THE UNIVERSITY OF CHICAGO

PROFILING B CELL RESPONSES TO A NOVEL CORONAVIRUS AND INVESTIGATING THE EVOLUTIONARY HISTORY OF THE B CELL RECEPTOR

A DISSERTATION SUBMITTED TO

THE FACULTY OF THE DIVISION OF THE BIOLOGICAL SCIENCES

AND THE PRITZKER SCHOOL OF MEDICINE

IN CANDIDACY FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

COMMITTEE ON IMMUNOLOGY

BY

CHRISTOPHER THOMAS STAMPER

CHICAGO, ILLINOIS

JUNE 2021

"Nature is always more subtle, more intricate, more elegant than what we are able to imagine."

Carl Sagan

Table of Contents

LIST OF FIGURES ..... **VII**

LIST OF TABLES ..... **IX**

ACKNOWLEDGEMENTS ..... **X**

ABSTRACT ..... **XVI**

CHAPTER I: INTRODUCTION **1**

1.2 The Humoral Immune Response **1**
1.2.1 Overview 1
1.2.2 B Cell Development and V(D)J Recombination 3
1.2.3 Antibody Responses 6
1.2.4 Naïve B cells 10
1.2.5 Memory B cells 11
1.2.6 Antibody Secreting Cells 13
1.2.7 Innate-like B cells 15
1.3 The Germinal Center **19**
1.3.1 Overview 20
1.3.2 Affinity Maturation 21
1.3.3 Somatic Hypermutation 23
1.4 Severe Acute Respiratory Syndrome Coronavirus 2 ..... **28**
1.4.1 Overview 28
1.4.2 Coronaviruses in Humans 29
1.4.3 COVID-19 31
1.4.4 The Immune System and SARS-CoV-2 35
1.4.5 Testing, Diagnostics and Therapeutics 36
1.5 High throughput single cell analysis methods **40**
1.6 Summary **41**

CHAPTER II: MATERIALS AND METHODS **43**

2.1 Peripheral Blood Mononuclear Cell Isolation **43**
2.2 Oligo-tagged antigen Probe Generation **44**
2.3 Fluorescence Activated Cell Sorting **45**
2.4 Single Cell RNA Sequencing **46**
2.5 Linq-View for Data Processing **47**
2.6 Monoclonal Antibody Cloning and Expression **49**
2.7 Enzyme Linked Immuno Absorbant Assay **50**
2.8 Polyreactivity ELISA **51**
2.9 Memory B Cell Stimulations and enzyme-linked immunospot assays (ELISpot) **51**

2. 10 Viral Plaque Assay - *In-vitro* neutrilization 52
2.11 *In-vivo* Infection Model in Hamsters 53
2.12 *In-vivo* Protection Model in hACE2 Mice 53
2.13 Mutational and BCR Analysis 54

CHAPTER III: PROFILING B CELL IMMUNODOMINANCE UPON SARS-COV-2 INFECTION REVEALS ANTIBODY EVOLUTION TO NON-NEUTRALIZING VIRAL TARGETS 55

3.1 Summary 55
3.2 Introduction 56
3.3 Single Cell RNA Sequencing of Endemic HCoV and SARS-CoV-2 Specific B Cells 58
3.4 The SARS-CoV-2-specific B cell landscape is defined by naïve-like and memory B cell subsets 64
3.5 Evolution of the B Cell Response to Endemic HCoV and SARS-CoV-2 Protein 69
3.5.1 B cell immunodominance and adaptability to SARS-CoV-2 and HCoVs changes with time after infection 69
3.5.2 Memory reactivity to intracellular SARS2 antigens increases over time 74
3.5.3 Memory by probe sorting does not correlate with serum titers 76
3.6 SARS-CoV-2 B Cells Display Unique Repertoire Features and Protective Abilities 78
3.6.1 B Cell Variable Gene Usage and Enrichment by Antigen 78
3.6.2 Neutralization Capacity and In-vivo Protective Ability of the Antibodies 82
3.7 SARS2 Antigen Specificity and B Cell Subset Distribution is Linked to Clinical Features 85
3.8 Novel Public Clones 91
3.9 Additional Analysis of Varibale Gene Usage by Antigen 93
3.10 Mutational Analysis Can Reveal Previously Unappreciated Signatures Within Large B Cell Datasets 95
3.11 Mutation Patterns Reveal Endemic HCoV Reactive B cells Are Highly Mutated in Unusual Patterns 100
3.12 Chapter Discussion 106
3.13 Chapter Acknowledgements 108
3.14 Funding 109
3.15 Declaration of Interests 110

CHAPTER IV: IMPROVED METHODS OF ANALYZING MULTI-OMICS SINGLE CELL DATA 111

4.1 LinQ-View combines single-cell mRNA and protein expression profiling data and enables multi-modal downstream analysis 114
4.2 LinQ-View identifies unique B cell subsets from COVID-19-infected patients 116
4.3 Discussion of LinQ-Seq 120

CHAPTER V: EXPLORING THE FUNCTIONAL IMPACT OF ACTIVATION INDUCED CYTIDINE DEAMINASE TARGETING 123

5.1 Introduction **123**
5.2 Characteristics of the monoclonal antibody 045-051310-2B06 ..... **125**
5.3 Considerations in the design of an AID untargeted BCR **127**
5.4 Generation of the mice **134**
5.5 Chapter Five Conclusions **137**

CHAPTER VI: DISCUSSION **139**

6.1 Overview **139**
6.2 The Immune Landscape During SARS-CoV-2 Infection and the Importance of Vaccination for Long Term Immunity to Protective Targets **141**
6.3 The application of oligo-tagged antigen bait sorting to other diseases **143**
6.4 The heterogeneity of B cell subsets **145**
6.5 The impact of endemic HCoV infection on COVID-19 and universal coronavirus vaccines. **146**
6.6 Rapid production of monoclonal antibodies to known antigens are a powerful tool for additional discovery **148**
6.7 The evolutionary history of AID and mutational targeting **150**
6.8 Summary of future directions and issues **154**
6.9 Final Conclusions **155**

TABLES **157**

REFERENCES **182**

List of Figures

Figure 1.1: Model of a Germinal Center Reaction……19

Figure 3.1: SARS-CoV-2-specific B cells comprise multiple distinct clusters……..58

Figure 3.2: B cell subset distribution and antigen specificity by subject……..62

Figure 3.3: B cell receptor and transcriptional analysis reveals cluster identities…….65

Figure 3.4: Expression maps of select genes…….68

Figure 3.5: B cell immunodominance and adaptability landscapes vary in acute infection in convalescence……70

Figure 3.6: Further analysis of antigen-specific B cell properties across distinct cohorts and timepoints……...73

Figure 3.7: Correlation between antigen-probe positive B cells and serum titers…….76

Figure 3.8: B cells targeting distinct antigens display unique variable gene usages…….79

Figure 3.9: Neutralization capacity and *in vivo* protective ability of mAbs to the SARS-CoV-2 spike and intracellular proteins……..82

Figure 3.10: Antigen-specificity and B cell subset distribution is linked to clinical features…...86

Figure 3.11: Additional analyses of antigen reactivity by clinical parameter……...90

Figure 3.12: Frequency of Vgene usage by antigen reactivity……..93

Figure 3.13: Mutational frequency by Vgene shows substantial maturation of HCoV S reactive BCRs…….96

Figure 3.14: Mutational frequency by Vgene reveal differences across intraviral antigens……..99

Figure 3.15: Endemic HCoV reactive cells have high levels of mutation in CDRs and FWRs and show evidence of strong antigenic selection……101

Figure 3.16: HCoV and SARS2 spike reactive B cells show varying amounts of selection pressure depending on the category of replacement mutation……105

Figure 4.1: Kernel process of the LinQ-View toolkit……..114

Figure 4.2: LinQ-View analysis identifies unique B-cell subsets from COVID-19 cohorts….117

Figure 5.1: Comparison of Germline and ‘De-evolved’ 045-051310-2B06 Heavy Chains…..128

Figure 5.2: Amino acid class usage patterns in computationally mutated heavy chain sequences……132

Figure 5.3: Approach to Generate and Characterize the BCR Knock-In Mice……..134

Figure 6.1: The immune landscape during infection with, and after recovery from, SARS-CoV-2……...141

Figure 6.2: CDR3 amino acid residue properties in all antigen reactive naive or effector cells.....150

Figure 6.3: FWR3 amino acid residue properties in all antigen reactive naive or effector cells……..153

List of Tables

Table 3.1: Public Clones……...91

Table 3.2: Convalescent patient information……...157

Table 3.3: Distribution of clinical parameters for convalescent patients included in the study..161

Table 3.4: Severe acute patient information……162

Table 3.5: Distribution of clinical parameters for severe acute patients included in the study...163

Table 3.6: Top 10 most differentially expressed genes per transcriptional cluster…….164

Table 3.7: Key genes used in the identification of B cell subsets……171

Table 3.8: Specific mAbs generated from single B cell heavy and light chain gene sequences…….176

Acknowledgements

Were I to thank everyone who made this work possible, it would surely stretch to an untenable length that rivals the rest of the document itself. Therefore, I must limit myself in who I list, but would like to dedicate some space here to all others who will not be mentioned but must be thanked.

First, I would like to thank my mentor and advisor, Patrick Wilson. His lab has been my home for nearly seven years and my time here has surely changed me forever. From the earliest days of my rotation in the lab it was obvious Patrick had a great passion for B cells, antibodies, and human immunology and how a better understanding of them could be applied to human disease. However, he still allowed me to pursue a project rooted fully in mouse models and basic questions of evolution. This project proved largely unsuccessful but taught me an incredible amount about the art of science. I also spent time working on a project related to universal influenza vaccines to see most of the data produced by it end up being negative. Patrick remained supportive through it all and my interactions with him have shown me how to be a better scientist, how to run an effective lab, and how to make great discoveries. I can only hope that in my long time in the lab it has gained something from me, as I have gained so much from it.

Next, I would like to thank all the members of my committee; Pete Savage, Albert Bendelac, Sarah Cobey, and Anne Sperling. You have all provided invaluable feedback and advice at committee meetings and WIPs and all of my work is better for it and I am a better scientist because of it. I would also like to specially recognize my chair Pete Savage, who has always been someone I can go to for advice, thank you for everything.

Several key collaborations made the COVID work possible. There would have been no project were it not for Lucia Madariaga and Kumaran Shanmugarajah, thank you for trusting our lab to handle such precious and valuable samples, and for sometimes fishing bags of blood out of trash cans for us. I must also thank all of the donors, who agreed to give their plasma to help others and consented for us to study their B cells, everything was only possible because of you. To the rest of the convalescent plasma project team, thank you for all of your hard work during even harder times.

In my time here I have had the good fortune to serve many years on Graduate Council in a variety of elected positions and on nearly endless committees. I enjoyed every second of it, even when conscience dictated that I take stands contrary to others and disagree. Through my years on GC I was able to meet a large number of graduate students from outside of the Biological Sciences Division. To all of you, I am so happy I had the good fortune for us to meet and to learn about your work on campus. I leave the University of Chicago fully convinced that its graduate students are the beating heart of the university and the source of its much-lauded life of the mind, may all your work one day be recognized.

Lab sciences can often involve long days and longer nights. For keeping me entertained in many monotonous hours at the bench I must thank several podcasters. To Arnie, Chunt, and Usidore- thank you for introducing me to the magical land of Foon and to podcasts in general, you helped me survive many midnight timepoints in Carlson. To the Clue Crew of Hey Riddle Riddle, you have kept me sane through many hard times, especially in the last year, and made my lambastes think I'm insane with how often you make me laugh out loud. Lastly, to the McElroys, thank you for always bringing joy to my days and serving as an inspiration and a reminder that Appalachians can do great things too.

To all the students of the Committee on Immunology, thank you for making this such an enriching and amazing environment to be a part of. It is inspirational to be surrounded by so many amazing scientists and people, I hope you all go on to achieve all the great things you deserve.

To the Wilson Lab, it has been the joy of a lifetime to meet and work with so many of you. I must specifically mention Carole Henry, one of the best scientists I have ever had the pleasure of working with and who taught me so much. To Jenna, who has also been a great friend and teacher, I know you will do amazing science in your new faculty position. To Karlynn, Denise, Noel, and Kaval who were there to welcome me into the lab when I was young and naive and had no idea what I was doing. To Ying, who has always been an invaluable source of advice and guidance. To Dustin, who always made me feel welcome when I rotated in the Jabri lab and has been a friend since. To Angela, Gloria, Greg, and Louise; your hard work makes it possible for us to do our science and I am eternally grateful for it. Your kindness and friendliness also make the days pass faster and make bad lab days not so bad, I will never be able to fairly repay you all.

I must specifically thank Lei Li and Haley Dugan, my co-leaders on the COID work. You are both phenomenal scientists and wonderful friends. We accomplished so much in such a short amount of time, and it has been the honor of a lifetime to work with you. I hope you both achieve everything you desire in science and life; you fully deserve it.

To Kipp, Joel, and Joe, thank you for being three of the most amazing friends someone could ask for, for all of these years. I am convinced I would have never even made it to graduate school without your support and companionship.

I fear I have made too many amazing and wonderful friends here to list them all. To Brendan and Lydia, thank you for all the great times and games, for all the delicious pizza, the too strong

cocktails, and insanely spicy chili. To Shan, thanks for having put up with me as a roommate for four years and for being such a great friend, you are one of the best and brightest scientists I have met and living with you has been a pleasure. When are we starting our immunology and DnD podcast? To Becca, thanks for always being the funniest person in the chat, all the great entertainment recommendations, and generally being amazing. To Liana, thanks for all the great times, all the cat chats, and always providing great conversation. To Alex, thanks for being one half of the best trivia duo in Chicago, for introducing me to Improv, and for being such a dear friend. To the Sensational Six and Matt, I feel as if I have been the worst member of the greatest cohort in BSD history, thank you for all of the amazing times. To Jaime, thanks for all the fun on DC together, all the Chicago sporting events, and somehow always being able to convince me that it is a good time to leave lab for a coffee, beer, or both. To Jess, thanks for always being down for a shot of Malört and for being such a great friend. To Steven, thanks for the introductions to so much good metal and for your endless generosity and patience in helping me so often with the generation of mice and other lab techniques. I was not lying when I said at your defense that I would have probably mastered out years ago if not for all of your assistance. As repayment, I preemptively extend an invitation for you to stay with me in Sweden when you travel to Stockholm to collect your future Nobel prize. Last, to Ryan, meeting you has made me a better person in innumerable ways. You have introduced me to an endless variety of new things, taught me so much, and have made me want to always strive to improve myself. I would remark on your intellectual and scientific abilities, but your kindness and empathy are all the more remarkable. I am who I am today because of your friendship, and I can only hope that I have been able to return one fraction of one percent of what I have taken from it.

To the research animals, words can never repay what we have taken from you nor convey the gratitude owed to you. I promise to always strive to make sure what we have learned from you has meaning and purpose.

Till min framtida svenska familj. Tack för all värme och vänlighet och ni alltid fått mig att känna mig välkommen. Jag har aldrig behövt oroa mig för min kommande invandring, för ni har alla varit så underbara mot mig. Jag ser fram emot många glada tillfällen tillsammans i framtiden.

To my family, I would not be who I am today without your endless love and support. To my siblings Derrick, Jon, and Laura; you have always made my life better, and I am convinced all three of you have already improved this world in so many ways and far more than I ever will. To Laura, specifically, you have been one of my best friends all my life and have helped me through so much, I do not think you even realize. I have always been impressed by your kind heart and keen mind, by your boundless energy, outgoing spirit, and your bravery. While I studied COVID-19 from the safety of a laboratory you battled it as a new resident in an Emergency Room, I hope some of our work can help prevent you and your colleagues from ever being so beleaguered again. To my parents, Jack and Jayne Stamper, I find words inadequate to thank you for all you have done. You have sacrificed so much for me, and your love and support has enabled me to achieve all I have. Your endless warmth, kindness and generosity with everyone you meet is my model of how to be towards others and it has helped me in this process more than you can know. Thank you both, endlessly.

Till Anna-Karin. Att träffa dig har utan tvekan varit det bästa som har hänt mig i livet och under forskarutbildningen. Den sista halvan av mitt avhandlingsarbete har haft sina högsta toppar och djupaste dalar, och jag är otroligt lycklig att ha haft dig vid min sida för att ta mig genom

dem. För att fira det goda och trösta mig efter det dåliga. Att möta dig har gjort musprojektet till det mest framgångsrika projektet någonsin. Din kärlek och ditt stöd.

Abstract

B cells are the central component of the humoral immune system and critical for long term immunity against all forms of pathogens. Antibodies represent one of the earliest lines of defense within the adaptive memory response and are the major effectors of vaccine mediated protection. By applying an oligo-linked bait sorting approach and single cell sequencing to investigate infection by a novel pathogen, severe acute respiratory coronavirus 2, we were able to elucidate that there was considerable heterogeneity in the B cell response. We found that while protective anti-spike protein responses were readily fixed into B cell memory after recovery from infection, memory towards the intracellular and intraviral viral ORF8 and NP proteins increased in relative abundance and maturation over time and found antibodies targeting these antigens to be exclusively non-protective. Additionally, we found that memory to endemic human coronaviruses is activated upon infection with SARS-CoV-2, and that the B cell receptors of these cells show remarkably high amounts of mutation and selective pressure, potentially revealing a rich source to be targeted for future universal coronavirus vaccine efforts. In exploring these data we have developed new computational tools to analyze high throughput, multi-omics data sets and show that it reveals further heterogeneity in B cells responding to SARS-CoV-2 than RNA sequencing alone. Further, our dataset of 13,000 antigen specific B cells with matched transcriptome and receptor sequence and 55,000 B cells total with receptor and transcriptome will provide a foundation for future studies. Ultimately, we have built new tools to investigate B cell responses and shown them capable of generating discoveries of immediate translational impact and of revealing previously unappreciated aspects of basic B cell biology.

Chapter I: Introduction

1.2 The Humoral Immune Response

1.2.1 Overview

In all the abundance of life in its endless forms most beautiful it is inevitable that various organisms will come into contact and conflict with one another. Therefore, most organisms have developed systems of varying complexities and intricacies that allow them to determine the danger, if any, posed by one another and how to appropriately respond. As multicellular life evolved from single-celled forbears, organisms increased in complexity and size their forms and organization. This was followed by a resultant increase in the numbers of smaller organisms that could live in, on, and about them which necessitated ever more complex systems to respond to and monitor them. Additionally, as organisms increase in complexity and numbers of specialized tissue and organ systems they are increasingly faced with the possibility of aberrant cellular multiplication and activation, which necessitates systems to monitor and maintain all others in a manner that is most conducive to successfully surviving to the age of reproduction. Approximately 500 million years ago, arising within vertebrates, a vast and complex system began to take shape that links multiple cells, tissues, and organs and organizes them to promote and protect the bodily homeostasis and wellbeing of the host.[1] This is the immune system, a fundamental requirement for the survival of vertebrate life.

The modern vertebrate immune system can be broadly split into two major arms, the innate and adaptive response, however it is increasingly apparent there is an area of broad overlap between them. The innate arm of the immune system is capable of responding to and resolving the majority of threats faced. The innate system is composed of various cell types including granulocytes, dendritic cells, macrophages, mast cells, natural killer cells, and a variety

of specialized epithelial and stromal cells that are found within lymphatic tissue or at mucosal surfaces. In addition to its cellular components the innate response also contains the complement system and anatomical barriers such as skin, mucosal surfaces, tears, and the blood brain barrier. The innate system largely functions by preventing entry or through recognition of conserved molecular patterns that mark something as non-self. Critically, long term and sustained adaptive immune responses require an initial innate immune activation.[2]

The adaptive immune response is the arm of the immune system responsible for generating responses to specific pathogens and developing memory towards them. It consists of two hematopoietic cell lines, T cells and B cells. T cells mature in the thymus and are responsible for cellular immunity, or the targeting and destruction of infected or unhealthy cells, and helper T cells which activate all other components of the adapt response. B cells are responsible for the adaptive humoral immune response, in addition to occasionally acting as antigen presenting cells (APCs).[3, 4] Both B and T cells generate their broad diversity of receptors through the process of V(D)J recombination, which allows the immune system to be ready to respond to nearly anything, including novel pathogens.[5]

B cells can be divided into two major developmental pathways, B1 and B2. B1 cells are innate-like B cells that develop from hematopoietic stem cells (HSCs) that derive from bone marrow or the fetal liver. Mature B1 B cells are a self-renewing population that can differentiate into short-lived plasmablasts and secrete natural IgM antibodies, which tend to recognize broadly conserved molecular epitopes. B1 B cells use a restricted repertoire of variable genes with which they produce their B cell receptors (BCRs).[6]

The other major B cell division are B2 B cells, all of which derive from HSC precursors in the bone marrow, where they also develop and mature. B2 B cells can be further divided into

two major pathways- marginal zone B cells (MZBs) and follicular B cells (FOBs). Both are named for where they reside, MZBs in the marginal zone of the spleen and lymph nodes and FOBs in the follicles of lymph nodes and other lymphatic sites such as the gastrointestinal associated lymphoid tissue (GALT) and bronchial-alveolar lymphoid tissue (BALT). All three types of B cells can also be found circulating in blood, to perform surveillance or to traffic from one site to another. FOBs are the major source of memory B cells and long-lived plasma cells, making them the key for the development of long term humoral immune memory. Follicular B cells are the most studied and best characterized of all B cell populations, especially in humans, with much more robust phenotypes and markers associated with them than the two more innate-like cell populations.[7]

1.2.2 B Cell Development and V(D)J Recombination

B cells are hematopoietic cells that mature from a common bone marrow precursor. They pass through several stages in their development; after beginning as a stem cell they progress to the pro-B cell stage, from there they enter the pre-B cell stage before continuing onto the immature B cell stage, and finally arriving at the end as mature B cells. Many developing cells will fail along the way for a variety of reasons, and this process is evolved to try to ensure that only healthy B cells with a stable, non-self-reactive BCR reach the mature B cells stage.[8, 9]

Central to the process of maturation is the formation and testing of the BCR. BCRs are generated randomly through the process of V(D)J recombination. All human B cell receptors possess and heavy and a light chain. The heavy chain is derived from the recombination of three gene segments: one variable (V), one diversity (D) and one joining (J) segment. The light chain is made of only a variable (V) and joining (J) segment.[5] There are approximately 56 functional

heavy chain V genes, 23 D genes, and 6 J genes. For light chains there are two separate sets of V and J genes that produce two separate sets of light chains, kappa(ϰ) and lambda(ƛ). There are approximately 40 functional Vk genes and five Jk and 30 functional Vl and four Jl. The recombination of these genes to produce immune receptors is mediated by the proteins RAG1 and RAG2.[10, 11]

To ensure that recombination takes place with genes in the proper order with one another and the constant region the RAG proteins recognize genomic sequences known as recombination signal sequences (RSS). The RSS consists of a conserved heptamer, followed by a non-conserved stretch of nucleotides that is either 12 or 23 basepairs long, which is itself followed by a conserved nonamer sequence. The spacer length corresponds to one or two turns of a DNA double helix, to always guarantee the heptamer and nonamer are aligned on one side. The recombination complex generally can only link a segment flanked by a 12 bp spacer with one flanked by 23 bp, this is known as the 12/23 rule and is evolved promote proper gene recombination, in effect it means that for a heavy chain it is possible to link a V and D gene or a D and J gene, but not a V and J gene directly.[10] However, it has now been shown that some BCRs violate this rule by producing D-D linkages.[12] The process of VDJ recombination has also been selected to be an imperfect process, sometimes deleting or inserting additional nucleotides at junction sites. This junctional diversity allows for greater overall diversity BCRs that can be produced. Finally, further increasing diversity, BCRs have been discovered which seem to be generated using genetic segments from elsewhere in the genome. All of these different mechanisms combine to produce a potential pool of approximately $3.5x10^{10}$ or $3.5x10^{11}$ possible BCR combinations.[10]

However, many of these combinations can prove either unstable or self-reactive. Heavy chain V(D)J recombination occurs during the pro-B cell stage then, using a surrogate light chain, the B cell only survives the early pre-B cell stage if it can signal through its BCR. If it successfully does so, it then rearranges the light chain VJ segments in the later pre-B stages. First a kappa light chain is produced and the BCR tested for signaling, if it is unsuccessful the B cell will then rearrange a lambda chain. If both are unsuccessful it will undergo apoptosis, however if a stable B cell is successfully produced the B cell will progress to the immature B cell stage where its BCR will be checked for self-reactivity, an autoreactive B cell using a kappa light chain will also try to rearrange a new lambda light chain to check if it is tolerant of self. If the B cell remains self-reactive, ideally this results in apoptosis.[9, 13]

The final pool of potential BCRs is definitively smaller than would be mathematically predicted. Some combinations are lost to tolerance against self-recognition while others are incapable of stably expressing at the surface of the B cell. Beyond these losses, large scale sequencing studies in mice and humans have confirmed that V,D, and J gene usage is not fully random and that some genes are used at much higher frequencies than others. The members of the VH3 family are used in nearly one half of all mature, peripheral B cells with VH3-23 being the most common. VH4 genes, especially VH4-34 and VH4-39 are also incredibly common.[14] Even with these biases, the full B cell repertoire is still comprised of millions of combinations that can effectively respond to anything the immune system may encounter allowing B cells to mount an effective and often protective antibody response.

1.2.3 Antibody Responses

Antibodies are the secreted form of a BCR and are the core component of the adaptive humoral response. Their structure, recognition of molecular patterns, and multiple isotypes allow them to be active in nearly any environment within an organism. Additionally, the large variety of effector functions they can perform or mediate make them capable of targeting any type of pathogen, from the smallest of viruses to multicellular parasites such as worms.[7, 10] This is why antibodies are one of the most effective immune molecules, and why synthetically produced or cloned antibodies represent an increasingly large percentage of the pharmaceutical and therapeutics industry.[15]

Antibodies, being secreted BCRs, are formed of a heavy and a light chain which can be further split into variable and constant regions. Each is formed of two copies of the same heavy and light chain combination. The variable region is responsible for antigen recognition and binding, and is comprised of four framework regions (FWRs) and three complementarity determining regions (CDRs). Structurally, the CDRs represent the loops that are directly responsible for forming the bonds with target antigens allow the antibodies to bind. The FWRs provide the structural support and stability for the CDRs to interact with other molecules and form those bonds. CDRs 1 and 2 are derived from the heavy or light V gene. CDR3 represent the greatest site of diversity within an antibodies variable region, as it is formed where the V, D, and J genes join together for the heavy chain or the V and J genes for the light chain.[10]

The constant region is so named as it is somatically coded for and remains constant for each antibody, apart from the five distinct isotypes; IgM, IgD, IgE, IgA, and IgG.[10] The constant region also provides stability and structure, and is responsible for many of the antibodies indirect

protective functions such as antibody dependent cellular cytotoxicity (ADCC), antibody-dependent cellular phagocytosis (ADCP), and complement-dependent cytotoxicity (CDC).[16, 17] Additionally signaling through Fc receptors can modulate overall immune responses, and recent work shows this can be influenced by the patterns of glycosylation on the constant regions of antibodies.[18] Antibodies have even been shown to have intracellular immune functions, related to the interactions of constant regions and the intracellular receptor TRIM21.[19] The five different isotypes of antibodies is dependent upon which constant region the heavy chain is using and are specialized for certain immune responses or sites of activity.[7, 10]

All B cells start with IgM as the isotype of their BCR and antibodies. IgM antibodies are secreted as a pentamer, which raises their avidity.[20, 21] The pentamer is created by linking five IgM antibodies together using a protein called the J chain.[22] IgM can be found in large amounts in the blood, but can also be transported across mucosal surfaces and be found in the mucus as secretory IgM. IgM probably plays a complementary role to IgA at mucosal surfaces while also acting as a layer of redundancy. Secretory IgM is also one of the immunoglobulins that can be found in human and murine breastmilk.[23] IgM is also capable of fixing complement onto pathogens it has bound and of neutralizing viruses by binding them and preventing cellular entry.[24] There is an Fc receptor, the FcμR that recognizes IgM and mediates various effector functions.[25] IgM is also the only antibody class produced by the innate-like B cells, B1s and MZBs, and often has distinct roles within the body. Some of these antibodies bind to conserved molecular patterns, similar to how a pattern recognition receptor functions, allowing them to target a broad range of pathogens. Additionally, both MZBs and especially B1s produce a class of IgM that recognizes and binds to cellular detritus and debris, these housekeeping antibodies

play a role in the maintenance of bodily homeostasis by marking the debris for removal by innate phagocytic cells.[6, 26, 27]

IgD is found on all mature, naïve follicular B cells and can sometimes be found on marginal zone and B1 B cells.[6, 28] IgD, as a class of BCR, is thought to play roles in signaling and development of B cells and maintenance of naive B cells. IgD as a class of secreted antibodies are little studied and poorly understood. Mice appear to lack any form of IgD antibodies, and in humans they seem to be largely confined to the lungs and respiratory tract where potentially unique subpopulations of B cells can secrete them. Here they help protect against respiratory infections through their binding of Fc receptors on innate cells, and especially mast cells, and driving inflammatory responses.[29] Much of the work concerning the role of IgD in respiratory infections has been performed specifically with regards to influenza infection, and the roles they play may differ for other infections.[30]

IgE is the rarest of all antibody isotypes, generally making up only 0.05% of all immunoglobulins present in adult human or murine serum.[31] IgE exists only within mammals, and is linked to immunity against multi-cellular parasites such as helminths and other worms. It has also been shown to play a role in immunity against protozoans such as those within the genus *Plasmodium*, which includes the causative agents of malaria.[32] IgE interacts with two Fc receptors, FcεRI and FcεRII, the 'strong' and 'weak' IgE receptors, respectively. These receptors are found on a variety of innate immune cells including eosinophils, basophils, mast cells, and dendritic cells.[33] Through binding its Fc receptors IgE is capable of driving stronger inflammatory responses than any other immunoglobulin, which has probably evolved because of their role in anti-parasitic immunity.[32] Improper binding of benign targets by IgE and the inflammatory response it produces is the major cause of type 1 hypersensitivy and the allergic

diseases related to them. More recent work indicates that IgE may play a critical role in the clearance of venoms and toxins, such as those from insect stings or snake bites.[34] Research also suggests IgE may be one of the components of the immune systems mechanisms for surveillance and early detection of cancerous cells, but this remains to be firmly established.[35]

IgA is the most abundantly produced protein on a daily basis, with humans producing upwards of 66mg/kg of body weight per day.[36] Much of this is in the form of secretory IgA at mucosal surfaces.[37, 38] IgA is secreted as a dimer, with two IgAs being linked by the J chain, the same protein responsible for the formation of the IgM pentamer.[22] Beyond the large amount produced and secreted at mucosal surfaces IgA is also the second most abundant immunoglobulin in serum, and altogether accounts for approximately 70% of an adult human's immunoglobulins. IgA is also found in tears, saliva, semen, and vaginal secretions and can cross the placental barrier.[36, 39] IgA is recognized by the Fc receptor FcαRI in addition to the asialoglycoprotein and transferrin receptors.[40] IgA is the first line of humoral defense, because of its large presence at mucosal sites. It can neutralize viruses and bacteria, preventing their entry through the epithelium and protecting against further infection. It has also been shown to bind and neutralize pathogen secreted toxins and effector molecules that assist with the cleavage of mucosal bonds and allow pathogenic entry through mucous secretions.[41] Interestingly, IgA deficiency causes no obvious major health defects in humans or mice. In fact, potentially as many as 3% of all caucasians are IgA deficient.[37] Rather than implying IgA is unimportant, the absence of negative effects is thought to represent a system of redundancy with secretory IgM fulfilling the role normally played by secretory IgA.

IgG is the most abundant immunoglobulin class in serum, making up approximately 75% of serum antibodies.[42] It is the smallest of all immunoglobulins, which allows it to easily pass

through the smallest blood and lymphatic vessels and perfuse into tissue. IgG binds the Fc receptors, FcγRI and FcRn, a unique neonatal Fc receptor that allows IgG to cross the placental barrier to provide immunity to developing fetuses and immediately after birth.[43] The highest affinity antibodies are often class switched to IgG, and it is one of the most important of all Ig subclasses. IgG deficiency in humans results in severe immunodeficiency disease, whereas inappropriate IgG binding is associated with type II and III hypersensitivity diseases and conditions.[44] IgG in humans can be subdivided into four subclasses, IgG1, 2, 3, and 4 that all bind differing Fc receptors with different affinities, giving them different patterns of immune activation and inflammatory properties. In mice the subclasses of IgG are IgG1, IgG2a, IgG2b, and IgG3. IgG can also fix all components of the classical pathway of the complement system and, along with IgE, has been shown to be capable of binding and neutralizing toxins.[42, 45, 46] Long term, robust IgG responses have been used as one of the core hallmarks of protective humoral immunity after infection or vaccination for decades.

1.2.4 Naïve B cells

After developing in the bone marrow and entering the periphery all follicular B cells start as mature, naïve B cells. Naïve B cells circulate thoughout the blood and lymphatic system surveilling for threats. Upon recognition of an antigen with their BCR and the second signal from a T cell, APC, or one of their own PRRs the naïve B cell activates. Upon activation a naïve B cell can begin dividing with many of its daughters differentiating into plasmablasts and immediately producing large amounts of low affinity IgM antibodies. Other daughters may differentiate into IgM memory B cells.[7]

Naïve B cells have been classically defined using a mix of surface markers. In humans these are the surface immunoglobulins IgM and IgD, the pan B cell marker CD19, and CD21, CD22, and CD23.[47] More recent microarray and sequencing work has also elucidated several genes that are strongly associated with naïve B cells in comparison to other cells, including *BACH2, ZBTB16, APBB2, SPRY1, TCL1A,* and *IKZF2*.[48-53]

Upon activation most naïve B cells will, rather than immediately differentiating into effectors or memory, enter into a germinal center response within a lymph node or other lymphatic tissue.[7] The study of naïve B cells in humans is largely complicated by the fact that in many infections they are difficult to sort, as the responses will be dominated by activated memory. Infection with a novel pathogen, such as SARS-CoV-2, presents an exciting possibility to better characterize naïve B cells in humans upon first exposure to a pathogen.

1.2.5 Memory B cells

Memory B cells are the descendants of previously activated naïve B cells and are one of two central components of long term humoral immune memory and are critical to overall immune memory in general. They can be either extrafollicular or follicular in nature, depending on whether they had been produced in a germinal center or not. Upon re-activation memory B cells can rapidly undergo clonal expansion to produce plasamblasts that are producing large amounts of high affinity antibodies, often IgG or IgA, though there are IgM memory B cells. In addition, some reactivated memory can reenter germinal centers to perform additional affinity maturation towards their cognate antigens, though how common this is or how frequently it occurs has recently been questioned.[7, 54]

Memory B cells are often defined with a mix of surface markers including IgG and IgA, or more simply by being defined as IgM-. They also will be positive for CD19, CD20, CD27, CD40, CD80, and PDL-2. They have moderate expression of CD35, so a simple gating strategy can be $CD19^{+}CD27^{Hi}CD45^{mid}$, with the addition of IgM^{-} if one is interested specifically in class switched memory B cells.[47, 55]

Genetically memory B cells have been defined by having differential expression of the genes *CD27, CD86, RASSF6, TOX, TRERF1, TRPV3, POU2AF1, RORA, TNFRSF13B, CD80, FCLR5, GDPD5, BAIAP3, TGM2,* and *MUC16*. Though *CD80* and *CD86* are both also differentially expressed in antibody-secreting cells, so are often used in combination with other differentially expressed genes between the two populations to determine cellular identities.[50, 53, 56-60]

Memory B cells are a core component of humoral memory, along with long lived plasma cells, a specialized class of antibody secreting cells. However, traditionally, plasma cells are much easier to evaluate through indirect methods, such as measurement of serum titer levels towards specific antigens. This has caused a shift towards the over-emphasis of long-lived plasma cells and relative de-emphasis of memory B cells when assessing the effectiveness of vaccines or long-term memory after infection. Memory B cells can be partially assessed through ELISPOT assays, but the polyclonal stimulation used in the procedure can produce results of varying quality and accuracy, and it requires a subjective determination of what constitutes a 'spot' which can impact precision, accuracy, and reproducibility.[61] Additionally, it has become recently appreciated that there is much more heterogeneity in memory B cells than previously realized.[7, 57, 62] The identities of these various memory populations and the roles they could play in infection remain to be fully resolved. Therefore, new techniques and technologies to better

characterize memory B cell responses will be critical to more fully and accurately understand long term humoral memory in humans.

1.2.6 Antibody Secreting Cells

Antibody secreting cells (ASCs) are activated B cells that are actively producing and secreting antibodies. ASC is an umbrella term that encompasses a wide variety B cells, at times including B1s and MZBs. Follicular ASCs can be defined into subgroupings, though the definitions remain inconsistent between humans and mice, and across labs within the same organism.[63] Plasmablast tends to refer to a recently differentiated naïve B cell that is actively secreting antibodies. Some papers, including all from our lab, define these as short-lived effector cells that are distinct from long lived plasma cells. Other work will more broadly refer to plasmablasts as any recently differentiated ASC, including recent germinal center emigrants that will become long lived plasma cells.[64]

Further complicating the matter is the propensity of follicular ASCs to shed many of their surface markers, especially long-lived plasma cells, making them more difficult to identify by techniques such as flow cytometry. Similar to memory B cells, they can be defined as IgM-, especially long-lived plasma cells with additional surface markers being CD27hi, CD38+, CD78+, CD138+, and CD319+ to varying degrees, with CD138 being a classic marker of long lived plasma cells. Notably, human long lived plasma cells are frequently CD19-, even though it is considered a pan B cell marker.[47]

Genetically, ASCs can be defined and subsetted by differentially expressed genes. A decrease in *BCL6* and a corresponding increase in *IRF4* is associated with long lived plasma cell fate, and knockout models have linked *TNFSR17* to PC fate decisions as well and is believed to

be the same in humans. The gene encoding for *CD38* is also used, along with proteins associated with endoplasmic reticulum stress and unfolded protein responses, *MANF* and *XBP1*, reflecting the massive amount of protein being produced by ASCs. *IL6R*, the receptor for IL-6, a cytokine that promotes differentiation into ASCs is another marker. Finally, the most reliable genetic marker is *PRDM1*, the transcription factor that controls and defines the entire lineage.[59, 65] Notably, the genetic markers for ASCs tend to be shared between both short-lived plasma blasts and long-lived plasma cells (LLPCs). This can make it difficult to determine between the two, though a temporal model of germinal center output has been propsed. [66] Therefore, it is possible for well-timed antigen bait sorting to potentially capture cells exiting germinal centers that are trafficking to mucosal sites or the bone marrow to become LLPCs, before they have had the opportunity to fully downregulate and lose their surface BCR expression.

ASCs are the B cells that are directly responsible for actively producing the follicular humoral immune response. As such they are critically important, both in pathogen clearance and protection against future infection. Long lived plasma cells traffic into the bone marrow or mucosal surfaces, such as the respiratory or GI tracts, and can secrete millions of antibodies per second.[7] These cells are termed long lived because antibodies produced by them can be found in serum for decades after infection, with some studies suggesting the antibody response can persist for 70 years or more.[63, 65, 67] It is not yet fully resolved whether the initial ASCs survive for decades, or are self-renewing.[65] ASCs have also played a central role in the history of immunizations and vaccinations, especially in humans. Vaccines have never been found to produce durable T cell responses in humans, perhaps because of limits on adjuvants that can be used, which has shifted even more importance onto the antibody response.[68] ASCs and the antibodies they produce have long been the marker of successful vaccination, though modern

techniques are beginning to allow for a better understanding of long term immunity by also probing the memory B cell response. As such, understanding the composition and antigenic targets of ASCs responding during infection with and after recovery from SARS-CoV-2 is necessary to make conclusions about long term immunity to the disease and future vaccination strategies against emerging variants.

1.2.7 Innate-like B cells

Innate-like B cells are two groups of B cells that produce IgM that binds to more conserved molecular patterns than their follicular counterparts. They can also be activated and differentiate into antibody secreting cells without T cell help, instead relying upon their own PRRs, allowing them to mount responses more rapidly. The ability to rapidly respond to conserved epitopes is the basis for their innate like classification. The two groups innate-like B cells are B1 B cells (B1B) and marginal zone B cells (MZBs).[6, 26]

B1 B cells are an entirely separate lineage of B cells that are distinct from B2 B cells. They are thought to potentially mature within and arise from the fetal liver, before seeding the rest of the body during fetal development, though evidence for them also maturing within bone marrow in adults exists as well.[69] They are mostly found within the peritoneum, rather than within lymphatic tissues.[70] B1 B cells seem to be the type of B cell that are most defined by their BCR, using a restricted repertoire that is distinct from the B2 repertoire.[71] Additionally, genetic manipulation has revealed that a B2 B cell made to express a classic B1 receptor will alter itself to adopt, as much as possible, the genotype and phenotype of a B1 B cell and vice versa.[62] B1 B cells, upon activation, differentiate into short lived ASCs that secrete IgM targeting conserved molecular patterns of pathogens or, interestingly, self-antigens. It is believed the self-recognizing

B1 B cells play a role in maintaining bodily homeostasis through a housekeeping function, by marking cellular debris for removal by phagocytic innate cells.[6]

B1 B cells are perhaps the least studied B cell population in humans, so surface markers remain to be firmly defined though they are generally considered to be positive for CD20, CD27, and CD43 and negative for CD70. Markers more generally associated with memory or ASCs, such as CD86 and CD138 are also expressed on the surface of some B1 B cells.[47] Lastly, CD5 was traditionally thought to be a marker of B cells but recent data suggests there are two populations that are CD5 positive and CD5 negative and that they may perform slightly different functions in the body.[72] Genetically, a classic B1 marker is the gene *SPN*, with recent work also suggesting that *MYO1D, PSTPIP2, CD300LF, LYSMD2,* and *IZUMO1R* are also upregulated in B cells, though the latter two have only been demonstrated in murine B1s. *AHR* can be expressed by all B cell populations but is most highly expressed by B1 B cells. Lastly, *PLSCR1* is associated with all B cells producing natural antibodies, including B1Bs and MZBs.[73-77]

Marginal Zone B cells are the other population of innate-like B cells. They are members of the B2 lineage, with follicular B cells, and are named because they tend to localize to the marginal zone of the spleen and lymph nodes. In humans MZBs can also be found recirculating through blood, and localizing to mucosal surfaces, though in lower numbers than in the marginal zone.[6, 26] MZBs, like B1s, secrete natural antibodies in a T-independent manner rapidly upon infection, though can also mount T-dependent responses. MZBs have been identified to be the largest source of IgM in serum.[78] MZBs play a critical role in immunity to blood borne infections, as an early responder to all infections, and also in housekeeping functions related to homeostasis. In addition to their role in humoral immunity, MZBs are potent APCs and are able to more effectively activate CD4+ T cells than any other B cell, increasing evidence suggest

many autoimmune diseases may rely upon an initial break in tolerance driven by MZBs during an infection.[26] Some MZBs undergo somatic hypermutation (SHM) in the periphery and in response to the natural host microbiota, whether this SHM is acting as a driver of affinity maturation or diversification remains unclear. Lastly, populations of memory marginal zone B cells have been found in humans and in mice.[79]

Surface markers of MZBs in humans include having high expression of IgM, CD21, CD9, CD27, and CD1 while having low to no expression of IgD, CD23, CD5, and CD11b.[47] However, markers to distinguish naïve marginal zone B cells and memory marginal zone B cells remain unclear and undetermined in humans. Genetically MZBs are also much less studied than their follicular counterparts, however microarray studies have provided genes that strongly associate with them. These include *SOX7, SLA, TMEM71, HS3STB1, RPS6KA2, RHOBTB3, AEBP1,* and *DMD*. Intriguingly, the gene *IGHD*, which encodes for IgD, appears to be highly expressed in MZBs even though surface expression is practically nonexistent.[80]

The impact of innate-like B cell populations in human disease is little studied and not well elucidated. [6, 26] However, it has been found that a large expansion of extra-follicular, innate-like responses correlates with more severe disease and worse outcomes in COVID-19. [81] Even when those same cells seem to be producing a large amount on neutralizing antibodies, they also appear to be driving severe inflammation. A better understanding of the heterogeneity of innate-like B cell populations and how they influence the pathology and severity of COVID-19 will be critical for developing new therapies and treatments and can shed on a light on the more fundamental role these B cells play in immunity to infectious diseases.

To date, all varieties of B2 follicular lineage B cells have been well characterized in mice and fairly well characterized in humans. Innate-like cells, including the B1 lineage and marginal

zone B cells are less well characterized, especially within humans. Many past studies have characterized the genes, surface markers, or BCRs of the various populations of these cells but future studies that combine all of these metrics for single cells will allow for new insights into their biology. Additionally, the roles each of these population play in humans, and the impact of heterogeneity within them, remains an outstanding question. Therefore, new tools and techniques that can better elucidate differences between and amongst these populations will be key to improving our understanding of them and the roles they play in human health. Finally, human studies are often limited by having to contend with the interference and influence of established memory responses and unknown histories of infection. The COVID-19 pandemic represents a rare opportunity to study and profile immune responses from an extensive number of humans all experiencing infection by a novel pathogen.

1.3 The Germinal Center

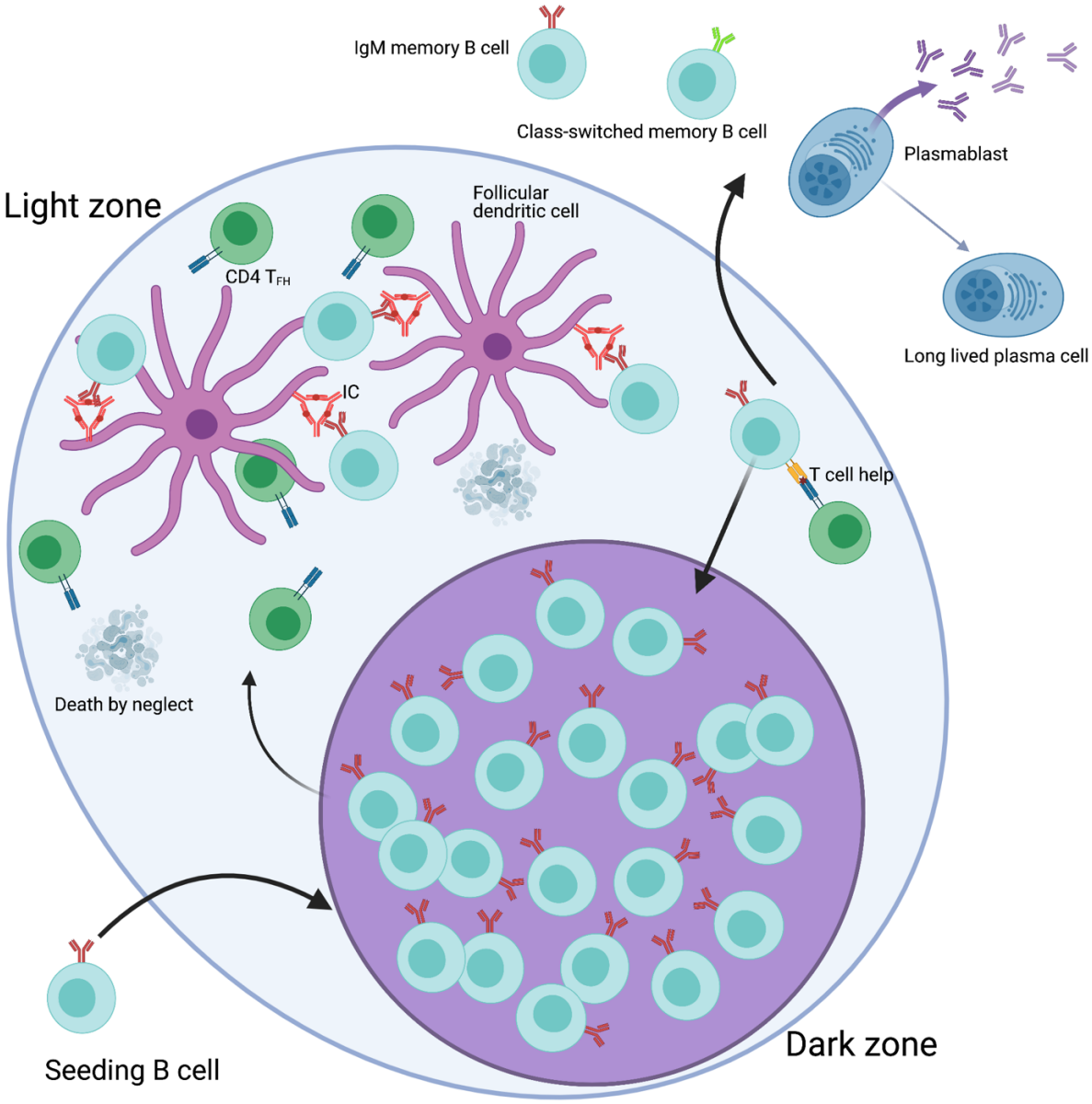


Figure 1.1: Model of a germinal center reaction.
Briefly, upon encounter with antigen naïve B cells (shown in light blue) migrate to the T:B border. Upon reception of co-stimulatory signals from CD_4 T cells they rapidly proliferate, with a collection of these cells coalescing into the nascent dark zone (DZ) of the GC (shown in purple). After expansion and mutation of their BCR, B cells will then traffic to the light zone where they compete for antigen in the form of immune complexes (shown in orange and red) presented by follicular dendritic cells (shown in pink). B cells able to acquire antigen can process and present it to CD4 T_{FH} cells (shown in green). A B cell incapable of acquiring antigen will undergo apoptosis, in a process called death by neglect. B cells that receive help can traffic back into the DZ and continue to divide and undergo SHM. This process continues for many cycles, with B

Figure 1.1, continued.
cells leaving at different stages either as memory B cells, plasmablasts, or long-lived plasma cells.

