THE INTERSECTION OF HOST IMMUNITY AND VIRAL EVOLUTION: SPECIFIC AND NON-SPECIFIC HOST IMMUNE PRESSURES HAVE DIFFERENTIAL INFLUENCES ON INFLUENZA VIRUS AND SARS-COV-2 ANTIGENIC CHANGES

by

Melissa L. Rioux

Submitted in partial fulfilment of the requirements
for the degree of Master of Science

at

Dalhousie University
Halifax, Nova Scotia
December 2021

Dalhousie University is located in Mi'kma'ki,
the ancestral and unceded territory of the Mi'kmaq.
We are all Treaty people.

DEDICATION

I dedicate this thesis to my siblings Meghan and Michael. Thank you for always believing in me! You are both getting copies of my thesis for Christmas this year.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Improvements to public health strategies surrounding vaccination against mutating respiratory pathogens requires a more in-depth understanding of the host factors driving viral evolution. In this thesis, I used informatic pipelines to investigate the impact of specific host immunity on viral mutation across influenza virus proteins and antigenic differences in the SARS-CoV-2 spike proteins across circulating variants. Using a mouse model of previous influenza virus infection and vaccination, I characterized viral mutations present in the lungs of mice at heterologous challenge and found that strain-specific immune responses facilitated the greatest degree of mutation. My *in-silico* analysis of SARS-CoV-2 antigenicity suggested changes in B cell epitopes and conservation of T cell epitopes. I propose that the infection and vaccination history of the host dictates the capacity for viral mutation at infection through elicitation of specific immune mechanisms. Additionally, I present an analysis pipeline that could be leveraged in next-generation vaccine design against respiratory viruses.

LIST OF ABBREVIATIONS USED

HA	hemagglutinin
NA	neuraminidase
M2	matrix protein 2
M1	matrix protein 1
NP	nucleoprotein (influenza virus)
NSP1	non-structural protein 1
NEP	nuclear export protein
PA	polymerase acidic protein
PB1	polymerase basic protein 1
vRNP	viral nucleoprotein
IIV	inactivated influenza vaccine
LAIV	live-attenuated influenza vaccine
CoV	coronavirus
hCoV	human coronavirus
MERS	Middle Eastern respiratory syndrome
SARS	severe acute respiratory syndrome
COVID-19	coronavirus disease of 2019
ORF	open reading frame
nsp	non-structural protein
S	spike protein
E	envelope protein
M	membrane protein
N	nucleocapsid protein (SARS-CoV-2)

IAV	influenza A virus
IBV	influenza B virus
RBD	receptor-binding domain
ACE2	angiotensin-converting enzyme 2
DPP4	dipeptidyl peptidase 4
ER	endoplasmic reticulum
PAMPs	pathogen-associated molecular patterns
PRRs	pattern recognition receptors
TLR	Toll-like receptor
NLR	nucleotide oligomerization domain (NOD)-like receptor
RLR	RIG-I-like receptor
IFN	interferon
PKR	protein kinase R
ZAP	zinc antiviral protein
APC	antigen-presenting cell
TCR	T cell receptor
Ig	immunoglobulin
Tfh	T follicular helper cell
MHC	major histocompatibility complex
HLA	human leukocyte antigen
MDA5	melanoma differentiation-associated gene 5
IL	interleukin
TNF	tumor necrosis factor
ISG	interferon-stimulated gene

H1N1pdm09	2009 pandemic H1N1
NGS	Next Generation sequencing
VOC	variant of concern
VOI	variant of interest
OAS	original antigenic sin
EID_{50}	median embryo infectious dose
$TCID_{50}$	median tissue culture infectious dose
HI	hemagglutinin inhibition
FM/47	A/Fort Monmouth/1/1947
Mex/09	A/Mexico/4108/2009
QIV	quadrivalent influenza vaccine
SNP	single nucleotide polymorphism
PDB	protein database
FPKM	fragments per kilobases mapped
DEG	differentially expressed gene
DAVID	Database for Annotation, Visualization, and Integrated Discovery
GO	gene ontology
GISAID	Global Initiative on Sharing All Influenza Data

ACKNOWLEDGEMENTS

First and foremost, I would like to acknowledge my brilliant lab members Magen Francis, Nicholas Dawe, Anni Ge, Anthony Yourkowski, and Mara McNeil who contributed significantly to this project and taught me many of the skills I needed to succeed. I would like to thank my supervisory committee members Dr. Francesca Di Cara, Dr. Jean Marshall, and Dr. Joanne Langley for their invaluable support and wisdom throughout my studies. I would like to thank my external examiner, Dr. Scott Napper, for his careful reading of this manuscript and for his thoughtful commentary. I would also like to thank my co-supervisor Dr. Chris Richardson for his ongoing guidance, advice, and enthusiasm toward my project. Most of all, I would like to thank my supervisor Dr. Alyson Kelvin for being the best mentor I could ask for. Thank you for sharing your endless knowledge and for always encouraging me. I truly could not have done this without you and appreciate all you have done to make this project the best it can possibly be.

Lastly, I would like to thank my loved ones for always being open to listen and for keeping my spirits high. Thank you to Meghan, Michael, Izzy, Grace, Kat, and Maddie for bringing me back to planet earth when needed!

CHAPTER 1 INTRODUCTION

Pathogens and their hosts are constantly engaged in an evolutionary arms race. The immune response of a host is continually at odds with the goal the pathogen – survival – which pushes the pathogen to adapt. In parallel, host immune mechanisms detect changes on the pathogen and mount novel immune responses, again generating new selective pressures for the pathogen. Such is the case for influenza viruses and coronaviruses, which continue to appear in seasonal and pandemic variations and are constantly adapting to host immunity. This has been a significant public health issue for many years that is yet to be solved. Although our knowledge of viral evolution and effective vaccine formulations has improved tremendously in recent decades, we are still unable to be one step ahead of emerging viruses, which can have devastating consequences on the order of millions of deaths.

Current public health strategies such as seasonal influenza virus surveillance focus on detecting antigenic changes on viral proteins as they appear [1]. Our knowledge of factors that drive a virus to evolve, together with our ability to associate novel mutations that help the virus evade the host immune response, is limited. In the following section, I introduce the knowledge related to how influenza viruses and coronaviruses evolve in the context of the host immune system.

1.1 Introduction to influenza viruses

Despite recurrent vaccination, infections caused by influenza viruses (commonly known as “the flu”) continue to place a heavy burden on public health. Seasonal epidemics caused by influenza virus infect an estimated 5-10% of adults and 20-30% of children annually [2]. Symptoms of infection are highly variable, ranging from mild upper respiratory tract illness to multi-

organ failure and death, and the ultimate outcome depends on a number of host and viral factors [3,4].

Influenza viruses are negative-sense, single-stranded RNA viruses belonging to the *Orthomyxoviridae* family of viruses [5]. Of the four types of influenza viruses (types A through D), [5], only types A and B circulate seasonally in humans, and only type A has historically caused pandemics [6]. Type A influenza viruses are further subtyped by the external and immunologically dominant glycoproteins hemagglutinin (HA) and neuraminidase (NA), of which 18 HA (H1-H18) and 11 NA (N1-N11) molecules have been identified [7]. Type B viruses are not categorized by subtypes but instead are defined by two lineages, B/Victoria and B/Yamagata, each of which is also comprised of antigenically evolving strains [8].

Influenza A viruses are spherical, enveloped viruses covered in membrane proteins HA, NA, and matrix 2 (M2) [5]. The viral membrane or envelope is supported by the matrix 1 (M1) protein, with the segmented RNA genome and nucleocapsids inside [9]. The eight segments of single-stranded, negative-sense RNA encode for at least eleven different viral proteins: HA, NA, M1, M2, nucleoprotein (NP), non-structural protein 1 (NSP1), nuclear export protein (NEP), polymerase acidic protein (PA), polymerase basic protein 1 (PB1), and polymerase basic protein 1-F2 (PB1-F2) [10]. Each segment is contained in a rod-shaped, viral ribonucleoprotein (vRNP) complex that contains viral RNA wrapped around copies of NP, as well as the heterotrimeric viral RNA polymerase made of PA, PB1, and PB2 [11]. A schematic of the influenza A virion is shown in **Figure 1.1**. The general function of the eleven proteins encoded on the eight influenza virus genome segments are summarized in **Table 1.1**.

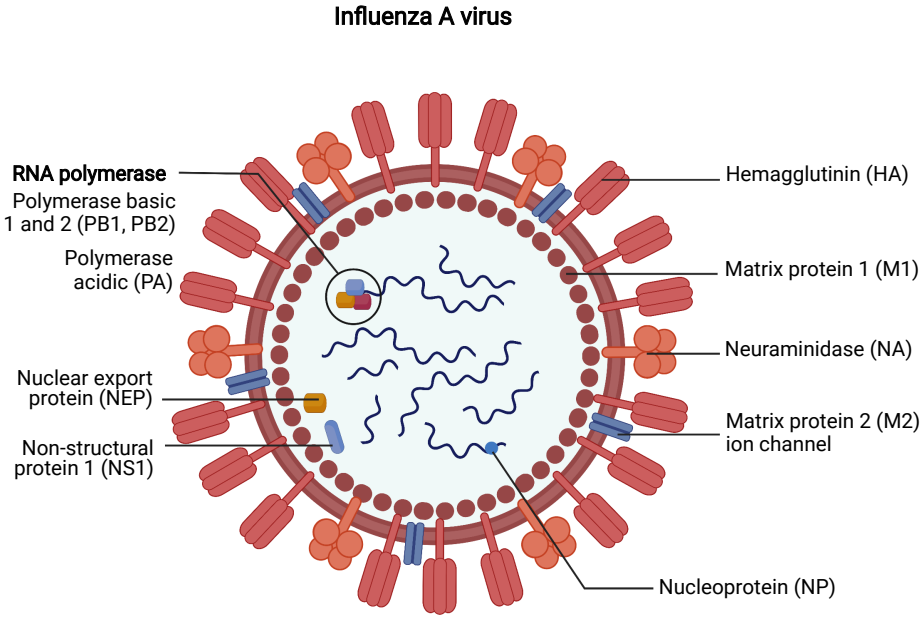


Figure 1.1 Schematic of influenza A virus structure.

Influenza A viruses are spherical, enveloped viruses covered with membrane proteins hemagglutinin (HA) neuraminidase (NA), and the matrix 2 ion channel (M2). The viral envelope is supported by the matrix 1 (M1) protein, and the eight segments of RNA are found inside wrapped around viral nucleoprotein (NP) to form the viral ribonucleoprotein (vRNP) complex. Influenza viral RNA polymerase is a complex comprised of polymerase basic 1 (PB1), polymerase basic 2 (PB2), and polymerase acidic (PA). Non-structural protein 1 (NSP1) and nuclear export protein (NEP) are essential for viral replication and protecting the virus from host immunity.

Table 1.1 Genes encoded on all eight influenza A virus segments and the function of the transcribed proteins [5].

Segment	ORF length (nt)	Protein	Main functions
1	2280	Polymerase basic 2 (PB2)	RNA polymerase component, cap binding, initiates transcription
2	2274	Polymerase basic 1 (PB1)	RNA polymerase component, RNA chain elongation, endonuclease, initiates transcription and replication
		Polymerase basic 1-F2 (PB1-F2)	RNA polymerase component, pro-apoptotic activity
3	2151	Polymerase acidic (PA)	RNA polymerase component, proteolytic activity
4	1701	Hemagglutinin (HA)	Receptor binding and fusion, major antigen
5	1497	Nucleoprotein (NP)	RNA binding, synthesis, and nuclear export
6	1410	Neuraminidase (NA)	Virion release
7	982	Matrix 1 (M1)	Interacts with vRNPs and surface glycoproteins, virion assembly and budding, nuclear export
		Matrix 2 (M2)	Ion channel, virion assembly and budding
8	844	Non-structural 1 (NS1)	Interferon antagonist
		Nuclear export protein (NEP)	Nuclear export of vRNPs

Hemagglutinin and neuraminidase glycoproteins dominate the virion surface and are present at an approximately four to one ratio, respectively [5]. Due to their essential role in the influenza viral infection process and their prevalence on the virion surface, HA and NA and are the main targets for host neutralizing antibodies, and HA is the main component of seasonal influenza virus vaccines [12–14]. Several influenza vaccine platforms are currently in use, including inactivated (IIV), live-attenuated (LAIV), recombinant, and adjuvanted vaccines.

1.2 Introduction to coronaviruses

Viruses of the *Coronaviridae* family are enveloped, positive-sense RNA viruses that are categorized into four lineages referred to as Alpha-, Beta-, Delta- and Gammacoronaviruses [15]. Like influenza viruses, human coronaviruses (hCoVs) circulate seasonally and cause a broad range of respiratory clinical symptoms [16]. In most cases, infection with seasonal coronaviruses such as hCoV-OC43 and hCoV-HKU1 (Betacoronaviruses) or hCoV-NL63 and hCoV-229E (Alphacoronaviruses) are generally mild for most people infected and lead to a disease that resembles the common cold [17]. Other coronaviruses have been shown to be more highly pathogenic: Middle Eastern respiratory syndrome coronavirus (MERS-CoV) and severe acute respiratory syndrome coronavirus (SARS-CoV) of the Betacoronavirus lineage can cause severe lung pathology and death in many cases [18]. More recently, a highly contagious cousin of SARS-CoV, coined SARS-CoV-2, has caused a major pandemic of a magnitude never seen before in human history. As of October of 2021, 242 million cases of coronavirus disease of 2019 (COVID-19) have been reported, and over 4.92 million deaths worldwide have been attributed to the virus [19].

Coronaviruses are single-stranded RNA viruses that, unlike influenza viruses which have a segmented genome, contain a single genomic segment [20]. Specifically, the SARS-CoV-2 genome is approximately 29.6 kilobases in length and encodes six open reading frames (ORFs) [20,21] (**Figure 1.2**). First is ORF1ab, which comprises nearly two-thirds of the entire genome and encodes a large polyprotein (polyprotein 1ab) that is later cleaved into 16 non-structural proteins (nsps), named nsp1 to nsp16 [21] (**Figure 1.2B**). The remainder of the genome encodes several structural proteins: the spike (S) protein, which exists as trimers on the virion surface and is responsible for binding host receptors [22,23]; the envelope (E) protein, which is also present on the virion surface and is required for virion assembly, release and pathogenesis [24,25]; the

membrane (M) protein, which helps shape the virion and bind the nucleocapsid (N) protein [26,27]; and N, which binds and packages viral RNA and is critical to viral replication [28] (**Figure 1.2**).

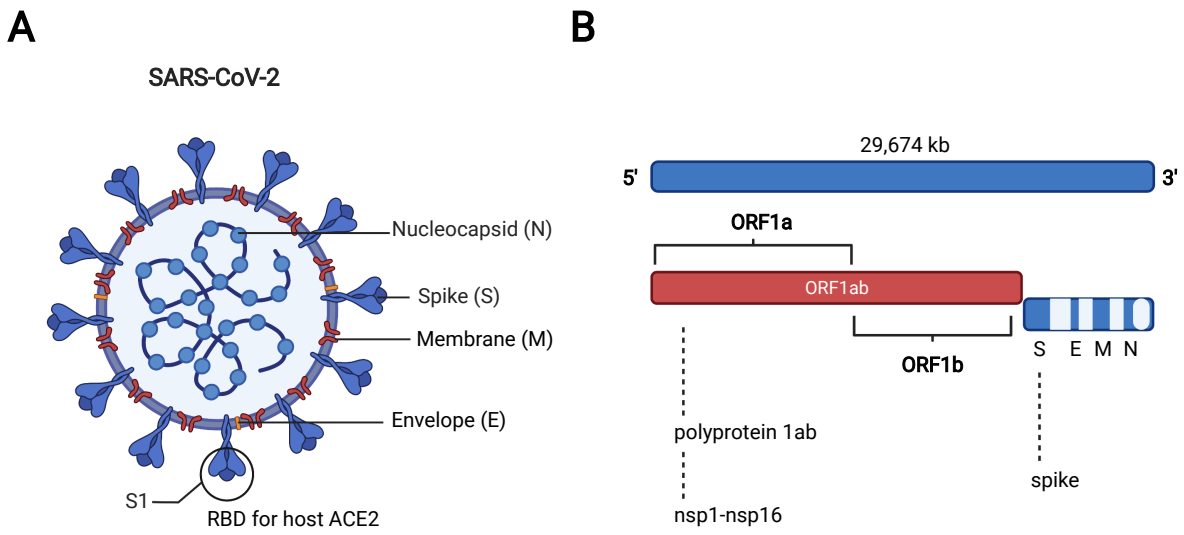


Figure 1.2 The structure of the SARS-CoV-2 virion and genome.

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is an enveloped, positive-sense RNA virus. (A) The virion is comprised of the host cell membrane-derived envelope with the spike (S), membrane (M), and envelope (E) proteins present on the surface. The spike protein is divided into two functionally distinct subunits, referred to as S1 and S2; S1 contains the receptor binding domain, and S2 comprises the transmembrane domain. The RNA genome is organized by binding of the nucleocapsid (N) to viral RNA. (B) The SARS-CoV-2 genome is 29.6 kilobases in length and encodes six open reading frames (ORFs). During viral infection, ORF1ab is transcribed first to yield polyprotein1ab that is later cleaved into 16 non-structural proteins (nsps), named nsp1 to nsp16. The remainder of the genome encodes several structural proteins which are S, E, M, and N. This figure is modified from [20].

Despite the prevalence of seasonal coronavirus infections in humans, vaccination against coronaviruses has not previously been a standard public health practice. However, the emergence of the COVID-19 pandemic stimulated the development and widespread distribution of several coronavirus vaccine platforms. Host infection with SARS-CoV-2 is initiated by a transmembrane, homotrimeric fusion glycoprotein referred to as the spike protein [29].

Vaccination efforts have greatly revolved around the spike protein as an immunogenic target [30,31], since antibody binding to spike effectively blocks viral entry into host cells and suppresses viral infection [32]. Several spike-based vaccine candidates are already in distribution, such as the Pfizer-BioNTech (BNT162b2) and Moderna (mRNA-1273) mRNA vaccines, as well as the adenovirus vector vaccines from Johnson and Johnson (JNJ-78436735/Ad26.COV2.S) and Oxford-AstraZeneca (AZD1222/ChAdOx1) [33,34]. Both vaccine platforms have been shown to elicit B cell and T cell specific immunity [35,36].

1.3 The first host-pathogen interaction: viral infection

The dynamic interplay between host and virus begins with viral infection. When a virus enters a susceptible host, the virus introduces its genetic material into host cells and takes control of the host's replication machinery to generate more virus particles. The specific proteins involved in viral replication, and the ways in which a virus can stimulate or evade the host immune system, are dependent on the type of virus. Both influenza viruses and coronaviruses encode several viral proteins that are necessary for viral replication while also being key players in the host-virus interaction. Thus, the process of viral replication with influenza viruses and coronaviruses, as well as the structure and function of key immunologically dominant proteins, will be described here.

1.3.1 Infection with influenza virus

Infection with influenza A viruses (IAVs) begins when N-acetylneuraminic (sialic) acid residues on host cells are recognized by the virus [5]. Infectious particles are endocytosed when HA spikes on the virion surface attach to the alpha-2,3- and alpha-2,6-glycosidic linkages present on human respiratory epithelial cells, which facilitates viral entry into the cell [5]. Once inside an endocytic vesicle, the low endosomal pH triggers conformational changes on viral HA that expose

the viral fusion peptide (FP) on the HA molecule, prompting the fusion of the viral envelope with the host endosomal membrane to form a pore [37]. In parallel, ions are pumped through the viral M2 ion channel into the virion, which acidifies the virion and releases vRNPs into the host cell [5,37]. Now the vRNPs follow nuclear localization signals to the host cell nucleus, where viral RNA polymerase synthesizes mRNA and complementary RNA from the negative-sense viral RNA [38]. The mRNA is then 5'-capped and 3'-polyadenylated and can be exported and translated in the same manner as host mRNAs [39]. The export of mRNA from the nucleus is regulated by viral matrix-1 (M1) and nuclear export protein (NEP) [39]. Synthesized viral proteins in the cytosol are transported to the host plasma membrane via the secretory pathway [40]. Accumulation of viral HA, NA, and M1 at the host plasma membrane induces a curvature in the membrane that ultimately "buds" off to form the virion once the internal viral proteins and the viral genome have been incorporated [9]. The virion is considered fully infectious once a full genome is acquired. Current findings suggest that IAVs use selective genome packaging to ensure a fully infectious virion [41].

After budding, sialidase activity in the NA protein facilitates release of the virion from the host cell [9]. The released virion moves between cells of the respiratory epithelium, where NA continues to play a role in viral infectivity by cleaving sialic acid residues from mucins to prevent IAV interactions that would limit cell binding [9].

In humans, seasonal influenza viruses lead to pathology primarily limited to the respiratory epithelium, mostly regulated by the host inflammatory response to the virus [4]. Clinical symptoms of seasonal influenza virus infection in humans vary widely from mild symptoms (fatigue, coughing, fever) to lower respiratory tract infection, such as bronchitis or complicated pneumonia, possibly lead to multi-organ failure and death [42]. Physiological failure of the lungs can occur due to airway obstruction, loss of alveolar structure, degradation of the lung extracellular matrix, and epithelial cell death [4].

As HA and NA are the most antigenically relevant influenza virus proteins with respect to influenza virus evolution and vaccine design, I will describe the structure of each glycoprotein in more detail. Hemagglutinin is a homotrimeric protein, and each subunit monomer is made up of two domains referred to as HA1 and HA2 [43]. Importantly, the HA1 domain contains the antigenic sites of HA [44]. The HA1 domain is mostly composed of antiparallel beta-sheets and makes up the globular "head" region of the protein, HA2 is comprised of three alpha helices, one from each monomer, and is described as the "stalk" of the protein [45]. The alpha helices are connected by a flexible loop referred to as Loop B, which spans residues 59 to 76 (H1 numbering) [46]. The HA monomer can be further defined by distinct subdomains: the globular head of HA1 contains the N-terminal F' subdomain (residues 1-41), a vestigial esterase subdomain (residues 42-109 and 263-272), and importantly, the receptor binding domain (RBD) of HA (residues 110-262) [43]. The stalk region of HA, which is highly conserved among influenza A viruses, is composed of the F' (273-330) and F (348-515) subdomains, the fusion peptide (331-347), a transmembrane domain (516-537), and a cytoplasmic domain at the C-terminus of HA2 (538-552) [43]. Due to the role of HA in the infection process, the HA protein is the main antigenic component of current influenza vaccines.

The other external glycoprotein, NA, has enzymatic activity and is mainly important in the final stage of infection where it functions to release the virion from the host cell surface through cleavage of sialic acid residues [47]. Neuraminidase is a homotetramer that resembles a mushroom shape. One influenza virion typically contains 40-50 tetramers [48,49]. The 3D structure of NA is comprised of a highly conserved cytoplasmic tail [50], a hydrophobic transmembrane domain predicted to form an alpha-helix [50,51], a variable "stalk" region, and a "head" domain that has enzymatic activity [22]. The NA head is comprised of anti-parallel beta-sheets that are connected by loops of variable lengths [48]. Although NA can be found in current

influenza virus vaccines, its presence is not regulated, as vaccine potency is based on the HA protein [52].

1.3.2 Infection with coronaviruses

Although the process of infection with coronaviruses varies slightly by taxa, all human coronavirus infections follow a similar sequence of events. Like influenza virus infections, coronavirus infection is initiated when an external viral glycoprotein binds a receptor present on the host cell. All human coronaviruses use the spike glycoprotein to bind host cell receptors, but the specific receptor depends on the specific coronavirus; for example, SARS-CoV-2 spike binds Angiotensin-converting enzyme-2 (ACE2), while MERS -CoV uses dipeptidyl peptidase 4 (DPP4) [53]. The expression of these specific host receptors on different cell and tissue types generally dictates viral tropism. However, this is not perfectly straightforward: although ACE2 is expressed throughout the entire respiratory tract, SARS-CoV-2 mostly replicates in the upper respiratory tract and is easily transmitted [54,55], while SARS-CoV mainly infects the lower respiratory tract and is less transmissible [56,57].

The spike protein of SARS-CoV-2 mainly binds ACE2 on bronchial epithelial cells and type II pneumocytes [58–60]. Spike is divided into two functionally distinct subunits, referred to as S1 and S2 [53]. The S1 subunit is exposed on the virion surface and contains the receptor binding domain, while S2 comprises the transmembrane domain which contains the heptad repeats and fusion peptide [53]. Once inside the host cell, proteases cleave the S protein to allow fusion to the endosome [61], after which the virus is uncoated and its genetic contents enter the host cytoplasm [17]. There the first open reading frame, ORF1ab, is transcribed to yield a polyprotein that is later cleaved into several non-structural proteins, namely nsp1 to nsp16 [17]. The remaining ORFs are then transcribed in double-membrane vesicles associated with the host endoplasmic reticulum

(ER), which help protect viral RNA and sub-genomic RNA from innate immune sensors prior to translation [62]. Finally, the viral proteins including the M, E, N, and S proteins are assembled at the ER-Golgi intermediate compartment, and the virion buds out of the host cell [63,64]. Studies on SARS-CoV have shown that the E, N, and M proteins are all required for the assembly and subsequent release of viral particles [65].

1.4 The host strikes back: immune responses to viral infection

The host recognizes the virus through activation of the immune system. The immune response to viruses is initially characterized by broad, non-specific antiviral mechanisms, followed by the development of antigen-specific adaptive immunity. Host immunity is in large part responsible for the pathology observed during viral infection and dictates the severity of infection. Furthermore, the specificity of the immune response, as discussed later in the context of influenza virus, can apply selective pressure on viral proteins that help drive viral evolution. It is therefore important to understand the immune mechanisms at play throughout a viral infection and how the host interacts with specific viral proteins.

1.4.1 Introduction to innate immunity

When a virus interacts with a host, it first activates the innate immune system. Innate immunity refers to non-specific defense mechanisms that are germline encoded and activated by conserved features of microbial pathogens, called pathogen-associated molecular patterns (PAMPS) [66]. These PAMPs are detected by pattern recognition receptors (PRRs), which initiates pro-inflammatory and antiviral innate immune responses [66,67]. Several types of PRRs exist which recognize different microbial components. In the context of a viral infection, PRRs

such as the Toll-like receptors (TLRs), the NOD-like receptors (NLRs), and the RIG-I-like receptors (RLRs) are capable of binding viral nucleic acids [67].

Activation of PRRs by viral components (such as nucleic acids) stimulates the production of type I interferons (IFN) [67]. Interferons are produced by several cell types including macrophages, dendritic cells, and pneumocytes and are essential to inducing an antiviral state in neighbouring cells [67]. Interferons stimulate the production of hundreds of interferon-stimulated genes (ISGs) such as MX1 [68], protein kinase R (PKR), and zinc antiviral protein (ZAP), which collectively interfere with viral RNA nuclear import, mRNA synthesis, protein synthesis, and cause viral RNA degradation [69]. Along with initiating the interferon pathway, PRR activation generates cytokines and chemokines that serve to recruit innate immune cells to the site of infection. Cells of the innate immune system include granulocytes (neutrophils, basophils, and eosinophils), and the mononuclear cells (monocytes, mast cells, and macrophages). These cells can have phagocytic activity that function to non-specifically ingest material and trigger an innate immune response such as inflammatory response or antiviral response. Granulocytes release granules, which have the capacity to kill pathogens, extracellularly or intracellularly. In response to influenza virus infection, neutrophils, monocytes, and natural killer cells typically migrate to the airway in response to chemokines to clear virally infected or dead cells [67,70]. Phagocytosis of infected cells by alveolar macrophages and neutrophils is an important mechanism of viral clearance [71]. Additionally, phagocytic cells can process viral components and present them on their surface to stimulate T cells, as described in the next section [72].

1.4.2 Introduction to adaptive immunity

In contrast to innate immunity, adaptive immune responses are highly specific to a given pathogen. The components of a pathogen that have the potential to stimulate the immune

response are called antigens, and the region of the antigen that specifically binds receptors on adaptive immune cells are called epitopes [72]. The adaptive immune system is primarily mediated by T and B lymphocytes and antigen presenting cells (APCs). [72]. Lymphocytes recognize antigens on their T cell receptors (TCRs) or immunoglobulins (Igs) for T cells and B cells, respectively. Antigen-presenting cells include cells capable of presenting antigen such as B cells, macrophages, and dendritic cells.

T cells play several roles in innate and adaptive immune activation, including recognizing and killing virally-infected cells, as opposed to simply removing extracellular pathogens [72]. T cells with cytotoxic function are generally characterized as CD8+ T cells, and are present at high proportions in peripheral and non-lymphoid tissues [73]. Another class of T cells, referred to as helper T cells (CD4+), play an essential role in lymphocyte maturation, macrophage activation, and activation of cytotoxic T cells [74]. Depending on specific signals, activated helper T cells differentiate into several sub-classes that interact with the immune system in different ways: naïve helper cells can become Th1 or Th2 effector cells, which mainly help defend against intracellular and extracellular pathogens, respectively [74]; Th17 cells help mediate infections that Th1 and Th2 cells are not well-suited for, such as some extracellular bacteria and fungi [75]; and follicular helper T cells (Tfh) are required for germinal centre reactions [76].

T cells interact closely with APCs, which can present peptide fragments on host major histocompatibility complex (MHC) molecules to engage with the T cell receptor. T cell receptor engagement initiates activation of the T cell leading to clonal expansion, release of specific cytokines, or T cell killing [73–75]. Together, the MHC molecules combined with the viral peptide form the epitope for T cell receptors [72]. There are two classes of MHC molecules, MHC class I and MHC class II. It is important to note that MHC molecules, otherwise known as human leukocyte antigens (HLAs), are highly polymorphic in the human population. For MHC class I, major HLAs are found on three loci: these are HLA-A, HLA-B, and HLA-C, of which thousands of

alleles exist for each [77]. According to the ImMunoGeneTics HLA database, there are 23,002 different possible HLA class I alleles [78]. For MHC class II, HLAs of types DP, DR, DQ, DO, and DM have approximately different 8673 alleles [78]. A single MHC molecule of either class can bind several microbe-derived peptides, leading to great variation in possible TCR epitopes that can be generated from a single MHC; however, peptides can only bind certain MHCs, meaning that certain microbial antigens are therefore HLA-restricted [72,77]. Despite this large variability in HLAs, peptide binding data is typically only available for common and well-documented HLA alleles, which is described as being alleles present at > 0.0001 per 1,500 people [79]. These are only a fraction of the HLA haplotypes; for example, for HLA-A, HLA-B, HLA-C, and DRB1 loci, only 27-30% of these were considered common and well-documented [80]. This HLA restriction along with limited peptide binding data makes it difficult to generalize peptide-based therapeutics across populations or predict T cell epitopes. A solution to this is to identify peptides present on HLA supertypes, which are clusters of several HLA alleles that can bind the same peptide sequence [81–83]. This allows one to reduce the variability in HLA molecules and determine microbial peptides that have maximum population coverage [82].

The source of the peptide determines the class of MHC molecule on which it is presented, and thereby the T cell subset that will interact with it: when viral proteins are phagocytosed, digested, and then presented on the cell surface, they bind MHC class II molecules and stimulate CD4+ T cells; when peptides are synthesized within an infected cell, they are presented on MHC class I molecules and stimulate CD8+ T cells [72]. In the context of viral infection, CD8+ T cells are cytotoxic in nature and their stimulation leads to lysis of the infected cell [72]. CD4+ T cells (T helper cells, as introduced previously) do not directly kill virus but are necessary for optimizing adaptive immune responses, such as antibody production [72].

Antibodies are produced by B lymphocytes, which are defined by their specific Igs [72]. The Igs on the B cell surface serve as functional antigen receptors that recognize specific

epitopes. Tremendous variation exists for B cell receptors, which are analogous to T cell receptors [72]. B cell receptors are comprised of antigen-specific membrane Ig regions, and heterodimers of Ig-alpha and Ig-beta, the intracellular domains that initiate signal transduction and antigen binding [84]. B cell receptors can mutate to bind antigen more specifically through the process of affinity maturation [85,86]. Along with better binding affinity, B cell maturation often involves class or isotype switching, in which the type of immunoglobulins produced by the B cell switch from the IgM or IgD present on naïve B cells to IgA, IgG, or IgE [72]. Different classes of immunoglobulins serve different functions in response to infection, and class-switched, affinity-matured B cells are essential to efficient humoral responses and establishing B cell memory [72]. Because production of IgM is typically associated with immature B cells, and IgG, IgA, and IgE are associated with mature B cells, a secondary exposure to an antigen typically shows early production of the latter antibody isotypes [72].

1.4.3 Immune response to influenza virus infection

The early stages of influenza virus infection are defined by the antiviral response and innate immunity [87]. When influenza viruses enter airway epithelial cells through binding of either a-2,3- or a-2,6-glycosidic linkages [88], intracellular host PRRs are activated. Within the host endosome, TLR3 and TLR7 recognize double-stranded and single-stranded RNA, respectively [89]. In the cytoplasm, RIG-I, melanoma differentiation-associated gene 5 (MDA5), and NOD2 also detect viral RNA [90–93] (**Figure 1.3**). Binding to PRRs stimulates the production of type I and type III IFNs as well as initiate other components of the immune response. For example, NLRs are key components of the inflammasome complex [92], and TLR7 activation helps mediate B cell responses against influenza virus [94]. In the first several hours after infection, airway epithelial cells also strongly upregulate pro-inflammatory cytokines such as interleukin (IL)-6,

CCL5, CXCL10, IL-8, TNF-a, and CCL2 [95,96]. Cytokine production recruits leukocytes, in particular monocytes from the bloodstream, to the airway microenvironment [97]. Additionally, alveolar macrophages primary producers of TNF-alpha during influenza virus infection amplify the expression of the other proinflammatory cytokines [98] (**Figure 1.3**).

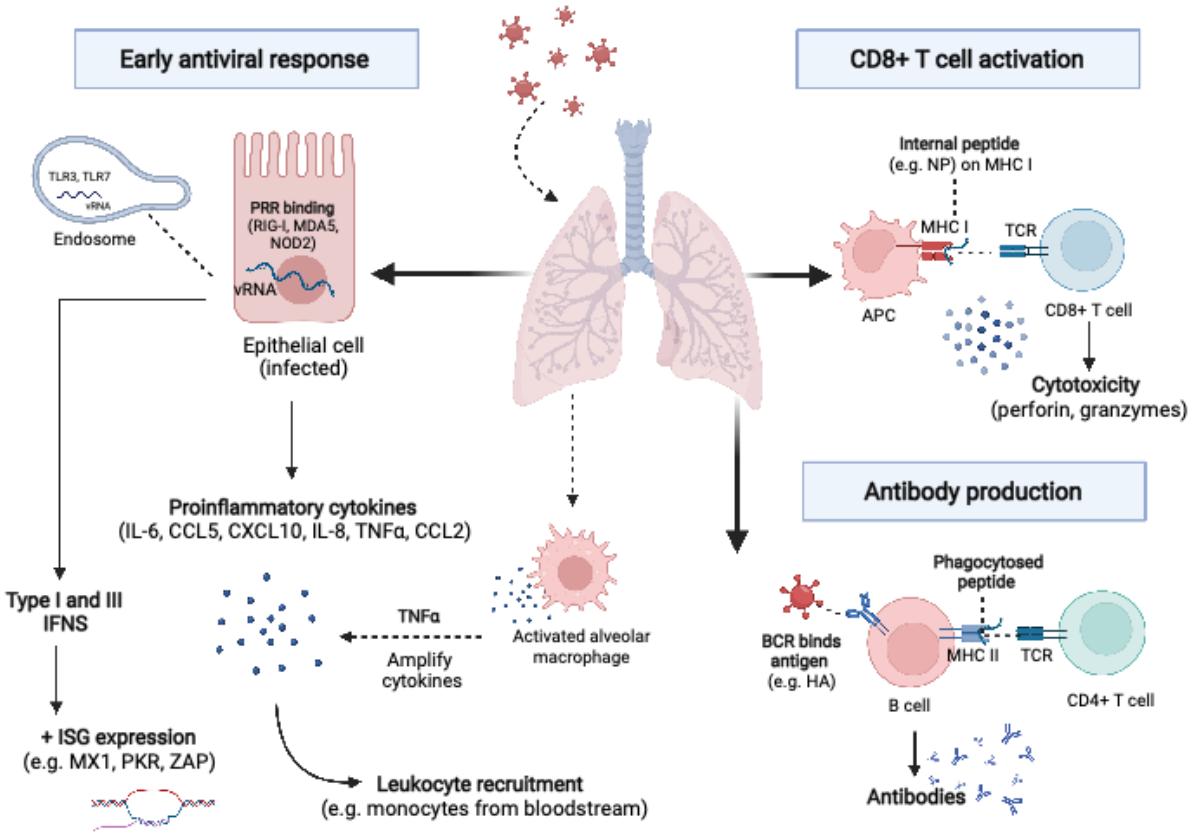


Figure 1.3 Features of the immune response during influenza virus infection.

Influenza viruses are first detected when viral RNA binds host PRRs in the host endosome (TLR3, TLR7) and cytoplasm (RIG-1, MDA5, NOD2). Within a few hours, infected epithelial cells strongly upregulate proinflammatory cytokines IL-6, CCL5, CXCL10, IL-8, TNF-alpha, and CCL2, which serve to recruit leukocytes to the site of infection. In parallel, alveolar macrophages produce TNF-alpha and amplify cytokine production. Type I and II interferons are also produced, and ISGs such as MX1, PKR, and ZAP are upregulated to impede viral replication. To stimulate adaptive immunity, antigen-presenting cells present viral peptides to T cells: internally derived peptides are presented on MHC class I and stimulate CD8+ T cells to perform cytotoxic effector functions, while phagocytosed antigens are presented on MHC class II and stimulate CD4+ T cells. B cell receptors bind viral antigen directly as well as presenting peptides on MHC class II. Signals from CD4+ T cells at antigen binding stimulates antibody production by B cells.

This early cytokine production is important to initiating adaptive immune mechanisms against influenza viruses. For example, type I interferon signalling directly stimulates B cells and CD8+ T cells [99,100]. Other studies have pointed to the NLRP3 inflammasome complex as necessary for optimal IgM, IgG, and IgA responses in the serum and nasal mucosa [101]. Neutralizing IgG antibodies against influenza viruses are often strain-specific and directed at the HA RBD, especially after seasonal vaccination [102]. Internal influenza virus proteins, which have greater similarity among strains, are also important targets of the host adaptive immune response. Immunity elicited toward the nucleocapsid protein of the virus can induce CD8+ and CD4+ T cell responses with the ability to cross-react with antigenically similar as well as distant influenza viruses [103–105].

