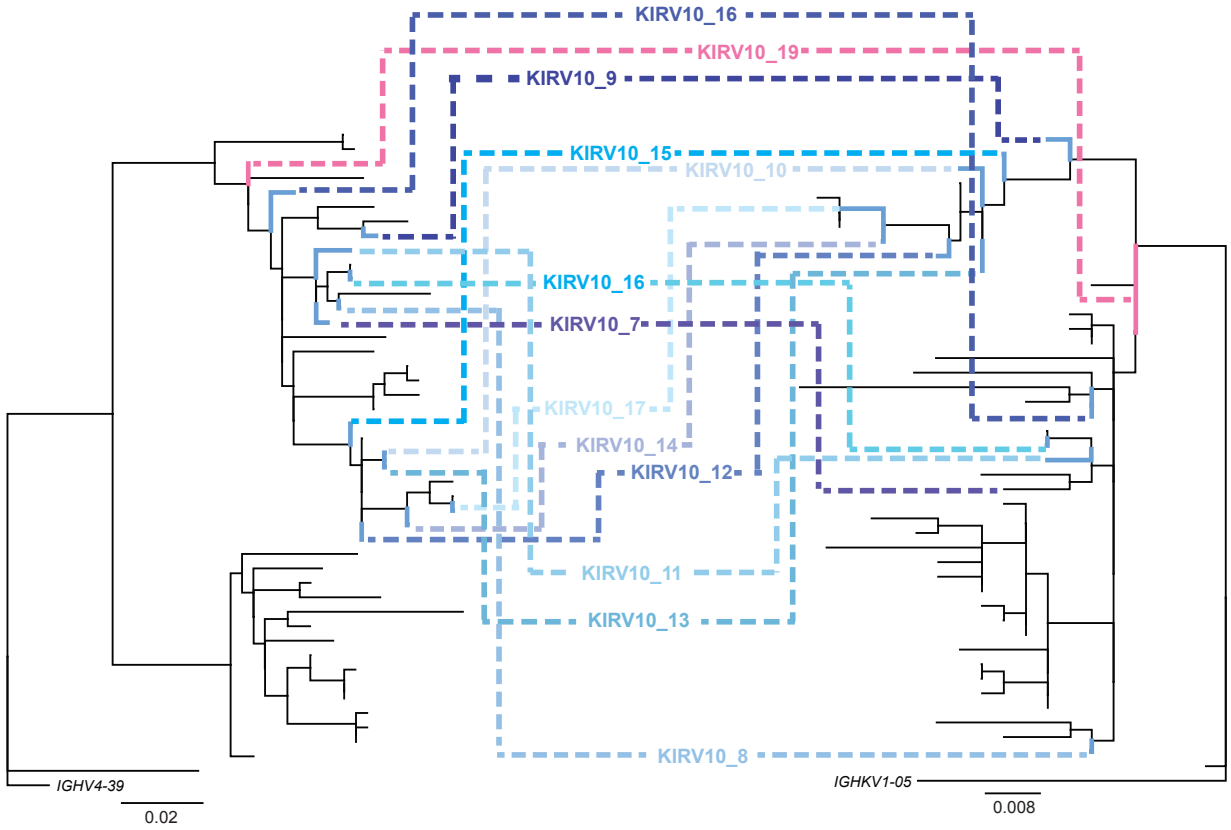


Figure III-12: Phylogenetic tree of NA reactive lineage within donor KIRV10

Phylogenetic tree constructed for NA specific antibodies. Dotted lines link the heavy and light chain trees. All antibodies indicated by a KIRV10_XX are clones that shared the same V3J clonotype in both plasmablast and bone marrow repertoires, and were therefore tested. Black lines indicate sequences that had not been tested yet. Plasmablast identified sequences are in blue and bone marrow identified sequence is in pink.

Revelation of a possible light chain driven influenza B light chain driven public clonotype

As cross-reactive antibodies were identified in both the acute influenza B donor study as well as the vaccination bone marrow plasma cells and plasmablast overlap study, we aimed to identify the number of sites these antibodies bin into. B/Singapore/INFTT16-0610/2016 was loaded onto anti-his biolayer interferometry tips, and blocking was detected. This study showed that the cross-reactive NA antibodies are binned into mainly three epitopes. With one slightly overlapping the two (**Figure III-13**).

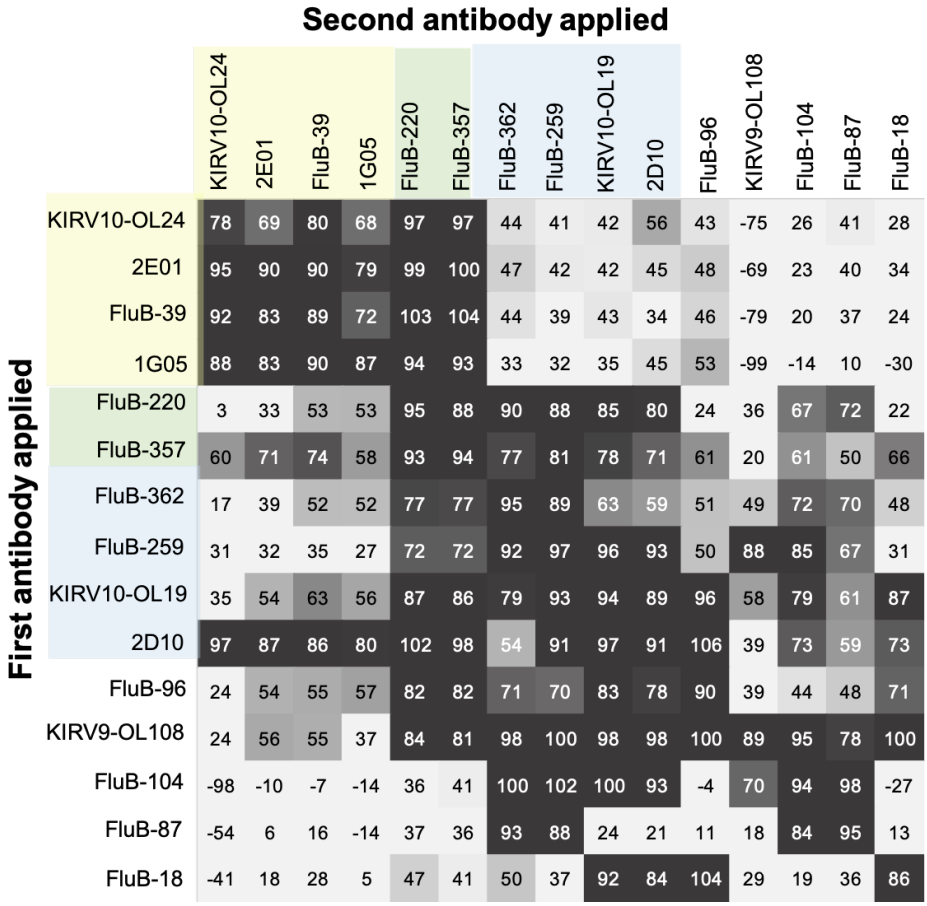
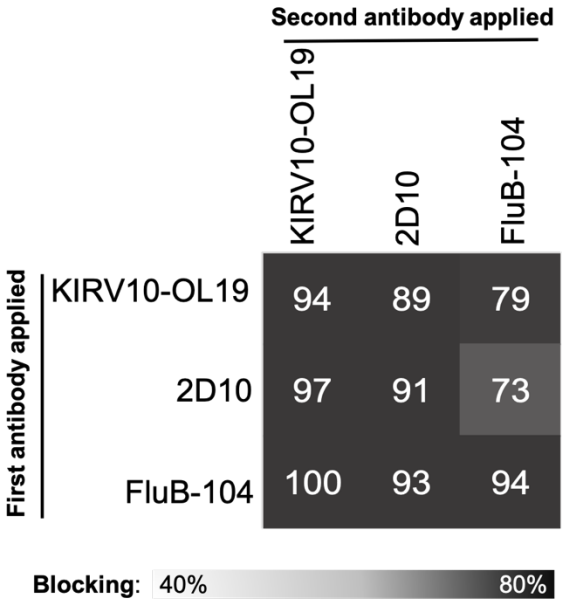


Figure III-13: Influenza B cross-reactive NA specific antibodies binning

Biolayer interferometry experiment readings normalized to percent blocking for all NA cross-reactive antibodies identified in the two experiments. The three projected epitope bins are identified in yellow, blue, and green. With the green bin likely nested between the yellow and blue due to some overlap in antibody blocking. Antigen loaded was B/Singapore/2016.

As previously mentioned, it is hypothesized that KIRV10-OL19 and 2D10 are public clonotypes as they share the same heavy and light chain variable gene usage. When comparing gene usages of other antibodies that bind a similar site, it was found that FluB-104 not only seems to bind a similar site but also utilizes the same light chain gene usage. It is possible that IGKV1-5 drives a light chain dominated public clonotype eliciting cross-reactive Influenza B antibodies (**Figure III-14**).



mAb	*IGHV*	*IGKV*
2D10	*IGHV4-39*	***IGKV1-5***
KIRV10-OL19	*IGHV4-39*	***IGKV1-5***
FluB-104	*IGHV4-31*	***IGKV1-5***

mAb	CDRL3 alignment	CDRL2 alignment
2D10	QQYHSYSGT	DAS
KIRV10-OL19	QQYNIYSWT	KAS
FluB-104	QQYSSFPGT	KAS

Figure III-14: Possible Influenza B NA specific public clonotype
A possible light chain driven public clonotype identified using *IGKV1-5*. Biolayer interferometry was used to identify binding of all three antibodies to a similar site. Two tables on the right side of the figure indicate gene usages for each antibody as well as the sequence alignments of the CDRs. Residues indicated in red are conserved across all three antibodies.

DISCUSSION

The components of this chapter are snippets of projects that are currently all in progress in the Crowe lab.

In the first section, I describe a project based on a large panel of antibodies derived from an H3N2 natural infection plasmablast response that had been previously characterized. In addition to this, there had been large amounts of bulk sequencing data generated from the same donor capturing the plasmablast response to vaccination for five consecutive years. This was a unique opportunity to utilize the already characterized antibodies to index the bulk sequencing, thereby not only creating the ability to understand the dynamics of the antibody response throughout five years but being able to assign a function to each clonal lineage. Utilizing a chaining method allows for the retention of the original cluster to remain while grabbing additional sequencing into the same clonal family. Utilization of such work allowed for the revelation of two different antibody lineages: an antibody lineage targeting the trimer interface as well as a lineage targeting a cross-reactive epitope between M1 and M2 proteins. Not only do this class of trimer interface antibodies reoccur every year upon vaccination, but they are also found in multiple individuals, constituting them as a light chain-driven public clonotype. When administered prophylactically or therapeutically, these antibodies had protected mice against challenges against multiple influenza A viruses, indicating they may play a role in protecting individuals from severe disease during seasonal circulation of influenza A(Zost et al., 2021a). And that this class of antibodies is likely found commonly in different individuals as well, therefore possibly playing a role in population immunity year to year(Zost et al., 2021b). Another lineage of antibodies characterized is one that cross-reacts with influenza A M1 matrix protein and the M2 ectodomain. This lineage is recurrent in every year of vaccination sequencing as well and is likely to infer protection based on previous studies done by Grandea and colleagues(Grandea et al., 2010).

Having large sets of sequences and a clustering method in place allows for systematic selection on clones to synthesize and therefore test downstream. This prevents having multiple clones per clonal family from being tested at the screening level, therefore, saving on cost and time. Such workflows were applied to two projects: antibody discovery to H1N1 and Influenza B. Clustering the repertoire sequences to identify clonal families and selecting clonally expanded families as a method to identify antibodies to synthesize and characterized allowed us to maintain a positive hit rate of 50-60% in both projects.

Many aspects of influenza B humoral immunity is understudied compared to influenza A. Therefore, the large panel of influenza B specific antibodies revealed many classes of antibodies to be further characterized. Within the cross-reactive HA antibodies, there were two loosely defined antigenic sites, and within the cross-reactive NA antibodies, there were three loosely defined antigenic sites. From the large panel identified, there were many antibodies with neutralizing activity as well. Further characterization of these antibodies would allow for a complete understanding of the humoral immune response to influenza B.

Within the panel of influenza A specific antibodies, we had identified a couple that fall into the public clonotype trimer interface class of antibodies that were previously identified from an H3N2 infected donor. Additionally, another public clonotype is hypothesized specific to Influenza B NA, with antibodies from two separate projects. This reveals that with an increased amount of sequencing and functional characterization of antibodies, more public clonotypes may come to light and inform next-generation vaccine designs.

METHODS

Clustering. All sequences processed through PyIR to identify gene usages and CDR sequences(Soto et al., 2020). Then sequences are first binned by V_H, J_H, and CDRH3 length. Sequences are following clustered by 80% sequence identity on the CDRH3 nucleotide sequence. Following, they are sub-binned again by V_L and J_L gene usages.

Antibody gene synthesis. Sequences of selected mAbs were synthesized using a rapid high-throughput cDNA synthesis platform (Twist Bioscience) and subsequently cloned into an IgG1 monocistronic expression vector (designated as pTwist-mCis_G1) for mammalian cell culture mAb secretion. This vector contains an enhanced 2A sequence and GSG linker that allows simultaneous expression of mAb heavy- and light-chain genes from a single construct upon transfection(Chng et al., 2015).

High-throughput antibody production. Antibodies were expressed in “micro-scale” method, where 1mL of CHO cell cultures were transfected for each antibody using the Gibco ExpiCHO Expression System and a protocol for deep 96-well blocks (Thermo Fisher Scientific), as previously described(Gilchuk et al., 2020a). Synthesized antibody-encoding lyophilized DNA was reconstituted in OptiPro serum-free medium (OptiPro SFM) and used for transfection of ExpiCHO cell cultures into 96-deep-well blocks. For antibody purification, clarified culture supernatants were incubated with MabSelect SuRe resin (Cytiva), washed with PBS, eluted, buffer-exchanged into PBS using Zeba Spin Desalting Plates (Thermo Fisher Scientific) and stored at 4 °C until use.

ELISA binding screening assays. 384-well plates were coated with antigens at 1μg/mL and incubated at 4˚c O/N. Next day, plates were blocked with 2% non-fat dry milk 2% goat serum in PBS containing 0.05% Tween-20 (blocking buffer). Microscale expressed antibodies were diluted 1:20 using blocking buffer and applied as primary antibody. Bound antibodies were detected using goat anti-human IgG conjugated with horseradish peroxidase (Southern Biotech)

and TMB substrate (Thermo Fisher Scientific). 1N hydrochloric acid was used to stop reaction and absorbance was measured at 450 nm using a spectrophotometer.

Neutralization RTCA assay to influenza B. To screen for neutralizing activity, we used the xCelligence real time cell analysis system (RTCA) which assesses cellular impedance as a measure of neutralization. MDCK cells are plated the night before with 20,000 cells per well in 100 μl. The morning of, antibody and virus dilutions are made in infection media (with 4 μg/mL of TPCK trypsin for a final concentration of 2 μg/mL). Antibodies are diluted 1:10 using infection media creating 60μl of antibody dilution n to add to 60 μl of virus in infection media. Virus and antibody mixture was incubated at 37˚c for one hour. Following, 100 μl of the mixture was plated onto the cells and incubated at 33˚c for the remaining of the experiment. Experiment was stopped once CPE was identified.

Cell surface binding to M2e. A/Colorado/15/2014 M2 construct was transfected into HEK293F cells according to the manufacturer's protocol (Gibco). Cells were then harvested at 36 hours post transfection. Microscale expressed antibodies were diluted 1:10 in FACS buffer (2% ultra low IgG FBS and 1mM EDTA in DPBS without calcium and magnesium). 293F cells are diluted to 2-5 x 10^6/mL, and 10μl is added to each well. 40 μl of dilute antibody is then added to each well and incubated for 1 hr at 4˚c. Cells are then washed and secondary antibody is used to detect primary antibody binding (1:1000 diluted anti-human PE Biolegend). Cells are resuspended in 50μl of secondary antibody dilution and incubated at room temperature for 1hr. After incubation, cells are washed and resuspended in 20 μl for iQue flow cytometry.

HAI assay. Virus is serially diluted 2-fold with 50 μl of virus in each well. Turkey red blood cells are diluted to 2% in PBS and 12.5 μl of cells were used in each well in a 96-well V-bottom plate. The virus and RBC mixture is incubated at room temperature for 1hr. The 4 times agglutination dose (4AD) is identified at the drop off point for which dilution virus agglutination stops. From this antibody was diluted 1:10 in PBS and 50 μl of antibody was added in each well.

Virus was diluted to the 4AD determined earlier and 50 μl of virus is then added to each well. This antibody virus mixture is then incubated at room temp for 1hr. After incubation, agglutination inhibition is identified by a pellet at the bottom of the plate.

Neuraminidase inhibition assay. NA activity was assessed using the NA Fluor kit. Viruses or antigens are first diluted serially to determine dynamic range for each virus or antigen. Substrate is then added and the mixture is incubated at 37˚c for 1 hour. Stop solution (by the kit) is used to terminate the reaction and plates are read by exciting at 360 nm and emission at 450 nm. Once a dilution of virus is determined, a stock of the virus dilution is made and plated then added to the antibody. The virus and antibody mixture is incubated for 30 minutes at 37˚c. Following substrate is added to each well then incubated for 60 minutes at 37˚c. The reaction is then terminated using stop solution, and plates are red by exciting at 360nm and emission at 450 nm.

Building phylogenetic trees. Clonal families were identified through clustering. Each clonal family were then aligned to their assigned germline gene using Clustal Omega v1.2.0. Then we used the PHYLIP phylogenetic software package v3.697 to generate a maximum-likelihood tree from the aligned sequences using the DNAML program, using the sequence of the germline *IGHV* or *IGKV* gene as an outgroup.

PBMC and plasmablast isolation and repertoire sequencing. Studies were approved by the Vanderbilt University Medical Center Institutional Review Board. Peripheral blood was collected from a healthy donor (Donor 269) with prior history of many seasonal influenza vaccinations. For longitudinal repertoire sequencing, PBMCs from the donor were isolated by density gradient separation on Ficoll, cryopreserved and stored in liquid nitrogen storage until use. Total RNA was extracted from 10 million PBMCs. In some instances, a one-step RT-PCR was performed for 25 cycles using heavy chain BIOMED-2 variable antibody gene-specific primers as previously described (1-3) and the OneStep SuperScript III with Platinum® Taq High

Fidelity kit (Invitrogen, 11304011). The Illumina- specific adapters were added using the Illumina TruSeq Library Preparation Kit (Illumina, FC- 121-3001) according to the manufacturer's recommendations. The final amplicon libraries were sequenced on an Illumina MiSeq instrument using the MiSeq PE-300 v3 reagent kit (Illumina, MS-102-3001). Sequence analysis was performed using IG-BLAST v1.4, and results were parsed to MongoDB for further study. In other instances, we followed a previously described 5' RACE approach incorporating unique molecular identifiers (UMIs) for bulk un-paired B cell repertoire sequencing (4). Final libraries generated using this approach were sequenced in a symmetric (r1:300 cycles and r2: 300 cycles) or asymmetric (r1:30 cycles and r2:270 cycles) fashion using the MiSeq PE-300 v3 reagent kit (Illumina, MS-102-3001) or NovaSeq 6000 S1 reagent kit (Illumina, 20012863), respectively. For sequencing the plasmablast response to H3N2 infection, PBMCs were isolated upon natural H3N2 infection on day 7 from symptom onset. Approximately 2.2 x 10^7 PBMCs were stained in FACS buffer (D-PBS supplemented with 2% FBS and 1mM EDTA) with the following phenotyping antibodies; anti-CD19-FITC (1:20 dilution, eBioscience, 11-0199-42), anti-CD27-APC (1:20 dilution, BD Biosciences, 558664), and anti-CD38-PE (1:25 dilution, BD Biosciences, 555460). Cells were resuspended in sc- V_H:V_LSeq sequencing buffer (D-PBS supplemented with 0.04% non-acetylated BSA) containing propidium iodide as a viability dye. Approximately 28,000 viable $CD19^{Low}$ $CD27^{high}$ $CD38^{high}$ cells were sorted into sc-V_H:V_LSeq sequencing buffer. ~20,000 plasmablasts were carried through single-cell RNA sequencing using the 10X Genomics Chromium platform with enrichment using the 5' VDJ amplification kit (10X Genomics) according to manufacturer instructions. Amplicons were sequenced on an Illumina Novaseq 6000, and data were processed using the CellRanger software v3.1.0 (10X Genomics).

Biolayer interferometry competition-binding assays. Biolayer interferometry on an Octet HTX instrument (FortéBio) was used to perform competition-binding assays. Briefly, antigen and antibodies were diluted in D-PBS with 1% BSA and 0.05% Tween20. We first loaded either trimeric recombinant NA (B/Singapore/INFTT16-0610/2016) or HA (B/Iowa06/2017) onto tips at a concentration of 20 µg/mL. We then tested binding of two successively applied mAbs at 50 µg/mL. Competition was analyzed using the Octet analysis software (Data Analysis 9, FortéBio). Binding values were normalized to the binding signal measured in the absence of the first antibody, and self-self competition values were subtracted.

Quantification and statistical analysis. The descriptive statistics mean ± SEM or mean ± SD were determined for continuous variables as noted. Curves for antibody binding and neutralization were fitted after log transformation of antibody concentrations using non-linear regression analysis. Statistical analyses were performed using Prism v8.4.3 (GraphPad). Violin plots were generated with python3 seaborn library.

CHAPTER IV

UNDERSTANDING THE PRIVATE ANTIBODY RESPONSE TO THE EBOLA GLYCOPROTEIN

This chapter is an adaptation of the following manuscripts:

Chen E.C., Gilchuk P., Zost S.J., Ilinykh P.A., Binshtein E., Huang K., Myers L., Bonissone S.R., Day S., Kona C.R., Trivette A., Reidy J.X., Sutton R.E., Gainza C., Diaz S.M., Williams J.K., Selverian C., Davidson E., Saphire E.O., Doranz B.J., Castellana N., Burkreyev A., Carnahan R.H., Crowe J.E., Systematic analysis of human antibody response to ebolavirus glycoprotein reveals a high prevalence of neutralizing public clonotypes. *Under review.*

Gilchuk P., Guthals A., Bonissone S., Ilinykh P.A., Huang K., Soto C., Bombardi R., Bryan A., Davidson E., **Chen E.C.**, Dornaz B.J., Bukreyev A., Zeitlin L., Castellana N., Crowe J.E. Molecular analysis of serum antibody repertoire from a human survivor identified prevalence of potent antibodies against base region of Ebola virus glycoprotein. *Frontiers in Immunology (2021)*

CHAPTER OVERVIEW

This chapter details work on understanding the private antibody response to the EBOV GP. It will start off with a data set of paired single cell sequencing done on a single donor specific to the EBOV GP to guide the understanding of the antibody repertoire to the EBOV GP. Following identification of antibodies present in serum and the memory B-cell repertoire are identified. Additionally, different cross-reactive antibody lineages are identified and germline gene encoded neutralization properties are elicited. This work would not have been possible without the help of many individuals: Dr. Pavlo Gilchuk and Dr. Seth Zost for their guidance and help on sorting and sequencing the memory B cells; Dr. Erica Saphire for the contribution of drosophila cell line produced glycoprotein for cell sorting; Dr. Cinque Soto, Luke Myers, and Sam day for their guidance on the network plot and diversity calculations.

INTRODUCTION

The genetic and functional diversity of the memory B cell response and prevalence of public clonotypes to the EBOV GP remains unknown despite previous repertoire analysis of the B cell response in individuals following vaccination with rVSV-ZEBOV (Ehrhardt et al., 2019) and following natural infection (Davis et al., 2019). Single-cell RNAseq methods now allow for isolation of authentically-paired heavy and light chain antibody variable genes, retaining the ability to functionally assay all B cells sequenced. This approach enables functional validation and profiling of antibodies at a large scale. The B-cell repertoire induced by EBOV vaccination or infection is likely to be diverse but has not been comprehensively characterized as a large data set from a single donor with paired heavy and light chain variable genes. Therefore, using single-cell paired heavy and light chain sequencing of the memory B cell response to EBOV GP allows us to (1) define the paired sequence repertoire and therefore accurately estimate clonal diversity, (2) systematically select and characterize the functional diversity of repertoires, (3) understand evolution both on a genetic and functional basis, and (4) identify antibodies shared in the memory B-cell repertoire and sera and the functions of those antibodies.

A comprehensive in-depth understanding of the protective humoral response to EBOV GP on the repertoire level is important for devising optimal immunization schemes and informing the development of vaccines for multiple strains of Ebola (Cohen-Dvashi et al., 2020). In tandem, large-scale antibody studies could identify next-generation therapeutic antibody candidates. Such studies also can identify commonly induced antibodies that do not contribute to neutralization or protection, which is useful for building tools to benchmark the immunogenicity of new vaccine candidates. Understanding the genetic and functional diversity of the antibody repertoire to EBOV GP also can be used for benchmarking vaccine candidates.

Previous studies suggest that potent antibody response to GP early in convalescence is low, as potently neutralizing antibody responses appear later after recovery from infection (Davis et al., 2019; Williamson et al., 2019). This observation suggests that the neutralizing potency of antibodies to EBOV evolves through multiple rounds of affinity maturation during the process of somatic hypermutation. The antigenic landscape recognized by neutralizing antibodies also may evolve during convalescence. For instance, at early time points after recovery from EBOV infection, most mAbs isolated target the glycan cap of GP, suggesting glycan-cap-specific antibodies may play a dominant role in the early human antibody response to Ebola virus disease (EVD). A class of glycan-cap-specific antibodies are encoded by the *IGHV1-69* heavy chain gene, which specifies a germline-encoded complementarity-determining region 2 (CDRH2) with hydrophobic residues that facilitates binding to the glycan cap region (Murin et al., 2021). Therefore, germline-encoded *IGHV1-69*- antibodies likely play a role in the initial response to the GP. It has been shown that the functional profiles of several antibodies are retained when the somatically mutated sequences of such antibodies are reverted to the germline-encoded sequences (usually with minimal or no somatic mutations). Retention of function also has been reported for mAbs reverted in this way to germline-encoded sequences for other viral pathogens (Dong et al., 2021; Pappas et al., 2014; Yuan et al., 2020a; Zhou et al., 2015; Zost et al., 2021b). Identification of germline antibody genes encoding immunoglobulins with antiviral functional characteristics reveals a critical component of the early response to viral pathogens. Additionally, understanding how potent and cross-reactive antibodies evolve from germline-gene-encoded forms of those antibodies may inform rational vaccine design for the sites of vulnerability recognized by those mAbs (Rappuoli et al., 2016). Such studies, for example, define the critical residues and structures that should be retained in an immunogen to induce antibodies with the desired functionality.

Humoral immunological memory is partially mediated by serum antibodies secreted by long-lived plasma cells, which usually live in the bone marrow. In addition, memory B cells persist in circulation and are defined as long-lived and quiescent cells poised to quickly respond to antigen upon recall. Many antibody discovery efforts focus on memory B cells. Little is known about the composition of the polyclonal antibody secreted IgG repertoire and its overlap with the B cell receptors of memory B cells in EVD survivors. Defining the overlap of these two immune compartments could identify clonal families that contribute to the maintenance of protective humoral immunity.

To address this gap in knowledge, we sorted 100,000 EBOV GP-reactive memory B cells from a previously infected individual and performed large-scale single-cell antibody gene sequencing. These sequences were used for in-depth analysis of the paired sequence repertoire, clonal expansion, the molecular basis for antibody evolution.

RESULTS

Identification of EBOV GP-specific memory B cells

To identify EBOV GP-specific memory B cells, we took PBMCs from a convalescent donor with a history of EBOV infection in West Africa during the 2014 epidemic and performed a pan-B cell enrichment. Memory cells were then collected by flow cytometric sorting for IgM^- IgD^- $CD19^+$ B cells (**Figure IV-1A**, **B**). From this class-switched B cell population, we identified EBOV GP-reactive cells using biotinylated EBOV GP and fluorescently labeled streptavidin (**Figure IV-1B**). From 7 x 10^7 total CD19+ B cells, we sorted ~100,000 GP-reactive class-switched B cells; roughly 3% of the class-switched B cell population bound EBOV GP (**Figure IV-1C**). These GP-reactive cells then were single-cell sequenced using a single-cell encapsulation automated system (Chromium Controller; 10X Genomics). A total of 15,191 paired antibody heavy and light chain

variable region sequences were obtained from the single-cell sequencing experiment. As a control experiment in parallel, a PBMC sample from a healthy adult with no history of EBOV exposure was also subjected to the same workflow. For the control subject, IgM^- IgD^- $CD19^+$ B cells were sorted and subjected to single-cell sequencing to obtain 10,960 total paired heavy and light chain sequences for this non-specified-antigen-specific B cell set (**Figure IV-1C**).

Genetic characteristics of the memory B cell repertoire to EBOV GP

Immunoglobulin features such as inferred variable gene use and percent identity to inferred germline genes were identified using the PyIR informatics pipeline based on IgBLAST (Soto et al., 2020; Ye et al., 2013). To examine the different variable genes used in EBOV GP-specific antibodies that were captured from our GP-reactive B cell sorting experiment, the frequency of each *IGHV* or *IGKV/IGLV* gene used was measured and normalized to the total number of sequences acquired, defined as 100%. The same analysis also was performed on the 10,960 total paired sequences for B cells obtained from the control subject (**Figure IV-1C**). *IGHV1-69* was the most frequently used heavy chain variable gene (9.2%), followed by *IGHV1-02* (6.3%) and *IGHV4-34* (7.5%) (**Figure IV-1E**). *IGKV1-39* was the most frequently used light chain variable gene (12.2%). Gene use for the non-specified-antigen-specific antibodies from the control subject are plotted for comparison. For the EBOV GP-specific repertoire the median amino acid length of CDRH3 was 17, and 9 for or CDRL3 . In comparison, the median amino acid lengths of CDRH3 or CDRL3 for the non-specified-antigen-specific repertoire were 15 or 9, respectively. For the EBOV GP-specific repertoire, the average identity to germline was 94.2% for the heavy chain and 96.1% for the light chain, with the median number of mutations being 16 or 10 nucleotides respectively. As a comparison, in the non-specified-antigen-specific repertoire, the average identity to germline was 95.6% for heavy chain and 97.2% for light chain, with the median number

of mutations being 12 or 7 nucleotides, respectively. These findings indicate that the EBOV GP-specific repertoire is slightly more mutated than the non-specified-antigen repertoire, with slightly longer CDR3s in heavy and light chains.

Identification of clonal families

To identify antibody clonal families , sequences were clustered by binning based on the inferred immunoglobulin heavy variable (*IGHV*) gene, immunoglobulin heavy joining (*IGHJ*) gene, and the amino acid length of the CDRH3. Then, sequences were clustered according to 80% nucleotide identity in the CDRH3 sequence. Next, sequences were binned further based on the inferred immunoglobulin light variable gene (*IGLV* or *IGKV*) and immunoglobulin light joining (*IGLJ* and *IGKJ*) genes and 80% nucleotide identity in the CDRL3 sequence. We identified 10,087 cloNal families from the 15,191 total paired sequences derived from our EBOV GP-specific sort. Of these, 6,923 were singlets (meaning no other sequences clustered with that single sequence). Additionally, 2,382 were doublets, meaning two sequences clustered together, but did not cluster with any other sequence. We defined clusters as clonally expanded families if they included five or more sequences and found 224 such clonal families. To compare the distribution of EBOV GP-reactive clonal families to those of the control subject, we applied the same clustering scheme to the non-antigen-specific sequence set (**Figure IV-2B**). From the 10,960 sequences in that individual, 10,527 clonal families were identified. From that set, 10,172 were singlets and 305 were doublets. Only 21 clonal families clonally expanded were observed in the control subject.

We next estimated the size and diversity of the EBOV GP-reactive memory B cell repertoire in the convalescent donor using rarefaction analysis (Saary et al., 2017). For this we used the clonal families determined through the clustering scheme described above as taxonomic units or species. We first plotted species richness that is present in the defined sample set; species richness measures the number of species, or in our case clonal families (**Figure IV-1D**).

The species richness curve continually increased but never plateaued, suggesting that even at this substantial depth of sequencing, we did not identify all clonal families in this sample. The same approach was used for the sequences from the control subject, for whom the species richness curve trended steeper than the EBOV GP-specific curve. This comparison indicates the EBOV GP-specific sequence set is less diverse than the non-GP-specific sequence set that was captured in our experiments.

Next, we calculated the Chao diversity index (Chao1) (Hsieh et al., 2016), which estimates the number of species there is likely to be in the sample set. The Chao diversity index for the EBOV GP-reactive antibody sequences at a sample depth of 90% of the repertoire was 20,329. This value suggests an estimated 20,329 clonal families are present in the EBOV GP-specific antibody repertoire within this donor, of which we identified 10,087 (~50% of the total). The estimated total of the non-antigen-specific antibody sequence set at a sample depth of 90% is 177,374, from which we have identified 10,527 clonal families. However, the number for the non-GP-specific repertoire is not a confident estimate as this value grows at every new sample depth, and it is likely that the number of clonotypes in a non-GP-specific repertoire is much higher than estimated here. In contrast, the estimated number of clonal families stayed consistent with increasing depth for the EBOV GP-reactive antibody sequence set estimate, giving confidence in the estimated number of total species. This outcome can be visualized in **Figure IV-1D,** as the EBOV GP-specific repertoire unique species plot exhibits plateauing at a much lower sample depth. With more than a ten-fold difference in estimated diversity, this finding indicates that the EBOV GP-reactive antibody repertoire is much less diverse and has fewer clonal families compared to a GP-antigen-specific repertoire. Therefore, through modeling, we predicted that we have sampled about half of the EBOV GP-specific memory B-cell repertoire in the convalescent donor, as we have identified 10,087 clonal families out of the 20,329 estimated.

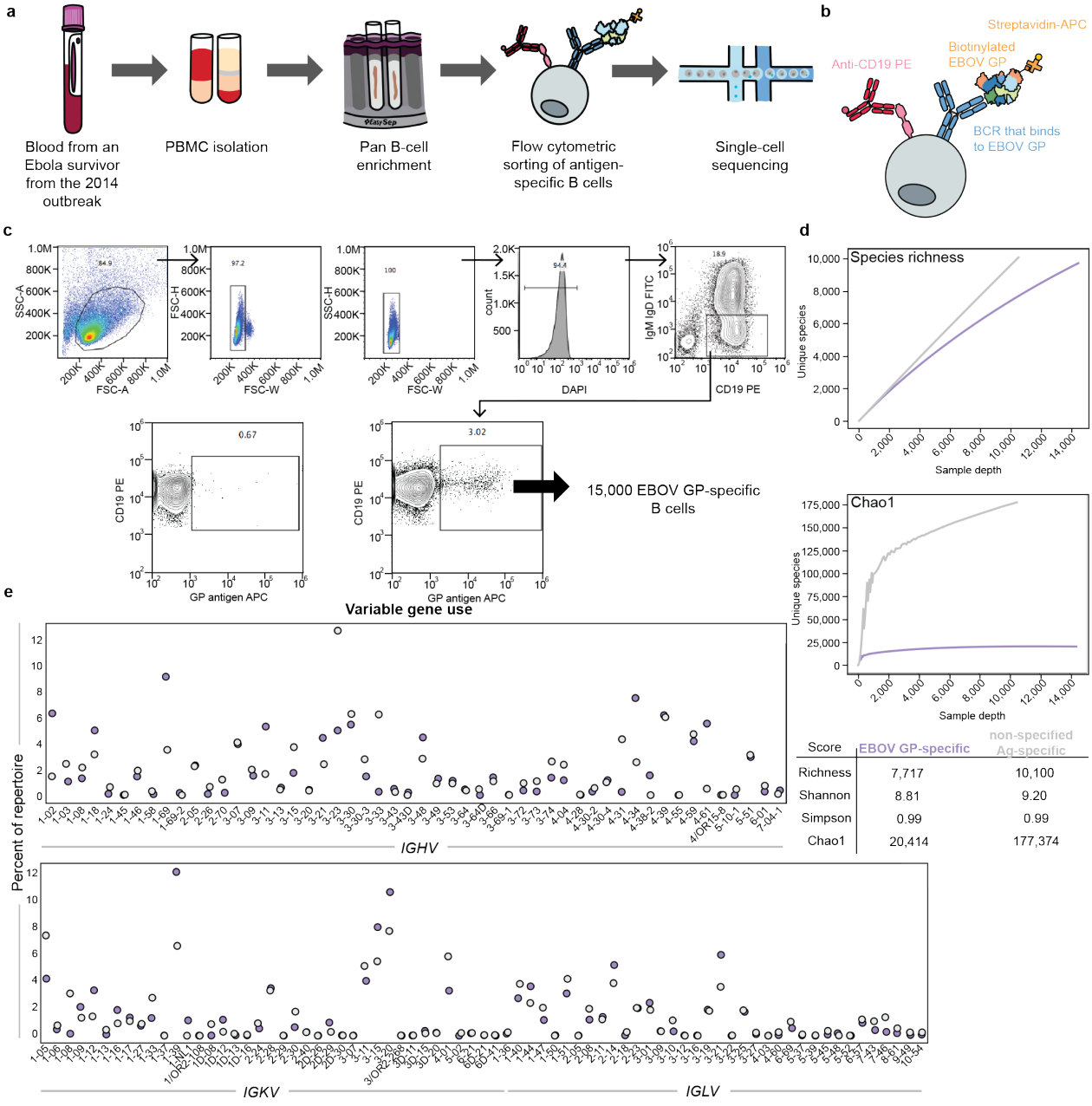


Figure IV-1: Identification of and diversity of EBOV GP specific memory B cells.

(A) Schematic of sample processing to identify and sort memory B cells. **(B)** Schematic of flow cytometric staining to identify EBOV GP specific B cells. **(C)** Gating for lymphocytes, singlets, live cells using DAPI, followed by class switched B cells. Cells were stained with anti-CD19 antibody conjugated to PE and anti-IgM and anti-IgD conjugated to FITC. EBOV GP was biotinylated and conjugated to streptavidin APC. FACS isolation of class-switched B cells ($CD19^+$ IgM^- IgD^-) specific to the EBOV GP (Antigen-APC) in a donor that has not previously been exposed to EBOV (left) or the convalescent donor (right) shown.
(Legend continued on next page.)

Figure IV-1 legend continued: **(D)** Diversity metrics calculated for the EBOV GP-specific repertoire (purple) compared to the non-antigen-specific repertoire (grey). The first plot shows species richness and the second shows Chao diversity. The sample depth on the x-axis indicates the number of sequences, and the unique species on the y-axis indicates the number of clonal families. Additional diversity metrics were calculated including Shannon entropy and Simpson index. **(E)** Variable gene usages in the heavy chain (top) or light chain (bottom) repertoire from sequencing. The number of sequences using each gene was calculated and normalized to a percentage using the total number of sequences as 100%. Purple dots indicate the EBOV GP-specific repertoire, grey dots indicate the non-antigen-specific repertoire.

Functional characterization of clonally expanded EBOV GP-specific repertoire

To understand the functionality of each of the 224 clonally-expanded families from the EBOV immune subject, the most somatically mutated member of each family was selected for antibody gene cDNA synthesis and recombinant IgG expression (**Figure IV-2A**), using previously described microscale production and purification approaches (Gilchuk et al., 2020a). Purified mAbs then were tested for binding in ELISA to recombinant trimeric EBOV GP, BDBV GP, SUDV GP, or MARV GP ΔTM proteins. Next, they were tested for binding to cleaved or intact GP displayed on Jurkat cells using a flow cytometric assay (Gilchuk et al., 2020b). The full-length membrane-bound EBOV GP molecules expressed on the surface of Jurkat cells likely are similar to the native form of GP on the surface of a viral particle or on naturally-infected cells (Davis et al., 2019). Lastly, neutralizing activity of the antibodies was assessed using real-time cell analysis (RTCA) assay, allowing for quantification of cytopathic effect induced by replication-competent VSV-EBOV GP (Gilchuk et al., 2020a; Zost et al., 2020b). The tests showed that 80% of mAbs recognized EBOV GP, 60% recognized BDBV GP, 30% recognized SUDV GP, and 2% recognized MARV GP. 11% of mAbs recognized the three lethal strains of ebolavirus: EBOV, BDBV, and SUDV. Also, 30% of mAbs preferentially bound to cleaved GP, while 34% of mAbs preferentially bound to intact GP. Lastly, 95 mAbs neutralized VSV-EBOV GP (**Figure IV-3**). In

conclusion, 80% of the clonally-expanded repertoire selected reacted with EBOV GP, with 42% showing neutralizing properties to VSV-EBOV.

Building a network of the clonally expanded population

As there are multiple antibodies within each clonal family/lineage, we sought to visualize the relationships of clones between and within lineages. We built a network combining the genetic similarities of antibodies within lineages, between different lineages, and the functional profile of each lineage through the experimental data determined from the experiments described above.

To retain paired heavy and light chain sequence information, we selected matching CDRH3 and CDRL3 amino acid sequences and linked the two with an arbitrary string, allowing for the resulting single string to be used as a node, representing a single antibody. A centroid also was computed for each lineage of antibodies using Vsearch to represent the average CDRH3 and CDRL3 sequence for each clonal family. Hamming distances were then calculated from each of the linked antibody CDR3 sequences to the calculated centroids to investigate the relationships of antibodies within clonal families. As using Levenshtein distance accommodates for the different CDR3 lengths that can be present between different clonal families/lineages, the Levenshtein distances were calculated between each centroid representative of each clonal family to investigate the relationship between clonal families (**Figure IV-2C**). The functional characteristics of each clonal family (**Figure IV-3**) then were mapped onto the network, with all the clonal families colored by functional profile. Here, we can visualize the diversity of functional phenotypes within the clonally expanded repertoire (**Figure IV-2C**) in conjunction with genetic similarities of antibodies within and between clonal families revealing that there is a large set of neutralizing antibodies that preferentially bound to the cleaved GP, but the CDR3 similarities of these antibodies vary. This plot revealed a cluster of cross-reactive neutralizing antibodies with similar

CDR3s on the upper left portion of the network plot (**Figure 2C**). Understanding such relationships could be useful in predicting antibody function through sequence analysis.

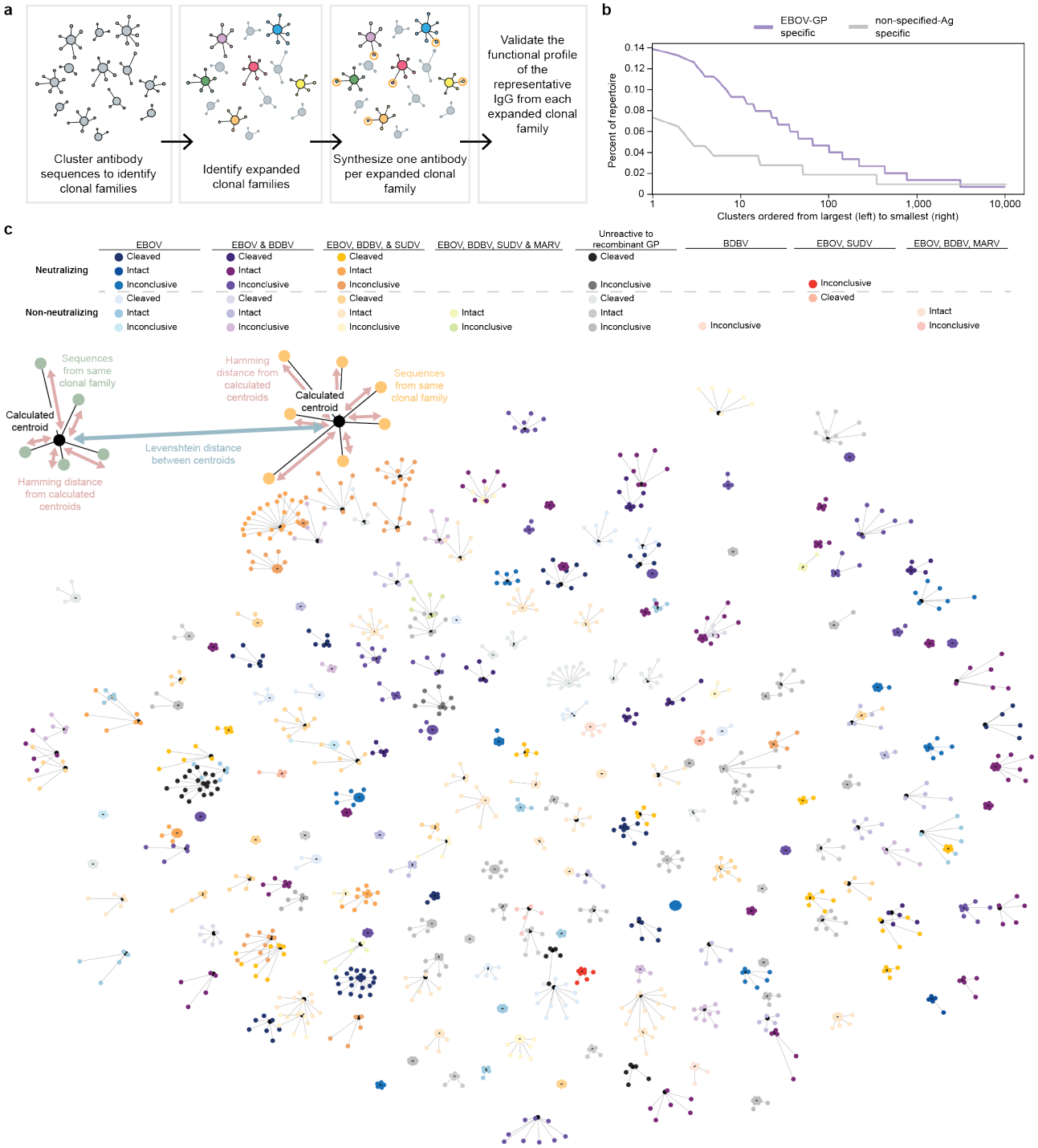


Figure IV-2: Analysis of the clonally expanded EBOV GP specific repertoire

(Legend on next page)

Figure IV-2 Analysis of the clonally expanded EBOV GP specific repertoire figure legend:
(A)Schematic of identification of clonally-expanded families and selection of one clone per family. (B)Graph showing the distribution of clonal families. After clustering was completed, clusters were ordered from the largest cluster to smallest cluster, then plotted in that order as the percent of the EBOV GP-specific repertoire (purple) and non-antigen-specific repertoire (grey). (C)Network plot of the clonally-expanded repertoire with the functional characteristics of each clonal family plotted on and schematic showing how calculations were derived to construct network diagram. Reactivity to each glycoprotein in ELISA is denoted by different colors. Blue indicates antibodies monospecific to EBOV, purple indicates antibodies specific to EBOV and BDBV, orange indicates antibodies specific to EBOV, BDBV, and SUDV. Green indicates antibodies specific to EBOV, BDBV, SUDV, and MARV. Grey antibodies did not react with any GP tested in ELISA. Salmon indicates antibodies specific to BDBV only. Red indicates specificity for EBOV and SUDV. Pink indicates specificity for EBOV, BDBV, and MARV. Different shades of each color indicate neutralizing capacity, with the darker dots indicating neutralizing antibodies for VSV-EBOV and the lighter dots indicating non-neutralizing antibodies. Different shades within each color represent whether the antibody preferably bound to the intact GP (GP_{ecto}) or cleaved GP (GP_{cl}), or if it bound well to both.

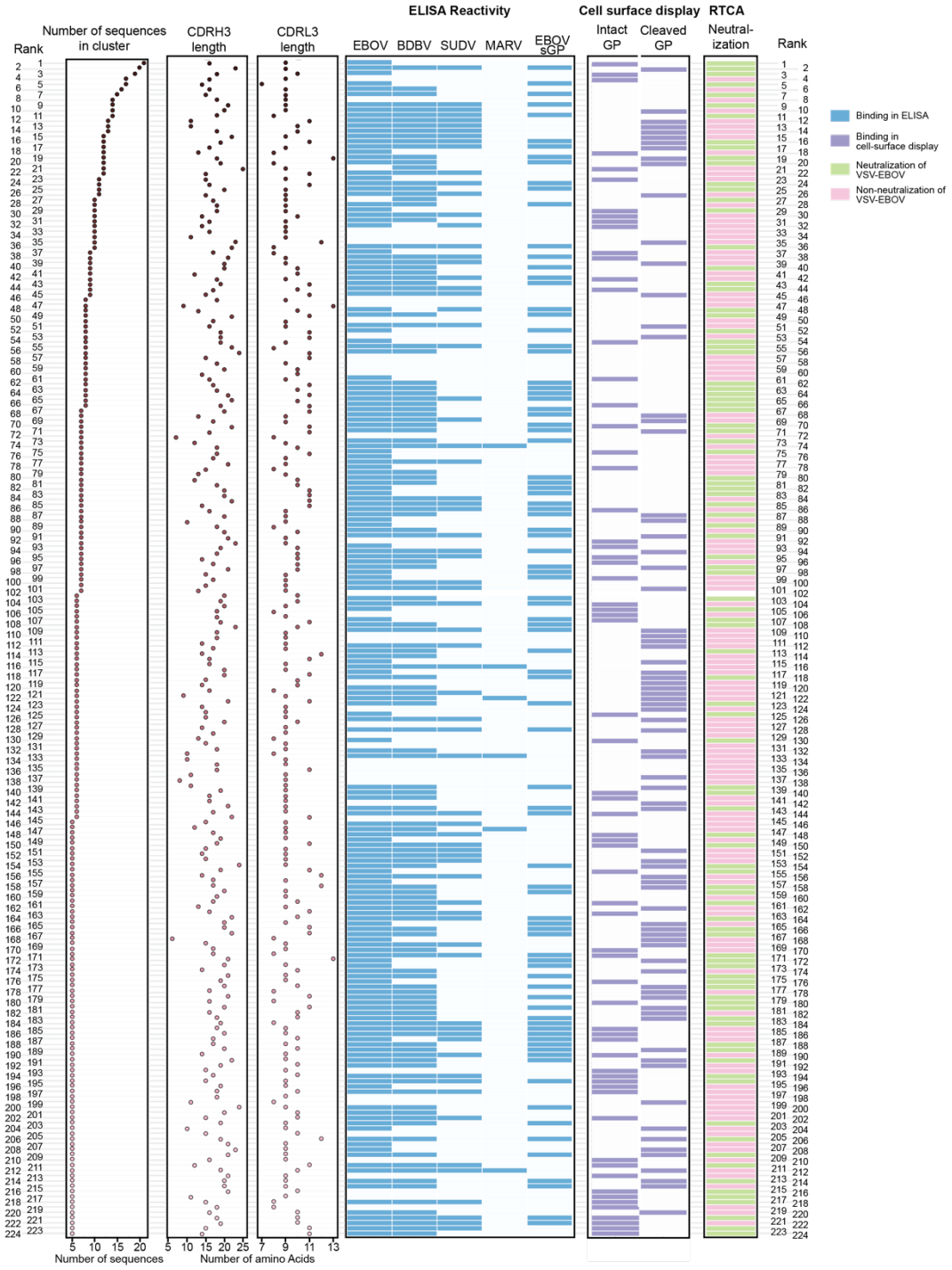


Figure IV-3: Functional characteristics of the EBOV GP-specific clonally-expanded repertoire.