1.3.1 Overview

One of the central components of humoral immunity and the generation of long term humoral immune memory is the germinal center reaction. Germinal centers are temporary cellular structures that arise within lymph nodes and other lymphatic tissues and, in humans and mice, are the major sites of affinity maturation of B cells.[82] B cells enter germinal centers after interacting with an APC, with most germinal centers being seeded by one or only a few B cells.[83] Once within a germinal center a B cell will begin rapidly proliferating and mutating its BCR, afterwards all of the cells will compete for antigen with one another in an iterative, Darwinian process with the majority of cells dying. A lineage of B cells will increase their affinity to their cognate antigen as the number of GC cycles they have been through increases. Germinal centers can be split into two major zones, the light and dark zone, named for their differing appearance in optical microscopy. The dark zone is the site of rapid clonal expansion of B cells and mutation of their BCRs, and the light zone is the site of competition for antigen and reception of T cell help.[82, 84]

A specialized subset of CD4 T cells called T_{FH} cells and specialized stromal cells called follicular dendritic cells (FDCs) also inhabit the light zone (LZ). FDCs are responsible for presenting antigens on their cellular surface, which competing B cells will bind with their BCR. If a B cell is capable of binding and removing antigens from an FDC it will then phagocytose and process it. After processing the B cell will then load the antigens into their MHCII molecules and present it to the T_{FH} cells in the light zone. Upon presentation the TFH will give the B cell

help and survival signals, these signals tell the B cell to return to the DZ and proliferate and mutate their BCR. A B cell that does not receive enough help from the TFHs will undergo apoptosis, in a process termed 'death by neglect.'[82, 84, 85]

Notably, while the germinal center reaction in humans and mice seems to primarily be a driver of affinity maturation in many other species this is not the case. Disparate species such as chickens, cattle, sheep, horses, and rabbits all appear to largely use their germinal centers as sites of repertoire diversification. The number of V, D, and J genes available in them for recombination is much lower than in humans and they rely upon mutations in GCs that arise in response to the microbiota in development and early life to generate new BCRs, allowing them to recognize a greater array of antigens. This would perhaps imply the conditions for survival aren't as stringent in these GCs as they are if those driving affinity maturation, though these questions remain unaddressed.[86-94]

1.3.2 Affinity Maturation

Affinity maturation is the name for the process in which mature B cells can, within a germinal center, improve their affinity to their cognate antigen. In humans and mice evidence indicates this is the major purpose of the germinal center reaction. This process is, in humans, critical to the formation of high affinity, protective antibody responses against pathogens. Germinal centers are also the major source of class-switched memory B cells and long-lived plasma cells, as such GCs underlie humoral immunity.[59, 82, 84]

Affinity maturation begins in the DZ, with the B cell(s) that have seeded the GC undergoing rapid expansion and introducing mutations into their BCRs. Evidence indicates that a B cell will, on average, introduce one BCR mutation per GC cycle.[95] This implies that during the

phase of rapid expansion each daughter B cell will only differ in their BCRs at one site from the parent. These rapidly dividing and mutating B cells are known as centroblasts, and express high levels of CXCR4. The expression of CXCL12, the ligand for CXCR4, is expressed at very high levels by DZ stromal cells. This is what maintains the structure of the GC and keeps dividing B cells properly segregated.[95] After proliferation and mutation, the centroblasts will down regulate CXCR4 and begin to upregulate CXCR5. FDCs in the light zone express large amounts of CXCL13, the ligand for CXCR5. The B cells, now known as centrocytes will follow this chemokine gradient into the light zone where they will begin to compete with one another for antigen. They do this by attempting to bind and phagocytose antigens that are presented on the surface of FDCs. FDCs, being specialized stromal cells and not true DCs derived from a hematopoietic precursor, don't process or present antigens in the traditional manner. Rather, they are specialized to capture immune complexes (ICs) that are in circulation in the body.[96] ICs are protein complexes made of antigen bound antibodies, and can also be recognized and bound by complement proteins. FDCs capture ICs with incredible efficiency by using a variety of FC receptors, especially FcγRIII and the complement receptors CR1 and CR2. Once an FDC captures an IC it can present it on its surface or hold it in special, non-degradative phagocytic vesicles for later presentation. FDCs have been shown to be capable of recycling ICs for months.[97, 98]

It is these ICs that GC B cells, in the form of centrocytes, are competing for. This means that B cells are not only competing with one another for antigen, but also with current antibodies in circulation that are forming the ICs. Thus, the process of competition should ensure that GC B cells that are capable of phagocytosing the most antigen are binding with greater affinity than current responses and their direct competitors in the GC. This process can increase antibody

affinities by several orders of magnitude, with reports of affinities increasing toward some antigens by as many as ten orders of magnitude.[99]

After capturing antigen from the surface of FDCs, the centrocytes will then internalize and process it, preparing to present it to cognate CD4 T_{FH} cells that are also present in the LZ. Upon successful presentation to these T cells the centrocyte receives survival signals and co-stimulation. All centrocytes exist in a state that borders on the edge of death, and if they do not receive any or enough survival signal, because of an inability to capture and present antigen, they will undergo apoptosis in the light zone.[82, 84, 85] This process allows for only the higher affinity B cell clones to survive, and insures that any B cells that have destabilized their BCR through mutation will also be unable to survive and propagate. The signals GC B cells receive from T_{FH} cells act in an analog, rather than digital, manner. As such, B cells, and their eventual daughter cells, that can capture and present more antigen dwell even longer in the DZ and progress through the cell cycle more rapidly.[100, 101] These three combined processes; survival, longer dwell times, and more rapid proliferation combine to rapidly and efficiently promote an increase in affinity within the pools of competing cells inside of a GC.

1.3.3 Somatic Hypermutation

The heart of the germinal center reaction, whether it is being used for affinity maturation or repertoire diversification, is somatic hypermutation (SHM), the process by which GC B cells introduce mutations into their BCRs. SHM can produce mutation rates that are 10^6 times higher than background somatic mutation rates in the genome.[102, 103] SHM is a complex and multilayered process that relies upon the cooperation between multiple molecular and cellular mechanisms. The major mechanisms and components underlying SHM are the protein

activation-induced cytidine deaminase (AID), multiple DNA repair pathways and polymerases including the error prone polymerase-eta, and finally, the intrinsic genetic sequences of the BCRs themselves.[104, 105]

AID is known as the somatic hypermutator and is the protein responsible for initiating and driving SHM. It is encoded, in humans, by the gene *AICDA* and is 24kDa in size. AID is a member of the APOBEC (apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like) family of enzymes, a class of proteins that are innate immune effectors.[106] APOBEC enzymes can target and deaminate the cytidine and deoxycytidine of cDNA being produced by reverse transcription by RNA viruses. Other members of the family appear to work on mRNA directly.[107] AID has evolved to specifically target self-DNA inside of B cells that are undergoing SHM. In-vitro studies have confirmed that a B cell undergoing SHM expresses multiple copies of AID and that multiple copies can bind a strand of DNA at once.[108] AID is proposed to operate using a 'sit and glide' mechanism, in which is binds a strand of single stranded DNA (ssDNA) and glides down it, potentially deaminating one or more cytidine before releasing and attaching again elsewhere.[109] AID is only capable of deaminating one cytidine per several minutes, a rate of activity that is incredibly slow for an enzyme.[108] In-vitro mutational studies suggest this may be a protective mechanism, to protect the genome from an abundance of mutations leading to double-stranded breaks and chromosomal destabilization.[108, 109] AID has never been successfully crystalized, thus no structures of it have been published. Using predictive modeling based off of the structures of other APOBEC family members, the active site of AID was predicted to be built around two cysteines, 87 and 90, critical to its activity. Mutational studies have confirmed this as the general area of the catalytic pocket, and its predicted structure also implies it probably required the cytidines it deaminase to be "flipped out," meaning it would be out of the stacking

alignment of all other bases on the strand.[108, 110] Whether AID actively flips out the cytidine itself or preferentially acts on bases that are already flipped out of alignment is not yet established. Though AID initiates the process of somatic hypermutation it relies upon downstream mechanisms to complete it, including a variety of DNA repair mechanisms.

After deamination of cytidine to uracil, there are six potential outcomes or pathways the altered DNA could follow. Five of these outcomes rely upon DNA repair pathways that are one of two families of repair mechanisms, known as base excision repair (BER) or mismatch repair (MMR). Four enzymes exist that can excise a base as part of the base excision repair (BER) machinery, however, only one of them is found to be active during the process of SHM; UNG.[104, 111] UNG can be spliced differently two generate two variants, UNG1 is active in the mitochondria and UNG2 in the nucleus, therefore UNG2 is the splice variant that is active in SHM.[104] The first potential outcome after deamination relies upon the uracil, which appears as a thymine to DNA repair machinery, being properly excised and replaced with cytidine resulting in no mutation. The other five outcomes all result in mutation, however many of them will still end up being silent.[104] The first outcome resulting in mutation is for the DNA to synthesize across the uracil, replacing the original opposite strand G with an A and the targeted C with a T, resulting in a transition mutation.[112] The next potential pathway is for the base to properly excised by UNG2, which is then followed by translesion synthesis (TLS) to repair the lesion in the DNA, which will remove complementary base from the opposite strand as well. It could replace the removed bases with their original nucleotides, or could also replace them with any other base and its complementary partner, leading to transition or transversion mutations.[113] The third possibility relies upon a non-canonical mismatch repair pathway, which utilizes the MSH2/MSH6 mismatch recognition complex. After recognition of the AID generated mismatch, the MSH2/MSH6

complex anchors to the site and recruits another complex, formed of the enzymes PMS2 and MLH1, which will introduce a break into the DNA next to the mismatch.[114, 115] After the introduction of the break the exonuclease, EXOI, is recruited and excises a single stranded section of DNA to allow for TLS. This process is generally performed in mammalian cells in a highly reliable and accurate manner, however within GC B cells TLS is specifically mediated by polymerase-eta (POLH).[116] POLH has no proof-reading mechanism, making it much more prone to errors than other polymerases.[117] This pathway introduces mutations not at the initial site of deamination, but most frequently at A/T sites next to it.[104, 105]

The final two pathways are both somewhat of hybrid pathways, with the fourth pathway being a combination of BER and MMR. After initial excision of the uracil by UNG2, the damaged site is recognized by the previously described MMR machinery which anchors in place and follows the third path. This means that this pathway is capable of generating transition and transversion mutations in both the C/G pair of initial deamination and adjoining A/T sites. The fifth and final pathway is the most recently described and seemingly rarest. It relies upon initial BER performed by UNG2, followed by recruitment of a separate TLS pathway, PCNA-ubiquitination.[114, 118] Proliferating cell nuclear antigen, PCNA, clamps onto the DNA at the site of excision and initiates additional excision, potentially by EXOI. Afterwards, it appears to recruit POLH to the site to perform repair, which can result both in mutations at the initial C/G site or adjacent A/T sites. However, datasets seem to indicate this final pathway is responsible for less than 10% of all A/T mutations and a negligible amount of C/G mutations, if any.[119]

If AID begins the process and repair machinery ends it, the third component is the scaffolding upon which the other two are erected, the underlying DNA sequences itself. Both AID and POLH exhibit biases in how they target DNA, for deamination or error prone repair

respectively. Some sites seem to be preferentially selected for mutation, while others are preferentially protected from it. This has been termed hotspot / coldspot targeting.[120-122] For AID hotspots, cytidines within these sites are preferentially deaminated which increases the chance of a mutation occurring within them.[121] Traditionally AID hotspots are defined as the C within the DNA motif WR**C**Y, and it's corresponding reverse complement R**G**YW. This can be additionally broadened to WR**C**H or even WR**C**, though mutational rates within them are not as high as the more specific WR**C**Y. AID coldspots include the C within a SY**C**, and reverse complement **G**RS and the C at the end of any of the trinucleotides TT**C**, CA**C**, GG**C**, and GA**C** and their reverse complements.[105, 120] POLH has its own hotspots in which a mutation is more likely to occur, defined as the A in a W**A**, and complement **T**W.[105] Recent data indicates that POLH hotspots are perhaps much more important than previously appreciated, being major sites of non-silent mutations within BCRs.[105, 123] Additionally it has been suggested that overlapping AID hotspots, or locations where hotspots occur on both the coding and noncoding strand in such a way they overlap, are also major sites of NS mutation.[124, 125] Large scale sequencing datasets have been combined with mathematical modeling to produce an updated system of hotspot and coldspot targeting that is known as the S5F system, based on silent mutations at any given nucleotide and taking into account the two nucleotides to either side of it. This model overlays nicely with traditionally defined hot- and coldspots and has revealed additional sites of differing mutational activity.[126]

All variable genes feature hotspot, coldspot targeting but to varying degrees, with VH1-69 having the highest number of hotspots and overlapping hotspots.[127] Mouse models have also provided proof that the BCR locus is specially enriched for increased mutational activity, even in the absence of hotspot / coldspot targeting.[128] Finally, in addition to the intricate targeting system

that has evolved for BCRs, variable genes have also been shown to have a differential codon usage to the rest of the human or murine genome, with evidence that this helps balance silent (S) and nonsilent (NS) mutations.[105]

The combination of AID activity, various error-prone DNA repair pathways, and intricate targeting mechanisms imply there has been years of selection and evolution to refine and improve the process of somatic hypermutation. However, the impacts of hotspot coldspot targeting and the role it can play in novel or repeat infection remain little studied. Additionally, it is well appreciated by veterinary immunologists and those working in animal models other than rodents that SHM largely, if not entirely, exists as a system of B cell repertoire diversification rather than as the driver of affinity maturation.[129, 130] The impact this evolutionary history has on species that have co-opted it for affinity maturation also remains unexplored.

1.4 Severe Acute Respiratory Syndrome Coronavirus 2

1.4.1 Overview

Coronaviruses are enveloped positive-sense single-stranded RNA viruses belonging to the subfamily *Coronavirinae* of the family *Coronaviridae* in the order of *Nidovirales*. They can be further subdivided into four genera: *Alphacoronavirus, Betacoronavirus, Gammacoronavirus,* and *Deltacoronavirus*. Coronaviruses have non-segmented genomes that are approximately 30kb in size, amongst the largest found in any RNA virus. These large genomes support some of the highest rates of recombination seen amongst RNA viruses and are, in part, made possible by another unique trait of *Nidovirales*, the presence of proof-reading exoribonucleases.[131, 132] This ability to proofread their genomes is both a benefit and hindrance for pathogenic coronaviruses

and their hosts, while they may mutate more slowly a mutation that increases their fitness can be more rapidly fixed into the overall population of the virus.

Alpha- and betacoronaviruses broadly circulate amongst many mammals, including humans, causing respiratory and gastrointestinal diseases that rage from mild to severe. Delta- and Gammacoronaviruses mainly infect and circulate in avian populations, though some also show the ability to infect mammals on rare occasions.[133] Phylogenetic studies trace all alpha- and betacoronaviruses back to either bats or rodents as their original hosts. From there they have spread and adapted to a variety of other hosts, both domestic and wild. They can present a large disease burden to livestock and spread from livestock into humans.

Coronaviruses derive their name from the large, class I fusion proteins that coat their surface called spike proteins. The spike protein mediates attachment to host cell surfaces and viral entry. Cleavage of the spike protein and viral and endosomal membrane fusion mediates the release of viral RNA into the intracellular cytosol.[134] In most coronaviruses the spike protein can be further subdivided into two subunits; S1 and S2. S1 contains the receptor binding domain (RBD), responsible for the recognition and binding to a host cell surface protein. Coronavirus genomes also encode for the viral replication machinery, structural proteins, and a variety of non-structural proteins (NSPs).

1.4.2 Coronaviruses in Humans

They are currently seven known coronaviruses that can infect humans. Four of them cause only mild disease, the common cold, amongst healthy, immunocompetent individuals. Two of these human coronaviruses (HCoV) are alphacoronaviruses: HCoV-229E and HCoV-NL63, and two are betacoronaviruses: HCoV-HKU1 and HCoV-OC43. HCoV-229E and HCoV-

NL63 originated in bats while HCoV-HKU1 and HCoV-OC43 likely originated in rodents, with all four spreading to humans through intermediate hosts. Alpacas and cattle are the likely intermediates for HCoV-229E and HCoV-OC43 respectively, with those for HCoV-NL63 and HCoV-HKU1 remaining unidentified.[135]

Over the past two decades, two betacoronaviruses, severe acute respiratory syndrome coronavirus (SARS-CoV-1) and Middle East respiratory syndrome coronavirus (MERS-CoV), were responsible for over 10,000 cases of infection.[135, 136] In December of 2019, the novel coronavirus SARS-CoV-2 (the causative agent of COVID-19) was discovered through deep sequencing analysis after a series of pneumonia cases were reported in Wuhan, Hubei, China with unknown viral origin.[137] Both SARS-CoV-1 and MERS-CoV likely originated in bats and spread to humans through intermediate hosts, a common transmission pathway of coronaviruses. Civets have been identified as the intermediate host for SARS-CoV-1 and dromedary camels for MERS-CoV, with serological studies indicating the latter has been spreading amongst them for approximately thirty years.[138] It is likely that SARS-CoV-2 also spread to humans through an intermediate host, though it remains to be identified.[139] Severe acute respiratory syndrome coronavirus two (SARS-CoV-2) is the most recent coronavirus to jump into humans, in late 2019.

Coronaviruses in humans use a variety of receptors to infect host cells, though all rely upon their spike proteins to do so. HCoV-OC43 and HCoV-HKU1 use 9-O-acetyl-sialic acid, HCoV-229E uses human aminopeptidase N, MERS-CoV uses dipeptidyl peptidase 4, and HCoV-NL63, SARS-CoV-1, and SARS-CoV-2 all use angiotensin-converting enzyme 2 (ACE2).[140-142] Because of its role in viral entry, a better understanding of the immune response to spike protein is critically important. Other proteins of interest, with regards to their impact on

disease course and human health, are open reading frame protein 8 (ORF8) and the nucleocapsid protein (NP). ORF8 is the second most common site of RNA recombination, behind the spike protein, and also showed the second highest rates of variation and mutation during the 2003 SARS-CoV-1 epidemic.[135] Previous work in coronaviruses also suggests that ORF8 may play in role in modulating host endoplasmic reticulum stress, interfering with interferon responses and preventing apoptosis.[143-145] NP presents another intriguing target for further study because it plays a key role in the protection and propagation of viral RNA and can also potentially modulate host responses to ensure viral survivability.[146]

The last three coronaviruses that have evolved to infect humans all cause severe disease and high rates of morbidity and mortality, and SARS-CoV-2 has currently caused nearly 140 million cases and over 2.9 million deaths in 192 countries. It was officially declared a pandemic by the World Health Organization on March 11, 2020 and represents the largest and most severe outbreak of a coronavirus ever, as such a clearer understanding of the immune responses to coronaviruses is more paramount than ever.

1.4.3 COVID-19

Upon the emergence of SARS-CoV-2 in humans in Wuhan, China in 2019 it soon became apparent it was transmitting from human to human and could cause severe disease, including acute reparatory distress syndrome (ARDS). The disease caused by SARS-CoV-2 was officially named Coronavirus Disease 19 (COVID-19), in recognition of its causative agent and the year 2019.[147] COVID-19 is largely marked by respiratory symptoms, though has been shown to impact other organ systems as well, with the bulk of its transmission being caused by droplets

and aerosols shed by infected individuals. Cases of COVID-19 fall along a broad spectrum of severity, from entirely asymptomatic to highly severe cases that require mechanical ventilation.

The respiratory symptoms caused by COVID-19 are heterogenous and can be grouped into four broad categories: mild, moderate, severe, and critical. Mild symptoms include those that are common to many respiratory infections such as fever, fatigue, myalgia, dry cough, sore throat, rhinorrhea (runny nose), and frequent sternutation (sneezing).[147] Notably, COVID-19 also causes a set of symptoms termed mild that are less common, including frequently feeling short of breath (dyspnea) and either greatly reduced ability to smell and taste (hyposmia and hypogeusia, respectively) or their complete loss (anosmia and ageusia). Which is potentially caused by viral infection and destruction of olfactory neurons.[148] The common occurrence of anosmia and ageusia, even using an otherwise asymptomatic disease course, have made them somewhat of a hallmark of the infection. Moderate respiratory symptoms include pneumonia without hypoxemia (a blood $SpO2 < 92\%$), more frequent and severe coughing, higher fever, and a computerized tomography (CT) scan that shows lesions on the lung. Severe cases feature pneumonia with hypoxemia, and CT scans showing a large number of lesions or, alternatively, X-rays that show a 'ground glass' pathology in the lungs. Critical cases are defined as having ARDS, which can lead to shock and multi-organ system failure.[147]

The second most common set of symptoms related to COVID-19 infection impact the gastrointestinal (GI) tract. The viral receptor, ACE2, is more highly expressed in GI tissue than anywhere else in the body, and concomitantly, GI tissues have the highest ratios of ACE2 positive to negative cells.[149, 150] Approximately 50% of patients who suffer a symptomatic COVID-19 infection will experience GI symptoms, though they are largely mild, including

nausea, vomiting, diarrhea, abdominal pain and discomfort, vomiting, and loss of appetite (anorexia).[151]

COVID-19 also has been found to cause a range of symptoms that impact other tissues and organ systems. These include a variety of hematological symptoms including lymphopenia, increased prothrombin time, elevated d-dimer levels, and heightened lactate dehydrogenase. These also coincide with dysregulated coagulation. Laboratory findings associated with this include, in addition to the aforementioned elevated d-dimer, increased amounts of Fibrin Degradation Products (FDP) and fibrinogen and correspondingly decreased levels of antithrombin. All of this combined can lead to disseminated intravascular coagulation (DIC) which is associated with higher mortality. Because ACE2 is also expressed in cardiovascular tissue, SARS-CoV-2 can infect these cells and some patients experience myocarditis and myocardial injury. Clinical studies showing the rates of myocardial injury in hospitalized COVID-19 patients varies, from a low of approximately 5% of all patients to a high of 35%.[152, 153] Biomarkers associated with myocardial injury were significantly associated with increased mortality and case fatality rates (CFR), and more generally myocarditis is associated with negative long term impacts on health.[152, 154, 155]

COVID-19 can also damage renal tissue, causing an acute kidney injury (AKI) both directly and indirectly. Infections of kidney tubules have been observed, with the ability to result in complete renal failure and, indirectly, renal damage has been linked to both inflammation caused by blood clots caused by COVID-19.[156, 157] In addition, AKI is commonly associated with severe COVID-19, as is a heightened risk of chronic kidney disease in those who recover from severe infection.[156] Similarly to myocarditis, patients hospitalized with COVID-19 who experience an AKI or have biomarkers of renal damage exhibit significantly higher rates of

mortality than those who do not. Renal damage has previously been associated with negative outcomes from other respiratory infections, notably SARS-CoV-1 and H1N1 influenza.[158-160]

Lastly, in very rare cases COVID-19 can also present with dermatological and neurological symptoms, excepting the aforementioned anosmia and ageusia. Dermatological symptoms include rash and, even more rarely, skin lesions.[161] Viral encephalitis and potentially accompanying seizures caused by COVID-19, along with cerebral hemorrhage, is linked to viral invasion of the endothelium and damage to the blood-brain barrier.[162]

Children tend to experience asymptomatic or very mild infections, marked by minor respiratory symptoms.[147] Incredibly rarely COVID-19 can also cause Multisystem inflammatory syndrome in children (MIS-C), reminiscent of Kawasaki's disease. MIS-C causes many complications and can be deadly, further studies will be needed to understand how it is caused by COVID-19.[163]

The pathology of COVID-19 is broad and results in a myriad array of symptoms and outcomes. A better understanding of how and why SARS-CoV-2 can generate such different disease courses and outcomes will be critical to developing better therapeutics and vaccines for this and future coronavirus infections. The highly varied nature of the disease, the differences between children and adults, and the role of aberrant inflammation all suggest the immune response has a critical role to play in disease pathology. A better understanding of the immune response to COVID-19 is paramount in understanding the disease, and for better understanding the impact of novel infections in humans.

1.4.4 The Immune System and SARS-CoV-2

Apart from SARS-CoV-1 and, to a lesser extent, MERS immune, and specifically B cell, responses to coronaviruses have been little studied. This is presumably because, historically, coronaviruses cause only mild disease in humans or, if they do cause severe disease, infect very limited numbers of people. As such, there are critical gaps in our knowledge of B cell and antibody responses directed against them.

Studies revealed that those infected with SARS-CoV-1 generated high serum titers of anti-SARS antibodies and, more specifically, neutralizing antibodies.[164] Initially high titers of neutralizing antibodies waned over time, as expected, but tended to drop lower compared to other viral infections, similar to antibody responses to endemic-HCoVs. Neutralizing titers saw a marked drop around sixteen months post infection, becoming undetectable in some patients.[164] Contrary to natural infection, preclinical studies showed vaccines using purified spike proteins or spike proteins bound to nanoparticles produced potently neutralizing and long lasting antibody responses.[165] Intriguingly, a retrospective study performed six years post SARS infection found memory T cells in 100% of subjects but no memory B cells in any subject, and only 9% had any detectable antibodies.[166]

MERS, having infected even fewer humans than SARS, has less data available concerning humoral immune responses. Previous work has shown that subjects who had recovered from MERS all have detectable neutralizing antibody titers at thirty-four months post infection, with 86% showing no measurable decrease in titers as compared to 13 months post infection.[167] Additionally, neutralizing antibodies generated in response to SARS infection have been identified as being cross-reactive to MERS but with many lacking the ability to neutralize it.[168]

Early work from us and others successfully identified potently neutralizing antibodies against the SARS-CoV-2 spike after infection.[169-174] Additionally, more simple methods to detect memory B cells such as ELISPOT confirm that MBCs are generated against the spike and other viral antigens, even after mild infections, and persist for at least for six to eight months.[175-178] Finally, some studies have found potential links between certain B cell subsets and clinical outcomes, with both the expansion of IgM antibody secreting cells and extra-follicular, innate-like 'double negative' IgM producing cells being linked to more severe disease courses.[81, 179]

Large gaps in our knowledge of how B cells respond to SARS-CoV-2, the role they play in the disease course of COVID-19, and whether antibody responses towards viral antigens beyond the spike are helpful or harmful persist. Furthermore, it remains unclear if past memory to endemic-HCoVs can respond to SARS-CoV-2 infection, and what impact they might have upon COVID-19 infection. These questions are also more broadly applicable to coronaviruses at large, a historically understudied group of viruses, and can help understand how B cells respond to novel infections.

1.4.5 Testing, Diagnostics and Therapeutics

The COVID-19 pandemic has driven rapid development of diagnostic and therapeutic strategies and reagents. Diagnostics can be further divided into testing related to diagnosing an active infection, surveilling a population for unknown infections, interrogating a population to better understand historical infection rates, and screening individuals for various purposes. Therapeutics, likewise, can be split into true therapeutics, given for the treatment of disease, and prophylactics which are intended to prevent disease. All of these are indispensable in combating

SARS-CoV-2 and corresponding reagents for future pathogens will be useful in containing and preventing potential pandemics.

The first of the three pillars is testing and diagnostics. There are several types of tests approved worldwide for use in screening for SARS-CoV-2 infection. Most testing strategies require a tradeoff between accuracy and speed, with those that excel in one or the other being better suited to different situations. Testing based on using a polymerase chain reaction (PCR) or quantitative polymerase chain reaction (qPCR) amplifies viral genes unique to SARS-CoV-2, and is the most common form of diagnostic testing. It is currently considered the gold standard of tests, especially for diagnostic confirmation, showing remarkably high sensitivity and specificity, however it has downsides when considered for other uses. PCR testing is expensive, requiring costly machines and reagents to perform, fairly low throughput, and requires healthcare workers and technicians trained in paper sample collection, handling, and test performance.[180] Loop-mediated isothermal testing (LAMP) is a more recent technology that aims to recreate the specificity and accuracy of PCR testing more efficiently and in a single tube at a constant temperature. By constantly amplifying complementary regions from viral RNA, DNA loops are formed that are recognized by a chemical in the reagent mixture which causes a color shift.[181] Modern CRISPR-Cas9 based testing solutions also show promise as a way to gain the specificity and sensitivity of targeting nucleic acids without the cost, time, and technique associated with PCR.[182] Nucleic acids based tests will continue to serve in invaluable roles, but there is a need for other, simple high throughput tests.

The other two major types of tests are complementary sides of the same coin, antigen and antibody testing. Antibody testing relies upon using viral antigens to search for the presence of anti-viral antibodies within a person's bodily fluids, most commonly serum from blood.

Antibody tests offer moderately high specificity, but their sensitivity can lessen as a person's antiviral antibodies can drop over time.[180] Antibody testing is appropriate for understanding the past levels of infection within a population, as it reveals a history of disease but is potentially unreliable for diagnosing active infection. Antigen testing is the reverse of antibody testing. In it, specific anti-viral antibodies are used to check bodily fluids, often saliva, mucous, or serum from blood for the presence of viral antigens. Though their specificity is often lower than nucleic acid base testing, antigen tests can be rapidly and cheaply massed produced, are simple to use, and can give results in a matter of seconds or minutes. It in tantamount for testing strategies for SARS-CoV-2 and other, future, pathogens to continue to be improved and expanded.

Common anti-viral therapeutics include small molecule synthetic drugs and monoclonal anitbodies or antibody cocktails. Small molecules are often designed to inhibit or interfere with viral processes such as cellular entry and replication, and can be highly effective.[183] To date only one small molecule therapy, Remdesivir, has been approved for the treatment of COVID-19 and shows only moderate effectiveness.[184] The next class of therapeutics relies upon mimicking the antibody responses of the immune system, by injecting antibodies that are expressed in cell culture. These are often given as a cocktail, a mixture of antibodies that target different protective epitopes. Antibody therapies can be created by isolating and sequencing the BCRs that produce the antibodies or by synthetically producing them in a lab, using techniques such as phage display. There are currently two such treatments approved, from Regeneron and Eli Lily.[185, 186] Antibody therapies can be given both prophylactically and therapeutically, adding to their versatility and usefulness.

The final prophylactic in the arsenal is the most important, vaccines. By generating long term immunity against a pathogen, vaccines play an irreplaceable role in the prevention and

resolution of pandemics. Vaccines against SARS-CoV-2 have been developed at a more rapid pace than any before and include new and novel vaccine technologies. Two emergent technologies that have played a key role in this rapidity are messenger RNA (mRNA) based vaccines and adenovirus vector DNA vaccines. Other candidates, using more traditional vaccine platforms such as inactivated virus and subunit vaccine are also currently in clinical trials or approved for usage.[187] Currently approved vaccines show very high effectiveness in preventing disease, and especially severe disease. Two mRNA vaccines showed initial efficacies in excess of 90%.[188, 189] Additionally, an adenovirus vector vaccine and subunit vaccine had serological indication they would be equally effective against Wuhan variant SARS-CoV-2, with interim phase 3 results confirming this.[190, 191] Though clinical trial results can be useful in understanding the various efficacies of different vaccine platforms, a deeper understanding of the underlying immunology and the B cells generating the protective antibody responses will be critical for the evaluation of these vaccine technologies and their use against future pathogens.

As such, a method that allows for the high-throughput discovery and characterization of antibodies along with transcriptomic characterization of the B cells producing them can greatly aid in the creation and validation of novel and emerging diagnostic, therapeutic, and vaccine technologies. These characterization techniques are also an effective and powerful bridge between basic science and a hospital bedside, generating quickly translatable discoveries through the characterization of antibodies while simultaneously creating incredibly rich datasets ripe to be mined concerning questions of basic B cell immunology.

1.5 High throughput single cell analysis methods

The increasing availability of high throughput single cell techniques, such as single cell RNA sequencing (scRNAseq) is generating ever larger and more complex data sets.[192, 193] Additionally, techniques such as CITE-seq or oligo-tagged antigen sequencing are continuing to add additional layers of complexity into the data as well.[194-196] These massive single cell datasets can represent amazing sources to be mined for biological insights and discoveries, both those with immediate translational impact and others that help uncover fundamental aspects of biology, such as previously unrealized heterogeneity within cell populations.

However, datasets such as these can also represent challenges in proper interpretation and analysis. Massive datasets require extra caution and care to be used to ensure that all conclusions are accurate and not merely 'noise' or something that only has the appearance of true statistical significance. Additionally, sources of batch effects must be carefully considered and any such effects must be removed for the data they represent to be accurate.[192, 197]

Opening new frontiers within scRNA sequencing is the use of surface labels, termed cellular indexing of transcriptomes and epitopes by sequencing (CITE-seq) that allows for the measurement of surface protein expression on cells, similar to traditional flow cytometry.[196] This can allow for the accurate scRNA sequencing of traditionally flow defined populations of cells, guaranteeing that recent transcriptomic work is accurately building upon decades of past experiments. Furthermore, it can aid in the discovery and definition of additional heterogeneity in populations previously considered to be a homogenous whole. However, tools for proper analysis of combined multi-comics datasets remain lagging and are often designed without biological experiments and the constraints of wet work in mind. Therefore, new tools to better

analyze and understand this data that properly take into account how these datasets most often look are of pressing importance in the field today.

1.6 Summary

The biomedical sciences often require an intricate balance between two worlds; one of basic science that is the foundation of all other work and an endless source of inspiration and the world of translational work that can have an immediate impact. New tools and techniques increasingly offer a way to achieve both concurrently. Herein we used an oligo-tagged antigen bait sorting approach to profile B cell responses and generate antibodies against the causative agent of the COVID-19 pandemic, Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2). Upon its emergence in 2019 and following development into a pandemic in 2020 it was critically important to develop a better understanding of immune responses against it. We were able to demonstrate that spike reactive cells are readily fixed in the memory, which remained robust and stable even as serum titers dropped. However, we also found that immunity towards other viral antigens, such as ORF8 and NP, increased over time and found them to be exclusively non-protective. Further, we showed that differences in cellular reactivity patterns and population frequencies correlates with disease severity and age.

This is all in addition to generating a large dataset that provides a remarkably deep and rich source of information to be mined concerning humoral immune responses to viral infections, and especially towards novel viral infections. It also reveals that there is substantial heterogeneity amongst B cells, even with the traditionally broad groupings of naïve and memory. In the course of the work, it became apparent that tools for analysis and understanding of such

datasets have not kept up with ability to generate them, and so we have built new pipelines and charted new paths in how to effectively understand and utilize data such as these.

Lastly, we found that memory towards endemic human coronaviruses is activated and differentiates into mucosal homing antibody secreting cells upon infection by SARS-CoV-2. These cells show incredibly high amounts of somatic hypermutation which does not follow patterns predicted by mutational modeling based on AID targeting. Some of them appear to fall into a category of cells reminiscent of broadly neutralizing HIV and influenza antibodies and therefore they may serve as more readily available models for better understanding such antibody responses while additionally helping us to understand the limits of affinity maturation. They could potentially be promising targets for future designs of universal coronavirus vaccines.

In conclusion, we have crafted a method and the accompanying framework to analyze and understand it that has allowed us to rapidly answer questions with immediate impacts of global concern while also generating a powerful dataset ripe to be applied in better understanding fundamental aspects of B cell biology.

Chapter II: Materials and Methods

2.1 Peripheral Blood Mononuclear Cell Isolation

All studies and sample collection were performed with the approval of the University of Chicago institutional review board IRB20-0523 and University of Chicago biosafety committees. Informed consent was obtained after the research applications and possible consequences of the studies were disclosed to study subjects. The clinical trial associated with this sample collection was registered at ClinicalTrials.gov with identifier NCT04340050, and clinical information for patients included in the study is detailed in Table 1 and Table 2. COVID-19 convalescent donors were 18 years of age or older, eligible to donate blood per standard University of Chicago Medicine Blood Donation Center guidelines, had a documented COVID-19 polymerase chain reaction (PCR) positive test, and complete resolution of symptoms at least 28 days prior to donation. Acute infected blood donors were 18 years of age or older and their blood was collected per standard University of Chicago Medical Center guidelines, they had a documented COVID-19 polymerase chain reaction (PCR) positive test, were hospitalized, and had been scheduled to receive an infusion of convalescent donor plasma. For the acute infected subjects four blood draws were collected both before and after plasma infusion. Peripheral Blood Mononuclear Cells (PBMCs) were collected from leukoreduction filters or blood draws within 2 hours post-collection and, if applicable, flushed from the filters using sterile 1X Phosphate-Buffered Saline (PBS, Gibco) supplemented with 0.2% Bovine Serum Albumin (BSA, Sigma). Lymphocytes were purified using a Lymphoprep Ficoll gradient (ThermoFisher) and contaminating red blood cells were lysed by ACK buffer (ThermoFisher). Cells were frozen at -80 celsius in Fetal Bovine Serum (FBS, Gibco) with 10% Dimethyl sulfoxide (DMSO, Sigma)

prior to downstream analysis. On the day of sorting, B cells were enriched using the human pan B cell EasySep™ enrichment kit (STEMCELL).

2.2 Oligo-tagged antigen Probe Generation

SARS-CoV-2 and Hanta PUUV proteins were obtained from the Krammer laboratory at Icahn School of Medicine at Mount Sinai, the Joachimiak laboratory at Argonne National Laboratory, and the Fremont laboratory at Washington University in St. Louis. In addition, the pCAGGS expression plasmids for the spike protein, spike RBD, and hanta PUUV were obtained from the Krammer laboratory at Icahn School of Medicine at Mount Sinai and produced in house in Expi293F suspension cells (ThermoFisher). Sequences for the spike and RBD proteins as well as details regarding their expression and purification have been previously described.[187, 198] For the expression and biiotinylation of ORF8 truncated cDNAs encoding the Ig-like domains of ORF8 were inserted into the bacterial expression vector pET-21(a) in frame with a biotin ligase recognition sequence at the c-terminus (GLNDIFEAQKIEWHE). Soluble recombinant proteins were produced as described previously.[199] In brief, inclusion body proteins were washed, denatured, reduced, and then renatured by rapid dilution following standard methods. The refolding buffer consisted of 400 mM arginine, 100 mM Tris-HCl, 2 mM EDTA, 200 µM ABESF, 5 mM reduced glutathione, and 500 µM oxidized glutathione at a final pH of 8.3. After 24 hr, the soluble-refolded protein was collected over a 10 kDa ultrafiltration disc (EMD Millipore, PLGC07610) in a stirred cell concentrator and subjected to chromatography on a HiLoad 26/60 Superdex S75 column (GE Healthcare). Site specific biotinylation with BirA enzyme was done following the manufacture's protocol (Avidity) except that the reaction buffer consisted of 100mM Tris-HCl (pH7.5) 150 mM NaCl, with 5mM MgCl2 in place of 0.5 M

Bicine at pH 8.3. Full-length SARS-CoV-2 NP was cloned into pET21a with a hexahistidine tag and expressed using BL21(DE3)-RIL E. coli in Terrific Broth (bioWORLD). Following overnight induction at 25°C, cells were lysed in 20 mM Tris-HCl pH 8.5, 1 M NaCl, 5 mM β-mercaptopethanol, and 5 mM imidazole for nickel-affinity purification and size exclusion chromatography. Proteins other than ORF8 were biotinylated for 2 hours on ice using EZ-Link™ Sulfo-NHS-Biotin, No-Weigh™ Format (ThermoFisher) according to the manufacturer's instructions, afterwards unreacted biotin was removed by passage through a 7K MWCO desalting column (Zeba spin, ThermoFisher). Biotinylated proteins were then conjugated to Biolegend TotalSeq™ PE streptavidin (PE-SAV) oligos at a 0.72:1 molar ratio of antigen to PE-SAV. The amount of antigen was chosen based on a fixed amount of 0.5 µg PE-SAV or APC-SAV and diluted in a final volume of 10 µL. PE-SAV or APC-SAV was then added gradually to 10 µl biotinylated proteins 5 times on ice, 1 µl PE-SAV or APC-SAV (0.1 mg/ml stock) was added every 20 minutes for a total of 5 µl (0.5 µg) PE-SAV or APC-SAV. The reaction was then quenched with 5 µl 4mM Pierce™ biotin (Thermo Fisher) for 30 minutes for a total probe volume of 20 µL. Probes were then used immediately for staining.

2.3 Fluorescence Activated Cell Sorting

PBMCs were thawed and B cells were enriched using EasySep™ pan B cell magnetic enrichment kit (STEMCELL). B cells were stained with a panel containing CD19 PE-Cy7 (Biolegend), IgM APC (Southern Biotech), CD27 BV605 (Biolegend), CD38 BB515 (BD Biosciences), and CD3 BV510 (BD Biosciences). B cells were stained with surface stain master mix and the oligo-tagged antigen probes for 30 minutes in the dark, on ice in 1X PBS supplemented with 0.2% BSA and 2 mM Biotin. Cells were stained with probe at a 1:100

dilution (NP, ORF8, RBD, PUUV, empty) or 1:200 dilution (spike, endemic HCoV spikes). Cells were subsequently washed with 1X PBS 0.2% BSA and stained with Live/Dead BV510 (Thermo Fisher) in 1X PBS for 15 minutes. Cells were washed again and re-suspended at a maximum of 4 million cells/mL in 1X PBS supplemented with 0.2% BSA and 2 mM Biotin for downstream cell sorting using the MACSQuantTyto cartridge sorting platform (Miltenyi). Cells that were viable/$CD19^+$/antigen-PE^+ or viable/$CD19^+$/antigen-APC^+ were sorted as probe positive. The PE^+ and APC^+ gates were drawn by use of FMO controls. Cells collected from the cartridge sorting chamber using a 1000µl micropipette and gel loading tips, the chamber was then flushed with an additional 500µL of PBS supplemented with 0.2% BSA and 2 mM Biotin which was collected and added to the original flow through. The collected cells were centrifuged at 400 RCF and 4 degrees Celsius to pellet them, afterwards all but 31.7 µl of buffer was removed. The sorted cells were then immediately used for scRNA sequencing.

2.4 Single Cell RNA Sequencing

VDJ, 5', and probe feature libraries were prepared using the 10X Chromium System (10X Genomics, Pleasanton, CA). All steps were performed using Chromium Single Cell V(D)J Reagent Kits (v1.1 Chemistry) including the Chromium Single Cell 5' Library and Gel Bead v1.1 Kit, Chromium Single Cell 5' Library Construction Kit, Human B Cell V(D)J Enrichment Kit, and Feature Barcode Library Kit. All steps were followed as listed in the manufacturer's instructions. Specifically, user guide CG000186 Rev D was used. Several acute infected samples were pooled post sort and hashtagged (Biolegend) and ran as a single sample, to account for low cell numbers. All bioanalysis and quality control steps in the protocol were performed using an Agilent 2100 Bioanalyzer using a High Sensitivity DNA Chip and Assay reagents and ran on the

machine using the "2100 expert_High Sensitivity DNA Assay setting." Final libraries were pooled according to concentration and cell number and sequenced using the NextSeq550 (Illumina, San Diego, CA) with 26 cycles apportioned for read 1, 8 cycles for the i7 index, and 134 cycles for read 2.

2.5 Linq-View for Data Processing

Data Processing of Single Cell Sequencing Data

We used Cell Ranger (version 3.0.2) for raw sequencing processing, including 5' gene expression analysis, antigen probe analysis, and immunoprofiling analysis of B cells. Based on the output from Cell Ranger, we performed downstream analysis using Seurat (version 3.2.0, an R package, for transcriptome, cell surface protein and antigen probe analysis) and IgBlast (version 1.15, for immunoglobulin gene analysis). For transcriptomic analysis; Seurat was used for cell quality control, data normalization, data scaling, dimension reduction (both linear and non-linear), clustering, differential expression analysis, batch effects correction, and data visualization. Unwanted cells were removed according to the number of detectable genes with cells having transcriptomes that numbered of less than 200 or more 2500 genes being eliminated. Percentage of mitochondrial genes was used to further identify and eliminate cells using a soft threshold set to the 95th percentile of the current dataset distribution, with the soft threshold subject to a sealing point defined as 10% of the maximum threshold in the case of particularly poor cell quality. Transcriptomic data was normalized by a log-transform function with a scaling factor of 10,000 whereas cell surface protein and antigen probe were normalized by a centered log-ratio (CLR) normalization. We used variable genes in principal component analysis (PCA) and used the top 15 principal components (PCs) in non-linear dimension reduction and

clustering. High-quality cells were then clustered by Louvain algorithm implemented in Seurat under the resolution of 0.6. Differentially expressed genes for each cell cluster were identified using a Wilcoxon rank-sum test implemented in Seurat. Batch effects correction analysis was performed using an Anchor method implemented in Seurat to remove batch effects across different datasets. All computational analyses were performed in R (version 3.6.3).

ROGUE Scoring

To assess the quality and purity of B cell subsets identified in this study we used a previously published method called ROGUE scoring, an entropy-based metric for assessing the purity of single cell populations.[200] The expression entropy for each gene was calculated using "SE_fun" from "ROGUE" package (version 1.0). Based on the expression entropy, ROGUE score for each cluster was calculated using "rogue" function from the same package with parameters "platform" set to "UMI" and "span" set to 0.6.

Antigen Reactivity Determination

Oligo-tagged antigen probe signals were normalized by centered log-ratio transformation individually for each subject. All B cells were subsequently clustered into multiple probe-specific groups according to their normalized probe signals. By investigating all normalized antigen-probe binding signals, we set an arbitrary threshold equal to 1 for all normalized probe signals to distinguish probe binding cells as 'positive' or 'negative.' Cells that were negative to all probes were clustered into the "Negative" group; those positive to only one probe were clustered into corresponding probe-specific groups; and those that were positive to multiple probes were further investigated. Only cells whose top hit probe value was at least two-fold greater than their second highest probe hit value were clustered into the top hit probe-specific group; others were clustered into the "Poly" group that indicates polyreactive cells. To account for the inclusion of

endemic HCoV spike proteins in some samples, cells positive to both SARS2 Spike and Endemic Spike were further clustered into a group designated 'spike crossRx.' For samples that included separate SARS-CoV-2 spike and RBD oligo-tagged probes we placed cells positive to both SARS2 Spike and SARS2 RBD into the 'Spike' group.