1.4.4 Immunity at SARS-CoV-2 infection

As described in the context of influenza virus immune responses, the immune response against SARS-CoV-2 is initiated by binding of viral RNA to host PRRs TLR3, TLR7, RIG-I, MDA5, and NLRP3 [17]. This binding typically viral infections elicit the production for type I and type III interferons, downstream interferon stimulated genes, and other proinflammatory cytokines which are key to regulating both the innate and adaptive immune response. For SARS-CoV-2 infection, the early host transcriptional response has been described as “imbalanced”, as production of proinflammatory IL-1, IL-6, and CXCLs are fairly robust, while induction of IFNs and ISGs is delayed [106].

Effective adaptive immune responses against SARS-CoV-2 are important indicators of disease outcomes. During a SARS-CoV-2 infection, CD8+ T cells have shown reactivity to the spike, membrane, and other ORF proteins [107], and these T cell responses are maintained over time [108]. The CD4+ T cell responses against SARS-CoV-2 are correlated with production of

neutralizing antibodies, as affinity-matured B cells require T cell help [76,109]. Neutralizing antibodies against SARS-CoV-2 are primarily directed against the spike RBD [110,111], as has been shown for SARS-CoV [112]. RBD-specific IgA and IgG are consistently produced in response to SARS-CoV-2 infection, while IgM production is inconsistent in acute and convalescent SARS-CoV-2 patients [109,113]. Antibodies (IgA, IgG, and IgM) are also typically produced against the nucleocapsid protein [109]

1.5 Evading the host: viral evolution

The basic principle of evolution– survival of the fittest – applies to viruses as well. Viruses possess a remarkable ability to mutate and evolve over time. When selection pressure is applied, such as antigen-specific immune responses or introduction into a new host, many viruses are capable of adaptation. This is an important part of host-pathogen interactions at the molecular level but also a substantial public health issue.

Influenza viruses are prone to genetic variability through two mutational mechanisms: antigenic shift and antigenic drift [5] (**Figure 1.4**). Antigenic drift is the accumulation and retention of viral mutations during replication, whereas antigenic shift is the swapping of viral genomic segments between different viruses as they infect the same cell at the same time [5]. Antigenic shift can yield novel influenza viruses with pandemic potential and cause increased morbidity in populations that have not previously been exposed [114–116]. Although antigenic drift only yields point mutations, it can cause extensive antigenic diversity and is responsible for recurring seasonal epidemics [117].

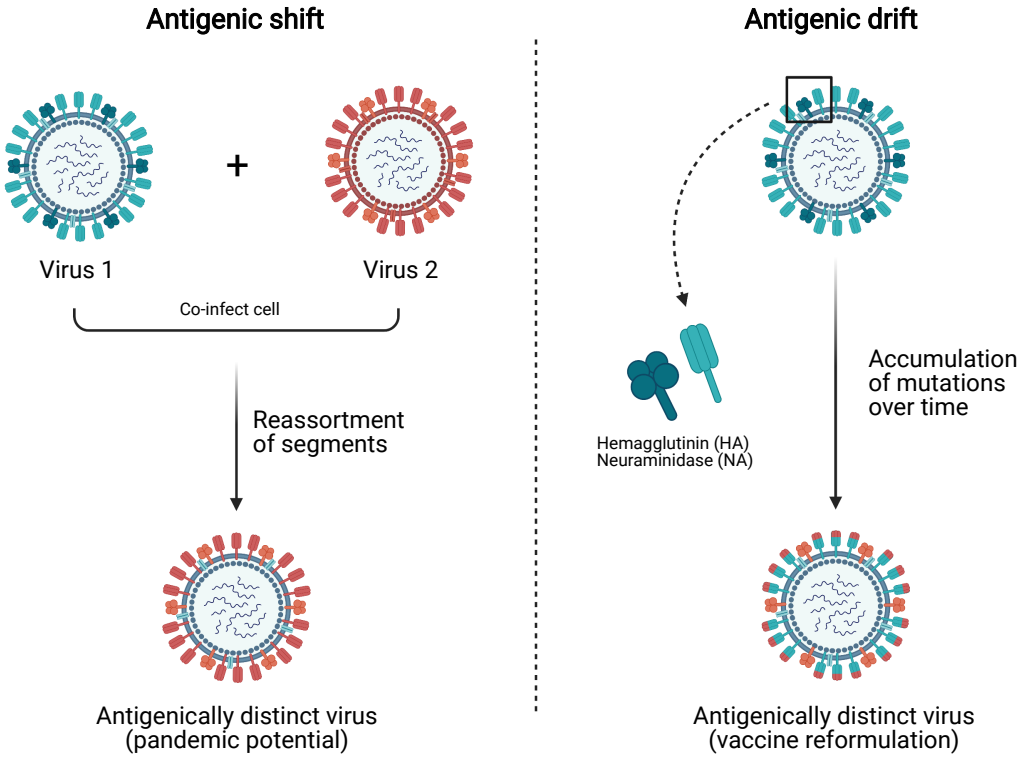


Figure 1.4 Influenza viruses evolve through antigenic shift and drift.

Genetic variability of influenza viruses occurs through two mutational mechanisms, antigenic shift and drift. The segmented genome of influenza viruses facilitates antigenic shift, which occurs when genome segments are swapped when two influenza viruses co-infect a cell. Antigenic shift has been observed in influenza A viruses and can lead to dramatic genomic changes that give the novel influenza virus pandemic potential and cause significant morbidity and mortality. Smaller-scale changes in the genome occur in antigenic drift, which refers to accumulation of point mutations on influenza virus proteins (HA and NA) that are retained and eventually change viral antigenicity. The consequence of antigenic drift is recurring seasonal epidemics and continuous reformulation of influenza virus vaccines.

For the HA protein, point mutations are concentrated across the HA1 segment [118,119]. Studies have shown that positive selection of HA mutations is ongoing and occurs rapidly, often resulting in simultaneous mutations at antigenic sites that quickly increase in frequency [44]. From 1968 to 2010, 85.5% of amino acid substitutions in the major epidemic influenza strains were clustered in the five antigenic sites [120], designated Sa, Sb, Ca1, Ca2, and Cb [121]. The HA gene of H3 tends to evolve more quickly compared to other hemagglutinin subtypes such as H1

strains, as mutations happen more frequently and new variants emerge more rapidly [122,123]. The resulting “drifted” virus can be antigenically distinct to allow reinfection of the host through evasion of previously acquired immunity. Drifted viruses can also elicit cross-protective responses if there are epitope similarities in viral proteins of the first and sequential virus [124]. Importantly, this means that the host and influenza viruses will interact several times over the host’s life along with potential vaccination events, leading to continual refinement of the immune system [125,126].

1.5.1 Measuring viral evolution through sequencing

Where viruses such as influenza type A and B viruses and SARS-CoV-2 are constantly evolving for various reasons, it is exceptionally important to detect changes in viral protein sequence. Nucleic acid sequencing technologies have improved tremendously in the past decade and during pandemics and seasonal epidemics. For example, several HA mutations were identified during the 2009 H1N1 pandemic, some of which were associated with increased fatality. A low-frequency amino acid substitution on the HA receptor binding domain was associated with increased disease severity in H1N1pdm09-infected patients [127,128]. Additionally, sequenced influenza viruses from patient isolates provide insight on the dominant viral strains the following season. The ability to understand and predict mutations across HA is therefore invaluable to designing well-matched vaccine formulations.

The use of viral genome sequencing has allowed scientists to understand the mechanisms driving mutations in the HA gene as well as the rest of the genome. These mutations may give insight into which viruses evade pre-existing immunity from previous infection of seasonal vaccination. Next-generation sequencing (NGS) has been employed to identify host immune factors that drive mutation of HA. For example, one study of sequenced H3N2 from infected patient isolates found significantly increased amino acid changes across HA1 in influenza

vaccinated patients compared to infected-unvaccinated patients [129]. Another study by Parisi and colleagues yielded a model to demonstrate how heterogeneity in HA-directed antibodies drives specific mutations in influenza A viruses following virus infection [130].

1.6 The consequences of viral evolution: variants of SARS-CoV-2

In January 2020, sequencing led to the identification of the SARS-CoV-2 virus within ten days of the novel virus being reported [131]. The importance of viral sequencing as an investigative and predictive tool has been shown clearly throughout the ongoing SARS-CoV-2 pandemic. Early in the pandemic, SARS-CoV-2 spike containing an aspartic acid to glycine switch at position 614 (D614G) in the receptor-binding domain (RBD) rapidly became the dominant circulating virus [132]. This substitution has since been shown to stabilize the protein trimer, thereby increasing viral infectivity and fitness [133,134]. Although vaccine research readily adapted to the mutation and incorporated D614G spike into vaccine formulations, the virus has continued to change. In the past year, several variants of SARS-CoV-2 have emerged that have raised concerns about the effectiveness of the developed SARS-CoV-2 vaccine platforms that are currently in use [135]. The World Health Organization currently designates a SARS-CoV-2 variant a “variant of concern” (VOC) when the variant demonstrates one of the following traits: increased virus transmission; increase in virulence; change in clinical disease presentation; or decrease in effectiveness of countermeasures (13). This is in contrast to a “variant of interest” (VOI), which is defined as a variant causing significant community transmission or appearing in multiple clusters, in addition to containing mutations that may affect transmissibility and disease severity, or effectiveness of therapeutics and diagnostics [136]. Some variants are also under surveillance, which means that they may pose a future risk, but phenotypic evidence of increased risk is lacking [136].

The first VOC identified was lineage B.1.1.7 reported in the UK in December of 2020 [137]. Now designated Alpha, the variant was modeled to have an increased reproduction number compared to original SARS-CoV-2 Wuhan virus, which accounts for an estimated 30%-90% increase in transmissibility [138,139]. Increased rates of hospitalization and mortality have also been reported for infections caused by the Alpha variant, although supporting evidence is limited [140,141]. Notably, Alpha contains more lineage-defining mutations than are typically seen [142]. The variant contains 17 mutations (14 non-synonymous point mutations and three deletions): eight mutations are contained within the spike protein and include D614G, a N501Y mutation in the receptor-binding domain of spike, a deletion at position 69-70, and a P681H mutation near the S1/S2 cleavage site [143]. Current evidence suggests that Alpha is still susceptible to neutralization by antibodies elicited by natural infection and ancestral spike vaccines [144–147], although some RBD-specific monoclonal antibody titers may be reduced [144,145].

A second VOC originally designated lineage B.1.351 and now referred to as the Beta variant was detected in South Africa in October of 2020 [148,149]. Similar to Alpha, Beta has shown a marked increase in transmissibility compared to original SARS-CoV-2, with current estimates around 25% [139]. Beta has eight mutations in the spike protein, including the lineage-defining mutations K417N, E484K, and N501Y [148,149]. In particular, this observation represents the first instance of the E484K mutation in a variant of concern, which is thought to impact the binding of neutralizing antibodies [150]. Beta has shown reduced susceptibility to antibodies elicited after natural infection [151], vaccination [147], and treatment with monoclonal antibodies bamlanivimab, estesevimab, and imdevimab [152]. However, individuals vaccinated with COVID-19 vaccines developed against the Wuhan virus appear to remain protected against severe disease [151].

Another significant VOC, referred to as Gamma or previously as lineage P.1, was identified in Brazil in late 2020 [153]. Sequencing analysis found 17 amino acid changes, three

deletions, and four synonymous mutations in the spike protein, along with a four-nucleotide insertion. The Gamma lineage shares several of the same spike mutations present in the Beta lineage, including RBD substitutions K417N, E484K, and N501Y, suggesting convergent evolution of these separate lineages [153]. Gamma is observed to be more transmissible than Alpha and Beta, and 38% more transmissible than the original Wuhan SARS-CoV-2 [139]. Some reports have also suggested increased rates of hospitalization and mortality in healthy individuals [154].

October of 2020 also saw the emergence of the B.1.617 lineage in India, which is comprised of three sub-lineages designated B.1.617.1 (Kappa), B.1.617.2 (Delta), and B.1.617.3 [155]. Preliminary data speculated that some B.1.617 sub-lineages had an increased transmission rate [139], reduced effectiveness of treatment with the monoclonal antibody bamlanivimab, and reduced neutralization from convalescent and Pfizer-BioNTech (BNT612b) vaccine sera [148,152]. A P681R mutation near the furin cleavage site of the spike protein in all B.1.617 lineages is suggested to increase the infectivity of the variant, resulting in enhanced transmissibility [156]. The other two characteristic spike substitutions L452R and E484K, found only in variants of interest B.1.617.1 (Kappa) and B.1.617.3, are hypothesized to reduce neutralization by monoclonal antibodies [155,156]. Of the B.1.617 sub-lineages, the Delta variant has raised concerns in recent months due to a significantly increased transmission rate (97%)[139], increased hospitalization and death rates [157], and an increased risk of reinfection [158]. The Delta variant has 15 lineage-defining mutations, including the spike mutations G142D, L452R, P681R, and T478K (the latter which is not found in the other B.1.617 sub-lineages) [159].

Other variants of interest include the Epsilon or B.1.429 lineage identified in California [160] and the mink-related (Cluster 5) lineage detected in Denmark in June of 2020, which also have several non-synonymous mutations compared to the original SARS-CoV-2 reference sequence [161]. The mink-related variant has been suggested as a potential spillover threat with

minks acting as viral reservoir, as mink farmers have frequent interaction with the animals. However, at this time, spread of the mink-related variant to humans has been rare [162].

The mutations identified so far across SARS-CoV-2 variants are predicted to affect several key functions of the virus, such as causing increased viral binding to the host receptor ACE2 or entry augmentry factors [163] thereby increasing viral infectivity [133], or by impacting viral susceptibility to host immune responses such as neutralizing antibodies [132]. At this time, it seems that current vaccines are still protective from severe disease in SARS-CoV-2 variants of concern. However, it is unclear whether future variants will be inhibited by previously acquried immunity. Furthermore, as vaccine effectiveness is dependent on the recognition of viral proteins by antigen-educated T and B cells, mutations influencing T and B cell epitopes in new SARS-CoV-2 variants may have significant effects on the performance of vaccines that were originally developed toward original SARS-CoV-2. The ability to rapidly detect SARS-CoV-2 mutations by viral sequencing during the pandemic has been essential to determining mutation-associated changes in transmissability, pathogenicity, and antigenicity.

1.7 Maintaining immunity against evolving viruses

Evolution of viruses in animal reservoirs has been responsible for the emergence of novel human viruses with pandemic potential. However, as demonstrated throughout the SARS-CoV-2 pandemic, the continual evolution of viruses already circulating in the human population poses several challenges. In a single “season”, mutation of immunodominant viral proteins can lead to changes in viral transmissibility and pathogenicity, as well as decreased effectiveness of vaccines. As with influenza viruses, continual antigenic changes on vaccine antigens necessitate new vaccine formulations each flu season, which is further complicated by the infection and vaccination history of the host.

1.7.1 Seasonal influenza virus vaccination

A recurring consequence of antigenic drift is the need to reformulate influenza virus vaccines. Current seasonal influenza vaccines contain concentrated HA protein, and even small changes at these antigenic sites can lead to evasion of the hosts' previously acquired neutralizing antibodies, allowing reinfection of a previously-immune host [124]. Influenza vaccine research continues to focus on development of universal vaccines that will provide broad protection across influenza viruses. However, the World Health Organization currently selects IAV and IBV strains to be included in seasonal formulations, basing their prediction on strains circulating in the previous flu season [164].

Several influenza vaccine platforms are currently in use, including inactivated, live-attenuated, recombinant, and adjuvanted vaccines. Inactivated vaccines are administered intramuscularly and are available in trivalent and quadrivalent formulations [12]. Trivalent inactivated influenza vaccines are composed of the two IAV strains currently circulating in humans (H1N1 and H3N2), and only one IBV strain (either Yamagata or Victoria lineage), while quadrivalent formulations contain two IAV strains and two IBV strains [165]. Quadrivalent inactivated vaccines have become increasingly popular to reduce mismatch between IBV strains in circulation and that which the population is vaccinated with, ultimately conferring better protection from influenza virus infection [166].

Annual vaccination is the best way to prevent morbidity and mortality from infection with emerging influenza strains and has been recommended by the Centers for Disease Control and Prevention since 2010 [167,168]. However, no vaccine platform offers complete coverage against influenza virus because of the virus' ability to rapidly mutate. As described in the next section, the immune response at vaccination is dependent on the specific immune memory of the host.

1.7.2 The role of previously acquired infections and vaccinations

The dynamics of repeated infections with antigenically divergent pathogens are complex and incompletely understood. It is known that immunity elicited from previous influenza virus infection or vaccination, or *preimmunity*, can be beneficial or deleterious for protection from disease during a subsequent virus infection or vaccination event [117]. Studies have shown that the immune response elicited at vaccination depends on the specific antibody profile of the host [169]. If an individual is vaccinated with antigens similar to antibodies already in circulation, the result is typically a boost of these existing antibodies [170]. It seems that the induction of novel versus memory immune responses to vaccination is dependent on the similarity of vaccine antigens to what is already present in the complex adult immune system [171]. The immune memory of the host also impacts the broadness and magnitude of HA-specific antibodies generated at vaccination. For example, a study by Andrews and colleagues showed that individuals with low pre-existing serological titers to the vaccination strain generated broadly reactive antibodies against the HA stalk, while those with high pre-existing titers had a strain-specific, HA head-dominated response [169]. Cross-reactivity studies of sequential influenza virus infection in mice have shown heterosubtypic cross-protection through influenza-specific non-neutralizing IgG and CD8+ T cells when previously infected or vaccinated [172]. These mechanisms are especially important for cross-reactivity against influenza virus proteins apart from HA. Influenza virus nucleoprotein is shown to constitute a large portion of the cytotoxic T lymphocyte response at infection, and is therefore a primary target for cross-reactive CD8+ T cells at repeat infection [173,174].

Preimmunity is further complicated by the concept of immune imprinting. When building an influenza virus immune history, not all infections are created equally. The significance of one's first influenza virus exposure in early life on subsequent immune responses to infection and

vaccination has been described for decades. The term *Original Antigenic Sin* (*OAS*) was coined in the 1950s by Thomas Francis to describe the dominating effect of the first IAV infection on immune responses later in life [175]. Specifically, OAS refers to the preference of the immune system to recall pre-existing antibodies, rather than elicit a new response against a novel antigen during infection [12]. The *sin* in OAS highlights the possible deleterious effect of the first infection on subsequent immune responses, but OAS is now frequently referred to as *antigenic seniority* to capture possible protective effects as well [176].

Imprinting has been demonstrated repeatedly in the real world as age cohorts are differentially impacted by emerging influenza viruses. For example, elders were least affected by the 2009 H1N1 pandemic due to imprinting with similar viruses in childhood [169,177]. In a study of pre-existing antibodies in human patients during the 2009 H1N1 pandemic, 34% of persons born before 1950 had relatively high antibody titers cross-reactive to the novel strain, compared to only 4% in those born after 1980 [178]. It is possible that we are thus able to predict which age population will be most affected by emerging viruses based on imprinting year [179]. Animal studies have shown that despite antibodies elicited during heterologous infection being primarily directed towards the original imprinting strain and not the challenge strain, these antibodies are still able to induce protection through non-neutralizing cross-reactions [180].

1.8 Thesis aim: the intersection of host-virus interactions and viral evolution

Humans have diverse immunological backgrounds composed of immunity acquired from previous infections and vaccinations, in varied combinations. This factor is highly relevant to emerging and re-emerging viruses, such as influenza viruses and coronaviruses. The capacity of the host to be reinfected with similar viruses through viral mutation or waning of short-lived adaptive immune responses, as well as recurrent vaccination (as seen in seasonal influenza virus

vaccines), means that the host and virus will have several interactions throughout a lifetime. This creates of complex web of several host and viral factors that intercept at a viral infection or vaccination event. The unique immune backgrounds present in the host contribute to the specific immune response elicited at an infection and vaccination, which in turn influences viral replication and selection of mutations. As the virus evolves, the immunodominant proteins can be recognized as antigenically different by the host adaptive immunity, leaving the host vulnerable to reinfection.

In this thesis, I aimed to investigate the bridge between the host immune background and the viral capacity for evolution. Immune backgrounds were developed in a C57Bl/6j mouse model by non-lethal infection with A/FortMonmouth/1/1947 H1N1 virus followed by vaccination with the 2018-2019 split virion quadrivalent vaccine, resulting in four experimental groups: naïve-mock vaccinated, naïve-vaccinated, preimmune mock-vaccinated, and preimmune-vaccinated mice. All animals were then challenged with the antigenically divergent A/Mexico/4108/2009 H1N1 2009 pandemic virus. Here I hypothesized that the polarizing immune responses elicited in each immune background would drive differential mutation of viral proteins. Through viral sequencing at infection followed by bioinformatic analysis of viral protein antigenicity, I show that differential vaccination and infection history in the mouse model drives different patterns in mutation across viral hemagglutinin, neuraminidase, nucleoprotein, and polymerase basic 1. Moreover, my host transcriptome analysis demonstrates early adaptive immune responses dominated by T cell-mediated pathways in preimmune animals, compared to non-specific antiviral responses in naïve animals. This study strengthens the important link between host infection and vaccination history and immune pressures that drive antigenic drift.

To further our understanding of mutation-associated antigenic changes, I applied my bioinformatic analysis pipeline to the circulating SARS-CoV-2 variants that have recently arisen. Using the published sequences for Wuhan SARS-CoV-2 and eight variants (Alpha (B.1.1.7), Beta (B.1.351), Gamma (P.1), Delta (B.1.617.2), Kappa (B.1.617.1), Epsilon (B.1.429), B.1.617.3, and

mink-related), I investigated how mutations across the spike protein may lead to changes in antigenicity with respect to B cell epitopes, CD4+ and CD8+ T cell epitopes, and spike glycosylation patterns. Here I show differences in predicted B cell epitopes in the S1 subunit across spike variants, but relative conservation of T cell epitopes. Moreover, I identified a different glycosylation pattern on the Gamma variant that may contribute to evasion from immunity established through natural infection or vaccination.

My study of immune background-driven selection of mutation of influenza virus in a mouse model, as well as my bioinformatic pipeline investigating mutation-associated changes in antigenicity on SARS-CoV-2, are a significant contribution to our understanding of viral antigenic changes and how changes may be recognized by the adaptive immune system. This study proposes a mechanism of influenza virus mutation selection and analysis pipeline that could be leveraged in next generation vaccine design for influenza viruses, but also for other respiratory viruses such as SARS-CoV-2.

CHAPTER 2 MATERIALS AND METHODS

2.1 Ethics statement

All animal work was completed in accordance with the Canadian Council of Animal Care guidelines. Animal use protocols were approved by the Animal Care Committee for the Dalhousie University Carleton Animal Care Facility (Halifax, NS, Canada; animal ethics number 18-091). Animal vaccinations and viral inoculations were performed under short-term 3% isoflurane anesthesia to minimize distress. Animals removed for sample collection or humane endpoint reasons were euthanized under 5% isoflurane.

2.2 Experimental animals

Female C57BL/6J mice (5-8 weeks old) were obtained from Jackson Laboratories (Bar Harbor, MN, USA) and used in the influenza sequential infection, vaccination, and challenge study (described in Chapter 3). All mice were housed in HEPA-filtered cage racks adherent to ABSL2 guidelines (Dalhousie University Carleton Animal Care Facility, Halifax, NS, Canada) in clean environmental conditions with controlled temperature and humidity and a 12/12 light/dark cycle. The mice were maintained on standard animal feed and water *ad libitum*. All animal procedures were performed in a certified class II biological safety cabinet. Following viral infection, vaccination, and challenge, mice were monitored daily for weight, survival, and other clinical signs of illness. In accordance with the Animal Care Committee, mice were humanely euthanized if 80% of baseline weight was reached.

2.3 Viruses and experimental timeline

All virus work was conducted in a BSL2+ facility. For use in preimmune inoculations, historical mouse adapted H1N1 strain A/Fort Monmouth/1/1947 was obtained from the American Type Culture Collection (ATCC) through Cedarlane (Burlington, ON, Canada). The challenge H1N1 strain A/Mexico/4108/2009 was also obtained through ATCC. Prior to use, viral median tissue culture infectious dose ($TCID_{50}$) and median embryo infectious dose (EID_{50}) were determined through titration in embryonated chicken eggs or Madin-Darby Canine Kidney cells. For animal vaccination and boost, the Sanofi FLUZONE quadrivalent influenza vaccine (QIV) for the 2018-2019 season was used (Sanofi Canada; North York, ON, Canada). The seasonal vaccine contained concentrated HA proteins from H1N1 (A/Michigan/45/2015 X-275), H3N2 (A/Brisbane/1/2018 X-311), influenza B Yamagata lineage (B/Phuket/3073/2013), and the influenza B Victoria lineage (B/Colorado/6/2017-like virus).

Preimmune inoculation with A/Fort Monmouth/1/1947 was designated day 0 of the study. Mice were inoculated intranasally with $10^{3.5}$ EID_{50} historical mouse adapted H1N1 (50 uL total). On days 60 and 74, mice were vaccinated and boosted with 50 uL of the human dose of Sanofi QIV in the hind caudal muscle. On day 105, mice were challenged intranasally with 10^{6} EID_{50} A/Mexico/4108/2009.

2.4 Antibody and viral load assessment

Prior to 2009 pandemic H1N1 challenge, hemagglutinin inhibition (HAI) titers in mouse serum were determined by HAI assay against preimmune (A/Fort Monmouth/1/1947) and vaccine (A/Mexico/4108/2009) strains. Mouse antisera were treated for a minimum of 4 hours with Receptor Destroying Enzyme at 37°C in a water bath. The sera were then serially diluted with PBS in a 96-well V-bottom plate with 8 hemagglutinin (HA) units/50 uL of antigen, then incubated for 15 minutes at room temperature. Turkey red blood cells were washed and diluted to 0.05%

(vol/vol) in PBS and 50 uL were added to each well. The plate was incubated at room temperature for 30 minutes in the dark and assessed for agglutination. The HAI titer was calculated as log base 2 of the highest serum dilution factor required to prevent agglutination.

At 3 days post-challenge, live viral titer in the lungs was measured by hemagglutination (HA) assay. The lungs were flash frozen in liquid nitrogen at collection and homogenized for use in $TCID_{50}$ assay prior to the HA assay. The viral load was calculated according to the Reed and Muench method [181].

2.5 RNA extraction and virus whole-genome sequencing

Total RNA was extracted from the lungs of influenza virus-infected mice 3 days post-infection with A/Mexico/4108/2009 using the RNeasy Mini Kit (QIAGEN, Cat. No 74004). According to manufacturer's instructions, 30 mg frozen mouse lungs were lysed and homogenized using a high-salt buffer. Ethanol was added to the lysate to separate the cell contents into organic and aqueous phases. Total RNA was bound to a silica-based membrane using the RNeasy spin column and eluted in RNase-free water. For stock A/Mexico/4108/2009, RNA was extracted three separate times from the stock aliquot using an RNA-binding spin column as described above using the QIAamp Viral RNA Mini Kit (QIAGEN, Cat. No. 52906). For downstream viral sequencing, the influenza virus genome segments from lung and stock RNA samples were reverse-transcribed using the iScript™ Reverse Transcription Supermix for RT-qPCR (Bio-Rad, Cat. No. 1708840) and the Uni12 primer (5'-AGC AAA AGC AGG-3') [182] with priming for 5 minutes at 25°C, reverse transcription for 20 minutes at 46°C, and reverse transcriptase inactivation for 1 minute at 95°C (Bio-Rad thermocycler, Cat. No. 186-1096). Viral complementary DNA (cDNA) was then synthesized by PCR for 40 cycles using *Taq* polymerase (New England Biolabs, Cat. No. M0273) and primers containing Illumina Nextera Transposase

adaptors. The primers were as follows: R1-Uni12 (5'-TCG TCG GCA GCG TCA GAT GTG TAT AAG AGA CAG AGC GAA AGC AGG-5') and R2-Uni13 (5'-GTC TCG TGG GCT CGG AGA TGT GTA TAA GAG ACA GAG TAG AAA CAA GG-3') (adaptor and barcode oligonucleotide sequences from Illumina, Inc., San Diego, CA, USA). Annealing and extension steps were performed for 30 seconds at 55°C and 7 minutes at 72°C respectively [183]. Prior to sequencing, PCR products were cleaned using the QIAquick PCR Purification Kit (QIAGEN, Cat. No. 28104), and up to 20 pM DNA samples were used for library preparation. Libraries were sequenced by CGEB-IMR (Centre for Genomics and Evolutionary Biology Integrated Microbiome Resource, Halifax, NS, Canada) (http://cgeb-imr.ca)) in a 300 + 300 bp paired-end MiSeq run (Illumina 600-cycle v3 kit, Cat. No. MS-102-3003).

2.6 Host transcriptome sequencing and analysis

Total mouse lung RNA was extracted at days 0 and 3 post-challenge using the RNeasy Mini Kit as described above (QIAGEN, Cat. No 74004) and purified with the QIAquick PCR Purification Kit (QIAGEN, Cat. No. 28104). Extracted RNA was sent to Novogene (Sacramento, CA, USA) for RNA sequencing. Gene expression was quantified in terms of fragments per kilobases mapped (FPKM). The gene expression fold-change per gene was calculated relative to uninfected mice, and genes were designated differentially expressed at analysis of variance (ANOVA) adjusted p-value < 0.05. Lists of significantly differentially expressed genes (DEGs) were compared across experimental groups to generate Venn diagrams using the R package VennDiagram (version 1.6.20).

To analyze biological processes enriched at challenge, lists of significantly differentially expressed genes per experimental group (naïve-mock vaccinated, naïve-vaccinated, preimmune-mock vaccinated, preimmune-vaccinated), were entered into the Database for Annotation,

Visualization and Integrated Discovery (DAVID) version 6.8 (https://david.ncifcrf.gov/home.jsp) functional annotation tool. The top Gene Ontology (GO) biological processes were selected and compared across experimental groups. Finally, GO biological processes related to antiviral response (GO term “defense response to virus” (GO:0051607)), T cell proliferation and cytotoxicity (terms “T cell-mediated cytotoxicity” (GO:0001913) and “activated t cell proliferation” (GO:0050798)), and the B cell response (GO terms “B cell chemotaxis” (GO:0035754), “B cell homeostasis” (GO:0001782), “B cell differentiation” (GO:0030183), “negative regulation of B cell differentiation” (GO:0045578), “B cell receptor complex” (GO:0019815), “negative regulation of B cell receptor signalling” (GO:0050859), “B cell apoptotic process” (GO:0001783), and “positive regulation of B cell proliferation” (GO:0030890)) were selected to build heatmaps in RStudio using the ggplot2 package (version 3.3.2). Differential expression was expressed in terms of log2(fold change).

2.7 Global SNP analysis and viral sequence alignment

For influenza virus sequence analysis, paired-end reads of variable length were imported into Geneious Version 2019.2.1 (Biomatters LTD) [184]. Reads were filtered and trimmed according to default settings, and primer sequences (R1-Uni12 and R2-Uni13 listed above) were removed. Within Geneious, paired reads were merged using the BBMerge too, and duplicate reads were removed using the DEDupe tool according to default settings. The remaining reads were then mapped to reference sequences for A/Mexico/4108/2009 viral segments. The reference sequences used for each segment alignment are as follows: PB2 (GQ379815), PB1 (GQ149652), PA (GQ149653), HA (GQ223112), NP (GQ149655), NA (GQ149650), M2 (GQ149657), and NEP (GQ149658). Following alignment, single nucleotide polymorphisms (SNPs) above 1% frequency [1] with a minimum of 5 reads supporting their discovery were

detected using the SNPs/variants tool in Geneious. I assessed group-specific SNPs by importing the consensus stock sequence into MEGAX (Molecular Evolutionary Genetics Analysis, Version 0.1) [185], and manually inputting SNPs that were found in the majority of experimental animals, regardless of frequency.

For the analysis of immune epitopes in the SARS-CoV-2 viral variants, I focused on analyzing the amino acid sequences acquired from public online databases. Amino acid sequences for the full Wuhan SARS-CoV-2 genome, as well as for SARS-CoV-2 variants D146G, B.1.1.7 (Alpha), B.1.351 (Beta), B.1.429 (Epsilon), P.1 (Gamma), B.1.617.1 (Kappa), B.1.617.2 (Delta), B.1.617.3, and mink-related (Cluster 5) SARS-CoV-2, were obtained from GISAID [186]. The accession numbers for each full amino acid sequence are as follows: Wuhan SARS-CoV-2 (EPI_ISL_402124), D614G (EPI_ISL_406862), B.1.1.7 (EPI_ISL_852526), B.1.351 (EPI_ISL_864621), B.1.429 (EPI_ISL_1017160), P.1 (EPI_ISL_1171653), B.1.617.1 (EPI_ISL_1841381), B.1.617.2 (EPI_ISL_1914598), B.1.617.3 (EPI_ISL_1939891), and mink-related (EPI_ISL_615652). Full genome sequences were truncated to restrict our analysis to the spike glycoprotein. I compared variant-specific mutations across the various spike proteins by aligning amino acid sequences in MEGAX, with deletions denoted by dashes. All alignment figures were generated in ClustalX Version 2.0 [187].

2.8 Modelling influenza virus and SARS-CoV-2 protein mutations

To generate folding models of variant influenza viral proteins and SARS-CoV-2 spike variants based on our generated or publicly-available amino acid sequences, I used the web-based bioinformatics server Phyre2 (Protein Homology/ analogY Recognition Engine Version 2.0 (http://www.sbg.bio.ic.ac.uk/~phyre2/html/page.cgi?id=index)) in intensive modelling mode [188]. For SARS-CoV-2 spike variants, the following templates were applied to model the reference and

variant spike sequences: c7dk3B, c6vsb, c2kncA, c7a4nA, c6vsbA, c2fxpA, c3j3bG, c6vybB, c6zowB, c1m0jA, c1s6wA, c5h36E, c4qaeP, c3h0tC, c6wbvB, c1m4fA, c2kefA, and c4qaeR (**Table 2.1**). The model coordinate files in Protein Database (PDB) format were downloaded for downstream visualization and manipulation. I created structural images of each influenza viral protein or spike variant protein with PyMOL (The PyMOL Molecular Graphics System, Version 2.3.5, Schrödinger, LLC). Maximum performance quality was selected, and proteins were viewed as "sphere" models with the variant-defining mutations highlighted.

Table 2.1 Protein database templates and % identity after modelling Wuhan SARS-CoV-2 and nine variants of concern/interest using the Phyre2 platform.

	Reference	B.1.1.7	B.1.351	B.1.429	D614G	Mink-related	P.1	B.1.617.1	B.1.617.2	B.1.617.3
c7dk3B	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
c6vsbB	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
c2kncA	68%	73%	66%	76%	-	76%	75%	73%	75%	-
c7a4nA	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
c6vsbA	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
c2fxpA	99%	99%	99%	99%	99%	99%	99%	99%	99%	99%
c3j3bG	69%	70%	70%	71%	68%	71%	70%	71%	-	67%
c6vybB	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
c6zowB	-	100%	100%	100%	-	100%	100%	100%	100%	100%
c1m0jA	-	66%	-	62%	-	62%	69%	66%	56%	-
c1s6wA	-	-	50%	-	63%	-	-	-	-	64%
c5h36E	-	-	-	-	55%	-	-	-	-	53%
c4qaeP	-	-	-	-	-	-	-	-	66%	-
c3h0tC	-	-	-	-	-	-	-	-	68%	-
c6wbvB	-	-	-	-	-	-	-	-	66%	-
c1m4fA	-	-	-	-	-	-	-	-	68%	-
c2kefA	-	-	-	-	-	-	-	-	68%	-
c4qaeR	-	-	-	-	-	-	-	-	66%	-

Next, the web-based software Missense3D (http://www.sbg.bio.ic.ac.uk/~missense3d/) [189] was employed to predict the structural impact of each individual variant amino acid

substitution. For this analysis, either the original influenza virus protein (i.e., HA, NA, NP, or PB1) or Wuhan spike sequence was set as the reference, and substitutions across all protein variants were input manually. A cavity volume expansion or contraction of ≥ 70 $Å^3$ was defined as structural damage to differentiate from minor expansions or contractions.

A second set of protein-folding predictions was generated for each SARS-CoV-2 spike variant using the Swiss-Model structural bioinformatics server (https://swissmodel.expasy.org/interactive). Models were built based on multiple computer-generated templates determined by BLAST-searching homologous sequences with known tertiary and quaternary structures [190]. The finished products were scored based on a Quantitative Model Energy ANalysis (QMEAN) and Global Model Quality Estimate or GMQE scoring system [190,191] as well as analyzed based on their agreement with Ramachandran plots. Models with the lowest absolute value of their QMEAN scores, highest GMQE scores, and highest Ramachandran agreement were kept in Protein Database (PDB) format to be further analyzed for each variant (**Table 2.2**). Each 3D model was viewed and edited in PyMol. Glycan additions predicted by Swiss-Model were removed from the rough trimeric models and the "A" chain monomer was isolated to be further analyzed from full, trimeric models.

Table 2.2 SWISS-Model parameters for SARS-CoV-2 spike sequences for Wuhan reference, D614G, Alpha, Beta, Gamma, Delta, Epsilon, Kappa, B.1.617.3, and mink-related variants.

GISAID ID tag	Variant	IQmeanI (Z score)	GMQE (/1)	Ramachandran agreement (%) chain "A"	# of subs	# AA's deleted	Model template SMTL ID#	Template coverage (%)
EPI_ISL_402124	WIV04 (ref)	1.94	0.64	93.11	0	0	7cwu.1.G	100
EPI_ISL_406862	D614G	1.58	0.71	93.84	1	0	7krs.1.A	100
EPI_ISL_852526	Alpha (B.1.1.7)	1.51	0.72	93.91	7	3	7krs.1.A	98.98
EPI_ISL_864621	Beta (B.1.351)	1.62	0.71	93.74	7	3	7krs.1.A	99.45
EPI_ISL_1171653	Gamma (P.1)	1.48	0.71	93.9	11	0	7krs.1.A	99.14
EPI_ISL_1914598	Delta (B.1.617.2)	1.55	0.72	94.15	8	2	7krs.1.A	99.37
EPI_ISL_1017160	Epsilon (B.1.429)	1.5	0.71	93.87	3	0	7krs.1.A	99.76
EPI_ISL_1841381	Kappa (B.1.617.1)	1.49	0.72	93.93	7	0	7krs.1.A	99.53
EPI_ISL_1939891	B.1.617.3	1.61	0.72	93.93	7	0	7krs.1.A	99.53
EPI_ISL_615652	Cluster 5 (mink related)	1.52	0.72	93.27	3	2	7krs.1.A	99.53

2.9 Prediction of B cell epitopes

Predicted B cell epitopes across influenza virus proteins and SARS-CoV-2 spike variants were assessed using the DiscoTope 2.0 server (http://www.cbs.dtu.dk/services/DiscoTope/) [192]. Following both Phyre2 and Swiss-Model predictions for protein folding, the corresponding PDB file of each sequence was input to DiscoTope 2.0 using the least conservative threshold score of -3.7 (0.47 sensitivity, 0.75 specificity). As I found the Swiss-Model to ultimately provide the more accurate folding model of the SARS-CoV-2 spike protein, output obtained from the Swiss-Model PDB files was used to generate heatmaps of B cell epitope probability across SARS-CoV-2 spike variant protein residues in PyMOL, with predicted epitopes in yellow. The amino acid number and DiscoTope score were used to map predicted B cell epitopes in RStudio (version 3.6.1) using the ggplot2 package (version 3.3.2) [193].