The functional characteristics of all 224 clonally-expanded antibodies are listed in this table. The first column shows the number of antibodies in the cluster, the second column shows the CDRH3 amino acid length, the third column shows the CDRL3 amino acid length. Blue boxes indicate binding to GPs in ELISA, purple indicates binding in a cell-surface GP display assay. Green indicates neutralizing activity for VSV-EBOV, and pink indicates lack of detectable neutralizing activity.

Overlap in the repertoire between B cell receptors encoded in the memory B cell population and immunoglobulins present in plasma.

We recently described a proteo-genomic analysis for identifying EBOV-specific immunoglobulin proteins in convalescent human plasma from the same donor we used in this study (Gilchuk et al., 2021). EBOV GP-specific polyclonal antibodies from the donor plasma were purified and subjected to high-resolution liquid chromatography coupled to tandem mass-spectrometry, yielding sequences of antibody proteins present in plasma. A subset of 1,512 EBOV GP-specific memory B cell antibody variable gene sequences was used for the original study. Here, since we obtained a much (10-fold) larger memory B cell repertoire from heavy and light chain paired sequences from this individual, we reinvestigated the portion of antibodies shared between the plasma immunoglobulin protein and memory B cell receptor repertoires.

Immunoglobulins in the plasma repertoire were identified as present if over 50% coverage in the CDR3 region and a general peptide coverage of over 100% of the CDR3, as previously described (Gilchuk et al., 2021). Despite the 10-fold increase of gene sequences against which we could search, we found only an additional 82 antibodies, giving a total of 153 antibodies present in the overlap of antibodies in the plasma and the memory B cell repertoire. These 153 clones belonged to 106 clonal families. A subset of 24 was from clonally-expanded clusters, and 44 were from singlets (**Figure IV-4A**). Of the 24 plasma antibodies that came from clonally-expanded families, 17 antibodies neutralized VSV-EBOV GP. It had been previously described that polyclonal IgG isolated from convalescent plasma demonstrates preferential binding to cleaved GP(Davis et al., 2019; Gilchuk et al., 2021). However, of the 24 plasma antibodies from clonally-expanded families, 16 of 24 bound preferentially to intact GP (**Figure IV-4B**). Therefore, it is likely that many of the polyclonal antibodies found in the plasma come from non-clonally-expanded memory B cell families. When identifying the V_H gene usages of these serum-identified antibodies, the highest used genes were *IGHV1-2* and *IGHV3-11* at 10%. Following, *IGHV4-34*,

IGHV3-21, *IGHV1-69*, and *IGHV4-59* at 8%, 8% 7%, and 6% of the total serum antibodies identified respectively. Additionally, despite the large number of gene sequences used here as a reference set for the proteomics studies, there appears to be a relatively small overlap between antibodies in the plasma and B cell receptors in the circulating memory B cell population.

Unmutated common ancestors of expanded clones reveal germline reactivity of clone encoded by *IGHV1-69* or *IGHV1-02*.

Although we obtained unprecedented depth for paired heavy and light chain variable gene sequencing from single antigen-specific B cells from this donor, the depth of sequencing that can be acquired by bulk heavy or light chain antibody variable gene sequencing (without pairing) is still far superior. Therefore, to investigate the evolution of cross-reactive antibodies, we clustered sequences obtained from single-cell paired sequencing with those obtained from bulk sequencing from the same leukapheresis sample PBMCs from this donor and constructed phylogenetic trees detailing the evolution of these cross-reactive antibodies versus that of monospecific antibodies.

Several clonally-expanded neutralizing antibodies with varying reactivities to the different GP and predicted epitopes were selected for further investigation of their evolution. To increase the amount of sequencing for phylogenetic analysis, heavy and light chain bulk sequencing was performed on PBMCs originating from this donor without antigen-specific sorting. The heavy chains from all clonal families previously identified in the single-cell sequencing were then clustered with those in the heavy chain bulk sequencing, and the light chains were clustered similarly. Sequences were clustered based on the same V and J gene usage as well as 80% identity of the CDR3 nucleotide sequences. Using the clustered sequences for each clonal family, maximum likelihood trees were constructed for both heavy and light chains, and the unmutated common ancestor (UCA) was inferred for each clonal family.

To determine whether the antiviral function of each clonal family was due to germline-encoded reactivity or due to somatic mutations that accumulated during antibody evolution, we investigated the binding and neutralization profile for each inferred UCA. Each UCA antibody was expressed and tested for binding to EBOV, BDBV, or SUDV GP, and neutralization against VSV-EBOV GP, VSV-BDBV GP, or VSV-SUDV GP. Of the 18 UCAs tested, 12 lost their ability to bind to the different GPs in comparison to their mutated counterparts (**Figure IV-4C**).

Two UCA antibodies not only still bound to the appropriate GPs but also maintained their ability to neutralize VSV-EBOV GP, albeit with lower potency. EBOV-888-UCA uses *IGHV3-11*/*IGKV1-39*. EBOV-874-UCA uses *IGHV4-39*/*IGKV3-15* and maintained capacity to neutralize VSV-EBOV GP with reduced potency but lost its ability to neutralize VSV-BDBV GP. Therefore, it is likely that these two gene combinations contribute to germline-encoded neutralization properties specific to EBOV.

Additionally, four UCA antibodies retained the ability to mediate cross-reactive neutralization: EBOV-879-UCA, EBOV-872-UCA, EBOV-591-UCA, and EBOV-967-UCA. EBOV-879-UCA is encoded by *IGHV1-69/IGKV3-20* and neutralized VSV-EBOV GP, -BDBV GP, and -SUDV GP. These data show that antibodies encoded by *IGHV1-69/IGKV3-20* possess germline-encoded capacity to neutralize across all three medically important ebolavirus species, and they acquired increased potency during the process of somatic hypermutation (**Figure IV-4D, E**). EBOV-872-UCA uses *IGHV1-2/IGKV3-20* and maintained neutralization against VSV-EBOV GP and VSV-BDBV GP. As EBOV-872-UCA lost its ability to neutralize VSV-SUDV GP, EBOV-872 likely acquired the capacity to neutralize SUDV by acquiring somatic mutations (**Figure IV-4D, F**). EBOV-967-UCA also maintained its cross-reactive neutralizing activity for both EBOV and BDBV, even though its neutralization potency for BDBV is relatively low at 50 μg/mL (**Figure IV-4D**). Therefore, it is likely that these antibodies initially neutralized EBOV with weak inhibition of BDBV but evolved to gain potency for the two strains (**Figure IV-4G**). EBOV-591-UCA also

retained its ability to neutralize all three strains, however, its potency to SUDV dropped substantially. EBOV-591 also is encoded by *IGHV1-2* but uses a different light chain gene, *IGKV4-1* (**Figure IV-4D, H**).

EBOV-967 uses *IGHV1-2/IGKV3-20*, the same V_H and J_H genes as EBOV-872, however they differ in their J_L gene usage. As both of these antibodies neutralized EBOV and BDBV, *IGHV1-2/IGHJ3* in combination with *IGKV3-20* likely encodes for neutralization of EBOV and BDBV (**Figure IV-4C, D**). We note all UCA antibodies that neutralized virus targeted the glycan cap region of the GP, since they competed for binding with the glycan cap antibody 13C6. These results indicate that germline-encoded structural features contribute to the ability of these antibodies to neutralize virus (**Figure IV-4C**).

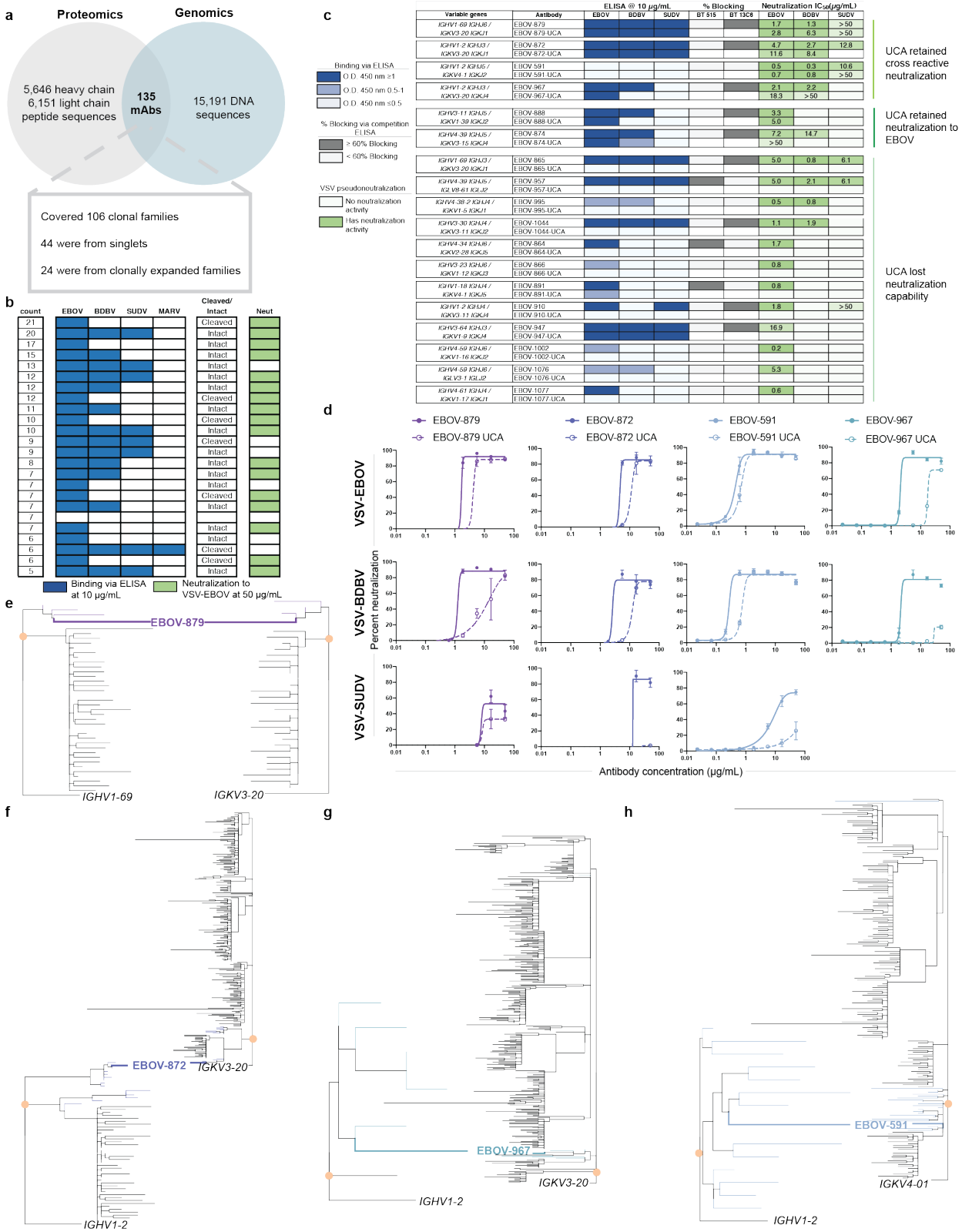


Figure IV-4: Characteristics of plasma antibodies and unmutated common ancestors of clonally-expanded antibodies.
(Legend on the next page.)

Figure IV-4 legend: **(A)** Venn diagram detailing the identification of antibodies present in both plasma and the memory B-cell repertoire. **(B)** Characteristics of antibodies present in both the plasma and memory B-cell repertoire that originated from clonally-expanded families. The number of antibodies present in the clonal family is shown in the first column, followed by blue color indicating binding reactivity in ELISA to the different GPs, followed by binding to cleaved or intact GP, followed by neutralization for VSV-EBOV indicated in green. Experiments were performed in biological duplicate, and the compilation of replicates is shown. **(C)** Antibodies and their inferred UCAs with the functional profiles of each antibody. Gene use is listed in the first column followed by antibody names. Antibodies are listed with the mutated version of the antibody on the top row and the unmutated common ancestor (UCA) version on the bottom. The blue boxes indicate binding in ELISA to the different GPs at 10 μg/mL of antibody. The grey boxes indicate percent blocking in competition-binding ELISA against biotinylated EBOV-515, a base-region-specific reference antibody or against the glycan cap reference antibody 13C6. The green boxes indicate neutralization for VSV-EBOV, -BDBV, or -SUDV. The numbers inside the boxes indicate the calculated IC_{50} value for each antibody. Experiments were performed in biological duplicate and technical triplicates with similar results. A biological replicate from a single experiment is shown. **(D)** Neutralization curves of unmutated common ancestor antibodies (dotted lines) that retained cross-reactive neutralization and their mutated counterparts (solid lines) against VSV-EBOV, -BDBV, or -SUDV. Experiments were performed in biological duplicate and technical triplicates with similar results. A biological replicate from a single experiment is shown. **(E)** Maximum likelihood phylogenetic tree of the EBOV-879 lineage. The inferred UCA is indicated in the orange circle on the heavy chain and light chain tree. Blue lines indicate antibody sequences that were found in paired chain sequencing; black lines indicate sequences that were found in unpaired chain bulk sequencing that clustered with the clonal family. **(F)** Maximum likelihood phylogenetic tree of the EBOV-872 lineage. The inferred UCA is indicated in the orange circle on the heavy chain and light chain tree. Blue lines indicate those antibodies found in paired sequencing; black lines are bulk sequences that clustered with the clonal family. **(G)** Maximum likelihood phylogenetic tree of the EBOV-967 lineage. The inferred UCA is indicated in the orange circle on the heavy chain and light chain tree. Blue lines indicate those antibodies found in paired sequencing; black lines are bulk sequences that clustered with the clonal family. **(H)** Maximum likelihood phylogenetic tree of the EBOV-591 lineage. The inferred UCA is indicated in the orange circle on the heavy chain and light chain tree. Blue lines indicate those antibodies found in paired sequencing; black lines are bulk sequences that clustered with the clonal family.

DISCUSSION

The size of the human B cell response to a pathogenic virus protective antigen has not been defined. Previous work has established that the overall circulating repertoire of each individual contains around 11 million or more B cell clonotypes defined by V_H, J_H, and CDR3 amino acid sequence using bulk sequencing data (Briney et al., 2019; Soto et al., 2019), however, the number and diversity of B cells specific to viral antigens is poorly understood at the paired

sequence level. Here we present the largest individual antigen-specific repertoire from a single sample reported, to estimate the size and complexity of an individual's response to a virus. After sorting 100,000 GP-specific B cells, we recovered paired antibody genes for over 15,000 clones and found over 10,000 clonotypes in that repertoire. Species richness calculations estimate that the individual's sample contained over 20,000 clonotypes reactive with EBOV GP at the time point tested, about 9 months after infection. It should be noted that each of those >20,000 clonotypes contains many somatic variants (for instance, the largest clonal family we recovered had 21 sequences within its single lineage in this study). Thus, the size and complexity of the response to a single viral protein is enormous.

A strength of the antibody discovery approach used here was that we not only obtained variable gene sequences, but also those sequences were authentically paired heavy and light chain sequences from single cells. This approach allowed us to express representative naturally occurring mAbs of each of the clonotypes of interest so that we could validate their specificity and define their cross-reactivity and neutralizing potency. Here, we observed that 45% of the clonally expanded antibody repertoire neutralized EBOV. About two-thirds of the neutralizing clones targeted the glycan cap region of the GP. This finding shows that, even though many have considered the glycan cap a poor target for protective responses, most neutralizing antibodies in the clonally expanded repertoire target the glycan cap. Additionally, we mapped functional characteristics of each clonally expanded family to genetic similarities of antibodies not only within clonal families but also between clonal families at a scale previously unseen, allowing for visualization of clustering of functionally similar antibodies with genetically similar CDR3s.

The nature of future ebolavirus epidemics cannot be predicted, and therefore it is important to understand how cross-reactive neutralizing antibodies arise in response to the virus. For the cross-reactive neutralizing antibodies identified that recognized multiple ebolavirus species, we investigated if that cross-reactive neutralizing activity was germline-encoded or

acquired through acquisition of somatic mutations. The studies of UCAs revealed that the *IGHV1-69* and *IGHV1-2* heavy chain variable gene segments can encode cross-reactive neutralizing antibodies, suggesting the origin of heterologous immunity in individuals infected with one ebolavirus species.

Antibodies are not secreted by circulating memory B cells but rather by long-lived plasma cells in the bone marrow. We have recently described the antibody response in convalescent plasma within the same donor (Gilchuk et al., 2021). However, little is known about the diversity of antibodies that overlap between the memory B cell repertoire and the plasma antibody repertoire. Previous data had shown that in plasma, there is preferential recognition of the cleaved EBOV GP. However, in our study, it is interesting that when narrowing the antibody characteristics of plasma antibodies identified from clonally expanded families, these antibodies preferentially recognized the intact GP. Therefore, it seems likely that the bulk of plasma antibodies could derive from specificities less common in circulating memory B cells (noted as singlets in our single cell RNAseq repertoire) where a lot of the reactivity for cleaved GP would reside. A technical limitation of plasma proteomic antibody studies is that likely only highly represented antibodies are detected, and therefore this dataset is likely lacking antibodies present at low levels in the plasma.

METHODS

Research participants. Human PBMCs and plasma were obtained at Vanderbilt University Medical Center in Nashville, TN, USA, from a survivor of the 2014 EVD epidemic after written informed consent. The studies were approved by the Vanderbilt University Medical Center Institutional Review Board. PBMCs and plasma were collected after the illness had resolved. The donor is a male human survivor of the 2014 EVD outbreak in Nigeria and was 31 years of age when infected, and 32 when PBMCs and plasma were collected 15 months later. At the time of blood collection, plasma samples were tested by qRT-PCR and found to be negative for the presence of viral RNA.

Cell lines. Vero-E6, Jurkat, Vero CCL-81, and THP-1 cells were obtained from the American Type Culture Collection (ATCC). Vero-E6 cells were cultured in Minimal Essential Medium (MEM) (Thermo Fisher Scientific) supplemented with 10% fetal bovine serum (FBS; HyClone) and 1% penicillin-streptomycin in 5% CO_2, at 37°C. ExpiCHO (hamster, female origin) and FreeStyle 293F cell lines were purchased from Thermo Fisher Scientific and cultured according to the manufacturer's protocol. The Jurkat-EBOV GP (Makona variant) cell line stably transduced to display EBOV GP on the surface(Davis et al., 2019) was a kind gift from Carl Davis (Emory University, Atlanta, GA). Jurkat-EBOV GP and THP-1 cells were cultured in RPMI 1640 (Gibco) medium supplemented with 10% FBS and 1% penicillin-streptomycin in 5% CO2, at 37°C.

Viruses. The mouse-adapted EBOV Mayinga variant (EBOV-MA, GenBank: AF49101) (Bray et al., 1998), authentic EBOV Mayinga variant expressing eGFP(Towner et al., 2005), infectious vesicular stomatitis virus rVSV/EBOV GP (Mayinga variant), rVSV/BDBV GP, rVSV/SUDV GP(Garbutt et al., 2004), and chimeric EBOV/BDBV-GP and EBOV/SUDV-GP(Ilinykh et al., 2016) were used for mouse challenge studies or neutralization assays. Viruses were grown and titrated in Vero-E6 cell monolayer cultures.

GP expression and purification. For B cell labeling, and flow cytometric sorting, we used EBOV GP produced in *Drosophila Schneider* 2 (S2) cells. Briefly recombinant ectodomain of EBOV GP ΔTM in a modified pMTpuro vector was transfected into S2 cells followed by stable selection of transfected cells with 6 μg/mL of puromycin. GP ectodomain expression was induced with 0.5 mM $CuSO_4$ for 4 days. Protein was purified using Step-Tactin resin (Qiagen) via an engineered strep II tag and purified further by Superdex 200 (S200) column chromatography. For ELISA studies, the ectodomains of EBOV GP ΔTM (residues 1-636; strain Makona; GenBank KM233070), BDBV GP ΔTM (residues 1- 643; strain 200706291 Uganda; GenBank: NC_014373), SUDV GP ΔTM (residues 1-637; strain Gulu; GenBank: NC_006432), and MARV GP ΔTM (residues 1-648; strain Angola2005; GenBank: DQ447653) were expressed using the FreeStyle 293F cell line and purified as described (Gilchuk et al., 2018).

Memory B cell isolation and flow cytometric analysis. PBMCs from a leukopak were isolated with Ficoll-Histopaque by density gradient centrifugation. The cells were cryopreserved in the vapor phase of liquid nitrogen until use. Total B cells were enriched by negative selection from PBMCs using EasySep Human Pan-B Cell Enrichment Kit (StemCell Technologies). Enriched cells were stained on ice in Robosep buffer containing the following phenotyping antibodies: anti-Human CD19-PE, anti-IgM-FITC, anti-IgD-FITC. The EBOV GP-reactive memory B cells were labeled with recombinant EBOV GP protein that was produced in Drosophila S2 cells as described above and purified by flow cytometric cell sorting using an SH800 cell sorter (Sony) as described previously (Gilchuk et al., 2020b). Approximately 100,000 cells were FACS-sorted in bulk for downstream paired antibody heavy and light chain variable gene sequence analysis.

Generation of antibody variable-gene libraries from single B cells. For paired antibody variable gene sequence analysis, cells were resuspended into DPBS containing 0.04% non-acetylated BSA, split into four replicates, and separately added to 50 µL of RT Reagent Mix, 5.9 µL of Poly-dt RT Primer, 2.4 µL of Additive A and 10 µL of RT Enzyme Mix B to complete the

Reaction Mix as per the vendor's protocol. The reactions then were loaded onto a Chromium chip (10x Genomics). Chromium Single Cell V(D)J B-Cell-enriched libraries were generated, quantified, normalized and sequenced according to the User Guide for Chromium Single Cell V(D)J Reagents kit (CG000086_REV C). Amplicons were sequenced on an Illumina Novaseq 6000, and data were processed using the CellRanger software v3.1.0 (10x Genomics).

Bulk sequence analysis of antibody variable region genes. Total RNA was extracted from approximately 5,000 B cells using the Qiagen RNeasy Micro kit following the manufacturer's recommendations (Qiagen). To maximize target enrichment recovery, we employed two separate library preparation approaches with three separate primer mixes to avoid any individual primer set's amplification bias. In the first library preparation approach, we used the OneStep SuperScript III Platinum®Taq High Fidelity kit (Thermo Fisher Scientific) in a one-step RT-PCR approach with 2 μL total RNA as input into separate reactions to enrich for B cell heavy- and light-chain transcripts. In the first set of one-step RT-PCR reactions we used a combination of previously published primer sets (Diss et al., 2002; Smith et al., 2009; van Dongen et al., 2003), while in the second set of reactions we used in-house designed heavy- and light-chain primers targeting the beginning or end of FR1 or FR4 of B cell transcripts, respectively. All primer sequences used for the one-step RT-PCR approach are listed below. The thermal cycling parameters for both sets of reactions were as follows: 50°C for 30 min; 94°C for 2 min; 24 cycles of 94°C for 15 s, 58°C for 30 s, and 68°C for 1 min; 68°C for 10 min. PCR products were purified using a 2% gel cassette on a PippinHT system (Safe Science) targeting 200-500 bp amplicons (Sage Science). Illumina indexing and adapter ligation was performed using the NEBNext® Ultra DNA Library Prep kit (NEB).

In the second library preparation approach, 4 μL total RNA was shipped to and processed by AbHelix, LLC (www.abhelix.com, South Plainfield, NJ, USA). Briefly, RNA samples were reverse-transcribed using oligo d(T) 18 and SuperScript IV Reverse

Transcriptase (Thermo Fisher Scientific) followed by Ampure XP bead purification (Beckman Coulter). The purified RT products were divided evenly for the first round of PCR amplification specific to human IgG, IgK, IgL, IgM, or IgA. The 5′ multiplex PCR primers are designed within the leader sequences of each productive V-gene and the 3′ primers within the constant regions but in close approximation to the J-C junctions. The resulting first-round PCR products were purified with magnetic beads and subjected to the second round of PCR amplification to add Illumina index and adapter sequences. The resulting PCR products were purified with Ampure XP (Beckman Coulter) magnetic beads and pooled. Phusion High-Fidelity DNA Polymerase (Thermo Fisher Scientific, CA) was used in all PCR amplification reactions and care was taken to minimize the number of cycles to achieve adequate amplification. Primer sequences used by Abhelix are proprietary and are not provided here.

The DNAs in the final resulting libraries from both library preparation approaches were quantified using the Qubit 3.0 fluorometer (Thermo Fisher Scientific) prior to size determination using a Bioanalyzer 2100 (Agilent). Libraries were re-quantified using the KAPA qPCR kit (KAPA Biosystems) before sequencing on an MiSeq instrument (Illumina) using two separate 2 x 300 bp flow cells (Illumina).

Source	**Chain specificity and orientation of the primer**	**Oligo name**	**DNA sequence, 5′ to 3′**	**Length (base pair)**
In-house	Heavy reverse	IgExp_Hconst	GGAGACGGTGACCAGGGT	18
	Heavy forward	IgExp_H1	CARRTNCAGCTGGTRCAGTC	20
		IgExp_H2	CAGRTCACCTTGARGGAGTC	20
		IgExp_H3	SARGTGCAGCTGGTGGAGTC	20
		IgExp_H4	CAGSTGCAGCTRSAGGAGTC	20

		IgExp_H5	GARGTGCAGCTGGTGCAGTC	20
		IgExp_H6	CAGGTACAGCTGCAGCAGTC	20
		IgExp_H7	CAGGTGCAGCTGGTGCAGTC	20
	Kappa reverse	IgExp_Kconst	AGATGGTGCGGCCGCAGT	18
	Kappa forward	IgExp_K1	GMCATCCRGWTGACCCAG	18
		IgExp_K2	GAKRTTGTGATGACYCAG	18
		IgExp_K3	GAAATWGTRWTGACRCAG	18
		IgExp_K4	GACATCGTGATGACCCAG	18
		IgExp_K5	GAAACGACACTCACGCAG	18
		IgExp_K6	GAWRTTGTGMTGACWCAG	18
	Lambda forward	IgExp_Lconst	TGGAGCGGCCTTAGGCTG	18
		IgExp_L1	CAGTCTGTSBTGACKCAG	18
		IgExp_L2	CARTCTGCCCTGACTCAG	18
		IgExp_L3	TCCTMTGDGCYRAYWCAG	18
		IgExp_L4	CWGCYTGTGCTGACTCAA	18
		IgExp_L5	CAGSCTGTGCTGACTCAG	18
		IgExp_L6	AATTTTATGCTGACTCAG	18
		IgExp_L7	CAGRCTGTGGTGACTCAG	18
		IgExp_L8	CAGWCTGTGGTGACCCAG	18
		IgExp_L9	CAGCCTGTGCTGACTCAG	18
		IgExp_L10	CAGGCAGGGCTGACTCAG	18
		IgExp_L11	CGGCCCGTGCTGACTCAG	18
Van Dongen, *et. al.* .	Heavy reverse	JH_Human	CTTACCTGAGGAGACGGTGACC	22
	Heavy forward	VH6-FR1	TCGCAGACCCTCTCACTCACCTGTG	25
		VH5-FR1	CGGGGAGTCTCTGAAGATCTCCTGT	25
		VH4-FR1	CTTCGGAGACCCTGTCCCTCACCTG	25
		VH3-FR1	CTGGGGGGTCCCTGAGACTCTCCTG	25
		VH2-FR1	GTCTGGTCCTACGCTGGTGAACCC	24
		VH1-FR1	GGCCTCAGTGAAGGTCTCCTGCAAG	25
Diss, *et. al.*	Kappa reverse	JK1	TTTGATATCCACCTTGGTCCC	21
		JK2	TTTAATCTCCAGTCGTGTCCC	21
	Lambda reverse	JL1	AGGACGGTGACCTTGGTCCC	20

		JL2	AGGACGGTCAGCTGGGTCCC	20
Smith, *et. al.*	Lambda forward	VL1	GGTCCTGGGCCCAGTCTGTGCTG	23
		VL2	GGTCCTGGGCCCAGTCTGCCCTG	23
		VL3	GCTCTGTGACCTCCTATGAGCTG	23
		VL4+5	GGTCTCTCTCSCAGCYTGTGCTG	23
		VL6	GTTCTTGGGCCAATTTTATGCTG	23
		VL7	GGTCCAATTCYCAGGCTGTGGTG	23
		VL8	GAGTGGATTCTCAGACTGTGGTG	23
	Kappa forward	VK1-2	ATGAGGSTCCCYGCTCAGCTGCTGG	25
		VK3	CTCTTCCTCCTGCTACTCTGGCTCCCAG	28
		VK4	ATTTCTCTGTTGCTCTGGATCTCTG	25

Table IV-1: Primers used for sequencing

Paired chain sequence clustering. To identify clonal families, paired sequences obtained from our antigen-specific sort was obtained. Sequences were then clustered based on genetic similarity. Sequences were first binned together if they shared the same heavy chain V and J gene as well as CDRH3 length. After, sequences were clustered according to 80% sequence similarity on the CDRH3 nucleotide sequence. Then, they were binned together if they shared the same light chain V and J gene as well as CDRL3 length. Lastly, sequences were clustered again according to 80% sequence similarity on the CDRL3 sequence. These resulting clusters of sequences were designated as clonal families.

To identify public clonotypes, publicly available paired sequence sets of antibody genes were obtained (Bornholdt et al., 2016; Davis et al., 2019; Ehrhardt et al., 2019; Rijal et al., 2019). Together with sequences derived from this paper, public clonotypes were determined by genetic

similarities of antibody sequences using the following clustering scheme. They were first binned by V_H and J_H gene and CDRH3 amino acid length. Sequences within each bin then were clustered according to 60% sequence similarity on their CDRH3 nucleotide sequence. Lastly, sequences were binned if they used the same light chain V and J gene. Clusters of sequences meeting the described criteria and contained sequences originating from two or more individuals were deemed public clonotypes.

Bulk sequence clustering. Sequences within a same paired sequence cluster was taken. These sequences were then used to search for sequences within the bulk sequence dataset. Sequences sharing the same V and J gene as well as 80% similarity on the CDR3 sequence were then clustered together.

Heat map generation. The heavy chain variable gene and light chain variable gene used for each public clonotype were tallied. The number of public clonotypes with corresponding V_H-V_L genes were counted. These frequency counts were then plotted onto the heatmap using Python Seaborn Library.

Network generation. Antibody sequences within the same clonal family was taken to compute a centroid sequence using Vsearch v2.7.1 to be used as a representative for that clonal family. The hamming distance of each antibody sequence within the clonal family to its respective centroid was then calculated. The distance between centroids belonging to different clonal families were then calculated using Levenshtein distance. Distances were calculated using the Python distance library (https://pypi.org/project/Distance/) for hamming distance. Levenshtein distance was calculated as described in literature(Miho et al., 2019). The graph was created with NetworkX and visualized using Matplotlib and PyGraphviz.

Species richness calculations. Clonal families identified as described above, were utilized as a taxonomic unit/species. Rarefaction curves were calculated based on clonal families and unique members as species and individuals respectively for 10,000 repetitions with RTK

(Saary et al., 2017). The mean values of these repetitions were plotted for species richness and the Chao1 estimate of abundance. Fluctuations and rise in Chao1 estimate for the non-antigen-specific data set are interpreted to mean that sequencing depth was inadequate to capture an accurate estimate.

Construction of maximum likelihood trees. Sequences belonging to each cluster/clonal family were aligned to their corresponding germline gene using Clustal Omega v1.2.0. We used the PHYLIP phylogenetic software package v3.697 to generate maximum-likelihood trees from the aligned sequences using the DNAML program, using the sequence of the germline IGHV or IGKV/IGLV as an out group. The resulting phylogenetic trees were visualized using Geneious Prime v2019.2.1. Branches were colored corresponding to the sequence set in which they were identified. The inferred unmutated common ancestor (UCA) was extracted from the PHYLIP-generated tree.

Antibody production and purification. Sequences of mAbs were synthesized using a rapid high-throughput cDNA synthesis platform (Twist Bioscience) and subsequently cloned into an IgG1 monocistronic expression vector (designated as pTwist-mCis_G1) for mAb secretion from mammalian cell culture. This vector contains an enhanced 2A sequence and GSG linker that allows simultaneous expression of mAb heavy- and light-chain genes from a single construct upon transfection (Chng et al., 2015). We performed transfections of ExpiCHO cell cultures using the Gibco ExpiCHO Expression System and protocol for 50 mL mini bioreactor tubes (Corning) as described by the vendor. Culture supernatants were purified using HiTrap MabSelect SuRe (Cytiva) on a 24-column parallel protein chromatography system (Protein Biosolutions). Purified mAbs were buffer-exchanged into PBS, concentrated using Amicon Ultra-4 50-kDa centrifugal filter units (Millipore Sigma) and stored at 4°C until use.

ELISA binding assays. Wells of 384-well microtiter plates were coated with purified recombinant GP at 4°C overnight. Plates were blocked with 2% non-fat dry milk and 2% normal

goat serum in DPBS containing 0.05% Tween-20 for 1 h. All antibodies were diluted to a concentration of either 0.4 μg/mL for the matured antibodies or 5 μg/mL for the germline-revertant antibodies. Antibodies were diluted in two-fold dilutions until binding was no longer detected. Bound antibodies were detected using goat anti-human IgG conjugated with horseradish peroxidase and TMB substrate. The reaction was quenched with 1N hydrochloric acid once color was developed. The absorbance was measured at 450 nm using a spectrophotometer (Biotek).

Real-time cell analysis (RTCA) neutralization assay. To determine neutralizing activity of purified antibodies or human serum, we used real-time cell analysis (RTCA) assay on an xCELLigence RTCA MP Analyzer (ACEA Biosciences Inc.) that measures virus-induced cytopathic effect (CPE)(Suryadevara et al., 2021; Zost et al., 2020b). Briefly, 50 µL of cell culture medium (DMEM supplemented with 2% FBS) was added to each well of a 96-well E-plate to obtain background reading. A suspension of 15,000 Vero cells in 50 µL of cell culture medium was seeded in each well, and the plate was placed on the analyzer. Measurements were taken automatically every 15 min, and the sensograms were visualized using RTCA software version 2.1.0 (ACEA Biosciences Inc). VSV-EBOV, VSV-BDBV, and VSV-SUDV were mixed 1:1 with a respective dilution of mAb using DMEM supplemented with 2% FBS as a diluent and incubated for 1 h at 37°C in 5% CO_2. At 16 h after seeding the cells, the virus-mAb mixtures were added in replicates to the cells in 96-well E-plates. Triplicate wells containing virus only (maximal CPE in the absence of mAb) and wells containing only Vero cells in medium (no-CPE wells) were included as controls. Plates were measured continuously (every 15 min) for 48 h to assess virus neutralization. Normalized cellular index (CI) values at the endpoint (48 h after incubation with the virus) were determined using the RTCA software version 2.1.0 (ACEA Biosciences Inc.). Results are expressed as percent neutralization (CI of wells divided by CI of cells only wells) in a presence of respective mAb relative to control wells with no CPE minus CI values from control wells with

maximum CPE. RTCA IC_{50} values were determined by nonlinear regression analysis using GraphPad Prism 9 software.

Competition-binding ELISA. Wells of 384-well microtiter plates were coated with purified recombinant EBOV GP at 4°C overnight. Plates were blocked with 2% bovine serum albumin (BSA) in DPBS containing 0.05% Tween-20 for 1 h. Each antibody was diluted to a concentration of 10 μg/mL. Next, biotinylated antibodies were diluted to 2.5 μg/mL and added to the primary antibody solution without washing the plate to a final concentration of 0.5 μg/mL. Biotinylated antibody binding was detected with horseradish peroxidase-conjugated avidin (Sigma) and developed with TMB. The reaction was quenched with 1N hydrochloric acid once color was developed. Absorbance was measured at 450 nm using a spectrophotometer.

Cell-surface binding to cleaved or intact GP. Alexa Fluor 647 NHS ester (Thermo Fisher Scientific) was used for antibody labeling. Binding of purified polyclonal or monoclonal antibodies to Jurkat-EBOV GP or Jurkat-EBOV GPCL cells was assessed by flow cytometry using an iQue Screener Plus high-throughput flow cytometer (Intellicyt Corp.) as described previously (Gilchuk et al., 2018; Gilchuk et al., 2020b). Briefly, 50,000 cells were added per each well of V-bottom 96-well plate (Corning) in 5 mL of the DPBS containing 2% heat-inactivated ultra-low IgG FBS (Gibco) (designated as incubation buffer). Serial dilutions of antibody were added to the cells in replicates for a total volume of 50 μL per well, followed by 1 h incubation at ambient temperature, or 4°C in some experiments. Unbound antibody was removed by washing with 200 μL of the incubation buffer. Staining of cells was measured by flow cytometric analysis using the IntelliCyt iQue Screener Plus. Data for up to 20,000 events were acquired, and data were analyzed with ForeCyt (Intellicyt Corp.) software. Dead cells were excluded from the analysis based on forward and side scatter gates to identify the viable cell population. Binding to un-transduced Jurkat cells or binding of dengue antigen-specific mAb DENV 2D22 served as negative controls for most experiments.

Cells that displayed cleaved GP were prepared as described previously (Davis et al., 2019; Gilchuk et al., 2018; Gilchuk et al., 2020b). Briefly, Jurkat-EBOV GP cells were washed with DPBS containing calcium and magnesium (DPBS++), resuspended at 10^6 cells/mL in DPBS containing 0.5 mg/mL of thermolysin (Promega), and incubated for 20 min at 37°C. The cleavage reaction was inhibited by washing cells with the incubation buffer containing DPBS, 2% of heat-inactivated FBS and 2 mM EDTA (pH 8.0). The GP cleavage was confirmed by loss of mAb 13C6 binding and high-level of binding that assessed with RBD-specific mAb MR78 relative to intact Jurkat-EBOV GP antibody binding. Antibody binding to un-transduced Jurkat (mock) cells served as a control for specificity of antibody staining. For screening of the micro-scale purified mAbs, cells were incubated with individual mAbs at a single 1:10 dilution, and the bound antibodies were detected using goat anti-human IgG antibody conjugated with PE (Southern Biotech).

Proteogenomic analysis. The immunoproteogenomic platform Alicanto (Bonissone, 2021) was used for identifying antibody sequences and visualizing proteomics results, similar to a previous study(Gilchuk et al., 2021). Briefly, the variable region sequences of antigen-sorted and sequenced B cells were analyzed and annotated by Alicanto. The tandem mass spectra were searched against this custom antibody database. Antibody clones were determined as present if unique peptide coverage exceeded 50% of the CDR3 region and general peptide coverage was 100%, while coverage over the entire variable region sequence was above 90%.

Quantification and statistical analysis. The descriptive statistics mean ± SEM or mean ± SD were determined for continuous variables as noted. Curves for antibody binding and neutralization were fitted after log transformation of antibody concentrations using non-linear regression analysis. Technical and biological replicates are indicated in the figure legends. Statistical analyses were performed using Prism v8.4.3 (GraphPad).

CHAPTER V

UNDERSTANDING THE PUBLIC ANTIBODY RESPONSE TO THE EBOLA GLYCOPROTEIN

This chapter is an adaptation of the following submitted manuscript:

Chen E.C., Gilchuk P., Zost S.J., Ilinykh P.A., Binshtein E., Huang K., Myers L., Bonissone S.R., Day S., Kona C.R., Trivette A., Reidy J.X., Sutton R.E., Gainza C., Diaz S.M., Williams J.K., Selverian C., Davidson E., Saphire E.O., Doranz B.J., Castellana N., Burkreyev A., Carnahan R.H., Crowe J.E., Systematic analysis of human antibody response to ebolavirus glycoprotein reveals a high prevalence of neutralizing public clonotypes. *Under review.*

CHAPTER OVERVIEW

This chapter describes the public antibody response to the EBOV GP. Due to the large set of sequencing achieved described in chapter IV, we were uniquely positioned to mine for public clonotypes. We not only identified a high frequency of public clonotypes but also functionally validated all the ones found, and saw that there are many public clonotypes that protect against lethal challenge *invivo* in mice. This would not have been possible without Dr. Philip Ilinykh and Dr. Kai Huang who conduced the BSL4 studies showing authentic neutralization in neutralization assays as well as the animal studies. Dr. Cinque Soto and Luke Myers were extremely helpful in their guidance to constructing circus plots and experimenting with clustering thresholds. Edgar Davidson and Dr. Benjamin Doranz assisted in the alanine scanning mutagenesis experiments. The electron microscopy studies were performed by Dr. Elad Binshein.

INTRODUCTION

As mining the human repertoire for public clonotypes requires very large numbers of antibody gene sequences, the prevalence of ebolavirus-reactive public antibody clonotypes has yet to be described. Three classes of antibodies were described as encoded by the variable genes *VH3-15/VL1-40, IGHV3-13,* or *IGHV3-23* in multiple individuals following vaccination with the rVSV-ZEBOV (Cohen-Dvashi et al., 2020; Ehrhardt et al., 2019) or ChAd3-ZEBOV(Rijal et al., 2019) vaccines, or natural infection (Bornholdt et al., 2016; Cagigi et al., 2018; Davis et al., 2019; Wec et al., 2017). However, these studies were not done on a scale large enough to understand the prevalence and functionality of public clonotypes. Therefore, information on common responses is limited. Additionally, for validating public clonotypes by testing recombinant immunoglobulins for specificity and antiviral function requires the antibody gene datasets need to containing authentically-paired heavy and light chain genes from single B cells. With larger paired-chain sequence datasets, the likelihood of identifying public clonotypes increases, allowing a functional understanding of the public antibody response to ebolavirus GPs. Mining for and understanding the properties of public clonotypes informs a deeper understanding of population immunity by revealing immunodominant B cell responses within immune populations, which may benefit the rational design of vaccines exhibiting immunogenicity in a broader segment of the population. Knowing the public clonotype profile following natural infection also can enhance experimental vaccine testing, since the immunogenicity for desirable public antibodies recognizing cross-reactive sites of vulnerability for potent neutralization can be recognized at the cDNA sequence level. We should also keep in mind, however, that the broad induction of public antibody clonotypes recognizing the protective antigen of an RNA virus can lead to a constant and collective pressure on certain epitopes, leading to rapid selection of escape mutant variant viruses.

To address this gap in knowledge, we utilized our large set of EBOV GP specific B cells identified in Chapter V to mine for and identify public clonotypes shared between infected or vaccinated individuals.

RESULTS

73 public clonotypes are identified

We curated a database containing Ebola-specific mAbs by combining the large set of sequences acquired here with several smaller sets of previously reported Ebola-specific antibodies (Davis et al., 2019; Ehrhardt et al., 2019; Rijal et al., 2019; Wec et al., 2017). Collectively, this database includes sequences from 12 individuals determined following either natural infection or vaccination. These sequences then were clustered to identify public clonotypes. Sequences were first binned by their V and J gene use and CDR3 length. Next, sequences were clustered by 60% on the CDR3 nucleotide sequence and binned by the light chain V and J genes. Clusters with sequences from two or more of the 12 individuals were identified as public clonotypes. A total of 73 public clonotypes were identified. One public clonotype was shared among 6 donors. Another was shared among four donors. Five public clonotypes were shared between three donors, and the remaining were all shared between two donors (**Figure V-1A**). All 294 members of the 73 public clonotypes were synthesized and expressed as recombinant IgGs as previously described (Gilchuk et al., 2020a) and tested by ELISA for binding to EBOV, BDBV, SUDV, MARV GP, or EBOV sGP. Next, they were tested for binding to cleaved EBOV GP (GP_{cl}) or intact EBOV GP (GP_{ecto}), and for neutralization of VSV-EBOV GP. As most members of each public clonotype were expected to share similar functional profiles due to genetic similarity, the predicted functional profile was determined by identifying the dominant functional phenotype in each public clonotype (**Figure V-1C**).

From the 73 public clonotypes, there was a diversity of variable heavy and light chain combinations used. However, the two most frequent combinations observed were *IGHV1-18/ IGKV3-20* and *IGHV3-07/IGKV3-15.* All public clonotypes that used *IGHV1-18/IGKV3-20* bound to EBOV, BDBV, and SUDV GP, but none neutralized VSV-EBOV GP. All but one of the public clonotypes that used *IGHV3-07/IGKV3-15* bound to EBOV, BDBV, and SUDV GP; the outlier bound to EBOV and SUDV but not BDBV GP. None of these public clonotypes exhibited neutralization to VSV-EBOV GP. Therefore, it is likely that GP-reactive antibodies reacting to EBOV, BDBV, and SUDV using *IGHV1-18/IGKV3-15* and *IGHV 3-07/IGKV3-15* are found in many individuals. Additionally, there were four public clonotypes that used *IGHV3-13/IGKV3-20*, and three that used *IGHV3-21/IGKV3-15*. The majority of these seven public clonotypes had neutralizing properties. Additionally, the bulk of the public clonotypes using these variable genes had similar functional profiles, hinting that these combinations of variable genes may encode the neutralization properties for VSV-EBOV GP (**Figure V-1B**).

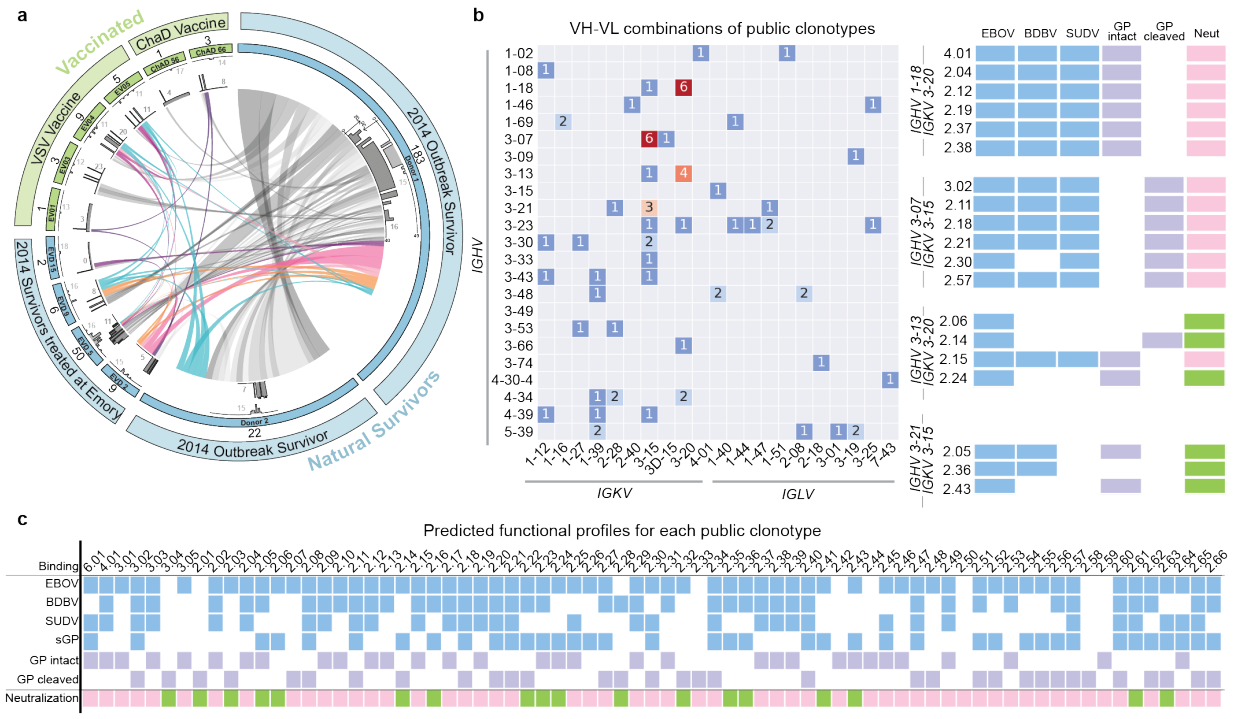


Figure V-1: Identification of public clonotypes.