Cellular Identification by Gene Module Scoring

Scores for B cell genotype related gene modules (e.g. MBC score, Naïve score, ASC score, and GC emigrant score) were calculated using "AddModuleScore" function from Seurat package, which was adapted from a previous study.[201] Naïve score was calculated based on the following genes: *BACH2, ZBTB16 ,APBB2, SPRY1, TCL1A, IKZF2*; MBC score was calculated based on the following genes: *CD27, CD86, RASSF6, TOX, TRERF1, TRPV3, POU2AF1, RORA, TNFRSF13B, CD80, FCRL5*; ASC score was calculated based on the following genes: *PRDM1, MANF, XBP1, IL6R, BCL6, IRF4, TNFRSF17, CD38*; GC emigrant score was calculated based on the following genes: *NT5E, MKI67, CD40, CD83, TNFRSF13B, MAP3K8, MAP3K1, FAS.*

2.6 Monoclonal Antibody Cloning and Expression

Selection of Candidate Monoclonal Antibodies

Representative antibodies from each subject were chosen for synthesis by choosing random samplings of B cells that bound to a given antigen probe with higher intensity relative to all other probes. B cells with varying ranges of probe-binding intensities were chosen for confirmation by ELISAs. In addition, B Cells representing select public clonal expansions were also chosen for cloning. B cells binding to all probes in a polyreactive manner were also chosen and validated for polyreactivity by polyreactivity ELISA.

Monoclonal Antibody Generation and Expression

Immunoglobulin heavy and light chain gene sequences were obtained by 10X Genomics VDJ sequencing analysis and synthesized by Integrated DNA Technologies. Cloning, transfection, and mAb purification have been previously described.[202] Briefly, sequences were cloned into human IgG1 expression vectors using Gibson assembly, and heavy and light genes were co-transfected into 293T cells (ThermoFisher). Secreted mAbs were then purified from the supernatant using protein A agarose beads (ThermoFisher).

Identification of public clones

Potential public clones were identified using Vgenes software (internal to Wilson Lab) by generating a list of B cells using the same V(D)J heavy chain pairing that appeared in multiple subjects. Afterwards, VJ light chain pairings for all cells from a potential pool were compared and those using different light chains were excluded. Next, sequences were aligned and only those with a 98% amino acid similarity in both heavy and light chains were considered true potential public clones. Afterwards, all potential public clones were expresses as monoclonal antibodies and tested against their cognate antigens to confirm binding, only pools that had cells that bound antigen from multiple subjects were called public clones.

2.7 Enzyme Linked Immuno Absorbant Assay

High-protein binding microtiter plates (Costar) were coated with recombinant SARS-CoV-2 proteins at 2 μg/ml in 1X PBS overnight at 4°C. Plates were washed the next morning with 1X PBS 0.05% Tween and blocked with 1X PBS containing 20% fetal bovine serum (FBS) for 1 hour at 37°C. Antibodies were then serially diluted 1:3 starting at 10 μg/ml and incubated for 1 hour at 37°C. Horseradish peroxidase (HRP)-conjugated goat anti-human IgG antibody

diluted 1:1000 (Jackson Immuno Research) was used to detect binding of mAbs, and plates were subsequently developed with Super Aquablue ELISA substrate (eBiosciences). Absorbance was measured at 405 nm on a microplate spectrophotometer (BioRad). To standardize the assays, control antibodies with known binding characteristics were included on each plate and the plates were developed when the absorbance of the control reached 3.0 OD_{405} units. All experiments were performed in duplicate two to three times.

2.8 Polyreactivity ELISA

High-protein binding microtiter plates (Costar) were coated with 10 μg/ml calf thymus dsDNA (ThermoFisher), 2 μg/ml Salmonella enterica serovar Typhimurium flagellin (Invitrogen), 5 μg/ml human insulin (Sigma-Aldrich), 10 μg/ml KLH (Invitrogen), and 10 μg/ml Escherichia coli LPS (Sigma-Aldrich) in 1X PBS. Plates were coated with 10 μg/ml cardiolipin in 100% ethanol and allowed to dry overnight. Plates were washed with water and blocked with 1X PBS/0.05%Tween/1mM EDTA. MAbs were diluted 1 μg/ml in PBS and serially diluted 4-fold and added to plates for 1.5 hours. Goat anti-human IgG-HRP (Jackson Immunoresearch) was diluted 1:2000 in PBS/0.05%Tween/1mM EDTA and added to plates for 1 hour. Plates were developed with Super Aquablue ELISA substrate (eBioscience) until the positive control mAb, $3H9^{203}$, reached an OD_{405} of 3. All experiments were performed in duplicate.

2.9 Memory B Cell Stimulations and enzyme-linked immunospot assays (ELISpot)

MBC stimulations were performed on PBMCs collected from subjects in the convalescent cohort. To induce MBC differentiation into antibody secreting cells, $1x10^6$ PBMCs were stimulated with 10 ng/ml Lectin Pokeweed Mitogen (Sigma-Aldrich), 1/100,000 Protein A

from Staphylococcus aureus, Cowan Strain (Sigma-Aldrich), and 6 μg/ml CpG (Invitrogen) in complete RPMI in an incubator at 37C/5% CO_2 for 5 days. After stimulation, cells were counted and added to ELISpot white polystyrene plates (ThermoFisher) coated with 4 μg/ml of SARS-CoV-2 spike that were blocked with 200 μl of complete RPMI. ELISpot plates were incubated with cells for 16 hours overnight in an incubator at 37C/5% CO_2. After the overnight incubation, plates were washed and incubated with anti-IgG-biotin and/or anti-IgA-biotin (Mabtech) for 2 hours at room temperature. After secondary antibody incubation, plates were washed and incubated with streptavidin-alkaline phosphatase (Southern Biotech) for 2 hours at room temperature. Plates were washed and developed with NBT/BCIP (Thermo Fisher Scientific) for 2–10 minutes, and reactions were stopped by washing plates with distilled water and allowed to dry overnight before counting. Images were captured with Immunocapture 6.4 software (Cellular Technology Ltd.), and spots were manually counted.

2. 10 Viral Plaque Assay - *In-vitro* neutrilization

The SARS-CoV-2/UW-001/Human/2020/Wisconsin (UW-001) virus was isolated from a mild case in February 2020 and used to assess neutralization ability of monoclonal antibodies (mAbs). Virus (~500 plaque-forming units) was incubated with each mAb at a final concentration of 10 μg/ml. After a 30-minute incubation at 37C, the virus/antibody mixture was used to inoculate Vero E6/TMPRSS2 cells seeded a day prior at 200,000 cells per well of a TC12 plate. After 30 minutes at 37°C, cells were washed three times to remove any unbound virus, and media containing antibody (10 μg/ml) was added back to each well. Two days after inoculation, cell culture supernatant was harvested and stored at -80°C until needed. A non-relevant Ebola virus GP mAb and PBS were used as controls.

To determine the amount of virus in the cell culture supernatant of each well, a standard plaque-forming assay was performed. Confluent Vero E6/TMPRSS2 cells in a TC12 plate were infected with supernatant (undiluted, 10-fold dilutions from 10^{-1} to 10^{-5}) for 30 minutes at 37C. After the incubation, cells were washed three times to remove unbound virus and 1.0% methylcellulose media was added over the cells. After an incubation of three days at 37C, the cells were fixed and stained with crystal violet solution in order to count the number plaques at each dilution and determine virus concentration given as plaque-forming units (PFU)/ml.

2.11 *In-vivo* Infection Model in Hamsters

To evaluate the efficacy of RBD and NP monoclonal antibodies (mAbs) in vivo, groups of 4–5 week old female Syrian golden hamsters (four animals in each group) were infected with SARS-CoV-2 at a dose of 10^3 PFU by intranasal inoculation. One day later, the hamsters were treated by intraperitoneal injection with one of the mAbs at 5 mg/kg. Control groups of hamsters were injected with either sterile PBS or a non-relevant mAb (Ebola glycoprotein 133/3.16). Weights were recorded daily. Four days after the infection, nasal turbinate and lung samples were collected to determine viral loads in these tissues by standard plaque assay on Vero E6/TMPRRSS2 cells. All animal studies were conducted under BSL-3 containment with an approved protocol reviewed by the Institutional Animal Care and Use Committee at the University of Wisconsin and Institutional Biosafety Committee.

2.12 *In-vivo* Protection Model in hACE2 Mice

Eight-week-old heterozygous female K18-hACE c57BL/6J mice (strain: 2B6.Cg-Tg(K18-ACE2)2Prlmn/J) received 200 μg of each indicated mAb by intraperitoneal injection one day prior to intranasal inoculation with 10^3 PFU of SARS-CoV-2 (n-

CoV/USA_WA1/2020 strain). Weight change was monitored daily, and lungs were harvested at 7 days post-infection. Viral RNA levels in the lung were determined qRT-PCR quantifying N gene copy number and compared to a standard curve. Studies were performed with the approval of the Washington University in St. Louis Institutional Biosafety Committee and Institutional Animal Care and Use Committee.

2.13 Mutational and BCR Analysis

Additional computational analysis was performed using the Immcantation Framework.[204] [205]. Clonal expansions were identified using the pRESTO package and Python 3.9.1, with public clones confirmed manually using JMP statistical software (SAS Institute). Analysis of clonal abundance, Vgene usage diversity, and amino acid physiochemical property analysis was performed using the Alakhazam R package from the Immcantation framework. Quantification of mutational load, Statistical models of SHM targeting patterns, Analysis of selection pressure, and model dependent distance calculations with performed using the SHazaM R package from the Immcantation framework. R version 4.0.3, "Bunny-Wunnies Freak Out" was used for all analyses.

Chapter III: Profiling B cell immunodominance upon SARS-CoV-2 infection reveals antibody evolution to non-neutralizing viral targets

This section adapted from
Dugan, Stamper, and Li et al. Profiling B cell immunodominance upon SARS-CoV-2 infection reveals antibody evolution to non-neutralizing viral targets. *Immunity*. DOI: https://doi.org/10.1016/j.immuni.2021.05.001

3.1 Summary

Dissecting the evolution of memory B cells (MBCs) against SARS-CoV-2 is critical for understanding antibody recall upon secondary exposure. Here, we utilized single-cell sequencing to profile SARS-CoV-2-reactive B cell subsets in 42 COVID-19 patients. We isolated thousands of B cells in multiple distinct subsets specific to the SARS-CoV-2 spike, endemic coronavirus (HCoV) spikes, nucleoprotein (NP), and open reading frame 8 (ORF8). Spike-specific cells were enriched in the memory compartment of acutely infected and convalescent patients 1.5–5 months post-infection. With severe acute infection, we identified substantial populations of endemic HCoV-reactive antibody secreting cells with highly mutated variable genes, indicative of preexisting immunity. Finally, MBCs exhibited maturation to NP and ORF8 over time relative to spike, especially in older patients. Monoclonal antibodies against these targets were non-neutralizing and non-protective in vivo. These findings reveal considerable antibody adaptation to non-neutralizing antigens during infection, emphasizing the importance of vaccination for inducing neutralizing spike-specific MBCs.

3.2 Introduction

Since the emergence of SARS-CoV-2 in December 2019, the World Health Organization has reported spread to nearly every country with over 132 million infections and 2.9 million deaths worldwide. Faced with such persistence, the prospect of re-infection or infection with newly emerging variants warrants the need for studies evaluating the generation of durable B cell memory upon infection.

Early in the pandemic, several independent groups identified potently neutralizing antibodies against the SARS-CoV-2 spike protein, the major antigenic glycoprotein of the virus.[169-174] Since then, there has been a dedicated interest in the identification of durable memory B cells (MBCs) that can provide protection from reinfection. Our group and others have identified MBCs against the spike, nucleoprotein (NP) and open reading frame 8 (ORF8) proteins in convalescence, and some studies show these populations persist up to 6 months post-infection.[175-178] Beyond the identification of durable MBCs, spike-specific MBCs continue to adapt to SARS-CoV-2 up to 6 months post infection, in a manner consistent with antigen persistence and ongoing germinal center reactions.[206] Finally, recent data suggest that particular B cell subsets may be associated with disease severity, including the expansion of antibody-secreting cells (ASCs) and double negative B cells lacking surface expression of CD27 and IgD.[81, 179]

Despite these advances, we lack a clear understanding of MBC immunodominance and adaptation to distinct SARS-CoV-2 antigens over time, and how this correlates with factors such as patient age or disease severity. Moreover, it remains to be determined whether MBCs to targets such as NP and ORF8 can provide protection from infection. Finally, the role of

preexisting immunity to endemic human coronaviruses (HCoV) in shaping MBC responses to SARS-CoV-2 is poorly understood.

To address these knowledge gaps, we characterized the SARS-CoV-2-specific B cell repertoire in 38 COVID-19 patients, both severe acute and convalescent, approximately 1.5–4.5 months post-symptom onset using single-cell RNA sequencing. Through this approach, we provide a tool for evaluating human B cell subsets, immunodominance, and antibody adaptation to SARS-CoV-2, and have made a repository of over 13,000 antibody sequences available to the SARS-CoV-2 research community.

Our studies reveal that MBCs display substantial reactivity toward NP and ORF8, and continue to expand and adapt over time, particularly in older patients. While SARSCoV-2 RBD-specific monoclonal antibodies (mAbs) were potently neutralizing and protective, we showed that anti-NP and anti-ORF8 mAbs failed to neutralize and provide protection in vivo. Analogous to findings in the influenza virus field, B cells to non-neutralizing viral targets may be preferentially induced in natural infection and shaped by factors such as preexisting immunity and age.[207, 208] Therefore, preexisting MBC bias to non-neutralizing targets in SARS-CoV-2 could impact susceptibility to or severity of re-infection. Together, these findings highlight the importance of SARS-CoV-2 vaccines, which are optimally formulated to develop protective MBC responses against the spike.

3.3 Single Cell RNA Sequencing of Endemic HCoV and SARS-CoV-2 Specific B Cells

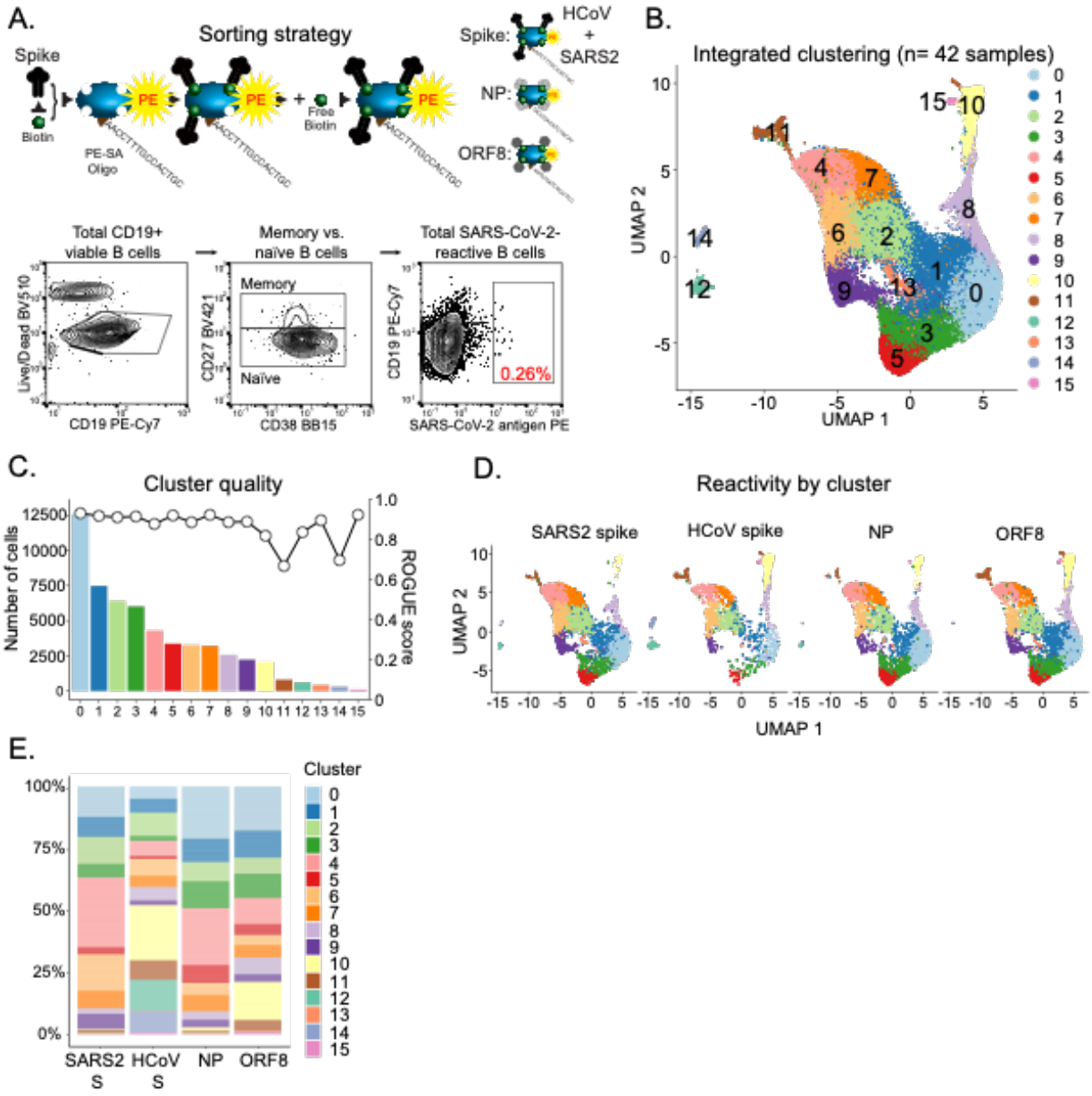


Figure 3.1: SARS-CoV-2-specific B cells comprise multiple distinct clusters.

Figure 3.1, continued.
(A) Model demonstrating antigen probe preparation and representative gating strategy for sorting antigen-positive B cells.
(B) Integrated transcriptional UMAP analysis of distinct B cell clusters (n=42 samples from severe acute (n=10), convalescent visit 1 (n=28), and convalescent visit 2 (n=4) cohorts; 55,656 cells)
(C) Cluster quality score determined by ROGUE analysis.
(D) UMAP projections showing CoV-specific cells used in all downstream analyses and the clusters they derive from.
(E) Quantitative visualization of CoV-specific cells and their distributions across distinct clusters.
Contributions for 3.1 and 3.2: Chris Stamper and Haley Dugan selected all samples. Chris generated all probes for every experiment while Haley thawed and stained cells. Chris and Haley performed all sorting and 10x genomics for convalescent samples, including construction of all cDNA libraries. Chris performed all bioanalysis on the libraries. Siri Changrob and Steven Erickson sorted acute infected samples. Chris analyzed all flow files in FlowJo to determine cell numbers for pooling and sequencing. Nick Asby pooled and sequenced libraries on early samples. Chris and Haley, with the assistance of Siri, pooled and sequenced libraries on later samples. Lei Li processed all raw data and produced figures in R. Chris and Haley analyzed the data and Lei helped write code for analysis. Chris and Haley made the figure in Illustrator and produced all figures not made in R.

MBCs have potential to act as an early line of defense against viral infection, as they rapidly expand into antibody-secreting cells (ASCs) upon antigen re-encounter. To determine the landscape of MBC reactivity toward distinct SARS-CoV-2 and endemic HCoV spike viral targets, we collected peripheral blood mononuclear cells (PBMCs) and serum between April and May of 2020 from 10 severely infected acute subjects and an additional 28 subjects upon recovery from SARS-CoV-2 viral infection. In addition, 4 convalescent subjects returned approximately 4.5 months post-symptom onset for a second blood draw, with similar volumes of whole blood processed across timepoints. Severe acute infected samples were collected days 0, 1, 3, 5, and 14 before (day 0) and after receiving convalescent plasma therapy (Table 3.1 and Table 3.2). All sampling timepoints were pooled from the same subjects for analysis due to low cell numbers.

To identify SARS-CoV-2-specific B cells, we used the SARS-CoV-2 (SARS2) spike protein, spike RBD, NP, and ORF8 to generate probes for bait-sorting enriched B cells for subsequent single-cell RNA sequencing analysis. This was done by conjugating distinct PE streptavidin (SAV)-oligos to individual biotinylated antigens (Figure 3.1A). To control for non-specific B cell reactivity and B cells reactive to PE, we included an empty PE-SAV-oligo, along with an irrelevant viral antigen control on APC, the Gn protein from the hantavirus, PuuV, to improve the specificity of sorting and downstream analysis. Finally, to understand the impacts of preexisting immunity to endemic HCoVs, which share up to 30% amino acid identity with the SARS2 spike, we included a cocktail of spike proteins from four coronavirus (CoV) strains that cause mild upper respiratory infections in the vast majority of individuals: HCoV-229E, HCoV-NL63, HCoV-HKU1, and HCoV-OC43, on an additional APC-SAV-oligo.

From a total of 38 subjects analyzed, including four matched follow-up visits ~4.5 months post-symptom onset, we detected small percentages (0.02–1.25%) of SARS-CoV-2-reactive total $CD19^+$ B cells, which were subsequently used to prepare 5' transcriptome, immunoglobulin (Ig) VDJ, and antigen-specific probe feature libraries for sequencing (Figure 3.1A). We sorted on total $CD19^+$ B cells with elevated mean fluorescence intensity in order to capture highly specific cells regardless of naïve-like or MBC origin, though a caveat of this approach may be the exclusion of lower affinity B cells. We integrated sequencing results from all 38 subjects (42 samples) using Seurat to remove batch effects and identified 16 transcriptionally distinct B cell clusters based on expression profiles (Figure 3.1B). Adopting the ROGUE scoring method, which compares how similar all transcriptomes within a cluster are to one another, we determined most clusters were highly pure, with the majority having a score over 0.9 (1.0 being 100% pure) (Figure 3.1C).[200] We ensured our feature libraries correlated with

single probe antigen-specific reactivity via a series of filtering steps to remove cells that were probe-negative, multi-reactive and non-specific, empty SAV-PE^{+}, or Hanta-PUUV^{+}. After all pre-filtering steps were complete, mapping only the cells that bound a single probe revealed that antigen-specific cells were enriched in distinct transcriptional clusters (Figure 3.1D and 3.1E), with considerable variation observed amongst individual subjects (Figure 3.2A and 3.2B). Notably, we did not identify obvious differences in B cell subset distribution or antigen reactivity in B cells from severe acute subjects analyzed early (days 0, 1, 3) or late (days 7, 14) post-convalescent plasma therapy (Figure 3.2C and 3.2D). In summary, method revealed substantial complexity in the B cell response to distinct CoV antigens, which we then further dissected by subset.

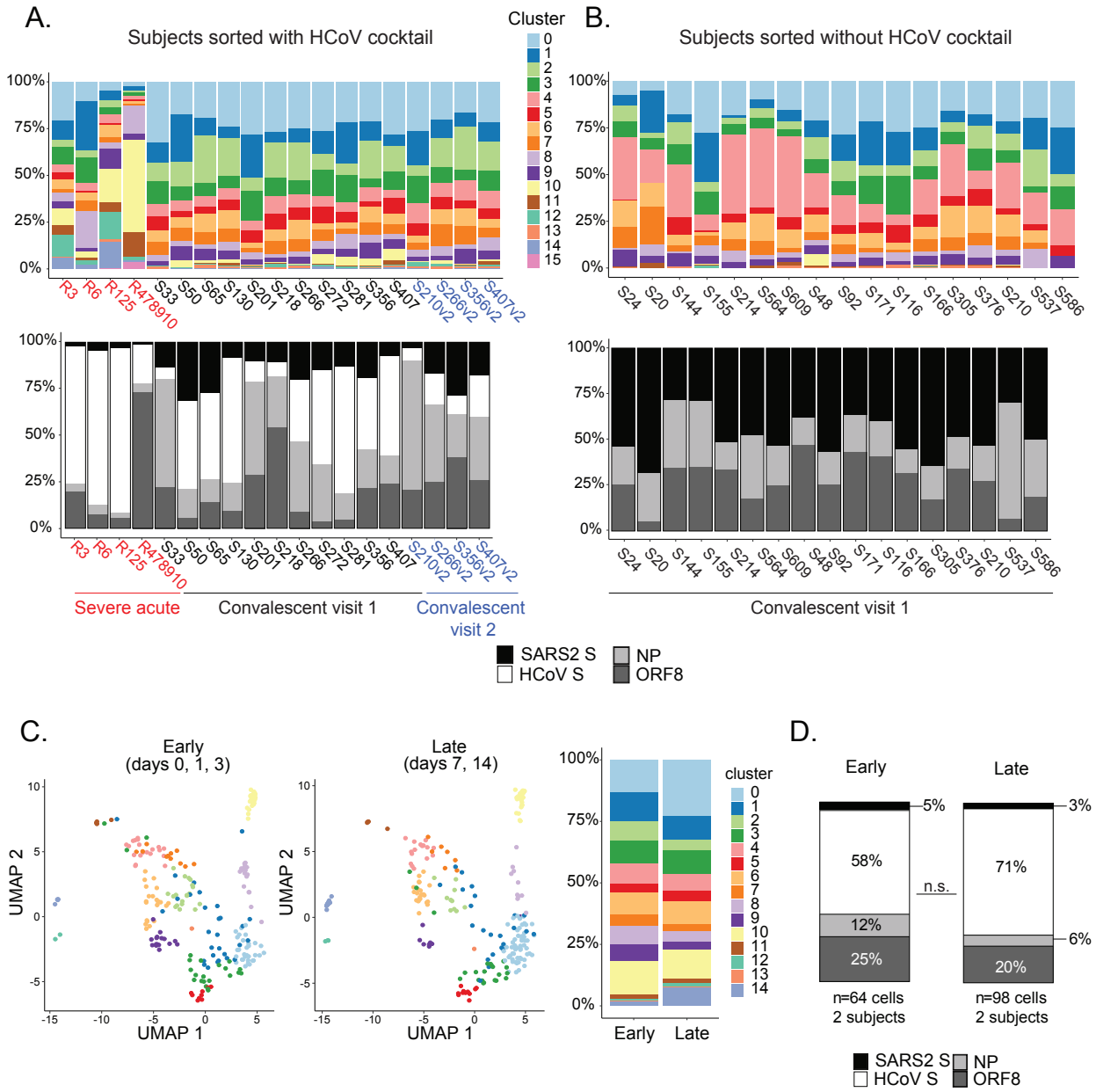


Figure 3.2. B cell subset distribution and antigen specificity by subject.

Figure 3.2, continued.

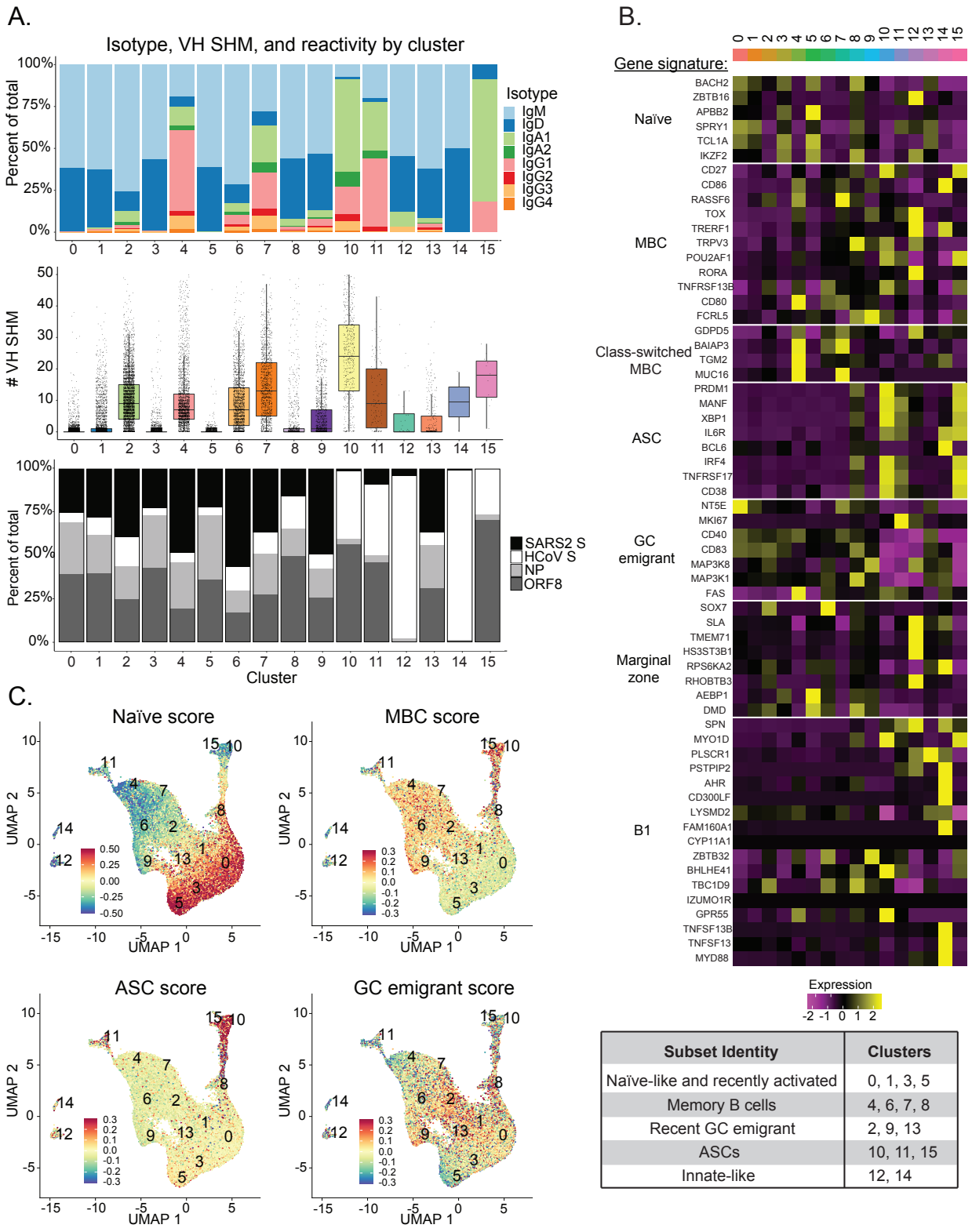
(A) Overall B cell subset distribution (top) and percentage of antigen-specific B cells targeting each antigen (bottom) for subjects sorted with SARS2 and endemic HCoV spike (S) proteins, as well as SARS2 RBD, NP, and ORF8.

(B) Overall B cell subset distribution (top) and percentage of antigen-specific B cells targeting each antigen (bottom) for subjects sorted only with SARS2 spike, RBD, NP, and ORF8. Study cohort is indicated below each graph.

(C) Integrated UMAP analysis showing cluster distribution for two severe acute subjects (R3 and R6) at pooled early (days 0, 1, 3) and late (days 7, 14) sampling time points post-convalescent plasma therapy (left) and summary of cluster distribution per timepoint (right).

(D) Distribution in antigen-reactivity for pooled early and late timepoints post-convalescent plasma therapy for severe acute subjects R3 and R6. Statistics are Chi square test, n.s.

3.4 The SARS-CoV-2-specific B cell landscape is defined by naïve-like and memory B cell subsets



Subset Identity	Clusters
Naïve-like and recently activated	0, 1, 3, 5
Memory B cells	4, 6, 7, 8
Recent GC emigrant	2, 9, 13
ASCs	10, 11, 15
Innate-like	12, 14

Figure 3.3: B cell receptor and transcriptional analysis reveals cluster identities.

Figure 3.3, continued.
(A) BCR isotype usage, VH SHM, and antigen reactivity by cluster for all integrated samples from each cohort. SHM data are plotted with the overlay indicating the median with interquartile range.
(B) Heatmap displaying differentially expressed genes across clusters. A summary of cluster identities is provided below.
(C) UMAP projections with cell color indicating gene module scoring for the indicated B cell subsets. Also see Table S5 and Table S6.
Contributions for 3.3 and 3.4: As before, in addition, Chris created the curated gene modules for each type of B cell, for the scoring algorithm, and what is shown in 3.4. Lei wrote the code to score the subsets. Chris and Haley made the figure.

To discern the identity and specificity of these B cell subset clusters, we analyzed Ig repertoire, variable heavy (VH) chain somatic hypermutation (SHM) rates, and differentially expressed genes. Different B cell clusters varied widely in their degree of class-switch recombination (CSR) and SHM, consistent with the presence of both naïve-like and memory-like B cell clusters (Figure 3.3A). Moreover, we quantitatively identified that targeting of viral antigens was variable across clusters (Figure 3.3A). To confirm B cell subset identities, we curated lists of differentially expressed genes across clusters associated with naïve B cells, MBCs, recent germinal center (GC) emigrant cells, ASCs, and innate-like B cells (including B1 B cells and marginal zone B cells) (Figure 3.3B). Clusters 0, 1, 3, and 5 expressed Ig genes with little to no SHM or CSR and gene signatures associated with naïve B cells, suggesting these subsets are comprised of naïve-like B cells or very recently activated B cells (Figure 3.3A and 3.3B). In addition, clusters with patterns of higher CSR and SHM were further investigated for memory gene signatures. Based on expression of key genes (Table 3.5 and Table 3.6), we identified clusters 4, 6, 7, and 8 as MBCs; clusters 2, 9, and 13 as recent memory or GC emigrants; clusters 10, 11, and 15 as ASCs; and clusters 12 and 14 as innate-like in nature, though genes for these subsets are not well-defined in humans (Figure 3.3A and 3.3B).

We generated scores for each cluster and projected them onto UMAP, allowing us to visualize how closely associated clusters relate to one another based on their B cell subset score (Figure 3.3C). By overlaying key gene signatures for MBCs, recent GC emigrants, and ASCs (Table 3.5), we can further visualize how the cells cluster based on identity and observe the plasticity of cells in an active immune response, as some cells seem to be outside of their home cluster, potentially capturing cells that are in the course of differentiation (Figure 3.4A–3.4C). We identified that ASC clusters 10, 11, and 15 displayed a high degree of SHM, suggesting they may derive from preexisting memory that was driven against endemic HCoV spike proteins (Figure 3.3A). These clusters also were predominantly class-switched to IgA, an isotype most associated with mucosal immunity. To explore this possibility, we mapped the expression of genes related to mucosal surface homing and found them be highly upregulated in ASC clusters, implying memory to past HCoV infection generates a large plasmablast response during SARSCoV-2 infection that re-circulates in the blood and should localize to mucosal surfaces (Figure 3.2D). In conclusion, we confirm that the landscape of B cell reactivity to SARSCoV-2 and HCoV antigens is defined by distinct naïve-like and MBC subsets.

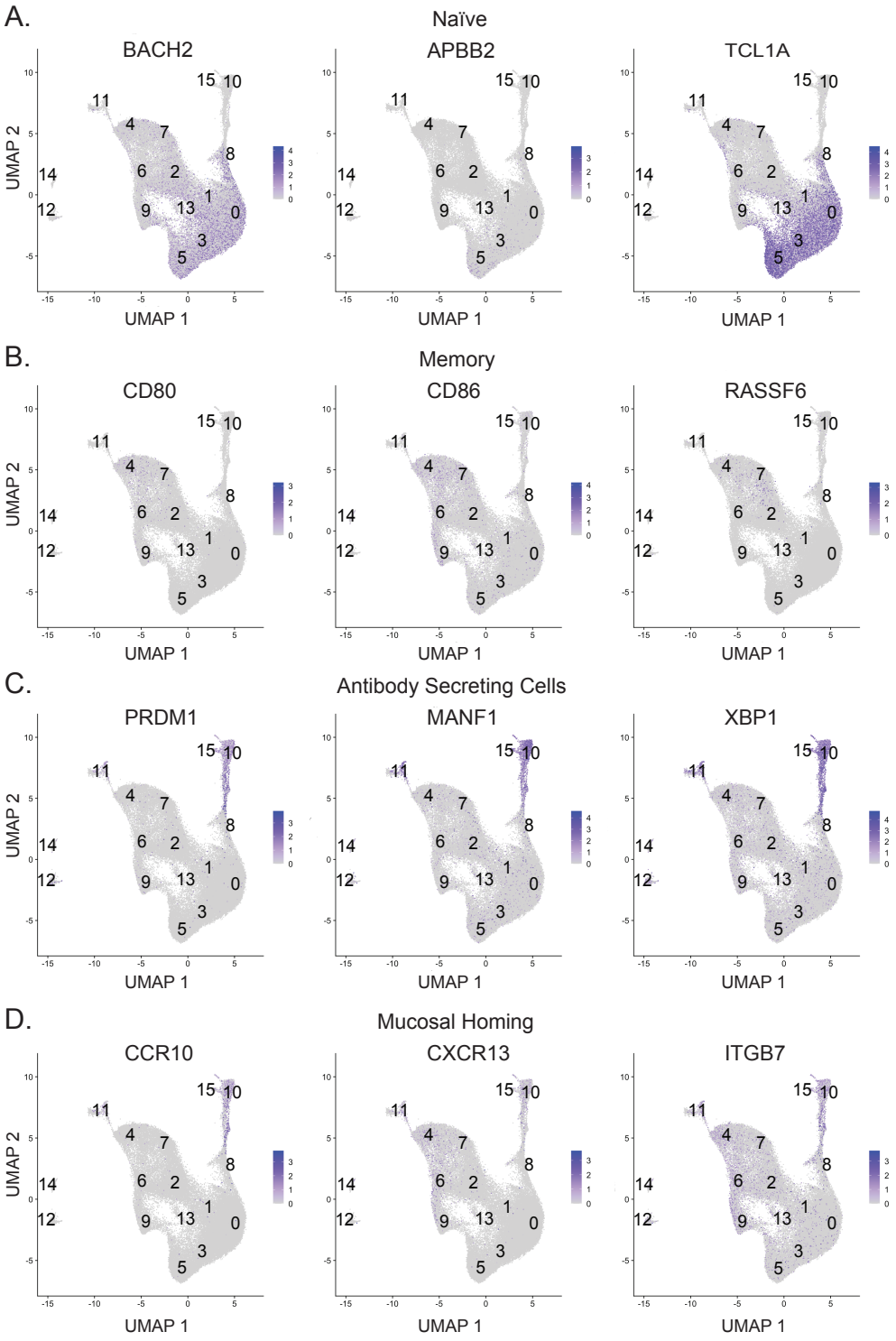


Figure 3.4: Expression maps of select genes.
(A–D) UMAP projections with cells colored by expression level of the indicated genes, associated with naïve (A), MBCs (B), ASCs (C), or mucosal homing (D). Also see Table 3.6 and Table 3.7

3.5 Evolution of the B Cell Response to Endemic HCoV and SARS-CoV-2 Protein

3.5.1 B cell immunodominance and adaptability to SARS-CoV-2 and HCoVs changes with time after infection

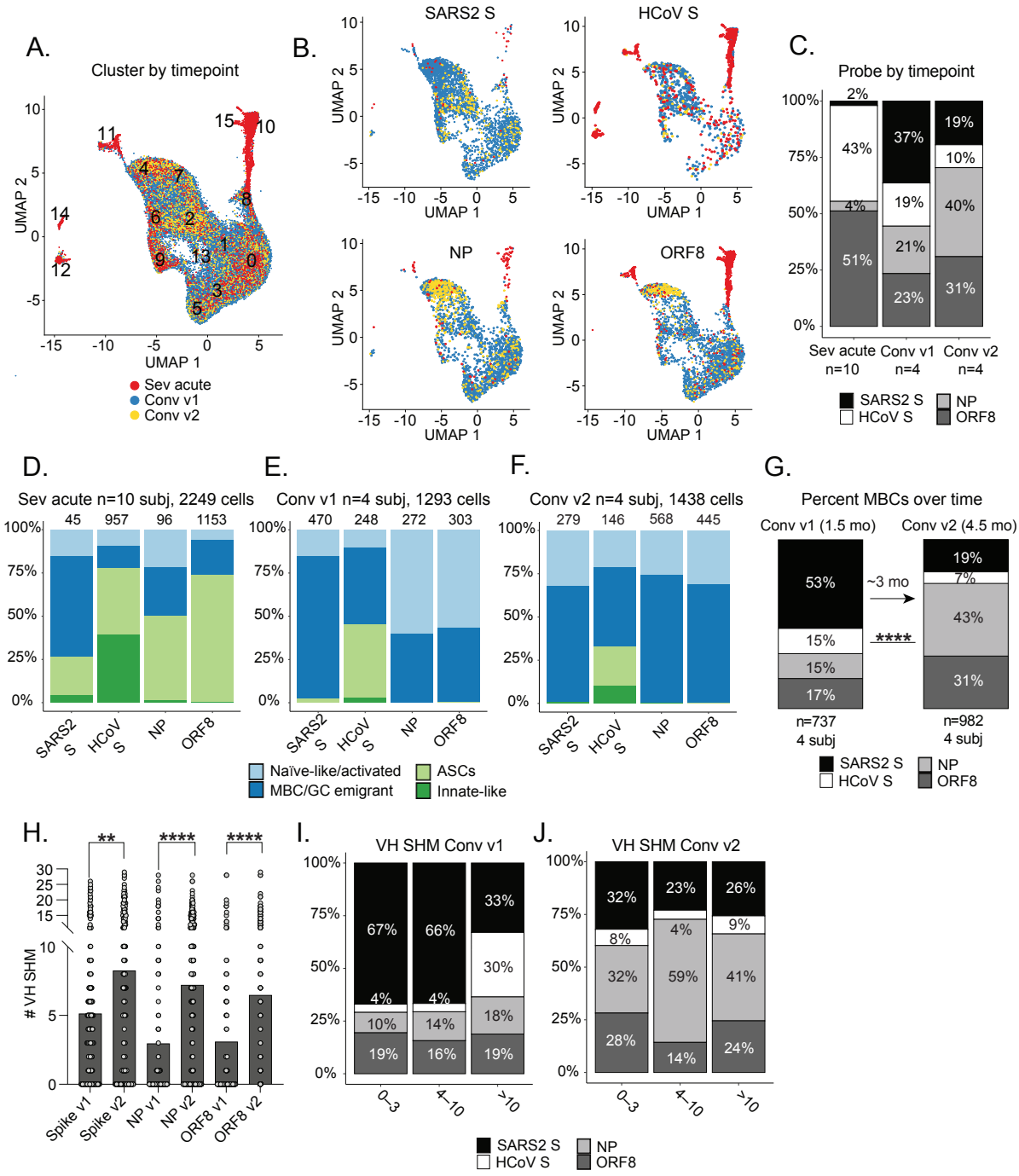


Figure 3.5: B cell immunodominance and adaptability landscapes vary in acute infection in convalescence.

Figure 3.5, continued.
(A) UMAP projection showing cells colored by cohort/timepoint of blood draw. Sev acute= Severe acute; Conv v1= convalescent visit 1; Conv v2= convalescent visit 2.
(B) UMAP projections showing only cells binding the specified antigens, colored by cohort/timepoint of blood draw.
(C) Percentage of B cells targeting distinct antigens by cohort. Four Conv v1 and Conv v2 subjects represent matched visits.
(D–F) Quantification of B cell subsets targeting distinct antigens across cohorts. Naïve-like/activated is comprised of clusters 0, 1, 3, and 5; MBC/GC emigrant is comprised of clusters 2, 4, 6, 7, 8, 9, and 13; ASCs comprise clusters 10, 11 and 15; and innate-like B cells are clusters 12 and 14. Numbers above bars indicate the number of specific cells isolated.
(G) Percentage of total antigen-specific memory B cells over time from ~1.5–3 months (mo) post-symptom onset in matched-convalescent subjects. Statistics are Chi-square test, ****$p<0.0001$.
(H) VH SHM of antigen-specific B cells across both convalescent timepoints for four matched subjects. Statistics used are unpaired non-parametric Mann-Whitney tests, ****$p<0.0001$; **$p=0.0021$.
(I and J) Antigen-specific MBCs divided by SHM tertiles at Conv v1 (I) and Conv v2 (J) for four matched subjects.
Contributions for 3.5 and 3.6: As before, in addition, Jenna Guthmiller measured the plasma / serum titers towards each antigen at timepoints 1 and 2. Chris and Haley contributed to all analysis and produced the figure.

The kinetics and evolution of B cells against the spike and non-spike antigens are poorly understood. We investigated whether there were notable changes in B cell subsets and their antigenic targets over time in acute and convalescent subjects. By color-coding cells belonging to the acute cohort (red), convalescent visit 1 (~1.5 months post-infection; blue), and convalescent visit 2 (~5 months post-infection; yellow) in the integrated UMAP, it became evident that distinct B cell subsets were enriched in different timepoints/cohorts. As expected, ASC clusters 10, 11, and 15 were derived almost entirely from acutely infected subjects (Figure 3.5A, Figure 3.6A). The two convalescent timepoints were largely comprised of naïve-like and MBC clusters, with convalescent visit 2 being the most enriched for canonical class-switched MBCs (clusters 4 and 7) (Figure 3.5A). The severe acute cohort exhibited minimal targeting of the SARS2 spike protein, and instead targeted HCoV spike and ORF8 (Figure 3.5B and 3.5C). As these ASCs

were activated by SARS-CoV-2, it appeared that these are boosted MBCs with higher affinity for the HCoV spikes, and therefore displayed B cell receptors (BCRs) predominately loaded with HCoV spike probe when stained. By contrast, convalescent visit 1 was most enriched for SARS2 spike binding, which subsequently declined in convalescent visit 2, in which the frequency of B cells to NP and ORF8 was increased in proportion (Figure 3.5B and 3.5C).

The dynamic change observed in antigen targeting over time led us to examine antigen reactivity within distinct B cell subsets for each cohort. For the severe acute cohort, B cells binding intracellular proteins were dominated by ASC clusters, whereas SARS2 spike-specific B cells were enriched in early memory and GC emigrant B cell clusters (Figure 3.5D). As previously noted, HCoV spike-specific B cells were enriched in ASCs of the severe acute cohort, indicative of re-activation of preexisting immune memory. Consistent with this, HCoV spike-specific B cells were highly mutated in the acute cohort compared to SARS2 spike-, NP-, and ORF8-specific B cells (Figure 3.6A).

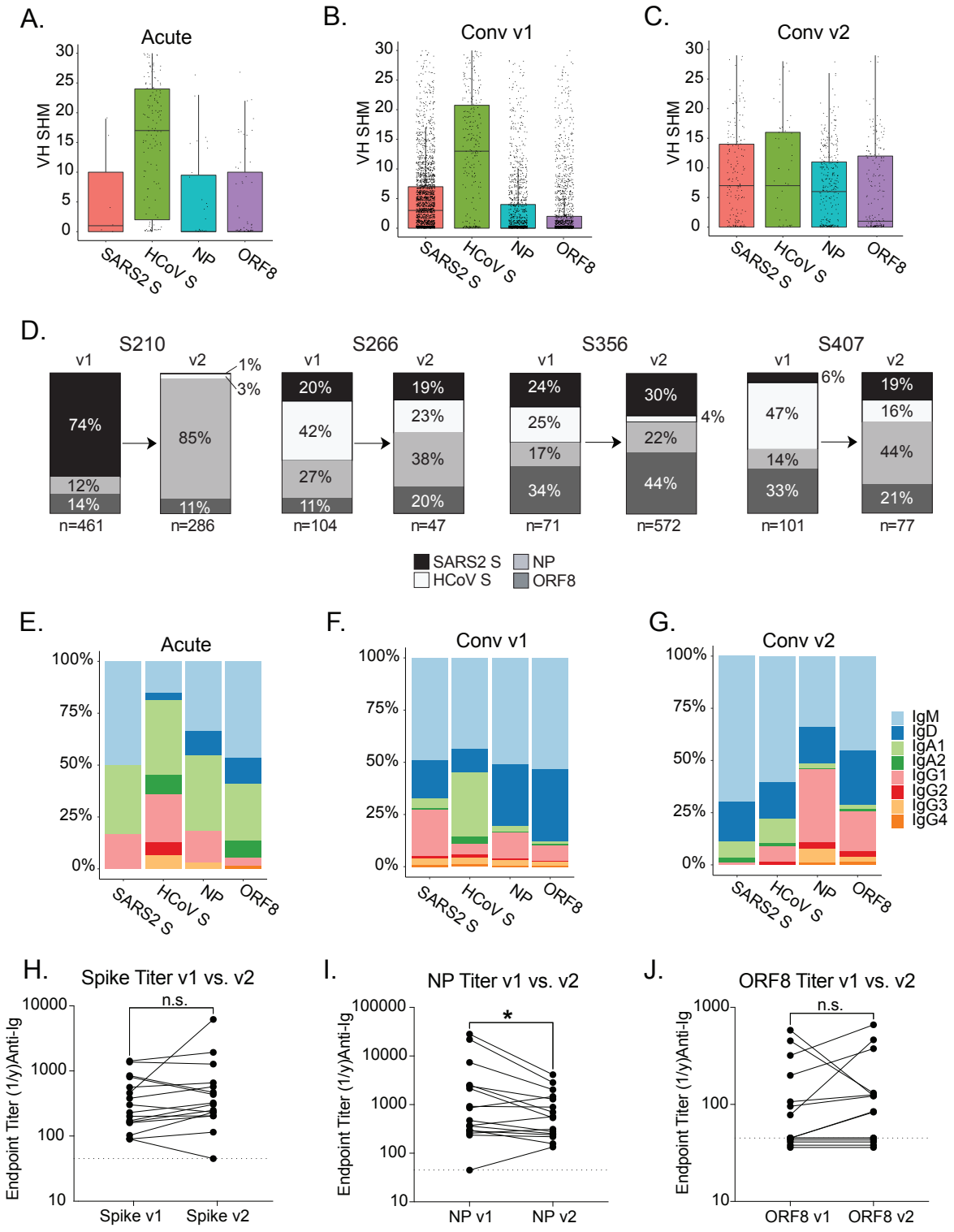


Figure 3.6: Further analysis of antigen-specific B cell properties across distinct cohorts and timepoints.