2.10 Variant spike glycosylation analysis

Original Wuhan and variant amino acid sequences for SARS-CoV-2 spike were uploaded to the NetNGlyc 1.0 server (http://www.cbs.dtu.dk/services/NetNGlyc/) [194] to assess potential N-linked glycosylation sites through detection of Asn-Xaa-Ser/Thr sequons. The default threshold of 0.5 was used. The predicted glycosylation sites across variants were modelled again using PyMOL. The Glycam Glycoprotein online tool (http://glycam.org/tools/molecular-dynamics/glycoprotein-builder/upload-pdb) [195] was used to attach representative $Man_9GlcNAc_2$ chains to each positively predicted glycosylation site.

2.11 Assessment of predicted T cell-binding peptides

The likelihood of influenza virus or SARS-CoV-2 peptides being presented on mouse or human MHC-I molecules, respectively, was predicted using the Immune Epitope Database (IEDB) Analysis Resource NetMHCpan EL 4.1 (http://tools.immuneepitope.org/main/). For influenza virus peptide analysis, the complete amino acid sequences for reference and mutated protein sequences were used. Our analysis was restricted to the two HLA alleles present in our mouse model, which are H-2-Kb and H-2-Db. For analysis of SARS-CoV-2 MHC-I-binding peptides, the complete amino acid sequences for the Wuhan and variant SARS-CoV-2 spike proteins were used, and predicted peptides were restricted to 27 common human HLA genes to encompass most of the human population [81,196]. Analysis of MHC-II peptides was only performed for SARS-CoV-2 sequences. I applied the NetMHCII 2.3 server from DTU Health Tech and selected HLA-DP, HLA-DQ, and HLA-DR alleles (http://www.cbs.dtu.dk/services/NetMHCII/) [197].

To determine immune background- or variant-specific epitopes in influenza proteins or SARS-CoV-2 spike respectively, T cell epitopes (defined as unique HLA allele*peptide combinations) were compared in RStudio using the tidyverse package (version 1.3.0) [198]. For

MHC-I and MHC-II analysis, high-affinity peptides were defined as those with scores falling below the 0.5 and 2.0 percentile ranks, respectively [199].

2.12 Statistical analysis

All statistical analyses were conducted in RStudio version 3.6.1. For normally distributed data, one-way analysis of variance (ANOVA) followed by Tukey's Honest Significant Difference (HSD) was used to determine differences between groups. A p-value threshold of 0.05 was used. For determination of differentially expressed genes by RNA-sequencing, an adjusted p-value threshold of 0.1 was used. All plots were generated in RStudio using the ggplot2 package version 3.3.2.

CHAPTER 3 INFLUENZA VIRUS INFECTION AND VACCINATION HISTORY DRIVES VIRAL MUTATIONS AT HETEROLOGOUS CHALLENGE

3.1 Introduction to the antigenic drift problem

Each person has a complex history of influenza virus infections and vaccinations. Although an individual will have several exposures to influenza viruses by adulthood, influenza viruses continue to evade pre-existing immunity and circulate across the globe [200]. This is due to the greatest challenge facing immunity against influenza viruses, which is antigenic drift [176,201–203]. Antigenic drift refers to the accumulation of mutations on viral proteins hemagglutinin and neuraminidase that lead to changes in antigenicity, which are eventually sufficient to escape host immune memory [204].

Antigenic drift is typically considered a cause-and-effect relationship from the virus perspective, where vaccines are updated with new antigens to reflect the newly circulating strains. However, less attention is given to the drivers of antigenic drift. It is generally known that antigenic drift is the result of immune selection of viral mutations, as the immunodominant viral proteins hemagglutinin and neuraminidase are those that primarily undergo mutation [205–208]. Indeed, some studies have established the important link between strain-specific neutralizing antibodies and mutation of influenza virus hemagglutinin [129,205,207]. Nonetheless, our understanding of how previous influenza virus infection or vaccination contributes to antigenic changes remains limited.

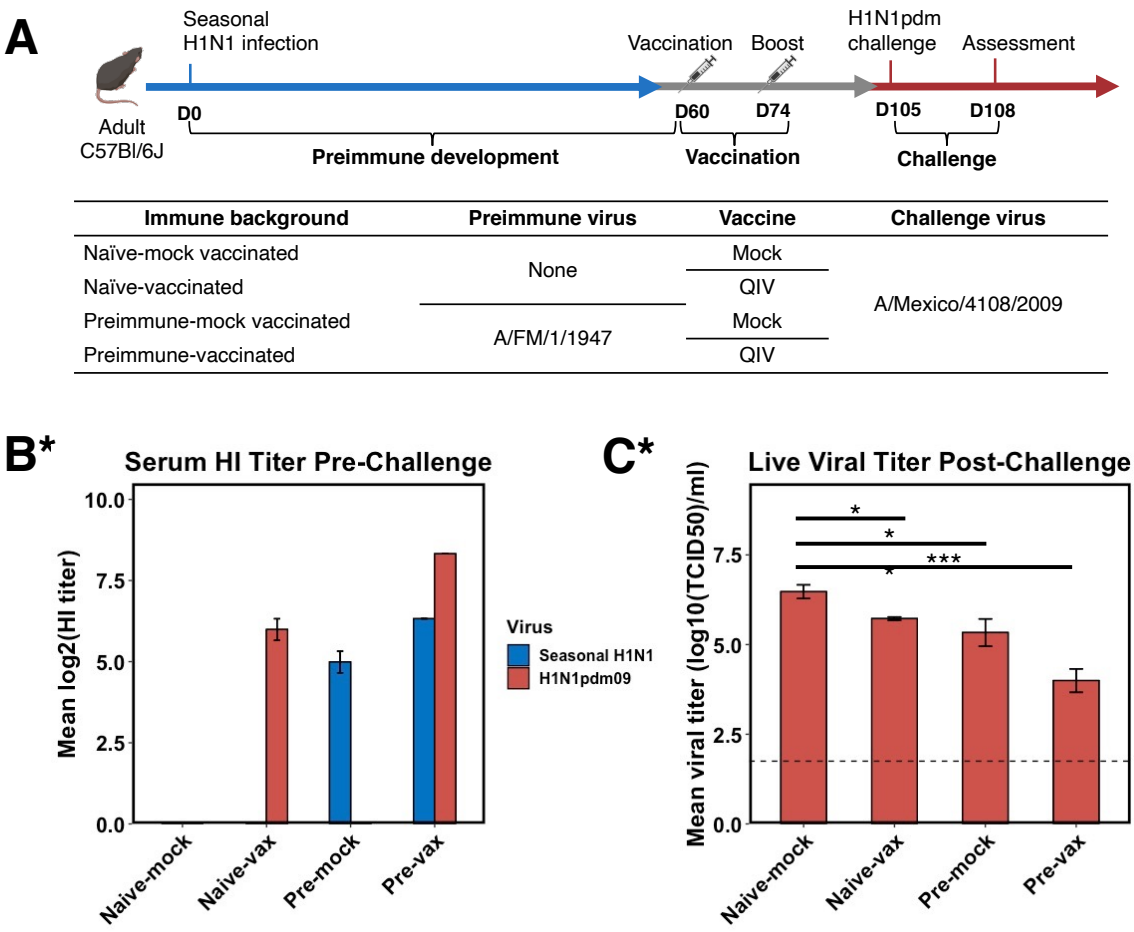
Given that adaptive immune responses are antigen-specific, and that each host has a unique immune history of influenza virus infections and vaccinations, I hypothesize that a host's specific immune background may dictate the capacity for antigenic drift at re-exposure. It is important to consider how the diverse immune backgrounds that exist in the population toward influenza viruses, as well as the differences in immunity established through natural infection

versus vaccination, may drive antigenic drift in different ways. Moreover, the eight segments of the influenza virus genome encode at least ten different proteins that have varying levels immunogenicity and conservation across influenza viruses. The external proteins hemagglutinin (HA) and neuraminidase (NA) are rapidly mutating antigenic proteins on the virion surface; matrix proteins M1 and M2 interact with the virion surface and have important roles in virion assembly and uncoating; RNA polymerase components polymerase basic 1 (PB1), polymerase basic 2 (PB2), polymerase basic 1-F2 (PB1-F2), and polymerase acidic (PA) are internal and responsible for viral transcription and replication; nucleoprotein (NP) binds viral RNA and helps facilitate synthesis; nuclear export protein (NEP) facilitates viral RNA export from the host nucleus; and non-structural protein 1 (NS1) antagonises the host interferon response. The location and functional roles of each protein have implications for their capacity to mutate without suffering a fitness cost, as well as their role as viral antigens. Thus, the goal of my study was to investigate how the immune background of the host contributes to patterns of mutation on the ten different influenza virus proteins at re-infection, and whether these mutations lead to detectable changes in antigenicity. Here I combined an *in vivo* challenge study with downstream computational approaches to evaluate how previous immunity can drive antigenic changes on influenza virus proteins. I hypothesized that at three days post-challenge, the specific infection and vaccination history of the mice would influence the mutations detected across influenza virus proteins, specifically on the HA protein.

3.2 Study design and the preimmune-vaccinated mouse model

To assess the influence of the host immune background on the emergence of influenza virus mutations during infection, I utilized lung samples generated prior to my arrival in the lab. We previously constructed immune backgrounds in adult C57Bl/6j mice through combinations of

sequential viral infection, vaccination, and heterologous challenge. A schematic of the experimental time course and group descriptions are shown in **Figure 3.1A**. For the purpose of this study, two distinct lineages of influenza virus were used: the first was a seasonal, historical H1N1 strain to be used in the initial infections (A/FortMonmouth/1/1947, herein referred to as FM/47), and the second was a 2009 pandemic H1N1 to be used in vaccination and challenge (A/Mexico/4108/2009, herein referred to as Mex/09). Previous studies by our group and others have confirmed that mice are susceptible to infection with Mex/09 [209–212], and FM/47 is itself a mouse-adapted influenza virus [213], making these appropriate strains for our purposes.



Immune background	Preimmune virus	Vaccine	Challenge virus
Naïve-mock vaccinated	None	Mock	A/Mexico/4108/2009
Naïve-vaccinated		QIV	
Preimmune-mock vaccinated	A/FM/1/1947	Mock	
Preimmune-vaccinated		QIV	

Figure 3.1 Developing influenza virus immune backgrounds in adult mice.

A) At day 0 of the study, adult C57Bl/6j mice were intranasally infected with seasonal H1N1 strain A/FortMonmouth/1/1947 (FM/47), followed by intramuscular vaccination and boost with the 2018-2019 quadrivalent vaccine on days 60 and 74. Mice were divided into four immune backgrounds depending on their infection and vaccination status: naïve-mock vaccinated, naïve-vaccinated, preimmune-mock vaccinated, and preimmune-vaccinated. On day 105 of the study, mice of all immune backgrounds were subject to a lethal dose of pandemic H1N1 strain A/Mexico/4108/2009 (Mex/09), followed by assessment of host and virus on day 108. **(B)** Three days prior to challenge, serum was isolated from collected blood samples to quantify neutralizing antibody titers against FM/47 and Mex/09 viruses using a hemagglutinin inhibition assay. These assays were done to confirm the establishment of immune memory offered by FM/47 infection and quadrivalent vaccination. **(C)** At three days post-challenge, live viral titer was measured in the mouse lungs by $TCID_{50}$ viral load assay with hemagglutination assessment. Figures B and C were generated using the ggplot2 package version 3.3.2. Groups were compared by one-way ANOVA (signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1) and Tukey multiple comparisons of means (95% family-wise confidence level). *I completed the analysis and generated the plot while the data itself were generated by M. Francis prior to my arrival in the lab.

On day 0 of the study, mice were infected intranasally with FM/47 at $10^{3.5}$ $TCID_{50}$ and left to develop immune memory over two months. Mice were then vaccinated intramuscularly with the Sanofi Fluzone quadrivalent influenza vaccine (QIV) from 2018-2019 on day 60 and boosted on day 74. Specifically, the QIV contains HA proteins from the following influenza viruses: H1N1 (A/Michigan/45/2015 X-275), H3N2 (A/Brisbane/1/2018 X-311), influenza B Yamagata lineage (B/Phuket/3073/2013), and influenza B Victoria lineage (B/Colorado/6/2017-like virus). One month after vaccination, mice were challenged intranasally with a lethal dose of Mex/09 (10^6 $TCID_{50}$) and assessed at three days post-challenge. Importantly, the challenge virus was antigenically similar to the H1 vaccine component of the vaccine, where the H1N1 FM/47 HA shares only 78% identity with the HA from the Mex/09 H1N1 virus.

Here I defined these FM/47-infected mice as *preimmune,* as they possessed pre-existing H1N1 immunity to influenza virus distinct from that elicited by influenza virus vaccination. Conversely, mice with no previous exposure to influenza viruses were defined as *naïve*, and those exposed to the 2018-2019 vaccine were *vaccinated.* At Mex/09 challenge, all mice fell into one of four groups or *immune backgrounds* that mirrored the possible influenza virus infection and vaccination statuses of the human population: naïve-mock vaccinated, naïve-vaccinated, preimmune-mock vaccinated, and preimmune-vaccinated.

As the infection and vaccination histories of each group would impose unique immune pressures at challenge, it was important to confirm the establishment of previous immunity from infection and/or vaccination by assessing the neutralizing antibody levels against the influenza viruses from each distinct lineage. Three days prior to Mex/09 challenge, the preimmune and vaccinated status of mice was confirmed by hemagglutination inhibition (HI) assay against preimmune (FM/47) and vaccine (Mex/09) influenza strains (**Figure 3.1B**). Confirming vaccination, naïve-vaccinated and preimmune-vaccinated mice had average log_2(HI titers) of 6 and 8 HAI units, respectively, toward Mex/09. Both preimmune groups had high HI titer against

the preimmune FM/47 strain (5 and 7 HAI units in preimmune-mock and preimmune-vaccinated groups, respectively). The presence of antibodies specific to each virus only in the mice exposed to the specific viral antigens demonstrated that adaptive immune responses specific to each influenza virus lineage had been established in the mouse model.

3.3 Preimmune-vaccinated mice still replicate H1N1 influenza Mex/09 virus in the lungs

Following confirmation of neutralizing antibody titers against FM/47 and Mex/09 in the preimmune and vaccinated animals, mice of all immune backgrounds were subject to the same infectious dose of Mex/09 by intranasal inoculation. Assessment of viral load in the lungs at three days post-challenge by $TCID_{50}$ assay followed by hemagglutination assessment showed that productive influenza viral replication still occurred in preimmune-vaccinated mice, as well as the mice with other immune backgrounds (**Figure 3.1C**). Live viral titer in preimmune-vaccinated mice reached an average of 4 ± 0.33 $\log_{10}TCID_{50}$ ml^{-1} at three days post-challenge. Additionally, all other mice with varying backgrounds were positive for live virus at greater degrees within the lung tissue (**Figure 3.1C**). These data suggested that despite a layered immune background, which included vaccination against a similar virus as the challenge strain Mex/09, replication of a subset of the initial viral population evaded sterilizing immunity in the preimmune-vaccinated mice.

3.4 Combined preimmunity and vaccination decreases the frequency and number of influenza virus single nucleotide polymorphisms post-challenge

Since viral replication was observed in the lungs of preimmune-vaccinated mice as well as in the mice of other immune backgrounds, I aimed to characterize the viral population that had evaded pre-existing immunity established in the mouse model with historical H1N1 virus infection of QIV vaccination. To catalog the viral quasispecies and variants that remained in the mice during

infection, Illumina MiSeq whole-viral genome sequencing was performed (Halifax, NS, CA) on viral RNA isolated from lung tissue at three days post-challenge. The viral sequences are available on the Sequence Read Archive (SRA; BioProject ID PRJNA787976; https://www.ncbi.nlm.nih.gov/biosample?Db=biosample&DbFrom=bioproject&Cmd=Link&LinkName=bioproject_biosample&LinkReadableName=BioSample&ordinalpos=1&IdsFromResult=787976, see **Appendix 1**). Using Geneious software [184], I aligned the 50-300 base-pair reads to reference Mex/09 sequences for each of eight influenza virus segments encoding ten different proteins (**Table 3.1**). The reference sequences and NCBI accession numbers used for each alignment are as follows: polymerase basic protein 2 (PB2, GQ379815); polymerase basic protein 1 (PB1, GQ149652); polymerase acidic protein (PA, GQ149653); hemagglutinin (HA, GQ223112); nucleoprotein (NP, GQ149655); neuraminidase (NA, GQ149650); matrix-2 protein (M2, GQ149657); and nuclear export protein (NEP, GQ149658). To determine the variation existing within the input virus, I sequenced our stock of Mex/09 used for mouse inoculation and mapped this to reference Mex/09 as well.

Although the highest number of reads were mapped toward the 3' and 5' ends of the viral segments, 100% coverage was achieved for all segments (**Table 3.1**). Across experimental groups, the number of reads per segment was generally correlated with reference length, except for segments NA, M1/M2, and non-structural protein 1 (NS1). Interestingly, the number of reads per segment did not appear to be correlated with the viral load trends shown in **Figure 3.1C**. Sequenced stock Mex/09 yielded the highest number of reads and coverage across most segments compared to virus extracted from mouse lungs, as expected.

Table 3.1 Mean number of reads and coverage per influenza virus gene following alignment of viral sequences to A/Mexico/4108/2009 reference segments in Geneious.

		Stock Mex/09		Naïve-mock vaccinated		Naïve-vaccinated		Preimmune-mock vaccinated		Preimmune-vaccinated	
Gene	**Length (nt)**	Mean reads	Mean coverage	Mean reads	Mean coverage	Mean reads	Mean coverage	Mean reads	Mean coverage	Mean reads	Mean coverage
PB2	2280	152,153	36,192	116,243	18,782	144,600	22,568	156,061	19,354	201,613	26,076
PB1	2274	111,003	18,718	78,144	10,244	122,704	17,736	94,685	9378	92,614	8949
PA	2151	75,936	13,124	52,570	5553	19,548	1564	46,844	3716	54,251	4773
HA	1701	42,205	5111	6787	799	11,001	1091	9939	899	5776	529
NP	1497	6591	771	1831	206	632	74	2249	185	1375	140
NA	1410	20,003	2382	1430	169	1904	119	1027	130	2935	410
M1, M2	982	43,187	7426	11,507	2095	5142	942	7939	1315	13,928	2335
NS1	844	60,291	11,449	16,024	3330	14,358	2999	17,260	3364	47,157	6245

Following alignment of viral sequences to reference Mex/09, I aimed to determine whether novel viral mutations had appeared on influenza virus genes extracted from mouse lungs, or whether variants already present in the stock virus had changed in frequency. To this end, I again employed Geneious software to catalog the insertions, deletions, and single nucleotide polymorphisms (SNPs) present on the gene sequences using the SNPs/variants tool. The complete bioinformatic analysis pipeline is shown in **Figure 3.2.**

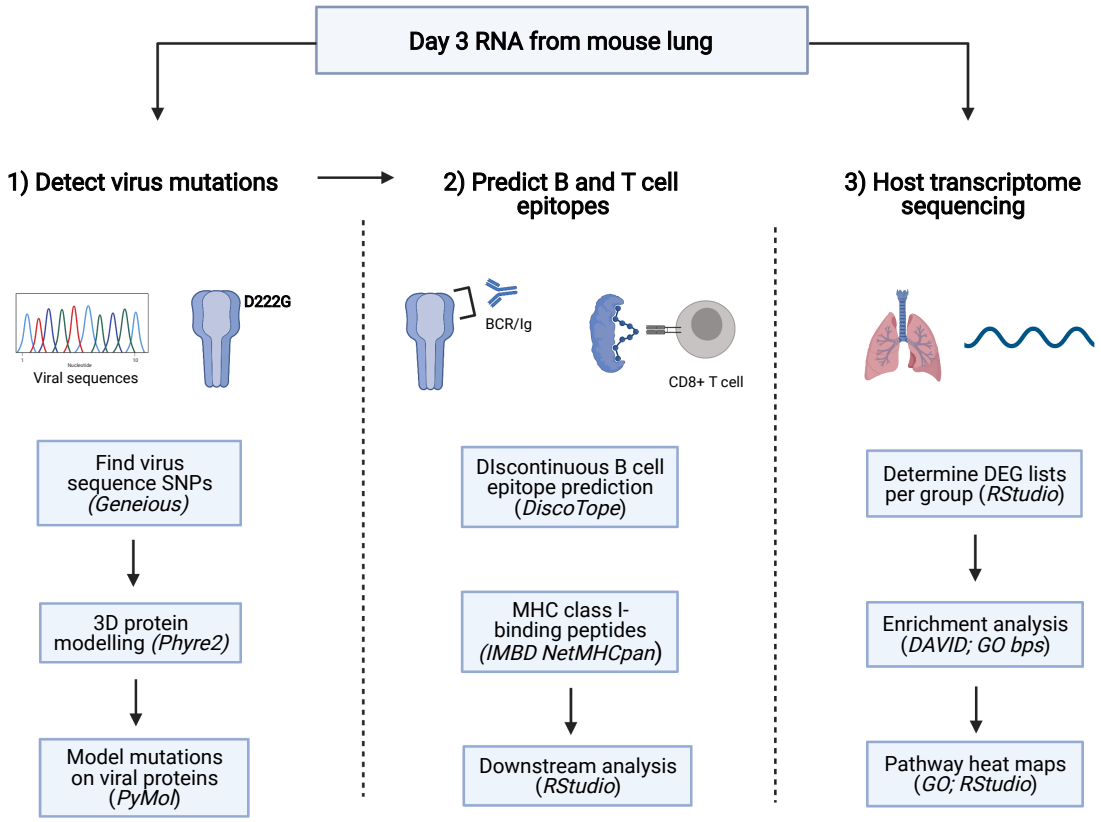


Figure 3.2 Bioinformatic analysis pipeline to determine influenza virus mutations, predicted B and T cell epitopes, and host transcriptome analysis after challenge.

At three days post-challenge with A/Mexico/4108/2009 (Mex/09), viral RNA from mouse lungs was sequenced using Illumina MiSeq. Viral sequences were aligned to reference Mex/09 using Geneious, and SNPs and variants above 1% frequency were detected for sequences extracted from mice of each immune background. The gene sequences were aligned using MEGAX, and only SNPs found in most animals of an experimental group were retained in the sequence. To predict the possible structural and downstream antigenic impact of each immune background-specific amino substitution, a representative 3D model of each HA protein was generated using the Phyre2 platform. Protein images were generated in PyMol to visualize the structural location of each mutation. The 3D model was used to predict discontinuous B cell epitopes along the viral proteins of each group using DiscoTope 2.0 (DTU Health Tech). Variant amino acid sequences were uploaded to the Immune Epitope Database NetMHCpan EL 4.1 tool to predict T cell epitopes binding MHC class I molecules in the mouse model. Finally, the host transcriptional response was investigated by RNA sequencing of mouse lung RNA. Lists of significantly differentially expressed genes (DEGs) for each immune background were uploaded to the Database for Annotation, Visualization and Integrated Discovery (DAVID) version 6.8 functional annotation tool to determine enriched immune pathways. Heatmaps were generated in for antiviral, T cell, and B cell-mediated immune response pathways using RStudio.

I restricted my analysis to gene variants above one percent frequency with a minimum of five reads supporting their discovery. Moreover, I selected only single nucleotide variants for downstream analysis to simplify comparisons across viral proteins. I generated Manhattan plots and frequency histograms to visualize the distribution of viral SNPs and their frequencies across groups and over viral genes (**Figure 3.3**). Here I detected highly significant ($p << 0.05$) SNPs across all ten influenza virus genes and across all immune backgrounds, except for M1/M2 segment of the stock virus, and M2 gene in the naïve-mock vaccinated group (see **Appendix 1** for a complete list of all protein variants found). For HA, naïve-mock vaccinated mice had markedly less SNPs than the stock virus used to inoculate animals, indicating selection of variants adaptive to the mouse model. Naïve-vaccinated mice showed more mutations across HA compared to all other groups, with an increased number of SNPs that were highly significant ($p << 0.05$). Although preimmune groups had a lower number of detectable SNPs across HA, a greater proportion of these SNPs were highly significant compared to the HA sequences from naïve mice.

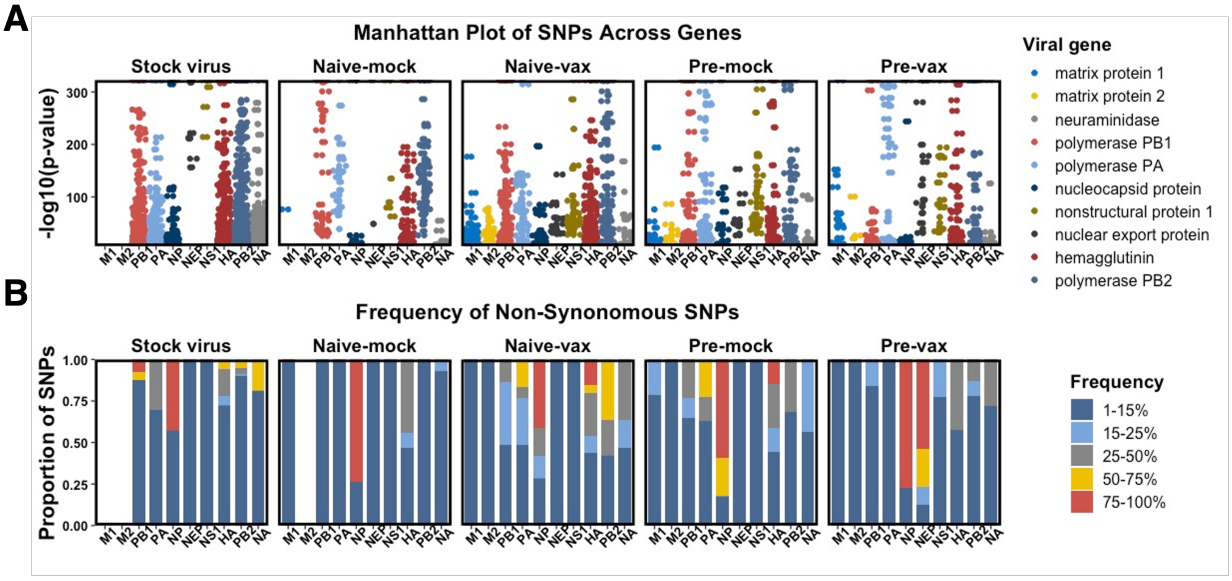


Figure 3.3 Influenza virus gene sequences extracted from mice of different immune backgrounds post-H1N1challenge have significant SNPs on all viral genes.

At three days post-challenge with A/Mexico/4108/2009 (Mex/09), viral RNA from mouse lungs was sequenced using Illumina MiSeq. Viral sequences were aligned to reference Mex/09 (accession numbers as follows: PB2 (GQ379815), PB1 (GQ149652), PA (GQ149653), HA (GQ223112), NP (GQ149655), NA (GQ149650), M2 (GQ149657), and NEP (GQ149658) using Geneious, and SNPs and variants above 1% frequency were detected for sequences extracted from mice of each immune background. Stock Mex/09 used to inoculate animals was sequenced in parallel to determine the existing sequence variation. **(A)** The SNPs found in all ten influenza virus genes were visualized using a Manhattan plot to determine the number and significance ($-\log_{10}$(p-value) of each mutation. **(B)** After translating the genetic sequence to protein sequence, the proportion of non-synonymous SNPs detected for each protein were categorized by their frequency, with frequency bins as follows: 1-15%, 15-25%, 25-50%, 50-75%, and 75-100%. The figure was generated using ggplot2 version 3.3.2.

The trend of increased detectable SNPs in naive-vaccinated mice was generally consistent for other influenza virus genes. Nucleocapsid protein (NP) and neuraminidase (NA) both showed the greatest number of mutations in the naïve-vaccinated and preimmune-mock vaccinated groups, with the least number of SNPs detected in preimmune-vaccinated animals (**Figure 3.3A**). Similarly, polymerase basic subunits 1 and 2 (PB1 and PB2) had the greatest number of SNPs in naïve-vaccinated animals. For NA, selection of mutations was largely evident in naïve-mock vaccinated animals, as the number of SNPs greatly decreased compared to the stock virus used to inoculate animals. In general, I observed the greatest variation in influenza viral gene sequences when mice had only one prior exposure, either through previous infection with FM/47, or vaccination with 2009 pandemic H1N1. When preimmunity and vaccination are combined, or when mice have no prior exposures (i.e., naïve), the variation in viral sequences decreased.

To address whether SNPs already present in the stock virus were differentially selected when encountering the previous immunity of each experimental group, I next examined the frequency distribution of SNPs across viral genes leading to amino acid changes. Since not all SNPs represent amino acid changes which may lead to changes in antigenicity, I used RStudio

to select only SNPs leading to amino acid substitution, truncation, or start codon loss across viral genes (**Figure 3.3B**). I then placed the remaining non-synonymous SNPs into one of five frequency bins (1-15%, 15-25%, 25-50%, 50-75%, and 75-100%) based on the percentage reported by Geneious and plotted the proportion falling into each bin (**Figure 3.3B**). Here the aim was to visualize whether viral genes had a smaller number of variants that increased in frequency in certain immune backgrounds (i.e., selective propagation of mutations), versus a greater number of variants that existed at lower frequencies with a relatively equal chance of existing in the viral population.

As shown in **Figure 3.3A**, there were no M1/M2 SNPs of any frequency detected in the stock virus or in naïve-mock vaccinated animals (**Figure 3.3B**). After replicating within the mouse host, detectable SNPs along M1/M2 appeared in naïve-vaccinated, preimmune-mock vaccinated, and preimmune-vaccinated groups. However, only preimmune-mock vaccinated animals had M1/M2 SNPs above 15% frequency (**Figure 3.3B**). For HA sequences, at least one quarter of the total SNPs fell into the higher-frequency ranges (> 15%) by three days post-challenge for all experimental groups (**Figure 3.3B**). This proportion was even higher in the naïve-mock vaccinated animals, with over half of all HA SNPs occurring above 15% frequency. In the naïve-vaccinated mice, the same proportion (approximately half) of SNPs fell into the higher frequency range, but unlike the naïve-mock vaccinated mice, a small subset of mutations appeared above 75% frequency. This subset above 75% frequency was also present in preimmune-mock vaccinated animals. Finally, less than half of the total SNPs in the HA sequences from preimmune-vaccinated animals fell into the 25-50% frequency range, which is comparatively lower than the other experimental groups (**Figure 3.3B**).

For the HA gene, it seems that the presence of previous infection combined with vaccination, or conversely, no previous immunity against the challenge strain, is associated with a decrease in high-frequency variants like the trend observed in **Figure 3.3A**. In the previous

analysis combining previous infection with vaccination led to less variation along viral genes. Overall, this preliminary analysis suggests a relationship between the infection and vaccination history of the host, and potential trends for viral mutation in terms of significance and SNP frequency throughout the course of an influenza virus infection.

3.5 The hemagglutinin protein shows immune background-specific mutations that may impact protein structure

The HA protein on the influenza virion surface facilitates viral entry into host cells, is a major immunogen, and is also prone to mutation [214]. I therefore hypothesized that pre-existing immunity from infection and vaccination would have the greatest impact on this immunologically dominant surface protein, leading to mutations that alter protein shape and downstream antigenicity. To investigate the presence of specific mutations evading immune detection, I plotted the SNPs that led to non-synonymous changes in the HA amino acid sequence (**Figure 3.4A)**. As expected, the stock Mex/09 virus used to inoculate animals had the greatest number of HA variants, and analysis after challenge in all backgrounds indicated selection in the animal model as evidenced by the SNP decrease in naïve-mock vaccinated mice (**Figure 3.4A**). However, novel mutations that were not detected in stock virus also emerged in naïve-mock vaccinated animals, suggesting the emergence due to errors in replication.

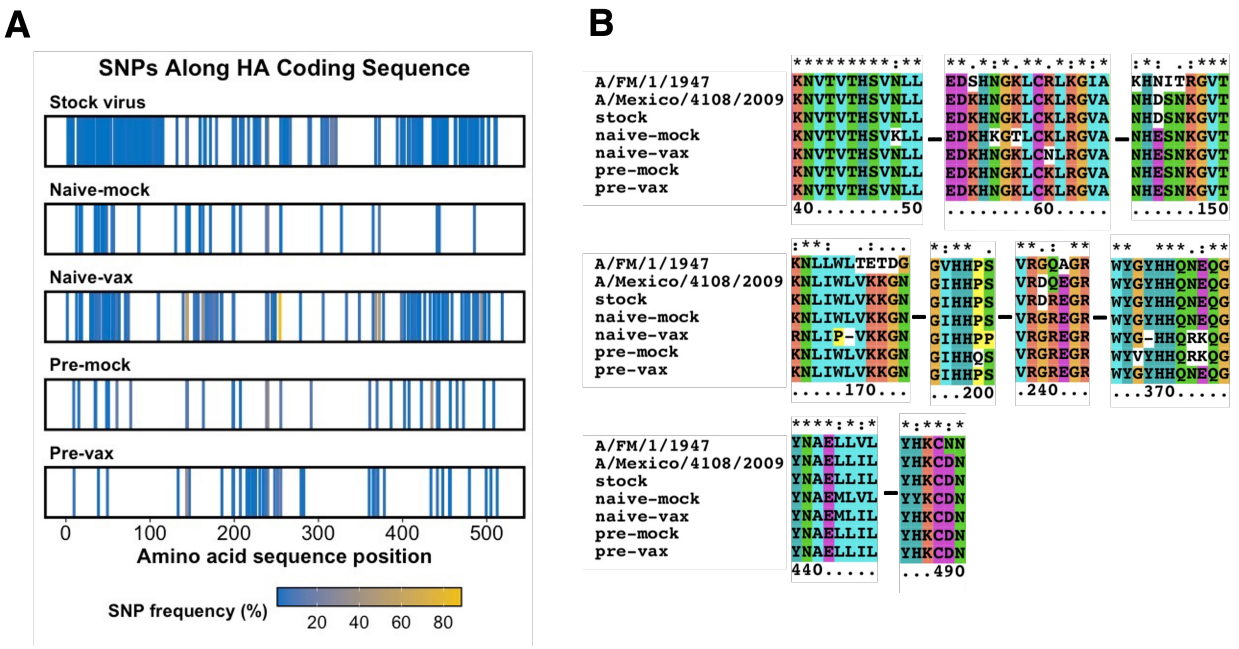


Figure 3.4 Naïve-vaccinated animals have the highest frequency of SNPs across influenza virus HA post-H1N1 challenge.

At three days post-challenge with A/Mexico/4108/2009 (Mex/09), viral RNA from mouse lungs was sequenced using Illumina MiSeq and aligned to reference Mex/09 HA using Geneious (NCBI accession no. GQ223112). The SNPs and variants above 1% frequency were detected in all influenza virus genes. Stock Mex/09 used to inoculate animals was sequenced in parallel to determine the existing sequence variation. **(A)** The non-synonymous SNPs found in HA gene sequences after introduction to naïve-mock vaccinated, naïve-vaccinated, preimmune-mock vaccinated, and preimmune-vaccinated mice were plotted on the amino acid sequence number of HA (1 band = 1 SNP). Each line representing a mutation found per experimental group generated a bar-code effect for each SNP profile. The color of the band represents the relative frequency of each mutation among the total population of viruses sequenced in the lung. The figure was generated using the ggplot2 package version 3.3.2. **(B)** Hemagglutinin gene sequences from mice of each immune background were aligned using MEGAX. To better determine selection of non-synonymous SNPs dependent on immune background, only SNPs that appeared in at least two of three mice from each group were retained in the gene sequence. The amino acid sequences containing immune background-specific substitutions were shown using ClustalX in the alignment.

Of the four immune backgrounds, the greatest number of SNPs on HA post-challenge were detected in naïve-vaccinated animals, with the majority in the 1-20% frequency range (**Figure 3.3B; Figure 3.4A**). Naïve-vaccinated animals had the highest density of HA mutations at the N-terminal F' subdomain (amino acids 1-41, H1 numbering), the receptor binding domain

(110-262), and the F subdomain in the stalk region (348-515) (**Figure 3.4A**) [215]. Preimmune-mock vaccinated animals had a relatively even distribution of mutations across the entire HA protein, while mutations in preimmune-vaccinated mice were most dense in the receptor binding domain (**Figure 3.4A**).

To further evaluate which mutations were the result of selection pressure by the host immune background, I aligned the HA sequences in MEGA [185] and filtered the mutations to include only those shared by the majority of animals of each group (**Figure 3.4B**). Interestingly, the number of shared mutations between animals of a group was not associated with the overall number of SNPs that persisted in that group; for example, although preimmune-mock and preimmune-vaccinated animals show a similar number of SNPs across the HA gene, preimmune-mock vaccinated animals have 5X as many shared SNPs within their group compared to preimmune-vaccinated mice (**Figure 3.4B;** summarized in **Table 3.2**). To determine how each of these mutations may impact protein structure and antigenicity, I next generated protein folding predictions for the HA amino acid sequences for each immune background using the Phyre2 platform [188]. Structural information in protein database (PDB) format from Phyre2 was uploaded into PyMOL (The PyMOL Molecular Graphics System, Version 2.3.5, Schrödinger, LLC) to generate a representative 3D model containing the immune background-specific mutations for each group (**Figure 3.5**). Since only mutations shared between animals of a group were included in the sequence from which the folding predictions were derived, only those shared mutations are shown in **Figure 3.5**.

One mutation in the receptor binding domain of HA, specifically D222G, was detected in mice of all immune backgrounds and was the only mutation common to all mice within the preimmune-vaccinated group (**Figure 3.5**). However, for each of the remaining immune backgrounds, group-specific mutations were identified in both the head and stalk regions of HA (**Table 3.2**). Naïve-mock vaccinated mice had several group-specific mutations in the stalk region

of the protein, including N38K, K40T, I427V, and H469Y, as well as L425M shared with naïve-vaccinated mice (**Table 3.2**). Group-specific mutations were also detected in the HA stalk region in naïve-vaccinated mice, such as K43N; however, three mutations (W150, L151P, and S183P) emerged in the receptor-binding domain of HA that were not found in naive-mock vaccinated animals (**Figure 3.5**; **Table 3.2**). Finally, preimmune-mock vaccinated mice shared only P182Q in the receptor-binding domain, and G350V in the protein stalk (**Figure 3.5; Table 3.2**).