(A) Clonal overlap between each vaccinated (green) and convalescent (blue) donors. Numbers inside the first outer circle indicate the number of sequences that were identified as public clonotypes from the respective donor. The light grey color shows the distribution and median CDR3 length. The dark grey color shows the distribution and median number of somatic mutations of the public antibodies from that donor. **(B)** Heavy and light chain variable gene usage combinations for all public clonotypes identified. Numbers inside boxes indicate the number of public clonotypes using that gene combination. Public clonotype groups using highly used genes are listed on the right. Blue indicates binding to GPs in ELISA, purple indicates binding via cell-surface GP display assay; pink indicates non-neutralizing, and green indicates neutralizing. **(C)** Functional profiles of each of the 73 public clonotypes after expression and functional testing of the 294 public clonotype antibodies and inferring a functional profile for each of the public clonotype groups deemed as “predicted functional profile”. Blue indicates binding to GPs in ELISA, purple indicates binding in a cell-surface GP display assay; pink indicates non-neutralizing, and green indicates neutralizing. Experiments were performed in biological duplicates. A compilation of the average of all experiments is shown.

15 of 73 public clonotypes neutralize EBOV

Of the 73 public clonotypes, 15 neutralized VSV-EBOV GP. Members of these 15 public clonotypes then were tested for binding to EBOV, BDBV, and SUDV GP. One of the 15 public

clonotypes bound to all three GPs, two of 15 bound to EBOV and BDBV GP, nine of 15 bound to only EBOV GP, and three of 15 did not exhibit binding to any GP, indicating that the majority of the neutralizing public antibody response is primarily monospecific (**Figure V-2A**).

Next, all mAbs were tested for neutralization of VSV-EBOV, -BDBV, or -SUDV GPs. Although most public clonotypes only exhibited neutralization to VSV-EBOV GP, Group 3.04 neutralized both VSV-EBOV and -SUDV GPs. Groups 2.61, 2.22, 2.23, and 2.28 neutralized VSV-EBOV and -BDBV GPs (**Figure V-2A**).

One mAb from each public clonotype group was tested for neutralization of authentic virus. The neutralization profile for each mAb previously established with VSV-EBOV GP was reflected in the authentic virus neutralization assay except for EBOV-854. EBOV-854 exhibited a low neutralization potency in the VSV neutralization experiment and did not show any neutralization in the authentic virus experiment. Together, these findings verified that we identified public clonotypes exhibiting neutralization properties for EBOV, BDBV, and SUDV (**Figure V-2B**).

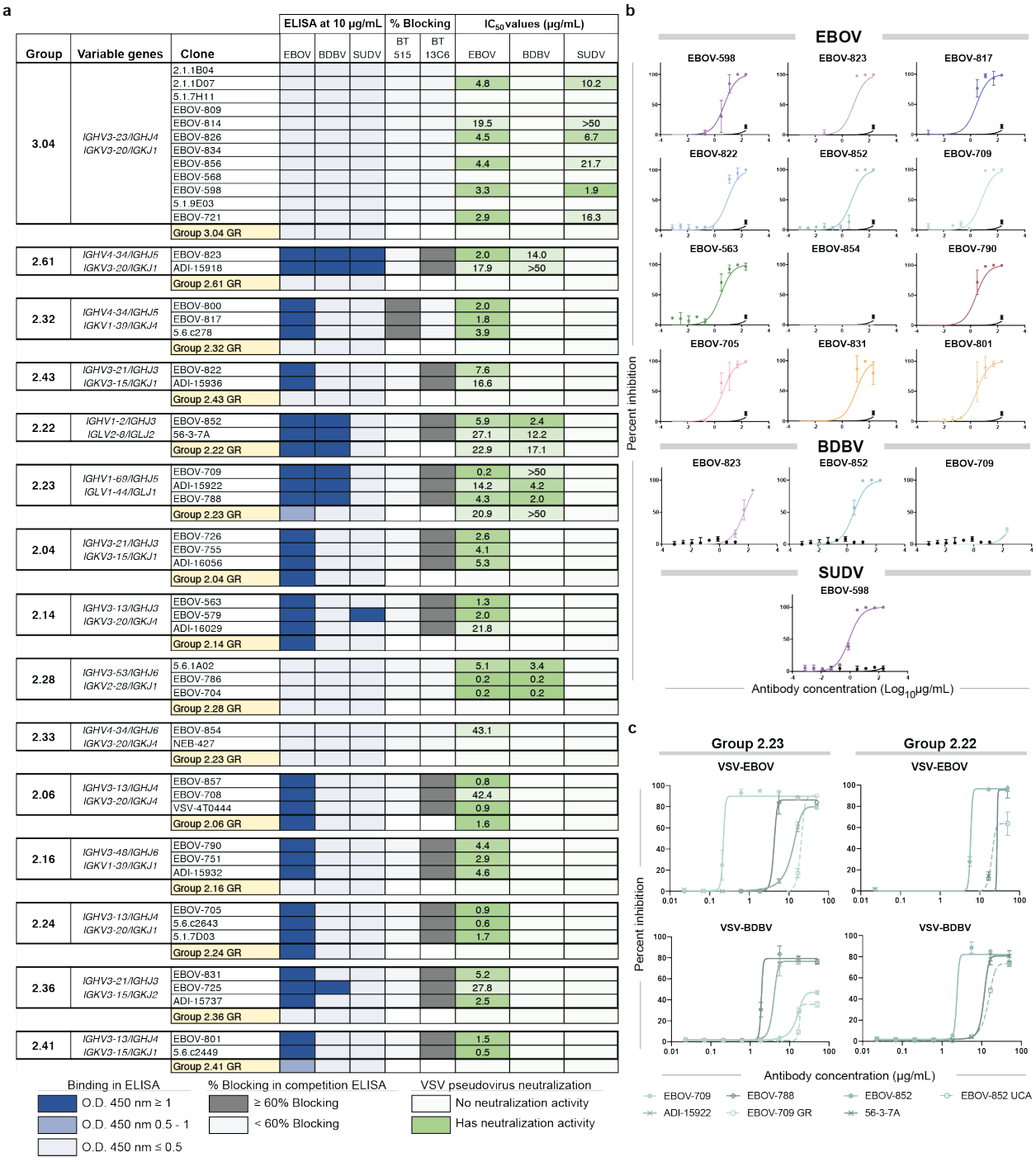


Figure V-2: Properties of neutralizing public clonotypes

(Legend on next page)

Figure V-2 legend: **(A)**Table showing all 15 neutralizing public clonotypes. The first column identifies the public clonotype group number, the second column details the variable gene usage. The third column indicates clone name, with all the public clonotype antibodies in the group indicated in white and the germline revertant version of that group's antibody in yellow. Blue boxes indicate binding in ELISA at an antibody concentration of 10 μg/mL. Grey indicates percent blocking in a competition-binding assay. Green indicates neutralization for VSV-EBOV, -BDBV, or -SUDV with the IC_{50} values written inside the boxes. Clone names highlighted in yellow at the bottom row of each section is the germline revertant version of that public clonotype and its respective functionality. **(B**)Authentic virus neutralization curves for a representative antibody of each public clonotype group. **(C)**Neutralization curves from antibodies that retained cross-reactive neutralization at the germline level. Group 2.23 antibodies to VSV-EBOV and -BDBV are shown on the left and Group 2.22 antibodies to VSV-EBOV and -BDBV are on the right. Dotted lines in each graph indicate the germline revertant antibody curve. Solid lines in varying colors indicate the matured versions of the antibodies in that public clonotype group.

Most of the neutralizing public clonotypes identified target the glycan cap

All mAbs within the public clonotype groups were competed against each other for binding to EBOV GP in a competition-binding ELISA for pairwise comparison (**Figure V-3A**). As this pairwise competition-binding ELISA was done with intact IgG, it is likely that the flexibility of the Fc region of the mAbs resulted in the asymmetric competition-binding grid in several public clonotype groups. Despite asymmetric competition (**Figure V-3A**), all mAbs within each public clonotype group competed against each other. All mAbs then were tested for competition-binding with the previously epitope-mapped mAbs EBOV-515 (a base antibody) or 13C6 (a glycan cap antibody). Of the 15 neutralizing public clonotypes, 11 targeted the glycan cap and 1 targeted the base region of GP (**Figure V-2A**) as concluded from competition-binding ELISA results. The remaining three antibodies did not bind to GP in ELISA and therefore, we used negative stain electron microscopy (EM) to identify the antigenic site recognized by these neutralizing antibodies.

Fab-EBOV GP complexes were imaged with a representative mAb from each public clonotype group (**Figure V-3B, V-4**). Although EBOV-598 (Group 3.04), EBOV-786 (Group 2.28),

and EBOV-854 (Group 2.33) and other members of their respective public clonotype groups did not exhibit binding to GP in ELISA they did show binding to GP_{cl} in a cell-surface display assay and had neutralization properties (**Figure V-1C, Figure V-2A**); these mAbs complexed with EBOV GP for EM studies, and 3D reconstructions were made. These low-resolution reconstructions show that all three of these mAbs as well as EBOV-817, which competed for binding with the reference antibody EBOV-515, bind the base region of GP. Although EBOV-852 was visualized on the grid, and we were able to obtain 2D images of it in complex with GP, we were unable to obtain a 3D reconstruction for it. Low-resolution reconstructions of the rest of the public clonotypes show that the public clonotypes bind diverse regions of the GP ranging from the glycan cap to the base (**Figure V-3B, Figure V-4**).

We then attempted to determine the critical binding residues at the amino acid level for a representative antibody from each of the 15 groups. Antibodies were screened for binding to alanine scanning mutant libraries of the EBOV GP. Screening was successful for EBOV-852, and we were able to identify single binding site residues for EBOV-598, EBOV-786, EBOV-709, and EBOV-823 (**Figure V-3C**). However, as all these antibodies are neutralizing antibodies and therefore bind very avidly to the EBOV GP, single residue alanine mutations failed to disrupt binding for the rest of the antibodies even after digestion and screening of the binding of the antibodies as Fabs.

Critical residues for EBOV-852 were P279, E303, S302, and K299, all residues which span the glycan cap. These results are consistent with the competition-binding ELISA results, in which EBOV-852 competed with the glycan cap antibody 13C6 (**Figure V-2A**). E303 is a conserved residue for not only EBOV, BDBV, and SUDV but also for TAFV and RESTV. S302 and P279 are also conserved between EBOV and BDBV. However, in SUDV and RESTV the serine is replaced with glycine and in SUDV only, the proline is replaced with alanine. These findings likely explain why EBOV-852 binds only to EBOV and BDBV GPs.

Critical residues for EBOV-598 are R89 and G149. The G149 residue is conserved across EBOV, BDBV, SUDV, TAFV, REST, and R89 is conserved across EBOV, BDBV, SUDV, TAFV, REST, and MARV. Although this residue sits in the conserved region of the receptor binding domain (RBD), the residue after it, S90 is only conserved between EBOV, SUDV, and REST. In BDBV, TAFV, and MARV, this residue is substituted to an alanine. Therefore, this finding likely explains the cross-reactive neutralization of EBOV and SUDV but not BDBV by EBOV-598, which was unexpected as EBOV and BDBV GP are generally more similar in sequence identity than EBOV and SUDV GP. A single critical residue was identified for EBOV-786, EBOV-709, and EBOV-823 (**Figure V-3C**). The one critical residue indicated for EBOV-786 was S46. This residue is conserved between EBOV and BDBV but not SUDV GP (in which the serine changes to a threonine). However, this residue is also conserved in TAFV. Lastly, the critical residue identified for EBOV-709 and EBOV-823 is W275. This residue is conserved across EBOV, BDBV, SUDV, TAFV, and REST, and sits in the glycan cap, also mirroring the results of the competition-binding ELISA data as these antibodies compete with mAb 13C6 (**Figure V-2A**) and the results of the negative stain EM studies (**Figure V-3B**). Therefore, it is likely that these findings explain the capacity of EBOV-709 and EBOV-852 to neutralize EBOV and BDBV but not SUDV.

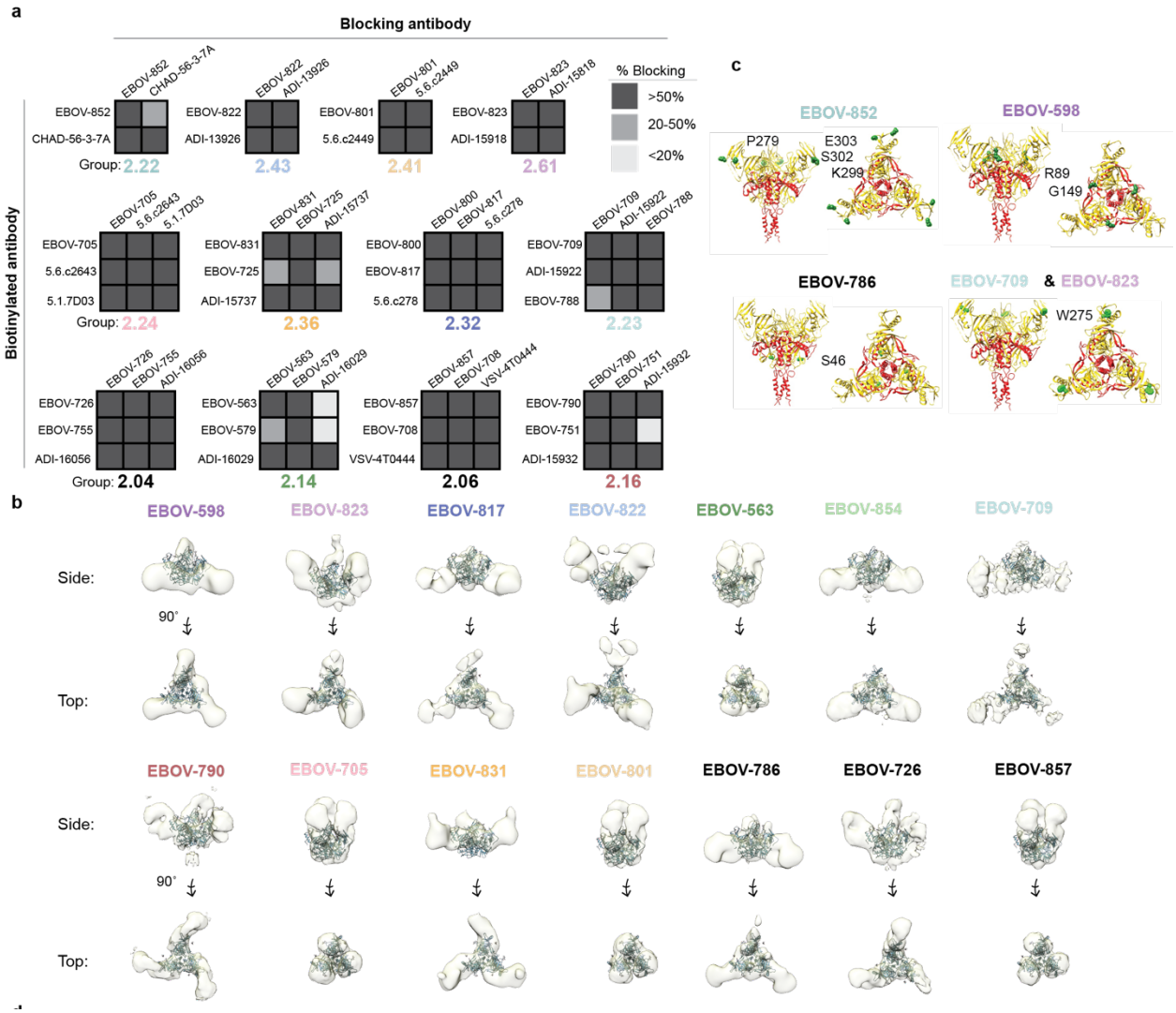


Figure V-3: Epitopes targeted by neutralizing public clonotypes.

(A) Competition-binding ELISA results for antibodies within each public clonotype group in competition with each other. Unlabeled blocking antibodies applied to the GP antigen first are listed across the top of each grid while the biotinylated antibodies that are added to the antigen-coated wells second are indicated on the left. The number in each box represents the percent un-competed binding of the biotinylated antibody in the presence of the indicated competed antibody. The experiment was performed in biological duplicate and technical triplicates with similar results. A biological replicate from a single experiment is shown. **(B)** Negative stain EM of EBOV GP in complex with Fab forms of different antibodies. 3D reconstructions are shown. The Fab (blue) is docked to a trimer of the EBOV GP (grey). **(C)** Critical binding residues for EBOV-598, EBOV-709, and EBOV-823 are indicated in green, as determined by loss of binding in alanine-scanning GP mutagenesis studies.

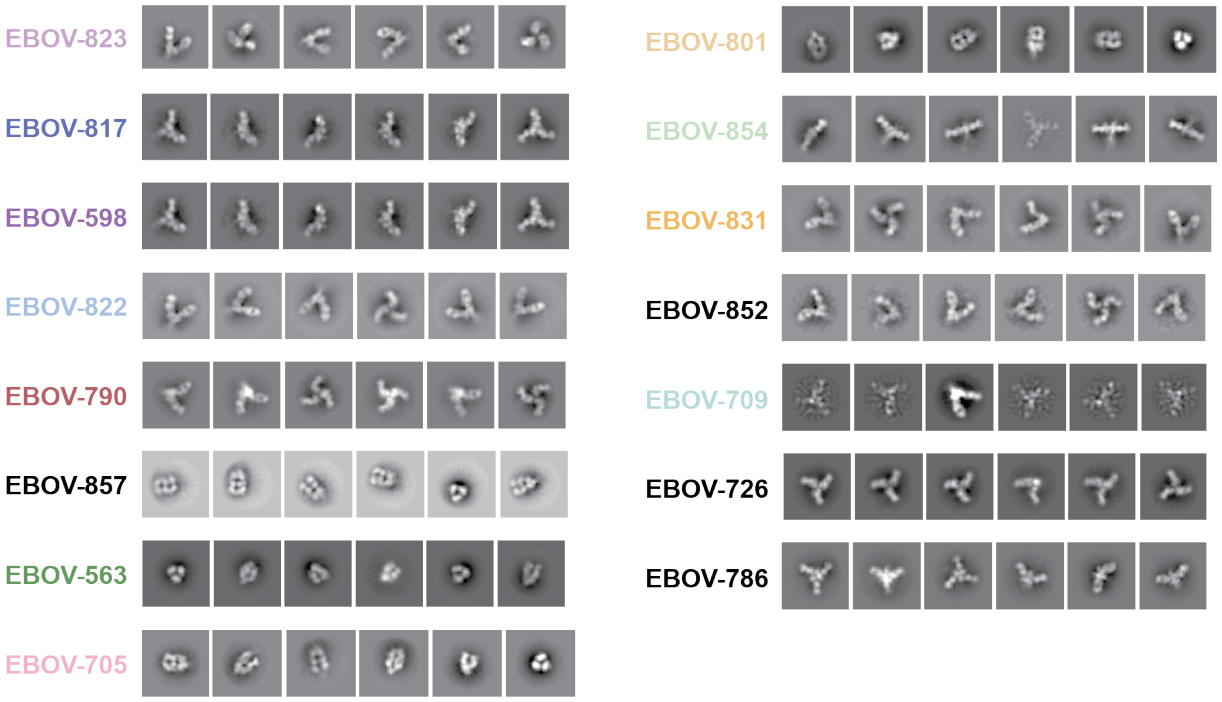


Figure V-4: Negative stain electron microscopy complexes of representative antibodies from each public clonotype.

IgG for a representative antibody for each public clonotype was digested to obtain the Fab form of antibody and complexed with EBOV GP.

Surveying the level of publicness in identified public clonotypes

There are many methods for identifying public clonotypes, using multiple identity thresholds or junction matching techniques. Our approach used a paired heavy and light chain gene sequence set and we characterized the functional phenotypes of all public clonotype antibodies identified, allowing us to use a sequence identity threshold on the lower end of common practice. Using the same clustering scheme of binning on the heavy chain V and J gene as well as CDR3 length, we next clustered the public clonotypes identified at 70% and 80% similarity on the CDR3 nucleotide sequence and binned them at the back end by matching on the light chain V and J gene. Next, we identified the antibodies that fell out of each public clonotype cluster at each threshold of 60%,

70%, and 80% and investigated if antibodies that fell out at each threshold shared similar functional phenotypes that would differentiate them from the main group. We did not detect a difference in the functional binning of antibodies when clustering at differing identity thresholds.

Our criteria for public clonotype identification requiring the same heavy and light chains could be considered conservative, as there are numerous examples of public clonotypes defined by a recurrent heavy chain that undergo promiscuous pairing with various light chains (Setliff et al., 2018; Tan et al., 2021). To determine the flexibility of sequences on the light chain, we tested if public clonotypes within the same group would express and function with light chains belonging to differing donors. Three public clonotypes were selected for which the heavy chain from one donor and the light chain from another donor were recombined to investigate if the reactivity of the public clonotype was preserved. EBOV-1182 uses the heavy chain from EBOV-826 and the light chain from 2.1.1D07 and maintains its ability to neutralize both EBOV and BDBV when the antibody chains were swapped. EBOV-1190 uses the heavy chain from EBOV-786 and the light chain from 5.6.1A02 and neutralized both EBOV and BDBV. Lastly EBOV-1187 which uses the heavy chain from EBOV-852 and the light chain from 56-3-7A, also neutralizes both EBOV and BDBV (**Figure V-5**). Together, we are confident that our approach of using a threshold of 60% on the CDRH3 sequence in conjunction with binning on the CDRH3 length and both heavy and light chain V and J genes is successful in identifying public clonotypes when using paired sequence sets.

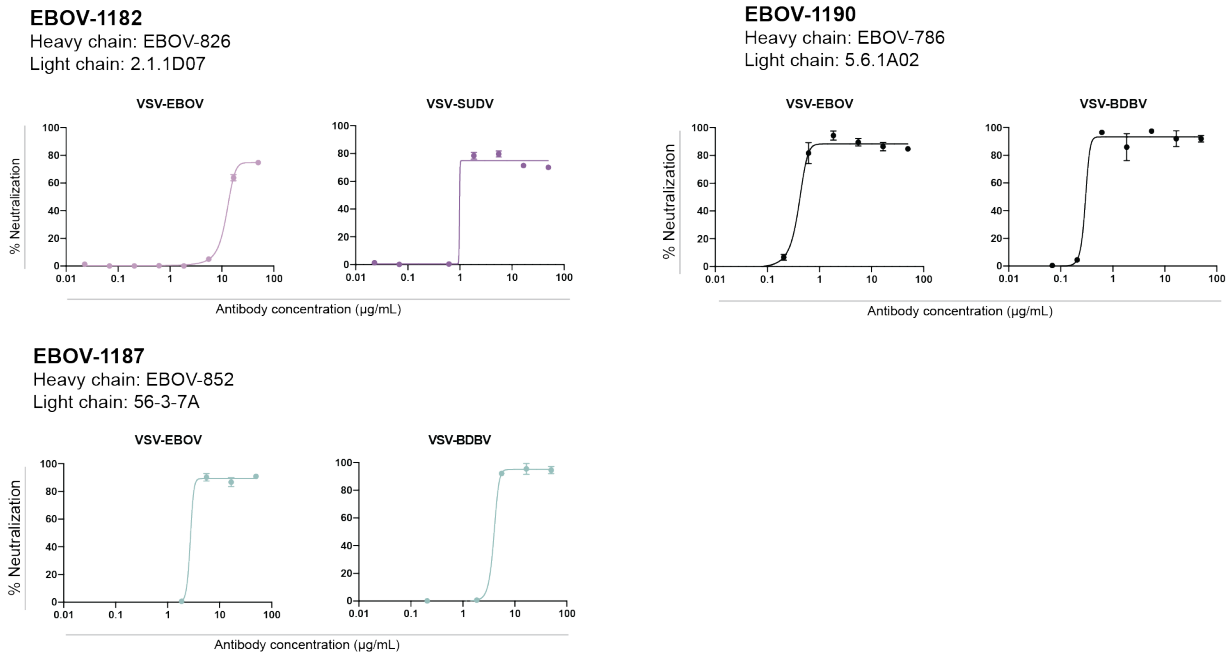


Figure V-5: Functionality of public clonotypes after swapping heavy and light chains. Heavy chains from a public clonotype antibody from one donor and light chain from that public clonotype observed in another donor were paired and expressed together and tested for neutralization of VSV-EBOV, -BDBV, or -SUDV.

Germline-encoded properties are retained in some public clonotypes

To investigate if the neutralizing activity of these public clonotypes was due to germline-encoded reactivity or the result of somatic mutations, we investigated the equivalent germline-encoded antibodies for each public clonotype. We aligned each heavy and light chain variable region sequences to its respective germline gene sequence and reverted residue that differed from the germline gene to the inferred germline residue. Each germline-revertant (GR) antibody then was tested to see if the GR version of the antibody shared similar properties to its mutated counterparts (**Figure V-2**). All GR antibodies were tested for binding to EBOV, BDBV, or SUDV GP. Additionally, they were tested for neutralization of VSV-EBOV, -BDBV, or -SUDV. Although most

GR antibodies did not retain binding to either GP or neutralize either virus, three GR antibodies retained functional activity compared to their mutated counterparts (**Figure V-2A**). EBOV-852-GR, encoded by *IGHV1-2/IGLV2-8* retained ability to bind and neutralize EBOV and BDBV. EBOV-709-GR, encoded by *IGHV1-69/IGLV1-44* retained ability to bind and neutralize EBOV but only partially to BDBV (**Figure V-2C**). EBOV-857, encoded by *IGHV3-13/IGKV3-20* retained its ability to bind and neutralize EBOV (**Figure V-2A**). These findings indicate that germline genes in these public clonotypes encode antibodies with critical residues that not only mediate binding but also neutralization.

EBOV public clonotypes protect *in vivo*

We then tested these public clonotypes and their level of protection *in vivo* in mice against EBOV (Mayinga strain). Antibodies were delivered at 5 mg/kg 1 day after inoculation with EBOV. Scores on protection from death, weight loss, and disease were measured for 28 days. Treatment with mAbs representing public clonotypes conferred protection against mortality. 100% of animals survived the infection after treatment with EBOV-598, EBOV-790, EBOV-852, EBOV-705, EBOV-709, EBOV-801, EBOV-817, or EBOV-831. 80% of animals survived after treatment with EBOV-823 or EBOV-563, and 40% – after treatment with EBOV-822 (**Figure V-6, Figure V-7**). Although EBOV-854 showed low levels of neutralization *in vitro* using VSV-EBOV, it did not show neutralization with authentic virus, and accordingly failed to protect animals *in vivo*. Overall, these findings show that there are public clonotypes specific to EBOV that protect *in vivo*.

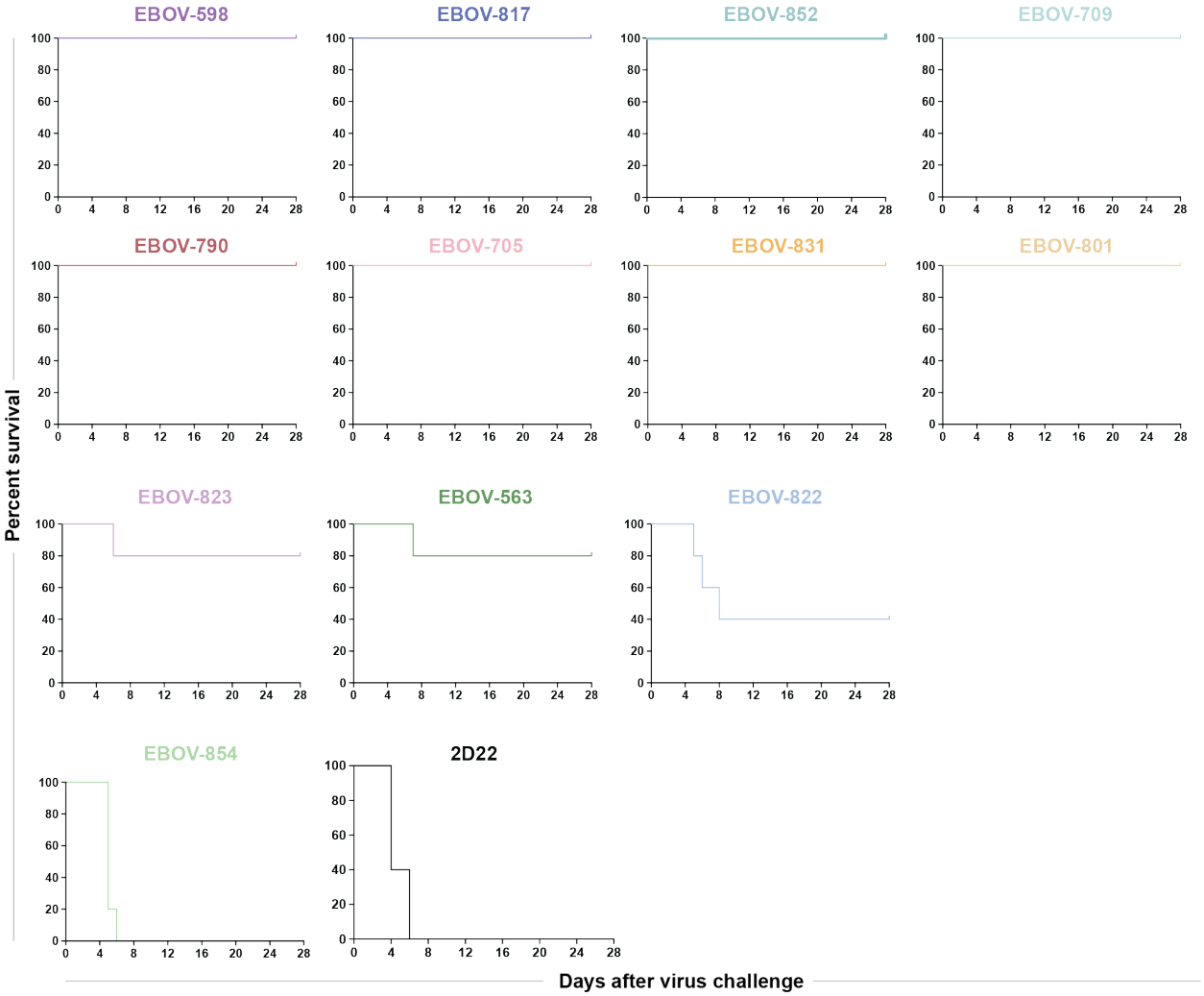


Figure V-6: In vivo protection using public clonotypes

Mice (n = 5) were treated i.p. with 100 μg (~5 mg/kg) of an individual antibody per mouse on day 1 post-challenge. Human antibody DENV 2D22 (specific to dengue virus) served as a negative control. Mice were monitored twice daily from day 0 to day 14 post challenge for survival and monitored daily from day 15 to 28 as described previously (Ilinykh et al., 2018).

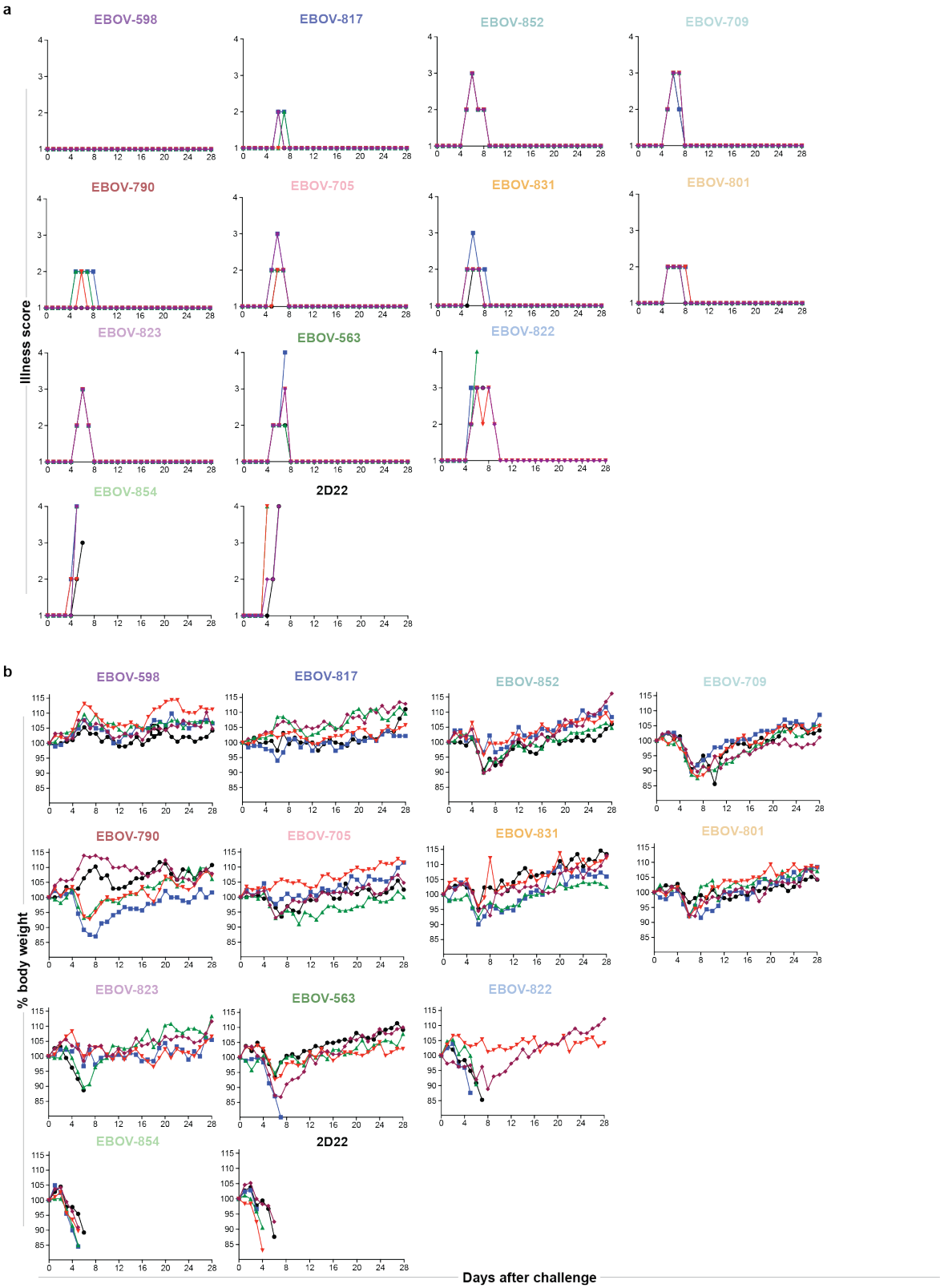


Figure V-7: In vivo efficacy of public clonotypes

(Legend on next page)

Figure V-7 legemd: Invivo efficacy of public clonotypes figure legend: Mice (n = 5) were treated i.p. with 100 μg (~5 mg/kg) of individual antibody per mouse on day 1 post-challenge. MAb DENV 2D22 was used as a negative control. Mice were monitored twice daily from day 0 to day 14 post-challenge for illness, survival, and weight loss, followed by once daily monitoring from day 15 to the end of the study at day 28. **(A)** Illness scores of mice treated with each public clonotype. **(B)** Body weight graphs of mice treated with each public clonotype.

DISCUSSION

Public clonotypes have been identified in human antibody repertoires in response to a variety of viral pathogens including influenza virus (Joyce et al., 2016; Pappas et al., 2014; Zost et al., 2021b), respiratory syncytial virus (Mukhamedova et al., 2021), hepatitis C virus (Bailey et al., 2017a), HIV (Setliff et al., 2018; Zhou et al., 2015), and SARS-CoV-2 (Chen et al., 2021a; Sakharkar et al., 2021; Schmitz et al., 2021; Tan et al., 2021) revealing selection of genetically similar B cell receptors in memory cells in circulation of diverse immune individuals. Understanding the prevalence of public clonotypes and their functionalities requires very large numbers of paired sequences.

Several public clonotypes have previously been reported that recognize the EBOV GP (Cagigi et al., 2018; Cohen-Dvashi et al., 2020; Davis et al., 2019; Rijal et al., 2019; Wec et al., 2017). But the large scale of sequencing obtained in this study uniquely positioned us to systematically identify a high prevalence of public clonotypes elicited to the EBOV GP, with 73 public clonotypes, a level of sharing that is unexpected since we required the public clonotypes to share not only heavy chain features but also the same light chain gene usage. This data collection is the largest set of B cell public clonotypes reported to date for a viral pathogen, and most of these are novel public clonotypes to EBOV that have not yet been described. By functionally characterizing every

antibody identified in the 73 public clonotypes, we found that roughly 20% of EBOV GP specific public clonotypes neutralized the virus. Most of the neutralizing public clonotypes also conferred therapeutic protection *in vivo* against lethal challenge. The studies using negative stain electron microscopy revealed that these 15 neutralizing public clonotypes target diverse regions of the GP ranging from the glycan cap to the base. Additionally, analysis into the germline-encoded functions of public clonotypes revealed that three of the 15 neutralizing public clonotypes retained neutralization when somatic mutations were reverted to the inferred germline gene segment sequence. Public clonotypes that neutralized virus as UCA antibodies were encoded by *IGHV1-69*, *IGHV1-02*, or *IGHV3-13*.

Mining for and understanding the properties of public clonotypes informs a deeper understanding of population immunity by revealing immunodominant B cell responses within immune populations, which may be of benefit for rational design of vaccines that may exhibit immunogenicity in a broader segment of the population. Knowing the public clonotype profile following natural infection also can enhance experimental vaccine testing, since the immunogenicity for desirable public antibodies recognizing cross-reactive sites of vulnerability for potent neutralization can be recognized at the cDNA sequence level. We should also keep in mind, however, that the broad induction of public antibody clonotypes recognizing the protective antigen of an RNA virus can lead to a constant and collective pressure on certain epitopes to viruses leading to rapid selection of escape mutant variants.

The large size of the data set of EBOV GP-reactive memory B cells created in this study provided an opportunity to mine for public clonotypes specific to EBOV GP but also posed technical challenges for data analysis. We sought to identify public clonotypes using both heavy and light chain sequences, a workflow that is only achievable with paired sequences from single B cells.

Identity thresholds used to identify public clonotypes in variable gene sequence sets vary greatly in the field. We chose an identity threshold for identification of public clonotypes on the lower end of common conventions, but our confidence in these assignments was supported by both heavy and light chain gene segment assignments and with functional testing of recombinant antibodies encoded by these sequences. As we tested all 294 members of the public clonotypes functionally, we were in a unique position to investigate different clustering thresholds for identifying public clonotypes and tested how those thresholds affected the grouping of antibodies with functional phenotypes. As clustering at higher thresholds did not necessarily bin antibodies into tidier functional phenotype bins, we conclude that when mining for public clonotypes within an antigen-specific sequence set using paired sequencing from single cells, a threshold of 60% identity in the CDRH3 is sufficient.

METHODS

Mouse challenge with EBOV. Mice were housed in microisolator cages and provided food and water *ad libitum*. Groups of 7-8-week-old BALB/c mice (Charles River Laboratories) were inoculated with 1,000 plaque-forming units of EBOV-MA by the intraperitoneal (i.p.) route. Mice (n = 5) were treated i.p. with 100 μg (~5 mg/kg) of individual mAb per mouse on day 1 post-challenge. Human mAb DENV 2D22 (specific to dengue virus) served as negative control. Mice were monitored twice daily from day 0 to day 14 post-challenge for illness, survival, and weight loss, followed by once daily monitoring from day 15 to the end of the study at day 28, as described elsewhere (Ilinykh et al., 2018). Moribund mice were euthanized as per the IACUC-approved protocol. All mice were euthanized on day 28 after EBOV challenge.

Neutralization assay. Neutralization was tested against GFP-expressing EBOV and chimeric EBOV/BDBV-GP and EBOV/SUDV-GP constructs in a high-throughput format, as

previously described (Ilinykh et al., 2016). The neutralization assays were performed using Vero-E6 cells. Neutralization assays were performed in triplicate, across 12 four-fold dilutions, starting from 200 µg/mL.

Electron microscopy sample and grid preparation, imaging and processing of EBOV GP–Fab complexes. For electron microscopy imaging of EBOV GP and Fabs, Fabs were produced by digesting recombinant chromatography-purified IgGs using resin-immobilized cysteine protease enzyme (FabALACTICA, Genovis). The digestion occurred in 100 mM sodium phosphate and 150 mM NaCl pH 7.2 (PBS) for around 16 h at ambient temperature. To remove cleaved Fc from intact IgG, the digestion mix was incubated with CaptureSelect Fc resin (Genovis) for 30 min at ambient temperature in PBS buffer. For screening and imaging of negatively-stained EBOV protein in complex with human Fabs, the proteins were incubated at a Fab:EBOV GP (trimer) molar ratio of 4:1 for about 1 hour at ambient temperature, and approximately 3 µL of the sample at concentrations of about 10 to 15 µg/mL was applied to a glow-discharged grid with continuous carbon film on 400 square mesh copper electron microscopy grids (Electron Microscopy Sciences). The grids were stained with 0.75% uranyl formate (Ohi et al., 2004) . Images were recorded on a Gatan US4000 4k×4k CCD camera using an FEI TF20 (TFS) transmission electron microscope operated at 200 keV and control with SerialEM (Mastronarde, 2005). All images were taken at 50,000× magnification with a pixel size of 2.18 Å per pixel in low-dose mode at a defocus of 1.5 to 1.8 μm. The total dose for the micrographs was around 30 e–/per Å^2. Image processing was performed using the cryoSPARC software package (Punjani et al., 2017). Images were imported, CTF-estimated with CTFFIND4 (Rohou and Grigorieff, 2015) and particles were picked automatically with template picker (a part of cryoSPARC). The particles were extracted with a box size of 160 pixels and binned to 80 pixels (pixel size of 4.36 Å/pix). Multiple 2D class averages were performed, and good classes

were selected for *ab-initio* 3D map reconstruction. At the final step, the data sets were refined. Maps were imaged using Chimera software (Pettersen et al., 2004).

Antibody production and purification. Sequences of mAbs were synthesized using a rapid high-throughput cDNA synthesis platform (Twist Bioscience) and subsequently cloned into an IgG1 monocistronic expression vector (designated as pTwist-mCis_G1) for mAb secretion from mammalian cell culture. This vector contains an enhanced 2A sequence and GSG linker that allows simultaneous expression of mAb heavy- and light-chain genes from a single construct upon transfection (Chng et al., 2015). We performed transfections of ExpiCHO cell cultures using the Gibco ExpiCHO Expression System and protocol for 50 mL mini bioreactor tubes (Corning) as described by the vendor. Culture supernatants were purified using HiTrap MabSelect SuRe (Cytiva) on a 24-column parallel protein chromatography system (Protein Biosolutions). Purified mAbs were buffer-exchanged into PBS, concentrated using Amicon Ultra-4 50-kDa centrifugal filter units (Millipore Sigma) and stored at 4°C until use.

ELISA binding assays. Wells of 384-well microtiter plates were coated with purified recombinant GP at 4°C overnight. Plates were blocked with 2% non-fat dry milk and 2% normal goat serum in DPBS containing 0.05% Tween-20 for 1 h. All antibodies were diluted to a concentration of either 0.4 μg/mL for the matured antibodies or 5 μg/mL for the germline-revertant antibodies. Antibodies were diluted in two-fold dilutions until binding was no longer detected. Bound antibodies were detected using goat anti-human IgG conjugated with horseradish peroxidase and TMB substrate. The reaction was quenched with 1N hydrochloric acid once color was developed. The absorbance was measured at 450 nm using a spectrophotometer (Biotek).

Real-time cell analysis (RTCA) neutralization assay. To determine neutralizing activity of purified antibodies or human serum, we used real-time cell analysis (RTCA) assay on an xCELLigence RTCA MP Analyzer (ACEA Biosciences Inc.) that measures virus-induced cytopathic effect (CPE)(Suryadevara et al., 2021; Zost et al., 2020b). Briefly, 50 µL of cell culture

medium (DMEM supplemented with 2% FBS) was added to each well of a 96-well E-plate to obtain background reading. A suspension of 15,000 Vero cells in 50 µL of cell culture medium was seeded in each well, and the plate was placed on the analyzer. Measurements were taken automatically every 15 min, and the sensograms were visualized using RTCA software version 2.1.0 (ACEA Biosciences Inc). VSV-EBOV, VSV-BDBV, and VSV-SUDV were mixed 1:1 with a respective dilution of mAb using DMEM supplemented with 2% FBS as a diluent and incubated for 1 h at 37°C in 5% CO_2. At 16 h after seeding the cells, the virus-mAb mixtures were added in replicates to the cells in 96-well E-plates. Triplicate wells containing virus only (maximal CPE in the absence of mAb) and wells containing only Vero cells in medium (no-CPE wells) were included as controls. Plates were measured continuously (every 15 min) for 48 h to assess virus neutralization. Normalized cellular index (CI) values at the endpoint (48 h after incubation with the virus) were determined using the RTCA software version 2.1.0 (ACEA Biosciences Inc.). Results are expressed as percent neutralization (CI of wells divided by CI of cells only wells) in a presence of respective mAb relative to control wells with no CPE minus CI values from control wells with maximum CPE. RTCA IC_{50} values were determined by nonlinear regression analysis using GraphPad Prism 9 software.

Competition-binding ELISA. Wells of 384-well microtiter plates were coated with purified recombinant EBOV GP at 4°C overnight. Plates were blocked with 2% bovine serum albumin (BSA) in DPBS containing 0.05% Tween-20 for 1 h. Each antibody was diluted to a concentration of 10 μg/mL. Next, biotinylated antibodies were diluted to 2.5 μg/mL and added to the primary antibody solution without washing the plate to a final concentration of 0.5 μg/mL. Biotinylated antibody binding was detected with horseradish peroxidase-conjugated avidin (Sigma) and developed with TMB. The reaction was quenched with 1N hydrochloric acid once color was developed. Absorbance was measured at 450 nm using a spectrophotometer.

Cell-surface binding to cleaved or intact GP. Alexa Fluor 647 NHS ester (Thermo Fisher Scientific) was used for antibody labeling. Binding of purified polyclonal or monoclonal antibodies to Jurkat-EBOV GP or Jurkat-EBOV GPCL cells was assessed by flow cytometry using an iQue Screener Plus high-throughput flow cytometer (Intellicyt Corp.) as described previously (Gilchuk et al., 2018; Gilchuk et al., 2020b). Briefly, 50,000 cells were added per each well of V-bottom 96-well plate (Corning) in 5 mL of the DPBS containing 2% heat-inactivated ultra-low IgG FBS (Gibco) (designated as incubation buffer). Serial dilutions of antibody were added to the cells in replicates for a total volume of 50 μL per well, followed by 1 h incubation at ambient temperature, or 4°C in some experiments. Unbound antibody was removed by washing with 200 μL of the incubation buffer. Staining of cells was measured by flow cytometric analysis using the IntelliCyt iQue Screener Plus. Data for up to 20,000 events were acquired, and data were analyzed with ForeCyt (Intellicyt Corp.) software. Dead cells were excluded from the analysis based on forward and side scatter gates to identify the viable cell population. Binding to un-transduced Jurkat cells or binding of dengue antigen-specific mAb DENV 2D22 served as negative controls for most experiments.

Cells that displayed cleaved GP were prepared as described previously (Davis et al., 2019; Gilchuk et al., 2018; Gilchuk et al., 2020b). Briefly, Jurkat-EBOV GP cells were washed with DPBS containing calcium and magnesium (DPBS++), resuspended at 10^6 cells/mL in DPBS containing 0.5 mg/mL of thermolysin (Promega), and incubated for 20 min at 37°C. The cleavage reaction was inhibited by washing cells with the incubation buffer containing DPBS, 2% of heat-inactivated FBS and 2 mM EDTA (pH 8.0). The GP cleavage was confirmed by loss of mAb 13C6 binding and high-level of binding that assessed with RBD-specific mAb MR78 relative to intact Jurkat-EBOV GP antibody binding. Antibody binding to un-transduced Jurkat (mock) cells served as a control for specificity of antibody staining. For screening of the micro-scale purified mAbs,

cells were incubated with individual mAbs at a single 1:10 dilution, and the bound antibodies were detected using goat anti-human IgG antibody conjugated with PE (Southern Biotech).

Clustering for identification of public clonotypes. Publicly available paired sequence sets of antibody genes were obtained and together with sequences described in chapter IV, public clonotypes were determined by genetic similarities of antibody sequences using the following clustering scheme. The sequences were first binned by the same heavy chain V and J genes. Following sequences then were clustered according to 60% sequence similarity on their CDRH3 nucleotide sequence. Lastly, sequences then were binned together again if they used the same light chain V and J genes. Clusters of sequences containing sequences from two or more donors were determined to be public clonotypes.

Epitope mapping of antibodies by alanine scanning. Epitope mapping was performed essentially as described previously (Davidson and Doranz, 2014) using EBOV GP shotgun mutagenesis mutation libraries, made using a full-length expression construct for EBOV GP. Residues were mutated individually to alanine, and alanine residues to serine. Mutations were confirmed by DNA sequencing, and clones arrayed in a 384-well plate, one mutant per well. Binding of mAbs to each mutant clone in the alanine scanning library was determined, in duplicate, by high-throughput flow cytometry. A plasmid encoding cDNA for each GP mutant was transfected into HEK-293T cells and allowed to express for 22 h. Cells were fixed in 4% (v/v) paraformaldehyde (Electron Microscopy Sciences), and permeabilized with 0.1% (w/v) saponin (Sigma-Aldrich) in PBS plus calcium and magnesium (PBS++) before incubation with mAbs diluted in PBS++, 10% normal goat serum (Sigma), and 0.1% saponin. MAb screening concentrations were determined using an independent immunofluorescence titration curve against cells expressing wild-type GP to ensure that signals were within the linear range of detection. Antibodies were detected using 3.75 µg/mL of Alexa-Fluor-488-labeled secondary antibodies (Jackson ImmunoResearch Laboratories) in 10% normal goat serum with 0.1%

saponin. Cells were washed three times with PBS++/0.1% saponin followed by two washes in PBS, and mean cellular fluorescence was detected using a high-throughput Intellicyte iQue flow cytometer (Sartorius). Antibody reactivity against each mutant GP clone was calculated relative to wild-type GP reactivity by subtracting the signal from mock-transfected controls and normalizing to the signal from wild-type S-transfected controls. Mutations within clones were identified as critical to the mAb epitope if they did not support reactivity of the test MAb but supported reactivity of other antibodies. This counter-screen strategy facilitates the exclusion of GP mutants that are locally misfolded or have an expression defect.