Figure 3.6, continued.
(A–C) VH SHM by antigen-specific B cells shown for severe acute (A), Conv v1 (B), or Conv v2 (C). Overlay shows median with interquartile range.
(D) Distribution of MBC specificity across timepoints for Conv v1 and Conv v2 subjects, sampled at approximately 1.5 and 4.5 months post-symptom onset. Also see Table S1.
(E–G) BCR isotype usage by antigen-specific B cells shown for severe acute (E), Conv v1 (F), or Conv v2 subjects (G).
(H–J) Total anti-Ig serum titers across timepoints for 16 matched convalescent subjects, shown for SARS2 spike (H), NP (I), and ORF8 (J). Dashed line at y= 45 indicates cutoff for positivity; values are staggered to avoid overlap. Statistics are paired non-parametric Wilcoxon test, *p=0.0386. Data are representative of two independent experiments.

3.5.2 Memory reactivity to intracellular SARS2 antigens increases over time

Across the two convalescent visits, B cells reactive to ORF8 and NP were increased in percentage and absolute numbers relative to spike B cells (Figure 3.5E–3.5G; total cell numbers indicated). While the degree of SHM for all antigen-specific B cells was increased across study visits (Figure 3.5H, Figure 3.6B and 3.6C), the B cells displaying the highest degree of SHM in convalescent visit 2 were majority NP-specific (Figure 3.5I and 3.5J). At the individual level, all four subjects displayed an increase in the percentage of MBCs to NP across time points, and half of the subjects displayed modest increases to ORF8. The change in percentage for spike-specific B cells across visits was negligible for three out of four subjects, with one subject displaying a substantial decrease (Figure 3.6D; S210). Previous groups have identified that spike-specific MBCs increase over time[175, 178, 209], and our study is limited in that this analysis was performed on only four subjects. However, our data support the claim that there is MBC maturation to NP, and to a lesser extent, ORF8 over time. While there were no major differences in isotype usage of total B cells between the two convalescent timepoints, the severe acute timepoint was comparatively enriched for IgG and IgA, largely owing to the ASC clusters (Figure 3.6C). Analyzing isotype frequencies by antigen specificity for each cohort revealed the majority of class-switched B cells were IgA in the acute cohort, regardless of antigen-reactivity (Figure

3.6D). By contrast, class-switching to IgG1 was prominent for SARS2 spike, NP, and ORF8-reactive B cells in convalescent visit 1, while HCoV spike-reactive B cells remained largely IgA (Figure 3.6E). Class-switched B cells specific to the SARS2 spike declined in convalescent visit 2, and IgG1 class-switched B cells to ORF8 and NP increased in proportion (Figure 3.6F and 3.6G).

3.5.3 Memory by probe sorting does not correlate with serum titers

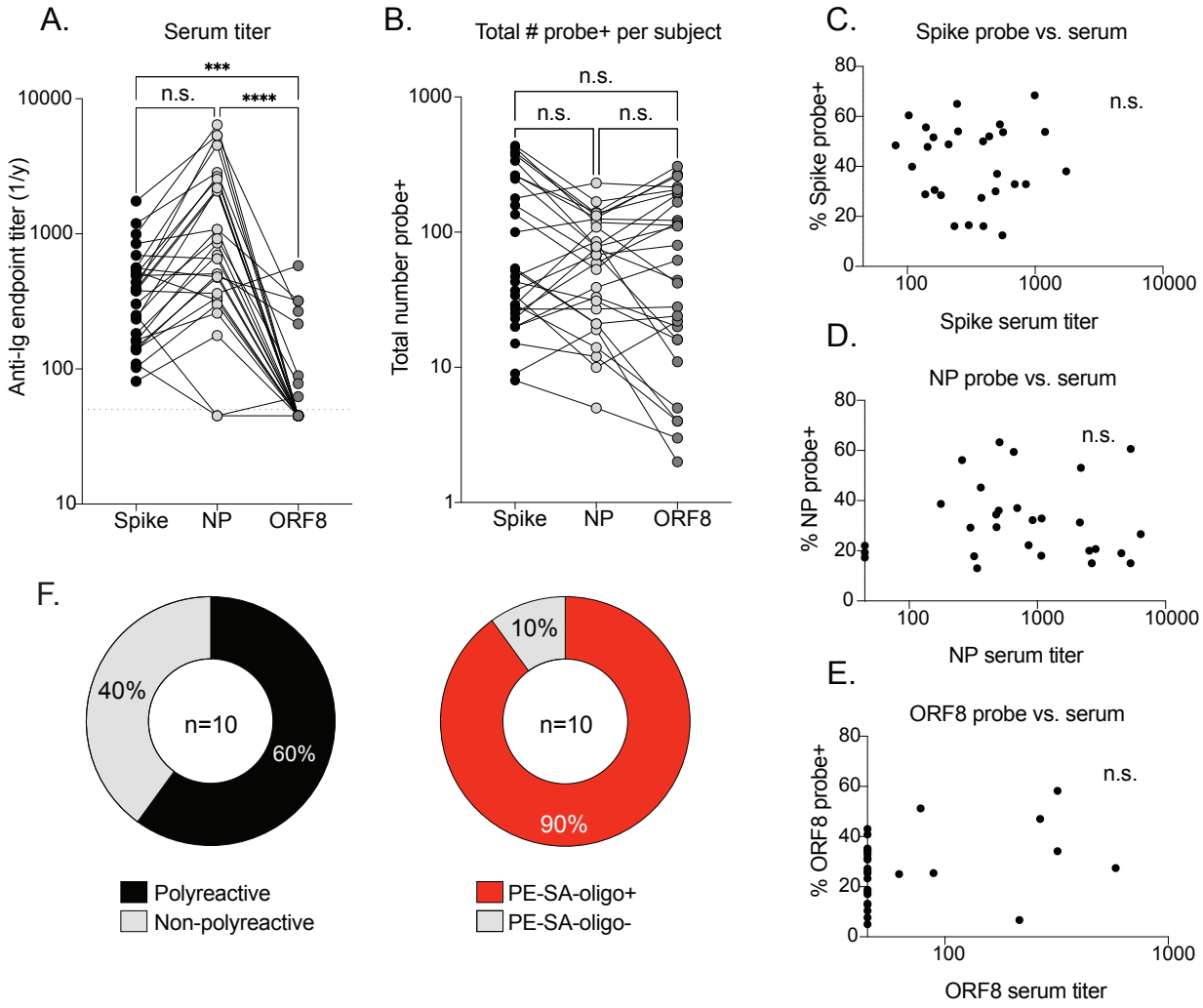


Figure 3.7: Correlation between antigen-probe positive B cells and serum titers.

Figure 3.7, continued.
(A) Matched serum titers for spike, NP, and ORF8 for Conv v1 subjects (n=28). Statistics are paired non-parametric Friedman test with Dunn's post-test for multiple comparisons, ****p<0.0001; ***p=0.0002; n.s.= not significant. Data are representative of two independent experiments.
(B) Matched antigen-specific probe hit per Conv v1 subject (n=28). Statistics are paired non-parametric Friedman test with Dunn's post-test for multiple comparisons, no differences were significant (n.s.).
(C–E) Percentage of B cells specific for spike (D), NP (E), or ORF8 (F) in Conv v1 subjects (n=28) compared to serum titer levels for the same antigen. Statistics are nonparametric Spearman correlation, two-tailed, CI = 95%. No correlations were significant (n.s.). Data are representative of two independent experiments.
(F) MAbs cloned from non-specific multi-reactive B cells tested for polyreactivity (left) and PE-SA binding (right) by ELISA (n=10). Data are representative of one independent experiment.
Contributions: Jenna measured serum titers. Chris and Haley performed the polyreactivity ELISAs. Chris performed analysis comparing probe and serum titers. Both produced the figure.

Finally, we did not identify substantial differences in serum titer to distinct antigens across convalescent visit time points (Figure 3.6H–3.6J). Similarly, reactivity patterns in serological titer and probe hit to distinct antigens in individual subjects did not appear to be correlated (Figure 3.7A–3.7E). This may be related to differences in B cell affinity to 3-dimensional probes in the bait sorting assay versus ELISA, or the fact that the cellular response is sampled at one snapshot in time (over 1 month post-symptom onset), with serology reflective of antibody that has accumulated since initial infection.

Together, our results point to differences in B cell immunodominance and adaptability landscapes across severe acute and convalescent cohorts, independent of serum titer. For both the severe acute cohort and convalescent visit 1 time point, SARS2 spike-specific B cells were initially the most enriched cells in memory. However, NP- and ORF8-reactive MBCs increased in proportion and showed signs of adaptation over time.

3.6 SARS-CoV-2 B Cells Display Unique Repertoire Features and Protective Abilities

3.6.1 B Cell Variable Gene Usage and Enrichment by Antigen

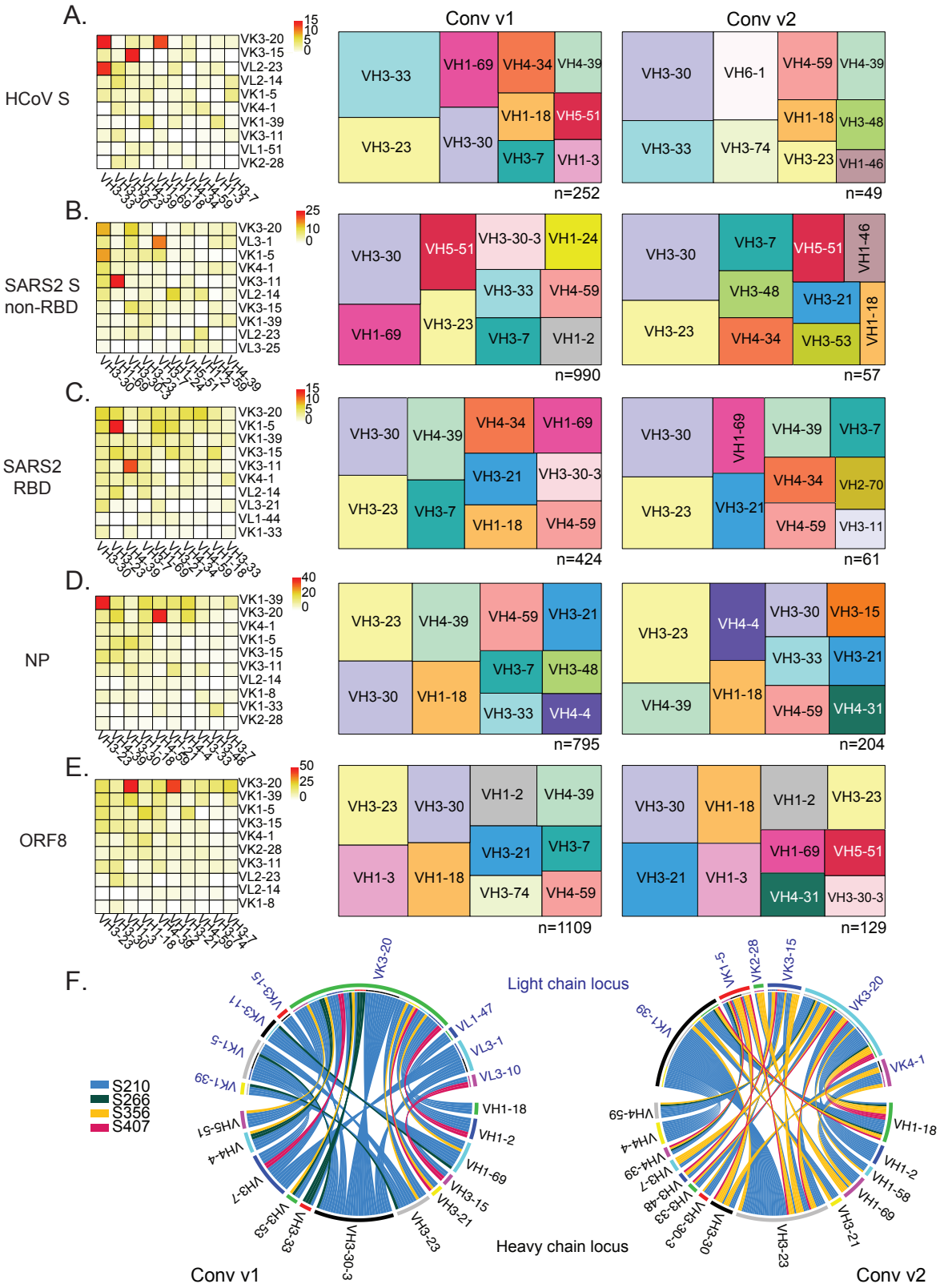


Figure 3.8: B cells targeting distinct antigens display unique variable gene usages.

Figure 3.8, continued.
(A–E) Heatmaps showing the frequency of heavy and light chain gene pairings for B cells binding the indicated antigens using integrated data from all cohorts (left; legend indicates number of cells per pairing), and dendrograms showing the top ten VH gene usages for Conv v1 (n=28) and Conv v2 (n=4) cohorts. The number of cells encompassing the top 10 VH genes represented per antigen is indicated below each dendrogram.
(F) Circos plots showing the top 10 heavy and light chain gene pairings shared across 4 matched Conv v1 (left; n=1293 cells) and Conv v2 (right; n=1438 cells) subjects. Total antigen-specific cells against SARS2 spike and RBD, HCoV spike, ORF8 and NP are shown.
Contributions: Chris and Haley originally produced dendrograms in JMP. Lei Li regenerated them in R. Lei generated heat maps and Circos plots on Chris and Haley's suggestion. Both produced the figure.

The identification of B cells against distinct antigens is typically associated with stereotypical VH and variable light chain (VK or VL) gene usages. Immunodominant and neutralizing spike and RBD epitopes are of interest, as they represent key targets for vaccine-induced responses. To investigate whether antigen-specific B cells displayed enriched variable gene usages, we analyzed VH and VK/VL pairs for B cells targeting HCoV spike, non-RBD spike epitopes, and RBD-specific epitopes. A B cell was considered non-RBD spike-specific if it bound full-length spike probe and not RBD probe, and a cell that bound both RBD and full-length spike was considered to be RBD-specific. Using this approach, we found that B cells against HCoV spike, non-SARS2 RBD spike epitopes, and the SARS2 RBD were enriched for VH1-69 gene usage (Figure 3.8A–3.8C). VH1-69 is commonly utilized by broadly neutralizing antibodies against the hemagglutinin stalk-domain of influenza viruses, as well as the gp120 co-receptor binding site of HIV-1 due to its ability to bind conserved hydrophobic regions of viral envelope glycoproteins .[210] VH1-69 usage by B cells that cross-react to SARS-CoV-2 and HCoV has also been indicated.[211] However, VH1-69 usage for B cells targeting HCoV spike and SARS2 spike non-RBD epitopes was predominantly enriched in convalescent visit 1 subjects and not convalescent visit 2, suggesting that the repertoire may continue to evolve months after

infection (Figure 3.8A and 3.8B; right). However, several VH gene usages were enriched in both convalescent visits, regardless of antigen specificity. For SARS2 spike-specific B cells that bound regions outside of the RBD, VH3-7 and VH1-24 were also commonly used, which we confirmed by characterizing cloned mAbs from our cohort (Figure 3.8B and Table 3.7). While NP-specific B cells utilized similar variable gene usages as RBD-specific B cells (Figure 3.8D), ORF8-specific B cells were enriched for VH1-2 and VH1-3 paired with VK3-20, and enrichment for these VH genes persisted across both convalescent timepoints (Figure 3.8E) Finally, by analyzing the frequency of the top 10 heavy and light chain gene pairings (total antigen-specific cells) shared across subjects for both convalescent timepoints, we observed variability amongst individual subjects and timepoints as expected (Figure 3.8F).

3.6.2 Neutralization Capacity and In-vivo Protective Ability of the Antibodies

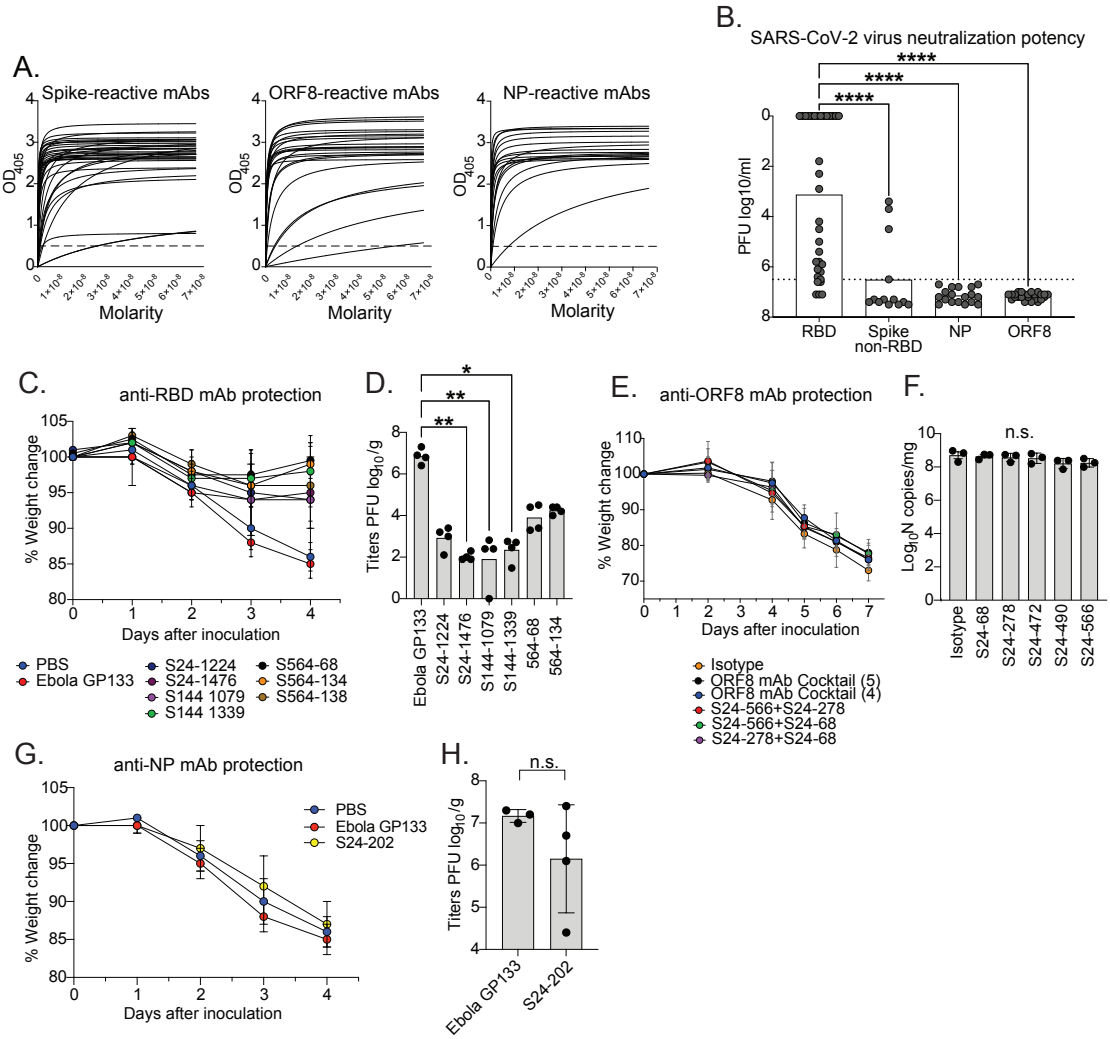


Figure 3.9: Neutralization capacity and in vivo protective ability of mAbs to the SARS-CoV-2 spike and intracellular proteins.

Figure 3.9, continued.
(A) Antigen binding curves by ELISA of antigen-specific mAbs. Dashed line at y= 0.5 on ELISA curves represents the OD_{405} cutoff of 0.5 for positivity. (spike, n= 48; RBD, n=36; NP, n=19; ORF8, n=24). Also see Table 3.7.
(B) Neutralization potency (log_{10} PFU/ml) of mAbs tested by SARS-CoV-2 virus plaque assay. Dashed line at x= 6.5 indicates the cutoff for neutralization. Statistics are non-parametric Kruskal-Wallis with Dunn's post-test for multiple comparisons, ****p<0.0001. Data are representative of one independent experiment.
(C) Weight change in hamsters intranasally challenged with SARS-CoV-2, followed by therapeutic intraperitoneal (I.P.) administration of anti-RBD antibodies (mean ± SD, n= four biological replicates for each mAb). Control conditions are PBS injection, or injection of an irrelevant Ebola virus anti-GP133 mAb.
(D) Viral titers of SARS-CoV-2 in lungs harvested from hamsters post-challenge in (C). Statistics are unpaired non-parametric Kruskal-Wallis with Dunn's post-test for multiple comparisons, **p=0.0011; **p=0075; *p=0.0135.
(E) Weight change of mice intranasally challenged with SARS-CoV-2, followed by therapeutic I.P. administration of anti-ORF8 antibody cocktails (mean ± SD, n= three biological replicates for each mAb).
(F) Viral titers of SARS-CoV-2 in lungs harvested from mice post challenge in (E). Titers are presented as N gene copy number compared to a standard curve. Statistics performed are non-parametric Kruskal-Wallis with Dunn's post-test for multiple comparisons, no differences were significant.
(G Weight change in hamsters intranasally challenged with SARS-CoV-2, followed by therapeutic intraperitoneal (I.P.) administration of an anti-NP antibody (mean ± SD, n= four biological replicates for each mAb).
(H) Viral titers of SARS-CoV-2 in lungs harvested from hamsters post challenge shown in (G). Statistics performed are non-parametric Mann-Whitney test, no differences were significant.
Contributions: Patrick Wilson, Steven, and Chris selected antibodies to be cloned and expressed. Min Huang performed cloning. Nai Ying Zheng, Jialong Wang, and Chris expressed and purified all antibodies. Chris and Haley performed all ELISAs. Plaque assays and hamster infections were performed at the University of Wisconsin-Madison. Mouse infections were performed at Washington University in St. Louis. Chris and Haley produced all panels and made the figure.

To better understand antigen specific BCRs and how antigenic reactivity relates to immune effectiveness, we next investigated the binding, neutralization potency, and in vivo protective ability of mAbs cloned from select BCRs. To do so, we expressed nearly 100 mAbs against the SARS2 spike, NP, and ORF8 from convalescent subjects, representing a multitude of clusters (Table 3.7). Cells from which to clone antibodies were chosen at random and were not chosen based on specific sequence features. However, we note that the results described herein

may be affected by sampling bias, as only a small subset of antigen-specific mAbs were cloned. We confirmed that cells designated as specific bound with moderate to high affinity to their corresponding antigens (Figure 3.9A), and cells identified as multi-reactive exhibited features of polyreactivity or bound to PE (Figure 3.7F). We next tested the antibodies for viral neutralization by SARS-CoV-2/UW-001/Human/2020/Wisconsin virus plaque assays, where lower plaque forming units (PFU)/ml equates to increased neutralization. Whereas 82% percent of mAbs to the RBD were neutralizing including 42% exhibiting complete inhibition, only 23% of mAbs to spike regions outside of the RBD were neutralizing, and these showed relatively low potency (Figure 3.9B). NP- and ORF8-specific mAbs were entirely non-neutralizing (Figure 3.9B). Using animal models of SARS-CoV-2 infection, we confirmed that anti-RBD antibodies were therapeutically protective in vivo, preventing weight loss and reducing lung viral titers relative to PBS control and an irrelevant Ebola anti-GP133 mAb (Figure 3.9C and 3.9D).

While mAbs to NP and ORF8 were non-neutralizing in vitro, they might still provide protection in vivo, potentially through Fc-mediated pathways if the proteins were exposed on the virus or cell surface at appreciable levels. However, neither ORF8-reactive mAbs nor NP-reactive mAbs conferred protection from weight loss or viral infection in the lung in vivo (Figure 3.9E–3.9H). Altogether, our data suggest that while B cells may continue to expand and evolve to intracellular antigens upon SARS-CoV-2 infection, B cell responses against these targets may not provide substantial protection from re-infection.

3.7 SARS2 Antigen Specificity and B Cell Subset Distribution is Linked to Clinical Features

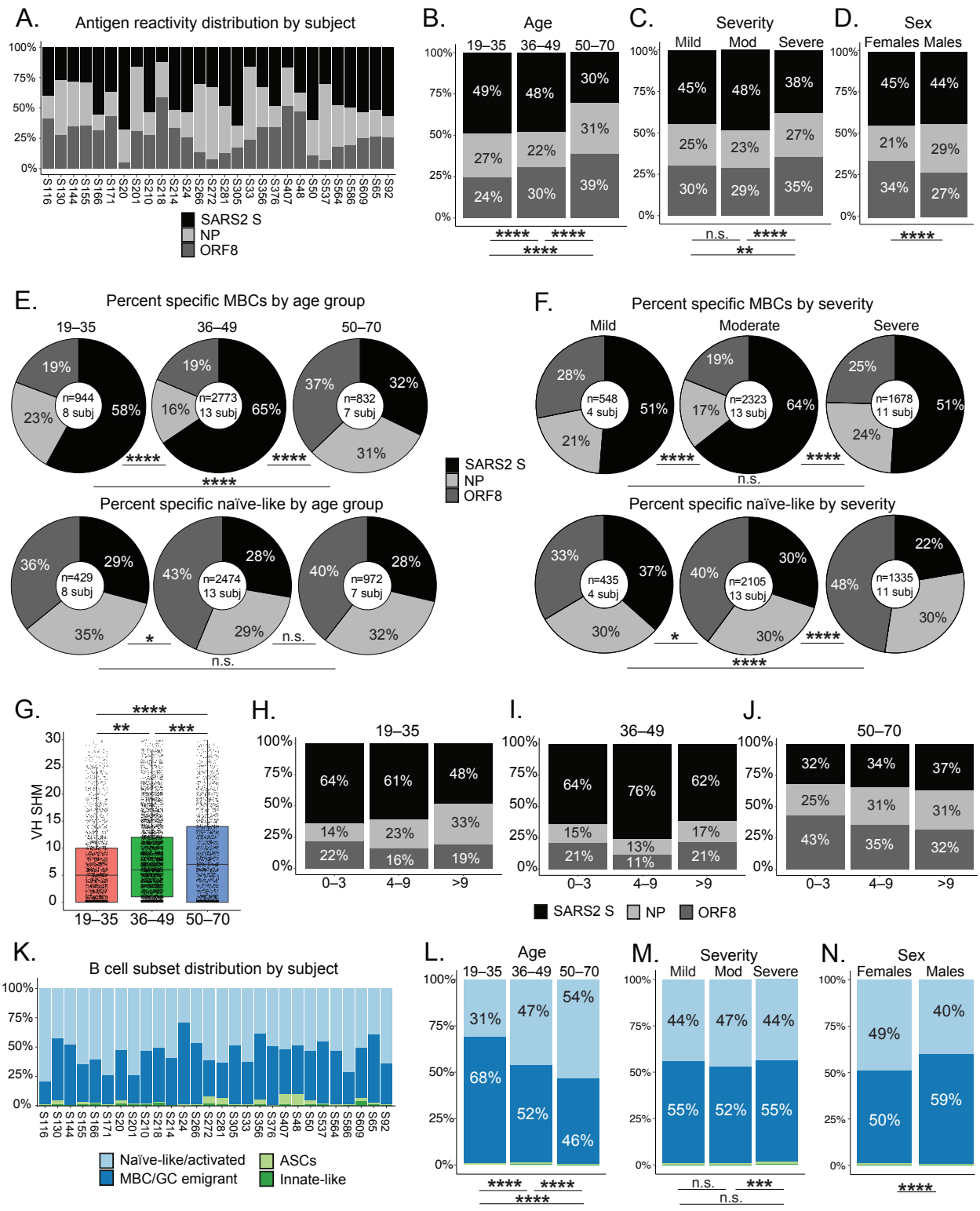


Figure 3.10: Antigen-specificity and B cell subset distribution is linked to clinical features.

Figure 3.10, continued.
(A) Reactivity distribution of total antigen-specific B cells by subject for the convalescent visit 1 cohort (n= 28).
(B–D) Reactivity distribution of total antigen-specific B cells by age (B), disease severity (C), and sex (D). Statistics are Chi-squared post-hoc tests with Holm-Bonferroni adjustment, ****p<0.0001; **p=0.0012; n.s.= not significant. For age groups, 19–35: n= 1382 cells, 8 subjects; 36–49: n= 5319 cells, 13 subjects; 50–70: n= 1813 cells, 7 subjects. For severity groups, mild: n= 990 cells, 4 subjects; moderate: n= 4462 cells, 13 subjects; severe: n= 3062 cells, 11 subjects. For sex, females: n= 5005 cells, 14 subjects; males: n= 3509 cells, 14 subjects.
(E) Reactivity of antigen-specific memory B cells (top) or naïve B cells (bottom) by age group. Statistics are Chi-squared post-hoc tests with Holm-Bonferroni adjustment, ****p<0.0001; *p=0.0145; n.s.= not significant.
(F) Reactivity of antigen-specific memory B cells (top) or naïve B cells (bottom) by disease severity. Statistics are Chi-squared post-hoc tests with Holm-Bonferroni adjustment, ****p<0.0001; *p=0.0143; n.s.= not significant.
(G) VH SHM for MBCs by age group (overlay shows median with interquartile range). Statistics are unpaired non-parametric ANOVA with Tukey's test for multiple comparisons, ****p<0.0001, ***p=0.0008, **p=0.002.
(H–J) Antigen-specific MBCs by age, divided by SHM tertiles.
(K) B cell subset distribution by subject. Naïve-like/activated subsets are comprised of clusters 0, 1, 3, and 5, MBC/GC emigrant is comprised of clusters 2, 4, 6, 7, 8, 9, and 13, ASCs are defined by clusters 10 and 11, Innate-like clusters are 12 and 14.
(L–N) B cell subset distribution by age (L), disease severity (M), and sex (N). Statistics are Chi-squared post-hoc tests with Holm-Bonferroni adjustment, ****p<0.0001; ***p=0.0007; n.s.= not significant. For each group, n is the same as in (B–D).
Contributions 3.10 and 3.11: Chris and Haley suggested all analysis and Lei wrote the code to perform it in R. Lei produced figures in R and Chris and Haley modified them (text and labels to match journal requirements) and made the figure in Illustrator. Chris performed the analysis in 3.12 of age vs ORF8 reactivity.

Serum antibody titers to the spike and intracellular proteins are shown to correlate with age, SARS-CoV-2 severity, and sex.[171, 176, 212] We therefore analyzed the distribution of B cell subsets and frequencies of B cells specific to the spike, NP, and ORF8 in convalescent subjects stratified by age, severity of disease, and sex. Disease severity was stratified into three categories: mild, moderate, and severe, based on symptom duration and symptoms experienced (Table 3.1). Our disease severity scoring method has been defined previously.[176]

We found that reactivity of total B cells toward different antigens varied widely by subject, likely reflecting host-intrinsic differences (Figure 3.10A). With age, we identified a

decrease in the generation of spike-specific B cells, and an increase in ORF8 and NP-specific B cells (Figure 3.10B). Similarly, the percentage of total spike-specific B cells was reduced in subjects with more severe disease, whereas ORF8 specific B cells were increased (Figure 3.10C). Lastly, we identified females had increased percentages of ORF8 reactive cells, whereas males showed slightly greater percentages of NP reactive cells (Figure 3.10D). To address whether differences in B cell reactivity with age and severity were associated with naïve-like or MBC subsets, we analyzed reactivity by subset. We observed a substantial decrease in spike-specific MBCs and an increase in NP and ORF8 reactive MBCs with age, while naïve-like B cell subsets were more evenly distributed in reactivity across age groups (Figure 3.10E, Figure 3.11A). Notably, we identified a significant correlation with age and the percentage of ORF8 reactive MBCs in females, but not in males (Figure 3.10B and 3.10C). By contrast, the generation of specific MBCs was not different between mild and severe cases, though naïve-like subsets targeting ORF8 were increased across mild, moderate, and severe disease (Figure 3.10F, Figure 3.11D).

While B cell memory to the spike was decreased in older patients, the overall median number of VH SHMs for antigen-specific MBCs was increased relative to younger patients (Figure 3.11G). However, the majority of MBCs harboring the most mutations targeted the SARS2 spike in younger age groups (Figure 3.10H and 3.10I), mutated MBCs against NP and ORF8 were proportionately increased relative to the spike in older patients (Figure 3.10J). Finally, we observed variability in the percentages of MBCs and naïve-like B cells across subjects (Figure 3.10K), with older patients, patients with severe disease, and female patients generating reduced percentages of MBCs (Figure 3.10L–3.10N). These findings strongly point to older patients exhibiting poorly adapted MBC responses to the spike, instead exhibiting

increased targeting and adaptation to intracellular antigens. These data are analogous to B cell responses to influenza virus vaccination in the elderly and may be attributed to the effects of immunosenescence impairing the ability to form new memory over time.[208, 213] Alternatively, these findings may reflect potential effects of preexisting immunity on the boosting of NP-specific cross-reactive MBCs.

In summary, our study highlights the diversity of B cell subsets expanded upon novel infection with SARS-CoV-2. Using this approach, we identified that B cells against the spike, ORF8, and NP differ in their ability to neutralize and derive from functionally distinct and differentially adapted B cell subsets; that memory B cell output overtime shifts from the spike to intracellular antigens; and that targeting of these antigens is impacted by age, sex, and disease severity.

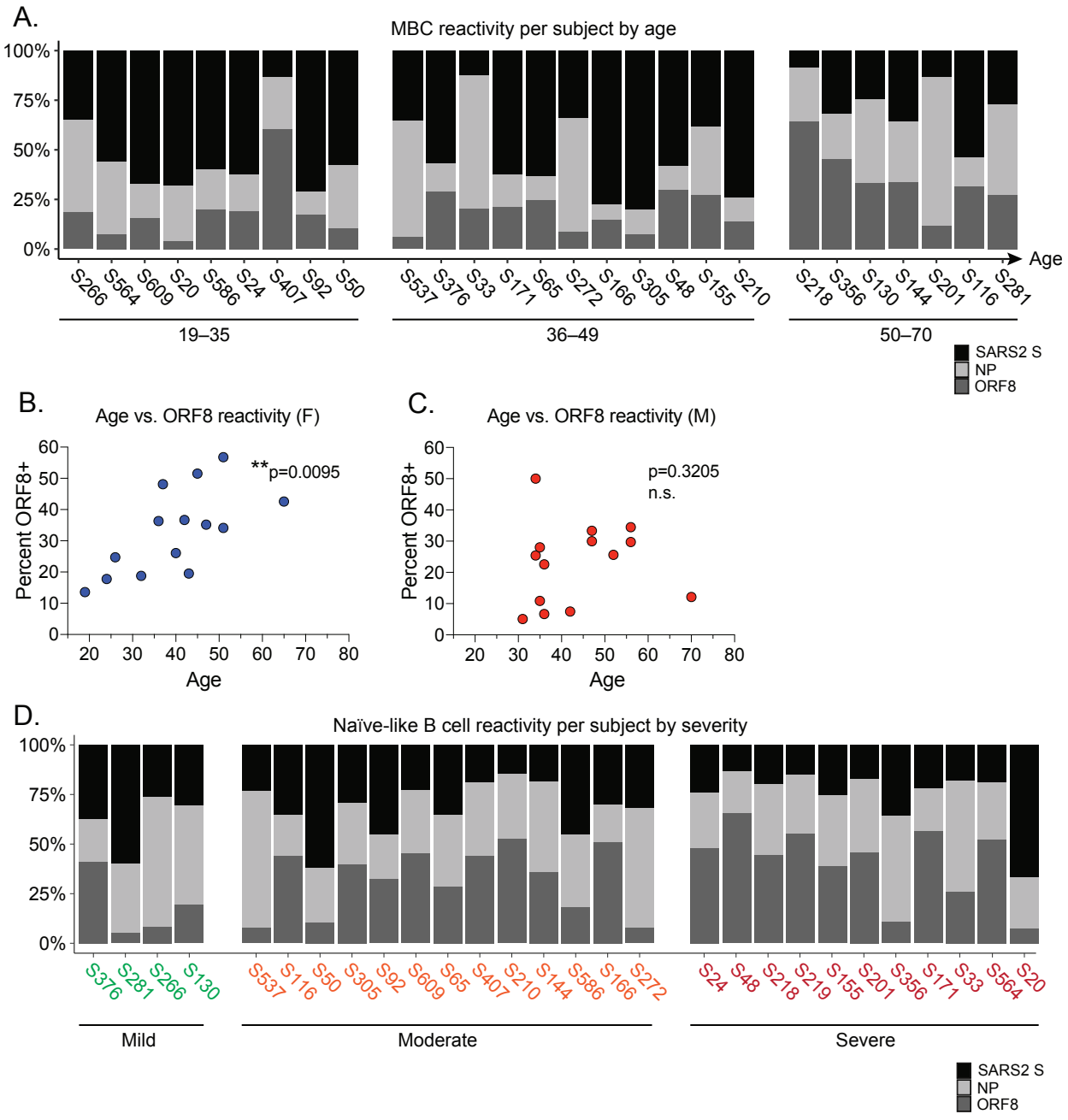


Figure 3.11: Additional analyses of antigen reactivity by clinical parameter

Figure 3.11, continued.
(A) Percentages of antigen-specific MBCs shown per conv v1 subject by age. Age increases left to right along the graph.
(B) Percentage of MBCs specific for ORF8 versus age for female (F) conv V1 subjects. Statistics are nonparametric Spearman correlation, two-tailed, CI = 95%.
(C) Percentage of MBCs specific for ORF8 versus age for male (M) conv V1 subjects. Statistics are nonparametric Spearman correlation, two-tailed, CI = 95%. P values are indicated.
(D) Percentages of antigen specific naïve-like B cells shown for each conv v1 subject by severity. Severity score increases left to right along the graph, also see Table S1 for severity score per subject.

3.8 Novel Public Clones

B cell clone	Clonal Pool	Antigen	VH gene	DH gene	JH gene	LC V gene	LC J gene	HC CDR3 AA sequence	LC CDR3 AA sequence
S144-121	1	Spike/RBD	3-23*01	N/A	4*02	k3-20*01	k1*01	AKGSSTARPYYFDY	QEYGSSRM
S155-37	1	Spike/RBD	3-23*01	6-13*01	4*02	k3-20*01	k1*01	VKGSAAARPYYFDY	QQYGNSRI
S24-1047	2	Spike/RBD	1-46*01	7-27*01	4*02	l2-8*01	l3*02	ARDGTHWDFDF	NSYKRGNTWV
S24-342	2	Spike/RBD	1-46*01	7-27*01	4*02	l2-8*01	l3*02	ARDGHHWDFDF	NSYTDRNKWV
S210-896	3	Spike	3-30-3*01	1-7*01	4*02	k3-20*01	k3*01	ARGHGNYLTYFDY	QQYGSSPLT
S376-2486	3	Spike	3-30-3*01	1-26*01	4*02	k3-20*01	k4*01	ARGRGNYFTYFDY	QQYGGSLT
S166-2620	4	Spike	3-7*03	6-19*01	4*02	l3-1*01	l2*01	ARDSIAVAGGLDY	QAWDSSTVV
S166-1318	4	Spike	3-7*03	6-19*01	4*02	l3-1*01	l2*01	ARDGIAVAGGFDY	QAWDSSTVV
S171-1150	4	Spike	3-7*01	6-19*01	4*02	l3-1*01	l2*01	ARDGIAVAGGLDY	QAWDSSTVV
S210-852	4	Spike	3-7*01	6-19*01	4*02	l3-1*01	l2*01	ARDGIAVAGGFDY	QAWDSSTSVV
S305-968	4	Spike	3-7*03	6-19*01	4*02	l3-1*01	l2*01	ARDSIAVAGGFDY	QAWDSSTNVV
S564-128	5	NP	3-7*01	1-26*01	4*02	k3-15*01	k2*01	ARGDGSNSGIYFDS	QQYNYWYT
S469-373	5	NP	3-7*03	6-6*01	4*02	k3-15*01	k2*01	ARGGGSSSGLYFES	QQYNYWYT
S144-292	6	Spike	5-51*01	2-21*02	4*02	k1-5*01	k1*01	ARLFCGGDCPFDY	QQYNTYPRT
S2141-65	6	Spike	5-51*01	2-21*02	4*02	k1-5*01	k1*01	ARQFCGGDCPFDY	QQYNSYPRT
S144-1364	6	Spike	5-51*01	3-10*01	4*02	k1-5*01	k2*01	ARPNYYGSGSPPGY	QQYNSYYT

Table 3.1: Novel Public Clones
B cells found in at least two subjects using BCRs of the same sequence that have registered as positive for the same probe. Sequences had to be at least 98% similar to be considered a public clone.
Contributions: Chris identified all potential public clones and ordered the antibodies. Min cloned them. Chris expressed them and performed the ELISAs.

From our dataset we have discovered five public clonal expansions. Public clones represent exciting targets of vaccination strategies, because it is assumed the germline VDJ

combinations they arise from must be more prevalent in the human population.[214] We attempted to clone and express all 5 clonal pools as monoclonal antibodies, however clonal pool 2 did not properly express. Additionally, clonal expansions 1 and 4 didn't bind by ELISA, however both are IgM, and it is possible that their affinity when expressed monoclonally is below the level of detection by ELISA. Anitbodies generally secreted as multimers appear weaker when expressed as monoclonals, and then has been confirmed for anti-SARS-CoV-2 antibodies as well.[215] Clonal pools 3, 5, and 6 all bound, but weakly. However, all are either lowly mutated, IgM, or both. These clones potentially represent the intriguing possibility that their affinity can be boosted with additional rounds of affinity maturation. Most notable is clonal pool 4, which binds a non-RBD target. Recent reports indicate non-RBD targets are often more broadly binding to emerging variants and can retain protectiveness. As such, this could represent a class of stalk binding public antibodies that are of interest for future efforts for universal coronavirus vaccination efforts.

3.9 Additional Analysis of Varibale Gene Usage by Antigen

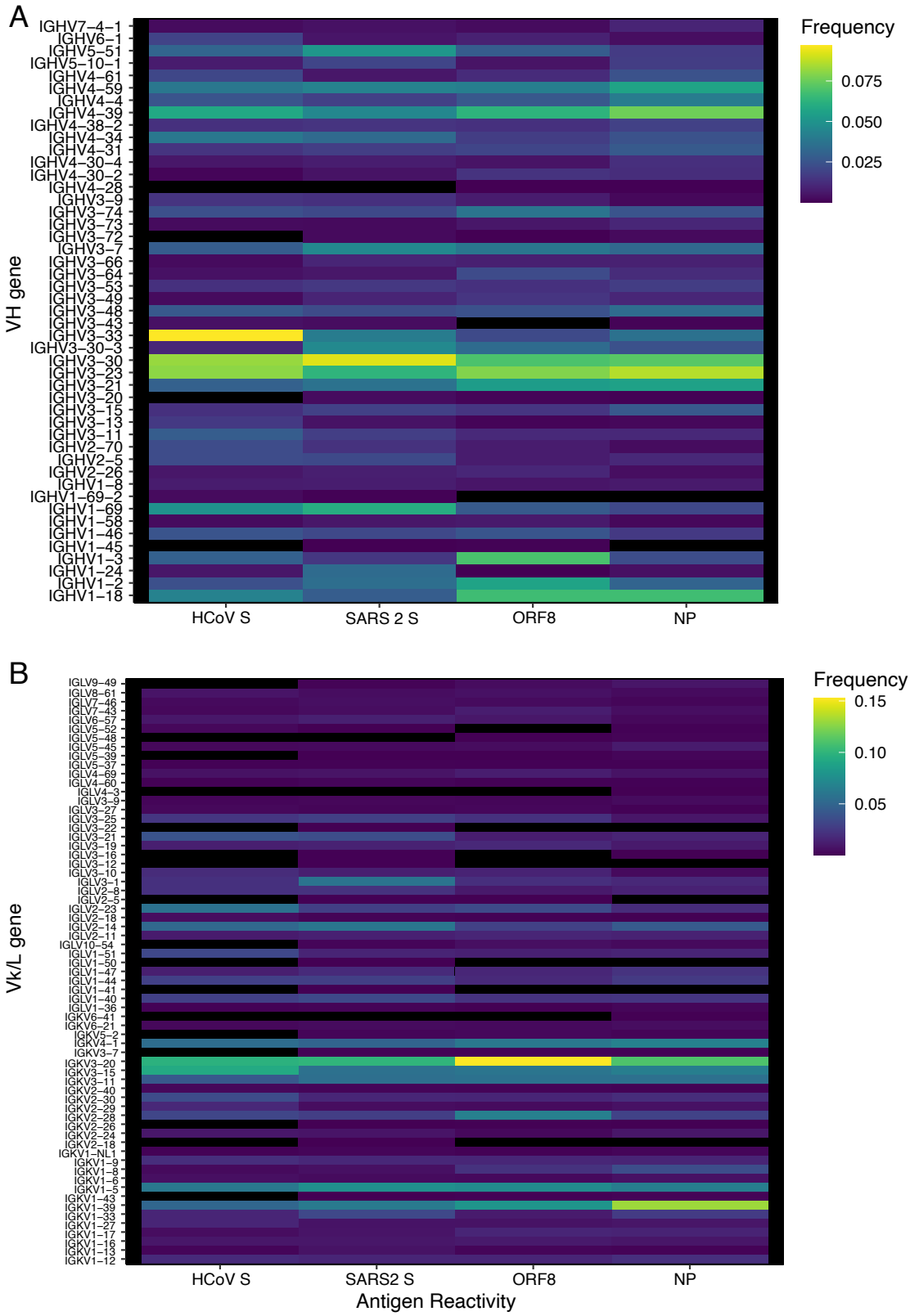


Figure 3.12: Frequency of Vgene usage by antigen reactivity.

Figure 3.12, continued.
(A) Frequency of heavy chain variable gene usage by antigen positive B cells.
(B) Frequency of light chain variable gene usage by antigen positive B cells.
Contributions: All figures in chapter 3 beyond this point were created solely by Chris. He wrote all code for analysis, produced all panels, and created all figures.

After having previously investigated top gene pairings by cluster, we decided to further investigate the repertoire within our dataset. As expected, B cells binding different probes have different patterns of V gene usage (Fig 3.12 A and B) while also showing some similarities in having a high frequency of BCRs utilizing VH3-23 and VH3-30, two of the most commonly used variable genes in human repertoires.[14] As previously discussed, HCoV and SARS2 spike reactive cells both show an enrichment for VH1-69, however SARS 2 reactive cells also show an increased usage of VH5-51 compared to other cells. A heightened frequency of VH5-51 usage in anti-S antibodies has not been previously reported. A neutralizing VH5-51 antibody has been crystalized when bound to spike, and the structure reveals it binds the RBD with high affinity.[216]

Because the endemic HCoV reactive B cells we were able to sort largely represent ASCs or MBCs activated in response to infection with COVID-19 it can be inferred that they are largely cross reactive and represent intriguing targets for additional study and as targets for potential universal vaccination efforts. As such, BCRs utilizing genes such as VH3-33, VH4-34, VH4-39, and VH3-30, in addition to VH1-69, should be further explored and characterized. The same is true of BCRs using light chains such as Vk3-15, and VL2-14 and 2-23.