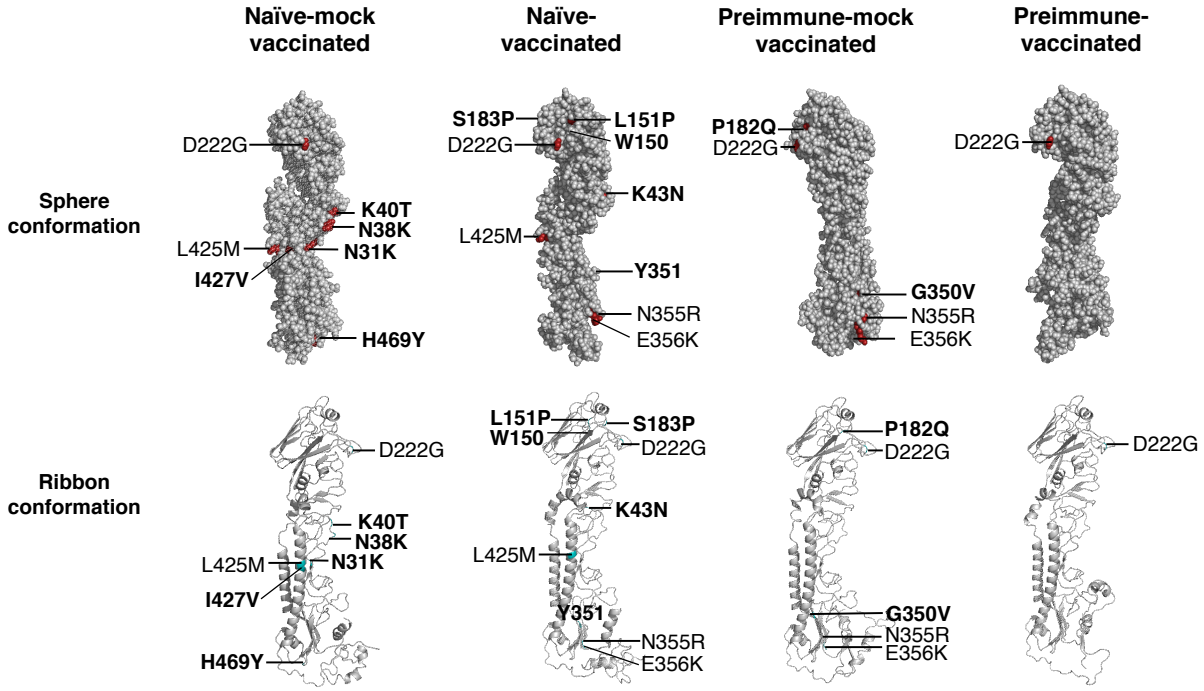


Figure 3.5 Folding models of immune background-specific SNPs on influenza virus hemagglutinin after Mex/09 challenge in mice show the highest group-specific SNPs in naïve-vaccinated animals.

At three days post-challenge with A/Mexico/4108/2009 (Mex/09), the presence of viral mutations on ten influenza virus genes was assessed by viral RNA extraction and sequencing. The HA gene sequences were aligned using MEGAX, and only SNPs found in most animals of an experimental group were retained in the sequence. To predict the possible structural and downstream antigenic impact of each immune background-specific amino substitution, a representative 3D model of each HA protein was generated using the Phyre2 platform (http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index). The protein images were generated in PyMol (The PyMOL Molecular Graphics System, Version 2.3.5, Schrödinger, LLC) to show both sphere (top) and ribbon (bottom) conformations. Only amino acid substitutions shared within

animals of a group were included in the protein folding predictions. Group-specific SNPs are highlighted in bold, and non-bolded text represents SNPs not unique to a group.

Considering that changes in protein shape can impact immune recognition by humoral immune defences, I next analyzed the potential effect of each amino acid substitution on HA protein structure (summarised in **Table 3.2**). Using the Phyre2-generated PDB file for the stock virus as the "wild-type" HA model, each amino acid substitution was input into Missense3D [189] manually to estimate structural damage. Here I found that several mutations found on HA across immune backgrounds had predicted structural damage with respect to protein folding (**Table 3.2**). In naïve-mock vaccinated mice, the H469Y mutation in the stalk region of HA led to contraction (>70Å) of the cavity found in this region. Also found in naïve-mock vaccinated mice, two substitutions in the F' subdomain (N31K, N38K) and one in the fusion subdomain (I427V) were each predicted to expand the cavity volume (**Table 3.2**). Of the nine substitutions found to be shared between naïve-vaccinated animals post-challenge, three had substantial predicted structural impacts; these were K43N in the vestigial esterase domain, which was predicted to cause a buried residue to become exposed; S183P in the receptor binding domain, which was predicted to cause a buried to exposed residue switch as well as disallowed psi/phi angles; and N355R, which lead to a cavity contraction < 70Å (**Table 3.2**) [215]. Finally, three of the five substitutions found in preimmune-mock vaccinated animals had reported structural damage. One of these fell in the receptor-binding domain (P182Q) and lead to a cavity contraction < 70Å. The remaining two substitutions occurred in the fusion domain; G350V lead to the replacement of a glycine residue, and N355R lead to cavity contraction < 70Å as in the naïve-vaccinated mice (**Table 3.2**).

Although shared mutations on HA were found within all four immune backgrounds, the number of shared mutations per immune background seems to follow the trend identified in the previous section relating to the number and frequency of SNPs. Even when examining non-

synonymous SNPs that are shared within animals of a group post-challenge, we see that the greatest number of mutations occurs in naïve-vaccinated and preimmune-mock vaccinated animals. The least mutations were observed when previous infection and vaccination are combined, or when an animal has no influenza virus exposure prior to challenge. Furthermore, the Missense3D prediction data suggested that at least some of these group-specific mutations may have a structural impact on the HA protein, which has implications on downstream protein function and antigenicity.

Table 3.2 Summary of immune background-specific single nucleotide polymorphisms on influenza virus hemagglutinin and potential structural damage predicted by Missense3D.

Group	SNP (H1 numbering)	Protein subdomain*	Missense3D prediction
Naïve-mock vaccinated	N31K	F' (N-term)	Expands cavity volume > 70Å
	N38K	F' (N-term)	Slightly expands cavity volume < 70Å
	K40T	F' (N-term)	None detected
	D222G	RB	Affects receptor binding
	L425M	F	None detected
	I427V	F	Expands cavity volume
	H469Y	F	Largely contracts cavity volume
Naïve-vaccinated	K43N	VE	Buried/exposed residue switch; slight cavity contraction
	W150	RB	NA
	L151P	RB	None detected
	S183P	RB	Disallowed psi/phi angles; buried/exposes residue switch; slight cavity contraction
	D222G	RB	Affects receptor binding
	Y351	F	NA
	N355R	F	Slight cavity contraction
	E356K	F	None detected
	L425M	F	None detected
Preimmune-mock vaccinated	P182Q	RB	Slight cavity contraction
	D222G	RB	Affects receptor binding (previously described)
	G350V	F	Replaces buried glycine residue
	N355R	F	Slight cavity contraction
	E356K	F	None detected
Preimmune-vaccinated	D222G	RB	Affects receptor binding (previously described)

*RB = receptor-binding domain; F = fusion domain; VE = vestigial esterase domain [215].

3.6 Immune background-specific mutations identified in the hemagglutinin protein alter predicted antigenicity

Amino acid mutations can lead to changes in antigenicity with respect to B cell epitopes and the specific peptides presented to host T cells. Specifically, T cell epitopes are linear peptides, consisting of approximately 8-11 amino acids for MHC class I, while B cell epitopes typically require a 3D confirmation of continuous or discontinuous amino acids for recognition [216]. Since both arms of the immune system are important for clearing a pathogen and establishing immune memory for future protection, I assessed the predicted B cell and T cell epitopes for the HA molecules sequenced from each immune background.

Starting with the group-specific HA folding models generated by Phyre2, I input each protein sequence into the DiscoTope 2.0 server from DTU Health Tech [192] to map the discontinuous B cell epitopes per immune background (**Figure 3.6**). For this analysis, I used the default DiscoTope threshold of -3.7 (0.47 sensitivity, 0.75 sensitivity), which is the least conservative threshold for selecting B cell epitopes. As shown in **Figure 3.6A**, peaks appearing above and below the prediction threshold indicate regions that are likely and unlikely to be B cell epitopes, respectively. The DiscoTope scoring along the amino acid sequence associated with each immune background suggests that HA, with respect to B cell epitopes, had potential antigenic changes at three days post-challenge. As shown in **Figure 3.6B**, there are marked differences in the predicted B cell epitopes, both in the "size" of the epitope (i.e., the number of amino acids included in an epitope), and in the DiscoTope score (i.e., propensity to be a B cell epitope) when compared to the reference Mex/09 HA sequence. Differences in predicted B cell epitopes were evident in both the head (amino acids 42-272) and stalk (331-552) regions of the protein (**Figure 3.6C**). Specifically, the epitope appearing at residues 137-147 in the receptor-binding domain of HA scores more highly in the naïve-vaccinated group compared to the Mex/09 reference or the other immune backgrounds (**Figure 3.6B**). Other regions of the RBD, specifically

at residues 197-215 and 227-241, had increased DiscoTope scores in naïve-vaccinated mice compared to the other sequences (**Figure 3.6B**). Another small epitope occurring at residue 261 was also apparent in HA from naïve-vaccinated animals only (**Figure 3.6B**). Other predicted epitope changes occurred outside of the RBD of HA. The predicted epitope at residues 405-416 in the fusion subdomain was larger with respect to the number of amino acids included in the epitope in naïve-vaccinated, preimmune-mock vaccinated, and preimmune-vaccinated groups. Another predicted epitope region in the fusion subdomain, specifically at residues 496-505, was expanded in the preimmune-vaccinated group (**Figure 3.6B**). Taken together, my analysis of predicted B cell epitopes using DiscoTope showed that immune background-specific mutations occurring on the HA protein after challenge may lead to antigenic differences. These differences were most prominent in the HA of naïve-vaccinated animals and occurred in both the globular head and stalk regions of HA.

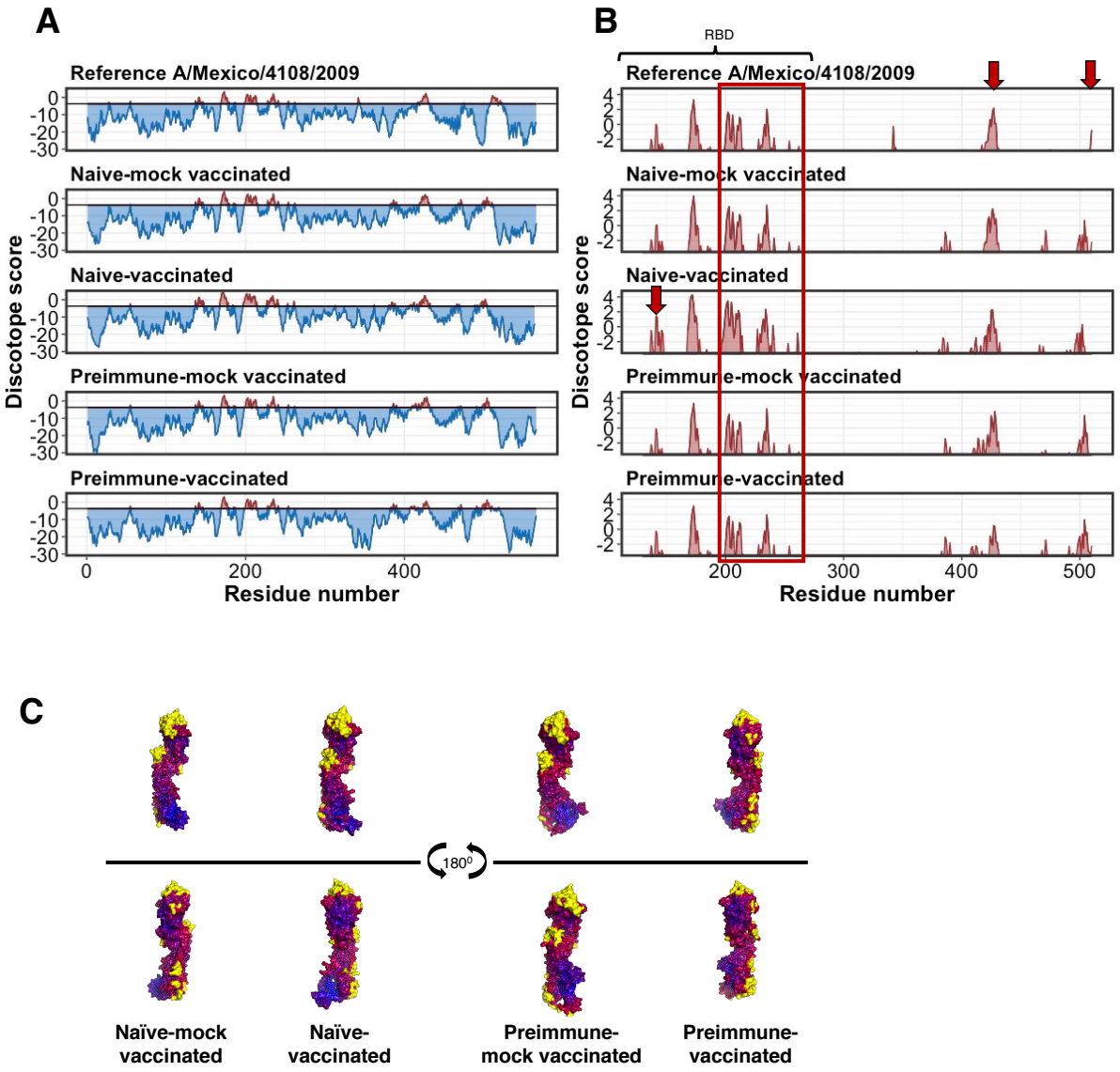


Figure 3.6 Influenza virus hemagglutinin proteins show changes in predicted B cell epitopes after viral replication in mice with varying immune backgrounds.

Immune memory toward influenza viruses was established in mice by differentially infecting or vaccinating (preimmunity) to generate naïve-mock vaccinated, naïve-vaccinated, preimmune-mock vaccinated, and preimmune-vaccinated animals. All mice were then subject to challenge with a lethal dose of Mex/09 and lung viral RNA was extracted. The RNA isolated at three days post-challenge was sequenced on the Illumina platform. Folding models of HA containing immune background-specific mutations were generated with the Phyre2 platform (http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index). Surface epitopes along HA were predicted using DiscoTope 2.0 (DTU Health Tech) (http://www.cbs.dtu.dk/services/DiscoTope/_). (A) DiscoTope scores falling below (blue) and above (red) the B cell epitope prediction threshold of -3.7 (0.47 sensitivity, 0.75 specificity) were mapped against amino acid sequence number for reference Mex/09 (NCBI accession no. GQ223112) and HA extracted from mice of varying immune backgrounds post-challenge. (B) Only the positive B cell epitope prediction results are shown for greater resolution of the B cell epitope regions. Arrows highlight notable differences in

DiscoTope scores compared to reference Mex/09, which fall within and outside of the RBD of HA. C) Representative folding models of the HA protein showing predicted B cell epitopes after challenge in mice of different immune backgrounds were made in PyMol (The PyMOL Molecular Graphics System, Version 2.3.5, Schrödinger, LLC). DiscoTope scores are shown as heatmaps along the folded protein, with residues colored according to their predicted score: yellow indicates positively predicted B cell epitopes (scores > -3.7 threshold), red indicates high-scoring amino acids, and blue indicates low scoring regions (i.e., unlikely B cell epitopes). Figures A and B were generated using the R ggplot2 package version 3.3.2.

Finally, I evaluated whether immune background-specific mutations occurred in regions of HA that were likely to be presented on MHC class I molecules in the mouse model, thus interacting with specific CD8+ T cell receptors. Using IEDB NetMHCpan EL 4.1, the highest scoring peptides were predicted to be on H2-Db and H2-Kb alleles that were specific to at least one immune background at three days post-challenge (summarized in **Table 3.3**). This analysis found several high-affinity, immune-background specific peptides across the HA protein that were not found in the stock Mex/09 virus. For example, the peptide LVLLENERTL from amino acids 443-442 (stalk region) was specific to naïve-mock vaccinated animals, while the peptide NAEMLILLENERTL (437-450) spanning the same region of HA scored more highly and was specific to naïve-vaccinated animals (**Table 3.3**). Additionally, the viral sequences emerging from naïve-vaccinated animals contained several novel peptides in the RBD of HA (specifically 157-168) that were not identified in the other immune backgrounds (**Table 3.3)**.

Table 3.3 High-affinity MHC class I-binding peptides derived from immune background-specific hemagglutinin protein sequences using NetMHCpan EL 4.1.

Group	Allele	Location	Peptide	Score	Percentile rank
Stock	H-2-Db	138-149	SSWPNHDSNKGV	0.223368	0.44
Naïve-mock vaccinated	H-2-Db	443-452	LVLLENERTL	0.13495	0.94
		439-452	NAEMLVLLENERTL	0.16571	0.7
		41-49	VTVTHSVKL	0.165951	0.7
		438-446	YNAEMLVLL	0.181361	0.6
		207-217	SIYQNADAYVF	0.201008	0.51
		207-215	SIYQNADAY	0.269355	0.33
		207-216	SIYQNADAYV	0.389948	0.17
		444-452	VLLENERTL	0.684796	0.04
	H-2-Kb	438-446	YNAEMLVLL	0.309654	0.81
		444-452	VLLENERTL	0.382233	0.55
		43-50	VTHSVKLL	0.435597	0.41
		41-49	VTVTHSVKL	0.794774	0.06
Naïve-vaccinated	H-2-Db	436-444	YNAEMLILL	0.150642	0.82
		437-450	NAEMLILLENERTL	0.16571	0.7
		160-168	SFYRNLIPV	0.323185	0.23
	H-2-Kb	436-444	YNAEMLILL	0.277149	0.95
		160-168	SFYRNLIPV	0.294991	0.87
		158-165	AKSFYRNL	0.430483	0.43
		156-165	AGAKSFYRNL	0.486693	0.34
		159-166	KSFYRNLI	0.531279	0.26
		157-165	GAKSFYRNL	0.795388	0.06
Preimmune-mock vaccinated	H-2-Db	359-367	TGMVDGWYV	0.498793	0.1
Naïve-mock vaccinated + naïve-vaccinated	H-2-Kb	434-442/432-440	DIWTYNAEM	0.255978	1.1
		435-442/433-440	IWTYNAEM	0.240496	1.2
		435-443/433-441	IWTYNAEML	0.105114	3
		436-443/434-441	WTYNAEML	0.290618	0.9
	H-2-Db	435-443/433-441	IWTYNAEML	0.035329	3.4
		436-443/434-441	WTYNAEML	0.077136	1.7
Stock + preimmune-mock vaccinated + preimmune vaccinated	H-2-Kb	435-442	IWTYNAEL	0.337446	0.68
		434-442	DIWTYNAEL	0.374151	0.57
	H-2-Db	439-452	NAELLILLENERTL	0.16571	0.7

Overall, the greatest number of high-affinity, group-specific peptides were found in viral sequences isolated from naïve-mock vaccinated animals, with the least found from both preimmune groups. The presence of group-specific peptides that were likely to be presented on MHC class I suggested immune background-specific activation of CD8+ T lymphocytes and the resulting variant viruses that emerge or are propagated in the lung have a different antigenic signature.

Together with my B cell epitope analysis, I found that specific immune backgrounds were associated with unique amino acid changes affecting predicted B cell and T cell epitopes on the virus. These results suggest that the immune pressure developed by specific previous encounters with influenza viral antigens, whether through previous infection or vaccination, can select for novel influenza virus variants.

3.7 Immune background-specific mutations occur on influenza viral proteins nucleoprotein (NP), neuraminidase (NA), and polymerase basic 1 (PB1)

Although HA is the immunodominant viral protein, other proteins on the surface as well as internal to the virion can contribute to virus antigenicity [217]. These viral proteins can be recognized by homologous and cross-reactive CD8+ T cells and antibodies during a subsequent encounter with the same or different influenza virus [173,105,218–220,103]. Because these proteins are potential antigenic targets at an influenza virus infection, I hypothesized that they may be subject to immune background-driven selection of mutations when the host is infected. To this end, I aimed to determine if the mutations acquired on other influenza viral proteins affected protein structure and antigenicity, as I had done for the HA protein in the previous section. My analysis focused on the nucleocapsid protein (NP), which is internal and highly conserved; neuraminidase (NA), an external major antigen; and polymerase basic 1 (PB1), a highly conserved polymerase subunit [9].

As in my previous analysis of HA mutations, I first mapped the NP, NA, and PB1 sequences extracted at three days post-challenge to Mex/09 reference segments in Geneious (described above). I then used the SNPs and variants tool in Geneious to identify all mutations occurring above 1% frequency. Next, all non-synonymous SNPs found in NP, NA, and PB1 were plotted along the amino acid sequence number of each protein to determine patterns across groups (**Figure 3.7**). With respect to the number of SNPs detected across immune backgrounds, NP showed a similar trend to HA, with the greatest number of SNPs found in naïve-vaccinated animals, and the least in preimmune-vaccinated animals (**Figure 3.7**). In contrast, NA showed the most detectable mutations in naïve-mock vaccinated animals, and PB1 mutations were most dense in naïve-vaccinated and preimmune-mock vaccinated groups (**Figure 3.7**).

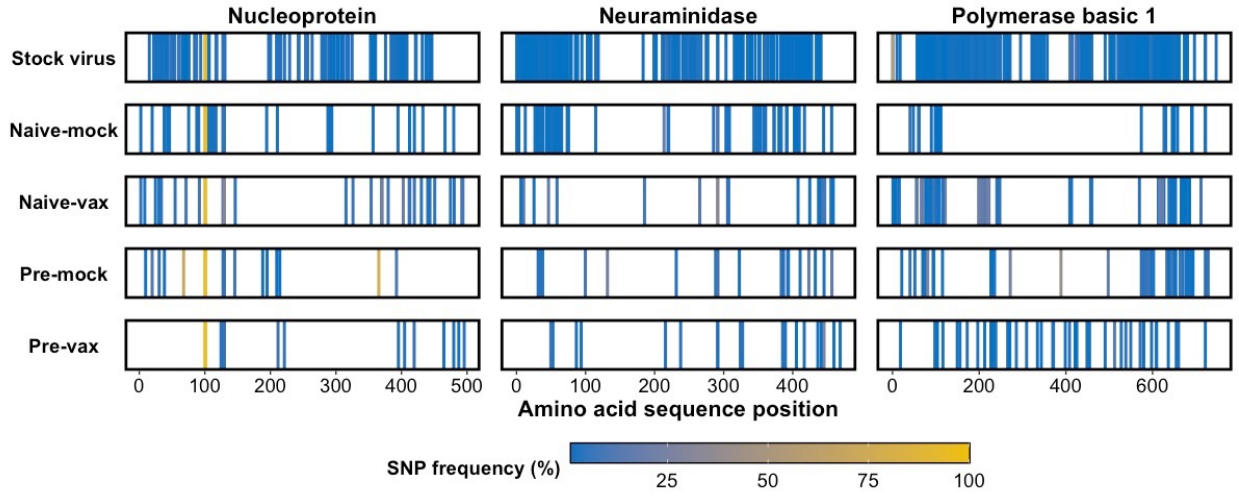


Figure 3.7 Location and frequency of SNPs on influenza virus nucleoprotein (NP), neuraminidase (NA), and polymerase basic 1 (PB1) amino acid sequences after Mex/09 challenge in mice of different immune backgrounds.

At three days post-challenge with Mex/09, viral RNA from infected mouse lungs was sequenced using Illumina MiSeq and aligned to reference Mex/09 NP, NA, and PB1 using Geneious (NCBI accession numbers GQ149655 (NP), GQ149650 (NA), and GQ149652 (PB1)). The SNPs and variants above 1% frequency were detected for all influenza virus genes. Stock Mex/09 used to inoculate animals was sequenced in parallel to determine the existing sequence variation at input. The non-synonymous SNPs found in NP, NA, and PB1 gene sequences after introduction to naïve-mock vaccinated, naïve-vaccinated, preimmune-mock vaccinated, and preimmune-

vaccinated mice were plotted against the amino acid sequence number (1 band = 1 SNP). A schematic of each protein is shown where a line represents an amino acid change, generating a bar-code effect for each SNP profile. The color of the band represents the relative frequency of each mutation among the total population of viruses sequenced. The figure was generated using the ggplot2 package version 3.3.2.

To further determine which mutations may have arisen from immune-background selection at challenge, I filtered the total mutations on NP, NA, and PB1 shown in **Figure 3.7** to include only those shared among animals in a group (**Figure 3.8**) and assessed each of these mutations for structural impact using Missense3D (described above) (**Table 3.4**). In NP, as with HA, I noted the greatest number of mutations in naïve-vaccinated animals (**Table 3.4**). These were R31G, D101G, D128E, and deletion of arginine at residue 441 (**Table 3.4**). Viral NP sequences from all four immune backgrounds contained a T130K mutation in the body region of the protein [221], which was predicted to largely expand the volume of the cavity found in this region (**Table 3.4**). Moreover, the D to E change at amino acid 128 in NP was shared among all groups, except for preimmune-vaccinated mice (**Figure 3.8**). No structural modifications were identified from D128E by Missense3D (**Table 3.4**).

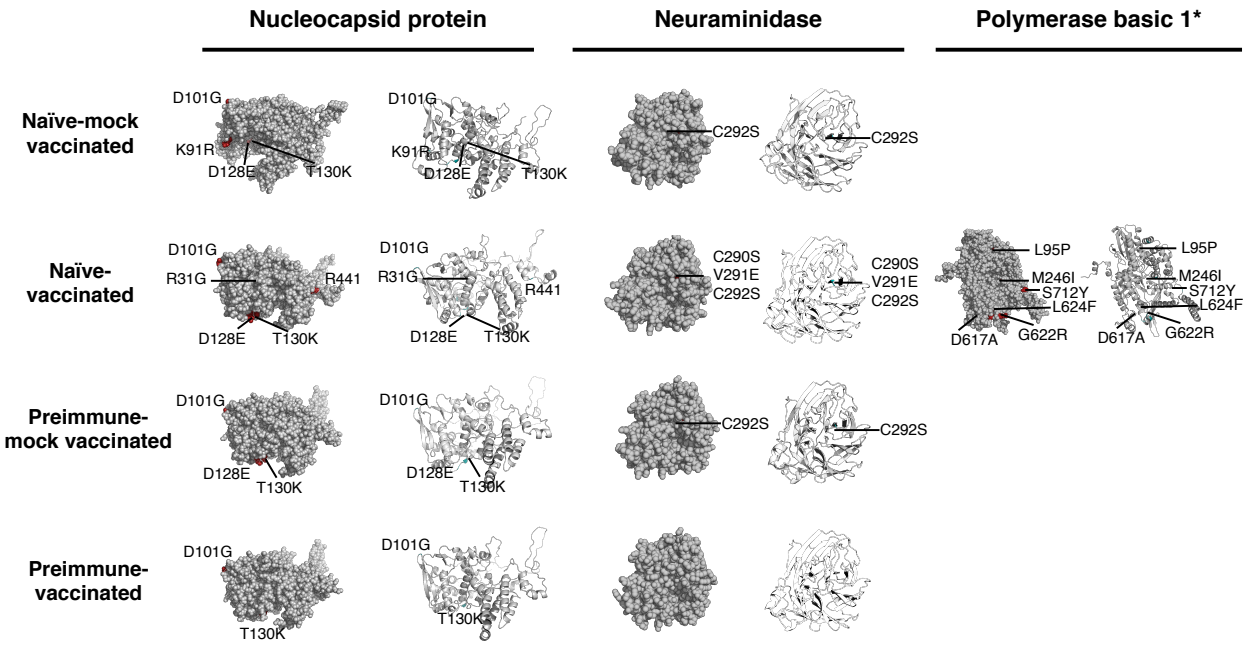


Figure 3.8 Influenza virus nucleoprotein (NP), neuraminidase (NA), and polymerase basic 1 (PB1) show immune background-specific mutations at Mex/09 challenge.

At three days post-challenge with A/Mexico/4108/2009 (Mex/09), the presence of viral mutations on ten influenza virus genes was assessed by viral RNA extraction and sequencing. Viral sequences were aligned to reference Mex/09 (accession no. GQ149655 (NP), GQ149650 (NA), and GQ149652 (PB1)) and variants above 1% frequency were detected. Only SNPs found in most animals of an experimental group were retained in the amino acid sequence for downstream analysis. To predict the possible structural and downstream antigenic impact of each immune background-specific amino substitution, a representative 3D model of each NP, NA, and PB1 for each immune background was generated using the Phyre2 platform (http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index). The protein images were generated in PyMol (The PyMOL Molecular Graphics System, Version 2.3.5, Schrödinger, LLC) to show both sphere (left) and ribbon (right) conformations. Only amino acid substitutions shared within animals of a group were used to generate folding predictions. *The transmembrane (stalk) domain of neuraminidase was not included in the model.

Table 3.4 Summary of immune-background specific mutations detected on influenza viral nucleoprotein (NP), neuraminidase (NA), and polymerase basic 1 (PB1) and predicted structural impact at three days post-challenge.

Nucleoprotein (NP)			
Group	**SNP**	**Protein Subdomain**	**Missense3D prediction**
Naïve-mock vaccinated	D101G	Body	None; present in stock virus
	K91R	Body	Slightly contracts cavity volume
	D128E	Body	None detected
	T130K	Body	Largely expands cavity volume
Naïve-vaccinated	R31G	Body	Slight cavity expansion
	D101G	Body	*
	D128E	Body	None detected
	T130K	Body	Largely expands cavity volume
	R441	Bottom/tail region	NA
Preimmune-mock vaccinated	D101G	Body	*
	D128E	Body	None detected
	T130K	Body	Largely expands cavity volume
Preimmune-vaccinated	D101G	Body	*
	T130K	Body	Largely expands cavity volume
Neuraminidase (NA)			
Group	**SNP (N1 Numbering)**	**Protein Subdomain**	**Missense3D prediction**
Naïve-mock vaccinated	**E47G**	Stalk	Contract cavity volume
	Q51E	Stalk	Expands cavity volume

Neuraminidase (NA)			
Group	**SNP (N1 Numbering)**	**Protein Subdomain**	**Missense3D prediction**
Naïve-mock vaccinated	**C292S**	Head	Disulfide bond breakage
	S350	Head	Deletion
Naïve-vaccinated	**C290S**	Head	Disulfide bond breakage and expands cavity volume
	V291E	Head	Introduces buried hydrophilic and charged residue and expands cavity volume
Preimmune-mock vaccinated	**C292S**	Head	Disulfide bond breakage
	I38T	Stalk	Contract cavity volume
	C292S	Head	Disulfide bond breakage
Polymerase basic 1 (PB1)			
Group	**SNP**	**Protein Subdomain**	**Missense3D prediction**
Naïve-vaccinated	L95P	cRNA promoter binding site	Buried proline introduced, cavity altered and buried/exposed switch; may affect the cRNA promoter binding
	M246I	Putative nucleotide binding site	No structural damage detected
	D617A	vRNA promoter binding site (C-term)	Salt bridge between ASP 617 and ARG 623 disrupted; may inhibit vRNA promoter binding
	G622R	vRNA promoter binding site (C-term)	Exposed charge introduced; may inhibit vRNA promoter binding
	L624F	vRNA promoter binding site (C-term)	May inhibit vRNA promoter binding
	S712Y	C-term (Core interaction with PB2)	Expand cavity volume by 4.968 Å^3

I noted several group-specific mutations across NA, including E47G and Q51E (N1 numbering) located in the stalk domain of sequences from naïve-mock vaccinated and challenged mice (**Table 3.4**). Additionally, a V291E mutation was identified in the head domain of NA

extracted from naïve-vaccinated animals, and I38T in the stalk domain of NA from preimmune-mock vaccinated animals (**Figure 3.8**; **Table 3.4**). As with HA, I detected the greatest number of group-specific mutations on NA from naive-vaccinated mice (**Figure 3.8**), where the number of shared mutations within a group was not associated with the total number of SNPs found in that group (**Figure 3.7**). Apart from preimmune-vaccinated mice, all groups shared NA sequences with a C to S substitution at amino acid 292 in the head domain. When examining structural impacts of each mutation using Missense3D, I found that certain group-specific substitutions on NA lead to possible structural damage (**Table 3.4**). For example, C292S present in all immune backgrounds was predicted to induce disulfide bond breakage; V291E in naïve-vaccinated mice would introduce buried hydrophilic and charged residues and greatly expand the predicted cavity volume; and I38T from the preimmune-mock vaccinated group would contract the NA stalk domain cavity (**Table 3.4**).

Interestingly, only PB1 extracted from preimmune-vaccinated mice had mutations evenly distributed across the entire sequence. As shown in **Figure 3.7**, PB1 sequences isolated from other immune backgrounds generally had most mutations concentrated at the 3' and 5' region of the PB1 sequence, and preimmune-vaccinated mice viral sequences lacked mutations in this region. When accounting only for mutations shared between animals of a particular immune-background, mutations on PB1 were only detected in the naïve-vaccinated animals (**Figure 3.8**). These mutations occurred in several functional domains of the protein, including L95P in the cRNA promoter binding site; M246I in the putative nucleotide binding site; and D617A in the vRNA promoter binding site (**Figure 3.8**; **Table 3.4**) [222]. Prediction of the amino acid substitution effect on protein structure using Missense3D determined that PB1 mutations induced structural alterations to the protein that may impact function. For example, L95P introduces a buried proline and causes a switch of buried/exposed residues, and D617A disrupts the salt bridge between ASP 617 and ARG 623 (**Table 3.4**).

Since the amino acid changes identified in NP, NA, and PB1 may impact protein structure and therefore recognition by the host immune system, I next aimed to determine whether the background-specific viral mutations would impact antigenicity with respect to B cell epitopes and MHC I peptide presentation (**Figure 3.9**). Using scores derived from DiscoTope, I plotted the B cell epitope score along the variant amino acid sequences for NP, NA, and PB1 for each immune background (**Figure 3.9, Figure 3.10**). For NP, there were clear differences in the height and width of peaks corresponding to predicted B cell epitopes in naïve-vaccinated and preimmune-mock vaccinated hosts, specifically at residues 0-25 and 450-500 (**Figure 3.9**). Differences in predicted B cell epitopes were not observed for NA, aside from an additional peak found in viruses from naïve-vaccinated, preimmune-mock vaccinated, and preimmune-vaccinated groups at the far 5' end of the NA sequence, which likely corresponds to the transmembrane region (**Figure 3.9**). I did not observe any differences in predicted B cell epitopes for PB1 (**Figure 3.9**).

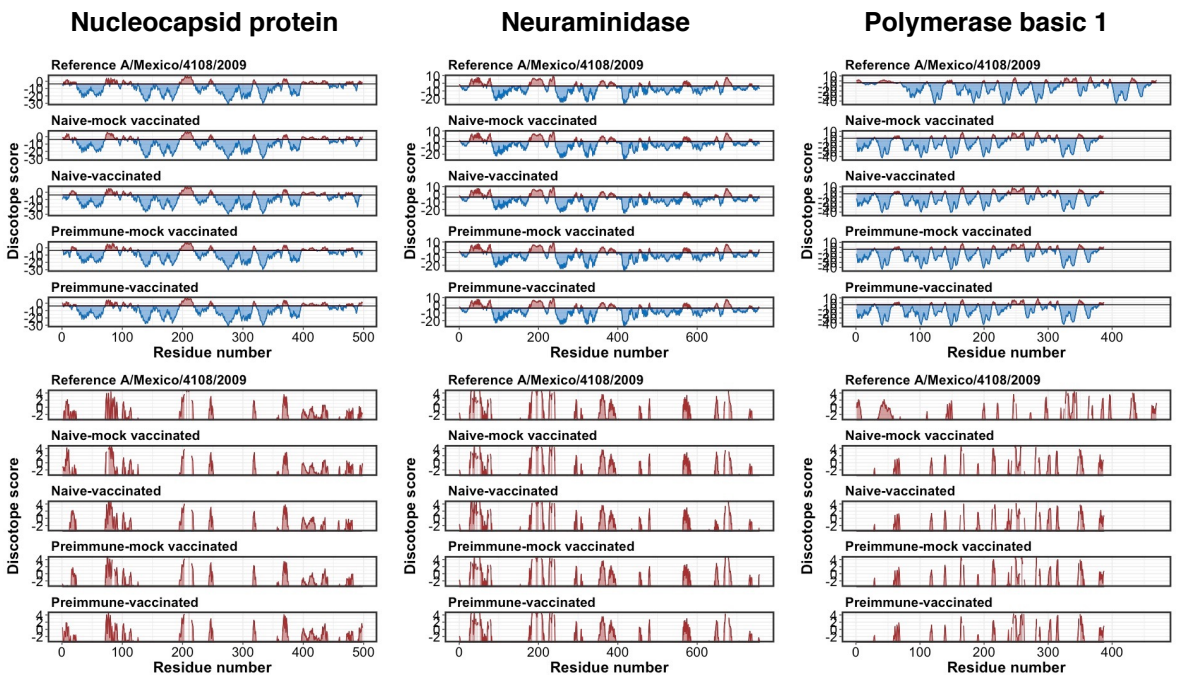


Figure 3.9 Predicted B cell epitopes on influenza virus nucleoprotein (NP), neuraminidase (NA), and polymerase basic 1 (PB1) after Mex/09 challenge in mice with varying immune backgrounds show minimal differences between groups.

At three days post-challenge with A/Mexico/4108/2009 (Mex/09), the presence of viral mutations on ten influenza virus genes was assessed by viral RNA extraction and sequencing. The nucleoprotein (NP), neuraminidase (NA), and polymerase basic 1 (PB1) gene sequences were aligned using MEGAX, and only SNPs found in most animals of an experimental group were retained in the sequence. To predict the possible structural and downstream antigenic impact of each immune background-specific substitution, a representative 3D model of each protein was generated using the Phyre2 platform (http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index). Surface epitopes along NP, NA, and PB1 were predicted using DiscoTope 2.0 (DTU Health Tech) (http://www.cbs.dtu.dk/services/DiscoTope/_). (A) DiscoTope scores falling below (blue) and above (red) the B cell epitope prediction threshold of -3.7 (0.47 sensitivity, 0.75 specificity) were mapped against amino acid sequence number for NP, NA, and PB1 to show regions of likely and unlikely B cell epitopes. Epitope maps for each immune background are compared to that of the respective Mex/09 reference protein, which are as follows: NP (GQ149655), NA (GQ149650), and PB1 (GQ149652). (B) Only the positive B cell epitope prediction results are shown for greater resolution of the B cell epitope regions. Arrows highlight any notable differences in DiscoTope scores compared to reference Mex/09. All figures were generated in using the ggplot2 package version 3.3.2.