Quantification and statistical analysis. The descriptive statistics mean ± SEM or mean ± SD were determined for continuous variables as noted. Curves for antibody binding and neutralization were fitted after log transformation of antibody concentrations using non-linear regression analysis. Technical and biological replicates are indicated in the figure legends. Statistical analyses were performed using Prism v8.4.3 (GraphPad).

CHAPTER VI

CONVERGENT ANTIBODY RESPONSES TO THE SARS-COV-2 SPIKE PROTEIN IN CONVALESCENT AND VACCINATED INDIVIDUALS

This chapter is an adaptation of the following manuscripts:

Chen E.C., Gilchuk P., Zost S.J., Suryadevara N., Winkler E.S., Cabel C.R., Binshtein E., Sutton R.E., Rodriguez J., Day S., Myers L., Trivette A., Williams J.K., Davidson E., Li S., Doranz B.J., Campos S.K., Carnahan R.H., Thorne C.A., Diamond M.S., and Crowe J.E. Convergent antibody responses to the SARS-CoV-2 spike protein in convalescent and vaccinated individuals. *Cell Reports (2021)*

Dong J., Zost S.J., Greaney A.J., Starr T.N., Dingens A.S., **Chen E.C.**, Chen R.E., Case J.B., Sutton R.E., Gilchuk P., Rodriguez J., Armstrong E., Gainza C., Nargi R.S., Binshtein E., Xie X., Zhang X., Shi P.Y., Logue J., Weston St., McGrath M., Frieman M., Brady T., Tuffy K., Bright H., Loo Y.M., McTamney P., Carnahan R.H., Diamond M.S., Bloom J.D., Crowe J.E. Genetic and structural basis for recognition of SARS-CoV-2 spike protein by a two antibody cocktail. *Nature Microbiology (2021)*

CHAPTER OVERVIEW

This chapter describes the work done on characterizing public clonotypes to SARS-CoV-2. The first part of the chapter details a study mining for public clonotypes to SARS-CoV-2, and describes 3 novel public clonotypes. The second part of the chapter describes work done on characterizing a public clonotype that is present in the AZD7442 antibody cocktail to SARS-CoV-2. This work could not have been done without the help and guidance of Dr. Pavlo Gilchuk, Dr. Seth Zost, and Dr. Naveenchandra Suryadevara on the experimental assays and troubleshooting for all the SARS-CoV-2 assays. Negative stain EM studies were performed by Dr. Elad Binshtein. Alanine scanning mutageneisis assays were done by Integral Molecular.

And the animal studies and authentic neutralization assays were done by Emma Winkler and Carley Cabel.

INTRODUCTION

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is the causative agent of COVID-19 and the ongoing worldwide pandemic. SARS-CoV-2 is a betacoronavirus, with other virus family members having caused global outbreaks including the 2003 SARS-CoV-1 and 2012 Middle East Respiratory Syndrome coronavirus (MERS-CoV) epidemics. The spike (S) protein is the principal antigen recognized by the protective antibody response against SARS-CoV-2(Jiang et al., 2020; Krammer, 2020). The S protein is cleaved into S1, which includes the receptor-binding domain (RBD) and the N-terminal domain (NTD), and S2, which contains the fusion peptide and heptad repeats HR1 and HR2 and mediates fusion between virus and host cell membrane(Bosch et al., 2003; Tortorici and Veesler, 2019). SARS-CoV-2 and SARS-CoV-1 share approximately 80% amino acid sequence identity, and both use human angiotensin-converting enzyme 2 (ACE2) as an entry receptor through binding mediated by the RBD(Hoffmann et al., 2020; Li et al., 2003; Wan et al., 2020).

Monoclonal antibodies (mAbs) targeting the SARS-CoV-2 S protein have been a focus for development of medical countermeasures against COVID-19. Many studies have identified antibodies to the S1 and S2 regions on the S protein, with the majority of neutralizing antibodies targeting the RBD in S1 and inhibiting ACE2 binding(Liu et al., 2020; Robbiani et al., 2020; Seydoux et al., 2020; Wec et al., 2020; Zost et al., 2020c). Multiple RBD-specific mAbs have been developed as monotherapies or cocktail therapeutics, and two (Lilly mAbs bamlanivimab [LY-CoV555] and etesevimab [LY-CoV016, also known as JS016] as well as Regeneron mAbs casirivimab and imdevimab) have received Emergency Use Authorization (EUA) (Company, 2020; Regeneron Pharmaceuticals, 2020). Additionally, multiple vaccines eliciting antibodies to

the S protein are being deployed globally under similar EUA(Johnson, 2021; Moderna, 2020; Pfizer, 2020).

In recent years, public B cell clonotypes have been identified in the human antibody repertoires formed in response to diverse viruses including Ebola(Cohen-Dvashi et al., 2020; Davis et al., 2019; Ehrhardt et al., 2019), influenza(Joyce et al., 2016; Pappas et al., 2014; Sui et al., 2009; Wheatley et al., 2015; Zost et al., 2021b), human immunodeficiency virus 1 (HIV-1)(Setliff et al., 2018; Williams et al., 2015; Wu et al., 2011; Zhou et al., 2015), hepatitis C(Bailey et al., 2017b; Giang et al., 2012), SARS-CoV-2(Nielsen et al., 2020; Yuan et al., 2020a), respiratory syncytial virus (RSV)(Mukhamedova et al., 2021), and in healthy individuals(Briney et al., 2019; Soto et al., 2019). These studies reveal a convergence of B cell selection resulting in circulating B cells clones with genetically similar antigen receptor genes in multiple individuals. The selection of public B cell clonotypes often has a structural basis mediated by low-affinity recognition of virus surface antigens by unmutated germline-encoded naïve B cell receptors that are preconfigured for binding and cell activation. Public clonotypes are of great interest, since the understanding of viral epitopes that commonly induce antibodies in humans has implications for predicting the most common responses to vaccines in large populations. With newer single-cell technologies, it is now possible to obtain paired heavy and light chain antibody variable gene sequences, allowing investigators to describe gene usage and study the function of recombinant antibodies expressed from synthesized cDNA in a large scale. This approach is powerful, since coupling genotype with function allows analysis of the role of public B cell clonotypes in the response to infection or vaccination.

There have been several efforts to characterize public clonotypes in the response to SARS-CoV-2, with most work focused on neutralizing public clonotypes(Dong et al., 2021; Robbiani et al., 2020; Tan et al., 2021; Yuan et al., 2020a) that target the S1 domain of the S trimer, more

specifically the RBD and NTD domains. However, it is less clear if public clonotypes are directed to other sites on the S trimer such as the S2 domain. Epitopes on the S2 domain may be of interest, as these sites may be more conserved than those in RBD in different strains of coronavirus due to functional constraints associated with the viral fusion mechanism. This sequence conservation reflects the fact that the S2 domain contains the HR1, HR2, and fusion loop of the S trimer, all of which are required for viral entry in coronaviruses. Given the importance of defining immune responses to SARS-CoV-2 infection or vaccination, we sought to identify the spectrum of public clonotypes, including less well studied ones directed to non-RBD regions or those lacking neutralizing activity. Understanding public clonotype recognition to all antigenic domains of the S trimer, and not just the RBD, delineates the B cell response to SARS-CoV-2 to regions that are more conserved in the S protein. In this study, we identified 37 total public clonotypes, 27 of which are shared between vaccinated and convalescent individuals. Of the public clones identified a detailed analysis of three public clonotypes (Groups 1, 2, and 3) not previously described and comparisons of public antibodies discovered from large-scale discovery efforts were investigated (**Figure VI-1**). We found that shared clonotypes comprise a substantial proportion of the elicited human B cell response to the S trimer. We also compared the response following infection or mRNA vaccination to investigate the genetic basis for the efficacy of mRNA vaccines in the population. These data show that many clonotypes are shared between convalescent and vaccinated individuals. Finally, as if diverse individuals independently make the same antibody in response to an antigen, it induces selective pressure on that epitope. And therefore, the frequent occurrence of public clonotypes recognizing sites of vulnerability on S protein that tolerate mutations may explain the rapid emergence of particular SARS-CoV-2 variant viruses in the field. The collective immunity mediated by the large number of public clonotypes described here on particular sites of vulnerability like drive the independent escape events leading to emergence of variants of concern in diverse geographic areas.

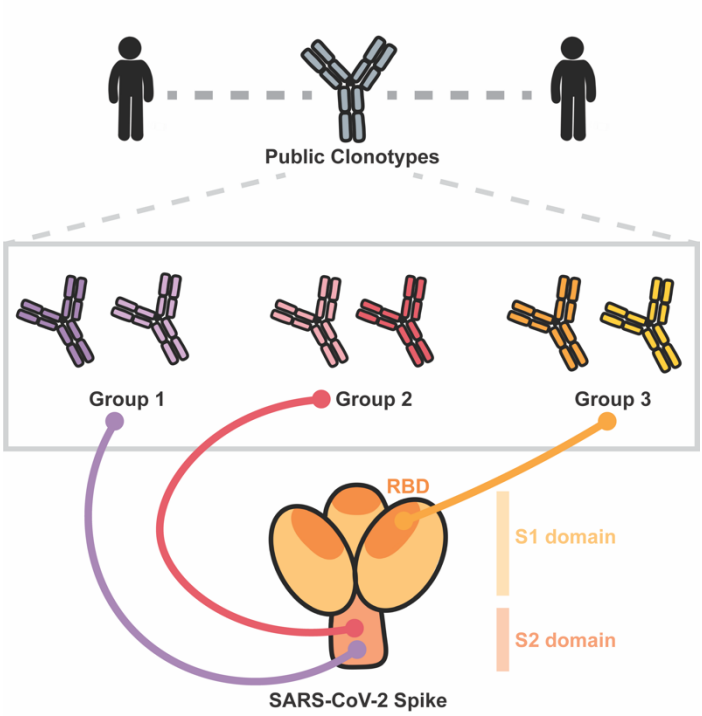


Figure VI-1: Three novel SARS-CoV-2 public clonotypes described

Three novel public clonotypes are described in this chapter: group 1 (purple), 2 (red), and 3 (orange). Two are directed to the S2 domain of the SARS-CoV-2 S protein. The last one is directed at the RBD.

RESULTS

Identification of public clonotypes

To identify a comprehensive set of public clonotypes in the B cell response to SARS-CoV-2, we first collected antibody variable gene sequences for SARS-CoV-2 human mAbs from existing publications that had isolated mAbs from individuals with a history of SARS-CoV-2 infection(Brouwer et al., 2020; Cao et al., 2020; Kreer et al., 2020; Liu et al., 2020; Robbiani et al., 2020; Rogers et al., 2020; Seydoux et al., 2020; Wec et al., 2020; Zost et al., 2020c). This search identified a panel of 2,865 paired heavy and light chain variable gene sequences for analysis. We clustered all sequences by binning the clones based on the inferred immunoglobulin

heavy variable (*IGHV*) gene, immunoglobulin heavy joining (*IGHJ*) gene, and the amino acid length of the heavy chain complementarity determining region 3 (CDRH3). These sequences then were clustered according to 70% nucleotide sequence identity in the DNA sequence encoding the CDRH3. Next, the sequences were binned further based on the inferred immunoglobulin light variable gene (*IGLV* or *IGKV*) and immunoglobulin light joining (*IGLJ* or *IGKJ*) genes. Clusters meeting these similarity criteria in both heavy and light chains with sequences originating from two or more individuals were deemed public clonotypes (**Figure VI-3**). Eleven public clonotypes were identified in the repertoires of subjects with prior natural infection (**Figure VI-2A, B, C**), and these clones are encoded by a variety of heavy and light chain variable genes. Of the 11 public clonotypes identified, five of the heavy chain genes have been reported previously to encode potently neutralizing SARS-CoV-2 antibodies that bind to the RBD: *IGHV3-53(Yuan et al., 2020a), IGHV1-58(Dong et al., 2021), IGHV3-30(Robbiani et al., 2020), IGHV3-30-3, IGHV3-66(Yuan et al., 2020a)*, whereas three have not been reported: *IGHV1-69, IGHV4-59, IGHV3-7*. *IGHV3-53* and *IGHV3-66* are commonly observed in antibodies in SARS-CoV-2 patients since the germline gene segments encode amino acid motifs that are preconfigured for RBD binding(Yuan et al., 2020a). *IGHV1-58* also commonly encodes antibodies that neutralize SARS-CoV-2, as this germline gene segment encodes motifs that mediate binding to the S protein. Notably, *IGHV1-58* encodes the mAb COV2-2196, which is the basis for one of the two antibodies in a cocktail currently in Phase III clinical trials(Zost et al., 2020a). Clonally expanded B cell populations containing potently neutralizing antibodies encoded by *IGHV3-30* also have been found in multiple individuals(Dong et al., 2021; Robbiani et al., 2020). However, the role of *IGHV1-69*, *IGHV4-59*, and *IGHV3-7* public clonotypes in SARS-CoV-2 responses remains unknown. In this paper, for clarity, we designated public clonotypes incorporating these additional three V_H gene segments as members of Group 1, 2, or 3 mAbs, respectively (**Figure VI-2C, D, E**). Group 1 is shared by two donors from the cohort we studied and includes mAbs COV2-2002 and COV2-2333. Group

2 is shared by a donor from our group and a previously described donor IDCnC2(Kreer et al., 2020) and includes antibodies COV2-2164 and CnC2t1p1_B10. Lastly, Group 3 is shared by a donor from our group and a previously described donor COV107(Robbiani et al., 2020) and includes antibodies COV2-2531 and C126 (**Figure VI-2C, D, E**). cDNAs for the antibody variable genes encoding each of the six antibodies from the three groups of public clonotypes were synthesized and cloned into an IgG1 expression vector, as previously described(Gilchuk et al., 2020a).

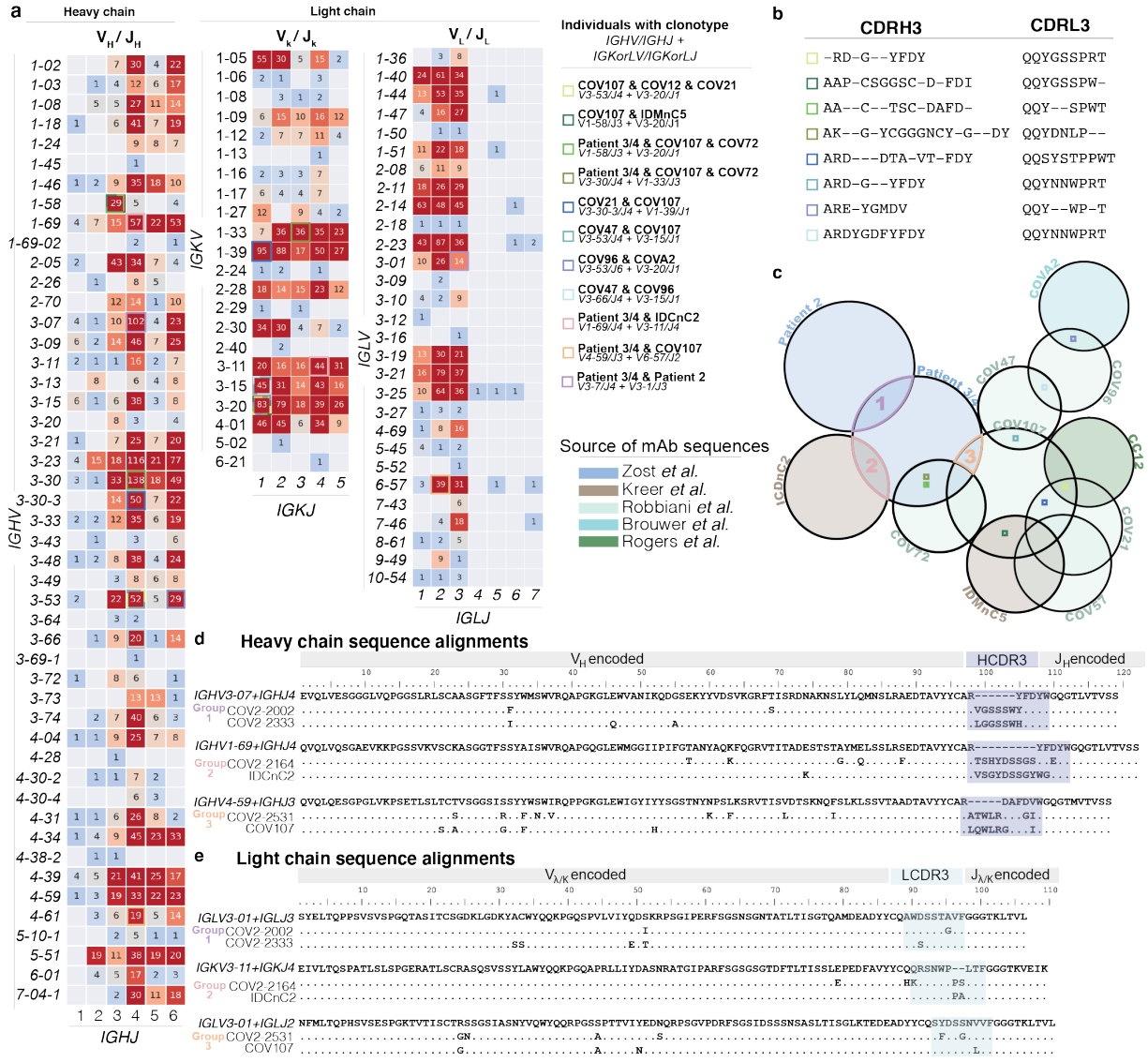


Figure VI-2: Sequence characteristics of monoclonal antibodies to SARS-CoV-2.
(Legend on next page)

Figure VI-2 legend: **(A)** Available sequences of mAbs to SARS-CoV-2 from multiple publications were obtained from public databases. Numbers inside each box represents the number of sequences with the indicated gene usage. Colored outlining boxes represent public clonotypes that are shared between the individuals listed in the key to the right side. The heat map is color coded so that red represents a higher number of sequences using the corresponding genes, and blue represents a lower number of sequences using the corresponding genes. **(B)** CDR3 sequences of the heavy and light chains of each of the remaining eight public clonotypes are shown. Dashes represent amino acids that differed in the public clonotype. Each box color correlates to the public clonotypes in **Fig. 1a** and **Fig. 1c**. **(C)** A Venn diagram illustrating all of the public clonotypes identified between naturally-infected individuals. The colored boxes in the Venn diagram overlaps represent the public clonotypes identified in Fig 1a. Novel public clonotypes, designated as Groups 1, 2, or 3, are highlighted in the purple, pink, or orange overlaps respectively. (D)Multiple sequence alignments of the heavy chain sequences for Groups 1, 2, or 3 to their respective inferred germline genes *IGHV 3-07/IGHJ4, IGHV1-69/IGHJ4, or IGHV4-59/IGHJ3.* The CDRH3 sequence is highlighted in dark blue. (E) Multiple sequence .alignments of the light chain sequences for Groups 1, 2, or 3 to their respective inferred germline genes *IGLV3-01/IGLJ3, IGKV3-11/IGKJ4, or IGHV3-01/IGLJ2.* The CDRL3 sequence is highlighted in light blue.

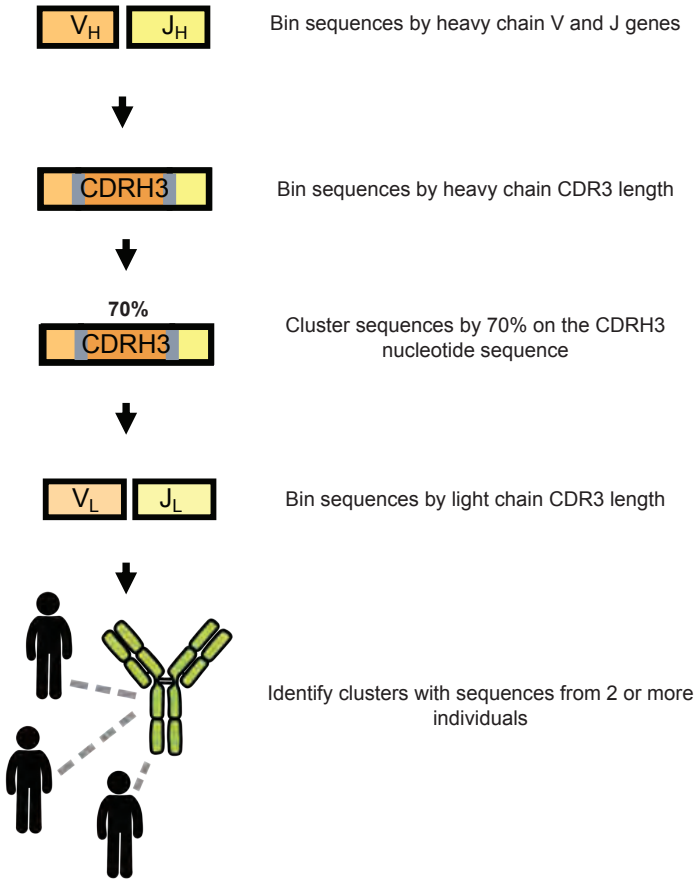


Figure VI-3: Clustering to identify public clonotypes.

Schematic of how sequences were binned and clustered to identify public clonotypes.

Functional properties of identified public clonotype antibodies.

To examine the binding properties of antibodies in these three new SARS-CoV-2 public clonotypes, we tested six recombinant purified antibodies, two for each public clonotype, for binding to recombinant stabilized trimeric prefusion ectodomain of the SARS-CoV-2 S protein ($S6P_{ecto}$), SARS-CoV-2 RBD, or recombinant stabilized trimeric prefusion ectodomain of the SARS-CoV-1 S protein ($S2P_{ecto}$) proteins by ELISA (**Figure VI-4**). The two Group 1 antibodies, COV2-2002 and COV2-2333, did not bind to SARS-CoV-2 RBD, but both bound to SARS-CoV-2 $S6P_{ecto}$ and SARS-CoV-1 $S2P_{ecto}$ proteins. However, they did not saturate in binding to SARS-CoV-1 $S2P_{ecto}$ at the maximum concentration tested (400 ng/mL) indicating relatively weak binding to recombinant SARS-CoV-1 $S2P_{ecto}$. Group 2 antibodies, which include COV2-2164 and CnC2t1p1_B10, did not bind to SARS-CoV-2 RBD, but both bound to SARS-CoV-2 $S6P_{ecto}$ and

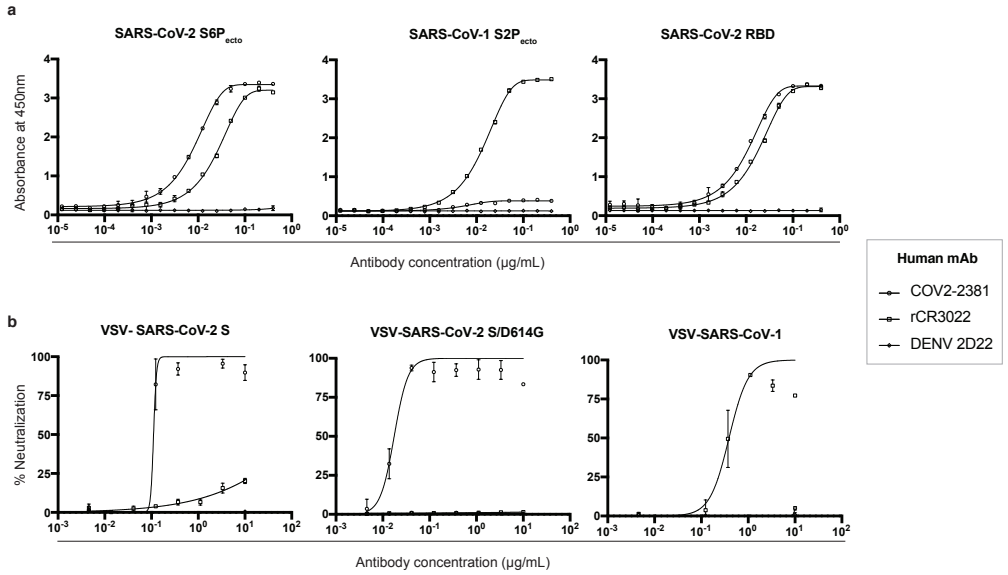


Figure VI-4: Controls for ELISA and neutralization assays.

(A) Positive or negative controls used for testing antibody binding in ELISA to SARS-CoV2-$S6P_{ecto}$, SARS-CoV-1 $S2P_{ecto}$, or SARS-CoV-2 RBD proteins. The positive control antibody COV2-2381 binds to SARS-CoV-2 $S2P_{ecto}$ and RBD but not to SARS-CoV-1 $S2P_{ecto}$, and the positive control antibody rCR3022 also binds to SARS-CoV-1 $S2P_{ecto}$. **(B)** Positive or negative controls used for replication-competent chimeric VSV neutralization assays. COV2-2381 was used as a positive control for SARS-CoV-2 WT and D614G, whereas rCR3022 was used as a positive control for SARS-CoV-1.

SARS-CoV-1 $S2P_{ecto}$ proteins. Group 3 antibodies, which include COV2-2531 and C126, bound to SARS-CoV-2 $S6P_{ecto}$ and SARS-CoV-2 RBD proteins (**Figure VI-5A, H**). However, antibodies from Group 3 did not bind SARS-CoV-1 $S2P_{ecto}$.

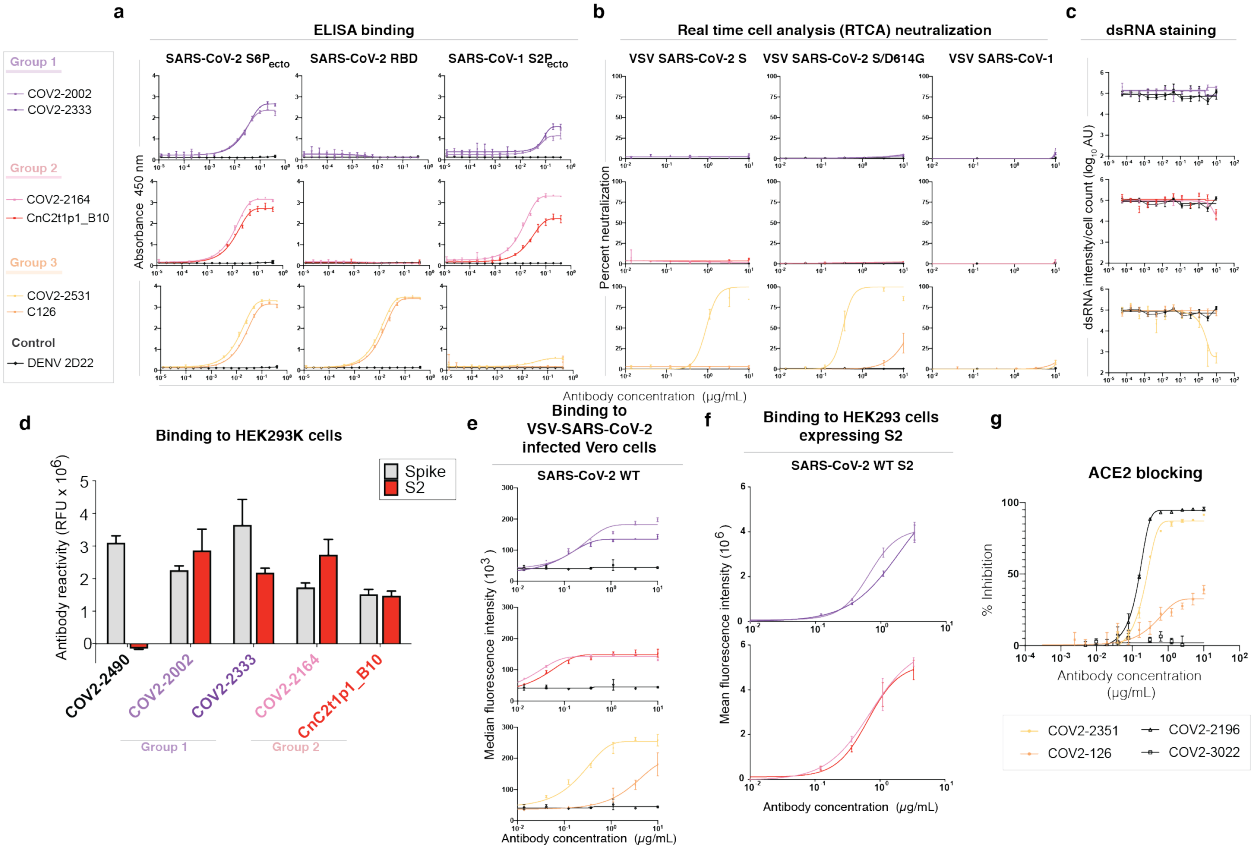
As antibodies from Groups 1 and 2 did not bind the RBD but cross-reacted to both SARS-CoV-2 S6P and SARS-CoV-1 S2P, we hypothesized that they might bind the S2 domain of the S trimer; SARS-CoV-2 infection can elicit antibodies that recognize cross-reactive epitopes on the S2 domain(Ladner et al., 2021). Antibodies were tested for binding against the S2 domain of SARS-CoV-2 S expressed on HEK-293T cells. An NTD-directed antibody COV2-2490 was used as a control. This experiment showed that Group 1 and 2 antibodies bound to S2 in a dose-dependent manner (**Figure VI-5D**, **F, H**), and revealed that public clonotypes can be elicited to the S2 domain of the S trimer.

Antibodies from each group then were tested for neutralizing activity using a previously described real-time cell analysis (RTCA) assay that measures cellular impedance(Gilchuk et al., 2020a; Zost et al., 2020c). We used recombinant, infectious vesicular stomatitis virus (VSV) expressing the S proteins from SARS-CoV-2 (WA1/2019 strain), SARS-CoV-2/D614G, or SARS-CoV-1 (Urbani strain) (**Figure VI-4**). In addition, we used authentic infectious SARS-CoV-2 (WA1/2019) virus and Calu3 (human lung epithelial adenocarcinoma) cell monolayer cultures, and neutralization was measured by staining for double-stranded RNA, which is produced in the cytoplasm in virus-infected cells (**Figure VI-6**)**.** Group 3 mAb COV2-2531 neutralized SARS-CoV-2 (VSV-SARS-CoV-2, and VSV-SARS-CoV-2/D614G (**Figure VI-5B, C, H**) and authentic SARS-CoV-2, but not SARS-CoV-1. In contrast, another Group 3 mAb, C126 partially neutralized SARS-CoV-2/D614G variant but did not neutralize the WT VSV-SARS-CoV-2, VSV-SARS-CoV-1, or authentic virus. Groups 1 and 2 antibodies did not exhibit neutralizing capacity for any of the viral strains tested.

As both Group 3 antibodies exhibited neutralizing capacity, we considered that they might bind to the RBD and block virus attachment to ACE2, a principal mechanism of inhibition by RBD-targeted antibodies against SARS-CoV-2(Hansen et al., 2020; Zost et al., 2020a). We tested whether each antibody could block binding of soluble trimeric S protein to recombinant human ACE2 protein in an ELISA. Only Group 3 antibodies blocked binding to ACE2 (**Figure VI-5H**). Similar to the pattern we observed for neutralization, COV2-2531 fully blocked ACE2 binding, whereas C126 partially blocked binding, with less than 50% inhibition at maximal effect (**Figure VI-5G**). Therefore, it is likely that COV2-2531 neutralizes virus infection at least in part by blocking binding to ACE2.

Additionally, we sought address the impact of mutations in recently emerged variants of SARS-CoV-2 on the binding of mAbs from each of our public clonotypes. Therefore, we assessed binding of all 6 antibodies to P.1 (Gamma), P.2 (Zeta), B.1.429 (Epsilon), B.1.1.7 (Alpha), B.1.135 (Beta), and B.1.1.298 (**Figure VI-7**). Antibody binding of both group 1 and 2 mAbs was largely affected by variant residues in B.1.1.7 or B.1.135. However, group 3 antibody COV-2531 maintained its ability to bind all variants tested. C126, the less mutated counterpart of the group 3 public clonotype exhibited a decrease in binding to P.2, B.1.1.7, and B.1.135 (**Figure VI-5I**).

As COV2-2531 maintained binding to all variants tested, we tested if it retained its neutralization capacity. Group 3 antibodies COV2-2531 and C126 were tested against clinical isolates for B.1.1.7 (Alpha), B.1.1.298, and B.1.429 (Epsilon) viruses. The mAbs also were tested against a chimeric WA1/2020 virus encoding the spike gene of B.1.1.28 (Gamma, Wash-B.1.1.28). These results were compared to neutralization of WA1/2020 and an isogenic mutant containing the D614G mutation (WA1/2020 D614G) (**Figure VI-2J**). C126 lacked inhibitory activity against authentic SARS-CoV-2 viruses. In comparison, COV2-2531 neutralized the panel of variant SARS-CoV-2 viruses comparably to WA1/2020, with IC_{50} values ranging from 85 to 522 ng/mL (**Figure VI-2K**)



h

Group	mAb	Binding to indicated antigen EC 50 (µg/mL)					Neutralization/ blocking IC 50 (µg/mL)				
		SARS-CoV-2 S6P ecto	SARS-CoV-2 RBD	SARS-CoV-1 S2P ecto	SARS-CoV-2 WT	SARS-CoV-2 S2	SARS-CoV-2 WT VSV	SARS-CoV-2 D614G VSV	SARS-CoV-1 VSV	dsRNA intensity	ACE2 blocking%
1	COV2-2002	0.017	-	> 0.400	0.16	0.67	-	-	-	-	13
	COV2-2333	0.024	-	> 0.400	0.08	1.00	-	-	-	-	22
2	COV2-2164	0.010	-	0.010	0.01	0.61	-	-	-	-	5
	CnC2t1p1_B10	0.012	-	0.020	0.03	0.63	-	-	-	-	11
3	COV2-2531	0.012	0.009	-	0.17	-	1.6	1.5	-	0.57	98
	C126	0.019	0.012	-	3.7	-	-	>10	-	-	53

i

Binding to spike protein of different SARS-CoV-2 variants

	P.1	P.2	B.1.429	B.1.1.7	B.1.135	B.1.1.298
COV2-2002	1.3 ± 0.13	1.2 ± 0.07	0.9 ± 0.11	0.3 ± 0.01	0.8 ± 0.03	0.8 ± 0.03
COV2-2333	1.0 ± 0.06	0.9 ± 0.08	0.9 ± 0.03	0.5 ± 0.01	0.8 ± 0.05	1.0 ± 0.03
COV2-2164	1.1 ± 0.02	1.0 ± 0.1	1.0 ± 0.03	0.6 ± 0.03	0.9 ± 0.1	1.1 ± 0.1
CnC2t1p1_B10	1.2 ± 0.09	1.1 ± 0.12	1.0 ± 0.05	0.5 ± 0.02	0.7 ± 0.03	1.0 ± 0.11
COV2-2531	1.3 ± 0.16	1.04 ± 0.1	1.1 ± 0.05	1.0 ± 0.1	1.0 ± 0.06	1.1 ± 0.06
C126	1.1 ± 0.04	0.9 ± 0.02	1.4 ± 0.03	0.8 ± 0.04	0.9 ± 0.02	1.3 ± 0.06

k

IC50 values of Group 3 antibodies against SARS-CoV-2 variants

	WA1/2020	D614G	B.1.1.7	P1	B.1.1.298	B.1.429
COV2-2196	4	3	9	24	5	9
COV2-2531	85	469	206	114	249	522
C126	>10,000	>10,000	>10,000	>10,000	>10,000	>10,000
2D22	>10,000	>10,000	>10,000	>10,000	>10,000	>10,000

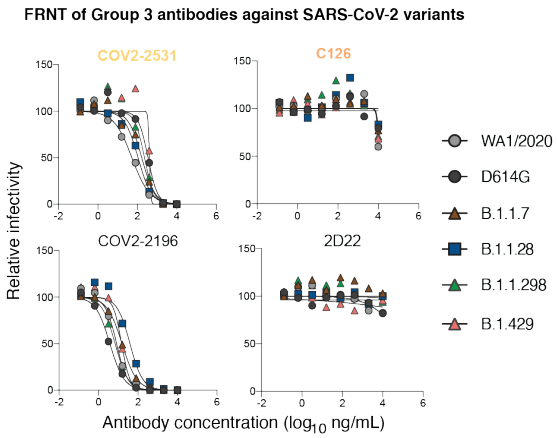


Figure VI-5: Reactivity and functional activity of Groups 1, 2, and 3 antibodies.
(Legend on next page)

Figure VI-5 legend: Group 1 antibodies are shown in light or dark purple, Group 2 antibodies are in red or pink, and Group 3 antibodies are in light or dark orange. MAb DENV 2D22 was used as a negative control antibody, as shown in the lines in black. All experiments are performed in biological replicates and technical triplicates. Biological replicate from representative single experiment shown. **(A)** ELISA binding to SARS-CoV-2 $S6P_{ecto}$, SARS-CoV-2 RBD, or SARS-CoV-1 $S2P_{ecto}$ was measured by absorbance at 450 nm. Antibody concentrations starting at 0.4 μg/mL were used and titrated two-fold. **(B)** Neutralization activity of antibodies to VSV-SARS-CoV-2, VSV-SARS-CoV-2/D614G, and VSV-SARS-CoV-1 determined by using real time cell analysis (RTCA) assay. The percent of neutralization is reported. Antibody concentrations started at 10 μg/mL and were titrated three-fold. **(C)** Neutralization activity of antibodies to authentic SARS-CoV-2 (USA-WA1/2020) determined by measuring dsRNA intensity per cell count after Calu3 lung epithelial cells were inoculated with SARS-CoV-2. Antibody concentrations started at 10 μg/mL and were titrated three-fold. **(D)** Antibody binding to full-length S (grey) or S protein C-terminus S2 region (red) expressed on the surface of HEK-293T cells that were fixed and permeabilized. Antibodies were screened at 1 μg/mL. Antibody reactivity was measured by flow cytometry and cellular florescence values were determined. COV2-2490, an NTD-directed antibody, was used as a control. **(E)** Binding to VSV-SARS-CoV-2-infected Vero cells (SARS-CoV-2 WT) was measured using flow cytometry and median florescence intensity values were determined for dose-response binding curves. Antibody was diluted 3-fold staring from 10 μg/mL. **(F)** Binding to S protein C-terminus S2 region expressed on HEK-293T cells (SARS-CoV-2 WT S2) was measured using flow cytometry and mean fluorescence intensity values were determined for dose-response binding curves. Antibody was diluted 3-fold starting from 10 μg/mL. **(G)** Inhibition of ACE2 binding curves for COV2-2531 or C126. Antibody concentrations started at 10 μg/mL and were titrated 3-fold to identify ACE2 blocking curves. COV2-2531 is shown in light orange, and C126 is shown in dark orange. **(H)** Binding EC_{50} and neutralization IC_{50} values for each of the assay curves in **Fig 3a, b, c, d, e**. All values are denoted as μg/mL. ACE2 blocking was determined by measuring amount of ACE2 with FLAG tag binding in the presence of each antibody, measured by binding of an anti-FLAG antibody. Percent blocking is shown, calculated by using ACE2 binding without antibody as 0% blocking. **(I)** Binding of each antibody to several variants of concern spike proteins. Binding of each antibody (at 1 μg/mL) to the SARS-CoV-2 spike variants is shown relative to the antibody's binding to the wild-type spike, defined as a value of 1.0. The relative amounts of each variant expressed in cells were estimated using the signal for antibody 1A9, normalizing the average 1A9 binding (at 1 μg/mL) for each variant to the average 1A9 binding with the wild-type construct. For each public clonotype antibody, the binding values to each variant spike were corrected for spike expression equivalent to that of the wild-type spike. Darker blue indicates less change is to binding of that antibody to the variant. Lighter blue color indicates more change for binding of that antibody to the variant. **(J)** Neutralization curves of group 3 antibodies against variant SARS-CoV-2 strains. Antibodies were tested for inhibition of infection of the indicated viruses on Vero-TMPRSS2 cell monolayer cultures using a focus reduction neutralization test. **(K)** Antibody neutralization IC_{50} values for group 3 antibodies against variant SARS-CoV-2 strains. One representative experiment of two performed in duplicate and mean IC_{50} values (ng/mL) from two independent experiments are shown.

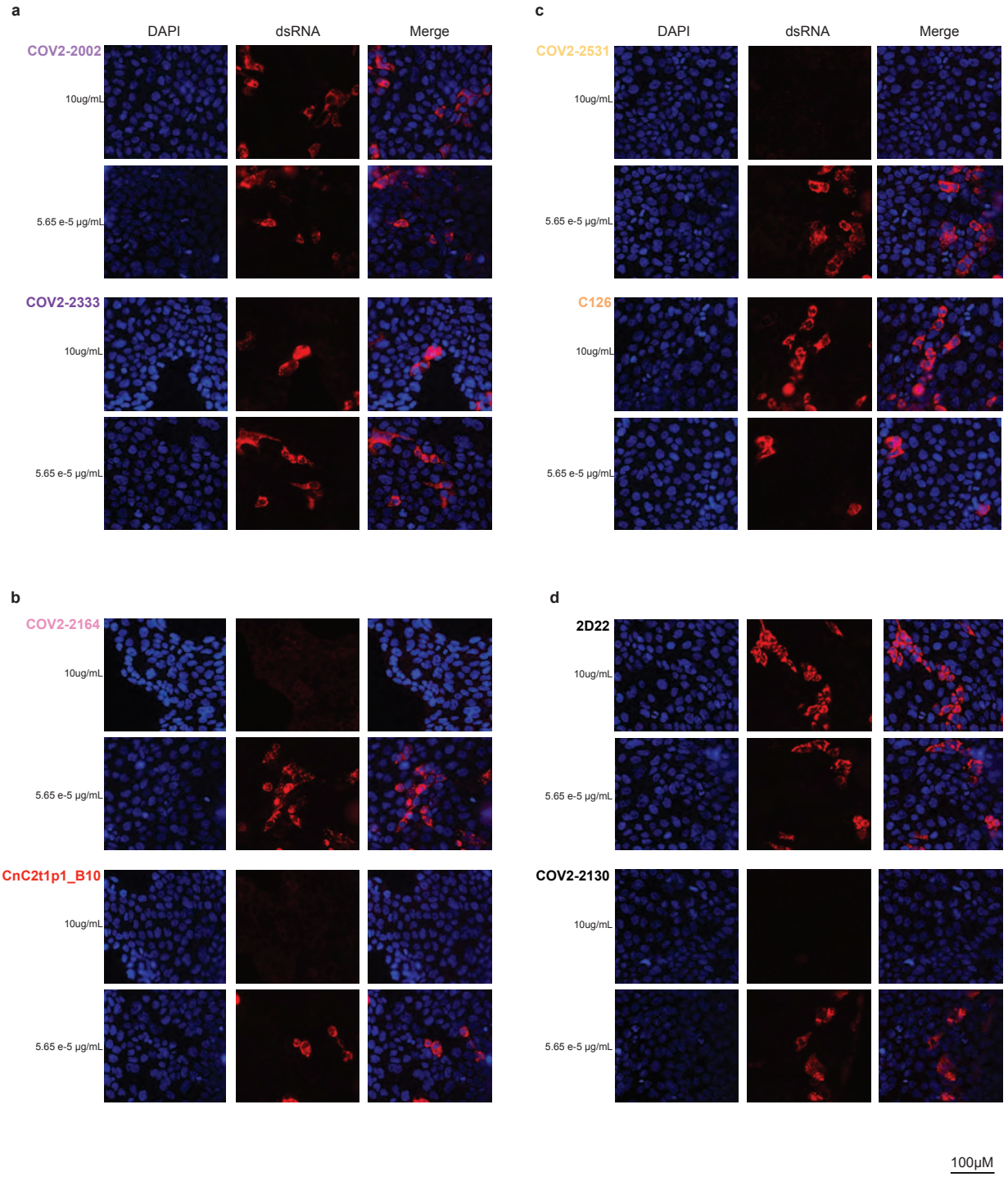


Figure VI-6: Staining of dsRNA intensity.

Staining of dsRNA intensity split into DAPI stain, dsRNA stain, and the two merged for each antibody group. **(A)** Staining for Group 1 antibodies. **(B)** Staining for Group 2 antibodies. **(C)** Staining for Group 3 antibodies.

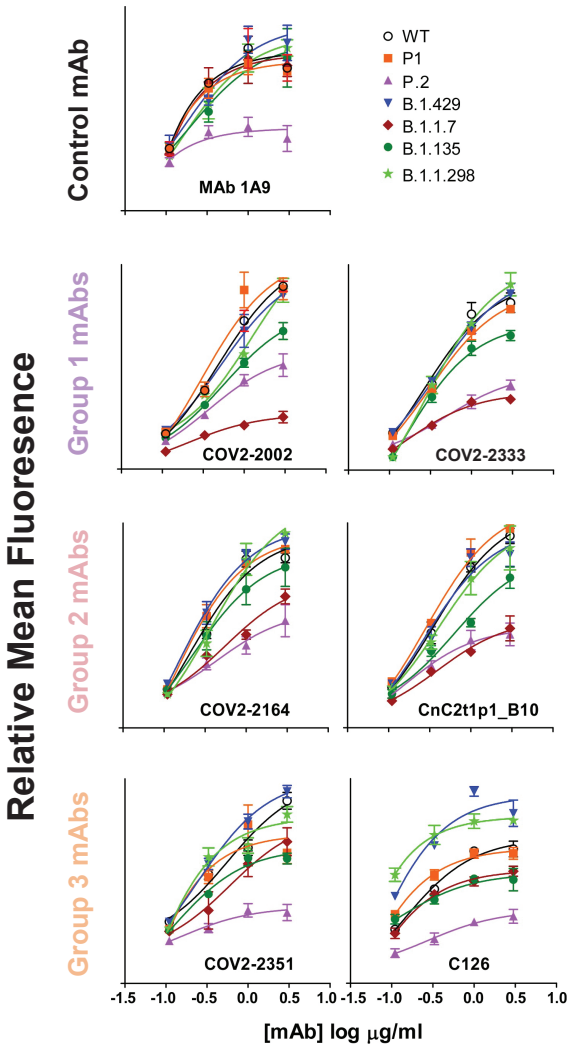


Figure VI-7: Antibody binding to cell surface displayed variant S protein.

All public clonotypes were tested against all variants listed, starting at 1 μg/mL. 1A9 from Genetex, an anti-S2 non-conformational antibody, was used as a control to test for spike protein expression.

Binding sites of identified clonotype antibodies.

We used negative stain electron microscopy (EM) to image Fab-SARS-CoV-2 $S6P_{ecto}$ complexes. Even though all of the antibodies bound to S protein in ELISA as IgG1, only the Group 3 antibody Fabs COV2-2531 and C126 formed complexes visualized on EM grids, suggesting

that some antibodies may require an IgG format for strong binding (**Figure VI-8A**, **Figure VI-9**). Low-resolution 3D reconstructions for COV2-2531 and C126 showed that these two antibodies bind the side of the RBD and recognize the cryptic face of the RBD that is accessible only in the "open" position of the RBD in the context of the S trimer (**Figure VI-8B,C**).

We then tested binding of antibodies to the full-length membrane-bound S protein using infected Vero cells that were inoculated with VSV-SARS-CoV-2 chimeric viruses. We used a dengue virus specific antibody (DENV 2D22)(Fibriansah et al., 2015) and SARS-CoV-2-reactive antibody (COV2-2381)(Zost et al., 2020c) as controls (**Figure VI-10**). Antibodies from each of the groups bound to infected cells dose-dependently, with the Group 3 RBD-reactive antibodies exhibiting greater binding than the Group 1 or 2 S2-reactive antibodies (**Figure VI-5E, H**). Binding of the Group 3 antibodies correlated with their neutralization capacity, as COV2-2531 showed greater binding than C126. The capacity to bind to infected cells also suggested that these antibodies could act *in vivo* not only by direct virus neutralization but also through Fc-mediated functions.

To identify if the antibodies within each discrete public clonotype group bind similar epitopes, we used competition-binding ELISA for pairwise comparison of antibodies binding to the $S6P_{ecto}$ protein (**Figure VI-8D**). As expected, members of each public clonotype group clustered with the other member of the same group by competition-binding. To begin to determine specific epitopes recognized by mAbs in each group, we competed the antibodies for binding against a larger group of epitope-mapped antibodies we previously described(Zost et al., 2020c), that covers various sites on the S protein, and against rCR3022(Yuan et al., 2020b), which bind less well to the RBD of SARS-CoV-2 compared to SARS-CoV-1 and does not block ACE2 binding (**Figure VI-5G**). Both Group 3 antibodies competed with rCR3022, with COV2-2531 exhibiting a higher level of competition than C126. None of the Group 1 or 2 antibodies competed with the reference antibodies tested (**Figure VI-8E**).