3.10 Mutational Analysis Can Reveal Previously Unappreciated Signatures Within Large B Cell Datasets

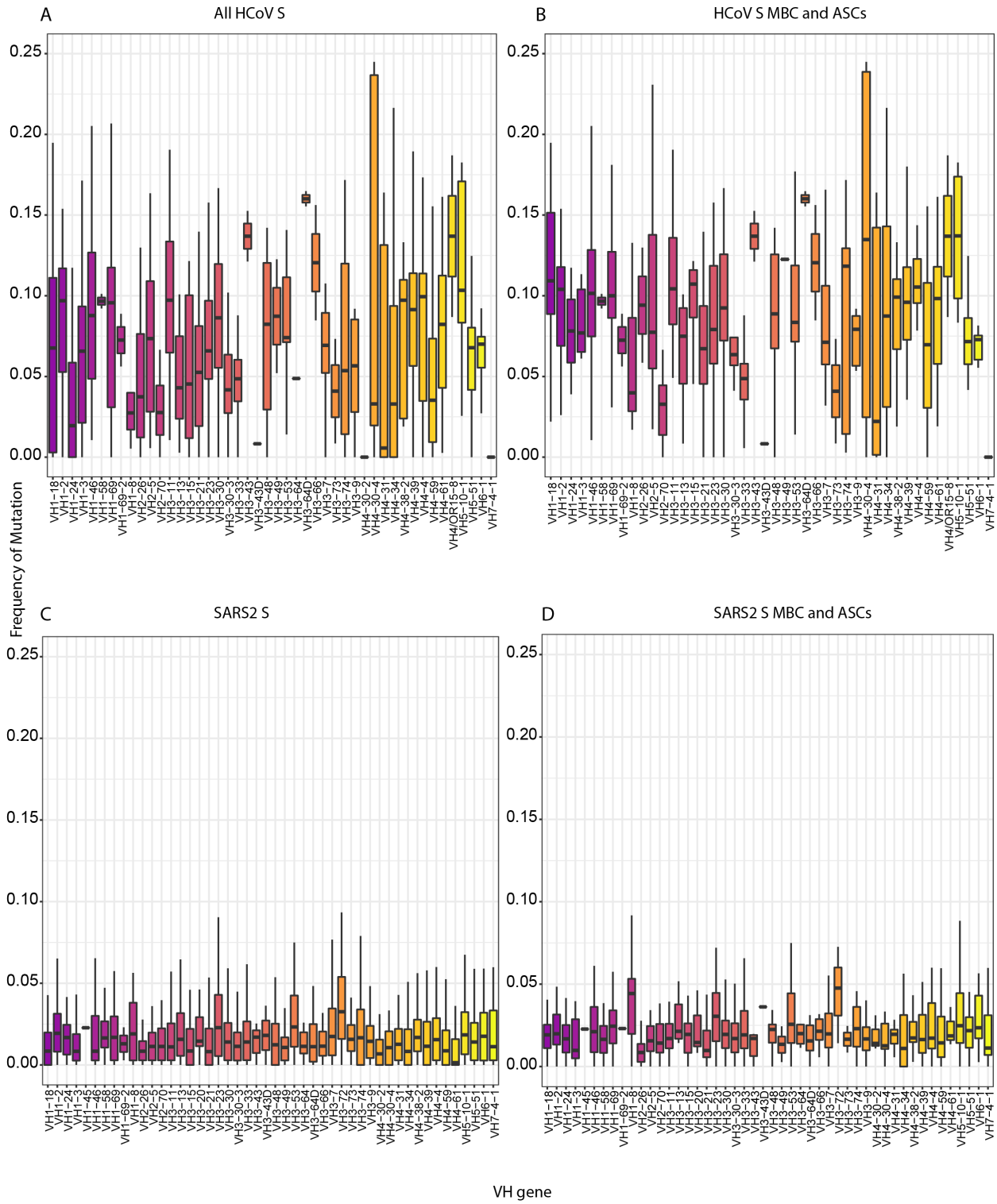


Figure 3.13: Mutational frequency by Vgene shows substantial maturation of HCoV S reactive BCRs

Figure 3.13, continued.
(A) Frequency of mutation, both silent and replacement, plotted by Vgene for all endemic HCoV S reactive B cells.
(B) Frequency of mutation, both silent and replacement, plotted by Vgene for endemic HCoV S reactive B cells transcriptionally identified as MBCs or ASC.
(C) Frequency of mutation, both silent and replacement, plotted by Vgene for all SARS2 S reactive B cells.
(D) Frequency of mutation, both silent and replacement, plotted by Vgene for SARS2 S reactive B cells transcriptionally identified as MBCs or ASC.

Though variable gene usage analysis can be informative, the addition of mutational analysis on top of it can more deeply reveal features of the response, in a way that could not previously be achieved with more traditional techniques of antibody cloning and characterization. To return to to previously mentioned genes, VH3-33 does not show a high frequency of mutation and therefore is probably not often selected for and its frequent usage is perhaps more reflective of how common it is in the repertoire in general. Alternatively, it could mean they represent a subset of B cells not previously activated that represent more of a blank canvas for vaccination strategies and maturation. VH4-39 and VH1-69 show substantial evidence of maturation, implying they are potentially high affinity (Fig 3.13 A and B). Whereas for SARS2 S reactive antibodies, VH5-51 shows very little evidence of maturation, even though it is enriched for usage, again potentially revealing a large population of naive B cells that could be targeted for additional maturation. Additionally, B cells using VH1-24 and VH3-72 show substantial amounts of maturation, even though both genes are so rarely used they have not been identified as variable genes of interest in analysis of the repertoire alone (Fig 3.13 C and D). Intriguingly, we cloned two VH1-24 anti-spike antibodies, one of which (S305-1456) is our most potently neutralizing anti-S, non-RBD antibody tested to date (Fig 3.9B).

The non-neutralizing, intraviral antigens also show interesting mutational patterns. NP is marked by a large accumulation of mutations several VH3 family genes (Fig 3.14 A and B).

VH4-30-2 and -4 immediately stand out for ORF8, showing a very high frequency of mutation within memory and ASC populations compared to all other variable genes. Indeed, there show a higher frequency of mutation than variable genes being used to bind any other antigen apart from HCoV S and deserve further scrutiny (Fig 3.14 C and D). VH4-61 and VH3-11 also show heightened mutation amongst B cells binding to ORF8.

One problem with such a high throughput method as ours is that while it can be a powerful tool for the discovery and characterization of antibodies, it is practically impossible to clone and test thousands of monoclonal antibodies in a short amount of time; if it is ever feasible or possible to clone the entire data set. Therefore, as this approach becomes more widely adopted and applied to other diseases it will be imperative to have reliable methods for selecting which antibodies to clone. We initially tried to focus on memory subsets and cells that had a high probe score, meaning they had a higher copy of oligo transcripts associated with them, assuming there was a possibility that it might imply greater affinity. This correlation has not been proved to hold true, though it is interesting some of them overlap with cells that seem use variable genes with signs of selection. However, future users of this technique should consider selecting antibodies based on mutational loads on the corresponding variable gene, as it could help to more rapidly identify what is being most selected within germinal centers.

Finally, it is clearly apparent that endemic HCoV reactive B cells are carrying substantially higher loads mutations than all others. This is interesting for several reasons. Recent work in the field has called into question how often memory B cells re-enter germinal centers, and this potentially implies it happens with a greater frequency in humans than murine models may indicate.[83] Especially when considered in light of the fact that these cells are risen against one of the most commonly encountered infections, further implying a cycle of infection

and germinal center re-entry of cells. These highly mutated HCoV reactive cells warrant further investigation.

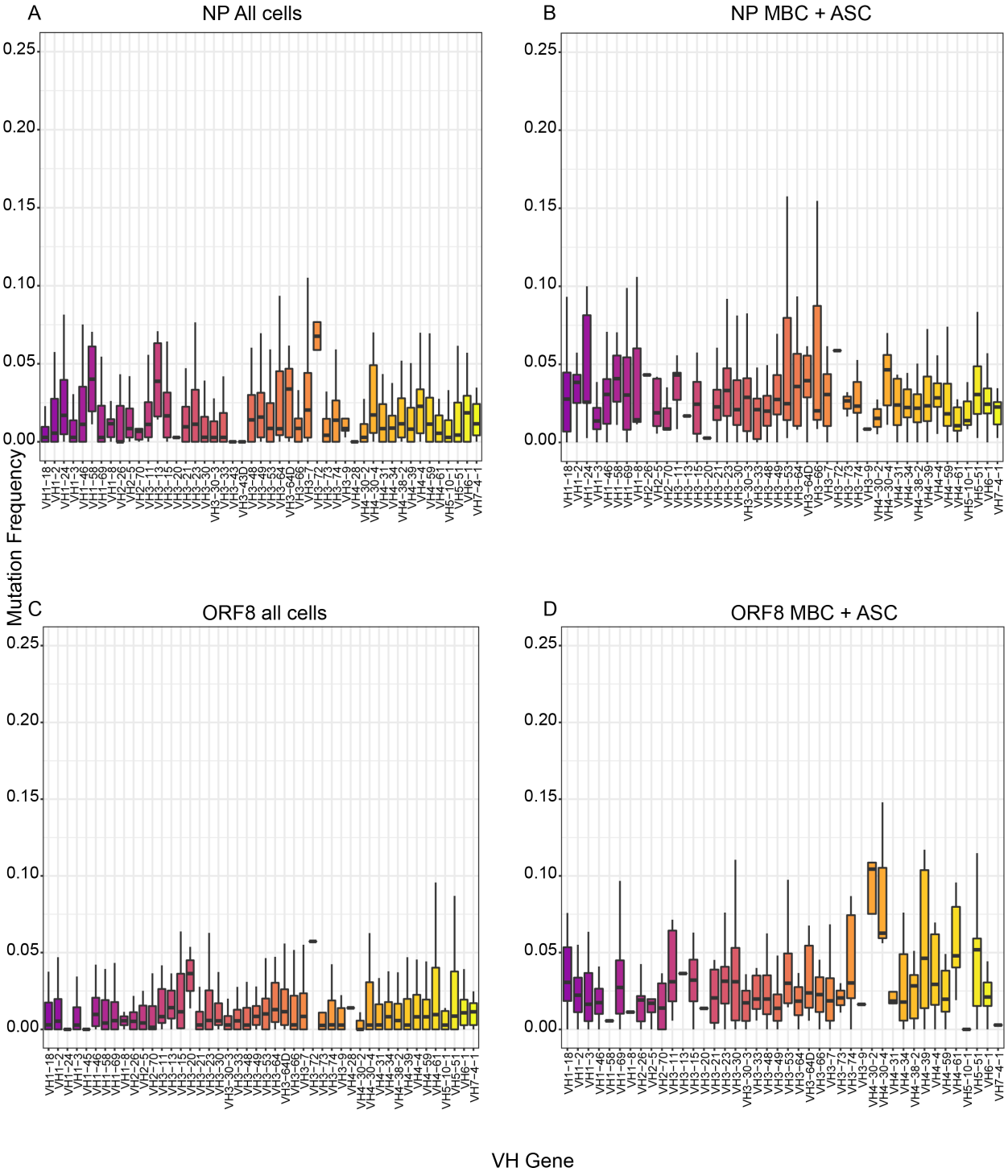


Figure 3.14: Mutational frequency by Vgene reveal differences across intraviral antigens

Figure 3.14, continued.
(A) Frequency of mutation, both silent and replacement, plotted by Vgene for all ORF8 reactive B cells.
(B) Frequency of mutation, both silent and replacement, plotted by Vgene for ORF8 reactive B cells transcriptionally identified as MBCs or ASC.
(C) Frequency of mutation, both silent and replacement, plotted by Vgene for all NP reactive B cells.
(D) Frequency of mutation, both silent and replacement, plotted by Vgene for NP reactive B cells transcriptionally identified as MBCs or ASC.

3.11 Mutation Patterns Reveal Endemic HCoV Reactive B cells Are Highly Mutated in Unusual Patterns

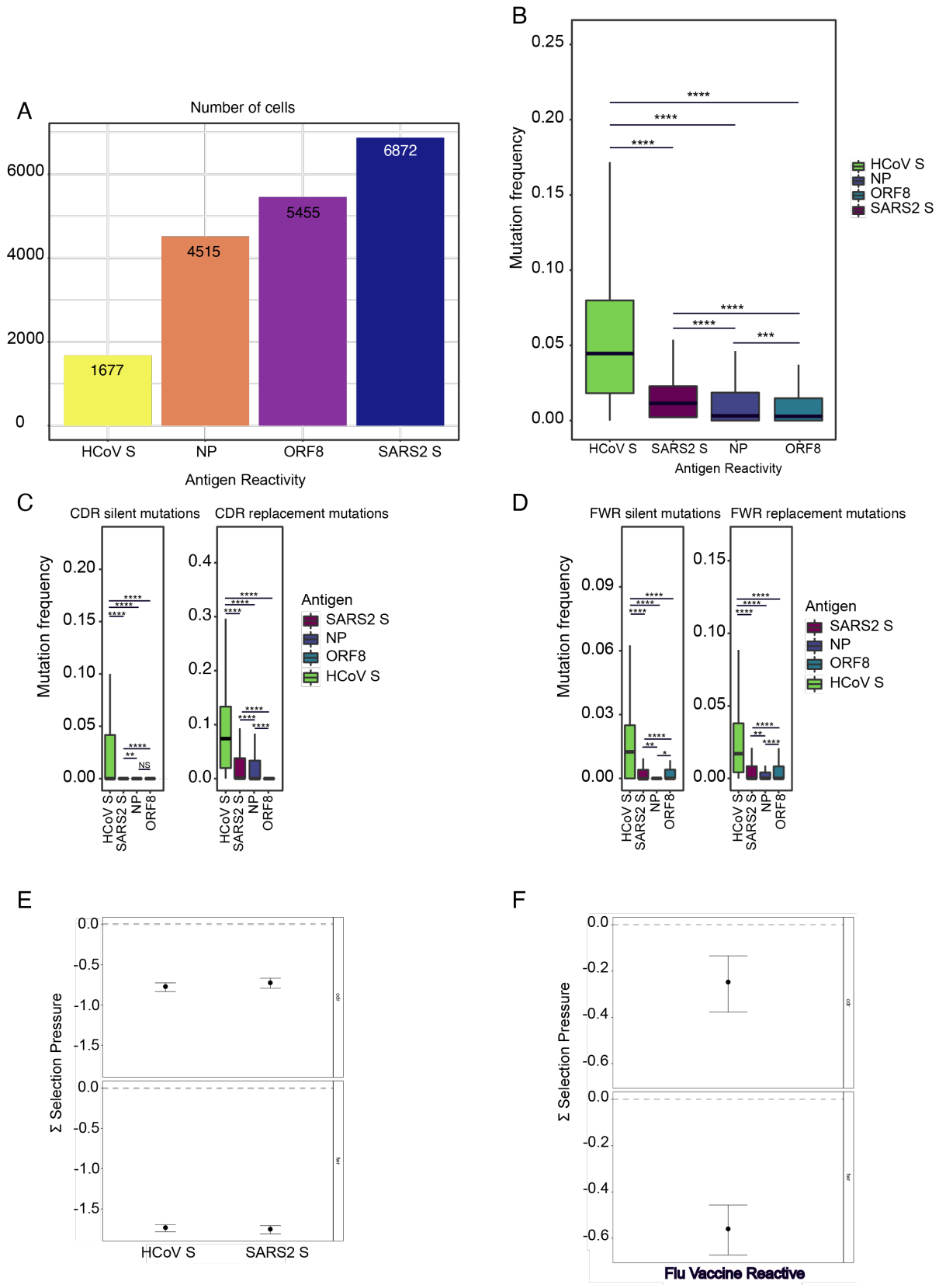


Figure 3.15: Endemic HCoV reactive cells have high levels of mutation in CDRs and FWRs and show evidence of strong antigenic selection.

Figure 3.15, continued.
(A) Total number of B cells reactive to each antigen.
(B) Mutational Frequency by antigen.
(C) Silent and replacement mutations occurring in the CDRs of antigen reactive B cells.
(D) Silent and replacement mutations occurring in the CDRs of antigen reactive B cells.
(E) Bayesian estimation of Antigen-driven selection of Endemic HCoV and SARS2 S reactive B cells.
(F) Bayesian estimation of Antigen-driven selection of flu vaccine reactive B cells from a public dataset.
Statistics for B, C, and D are One-Way ANOVA with Tukey's Multiple comparison test. Adj P value, **** = <0.0001, *** = 0.0002, **=0.0021, *=0.0332, NS= no significance

To further confirm the unusual amounts of mutation observed for HCoV S reactive B cells was not just an effect of sample size or a small population of cells skewing results we insured there was a large number of cells for all antigens (Fig 3.15A). While there were less than all other antigen reactive B cells, which is to be expected because of the nature of the samples, there were still over 1500 B cells included in the analysis. Total mutation frequency reaffirms what had been previously observed by variable gene, that HCoV reactive B cells are substantially mutated (Fig 3.15B). We next wanted to calculate the frequency of mutation by region, to see if such a high level of mutation requires an increasing push of mutations into the FWRs. All antigens but ORF8 showed a higher frequency of replacement mutations within CDRs than silent mutations, as expected, with HCoV potentially accumulating silent mutations because of how mutated they already are (Fig 3.15C). Another possibility is that the silent mutations which accumulate in highly mutated cells may be part of an intricate, second layer of mutational targeting. Such that, while they cause no amino acid changes, they do impact the sequence and change the overall mutability landscape before additional rounds of SHM.

Interestingly, ORF8 reactive cells showed only negligible amounts of mutation in CDRs, both silent and replacement but did show a low level of replacement mutations in the FWR (Fig 3.15D). However, HCoV reactive cells once again stand out. They show a markedly higher

amount of FWR replacement mutations than all other antigens (Fig 3.15D). To investigate this further we next used Bayesian Estimation of Antigen-Driven Selection in Immunoglobulin Sequences (BASELINe) to quantify selection pressure acting upon the populations of cells.[217, 218] By comparing all mutations to mutations predicted to arise from the S5F AID targeting model it can generate a numerical score for selection, with values further from zero implying greater amounts of selective pressure.[126, 205] Any B cell having been through a germinal center is expected to show evidence of antigen selection, however HCoV and SARS2 S reactive cells showed strong evidence of selection in both CDRs and FWRs (Fig 3.15E). To better understand the amount of selection acting on these cells we calculated the selective pressure acting on B cells responding to influenza vaccination from a publicly available data set, and found the flu reactive cells to show much weaker selection in both CDRs and FWRs (Fig 3.15 F).[219] This warrants further study, to determine why spike reactive cells are so strongly selected, or rather to understand why spike reactive cells seem to require a high number of rare or unusual mutations to improve their affinity. We next estimated selection for specific classes of replacement mutations, including those impacting the volume, polarity, hydrophobicity, or charge of the amino acid side chains (Fig 3.16 A-D). All FWRs still showed substantial amounts of selection, though differences began to emerge between SARS2 and HCoV S reactive cells, particularly in CDRs. SARS2 reactive cells shows slight amounts of positive selection for charge and polarity, while HCoV reactive cells do not (Fig 3.16 B and D). This difference perhaps stems from the SARS2 reactive cells being activated for the first time, whereas the HCoV reactive have been activated many times. Previous studies have shown that after early amounts of positive selection to raise affinity many B cells will begin to show signs of negative selection, as additional rounds of maturation purify out deleterious mutations.[220, 221]

These data confirm the need for further characterization and study of endemic HCoV reactive B cells. Influenza and HIV studies have generated great interest in highly mutated B cells which also show high amounts of FWR mutations, as these cells often prove to have broadly binding and neutralizing capabilities.[222-225] However, they can be difficult to study as they are fairly rare. HCoV S reactive cells could potentially provide a source of cells with these same characteristics for further study. They have the advantage of potentially being more common, because of how often common cold infections occur, and may also be of interest for universal coronavirus vaccine efforts. This makes them doubly valuable as both a source of better understanding coronavirus immunity and elucidating basic underlying principles of these classes of antibodies.

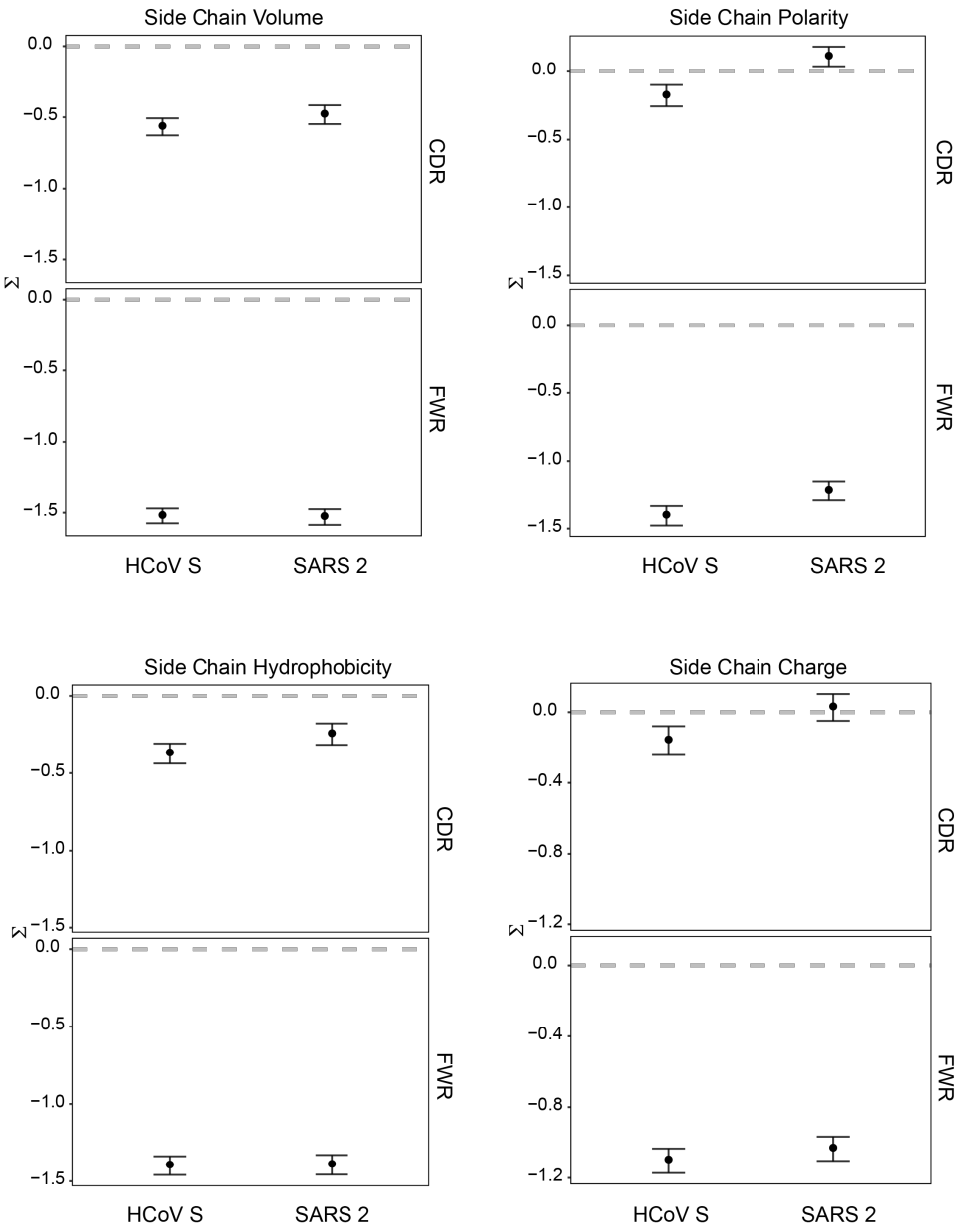


Figure 3.16: HCoV and SARS2 spike reactive B cells show varying amounts of selection pressure depending on the category of replacement mutation.
(A) Bayesian estimation of Antigen-driven selection of Endemic HCoV and SARS2 S reactive B cells for replacement mutations that change the side chain volume.
(B) Bayesian estimation of Antigen-driven selection of Endemic HCoV and SARS2 S reactive B cells for replacement mutations that change the side chain polarity.
(C) Bayesian estimation of Antigen-driven selection of Endemic HCoV and SARS2 S reactive B cells for replacement mutations that change the side chain hydrophobicity.
(D) Bayesian estimation of Antigen-driven selection of Endemic HCoV and SARS2 S reactive B cells for replacement mutations that change the side chain charge.

3.12 Chapter Discussion

The COVID-19 pandemic continues to pose one of the greatest public health and policy challenges in modern history, and robust data on long-term immunity is critically needed to evaluate future decisions regarding COVID-19 responses. Our approach combines three powerful aspects of B cell biology to address human immunity to SARSCoV-2: B cell transcriptome, Ig sequencing, and recombinant mAb characterization. Our approach enables the identification of potently neutralizing antibodies and the characteristics of the B cells that generate them. Importantly, we show that antibodies targeting key protective spike epitopes are enriched within MBC populations, but over time the MBC pool shifts towards non-protective intracellular antigens which could be a molecular hallmark of waning spike-mediated protection. This is further evidence that widespread vaccination, which only elicits a response to the spike, may be critical to end the pandemic.

Identification of multiple distinct subsets of naïve, innate-like B cells, recent and older MBCs, and ASCs illustrates the complexity of the B cell response to SARS-CoV-2, revealing an important feature of the immune response against any novel pathogen. The approximately 55,000 B cell transcriptomes with corresponding B cell receptor sequences herein may provide biomarkers or new gene signatures to help identify distinct B cell populations. Such populations can be used to evaluate future responses to infections of various kinds, vaccine formulations, and aid in a better understanding of human B cell biology and how it shifts over the course of the immune response. Thus, future studies further elucidating distinct identities and functions of these subsets are necessary and will provide key insights into various aspects B cell immunology.

Through this data, we revealed that the landscape of antigen targeting and B cell subsets varied widely across severe acute subjects and convalescent subjects between 1.5–5-months post-infection. Severe acute patients mounted a large ASC response toward HCoV spike and ORF8, largely derived from IgA ASC populations. Plasmablasts are reflective of ongoing germinal center or extrafollicular responses to infection and are often re-activated from preexisting immune memory. The expansion of highly mutated plasmablasts to HCoV spike in severe acute patients suggests that the early response to SARS-CoV-2 in some patients may be dominated by an original antigenic sin response. It remains unclear whether such responses worsen the severity of disease or reflect an inability to adapt to novel SARS2 spike epitopes. Alternatively, whether HCoV spike binding B cells adapt to the SARS2 spike and can provide protection is of interest for the potential generation of a universal coronavirus vaccine. Further investigation into the protection afforded by cross-reactive antibodies are warranted, as previous studies have identified cross-reactive HCoV and SARS1 binding antibodies can neutralize SARSCoV-2.[211, 226] Importantly, vaccine-induced responses to the spike are likely similarly shaped by pre-existing immunity, and should be investigated.

While SARS2 spike-specific B cells were enriched in memory in our convalescent cohort, we also identified MBCs and ASCs to HCoV spike, which appeared to wane 5 months after infection in the same subjects. This later timepoint coincided instead with an increase in the overall numbers and percentage of ORF8- and NP-specific MBCs, which displayed a marked increase in SHM. This phenotype was particularly evident in older patients, which exhibited reduced MBC targeting of the spike. Notably, older, female, and more severe patients had increased percentages of total ORF8-specific B cells; and older patients tended to create more memory to intracellular proteins over time. We identified B cells targeting these intracellular

proteins as exclusively non-neutralizing and nonprotective. Mechanistically, these observations may be explained by reduced adaptability of B cells or increased reliance on CD4 T cell help for B cell activation, which have been observed in aged individuals upon viral infections and are dysregulated in aged patients.[208, 213] Furthermore, T cell responses to SARS-CoV-2 intracellular proteins are prevalent in convalescent COVID-19 patients (Grifoni et al., 2020; Le Bert et al., 2020; Peng et al., 2020). The shift in memory output during convalescence may also reflect the massive difference in protein availability, with each virion producing only dozens of spikes but thousands of intracellular proteins.[227, 228]

More research is warranted to determine whether B cell targeting of distinct SARSCoV-2 antigens directly impacts susceptibility and disease severity, and conversely, whether age or disease severity shape memory formation. Addressing these questions will be critical for understanding the disease course, determining correlates of protection, and developing vaccines capable of protecting the most vulnerable individuals against SARS-CoV-2 and future CoV.

3.13 Chapter Acknowledgements

We kindly thank the University of Chicago CAT Facility (RRID: SCR_017760) for their assistance and allowing us to use their facilities. We thank Dr. Nicholas Chevrier for allowing us to use the Pritzker School of Molecular Engineering's sequencing facility. We are thankful to the University of Chicago Genomics Facility (RRID: SCR_019196) for their assistance with sequencing several samples. We thank John Bivona and the staff of the Howard Taylor Ricketts Laboratory for prompt training in BSL-3 practices and procedures and for use of their facilities for acute infected sample collection. We are grateful for the training and support provided by Megan Ciarlo and Allison Guiang of Miltenyi Biotech for use of the MACSQuant Tyto. Lastly,

we are most appreciative of the patients who donated samples for our research purposes upon recovery from, or during infection with, COVID-19.

3.14 Funding

This project was funded in part by the National Institute of Allergy and Infectious Disease (NIAID); National Institutes of Health (NIH) grant numbers U19AI082724 (P.C.W.), U19AI109946 (P.C.W.), U19AI057266 (P.C.W.), the NIAID Centers of Excellence for Influenza Research and Surveillance (CEIRS) grant numbers HHSN272201400005C (P.C.W.). N.W.A. was supported by the Multi-disciplinary Training program in Cancer Research (MTCR) - NIH T32 CA009594. A.J. and R.P.J were supported by federal funds from the NIAID, NIH, and Department of Health and Human Services under Contract HHSN272201700060C. F.K and F.A. were funded by the NIAID CEIRS contract HHSN272201400008C, Collaborative Influenza Vaccine Innovation Centers (CIVIC) contract 75N93019C00051 and the generous support of the JPB foundation, the Open Philanthropy Project (#2020-215611) and other philanthropic donations. Y.K. and P.H. were funded by the Research Program on Emerging and Re-emerging Infectious Disease grant (JP19fk0108113) and the Japan Program for Infectious Diseases Research and Infrastructure (JP20fk0108272) from the Japan Agency for Medical Research and Development (AMED), NIAID CEIRS contract HHSN272201400008C, and CIVIC contract 75N93019C00051. D.F., C.N, Y.D., and P.D.H, were supported by NIAID contracts HHSN272201700060C and 75N93019C00062. M.S.D. and E.S.W. were supported by NIH grants R01 AI157155 and F30 AI152327, respectively.

3.15 Declaration of Interests

The University of Chicago has filed a patent application relating to anti-SARS-CoV-2 antibodies generated in this work with P.C.W., H.L.D., and C.T.S. as co-inventors. Several antibodies generated from this work are being used by Now Diagnostics in Springdale, AR for the development of a diagnostic test. M.S.D. is a consultant for Inbios, Vir Biotechnology, NGM Biopharmaceuticals, and Carnival Corporation, and on the Scientific Advisory Boards of Moderna and Immunome. The Diamond laboratory has received funding support in sponsored research agreements from Moderna, Vir Biotechnology, and Emergent BioSolutions. The Icarhn School of Medicine at Mount Sinai has filed patent applications relating to SARS-CoV-2 vaccines which list F.K. as co-inventor. Mount Sinai has spun out a company, Kantaro, to market serological tests for SARS-CoV-2. F.K. has consulted for Merck and Pfizer (before 2020), and is currently consulting for Pfizer, Seqirus, and Avimex. The Krammer laboratory is also collaborating with Pfizer on animal models of SARS-CoV-2.

Chapter IV: Improved methods of analyzing multi-omics single cell data

In the past decade, advances in computing and single cell sequencing have ushered in a new era of discovery in biology and medicine. Improved sensitivity, reliability, throughput, and sequencing depth have provided researchers immense power to study diverse biological systems at single cell resolution.[229-237] The immune system, for example, encompasses a remarkable degree of cellular and phenotypic diversity, with more cells and functions yet to be discovered. As a result, the field of immunology has benefited significantly from the adoption of single cell RNA sequencing (scRNAseq) in order to characterize the molecular and cellular heterogeneity of immune cells during their developmental pathways, activation, and migration[238]. Furthermore, as the role of the immune system becomes more appreciated in different organs and diseases, single cell sequencing has become instrumental in characterizing the immune microenvironment in diverse contexts, such as cancer and the brain.[231, 239-242].However, despite tremendous progress, a full understanding of the phenotypic diversity and population structure of immune cells during complex and dynamic immune responses remains elusive.

Recent technological advances have enabled the simultaneous measurement of multiple cell properties in combination with scRNAseq, allowing for an unprecedented view of immune cell heterogeneity. For example, cellular indexing of transcriptomes and epitopes by sequencing (CITE-seq), and RNA expression and protein sequencing assay (REAPseq) are two multimodal technologies that are capable of measuring the expression of surface protein markers alongside the transcriptome at single-cell resolution.[196, 243] Much of foundational immunology has been elaborated through the use of specific lineage-restricted cell surface proteins to define immune cell populations. However, mRNA expression alone has demonstrated modest correlation with

the expression of lineage-defining protein markers, suggesting that transcriptional data alone cannot fully characterize the population structure of immune cells. Thus, direct estimation of protein expression may enable improved methods for delineating cell populations.

In order to gain a better understanding of immune functions at single cell resolution, the integration of surface protein expression and transcriptional profiling data is critical. Existing single-cell analysis software packages are typically designed for the analysis of unimodal data and provide limited integration of additional profiling modalities, such as cell surface protein expression data. In CITE-seq, protein expression is estimated through use of antibody derived tags (ADTs), which promise to better characterize immune cell heterogeneity (Peterson et al., 2017; Stoeckius et al., 2017; Stuart and Satija, 2019). For example, CITE-seq based protein expression information is often used in a post-hoc setting to annotate the population structure of single cells visualized based on transcriptional profiling. While some single-cell analysis pipelines have been designed to be compatible with multimodal data, these pipelines identify cell heterogeneity from transcriptome or surface protein markers separately instead of integrating information from both modalities.[244-246] Recently, a few methods have been published or pre-released for integrative analysis of CITE-seq data.[247-249] While these tools are valuable, improvements to the workflow and imbalances between transcriptome and surface protein expression are warranted. Additionally, there is also a lack of a quantitative metric of cluster purity for CITE-seq data. Thus, there remains a need for efficient methods capable of integrating transcriptome and surface protein expression for single cell CITE-seq data that allow downstream analyses with deeper resolution.

To overcome these challenges, we developed a single-cell analysis toolkit called "LinQView" for integrative analysis of CITE-seq data. LinQ-View integrates information from

transcriptome and cell surface protein expression levels by fusing pairwise distance matrices with an L_∞ norm. Based on the integrative kernel, we developed a quantitative metric (purity score) to access the purity of single cell populations on both the transcriptome modality and surface protein modality of CITE-seq data. Because the resulting matrix is also a distance matrix, subsequent downstream analyses can seamlessly utilize the protein and transcriptome linked views of single cells. By applying our integrated method to the analysis of immune cells from public datasets as well as our own datasets, derived from COVID-19-infected and influenza-immunized patients, we show the utility of this method in resolving more complex phenotypic clusters than conventional unimodal based methods alone. By comparing to existing methods, we demonstrated that LinQ-View is not only effective and efficient, but also superior on CITEseq datasets with small numbers of surface protein features.

4.1 LinQ-View combines single-cell mRNA and protein expression profiling data and enables multi-modal downstream analysis

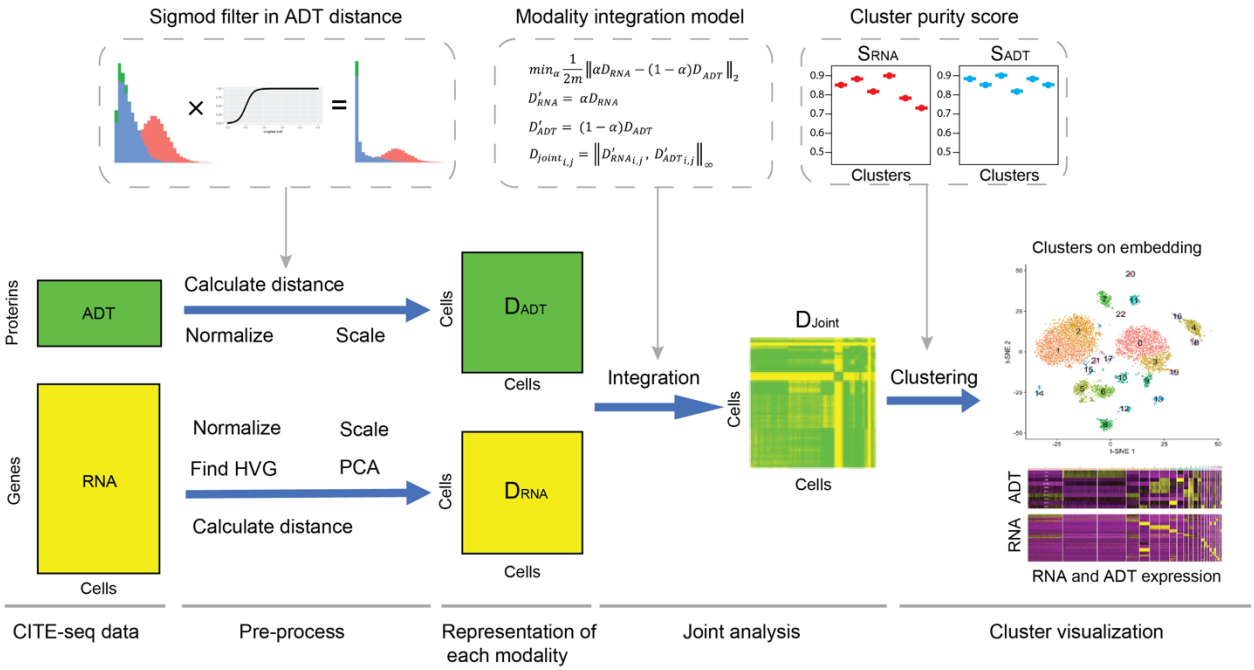


Figure 4.1: Kernel process of the LinQ-View toolkit. Cell-cell distances of each modality were calculated after several pre-processing steps adopted from conventional scRNA-seq data processing workflows. Next, joint distances were calculated by a linear transformation to rescale distances from two modalities into the same level. The scale factor was learned by minimizing the average sum of squared differences between the dissimilarity matrices. Cell heterogeneity was then inferred from variations in both transcriptome and cell surface proteins, which were represented by joint cell-cell distances. Purity scores for RNA and ADT modalities in CITE-seq data were used to assess the quality of clustering results and determine proper clustering methods/parameters.
Contributions: Lei Li developed the model with suggestions from Linda Lan, Chris, and Haley. Chris and Haley produced the COVID dataset and assisted with writing the text.

One possible representation of single cell profiling data is based on pair-wise comparisons of cells. Specifically, mRNA or protein expression profiles may be chosen as feature representations and pair-wise distances (or dissimilarities) can be calculated between all cells. The primary idea behind LinQ-View for CITE-seq is to calculate the dissimilarity representations for each single cell profiling modality (i.e. mRNA and surface protein expression) and then combine the dissimilarities from each modality. We reason that combining

dissimilarity representations may help to naturally emphasize different information from each modality and maximally capture and infer cell heterogeneity (Figure 4.1).

To calculate and integrate dissimilarity matrices, we first calculated pairwise Euclidean distances among cells from preprocessed gene or protein expression profiling data. To eliminate potential biases introduced from random noise in ADT signals, we specifically applied a sigmoid filter to distances calculated from protein expression profiles. Second, because the feature scale and dimensions differ between gene and protein expression profiles, we re-scaled each pairwise dissimilarity matrix using a linear transformation to minimize the average sum of squared differences between the dissimilarity matrices (Figure 4.1). Finally, we integrated the dissimilarity matrices using the L_∞ norm (Figure 4.1). Broadly, this is equivalent to choosing the maximal distance between any pair of cells based on either protein or gene expression. In conclusion, we simply maximize the cell-cell dissimilarity using information both from single cell mRNA and protein expression profiling data.

To quantitatively evaluate the quality of clustering, we developed a measurement to assess the purity of clusters on both modalities of CITE-seq (Figure 4.1). Because the RNA modality is on the scale of the whole genome whereas features of the ADT modality is highly variable and dataset-dependent, we adopted an entropy-based metric (ROGUE score) for assessing the purity of single cell populations using transcriptome profiles from a previous study, and developed a quantitative metric for assessing the purity of surface protein profiles.[250] In brief, our ADT scoring method uses sum of standard deviation (SD) of all ADT features of entire dataset as benchmark, quantifies purity of a specific cluster by evaluating the its sum of SDs of all ADT features against the benchmark. To better quantify cluster purity for CITE-seq data with different numbers of ADT features, we designed an adjustable parameter termed the "rank" of

the ADT score. The purity scoring for CITE-seq proposed in this study allows users to confidently evaluate the quality of their clusters and efficiently determine the best clustering algorithm for optimized resolution (Figure 4.1).

Finally, because the resulting cell-cell dissimilarity matrix is equivalent to a distance matrix, LinQ-View is compatible with most conventional clustering methods (e.g. k-means, hierarchical clustering, Louvain, Fuzzy c-means[244, 251, 252] and dimension reduction methods (e.g. multidimensional scaling [MDS], tdistributed stochastic neighbor embedding [t-SNE] and uniform manifold approximation and projection [UMAP].[253-255] The resulting dissimilarity matrix is also compatible with various batch effects correction methods (e.g. Mutual nearest neighbors [MNN], Canonical correspondence analysis [CCA] and Seurat anchors.[244, 245, 256]

4.2 LinQ-View identifies unique B cell subsets from COVID-19-infected patients

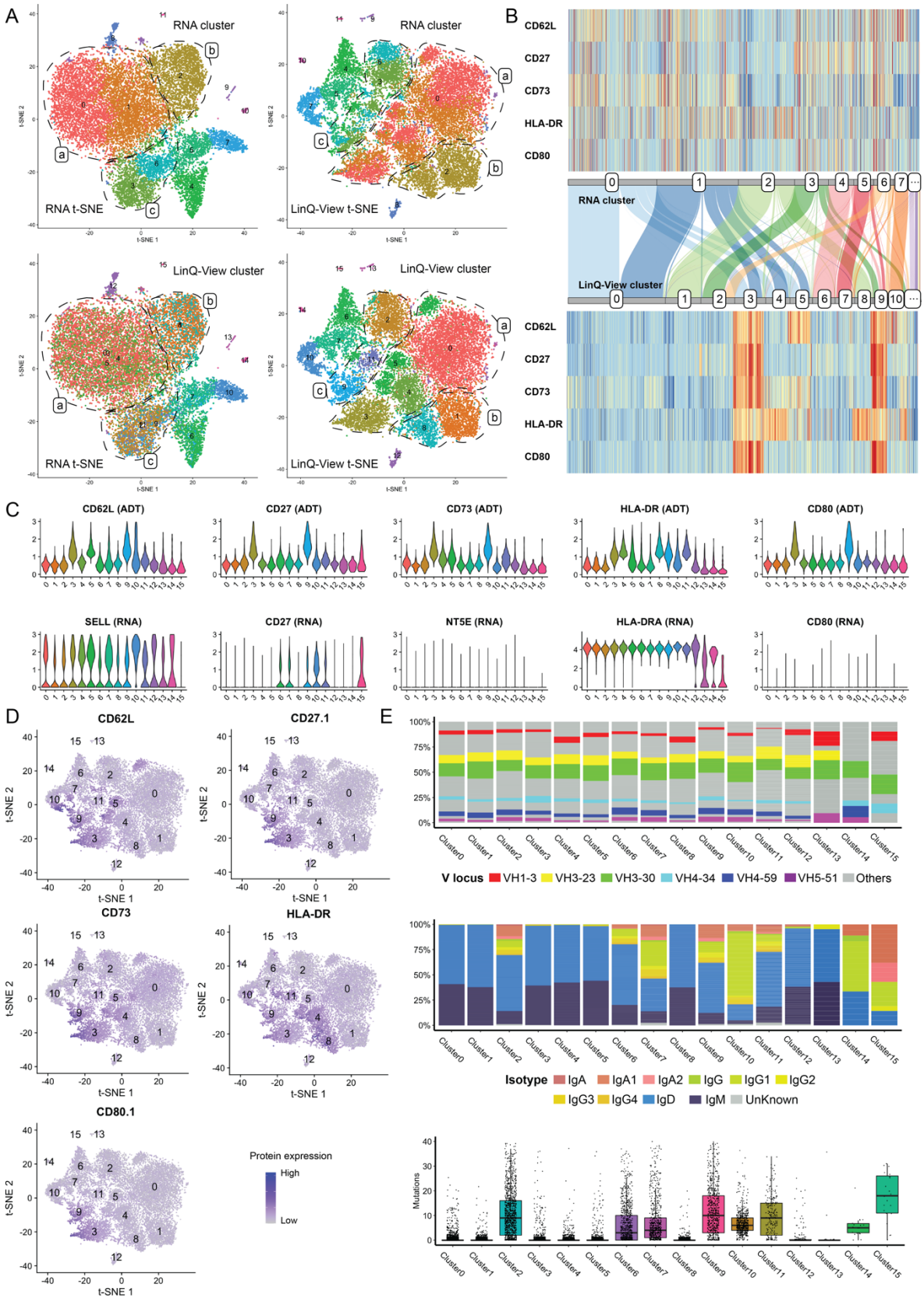


Figure 4.2: LinQ-View analysis identifies unique B-cell subsets from COVID-19 cohorts

Figure 4.2, continued.
(A) B cell subsets identified by transcriptome only (top two panels) and LinQ-View analysis (bottom two panels), displayed on RNA t-SNE (left two panels) and on LinQ-View t-SNE (right two panels). Unique B cell subsets identified by LinQ-View analysis were roughly highlighted by dashed circles and indicated by cluster ID. Three major groups of B-cell subsets were indicated by dashed circles with labels a, b, and c.
(B) Comparison between unimodal RNA clusters and LinQ-View clusters. Only the largest eight RNA clusters and largest nine LinQ-View clusters were indicated by their IDs due to limited space. A Sankey diagram showed the relationships between RNA clusters and LinQ-View clusters.
(C) Expression levels of five tested cell surface proteins and their corresponding genes for 16 cell clusters identified by LinQ-View.
(D) Visualization of cell surface protein expression on t-SNE embedding. t-SNE embeddings in this sub-panel were generated by LinQ-View. LinQ-View cluster IDs were labeled on the t-SNE embedding to roughly indicate distinct cell clusters.
(E) Three bar plots, from top to bottom, show the proportion of V gene usage, proportion of isoform and number of somatic hypermutation in each cluster, respectively.
Contributions: Lei Li developed the model with additional suggestions from Linda, Chris, and Haley. Chris and Haley produced the COVID dataset and assisted with writing the text.

To further demonstrate the utility of LinQ-View, we applied this pipeline to peripheral blood CD19+ B cells from ten COVID-19-infected subjects. We measured the expression of five cell surface proteins and binding to seven COVID-19 protein probes by flow cytometry: spike protein, spike receptor binding domain (RBD), open reading frames 7a and 8 (ORF8), nucleoprotein (NP) and non-structural proteins (NSPs), B cell immunoglobulin genes, as well as 5' gene expression in the same immune cells simultaneously. We used a pre-titrated cell staining panel that contained oligo-conjugated antibodies against CD62L, CD27, CD73, HLA-DR, and CD80. Batch effects among ten subjects in this dataset were removed by Seurat 3, then the integrated transcriptome modality and cell surface protein modality of this dataset were integrated by LinQ-View to identify unique B cell subsets.