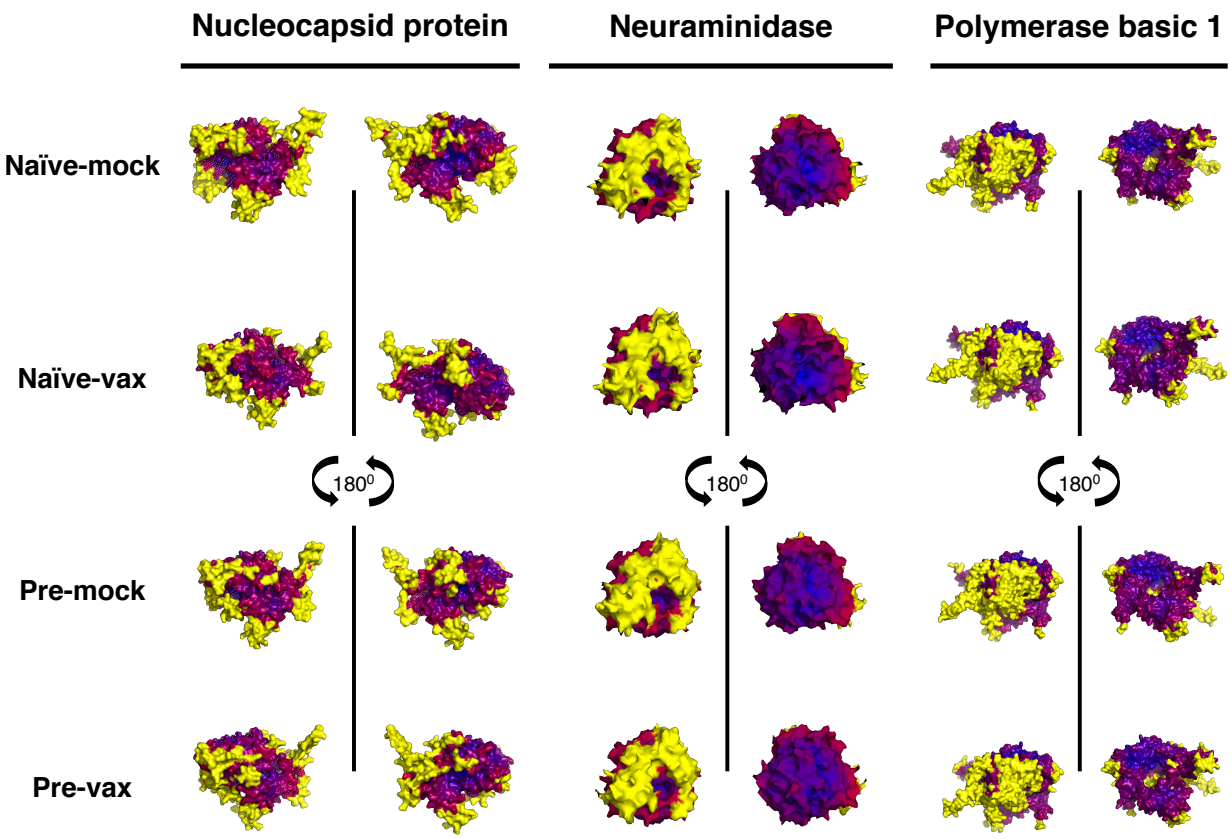


Figure 3.10 Structural models of predicted B cell epitopes for influenza virus nucleoprotein (NP), neuraminidase (NA), and polymerase basic 1 (PB1) after challenge in mice of varying immune backgrounds show minimal differences in predicted epitopes.

Amino acid substitutions specific to immune background were assessed by aligning viral sequences to reference Mex/09 (NCBI accession numbers as follows: NP (GQ149655), NA (GQ149650), and PB1 (GQ149652)) and detecting variants above 1% frequency. Folding models of NP, NA, and PB1 were generated using the Phyre2 platform (http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index), with only amino acid substitutions shared within animals of a group used to generate folding predictions for each viral protein. Surface epitopes along NP, NA, and PB1 were predicted using DiscoTope 2.0 (DTU Health Tech) (http://www.cbs.dtu.dk/services/DiscoTope/) with a non-conservative prediction threshold of -3.7 (0.47 sensitivity, 0.75 specificity). DiscoTope scores were mapped against amino acid sequence number of reference Mex/09 NP, NA, and PB1. DiscoTope scores are shown as heatmaps along the folded protein, with residues colored according to their predicted score: yellow indicates positively predicted B cell epitopes (scores > -3.7 threshold), red indicates high-scoring residues, and blue indicates low scoring regions (i.e., unlikely B cell epitopes). The protein images were generated in PyMol (The PyMOL Molecular Graphics System, Version 2.3.5, Schrödinger, LLC).

As with my analysis of HA, I then used IEDB NetMHCpan to determine changes in CD8+ T cell receptor recognition of NP, NA, and PB1 by analyzing the highest-scoring peptides presented on H2-Db and H2-Kb alleles. I found several group-specific peptides across NP, NA, and PB1, suggesting that variant amino acid sequences may be presented on MHC I molecules and differentially activate CD8+ T cells, giving the variant proteins a unique antigenic presentation.

3.8 Transcriptomic analysis of host immunity in the lung shows polarizing immune responses per immune background that may drive differential viral mutation at challenge

To parallel my analysis of influenza viral protein mutations at three days post-challenge, I characterized the host immune response in the lungs at this time point to identify the immune pressure exerted by each pre-existing immune background. I hypothesized that immune responses defined by transcriptomic analysis may give insight into the origins of the mutations noted in the viral proteins. To this end, RNA sequencing using the Illlumina NovaSeq 6000 Sequencing System at Novogene (Sacramento, CA, USA) was performed on host lung RNA extracted three days post-challenge to sequence the transcriptome. Like the viral sequences, host lung transcripts were uploaded to the SRA under BioProject ID PRJNA787976 (https://www.ncbi.nlm.nih.gov/biosample?Db=biosample&DbFrom=bioproject&Cmd=Link&LinkName=bioproject_biosample&LinkReadableName=BioSample&ordinalpos=1&IdsFromResult=787976, see **Appendix 1**). Lung gene expression was quantified in terms of fragments per kilobases mapped (FPKM) to *Mus musculus* mm10 exons to calculate gene expression fold-change. I initially compared the significant (p-value < 0.05) differentially expressed genes (DEGs) between the four immune backgrounds to assess broad similarities and/or differences in the immune response profiles as per numbers of shared genes expressed (**Figure 3.11A**). At three days post-challenge, a core 625 genes were significantly up- or down-regulated, regardless of immune

background, as shown in the Venn diagram (**Figure 3.11A**). Mice of each immune background had significant DEGs that were not shared with the other groups, with 227 DEGs in naïve-mock vaccinated, 1822 in naïve-vaccinated, 305 in preimmune-mock vaccinated, and 872 in preimmune-vaccinated mice (**Figure 3.11A**).

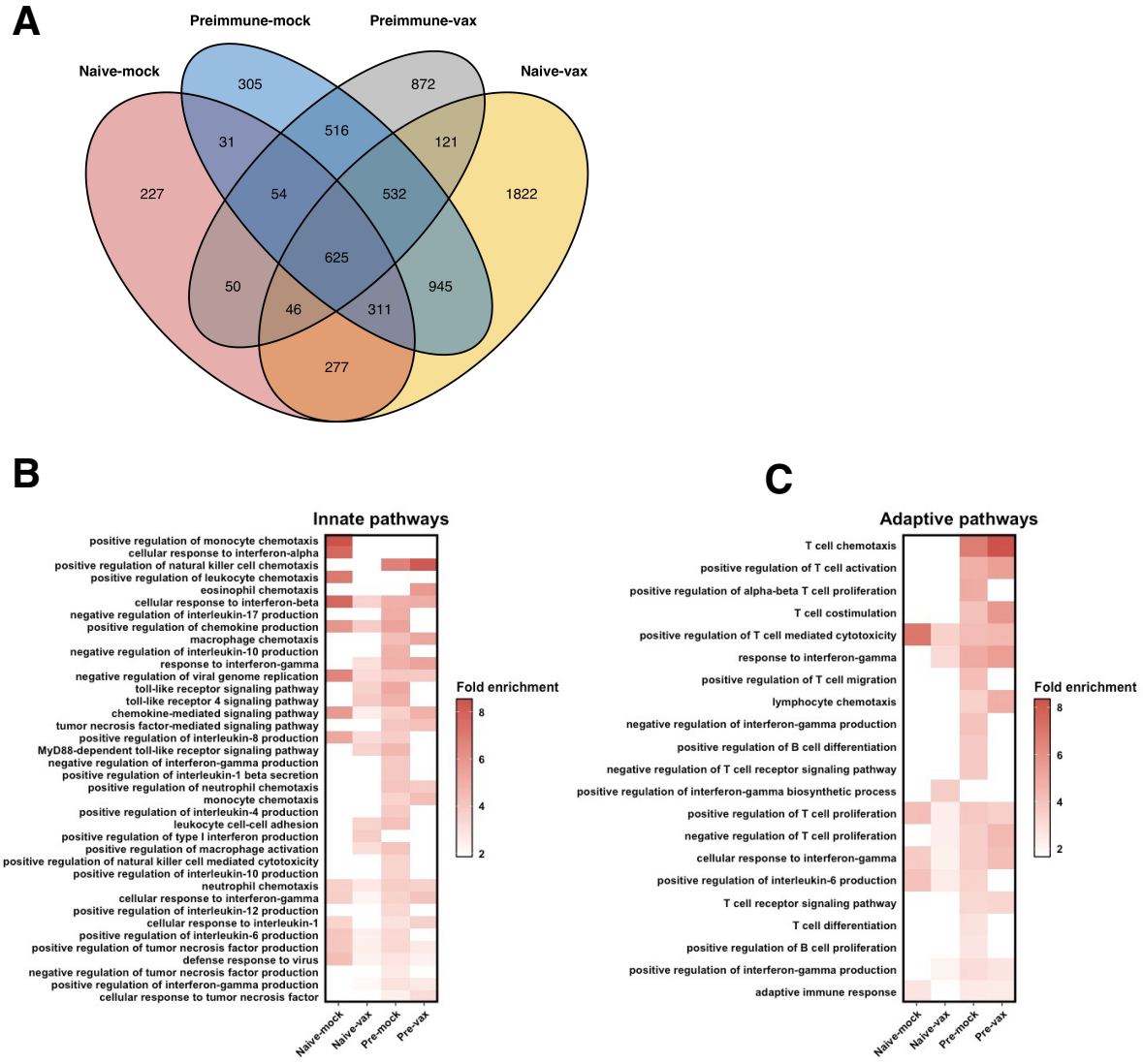


Figure 3.11 Lungs of mice of different immune backgrounds have different gene enrichment profiles at three days post-Mex/09 challenge.

Adult C57Bl/6j mice of varying immune backgrounds (naïve-mock vaccinated, naïve-vaccinated, preimmune-mock vaccinated, and preimmune-vaccinated) were infected with a lethal dose of influenza virus A/Mexico/4108/2009 (Mex/09). At three days post-inoculation, total RNA was extracted from mouse lungs and the host transcripts were sequenced using the Illumina platform. Host gene expression was quantified as fragments per kilobases mapped (fpkm), and the $\log_2$(fold

change) was calculated relative to non-infected mice. Differentially expressed genes for each immune background were calculated by analysis of variance comparing the fold change to non-infected control lung RNA (significance threshold $p < 0.05$). (A) Lists of significant differentially regulated genes were compared across immune backgrounds. Similarities and differences in gene regulation between groups are shown in a Venn diagram, indicating that the naïve-vaccinated animals have the most non-shared differentially regulated genes. The figure was generated in R using the VennDiagram package version 1.6.20. (B) The lists of differentially regulated genes for each immune background were uploaded to The Database for Annotation, Visualization and Integrated Discovery (DAVID) version 6.8 functional annotation tool (https://david.ncifcrf.gov/home.jsp). Lists of enriched gene ontology (GO) biological processes were downloaded for each immune background, and the fold enrichment metric calculated by DAVID was plotted for each pathway. GO biological processes were loosely separated into "innate" or (C) "adaptive" immune pathways to more easily compare gene enrichment profiles across immune backgrounds. The figure was generated using ggplot2 version 3.3.2.

Overall, differential gene expression at challenge was least observed in naïve-mock vaccinated animals, with only 1621 total DEGs detected, compared to 3000-4600 DEGs in the other groups (**Figure 3.11A**). However, the majority of these DEGs (78%) were also differentially expressed in naïve-vaccinated animals, indicating a high degree of similarity in the immune response profiles, despite a large difference in the number of DEGs overall. In contrast, only 48% of the DEGs found in naïve-mock vaccinated animals were also found in preimmune-vaccinated mice, pointing to a core innate immune response shared among the animals. Naïve-vaccinated mice had the greatest differential gene expression at challenge, with 4680 genes differentially regulated relative to the baseline. When comparing the DEGs present in naïve-vaccinated animals to those found in the other immune backgrounds, naïve-vaccinated mice were most similar to preimmune-mock vaccinated animals, with 2413 (52%) of the DEGs shared between both groups (**Figure 3.11A**).

Next, I input the DEG lists from each experimental group into The Database for Annotation, Visualization, and Integrated Discovery (DAVID, https://david.ncifcrf.gov/home.jsp). To characterize the dominant immune response pathways of each group at challenge, I used the functional annotation tool to perform enrichment analysis. I looked specifically at gene ontology

(GO) biological processes that were enriched for each group, categorizing each pathway as either an innate (**Figure 3.11B**) or adaptive (**Figure 3.11C)** immune response. With respect to enriched immune pathways at challenge, I noted distinct immune response profiles for each of the four immune backgrounds. With respect to innate signalling, naïve-mock vaccinated mice had a strong proinflammatory and antiviral response at challenge, highly enriched by monocyte chemotaxis, chemokine (e.g., IL-8) signalling, proinflammatory signaling (e.g., IL-1beta and IL-6), and type I and type II interferon response genes (**Figure 3.11B**). Similar to naïve-mock vaccinated animals, naïve-vaccinated mice had enrichment of antiviral (type 1 interferons) and inflammatory response genes and pathways (IL-8, IL-6, and TNF), marked by chemotaxis and cell adhesion (**Figure 3.11B**). However, naïve-vaccinated mice were also enriched in toll-like receptor (TLR) signalling and MyD88-dependent TLR signalling pathways, which were not observed in naïve-mock vaccinated animals. Preimmune mice had several commonalities as well as key differences in innate immune involvement at challenge compared to naïve-mock and naïve-vaccinated groups. As found in the naïve animals, I noted enrichment of IFN-beta and -gamma production, leukocyte chemotaxis and cell-cell adhesion, macrophage activation, and TNF production in the preimmune animals (**Figure 3.11B**). Like the naïve-vaccinated mice, TLR signalling and MyD88-dependent TLR signalling transcripts were enriched in the preimmune-vaccinated animals.

Differences were also found between the innate immune responses of the preimmune groups. Additional enriched pathways such as positive regulation of natural killer (NK) cell activation, IL-1beta, and IL-17 were noted, with concomitant negative regulation of IL-10 found only in preimmune-mock vaccinated mice (**Figure 3.11B**). Interestingly, preimmune-vaccinated mice displayed enrichment profiles that were distinct from the other groups and marked by monocyte, macrophage, neutrophil, and eosinophil chemotaxis gene enrichment (**Figure 3.11B**). Preimmune-vaccinated mice had IFN-beta enrichment at a similar magnitude to the preimmune-mock vaccinated mice. However, the IFN-gamma and proinflammatory cytokines IL-1 and TNF

as well as gene for NK cell chemotaxis were more highly enriched in the preimmune-vaccinated animals compared to the naive-vaccinated mice (**Figure 3.11B**).

Enrichment of genes for adaptive immune pathways at challenge was minimal in naïve-mock vaccinated and naive-vaccinated animals (**Figure 3.11C**). This involvement was restricted to positive regulation of T cell-mediated cytotoxicity and T cell proliferation transcripts, which was most prominent in the naive-mock vaccinated animals and only moderately enriched in the naive-vaccinated mice. In contrast, preimmune-mock vaccinated mice had widespread involvement of adaptive immunity (**Figure 3.11C**) with notable enrichment of lymphocyte chemotaxis, B and T cell differentiation and proliferation, and T cell-mediated cytotoxicity. Additionally, I found genes for cytokines regulating adaptive immune responses such as IL-12 and IL-4 to also be enriched in preimmune-mock vaccinated mice (**Figure 3.11B** and **3.11C**). Preimmune-vaccinated mice had adaptive gene regulation that was less diverse than the other experimental groups, but more highly enriched than preimmune-mock vaccinated mice (**Figure 3.11C**). Specifically, preimmune-vaccinated mice had strong enrichment of lymphocyte chemotaxis, positive regulation of T cell activation (TCR signaling and co-stimulation) and cytotoxicity, and regulation of T cell proliferation (**Figure 3.11C**). My analysis did not detect enrichment of pathways regulating B cell differentiation and proliferation in preimmune-vaccinated mice.

Finally, I aimed to further describe the specific innate and adaptive immune responses in each group at challenge by comparing expression of genes involved in antiviral responses (**Figure 3.12A**), B cell-mediated immunity (**Figure 3.12B**), T cell-mediated cytotoxicity, and activated T cell proliferation (**Figure 3.13**). As with my enrichment analysis, I observed four unique immune response gene profiles across each of these pathways.

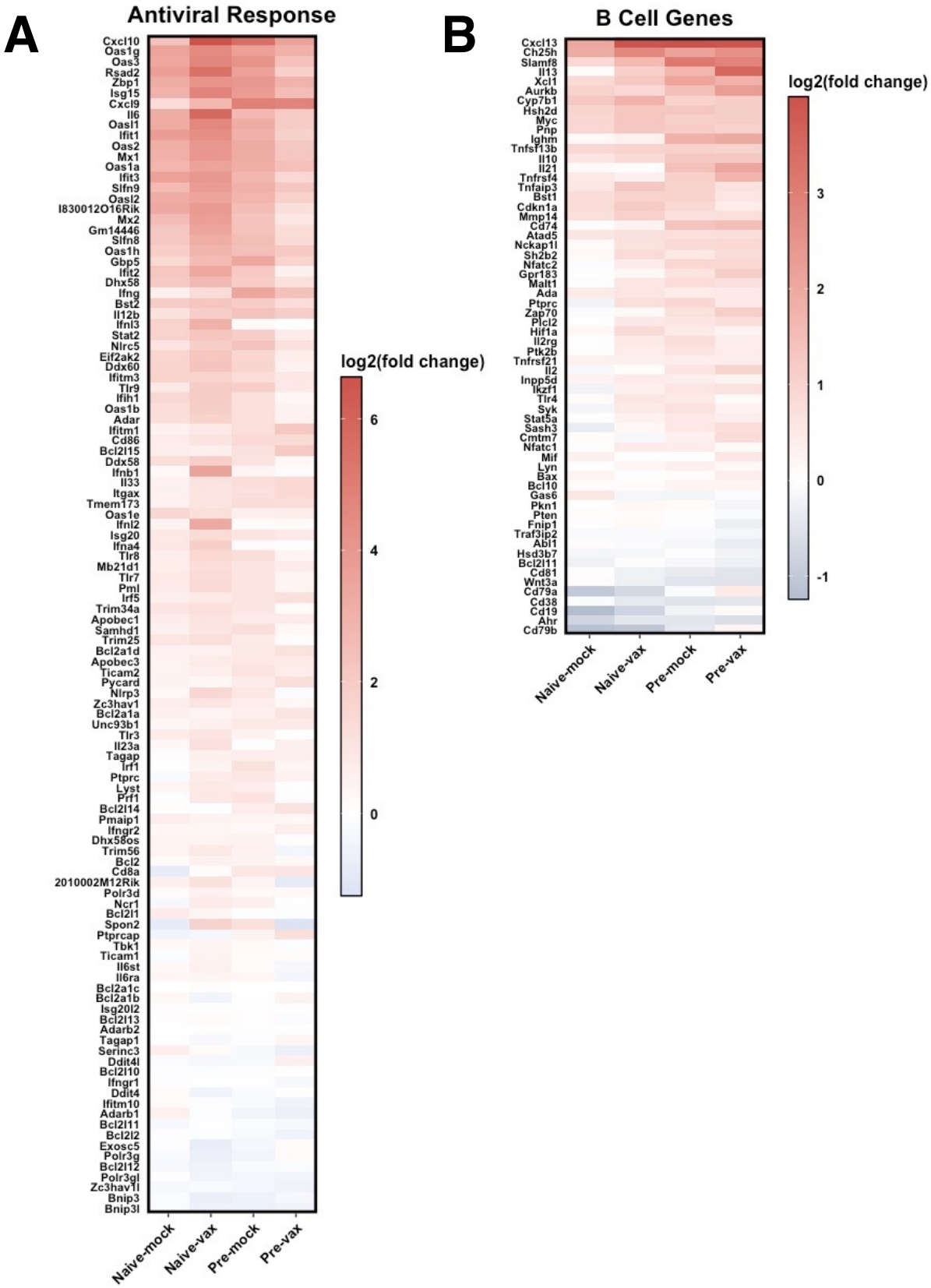

A
Antiviral Response
Cxcl10
Oas1g
Oas3
Rsad2
Zbp1
Isg15
Cxcl9
Il6
Oasl1
Ifit1
Oas2
Mx1
Oas1a
Ifit3
Slfn9
Oasl2
I830012O16Rik
Mx2
Gm14446
Slfn8
Oas1h
Gbp5
Ifit2
Dhx58
Ifng
Bst2
Il12b
Ifnl3
Stat2
Nlrc5
Eif2ak2
Ddx60
Ifitm3
Tlr9
Ifih1
Oas1b
Adar
Ifitm1
Cd86
Bcl2l15
Ddx58
Ifnb1
Il33
Itgax
Tmem173
Oas1e
Ifnl2
Isg20
Ifna4
Tlr8
Mb21d1
Tlr7
Pml
Irf5
Trim34a
Apobec1
Samhd1
Trim25
Bcl2a1d
Apobec3
Ticam2
Pycard
Nlrp3
Zc3hav1
Bcl2a1a
Unc93b1
Tlr3
Il23a
Tagap
Irf1
Ptprc
Lyst
Prf1
Bcl2l14
Pmaip1
Ifngr2
Dhx58os
Trim56
Bcl2
Cd8a
2010002M12Rik
Polr3d
Ncr1
Bcl2l1
Spon2
Ptprcap
Tbk1
Ticam1
Il6st
Il6ra
Bcl2a1c
Bcl2a1b
Isg20l2
Bcl2l13
Adarb2
Tagap1
Serinc3
Ddit4l
Bcl2l10
Ifngr1
Ddit4
Ifitm10
Adarb1
Bcl2l11
Bcl2l2
Exosc5
Polr3g
Bcl2l12
Polr3gl
Zc3hav1l
Bnip3
Bnip3l
Naive-mock
Naive-vax
Pre-mock
Pre-vax
log2(fold change)
6
4
2
0
B
B Cell Genes
Cxcl13
Ch25h
Slamf8
Il13
Xcl1
Aurkb
Cyp7b1
Hsh2d
Myc
Pnp
Ighm
Tnfsf13b
Il10
Il21
Tnfrsf4
Tnfaip3
Bst1
Cdkn1a
Mmp14
Cd74
Atad5
Nckap1l
Sh2b2
Nfatc2
Gpr183
Malt1
Ada
Ptprc
Zap70
Plcl2
Hif1a
Il2rg
Ptk2b
Tnfrsf21
Il2
Inpp5d
Ikzf1
Tlr4
Syk
Stat5a
Sash3
Cmtm7
Nfatc1
Mif
Lyn
Bax
Bcl10
Gas6
Pkn1
Pten
Fnip1
Traf3ip2
Abl1
Hsd3b7
Bcl2l11
Cd81
Wnt3a
Cd79a
Cd38
Cd19
Ahr
Cd79b
Naive-mock
Naive-vax
Pre-mock
Pre-vax
log2(fold change)
3
2
1
0
-1

Figure 3.12 Antiviral and B cell-mediated immune pathway regulation in mouse lungs at Mex/09 challenge depends on immune background.

Adult C57Bl/6j mice of varying immune backgrounds (naïve-mock vaccinated, naïve-vaccinated, preimmune-mock vaccinated, and preimmune-vaccinated) were infected with a lethal dose of influenza virus A/Mexico/4108/2009 (Mex/09). At three days post-inoculation, total RNA was extracted from mouse lungs and sequenced using the Illumina platform. Gene expression was quantified as fragments per kilobases mapped (fpkm), and the $\log_2$(fold change) was calculated relative to non-infected mice. (A) The $\log_2$(fold change) was plotted for the top differentially regulated genes for gene ontology (GO) biological process (BP) terms "defense response to virus" (GO:0051607), including antiviral innate immune response genes (GO:0140374). (B) Heatmaps were generated for genes falling under several B cell-mediated immune response pathways. The gene list was generated by combining genes for GO BP terms "B cell chemotaxis" (GO:0035754), "B cell homeostasis"(GO:0001782), "B cell differentiation" (GO:0030183), "negative regulation of B cell differentiation" (GO:0045578), "B cell receptor complex" (GO:0019815), "negative regulation of B cell receptor signalling" (GO:0050859), "B cell apoptotic process" (GO:0001783), and "positive regulation of B cell proliferation" (GO:0030890). The heatmaps were generated with ggplot2 package version 3.3.2.

In general, the antiviral response was most uniformly up-regulated in naïve-vaccinated mice, with naïve-mock vaccinated and preimmune-mock vaccinated mice showing similar levels of antiviral gene expression to each other (**Figure 3.12A**). In response to challenge, all groups showed up-regulation of inflammatory cytokines Cxcl10, Cxcl9, and IL-6. All groups also showed upregulation of key interferon-stimulated genes (ISGs), such as Oas1a and Oas3, Isg15, Rsad2, Mx1, Ifit2 and Ifit3, and Eif2ak2 (**Figure 3.12A).** However, I observed differences between groups in expression of certain antiviral response genes. Notably, the IFN-gamma-induced chemokine Cxcl10, which signals chemoattraction of macrophages, T cells, NK cells, and dendritic cells, was most highly up-regulated in naïve-vaccinated mice. Interferon gamma (Ifng) and IFN-gamma induced chemokine Cxcl9 [223,224] was most strongly up-regulated in preimmune mice, with little differential expression in naïve mice. Interestingly, the M-1 macrophage-recruiting protein Spon2 [225] was highly down-regulated in naïve-mock vaccinated and preimmune-vaccinated mice. I also found the inflammasome component Nlrp3 [226,92] to be upregulated in naïve-vaccinated mice, with only minor up-regulation in preimmune-mock vaccinated mice. Other apparent differences in antiviral response genes may point to increased adaptive immune involvement. For

example, Ptrpc/Ptrcap, a regulator of B and T cell activation [227], was up-regulated in preimmune groups, but down-regulated or unchanged in naïve mice (**Figure 3.12A**).

At challenge, all immune backgrounds showed differential expression of genes involved in B-cell mediated immunity (**Figure 3.12B**). For example, B cell chemoattractants Cxcl13 [228] and Xcl1 [229] were highly up-regulated in all groups, although the least up-regulation was observed in naïve-mock vaccinated mice. However, I found B cell genes to be most uniformly upregulated in preimmune animals. Both preimmune groups showed upregulation of Ighm, encoding the heavy chain of IgM; IL-21, which plays a key role in B cell differentiation to plasma cells and promotes germinal centre reactions [230]; and CD74, which regulates B cell survival [231] (**Figure 3.12B**). Interestingly, preimmune-vaccinated mice showed greatest up-regulation of IL-13, a central regulator of IgE synthesis [232] (**Figure 3.12B**). Preimmune-vaccinated mice showed upregulated Cd79a/b, a component of the B cell receptor complex [233], suggesting increased BCR signalling which was not observed in the other groups.

Finally, I assessed expression of genes involved in T cell-mediated cytotoxicity and activation (**Figure 3.13**). All groups showed up-regulation of genes involved in T cell cytotoxicity. For example, I found IL-12b, a subunit of IL-12, to be highly upregulated in all groups, as well as granzyme B (Gzmb). All groups also showed up-regulation of H2-M2, an MHC class I gene, as well as beta-2 microglobulin (B2m), an MHC class I subunit [234] where the greatest fold-change occurred in preimmune-vaccinated mice. However, I found key differences between groups with respect to genes involved in activated T cell proliferation (**Figure 3.13**). Activated T cell proliferation at challenge was most uniformly upregulated in preimmune mice. Preimmune-mock vaccinated mice showed strong upregulation of Ido1, which plays a role in suppressing T and NK cells [235,236] and generating regulatory T cells [237,238]. Both preimmune groups showed upregulation of Pdcd1lg2, encoding programmed cell death 1 ligand 2 which is a key costimulatory molecule for T cell proliferation [239]. I also noted Il27ra, which contributes to CD4+ T cell

differentiation to Th1 cells [240], to be upregulated in preimmune mice, and down-regulated in naïve-mock vaccinated mice (**Figure 3.13**). Cd24a, which is highly expressed in mouse activated T cells [241], was also upregulated in preimmune mice only.

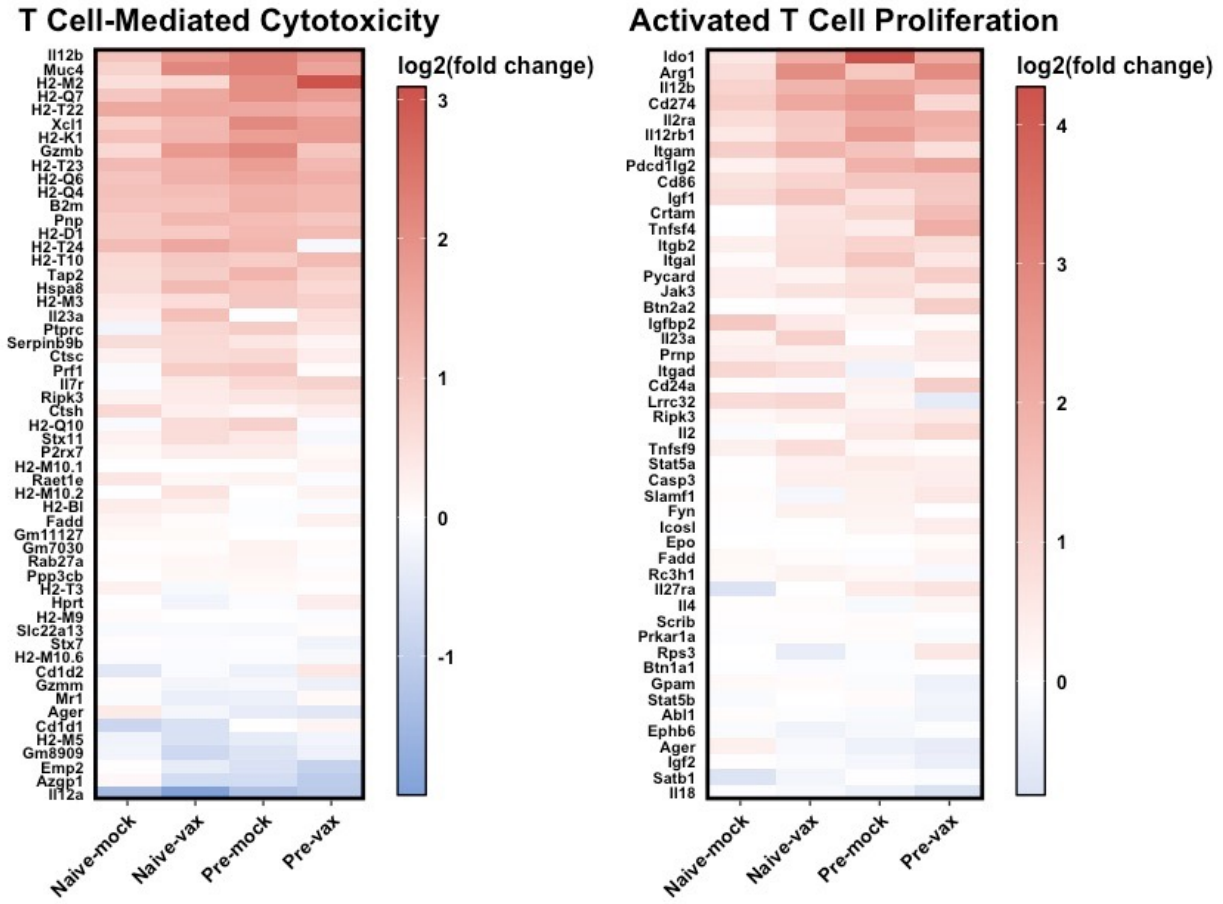


Figure 3.13 Preimmune-mock vaccinated animals have highest differential expression of genes regulating T cell-mediated cytotoxicity and activated T cell proliferation at three days post-Mex/09 challenge.

Adult C57Bl/6j mice of varying immune backgrounds (naïve-mock vaccinated, naïve-vaccinated, preimmune-mock vaccinated, and preimmune-vaccinated) were infected with a lethal dose of influenza virus A/Mexico/4108/2009 (Mex/09). At three days post-inoculation, total RNA was extracted from mouse lungs and sequenced using the Illumina platform to define the host transcriptome. Gene expression was quantified as fragments per kilobases mapped (fpkm), and the $\log_2$(fold change) was calculated relative to non-infected mice. The $\log_2$(fold change) was plotted for gene ontology (GO) biological process terms “T cell-mediated cytotoxicity” (GO:0001913) and “activated t cell proliferation” (GO:0050798). The heatmaps were generated with ggplot2 package version 3.3.2.

The results of this enrichment and immune pathway-specific gene expression analysis showed polarizing immune responses across the four immune backgrounds. Here I found that the immune response of naïve mice at three days post-challenge is dominated by proinflammatory and antiviral response pathways, with naïve-vaccinated mice having the highest antiviral response activation of all groups. While the antiviral response is also initiated in preimmune animals, they quickly shift to adaptive immune mechanisms that are highly skewed toward T cell activation and cytotoxicity, although B cell activation was also observed. Based on these differences in immune response, combined with the unique SNP profiles observed on the viral proteins of each group, it seems that the specific infection and vaccination status of the mice imposed distinct immune pressures corresponding to the mutation on HA, NP, NA, and PB1.

3.9 Summary of Chapter 3

Changing antigenicity of viral proteins is the single greatest challenge surrounding immunity to influenza viruses. With the goal of understanding the host's role in antigenic drift, our group developed differentially infected and vaccinated mouse models to reflect the diverse immune backgrounds in the human population. When mice of all immune backgrounds were infected with a H1N1 virus heterologous to the initial historical H1N1 viral infection, but homologous to the H1 vaccine component, the RNA from the virus populations extracted from the mouse lungs three days later showed differences in mutations across both internal (NP, PB1) and external (HA, NA) proteins that were specific to their infection and vaccination history (**Figure 3.14**). This pattern in mutation was related to the polarizing immune responses observed through sequencing of the host lung transcriptome. At challenge, naïve-mock vaccinated mice relied on non-specific proinflammatory immune mechanisms while facilitating the highest viral load and

driving moderate mutation of viral HA and NP. Previous vaccination primed the antiviral immune response in naïve-vaccinated animals, as well as improving B and T cell immune activation at challenge. However, the greatest number of mutations were observed across HA, NA, NP, and PB1 in naïve-vaccinated animals. Finally, preimmune mice have the greatest activation of adaptive immune mechanisms at challenge. These immune pathways were skewed toward T cell activation and proliferation along with early production of antibodies. Mice that are preimmune, but unvaccinated, showed fewer viral mutations than naïve-vaccinated mice, while combined preimmunity and vaccination lead to no shared mutations across viral proteins. Moreover, my computational analysis of predicted B cell and T cell epitopes showed that the emerged mutations may impact viral antigenicity. The results of this study highlight the important link between the infection and vaccination history of the host and the ever-changing antigenicity of influenza viruses. Immune memory toward specific antigens encountered through viral infection and vaccination form the unique immune pressures that may drive antigenic drift, possibly in predictable ways.

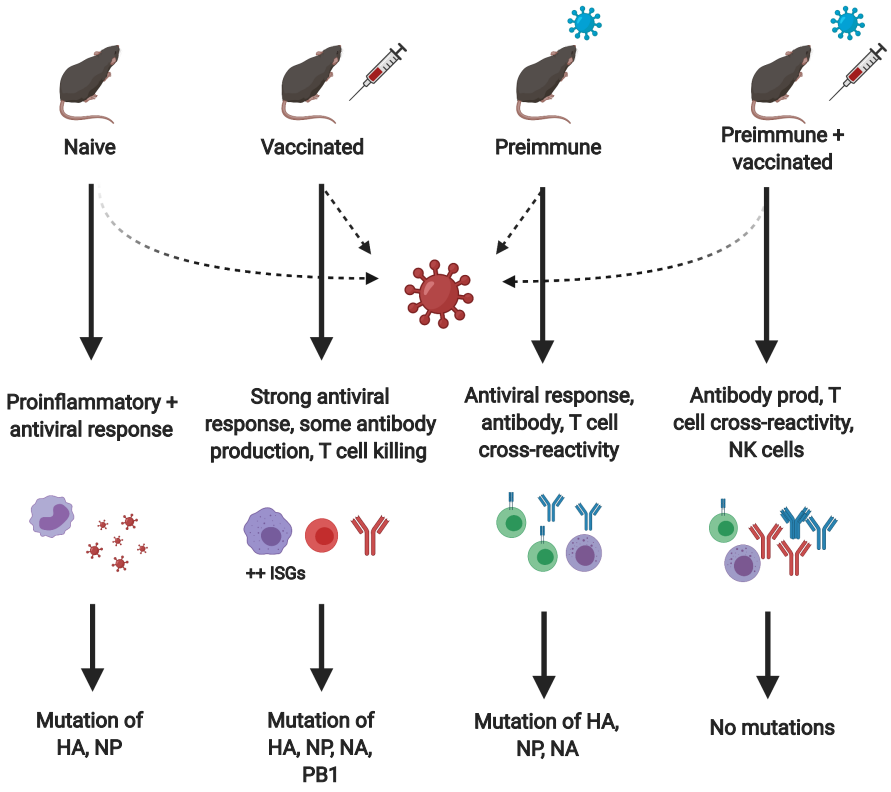


Figure 3.14 Previous influenza virus infection and vaccination generates selection pressure for influenza virus mutation at challenge.