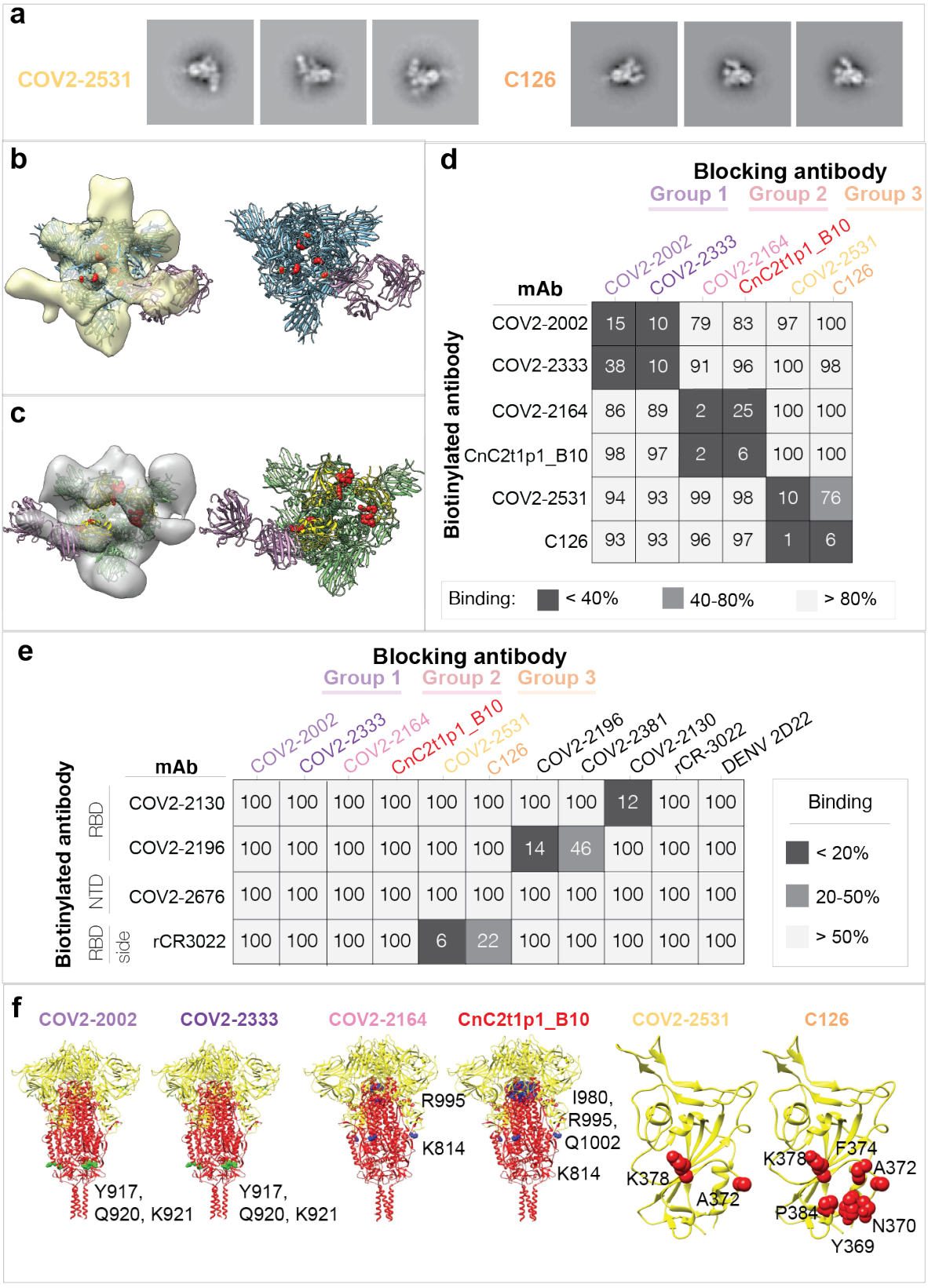


Figure VI-8: Epitope identification and structural characterization of antibodies.
(Legend on next page)

Figure VI-8 legend: **(A)** Negative stain EM of SARS-CoV-2 $S6P_{ecto}$ protein in complex with Fab forms of different mAbs. Negative stain 2D classes of SARS-CoV-2 S protein incubated with COV2-2531 or C126. Box size is 128 pix at 4.36 Å/pix. **(B)** MAb COV2-2531 3D volume with critical residues 372 and 275 shown in red on the S protein (blue). RBD is in the open position. Density corresponds to three Fabs, as we docked a single Fab structure onto the EM density map, shown in magenta. **(C)** MAb C126 3D volume with critical binding residues shown in red. The Fab is docked to a protomer of SARS-CoV-2 S protein in the open conformation. Top left is the RBD positioned in open conformation, with the other two protomers in the trimer in closed position. The S protein is shown in green, with the RBD in yellow. The Fab is shown in magenta. **(D)** Competition-binding ELISA results for mAbs within each clonotype group. Unlabeled blocking antibodies applied to antigen first are listed across the top, while biotinylated antibodies that are added to antigen-coated wells second are indicated on the left. The number in each box represents percent un-competed binding of the biotinylated antibody in the presence of the indicated competing antibody. Heat map colors range from dark grey (<40% binding of the biotinylated antibody) to light grey (>80% binding of the biotinylated antibody). Experiment was performed in biological replicate and technical triplicate. Biological replicate from representative single experiment shown. **(E)** Competition-binding ELISA data using Group 1, 2, or 3 antibodies against epitope-mapped reference antibodies. Biotinylated antibodies are indicated on the left, and the unlabeled antibodies applied to antigen first are indicated across the top. Heat map colors range from dark grey (<20% binding of the biotinylated antibody) to light grey (>50% binding of the biotinylated antibody). Experiment was performed in biological replicate and technical triplicates. Biological replicate from representative single experiment shown. **(F)** Alanine scanning mutagenesis results for Group 1, 2 or 3 antibodies. S2 epitope residues are shown (green spheres or blue spheres) on the S protein structure (PDB 6XR8), S1 is colored yellow, S2 red. RBD epitopes are shown in red on the RBD structure (PDB 6XR8). Primary data shown in **Fig. S5**.

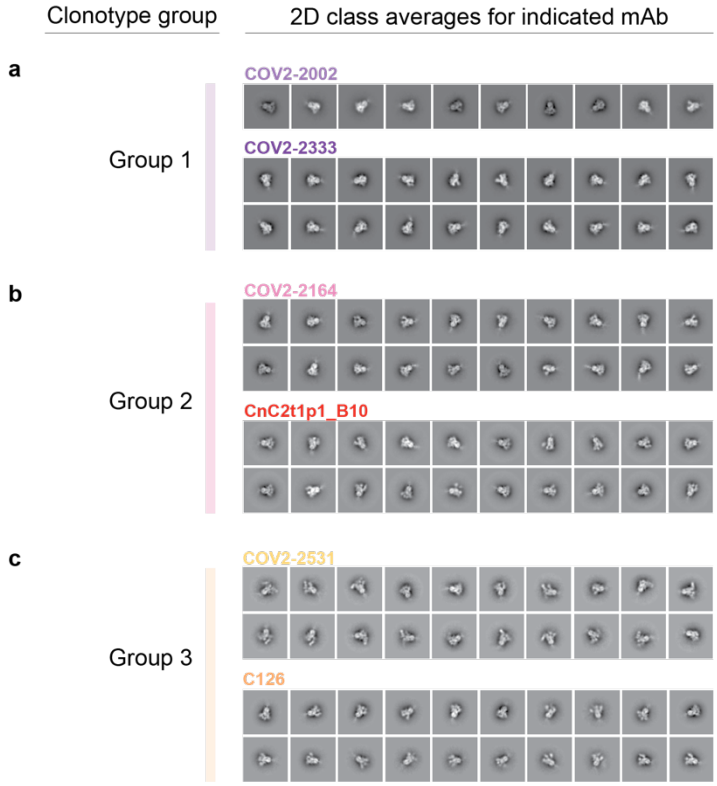


Figure VI-9: Negative stain complexes of each public clonotype

(A) Negative stain EM of SARS-CoV-2 $S6P_{ecto}$ protein in complex with Fab forms of COV2-2002 or COV2-2333. (B) Negative stain EM of SARS-CoV-2 $S6P_{ecto}$ protein in complex with Fab forms of COV2-2164 or CnC2t1p1_B10. (C) Negative stain EM of SARS-CoV-2 $S6P_{ecto}$ protein in complex with Fab forms of COV2-2531 or C126.

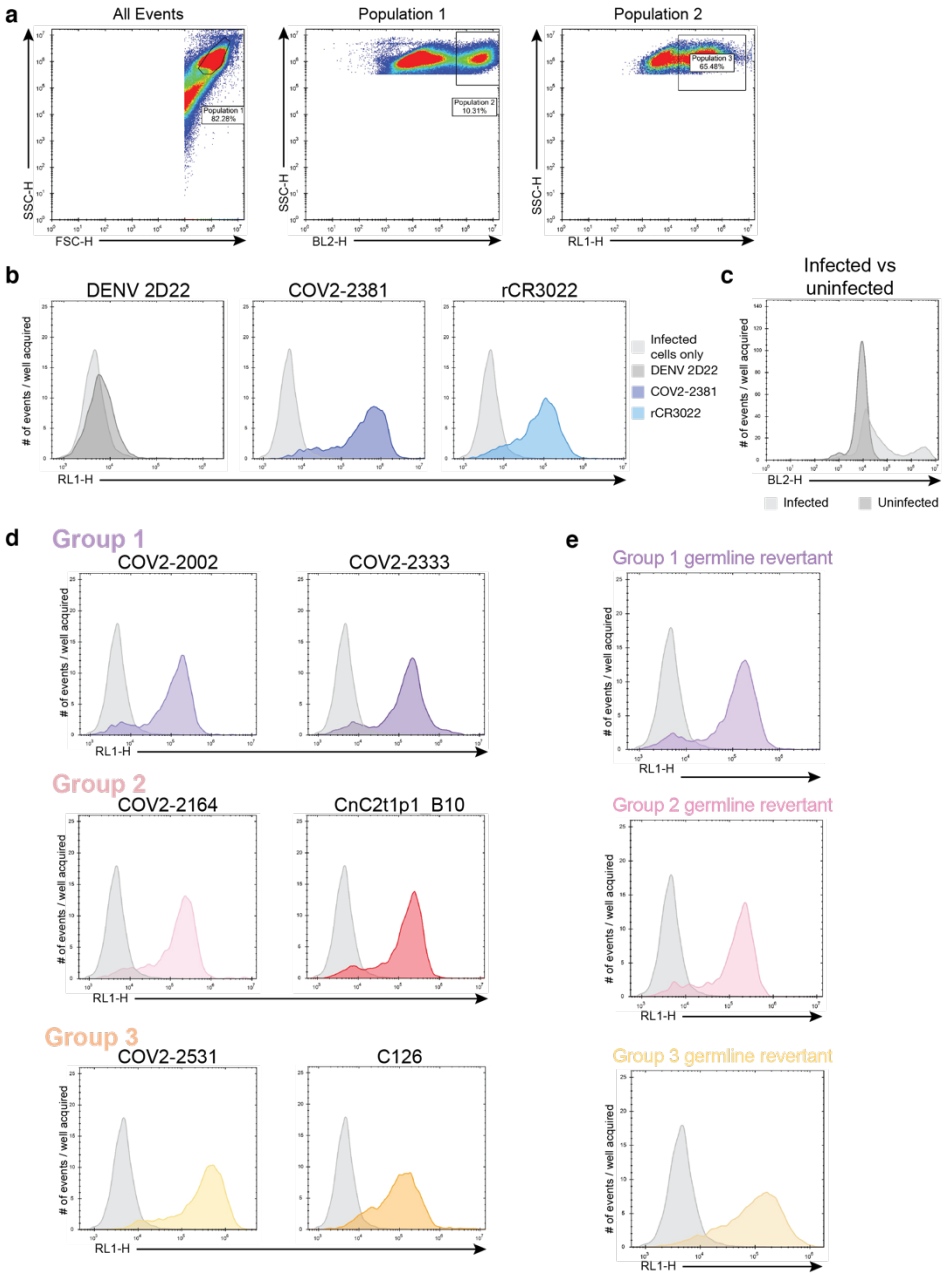


Figure VI-10: Control reagents for detection of antibody binding to membrane-anchored S protein in cell-surface antigen-display assays

(A) Gating strategy used for cell-surface antigen-display experiment. The first gate is for all cells, the second gate is for infected cells, and the third gate is for antibody binding to infected cells. (B) Controls used for cell-surface antigen-display antibody binding experiment. Cell-only control in shown in light grey. The unrelated mAb DENV 2D22 was used as an antibody negative control, shown in dark grey. The mAb COV2-2381 shown in dark blue and mAb rCR3022 shown in turquoise were used as positive antibody controls. (C) Histogram of data obtained using infected or uninfected cells. Infected cells are shown in light grey, and uninfected cells are shown in dark grey. (D) Group 1, 2, or 3 antibody binding to infected cells. The antibody concentration used was 10 μg/mL for all antibodies. (E) Group 1, 2, or 3 germline-revertant antibody binding to infected cells. The antibody concentration used was 10 μg/mL for all antibodies.

We then determined the critical binding residues at the amino acid level for each of the public clonotype antibodies by screening for binding to alanine-scanning mutant libraries of the SARS-CoV-2 S protein. Screening the RBD library revealed A372 and K378 as critical residues for COV2-2531 binding. For C126, we also identified A372 and K378, but with additional critical residues Y369, N370, F374, and P384 (**Figure VI-8F**, **Figure VI-11**). Notably, the identified residues are consistent with the binding site identified in the negative stain EM analyses and overlap with the epitope of CR3022(Yuan et al., 2020b). It was curious that several SARS-CoV-2-specific neutralizing antibodies competed with CR3022, which also binds to SARS-CoV-1 but is non-neutralizing. It is of note that SARS-CoV-1 has an N-glycosylation site at N370, in the binding site for these two mAbs, which SARS-CoV-2 lacks(Yuan et al., 2020b). This difference in glycosylation likely explains why COV2-2531 and C126 do not bind or neutralize SARS-CoV-1, even though they recognize the relatively conserved cryptic face of the RBD (**Figure VI-12**). In the alanine scanning libraries, native alanine residues are changed to serine. It is possible that

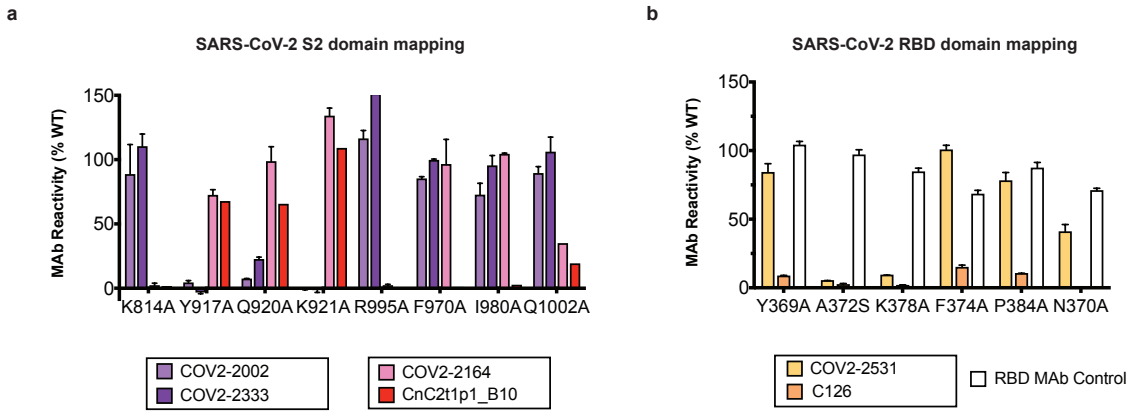


Figure VI-11: Primary data for alanine mutagenesis screening.

Binding values for mAbs on the SARS-CoV-2 S protein alanine scan library. The binding values at critical mutant clones for (A) Group 1 (COV2-2002 in light purple and COV2-2333 in dark purple) and Group 2 (COV2-2164 in pink and CnC2t1p1_B10 in red)antibodies are shown as a percentage of mAb binding to wild-type (WT) SARS-CoV-2 spike protein and are plotted with the range (highest-minus lowest binding value) of at least two measurements. (B) Group 3 (COV2-2531 in light orange and C126 in dark orange) antibodies are shown as a percentage of mAb binding to wild-type (WT) SARS-CoV-2 spike protein and are plotted with the range (highest-minus lowest binding value) of at least two measurements.

A372 was identified as critical for binding by COV2-2531 and C126 because the A372S mutation results in the introduction of N-linked glycosylation of N370, rather than making direct side-chain contact with the antibodies.

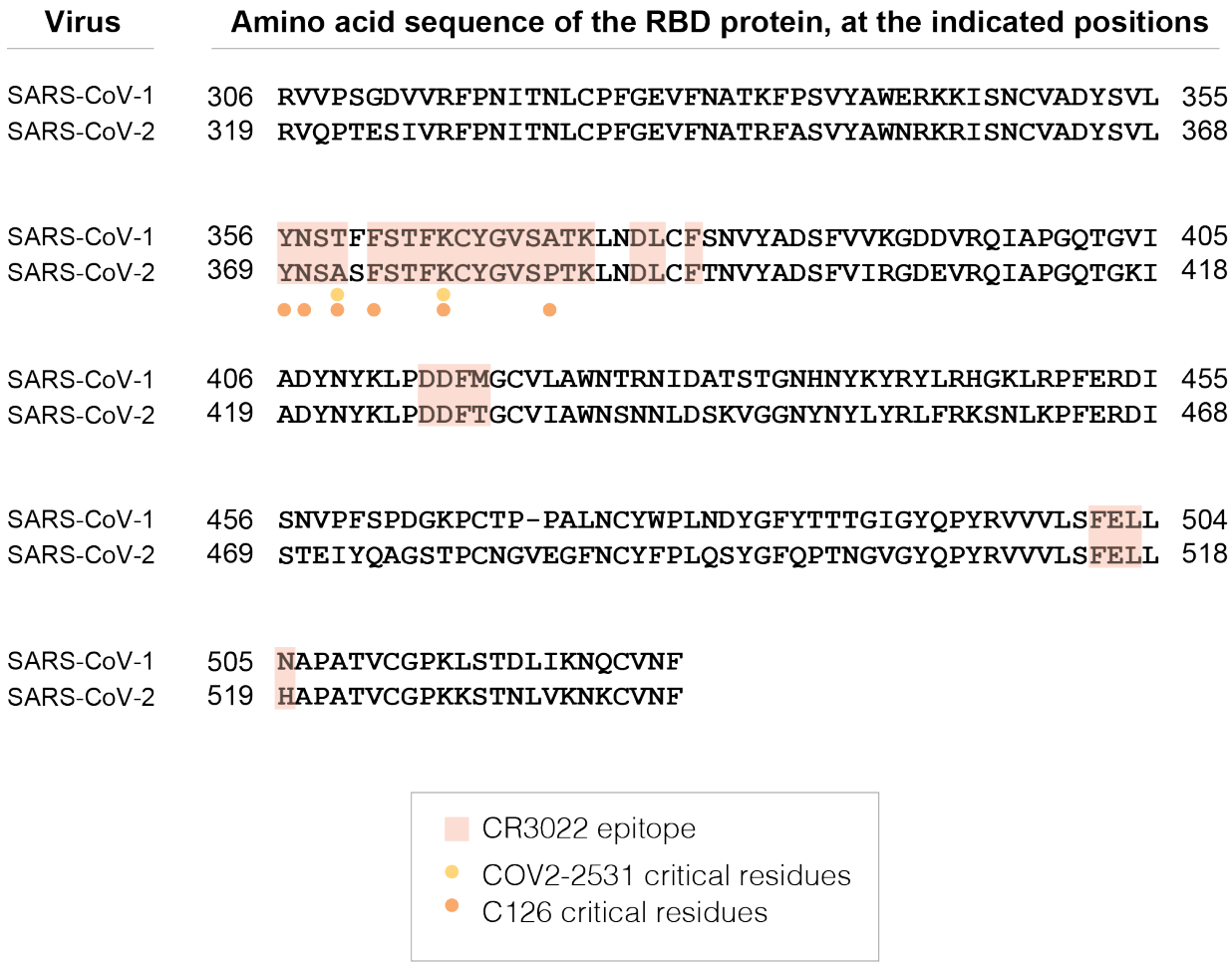


Figure VI-12: Overlay of CR3022 structure with Group 3 antibodies when bound to RBD

The structures for the RBD domains for both SARS-CoV-2 and SARS-CoV-1 were overlaid. The epitope of rCR3022 is highlighted in orange (from Yuan *et al.*). Light orange dots denote the binding residues for mAb COV2-2531, and dark orange dots denote the binding residues for C126. Figure adapted from previous study(Yuan et al., 2020b).

Screening the Group 1 and 2 antibodies against the SARS-CoV-2 alanine scanning mutation library confirmed that they bound to the S2 domain. For the Group 1 antibodies COV2-2002 and COV2-2333, we identified critical residues for both antibodies (Y917, Q920, K921) in

the heptad repeat (HR1) region of S2. These residues are conserved between SARS-CoV-2 and SARS-CoV-1. For the group 2 antibodies, screens identified two regions of residues that were specifically critical for binding. For both COV2-2164 and CnC2t1p1_B10, we identified K814 (also conserved between SARS-CoV-2 and SARS-CoV-1) as critical for binding. In addition, for both antibodies we also identified R995, and additionally for CnC2t1p1_B10, I980 and Q1002 (**Figure VI-8F**, **Figure VI-11**). K814 is not close to I980, R995, or Q1002 on the S protein structure. However, inspection of the available S protein structures (PDB: 6XR8 and 7C2L) suggested that residues I980, R995, and Q1002 are not readily accessible to antibodies in the full S protein, or even in the absence of S1. These residues make interactions that likely help maintain S2 structure, and so their mutation could indirectly affect Group 2 antibody binding. We conclude that K814 is an epitope residue for Group 1 antibodies, COV2-2002 and COV2-2333, as well as Group 2 antibodies, COV2-2164 and CnC2t1p1_B10. These results suggest that the mAbs in each public clonotype group have the essentially identical critical epitope residues.

Functional properties of germline-revertant forms of antibodies from each identified public clonotype

To determine if the function of each antibody group was due to germline-encoded reactivity or the result of somatic mutations, we investigated the equivalent germline-encoded antibodies. Heavy and light chain variable region sequences of antibodies COV2-2002, COV2-2164, and COV2-2531 were aligned with the germline sequences of [*IGHV3-7/IGHJ4*/ + *IGLV3-1/IGLJ3*], [*IGHV1-69*/IGHJ4 + *IGKV3-11/IGKJ4*], or [*IGHV4-59/IGHJ3* + *IGLV6-57/IGLJ2*], respectively. Each residue that differed from the germline gene was reverted back to the inferred germline residue (**Figure VI-13**). We then tested if the germline revertants of the antibodies in

each group shared similar functional properties with their somatically-mutated counterparts. Each

Group 1 Germline Revertant:

HC: EVQLVESGGGLVQPGGSLRLSCAASGFTFSSYWMSWVRQAPGKGLEWVANIKQDGSEKYY
VDSVKGRFTISRDNAKNSLYLQMNSLRAEDTAVYYCARVGSSSWYFDYWGQGTLVTVSS

LC: SYELTQPPSVSVSPGQTASITCSGDKLGDKYACWYQQKPGQSPVLVIYQDSKRPSGIPER
FSGSNSGNTATLTISGTQAMDEADYYCQAWDSSTGVFGGGTKLTV

Group 2 Germline Revertant:

HC: QVQLVQSGAEVKKPGSSVKVSCKASGGTFSSYAISWVRQAPGQGLEWMGGIIPIFGTANY
AQKFQGRVTITADESTSTAYMELSSLRSEDTAVYYCARTSHYDSSGSYFEYWGQGTLVTVSS

LC: EIVLTQSPATLSLSPGERATLSCRASQSVSSYLAWYQQKPGQAPRLLIYDASNRATGIPA
RFSGSGSGTDFTLTISSLEPEDFAVYYCQQRSNWPPSLTFGGGTKVEI

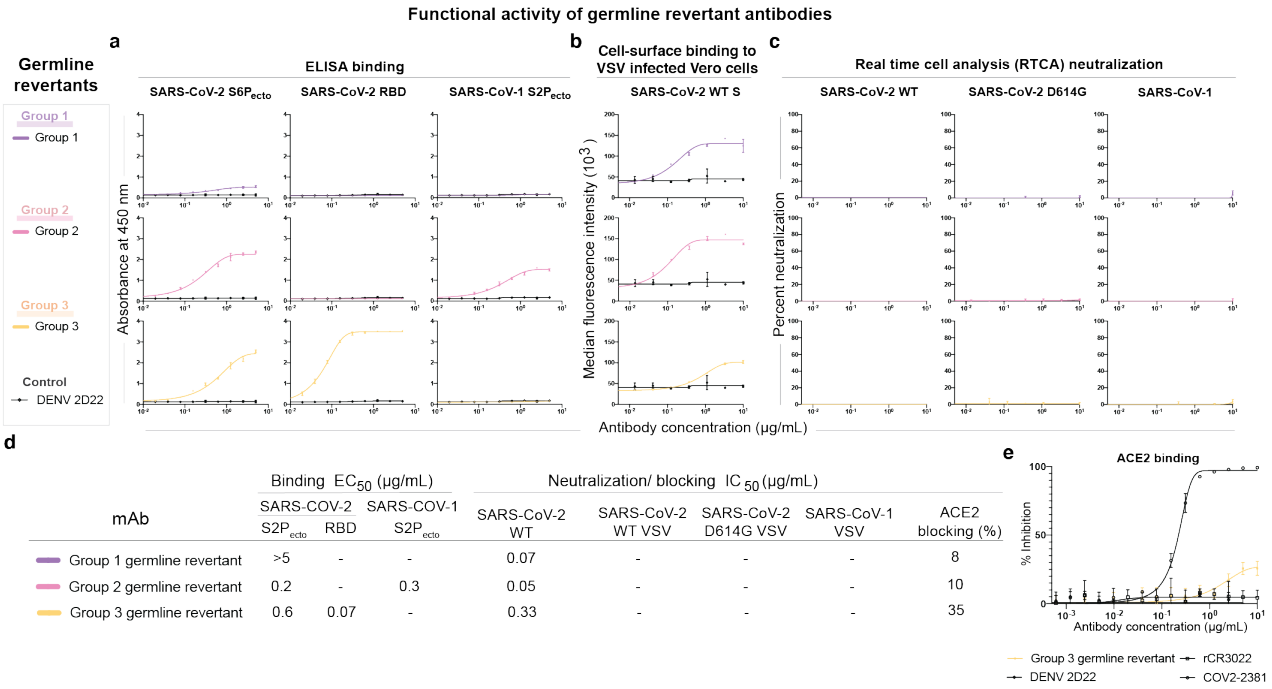
Group 3 Germline Revertant:

HC: QVQLQESGPGLVKPSETLSLTCTVSGGSISSYYWSWIRQPPGKGLEWIGYIYYSGSTNYN
PSLKSRVTISVDTSKNQFSLKLSSVTAADTAVYYCARATWLRDAFGIWGQGTMVTVSS

LC: NFMLTQPHSVSESPGKTVTISCTGSSGSIASNYVQWYQQRPGSAPTTVIYEDNQRPSGVP
DRFSGSIDSSSNSASLTISGLKTEDEADYYCQSYDSSNVVFGGGTKLTVL

Figure VI-13: Germline revertant sequences for each public clonotype group.
The corresponding heavy and light chain germline revertant sequences of each public clonotype are listed.

germline-revertant antibody was tested for binding to SARS-CoV-2 $S6P_{ecto}$, SARS-CoV-2 RBD, or SARS-CoV-1 $S2P_{ecto}$ proteins. The Group 1 germline revertant did not bind to SARS-CoV-2 $S6P_{ecto}$ or SARS-CoV-1 $S2P_{ecto}$. The Group 2 germline revertant maintained binding to both SARS-CoV-2 and SARS-CoV-1 proteins but exhibited lower binding avidity (higher EC_{50} values) than its matured counterparts COV2-2164 or CnC2t1p1_B10. The Group 3 germline revertant maintained binding to SARS-CoV-2 $S6P_{ecto}$ and RBD proteins (**Figure VI-14A, D**). Each germline revertant also bound to the surface of virus-infected cells (**Figure VI-14B, D**). While none of the germline revertants exhibited neutralizing capacity (**Figure VI-14C, D**), the Group 3 germline revertant showed a low level of ACE2 blocking (**Figure VI-14D, E**).



mAb	Binding EC_{50} (μg/mL)			Neutralization/ blocking IC_{50} (μg/mL)				
	SARS-COV-2 S2P$_{ecto}$	SARS-COV-2 RBD	SARS-COV-1 S2P$_{ecto}$	SARS-CoV-2 WT	SARS-CoV-2 WT VSV	SARS-CoV-2 D614G VSV	SARS-CoV-1 VSV	ACE2 blocking (%)
Group 1 germline revertant	>5	-	-	0.07	-	-	-	8
Group 2 germline revertant	0.2	-	0.3	0.05	-	-	-	10
Group 3 germline revertant	0.6	0.07	-	0.33	-	-	-	35

Figure VI-14: Germline-revertant antibody reactivity and functional activity

Group 1, 2, or 3 germline-revertant antibodies are shown in purple, pink, or yellow, respectively. DENV 2D22 was used as a control antibody for all assays, as shown in the lines in black. All experiments were performed in biological replicate and technical triplicate. Biological replicate from representative single experiment shown. **(A)** Binding to SARS-CoV-2 S6P$_{ecto}$, SARS-CoV-2 RBD, or SARS-CoV-1 S2P$_{ecto}$ were measured by absorbance at 450 nm, as shown in the first three columns. **(B)** Binding to Vero cells infected with VSV-SARS-CoV-2, measured by flow cytometric analysis and reported as median florescence intensity. **(C)** Results for neutralization curves for replication-competent VSV chimeric viruses in real time cell analysis (RTCA) are shown in the next three columns, measured by percent neutralization calculated by normalized cell index. **(D)** Binding EC_{50} and neutralization IC_{50} values for each of the assay curves in **Fig 5a**. All values are denoted as *μ*g/mL. ACE2 blocking was determined by measuring amount of ACE2 with FLAG tag binding in the presence of each antibody, measured by binding of an anti-FLAG antibody. Percent blocking is shown, calculated by using ACE2 binding without antibody as 0% blocking. **(E)** Inhibition binding curves for the Group 3 germline-revertant antibody. The starting antibody concentration used was 10 *μ*g/mL and was titrated three-fold serially to obtain ACE2-blocking curves.

COV2-2531 confers protection *in vivo*

MAbs can act by direct virus inactivation, but binding of some mAbs to the surface of virus-infected cells (**Figure VI-5E, H**) suggested that these antibodies also might act through Fc-mediated functions. Therefore, it was important to test some public clonotypes *in vivo*. We tested

the efficacy of these antibodies against SARS-CoV-2 *in vivo*. We used K18-hACE2 transgenic mice, which develop severe lung infection and disease after intranasal inoculation(Golden et al., 2020; Winkler et al., 2020; Zheng et al., 2021). K18-hACE2 transgenic mice received either one antibody from Group 2 (COV2-2164), one antibody from Group 3 (COV2-2531), or an isotype-control antibody (DENV 2D22) via intraperitoneal injection (200 μg, 10 mg/kg) one day prior to intranasal inoculation with 10^3 PFU of SARS-CoV-2 (WA1/2020). Mice treated with COV2-2531 were protected completely from weight loss (**Figure VI-15A**) and showed reduced viral infection in the lung, nasal wash, heart, and brain (**Figure VI-14B, C, D**) compared to the isotype-control antibody-treated group. However, mice treated with COV2-2164 were not protected from weight loss yet showed a reduction in viral load in the lung and brain (**Figure VI-14B, E**) but not in the nasal wash and heart (**Figure VI-14C, D**). Thus, antibodies that compete for binding with the SARS-CoV-1 mAb rCR3022 can be elicited after SARS-CoV-2 infection, some of which can confer protection.

Public clonotypes shared between vaccine and convalescent responses to SARS-CoV-2 S protein

We hypothesized that SARS-CoV-2 mRNA vaccines might induce public clonotypes that are shared with those seen in convalescent individuals after natural infection. We obtained peripheral blood mononuclear cells from a volunteer 10 after first vaccine dose and 7 days after second vaccine dose with the Pfizer-BioNTech vaccine. Circulating plasmablasts were enriched directly from blood by negative selection using paramagnetic beads and purified further by flow cytometric sorting (**Figure VI-16A, B**). Sorted plasmablasts were loaded on a Beacon microfluidics instrument for single-cell secreted antibody binding screening and antibody gene sequencing or in a Chromium single-cell microfluidics device (10X Genomics) followed by reverse transcription with PCR and sequence analysis to obtain paired antibody sequences. These antibody discovery

workflows were described in detail previously(Zost et al., 2020c). Enzyme-linked immunospot (ELISpot) assay analysis revealed large increase in the frequency of S-reactive cells in the enriched plasmablast cell fraction on day 7 after the second vaccination compared to that on day 10 after the first vaccine dose, confirming induction of target-specific responses in this individual. SARS-CoV-2 $S6P_{ecto}$-specific secreted antibodies were of IgG and IgA classes and accounted for >10% of total plasmablasts (**Figure VI-16C**). Further, single-cell antibody secretion analysis of a total of 4,797 purified plasmablasts loaded on a Beacon microfluidics instrument (Berkeley Lights Inc.) revealed that a large fraction of SARS-CoV-2-reactive clones (included $S6P_{ecto}$- and/or RBD-reactive clones) secreted RBD-specific IgG (**Figure VI-16D**).

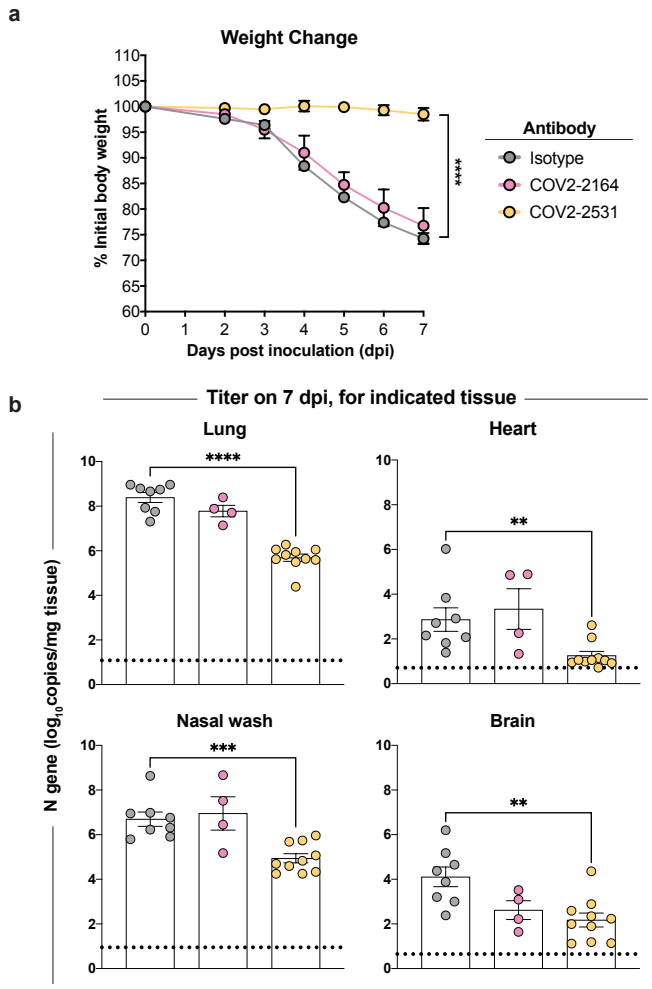


Figure VI-15: Antibody-mediated protection against SARS-CoV-2 challenge in mice

Eight-week-old male K18-hACE2 transgenic mice were inoculated by the intranasal route with 10^3 PFU of SARS-CoV-2 (WA1/2020 strain). One day prior to infection, mice were given a single 200 μg dose of COV2-2351 or COV2-2164 by intraperitoneal injection. **(A)**Weight change. Statistical analysis was performed only between isotype- and COV2-2351-treated groups. For isotype and COV2-2531 (mean ± SEM; n = 8-10, two experiments: unpaired t-test of area under the curve; **** $P < 0.0001$). For COV2-2164 (mean ± SEM; n = 8, two experiments) **(B)** Viral RNA levels at 7 days post-infection in the lung, nasal wash, heart, and brain as determined by qRT-PCR. For isotype and COV2-2531 (mean ± SEM; n = 8-10, two experiments: one-way ANOVA with Turkey's post-test: ns not significant, * $P < 0.05$, *** $P < 0.001$, **** $P < 0.0001$, comparison to the isotype control mAb-treated group). For COV2-2164 (mean ± SEM; n = 8, two experiment).

We also analyzed antibody reactivity and neutralization in serum collected on the day before vaccination (day 0), on day 10 after the first vaccine dose, on day 7 after the second vaccine dose, and on day 28 after the second vaccine dose. The reactivity of serum antibodies to both SARS-CoV-2 $S6P_{ecto}$ and SARS-CoV-2 RBD was measured by ELISA for binding (**Figure VI-16F**) and by VSV-SARS-CoV-2 neutralizing assay (**Figure VI-16G**). Binding and neutralizing activities steadily increased over time, with maximal activity detected on day 28 after the second vaccine dose.

From single-cell antibody variable gene sequencing analysis, we obtained 725 paired heavy and light chain sequences from plasmablasts following primary immunization and 8,298

paired sequences from plasmablasts following the second dose of vaccine. The same procedure was carried out on a sample collected 35 days after onset of symptoms from a convalescent individual with confirmed SARS-CoV-2 infection. This individual's serum had been determined previously to contain neutralizing antibodies(Zost et al., 2020c). Single-cell antibody secretion analysis revealed that a minor fraction (0.5%) of total plasmablasts produced S-protein-reactive antibodies. We identified 1,883 paired heavy and light chain antibody sequences for this specimen.

Antibody sequences identified in these new studies and sequences we collected from previous SARS-CoV-2 antibody discovery studies were clustered as described in **Figure VI-2**. We identified a total of 37 public clonotypes, 26 of which represented clonotypes shared between antibodies isolated from the vaccinee and individuals with exposure history to natural SARS-CoV-2 infection (**Figure VI-16H**). The antigen-binding specificity of each group was inferred through review of data in each respective publication in which the antibodies were reported. We determined that 14 of the 26 newly-identified shared clonotypes encoded antibodies specific to the SARS-CoV-2 S protein. Within that panel of mAbs, 8 of 26 clonotypes reacted with SARS-CoV-2 RBD protein, and 6 of the 26 public clonotypes cross-reacted with both SARS-CoV-1 and SARS-CoV-2 (**Figure VI-17)**. Most antibodies shared in public clonotypes were IgG, with a subset of IgAs noted. This finding shows that the Pfizer-BioNTech vaccine induces many antibodies that are genetically similar to ones elicited through natural SARS-CoV-2 infection, including multiple public clonotypes in convalescent donors encoded by commonly used V_H genes such as *IGHV3-53*, *IGHV3-66*, *IGHV1-58*, *IGHV3-30*, and *IGHV3-30-3*. Additionally, of the 37 total public clonotypes, 16 bound to RBD, and of these, 11 of 16 were neutralizing. All neutralizing public clonotypes recognized RBD. However, of the 37 public clonotypes identified, 21 are directed to antigenic sites other than the RBD, including ones described here directed to the S2 domain. It is likely that although a substantial portion of neutralizing public clonotypes is directed to the RBD,

non-RBD-targeted and non-neutralizing public clonotypes may make up an even larger portion of an individual's response to either vaccination or infection. Overall, these results suggest that many of the public clonotypes observed in previously infected individuals likely are found in vaccinated individuals.

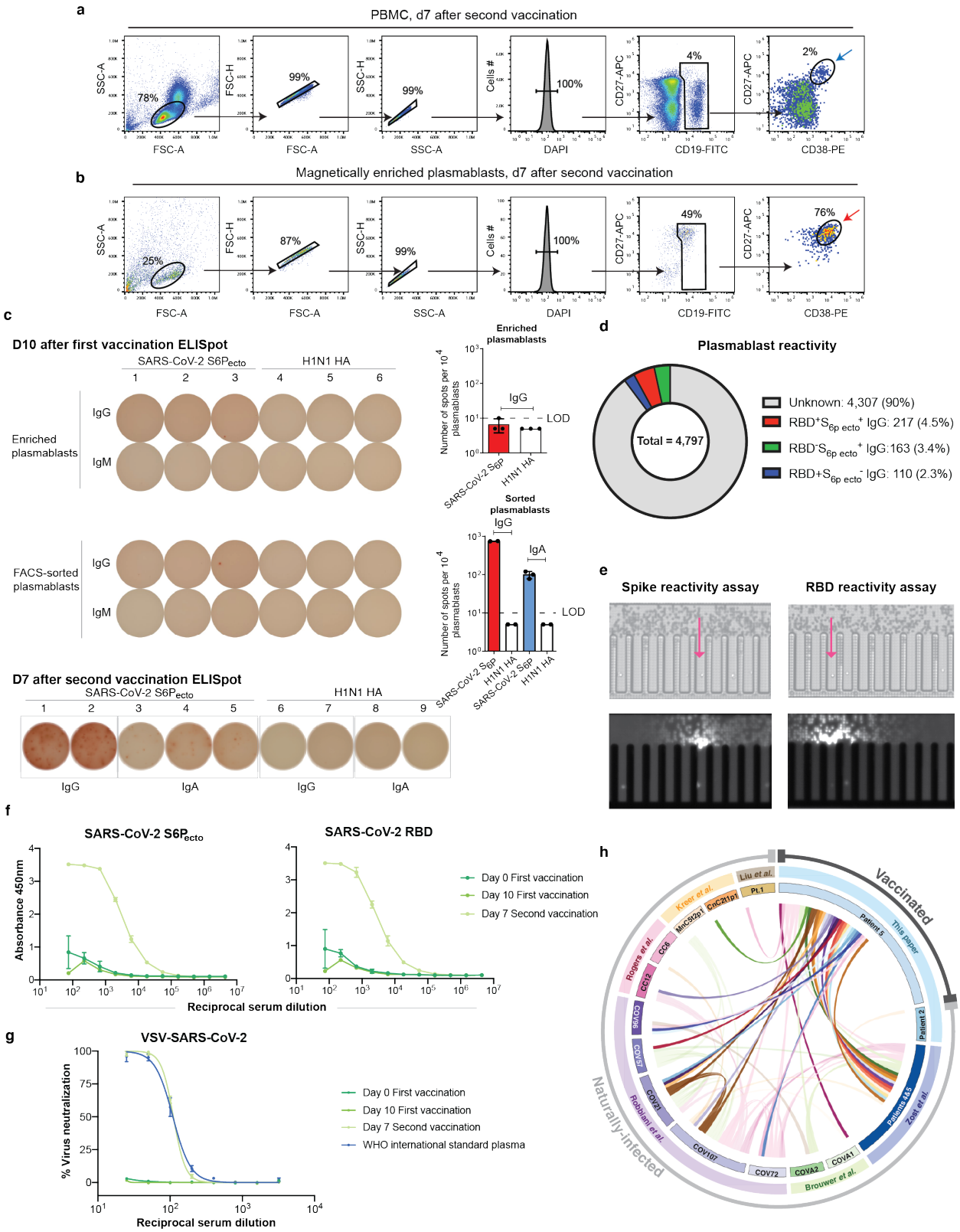


Figure VI-16: Analysis of vaccinated donor antibody response.
(Legend on next page)

Figure VI-16 legend: (A) Flow cytometric plots showing gating strategy to identify plasmablasts in total PBMC sample collected on day 7 after second vaccine dose (top panel) or identification of plasmablasts after direct enrichment from whole blood at the same time point using negative selection with paramagnetic beads (bottom panel). Blue arrow indicate enriched plasmablasts that were used for ELISpot analysis as in (b), and red arrow indicate plasmablasts ($DAPI^{-}CD19^{lo}CD27^{hi}CD38^{hi}$) that were FACS-sorted for single cell secretion and paired antibody sequencing studies. (B) ELISpot analysis of SARS-CoV-2 $S6P_{ecto}$-specific antibody secretion using enriched plasmablasts from blood collected on day 10 after the first vaccine dose (IgG), and day 7 after the second vaccine dose (IgG and IgA). A/Darwin/42/2020 H1N1 influenza virus hemagglutinin (HA) was used as a control for specificity of the plasmablast response. Wells with spots (left) and number of SARS-CoV-2 $S6P_{ecto}$-specific responses detected (right) is shown. Dotted line indicates values below limit of the detection (LOD=10 spots per 10^4 cells), that were set up to 5 spots per 10^4 cells. (C) FACS-sorted plasmablasts were loaded on a Beacon optofluidic instrument and assessed for binding to $S6P_{ecto}$ or RBD-coated beads using single cell antibody secretion analysis. Bright field images of the Beacon instrument chip with individual plasmablasts loaded into the pens of the chip are shown for the selected fields of view for each screening condition. False-color fluorescent images from the same fields of view showing binding of the detection anti-human Alexa-Fluor-568-labeled antibody to the $S6P_{ecto}$ or RBD-coated beads that captured human antibodies secreted by single plasmablasts (visualized as a plume from the beads that loaded into the channel of the chip). (D) Pie chart representation showing frequency of RBD and SARS-COV-2 $S6P_{ecto}$ reactive (red), SARS-COV-2 $S6P_{ecto}$ reactive only (green), or RBD reactive only (blue) plasmablasts identified as in c. Fraction of cells that did not react to either SARS-COV-2 RBD or $S6P_{ecto}$ is shown in grey. (E) Flow-cytometry-sorted plasmablasts were loaded on a Beacon instrument and assessed for binding to $S6P_{ecto}$ or RBD-coated beads using single-cell antibody secretion analysis. Bright field images of the Beacon instrument chip with individual plasmablasts loaded into the pens of the chip are shown for the selected fields of view for each screening condition (top). False-color fluorescent images from the same fields of view (bottom) showing binding of the detection anti-human Alexa-Fluor-568-labeled antibody to the $S6P_{ecto}$ or RBD-coated beads that captured human antibodies secreted by single plasmablasts (visualized as a plume from the beads that loaded into the channel of the chip). Arrow indicates cells secreting antigen-reactive IgG antibodies. (F) ELISA binding to SARS-CoV-2 $S6P_{ecto}$, of serum from patient 5 at days 0 of first vaccination, day 10 after first vaccination, day 7 after second vaccination, or day 28 after second vaccination were measured by absorbance at 450 nm. Serum was diluted 1:75 and then diluted serially three-fold. Experiment performed in biological replicate and technical triplicate. Biological replicate from representative single experiment shown. (G) Neutralization curves of serum from patient 5 at days 0 of first vaccination, day 10 after first vaccination, or day 7 after second vaccination. A WHO International standard for anti-SARS-CoV-2 human immunoglobulin was used as the positive control. Serum was diluted starting at a 1:25 dilution, then diluted serially two-fold. Experiment performed in technical triplicate. (H) Circos plot indicating public clonotypes identified in this paper. The more opaque ribbons within the circle represent public clonotypes that are shared between the vaccinated donor and convalescent donors after natural infection. Translucent ribbons indicate public clonotypes shared between convalescent infection individuals. The individuals from whom sequences were derived are indicated on the inner circle. The published sources from which the sequences were obtained are shown on the second circle. The outside circle indicates whether the individuals were naturally-infected or vaccinated.