LinQ-View successfully identified several unique B cell subsets with similar transcriptional expression but distinct profiles on the five tested surface proteins. RNA clusters 0 and 1 (indicated by dashed circle with label "a"), two naïve-like B cell subsets, were further

divided into four subsets by LinQ-View: they are LinQ-View clusters 0, 3, 4, 5 (Figure 4.2A). LinQ-View cluster 0 is negative to all tested ADTs (non-responder group), cluster 3 is positive to all tested ADTs (poly-reactive group), cluster 4 represents an HLA-DR+ group, and cluster 5 is CD62L+ (Figure 4.2B). These four subsets are indistinguishable on RNAbased t-SNE because they have highly similar transcriptional expression. In addition, RNA cluster 2 (indicated by dashed circles with label "b") was divided into two subsets: LinQ-View cluster 1 that is HLA-DR- and cluster 8 that is HLA-DR+ (Figure 4.2B). These two subsets also are indistinguishable by transcriptional expression (Figure 4.2A). Furthermore, LinQ-View identified three unique B cell subsets with distinct ADT profiles (LinQ-View clusters 2, 9, and 11) from RNA clusters 3 and 6 (Figure 4.2A, indicated by dashed circles with label "c"). LinQ-View cluster 2 is negative to all tested ADTs, cluster 9 is positive for all tested ADTs and cluster 11 is HLA-DR+ (Figure 4.2B). As indicated by the dashed circles in t-SNE embeddings, cells in group "a" (LinQ-View clusters 0, 3, 4, and 5), group "b" (LinQ-View clusters 1 and 8) and group "c" (LinQ-View clusters 2, 9, and 11) are mixed together on RNA t-SNE and cannot be distinguished using RNA only, even with higher resolution.

Among these unique B cell subsets identified by LinQ-View, clusters 4, 8, and 11 exhibited HLA-DR protein expression and cluster 5 exhibited CD62L protein expression distinguishing them from transcriptionally similar subsets, which indicates significant phenotypic differences. Notably, some B cells express B cell receptors (BCRs) that nonspecifically bind to multiple antigens. Clusters 3 and 9 exhibited strong binding to all tested ADTs and COVID-19 probes, suggesting that these two clusters are most likely composed of highly polyreactive B cells. We verified the polyreactive nature of a sampling of monoclonal antibodies cloned from cells predicted to be poly-reactive by testing them on a panel of variant antigens (Figure 4.2F).

Since cluster 3 and 9 have normal percentages of mitochondrial gene expression and similar transcriptional expression profiles compared to other B cells, they are indistinguishable for conventional transcriptome based unimodal methods (Figure 4.2C, D). However, because binding to the ADT probes cannot be distinguished from binding non-specifically to the BCRs, we concluded that these clusters should be excluded from analysis.

By investigating immunoglobulin repertoires from these subsets, we identified additional differences based on variable (V) gene locus, isotype, and degree of somatic hypermutation (SHM), which could further discern the identity of these B cell subsets (Figure 4.2E). Clusters 0, 1, 3, 4, 5, 8, 12, and 13 had the lowest median number of SHM (0) and highest IgM/D proportion (> 95%) and are thus most likely to be naïve-innate-like B cells. In contrast, clusters 2, 6, 7, 9, and 11 displayed an intermediate number of mutations and increased class-switching to IgA/G (20% to 50%) and so are most likely class-switched memory B cells from more recent immune responses to the SARS-CoV-2 proteins. Clusters 10, 14, and 15 displayed an increased median number of mutations and high IgA/G proportion (> 50%) and are most likely longer-lived memory B cells cross-reacting to the SARS-CoV-2 proteins or that had reacted early in the immune response and so have become more terminally differentiated.

4.3 Discussion of LinQ-Seq

Due to the large difference between the numbers of genes and proteins analyzed (thousands of genes versus approximately ten surface proteins) in CITE-seq, conventional dimension reduction methods fail to capture features of cell heterogeneity from CITE-seq data with two modalities merged together. In particular, variations in surface protein modality will be saturated by transcriptome data due to the highly imbalanced feature dimensions. While it has

been previously suggested that a joint dimension reduction may be a possible solution, this requires a non-trivial algorithmic structures to handle the highly imbalanced sizes of feature space between transcriptome and surface protein modalities.[238]

In this study, we use a simple and intuitive approach based on pairwise distances to represent variations between cells in multiple modalities. The L_∞ norm we introduce is sensitive to capturing dissimilarities in both modalities. By applying LinQ-View to several CITE-seq datasets, we prove the effectiveness of harnessing variations in pair-wise distances for maximally inferring cell heterogeneity. In particular, we introduce the efficient capability of the L_∞ norm for integrating features from multiple distance matrices. We demonstrate that LinQ-View has better resolution in determining cell population clusters compared with unimodal analysis that only use surface protein or transcriptome profiles. To facilitate the broad applicability of our approach, we developed an R-based toolkit that allows users to both identify cell populations by adopting information from both transcriptome and surface protein expression levels and integrate multiple datasets to eliminate batch effects in multi-modal data.

Assessment of cluster purity is paramount to method selection, resolution determination and subsequent data interpretation. Recently, an entropy-based purity metric was proposed to assess the purity for single cell populations using their gene expression profiles for conventional scRNA-seq data. However, development of CITE-seq technique revealed a need for assessing the purity of single cell populations using surface protein expression. To fill this gap, we propose a variance-based purity metric and demonstrate that this ADT scoring is capable of accurately quantifying purity of single cell populations based on their ADT variances, and thus could be used in the determination of the best clustering algorithms and optimizing resolutions. Through tests on both real and simulated datasets, we demonstrate that our ADT scoring can correctly

quantify variation levels among single cell clusters. We also demonstrate that a proper rank of ADT score is able to balance both global, difference between high-variant clusters and low-variant clusters, and local, difference within high-variant clusters and low-variant clusters, variances and, further, that it should be determined by the number of ADT features.

The collective development, maturation and activation states of immune cells results in an enormous diversity of cell types and functions. Multimodal profiling is thus key to deciphering accurate cell phenotypes and functions. Using 5 ADTs, LinQ-View successfully identified unique B cell subsets with identical transcriptional expression but distinct surface protein expression profiles from COVID-19-infected patients (e.g. $CD62L^+$ naïve-like B cell subsets and $HLA\text{-}DR^+$ memory-like B cell subsets). These results indicate that LinQ-View is capable of identifying intermediate states of cell differentiation and describing more accurate immune cell developmental histories than platforms that do not utilize multiple modalities.

The ultimate goal of single-cell methods is to elucidate both temporal and spatial aspects of cells, including spatial distribution, growth history, and interactions with other cells/molecules. In the future it is likely that more modalities will be accessed within a single experiment. To date, six modalities, including transcriptome, surface protein, T cell receptor (TCR) αβ , TCR γδ, sample identity from cell hashtag, and sample identity from single guide RNA, can be identified by the expanded CRISPR-compatible CITE-seq (ECCITE-seq) for single immune cells.[257] However, such multimodal profiling technologies pose new challenges for integrative computational analysis. A robust and extendable system is critical for handling the growing number of modalities. LinQ-View captures variations among cells by pairwise distances, which can be readily extended to multiple new modalities

Chapter V: Exploring the Functional Impact of Activation Induced Cytidine Deaminase Targeting

5.1 Introduction

Since it was first recognized that DNA bases neighboring a cytidine could influence the activity of AID there has been interest in better defining what those sequences are, how they function, and how they could impact the process of affinity maturation.[104, 121] The process of SHM is both powerful and dangerous. The mutation of BCRs can expand B cell repertoires beyond what gene conversion and VDJ recombination can achieve, and it can greatly improve the affinity of a BCR towards an antigen it binds.[99, 104, 129, 258] However, it can also introduce mutations that destabilize the BCR entirely, making it incapable of expressing at the cell surface and leading to the death of the B cell.[259, 260] Finally, AID has been shown to occasionally aberrantly act outside of the variable gene region, and is thought to be the cause of some B cell lymphomas.[261, 262] As such, the process of SHM and how it is targeted and performed must strike a delicate balance between mutating sites to expand diversity and improve affinity while protecting other sites from harmful damage.

Traditionally it had been thought that mutations within CDRs were desirable and those within FWRs were undesirable but unavoidable. As such, it was proposed that CDRs were enriched for hotspots and FWRs for coldspots to allow for a high amount of SHM within CDRs while protecting the structural stability of the antibody. Sequencing has proven the distribution of hot- and colspots to largely follow this pattern, however, there is also data which contradicts it. One is a group of frequently mutated hotpots within FWR3 that are sometimes even referred to as hypervariable region four or CDR4. Overtime, improved methods for monoclonal antibody

production and high-throughput sequencing has led to the discovery of increasing numbers of antibodies that not only had highly mutated FWRs, and also showed that these antibodies lost their binding or neutralization capacities if those mutations were reverted. Further structural and biophysical work has now revealed that some classes of antibodies, notably broadly neutralizing antibodies (bnAbs) targeting HIV or influenza, often rely on FWR mutations for their key characteristics. It is now appreciated that earlier mutations in the FWR can allow for increased flexibility, allowing an antibody to bind new epitopes, before latter mutations lock in the confirmation or, if a germline antibody already binds strongly then early FWR mutations can act to increase its rigidity. FWR mutations have also been shown to be capable of stabilizing Ab-epitope interactions indirectly through hydrogen bonds, charged side chains, and other biochemical means that don't involve them directly participating in binding. It has also been shown that some FWR mutations can play compensatory roles, re-establishing thermodynamic stability after critical CDR mutations that increase affinity have decreased stability.

All of this indicates that the overall system of SHM, and especially the sequence-based targeting of AID and polymerase eta, is intricately evolved and incredibly complex. Therefore, it is likely that there is no single role that targeting plays but perhaps several, which could potentially vary from species to species or even from one type of B cells to another within an individual. However, if one takes into account the evolutionary history of AID, what is known of its structure and function, and the role it plays in different species and immune responses it is likely that the most important function of mutational targeting is one of efficiency and protection. That the overall amount and distribution of hotspots work to encourage the binding of AID and its deaminase activity at a higher amount than would otherwise occur in their absence, thus increasing the speed at which diversification or affinity maturation occurs within a germinal

center while concurrently acting to target the activity in such a way that works to protect against destabilizing or deadly mutations.

With an ever-growing appreciation for the complexity and intricacy of AID mutational targeting it has become apparent that outstanding questions remain, especially with regards to the true functional impact of this targeting. Now that it is well appreciated that FWR mutations can actively influence binding and neutralization it can no longer be assumed that hotspot coldspot targeting exists only to protect the structural integrity of the BCR. It is possible, and perhaps likely, that this is still the major function of targeting, and the evolutionary tradeoff is sacrificing the ability to generate certain classes of antibodies in large amounts to protect all others. Another possibility is that targeting is evolved to improve the efficiency of affinity maturation, resulting in antibodies of a similar affinity being generated in less cycles than if the germline sequence were "untargeted."

We deemed the best way to address these questions would be the generation of knock-in (KI) BCR mice. By creating one mouse line that expresses the predicted or actual germline version of a BCR and a second mouse line that expresses a version designed to encode the same protein but using a sequence that is altered to remove hotspots and coldspots we could directly probe the impact of AID targeting during the process of affinity maturation.

5.2 Characteristics of the monoclonal antibody 045-051310-2B06

The first decision that needed to be made was which antibody to use as the basis of the mouse. After careful consideration we decided to generate the mouse using a monoclonal antibody previously isolated in the Wilson laboratory, 045-051310-2B06, which had several features that made it a strong candidate for this project.

045-051310-2B06, herein referred to as 045-2B06, was cloned from a plasmablast isolated from a young male subject 7 days post vaccination with a single strain vaccine developed against the 2009 H1N1 pandemic strain (A/california/7/2009 H1N1pdm09-like virus). Even though it was cloned from a cell responding to pH1N1 it had been found capable of binding a train of H3N2 influenza as well. Further studies revealed it was also capable of binding two pandemic potential H7 avian influenza strains, A/Shanghai/1/2013 (H7N9) and A/Anhui/1/2013 (H7N9). Thus, it can be classified as a broadly neutralizing anti-influenza antibody.[263]

Further characterization revealed it to be very broadly reactive, showing the ability to bind 14 of 18 different influenzas it was tested against. It was also found to be capable of neutralizing H7 in-vitro and protecting mice against H7 infections in-vivo. Competition ELISAs showed it bound at the same epitope as CR9114, a class defining HA stalk binder that binds a well characterized broadly neutralizing stalk epitope that is a site of great interest for vaccination. Escape mutants further confirmed this as the site of 045-2B06's binding.

045-2B06 uses VH1-18 as its heavy chains V gene frequently found in use in HA stalk binding bnAB, and Vk3-11 for its light chain. It is a highly mutated antibody, with 37 mutations, with 22 of them being non-silent, in its heavy chain and 19 mutations, 14 non-silent, in its kappa chain. Interestingly, nearly half of its heavy chain mutations, 18, are located within FWR3. The heavy chain has 17 mutations that have occurred within a traditional hotspot and the light chain has 11 hotspot targeted mutations. Critically, the predicted germline sequence, when expressed as an antibody has very weak binding for pH1N1 and weak or non-detectable binding for some of the other strains it previously bound, confirming its broad and strong reactivity is a product of affinity maturation.

The decision to use 045-2B06 was predicated on these established characteristics of it, and because it also represents a class of antibody that is of great interest in strategies to design universal influenza vaccines. Therefore, we decided to create an additional line of mice, using the fully affinity matured version of the BCR. We could then use the system to design experiments also related to the limits of affinity maturation and how far a BCR can be pushed, to see if there is some limit wherein it becomes a self-defeating process which is more likely to destabilize an effective BCR than improve it. The affinity matured mouseline would also allow for vaccination studies and more specifically to determine if a highly mutated bnAb could acquire more mutations that allow it to expand the number of influenza strains it could bind and neutralize.

The three proposed mouse lines; AM-045-2B06 (affinity matured), GL-045-2B06 (germline), and DE-045-2B06 ('de-evolved,' meaning AID-untargeted) could be powerful models to explore questions concerning both fundamental basic biology and those of immediate translational interest.

5.3 Considerations in the design of an AID untargeted BCR

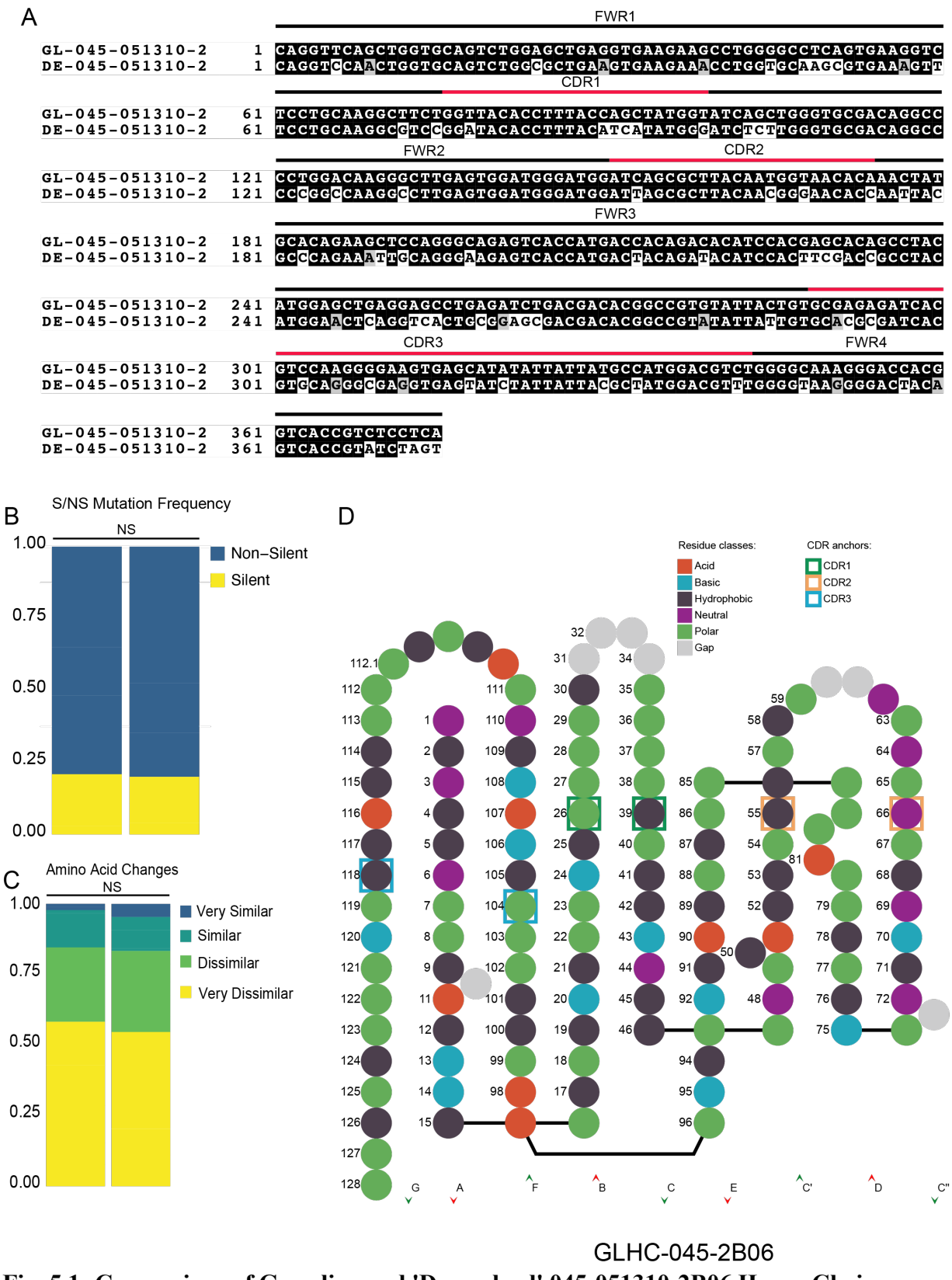


Fig. 5.1: Comparison of Germline and 'De-evolved' 045-051310-2B06 Heavy Chains.

Fig. 5.1, continued.
(A) Boxshade alignment of germline and 'de-evolved-045-051310-2B06 heavy chains with framework and complementarity determining regions marked.
(B) Ratio of non-silent to silent mutations in a set of computationally generated germline and 'de-evolved'-045-051310-2B06 heavy chain sequences with mutations introduced according to the S5F model.
(C) Frequency of IMGT defined amino acid substitutions in a set of computationally generated germline and 'de-evolved'-045-051310-2B06 heavy chain sequences with mutations introduced according to the S5F model.
(D) *Collier des Diamants* representation of amino acid class usage in the variable region of germline-045-051310-2B06. Numbers represent IMGT amino acid numbering. Each position is a pie chart showing the frequency of amino acid class usage at that position. CDR anchors are marked by colored boxes.
NS = No significance. Statistics for C are Multiple Mann-Whitney Tests

Hotspots and coldspots, defined traditionally, encompass a large and varied number of possible sequences. This means it is impossible to fully remove all hot- and coldspots from any given BCR sequence, and thus decisions must be made on how to prioritize their removal. Considerations could include their location (CDR vs FWR) but this risks biasing any data, so a better approach is to try and remove them from the entire BCR in an unbiased manner. In wanting to make our genes as neutral as possible we decided to prioritize by the amount of activity at each defined hot- or coldspot. Using a collection of 28,307 somatic mutations from a large collection of BCR sequences we determined a score for the likelihood for a mutation to occur in a given spot, setting "1" as equal to the chance of a mutation occurring in a neutral "C." Hotspot and coldspot reversion was then prioritized to remove the sites with scores the farthest from neutral.

In addition to classically defined hot- and coldspots, more recent work has used large sequencing datasets and mathematical modeling to create a model of mutational targeting known as the S5F model.[126] This modeling seems more accurate and allows for the identification of sites of enriched or dampened mutation outside of traditionally defined sequences, so any future

studies concerned with the generation of modified AID targeting mice should consider using this model to alter the BCR.

Finally, previous work in our laboratory had shown that variable gene sequences have differential codon usage rates than the rest of the human genome. All possible C to T mutations in the entire genome would generate silent mutations 42% of the time, but within CDRs C to T mutations would be silent 67% of the time, and frequently generate a conservative mutation, meaning causing a shift to an amino acid defined as similar, if non-silent. This was similar to FWRs, where 55% of all C to T transitions would be silent.[105]

Therefore, in addition to removing the sites of greatest AID activity we also designed the DE-045-2B06 antibody to have a codon usage pattern more similar to that of the overall human genome, raising the level of non-silent C to T mutations occurring within it.

One potential pitfall of any human gene knocked into a mouse, and especially one with an additionally manipulated codon usage, is that it will not properly express. This is because different species have differential codon usages, and thus their corresponding cellular machinery can more efficiently translate and transcribe genes that reflect this usage.[264, 265] We used the GenScript Rare Codon Analysis Tool to check our sequences.[266] A score of 1.0 is the highest, and 0.80 considered the cutoff for reliable expression. The affinity matured and germline sequences of 045-2B06 scored 0.80, with the DEHC scoring below at 0.78. However, because this was close to the cutoff and the two other genes, it was left as is to maintain it in as much of an 'un-targeted' state as possible.

Though we were unable to test the mutation in mice, we used the S5F model to generate a series of computationally mutated sequences for both germline- and 'de-evolved'-045-2B06 to explore their differences in-silico.[102, 123, 126] We set the model to introduce approximately 37

mutations into each sequence, to match the level of maturation seen in the originally sorted monoclonal antibody, 045-051310-2B06. We generated a set of 5 mutant sequences for each BCR and compared them. No significant differences were found between the ratio of silent to non-silent mutations (Fig 5.1B). However, we did observe differences in the type of non-silent mutation, when gauged by how similar or dissimilar the amino acid replacements are (Fig 5.1C). The 'de-evolved' genes mutated in patterns that generated very dissimilar replacements at a higher frequency, and very similar at a lower frequency (Fig 5.1C). To further explore these differences, we generated *Collier des Diamants* charts (Figs. 5.1D and 5.2A-B).[267] The chains represent, in two dimensional space, the variable regions and binding loops of BCRs. Each circle on the chain both marks the site of a residue and also serves as a pie chart showing the frequency of class usage at that site. The 'de-evolved' genes showed mutational patterns that generated residues of different classes, when compared to germline sequences, in both FWRs and CDRs (Fig. 5.2B). This presents an intriguing possibility that not only are variable genes evolved for intricate targeting at the nucleic acid level, but that they have also been selected for mutations that favor certain classes of amino acid residues at each site. This further shows there remains unappreciated and unstudied complexity in the intricate mutational targeting within variable genes encoding B cell receptors.

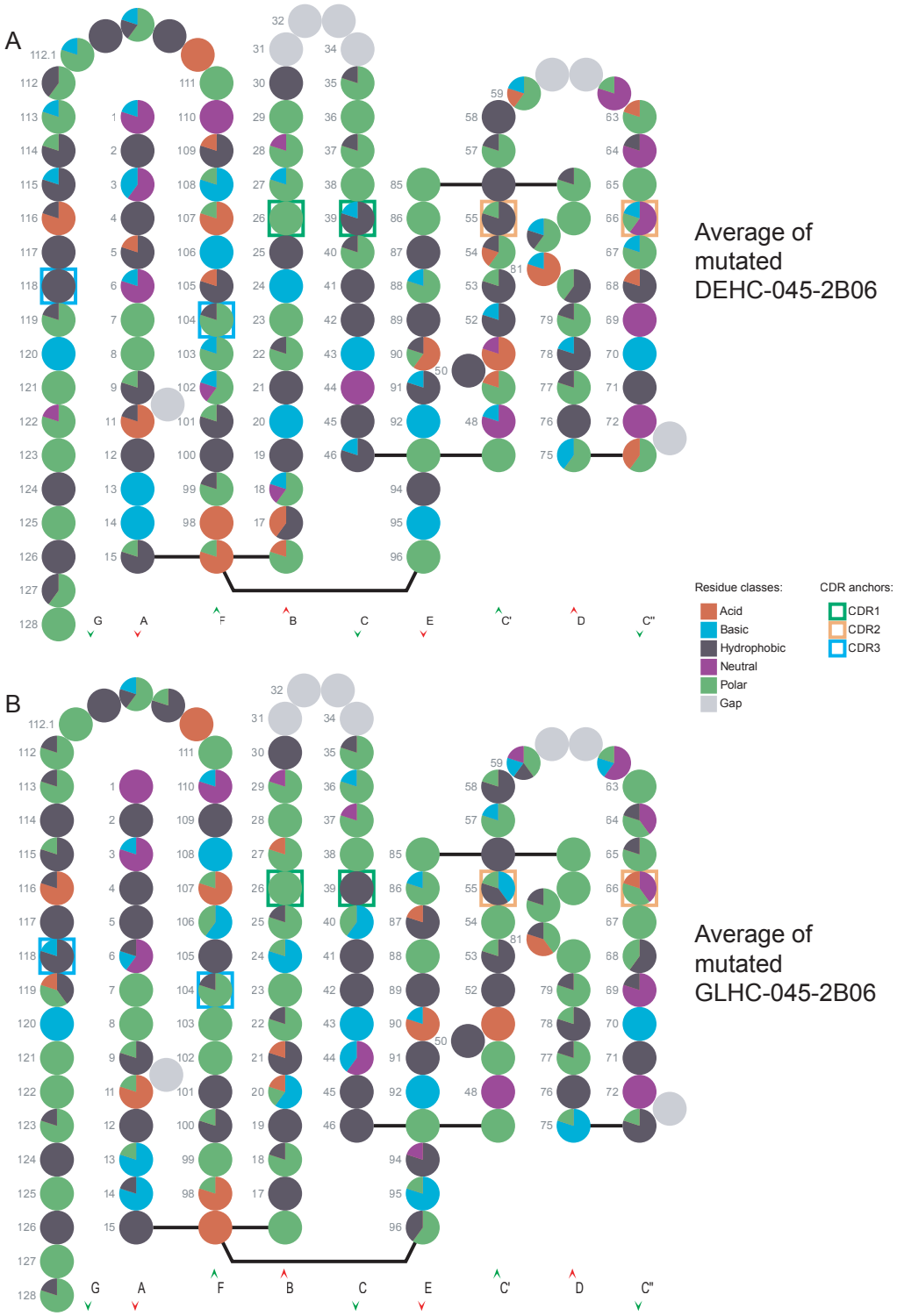


Fig. 5.2: Amino acid class usage patterns in computationally mutated heavy chain sequences.

Fig. 5.2, continued.
(A) *Collier des Diamants* representation of amino acid class usage in the variable region of a set of computationally generated 'de-evolved'-045-051310-2B06 heavy chain sequences with mutations introduced according to the S5F model. Numbers represent IMGT amino acid numbering. Each position is a pie chart showing the frequency of amino acid class usage at that position. CDR anchors are marked by colored boxes.
(B) *Collier des Diamants* representation of amino acid class usage in the variable region of a set of computationally generated germline-045-051310-2B06 heavy chain sequences with mutations introduced according to the S5F model. Numbers represent IMGT amino acid numbering. Each position is a pie chart showing the frequency of amino acid class usage at that position. CDR anchors are marked by colored boxes.

5.4 Generation of the mice

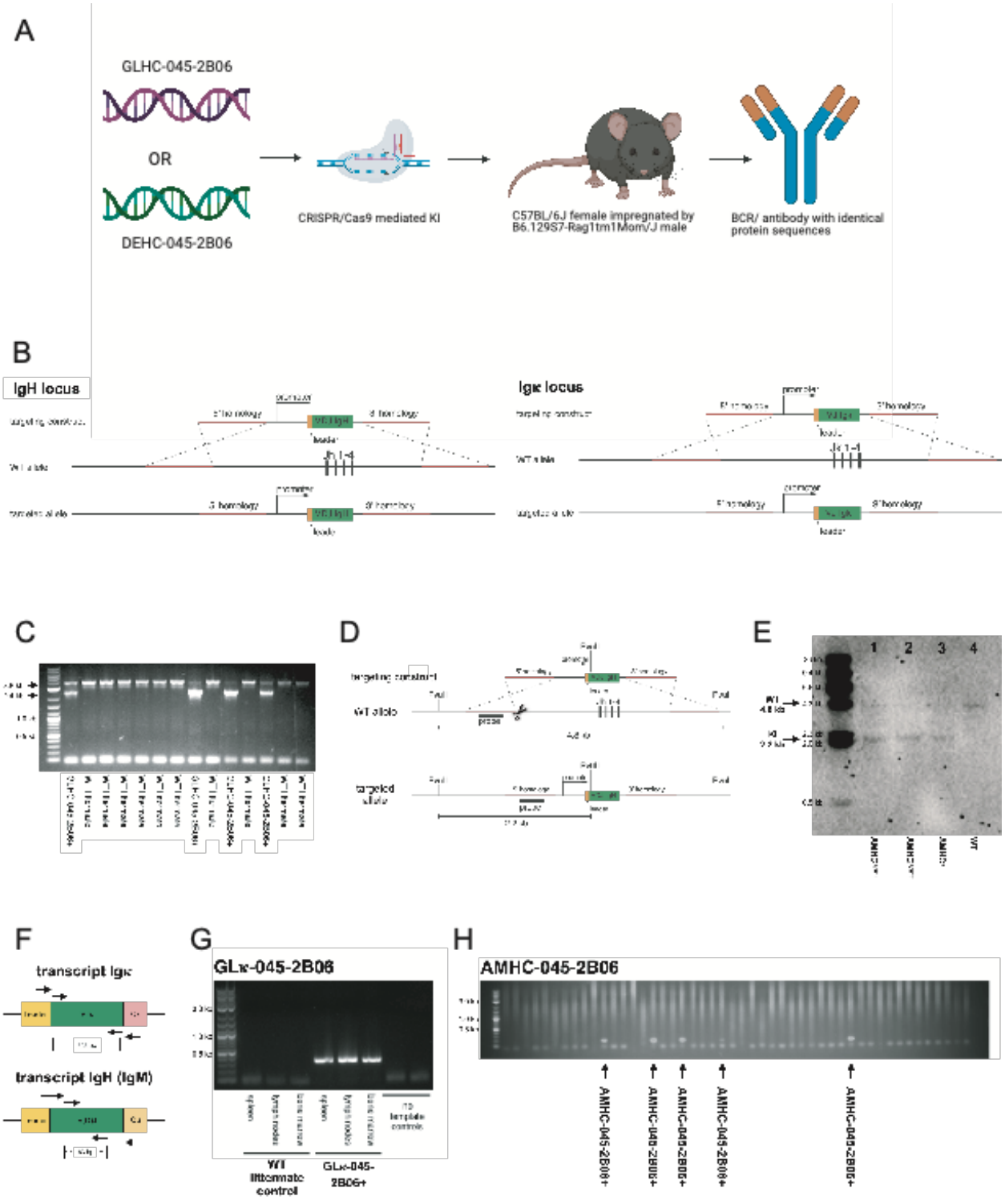


Figure 5.3: Approach to Generate and Characterize the BCR Knock-In Mice

Figure 5.3, continued.
(A) Graphical schematic showing the broad strategy to generate the untargeted mice.
(B) Schematic drawing of the targeting strategy used for the generation of bnAb knock-in mice.
(C) DNA agarose gel showing the IgH WT band at 3.8 kb and the transgene (GLHC) at 2.4 kb.
(D) The WT and targeted alleles shown schematically with cut sites for the enzyme and binding sites for the probes used for Southern blot, as well as the size of the WT and KI bands.(E) Southern blot of F2 AMCH-045-2B06 mice (2 heterozygotes (lanes 1 and 2), one homozygote (lane 3), and 1 WT littermate (lane 4)).
(F) Schematic drawings of the transcripts of the targeted Igκ (top) and IgH (bottom) genes are shown with the primer binding sites for nested RT-PCRs indicated with arrows.
(G) Separation of the PCR fragments on agarose gels show a specific 0.3 kb band for the Igκ transgene amplified from mRNA taken from spleen, lymph nodes, and bone marrow.
(H) Agarose gel showing a specific 0.16 kb band for the IgH transgene amplified from mRNA from single cell sorted peripheral blood B cells.

#

After deciding on which genes to use as the basis of our knock-in we next decided on a knock in strategy (Fig 5.3A). After a pilot attempt with full Cas9 showed substantial off target effects, we determined a system using nickase Cas9 using two separate gRNAs would still give a high efficiency of knock-in with a reduction of off-target effects, as it had been reported to do so.[268] We designed our knock-in constructs to knock-out the native mouse J chains, to reduce the chances of recombination and force B cells to use the knocked-in gene (Fig 5.3B). PCR genotyping, followed by gel purification and sequencing confirmed the gene could be amplified from the mouse genome, though all primers that successfully gave bands primed from within the homology arms, and no bands could be generated from primer pairs including a primer located outside of them the endogenous locus (Fig 5.3C). However, southern blotting showed bands of the expected size if the gene were knocked into the proper location, which could not be generated from WT controls (Fig 5.3D-E). In an attempt to clarify the disparity between PCR genotyping and southern blotting, we collected mRNA samples from spleen, inguinal lymph nodes, and bone marrow and used a nested rtPCR strategy in an attempt to amplify cDNA for the transcript of our knock-in BCR genes. We were able to successfully amplify bands of the proper

size, and confirm them by sequencing to be the expected gene (Fig 5.3G). Furthermore, the same rtPCR strategy successfully amplified the transcripts from single cell sorted B cells as well (Fig 5.3H).

However, even with this promising data other experiments proved more inconclusive. No genotyping PCRs to confirm the location of the gene were ever able to generate a band from gDNA, perhaps implying there was damage at the site of insertion, or the homology was not quite as expected. Additionally, transfer experiments into congenic B6 mice followed by immunization with influenza hemagglutinin (HA) showed no preferential expansion of B cells from the KI mice, which would be expected from the AMHC-045-2B06 mice, which should have been expressing a high affinity HA BCR. Because of this and other inconclusive or contradictory data we decided to regenerate the mice using a new approach, Speed-Ig, developed at the University of Chicago that had shown high success rates[269].

5.5 Chapter Five Conclusions

After re-generation of the mice using Speed-Ig system we again confirmed what appeared to be proper knock-in and targeting of the gene. However, full BCR knock-ins bred onto a RAG1$^{-/-}$ background have no to very few B cells in their periphery, indicating improper expression of the BCR. This is the case for all three stains of mice: AMHC-045-2B06, GLHC-045-2B06, and DEHC-045-2B06. A single cell RNA sequencing pilot experiment was planned using mice with either the affinity matured or germline 045-2B06 heavy chain knocked-in. The mice were immunized with chimeric H5/1 HA, followed by a boost at day 21 with chimeric H8/1 HA. These chimeric HAs have the heads from HAs from rare, avian strains of influenza and the stalk of HA from the a/cal/09 pandemic H1N1 strain of influenza. Because the knocked-in BCRs are reactive to the HA stalk of H1N1 they should outcompete endogenous mouse B cells. However, scRNA sequencing of B cells taken from the draining lymph node 14 days after boost found no B cells using the knocked-in genes. Because the BCRs are a combination of mouse and human genes there was uncertainty in which reference genome to use, therefore, to confirm the absence of our BCR sequences we worked in collaboration with 10x genomics to generate a reference genome featuring human variable regions and mouse constant regions, reflecting the construction of our KI BCRs.

It is clear the BCRs are not properly expressing in our generated mouse strains. They could potentially be auto-reactive in mice. This possibility could be explored by performing flow cytometry on bone marrow from the full BCR knock-in mice, which could indicate the presence of a developmental block. Additionally, humans and mice have slightly different systems for how nucleotides are donated at the junction of the variable and constant region, and this could prevent proper expression of the BCR. Lastly, the promoters used in the Speed-Ig system found

to drive strong expression of knocked-in murine BCRs might not be capable of driving strong human/mouse hybrid BCR expression, and so the issue could stem from weak expression rather than a frameshift caused by problems at the splice site or autoreactivity.

In the time since the project was conceived and began, there has been much research that further elucidated and explored the role of AID targeting and how it functions. However, the majority of this has been performed using high throughput sequencing and mathematical modeling. As such, there remains many interesting and unexplored question which could be effectively answered by the generation of 'untargeted' or 'de-evolved' AID knock-in mice.

Chapter VI: Discussion

6.1 Overview

The COVID-19 pandemic is the greatest public health crisis in a century, causing devastation globally. To date is has infected over 135.9 million globally, and caused over 2.9 million confirmed deaths.[270] In addition to the toll on human health and wellness, it has also inflicted massive economic harm and greatly undermined global happiness and mental wellbeing.[271, 272] Therefore it is of critical importance to better understand the virus, SARS-CoV-2, that is driving the pandemic and human immune responses to it. We improved upon established protocols and developed novel techniques, which allowed us to rapidly and thoroughly characterize B cell responses to COVID-19. In the course of this investigation, we have revealed substantial heterogeneity within populations of human B cells, both on a transcriptional level and in between individuals. Further, we were able to demonstrate the differences between individuals correlated with disease severity.

From the initial declaration of the COVID-19 pandemic the greatest hope for ending it without untenable loss has rested upon vaccines. After being developed and deployed in record time, COVID-19 vaccines are proving incredibly effective at preventing disease and slowing the pandemic.[188, 189] Though we have not yet profiled B cell responses from vaccinated individuals, we showed that infected individuals show substantial increases in memory toward non-neutralizing and non-protective antigens over time, further emphasizing the need for rapid, mass vaccination.

The tools and techniques pioneered in the course of this work can serve as a foundation for similar studies to other infections and diseases. This will be tantamount in growing knowledge from more to more and enriching and improving human life. Indeed, the burden of

COVID-19 pales in comparison to past, and potentially future, pandemics and endemic diseases that claim many lives every year.

Finally, the data generated in the process of profiling B cell responses to SARS-CoV-2 represents a rich and bountiful source that can continue to be mined for new insights into basic B cell biology, the transcriptional profiles of different families of B cells, and the mechanics and role of mutation during affinity maturation.

The work presented within this dissertation represents a confluence of years of study and training suddenly sharpened and forged into something different within the perilous fires of a pandemic. It has proven enormously beneficial in the short term, revealing key insights into the immune response against COVID-19 and generating reagents of immediate medical impact. It has uncovered a previous unappreciated and unknown antibody response to common colds that are marked by remarkably high amounts of affinity maturation and is poised to continue to produce new insights into the intricacies of the human immune system.

6.2 The Immune Landscape During SARS-CoV-2 Infection and the Importance of Vaccination for Long Term Immunity to Protective Targets

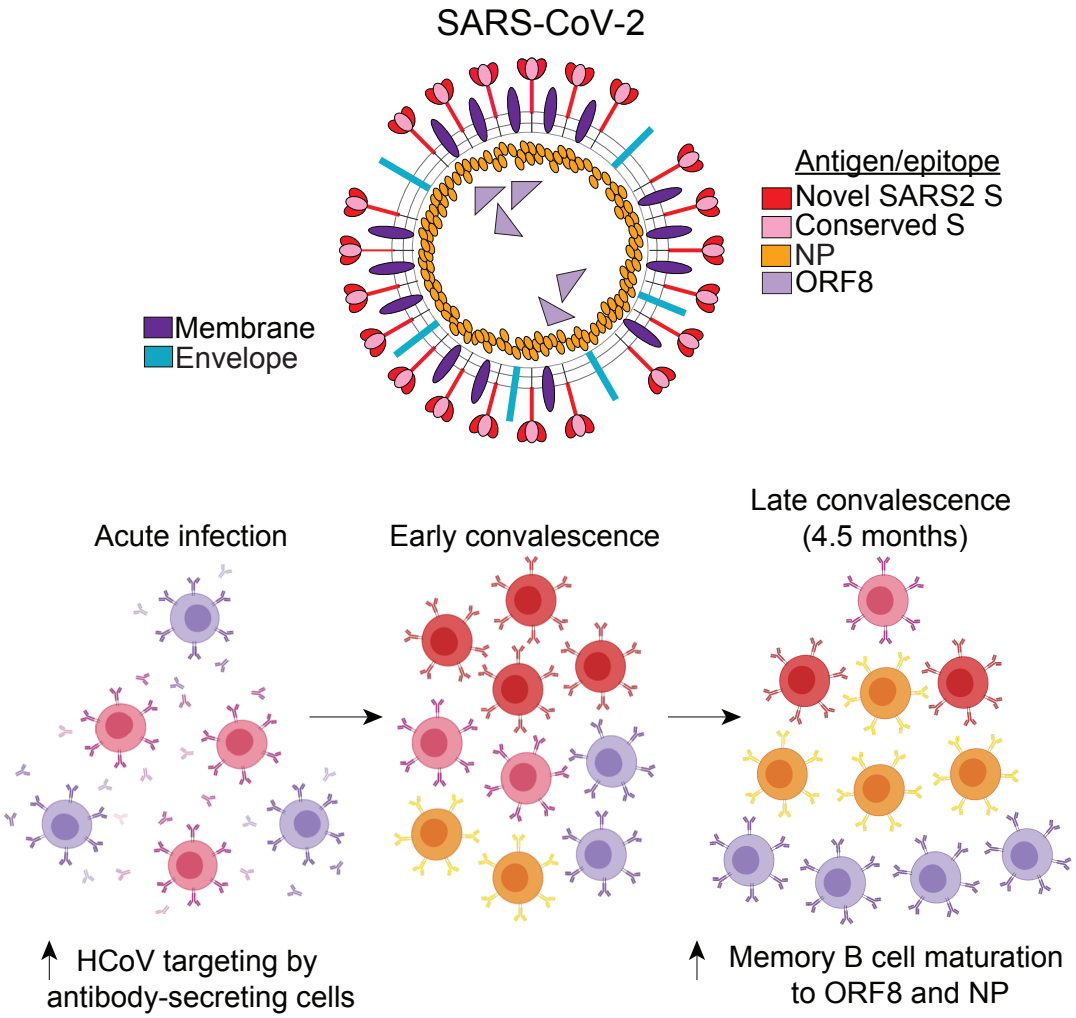


Figure 6.1: The immune landscape during infection with, and after recovery from, SARS-CoV-2
Early in the response a large burst of B cells reactive to endemic HCoVs is observed, however after infection their numbers have fallen and been replaced by SARS2 antigen specific cells, including protective responses against the spike protein. In late convalescence, even less memory to HCoVs can be observed, with non-protective memory having undergone a substantial expansion.

By sampling the immune responses of multiple subjects at various points in infection we can begin to understand the shifting focus of the humoral immune response to SARS-CoV-2. Surprisingly, we found there to be a substantial response from B cells that bind the spike proteins of endemic human coronaviruses in all subjects. Serological studies have previously suggested this could be the case, though the nature of these antibodies and their source remained unclear.[273, 274] We identified the source of such antibodies to presumably be recently activated, short lived antibody secreting cells. These B cells showed remarkably high amounts of affinity maturation, confirming they had differentiated from memory. Additionally, we found them all to have high expression of well-established mucosal homing genes, such as *CCR10*, *CXCR3*, and *ITGB7*. Because these cells are in a highly active state during infection with SARS-CoV-2 we conclude they are all cross-reactive between endemic HCoVs and SARS2. Indeed, we even identified several that bound probes for both sets of viruses in equal amounts and believe that the large number that did not can be explained by the nature of the staining and sorting. Ll probes were added at an equal concentration, as such they were four times as many endemic probes as SARS2 in the staining mixtures. The combination of this, the assumption their affinity will be higher for the endemic strains they are risen against, and the relatively low surface Ig expression of ASCs can reasonably explain the low amount of SARS2 spike binding by many of these cells.

During early convalescence we found spike reactive B cells to form the majority of the B cell response, showing the spike to be a powerfully immunodominant antigen. We also provided one of the earliest demonstrations that maintenance and amount of spike memory B cells did not correlate with serum titers, further demonstrating the importance of techniques that can better gauge memory responses on a cellular level. We also identified that B cells reacting to

intracellular antigens, such as ORF8 and NP, were largely naïve-like in character but could be found within memory as well.

However, by late convalescence there was a large expansion towards these internal, non-protective epitopes in all four subjects with matched blood draws. All four subjects also showed large decreases in the amount of endemic HCoV reactive memory B cells, potentially representing their loss in favor of memory formed against the more recent SARS2 infection. With regards to spike reactive memory, we found increases in two of four subjects, consistent amounts in one subject, and a massive loss of nearly all spike activity in the final subject. We hypothesize that, owing to the differences in the amount of these proteins within each vision, later germinal centers shift more of their focus onto the intraviral antigens and show a concurrent output of cells reactive to them.

However, these data represent only four subjects with a non-equal number of cells sorted from each. This raises the possibility of sampling bias compounded by any technical deficiencies that could account for the loss seen in the one subject. Notably, all four did show consistent increases in memory towards ORF8 and NP and decreases in memory to endemic HCoV spike proteins. Therefore, though this is a novel and important finding, it is worthy of further study to expand upon and better characterize the observed increases seen towards non-spike antigens.

6.3 The application of oligo-tagged antigen bait sorting to other diseases

The twentieth century has seen remarkable progress in combating and conquering infectious diseases. The dual advent of antibiotics and vaccines has been one of the greatest public health triumphs in human history, each saving countess numbers of lives every year. Developed nations have increasingly seen infectious disease drop farther down the list of leading

causes of death, being increasingly replaced with diseases of age and time such as cancer and cardiovascular disease. The eradication of smallpox in 1979 is a truly remarkable milestone in the long story of human history, ending a threat that must have easily killed more than a billion people throughout history and is estimated to have claimed 500 million lives in its last century of circulation.

The COVID-19 pandemic has been a stark reminder of the terrible power of infectious disease and the importance critical importance in continuing to study them and the immune responses against them. Though COVID-19 has exacted a major toll upon a weary world, it has only resulted in 15 to 100 times less deaths than the 1918 influenza pandemic. Emerging strains of influenza continue to pose the threat of causing another pandemic, especially avian strains that occasionally infect humans. Any one of which has the potential to prove far more deadly than COVID-19. Additionally, diseases that have faded from concern in much of the developed world, such as yellow fever, dengue, and tuberculosis, are gaining ground as new antibiotic resistant strains emerge and rising global temperatures allow their vectors to inhabit wider territories.

Beyond their re-emergence into locales from which they had previously been eradicated, many of these diseases have always and continue to pose large burdens across much of the southern hemisphere. Diseases such as malaria and tuberculosis cause as many as 450,000 and 1.4 million deaths per year, respectively. Malaria, caused by protists of the genus plasmodium, is known to require largely B cell mediated immunity for clearance and protection and there is growing evidence for a critical role for humoral immunity in the pathology and etiology of tuberculosis as well.

The prevalence of autoimmune diseases continues to rise with major consequences on the quality of life for many sufferers. B cells are implicated in many autoimmune disorders, both as the major effectors in those humorally mediated and potentially as the APCs that can break tolerance in T cell mediated autoimmunity.

Therefore, the technique and analysis pipelines presented in this body of work can be foundational for future studies of many other diseases. A better understanding of the types of B cells, their transcriptional profiles, and their antigenic targets involved in any infectious or autoimmune disease can be utilized to improve treatments and vaccination strategies, create monoclonal therapies, and in the generation of powerful new research reagents.

6.4 The heterogeneity of B cell subsets

In the course of profiling B cell responses to SARS-CoV-2 we uncovered a large amount of transcriptional heterogeneity within responding cells. Using transcriptional profiles alone we were able to identify 15 separate clusters across three distinct timepoints. For the purposes of our analyses, we grouped these clusters into five larger categories: naïve and recently activated B cells, memory B cells, antibody secreting cells, recent germinal center emigrants, and innate-like B cells. However, future work focusing on differences between clusters within the larger categories holds much promise for new discoveries in basic B cell biology.

One notable finding warranting further exploration within our own dataset is memory B cell cluster 9, in which the gene *FCRL5* is most differentially expressed compared to all other clusters. This gene has previously been associated with non-classical memory B cells following plasmodium infection, with conflicting reports of their behavior. Some work indicates them to be impaired and incapable of future responses, while other work finds they respond more robustly

upon re-infection than other memory B cells.[275, 276] Other work suggests they are transcriptionally poised to rapidly differentiate into plasmablasts upon re-infection, agreeing with the assessment of them as robust responders.[277] Our dataset could be used to further explore these possibilities and the gene expression profiles of FCRL5$^+$ B cells to better understand their roles within the body and during immune responses.

Many clusters also had very significant and highly differential expression of little studied genes. Many of these are lncRNAs or genes predicted to be protein coding from automatic annotation analyses of the human genome. Further in-vitro work, or potentially the identification of murine homologs and generation of knock-out mice, could better establish the roles of these genes within these populations leading to new discoveries in both immunology and genetics.

Lastly, on several subjects we were able to combine CITE-Seq data to measure the protein expression of several surface markers in conjunction with their transcriptional profiles. By developing new data processing models, we were able to distinguish additional clusters that were distinct from those generated by transcriptional profiling alone. Thus, future studies into the heterogeneity of B cells should consider using these additional reagents and computational pipelines to better characterize B cells at increasingly granular levels.