Immune backgrounds toward influenza viruses were established in mice through sequential infection with seasonal H1N1 virus A/Fort Monmouth/1/1947 followed by vaccination with 2009 pandemic H1N1 to generate naïve-mock vaccinated, naïve-vaccinated, preimmune-mock vaccinated, and preimmune-vaccinated animals. Mice of all immune backgrounds were subject to challenge with a lethal dose of A/Mexico/4108/2009 (Mex/09), and lung RNA was extracted at three days post-challenge for viral and host transcriptome sequencing. Viral proteins HA, NP, NA, and PB1 show patterns in mutation what were specific to immune background. At challenge, naïve-mock vaccinated mice rely on proinflammatory innate immune mechanisms, which facilitated several mutations on HA (external) and NP (internal). Naïve-vaccinated mice show a more targeted antiviral response, greater involvement T cell cytotoxicity, and some B cell activation; however, this response facilitates the greatest number of mutations on external proteins HA and NA, as well as internal NP and PB1. Finally, preimmune mice have the greatest activation of adaptive immune mechanisms which are skewed toward T cell activation and proliferation, although antibody production was also observed. Mice that are preimmune but unvaccinated showed a similar pattern of viral mutation to the naïve-vaccinated mice. Combined preimmunity and vaccination lead to no shared mutations across viral proteins.

CHAPTER 4 COMPUTATIONAL ANALYSIS OF SAR-COV-2 SPIKE ANTIGENICITY SHOWS DIFFERENCES IN PREDICTED B CELL EPITOPES AND GLYCOSYLATION SITES ACROSS VARIANTS

4.1 Introduction to vaccine effectiveness against SARS-CoV-2 variants

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is the highly contagious respiratory virus responsible for the coronavirus disease of 2019 (COVID-19) pandemic. Efforts to develop effective vaccines against SARS-CoV-2 have greatly revolved around the spike glycoprotein as the immunogenic target [30,31], since antibody binding to spike can effectively block viral entry into host cells and suppresses viral infection [32]. Several spike-based vaccine candidates are already in distribution, such as the Pfizer-BioNTech (BNT162b2) and Moderna (mRNA-1273) mRNA vaccines, as well as the adenovirus vector vaccines from Johnson and Johnson (JNJ-78436735/Ad26.COV2.S) and Oxford-AstraZeneca (AZD1222/ChAdOx1), each of which features the original SARS-CoV-2 spike protein sequence derived from the early cases of infection [33,34].

Over the past year, several variants of SARS-CoV-2 have emerged that have raised concerns about the effectiveness of SARS-CoV-2 vaccine platforms being utilized [135]. The World Health Organization categorizes variants as variant of concern; variant of interest; and variant under monitoring. A SARS-CoV-2 variant a “variant of concern” (VOC) is defined as a variant that demonstrates one of the following traits: increased virus transmission; increase in virulence; change in clinical disease presentation; or decrease in effectiveness of countermeasures [136]. This is in contrast to a “variant of interest” (VOI), which is defined as a variant causing significant community transmission or appearing in multiple clusters, in addition to containing mutations that may affect transmissibility and disease severity, or effectiveness of therapeutics and diagnostics [136]. Variants under monitoring (VUM) are that have genetic signatures that suggest they pose a possible risk.

The first VOC identified was lineage B.1.1.7 reported in the UK in December of 2020 [137]. Now designated Alpha, the variant was modeled to have an increased reproduction number compared to original SARS-CoV-2 Wuhan virus [138,139]. Current evidence suggests that Alpha is still susceptible to antibodies elicited by natural infection and ancestral spike vaccines [144–147], although some RBD-specific antibody titers may be reduced [144,145]. A second VOC originally designated lineage B.1.351 and now referred to as the Beta variant was detected in South Africa in October of 2020 [148,149]. In contrast to Alpha, Beta has shown reduced susceptibility to antibodies elicited after natural infection [151], vaccination [147], and treatment with monoclonal antibodies bamlanivimab, estesevimab, and imdevimab [152]. Another significant VOC referred to as Gamma (lineage P.1) was identified in Brazil in late 2020 [153]. Gamma was observed to be more transmissible than Alpha and Beta [139] and showed reduction in neutralizing antibodies obtained after natural infection, but retention of antibodies induced by vaccination [242]. October of 2020 also saw the emergence of the B.1.617 lineage in India, which is comprised of three sub-lineages designated B.1.617.1 (Kappa), B.1.617.2 (Delta), and B.1.617.3 [155]. Preliminary data speculated that some B.1.617 sub-lineages had reduced effectiveness of treatment with the monoclonal antibody bamlanivimab, and reduced neutralization from convalescent and Pfizer-BioNTech (BNT612b) vaccine sera [152,156]. Variants of interest include the Epsilon or B.1.429 lineage identified in California [160] and the mink-related (Cluster 5) lineage detected in Denmark in June of 2020, which also have several non-synonymous mutations compared to the original SARS-CoV-2 reference sequence [161]. The mink-related variant has been suggested as a potential spillover threat with minks acting as viral reservoir, although spread of the mink-related variant to humans has been rare [162].

At this time, it is still unclear whether SARS-CoV-2 variants of concern or interest are able to be managed by previously-acquired host adaptive immune responses, either through natural infection or through vaccination. Furthermore, as vaccine effectiveness is dependent on the

recognition of viral proteins by antigen-educated T and B cells, mutations influencing T and B cell epitopes in new SARS-CoV-2 variants may have significant effects on the performance of vaccines that were originally developed toward original SARS-CoV-2. Here I investigated the predicted immune epitope changes in the identified SARS-CoV-2 variants by leveraging my established informative pipeline used for influenza virus analysis. I hypothesized that the variant SARS-CoV-2 spike sequences would show differences in predicted surface epitopes that correspond to the observed differences in vaccine effectiveness compared to the original Wuhan SARS-CoV-2. The goal of this analysis was to determine if vaccine effectiveness could be predicted informatically by analyzing how the specific amino acid substitutions across the SARS-CoV-2 spike protein variants will impact antigenicity, and which arms of the immune response may be most affected by viral mutation.

4.2 A computational approach to screening antigenic differences on spike

Conducting immunological and virological analysis of a SARS-CoV-2 variant as it emerges to determine the effectiveness of pre-existing immunity and vaccine protection is time-consuming, expensive, and labor intensive. There is currently a need to perform immunological analysis of SARS-CoV-2 variants quickly to determine the effectiveness of existing vaccines and therapeutics. Therefore, I developed a bioinformatic antigenic analysis pipeline to gain a molecular understanding of potential vaccine performance against the new variants as a rapid way to screen the variants prior to complete immunological and virological analysis for the assessment of the threat variants may have globally. Using the Global Initiative on Sharing All Influenza Data (GISAID) [186] to access SARS-CoV-2 genome sequences, I compared the predicted antigenic profiles for Alpha, Beta, Gamma, Delta, Epsilon, Kappa, B.1.617.3, and mink-related variant spike proteins. Because the impact of specific SARS-CoV-2 spike mutations on

different arms of the immune response is unclear, my analysis included: 1) prediction of the structural impact of variant amino acid subsitutions on the original SARS-CoV-2 spike protein; 2) comparision of predicted B cell epitopes along the variant spike sequences; 3) determination of possible glycosylation patterns along variant spike proteins; and 4) comparison of MHC class I and II epitopes across spike variants. I followed a similar pipeline described in Chapter 3 with modifications for SARS-CoV-2 specifically, including a second strategy for predicting the folding structure of variant spike sequences, analysis of MHC class II- binding peptides to encompass a possible CD4+ T cell eptiopes, and prediciton of spike glycoslation patterns that may interfere with antibody binding. An overview of the bioinformatic pipeline and tools used is shown in **Figure 4.1**.

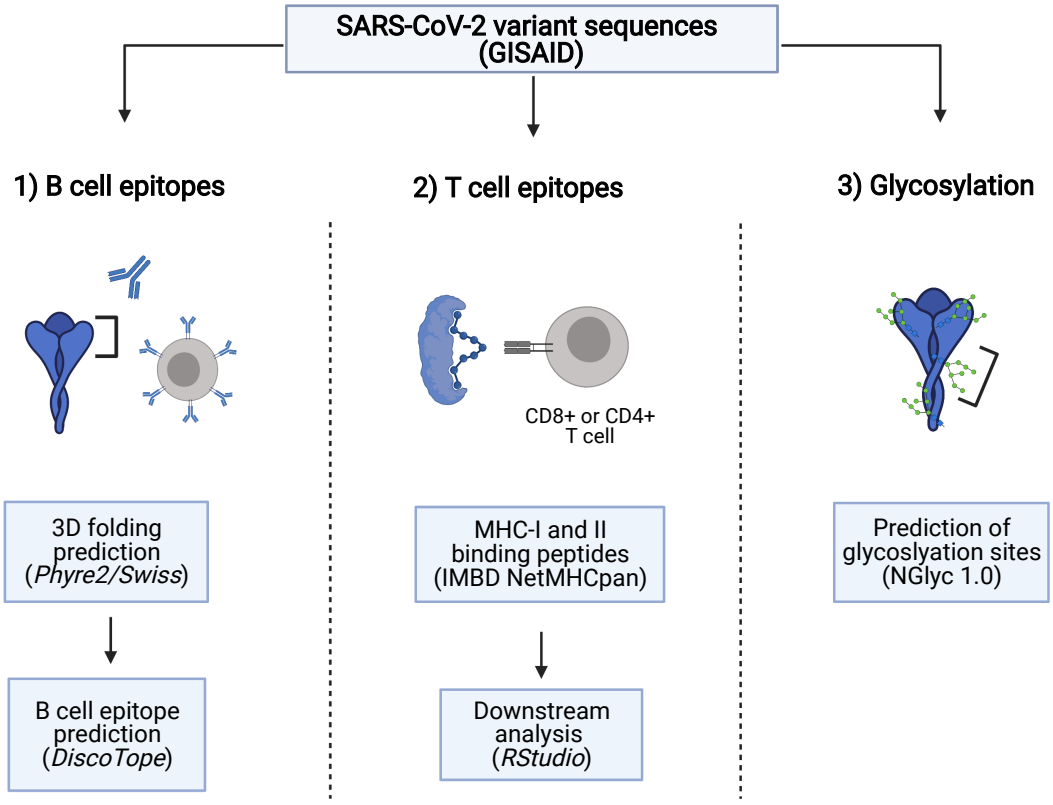


Figure 4.1 Bioinformatic analysis pipeline to predict variant SARS-CoV-2 spike epitopes.

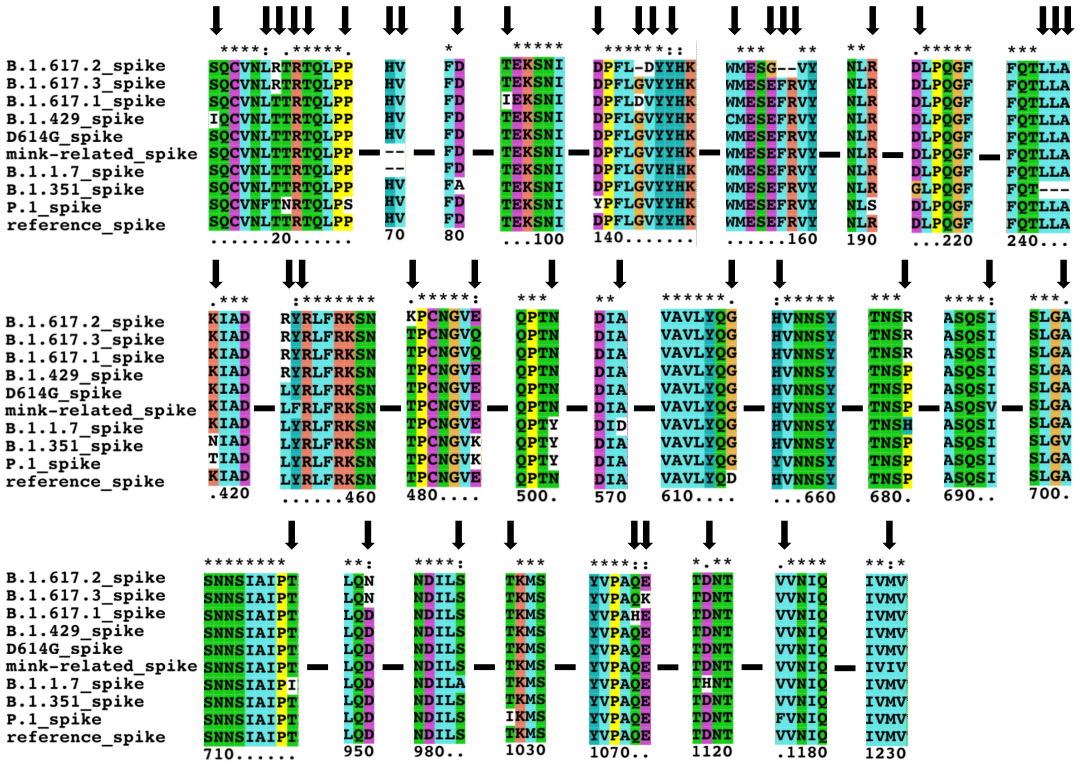
To compare the antigenicity and glycosylation of SARS-CoV-2 spike variants, the original Wuhan (reference) spike sequence along with nine SARS-CoV-2 variants were obtained from the GISAID database. The accession numbers are as follows: the original Wuhan SARS-CoV-2 reference sequence (EPI_ISL_402124); B.1.1.7 (EPI_ISL_852526); B.1.351 (EPI_ISL_864621); B.1.429 (EPI_ISL_1017160); D614G (EPI_ISL_406862), P.1 (EPI_ISL_1171653), B.1.617.1 (EPI_ISL_1841381), B.1.617.2 (EPI_ISL_1914598), B.1.617.3 (EPI_ISL_1939891), and the mink-related lineage (EPI_ISL_615652). Protein folding models were generated for each spike variant using Phyre2 and Swiss-Model platforms, after which discontinuous B cell epitopes were predicted using DiscoTope 2.0 from DTU HealthTech. To predict CD4+ or CD8+ T cell epitopes, the Immune Epitope Database (IMDB) NetMHC pan tools were used. Conserved and variant-specific high-affinity MHC I and MHC II-binding peptides were identifed in RStudio. Finally, variant spike glycosulation was predicted using NGlyc 1.0 to determine differences in glycan binding sites across variants.

4.3 Predicting structural impact of substitutions across variants

Nine distinct SARS-CoV-2 variants, specifically D614G, Alpha (B.1.1.7), Beta (B.1.351), Gamma (P.1), three B.1.617 sub-lineages (includes Delta and Kappa), Epsilon (B.1.429), and the mink-related variant were selected along with the original Wuhan SARS-CoV-2 virus (herein referred to as “reference” spike) to explore antigenic differences in response to mutations within their surface glycoprotein. Variants were chosen for analysis according to known or predicted high prevalence in communities (D614G, Alpha (B.1.1.7), Beta (B.1.351), Gamma (P.1), Delta and Kappa), or the suspected ability to cause future outbreaks (Epsilon (B.1.429), and the mink-related variant). Using truncated full genome sequences from GISAID, I first aligned reference and variant spike sequences using the MEGAX alignment tool [185] to better visualize the relative location of sequence mutations (**Figure 4.2A**). Across variants, thirty-seven substitutions (S13I, L18F, T19R, T20N, P26S, D80A, T95I, D138Y, G142D, W152C, E156G, R190S, D215G, K417T, K417N, L452R, Y453F, T478K, E484K, E484Q, N501Y, A570D, D614G, H655Y, P681H, P681R, I692V, A701V, T716I, D950N, S982A, T1027I, Q1071H, E1072K, D1118H, V1176F, and M1229I) and deletions at positions 69-70, 143, 145, 157-158, and 241-243, were noted (**Figure 4.2**). A common D614G mutation was identified in all variants, and the substitution at position 501 (N501Y) within the receptor-binding domain was shared among Alpha, Beta, and Gamma

lineages (**Figure 4.2B**). The mink-related variant contained a deletion that was also recognized in Alpha (B.1.1.7) at positions 69-70. Moreover, a leucine to arginine substitution at amino acid 452 (L452R) was found in Epsilon (B.1.429) and all B.1.617 sub-lineages (**Figure 4.2**). Another mutation at position 484 substituted glutamic acid for either lysine (E484K) in Beta and Gamma variants, or glutamine (E484Q) in B.1.617.1 and B.1.617.3 sub-lineages. Similarly, different substitutions at amino acid 681 were observed in Alpha (P681H) and B.1.617 (P681R) variants (**Figure 4.2B**).

A


B.1.617.2_spike
B.1.617.3_spike
B.1.617.1_spike
B.1.429_spike
D614G_spike
mink-related_spike
B.1.1.7_spike
B.1.351_spike
P.1_spike
reference_spike

B

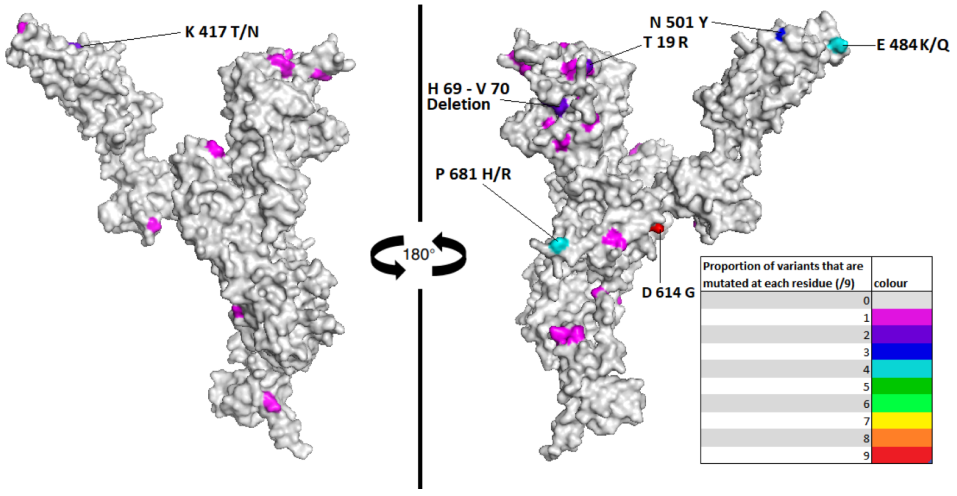

K 417 T/N
N 501 Y
T 19 R
E 484 K/Q
H 69 - V 70
Deletion
P 681 H/R
D 614 G
180°
Proportion of variants that are mutated at each residue (/9)
colour

Figure 4.2 Amino acid sequence changes in variant and Wuhan SARS-CoV-2 spike protein illustrate shared and unique mutations among the variants.

Sequences for Wuhan (reference) SARS-CoV-2 spike, as well as from nine variants (D614G, B.1.1.7, B.1.351, B.1.429, P.1, B.1.617 sub-lineages, and mink-related SARS-CoV-2) were obtained from the GISAID database and are as follows: the original Wuhan SARS-CoV-2 reference sequence (EPI_ISL_402124); B.1.1.7 (EPI_ISL_852526); B.1.351 (EPI_ISL_864621); B.1.429 (EPI_ISL_1017160); D614G (EPI_ISL_406862), P.1 (EPI_ISL_1171653), B.1.617.1 (EPI_ISL_1841381), B.1.617.2 (EPI_ISL_1914598), B.1.617.3 (EPI_ISL_1939891), and the mink-related lineage (EPI_ISL_615652). **(A)** Complete spike sequences were aligned using MEGAX. Amino acid substitutions relative to the Wuhan spike sequence are indicated with arrows, and sequence deletions are denoted by dashes. The figure was generated using the ClustalX platform. **(B)** A three-dimensional model of the Wuhan SARS-CoV-2 spike was generated using the Swiss-Model structural bioinformatics server (https://swissmodel.expasy.org/interactive) and shown in PyMOL (The PyMOL Molecular Graphics System, Version 2.3.5, Schrödinger, LLC). The frequency of amino acid substitutions across the nine variants are indicated by residue colors on the model to highlight convergent mutations.

It is unclear how the amino acid substitutions or deletions across variants will impact SARS-CoV-2 spike structure and function. I therefore aimed to predict the structural impact of each mutation using the web-based server Missense 3D using the PDB files generated by Phyre2 [189] (**Table 4.1**). Using the Wuhan SARS-CoV-2 spike as a reference, the individual amino acid substitutions were assessed for structural changes such as buried charged residues, disruption of salt bridges, and major cavity expansions or contractions. Of the thirty-seven substitutions spanning the nine spike variants, notable structural damage was predicted for six mutations. First, the aspartate to alanine switch at residue 80 (D80A) in the Beta variant was predicted to change a buried, charged residue to an uncharged residue and disrupt the corresponding salt bridge. Second was the proline to histidine/arginine switch at residue 681 (P681H/R) in the Alpha variant or B.1.617 sub-lineages, which was predicted to switch an uncharged, buried amino acid residue to an exposed, positively charged residue. Two mutations exclusive to the Gamma variant revealed structural damage: L18F was predicted to contract the cavity volume by 104.976 $Å^3$, while D138Y contracted the cavity volume by 91.584 $Å^3$ and disrupted the buried H-bonding, causing a buried amino acid to be switched to an exposed residue. Finally, two mutations

exclusive to the B.1.617 sub-lineages, which are E484Q in Kappa and B.1.617.3, and G142D in Kappa and Delta, had predicted structural damage (**Table 4.1**). The remaining amino acid substitutions were not predicted to cause structural changes. Because Missense 3D evaluates missense mutations, the deletions appearing in Alpha, Beta, and mink-related spike were not assessed in this analysis.

Table 4.1 Missense3D-predicted structural damage and description of amino acid changes for spike amino acid substitutions across SARS-CoV-2 variants of concern (Alpha, Beta, Gamma, and Delta) and variants of interest (Epsilon, Kappa, B.1.617.3, and mink-related).

Variant	Substitution	Location*	Missense3D structural damage	Amino acid change
Alpha (B.1.1.7)	N501Y	RBD	None detected	Exposed hydrophilic to neutral
	A570D	SD1	None detected	Exposed hydrophobic to hydrophilic
	D614G	SD2	None detected	Exposed hydrophilic to neutral
	P681H	Upstream of S1/S2 cleavage site	Switched from buried to exposed	Uncharged to positively charged
	T716I		None detected	Exposed neutral to hydrophobic
	S982A	HR1	None detected	
	D1118H	SD3	None detected	Both exposed hydrophilic
Beta (B.1.351)	D80A	NTD	Buried charged residue replaced with an uncharged; salt bridge disrupted	Negatively charged switch to uncharged
	D215G	NTD	None detected	
	D614G	SD2	None detected	Exposed hydrophilic to neutral
	N501Y	RBD	None detected	Both exposed hydrophilic
	K417N		None detected	
	E484K		None detected	
	A701V	Cleavage site	None detected	Both exposed hydrophobic
Gamma (P.1)	L18F	NTD	Leads to contraction of cavity volume by 104.976 Å^3	Both exposed hydrophobic
	T20N	NTD	None detected	Exposed neutral to exposed hydrophilic (both uncharged)
	P26S	NTD	None detected	Both exposed neutral
	D138Y	NTD	Disrupt buried H-bond; contraction of cavity volume by 91.584 Å^3; buried to exposed switch	Buried negatively charged hydrophilic to exposed neutral
	R190S	NTD	None detected	Exposed positively charged hydrophilic to buried neutral

Variant	Substitution	Location*	Missense3D structural damage	Amino acid change
	K417T	RBD	None detected	Exposed positively charged hydrophilic to exposed neutral
	E484K	RBD	None detected	Both exposed hydrophilic
	N510Y	RBD	None detected	Exposed hydrophilic to neutral
	D614G	SD2	None detected	Exposed hydrophilic to neutral
	H655Y	SD2	None detected	Exposed positively charged hydrophilic to neutral
	T1027I	CH	None detected	Exposed neutral to hydrophobic
	V1176F	HR2	None detected	Both exposed hydrophobic
Delta (B.1.617.2)	T19R	NTD	No structural damage detected	Exposed neutral to hydrophilic (uncharged to charged)
	G142D	NTD	Buried Gly replaced with a charged residue; cavity volume contracted by 71.064 Å^3	Buried neutral to buried hydrophilic
	E156G	NTD	No structural damage detected	Exposed hydrophilic to neutral
	L452R	RBD(RBM)	No structural damage detected	Exposed hydrophobic to hydrophilic
	T478K	RBD(RBM)	No structural damage detected	Exposed neutral to hydrophilic
	D614G	SD2	No structural damage detected	Exposed hydrophilic to neutral
	P681R	SD2	Buried residue switched to exposed	Buried neutral to exposed hydrophilic (positive-charged)
	D950N	HR1	No structural damage detected	Both exposed hydrophilic (charged to uncharged)
Epsilon (B.1.429)	S13I	SP	None detected	Exposed neutral to hydrophobic
	W152C	NTD	None detected	Both exposed hydrophobic
	L452R	RBD	None detected	Exposed hydrophobic to hydrophilic
	D614G	SD2	None detected	Exposed hydrophilic to neutral
Kappa (B.1.617.1)	T95I	NTD	Buried H-bond breakage	Buried neutral to hydrophobic; both uncharged
	G142D	NTD	Buried Gly replaced with a charged residue; cavity volume contracted by 71.064 Å^3	Buried neutral to buried hydrophilic

Variant	Substitution	Location*	Missense3D structural damage	Amino acid change
	L452R	RBD(RBM)	No structural damage detected	Exposed hydrophobic to hydrophilic
	E484Q	RBD(RBM)	Triggers clash alert	Exposed negatively charged to uncharged (both hydrophilic)
	D614G	SD2	No structural damage detected	Exposed hydrophilic to neutral
	P681R	SD2	Buried residue switched to exposed	Buried neutral to exposed hydrophilic (positive-charged)
	Q1071H	BH	No structural damage detected	Both exposed hydrophilic (uncharged to charged)
	T19R	NTD	No structural damage detected	Exposed neutral to hydrophilic (uncharged to charged)
	L452R	RBD(RBM)	No structural damage detected	Exposed hydrophobic to hydrophilic
	E484Q	RBD(RBM)	Triggers clash alert	Exposed negatively charged to uncharged (both hydrophilic)
B.1.617.3	D614G	SD2	No structural damage detected	Exposed hydrophilic to neutral
	P681R	SD2	Buried residue switched to exposed	Buried neutral to exposed hydrophilic (positive-charged)
	D950N	HR1	No structural damage detected	Both exposed hydrophilic (charged to uncharged)
	E1072K	BH	No structural damage detected	Both exposed charged hydrophilic
	Y453F	RBD	None detected	Exposed neutral to buried hydrophobic
Cluster 5 (Mink-related)	D614G	SD2	None detected	Exposed hydrophilic to neutral
	I692V	Cleavage region	None detected	Both buried hydrophobic
	M1229I	CT	None detected	Both exposed hydrophobic

*SP = signal peptide; NTD = N-terminal domain; RBD = receptor-binding domain; RBM = receptor-binding motif; SD = subdomain; HR = heptad repeat; CH = central helix; CT = cytoplastic tail; BH = b-hairpin [243].

4.4 Predicted B cell epitopes differ across variants

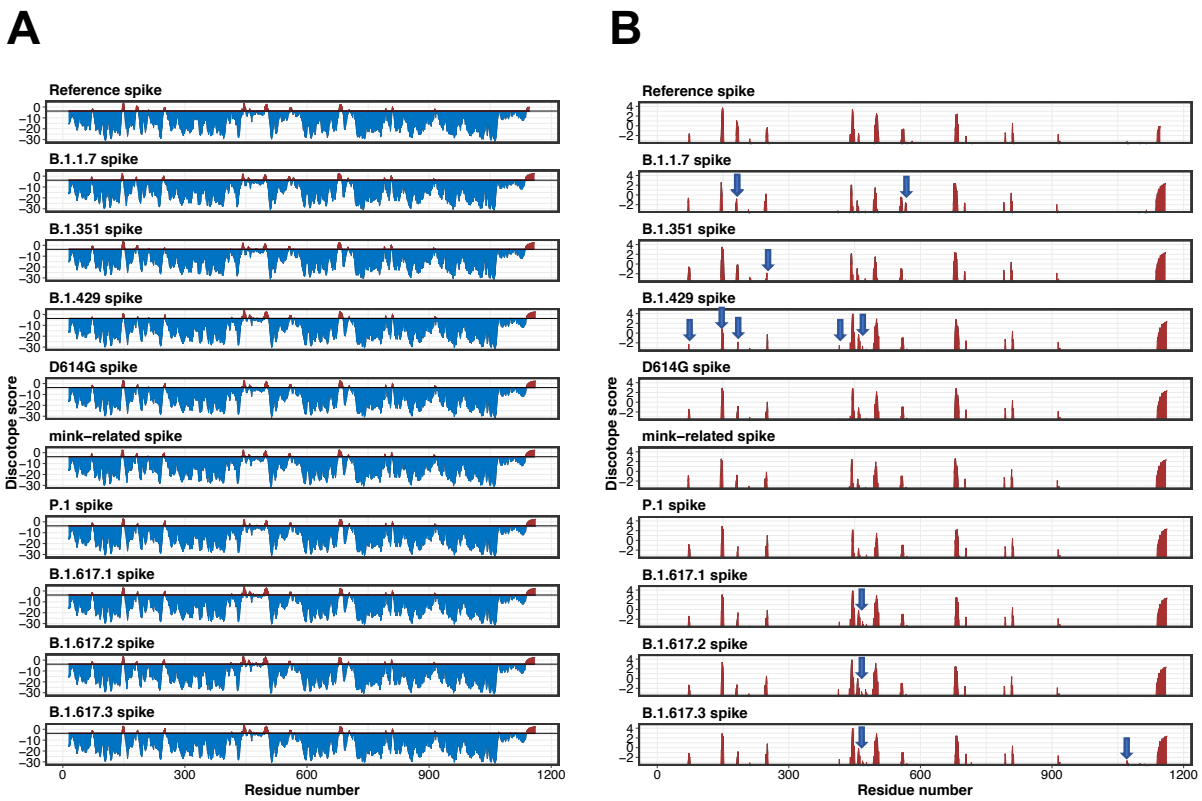
Although predicted changes in SARS-CoV-2 spike structure can help elucidate receptor docking as well as protein function such as membrane fusion, my interest was in understanding the immune system recognition of the viral variants and the stability of antigenicity across these

variants. Therefore, the primary goal of my analysis was to evaluate changes in spike antigenicity. Antibodies produced by B cells are essential to the immune memory response protecting against secondary exposures of a pathogen offering either complete sterilizing protection or decrease of disease severity. To this end, I compared the predicted B cell epitopes across reference and variant spike proteins to determine whether the variant viruses may present differently to host B cell receptors and thereby have differential recognition by secreted antibodies, which may have been elicited from previous exposures. Determination of discontinuous B cell epitopes on a protein surface first requires a three-dimensional model of a protein. Similarly as done in Chapter 3, I generated protein-folding models for reference and variant SARS-CoV-2 spikes using both Phyre2 [188] and Swiss-Model [190] web-based platforms. In comparing each model prediction with a previously published 3D spike model from the Protein Data Bank (PDB ID: 6VSB), I found that Swiss-Models best approximated the actual folding of spike. Therefore, although I generated B cell epitope predictions for both Phyre2 and Swiss-Model, only data generated using the Swiss-Model will be described in detail.

To compare discontinuous B cell epitopes across SARS-CoV-2 spike variants, I used scores generated by the DiscoTope 2.0 server from DTU Health Tech [192]. I used the default DiscoTope score threshold of -3.7 (0.47 sensitivity, 0.75 specificity), which is the least conservative threshold for selecting B cell epitopes. In examining the positively identified epitopes at the -3.7 threshold, I identified a total of 26 B cell epitope sequences across all variants with differing predicted antigenicity and levels of conservation between the new variants of concern (**Figure 4.3**; **Table 4.2;** see **Appendix 1**). I found that changes to specific B cell epitopes occurred almost exclusively within the S1 subunit and protease cleavage regions of the spike protein in all variants. First, I noted several predicted B cell epitopes with lowered DiscoTope scores in certain variants, possibly suggesting decreased antigenicity in those regions. In the Alpha (B.1.1.7) and Epsilon (B.1.429) variants, the predicted epitope spanning residues 180-185 in the reference with

sequence E-G-K-Q-G-N has shrunk to include only residues 183-184 and has scored lower (**Figure 4.3B**). Two other predicted epitopes beginning at residues 72 (G-T-N) and 146 (Y-H-K-N-N-K-S) scored lower in the Epsilon variant (**Figure 4.3B**). The Beta (B.1.351) variant also showed a lower score in the predicted epitope spanning residues 248-254 with sequence Y-L-T-P-G-D-S (**Figure 4.3B**).

I also noted regions of increased antigenicity in certain variants. The Alpha variant showed an additional predicted epitope spanning residues 565-568 with sequence D-I-D-D that was not noted in the reference spike or other variants in my analysis (**Figure 4.3B**). In Epsilon (B.1.429) and B.1.617 sub-lineages, the region spanning 455-480 showed several positively predicted residues that were not detected as antigenic regions in the reference spike (**Figure 4.3B**). Taken together, these results suggested that the specific amino acid substitutions in the S1 subunit of spike may lead to differential recognition or evasion of SARS-CoV-2 variants by specific B cell clonotypes elicited after infection.


A
B
Reference spike
B.1.1.7 spike
B.1.351 spike
B.1.429 spike
D614G spike
mink–related spike
P.1 spike
B.1.617.1 spike
B.1.617.2 spike
B.1.617.3 spike
Discotope score
Residue number
0
300
600
900
1200

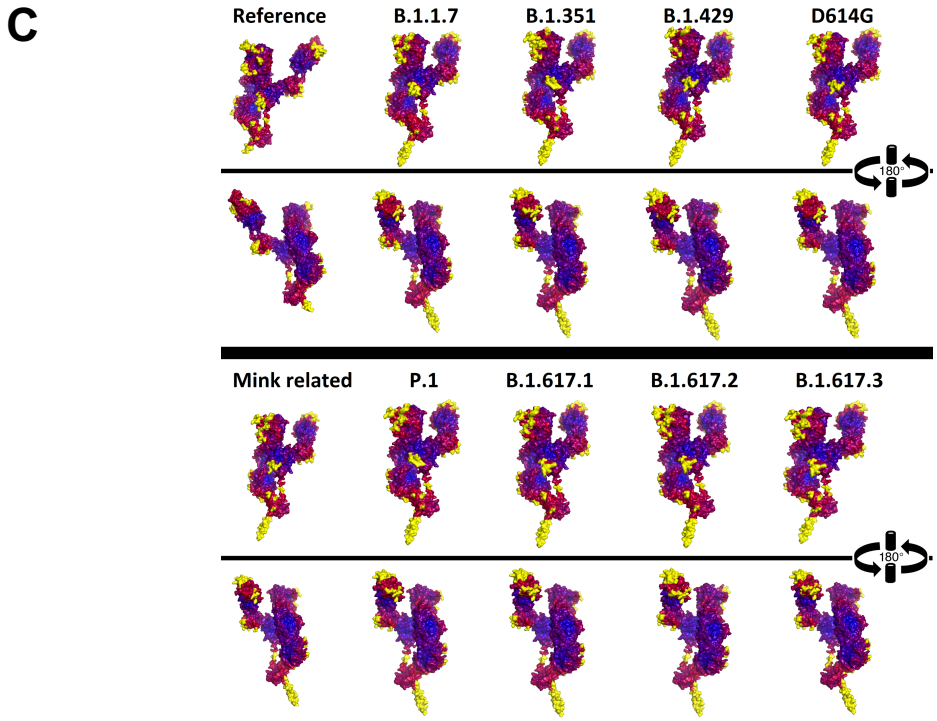

C
Reference
B.1.1.7
B.1.351
B.1.429
D614G
180°
Mink related
P.1
B.1.617.1
B.1.617.2
B.1.617.3

Figure 4.3 SARS-CoV-2 variant spike proteins show differences in predicted B cell epitopes spanning the S1 subunit compared to the Wuhan SARS-CoV-2 sequence.

Antigenic analysis of the spike protein of SARS-CoV-2 variants was performed using the DiscoTope 2.0 platform (DTU Health Tech) (http://www.cbs.dtu.dk/services/DiscoTope/). Sequences were obtained from the GISAID database and are as follows: the original Wuhan SARS-CoV-2 reference sequence (EPI_ISL_402124); B.1.1.7 (EPI_ISL_852526); B.1.351 (EPI_ISL_864621); B.1.429 (EPI_ISL_1017160); D614G (EPI_ISL_406862), P.1 (EPI_ISL_1171653), B.1.617.1 (EPI_ISL_1841381), B.1.617.2 (EPI_ISL_1914598), B.1.617.3 (EPI_ISL_1939891), and the mink-related lineage (EPI_ISL_615652). (A) DiscoTope scores falling below (blue) and above (red) the B cell epitope prediction threshold of -3.7 (0.47 sensitivity, 0.75 specificity) were mapped along the amino acid sequences of reference (Wuhan) and variant (D614G, B.1.1.7, B.1.351, B.1.429, P.1, B.1.617 sub-lineages, and mink-related) SARS-CoV-2 spike protein. (B) Only the positive B cell epitope prediction results are shown for greater resolution of the B cell epitope regions. The arrows highlight differences in predicted epitopes within the variants of concern compared to the reference spike protein. (C) Spike monomers with DiscoTope scores represented as a heatmaps were mapped along the spike protein. Predicted epitopes for each group are highlighted in yellow, with high scoring predicted B cell epitope regions shown in red, and lower predicted B cell epitope scores in blue. The image was generated using PyMOL (The PyMOL Molecular Graphics System, Version 2.3.5, Schrödinger, LLC). Figures A and B were generated using the R ggplot2 package version 3.3.2.

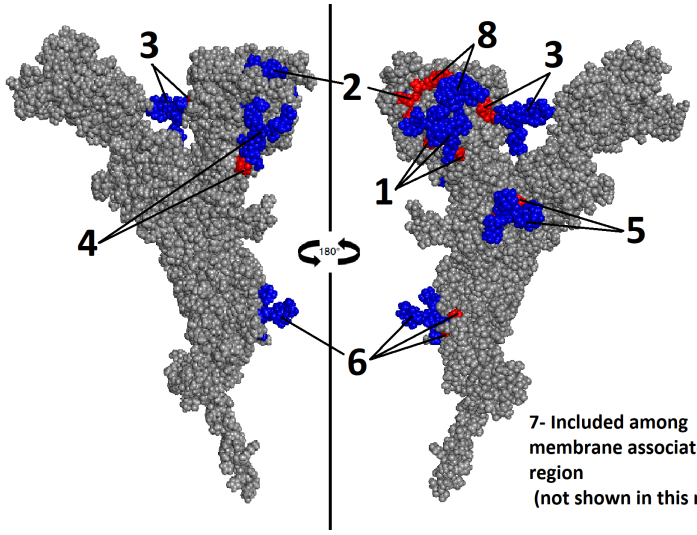
Table 4.2 Predicted B cell epitope sequences for SARS-CoV-2 spike protein using DiscoTope 2.0, proximity to glycosylation sites predicted using NetNGlyc 1.0, and sequence conservation among variants under investigation.