	Heavy chain		Light chain					Predicted Functional Profile							
Public clonotype number	*IGHV*	*IGHJ*	*IGLV*	*IGLJ*	*IGKV*	*IGKJ*	CDRH3 length	Spike reactivity	SARS-CoV-2 RBD binding	SARS-CoV binding	Neutralization of SARS-CoV-2	Isotype of clones	Shared between vaccinated & infected	Donors sharing public clone	Publication from which reactivity data originated
1	3-7	4	3-1	3	-	-	12					IgG		Patient 3/4, Patient 2	Zost et al.
19	3-7	4	3-1	1	-	-	12					IgG		Patient 3/4, Patient 5	Zost et al.
20	3-7	4	3-1	2	-	-	12					IgG		Patient 3/4, Patient 5	Zost et al.
29	3-7	3	-	-	2-30	2	15	N/A	N/A	N/A	N/A	IgG, IgA		Patient 3/4, Patient 5	This paper
2	1-69	4	-	-	3-11	4	15					IgG		CnC2t1p1, Patient 3/4, Patient 5	Zost et al.
23	1-69	4	-	-	3-11	4	15					IgG		COVA2, Patient 5	Brouwer et al.
3	4-59	3	6-57	2	-	-	12					IgG		COV107, Patient 3/4	Zost et al.
4	3-53	4	-	-	3-20	1	11					IgG		COV21, COV107, CC12, Patient 5	Roger et al., Robbiani et al.
14	3-53	6	-	-	3-20	1	9							COVA2, COV96	Brouwer et al.
31	3-53	4	-	-	3-20	5	11	N/A	N/A	N/A	N/A	IgG		COV107, Patient 5	This paper
16	3-53	3	-	-	1-9	2	11					IgG		Candian, Patient 5	Zost et al.
13	3-53	6	-	-	1-9	3	11					IgG		Patient 5, CC12	Rogers et al.
8	3-53	6	-	-	1-9	3	11					IgG		Patient 1, Patient 5	Liu et al.
36	3-53	6	-	-	1-9	2	11	N/A	N/A	N/A	N/A	IgG		COV72, Patient 5	This paper
37	3-53	6	-	-	1-9	5	11	N/A	N/A	N/A	N/A	IgG		COV96, Patient 5	This paper
11	3-53	3	1-40	2	-	-	16					IgG		Patient 3/4, Patient 5	Zost et al.
5	3-9	4	3-21	1	-	-	14							CC6, COV107	Rogers et al.
6	3-30	4	-	-	1-5	1	20					IgG		Patient 5, COV107, COV21	Robbiani et al.
15	3-30	4	-	-	1-33	3	20					IgG		Patient 3/4, Patient 5, COV107, COV72	Zost et al.
21	3-30	3	-	-	3-20	2	14					IgG		Patient 2, Patient 3/4	Zost et al.
25	3-30	3	-	-	4-1	2	14				N/A	IgG		Patient 3/4, Patient 5	Zost et al.
22	3-30	6	1-40	1	-	-	19					IgG		Patient 3/4, Patient 5	Zost et al.
33	3-30	4	6-57	2	-	-	16	N/A	N/A	N/A	N/A	IgG		COV96, Patient 5	This paper
7	1-58	3	-	-	3-20	1	16					IgG		COV72, COV107, Patient 3/4	Zost et al.
9	1-58	3	-	-	3-20	1	16							MnC5tp2p1, COV21, COV57, COV107	Kreer et al., Robbiani et al.
10	3-66	6	-	-	1-9	5	11					IgG		Patient 3/4, COV57	Zost et al., Robbiani et al.
12	3-30-3	4	2-14	3	-	-	14							COVA1, COV107	Brouwer et al., Robbianni et al.
18	3-30-3	4	6-57	3	-	-	17					IgG		COV72, Patient 5	Robbiani et al.
34	3-30-3	4	-	-	1-39	1	13	N/A	N/A	N/A	N/A	IgG		COV21, Patient 5	This paper
17	4-31	4	-	-	2-28	4	15					IgG		COV21, Patient 5	Robbiani et al.
24	3-21	3	-	-	3-15	1	20				N/A	IgG, IgA		COVA1, Patient 5	Brouwer et al.
28	3-21	4	-	-	1-9	4	11	N/A	N/A	N/A	N/A	IgG, IgA		Patient 2, Patient 5	This paper
26	1-8	4	5-45	1	-	-	10	N/A	N/A	N/A	N/A	IgG, IgA		Patient 3/4, Patient 5	This paper
27	1-46	5	3-19	3	-	-	15	N/A	N/A	N/A	N/A	IgG, IgA		Patient 3/4, Patient 5	This paper
30	3-11	4	6-57	3	-	-	12	N/A	N/A	N/A	N/A	IgG		COV21, Patient 5	This paper
32	4-59	5	1-51	3	-	-	15	N/A	N/A	N/A	N/A	IgG		COV57, Patient 5	This paper
35	3-48	6	-	-	1-39	1	16	N/A	N/A	N/A	N/A	IgG		COV96, Patient 5	This paper

Figure VI-17: Identification of public clonotypes shared between naturally-infected individuals and a vaccinated donor

Table showing all public clonotypes identified. Gene usage for each clone or CDRH3 length are shown in columns 2 or 3. Reactivity profiles obtained from published sources are shown for comparative purposes. Blue indicates positive reactivity, while white indicates that binding reactivity or neutralization was not detected. Grey indicates reactivity profile was not found in either publication and therefore is unknown. Isotypes of antibodies in each group are listed in the eighth column. If the group contained sequences from both vaccinated and infected individuals, it was denoted in yellow. White was used for clonotypes that were shared only between convalescent individuals following natural infection.

COV2-2196 is a public clonotype

COV2-2196, also known as AZD7442 commercially, is a part of an antibody cocktail known as Evusheld targeting COVID-19. It is made up of two antibodies, AZD8895 and AZD1061. We had found that COV2-2196 is a member of a public clonotype family. We determined whether we could identify potential precursors of this public clonotype in the antibody variable gene repertoires of circulating B cells from SARS-CoV-2-naïve individuals. We searched for the V-D-J and V-J genes in previously described comprehensive repertoire datasets originating from three healthy human donors without a history of SARS-CoV-2 infection and in datasets from cord blood collected before the COVID-19 pandemic(Soto et al., 2019). A total of 386, 193, 47 or 7 heavy chain sequences for this SARS-CoV-2-reactive public clonotype was found in each donor or cord blood repertoire, respectively (**Figure VI-18**). Additionally, we found 516,738 human antibody sequences with the same light chain V-J recombination (*IGKV3–20–IGKJ1*01*). A total of 103,534, 191,039 or 222,165 light chain sequences were found for this public clonotype in each donor respectively. Due to the large number of sequences, the top five abundant sequences were aligned from each donor. Multiple sequence alignments were generated for each donor's sequences and logo plots were generated. The top five sequences with the same recombination event in each donor were identical, resulting in the same logo plots (**Figure VI-18**).

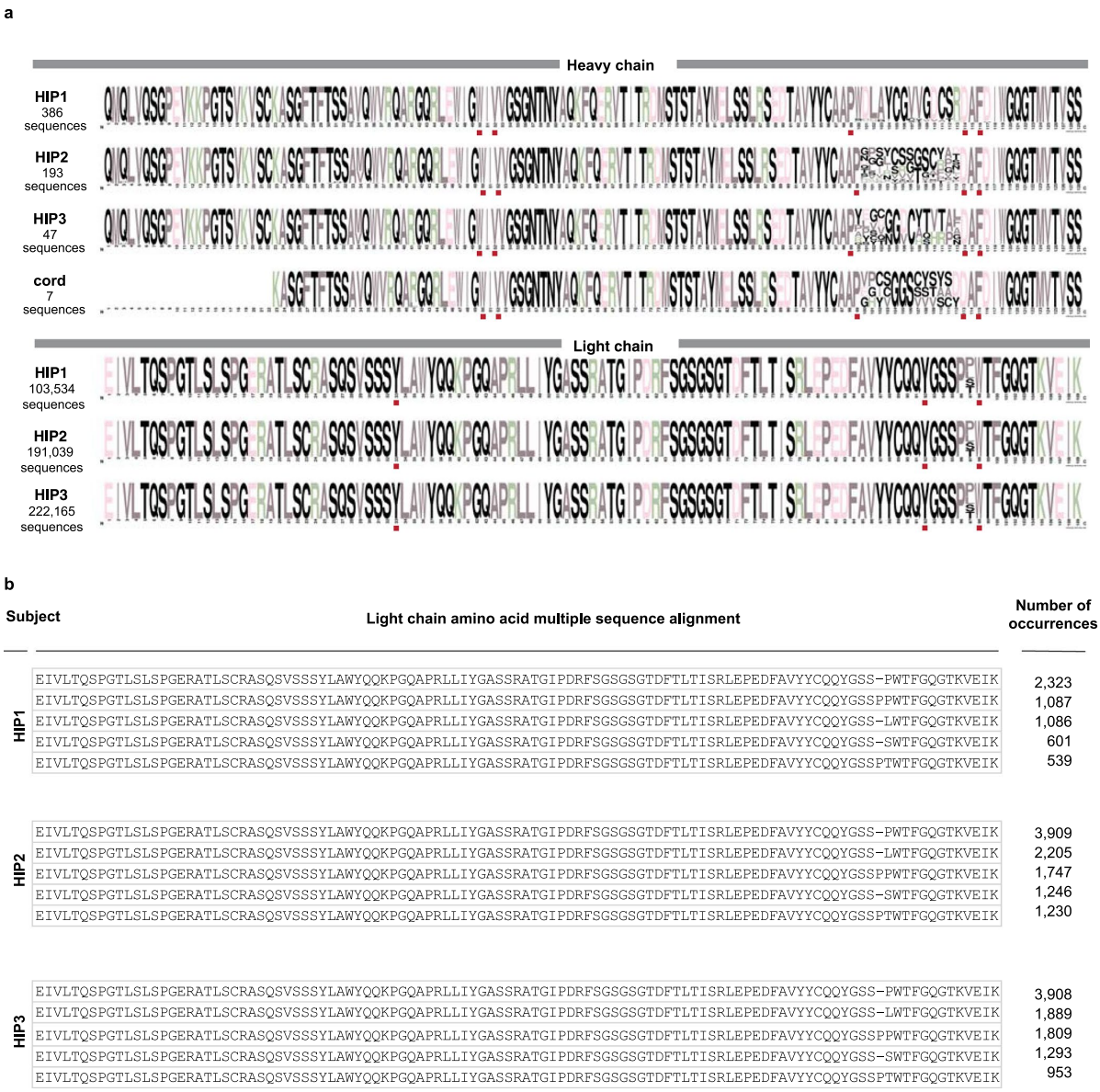


Figure VI-18: Identification of putative public clonotype members genetically similar to COV2-2196 in the antibody variable gene repertoires of virus-naïve individuals.

Antibody variable gene sequences collected from healthy individuals (HIP1, 2, or 3) prior to the pandemic with the same sequence features as AZD8895 heavy chain and light chain are aligned. **(A)** WebLogo plots of heavy chain (top) and light chain (bottom) sequences from three different adult donors and cord blood samples with the features of the public clonotype. The sequence features and contact residues used in AZD8895 are highlighted in red boxes below each multiple sequence alignment. **(B)** Since the light chain plots in **(A)** showed restricted diversity, here we show amino acid alignments for the top five representative light chains that occurred most frequently in the three adult donors studied (HIP1, 2, or 3).

We noted that eight of the nine common residues important for RBD binding in the antibody were encoded by germ line gene sequences. Interestingly, these residues were present in all 14 members of the public clonotype that we or others have described (**Figure VI-19A**). To validate the importance of these features, we expressed variant antibodies with point mutations in the HCDR3 of the paratope to determine the effect of variation at conserved residues on antibody binding to RBD (**Figure VI-19B**). We focused site-directed mutagenesis efforts on the P99 and D108 residues since these positions could be impacted at the stage of V-D-J recombination. Altering the D108 residue to A, N or E had little effect but removing the disulphide bond in the HCDR3 through cysteine to alanine substitutions greatly reduced binding. While altering the P99 residue to V or N (observed in other mature antibodies) had little effect, a P99G substitution had a dramatic effect on binding. Additionally, we made two germ line revertants (GRevs) of the AZD8895 antibody. The P99 residue is not templated by the V gene *IGHV1–58* nor by the D gene *IGHD2–2*. However, *IGHD2–2* has a likely templated G at position 99. Therefore, two GRevs were tested, one with P99 and the other with G99. Since the P99 residue orients the HCDR3 loop away from the interaction site with antigen, the G99 GRev exhibited reduced binding, whereas the P99 GRev bound antigen equivalently to wild-type (WT) AZD8895 (**Figure VI-19B**).

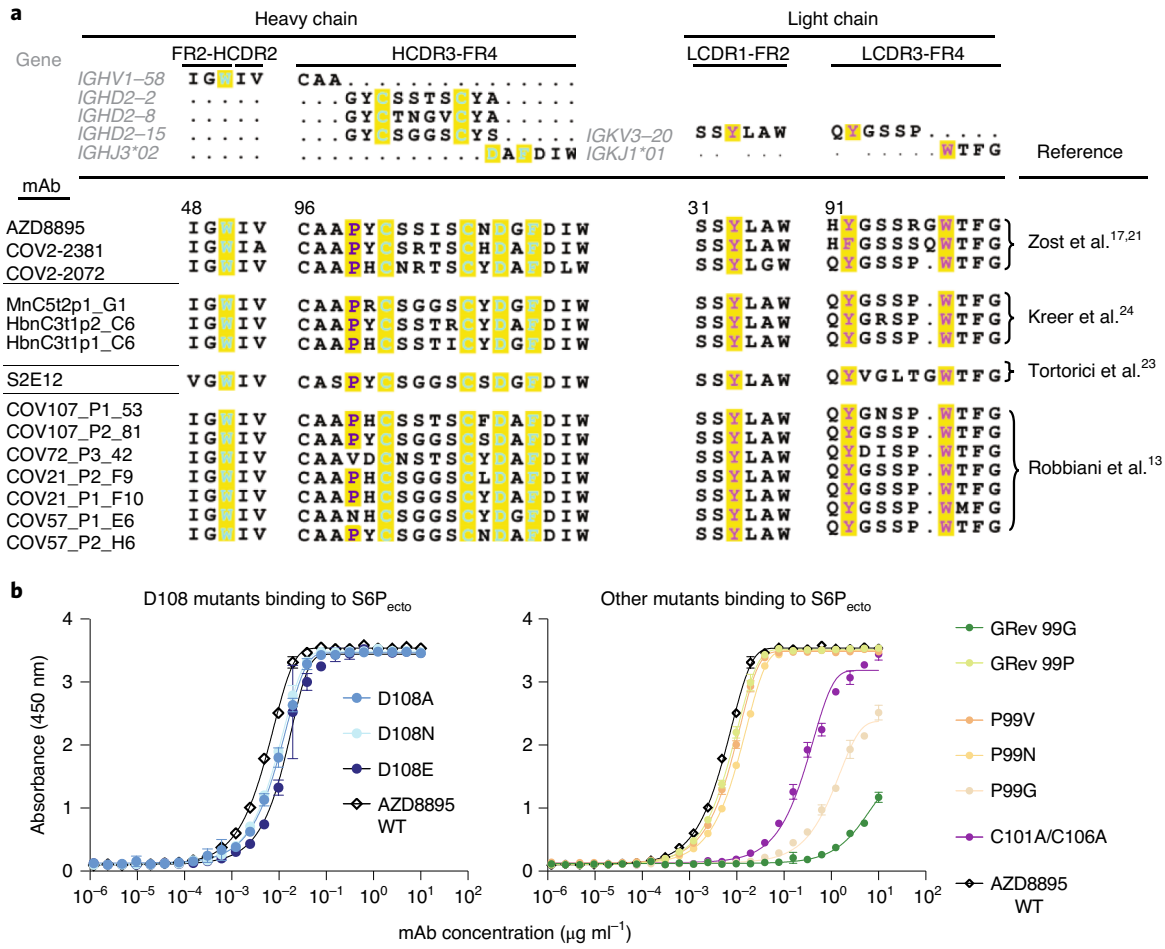


Figure VI-19: Characterization of important sequence features of the COV2-2196 public clonotype

(a) IMGT/DomainGapAlign results of the COV2-2196/AZD8895 heavy and light chains with germ line V (*IGHV1–58 and IGLV3–20*), D (*IGHD2–2, IGHD 2–8 or IGHD 2–15*) or J (*IGHJ3*02 and IGKJ1*01*) gene segments and with representative variable gene sequences of mAbs in this public clonotype. Key interacting residues and their corresponding residues in germ line genes are highlighted in yellow and coloured in blue except for P99 in purple (heavy chain) or red (light chain). (**b**) Binding curves of point mutants of COV2-2196/AZD8895. Mutants of the D108 residue are in blue, GRev of inferred somatic mutations to germ line sequence are in green, P99 mutants are in orange and the C101A/C106A mutations removing the disulphide bond in HCDR3 are in purple. The data points show the mean ± s.d. for each tested antibody dilution. experiments were performed in technical triplicate with data shown from a single experiment repeated twice. Data for the COV2-2196 WT binding curve shown in both panels are from the same experiment.

DISCUSSION

The high number of identified public B cell clonotypes in the response to SARS-CoV-2 infection or vaccination is striking, and the frequency of public clonotypes identified here is much higher than in randomly sampled B cells in a convalescent donor[8]. Many public clonotypes are shared between both infected and vaccinated individuals. Public clonotypes appear to be induced by each of the currently known antigenic sites on the S protein and are found in both the neutralizing and non-neutralizing repertoires. Some clonotypes in the shared SARS-CoV-2 response appear preconfigured in the germline state to recognize particular S epitopes, and this recognition likely is driven by particular structural features on S. The scale of available relatively large repertoire data for SARS-CoV-2 enabled us to identify many public clonotypes. The SARS-CoV-2 pandemic has resulted in comprehensive studies of antibody responses to SARS-CoV-2 S protein, with many groups identifying large panels of mAbs, including potently neutralizing ones(Brouwer et al., 2020; Cao et al., 2020; Liu et al., 2020; Robbiani et al., 2020; Rogers et al., 2020; Seydoux et al., 2020; Wec et al., 2020; Zost et al., 2020b). These discovery efforts have led to the identification of large paired heavy and light chain antibody variable gene sequence data sets for B cells specific to SARS-CoV-2, and the data has been made public at a scale unlike that for any other virus. In this study we compared the sequences of more than 14,000 paired B cell sequences encoding antibodies to S protein of SARS-CoV-2. Likely, this influx in the availability of paired antibody gene sequences from a multitude of donors contributed to our ability to identify an unexpectedly high number of paired-sequence public clonotypes. It will be interesting in future to use paired sequencing to determine if the scale of shared repertoire we observed here for SARS-CoV-2 is a more pronounced feature of the response to this particular virus than that of other viruses that have been studied for shared clones, such as HIV-1, influenza, and hepatitis C. Previous studies have identified public clonotypes in the response to these other viral pathogens,

for example a recent study of response to HIV(Setliff et al., 2018; Williams et al., 2015; Wu et al., 2011; Zhou et al., 2015) in which 27 public clonotypes were described with unpaired sequencing using only the heavy chain CDR3 sequence and V_H/J_H gene usage[26].

Several neutralizing public clonotypes have be identified previously, most commonly clonotypes encoded by the closely related heavy chain genes *IGHV3-53*, I*GHV3-66*(Tan et al., 2021; Yuan et al., 2020a), *IGHV1-2(Rapp et al., 2021), and IGHV3-30(Robbiani et al., 2020)*. Structural features of these public clonotypes likely drive the frequent selection of such clones, such as the canonical configuration of aromatic residues in the public clonotype *IGHV1-58 + IGHJ3* and *IGKV3-20 + IGKJ1* that engages the SARS-CoV-2 RBD F486 residue. Members of this public clonotype, such as COV2-2196, engage the RBD using predominantly germline-encoded residues in both the heavy and light chain*(Dong et al., 2021; Kreer et al., 2020; Nielsen et al., 2020; Robbiani et al., 2020; Tortorici et al., 2020)*. Identification of public clonotypes from multiple donors suggest these antibodies could contribute to humoral responses that mediate protection if they appear not only in memory B cells but also as antibodies from plasma cells secreted into the serum. The high prevalence of public clonotypes elicited to the SARS-CoV-2 S trimer may contribute to the high efficacy of S-encoding mRNA vaccines in large populations.

The recognition pattern of public clonotypes may predict the emergence of particular antibody escape virus variants. If diverse individuals independently make the same antibody in response to an antigen, there could be a constant and collective selective pressure on that epitope, resulting in a high potential for escape variants at that site. For example, while *IGHV3-53*- and *IGHV3-66*-encoded public clonotypes have been described in numerous individuals, neutralization of these antibodies is impacted adversely by the K417N or K417T substitutions present in the B.1.351 or P.1 SARS-CoV-2 variants of concern, respectively(Yuan et al., 2021). A similar case was described for *IGHV1-2*-encoded antibodies that target the RBD and *IGHV1-24*-encoded antibodies that target the NTD. These antibodies are found in the serum of

convalescent individuals(Voss et al., 2021), but neutralization of these antibodies is negatively affected for 501Y.V2 variant viruses(Wibmer et al., 2021). A possible explanation for the selective pressure that led to the emergence and propagation of these variants is the humoral immunity mediated by these public clonotypes.

The new Group 3 public clonotype neutralizing and protective antibodies described here bind to the cryptic face of the RBD and compete with the SARS-CoV-2 non-neutralizing mAb CR3022. Neutralizing antibodies that bind to the more conserved base of the RBD are of interest, as these sites are largely unaffected by common mutations in the variants of concern such as E484K, K417N, and N501Y(Yuan et al., 2021). Importantly, recent work has identified a B.1.1.7 variant with a deletion of RBD residues 375-377. This deletion disrupts the epitope of CR3022, yet appears to be functionally tolerated(Linda J. Rennick, 2021). As Group 3 antibodies share a similar epitope, with critical residues of COV2-2531 and C126 being K378 and A372, but with additional critical residues of Y369, N370, F374, and P384 identified for C126, this deletion might abrogate binding of antibodies from this public clonotype. Further study of public clonotypes may give insight into the most likely sites of future major antigenic changes in circulating field strains. Additionally, group 3 antibody characterization reveals additional insights on antibodies binding to epitopes similar to CR3022. Recently published literature suggests that there is an avidity threshold that affects antibody neutralization at the cryptic face of the RBD(Wu et al., 2020). Investigators showed that CR3022 binds more tightly to SARS-CoV than to SARS-CoV-2, and that although CR3022 does not neutralize SARS-CoV-2, the increased affinity to SARS-CoV-2 P384A enables neutralization with a similar potency to SARS-CoV. Therefore, differences in affinity are likely the reason why CR3022 neutralizes SARS-CoV but not SARS-CoV-2. This phenomenon also has been shown recently(Fangzhu Zhao, 2021), when investigators reengineered CR3022 to bind to the SARS-CoV-2 RBD with a >1,000-fold improved affinity, and subsequently this antibody provided prophylactic protection from viral infection in a small animal

model. This observation seems to be a general feature of antibodies that bind this cryptic face of the RBD, and therefore we hypothesize that affinity improvements via somatic hypermutation play an important role in COV2-2531's ability to neutralize, particularly since the germline revertant of group 3 does not neutralize virus. Additionally C126 is closer to germline than COV2-2531, further supporting our hypothesis that maturation through somatic hypermutation is required for antibodies that target this epitope on the spike protein.

While public clonotypes have been described that recognize the RBD,(Dong et al., 2021; Rapp et al., 2021; Robbiani et al., 2020; Tan et al., 2021; Yuan et al., 2020a) and NTD(Cerutti et al.; Suryadevara et al., 2021; Voss et al., 2021) of the S trimer, to our knowledge, those specific to the S2 domain have not been described. In this study, we identified two public clonotypes, designated here as Groups 1 and 2, which target the S2 domain of the S trimer. These mAbs do not neutralize, but they react with S proteins of both SARS-CoV-2 and SARS-CoV-1. It is likely that these S2 epitopes are the target of non-neutralizing antibodies in multiple individuals following infection or vaccination. Previous studies have identified broadly immunogenic epitopes that are conserved in the functional domains of the SARS-CoV-2 S trimer S2 domain, including cross-reactivity to endemic coronaviruses, and therefore these findings have important implications for antibody and vaccine design(Ladner et al., 2021). Rational reverse vaccinology approaches such as structure-based design of targeted antibody epitopes offer an opportunity to elicit or prevent boosting of neutralizing or non-neutralizing antibodies as desired(Rappuoli et al., 2016). The S2 region of the S trimer may be more capable of recruiting preexisting memory B cells for diverse coronaviruses, since the S2 domain is more conserved for functionally important sites such as the heptad repeat regions and fusion loop(Anderson et al., 2021). With a variety of public clonotype reactivities occurring to regions other than the RBD, it is likely that there are many additional public clonotypes that recognize the S2 domain or other regions of the S trimer. Although the S2-reactive public clonotypes described here (Groups 1 and 2) did not neutralize or

protect, future studies should investigate if additional non-RBD targeted public clonotypes can induce protection.

We propose that there are essentially four classes of public clonotypes (**Figure VI-20**): (1) neutralizing public clonotypes that bind to relatively invariant sites on S, (2) neutralizing public clonotypes that bind to sites that tolerate high sequence variability, (3) non-neutralizing public clonotypes that target relatively invariant sites, and (4) non-neutralizing public antibodies that target variable sites. The first class of antibodies is likely the most protective class in a population, as these mAbs neutralize and recognize residues unlikely to be sustained with mutations due to loss of viral fitness. An example of this class would be *IGHV1-58*-encoded antibodies as described previously(Dong et al., 2021). Many public clones currently identified for SARS-CoV-2 are categorized in the second class. While these clones initially offer protection, this property could be lost as widespread selective pressure on the virus is exerted on a region with genetic and structural plasticity. Examples of this group were discussed here, such as *IGHV3-53*- and *IGHV3-66*-encoded antibodies that target the RBD (Yuan et al., 2021). Here, we described three new public clonotypes following natural infection (Groups 1, 2, and 3) and a total of 29 new clonotypes after mRNA vaccination. Public clonotype Groups 1 and 2 fall into the third class of antibodies described here (non-neutralizing antibodies that target invariant sites), and public clonotype Group 3 antibodies falls into the second class (neutralizing public clonotypes that bind to variable sites). Future public clonotypes to SARS-CoV-2 could be binned with this four-quadrant scheme to better understand how public clonotypes contribute to humoral immunity against COVID-19.

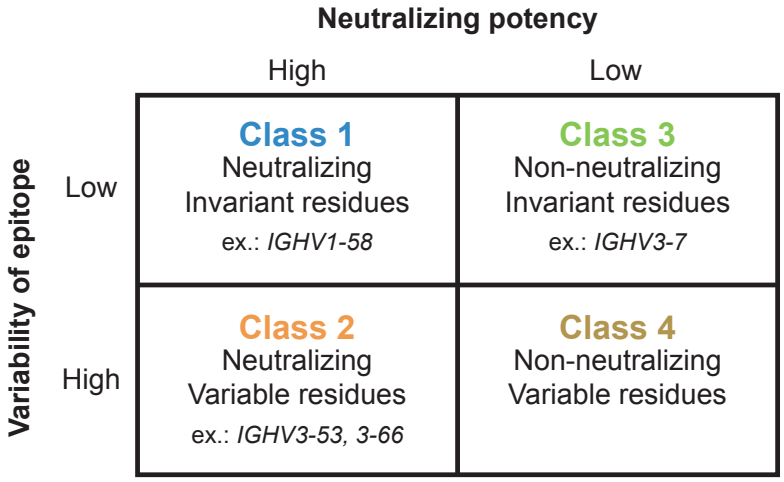


Figure VI-20: Proposed classes of public clonotypes to SARS-CoV-2

There are four proposed classes of public clonotypes to SARS-CoV-2, separated by the relationship between variability of targeted epitope (y-axis) and the neutralizing potency (x-axis) of each antibody clonotype.

Understanding the antibody response that is shared between convalescent and vaccinated individuals also will be of continued interest as the percentage of vaccinated individuals increases in the facing of emergence of new viral variants of concern. The understanding of viral epitopes that induce protective antibodies in multiple individuals has implications for predicting the most common responses to new vaccines in large populations. The emergence of SARS-CoV-2 variants with acquired mutations in epitopes for neutralizing antibodies, including antibody regimens currently authorized for EUA, is a cause for concern(Collier et al., 2021; Tegally et al., 2021; Wang et al., 2021a; Wang et al., 2021c). Our analyses of public clonotypes after natural infection and vaccination and their shared epitope targets may predict sites of future major antigenic changes in the S trimer.

METHODS

Research participants. We studied three patients (patient 2, 3 and 4) with in North America with laboratory-confirmed symptomatic SARS-CoV-2 infections that we have described previously(Zost et al., 2020c). We studied one patient (a 59-year-old male) who received Pfizer-BioNTech vaccine. The studies were approved by the Institutional Review Board of Vanderbilt University Medical Center.

Cell lines. Vero E6 (ATCC, CRL-1586) cells were maintained at 37°C in 5% CO_2 in Dulbecco's minimal essential medium (DMEM) containing 10% heat inactivated fetal bovine serum (FBS), 10 mM HEPES pH 73, 1 mM sodium pyruvate, 1× non-essential amino acids, and 100 U/mL of penicillin-streptomycin. ExpiCHO cells (Thermo Fisher Scientific, A29127) were maintained at 37°C in 8% CO_2 in ExpiCHO Expression Medium (Thermo Fisher Scientific, A2910002). Mycoplasma testing of cell lines was performed on monthly basis using a PCR-based mycoplasma detection kit (ATCC, 30-1012K), with negative results at each testing. Calu-3 (ATCC, HTB-55) cells were maintained at 37°C in 5% CO_2 in DMEM with high glucose and L-glutamine (Gibco 11965092), containing 10% heat inactivated fetal bovine serum (FBS), and 100 U/mL of penicillin-streptomycin. Vero-TMPRSS2 cells(Zang et al., 2020) were cultured at 37°C in Dulbecco's modified Eagle medium (DMEM) supplemented with 10% fetal bovine serum (FBS), 10 mM HEPES pH 7.3, 1 mM sodium pyruvate, 1× non-essential amino acids, 100 U/mL of penicillin–streptomycin, and 5 µg/mL of blasticidin.

Viruses. The generation of a replication-competent VSV expressing SARS-CoV-2 S protein with a 21 amino acid C-terminal deletion that replaces the VSV G protein (VSV-SARS-CoV-2) was described previously(Case et al., 2020b). The S protein-expressing VSV virus was propagated in MA104 cell culture monolayers (African green monkey, ATCC CRL-2378.1)(Case et al., 2020b). Viral stocks were titrated on Vero E6 cell monolayer cultures by visualizing VSV plaques using neutral red staining. VSV-SARS-CoV-2/D614G was introduced by site directed

mutagenesis. The 2019n-CoV/USA_WA1/2019 isolate of SARS-CoV-2 was obtained from the US Centers for Disease Control (CDC). Infectious stocks were propagated by inoculating Vero CCL81 cells. Supernatant was aliquoted and stored at -80^{o}C. The University of Arizona group obtained the USA-WA1/2020 isolate of SARS-CoV-2 from WRCEVA. Early passage virus stock was generated by a single passage on Vero CCL81 for 48 h. Infected cell lysate and culture supernatant was combined, subjected to one freeze-thaw, and then centrifuged to pellet cell debris. The stock was titered to ~3 x10^{6} PFU/mL by standard plaque assay on Vero CCL81 cells. Nanopore sequencing of these early passages confirmed the genome sequence was identical to the Genbank WA1/2020 sequence (MN985325.1), with no mutations in the spike furin cleavage site. All work with infectious SARS-CoV-2 was performed in Institutional Biosafety Committee-approved BSL3 or A-BSL3 facilities at Washington University School of Medicine or University of Arizona, using appropriate positive pressure air respirators and protective equipment. The WA1/2020 recombinant strains with substitutions (D614G or N501Y/D614G) or a chimeric spike gene (B.1.1.28) were obtained from an infectious cDNA clone of the 2019n-CoV/USA_WA1/2020 strain, as previously described(Plante et al., 2021). The B.1.1.7, B.1.429, and B.1.1.298 isolates were obtained from nasopharyngeal swabs, and all viruses were passaged once in Vero-TMPRSS2 cells and subjected to next-generation sequencing as previously described(Chen et al., 2021b) to confirm the introduction and stability of substitutions. Substitutions for each variant were as follows: B.1.1.7: deletion of 69 and 70 and 144 and 145, N501Y, A570D, D614G, P681H, T716I, S982A, and D1118H; B.1.1.28: L18F, T20N, P26S, D138Y, R190S, K417T, E484K, N501Y, D614G, H655Y, and T1027I; B.1.429: S13I, W152C, L452R, and D614G; B.1.1.298: deletion of 69 and 70, Y453F, D614G, I692V, and M1229I. All virus experiments were performed in an approved biosafety level 3 facility.

Clustering for identification of public clonotypes. Publicly available paired sequence sets of antibody genes were obtained(Brouwer et al., 2020; Kreer et al., 2020; Liu et al., 2020;

Rogers et al., 2020; Seydoux et al., 2020; Wec et al., 2020; Zost et al., 2020c). Together with sequences derived from this paper, public clonotypes were determined by genetic similarities of antibody sequences using the following clustering scheme. The sequences were first binned by the same heavy chain V and J genes. Following sequences then were clustered according to 70% sequence similarity on their CDRH3 nucleotide sequence. Lastly, sequences then were binned together again if they used the same light chain V and J genes. Clusters of sequences containing sequences from two or more donors were determined to be public clonotypes. This clustering allowed us to identify a total of 11 public clonotypes, three of which had not been previously described. Below are the number of sequences we derived from each publication.

Publication	**Number of sequences identified**	**Number of public clonotypes that are SARS1-reactive**
Liu *et al*., Nature 2020	19	
Kreer *et al*., Cell 2020	18	
Robbiani *et al*., Nature 2020	533	2
Rogers *et al*., Science 2020	33	1
Seydoux *et al*., Immunity 2020	32	
Brouwer *et al*., Science 2020	84	
Hansen *et al*., Science 2020	8	
Wec *et al*., Science 2020	54	
Zost *et al*., Nature Med 2020	389	4

Table VI-1: Number of sequenced identified from each publication

Heat map generation. All sequences that were identified to be public clonotypes were analyzed with PyIR(Soto et al., 2020) to identify the V and J genes. The number of sequences with corresponding V and J genes on the heavy and light chains were counted. These frequency counts then were plotted onto the heatmap using Python Seaborn Library.

Antibody production and purification. Sequences of mAbs were synthesized using a rapid high-throughput cDNA synthesis platform (Twist Bioscience) and subsequently cloned into an IgG1 monocistronic expression vector (designated as pTwist-mCis_G1) for mAb secretion

from mammalian cell culture. This vector contains an enhanced 2A sequence and GSG linker that allows simultaneous expression of mAb heavy- and light-chain genes from a single construct upon transfection(Chng et al., 2015). We performed transfections of ExpiCHO cell cultures using the Gibco ExpiCHO Expression System and protocol for 50mL mini bioreactor tubes (Corning) as described by the vendor. Culture supernatants were purified using HiTrap MabSelect SuRe (Cytiva, formerly GE Healthcare Life Sciences) on a 24-column parallel protein chromatography system (Protein Biosolutions). Purified monoclonal antibodies were buffer exchanged into PBS, concentrated using Amicon Ultra-4 50-kDa centrifugal filter units (Millipore Sigma) and stored at 4˚c until use.

Expression and purification of recombinant receptor binding domain (RBD) of SARS-CoV-2 S protein. For electron microscopy imaging of S protein in complex with Fab forms of human mAbs, we expressed a variant of $S6P_{ecto}$ protein containing a C-terminal Twin-Strep-tag, similar to that described previously(Zost et al., 2020c). Expressed protein was incubated with BioLock (IBA Lifesciences) and then isolated by Strep-tag affinity chromatography on StrepTrap HP columns (GE Healthcare), followed by size-exclusion chromatography on TSKgel G4000SWXL (TOSOH) if needed.

ELISA binding assays. Wells of 384-well microtiter plates were coated with purified recombinant SARS-CoV-2 $S6P_{ecto}$, SARS-CoV-2 RBD, or SARS-CoV $S2P_{ecto}$ at 4°C overnight. Plates were blocked with 2% non-fat dry milk and 2% normal goat serum in DPBS containing 0.05% Tween-20 for 1 h. All antibodies were diluted to a concentration of either 0.4 μg/mL for the matured antibodies or 5 μg/mL for the germline-revertant antibodies. Antibodies were diluted in two-fold dilutions until binding was no longer detected. Bound antibodies were detected using goat anti-human IgG conjugated with horseradish peroxidase and TMB substrate. The reaction was quenched with 1N hydrochloric acid once color was developed. The absorbance was measured at 450 nm using a spectrophotometer (Biotek).

Cell-surface antigen-display assay. Vero cell monolayers were monitored until 80% confluent and then inoculated with VSV-SARS-CoV-2 V (WA1/2020 strain) at an MOI of 0.5 in culture medium (DMEM with 2% FBS). For a T-225 flask, 10 mL of diluted VSV-SARS-CoV-2 virus was added to the monolayer, then incubated for 40 min. During the incubation, the flask was gently rocked back and forth every 10 min to ensure even infection. Following, the incubation the flask volume was topped off to 30 mL with 2% FBS containing DMEM and incubated for 14 h. Cells were monitored for CPE under a microscope, were trypsinized and washed in fluorescence activated cell sorting (FACS) buffer. 100,000 infected cells were seeded per well to stain with respective antibodies. All antibody was diluted to 10 μg/mL in FACS buffer, and then serially diluted 3-fold 7 times to stain for antibodies that react to cell-surface-displayed S protein. Infected cells then were resuspended in 50 μL of diluted antibody. Antibody binding was detected with anti-IgG Alexa-Fluor-647-labelled secondary antibodies. Cells were analyzed on an iQue cytometer for staining first by gating to identify infected cells as indicated by GFP-positive cells, and then gated for secondary antibody binding.

Real-time cell analysis (RTCA) neutralization assay. To determine neutralizing activity of purified antibodies or human serum, we used real-time cell analysis (RTCA) assay on an xCELLigence RTCA MP Analyzer (ACEA Biosciences Inc.) that measures virus-induced cytopathic effect (CPE)(Gilchuk et al., 2020a; Suryadevara et al., 2021; Zost et al., 2020c). Briefly, 50 µL of cell culture medium (DMEM supplemented with 2% FBS) was added to each well of a 96-well E-plate to obtain background reading. A suspension of 18,000 Vero cells in 50 µL of cell culture medium was seeded in each well, and the plate was placed on the analyzer. Measurements were taken automatically every 15 min, and the sensograms were visualized using RTCA software version 2.1.0 (ACEA Biosciences Inc). SARS-CoV-2 S VSV, SARS-CoV-2 S D614G VSV, or SARS-CoV-1 (~0.02 MOI, ~120 PFU per well) was mixed 1:1 with a respective dilution of mAb or heat-inactivated human serum in a total volume of 100 µL using DMEM

supplemented with 2% FBS as a diluent and incubated for 1 h at 37°C in 5% CO_2. At 16 h after seeding the cells, the virus-mAb mixtures were added in replicates to the cells in 96-well E-plates. Triplicate wells containing virus only (maximal CPE in the absence of mAb) and wells containing only Vero cells in medium (no-CPE wells) were included as controls. Plates were measured continuously (every 15 min) for 48 h to assess virus neutralization. Normalized cellular index (CI) values at the endpoint (48 h after incubation with the virus) were determined using the RTCA software version 2.1.0 (ACEA Biosciences Inc.). Results are expressed as percent neutralization in a presence of respective mAb relative to control wells with no CPE minus CI values from control wells with maximum CPE. RTCA IC_{50} values were determined by nonlinear regression analysis using Prism software.

Competition-binding ELISA. Wells of 384-well microtiter plates were coated with purified recombinant SARS-CoV-2 $S6P_{ecto}$ protein at 4°C overnight. Plates were blocked with 2% bovine serum albumin (BSA) in DPBS containing 0.05% Tween-20 for 1 h. Each antibody was diluted to a concentration of 10 μg/mL. Next, biotinylated antibodies were diluted to 2.5 μg/mL and added to the primary antibody solution without washing the plate to a final concentration of 0.5μg/mL. Biotinylated antibody binding was detected with horseradish peroxidase-conjugated avidin (Sigma) and developed with TMB. The reaction was quenched with 1N hydrochloric acid once color was developed. Absorbance was measured at 450 nm using a spectrophotometer.

ACE2 blocking assay. Wells of 384-well microtiter plates were coated with purified recombinant SARS-CoV-2 $S6P_{ecto}$ protein at 4°C overnight. Plates were blocked with 2% nonfat dry milk in DPBS containing 0.05% Tween-20 for 1 h. Each antibody was diluted to a concentration of 10 μg/mL. Next, recombinant human ACE2 protein with a C-terminal FLAG tag was diluted to 2 μg/mL and added to the antibody solution without washing the plate to a final concentration of ACE2 of 0.4 μg/mL. ACE2 binding was detected using HRP-conjugated anti-FLAG antibodies and developed with TMB substrate. The reaction was quenched with 1 N

hydrochloric acid once color was developed. Absorbance was measured at 450 nm using a spectrophotometer.

dsRNA staining neutralization assay. Calu-3 cells were seeded at 5,000 cells per well in SCREENSTAR 384-well black plates (Greiner) and allowed to adhere overnight. The cells then were treated with antibodies in 12 concentrations spanning from 5.65 x 10^{-5} μg/mL to 10 μg/mL and immediately transferred to a BSL-3 facility where they were inoculated with SARS-CoV-2 at an approximate MOI of 1 PFU/cell in 50 μL medium, and incubated for 48 h. At the end of the incubation, plates were submerged in PBS with 4% paraformaldehyde and 4% sucrose solution for 30 minutes to fix. Cells then were permeabilized with 0.2% Triton-X-100/PBS for 10 min and blocked with 5% BSA/PBS for 1 h. Primary J2 anti-dsRNA (Scicons #10010500) antibody solution at a 1:1,000 dilution was placed on the cells overnight at 4°C. Cells were washed with 0.1% Tween-20/PBS (PBST) three times and plates were incubated with secondary goat anti-mouse Alexa-Fluor-546-labeled antibody at 1:1,000 dilution (Thermo Fisher Scientific) for 2 h at room temperature in the dark. Plates were washed three times with PBST and incubated with DAPI for 30 min at room temperature in the dark. Plates were then imaged with fluorescent microscopy on a Nikon Eclipse TI2 automated microscopy system with a 20× objective. Six frames per well were imaged and sum dsRNA fluorescence intensity, normalized to cell count by DAPI, was measured by Nikon Elements imaging software.

Focus reduction neutralization test. Serial dilutions of mAbs (starting at 10 µg/mL dilution) were incubated with 100 FFU of different SARS-CoV-2 strains for 1 h at 37°C. Antibody–virus complexes were added to Vero-TMPRSS2 cell monolayers in 96-well plates and incubated at 37°C for 1 h. Subsequently, cells were overlaid with 1% (w/v) methylcellulose in MEM. Plates were collected 30 h later by removing overlays and fixed with 4% PFA in PBS for 20 min at room temperature. Plates were washed and sequentially incubated with an oligoclonal pool of SARS2-2, SARS2-11, SARS2-16, SARS2-31, SARS2-38, SARS2-57 and SARS2-71(Case et al., 2020a)

anti-S antibodies and HRP-conjugated goat anti-mouse IgG (Sigma, 12-349) in PBS supplemented with 0.1% saponin and 0.1% bovine serum albumin. SARS-CoV-2-infected cell foci were visualized using TrueBlue peroxidase substrate (KPL) and quantitated on an ImmunoSpot microanalyzer (Cellular Technologies).

Mouse experiments. Animal studies were carried out in accordance with the recommendations in the Guide for the Care and Use of Laboratory Animals of the National Institutes of Health. The protocols were approved by the Institutional Animal Care and Use Committee at the Washington University School of Medicine (assurance number A3381–01). Virus inoculations were performed under anesthesia that was induced and maintained with ketamine hydrochloride and xylazine, and all efforts were made to minimize animal suffering. Heterozygous K18-hACE c57BL/6J mice (strain: 2B6.Cg-Tg(K18-ACE2)2Prlmn/J) were obtained from The Jackson Laboratory. Animals were housed in groups and fed standard chow diets. One day prior to infection, mice were given a single 200 □g dose of COV2-2351 or COV2-2164 by intraperitoneal injection. Eight- to nine-week-old mice were administered 10^3 PFU of SARS-CoV-2 by intranasal administration.

Measurement of viral burden in mouse tissues. Tissues were weighed and homogenized with zirconia beads in a MagNA Lyser instrument (Roche Life Science) in 1,000 µL of DMEM medium supplemented with 2% heat-inactivated FBS. Tissue homogenates were clarified by centrifugation at 10,000 rpm for 5 min and stored at –80°C. RNA was extracted using the MagMax mirVana Total RNA isolation kit (Thermo Fisher Scientific) on the Kingfisher Flex extraction robot (Thermo Fisher Scientific). RNA was reverse transcribed and amplified using the TaqMan RNA-to-CT 1-Step Kit (Thermo Fisher). Reverse transcription was carried out at 48°C for 15 min followed by 2 min at 95°C. Amplification was accomplished over 50 cycles as follows: 95°C for 15 s and 60°C for 1 min. The number of copies of SARS-CoV-2 N gene RNA in samples was determined using a previously published assay(Case et al., 2020a). Briefly, a TaqMan assay

was designed to target a highly conserved region of the N gene (forward primer: ATGCTGCAATCGTGCTACAA; Reverse primer:m GACTGCCGCCTCTGCTC; Probe: /56-FAM/TCAAGGAAC/ZEN/AACATTGCCAA/3IABkFQ/). This region was included in an RNA standard to allow for copy number determination down to 10 copies per reaction. The reaction mixture contained final concentrations of primers or probe of 500 or 100 nM, respectively.

Electron microscopy sample and grid preparation, imaging and processing of $S6P_{ecto}$–Fab complexes. Fabs were produced by digesting recombinant chromatography-purified IgGs using resin-immobilized cysteine protease enzyme (FabALACTICA, Genovis). The digestion occurred in 100 mM sodium phosphate and 150 mM NaCl pH 7.2 (PBS) for around 16 h at ambient temperature. To remove cleaved Fc from intact IgG, the digestion mix was incubated with CaptureSelect Fc resin (Genovis) for 30 min at ambient temperature in PBS buffer.

For screening and imaging of negatively-stained SARS-CoV-2 $S6P_{ecto}$ protein in complex with human Fabs, the proteins were incubated at a Fab:S molar ratio of 4:1 for about 1 h at ambient temperature or overnight at 4°C. Approximately 3 µL of the sample at concentrations of about 10 to 15 µg/mL was applied to a glow-discharged grid with continuous carbon film on 400 square mesh copper electron microscopy grids (Electron Microscopy Sciences). The grids were stained with 0.75% uranyl formate(Ohi et al., 2004). Images were recorded on a Gatan US4000 4k × 4k CCD camera using an FEI TF20 (TFS) transmission electron microscope operated at 200 keV and control with Serial EM. All images were taken at 50,000× magnification with a pixel size of 2.18 Å per pixel in low-dose mode at a defocus of 1.5 to 1.8 µm. The total dose for the micrographs was around 30 e– per $Å^2$. Image processing was performed using the cryoSPARC software package. Images were imported, CTF-estimated, and particles were picked. The particles were extracted with a box size of 256 pixels and binned to 128 pixels (pixel size of 4.36 Å/pix) and 2D class averages were performed.

Epitope mapping of antibodies by alanine scanning. Epitope mapping was performed essentially as described previously (Davidson and Doranz, 2014) using SARS-CoV-2 (Wuhan-Hu-1 strain) S protein RBD and S2 shotgun mutagenesis mutation libraries, made using a full-length expression construct for S protein. 184 residues of the RBD (between S residues 335 and 526), and 513 S2 residues (between residues 689 -1247) were mutated individually to alanine, and alanine residues to serine. Mutations were confirmed by DNA sequencing, and clones arrayed in a 384-well plate, one mutant per well. Binding of mAbs to each mutant clone in the alanine scanning library was determined, in duplicate, by high-throughput flow cytometry. A plasmid encoding cDNA for each S protein mutant was transfected into HEK-293T cells and allowed to express for 22 h. Cells were fixed in 4% (v/v) paraformaldehyde (Electron Microscopy Sciences), and permeabilized with 0.1% (w/v) saponin (Sigma-Aldrich) in PBS plus calcium and magnesium (PBS++) before incubation with mAbs diluted in PBS++, 10% normal goat serum (Sigma), and 0.1% saponin. MAb screening concentrations were determined using an independent immunofluorescence titration curve against cells expressing wild-type S protein to ensure that signals were within the linear range of detection. Antibodies were detected using 3.75 µg/mL of Alexa-Fluor-488-labeled secondary antibodies (Jackson ImmunoResearch Laboratories) in 10% normal goat serum with 0.1% saponin. Cells were washed three times with PBS++/0.1% saponin followed by two washes in PBS, and mean cellular fluorescence was detected using a high-throughput Intellicyte iQue flow cytometer (Sartorius). Antibody reactivity against each mutant S protein clone was calculated relative to wild-type S protein reactivity by subtracting the signal from mock-transfected controls and normalizing to the signal from wild-type S-transfected controls. Mutations within clones were identified as critical to the mAb epitope if they did not support reactivity of the test MAb but supported reactivity of other SARS-CoV-2 antibodies. This counter-screen strategy facilitates the exclusion of S protein mutants that are locally misfolded or have an expression defect.

Cell-surface binding to full- length S protein, variant S proteins, or S2 domain protein. A plasmid encoding the S protein C-terminus S2 region (starting at residue S685) was transfected into HEK-293T cells arrayed in a 384-well plate and allowed to express for 22 h. Cells transfected with vector alone acted as negative controls. MAbs were screened over a range of concentrations, 4 replicates for each mAb concentration, as described for epitope mapping. Fluorescence values were background subtracted.

ELISA binding assay for serum analysis. To assess serum reactivity, 384-well microtiter plates were coated with purified recombinant SARS-CoV-2 $S6P_{ecto}$ at 4˚c overnight. Plates were blocked with blocking buffer (2% non-fat dry milk and 2% normal goat serum in DPBS containing 0.05% Tween-20) for 1 h. Serum was diluted 1:75 in blocking buffer, and then diluted three-fold serially 15 times, and added to wells. Binding was detected with goat anti-human IgG conjugated with horseradish peroxidase and TMB substrate. The reaction was quenched with 1N hydrochloric acid once color was developed. The absorbance was measured at 450 nm using a spectrophotometer (Biotek).

Plasmablasts isolation and flow cytometric analysis. Blood was collected into tubes containing heparin. To assess plasmablasts frequency in PBMCs for analytical flow cytometric studies, PBMCs were enriched from whole blood (day 10 after first, and day 7 after second vaccination) using direct PBMCs isolation kit (StemCell Technologies). For singe-cell antibody secretion and paired antibody sequencing studies, plasmablasts were enriched from the whole blood (day 7 after second vaccination) by negative selection using custom direct human plasmablasts isolation kit containing paramagnetic beads and antibodies for negative selection (StemCell Technologies). Enriched cells were stained 30 min on ice in a RoboSep buffer (StemCell Technologies) containing following phenotyping antibodies; anti-CD19-FITC (1:20 dilution, eBioscience), anti-CD27-APC (1:20 dilution), and anti-CD38-PE (1:25 dilution, BD Biosciences), and then analyzed by flow cytometry using an SH800 cell sorter (Sony). A DAPI

stain was used as a viability dye to exclude dead cells. Plasmablasts were identified as DAPI-$CD19^{lo}CD27^{hi}CD38^{hi}$ cells. Approximately 40,000 and ~6,000 plasmablasts were FACS-sorted in a bulk for paired antibody sequencing and single-cell antibody secretion studies, respectively.