6.5 The impact of endemic HCoV infection on COVID-19 and universal coronavirus vaccines.

The impact of previous endemic human coronavirus infections remains to be fully investigated. Recent studies have conclusively shown cross-reactive serum that predates the emergence of SARS-CoV-2 to be entirely non-neutralizing.[278] However, this study was designed to address whether antibodies risen against past infections could be protective against SARS2 upon initial infection and did not address the question of whether they can continue to affinity mature towards neutralization of SARS2. It is well appreciated in the context of influenza that

novel strains, while largely escaped from previous humoral protection, can also boost rare cross-protective antibodies or lead to additional affinity maturation towards protection from antibodies not initially found to be so.[263, 279]

This leaves unaddressed the question of continued maturation of rare cross-protective antibodies, and potential strategies to boost them. There is a possibility that after infection and continued maturation several of these antibodies acquire cross-protective binding capability. In the light of continuing emerging variants of SARS2 this is a tantalizing possibility. It has been reported that vaccination after infection with SARS2 or a second boost of vaccine can drive stronger responses to epitopes not targeted upon initial infection or vaccination, and this also could be a potential pathway to protectiveness for these endemic reactive antibodies.

It had previously been thought that the pandemic potential of coronaviruses was rather low, especially when compared to other viruses such as influenza.[135, 280] SARS-CoV-1 had caused a serious epidemic in 2003 it was able to be contained by rapid public health measures.[281] MERS continues to infect humans in North Africa and the Arabian peninsula but has been determined to be unlikely to mutate in such a way to facilitate their rapid spread from human to human.[138, 282] However, the COVID-19 pandemic has proven several assumptions concerning coronaviruses to be incorrect, and even caused reexamination of past pandemics. Some epidemiologists now hypothesize that the pandemic of 1890, long believed to have been caused by a strain of influenza, may have actually been the effect of HCoV-OC43 gaining the ability to infect humans.[283] Intriguingly, phylogenetic predictions based on mutation rates places the emergence of HCoV-OC43 into humans around 1890.[284] This all highlights the need to create prophylactic, universal coronavirus vaccines to protect against future coronavirus pandemic threats, an area of active study.[285] This goal will require the identification of broadly cross-

reactive and strongly cross-protective antibodies biding highly conserved epitopes, and our collection of over 1000 cross-reactive antibodies and several novel public clones could serve as a powerful starting point for such a goal.

Our study lays a promising groundwork for these lines of inquiry, especially by identifying promising variable genes to home in on for cloning and testing through mutational and maturation analysis. However, this raises the major issue of these assumptions. Until antibodies from these ASCs can be expressed and tested, their cross-reactivity to SARS2 remains speculative, no matter how well founded in logic and reason the assumption is. Additionally, all of the novel public clones we discovered bind very weakly to the antigen our bait sorting approach identifies them as being reactive to. These antibodies should be further characterized by more sensitive techniques, such as surface plasmon resonance or biolayer interferometry.

6.6 Rapid production of monoclonal antibodies to known antigens are a powerful tool for additional discovery

Antibodies are famously powerful reagents, forming the basis of many modern laboratory and clinical techniques, and monoclonal antibodies are one of the most common sources of modern therapeutics. The techniques and analysis pipelines presented herein present a rapid way to discover and clone antibodies to specific antigenic targets on a protein, or even sub-protein level. The antibodies presented within this work continue to drive forward new discoveries and show promise in diagnostic and therapeutic settings. They are already being utilized as the basis of rapid COVID-19 saliva testing and are being tested by Meiji Pharmaceuticals to create a COVID-19 antibody therapy that is resistant to the emerging variants of concern. Active work in the Wilson laboratory is using them to characterize epitopes on the spike protein and identify which are protective against emerging variants. Structural work at Argonne National Laboratory

has used them in the generation of multiple crystal structures of NP and are using these structures to predict promising targets for small molecule inhibitors. Indeed, these antibodies are now the foundation of the largest patent in the history of the University of Chicago. This further highlights the impressive power of our technique to have rapid, positive impacts on human health.

6.7 The evolutionary history of AID and mutational targeting

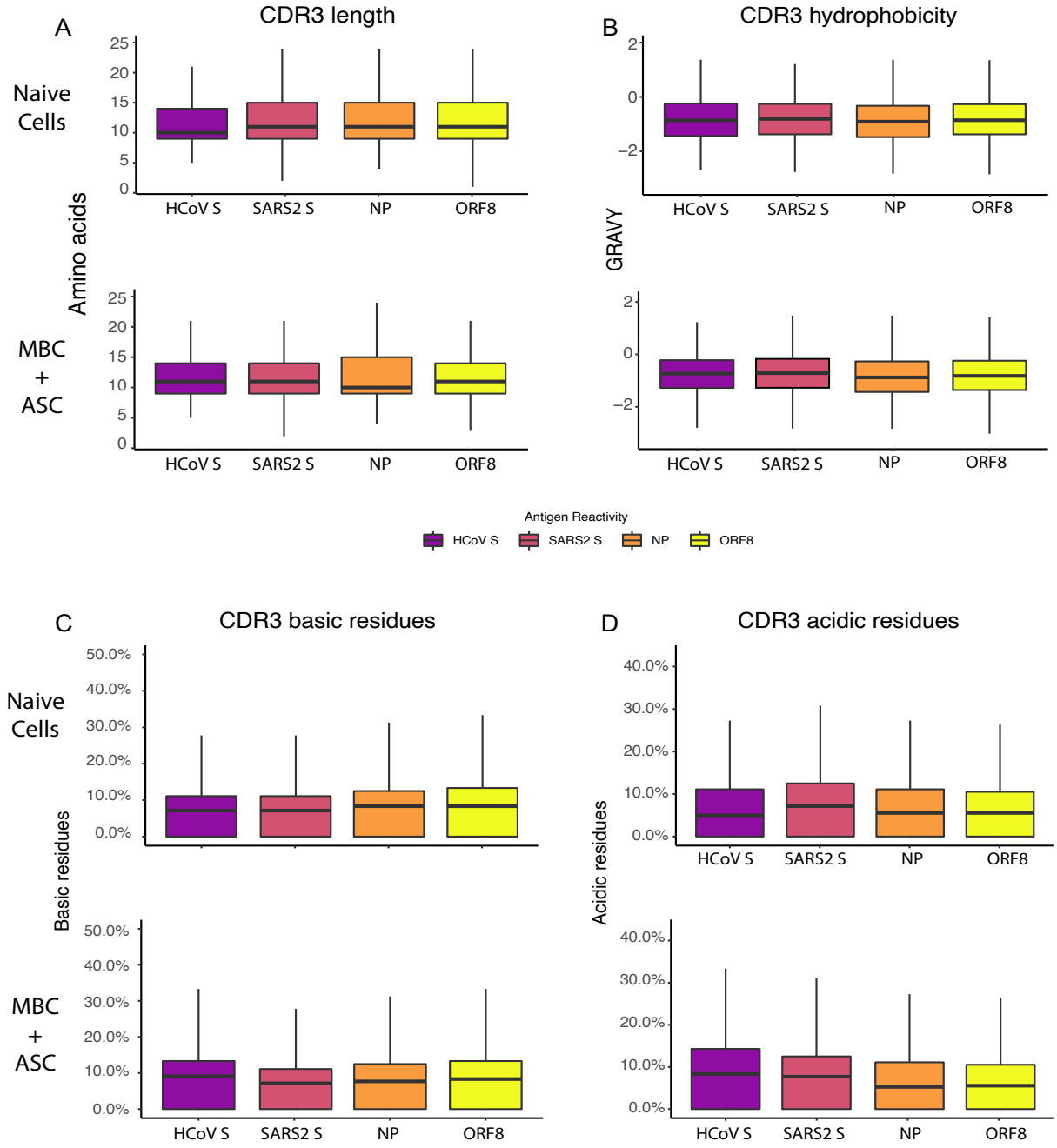


Figure 6.2: CDR3 amino acid residue properties in all antigen reactive naive or effector cells.

Figure 6.2, continued.
(A) CDR3 length for antigen specific cells. The top shows all naive cells and the bottom all memory B cells and antibody secreting cells.
(B) Grand average of hydrophobicity (GRAVY) of CDR3s for antigen specific cells. The top shows all naive cells and the bottom all memory B cells and antibody secreting cells.
(C) Frequency of residues which are basic within CDR3s for antigen specific cells. The top shows all naive cells and the bottom all memory B cells and antibody secreting cells.
(D) Frequency of residues which are acidic within CDR3s for antigen specific cells. The top shows all naive cells and the bottom all memory B cells and antibody secreting cells.

Activation induced cytidine deaminase and its associated variable gene sequence based mutational targeting is an intricate and complex system shaped by eons of evolution. Within humans and mice, the most common animal model used in immunology studies, AID plays the role of the affinity maturer. In these organisms somatic hypermutation drives a rapid Darwinian process within germinal centers that can increase the affinity of B cell receptors many times over.[95, 99] However, for many other vertebrates it appears that the process of somatic hypermutation largely exists to diversify their B cell repertoires by generating large amounts of, presumably, non-selected mutations during germinal center reactions that occur in development and early adolescence.[129, 130] Therefore it is conceivable that this process has left its evolutionary mark in those species which use SHM to drive affinity maturation.

This study generated a large collection of thousands of BCRs, both mutated and un-mutated, from a variety of B cell types. This allows for high throughput analysis of mutational patters combined with antigen specificity. We decided to compare the CDR3s, a site of frequent mutation during affinity maturation that is critical for binding antigens, from all antigen specific cells divided by transcriptional identity. Some differences are apparent, such as a selection for longer CDR3s for HCoV S and NP reactive B cells (Fig 6.2 A) or a decrease in the acidity of CDR3s for HCoV S reactive B cells (Fig 6.2 D) but, overall, there appears to be little change

between naive, unmutated BCRs and those that have survived germinal centers (Fig 6.2 A-D). This suggests that it is possible that, for the most part, the process of affinity maturation 'locks in' the features that first allows a B cell to bind its antigen. Further, this could also play a role in other vertebrates, in which mutations might impact the types of antigens a BCR targets and can bind, while maintaining the overall residue patterns at use so their usage stays largely consistent across the entire repertoire. Because we had observed extensive amounts of mutations accumulating in the FWR3s of BCRs reacting to HCoV S we decided to perform this analysis on FWR3 as well. Again, the results show little change between naive and effector cells (Fig. 6.3 A-C). This, combined with computational mutational modeling of 'de-evolved' BCRs suggests that the overall preservation of BCR structure and residue characteristics is another facet evolved into the intricate targeting of AID.

Future studies should consider applying this method to additional datasets. Immunization studies in mice could allow for a more controlled approach and would be of value to the field. Additionally, we had wanted to see if they were observed differences in mutational patterns between innate-like B cells and other human B cells, as they too can enter into microbiome driven GCs in the mucosa where they potentially mutate without selection.[26, 286] However, our dataset did not contain enough of these cells to perform proper analysis, and as such future studies should consider specifically sorting on them in addition to follicular B cells, to compare the two.

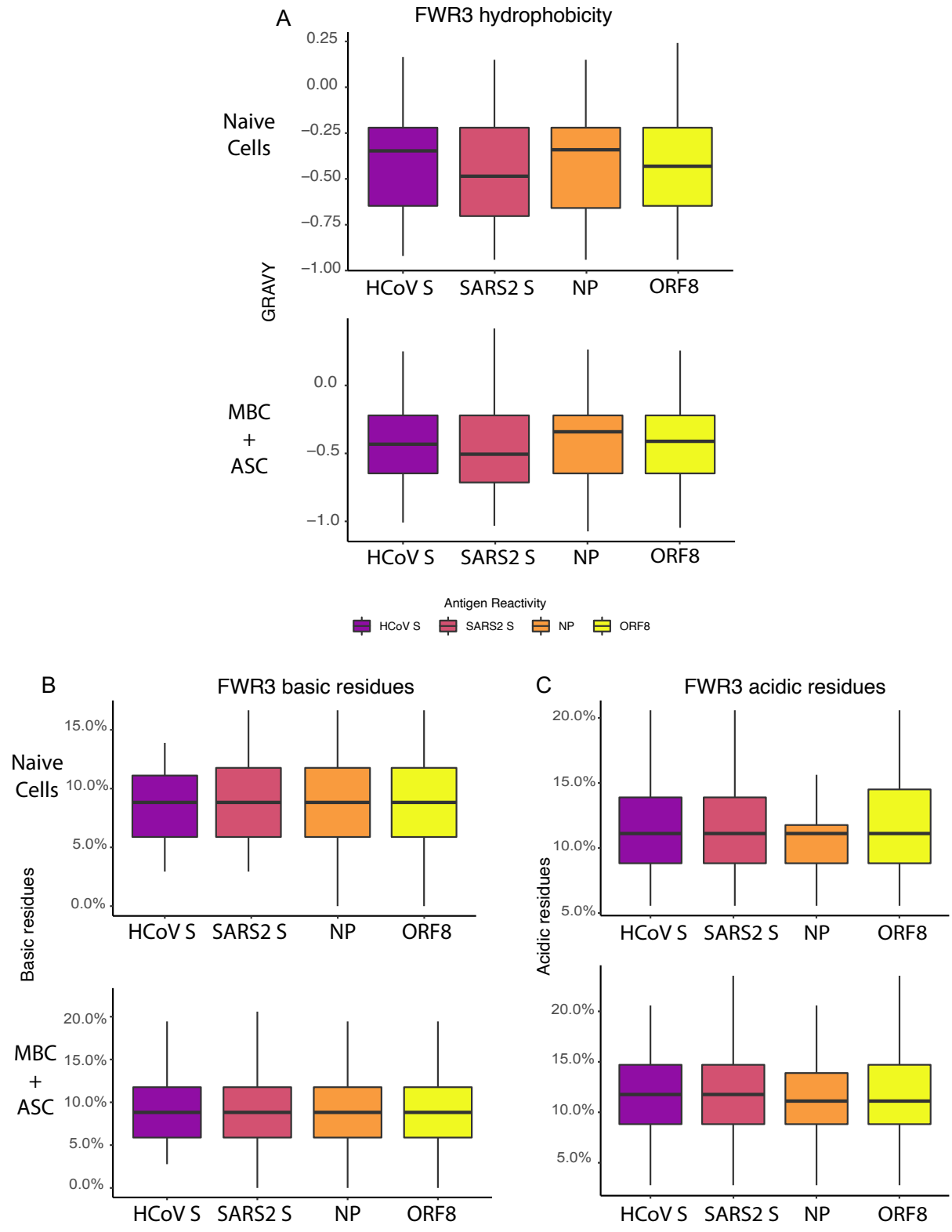


Figure 6.3: FWR3 amino acid residue properties in all antigen reactive naive or effector cells.

Figure 6.3, continued
(A) Grand average of hydrophobicity (GRAVY) of FWR3s for antigen specific cells. The top shows all naive cells and the bottom all memory B cells and antibody secreting cells.
(C) Frequency of residues which are basic within FWR3s for antigen specific cells. The top shows all naive cells and the bottom all memory B cells and antibody secreting cells.
(D) Frequency of residues which are acidic within FWR3s for antigen specific cells. The top shows all naive cells and the bottom all memory B cells and antibody secreting cells.

6.8 Summary of future directions and issues

- Our matched subjects show an increase in the targeting of internal antigens over time, but it is limited to only four subjects and an approximate maximum of 4.5 months post-infection. Additional subjects at additional time points could further clarify the dynamics of the immune response to COVID-19.
- We have discovered a large amount of heterogeneity in B cell responses during and after infection. Some of what we have found aligns with previous studies while other gene markers are potentially novel. Additional study should be performed to better characterize and confirm the identification of the clusters, and any potential impact in their maintenance or functions from genes of unknown function associated with them. This could be extended into animal models if homologs can be identified in their B cell subsets.
- We found a substantial number of B cells reactive to endemic human coronaviruses to be activated and largely trafficking to mucosal sites in severe, acute cases of COVID-19. These are worthy of additional consideration as a source of cross-protective antibodies and targets of universal coronavirus vaccination strategies. The most critical next step is to express and test the binding of some of these antibodies, to confirm they are cross-reactive. Additionally, if they are truly cross-reactive, without testing the monoclonals themselves it remains speculative whether they are helpful, harmful, or neutral.

- Additional antibodies to all targets are worth cloning and testing. We initially chose many antibodies in a semi-random fashion, with a focus on memory clusters and those with high probe scores. Though we found probe score and affinity not to have a clear correlation. Additional cloning of antibodies would also allow us to test out a system designed to use mutational profiles and variable gene usage patterns to aid in the selection of monoclonal expression.
- Several hypotheses about mutation and selection have been put forward based on the work, however, these all rely upon a single dataset that is based on a novel pathogen. The novel nature of SARs-CoV-2 could skew results in a way not indicative of immune responses to other pathogens or upon repeat exposure, and as such much of this work should be applied to other datasets generated in a similar fashion.

6.9 Final Conclusions

The data presented here advance our understanding of B cell responses to infection with SARS-CoV-2, provide new insights into fundamental B cell biology, and introduce novel techniques to generate and analyze highly informative multidimensional data. We have shown the power of these techniques to have an immediate impact on human health. In only a few months after the declaration of a pandemic we had generated data critical to the understanding of immunity against SARS-CoV-2, provided resources that are being used to develop new therapeutics and diagnostics, and were the first in the world to express and test monoclonal antibodies to antigenic targets beyond the spike protein. We also demonstrate an increase in memory to non-protective targets over time, further confirming the importance of vaccination to finally end the COVID-19 pandemic. Finally, we feel that our transcriptional data can be foundational for future studies of antiviral immunity, and our approach to generating and

analyzing such data applied to a wide array of human disease to produce key new insights and discoveries.

Tables

Subject ID	Age	Sex	Reported symptoms*	Severity Score	Severity Category	Duration of symptoms (days)	Symptom start to donation (days)
24	34	M	Fatigue, cough, SOB, SC, fever, headache, BAP, diarrhea, LOS, LOT	19	Severe	12	41
20	31	M	Fatigue, cough, SOB, SC, fever, headache, BAP, LOS, LOT	29	Critical	19	48
564	24	F	Fatigue, cough, SOB, SC, ST, fever, headache, BAP, diarrhea, LOS, LOT	24	Severe	32	60
144	56	M	Fatigue, cough, SC, ST, headache, BAP, LOS	17	Moderate	23	54
214	47	M	Fatigue, cough, SOB, SC, ST, headache, BAP, LOS, LOT	20	Severe	24	59

Table 3.2: Convalescent patient information

171	37	F	Fatigue, cough, SOB, SC, fever, headache, BAP, diarrhea, LOS, LOT	21	Severe	16	44
92	35	M	Fatigue, cough, SC, ST, fever, headache, BAP	16	Moderate	16	47
48	45	F	Fatigue, cough, SOB, SC, ST, fever, headache, AP, diarrhea, LOS, LOT	19	Severe	8	40
537	36	M	Fatigue, cough, fever, BAP	13	Moderate	14	59
586	32	F	Fatigue, cough, SOB, SC, headache, BAP, AP, diarrhea	18	Moderate	17	61
376	36	F	Diarrhea, LOS, LOT	8	Mild	7	48
305	43	F	Fatigue, cough, SC, ST, fever, headache, BAP, LOS, LOT	14	Moderate	4	47
116	65	F	Cough, SOB, fever, LOS, LOT	13	Moderate	18	49

Table 3.2, continued.

166	42	F	Fatigue, cough, SOB, SC, fever, headache, BAP, diarrhea, LOS, LOT	18	Moderate	17	55
155	47	F	Fatigue, cough, SOB, ST, fever, BAP, LOS, LOT	20	Severe	29	64
609	26	F	Fatigue, SOB, ST, fever, headache, BAP, LOS, LOT	16	Moderate	7	57
130	52	M	Fatigue, SC, headache, LOS, LOT	10	Mild	7	35
281	70	M	Cough, fever, BAP	9	Mild	7	48
272	42	M	Fatigue, cough, SOB, fever, headache, BAP, LOS, LOT	18	Moderate	14	43
50	35	M	Fatigue, SC, fever, BAP, LOS, LOT	13	Moderate	10	40
65	40	F	Fatigue, SC, fever, headache, BAP, diarrhea, LOS, LOT	16	Moderate	13	47

Table 3.2, continued.

33	36	M	Fatigue, cough, SOB, SC, fever, headache, BAP, AP, diarrhea, LOS, LOT	22	Severe	14	48
201	56	M	Fatigue, cough, SOB, SC, ST, fever, headache, BAP, LOS, LOT	20	Severe	18	58
218	51	F	Fatigue, cough, SOB, fever, headache, BAP, AP, diarrhea	19	Severe	19	48
266	19	F	Fatigue, cough, SC, headache, BAP	9	Mild	4	32, 137 V2*
356	51	F	Fatigue, cough, ST, fever, headache, BAP, AP, diarrhea, LOS, LOT	20	Severe	14	43, 137 V2*
407	34	M	Fatigue, cough, SC, fever, BAP, AP, diarrhea, LOS, LOT	16	Moderate	11	43, 131 V2*
210	47	M	Fatigue, cough, SOB, fever, headache, BAP, LOS, LOT	16	Moderate	7	41, 125 V2*

Table 3.2, continued.

Table 3.2, continued: *SOB = shortness of breath; SC = sinus congestion; ST = sore throat; BAP = body aches and pain; AP = abdominal pain; LOS = loss of smell; LOT = loss of taste. Starred symptom start to donation values indicate the value for follow-up visit donation (V2). Severity scoring method has been described previously (Guthmiller et al., 2021).

Median Age	41
Mean Age	42
Mode Age	47
Range Age	19–70
Number of Males	14
Number of Females	14
Median Duration of Symptoms (days)	14
Mean Duration of Symptoms (days)	14
Mode Duration of Symptoms (days)	7
Range Duration of Symptoms (days)	4–32
Median symptom start to donation (days)	48
Mean symptom start to donation (days)	49
Mode symptom start to donation (days)	48
Range symptom start to donation (days)	32–64

Table 3.3: Distribution of clinical parameters for convalescent patients included in the study.

Subject ID	Age	Sex	Reported symptoms*	Symptom start to first donation (days)	Co-morbidities*	COVID treatment
R1	57	M	Fever, cough, nausea	3	HTN, DM, NAFLD	Tocilizumab, mechanical ventilation
R2	61	M	Cough, weakness, hiccups, altered mental status	16	None	Hydroxychloroquine, nasal cannula
R3	51	F	Fever, cough, dyspnea	21	HTN, DM, PE, asthma	Remdesivir, tocilizumab, venovenous ECMO*
R4	70	F	Fever, altered mental status	2	HTN, Alzheimer's disease	Nasal cannula
R5	66	F	Altered mental status, dyspnea	9	HTN, PE/DVT, recent hospitalization for orthopedic procedure	Nasal cannula
R6	59	M	Fever, chills, decreased appetite, dizziness	20	HTN, DM	Remdesivir, tocilizumab, Venovenous ECMO
R7	57	M	Dyspnea	9	HTN, Myelodysplastic syndrome s/p stem cell transplant	Tocilizumab, anakinra, nasal cannula

Table 3.4: Severe acute patient information

R8	30	M	Fever, chills, fatigue, LOT*	13	Cystic fibrosis s/p bilateral lung transplant, DM	Room air
R9	78	M	Fever, cough	14	HTN, prostate cancer	High-flow nasal cannula
R10	86	F	Dyspnea, abdominal pain	6	ESRD on HD, stroke, PVD s/p AKA, DM, PE/DVT, CHF	Nasal cannula

Table 3.4, continued:
*LOT: Loss of taste; ECMO: Extracorporeal membrane oxygenation.
*AKA, above the knee amputation; CHF, congestive heart failure; DM, diabetes mellitus; DVT, deep venous thrombosis; ESRD, end-stage renal disease; HTN, hypertension; NAFLD, non-alcoholic fatty liver disease; PE, pulmonary embolism; PVD, peripheral vascular disease.

Median Age	60
Mean Age	61.5
Mode Age	57
Range Age	30–86
Number of Males	6
Number of Females	4
Median symptom start to first donation* (days)	11
Mean symptom start to first donation* (days)	11
Mode symptom start to first donation* (days)	9
Median symptom start to last donation* (days)	25
Mean symptom start to last donation* (days)	25
Mode symptom start to last donation* (days)	23

Table 3.5: Distribution of clinical parameters for severe acute patients included in the study.

Gene	q-val (adj p val)	Cluster
SELL	2.27E-284	0
HLA-DRA	5.20E-275	0
CCR7	6.42E-261	0
PFDN5	3.78E-217	0
HLA-DQB1	2.49E-208	0
MEF2C	4.29E-208	0
HVCN1	2.99E-199	0
HLA-DRB5	3.33E-192	0
FAM26F	2.90E-175	0
SARAF	1.64E-171	0
BTG1	2.00E-104	1
TSC22D3	2.14E-67	1
DUSP1	2.70E-64	1
JUN	1.71E-37	1
ZFP36L2	1.02E-17	1
RPS26	1.07E-14	1
RPS4Y1	9.15E-08	1
MT-ND6	0.02353042	1
SMARCB1	2.80E-286	2

Table 3.6: Top 10 most differentially expressed genes per transcriptional cluster. Significance is indicated by adjusted p value.
*Cluster 1 only expressed 8 genes above the significance threshold.

TNFRSF13B	1.36E-285	2
SAT1	8.77E-261	2
EEF1B2	2.92E-259	2
RPL9	1.16E-257	2
UBC	9.26E-256	2
CD82	3.50E-254	2
H3F3B	3.51E-244	2
DUSP2	1.32E-239	2
CTSH	7.68E-239	2
RPS24	2.13E-280	3
FOS	2.22E-224	3
YBX3	6.57E-210	3
BTG1	8.51E-189	3
FAM129C	7.36E-183	3
CD79B	6.60E-170	3
H3F3A	3.45E-165	3
CD79A	1.33E-162	3
DUSP1	3.20E-159	3
CXCR4	2.99E-149	3
C4orf48	3.48E-303	4
MTMR14	1.41E-300	4

Table 3.6, continued.

ARL6IP5	3.78E-300	4
PDLIM1	6.07E-300	4
ATP5B	1.01E-299	4
POU2AF1	1.25E-299	4
TMEM156	2.81E-299	4
PPP1CA	1.13E-294	4
RPLP1	3.43E-287	4
PKM	5.02E-287	4
FCRLA1	2.29E-293	5
HCK	1.59E-279	5
GADD45B1	5.30E-258	5
VPREB31	6.56E-242	5
ACTG1	2.26E-228	5
BCL7A1	2.37E-217	5
UCP22	5.88E-212	5
ACTB	9.61E-211	5
CD79A1	1.40E-196	5
RPS11	1.65E-187	5
GAPDH	5.19E-290	6
ITGB1	5.43E-272	6
SYK1	1.26E-271	6
FCRL2	3.00E-262	6

Table 3.6, continued.

CD27	5.44E-261	6
PLAC8	7.41E-240	6
ANXA4	1.65E-239	6
AC079767.4	1.66E-239	6
AP2S11	2.27E-237	6
CD241	4.50E-236	6
RPL14	3.83E-301	7
PDE4D	8.43E-301	7
RPS25	4.39E-284	7
RPS15A	1.62E-221	7
RPS3	3.64E-214	7
TPT1	3.28E-212	7
RPL32	8.50E-211	7
RPLP2	1.88E-210	7
RPS6	7.54E-196	7
RPS181	4.77E-191	7
MT-CO1	3.28E-256	8
MT-CO2	2.24E-177	8
MT-CYB	3.85E-147	8
MT-ND4L	6.61E-120	8
MT-CO3	4.50E-108	8
MT-ND3	2.45E-87	8

Table 3.6, continued.

MT-ND5	1.62E-76	8
MT-ATP8	1.02E-62	8
MT-ND1	9.47E-57	8
MZB1	5.49E-45	8
FCRL5	1.93E-301	9
HLA-DRB1	1.33E-294	9
HLA-DPA1	7.60E-247	9
RGS2	2.48E-245	9
HLA-DPB1	1.47E-237	9
B2M	4.95E-231	9
HCK	2.78E-222	9
HLA-DQA1	4.05E-222	9
LITAF	3.32E-217	9
RHOB	7.50E-217	9
MRPL27	7.31E-303	10
FAM173A	4.65E-300	10
OAS1	1.41E-299	10
COA3	9.63E-296	10
TUFM	3.86E-294	10
GOLGB1	1.66E-293	10
MFF	1.89E-293	10
UQCRH	8.40E-293	10

Table 3.6, continued.

NDUFA4	1.38E-291	10
MRPL24	1.87E-291	10
SRPRB	3.60E-302	11
CALU	2.84E-301	11
TMEM106C	8.66E-300	11
SLC35B1	2.73E-298	11
PRDX4	4.73E-291	11
MT2A	1.77E-290	11
YWHAE	4.34E-290	11
OSTC	1.63E-287	11
CD38	1.34E-286	11
DNAJB11	2.82E-286	11
LCP2	2.59E-299	12
RHOC	8.46E-296	12
RARRES3	1.81E-289	12
CTSD	1.62E-286	12
AAK1	4.65E-276	12
IFITM2	3.02E-248	12
B2M	1.97E-233	12
DDIT4	3.48E-233	12
S100A11	4.36E-230	12
TPST2	7.72E-228	12

Table 3.6, continued.

TPST2	1.01E-275	13
XAF1	3.67E-223	13
MX1	4.71E-206	13
IFITM1	5.27E-163	13
EPSTI1	1.44E-160	13
EIF2AK2	7.68E-157	13
IFI6	1.65E-149	13
PLSCR1	1.33E-147	13
ISG15	1.32E-146	13
STAT1	1.83E-140	13
MX2	2.00E-121	13
MT2A	6.19E-302	14
MS4A7	4.81E-301	14
CCDC88A	9.92E-300	14
MYD88	1.02E-284	14
NAGA	1.47E-279	14
FAM45A	8.59E-262	14
RNF144B	7.72E-257	14
ALDH2	5.68E-254	14
SLC7A7	5.93E-250	14
GBP2	6.72E-242	14
AQP3	4.41E-291	15

Table 3.6, continued.

TXNDC5	1.42E-256	15
LGALS3	2.35E-233	15
FKBP11	3.31E-229	15
PRDM1	4.18E-225	15
FNDC3B	5.54E-204	15
IFI27L1	4.57E-196	15
SDF2L1	9.78E-193	15
DERL3	2.11E-188	15
IFI27	1.05E-187	15

Table 3.6, continued.

Gene	B Cell Subset	Rationale	Citation
BACH2	Naïve	Promotes B cell development, maintains mature B cells	(Itoh-Nakadai et al., 2014)
ZBTB16	Naïve	Downregulated in memory compared to naïve	(Moroney et al., 2020)

Table 3.7: Key genes used in the identification of B cell subsets, Related to Figure 3.3 and Figure 3.4

APBB2	Naïve	Foxp1 target important for mature FO B cell survival	(Patzelt et al., 2018); The Human Protein Atlas (Uhlen et al., 2015)
SPRY1	Naïve	Proliferation inhibitor, differentially expressed (DE) between naïve and memory	(Frank et al., 2009) The Human Protein Atlas (Uhlen et al., 2015)
TCL1A	Naïve	DE between B cell pop. High in Naïve, low in GC, absent in memory and ASC	(Said et al., 2001)
IKZF2	Naïve	DE between memory and naïve, higher in naïve	(Moroney et al., 2020)
CD27	Memory	Classic memory marker	(Palm and Henry, 2019)
CD86	Memory	DE between memory and naïve, higher in memory	(Axelsson et al., 2020)
RASSF6	Memory	Increased in memory	(Moroney et al., 2020)
TOX	Memory	Increased in memory	(Moroney et al., 2020)
TRERF1	Memory	Increased in memory	(Moroney et al., 2020)
TRPV3	Memory	Increased in memory	(Moroney et al., 2020)
POU2AF1	Memory	B cell-specific TF	(Zhao et al., 2008)

Table 3.7, continued.

RORA	Memory	Increased in memory	(Moroney et al., 2020)
TNFRSF13B	Memory	BAFF-binding receptor expressed in memory and ASC	(Muller-Winkler et al., 2021)
CD80	Memory	High affinity memory marker	(Palm and Henry, 2019)
FCLR5	Memory	Atypical memory marker	(Kim et al., 2019)
GDPD5	Class-switched Memory	Highest in class-switched memory B cells	The Human Protein Atlas (Uhlen et al., 2015)
BAIAP3	Class-switched Memory	DE in switched memory, ion channel Ca^{2+} flux	(Moroney et al., 2020)
TGM2	Class-switched Memory	DE in switched memory, Ca^{2+} signal transduction	(Moroney et al., 2020)
MUC16	Class-switched Memory	DE in class-switched memory, membrane adhesion	(Moroney et al., 2020)
PRDM1	ASC	Lineage-defining TF	(Lightman et al., 2019)
MANF	ASC	ER stress	(Lightman et al., 2019)
XBP1	ASC	Unfolded protein response	(Lightman et al., 2019)
IL6R	ASC	Receptor for IL6, promotes PC fate and mAb production	(Dienz et al., 2009)

Table 3.7, continued.

BCL6	ASC	Drops in GCs to promote PC fate	(Palm and Henry, 2019)
IRF4	ASC	Rises as BCL6 drops to promote PC fate	(Palm and Henry, 2019)
TNFSR17	ASC	Genetic KOs experience sig PC reduction	(Lightman et al., 2019)
CD38	ASC	Classic PC marker	(Lightman et al., 2019)
NT5E	GC emigrant / recent MBC	Important for class-switch	(Schena et al., 2013)
MKI67	GC emigrant / recent MBC	Proliferation marker	(Scholzen and Gerdes, 2000)
CD40	GC emigrant / recent MBC	Required for memory formation	(Basso et al., 2004)
CD83	GC emigrant / recent MBC	GC composition	(Krzyzak et al., 2016)
MAP3K8	GC emigrant / recent MBC	DE during GC reaction	(Wohner et al., 2016)
MAP3K1	GC emigrant / recent MBC	Required for CD40 signaling	(Gallagher et al., 2007)

Table 3.7, continued.

FAS	GC emigrant / recent MBC	DE during GC reaction	(Smith et al., 1995)
Marginal Zone genes	Marginal Zone B cells	DE in MZBs	(Descatoire et al., 2014)
SPN	B1 B Cells	Classic B1 marker	(Rothstein et al., 2013)
MYO1D	B1 B Cells	DE in B1s	(Macias-Garcia et al., 2016)
PLSCR1	B1 B Cells	Expressed in natural ASCs	Cordero et al
PSTPIP2	B1 B Cell	DE during activation	(Ochiai et al., 2020)
AHR	B1 B Cell	Highest expression in B1	(Villa et al., 2017)
CD300LF	B1 B Cell	DE in B1s	(Macias-Garcia et al., 2016)
LYSMD2-GPR55	B1 B Cell	DE in mouse B1s	(Mabbott and Gray, 2014)
IZUMO1R	B1 B Cell	DE in B1s	(Macias-Garcia et al., 2016)
TNFSF13B-MYD88	Innate-like B cells	Highly expressed in MZB and B1	(Smulski and Eibel, 2018)

Table 3.7, continued.

mAb ID	Specificity	Isotype	Cluster	# HC SHM	VH Gene	# LC SHM	VK/L gene
S20-15	Spike/RBD	IgG1	4	8	VH 4-59	1	VL 3-21
S20-31	NP	IgG4	4	30	VH 1-24	22	VK 3-20
S20-40	NP	IgM	5	0	VH 4-4	1	VL 2-14
S20-58	Spike/RBD	IgG1	7	5	VH 4-30	2	VK 2-24
S20-74	Spike/RBD	IgG1	6	6	VH 4-59	3	VL 2-8
S20-86	Spike	IgG1	3	9	VH 3-9	2	VL 2-14
S24-68	ORF8	IgG1	4	4	VH 4-59	3	VL 1-44
S24-105	ORF8	IgG1	4	6	VH 3-48	4	VK 3-20
S24-178	NP	IgG1	4	2	VH 3-33	7	VL 2-14
S24-188	NP	IgG3	4	2	VH 1-69	3	VL 2-14
S24-202	NP	IgG1	6	3	VH 5-10	6	VK 3-11
S24-278	ORF8	IgG1	4	3	VH 1-2	1	VK 3-20
S24-339	Spike/RBD	Unknown	4	5	VH 3-49	1	VK 3-15
S24-472	ORF8	IgG1	4	5	VH 4-4	4	VL 4-16
S24-490	ORF8	IgM	7	2	VH 1-46	4	VK 3-20
S24-494	Spike/RBD	IgG3	2	0	VH 4-39	0	VK 1-39
S24-566	ORF8	IgG1	4	3	VH 3-49	1	VK 2-28
S24-636	ORF8	IgD	7	1	VH 3-7	4	VL 8-61
S24-740	ORF8	IgG1	4	5	VH 1-3	1	VK 4-1
S24-791	NP	IgG1	4	4	VH 4-59	6	VK 3-20

Table 3.8: Specific mAbs generated from single B cell heavy and light chain gene sequences, Related to Figure 4 and Figure 5.

S24-902	Spike/RBD	IgG1	6	0	VH 1-69	0	VL 7-46
S24-921	NP	IgG1	4	8	VH 4-59	7	VK 1-39
S24-1063	NP	IgG1	4	3	VH 4-59	1	VK 3-20
S24-1224	Spike/RBD	IgG1	4	7	VH 1-46	7	VL 1-40
S24-1271	Spike/RBD	IgM	4	6	VH 3-66	6	VL 3-1
S24-1339	Spike/RBD	IgG1	7	1	VH 3-53	1	VK 3-20
S24-1345	ORF8	IgD	3	0	VH 4-39	0	VK 1-13
S24-1378	ORF8	IgM	3	0	VH 3-53	0	VL 8-61
S24-1379	NP	IgG1	0	0	VH 4-59	0	VL 1-47
S24-1384	Spike/RBD	IgG1	4	2	VH 3-48	4	VL 3-21
S24-1476	Spike/RBD	IgG	1	2	VH 3-49	0	VK 3-15
S24-1564	NP	IgG1	4	10	VH 4-59	4	VK 1-39
S24-1636	NP	IgG1	2	3	VH 3-33	0	VK 3-11
S24-1002	Spike/RBD	IgM	3	3	VH 3-30	5	VK 1-13
S24-1301	Spike	IgG1	3	4	VH 1-24	4	VL 10-54
S24-223	Spike/RBD	IgM	2	1	VH 2-5	3	VL 2-14
S24-461	Spike/RBD	IgG1	4	7	VH 4-59	3	VL 3-16
S24-511	NP	IgD	2	0	VH 3-30	0	VL 3-1
S24-788	Spike/RBD	IgM	6	0	VH 3-33	1	VL 3-1
S24-821	Spike/RBD	IgM	4	4	VH2-70	0	VK 1-5
S144-67	Spike/RBD	IgG1	4	7	VH 5-51	5	VL 1-40
S144-69	Spike/RBD	IgG1	4	2	VH 5-51	3	VK 1-5

Table 3.8, continued.

S144-94	ORF8	IgG3	4	11	VH 3-30	0	VK 2-28
S144-113	ORF8	IgG1	4	9	VH 3-23	6	VK 1-39
S144-175	ORF8	IgG1	4	9	VH 1-2	1	VL 1-47
S144-208	ORF8	IgG1	4	6	VH 1-2	7	VL 2-11
S144-339	NP	IgG1	4	11	VH 3-21	7	VK 3-20
S144-359	ORF8	IgG3	4	5	VH 3-23	5	VK 1-39
S144-460	Spike/RBD	IgA1	4	34	VH 3-15	24	VK1D-17
S144-466	Spike/RBD	IgG3	4	6	VH 5-51	6	VK 1-5
S144-469	ORF8	IgG1	4	3	VH 4-59	2	VK 2-28
S144-509	Spike/RBD	IgG1	4	3	VH 5-51	1	VK 1-5
S144-516	ORF8	IgG1	4	5	VH 1-2	7	VL 1-40
S144-568	Spike/RBD	IgA2	7	11	VH 4-59	11	VK 3-20
S144-576	Spike/RBD	IgG1	4	3	VH 1-69	2	VK 1-5
S144-588	ORF8	IgG1	4	1	VH 4-39	3	VL 3-1
S144-628	Spike/RBD	IgA1	6	9	VH 5-51	10	VL 1-40
S144-740	ORF8	IgG1	4	1	VH 1-2	5	VK 3-20
S144-741	ORF8	IgG1	4	5	VH 1-2	1	VL 1-44
S144-803	Spike/RBD	IgG1	4	5	VH 5-51	3	VK 1-5
S144-843	ORF8	Unknown	6	20	VH 3-30	8	VK 3-20
S144-877	Spike/RBD	IgG1	4	2	VH 3-30	6	VK 1-33
S144-952	NP	IgM	4	4	VH 1-18	2	VK 4-1

Table 3.8, continued.

S144-971	ORF8	IgG1	4	6	VH 3-64	3	VK 4-1
S144-1036	NP	IgG1	4	2	VH 4-34	5	VK 4-1
S144-1079	Spike/RBD	IgG1	4	7	VH 1-69	3	VK 3-20
S144-1299	ORF8	IgG1	4	5	VH 4-59	0	VL 1-47
S144-1339	Spike/RBD	IgG1	4	12	VH 1-2	5	VL 2-14
S144-1406	Spike/RBD	IgG2	4	3	VH 1-3	0	VK 1-5
S144-1407	Spike/RBD	IgG1	4	6	VH 1-69	2	VK 1-5
S144-1569	ORF8	IgG1	4	7	VH 1-18	1	VL 9-49
S144-1827	Spike/RBD	IgM	2	20	VH 3-7	5	VK 3-20
S144-1848	NP	IgG1	4	4	VH 3-21	8	VL 1-47
S144-1850	Spike/RBD	IgG1	4	2	VH 3-23	3	VK 1-5

Table 3.8, continued.

S144-2234	ORF8	IgG1	4	4	VH 1-69	3	VK 4-1
S564-105	NP	IgG1	4	5	VH 4-61	2	VL 2-14
S564-14	Spike/RBD	IgD	6	3	VH 3-7	0	VK 3-21
S564-68	Spike/RBD	IgG1	4	6	VH 1-2	2	VL 2-8
S564-98	NP	IgG3	4	0	VH 4-59	3	VK 1-39
S564-105	NP	IgG1	4	5	VH 4-61	2	VL 2-14
S564-134	Spike/RBD	IgG1	4	2	VH 1-2	6	VL 2-8
S564-138	Spike/RBD	IgG1	4	8	VH 1-2	1	VL 2-14
S564-152	Spike/RBD	IgG1	11	4	VH 3-33	4	VK 1-33
S564-249	NP	IgA1	7	19	VH 3-64	19	VL 2-14
S564-265	Spike/RBD	IgG1	4	4	VH 1-2	3	VL 2-8
S564-275	NP	IgM	4	3	VH 4-59	6	VK 1-39
S116-2822	Spike	IgM	0	0	VH 3-30	0	VK 1-5
S166-32	Spike	IgG1	6	9	VH 3-11	2	VK 1-5
S166-2395	Spike	IgD	4	2	VH 4-4	5	VL 3-21
S166-2620	Spike	IgM	6	1	VH 3-7	1	VL 3-1
S210-852	Spike	IgM	6	8	VH 3-7	2	VL 3-1

Table 3.8, continued.

S210-896	Spike	IgM	9	1	VH 3-30	2	VK 3-21
S210-1262	Spike	IgD	1	0	VH 4-4	0	VK 3-20
S305-399	Spike	IgM	7	1	VH 1-24	4	VK 3-15
S305-1456	Spike	IgG2	7	2	VH 1-24	3	VK 3-15
S376-780	Spike	IgM	3	0	VH 3-30	0	VK 1-27

Table 3.8, continued.

References

1. Cooper, M.D. and M.N. Alder, *The Evolution of Adaptive Immune Systems.* Cell, 2006. **124**(4): p. 815-822.
2. Gayed, P.M., *Toward a modern synthesis of immunity: Charles A. Janeway Jr. and the immunologist's dirty little secret.* The Yale journal of biology and medicine, 2011. **84**(2): p. 131-138.
3. Hong, S., et al., *B Cells Are the Dominant Antigen-Presenting Cells that Activate Naive CD4(+) T Cells upon Immunization with a Virus-Derived Nanoparticle Antigen.* Immunity, 2018. **49**(4): p. 695-708 e4.
4. Popi, A.F., I.M. Longo-Maugéri, and M. Mariano, *An Overview of B-1 Cells as Antigen-Presenting Cells.* Frontiers in Immunology, 2016. **7**(138).
5. Tonegawa, S., *Somatic generation of antibody diversity.* Nature, 1983. **302**(5909): p. 575-581.
6. Martin, F. and J.F. Kearney, *B-cell subsets and the mature preimmune repertoire. Marginal zone and B1 B cells as part of a "natural immune memory".* Immunol Rev, 2000. **175**: p. 70-9.
7. Cyster, J.G. and C.D.C. Allen, *B Cell Responses: Cell Interaction Dynamics and Decisions.* Cell, 2019. **177**(3): p. 524-540.
8. Baumgarth, N., *A Hard(y) Look at B-1 Cell Development and Function.* The Journal of Immunology, 2017. **199**(10): p. 3387-3394.
9. Hardy, R.R. and K. Hayakawa, *B Cell Development Pathways.* Annual Review of Immunology, 2001. **19**(1): p. 595-621.
10. Murphy, K., et al., *Janeway's immunobiology*. 2012, New York: Garland Science.
11. Watson, C.T. and F. Breden, *The immunoglobulin heavy chain locus: genetic variation, missing data, and implications for human disease.* Genes & Immunity, 2012. **13**(5): p. 363-373.
12. D'Angelo, S., et al., *Many Routes to an Antibody Heavy-Chain CDR3: Necessary, Yet Insufficient, for Specific Binding.* Frontiers in Immunology, 2018. **9**(395).
13. Loder, F., et al., *B cell development in the spleen takes place in discrete steps and is determined by the quality of B cell receptor-derived signals.* J Exp Med, 1999. **190**(1): p. 75-89.
14. Rao, S.P., et al., *Biased VH gene usage in early lineage human B cells: evidence for preferential Ig gene rearrangement in the absence of selection.* J Immunol, 1999. **163**(5): p. 2732-40.
15. Singh, S., et al., *Monoclonal Antibodies: A Review.* Curr Clin Pharmacol, 2018. **13**(2): p. 85-99.
16. Sarmay, G., et al., *Mapping and comparison of the interaction sites on the Fc region of IgG responsible for triggering antibody dependent cellular cytotoxicity (ADCC) through different types of human Fcγ receptor.* Molecular Immunology, 1992. **29**(5): p. 633-639.
17. van Erp, E.A., et al., *Fc-Mediated Antibody Effector Functions During Respiratory Syncytial Virus Infection and Disease.* Frontiers in Immunology, 2019. **10**(548).
18. Jennewein, M.F. and G. Alter, *The Immunoregulatory Roles of Antibody Glycosylation.* Trends Immunol, 2017. **38**(5): p. 358-372.