#	Epitope sequence	Residue range on reference	Proximal glycosylation sites	Glycosylation prediction score (NetNGlyc)	Associated mutation sites	Sequence conserved in variant models?
1	VS[GTN]G	70 - 75	N74	++	V57	Yes
2	YH[KNNKS]WM	145-153	N149	+	Y132, W139	Yes
3	(LXDX)EGK [QG]N	176-185	NONE	N/A	NONE	Yes
4	(PX)[NXV]	209-213	NONE	N/A	NONE	Yes
5	SYLT[PG]DXS	247-255	NONE	N/A	NONE	Yes
6	T	415	NONE	N/A	T402	Not predicted in reference, P.1 and mink
7	NNXXS[KVGG]N[Y]N	439-450	NONE	N/A	NONE	Yes
8	LFR[KSN]XK	455-462	NONE	N/A	NONE	Yes
9	IXTE	468-471	NONE	N/A	NONE	Only in B.1.617 sub-lineages and B.1.429
10	SXPXN	477-481	NONE	N/A	T465 (not predicted epitope)	Not in reference
11	QSXGF[QPTYGV]GY	493-505	NONE	N/A	Y488	Yes
12	[NXK]F[L]PFQQ	556-564	NONE	N/A	NONE	Yes
13	D[I]AD	568-571	NONE	N/A	A557	Yes
14	TL	581-582	NONE	N/A	NONE	Only in reference
15	T[NSPRRARSV]A	678-688	NONE	N/A	P668	Yes
16	[NSV]XY	703-707	NONE	N/A	NONE	Yes
17	[T]	716	N717	++	T703	Not predicted in B.1.1.7
18	[PI]	793-794	Close to N801	+	NONE	Yes
19	PSXP	809-812	Close to N802	+	NONE	Yes
20	[NXXYE]	914-918	NONE	N/A	NONE	Yes
21	QXXN	1071-1074	NONE	N/A	Q1058, E 1059	Not predicted in B.1.1.7 mutated in B.1.617.2 (Q1069H)
22	[T]H**	1100-1101	close to N1098	+	NONE	Yes
23	E	1111	NONE	N/A	NONE	Only in B.1.617.3
24	I	1114	NONE	N/A	NONE	Only in B.1.1.7 and B.1.617.3
25	[D]	1118	NONE	N/A	D1105	Yes
26	[PLQ...]	1140-?	close to N1134	+	NONE	Yes

4.5 The Gamma variant has additional predicted N-glycosylation sites

Antigen glycosylation is an important factor in host immunity to pathogens. Glycosylation of viral envelope proteins can modify the protein such that it can increase infectivity [244] and shield it from host immune responses [245–247]. The spike protein of SARS-CoV-2 is known to be heavily glycosylated [195,248]. To assess whether spike mutations in nine SARS-CoV-2 variants impact the predicted N-linked glycosylation sites (N-X-S/T, X ≠P), and thus possibly antigenicity, I used the NetNGlyc 1.0 server from DTU Health Tech (http://www.cbs.dtu.dk/services/NetNGlyc/) [194] with the default threshold of 0.5. The data derived from the server suggested that seventeen conserved locations that scored higher than the detection threshold were found in all variants with similar scores at sites 17, 61, 74, 122, 149, 165, 234, 282, 331, 343, 603, 616, 717, 801, 1098, 1134, and 1194 (**Table 4.3**). These sites, excluding residue 17, have previously exhibited glycosylation in investigations of spike peptide glycosylation using liquid chromatography-tandem mass spectrometry [249,250]. The predicted site at residue 17 was not found in the Delta or B.1.617.3 variants. Notably, three additional sites were detected in the Gamma (lineage P.1) variant only, scoring (+) N-glycosylation results at positions 20, 188, and 657. Along with the three additional predicted sites, the Gamma variant had a higher potential glycosylation score at position 17 compared to the other variants (++ versus +) (**Figure 4.4, Table 4.3**). Of the seventeen glycosylation sites conserved across all variants, only 7 sites (61, 74, 234, 282, 616, 717, and 1194) had (++) or better prediction scores and were selected to generate the 3D structure of the spike protein with glycans for the highest specificity (**Figure 4.4).** Each sequon confirmed to carry glycosylation so far has been found to present multiple modified versions of Man_9GlcNA_c chains, with both branched and unbranched configurations [249,250]. No other differences were observed between the other eight variants of concern or interest, including D614G, Alpha, Beta, Epsilon, B.1.617 sub-lineages, and the mink-

related variant. Based on these predictions, it seems that mutations present exclusively in the Gamma variant could alter the glycosylation process and possibly contribute to immune evasion at infection.



Site #	Position	Sequence	N-Glyc result
1	61	NVTW	+++
2	74	NGTK	++
3	234	NITR	+++
4	282	NGTI	++
5	616	NCTE	++
6	717	NFTI	++
7	1194	NESL	++
8*	17	NFTN	++

Figure 4.4 Glycosylation sites mostly conserved among spike variants except for an additional site in the Gamma variant.

N-glycosylation sites on Wuhan reference and nine variant SARS-CoV-2 amino acid sequences were predicted using the NetNGlyc 1.0 platform. Predicted sites were filtered to show only those with (++) and (+++) NetNGlyc scores, revealing seven high-scoring sites conserved across variants Alpha, Beta, Gamma, Delta, Epsilon, Kappa, B.1.617.3, and mink-related spike. Along with three additional sites not found in the other variants, the Gamma variant had a higher potential glycosylation score at site 17. The figure shows a representative spike monomer with glycans in blue, and the corresponding peptide sequence residues in red at the predicted sites. The model was generated in PyMOL (The PyMOL Molecular Graphics System, Version 2.3.5, Schrödinger, LLC). Glycosylation site 7 at residue 1194 is associated with the membrane-spanning region of the spike protein and is not shown in this model. *Site 8 describes the additional high-scoring glycosylation site in the Gamma variant only.

Table 4.3 N-glycosylation sites and scores predicted by NetNGlyc server (threshold 0.5) across wild-type and SARS-CoV-2 variant spike sequences, showing a different predicted glycosylation pattern on the P.1 variant.

Site	Wuhan	D614G	*Alpha/ B.1.1.7	*Beta/ B.1.351	Gamma/ P.1	Delta/ B.1.617.2	Epsilon/ B.1.429	Kappa/ B.1.617.1	B.1.617.3	*Mink-related
17	0.6606 (+)	0.6606 (+)	0.6608 (+)	0.6607 (+)	0.6846 (++)	-	0.6508 (+)	0.6607 (+)	-	0.6607 (+)
20	-	-	-	-	0.5781 (+)	-	-	-	-	-
61	0.7820 (+++)	0.7820 (+++)	0.7798 (+++)	0.7821 (+++)	0.7820 (+++)	0.7820 (+++)	0.7820 (+++)	0.7820 (+++)	0.7820 (+++)	0.7798 (+++)
74	0.7192 (++)	0.7192 (++)	*0.6976 (++)	0.7292 (++)	0.7192 (++)	0.7192 (++)	0.7192 (++)	0.7192 (++)	0.7192 (++)	0.6976 (++)
122	0.6781 (+)	0.6782 (+)	0.6786 (+)	0.6782 (+)	0.6781 (+)	0.6782 (+)	0.6782 (+)	0.6780 (+)	0.6780 (+)	0.6786 (+)
149	0.6318 (+)	0.6318 (+)	0.6260 (+)	0.6321 (+)	0.6320 (+)	0.6264 (+)	0.6604 (+)	0.6339 (+)	0.6339 (+)	0.6321 (+)
165	0.6220 (+)	0.6220 (+)	0.6223 (+)	0.6221 (+)	0.6219 (+)	0.6446 (+)	0.6220 (+)	0.6220 (+)	0.6220 (+)	0.6223 (+)
188	-	-	-	-	0.5272 (+)	-	-	-	-	
234	0.7613 (+++)	0.7613 (+++)	0.7616 (+++)	0.7596 (+++)	0.7613 (+++)	0.7616 (+++)	0.7613 (+++)	0.7613 (+++)	0.7613 (+++)	0.7613 (+++)
282	0.7378 (++)	0.7379 (++)	0.7381 (++)	0.7381 (++)	0.7379 (++)	0.7381 (++)	0.7379 (++)	0.7377 (++)	0.7377 (++)	0.7380 (++)
331	0.5970 (+)	0.5970 (+)	0.5973 (+)	0.5976 (+)	0.5968 (+)	0.5974 (+)	0.5970 (+)	0.5972 (+)	0.5972 (+)	0.5971 (+)
343	0.5671 (+)	0.5671 (+)	0.5674 (+)	0.5674 (+)	0.5671 (+)	0.5674 (+)	0.5672 (+)	0.5671 (+)	0.5671 (+)	0.5674 (+)
603	0.5783 (+)	0.5783 (+)	0.5786 (+)	0.5786 (+)	0.5783 (+)	0.5786 (+)	0.5782 (+)	0.5783 (+)	0.5783 (+)	0.5784 (+)
616	0.7163 (++)	0.7279 (++)	0.7280 (++)	0.7281 (++)	0.7279 (++)	0.7281 (++)	0.7279 (++)	0.7279 (++)	0.7279 (++)	0.7279 (++)
657	-	-	-	-	0.5140 (+)	-	-	-	-	-
717	0.6426 (++)	0.6426 (++)	0.6086 (++)	0.6426 (++)	0.6426 (++)	0.6426 (++)	0.6426 (++)	0.6426 (++)	0.6426 (++)	0.6426 (++)
801	0.6146 (+)	0.6146 (+)	0.6149 (+)	0.6148 (+)	0.6146 (+)	0.6149 (+)	0.6146 (+)	0.6147 (+)	0.6147 (+)	0.6148 (+)
1098	0.5496 (+)	0.5494 (+)	0.5498 (+)	0.5496 (+)	0.5496 (+)	0.5498 (+)	0.5494 (+)	0.5496 (+)	0.5496 (+)	0.5497 (+)
1134	0.5800 (+)	0.5800 (+)	0.5802 (+)	0.5802 (+)	0.5800 (+)	0.5801 (+)	0.5802 (+)	0.5800 (+)	0.5800 (+)	0.5801 (+)
1194	0.6791 (++)	0.6790 (++)	0.6792 (++)	0.6791 (++)	0.6792 (++)	0.6791 (++)	0.6791 (++)	0.6791 (++)	0.6791 (++)	0.6790 (++)

* Site numbers may be shifted by 2-3 residues due to deletions

4.6 SARS-CoV-2 glycosylation sites may interfere with antibody binding

Along with modulating viral infectivity and sensitivity to innate immune mechanisms, antigen glycosylation can impact adaptive immune responses. Influenza virus hemagglutinin is a well-known example of antigen glycosylation interfering with the binding of neutralizing antibodies,

thus providing a fitness advantage to the virus [247]. With this under consideration, I aimed to determine whether predicted glycosylation patterns on the SARS-CoV-2 variant spike proteins could overlap with predicted B cell epitopes and possibly deter antibody binding. Both the 3D models and scored sequences obtained from DiscoTope 2.0 were observed side by side with the NetNGlyc glycosylation maps to examine possible interactions between regions where B cell epitopes and N-glycosylation sites were predicted to exist. Of the 26 B cell epitope regions identified across all variants, 7 of these 26 regions fell on top or close to a predicted glycosylation site (**Table 4.3**). For the purpose of this analysis, the sites 17, 61, 74, 122, 149, 165, 234, 282, 331, 343, 603, 616, 717, 801, 1098, 1134, and 1194 identified in the previous section were selected as possible candidates for glycosylation sites. Across the variants, twelve mutations were predicted to fall within these B cell epitope regions, and only one site was shown to carry a mutation within a potential glycosylation site, which were mutations at sites 18 and 20 of the Gamma variant (**Table 4.3**). Other mutations within site 19 were also recorded among the Delta and B.1.617.3 variants; however, glycosylation of this region was shown to be unlikely. Predicted B cell epitopes were compared to present immunoinformatic and immunological data, where four were found to have confirmed antibody binding affinity including sites adjacent to residues 229, 556, 581, and 809 [251]. All other residues predicted in this study have partial or complete agreement with immunoinformatic studies that used either DiscoTope 2.0 or Bepipred ([252], which includes [253,254] and others in a review) except for the predicted epitope next to residue 1100.

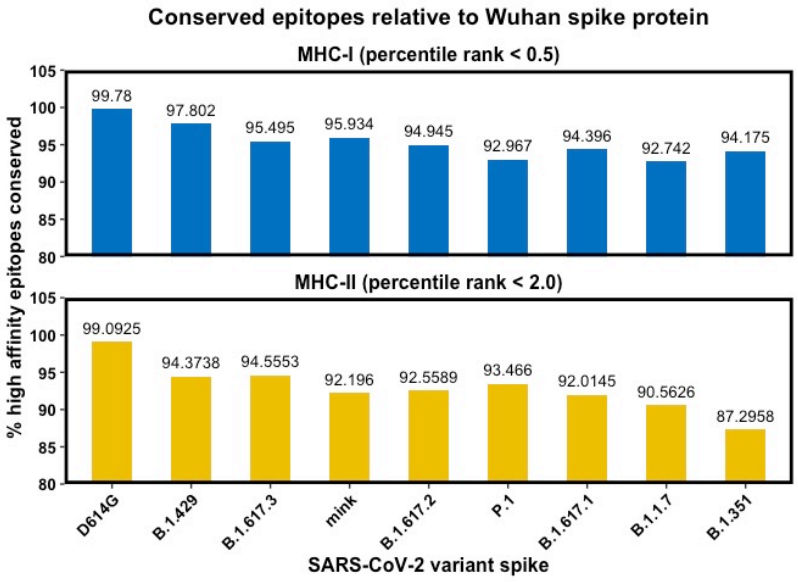
4.7 High-affinity MHC-I and MHC-II binding peptides are conserved across variants

In addition to altering B cell epitopes, amino acid mutations can lead to changes in antigenicity with respect to the specific peptides presented to host T cells. Specifically, T cell

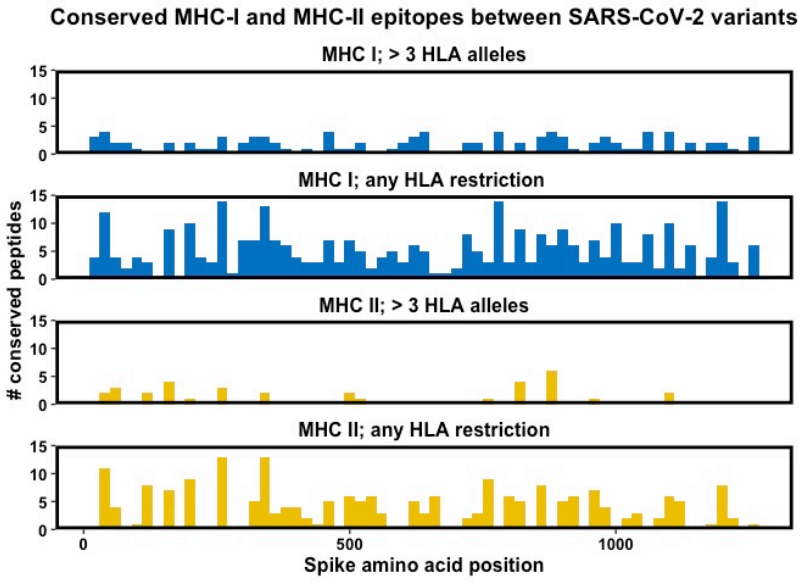
epitopes are linear peptides, consisting of approximately 8-11 amino acids for MHC-I and 13-17 amino acids for MHC-II, while B cell epitopes typically require a 3D confirmation of continuous or discontinuous amino acids for recognition [216]. Since both arms of the immune system are important during antigen encounter, I next assessed the amino acid substitutions and deletions across SARS-CoV-2 variants for changes in predicted T cell epitopes that may differentially stimulate host CD8+ and CD4+ T cells. Using the complete amino acid sequences for reference and variant spike, I generated MHC-I and MHC-II binding predictions using the Immune Epitope Database (IEDB) Analysis Resource NetMHCpan EL 4.1 and NetMHCII 2.3, respectively. For MHC-I peptide analysis, I restricted my results to 27 HLA-A and HLA-B alleles to encompass most of the human population [196]. Human HLA-DP, HLA-DR, and HLA-DQ alleles were selected for MHC-II analysis (see **Appendix 1** for NetMHCpan and NetMHCII output).

Overall, I found variants to differ slightly in the number of peptides predicted to bind to host MHC class I molecules, but that high-affinity epitopes were mostly conserved. After filtering the data such that peptides with the highest MHC-I binding affinity were selected for each group (percentile rank < 0.5 as done in [199]), I found that the number of HLA*peptide combinations ranged from 903-922 across the variants, which corresponded to 420-429 unique peptides. I found that despite this small variability in MHC-I epitopes, the variants ranged from 93% (Alpha) to 99% (D614G) conservation of high-affinity MHC-I epitopes **(Figure 4.5A)**, with 330 unique conserved peptides found. For relevance in the human population, I again filtered these conserved, high-affinity MHC-I peptides to reveal only those restricted to a minimum of three HLA alleles. Here I found 99 unique peptides spanning both the S1 and S2 subunits of the spike protein that are shared between all variants in my analysis and are predicted to bind several HLA alleles (**Figure 4.5B**) (**Table 4.4**).

A


Conserved epitopes relative to Wuhan spike protein
MHC-I (percentile rank < 0.5)
99.78
97.802
95.495
95.934
94.945
92.967
94.396
92.742
94.175
MHC-II (percentile rank < 2.0)
99.0925
94.3738
94.5553
92.196
92.5589
93.466
92.0145
90.5626
87.2958
% high affinity epitopes conserved
D614G
B.1.429
B.1.617.3
mink
B.1.617.2
P.1
B.1.617.1
B.1.1.7
B.1.351
SARS-CoV-2 variant spike

B


Conserved MHC-I and MHC-II epitopes between SARS-CoV-2 variants
MHC I; > 3 HLA alleles
MHC I; any HLA restriction
MHC II; > 3 HLA alleles
MHC II; any HLA restriction
conserved peptides
Spike amino acid position

C

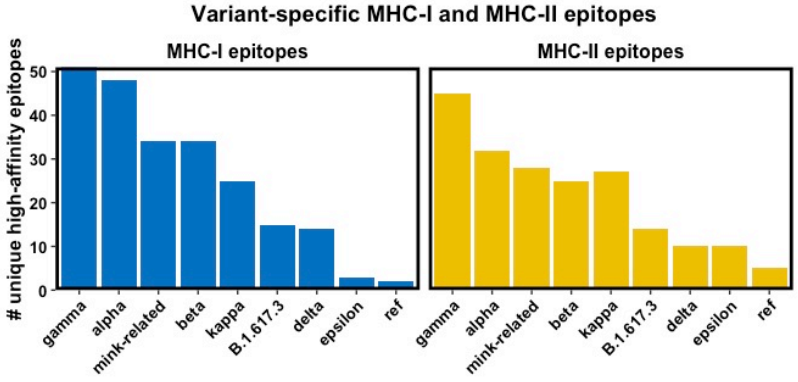

Variant-specific MHC-I and MHC-II epitopes
MHC-I epitopes
MHC-II epitopes
unique high-affinity epitopes
gamma
alpha
mink-related
beta
kappa
B.1.617.3
delta
epsilon
ref

Figure 4.5 High MHC I and II peptide conservation among SARS-CoV-2 variants, with the Gamma variant exhibiting the most diversity.

MHC class I and II epitopes were predicted using the NetMHCpan EL 4.1 and NetMHCII 2.3 servers on the Immune Epitope Database, respectively. High-affinity epitopes (HLA*peptide combinations) were defined as scores falling below percentile rank 0.5 for MHC-I, and scores below 2.0 for MHC-II epitopes. **(A)** The percentage of high-affinity epitopes shared between each variant and the Wuhan SARS-CoV-2 reference was calculated. **(B)** Of the conserved MHC-I and MHC-II epitopes identified, the frequency of unique peptides falling within 20-amino acid bins was plotted against the spike protein sequence to show epitope conservation across the protein. **(C)** The number of high-affinity epitopes specific to each variant (i.e., not shared with the Wuhan sequence or other variants) was identified, indicating that the Gamma and Alpha variants may have the most unique T cell epitope signature. The figures were generated in RStudio using the ggplot2 package version 3.3.2.

Table 4.4 High-affinity conserved peptides (percentile rank < 0.5 for MHC-I and < 2.0 for MHC-II) occurring on a minimum of three HLA alleles for SARS-CoV-2 spike variants Alpha, Beta, Gamma, Delta, Kappa, Epsilon, B.1.617.3, and mink-related spike.

MHC I peptides, minimum 3 HLA alleles					
#	Start residue	Peptide	Score	Percentile rank	Alleles
1	1016	AEIRASANL	0.676171-0.976223	0.01-0.11	HLA-B*40:01, HLA-B*44:03, HLA-B*44:02
2	1099	GTHWFVTQR	0.422018-0.927106	0.01-0.36	HLA-A*31:01, HLA-A*68:01, HLA-A*11:01, HLA-A*33:01, HLA-A*03:01
3	625	HADQLTPTW	0.249722-0.990462	0.01-0.45	HLA-B*58:01, HLA-B*53:01, HLA-B*57:01, HLA-B*35:01, HLA-A*32:01, HLA-A*01:01
4	896	IPFAMQMAY	0.328521-0.994479	0.01-0.43	HLA-B*35:01, HLA-B*53:01, HLA-B*51:01
5	458	KSNLKPFER	0.283834-0.935717	0.01-0.47	HLA-A*31:01, HLA-A*11:01, HLA-A*30:01
6	56	LPFFSNVTW	0.488507-0.98139	0.01-0.33	HLA-B*53:01, HLA-B*35:01, HLA-B*51:01, HLA-B*58:01
7	84	LPFNDGVYF	0.329782-0.984594	0.01-0.43	HLA-B*35:01, HLA-B*53:01, HLA-B*51:01, HLA-B*07:02
8	865	LTDEMIAQY	0.169613-0.997196	0.01-0.5	HLA-A*01:01, HLA-A*30:02, HLA-A*26:01, HLA-B*35:01, HLA-B*58:01, HLA-B*53:01
9	777	NTQEVFAQV	0.162889-0.975719	0.01-0.45	HLA-A*68:02, HLA-A*02:06, HLA-A*26:01
10	454	RLFRKSNLK	0.576303-0.978678	0.01-0.26	HLA-A*03:01, HLA-A*30:01, HLA-A*11:01, HLA-A*31:01
11	1264	VLKGVKLHY	0.153447-0.943479	0.01-0.48	HLA-B*15:01, HLA-A*30:02, HLA-A*32:01, HLA-A*03:01, HLA-A*30:01, HLA-A*26:01
12	1052	FPQSAPHGV	0.380916-0.853165	0.02-0.29	HLA-B*51:01, HLA-B*53:01, HLA-B*07:02, HLA-B*35:01

MHC I peptides, minimum 3 HLA alleles					
#	Start residue	Peptide	Score	Percentile rank	Alleles
13	718	FTISVTTEI	0.332779-0.923585	0.02-0.4	HLA-A*68:02, HLA-A*02:06, HLA-A*26:01, HLA-A*02:03, HLA-B*51:01, HLA-A*02:01
14	787	QIYKTPPIK	0.631297-0.947791	0.02-0.07	HLA-A*03:01, HLA-A*30:01, HLA-A*11:01
15	321	QPTESIVRF	0.472976-0.947034	0.02-0.27	HLA-B*35:01, HLA-B*53:01, HLA-B*51:01, HLA-B*07:02
16	349	SVYAWNRKR	0.596094-0.919617	0.02-0.21	HLA-A*31:01, HLA-A*33:01, HLA-A*68:01, HLA-A*03:01, HLA-A*11:01
17	302	TLKSFTVEK	0.491043-0.926974	0.02-0.43	HLA-A*03:01,HLA-A*30:01, HLA-A*11:01, HLA-A*31:01, HLA-A*68:01
18	604	TSNQVAVLY	0.207735-0.917927	0.02-0.41	HLA-A*30:02, HLA-A*01:01, HLA-A*26:01, HLA-B*35:01, HLA-B*58:01, HLA-B*57:01, HLA-B*15:01, HLA-B*53:01
19	269	YLQPRTFLL	0.284271-0.971198	0.02-0.33	HLA-A*02:01, HLA-B*08:01, HLA-A*02:03, HLA-A*02:06, HLA-A*32:01, HLA-A*23:01, HLA-A*24:02
20	372	ASFSTFKCY	0.29699-0.77457	0.03-0.46	HLA-A*30:02, HLA-B*15:01, HLA-A*01:01, HLA-B*57:01, HLA-B*58:01
21	780	EVFAQVKQI	0.374701-0.865844	0.03-0.34	HLA-A*68:02, HLA-A*26:01, HLA-B*51:01
22	192	FVFKNIDGY	0.232971-0.869689	0.03-0.49	HLA-A*26:01, HLA-B*35:01, HLA-B*15:01, HLA-A*30:02, HLA-A*01:01
23	786	KQIYKTPPIK	0.414021-0.916015	0.03-0.4	HLA-A*03:01, HLA-A*30:01, HLA-A*11:01
24	821	LLFNKVTLA	0.560973-0.890132	0.03-0.19	HLA-A*02:03, HLA-A*02:01, HLA-A*02:06
25	229	LPIGINITRF	0.329148-0.828081	0.03-0.43	HLA-B*53:01, HLA-B*35:01, HLA-B*51:01
26	109	TLDSKTQSL	0.730577-0.914998	0.03-0.1	HLA-A*02:01, HLA-B*08:01, HLA-A*02:03, HLA-A*02:06
27	36	VYYPDKVFR	0.292317-0.892155	0.03-0.44	HLA-A*31:01, HLA-A*33:01, HLA-A*30:01
28	258	WTAGAAAYY	0.331721-0.842739	0.03-0.39	HLA-A*26:01, HLA-A*01:01, HLA-A*30:02, HLA-B*35:01
29	989	AEVQIDRLI	0.758249-0.89662	0.04-0.13	HLA-B*44:03, HLA-B*44:02, HLA-B*40:01
30	361	CVADYSVLY	0.456493-0.813057	0.04-0.36	HLA-A*26:01, HLA-A*01:01, HLA-A*30:02, HLA-B*35:01, HLA-B*15:01
31	725	EILPVSMTK	0.269327-0.945141	0.04-0.49	HLA-A*68:01, HLA-A*11:01, HLA-A*03:01, HLA-A*33:01
32	815	RSFIEDLLF	0.479418-0.933645	0.04-0.11	HLA-B*58:01, HLA-B*57:01, HLA-A*32:01

MHC I peptides, minimum 3 HLA alleles					
#	Start residue	Peptide	Score	Percentile rank	Alleles
33	634	RVYSTGSNVF	0.291524-0.851186	0.04-0.3	HLA-B*15:01, HLA-A*32:01, HLA-A*24:02, HLA-A*23:01
34	1137	VYDPLQPEL	0.479112-0.851829	0.04-0.16	HLA-A*24:02, HLA-A*23:01, HLA-B*08:01
35	958	ALNTLVKQL	0.165467-0.84284	0.05-0.5	HLA-A*02:03, HLA-A*02:01, HLA-A*32:01, HLA-A*02:06
36	1262	EPVLKGVKL	0.261209-0.740861	0.05-0.42	HLA-B*08:01, HLA-B*07:02, HLA-B*51:01, HLA-B*53:01, HLA-B*35:01
37	1048	HLMSFPQSA	0.601563-0.845506	0.05-0.17	HLA-A*02:03, HLA-A*02:01, HLA-A*02:06
38	41	KVFRSSVLH	0.359703-0.859868	0.05-0.33	HLA-A*03:01, HLA-A*30:01, HLA-A*11:01, HLA-A*30:02
39	1000	RLQSLQTYV	0.151972-0.87376	0.05-0.49	HLA-A*02:01, HLA-A*02:03, HLA-A*02:06, HLA-A*32:01
40	1060	VVFLHVTYV	0.162221-0.850041	0.05-0.46	HLA-A*02:03, HLA-A*02:06, HLA-A*68:02, HLA-A*02:01", HLA-B*51:01, HLA-A*32:01
41	28	YTNSFTRGV	0.346-0.823738	0.05-0.42	HLA-A*68:02, HLA-A*02:06, HLA-A*02:03
42	344	ATRFASVYAW	0.515027-0.954826	0.06-0.16	HLA-B*57:01, HLA-A*32:01, HLA-B*58:01
43	35	GVYYPDKVFR	0.330452-0.811444	0.06-0.37	HLA-A*31:01, HLA-A*11:01, HLA-A*68:01, HLA-A*33:01, HLA-A*03:01
44	878	LAGTITSGW	0.532125-0.908875	0.06-0.12	HLA-B*58:01, HLA-B*57:01, HLA-B*53:01
45	962	LVKQLSSNF	0.177684-0.818281	0.06-0.43	HLA-B*15:01, HLA-A*26:01, HLA-A*32:01
46	983	RLDKVEAEV	0.403746-0.825045	0.06-0.34	HLA-A*02:01, HLA-A*02:06, HLA-A*02:03
47	1185	RLNEVAKNL	0.513013-0.788651	0.06-0.16	HLA-A*02:03, HLA-A*32:01, HLA-A*02:01
48	1136	TVYDPLQPEL	0.210208-0.758137	0.06-0.42	HLA-A*68:02, HLA-A*02:06, HLA-A*02:03, HLA-A*02:01, HLA-A*26:01, HLA-A*24:02
49	1094	VFVSNGTHW	0.168595-0.754597	0.06-0.5	HLA-A*23:01, HLA-A*24:02, HLA-B*58:01, HLA-B*53:01
50	1220	FIAGLIAIV	0.501049-0.76346	0.07-0.19	HLA-A*02:03, HLA-A*02:06, HLA-A*02:01, HLA-A*68:02
51	339	GEVFNATRF	0.682004-0.828162	0.07-0.16	HLA-B*44:03, HLA-B*44:02, HLA-B*40:01
52	624	IHADQLTPTW	0.199039-0.886293	0.07-0.43	HLA-B*58:01, HLA-B*57:01, HLA-B*53:01, HLA-A*23:01
53	202	KIYSKHTPI	0.324637-0.553766	0.07-0.32	HLA-A*32:01, HLA-A*30:01, HLA-A*02:03, HLA-B*08:01

MHC I peptides, minimum 3 HLA alleles					
#	Start residue	Peptide	Score	Percentile rank	Alleles
54	637	STGSNVFQTR	0.343407-0.89597	0.07-0.43	HLA-A*68:01, HLA-A*31:01, HLA-A*33:01, HLA-A*11:01
55	83	VLPFNDGVYF	0.387438-0.63303	0.07-0.44	HLA-B*53:01, HLA-B*35:01, HLA-B*15:01
56	320	VQPTESIVRF	0.267127-0.612597	0.07-0.39	HLA-B*53:01, HLA-B*15:01, HLA-A*23:01, HLA-A*24:02, HLA-B*35:01
57	880	GTITSGWTF	0.209008-0.811974	0.08-0.41	HLA-A*32:01, HLA-B*58:01, HLA-B*57:01, HLA-B*15:01, HLA-A*23:01
58	603	NTSNQVAVLY	0.323431-0.744843	0.08-0.38	HLA-A*01:01, HLA-A*26:01, HLA-A*30:02
59	464	FERDISTEI	0.23108-0.84839	0.09-0.44	HLA-B*40:01, HLA-B*44:03, HLA-B*44:02
60	894	LQIPFAMQM	0.20736-0.763797	0.09-0.37	HLA-B*15:01, HLA-A*02:06, HLA-A*32:01
61	628	QLTPTWRVY	0.177798-0.724015	0.09-0.43	HLA-A*30:02, HLA-B*15:01, HLA-A*32:01
62	162	SANNCTFEY	0.570386-0.753348	0.09-0.15	HLA-A*30:02, HLA-B*35:01, HLA-A*01:01
63	987	VEAEVQIDRL	0.205553-0.832715	0.09-0.48	HLA-B*40:01, HLA-B*44:03, HLA-B*44:02
64	574	DAVRDPQTL	0.336628-0.639281	0.1-0.32	HLA-B*51:01, HLA-B*53:01, HLA-B*35:01
65	898	FAMQMAYRF	0.226168-0.68794	0.1-0.38	HLA-B*53:01, HLA-B*35:01, HLA-B*58:01, HLA-A*23:01
66	1181	KEIDRLNEV	0.389617-0.808555	0.1-0.4	HLA-B*40:01, HLA-B*44:02, HLA-B*44:03, HLA-A*02:06
67	733	KTSVDCTMY	0.34981-0.642966	0.1-0.47	HLA-A*30:02, HLA-A*01:01, HLA-B*58:01
68	56	LPFFSNVTWF	0.292389-0.648367	0.1-0.47	HLA-B*53:01, HLA-B*35:01, HLA-B*51:01
69	249	LTPGDSSSGW	0.530312-0.6494	0.1-0.35	HLA-B*53:01, HLA-B*58:01, HLA-B*57:01
70	1054	QSAPHGVVF	0.219646-0.815153	0.1-0.4	HLA-B*15:01, HLA-B*58:01, HLA-B*35:01, HLA-B*57:01, HLA-A*32:01, HLA-A*26:01, HLA-B*53:01
71	634	RVYSTGSNV	0.23705-0.543326	0.1-0.35	HLA-A*30:01, HLA-A*02:03, HLA-A*32:01, HLA-A*02:06
72	28	YTNSFTRGVY	0.391325-0.661513	0.1-0.18	HLA-A*01:01, HLA-A*30:02, HLA-A*26:01
73	304	KSFTVEKGIY	0.378169-0.653254	0.11-0.43	HLA-A*30:02, HLA-B*57:01, HLA-B*58:01
74	1261	SEPVLKGVKL	0.241206-0.718257	0.11-0.45	HLA-B*07:02, HLA-B*40:01, HLA-B*44:02, HLA-B*44:03
75	30	NSFTRGVYY	0.492899-0.597285	0.12-0.22	HLA-A*30:02, HLA-A*26:01, HLA-A*01:01, HLA-B*35:01
76	915	VLYENQKLI	0.171258-0.648983	0.12-0.44	HLA-A*02:03, HLA-A*02:01, HLA-A*32:01

MHC I peptides, minimum 3 HLA alleles					
#	Start residue	Peptide	Score	Percentile rank	Alleles
77	529	KSTNLVKNK	0.495913-0.588891	0.14-0.26	HLA-A*30:01, HLA-A*11:01, HLA-A*03:01
78	1095	FVSNGTHWF	0.210819-0.431696	0.15-0.41	HLA-A*26:01, HLA-B*35:01, HLA-B*53:01
79	861	LPPLLTDEM	0.257301-0.608235	0.16-0.49	HLA-B*35:01, HLA-B*53:01, HLA-B*51:01
80	487	NCYFPLQSY	0.315549-0.605847	0.16-0.24	HLA-B*35:01, HLA-A*26:01, HLA-A*30:02
81	1192	NLNESLIDL	0.343239-0.618877	0.18-0.48	HLA-A*02:01, HLA-A*02:03, HLA-A*02:06
82	995	RLITGRLQSL	0.208067-0.576159	0.18-0.37	HLA-A*02:03, HLA-A*02:01, HLA-A*32:01
83	748	ECSNLLLQY	0.336613-0.378678	0.19-0.34	HLA-A*26:01, HLA-A*01:01, HLA-B*35:01
84	464	FERDISTEIY	0.437515-0.586808	0.2-0.35	HLA-B*44:03, HLA-B*44:02, HLA-B*15:01
85	424	KLPDDFTGCV	0.532999-0.562841	0.2-0.23	HLA-A*02:03, HLA-A*02:06, HLA-A*02:01
86	366	SVLYNSASF	0.269331-0.470989	0.21-0.38	HLA-A*32:01, HLA-A*26:01, HLA-B*15:01, HLA-B*35:01
87	47	VLHSTQDLF	0.191875-0.609724	0.21-0.5	HLA-B*15:01, HLA-A*32:01, HLA-A*24:02
88	1059	GVVFLHVTY	0.200689-0.585418	0.22-0.38	HLA-B*15:01, HLA-A*30:02, HLA-A*26:01, HLA-A*32:01
89	311	GIYQTSNFR	0.540742-0.587274	0.25-0.5	HLA-A*31:01, HLA-A*11:01, HLA-A*03:01, HLA-A*68:01
90	1093	GVFVSNGTHW	0.251174-0.757943	0.25-0.31	HLA-B*57:01, HLA-B*58:01, HLA-A*32:01
91	257	GWTAGAAAYY	0.218071-0.411498	0.26-0.34	HLA-A*30:02, HLA-A*01:01, HLA-A*26:01
92	638	TGSNVFQTR	0.397693-0.603019	0.27-0.47	HLA-A*33:01, HLA-A*31:01, HLA-A*68:01
93	1200	LQELGKYEQY	0.193472-0.505374	0.29-0.49	HLA-B*15:01, HLA-B*44:03, HLA-B*44:02
94	506	QPYRVVVLSF	0.280366-0.451783	0.29-0.45	HLA-B*07:02, HLA-B*53:01, HLA-B*35:01
95	879	AGTITSGWTF	0.169346-0.689438	0.31-0.44	HLA-B*57:01, HLA-B*58:01, HLA-A*32:01
96	869	MIAQYTSAL	0.33241-0.360459	0.31-0.43	HLA-B*08:01, HLA-A*68:02, HLA-A*02:03, HLA-B*07:02
97	516	ELLHAPATV	0.317281-0.356044	0.33-0.48	HLA-A*68:02, HLA-A*02:03, HLA-A*02:01
98	877	LLAGTITSGW	0.185323-0.633888	0.36-0.42	HLA-B*57:01, HLA-B*58:01, HLA-A*32:01
99	168	FEYVSQPFLM	0.21741-0.359256	0.39-0.45	HLA-A*23:01, HLA-B*40:01, HLA-A*24:02

MHC II peptides, minimum 3 HLA alleles					
#	Start residue	Peptide	Affinity (nM)	Percentile rank	Alleles
1	113	KTQSLLIVNNATNVV	1.5-44.3	0.01-1.9	DRB1_1302, DRB3_0202, DRB1_0405
2	114	TQSLLIVNNATNVVI	1.4-44.1	0.01-1.9	DRB1_1302, DRB3_0202, DRB1_0405