Generation of antibody variable-gene libraries from single plasmablasts. For paired antibody sequencing, cells were resuspended into DPBS containing 0.04% non-acetylated BSA, split into four replicates, and separately added to 50 µL of RT Reagent Mix, 5.9 µL of Poly-dt RT Primer, 2.4 µL of Additive A and 10 µL of RT Enzyme Mix B to complete the Reaction Mix as per the vendor's protocol. The reactions then were loaded onto a Chromium chip (10x Genomics). Chromium Single Cell V(D)J B-Cell-enriched libraries were generated, quantified, normalized and sequenced according to the User Guide for Chromium Single Cell V(D)J Reagents kits (CG000086_REV C). Amplicons were sequenced on an Illumina Novaseq 6000, and data were processed using the CellRanger software v3.1.0 (10X Genomics).

Single-cell antibody secretion analysis using Beacon instrument. FACS-purified plasmablasts were resuspended in plasmablast survival medium that promotes antibody secretion and assessed for reactivity of secreted antibodies using the 11k chip on Beacon optofluidic instrument (Berkley Lights) as previously described(Zost et al., 2020c). Single cell-antibody secretion binding assay was performed as previously described(Zost et al., 2020c) using SARS-CoV-2 $S6P_{ecto}$- and SARS-CoV-2 RBD-coated beads.

ELISpot assay. Direct enzyme-linked immunosorbent spot (ELISpot) assay was performed to enumerate plasmablasts present in the PBMC samples secreting IgG, IgM, or IgA antibodies reacting with either SARS-CoV-2-$S6P_{ecto}$ protein or influenza A/Darwin/42/2020 H1N1 hemagglutinin protein (as a negative control). Briefly, 96-well ELISpot MSIP plates (Millipore) were activated with 100 μL 100% methanol/well for 10 sec, washed three times with 1× DPBS, coated overnight either with 100 μL of 2 μg/mL of SARS-CoV-2-$S6P_{ecto}$ or influenza HA protein in PBS overnight at 4°C.

Plates were washed three times with 1× DPBS and blocked by incubation with RPMI containing 10% FCS at 37°C for 2 h. Enriched plasmablasts or FACS-sorted plasmablasts were added to the plates and incubated 18-24 h at 37°C. Plates were washed with PBS and then PBS containing 0.05% Tween, and then incubated with either goat anti-human IgG-HRP conjugated antibodies (Southern Biotech), goat anti-human IgA-HRP conjugated antibodies (Southern Biotech), or goat anti-human IgM-HRP conjugated antibodies (Southern Biotech) for 2 h at room temperature. After washing three times with PBS containing 0.05% Tween/1% BSA, plates were developed using 3-amino-9-ethyl-carbazole (AEC) substrate (Sigma). The developed plates were scanned, and spots were analyzed using an automated ELISpot counter (Cellular Technologies Ltd.). Plasmablasts or sorted plasmablasts from PBMCs were added to the plates and incubated 18-24 h at 37°C. Plates were washed with PBS and then PBS containing 0.05% Tween, and then incubated with either goat anti-human IgG-HRP conjugated antibodies (Southern Biotech, catalog no. 2040-05), goat anti-human IgA-HRP conjugated antibodies (Southern Biotech, catalog no. 2050-05), or goat anti-human IgM-HRP conjugated antibodies (Southern Biotech, catalog no. 2020-05) for 2 h at room temperature. After washing three times with PBS containing 0.05% Tween/1% BSA, plates were developed using 3-amino-9-ethyl-carbazole (AEC) substrate (Sigma). The developed plates were scanned and spots were analyzed using an automated ELISpot counter (Cellular Technologies Ltd.).

Quantification and statistical analysis. The descriptive statistics mean ± SEM or mean ± SD were determined for continuous variables as noted. Virus titers in the tissues were compared using one-way ANOVA with Turkey's post-test. Curves for antibody binding and neutralization were fitted after log transformation of antibody concentrations using non-linear regression analysis. Technical and biological replicates are indicated in the figure legends. Statistical analyses were performed using Prism v8.4.3 (GraphPad).

CHAPTER VII

SUMMARY AND FUTURE DIRECTIONS

Thesis Summary

The foundation of antibody repertoire analysis was set in large sequence datasets obtained from bulk sequencing. From this, groups including our own have estimated the diversity of the circulating antibody repertoire, identified public clonotypes to different infectious disease targets, investigated the prevalence and dynamics of different clonal lineages, and much more. However, these studies often lacked the ability to validate the antibody clone's functionality and, therefore, by extension, investigate the clone's contribution to humoral immunity. On the contrary, many panels of antibodies have been characterized to a variety of pathogenic targets, revealing reactivity, antiviral activity, essential binding residues, or cross-reactive epitopes of antibodies. This body of work has progressed the field of antibody repertoire studies by merging the two worlds of (1) utilizing large sequencing datasets analyzed with bioinformatics methods to investigate repertoire sequence characteristics (2) functional antibody characterization on a large scale. This theme of utilizing the best of both worlds is carried on throughout all chapters of this dissertation.

Work described in chapter II of this dissertation details using sequence analysis to investigate antibody repertoires specific to AL amyloidosis patients. Working through this project helped the familiarization of bioinformatic processing of immunoglobulin sequence analysis and build a toolbox of scripts to use when investigating gene usages and sequence features. Through this work, we were able to identify the dysplastic clone within each patient's sequencing data. Additionally, we revealed diversity in the genetic features of the clones and

repertoires in AL amyloidosis patients. I also had the opportunity to investigate the genetic features of SARS-CoV-2 specific antibodies, revealing that many of the antibodies isolated early in the pandemic had relatively high levels of similarity to germline genes, contrasting to the much higher levels of somatic hypermutation seen in B-cell recall responses against other pathogens(Klein et al., 2013; West et al., 2014; Wrammert et al., 2011). This indicates that antibodies might have the ability to quickly react to a SARS-CoV-2 infection as it doesn't take multiple rounds of SHM to achieve strong recognition of the antigen as well as neutralization of the virus which was later identified by multiple groups (Seydoux et al., 2020). Additionally, I worked on identifying TCR specificities without being able to personally functionally validate them on the bench by identifying TCR sequences in our dataset within the Genbank repository. There was a prevalence of TCR sequences within our dataset that matched to previously described TCRs in Genbank, allowing for assignment of specificities to TCR sequences that would have otherwise remained unknown. A platform of such can have further implications for curating sequence databases with specificity assignments which could be useful as machine learning training datasets as well as legal implications to check if certain sequences have been previously patented as well as for the use of data mining within patented sequences(Krawczyk et al., 2021; Krawczyk et al., 2019).

At the beginning stages of my graduate work, I had worked on optimizing and building a clustering pipeline along with Dr. Cinque Soto, which was applied to numerous projects within this dissertation. Alongside, I was trained by Rachel Nargi and Dr. Pavlo Gilchuk on the microscale expression/ purification workflow and high-throughput antibody characterization. Merging the two workflows, I executed large-scale studies investigating the antibody repertoire to several disease indications described in chapters III, IV, V, and VI. Doing so allows for (1) systematic selection of antibodies to synthesize and characterize by identifying one antibody per

clonal family and (2) mining for and understanding the contributions of public clonotypes on humoral immunity.

Work described in chapter III of this dissertation exhibits the scenarios specific to influenza-specific repertoires in which I had utilized the clustering pipeline we had built and genetic analysis along with high-throughput antibody characterization to assist in projects. Different classes of antibody lineages, public clonotypes, as well as therapeutic antibody candidates, have been identified through this work, all of which are currently ongoing projects in the Crowe lab.

Chapter IV of this dissertation details the private antibody repertoire to EBOV GP. This work is the first of its kind, with a large scale of B cells isolated within a singular donor to a protein, allowing us to estimate the genetic and functional diversity of B cells specific to an antigen. Having a large amount of single B-cell sequencing within a single donor serves as an optimal database for mining antibody lineages as well as identification of sequences of antibodies present in sera. Chapter V of this dissertation is an extension of chapter IV. In chapter V, the large set of paired single-cell sequences is used to mine for public clonotypes, revealing an unprecedented prevalence of public clonotypes. The prevalence of neutralizing and protective public clonotypes identified to EBOV was previously untapped and surprising, thereby providing insight on population immunity to EBOV.

Within chapters IV and V, we identified two V_H genes that encode cross-reactive antibodies at the germline level within both private and public antibody repertoires: *IGHV1-69* and *IGHV1-02*. For *IGHV1-69*, a structural explanation is now possible. *IGHV1-69/IGHJ6*-encoded antibodies have been described to target the mucin-like domain (MLD) cradle, exploiting hydrophobic residues encoded by the germline of these gene segments by binding and destabilizing the GP quaternary structure and therefore blocking cleavage required for receptor binding (Murin et al., 2021). Within the antibodies described, potency was acquired through

somatic hypermutation. However, antibodies in this class had cross-reactive neutralizing proprieties regardless since they target the conserved MLD anchor and cradle. The molecular and structural determinants of the cross-reactive activities associated with *IGHV1-02*-encoded antibodies have yet to be described. Here, the data reveal that the cross-reactive neutralizing properties of antibodies encoded by *IGHV1-02* are germline-encoded. *IGHV1-02* also has been shown to encode broadly neutralizing antibodies for HIV due to hydrophobic residues in the CDR2 similar to those encoded by the CDR2 of some alleles of the germline gene *IGHV1-69* (Lee et al., 2021). As all the *IGHV1-02*-encoded antibodies discovered here competed for binding with the glycan cap mAb 13C6, we predict that *IGHV1-02* encodes for cross-reactive neutralizing antibodies targeting the glycan cap or MLD region of the GP using a similar mechanism to that of *IGHV1-69*-encoded antibodies. Antibodies encoded by *IGHV1-69* and *IGHV1-02* may represent a substantial portion of the first line of defense during ebolavirus infection. We speculate that the earliest neutralizing response to ebolavirus infection is likely encoded *IGHV1-69* and *IGHV1-02* since all antibodies discovered here for which the UCA antibodies neutralized virus competed for binding with the glycan cap mAb 13C6.

It has been reported that B cells circulating early in convalescence target the glycan cap region of the GP (Williamson et al., 2019). We found that 9.2% of EBOV GP-specific antibodies in this large repertoire were encoded by *IGHV1-69*, causing it to be the most heavily used variable gene, with the *IGHV4-34* or *IGHV1-02* genes also used frequently, at 7.5% or 6.3%, respectively. This finding further demonstrates a substantial reliance on germline-encoded antibody responses in the humoral immune response to EBOV. The intrinsic hydrophobic properties of antibodies encoded by these genes likely play a vital role in immunity to ebolaviruses and other viruses. It is highly desirable to identify germline-encoded pan-ebolavirus cross-reactive antibodies to support rational vaccine design and testing efforts(Chen et al., 2019; Rappuoli et al., 2016).

Lastly, chapter VI of this dissertation investigates the public antibody response to SARS-CoV-2. Throughout the duration of the SARS-CoV-2 pandemic, there have been more antibody sequencing studies published to the SARS-CoV-2 S protein than to any other target worldwide. Therefore, this served as a pristine opportunity to mine for and understand public clonotypes. Three novel public clonotypes are described in this chapter, including the first public clonotype described to the S2 domain of the S protein. Understanding public clonotypes to SARS-CoV_2 revealed parallels relationships between public clonotypes and viral evolution, therefore providing a method to predict the emergence of viral variants, one of the keys to ending the pandemic.

Caveats

There are several caveats to the studies described throughout this thesis document. The first is the small sample size for each study. Throughout all chapters, there is a small sample size for each study. With the AL amyloidosis sequencing study, seven patients were sequenced. However, within the seven patients, there was only organ involvement noted for three patients, all of which had cardiac involvement indicating these patients were already far progressed in disease prognosis. Additionally, the number of cells sequenced from each patient varied greatly. This study could have been strengthened with an increase in patient samples with varying stages of disease as well as regulating the number of cells sequenced.

Although we had generated a large sequence dataset of memory B cells specific to EBOV, this was only done within a singular donor. The study could be strengthened if the same scale of sequencing was performed on multiple EBOV convalescent donors, creating a larger sample set to estimate the diversity of memory B cells to a specific antigen. Studies of such may be possible when the cost of paired single-cell sequencing decreases in the future. Additionally, comparisons on gene usages between an EBOV-specific repertoire and a non-EBOV-specific

repertoire would have been strengthened if the non-EBOV-specific memory B cells were sorted within the same donor.

Additionally, having large sequence datasets from multiple donors is useful for having a more accurate assessment of the frequency of public clonotypes shared between individuals. Current studies described in this dissertation use deep sequence datasets to mine for public clonotypes within published antibodies, which often range from tens to a couple of hundred antibodies per donor. The 100 to 1,000-fold difference in the number of sequences between donors is important for assessing the frequency in which public clonotypes may be shared between donors. Although we have identified a high frequency of public clonotypes compared to previous studies, this work may be important for considerations of public clonotypes in vaccine design.

The clustering pipeline used throughout this dissertation has served us very well on multiple platforms and in many projects. However, one of the major caveats is when new sequences are pulled into the database, the entire dataset has to be re-clustered and curated to identify if new public clonotypes arose. This was especially pertinent as through the pandemic, new panels of antibodies are published consistently. Therefore integration of continually clustering new antibody sequences to already existing clusters would be useful for future work on public clonotypes. Additionally, the clustering pipeline currently utilizes a single distance threshold. This is not optimal for sequences with shorter CDR3s, as shorter sequences would be penalized more heavily than longer sequences when clustering utilizes a fixed threshold. All studies described in chapters III, IV, V, and VI would benefit from a clustering algorithm that utilizes a flexible threshold that takes sequence length into account.

Future Directions

Antibody sequencing for use in diagnostics for AL Amyloidosis

The work in chapter II illustrating a study on identifying the dysplastic clone within AL amyloidosis patients may be used as a proof of concept for diagnostic techniques. Future work includes repeating the study on a larger dataset on samples from patients with varying disease burdens and organ involvement. Additionally, understanding the sensitivity of using mRNA sequencing on detecting immunoglobulins for diagnostic purposes may be useful in tracking a patient's disease load for AL amyloidosis and similar diseases such as multiple myeloma.

Antibody repertoires and implications for vaccine design and therapeutics discovery

Large-scale studies on understanding the genetic and functional diversity of antibody repertoires in response to different antigens have implications for benchmarking efficacious vaccines. Creating a "yellow pages" of the antibody repertoire categorizing every clonal family to different functional profiles indexes the antibody response to pathogens. Doing so builds a toolbox to survey elicited responses to different vaccine designs. The studies described in chapter IV illustrate the computed genetic diversity of the antibody repertoire to the EBOV GP at roughly 20,000 unique clonal families. Although a large number, it was much smaller than expected, indicating that indexing the entire antibody repertoire to EBOV might be feasible. As the studies described in chapter IV detail the clonally expanded repertoire, future work can detail the non-clonally expanded repertoire as well to get a complete understanding of the human antibody response to the EBOV GP.

Identification of therapeutic antibodies is a needle in a haystack scenario. Thereby, characterizing and screening large panels of antibodies increases the likelihood of capturing a therapeutically interesting antibody. As indicated in the work in chapter III, screening large panels of antibodies from different clonal families increases the probability of identifying cross-

reactive and protective antibodies. Within the filovirus field, there has yet to be a cross-reactive and neutralizing antibody identified to EBOV, BDBV, SUDV, and MARV. Through characterizing the clonally expanded antibody repertoire to the EBOV GP, there was a higher frequency of cross-reactive antibodies than expected in comparison to previous work. Therefore as an extension of further understanding the antibody repertoire to the EBOV GP, future work on replicating a similar analysis and characterization on the antibody repertoire to MARV GP may increase the chances for identifying a therapeutic pan-filovirus antibody.

Several germline-encoded cross-reactivity to EBOV, BDBV, and SUDV was described through the work described in chapters IV and V. However, several cross-reactive antibodies did not possess cross-reactivity at the germline level and instead acquired cross-reactivity through a series of mutations. Future studies focused on understanding how these cross-reactive antibodies evolved by validating intermediate ancestors of these cross-reactive antibodies can be informative for rational vaccine design efforts.

Future directions for the work described in chapter III include mapping out and investigating the functional and genetic diversity of the antibody repertoire in response to H3N2, H1N1, and Influenza B. With the revelation of several interesting clonal lineages, characterization of those antibodies through structural studies, *in vivo* work, and epitope mapping will be critical for therapeutic considerations.

Takeaways

Public clonotypes to antigenically variable epitopes

As the SARS-CoV-2 pandemic progressed, SARS-CoV-2 antibodies were being discovered at a scale previously unseen. Through the duration of the pandemic, there have been more published paired antibody sequences to SARS-CoV-2 than to any other pathogens, providing a perfect platform to better understand public clonotypes and how it contributes to

population immunity. From all the public clonotypes identified to SARS-CoV-2, we had proposed that there are four classes of public clonotypes: (1) neutralizing public clonotypes that bind to relatively invariant sites on S, (2) neutralizing public clonotypes that bind to sites that tolerate high sequence variability, (3) non- neutralizing public clonotypes that target relatively invariant sites, and (4) non-neutralizing public antibodies that target variable sites(Chen et al., 2021a).

Early on in the pandemic, many groups described potent antibodies isolated from a variety of individuals targeting the receptor-binding domain (RBD) of the SARS-CoV-2 spike. When investigating the sequence features and characteristics, it was found that many of these antibodies were public clonotypes using variable genes *IGHV3-53/3-66*(Yuan et al., 2020a), *IGHV 1-2*, and *IGHV 3-30*(Robbiani et al., 2020). As the pandemic progressed, so came the rise of escape variants. Interestingly, the neutralization of these public clonotypes encoded by *IGHV3-53/33-66* was impacted adversely by substitutions present in beta and gamma strains of SARS-CoV-2. A similar case was described to public clonotypes encoded by *IGHV1-24* targeting the N-terminal domain (NTD) of the SARS-CoV-2 spike, which are negatively affected in alpha, beta, and gamma strains(Voss et al., 2021). A public clonotype described in chapter VI, group 3, encoded by *IGHV4-59/IGLV3-01* binds to the side of the RBD, but deletions on the spike can be functionally tolerated at that site, which had also been found in a transmitted variant(Linda J. Rennick, 2021). All these above public clonotypes bins into “group 2”, indicating that public clonotypes that bind to sites of high sequence variability may be useful in predicting or mirroring escape variants of viruses. This phenomenon has only been observed in the SARS-CoV-2 space but could likely apply to other viruses that tolerate high sequence variability in their antigenic targets, such as common cold coronaviruses, influenza virus (Joyce et al., 2016; Pappas et al., 2014; Zost et al., 2021b), hepatitis C virus (Bailey et al., 2017a), and HIV (Setliff et al., 2018; Zhou et al., 2015) where public clonotypes have also been identified.

An example of class 1 public clonotype is one encoded by *IGHV1-58/IGHJ3* and *IGKV3-20/IGKJ1*(Dong et al., 2021). This public clonotype engages the RBD at a relatively invariant site using predominantly germline-encoded residues in both the heavy and light chains. Therefore, this antibody has retained its ability to neutralize multiple variants, including omicron. Identification of public clonotypes of such is important as there is a higher likelihood of multiple individuals making similar antibodies in response to antigen exposure. Therefore, such antibodies may be targeted when designing antigens for rational vaccine design.

Public clonotypes to highly conserved epitopes

Public clonotypes have also been identified for a variety of other viral pathogens that do not display as much antigenic variation, including respiratory syncytial virus (RSV) (Mukhamedova et al., 2021), and Ebola (EBOV) virus. As there is not as much of a concern for antigenic drift with these viruses, public clonotypes with neutralizing properties targeting these antigens can be utilized for rational vaccine design. And in the case of EBOV, public clonotypes with neutralizing properties can be leveraged for designing a cross-reactive vaccine covering multiple strains of Ebolavirus.

One of the explanations for the high efficacy of the EBOV monospecific vaccine may be due to the high prevalence of public clonotypes elicited by the GP. Antibodies encoded by *IGHV3-15/IGLV1-40* occur in both vaccinated and convalescent donors targeting the receptor-binding domain (RBD) of the GP (Cohen-Dvashi et al., 2020; Rijal et al., 2019), and the cross-reactivity for EBOV and BDBV occurs at the germline gene level. Convergent antibody responses encoded by *IGHV3-13* also have been described as paired with diverse light chains, suggesting the specificity of binding of these antibodies to the RBD is mediated principally by the heavy chain. These antibodies were found to share a common mutation in the CDR1, hinting at convergent down-selection of clones during the process of somatic hypermutation(Cagigi et al., 2018; Davis

et al., 2019). *IGHV3-23*-encoded public antibodies that target the GP1/2 interface also have been described(Bornholdt et al., 2016; Davis et al., 2019; Wec et al., 2017). Additionally, we recently described a high frequency of public clonotypes with neutralizing properties to EBOV, including several with cross-reactive neutralizing properties. Interestingly, public clonotypes encoded by *IGHV1-2* and *IGHV1-69* retained their cross-reactive neutralizing activity even when reverted back to the germline sequence. Therefore, this indicates that some of these public clonotypes may be useful for rational vaccine design of a pan-ebolavirus vaccine.

Non-neutralizing public clonotypes

Many antibody discovery work is primarily focused on neutralizing antibody responses as it is the most straightforward implication for protection. However, a large portion of the antibody repertoire is often neglected – the non-neutralizing antibodies. This is in part due to the lack of functional assays available to survey these antibody's ability for protection *invitro*. However, several non-neutralizing antibodies have exhibited protective properties by Fc effector functions. The same is reflected in public clonotypes. In chapter III, we described a public clonotype antibody to the trimer-interface of influenza A hemagglutinin (HA) head domain(Zost et al., 2021b). These antibodies split the HA trimer apart at an epitope that is conserved across influenza A HAs and do not show activity in neutralization assays but have the ability to protect animals *in vivo*. As these antibodies have been identified in multiple individuals, it serves as a prime example of how non-neutralizing public clonotypes may also contribute to population immunity to viruses. Therefore, further study into understanding the non-neutralizing public antibody response will also be incredibly influential in understanding the evolutionary processes behind how population immunity impacts viral evolution.

Gene usages of public clonotypes and germline-encoded properties

If one were to accumulate a list of all public clonotypes described to every pathogen and their corresponding gene usages, several genes pop up at the top of the list as commonly used. One of the most commonly used gene usages in public clonotypes is *IGHV1-69*, of which antibodies have been described to influenza(Pappas et al., 2014), SARS-CoV-2, Ebola, Hepatitis-C, and HIV(Setliff et al., 2018). *IGHV1-69* encodes two hydrophobic residues at the tip of the CDRH2 loop, enabling it to recognize conserved hydrophobic regions of viral envelope glycoproteins, therefore setting the basis for many broadly neutralizing antibodies. With frequent occurrences of *IGHV1-69* antibodies against a variety of viral pathogens, it does not come as a surprise that many public clonotypes also utilize this variable gene. Multiple public clonotype antibodies encoded by *IGHV1-02* have also been described to Ebola, SARS-CoV-2, and HIV(Setliff et al., 2018). Antibodies encoded by *IGHV1-02* have been shown to possess motifs in the variable region that enhance binding to viral glycoproteins. Germline revertant versions of both *IGHV1-69* and *IGHV1-02* antibodies have retained their ability to bind antigens and neutralize viruses, further demonstrating germline-encoded neutralizing properties. *IGHV1-69* and *IGHV1-02* encoded antibodies are just two examples of which the germline sequence encodes for antiviral neutralizing properties(Lee et al., 2021). There are many more genes to be further discovered that may encode for antiviral activity and are found in multiple public clonotypes as well.

Through chapters IV, V, and VI of this dissertation, we identify germline-encoded reactivity of several gene combinations by characterizing UCA or germline revertant versions of multiple antibodies specific to EBOV-GP as well as SARS-CoV-2 spike through measuring binding and neutralization. Multiple UCA or germline revertant antibodies retained not only the ability to bind

antigen but also neutralize virus. Additionally, several have maintained cross-reactive binding and neutralization.

From reversion of EBOV-specific antibodies, the majority of *IGHV1-69* and *IGHV1-2* encoded antibodies maintained not only binding but neutralization to EBOV, BDBV, and SUDV. *IGHV3-11/IGKV1-39*, *IGHV4-39/IGKV3-15*, and *IGHV3-13/IGKV3-20* antibodies maintained binding and neutralization to EBOV. From reversion of SARS-CoV-2 specific antibodies, *IGHV1-69/IGKV3-11* encodes for cross-reactivity against both SARS-CoV and SARS-CoV-2 S proteins. Additionally, *IGHV4-59/IGLV3-01* encodes for reactivity to the SARS-CoV-2 RBD and minor ACE2 blocking abilities. However, neither encoded for neutralizing properties to SARS-CoV-2. Together, this suggests that there are germline-encoded antibodies that are preconfigured for avid recognition, therefore likely contributing to the early humoral response to pathogens.

Understanding pairs of genes that encode for binding or neutralization activity prior to the introduction of mutations by somatic hypermutation reveals that these antibodies are more straightforward to elicit during antigen exposure as at the naïve B cell state, they are “pre-configured” to recognize antigens. Furthermore, one of the challenges to reverse vaccinology is that naïve B cells that recognize the correct epitope might be rare and buried deep in the antibody repertoire(Havenar-Daughton et al., 2018). Therefore public clonotypes encoded by genes with germline-encoded properties are of high value as there is not only a higher probability of eliciting these antibodies that have germline-encoded properties than ones that require the accumulation of multiple mutations contributing to the antibody’s configuration for binding and neutralization, but it is also more likely to be elicited in multiple individuals. Therefore, such antibodies are important to take into consideration for rational vaccine design efforts.

Public clonotypes and antibody specificity predictions

Wang and colleagues have identified a series of sequence features within public clonotypes directed to different domains to the SARS-CoV-2 S protein and used the dataset on public clonotypes to train a deep learning model to identify SARS-CoV-2 S protein versus influenza HA directed antibodies(Wang et al., 2021b). Identifying sequence features includes sequence motifs as well as gene usages, demonstrating that the possibility of predicting antibody specificity solely based on primary sequence is viable. Continual identification and characterization of public clonotypes will allow for the building of databases storing sequences along with extensive characterization of antibodies. Databases like such will allow us to address some fundamental questions about humoral immunity, which can have large impacts on many antibody discovery efforts/pipelines in the future.

Additional Methods

All cartoons in this dissertation are drawn by Elaine Chang Chen in Sketchbook (2021) on an iPad and assembled in either Adobe Illustrator (2022) or Microsoft PowerPoint(2022).

LIST OF PUBLICATIONS

Articles related to Dissertation:

1. **Chen EC**, Rubinstein S, Soto C, Bombardi RG, Day SB, et al. (2020) Diverse patterns of antibody variable gene repertoire disruption in patients with amyloid light chain (AL) amyloidosis. *PLOS ONE* 15(7): e0235713. https://doi.org/10.1371/journal.pone.0235713
2. Zost, S.J., Gilchuk, P., Chen, R.E. Case, J.B., Riedy J.X., Trivette, A., Nargi, R.S., Sutton, R.E., Suryadevara N., **Chen E.C.**, Binshtein E., Shrihari S., Chu, H.Y., Didier, J.E., MacRenaris, K.W., Jones, T., Day, S., Myers, L., Lee., F.E., Nguyen, D.C., Sanz, I., Martinez, D.R., Rothlauf, P.W., Bloyet, L.M., Whelan, S.P.J., Baric, R.S., Thackray, L.B., Diamond, M.S., Carnahan, R.H., and Crowe, J.E. Rapid isolation and profiling of a diverse panel of human monoclonal antibodies targeting the SARS-CoV-2 spike protein. *Nat Med* (2020). https://doi.org/10.1038/s41591-020-0998-x
3. Soto, C.S., Bombardi, R.G., Kozhevnikov, M., Sinkovits, R.S., **Chen E.C.**, Branchizio, A., Kose, N., Day, S.B., Pikinton, M., Gujral, M., Mallal, S., and Crowe, J.E.. High Frequency of Shared Clonotypes in Human T Cell Receptor Repertoires. *Cell Reports* (2020). https://doi.org/10.1016/j.celrep.2020.107882
4. Zost S.J., Dong J., Gilchuk I.M., Gilchuk P., Thornburg N.J., Bangaru S., Kose N., Finn J.A., Bombardi R., Soto C., **Chen E.C.**, Nargi R.S., Sutton R.E., Irving R.P., Suryadevara N., Westover J.B., Carnahan R.H., Turner H.L., Li S., Ward A.B., Crowe J.E. Canonical features of human antibodies recognizing the influenza heamgltinin trimer interface. *The Journal of Clinical Investigation* (2021) https://doi.org/10.1172/jci146791

5. Gilchuk P., Guthals A., Bonissone S., Ilinykh P.A., Huang K., Soto C., Bombardi R., Bryan A., Davidson E., **Chen E.C.**, Dornaz B.J., Bukreyev A., Zeitlin L., Castellana N., Crowe J.E. Molecular analysis of serum antibody repertoire from a human survivor identified prevalence of potent antibodies against base region of Ebola virus glycoprotein. *Frontiers in Immunology (2021)* https://doi.org/10.3389/fimmu.2021.706757
6. **Chen E.C.,** Gilchuk P., Zost S.J., Suryadevara N., Winkler E.S., Cabel C.R., Binshtein E., Sutton R.E., Rodriguez J., Day S., Myers L., Trivette A., Williams J.K., Davidson E., Li S., Doranz B.J., Campos S.K., Carnahan R.H., Thorne C.A., Diamond M.S., and Crowe J.E. Convergent antibody responses to the SARS-CoV-2 spike protein in convalescent and vaccinated individuals. *Cell Reports (2021)* https://doi.org/10.1016/j.celrep.2021.109604
7. Dong J., Zost S.J., Greaney A.J., Starr T.N., Dingens A.S., **Chen E.C.**, Chen R.E., Case J.B., Sutton R.E., Gilchuk P., Rodriguez J., Armstrong E., Gainza C., Nargi R.S., Binshtein E., Xie X., Zhang X., Shi P.Y., Logue J., Weston St., McGrath M., Frieman M., Brady T., Tuffy K., Bright H., Loo Y.M., McTamney P., Carnahan R.H., Diamond M.S., Bloom J.D., Crowe J.E. Genetic and structural basis for recognition of SARS-CoV-2 spike protein by a two antibody cocktail. *Nature Microbiology (2021)* https://doi.org/10.1038/s41564-021-00972-2
8. **Chen E.C.,** GilchukP., Zost S.J., Ilinykh P.A., Binshtein E., Huang K., Myers L., Bonissone S.R., Day S., Kona C.R., Trivette A., Reidy J.X., Sutton R.E., Gainza C., Diaz S.M., Williams J.K., Selverian C., Davidson E., Saphire E.O., Doranz B.J., Castellana N., Burkreyev A., Carnahan R.H., Crowe J.E., Systematic analysis of human antibody response to ebolavirus glycoprotein reveals a high prevalence of neutralizing public clonotypes. *Under review.*

Additional co-Authored Articles

9. Zost, S.J., Gilchuk P., Case, J.B., Binshtein E., Chen R.E., Reidy, J.X., Trivette, A., Nargi, R.S., Sutton, R.E., Suryadevara, N., Williamson, L.E**., Chen, E.C.**, Jones, T., Day, S., Myers, L., Hassan, A.O., Kafai, N.M., Winkler, E.S., Fox, J.M., Steinhardt, J.J., Ren, K., Loo, Y., Kallewaard, N.L., Martinez, D.R., Schafer, A., Gralinski L.E., Baric, R.S., Thackray L.B., Diamond, M.S., Carnahan, R.H., and Crowe, J.E. Potently neutralizing human antibodies that block SARS-CoV-2 receptor binding nad protect animals. *Nature* (2020). https://doi.org/10.1038/s41586-020-2548-6
10. Suryadevara N., Shrihari S., Gilchuk P., VanBlargan L.A., Binshtein E., Zost S.J., Nargi R.S., Sutton R.E., Winkler E.S., **Chen E.C.**, Fouch M.E., Davidson E., Doranz B.J., Carnahan R.H., Thackray L.B., Diamond M.S., and Crowe, Jr. J.E. Neutralizing and protective human monoclonal antibodies recognizing the N-terminal domain of the SARS-CoV-2 spike protein. *Cell* (2021) https://doi.org/10.1101/2021.01.19.427324
11. Rosenfeld A.M., Meng W., Horne K.I., **Chen E.C.**, Bagnara D., Prak E.T.L. Bulk gDNA sequencing of Antibody Heavy Chain Gene Rearrangements for Detection and Analysis of B cell Clone Distribution; a method by the AIRR Community. *Methods in Molecular Biology* (2021)
12. Gupta N., Marquez S., Soto C., **Chen E.C.**, Bostick M.L., Stervbo U., Farmer A. Bulk sequencing from mRNA with UMI for Evaluation of B Cell Isotype and Clonal Evolution; a method by the AIRR community. *Methods in Molecular Biology* (2021)

REFERENCES

PyIR: A scalable wrapper for processing billions of immunoglobulin and T cell receptor sequences using IgBLAST.

Abed, Y., Hardy, I., Li, Y., and Boivin, G. (2002). Divergent evolution of hemagglutinin and neuraminidase genes in recent influenza A:H3N2 viruses isolated in Canada. J Med Virol *67*, 589-595.

Anderson, E.M., Goodwin, E.C., Verma, A., Arevalo, C.P., Bolton, M.J., Weirick, M.E., Gouma, S., McAllister, C.M., Christensen, S.R., Weaver, J.*, et al.* (2021). Seasonal human coronavirus antibodies are boosted upon SARS-CoV-2 infection but not associated with protection. Cell.

Bahadoran, A., Lee, S.H., Wang, S.M., Manikam, R., Rajarajeswaran, J., Raju, C.S., and Sekaran, S.D. (2016). Immune Responses to Influenza Virus and Its Correlation to Age and Inherited Factors. Front Microbiol *7*, 1841.

Bahlis, N.J., and Lazarus, H.M. (2006). Multiple myeloma-associated AL amyloidosis: is a distinctive therapeutic approach warranted? Bone Marrow Transplant *38*, 7-15.

Bailey, J.R., Flyak, A.I., Cohen, V.J., Li, H., Wasilewski, L.N., Snider, A.E., Wang, S., Learn, G.H., Kose, N., Loerinc, L.*, et al.* (2017a). Broadly neutralizing antibodies with few somatic mutations and hepatitis C virus clearance. JCI Insight *2*, e92872.

Bailey, J.R., Flyak, A.I., Cohen, V.J., Li, H., Wasilewski, L.N., Snider, A.E., Wang, S., Learn, G.H., Kose, N., Loerinc, L.*, et al.* (2017b). Broadly neutralizing antibodies with few somatic mutations and hepatitis C virus clearance. JCI Insight *2*.

Bangaru, S., Lang, S., Schotsaert, M., Vanderven, H.A., Zhu, X., Kose, N., Bombardi, R., Finn, J.A., Kent, S.J., Gilchuk, P.*, et al.* (2019). A Site of Vulnerability on the Influenza Virus Hemagglutinin Head Domain Trimer Interface. Cell *177*, 1136-1152 e1118.

Barnes, C.O., Jette, C.A., Abernathy, M.E., Dam, K.A., Esswein, S.R., Gristick, H.B., Malyutin, A.G., Sharaf, N.G., Huey-Tubman, K.E., Lee, Y.E.*, et al.* (2020). SARS-CoV-2 neutralizing antibody structures inform therapeutic strategies. Nature *588*, 682-687.

Bashford-Rogers, R.J., Palser, A.L., Huntly, B.J., Rance, R., Vassiliou, G.S., Follows, G.A., and Kellam, P. (2013). Network properties derived from deep sequencing of human B-cell receptor repertoires delineate B-cell populations. Genome Res *23*, 1874-1884.

Beerli, R.R., Bauer, M., Schmitz, N., Buser, R.B., Gwerder, M., Muntwiler, S., Renner, W.A., Saudan, P., and Bachmann, M.F. (2009). Prophylactic and therapeutic activity of fully human monoclonal antibodies directed against influenza A M2 protein. Virol J *6*, 224.

Beniac, D.R., and Booth, T.F. (2017). Structure of the Ebola virus glycoprotein spike within the virion envelope at 11 A resolution. Sci Rep *7*, 46374.

Blancas-Mejia, L.M., and Ramirez-Alvarado, M. (2013). Systemic amyloidoses. Annu Rev Biochem *82*, 745-774.

Bonissone, S.R.L., T.; Harris, K.; Davison, L; Avanzino, B.; Trinklein, N.; Castellana, N.; Patel, A. (2021). Serum proteomics expands on high-affinity antibodies in immunized rabbits than deep B-cell repertoire sequencing alone. BioRxiv, 833871.

Bonsignori, M., Scott, E., Wiehe, K., Easterhoff, D., Alam, S.M., Hwang, K.K., Cooper, M., Xia, S.M., Zhang, R., Montefiori, D.C.*, et al.* (2018). Inference of the HIV-1 VRC01 Antibody Lineage Unmutated Common Ancestor Reveals Alternative Pathways to Overcome a Key Glycan Barrier. Immunity *49*, 1162-1174 e1168.

Bornholdt, Z.A., Herbert, A.S., Mire, C.E., He, S., Cross, R.W., Wec, A.Z., Abelson, D.M., Geisbert, J.B., James, R.M., Rahim, M.N.*, et al.* (2019). A two-antibody pan-ebolavirus cocktail confers broad therapeutic protection in ferrets and nonhuman primates. Cell Host Microbe *25*, 49-58 e45.

Bornholdt, Z.A., Turner, H.L., Murin, C.D., Li, W., Sok, D., Souders, C.A., Piper, A.E., Goff, A., Shamblin, J.D., Wollen, S.E.*, et al.* (2016). Isolation of potent neutralizing antibodies from a survivor of the 2014 Ebola virus outbreak. Science *351*, 1078-1083.

Bosch, B.J., van der Zee, R., de Haan, C.A., and Rottier, P.J. (2003). The coronavirus spike protein is a class I virus fusion protein: structural and functional characterization of the fusion core complex. J Virol *77*, 8801-8811.

Bray, M., Davis, K., Geisbert, T., Schmaljohn, C., and Huggins, J. (1998). A mouse model for evaluation of prophylaxis and therapy of Ebola hemorrhagic fever. J Infect Dis *178*, 651-661.

Briney, B., Inderbitzin, A., Joyce, C., and Burton, D.R. (2019). Commonality despite exceptional diversity in the baseline human antibody repertoire. Nature *566*, 393-397.

Briney, B.S., Willis, J.R., Finn, J.A., McKinney, B.A., and Crowe, J.E., Jr. (2014). Tissue-specific expressed antibody variable gene repertoires. PLoS One *9*, e100839.

Brochet, X., Lefranc, M.P., and Giudicelli, V. (2008). IMGT/V-QUEST: the highly customized and integrated system for IG and TR standardized V-J and V-D-J sequence analysis. Nucleic Acids Res *36*, W503-508.

Brouwer, P.J.M., Caniels, T.G., van der Straten, K., Snitselaar, J.L., Aldon, Y., Bangaru, S., Torres, J.L., Okba, N.M.A., Claireaux, M., Kerster, G.*, et al.* (2020). Potent neutralizing antibodies from COVID-19 patients define multiple targets of vulnerability. Science *369*, 643-650.

Burkholder, W.F., Newell, E.W., Poidinger, M., Chen, S., and Fink, K. (2017). Deep Sequencing in Infectious Diseases: Immune and Pathogen Repertoires for the Improvement of Patient Outcomes. Front Immunol *8*, 593.

Cagigi, A., Misasi, J., Ploquin, A., Stanley, D.A., Ambrozak, D., Tsybovsky, Y., Mason, R.D., Roederer, M., and Sullivan, N.J. (2018). Vaccine generation of protective Ebola antibodies and identification of conserved B-cell signatures. J Infect Dis *218*, S528-S536.

Cao, Y., Su, B., Guo, X., Sun, W., Deng, Y., Bao, L., Zhu, Q., Zhang, X., Zheng, Y., Geng, C.*, et al.* (2020). Potent neutralizing antibodies against SARS-CoV-2 identified by high-throughput single-cell sequencing of convalescent patients' B cells. Cell *182*, 73-84 e16.

Case, J.B., Bailey, A.L., Kim, A.S., Chen, R.E., and Diamond, M.S. (2020a). Growth, detection, quantification, and inactivation of SARS-CoV-2. Virology *548*, 39-48.

Case, J.B., Rothlauf, P.W., Chen, R.E., Liu, Z., Zhao, H., Kim, A.S., Bloyet, L.M., Zeng, Q., Tahan, S., Droit, L.*, et al.* (2020b). Neutralizing antibody and soluble ACE2 inhibition of a replication-competent VSV-SARS-CoV-2 and a clinical isolate of SARS-CoV-2. Cell Host Microbe *28*, 475-485 e475.

CDC (2016). SARS (10 Years After).

CDC (2022a). Ebola Virus Disease.

CDC (2022b). Types of Influenza Viruses.

Cerutti, G., Guo, Y., Zhou, T., Gorman, J., Lee, M., Rapp, M., Reddem, E.R., Yu, J., Bahna, F., Bimela, J.*, et al.* Potent SARS-CoV-2 neutralizing antibodies directed against spike N-terminal domain target a single supersite. Cell Host & Microbe.

Cerutti, G., Guo, Y., Zhou, T., Gorman, J., Lee, M., Rapp, M., Reddem, E.R., Yu, J., Bahna, F., Bimela, J.*, et al.* (2021). Potent SARS-CoV-2 neutralizing antibodies directed against spike N-terminal domain target a single supersite. Cell Host Microbe *29*, 819-833 e817.

Chaudhary, N., and Wesemann, D.R. (2018). Analyzing Immunoglobulin Repertoires. Front Immunol *9*, 462.

Chen, E.C., Gilchuk, P., Zost, S.J., Suryadevara, N., Winkler, E.S., Cabel, C.R., Binshtein, E., Chen, R.E., Sutton, R.E., Rodriguez, J.*, et al.* (2021a). Convergent antibody responses to the SARS-CoV-2 spike protein in convalescent and vaccinated individuals. Cell Rep *36*, 109604.

Chen, F., Tzarum, N., Wilson, I.A., and Law, M. (2019). VH1-69 antiviral broadly neutralizing antibodies: genetics, structures, and relevance to rational vaccine design. Curr Opin Virol *34*, 149-159.

Chen, R.E., Zhang, X., Case, J.B., Winkler, E.S., Liu, Y., VanBlargan, L.A., Liu, J., Errico, J.M., Xie, X., Suryadevara, N.*, et al.* (2021b). Resistance of SARS-CoV-2 variants to neutralization by monoclonal and serum-derived polyclonal antibodies. Nat Med *27*, 717-726.

Chen, Y.Q., Wohlbold, T.J., Zheng, N.Y., Huang, M., Huang, Y., Neu, K.E., Lee, J., Wan, H., Rojas, K.T., Kirkpatrick, E.*, et al.* (2018). Influenza Infection in Humans Induces Broadly Cross-Reactive and Protective Neuraminidase-Reactive Antibodies. Cell *173*, 417-429 e410.

Chen, Z., Collins, A.M., Wang, Y., and Gaeta, B.A. (2010). Clustering-based identification of clonally-related immunoglobulin gene sequence sets. Immunome Res *6 Suppl 1*, S4.

Chng, J., Wang, T., Nian, R., Lau, A., Hoi, K.M., Ho, S.C., Gagnon, P., Bi, X., and Yang, Y. (2015). Cleavage efficient 2A peptides for high level monoclonal antibody expression in CHO cells. MAbs *7*, 403-412.

Clark, K., Karsch-Mizrachi, I., Lipman, D.J., Ostell, J., and Sayers, E.W. (2016). GenBank. Nucleic Acids Res *44*, D67-72.

Cohen-Dvashi, H., Zehner, M., Ehrhardt, S., Katz, M., Elad, N., Klein, F., and Diskin, R. (2020). Structural basis for a convergent immune response against Ebola virus. Cell Host Microbe *27*, 418-427 e414.

Collier, D.A., De Marco, A., Ferreira, I.A.T.M., Meng, B., Datir, R., Walls, A.C., Kemp S, S.A., Bassi, J., Pinto, D., Fregni, C.S.*, et al.* (2021). Sensitivity of SARS-CoV-2 B.1.1.7 to mRNA vaccine-elicited antibodies. Nature.

Comenzo, R.L., Reece, D., Palladini, G., Seldin, D., Sanchorawala, V., Landau, H., Falk, R., Wells, K., Solomon, A., Wechalekar, A.*, et al.* (2012). Consensus guidelines for the conduct and reporting of clinical trials in systemic light-chain amyloidosis. Leukemia *26*, 2317-2325.

Company, E.L.a. (2020). A phase 3 randomized, double-blind, placebo-controlled trial to evaluate the efficacy and safety of LY3819253 alone and in combination with LY3832479 in preventing SARS-CoV-2 infection and COVID-19 in skilled nursing and assisted living facility residents and staff; a NIAID and Lilly Collaborative Study. In Clinical trial registration NCT04497987, clinicaltrialsgov, 2020.

Corti, D., Misasi, J., Mulangu, S., Stanley, D.A., Kanekiyo, M., Wollen, S., Ploquin, A., Doria-Rose, N.A., Staupe, R.P., Bailey, M.*, et al.* (2016). Protective monotherapy against lethal Ebola virus infection by a potently neutralizing antibody. Science *351*, 1339-1342.

Dai, K., Khan, S.N., Wang, Y., He, L., Guenaga, J., Ingale, J., Sundling, C., O'Dell, S., McKee, K., Phad, G.*, et al.* (2016). HIV-1 Vaccine-elicited Antibodies Reverted to Their Inferred Naive Germline Reveal Associations between Binding Affinity and in vivo Activation. Sci Rep *6*, 20987.

Davidson, E., Bryan, C., Fong, R.H., Barnes, T., Pfaff, J.M., Mabila, M., Rucker, J.B., and Doranz, B.J. (2015). Mechanism of Binding to Ebola Virus Glycoprotein by the ZMapp, ZMAb, and MB-003 Cocktail Antibodies. J Virol *89*, 10982-10992.

Davidson, E., and Doranz, B.J. (2014). A high-throughput shotgun mutagenesis approach to mapping B-cell antibody epitopes. Immunology *143*, 13-20.

Davis, C.W., Jackson, K.J.L., McElroy, A.K., Halfmann, P., Huang, J., Chennareddy, C., Piper, A.E., Leung, Y., Albarino, C.G., Crozier, I.*, et al.* (2019). Longitudinal analysis of the human B cell response to Ebola virus infection. Cell *177*, 1566-1582 e1517.

DeKosky, B.J., Ippolito, G.C., Deschner, R.P., Lavinder, J.J., Wine, Y., Rawlings, B.M., Varadarajan, N., Giesecke, C., Dorner, T., Andrews, S.F.*, et al.* (2013). High-throughput sequencing of the paired human immunoglobulin heavy and light chain repertoire. Nat Biotechnol *31*, 166-169.

Desport, E., Bridoux, F., Sirac, C., Delbes, S., Bender, S., Fernandez, B., Quellard, N., Lacombe, C., Goujon, J.M., Lavergne, D.*, et al.* (2012). Al amyloidosis. Orphanet J Rare Dis *7*, 54.

Diss, T.C., Liu, H.X., Du, M.Q., and Isaacson, P.G. (2002). Improvements to B cell clonality analysis using PCR amplification of immunoglobulin light chain genes. Mol Pathol *55*, 98-101.

Dong, J., Zost, S.J., Greaney, A.J., Starr, T.N., Dingens, A.S., Chen, E.C., Chen, R.E., Case, J.B., Sutton, R.E., Gilchuk, P.*, et al.* (2021). Genetic and structural basis for SARS-CoV-2 variant neutralization by a two-antibody cocktail. Nat Microbiol *6*, 1233-1244.

Edgar, R.C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res *32*, 1792-1797.

Edgar, R.C. (2010). Search and clustering orders of magnitude faster than BLAST. Bioinformatics *26*, 2460-2461.

Ehrhardt, S.A., Zehner, M., Krahling, V., Cohen-Dvashi, H., Kreer, C., Elad, N., Gruell, H., Ercanoglu, M.S., Schommers, P., Gieselmann, L.*, et al.* (2019). Polyclonal and convergent antibody response to Ebola virus vaccine rVSV-ZEBOV. Nat Med *25*, 1589-1600.

Elsayed, M., Usher, S., Habib, M.H., Ahmed, N., Ali, J., Begemann, M., Shabbir, S.A., Shune, L., Al-Hilli, J., Cossor, F.*, et al.* (2020). Current Updates on the Management of AL Amyloidosis. J Hematol *10*, 147-161.

Fangzhu Zhao, M.Y., Celina Keating, Namir Shaabani, Oliver Limbo, Collin Joyce, Jordan Woehl, Shawn Barman, Alison Burns, Xueyong Zhu, Michael Ricciardi, Linghang Peng, Jessica Smith, Deli Huang, Bryan Briney, Devin Sok, David Nemazee, John R. Teijaro, Ian A. Wilson, Dennis R. Burton, Joseph G. Jardine (2021). Broadening a SARS-CoV-1 neutralizing antibody for potent SARS-CoV-2 neutralization through directed evolution. BioRxiv.