19. Foss, S., et al., *TRIM21—From Intracellular Immunity to Therapy.* Frontiers in Immunology, 2019. **10**(2049).
20. Davis, A.C., K.H. Roux, and M.J. Shulman, *On the structure of polymeric IgM.* Eur J Immunol, 1988. **18**(7): p. 1001-8.
21. Morrison, S.L. and M.E. Koshland, *Characterization of the J chain from polymeric immunoglobulins (IgA-IgM-immunological specificity-primary structure).* Proc Natl Acad Sci U S A, 1972. **69**(1): p. 124-8.
22. Randall, T.D., J.W. Brewer, and R.B. Corley, *Direct evidence that J chain regulates the polymeric structure of IgM in antibody-secreting B cells.* J Biol Chem, 1992. **267**(25): p. 18002-7.
23. Van de Perre, P., *Transfer of antibody via mother's milk.* Vaccine, 2003. **21**(24): p. 3374-6.
24. Garson, J.A., E.A. Quindlen, and P.L. Kornblith, *Complement fixation by IgM and IgG autoantibodies on cultured human glial cells.* J Neurosurg, 1981. **55**(1): p. 19-26.
25. Koenderman, L., *Inside-Out Control of Fc-Receptors.* Frontiers in Immunology, 2019. **10**(544).
26. Palm, A.E. and S. Kleinau, *Marginal zone B cells: From housekeeping function to autoimmunity?* J Autoimmun, 2021. **119**: p. 102627.
27. Schreiber, A.D. and M.M. Frank, *Role of antibody and complement in the immune clearance and destruction of erythrocytes. II. Molecular nature of IgG and IgM complement-fixing sites and effects of their interaction with serum.* J Clin Invest, 1972. **51**(3): p. 583-9.
28. Weill, J.-C., S. Weller, and C.-A. Reynaud, *Human Marginal Zone B Cells.* Annual Review of Immunology, 2009. **27**(1): p. 267-285.
29. Shan, M., et al., *Secreted IgD Amplifies Humoral T Helper 2 Cell Responses by Binding Basophils via Galectin-9 and CD44.* Immunity, 2018. **49**(4): p. 709-724.e8.
30. Flemming, A., *Uncovering the mystery of secreted IgD.* Nature Reviews Immunology, 2018. **18**(11): p. 668-669.
31. Corry, D.B. and F. Kheradmand, *Induction and regulation of the IgE response.* Nature, 1999. **402**(6760): p. 18-23.
32. Kelly, B.T. and M.H. Grayson, *Immunoglobulin E, what is it good for?* Ann Allergy Asthma Immunol, 2016. **116**(3): p. 183-7.
33. Novak, N., S. Kraft, and T. Bieber, *IgE receptors.* Current Opinion in Immunology, 2001. **13**(6): p. 721-726.
34. Oettgen, H.C., *Fifty years later: Emerging functions of IgE antibodies in host defense, immune regulation, and allergic diseases.* The Journal of allergy and clinical immunology, 2016. **137**(6): p. 1631-1645.
35. Karagiannis, S.N., et al., *Activity of human monocytes in IgE antibody-dependent surveillance and killing of ovarian tumor cells.* Eur J Immunol, 2003. **33**(4): p. 1030-40.
36. Woof, J.M. and M.A. Kerr, *The function of immunoglobulin A in immunity.* J Pathol, 2006. **208**(2): p. 270-82.
37. Cerutti, A., et al., *Regulation of mucosal IgA responses: lessons from primary immunodeficiencies.* Ann N Y Acad Sci, 2011. **1238**: p. 132-44.
38. Cerutti, A. and M. Rescigno, *The biology of intestinal immunoglobulin A responses.* Immunity, 2008. **28**(6): p. 740-50.

39. Kerr, M.A., *The structure and function of human IgA.* Biochem J, 1990. **271**(2): p. 285-96.
40. Morton, H.C., M. van Egmond, and J.G. van de Winkel, *Structure and function of human IgA Fc receptors (Fc alpha R).* Crit Rev Immunol, 1996. **16**(4): p. 423-40.
41. Maurer, M.A., et al., *Glycosylation of Human IgA Directly Inhibits Influenza A and Other Sialic-Acid-Binding Viruses.* Cell Rep, 2018. **23**(1): p. 90-99.
42. Vidarsson, G., G. Dekkers, and T. Rispens, *IgG subclasses and allotypes: from structure to effector functions.* Front Immunol, 2014. **5**: p. 520.
43. Ravetch, J.V. and S. Bolland, *IgG Fc Receptors.* Annual Review of Immunology, 2001. **19**(1): p. 275-290.
44. Schwab, I. and F. Nimmerjahn, *Intravenous immunoglobulin therapy: how does IgG modulate the immune system?* Nature Reviews Immunology, 2013. **13**(3): p. 176-189.
45. Irani, V., et al., *Molecular properties of human IgG subclasses and their implications for designing therapeutic monoclonal antibodies against infectious diseases.* Molecular Immunology, 2015. **67**(2, Part A): p. 171-182.
46. Shakib, F. and D.R. Stanworth, *Human IgG subclasses in health and disease(A review) Part II.* Ricerca in clinica e in laboratorio, 1980. **10**(4): p. 561-580.
47. Kaminski, D.A., et al., *Advances in human B cell phenotypic profiling.* Front Immunol, 2012. **3**: p. 302.
48. Frank, M.J., et al., *Expression of sprouty2 inhibits B-cell proliferation and is epigenetically silenced in mouse and human B-cell lymphomas.* Blood, 2009. **113**(11): p. 2478-87.
49. Itoh-Nakadai, A., et al., *The transcription repressors Bach2 and Bach1 promote B cell development by repressing the myeloid program.* Nat Immunol, 2014. **15**(12): p. 1171-80.
50. Moroney, J.B., et al., *Integrative transcriptome and chromatin landscape analysis reveals distinct epigenetic regulations in human memory B cells.* Nat Commun, 2020. **11**(1): p. 5435.
51. Patzelt, T., et al., *Foxp1 controls mature B cell survival and the development of follicular and B-1 B cells.* Proc Natl Acad Sci U S A, 2018. **115**(12): p. 3120-3125.
52. Said, J.W., et al., *TCL1 oncogene expression in B cell subsets from lymphoid hyperplasia and distinct classes of B cell lymphoma.* Lab Invest, 2001. **81**(4): p. 555-64.
53. Uhlen, M., et al., *Proteomics. Tissue-based map of the human proteome.* Science, 2015. **347**(6220): p. 1260419.
54. McHeyzer-Williams, L.J., et al., *Class-switched memory B cells remodel BCRs within secondary germinal centers.* Nat Immunol, 2015. **16**(3): p. 296-305.
55. Zuccarino-Catania, G.V., et al., *CD80 and PD-L2 define functionally distinct memory B cell subsets that are independent of antibody isotype.* Nat Immunol, 2014. **15**(7): p. 631-7.
56. Axelsson, S., et al., *A combination of the activation marker CD86 and the immune checkpoint marker B and T lymphocyte attenuator (BTLA) indicates a putative permissive activation state of B cell subtypes in healthy blood donors independent of age and sex.* BMC Immunol, 2020. **21**(1): p. 14.
57. Kim, C.C., et al., *FCRL5(+) Memory B Cells Exhibit Robust Recall Responses.* Cell Rep, 2019. **27**(5): p. 1446-1460 e4.
58. Muller-Winkler, J., et al., *Critical requirement for BCR, BAFF, and BAFFR in memory B cell survival.* J Exp Med, 2021. **218**(2).

59. Palm, A.E. and C. Henry, *Remembrance of Things Past: Long-Term B Cell Memory After Infection and Vaccination.* Front Immunol, 2019. **10**: p. 1787.
60. Zhao, C., et al., *POU2AF1, an amplification target at 11q23, promotes growth of multiple myeloma cells by directly regulating expression of a B-cell maturation factor, TNFRSF17.* Oncogene, 2008. **27**(1): p. 63-75.
61. Slota, M., et al., *ELISpot for measuring human immune responses to vaccines.* Expert Rev Vaccines, 2011. **10**(3): p. 299-306.
62. Graf, R., et al., *BCR-dependent lineage plasticity in mature B cells.* Science, 2019. **363**(6428): p. 748-753.
63. Brynjolfsson, S.F., et al., *Long-Lived Plasma Cells in Mice and Men.* Frontiers in Immunology, 2018. **9**(2673).
64. Sanz, I., et al., *Challenges and Opportunities for Consistent Classification of Human B Cell and Plasma Cell Populations.* Frontiers in Immunology, 2019. **10**(2458).
65. Lightman, S.M., A. Utley, and K.P. Lee, *Survival of Long-Lived Plasma Cells (LLPC): Piecing Together the Puzzle.* Frontiers in Immunology, 2019. **10**(965).
66. Weisel, F.J., et al., *A Temporal Switch in the Germinal Center Determines Differential Output of Memory B and Plasma Cells.* Immunity, 2016. **44**(1): p. 116-130.
67. Bortnick, A. and D. Allman, *What is and what should always have been: long-lived plasma cells induced by T cell-independent antigens.* J Immunol, 2013. **190**(12): p. 5913-8.
68. Gilbert, S.C., *T-cell-inducing vaccines - what's the future.* Immunology, 2012. **135**(1): p. 19-26.
69. Montecino-Rodriguez, E. and K. Dorshkind, *B-1 B cell development in the fetus and adult.* Immunity, 2012. **36**(1): p. 13-21.
70. Davis, R.S., *FCRL regulation in innate-like B cells.* Ann N Y Acad Sci, 2015. **1362**(1): p. 110-6.
71. Wong, J.B., et al., *B-1a cells acquire their unique characteristics by bypassing the pre-BCR selection stage.* Nature Communications, 2019. **10**(1): p. 4768.
72. Rodriguez-Zhurbenko, N., et al., *Human B-1 Cells and B-1 Cell Antibodies Change With Advancing Age.* Frontiers in Immunology, 2019. **10**(483).
73. Mabbott, N.A. and D. Gray, *Identification of co-expressed gene signatures in mouse B1, marginal zone and B2 B-cell populations.* Immunology, 2014. **141**(1): p. 79-95.
74. Macias-Garcia, A., et al., *Ikaros Is a Negative Regulator of B1 Cell Development and Function.* J Biol Chem, 2016. **291**(17): p. 9073-86.
75. Ochiai, K., et al., *Chromatin Protein PC4 Orchestrates B Cell Differentiation by Collaborating with IKAROS and IRF4.* Cell Rep, 2020. **33**(12): p. 108517.
76. Rothstein, T.L., et al., *Human B-1 cells take the stage.* Ann N Y Acad Sci, 2013. **1285**: p. 97-114.
77. Villa, M., et al., *Aryl hydrocarbon receptor is required for optimal B-cell proliferation.* EMBO J, 2017. **36**(1): p. 116-128.
78. Woods, K.M., et al., *CR2+ marginal zone B cell production of pathogenic natural antibodies is C3 independent.* J Immunol, 2011. **186**(3): p. 1755-62.
79. Hendricks, J., N.A. Bos, and F.G.M. Kroese, *Heterogeneity of Memory Marginal Zone B Cells.* Crit Rev Immunol, 2018. **38**(2): p. 145-158.

80. Descatoire, M., et al., *Identification of a human splenic marginal zone B cell precursor with NOTCH2-dependent differentiation properties.* J Exp Med, 2014. **211**(5): p. 987-1000.
81. Woodruff, M.C., et al., *Extrafollicular B cell responses correlate with neutralizing antibodies and morbidity in COVID-19.* Nature Immunology, 2020. **21**(12): p. 1506-1516.
82. Klein, U. and R. Dalla-Favera, *Germinal centres: role in B-cell physiology and malignancy.* Nature Reviews Immunology, 2008. **8**(1): p. 22-33.
83. Mesin, L., et al., *Restricted Clonality and Limited Germinal Center Reentry Characterize Memory B Cell Reactivation by Boosting.* Cell, 2020. **180**(1): p. 92-106.e11.
84. Victora, G.D. and M.C. Nussenzweig, *Germinal Centers.* Annual Review of Immunology, 2012. **30**(1): p. 429-457.
85. Stebegg, M., et al., *Regulation of the Germinal Center Response.* Frontiers in Immunology, 2018. **9**(2469).
86. Butler, J.E. and N. Wertz, *The porcine antibody repertoire: variations on the textbook theme.* Frontiers in immunology, 2012. **3**: p. 153-153.
87. Di Noia, J.M. and M.S. Neuberger, *Molecular mechanisms of antibody somatic hypermutation.* Annu Rev Biochem, 2007. **76**: p. 1-22.
88. Haakenson, J.K., R. Huang, and V.V. Smider, *Diversity in the Cow Ultralong CDR H3 Antibody Repertoire.* Frontiers in immunology, 2018. **9**: p. 1262-1262.
89. Klein, U., et al., *Transcriptional analysis of the B cell germinal center reaction.* Proceedings of the National Academy of Sciences, 2003. **100**(5): p. 2639-2644.
90. Li, Z., et al., *The generation of antibody diversity through somatic hypermutation and class switch recombination.* Genes Dev, 2004. **18**(1): p. 1-11.
91. Mage, R.G., et al., *Gene-conversion in rabbit B-cell ontogeny and during immune responses in splenic germinal centers.* Vet Immunol Immunopathol, 1999. **72**(1-2): p. 7-15.
92. Opstelten, D., et al., *Germinal centres and the B cell system. IV. Functional characteristics of rabbit appendix germinal centre (-derived) cells.* Virchows Arch B Cell Pathol Incl Mol Pathol, 1980. **34**(1): p. 53-62.
93. Tallmadge, R.L., et al., *Developmental progression of equine immunoglobulin heavy chain variable region diversity.* Developmental and comparative immunology, 2013. **41**(1): p. 33-43.
94. Verma, S. and R. Aitken, *Somatic hypermutation leads to diversification of the heavy chain immunoglobulin repertoire in cattle.* Veterinary immunology and immunopathology, 2012. **145**(1-2): p. 14-22.
95. Zhang, J. and E.I. Shakhnovich, *Optimality of mutation and selection in germinal centers.* PLoS computational biology, 2010. **6**(6): p. e1000800-e1000800.
96. Kranich, J. and N.J. Krautler, *How Follicular Dendritic Cells Shape the B-Cell Antigenome.* Frontiers in Immunology, 2016. **7**(225).
97. Heesters, B.A., et al., *Endocytosis and recycling of immune complexes by follicular dendritic cells enhances B cell antigen binding and activation.* Immunity, 2013. **38**(6): p. 1164-1175.
98. Heesters, B.A., R.C. Myers, and M.C. Carroll, *Follicular dendritic cells: dynamic antigen libraries.* Nature Reviews Immunology, 2014. **14**(7): p. 495-504.

99. Kuraoka, M., et al., *Complex Antigens Drive Permissive Clonal Selection in Germinal Centers.* Immunity, 2016. **44**(3): p. 542-552.
100. Gitlin, A.D., et al., *HUMORAL IMMUNITY. T cell help controls the speed of the cell cycle in germinal center B cells.* Science (New York, N.Y.), 2015. **349**(6248): p. 643-646.
101. Gitlin, A.D., Z. Shulman, and M.C. Nussenzweig, *Clonal selection in the germinal centre by regulated proliferation and hypermutation.* Nature, 2014. **509**(7502): p. 637-640.
102. Nouri, N. and S.H. Kleinstein, *Somatic hypermutation analysis for improved identification of B cell clonal families from next-generation sequencing data.* PLOS Computational Biology, 2020. **16**(6): p. e1007977.
103. Oprea, M. and T.B. Kepler, *Genetic plasticity of V genes under somatic hypermutation: statistical analyses using a new resampling-based methodology.* Genome research, 1999. **9**(12): p. 1294-1304.
104. Pilzecker, B. and H. Jacobs, *Mutating for Good: DNA Damage Responses During Somatic Hypermutation.* Frontiers in Immunology, 2019. **10**(438).
105. Zheng, N.-Y., et al., *Intricate targeting of immunoglobulin somatic hypermutation maximizes the efficiency of affinity maturation.* The Journal of experimental medicine, 2005. **201**(9): p. 1467-1478.
106. Kumar, R., et al., *Biological function of activation-induced cytidine deaminase (AID).* Biomed J, 2014. **37**(5): p. 269-83.
107. Salter, J.D., R.P. Bennett, and H.C. Smith, *The APOBEC Protein Family: United by Structure, Divergent in Function.* Trends Biochem Sci, 2016. **41**(7): p. 578-594.
108. Larijani, M. and A. Martin, *The biochemistry of activation-induced deaminase and its physiological functions.* Seminars in Immunology, 2012. **24**(4): p. 255-263.
109. Bransteitter, R., et al., *Activation-induced cytidine deaminase deaminates deoxycytidine on single-stranded DNA but requires the action of RNase.* Proceedings of the National Academy of Sciences, 2003. **100**(7): p. 4102-4107.
110. Deng, L., et al., *A structural basis for antigen recognition by the T cell-like lymphocytes of sea lamprey.* Proc Natl Acad Sci U S A, 2010. **107**(30): p. 13408-13.
111. Ciccia, A. and S.J. Elledge, *The DNA Damage Response: Making It Safe to Play with Knives.* Molecular Cell, 2010. **40**(2): p. 179-204.
112. Neuberger, M.S. and C. Milstein, *Somatic hypermutation.* Current Opinion in Immunology, 1995. **7**(2): p. 248-254.
113. Noia, J.M.D. and M.S. Neuberger, *Molecular Mechanisms of Antibody Somatic Hypermutation.* Annual Review of Biochemistry, 2007. **76**(1): p. 1-22.
114. Krijger, P.H., et al., *Dependence of nucleotide substitutions on Ung2, Msh2, and PCNA-Ub during somatic hypermutation.* J Exp Med, 2009. **206**(12): p. 2603-11.
115. Rada, C., J.M. Di Noia, and M.S. Neuberger, *Mismatch Recognition and Uracil Excision Provide Complementary Paths to Both Ig Switching and the A/T-Focused Phase of Somatic Mutation.* Molecular Cell, 2004. **16**(2): p. 163-171.
116. Kano, C., F. Hanaoka, and J.-Y. Wang, *Analysis of mice deficient in both REV1 catalytic activity and POLH reveals an unexpected role for POLH in the generation of C to G and G to C transversions during Ig gene hypermutation.* International Immunology, 2012. **24**(3): p. 169-174.
117. Zhang, Y., et al., *Error-prone lesion bypass by human DNA polymerase eta.* Nucleic Acids Res, 2000. **28**(23): p. 4717-24.

118. Krijger, P.H., et al., *Rev1 is essential in generating G to C transversions downstream of the Ung2 pathway but not the Msh2+Ung2 hybrid pathway.* Eur J Immunol, 2013. **43**(10): p. 2765-70.
119. Schrader, C.E., et al., *The roles of APE1, APE2, DNA polymerase β and mismatch repair in creating S region DNA breaks during antibody class switch.* Philosophical Transactions of the Royal Society B: Biological Sciences, 2009. **364**(1517): p. 645-652.
120. Neuberger, M.S., et al., *Monitoring and interpreting the intrinsic features of somatic hypermutation.* Immunological Reviews, 1998. **162**(1): p. 107-116.
121. Rogozin, I.B. and N.A. Kolchanov, *Somatic hypermutagenesis in immunoglobulin genes: II. Influence of neighbouring base sequences on mutagenesis.* Biochimica et Biophysica Acta (BBA) - Gene Structure and Expression, 1992. **1171**(1): p. 11-18.
122. Shapiro, G.S. and L.J. Wysocki, *DNA Target Motifs of Somatic Mutagenesis in Antibody Genes.* 2002. **22**(3): p. 18.
123. Cohen, R.M., S.H. Kleinstein, and Y. Louzoun, *Somatic hypermutation targeting is influenced by location within the immunoglobulin V region.* Molecular immunology, 2011. **48**(12-13): p. 1477-1483.
124. Tang, C., et al., *AID Overlapping and Polη Hotspots Are Key Features of Evolutionary Variation Within the Human Antibody Heavy Chain (IGHV) Genes.* Frontiers in Immunology, 2020. **11**(788).
125. Wei, L., et al., *Overlapping hotspots in CDRs are critical sites for V region diversification.* Proceedings of the National Academy of Sciences, 2015. **112**(7): p. E728-E737.
126. Yaari, G., et al., *Models of Somatic Hypermutation Targeting and Substitution Based on Synonymous Mutations from High-Throughput Immunoglobulin Sequencing Data.* Frontiers in Immunology, 2013. **4**(358).
127. Yuan, C., et al., *The Number of Overlapping AID Hotspots in Germline IGHV Genes Is Inversely Correlated with Mutation Frequency in Chronic Lymphocytic Leukemia.* PLOS ONE, 2017. **12**(1): p. e0167602.
128. Yeap, L.-S., et al., *Sequence-Intrinsic Mechanisms that Target AID Mutational Outcomes on Antibody Genes.* Cell, 2015. **163**(5): p. 1124-1137.
129. Feng, Y., et al., *AID in Antibody Diversification: There and Back Again.* Trends Immunol, 2020. **41**(7): p. 586-600.
130. Malecek, K., et al., *Somatic Hypermutation and Junctional Diversification at Ig Heavy Chain Loci in the Nurse Shark.* The Journal of Immunology, 2005. **175**(12): p. 8105-8115.
131. Lai, M.M., *RNA recombination in animal and plant viruses.* Microbiol Rev, 1992. **56**(1): p. 61-79.
132. Ogando, N.S., et al., *The Curious Case of the Nidovirus Exoribonuclease: Its Role in RNA Synthesis and Replication Fidelity.* Front Microbiol, 2019. **10**: p. 1813.
133. Woo, P.C., et al., *Discovery of seven novel Mammalian and avian coronaviruses in the genus deltacoronavirus supports bat coronaviruses as the gene source of alphacoronavirus and betacoronavirus and avian coronaviruses as the gene source of gammacoronavirus and deltacoronavirus.* J Virol, 2012. **86**(7): p. 3995-4008.
134. Fehr, A.R. and S. Perlman, *Coronaviruses: An Overview of Their Replication and Pathogenesis*, in *Coronaviruses: Methods and Protocols*, H.J. Maier, E. Bickerton, and P. Britton, Editors. 2015, Springer New York: New York, NY. p. 1-23.

135. Cui, J., F. Li, and Z.L. Shi, *Origin and evolution of pathogenic coronaviruses.* Nat Rev Microbiol, 2019. **17**(3): p. 181-192.
136. Adachi, S., et al., *Commentary: Origin and evolution of pathogenic coronaviruses.* Front Immunol, 2020. **11**: p. 811.
137. Huang, C., et al., *Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China.* Lancet, 2020. **395**(10223): p. 497-506.
138. Muller, M.A., et al., *MERS coronavirus neutralizing antibodies in camels, Eastern Africa, 1983-1997.* Emerg Infect Dis, 2014. **20**(12): p. 2093-5.
139. Wu, D., et al., *The SARS-CoV-2 outbreak: What we know.* International Journal of Infectious Diseases, 2020. **94**: p. 44-48.
140. Li, Z., et al., *The human coronavirus HCoV-229E S-protein structure and receptor binding.* Elife, 2019. **8**.
141. Tortorici, M.A., et al., *Structural basis for human coronavirus attachment to sialic acid receptors.* Nat Struct Mol Biol, 2019. **26**(6): p. 481-489.
142. Wang, N., et al., *Structure of MERS-CoV spike receptor-binding domain complexed with human receptor DPP4.* Cell Res, 2013. **23**(8): p. 986-93.
143. Le, T.M., et al., *Expression, post-translational modification and biochemical characterization of proteins encoded by subgenomic mRNA8 of the severe acute respiratory syndrome coronavirus.* FEBS J, 2007. **274**(16): p. 4211-22.
144. Sung, S.C., et al., *The 8ab protein of SARS-CoV is a luminal ER membrane-associated protein and induces the activation of ATF6.* Virology, 2009. **387**(2): p. 402-13.
145. Wong, H.H., et al., *Accessory proteins 8b and 8ab of severe acute respiratory syndrome coronavirus suppress the interferon signaling pathway by mediating ubiquitin-dependent rapid degradation of interferon regulatory factor 3.* Virology, 2018. **515**: p. 165-175.
146. Surjit, M. and S.K. Lal, *The Nucleocapsid Protein of the SARS Coronavirus: Structure, Function and Therapeutic Potential.* Molecular Biology of the SARS-Coronavirus, 2009: p. 129-151.
147. Yuki, K., M. Fujiogi, and S. Koutsogiannaki, *COVID-19 pathophysiology: A review.* Clinical Immunology, 2020. **215**: p. 108427.
148. Han, A.Y., et al., *Anosmia in COVID-19: Mechanisms and Significance.* Chemical Senses, 2020. **45**(6): p. 423-428.
149. Hamming, I., et al., *Tissue distribution of ACE2 protein, the functional receptor for SARS coronavirus. A first step in understanding SARS pathogenesis.* J Pathol, 2004. **203**(2): p. 631-7.
150. Xu, J., et al., *Digestive symptoms of COVID-19 and expression of ACE2 in digestive tract organs.* Cell Death Discovery, 2020. **6**(1): p. 76.
151. AlSamman, M., et al., *Non-respiratory presentations of COVID-19, a clinical review.* Am J Emerg Med, 2020. **38**(11): p. 2444-2454.
152. Lala, A., et al., *Prevalence and Impact of Myocardial Injury in Patients Hospitalized With COVID-19 Infection.* Journal of the American College of Cardiology, 2020. **76**(5): p. 533-546.
153. Ruan, Q., et al., *Clinical predictors of mortality due to COVID-19 based on an analysis of data of 150 patients from Wuhan, China.* Intensive care medicine, 2020. **46**(5): p. 846-848.
154. Blauwet, L.A. and L.T. Cooper, *Myocarditis.* Progress in cardiovascular diseases, 2010. **52**(4): p. 274-288.

155. Long, B., et al., *Cardiovascular complications in COVID-19.* The American journal of emergency medicine, 2020. **38**(7): p. 1504-1507.
156. Khoshdel-Rad, N., et al., *Outbreak of chronic renal failure: will this be a delayed heritage of COVID-19?* Journal of nephrology, 2021. **34**(1): p. 3-5.
157. Perico, L., A. Benigni, and G. Remuzzi, *Should COVID-19 Concern Nephrologists? Why and to What Extent? The Emerging Impasse of Angiotensin Blockade.* Nephron, 2020. **144**(5): p. 213-221.
158. Chu, K.H., et al., *Acute renal impairment in coronavirus-associated severe acute respiratory syndrome.* Kidney Int, 2005. **67**(2): p. 698-705.
159. Jung, J.Y., et al., *Acute kidney injury in critically ill patients with pandemic influenza A pneumonia 2009 in Korea: a multicenter study.* J Crit Care, 2011. **26**(6): p. 577-85.
160. Pei, G., et al., *Renal Involvement and Early Prognosis in Patients with COVID-19 Pneumonia.* J Am Soc Nephrol, 2020. **31**(6): p. 1157-1165.
161. Recalcati, S., *Cutaneous manifestations in COVID-19: a first perspective.* J Eur Acad Dermatol Venereol, 2020. **34**(5): p. e212-e213.
162. Wu, Y., et al., *Nervous system involvement after infection with COVID-19 and other coronaviruses.* Brain Behav Immun, 2020. **87**: p. 18-22.
163. Rubens, J.H., et al., *Acute covid-19 and multisystem inflammatory syndrome in children.* BMJ, 2021. **372**: p. n385.
164. Liu, W., et al., *Two-year prospective study of the humoral immune response of patients with severe acute respiratory syndrome.* J Infect Dis, 2006. **193**(6): p. 792-5.
165. Coleman, C.M., et al., *Purified coronavirus spike protein nanoparticles induce coronavirus neutralizing antibodies in mice.* Vaccine, 2014. **32**(26): p. 3169-3174.
166. Tang, F., et al., *Lack of peripheral memory B cell responses in recovered patients with severe acute respiratory syndrome: a six-year follow-up study.* J Immunol, 2011. **186**(12): p. 7264-8.
167. Payne, D.C., et al., *Persistence of Antibodies against Middle East Respiratory Syndrome Coronavirus.* Emerging infectious diseases, 2016. **22**(10): p. 1824-1826.
168. Chan, K.H., et al., *Cross-reactive antibodies in convalescent SARS patients' sera against the emerging novel human coronavirus EMC (2012) by both immunofluorescent and neutralizing antibody tests.* J Infect, 2013. **67**(2): p. 130-40.
169. Chen, X., et al., *Human monoclonal antibodies block the binding of SARS-CoV-2 spike protein to angiotensin converting enzyme 2 receptor.* Cellular & Molecular Immunology, 2020. **17**(6): p. 647-649.
170. Lan, J., et al., *Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor.* Nature, 2020. **581**(7807): p. 215-220.
171. Robbiani, D.F., et al., *Convergent antibody responses to SARS-CoV-2 in convalescent individuals.* Nature, 2020. **584**(7821): p. 437-442.
172. Wang, C., et al., *A human monoclonal antibody blocking SARS-CoV-2 infection.* Nat Commun, 2020. **11**(1): p. 2251.
173. Yan, R., et al., *Structural basis for the recognition of SARS-CoV-2 by full-length human ACE2.* Science, 2020. **367**(6485): p. 1444-1448.
174. Yi, C., et al., *Key residues of the receptor binding motif in the spike protein of SARS-CoV-2 that interact with ACE2 and neutralizing antibodies.* Cellular & Molecular Immunology, 2020. **17**(6): p. 621-630.

175. Dan, J.M., et al., *Immunological memory to SARS-CoV-2 assessed for up to 8 months after infection.* Science, 2021. **371**(6529): p. eabf4063.
176. Guthmiller, J.J., et al., *SARS-CoV-2 Infection Severity Is Linked to Superior Humoral Immunity against the Spike.* mBio, 2021. **12**(1).
177. Hartley, G.E., et al., *Rapid generation of durable B cell memory to SARS-CoV-2 spike and nucleocapsid proteins in COVID-19 and convalescence.* Science Immunology, 2020. **5**(54): p. eabf8891.
178. Rodda, L.B., et al., *Functional SARS-CoV-2-Specific Immune Memory Persists after Mild COVID-19.* Cell, 2021. **184**(1): p. 169-183 e17.
179. Sosa-Hernandez, V.A., et al., *B Cell Subsets as Severity-Associated Signatures in COVID-19 Patients.* Front Immunol, 2020. **11**: p. 611004.
180. Mina, M.J. and K.G. Andersen, *COVID-19 testing: One size does not fit all.* Science, 2021. **371**(6525): p. 126-127.
181. Ben-Assa, N., et al., *Direct on-the-spot detection of SARS-CoV-2 in patients.* Exp Biol Med (Maywood), 2020. **245**(14): p. 1187-1193.
182. Fozouni, P., et al., *Amplification-free detection of SARS-CoV-2 with CRISPR-Cas13a and mobile phone microscopy.* Cell, 2021. **184**(2): p. 323-333.e9.
183. Wang, D., Z. Li, and Y. Liu, *An overview of the safety, clinical application and antiviral research of the COVID-19 therapeutics.* Journal of Infection and Public Health, 2020. **13**(10): p. 1405-1414.
184. Beigel, J.H., et al., *Remdesivir for the Treatment of Covid-19 — Final Report.* New England Journal of Medicine, 2020. **383**(19): p. 1813-1826.
185. Chen, P., et al., *SARS-CoV-2 Neutralizing Antibody LY-CoV555 in Outpatients with Covid-19.* New England Journal of Medicine, 2020. **384**(3): p. 229-237.
186. Weinreich, D.M., et al., *REGN-COV2, a Neutralizing Antibody Cocktail, in Outpatients with Covid-19.* New England Journal of Medicine, 2020. **384**(3): p. 238-251.
187. Amanat, F. and F. Krammer, *SARS-CoV-2 Vaccines: Status Report.* Immunity, 2020. **52**(4): p. 583-589.
188. Baden, L.R., et al., *Efficacy and Safety of the mRNA-1273 SARS-CoV-2 Vaccine.* New England Journal of Medicine, 2020. **384**(5): p. 403-416.
189. Polack, F.P., et al., *Safety and Efficacy of the BNT162b2 mRNA Covid-19 Vaccine.* New England Journal of Medicine, 2020. **383**(27): p. 2603-2615.
190. Keech, C., et al., *Phase 1–2 Trial of a SARS-CoV-2 Recombinant Spike Protein Nanoparticle Vaccine.* New England Journal of Medicine, 2020. **383**(24): p. 2320-2332.
191. Sadoff, J., et al., *Interim Results of a Phase 1–2a Trial of Ad26.COV2.S Covid-19 Vaccine.* New England Journal of Medicine, 2021.
192. Hwang, B., J.H. Lee, and D. Bang, *Single-cell RNA sequencing technologies and bioinformatics pipelines.* Experimental & Molecular Medicine, 2018. **50**(8): p. 1-14.
193. Haque, A., et al., *A practical guide to single-cell RNA-sequencing for biomedical research and clinical applications.* Genome Medicine, 2017. **9**(1): p. 75.
194. Bentzen, A.K., et al., *Large-scale detection of antigen-specific T cells using peptide-MHC-I multimers labeled with DNA barcodes.* Nature Biotechnology, 2016. **34**(10): p. 1037-1045.
195. Setliff, I., et al., *High-Throughput Mapping of B Cell Receptor Sequences to Antigen Specificity.* Cell, 2019. **179**(7): p. 1636-1646.e15.

196. Stoeckius, M., et al., *Simultaneous epitope and transcriptome measurement in single cells.* Nature Methods, 2017. **14**(9): p. 865-868.
197. Wu, Y. and K. Zhang, *Tools for the analysis of high-dimensional single-cell RNA sequencing data.* Nature Reviews Nephrology, 2020. **16**(7): p. 408-421.
198. Stadlbauer, D., et al., *SARS-CoV-2 Seroconversion in Humans: A Detailed Protocol for a Serological Assay, Antigen Production, and Test Setup.* Current Protocols in Microbiology, 2020. **57**(1): p. e100.
199. Nelson, C.A., C.A. Lee, and D.H. Fremont, *Oxidative refolding from inclusion bodies.* Methods Mol Biol, 2014. **1140**: p. 145-57.
200. Liu, B., et al., *An entropy-based metric for assessing the purity of single cell populations.* Nat Commun, 2020. **11**(1): p. 3155.
201. Tirosh, I., et al., *Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq.* Science, 2016. **352**(6282): p. 189-96.
202. Guthmiller, J.J., et al., *An Efficient Method to Generate Monoclonal Antibodies from Human B Cells.* Methods Mol Biol, 2019. **1904**: p. 109-145.
203. Shlomchik, M.J., et al., *Structure and function of anti-DNA autoantibodies derived from a single autoimmune mouse.* Proc Natl Acad Sci U S A, 1987. **84**(24): p. 9150-4.
204. Vander Heiden, J.A., et al., *pRESTO: a toolkit for processing high-throughput sequencing raw reads of lymphocyte receptor repertoires.* Bioinformatics, 2014. **30**(13): p. 1930-2.
205. Gupta, N.T., et al., *Change-O: a toolkit for analyzing large-scale B cell immunoglobulin repertoire sequencing data.* Bioinformatics, 2015. **31**(20): p. 3356-8.
206. Gaebler, C., et al., *Evolution of antibody immunity to SARS-CoV-2.* Nature, 2021. **591**(7851): p. 639-644.
207. Dugan, H.L., et al., *Preexisting immunity shapes distinct antibody landscapes after influenza virus infection and vaccination in humans.* Sci Transl Med, 2020. **12**(573).
208. Henry, C., et al., *Influenza Virus Vaccination Elicits Poorly Adapted B Cell Responses in Elderly Individuals.* Cell Host Microbe, 2019. **25**(3): p. 357-366 e6.
209. Sokal, A., et al., *Maturation and persistence of the anti-SARS-CoV-2 memory B cell response.* Cell, 2021. **184**(5): p. 1201-1213.e14.
210. Chen, F., et al., *VH1-69 antiviral broadly neutralizing antibodies: genetics, structures, and relevance to rational vaccine design.* Curr Opin Virol, 2019. **34**: p. 149-159.
211. Wec, A.Z., et al., *Broad neutralization of SARS-related viruses by human monoclonal antibodies.* Science, 2020. **369**(6504): p. 731-736.
212. Atyeo, C., et al., *Distinct Early Serological Signatures Track with SARS-CoV-2 Survival.* Immunity, 2020. **53**(3): p. 524-532 e4.
213. Dugan, H.L., C. Henry, and P.C. Wilson, *Aging and influenza vaccine-induced immunity.* Cell Immunol, 2020. **348**: p. 103998.
214. Schmitz, A.J., et al., *A public vaccine-induced human antibody protects against SARS-CoV-2 and emerging variants.* bioRxiv, 2021.
215. Wang, Z., et al., *Enhanced SARS-CoV-2 neutralization by dimeric IgA.* Science Translational Medicine, 2021. **13**(577): p. eabf1555.
216. Barnes, C.O., et al., *SARS-CoV-2 neutralizing antibody structures inform therapeutic strategies.* Nature, 2020. **588**(7839): p. 682-687.
217. Uduman, M., et al., *Detecting selection in immunoglobulin sequences.* Nucleic Acids Res, 2011. **39**(Web Server issue): p. W499-504.

218. Yaari, G., M. Uduman, and S.H. Kleinstein, *Quantifying selection in high-throughput Immunoglobulin sequencing data sets.* Nucleic Acids Res, 2012. **40**(17): p. e134.
219. Laserson, U., et al., *High-resolution antibody dynamics of vaccine-induced immune responses.* Proceedings of the National Academy of Sciences, 2014. **111**(13): p. 4928-4933.
220. Uduman, M., et al., *Integrating B cell lineage information into statistical tests for detecting selection in Ig sequences.* Journal of immunology (Baltimore, Md. : 1950), 2014. **192**(3): p. 867-874.
221. Yaari, G., et al., *The mutation patterns in B-cell immunoglobulin receptors reflect the influence of selection acting at multiple time-scales.* Philosophical transactions of the Royal Society of London. Series B, Biological sciences, 2015. **370**(1676): p. 20140242.
222. Klein, F., et al., *Somatic mutations of the immunoglobulin framework are generally required for broad and potent HIV-1 neutralization.* Cell, 2013. **153**(1): p. 126-138.
223. Ovchinnikov, V., et al., *Role of framework mutations and antibody flexibility in the evolution of broadly neutralizing antibodies.* Elife, 2018. **7**.
224. Stamper, C.T. and P.C. Wilson, *What Are the Primary Limitations in B-Cell Affinity Maturation, and How Much Affinity Maturation Can We Drive with Vaccination? Is Affinity Maturation a Self-Defeating Process for Eliciting Broad Protection?* Cold Spring Harb Perspect Biol, 2018. **10**(5).
225. Zhou, J.O., et al., *The Effects of Framework Mutations at the Variable Domain Interface on Antibody Affinity Maturation in an HIV-1 Broadly Neutralizing Antibody Lineage.* Frontiers in Immunology, 2020. **11**(1529).
226. Ng, K.W., et al., *Preexisting and de novo humoral immunity to SARS-CoV-2 in humans.* Science, 2020. **370**(6522): p. 1339-1343.
227. Grifoni, A., et al., *Targets of T Cell Responses to SARS-CoV-2 Coronavirus in Humans with COVID-19 Disease and Unexposed Individuals.* Cell, 2020. **181**(7): p. 1489-1501 e15.
228. Yao, H., et al., *Molecular Architecture of the SARS-CoV-2 Virus.* Cell, 2020. **183**(3): p. 730-738 e13.
229. Cao, J., et al., *Comprehensive single-cell transcriptional profiling of a multicellular organism.* Science, 2017. **357**(6352): p. 661.
230. Hashimshony, T., et al., *CEL-Seq: single-cell RNA-Seq by multiplexed linear amplification.* Cell Rep, 2012. **2**(3): p. 666-73.
231. Klein, A.M., et al., *Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells.* Cell, 2015. **161**(5): p. 1187-1201.
232. Macosko, E.Z., et al., *Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets.* Cell, 2015. **161**(5): p. 1202-1214.
233. Picelli, S., et al., *Smart-seq2 for sensitive full-length transcriptome profiling in single cells.* Nature Methods, 2013. **10**(11): p. 1096-1098.
234. Ramsköld, D., et al., *Full-length mRNA-Seq from single-cell levels of RNA and individual circulating tumor cells.* Nature Biotechnology, 2012. **30**(8): p. 777-782.
235. Rosenberg, A.B., et al., *Single-cell profiling of the developing mouse brain and spinal cord with split-pool barcoding.* Science, 2018. **360**(6385): p. 176-182.
236. Tang, F., et al., *mRNA-Seq whole-transcriptome analysis of a single cell.* Nature Methods, 2009. **6**(5): p. 377-382.

237. Zheng, G.X.Y., et al., *Massively parallel digital transcriptional profiling of single cells.* Nature Communications, 2017. **8**(1): p. 14049.
238. Stuart, T. and R. Satija, *Integrative single-cell analysis.* Nat Rev Genet, 2019. **20**(5): p. 257-272.
239. Lake, B.B., et al., *Integrative single-cell analysis of transcriptional and epigenetic states in the human adult brain.* Nature Biotechnology, 2018. **36**(1): p. 70-80.
240. Luo, C., et al., *Single-cell methylomes identify neuronal subtypes and regulatory elements in mammalian cortex.* Science, 2017. **357**(6351): p. 600-604.
241. Navin, N., et al., *Tumour evolution inferred by single-cell sequencing.* Nature, 2011. **472**(7341): p. 90-94.
242. Welch, J.D., et al., *Single-Cell Multi-omic Integration Compares and Contrasts Features of Brain Cell Identity.* Cell, 2019. **177**(7): p. 1873-1887.e17.
243. Peterson, V.M., et al., *Multiplexed quantification of proteins and transcripts in single cells.* Nature Biotechnology, 2017. **35**(10): p. 936-939.
244. Butler, A., et al., *Integrating single-cell transcriptomic data across different conditions, technologies, and species.* Nature Biotechnology, 2018. **36**(5): p. 411-420.
245. Stuart, T., et al., *Comprehensive Integration of Single-Cell Data.* Cell, 2019. **177**(7): p. 1888-1902.e21.
246. Wolf, F.A., P. Angerer, and F.J. Theis, *SCANPY: large-scale single-cell gene expression data analysis.* Genome Biology, 2018. **19**(1): p. 15.
247. Argelaguet, R., et al., *MOFA+: a statistical framework for comprehensive integration of multi-modal single-cell data.* Genome Biology, 2020. **21**(1): p. 111.
248. Hao, Y., et al., *Integrated analysis of multimodal single-cell data.* bioRxiv, 2020: p. 2020.10.12.335331.
249. Kim, N., et al., *Single-cell RNA sequencing demonstrates the molecular and cellular reprogramming of metastatic lung adenocarcinoma.* Nature Communications, 2020. **11**(1): p. 2285.
250. Liu, B., et al., *An entropy-based metric for assessing the purity of single cell populations.* Nature Communications, 2020. **11**(1): p. 3155.
251. Blondel, V.D., et al., *Fast unfolding of communities in large networks.* Journal of Statistical Mechanics: Theory and Experiment, 2008. **2008**(10): p. P10008.
252. Dunn, J.C., *A Fuzzy Relative of the ISODATA Process and Its Use in Detecting Compact Well-Separated Clusters.* Journal of Cybernetics, 1973. **3**(3): p. 32-57.
253. Buja, A., et al., *Statistical inference for exploratory data analysis and model diagnostics.* Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 2009. **367**(1906): p. 4361-4383.
254. Maaten, L.V.D., *Accelerating t-SNE using tree-based algorithms.* J. Mach. Learn. Res., 2014. **15**(1): p. 3221–3245.
255. Leland McInnes, J.H., Nathaniel Saul, Lukas Großberger, *UMAP: Uniform Manifold Approximation and Projection.* Journal of Open Source Software, 2018. **3**(29).
256. Haghverdi, L., et al., *Batch effects in single-cell RNA-sequencing data are corrected by matching mutual nearest neighbors.* Nature Biotechnology, 2018. **36**(5): p. 421-427.
257. Mimitou, E.P., et al., *Multiplexed detection of proteins, transcriptomes, clonotypes and CRISPR perturbations in single cells.* Nature Methods, 2019. **16**(5): p. 409-412.
258. French, D.L., R. Laskov, and M.D. Scharff, *The role of somatic hypermutation in the generation of antibody diversity.* Science, 1989. **244**(4909): p. 1152-7.

259. Mayer, C.T., et al., *The microanatomic segregation of selection by apoptosis in the germinal center.* Science, 2017. **358**(6360): p. eaao2602.
260. Stewart, I., et al., *Germinal Center B Cells Replace Their Antigen Receptors in Dark Zones and Fail Light Zone Entry when Immunoglobulin Gene Mutations are Damaging.* Immunity, 2018. **49**(3): p. 477-489.e7.
261. Donisi, P.M., et al., *Pattern and distribution of immunoglobulin VH gene usage in a cohort of B-CLL patients from a Northeastern region of Italy.* Diagn Mol Pathol, 2006. **15**(4): p. 206-15.
262. Migliazza, A., et al., *Frequent somatic hypermutation of the 5' noncoding region of the BCL6 gene in B-cell lymphoma.* Proc Natl Acad Sci U S A, 1995. **92**(26): p. 12520-4.
263. Henry Dunand, C.J., et al., *Preexisting human antibodies neutralize recently emerged H7N9 influenza strains.* The Journal of clinical investigation, 2015. **125**(3): p. 1255-1268.
264. Athey, J., et al., *A new and updated resource for codon usage tables.* BMC Bioinformatics, 2017. **18**(1): p. 391.
265. Plotkin, J.B. and G. Kudla, *Synonymous but not the same: the causes and consequences of codon bias.* Nature Reviews Genetics, 2011. **12**(1): p. 32-42.
266. Inc, G. [Website] [cited 2015; Available from: https://www.genscript.com/tools/rare-codon-analysis.
267. Carvalho, M.B., F. Molina, and L.F. Felicori, *Yvis: antibody high-density alignment visualization and analysis platform with an integrated database.* Nucleic Acids Research, 2019. **47**(W1): p. W490-W495.
268. Ran, F.A., et al., *Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity.* Cell, 2013. **154**(6): p. 1380-9.
269. Erickson, S.A., *A Novel B-1b Precursor Stage in the Perotineal Cavity*, in *Committee on Immunology*. 2020, University of Chicago: The University of Chicago. p. 99.
270. Dong, E., H. Du, and L. Gardner, *An interactive web-based dashboard to track COVID-19 in real time.* Lancet Infect Dis, 2020. **20**(5): p. 533-534.
271. Javed, B., et al., *The coronavirus (COVID-19) pandemic's impact on mental health.* Int J Health Plann Manage, 2020. **35**(5): p. 993-996.
272. Polyakova, M., et al., *Initial economic damage from the COVID-19 pandemic in the United States is more widespread across ages and geographies than initial mortality impacts.* Proceedings of the National Academy of Sciences, 2020. **117**(45): p. 27934-27939.
273. Capoor, M.N., et al., *Is the "Common Cold" Our Greatest Ally in the Battle Against SARS-CoV-2?* Frontiers in Cellular and Infection Microbiology, 2020. **10**: p. 797.
274. Mateus, J., et al., *Selective and cross-reactive SARS-CoV-2 T cell epitopes in unexposed humans.* Science, 2020. **370**(6512): p. 89-94.
275. Kim, C.C., et al., *FCRL5(+) Memory B Cells Exhibit Robust Recall Responses.* Cell reports, 2019. **27**(5): p. 1446-1460.e4.
276. Sullivan, R.T., et al., *FCRL5 Delineates Functionally Impaired Memory B Cells Associated with Plasmodium falciparum Exposure.* PLoS Pathog, 2015. **11**(5): p. e1004894.
277. Li, H. and M. Tolnay, *FCRL4 and FCRL5 expression distinguishes three human memory B cell subsets in tonsils.* The Journal of Immunology, 2017. **198**(1 Supplement): p. 144.11-144.11.

278. Anderson, E.M., et al., *Seasonal human coronavirus antibodies are boosted upon SARS-CoV-2 infection but not associated with protection.* Cell, 2021. **184**(7): p. 1858-1864.e10.
279. Guthmiller, J.J., et al., *A public broadly neutralizing antibody class targets a membrane-proximal anchor epitope of influenza virus hemagglutinin.* bioRxiv, 2021: p. 2021.02.25.432905.
280. Zumla, A.I. and Z.A. Memish, *Middle East respiratory syndrome coronavirus: epidemic potential or a storm in a teacup?* European Respiratory Journal, 2014. **43**(5): p. 1243-1248.
281. Cherry, J.D. and P. Krogstad, *SARS: the first pandemic of the 21st century.* Pediatric research, 2004. **56**(1): p. 1-5.
282. Ramadan, N. and H. Shaib, *Middle East respiratory syndrome coronavirus (MERS-CoV): A review.* Germs, 2019. **9**(1): p. 35-42.
283. King, A., *An uncommon cold.* New scientist (1971), 2020. **246**(3280): p. 32-35.
284. Vijgen, L., et al., *Complete genomic sequence of human coronavirus OC43: molecular clock analysis suggests a relatively recent zoonotic coronavirus transmission event.* Journal of virology, 2005. **79**(3): p. 1595-1604.
285. Koff, W.C. and S.F. Berkley, *A universal coronavirus vaccine.* Science, 2021. **371**(6531): p. 759-759.
286. Roy, B., et al., *Somatic hypermutation in peritoneal B1b cells.* Mol Immunol, 2009. **46**(8-9): p. 1613-9.