Fibriansah, G., Ibarra, K.D., Ng, T.S., Smith, S.A., Tan, J.L., Lim, X.N., Ooi, J.S., Kostyuchenko, V.A., Wang, J., de Silva, A.M.*, et al.* (2015). Cryo-EM structure of an antibody that neutralizes dengue virus type 2 by locking E protein dimers. Science *349*, 88-91.

Flyak, A.I., Ilinykh, P.A., Murin, C.D., Garron, T., Shen, X., Fusco, M.L., Hashiguchi, T., Bornholdt, Z.A., Slaughter, J.C., Sapparapu, G.*, et al.* (2015). Mechanism of human antibody-mediated neutralization of Marburg virus. Cell *160*, 893-903.

Flyak, A.I., Shen, X., Murin, C.D., Turner, H.L., David, J.A., Fusco, M.L., Lampley, R., Kose, N., Ilinykh, P.A., Kuzmina, N.*, et al.* (2016). Cross-reactive and potent neutralizing sntibody responses in human survivors of natural Ebolavirus infection. Cell *164*, 392-405.

Furuyama, W., Marzi, A., Nanbo, A., Haddock, E., Maruyama, J., Miyamoto, H., Igarashi, M., Yoshida, R., Noyori, O., Feldmann, H.*, et al.* (2016). Discovery of an antibody for pan-ebolavirus therapy. Sci Rep *6*, 20514.

Gamblin, S.J., and Skehel, J.J. (2010). Influenza hemagglutinin and neuraminidase membrane glycoproteins. J Biol Chem *285*, 28403-28409.

Garbutt, M., Liebscher, R., Wahl-Jensen, V., Jones, S., Moller, P., Wagner, R., Volchkov, V., Klenk, H.D., Feldmann, H., and Stroher, U. (2004). Properties of replication-competent vesicular stomatitis virus vectors expressing glycoproteins of filoviruses and arenaviruses. J Virol *78*, 5458-5465.

Georgiou, G., Ippolito, G.C., Beausang, J., Busse, C.E., Wardemann, H., and Quake, S.R. (2014). The promise and challenge of high-throughput sequencing of the antibody repertoire. Nat Biotechnol *32*, 158-168.

Gertz, M.A. (2016). Immunoglobulin light chain amyloidosis: 2016 update on diagnosis, prognosis, and treatment. Am J Hematol *91*, 947-956.

Giang, E., Dorner, M., Prentoe, J.C., Dreux, M., Evans, M.J., Bukh, J., Rice, C.M., Ploss, A., Burton, D.R., and Law, M. (2012). Human broadly neutralizing antibodies to the envelope glycoprotein complex of hepatitis C virus. Proc Natl Acad Sci U S A *109*, 6205-6210.

Gilchuk, P., Bombardi, R.G., Erasmus, J.H., Tan, Q., Nargi, R., Soto, C., Abbink, P., Suscovich, T.J., Durnell, L.A., Khandhar, A.*, et al.* (2020a). Integrated pipeline for the accelerated discovery of antiviral antibody therapeutics. Nat Biomed Eng *4*, 1030-1043.

Gilchuk, P., Guthals, A., Bonissone, S.R., Shaw, J.B., Ilinykh, P.A., Huang, K., Bombardi, R.G., Liang, J., Grinyo, A., Davidson, E.*, et al.* (2021). Proteo-genomic analysis identifies two major sites of vulnerability on Ebolavirus glycoprotein for neutralizing antibodies in convalescent human plasma. Front Immunol *12*, 706757.

Gilchuk, P., Kuzmina, N., Ilinykh, P.A., Huang, K., Gunn, B.M., Bryan, A., Davidson, E., Doranz, B.J., Turner, H.L., Fusco, M.L.*, et al.* (2018). Multifunctional pan-Ebolavirus antibody recognizes a site of broad vulnerability on the Ebolavirus glycoprotein. Immunity *49*, 363-374 e310.

Gilchuk, P., Murin, C.D., Milligan, J.C., Cross, R.W., Mire, C.E., Ilinykh, P.A., Huang, K., Kuzmina, N., Altman, P.X., Hui, S.*, et al.* (2020b). Analysis of a therapeutic antibody cocktail reveals determinants for cooperative and broad Ebolavirus neutralization. Immunity *52*, 388-403 e312.

Golden, J.W., Cline, C.R., Zeng, X., Garrison, A.R., Carey, B.D., Mucker, E.M., White, L.E., Shamblin, J.D., Brocato, R.L., Liu, J.*, et al.* (2020). Human angiotensin-converting enzyme 2 transgenic mice infected with SARS-CoV-2 develop severe and fatal respiratory disease. JCI Insight *5*.

Goldstein, L.D., Chen, Y.J., Wu, J., Chaudhuri, S., Hsiao, Y.C., Schneider, K., Hoi, K.H., Lin, Z., Guerrero, S., Jaiswal, B.S.*, et al.* (2019). Massively parallel single-cell B-cell receptor sequencing enables rapid discovery of diverse antigen-reactive antibodies. Commun Biol *2*, 304.

Gordon, D.E., Jang, G.M., Bouhaddou, M., Xu, J., Obernier, K., White, K.M., O'Meara, M.J., Rezelj, V.V., Guo, J.Z., Swaney, D.L.*, et al.* (2020). A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. Nature *583*, 459-468.

Grandea, A.G., 3rd, Olsen, O.A., Cox, T.C., Renshaw, M., Hammond, P.W., Chan-Hui, P.Y., Mitcham, J.L., Cieplak, W., Stewart, S.M., Grantham, M.L.*, et al.* (2010). Human antibodies reveal a protective epitope that is highly conserved among human and nonhuman influenza A viruses. Proc Natl Acad Sci U S A *107*, 12658-12663.

Greaney, A.J., Starr, T.N., Barnes, C.O., Weisblum, Y., Schmidt, F., Caskey, M., Gaebler, C., Cho, A., Agudelo, M., Finkin, S.*, et al.* (2021). Mapping mutations to the SARS-CoV-2 RBD that escape binding by different classes of antibodies. Nat Commun *12*, 4196.

Gupta, N.T., Adams, K.D., Briggs, A.W., Timberlake, S.C., Vigneault, F., and Kleinstein, S.H. (2017). Hierarchical clustering can identify B cell clones with high confidence in Ig repertoire sequencing data. J Immunol *198*, 2489-2499.

Hansen, J., Baum, A., Pascal, K.E., Russo, V., Giordano, S., Wloga, E., Fulton, B.O., Yan, Y., Koon, K., Patel, K.*, et al.* (2020). Studies in humanized mice and convalescent humans yield a SARS-CoV-2 antibody cocktail. Science *369*, 1010-1014.

Havenar-Daughton, C., Abbott, R.K., Schief, W.R., and Crotty, S. (2018). When designing vaccines, consider the starting material: the human B cell repertoire. Curr Opin Immunol *53*, 209-216.

Hoenen, T., Groseth, A., and Feldmann, H. (2019). Therapeutic strategies to target the Ebola virus life cycle. Nat Rev Microbiol *17*, 593-606.

Hoffmann, M., Kleine-Weber, H., Schroeder, S., Kruger, N., Herrler, T., Erichsen, S., Schiergens, T.S., Herrler, G., Wu, N.H., Nitsche, A.*, et al.* (2020). SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor. Cell *181*, 271-280 e278.

Hsieh, T.C., Ma, K.H., and Anne, C. (2016). iNEXT: an R package for rarefaction and extrapolation of species diversity (Hill numbers). Methods in Ecology and Evolution, 1451-1456.

Huang, C.C., Venturi, M., Majeed, S., Moore, M.J., Phogat, S., Zhang, M.Y., Dimitrov, D.S., Hendrickson, W.A., Robinson, J., Sodroski, J.*, et al.* (2004). Structural basis of tyrosine sulfation and VH-gene usage in antibodies that recognize the HIV type 1 coreceptor-binding site on gp120. Proc Natl Acad Sci U S A *101*, 2706-2711.

Ilinykh, P.A., Graber, J., Kuzmina, N.A., Huang, K., Ksiazek, T.G., Crowe, J.E., Jr., and Bukreyev, A. (2018). Ebolavirus chimerization for the development of a mouse model for screening of Bundibugyo-specific antibodies. J Infect Dis *218*, S418-S422.

Ilinykh, P.A., Shen, X., Flyak, A.I., Kuzmina, N., Ksiazek, T.G., Crowe, J.E., Jr., and Bukreyev, A. (2016). Chimeric filoviruses for identification and characterization of monoclonal antibodies. J Virol *90*, 3890-3901.

Jackson, K.J., Liu, Y., Roskin, K.M., Glanville, J., Hoh, R.A., Seo, K., Marshall, E.L., Gurley, T.C., Moody, M.A., Haynes, B.F.*, et al.* (2014). Human responses to influenza vaccination show seroconversion signatures and convergent antibody rearrangements. Cell Host Microbe *16*, 105-114.

Jiang, S., Hillyer, C., and Du, L. (2020). Neutralizing antibodies against SARS-CoV-2 and other human coronaviruses. Trends Immunol *41*, 355-359.

Johnson, J.a. (2021). Johnson & Johnson announces U.S. CDC advisory committee recommends first single-shot COVID-19 vaccine for adults 18 and older in U.S.

Joyce, M.G., Wheatley, A.K., Thomas, P.V., Chuang, G.Y., Soto, C., Bailer, R.T., Druz, A., Georgiev, I.S., Gillespie, R.A., Kanekiyo, M.*, et al.* (2016). Vaccine-induced antibodies that neutralize group 1 and group 2 Influenza A viruses. Cell *166*, 609-623.

Kastritis, E., Kostopoulos, I.V., Terpos, E., Paiva, B., Fotiou, D., Gavriatopoulou, M., Kanellias, N., Ziogas, D.C., Roussou, M., Migkou, M.*, et al.* (2018). Evaluation of minimal residual disease using next-generation flow cytometry in patients with AL amyloidosis. Blood Cancer J *8*, 46.

Katoh, K., Misawa, K., Kuma, K., and Miyata, T. (2002). MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res *30*, 3059-3066.

Keck, Z.Y., Enterlein, S.G., Howell, K.A., Vu, H., Shulenin, S., Warfield, K.L., Froude, J.W., Araghi, N., Douglas, R., Biggins, J.*, et al.* (2016). Macaque Monoclonal Antibodies Targeting Novel Conserved Epitopes within Filovirus Glycoprotein. J Virol *90*, 279-291.

Kenneth, M.T., P.; Walport, M.; and Janeway, C. (2008). Janeway's immunobiology (New York: Garland Science).

Khodadadi, L., Cheng, Q., Radbruch, A., and Hiepe, F. (2019). The Maintenance of Memory Plasma Cells. Front Immunol *10*, 721.

Kim, H., Webster, R.G., and Webby, R.J. (2018). Influenza Virus: Dealing with a Drifting and Shifting Pathogen. Viral Immunol *31*, 174-183.

Klein, F., Diskin, R., Scheid, J.F., Gaebler, C., Mouquet, H., Georgiev, I.S., Pancera, M., Zhou, T., Incesu, R.B., Fu, B.Z.*, et al.* (2013). Somatic mutations of the immunoglobulin framework are generally required for broad and potent HIV-1 neutralization. Cell *153*, 126-138.

Kohler, G., and Milstein, C. (1975). Continuous cultures of fused cells secreting antibody of predefined specificity. Nature *256*, 495-497.

Kovaltsuk, A., Leem, J., Kelm, S., Snowden, J., Deane, C.M., and Krawczyk, K. (2018). Observed Antibody Space: A Resource for Data Mining Next-Generation Sequencing of Antibody Repertoires. J Immunol *201*, 2502-2509.

Krammer, F. (2019). The human antibody response to influenza A virus infection and vaccination. Nat Rev Immunol *19*, 383-397.

Krammer, F. (2020). SARS-CoV-2 vaccines in development. Nature *586*, 516-527.

Krammer, F., and Palese, P. (2013). Influenza virus hemagglutinin stalk-based antibodies and vaccines. Curr Opin Virol *3*, 521-530.

Krammer, F., Smith, G.J.D., Fouchier, R.A.M., Peiris, M., Kedzierska, K., Doherty, P.C., Palese, P., Shaw, M.L., Treanor, J., Webster, R.G*., et al.* (2018). Influenza. Nat Rev Dis Primers *4*, 3.

Krawczyk, K., Buchanan, A., and Marcatili, P. (2021). Data mining patented antibody sequences. MAbs *13*, 1892366.

Krawczyk, K., Raybould, M.I.J., Kovaltsuk, A., and Deane, C.M. (2019). Looking for therapeutic antibodies in next-generation sequencing repositories. MAbs *11*, 1197-1205.

Kreer, C., Zehner, M., Weber, T., Ercanoglu, M.S., Gieselmann, L., Rohde, C., Halwe, S., Korenkov, M., Schommers, P., Vanshylla, K*., et al.* (2020). Longitudinal isolation of potent near-germline SARS-CoV-2-neutralizing antibodies from COVID-19 patients. Cell *182*, 843-854 e812.

Kumar, S., Dispenzieri, A., Lacy, M.Q., Hayman, S.R., Buadi, F.K., Colby, C., Laumann, K., Zeldenrust, S.R., Leung, N., Dingli, D*., et al.* (2012). Revised prognostic staging system for light chain amyloidosis incorporating cardiac biomarkers and serum free light chain measurements. J Clin Oncol *30*, 989-995.

Ladner, J.T., Henson, S.N., Boyle, A.S., Engelbrektson, A.L., Fink, Z.W., Rahee, F., D'Ambrozio, J., Schaecher, K.E., Stone, M., Dong, W*., et al.* (2021). Epitope-resolved profiling of the SARS-CoV-2 antibody response identifies cross-reactivity with endemic human coronaviruses. Cell Rep Med *2*, 100189.

Lan, J., Ge, J., Yu, J., Shan, S., Zhou, H., Fan, S., Zhang, Q., Shi, X., Wang, Q., Zhang, L*., et al.* (2020). Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. Nature *581*, 215-220.

Lee, J.E., Fusco, M.L., Hessell, A.J., Oswald, W.B., Burton, D.R., and Saphire, E.O. (2008). Structure of the Ebola virus glycoprotein bound to an antibody from a human survivor. Nature *454*, 177-182.

Lee, J.E., and Saphire, E.O. (2009). Ebolavirus glycoprotein structure and mechanism of entry. Future Virol *4*, 621-635.

Lee, J.H., Toy, L., Kos, J.T., Safonova, Y., Schief, W.R., Havenar-Daughton, C., Watson, C.T., and Crotty, S. (2021). Vaccine genetics of IGHV1-2 VRC01-class broadly neutralizing antibody precursor naive human B cells. NPJ Vaccines *6*, 113.

Levine, M.M. (2019). Monoclonal antibody therapy for Ebola virus disease. N Engl J Med *381*, 2365-2366.

Li, W., Moore, M.J., Vasilieva, N., Sui, J., Wong, S.K., Berne, M.A., Somasundaran, M., Sullivan, J.L., Luzuriaga, K., Greenough, T.C.*, et al.* (2003). Angiotensin-converting enzyme 2 is a functional receptor for the SARS coronavirus. Nature *426*, 450-454.

Linda J. Rennick, L.R.R.-M., Sham Nambulli, W. Paul Duprex, Kevin R. McCarthy (2021). Deletion disrupts a conserved antibody epitope in a SARS-CoV-2 variant of concern. bioRxiv.

Liu, L., Wang, P., Nair, M.S., Yu, J., Rapp, M., Wang, Q., Luo, Y., Chan, J.F., Sahi, V., Figueroa, A.*, et al.* (2020). Potent neutralizing antibodies against multiple epitopes on SARS-CoV-2 spike. Nature *584*, 450-456.

Madsen, A., Dai, Y.N., McMahon, M., Schmitz, A.J., Turner, J.S., Tan, J., Lei, T., Alsoussi, W.B., Strohmeier, S., Amor, M.*, et al.* (2020). Human Antibodies Targeting Influenza B Virus Neuraminidase Active Site Are Broadly Protective. Immunity *53*, 852-863 e857.

Malvy, D., McElroy, A.K., de Clerck, H., Gunther, S., and van Griensven, J. (2019). Ebola virus disease. Lancet *393*, 936-948.

Manwani, R., Cohen, O., Sharpley, F., Mahmood, S., Sachchithanantham, S., Foard, D., Lachmann, H.J., Quarta, C., Fontana, M., Gillmore, J.D.*, et al.* (2019). A prospective observational study of 915 patients with systemic AL amyloidosis treated with upfront bortezomib. Blood *134*, 2271-2280.

Mastronarde, D.N. (2005). Automated electron microscope tomography using robust prediction of specimen movements. J Struct Biol *152*, 36-51.

Matrosovich, M.N., Matrosovich, T.Y., Gray, T., Roberts, N.A., and Klenk, H.D. (2004). Neuraminidase is important for the initiation of influenza virus infection in human airway epithelium. J Virol *78*, 12665-12667.

Mesin, L., Ersching, J., and Victora, G.D. (2016). Germinal Center B Cell Dynamics. Immunity *45*, 471-482.

Miho, E., Roskar, R., Greiff, V., and Reddy, S.T. (2019). Large-scale network analysis reveals the sequence space architecture of antibody repertoires. Nat Commun *10*, 1321.

Milani, P., Basset, M., Russo, F., Foli, A., Merlini, G., and Palladini, G. (2017). Patients with light-chain amyloidosis and low free light-chain burden have distinct clinical features and outcome. Blood *130*, 625-631.

Misasi, J., Gilman, M.S., Kanekiyo, M., Gui, M., Cagigi, A., Mulangu, S., Corti, D., Ledgerwood, J.E., Lanzavecchia, A., Cunningham, J.*, et al.* (2016). Structural and molecular basis for Ebola virus neutralization by protective human antibodies. Science *351*, 1343-1346.

Moderna (2020). Moderna's COVID-19 Vaccine Candidate Meets its Primary Efficacy Endpoint in the First Interim Analysis of the Phase 3 COVE Study.

Moller-Tank, S., and Maury, W. (2015). Ebola virus entry: a curious and complex series of events. PLoS Pathog *11*, e1004731.

Mukhamedova, M., Wrapp, D., Shen, C.H., Gilman, M.S.A., Ruckwardt, T.J., Schramm, C.A., Ault, L., Chang, L., Derrien-Colemyn, A., Lucas, S.A.M.*, et al.* (2021). Vaccination with prefusion-stabilized respiratory syncytial virus fusion protein induces genetically and antigenically diverse antibody responses. Immunity *54*, 769-780 e766.

Murin, C.D., Gilchuk, P., Ilinykh, P.A., Huang, K., Kuzmina, N., Shen, X., Bruhn, J.F., Bryan, A.L., Davidson, E., Doranz, B.J.*, et al.* (2021). Convergence of a common solution for broad ebolavirus neutralization by glycan cap-directed human antibodies. Cell Rep *35*, 108984.

Nanbo, A., Imai, M., Watanabe, S., Noda, T., Takahashi, K., Neumann, G., Halfmann, P., and Kawaoka, Y. (2010). Ebolavirus is internalized into host cells via macropinocytosis in a viral glycoprotein-dependent manner. PLoS Pathog *6*, e1001121.

Nielsen, S.C.A., Yang, F., Hoh, R.A., Jackson, K.J.L., Roeltgen, K., Lee, J.Y., Rustagi, A., Rogers, A.J., Powell, A.E., Kim, P.S.*, et al.* (2020). B cell clonal expansion and convergent antibody responses to SARS-CoV-2. Res Sq.

Noda, T., Sagara, H., Suzuki, E., Takada, A., Kida, H., and Kawaoka, Y. (2002). Ebola virus VP40 drives the formation of virus-like filamentous particles along with GP. J Virol *76*, 4855-4865.

Nojima, T., Haniuda, K., Moutai, T., Matsudaira, M., Mizokawa, S., Shiratori, I., Azuma, T., and Kitamura, D. (2011). In-vitro derived germinal centre B cells differentially generate memory B or plasma cells in vivo. Nat Commun *2*, 465.

Ohi, M., Li, Y., Cheng, Y., and Walz, T. (2004). Negative staining and image classification - powerful tools in modern electron microscopy. Biol Proced Online *6*, 23-34.

Palladini, G., and Merlini, G. (2019). When should treatment of AL amyloidosis start at relapse? Early, to prevent organ progression. Blood Adv *3*, 212-215.

Palladini, G., Milani, P., and Merlini, G. (2015). Novel strategies for the diagnosis and treatment of cardiac amyloidosis. Expert Rev Cardiovasc Ther *13*, 1195-1211.

Palm, A.E., and Henry, C. (2019). Remembrance of Things Past: Long-Term B Cell Memory After Infection and Vaccination. Front Immunol *10*, 1787.

Pappas, L., Foglierini, M., Piccoli, L., Kallewaard, N.L., Turrini, F., Silacci, C., Fernandez-Rodriguez, B., Agatic, G., Giacchetto-Sasselli, I., Pellicciotta, G.*, et al.* (2014). Rapid development of broadly influenza neutralizing antibodies through redundant mutations. Nature *516*, 418-422.

Pascal, K.E., Dudgeon, D., Trefry, J.C., Anantpadma, M., Sakurai, Y., Murin, C.D., Turner, H.L., Fairhurst, J., Torres, M., Rafique, A.*, et al.* (2018). Development of clinical-stage human monoclonal antibodies that treat advanced Ebola virus disease in nonhuman primates. J Infect Dis *218*, S612-S626.

Pedrioli, A., and Oxenius, A. (2021). Single B cell technologies for monoclonal antibody discovery. Trends Immunol *42*, 1143-1158.

Pejchal, R., Doores, K.J., Walker, L.M., Khayat, R., Huang, P.S., Wang, S.K., Stanfield, R.L., Julien, J.P., Ramos, A., Crispin, M.*, et al.* (2011). A potent and broad neutralizing antibody recognizes and penetrates the HIV glycan shield. Science *334*, 1097-1103.

Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., and Ferrin, T.E. (2004). UCSF Chimera--a visualization system for exploratory research and analysis. J Comput Chem *25*, 1605-1612.

Pfizer (2020). Pfizer and BioNTech announce vaccine candidate against COVID-19 acheived success in first interim analysis from phase 3 study.

Pinto, D., Sauer, M.M., Czudnochowski, N., Low, J.S., Tortorici, M.A., Housley, M.P., Noack, J., Walls, A.C., Bowen, J.E., Guarino, B.*, et al.* (2021). Broad betacoronavirus neutralization by a stem helix-specific human antibody. Science *373*, 1109-1116.

Planas, D., Saunders, N., Maes, P., Guivel-Benhassine, F., Planchais, C., Buchrieser, J., Bolland, W.-H., Porrot, F., Staropoli, I., Lemoine, F.*, et al.* (2021). Considerable escape of SARS-CoV-2 Omicron to antibody neutralization. Nature.

Plante, J.A., Liu, Y., Liu, J., Xia, H., Johnson, B.A., Lokugamage, K.G., Zhang, X., Muruato, A.E., Zou, J., Fontes-Garfias, C.R.*, et al.* (2021). Spike mutation D614G alters SARS-CoV-2 fitness. Nature *592*, 116-121.

Punjani, A., Rubinstein, J.L., Fleet, D.J., and Brubaker, M.A. (2017). cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. Nat Methods *14*, 290-296.

Qiu, X., Wong, G., Audet, J., Bello, A., Fernando, L., Alimonti, J.B., Fausther-Bovendo, H., Wei, H., Aviles, J., Hiatt, E.*, et al.* (2014). Reversion of advanced Ebola virus disease in nonhuman primates with ZMapp. Nature *514*, 47-53.

Quast, I., and Tarlinton, D. (2021). B cell memory: understanding COVID-19. Immunity *54*, 205-210.

Rapp, M., Guo, Y., Reddem, E.R., Yu, J., Liu, L., Wang, P., Cerutti, G., Katsamba, P., Bimela, J.S., Bahna, F.A.*, et al.* (2021). Modular basis for potent SARS-CoV-2 neutralization by a prevalent VH1-2-derived antibody class. Cell Rep *35*, 108950.

Rappuoli, R., Bottomley, M.J., D'Oro, U., Finco, O., and De Gregorio, E. (2016). Reverse vaccinology 2.0: Human immunology instructs vaccine antigen design. J Exp Med *213*, 469-481.

Regeneron Pharmaceuticals, I. (2020). Regeneron's casirivimab and imdevimab antibody cocktail for COVID-19 is first combination therapy to receive FDA Emergency Use Authorization.

Rijal, P., Elias, S.C., Machado, S.R., Xiao, J., Schimanski, L., O'Dowd, V., Baker, T., Barry, E., Mendelsohn, S.C., Cherry, C.J.*, et al.* (2019). Therapeutic monoclonal antibodies for Ebola virus infection derived from vaccinated humans. Cell Rep *27*, 172-186 e177.

Robbiani, D.F., Gaebler, C., Muecksch, F., Lorenzi, J.C.C., Wang, Z., Cho, A., Agudelo, M., Barnes, C.O., Gazumyan, A., Finkin, S.*, et al.* (2020). Convergent antibody responses to SARS-CoV-2 in convalescent individuals. Nature *584*, 437-442.

Robins, H. (2013). Immunosequencing: applications of immune repertoire deep sequencing. Curr Opin Immunol *25*, 646-652.

Robinson, W.H. (2015). Sequencing the functional antibody repertoire--diagnostic and therapeutic discovery. Nat Rev Rheumatol *11*, 171-182.

Rodriguez, M.Z., Comin, C.H., Casanova, D., Bruno, O.M., Amancio, D.R., Costa, L.D.F., and Rodrigues, F.A. (2019). Clustering algorithms: A comparative approach. PLoS One *14*, e0210236.

Rogers, T.F., Zhao, F., Huang, D., Beutler, N., Burns, A., He, W.T., Limbo, O., Smith, C., Song, G., Woehl, J.*, et al.* (2020). Isolation of potent SARS-CoV-2 neutralizing antibodies and protection from disease in a small animal model. Science *369*, 956-963.

Rohou, A., and Grigorieff, N. (2015). CTFFIND4: Fast and accurate defocus estimation from electron micrographs. J Struct Biol *192*, 216-221.

Rota, P.A., Wallis, T.R., Harmon, M.W., Rota, J.S., Kendal, A.P., and Nerome, K. (1990). Cocirculation of two distinct evolutionary lineages of influenza type B virus since 1983. Virology *175*, 59-68.

Rysava, R. (2019). AL amyloidosis: advances in diagnostics and treatment. Nephrol Dial Transplant *34*, 1460-1466.

Saary, P., Forslund, K., Bork, P., and Hildebrand, F. (2017). RTK: efficient rarefaction analysis of large datasets. Bioinformatics *33*, 2594-2595.

Sakharkar, M., Rappazzo, C.G., Wieland-Alter, W.F., Hsieh, C.L., Wrapp, D., Esterman, E.S., Kaku, C.I., Wec, A.Z., Geoghegan, J.C., McLellan, J.S.*, et al.* (2021). Prolonged evolution of the human B cell response to SARS-CoV-2 infection. Sci Immunol *6*, eabg6916.

Saphire, E.O., and Aman, M.J. (2016). Feverish Quest for Ebola Immunotherapy: Straight or Cocktail? Trends Microbiol *24*, 684-686.

Schatz, D.G., Oettinger, M.A., and Baltimore, D. (1989). The V(D)J recombination activating gene, RAG-1. Cell *59*, 1035-1048.

Schmidt, A.G., Do, K.T., McCarthy, K.R., Kepler, T.B., Liao, H.X., Moody, M.A., Haynes, B.F., and Harrison, S.C. (2015). Immunogenic Stimulus for Germline Precursors of Antibodies that Engage the Influenza Hemagglutinin Receptor-Binding Site. Cell Rep *13*, 2842-2850.

Schmitz, A.J., Turner, J.S., Liu, Z., Zhou, J.Q., Aziati, I.D., Chen, R.E., Joshi, A., Bricker, T.L., Darling, T.L., Adelsberg, D.C.*, et al.* (2021). A vaccine-induced public antibody protects against SARS-CoV-2 and emerging variants. Immunity *54*, 2159-2166 e2156.

Schramm, C.A., and Douek, D.C. (2018). Beyond Hot Spots: Biases in Antibody Somatic Hypermutation and Implications for Vaccine Design. Front Immunol *9*, 1876.

Setliff, I., McDonnell, W.J., Raju, N., Bombardi, R.G., Murji, A.A., Scheepers, C., Ziki, R., Mynhardt, C., Shepherd, B.E., Mamchak, A.A.*, et al.* (2018). Multi-donor longitudinal antibody repertoire sequencing reveals the existence of public antibody clonotypes in HIV-1 infection. Cell Host Microbe *23*, 845-854 e846.

Seydoux, E., Homad, L.J., MacCamy, A.J., Parks, K.R., Hurlburt, N.K., Jennewein, M.F., Akins, N.R., Stuart, A.B., Wan, Y.H., Feng, J.*, et al.* (2020). Analysis of a SARS-CoV-2-Infected Individual Reveals Development of Potent Neutralizing Antibodies with Limited Somatic Mutation. Immunity *53*, 98-105 e105.

Sidana, S., Muchtar, E., Sidiqi, M.H., Jevremovic, D., Dispenzieri, A., Gonsalves, W., Buadi, F., Lacy, M.Q., Hayman, S.R., Kourelis, T.*, et al.* (2020). Impact of minimal residual negativity using next generation flow cytometry on outcomes in light chain amyloidosis. Am J Hematol *95*, 1-6.

Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Soding, J.*, et al.* (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol Syst Biol *7*, 539.

Smith, K., Garman, L., Wrammert, J., Zheng, N.Y., Capra, J.D., Ahmed, R., and Wilson, P.C. (2009). Rapid generation of fully human monoclonal antibodies specific to a vaccinating antigen. Nat Protoc *4*, 372-384.

Smith, S.A., and Crowe, J.E., Jr. (2015). Use of Human Hybridoma Technology To Isolate Human Monoclonal Antibodies. Microbiol Spectr *3*, AID-0027-2014.

Soto, C., Bombardi, R.G., Branchizio, A., Kose, N., Matta, P., Sevy, A.M., Sinkovits, R.S., Gilchuk, P., Finn, J.A., and Crowe, J.E., Jr. (2019). High frequency of shared clonotypes in human B cell receptor repertoires. Nature *566*, 398-402.

Soto, C., Finn, J.A., Willis, J.R., Day, S.B., Sinkovits, R.S., Jones, T., Schmitz, S., Meiler, J., Branchizio, A., and Crowe, J.E., Jr. (2020). PyIR: a scalable wrapper for processing billions of immunoglobulin and T cell receptor sequences using IgBLAST. BMC Bioinformatics *21*, 314.

Stadlbauer, D., Zhu, X., McMahon, M., Turner, J.S., Wohlbold, T.J., Schmitz, A.J., Strohmeier, S., Yu, W., Nachbagauer, R., Mudd, P.A.*, et al.* (2019). Broadly protective human antibodies that target the active site of influenza virus neuraminidase. Science *366*, 499-504.

Starr, T.N., Greaney, A.J., Addetia, A., Hannon, W.W., Choudhary, M.C., Dingens, A.S., Li, J.Z., and Bloom, J.D. (2021). Prospective mapping of viral mutations that escape antibodies used to treat COVID-19. Science *371*, 850-854.

Stokes, J., Jr.; Farquhar, J. A.; Drake, M. E.; Capps, R. B.; Ward, C. S., Jr.; Mills, O., and Kitts, A. W. (1951). Infectious Hepatitis: Length of Protection by Immune Serum Globulin (Gamma Globulin) During Epidemics. JAMA *147*, 714–719.

Sui, J., Hwang, W.C., Perez, S., Wei, G., Aird, D., Chen, L.M., Santelli, E., Stec, B., Cadwell, G., Ali, M.*, et al.* (2009). Structural and functional bases for broad-spectrum neutralization of avian and human influenza A viruses. Nat Struct Mol Biol *16*, 265-273.

Suryadevara, N., Shrihari, S., Gilchuk, P., VanBlargan, L.A., Binshtein, E., Zost, S.J., Nargi, R.S., Sutton, R.E., Winkler, E.S., Chen, E.C.*, et al.* (2021). Neutralizing and protective human monoclonal antibodies recognizing the N-terminal domain of the SARS-CoV-2 spike protein. Cell *184*, 2316-2331 e2315.

Tan, G.S., Lee, P.S., Hoffman, R.M., Mazel-Sanchez, B., Krammer, F., Leon, P.E., Ward, A.B., Wilson, I.A., and Palese, P. (2014). Characterization of a broadly neutralizing monoclonal antibody that targets the fusion domain of group 2 influenza A virus hemagglutinin. J Virol *88*, 13580-13592.

Tan, T.J.C., Yuan, M., Kuzelka, K., Padron, G.C., Beal, J.R., Chen, X., Wang, Y., Rivera-Cardona, J., Zhu, X., Stadtmueller, B.M.*, et al.* (2021). Sequence signatures of two public antibody clonotypes that bind SARS-CoV-2 receptor binding domain. Nat Commun *12*, 3815.

Taubenberger, J.K., and Kash, J.C. (2010). Influenza virus evolution, host adaptation, and pandemic formation. Cell Host Microbe *7*, 440-451.

Taubenberger, J.K., and Morens, D.M. (2008). The pathology of influenza virus infections. Annu Rev Pathol *3*, 499-522.

Tegally, H., Wilkinson, E., Giovanetti, M., Iranzadeh, A., Fonseca, V., Giandhari, J., Doolabh, D., Pillay, S., San, E.J., Msomi, N.*, et al.* (2021). Emergence of a SARS-CoV-2 variant of concern with mutations in spike glycoprotein. Nature.

Thompson, J.D., Higgins, D.G., and Gibson, T.J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res *22*, 4673-4680.

Tong, S., Li, Y., Rivailler, P., Conrardy, C., Castillo, D.A., Chen, L.M., Recuenco, S., Ellison, J.A., Davis, C.T., York, I.A.*, et al.* (2012). A distinct lineage of influenza A virus from bats. Proc Natl Acad Sci U S A *109*, 4269-4274.

Tortorici, M.A., Beltramello, M., Lempp, F.A., Pinto, D., Dang, H.V., Rosen, L.E., McCallum, M., Bowen, J., Minola, A., Jaconi, S.*, et al.* (2020). Ultrapotent human antibodies protect against SARS-CoV-2 challenge via multiple mechanisms. Science *370*, 950-957.

Tortorici, M.A., and Veesler, D. (2019). Structural insights into coronavirus entry. Adv Virus Res *105*, 93-116.

Towner, J.S., Paragas, J., Dover, J.E., Gupta, M., Goldsmith, C.S., Huggins, J.W., and Nichol, S.T. (2005). Generation of eGFP expressing recombinant Zaire ebolavirus for analysis of early pathogenesis events and high-throughput antiviral drug screening. Virology *332*, 20-27.

Trepel, F. (1974). Number and distribution of lymphocytes in man. A critical analysis. Klin Wochenschr *52*, 511-515.

Turchaninova, M.A., Davydov, A., Britanova, O.V., Shugay, M., Bikos, V., Egorov, E.S., Kirgizova, V.I., Merzlyak, E.M., Staroverov, D.B., Bolotin, D.A.*, et al.* (2016). High-quality full-length immunoglobulin profiling with unique molecular barcoding. Nat Protoc *11*, 1599-1616.

van Dongen, J.J., Langerak, A.W., Bruggemann, M., Evans, P.A., Hummel, M., Lavender, F.L., Delabesse, E., Davi, F., Schuuring, E., Garcia-Sanz, R.*, et al.* (2003). Design and standardization of PCR primers and protocols for detection of clonal immunoglobulin and T-cell receptor gene recombinations in suspect lymphoproliferations: report of the BIOMED-2 Concerted Action BMH4-CT98-3936. Leukemia *17*, 2257-2317.

VanBlargan, L.A., Errico, J.M., Halfmann, P.J., Zost, S.J., Crowe, J.E., Jr., Purcell, L.A., Kawaoka, Y., Corti, D., Fremont, D.H., and Diamond, M.S. (2022). An infectious SARS-CoV-2 B.1.1.529 Omicron virus escapes neutralization by therapeutic monoclonal antibodies. Nat Med.

Vanderven, H.A., and Kent, S.J. (2020). The protective potential of Fc-mediated antibody functions against influenza virus and other viral pathogens. Immunol Cell Biol *98*, 253-263.

Viant, C., Weymar, G.H.J., Escolano, A., Chen, S., Hartweger, H., Cipolla, M., Gazumyan, A., and Nussenzweig, M.C. (2020). Antibody Affinity Shapes the Choice between Memory and Germinal Center B Cell Fates. Cell *183*, 1298-1311 e1211.

Victora, G.D., Schwickert, T.A., Fooksman, D.R., Kamphorst, A.O., Meyer-Hermann, M., Dustin, M.L., and Nussenzweig, M.C. (2010). Germinal center dynamics revealed by multiphoton microscopy with a photoactivatable fluorescent reporter. Cell *143*, 592-605.

Voss, W.N., Hou, Y.J., Johnson, N.V., Delidakis, G., Kim, J.E., Javanmardi, K., Horton, A.P., Bartzoka, F., Paresi, C.J., Tanno, Y.*, et al.* (2021). Prevalent, protective, and convergent IgG recognition of SARS-CoV-2 non-RBD spike epitopes. Science *372*, 1108-1112.

Wan, Y., Shang, J., Graham, R., Baric, R.S., and Li, F. (2020). Receptor recognition by the novel coronavirus from Wuhan: an analysis based on decade-long structural studies of SARS coronavirus. J Virol *94*.

Wang, H., Shi, Y., Song, J., Qi, J., Lu, G., Yan, J., and Gao, G.F. (2016). Ebola Viral Glycoprotein Bound to Its Endosomal Receptor Niemann-Pick C1. Cell *164*, 258-268.

Wang, P., Nair, M.S., Liu, L., Iketani, S., Luo, Y., Guo, Y., Wang, M., Yu, J., Zhang, B., Kwong, P.D.*, et al.* (2021a). Antibody resistance of SARS-CoV-2 variants B.1.351 and B.1.1.7. Nature.

Wang, Y., Yuan, M., Peng, J., Wilson, I.A., and Wu, N.C. (2021b). A large-scale systematic survey of SARS-CoV-2 antibodies reveals recurring molecular features. bioRxiv.

Wang, Z., Schmidt, F., Weisblum, Y., Muecksch, F., Barnes, C.O., Finkin, S., Schaefer-Babajew, D., Cipolla, M., Gaebler, C., Lieberman, J.A.*, et al.* (2021c). mRNA vaccine-elicited antibodies to SARS-CoV-2 and circulating variants. Nature.

Wec, A.Z., Bornholdt, Z.A., He, S., Herbert, A.S., Goodwin, E., Wirchnianski, A.S., Gunn, B.M., Zhang, Z., Zhu, W., Liu, G.*, et al.* (2019). Development of a Human Antibody Cocktail that Deploys Multiple Functions to Confer Pan-Ebolavirus Protection. Cell Host Microbe *25*, 39-48 e35.

Wec, A.Z., Herbert, A.S., Murin, C.D., Nyakatura, E.K., Abelson, D.M., Fels, J.M., He, S., James, R.M., de La Vega, M.A., Zhu, W.*, et al.* (2017). Antibodies from a human survivor define sites of vulnerability for broad protection against Ebolaviruses. Cell *169*, 878-890 e815.

Wec, A.Z., Wrapp, D., Herbert, A.S., Maurer, D.P., Haslwanter, D., Sakharkar, M., Jangra, R.K., Dieterle, M.E., Lilov, A., Huang, D.*, et al.* (2020). Broad neutralization of SARS-related viruses by human monoclonal antibodies. Science *369*, 731-736.

Wei, L., Chahwan, R., Wang, S., Wang, X., Pham, P.T., Goodman, M.F., Bergman, A., Scharff, M.D., and MacCarthy, T. (2015). Overlapping hotspots in CDRs are critical sites for V region diversification. Proc Natl Acad Sci U S A *112*, E728-737.

West, A.P., Jr., Scharf, L., Scheid, J.F., Klein, F., Bjorkman, P.J., and Nussenzweig, M.C. (2014). Structural insights on the role of antibodies in HIV-1 vaccine and therapy. Cell *156*, 633-648.

Wheatley, A.K., Whittle, J.R., Lingwood, D., Kanekiyo, M., Yassine, H.M., Ma, S.S., Narpala, S.R., Prabhakaran, M.S., Matus-Nicodemos, R.A., Bailer, R.T.*, et al.* (2015). H5N1 vaccine-elicited memory B cells are genetically constrained by the IGHV locus in the recognition of a neutralizing epitope in the hemagglutinin stem. J Immunol *195*, 602-610.

WHO Global Influenza Programme.

WHO (2022a). Ebola Virus Disease.

WHO (2022b). Middle East respiratory syndrome coronavirus (MERS-CoV).

Wibmer, C.K., Ayres, F., Hermanus, T., Madzivhandila, M., Kgagudi, P., Oosthuysen, B., Lambson, B.E., de Oliveira, T., Vermeulen, M., van der Berg, K.*, et al.* (2021). SARS-CoV-2 501Y.V2 escapes neutralization by South African COVID-19 donor plasma. Nat Med *27*, 622-625.

Wiehe, K., Bradley, T., Meyerhoff, R.R., Hart, C., Williams, W.B., Easterhoff, D., Faison, W.J., Kepler, T.B., Saunders, K.O., Alam, S.M.*, et al.* (2018). Functional Relevance of Improbable Antibody Mutations for HIV Broadly Neutralizing Antibody Development. Cell Host Microbe *23*, 759-765 e756.

Williams, W.B., Liao, H.X., Moody, M.A., Kepler, T.B., Alam, S.M., Gao, F., Wiehe, K., Trama, A.M., Jones, K., Zhang, R.*, et al.* (2015). HIV-1 VACCINES. Diversion of HIV-1 vaccine-induced immunity by gp41-microbiota cross-reactive antibodies. Science *349*, aab1253.

Williamson, L.E., Flyak, A.I., Kose, N., Bombardi, R., Branchizio, A., Reddy, S., Davidson, E., Doranz, B.J., Fusco, M.L., Saphire, E.O.*, et al.* (2019). Early human B cell response to ebola virus in four U.S. survivors of infection. J Virol *93*, e01439-01418.

Winkler, E.S., Bailey, A.L., Kafai, N.M., Nair, S., McCune, B.T., Yu, J., Fox, J.M., Chen, R.E., Earnest, J.T., Keeler, S.P.*, et al.* (2020). SARS-CoV-2 infection of human ACE2-transgenic mice causes severe lung inflammation and impaired function. Nat Immunol *21*, 1327-1335.

Wong, G., Audet, J., Fernando, L., Fausther-Bovendo, H., Alimonti, J.B., Kobinger, G.P., and Qiu, X. (2014). Immunization with vesicular stomatitis virus vaccine expressing the Ebola glycoprotein provides sustained long-term protection in rodents. Vaccine *32*, 5722-5729.

Wrammert, J., Koutsonanos, D., Li, G.M., Edupuganti, S., Sui, J., Morrissey, M., McCausland, M., Skountzou, I., Hornig, M., Lipkin, W.I.*, et al.* (2011). Broadly cross-reactive antibodies dominate the human B cell response against 2009 pandemic H1N1 influenza virus infection. J Exp Med *208*, 181-193.

Wu, N.C., and Wilson, I.A. (2020). Structural Biology of Influenza Hemagglutinin: An Amaranthine Adventure. Viruses *12*.

Wu, N.C., Yuan, M., Bangaru, S., Huang, D., Zhu, X., Lee, C.D., Turner, H.L., Peng, L., Yang, L., Burton, D.R.*, et al.* (2020). A natural mutation between SARS-CoV-2 and SARS-CoV determines neutralization by a cross-reactive antibody. PLoS Pathog *16*, e1009089.

Wu, X., Zhou, T., Zhu, J., Zhang, B., Georgiev, I., Wang, C., Chen, X., Longo, N.S., Louder, M., McKee, K.*, et al.* (2011). Focused evolution of HIV-1 neutralizing antibodies revealed by structures and deep sequencing. Science *333*, 1593-1602.

Yaari, G., and Kleinstein, S.H. (2015). Practical guidelines for B-cell receptor repertoire sequencing analysis. Genome Med *7*, 121.

Ye, J., Ma, N., Madden, T.L., and Ostell, J.M. (2013). IgBLAST: an immunoglobulin variable domain sequence analysis tool. Nucleic Acids Res *41*, W34-40.

Yermanos, A.D., Dounas, A.K., Stadler, T., Oxenius, A., and Reddy, S.T. (2018). Tracing Antibody Repertoire Evolution by Systems Phylogeny. Front Immunol *9*, 2149.

Yu, L., and Guan, Y. (2014). Immunologic Basis for Long HCDR3s in Broadly Neutralizing Antibodies Against HIV-1. Front Immunol *5*, 250.

Yuan, M., Huang, D., Lee, C.D., Wu, N.C., Jackson, A.M., Zhu, X., Liu, H., Peng, L., van Gils, M.J., Sanders, R.W.*, et al.* (2021). Structural and functional ramifications of antigenic drift in recent SARS-CoV-2 variants. Science *373*, 818-823.

Yuan, M., Liu, H., Wu, N.C., Lee, C.D., Zhu, X., Zhao, F., Huang, D., Yu, W., Hua, Y., Tien, H.*, et al.* (2020a). Structural basis of a shared antibody response to SARS-CoV-2. Science *369*, 1119-1123.

Yuan, M., Wu, N.C., Zhu, X., Lee, C.D., So, R.T.Y., Lv, H., Mok, C.K.P., and Wilson, I.A. (2020b). A highly conserved cryptic epitope in the receptor binding domains of SARS-CoV-2 and SARS-CoV. Science *368*, 630-633.

Zang, R., Gomez Castro, M.F., McCune, B.T., Zeng, Q., Rothlauf, P.W., Sonnek, N.M., Liu, Z., Brulois, K.F., Wang, X., Greenberg, H.B.*, et al.* (2020). TMPRSS2 and TMPRSS4 promote SARS-CoV-2 infection of human small intestinal enterocytes. Sci Immunol *5*.

Zhang, Q., Xiang, R., Huo, S., Zhou, Y., Jiang, S., Wang, Q., and Yu, F. (2021). Molecular mechanism of interaction between SARS-CoV-2 and host cells and interventional therapy. Signal Transduct Target Ther *6*, 233.

Zheng, J., Wong, L.R., Li, K., Verma, A.K., Ortiz, M.E., Wohlford-Lenane, C., Leidinger, M.R., Knudson, C.M., Meyerholz, D.K., McCray, P.B., Jr.*, et al.* (2021). COVID-19 treatments and pathogenesis including anosmia in K18-hACE2 mice. Nature *589*, 603-607.

Zhou, J.Q., and Kleinstein, S.H. (2019). Cutting Edge: Ig H Chains Are Sufficient to Determine Most B Cell Clonal Relationships. J Immunol *203*, 1687-1692.

Zhou, T., Lynch, R.M., Chen, L., Acharya, P., Wu, X., Doria-Rose, N.A., Joyce, M.G., Lingwood, D., Soto, C., Bailer, R.T.*, et al.* (2015). Structural repertoire of HIV-1-neutralizing antibodies targeting the CD4 supersite in 14 donors. Cell *161*, 1280-1292.

Zost, S.J., Dong, J., Gilchuk, I.M., Gilchuk, P., Thornburg, N.J., Bangaru, S., Kose, N., Finn, J.A., Bombardi, R., Soto, C.*, et al.* (2021a). Canonical features of human antibodies recognizing the influenza hemagglutinin trimer interface. J Clin Invest *131*.

Zost, S.J., Dong, J., Gilchuk, I.M., Gilchuk, P., Thornburg, N.J., Bangaru, S., Kose, N., Finn, J.A., Bombardi, R., Soto, C.*, et al.* (2021b). Canonical features of human antibodies recognizing the influenza hemagglutinin trimer interface. J Clin Invest *131*, e146791.

Zost, S.J., Gilchuk, P., Case, J.B., Binshtein, E., Chen, R.E., Nkolola, J.P., Schafer, A., Reidy, J.X., Trivette, A., Nargi, R.S.*, et al.* (2020a). Potently neutralizing and protective human antibodies against SARS-CoV-2. Nature *584*, 443-449.

Zost, S.J., Gilchuk, P., Chen, R.E., Case, J.B., Reidy, J.X., Trivette, A., Nargi, R.S., Sutton, R.E., Suryadevara, N., Chen, E.C.*, et al.* (2020b). Rapid isolation and profiling of a diverse panel of human monoclonal antibodies targeting the SARS-CoV-2 spike protein. Nature Medicine *26*, 1422-1427.

Zost, S.J., Gilchuk, P., Chen, R.E., Case, J.B., Reidy, J.X., Trivette, A., Nargi, R.S., Sutton, R.E., Suryadevara, N., Chen, E.C.*, et al.* (2020c). Rapid isolation and profiling of a diverse panel of human monoclonal antibodies targeting the SARS-CoV-2 spike protein. Nature Medicine.

Zost, S.J., Wu, N.C., Hensley, S.E., and Wilson, I.A. (2019). Immunodominance and Antigenic Variation of Influenza Virus Hemagglutinin: Implications for Design of Universal Vaccine Immunogens. J Infect Dis *219*, S38-S45.