MHC II peptides, minimum 3 HLA alleles					
#	Start residue	Peptide	Affinity (nM)	Percentile rank	Alleles
3	338	FGEVFNATRFASVYA	11.6-4.7	0.02-1.5	HLA-DPA10201-DPB10101, HLA-DPA10103-DPB10401, HLA-DPA10301-DPB10402
4	339	GEVFNATRFASVYAW	15.3-5.5	0.05-1.5	HLA-DPA10201-DPB10101, HLA-DPA10103-DPB10401, HLA-DPA10301-DPB10402
5	885	GWTFGAGAALQIPFA	11.8-73.9	0.08-1.9	HLA-DQA10501-DQB10301, DRB1_0901, DRB1_0101, HLA-DQA10301-DQB10302, HLA-DQA10102-DQB10602, DRB1_0701
6	511	VVLSFELLHAPATVC	2.9-41.2	0.08-1.8	DRB1_0101, DRB1_0405, HLA-DPA10301-DPB10402
7	884	SGWTFGAGAALQIPF	11-84.9	0.09-1.9	HLA-DQA10501-DQB10301, DRB1_0901, HLA-DQA10301-DQB10302, DRB1_0101, HLA-DQA10401-DQB10402, DRB1_0701, HLA-DQA10102-DQB10602
8	507	PYRVVVLSFELLHAP	11.9-70.7	0.15-1.9	HLA-DPA10301-DPB10402, HLA-DPA10201-DPB11401, HLA-DPA10201-DPB10101, HLA-DPA10103-DPB10201
9	883	TSGWTFGAGAALQIP	11.5-8.5	0.15-1.9	HLA-DQA10501-DQB10301, DRB1_0901, DRB1_0701, DRB1_0101
10	886	WTFGAGAALQIPFAM	15.1-66.8	0.15-1.8	HLA-DQA10501-DQB10301, DRB1_0901, DRB1_0101, HLA-DQA10102-DQB10602, HLA-DQA10301-DQB10302
11	814	KRSFIEDLLFNKVTL	14.4-93.3	0.25-1.2	HLA-DPA10301-DPB10402, HLA-DPA10201-DPB10101, HLA-DPA10103-DPB10401, HLA-DPA10201-DPB10501
12	506	QPYRVVVLSFELLHA	15.2-63.8	0.25-1.4	HLA-DPA10301-DPB10402, HLA-DPA10201-DPB11401, HLA-DPA10201-DPB10101
13	34	RGVYYPDKVFRSSVL	36.9-6.9	0.25-1.8	DRB3_0101, DRB1_0401, DRB1_0301
14	255	SSGWTAGAAAYYVGY	11.3-58.9	0.25-1	DRB1_0901, HLA-DQA10501-DQB10301, HLA-DQA10102-DQB10602
15	254	SSSGWTAGAAAYYVG	11-72.7	0.25-1.5	DRB1_0901, HLA-DQA10501-DQB10301, HLA-DQA10102-DQB10602

MHC II peptides, minimum 3 HLA alleles					
#	Start residue	Peptide	Affinity (nM)	Percentile rank	Alleles
16	761	TQLNRALTGIAVEQD	207.2-94.9	0.25-1.7	HLA-DQA10401-DQB10402, HLA-DQA10301-DQB10302, HLA-DQA10501-DQB10201
17	33	TRGVYYPDKVFRSSV	38.9-7.1	0.25-1.7	DRB3_0101, DRB1_0401, DRB1_0301
18	815	RSFIEDLLFNKVTLA	15.4-66.4	0.3-1.5	HLA-DPA10301-DPB10402, HLA-DPA10201-DPB10101, HLA-DPA10103-DPB10401
19	882	ITSGWTFGAGAALQI	13.9-293.4	0.4-1.8	DRB1_0901, HLA-DQA10501-DQB10301, DRB1_0701, HLA-DQA10301-DQB10302
20	256	SGWTAGAAAYYVGYL	13.2-67.2	0.4-1.3	DRB1_0901, HLA-DQA10501-DQB10301, HLA-DQA10102-DQB10602
21	813	SKRSFIEDLLFNKVT	17.5-92.8	0.4-1.3	HLA-DPA10301-DPB10402, HLA-DPA10201-DPB10101, HLA-DPA10201-DPB10501, HLA-DPA10103-DPB10401
22	887	TFGAGAALQIPFAMQ	21.3-86.3	0.4-1.9	HLA-DQA10501-DQB10301, DRB1_0901, HLA-DQA10102-DQB10602
23	166	CTFEYVSQPFLMDLE	18.8-65.2	0.5-1.6	HLA-DPA10301-DPB10402, HLA-DPA10103-DPB10401, HLA-DQA10501-DQB10201, HLA-DPA10201-DPB10101, HLA-DPA10103-DPB10201
24	165	NCTFEYVSQPFLMDL	20.6-64.7	0.5-1.4	HLA-DPA10301-DPB10402, HLA-DQA10501-DQB10201, HLA-DPA10103-DPB10401, HLA-DPA10103-DPB10201, HLA-DPA10201-DPB10101
25	167	TFEYVSQPFLMDLEG	19-81.0	0.5-1.7	HLA-DPA10301-DPB10402, HLA-DPA10103-DPB10401, HLA-DQA10501-DQB10201, HLA-DPA10201-DPB10101
26	958	ALNTLVKQLSSNFGA	21.4-85.8	0.6-1.7	DRB1_0401, DRB1_0101, DRB1_0802
27	199	GYFKIYSKHTPINLV	11.5-8.5	0.6-1.6	DRB1_1101, DRB1_0701, HLA-DPA10201-DPB11401
28	812	PSKRSFIEDLLFNKV	104.1-42.9	0.6-1.5	HLA-DPA10301-DPB10402, HLA-DPA10201-DPB10101, HLA-DPA10103-DPB10401, HLA-DPA10201-DPB10501

MHC II peptides, minimum 3 HLA alleles					
#	**Start residue**	**Peptide**	**Affinity (nM)**	**Percentile rank**	**Alleles**
29	164	NNCTFEYVSQPFLMD	29.4-56.0	0.8-1.6	HLA-DQA10501-DQB10201, HLA-DPA10301-DPB10402, HLA-DPA10103-DPB10401
30	1091	REGVFVSNGTHWFVT	17.7-7.4	1-1.5	DRB3_0101, DRB1_1302, DRB3_0202
31	53	DLFLPFFSNVTWFHA	29.1-69.5	1.1-1.6	DRB1_0401, HLA-DPA10103-DPB10401, HLA-DPA10201-DPB10101, HLA-DPA10103-DPB10201
32	1092	EGVFVSNGTHWFVTQ	17-8.9	1.1-1.7	DRB3_0101, DRB3_0202, DRB1_1302
33	54	LFLPFFSNVTWFHAI	13.3-67.6	1.1-1.8	DRB1_0401, HLA-DPA10103-DPB10401, DRB1_0701, HLA-DPA10201-DPB10101, HLA-DPA10103-DPB10201
34	52	QDLFLPFFSNVTWFH	27.1-69.2	1.1-1.7	HLA-DPA10103-DPB10401, HLA-DPA10201-DPB10101, DRB1_0401

Although there was a high amount of conservation for T cell epitopes, my analysis did detect several unique MHC-I-binding peptides per variant that were not predicted in the Wuhan reference sequence or other variants (**Table 4.5**). The Alpha variant had 14 unique peptides, corresponding to 48 predicted epitopes; the Beta variant had 17 peptides, which corresponded to 34 epitopes; Gamma had 17 peptides, corresponding to 51 epitopes; Delta had 7 unique peptides (14 epitopes); Kappa had 14 unique peptides (25 epitopes); Epsilon had only 3 peptides (3 epitopes); B.1.617.3 had 5 peptides (15 epitopes); and mink-related spike had 34 (15 epitopes) (**Figure 4.5C**, **Table 4.5**). I noted two epitopes occurring in the Wuhan reference sequence only, which corresponded to the D614G substitution present in the other spike sequences. Across variants, these lineage-specific peptides spanned both S1 and S2 subunits of spike. Of the high-affinity peptides specific to the Alpha variant, less than half were in the S1 subunit, and none spanned the receptor binding domain of spike (**Table 4.5**). The B.1.617.3 lineage had all five of its lineage-specific MHC-I peptides falling in the S2 subunit (**Table 4.5**). In contrast, the remaining

variant- specific MHC-I-binding peptides fell heavily in the S1 subunit, with several occurring in the receptor-binding domain. The Beta, Kappa, Epsilon, and mink-related variants all had most of their lineage-specific peptides in S1, with 3/17 peptides falling in the RBD of Beta; 7/26 in the RBD of Gamma; 8/15 in the RBD of mink-related spike; but none in the RBD of Kappa and Epsilon variants (**Table 4.5**). Although all seven of Delta's lineage-specific peptides fell in S1, none spanned the RBD. With respect to potential CD8+ T cell activation, the Alpha, Epsilon, and Kappa variants appear antigenically similar to reference SARS-CoV-2 across the RBD of spike. However, Beta, Gamma, and mink-related lineages show RBDs that may have antigenic distinctions from the reference. With the exception of Alpha and B.1.617.3, I found the MHC-I binding peptides derived from the S2 subunit of spike to be more conserved between variant and reference spike proteins; that is, SARS-CoV-2 spike variants may have more divergent antigenic presentations with respect to MHC-I epitopes derived from the S1 subunit.

Table 4.5 High-affinity, lineage-specific MHC-I and MHC-II peptides predicted for SARS-CoV-2 spike variants using Immune Epitope Database tools NetMHCII 3.2 and NetMHCpan EL 4.1.

	MHC class I (percentile rank < 0.5)		MHC class II (percentile rank < 2.0)	
	Site	**Peptide**	**Site**	**Peptide**
Wuhan reference spike	612	YQDVNCTEV	606	NQVAVLYQDVNCTEV
			600	PGTNTSNQVAVLYQD
			601	GTNTSNQVAVLYQDV
			602	TNTSNQVAVLYQDVN
			603	NTSNQVAVLYQDVNC
	Site	**Peptide**	**Site**	**Peptide**
Alpha (B.1.1.7)	973	VLNDILARL	965	SNFGAISSVLNDILA
	711	IPINFTISV	709	IAIPINFTISVTTEI
	709	IAIPINFTI	710	AIPINFTISVTTEIL
	672	QTQTNSHRR	711	IPINFTISVTTEILP
	141	VYHKNNKSW	712	PINFTISVTTEILPV
	972	SVLNDILAR	713	INFTISVTTEILPVS
	674	QTNSHRRAR	699	ENSVAYSNNSIAIPI
	1110	QIITTHNTF	700	NSVAYSNNSIAIPIN
	140	GVYHKNNKSW	701	SVAYSNNSIAIPINF
	707	NSIAIPINF	136	DPFLGVYHKNNKSWM
	972	SVLNDILARL	137	PFLGVYHKNNKSWME
	676	NSHRRARSV	138	FLGVYHKNNKSWMES
	670	SYQTQTNSHR	139	LGVYHKNNKSWMESE
	708	SIAIPINFTI	140	GVYHKNNKSWMESEF
			702	VAYSNNSIAIPINFT
			678	HRRARSVASQSIIAY
			1101	VTQRNFYEPQIITTH
			1102	TQRNFYEPQIITTHN
			129	CEFQFCNDPFLGVYH
			130	EFQFCNDPFLGVYHK
	Site	**Peptide**	**Site**	**Peptide**
Beta (B.1.351)	696	LGVENSVAY	688	SIIAYTMSLGVENSV
	235	ITRFQTLHR	689	IIAYTMSLGVENSVA
	78	RFANPVLPF	690	IAYTMSLGVENSVAY

	MHC class I (percentile rank < 0.5)		MHC class II (percentile rank < 2.0)	
	Site	**Peptide**	**Site**	**Peptide**
Beta (B.1.351)	237	RFQTLHRSY	684	VASQSIIAYTMSLGV
	695	SLGVENSVAY	685	ASQSIIAYTMSLGVE
	208	TPINLVRGL	686	SQSIIAYTMSLGVEN
	414	NIADYNYKL	687	QSIIAYTMSLGVENS
	410	GQTGNIADY	691	AYTMSLGVENSVAYS
	212	LVRGLPQGF	207	HTPINLVRGLPQGFS
	690	IAYTMSLGV	697	GVENSVAYSNNSIAI
	215	GLPQGFSAL	698	VENSVAYSNNSIAIP
	411	QTGNIADYNY	215	GLPQGFSALEPLVDL
	694	MSLGVENSV	73	TNGTKRFANPVLPFN
	236	TRFQTLHRSY	74	NGTKRFANPVLPFND
	238	FQTLHRSYL	201	FKIYSKHTPINLVRG
	234	NITRFQTLHR		
	77	KRFANPVLPF		
	Site	**Peptide**	**Site**	**Peptide**
Gamma (P.1)	24	LPSAYTNSF	6	EMFVFLVLLPLVSSQ
	26	SAYTNSFTR	10	FLVLLPLVSSQCVNF
	1173	NASFVNIQK	1020	RAAEIRASANLAAIK
	416	GTIADYNYK	1021	AAEIRASANLAAIKM
	1020	ASANLAAIK	11	LVLLPLVSSQCVNFT
	1168	DISGINASF	12	VLLPLVSSQCVNFTN
	138	YPFLGVYYH	13	LLPLVSSQCVNFTNR
	413	GQTGTIADY	196	SEFVFKNIDGYFKIY
	417	TIADYNYKL	1019	IRAAEIRASANLAAI
	652	GAEYVNNSY	1025	RASANLAAIKMSECV
	415	TGTIADYNYK	1026	ASANLAAIKMSECVL
	133	FQFCNYPFL	1027	SANLAAIKMSECVLG
	414	QTGTIADYNY	189	QGNFKNLSEFVFKNI
	13	SQCVNFTNR	132	VVIKVCEFQFCNYPF
	1168	DISGINASFV	133	VIKVCEFQFCNYPFL
	651	IGAEYVNNSY	134	IKVCEFQFCNYPFLG
	411	APGQTGTIA	135	KVCEFQFCNYPFLGV
	1021	SANLAAIKM	136	VCEFQFCNYPFLGVY

	MHC class I (percentile rank < 0.5)		MHC class II (percentile rank < 2.0)	
	Site	**Peptide**	**Site**	**Peptide**
Gamma (P.1)	23	QLPSAYTNSF	137	CEFQFCNYPFLGVYY
	647	AGCLIGAEY	138	EFQFCNYPFLGVYYH
	129	KVCEFQFCNY	139	FQFCNYPFLGVYYHK
	137	NYPFLGVYY	140	QFCNYPFLGVYYHKN
	410	IAPGQTGTI	141	FCNYPFLGVYYHKNN
	136	CNYPFLGVY	142	CNYPFLGVYYHKNNK
	646	RAGCLIGAEY	143	NYPFLGVYYHKNNKS
	187	KNLSEFVFK	188	KQGNFKNLSEFVFKN
			190	GNFKNLSEFVFKNID
			191	NFKNLSEFVFKNIDG
			192	FKNLSEFVFKNIDGY
			195	LSEFVFKNIDGYFKI
			1022	AEIRASANLAAIKMS
	Site	**Peptide**	**Site**	**Peptide**
Delta (B.1.617.2)	135	FCNDPFLDYY	139	PFLDYYHKNNKSWME
	149	KSWMESGVY	140	FLDYYHKNNKSWMES
	136	CNDPFLDYY	141	LDYYHKNNKSWMESG
	135	FCNDPFLDY	142	DYYHKNNKSWMESGV
	134	QFCNDPFLDY	130	VCEFQFCNDPFLDYY
	142	DYYHKNNKSW	131	CEFQFCNDPFLDYYH
	475	KPCNGVEGF	132	EFQFCNDPFLDYYHK
			129	KVCEFQFCNDPFLDY
	Site	**Peptide**	**Site**	**Peptide**
Epsilon (B.1.429)	8	LPLVSIQCV	140	FLGVYYHKNNKSCME
	5	LVLLPLVSI	141	LGVYYHKNNKSCMES
	152	CMESEFRVY	142	GVYYHKNNKSCMESE
			143	VYYHKNNKSCMESEF
			3	VFLVLLPLVSIQCVN
			4	FLVLLPLVSIQCVNL
			5	LVLLPLVSIQCVNLT
			6	VLLPLVSIQCVNLTT
			7	LLPLVSIQCVNLTTR
			8	LPLVSIQCVNLTTRT

	MHC class I (percentile rank < 0.5)		MHC class II (percentile rank < 2.0)	
	Site	**Peptide**	**Site**	**Peptide**
Kappa (B.1.617.1)	1065	VTYVPAHEK	1058	HGVVFLHVTYVPAHE
	1066	TYVPAHEKNF	1059	GVVFLHVTYVPAHEK
	89	GVYFASIEK	1060	VVFLHVTYVPAHEKN
	1064	HVTYVPAHEK	1070	AHEKNFTTAPAICHD
	142	DVYYHKNNK	1071	HEKNFTTAPAICHDG
	95	IEKSNIIRGW	89	GVYFASIEKSNIIRG
	88	DGVYFASIEK	140	FLDVYYHKNNKSWME
	135	FCNDPFLDVY	141	LDVYYHKNNKSWMES
	136	CNDPFLDVY	142	DVYYHKNNKSWMESE
	136	CNDPFLDVYY	84	LPFNDGVYFASIEKS
	138	DPFLDVYYHK	129	KVCEFQFCNDPFLDV
	92	FASIEKSNI	130	VCEFQFCNDPFLDVY
	138	DPFLDVYYH	131	CEFQFCNDPFLDVYY
	1067	YVPAHEKNF	132	EFQFCNDPFLDVYYH
			133	FQFCNDPFLDVYYHK
	Site	**Peptide**	**Site**	**Peptide**
B.1.617.3	1064	HVTYVPAQK	1059	GVVFLHVTYVPAQKK
	1065	VTYVPAQKK	1060	VVFLHVTYVPAQKKN
	1066	TYVPAQKKNF	1061	VFLHVTYVPAQKKNF
	1064	HVTYVPAQKK	1070	AQKKNFTTAPAICHD
	1070	AQKKNFTTA	1071	QKKNFTTAPAICHDG
			1072	KKNFTTAPAICHDGK
			1062	FLHVTYVPAQKKNFT
			1063	LHVTYVPAQKKNFTT
			1069	PAQKKNFTTAPAICH
	Site	**Peptide**	**Site**	**Peptide**
Mink-related	685	VASQSVIAY	689	SVIAYTMSLGAENSV
	446	NYNYLFRLF	690	VIAYTMSLGAENSVA
	689	SVIAYTMSL	448	NYLFRLFRKSNLKPF
	684	SVASQSVIAY	449	YLFRLFRKSNLKPFE
	446	NYNYLFRLFR	450	LFRLFRKSNLKPFER
	447	YNYLFRLFR	685	VASQSVIAYTMSLGA

	MHC class I (percentile rank < 0.5)		MHC class II (percentile rank < 2.0)	
	Site	Peptide	Site	Peptide
Mink-related	451	FRLFRKSNLK	686	ASQSVIAYTMSLGAE
	687	SQSVIAYTM	687	SQSVIAYTMSLGAEN
	448	NYLFRLFRK	688	QSVIAYTMSLGAENS
	451	FRLFRKSNL	442	KVGGNYNYLFRLFRK
	1219	IAGLIAIVI	443	VGGNYNYLFRLFRKS
	1223	IAIVIVTIM	444	GGNYNYLFRLFRKSN
	681	RARSVASQSV	445	GNYNYLFRLFRKSNL
	442	KVGGNYNYLF	446	NYNYLFRLFRKSNLK
	445	GNYNYLFRLF	447	YNYLFRLFRKSNLKP

With respect to MHC class II epitopes derived from the spike protein, I found similar numbers of high-affinity peptides across the variants (percentile rank < 2.0 as in [199]), ranging from 510-564 epitopes, which corresponded to 325-348 unique peptides. As with the MHC-I epitope analysis, the variants showed a high degree of predicted epitope conservation in MHC-II epitopes compared to the Wuhan reference spike, ranging from 87% conservation for Beta, to 99% conservation for D614G (**Figure 4.5A).** This corresponded to 241 unique conserved peptides, regardless of HLA restriction. When filtered to show only epitopes presented on at least three HLA alleles, 34 unique conserved peptides were found which collectively spanned both the S1 and S2 subunits of spike (**Figure 4.5B**, **Table 4.4**). In general, I found variants to have slightly greater epitope conservation for MHC-I epitopes compared to MHC-II, except for the Gamma variant **(Figure 4.5A).** Notably, the Beta variant had seven percent less conservation in predicted high-affinity peptides for MHC-II compared to MHC-I.

I identified several high-affinity MHC-II epitopes that were specific to each variant (**Figure 4.5C, Table 4.5**). The Alpha variant had 20 unique peptides, corresponding to 32 epitopes; Beta had 15, corresponding to 25 epitopes; Kappa had 15, corresponding to 27 epitopes; and Gamma had 31, which lead to 45 unique epitopes (**Figure 4.5C**). The other variants had fewer lineage-

specific peptides, with Delta, B.1.617.3, Epsilon, and mink-related spike having 8, 9, 10, and 15 peptides respectively (**Table 4.5**). As noted in my MHC-I analysis, most variant-specific MHC-II epitopes occurred in the S1 unit of spike. However, in contrast to my MHC-I analysis, none of the peptides occurred in the RBD apart from the mink-related variant. Notably, the mink-related spike had 9/15 lineage-specific peptides occurring in the RBD (**Table 4.5**).

Conservation of the top spike peptides across human HLA alleles among the reference and variant SARS-CoV-2 viruses suggests that cellular adaptive immune responses may still be effective despite the mutations. However, my analysis also indicated that lineage-defining mutations span sequences likely to be presented on common human HLA alleles, possibly leading to some variant-specific CD8+ and CD4+ T cell immunity following exposure.

4.8 Summary of Chapter 4

Variants of SARS-CoV-2 have several amino acid substitutions along the spike protein that may impact effectiveness of vaccines currently in circulation. In this study, I enlisted my previously developed bioinformatic pipeline to screen antigenic changes across SARS-CoV-2 spike proteins that may impact human B cell immune responses, glycosylation patterns, and T cell immunity. Overall, I found broad conservation of high-affinity T cell epitopes and minimal B cell changes in antigenicity, but it is important to consider how these small-scale antigenic changes in viral proteins can still lead to evasion of the host immune system. In particular, the Alpha, Beta, Epsilon, and B.1.617 sub-lineages showed lowered antigenic scores for epitopes present in the reference spike, as well as novel antigenic sites in the N-terminal and receptor-binding domains, which coincides with existing antibody neutralization data. I did not note differences in predicted B cell antigenicity for the Gamma variant but detected additional glycosylation sites in the S1 subunit that may explain reduced antibody titers. Moreover, the

Gamma variant had the greatest number of lineage-specific MHC-I and MHC-II epitopes in my study, which may contribute to a unique antigenic signature despite broad T cell epitope conservation. Finally, I noted that lineage-specific MHC-I and MHC-II peptides are heavily skewed toward the RBD in the mink-related variant, which may be indicative of virus adaptation to mink and possible escape from pre-existing immunity. These observed changes in the variants investigated may affect immunodominance, leading to immune skewing during reinfection or the generation of immunodominant B cell clonotypes, which may be unfavorable in for vaccine-induced protection. My established pipeline can be used for the analysis of future SARS-CoV-2 variants.

CHAPTER 5 DISCUSSION

5.1 Host-pathogen interactions are a relevant public health problem

The evolutionary battle between virus and host can have devastating consequences. Viruses have the capacity to gain mutations that allow evasion from previously acquired adaptive immunity, which is reflected in the need to reformulate influenza virus vaccinations annually, and potentially in the decreased neutralizing antibody titers found in some SARS-CoV-2 variants. Improving our understanding of the many factors driving viral mutation, as well as enhancing strategies to predict the specific impact of these mutations, is imperative to reducing the harmful impact of viral evolution. In this thesis, I investigated the impact of specific host immunity on influenza viral mutation across both the external and internal influenza virus proteins. Using the sequence and antigenic analysis pipeline generated in my study of influenza viral mutations, I applied this strategy to predict the antigenicity of SARS-CoV-2 spike proteins across circulating viral variants.

Despite the development of the first influenza virus vaccine many decades ago, influenza viruses still cause significant morbidity and mortality in the human population. Antigenic drift remains the greatest challenge facing influenza virus vaccine development, and although it is generally known that viral mutation is driven by several factors, the exact mechanisms surrounding viral evolution remain largely unexplored in the context of host immunity driving viral mutation selection. My thesis provides one of the first comprehensive studies of specific host infection and vaccination history as a driver of influenza viral mutation. Using mice either infected with a historical H1N1 virus, vaccinated against 2009 pandemic H1N1, or a combination of both, I showed that differential stimulation of the immune system during contemporary pandemic H1N1 infection leads to positive selection of specific mutations across influenza viral HA, NA, NP, and PB1. Specifically, I show that immune responses that favor unrestricted viral replication combined

with strain-specific humoral responses facilitates the highest rate of mutation across influenza virus proteins. My study informs an important aspect of host-pathogen interaction that is not considered in traditional vaccine design.

My developed analysis pipeline was subsequently applied to screen antigenic changes in the SARS-CoV-2 spike protein between the original Wuhan viral sequence and several identified variants of concern or interest. I found that the nine SARS-CoV-2 variants in my investigation had minor, yet notable, changes in predicted antigenicity from the amino acid sequence changes. Notable changes were identified in the B cell epitopes of the Alpha (B.1.1.7), Beta (B.1.351), Epsilon (B.1.429), and B.1.617 sub-lineages, and in predicted glycosylation sites of the Gamma (P.1) variant. These changes may account for the decrease in neutralizing ability of convalescent and vaccine serum against new variants, as many of the B cell antigenic changes were identified in the receptor-binding domain. Additionally, I determined that the majority of predicted high-affinity MHC class I and II peptides were conserved across the variants. These data are important for assessing the performance of current COVID-19 vaccines against changes in the SARS-CoV-2 virus as they emerge, considering new vaccine formulations, and developing next-generation vaccine platforms.

5.2 Influenza virus mutations are driven by specific host immunity

Influenza virus is a single-stranded RNA virus, which mutates more rapidly than double-stranded DNA viruses [255]. Genetic variants are first generated by errors in replication, but then propagated through various mechanisms such as natural selection or random genetic drift [255]. Selection of virus mutations is described as positive when the rate of non-synonymous mutation surpasses that of synonymous mutation [256]. Here I investigated non-synonymous mutations on influenza virus HA, NA, NP, and PB1 at homosubtypic heterologous H1N1 challenge across

infected and vaccinated mice. Some of these viral mutations have been previously cited as mutations associated with adaptation to the mouse model and were present in mice of all immune backgrounds. However, I present several mutations specific to infection and vaccination histories that are not accounted for by animal model adaptation. Given that influenza virus antigenic drift is the result of interaction between viral proteins and the host immune system [257], I propose that the specific immune memory, or lack thereof, elicited at viral challenge contributes to specific patterns in viral mutations.

Along with characterization of the influenza viral SNPs and antigenicity profiles, I analyzed the host transcriptome by RNA sequencing to understand the corresponding host immune response, with the aim of characterizing immune response mechanisms that promote or restrict viral mutation. This immune characterization was essential to confirming the preimmune-vaccinated mouse model along with establishing immune response profiles for each immune background. Before in-depth comparisons of each immune background, I confirmed that our mouse model of primary and sequential infection with influenza viruses share several transcriptional landmarks with previously published models. At primary infection with Mex/09, our naïve-mock vaccinated animals showed upregulated ISGs including Cxcl10, Oas1, Irf1, Rsad2, and proinflammatory cytokines IL-6, IL-8, and IFN-g, which was shown previously in ferrets infected with Mex/09 and signifies activation of antiviral responses [258]. At heterologous re-infection, our preimmune animals showed upregulation of IgM and Cd8a. This suggests early induction of B and T cell-mediated immune mechanisms, which was shown at homologous re-infection in ferrets and indicates recall of previously acquired immune memory [258].

In the next sections, I will further describe the transcriptome response of each group and discuss how differential activation of innate and adaptive immune response pathways was associated with patterns in influenza virus mutations.

5.2.1 Previous vaccination drives mutation of immunodominant influenza virus proteins HA and NA through strain-specific adaptive immunity

Influenza viruses have both external and internal proteins that differentially stimulate or suppress the host immune response (reviewed here: [259]). The viral proteins HA and NA are prevalent on the virion surface and are major influenza virus antigens that undergo frequent mutational changes [260]. The immunodominant HA protein is the primary target for host neutralizing antibodies generated at vaccination, and because of this host-pathogen interaction, HA is subject to repeated immune pressures that lead to positive selection of mutations over time [44]. These mutations are frequently present at well-defined antigenic sites across HA [44,261,262]. Several NA mutations, including G41R, V241I, and N369K have been previously reported in circulating human seasonal H1N1 viruses [263]. Escape mutations have also been well-characterized in NA in response to antiviral drugs, in particular the H274Y mutation [264,265]. Given their immunodominant nature and propensity to mutate, I hypothesized that HA and NA would show the greatest degree of viral mutation, particularly in vaccinated animals who had strain-specific adaptive immunity.

The impact of previous vaccination can be best understood by comparing the transcriptional responses of naïve-vaccinated and naïve-mock vaccinated animals. Typically, influenza subunit vaccinations imbalance of more humoral activation and antibody elicitation with little specific cytotoxic T cell immunity gained after. In my immune analysis of the various groups, at three days post-H1N1 challenge, naïve-vaccinated mice had more robust activation of antiviral immune mechanisms compared to naïve mice. This may be explained by innate immune training, which describes reprogramming of innate immune cells that prime them for subsequent antigen encounters [39]. Innate immune training has been described in mouse models and can lead to enhanced activation of non-specific T lymphocytes and IFN-gamma production [39], which was enriched in naïve-vaccinated mice. Along with T cell cytotoxicity, naïve-vaccinated animals also

showed activation of humoral immunity at challenge. This adaptive immune activation from vaccination significantly decreased viral load in the lungs compared to mock-vaccinated animals. However, it seems that strain-specific adaptive immunity, in conjunction with only moderate restriction of viral replication, was associated with mutation of viral HA and NA.

On HA, naïve-vaccinated animals had the highest degree of viral mutation overall as well as in several groups where group-specific mutations were acquired. These were most concentrated in the RBD of HA and were W150, L151P, and S183P, of which only S183P has been cited previously as a mouse-adaptive mutation [266,267]. In contrast, the RBD mutation P182Q was shared among preimmune-mock vaccinated animals only, indicating a mutational pattern distinct from naïve-vaccinated animals. Non-RBD mutations may also be of importance, as influenza virus cross-protection through antibodies against conserved regions of HA have been reported previously [268–271]. Outside of the RBD, I found two mutations specific to naïve-vaccinated animals (K43N and Y351), and only one in preimmune-mock vaccinated animals (G350V).

Given that the most RBD-specific mutations were from naïve-vaccinated animals, it seems that the presence of strain-specific humoral immune memory only may contribute to mutation of HA. These mice were vaccinated with the 2018-2019 QIV, which contains concentrated HA protein that is similar in sequence to the challenge strain Mex/09. In a study of H3N2 mutations after seasonal vaccination in humans, variants of HA were selected most often in vaccinated individuals [129], supporting my hypothesis that highly specific antibodies can drive mutation. Another study found that treatment with a monoclonal antibody against H1N1pdm09 HA drives mutation of HA in mice [207]. In contrast, naïve-mock vaccinated mice with no previous antibodies against Mex/09 showed the most mutations outside of the HA RBD, which suggests that unrestricted viral replication is a contributing factor to viral mutation but not in the cellular receptor binding site.

The importance of strain-specific adaptive immunity is evident when comparing HA mutations from naïve-vaccinated mice to those found in preimmune-mock vaccinated mice. Preimmune-mock vaccinated mice, which have antibodies specific to the FM/47 seasonal H1N1 virus as well as activation of specific T cell immunity which might be more cross-reactive to various viral proteins, showed fewer RBD-specific mutations than naïve-vaccinated animals. It is important to consider that historical FM/47 and Mex/09 virus strains are not antigenically identical, where the two viruses only share 78% amino acid sequence similarity on HA. Drawing from classical immunology, it is likely that the antibodies elicited at three days post-challenge are those that cross-react with Mex/09, or in other words those that arise from B cell clonotypes with similar antigenic specificity that may or may not have now undergone affinity maturation in a germinal center reaction [272]. In this scenario, the FM/47 preimmune mice are limited to clonotypes that can cross-react with the Mex/09 RBD, resulting in a smaller repertoire compared to the BCR repertoires elicited in animals vaccinated against the Mex/09 RBD. A narrower repertoire of antibodies capable of binding the Mex/09 RBD may lead to fewer mutations on the Mex/09 RBD due to decreased immune pressure on the RBD.

Compared to viral HA, NA showed less mutation overall at Mex/09 challenge. First, I found a C292S mutation on NA to appear in all immune backgrounds, except for preimmune-vaccinated mice. This mutation is in the head domain of NA, which contains the active sites for NA enzymatic activity. I also found that naïve-vaccinated animals had the most mutations in the NA head compared to the other groups. Although QIV vaccine efficacy is based on HA content, NA can still be present, although its presence is not regulated [52,273,274]. Therefore, it is possible that antibodies generated against the NA head in naïve-vaccinated animals to then be recalled at challenge, leading to viral selection. However, because HA and NA are both necessary for viral infection, it is possible that NA mutations after vaccination exist as a compensatory mechanism for targeting of HA by neutralizing antibodies. The opposite scenario has been shown in the

literature, where mutations on HA were associated with the use of NA inhibitors [275]. Interestingly, in my work mutations in the NA stalk were detected in unvaccinated animals only (E47G and Q51E in naïve-mock vaccinated; I38T in preimmune-mock vaccinated). For preimmune-mock vaccinated animals, it could be the elicitation of cross-reactive immunity between NA of FM/47 and Mex/09 that select different mutations in the NA stalk, analogous to production of HA stalk-directed antibodies at heterologous influenza virus infection [218,276].

5.2.2 Previous infection against divergent H1N1 strain led to a high number of mutations but also restricted mutation of conserved influenza virus proteins NP and PB1

The influenza viral nucleocapsid protein is abundant inside the virion and is essential for vRNA transcription [5]. As NP is relatively conserved and has previously been shown to induce strongly cross-reactive CD4+ and CD8+ T cells [103–105], I hypothesized that heterologous H1N1 challenge would elicit NP-specific immune memory that could lead to immune selection of mutations. Non-synonymous NP mutations have been previously reported in circulating influenza viruses, including V217I in H1N1pdm09 during the 2011-2012 season [263], and mutations at several residue positions across both H1N1 and H3N2 lineages including 33, 100, 305, and 357 [277]. For example, the V100I mutation was shown to increase in frequency in H1N1pdm09 from ~10% to over 80% in just seven months, indicating positive selection [277]. In my analysis, I noted the mutations D101G and T130K to be present in most animals at three days post-H1N1 challenge, regardless of immune background. Like HA, NP can experience mutations when adapting to the mouse model; D101G represents a known adaptation of pandemic H1N1 in mice [267], while T130K has not been characterized previously. In my analysis, the greatest degree of immune background-specific mutation of NP occurred in naïve-vaccinated animals. Notably, the R31G and R441 persisted in naïve-vaccinated mice only, while only K91R was detected in naïve-

mock vaccinated mice. Although I did not find these mutations previously reported, this provides evidence for immune selection of mutations across internal, conserved proteins, and not solely the external proteins cited in antigenic drift. Moreover, NP may mutate as a compensatory mechanism to neutralization of HA in vaccinated animals.

Mutations across pandemic H1N1 PB1 have also been previously reported. Starting from the prototype isolate A/California/04/2009, the PB1 mutations T20I, I397M, and I435V appeared worldwide during the 2011-2012 season [263]. Of the four immune backgrounds in our study, I found only naïve-vaccinated mice to have shared non-synonymous mutations across PB1. I noted the L95P mutation in the cRNA promoter binding site of PB1 [278], as well as D617A, G622R, and L624F in the vRNA promoter binding site [279], and S712Y at a core interaction site with PB2 [280]. I did not find these mutations on PB1 to be previously reported in the literature, and pandemic H1N1 PB1 generally does not show adaptation in the mouse model [267]. These findings suggests that prior vaccination, but not previous infection, somehow influences mutations on PB1. It is unclear why this happens but given that mutations were found in cRNA and vRNA promoter binding sites, they may provide a fitness advantage when HA is targeted after vaccination.

Previous infection with FM/47 seems to limit mutation of NP and PB1. At Mex/09 challenge, preimmune-mock vaccinated animals had early activation of antiviral defense mechanisms, followed by strong upregulation of genes associated with specific T cell activation and B cell proliferation, which are thought to be protective. Having experienced intranasal infection with FM/47 previously, these animals likely have adaptive immune cells in the respiratory tract, such as resident memory T cells in the lungs, which are shown to be activated at influenza virus infection [105,281]. Additionally, robust T cell activation and proliferation gene signatures in preimmune mice are likely directed toward conserved influenza virus proteins that are targets of cross-reactive T cell responses, such as NP as previously shown [104,282]. Clearance by T cells

seems to result in lower variation to be selected by host immunity. Mutation may also be limited by the absence of strain-specific antibodies toward these proteins.

5.2.3 Limited viral replication and combined humoral and T cell-mediated immunity restricts viral mutation

Despite similar viral titer, preimmune-mock vaccinated and preimmune-vaccinated mice showed differences in immune response activation at lethal H1N1 challenge. Of all groups, preimmune-vaccinated mice had the greatest activation of B cell-mediated immune response pathways including B cell receptor signaling and B cell proliferation, as well as activated T cell proliferation. Given previous findings on immune imprinting, we can speculate that antibodies toward the FM/47 HA are still elicited despite the subsequent vaccination toward Mex/09 [283]. Like preimmune-mock vaccinated mice, T cell responses are likely directed toward internal proteins such as NP that are targets of cross-reactivity. It seems that this broader antibody repertoire, consisting of antibodies toward FM/47 as well as virus like Mex/09, combined with activation of cross-reactive T cells toward conserved proteins first encountered during FM/47 infection boosted by secondary infection, was effective in preventing escape mutations in HA, NA, NP, and PB1. This finding highlights the importance of engaging T cell immunity as a means of removing viral variants, in addition to generating broadly cross-reactive antibodies rather than a narrow repertoire of strain-specific antibodies that can apply greater selection pressure.

5.2.4 Lack of immune memory limits viral mutation in naïve animals

Finally, my analysis of viral mutations occurring in naïve-mock vaccinated animals demonstrates the contribution of adaptive immune mechanisms to influenza viral evolution. At three days post-pandemic H1N1 challenge, naïve-mock vaccinated mice had a highly

proinflammatory immune response driven by macrophages and cytotoxic T lymphocytes, and high viral load in the lungs. Adaptive immune mechanisms, such as the elicitation of antigen-specific B cells, are yet not activated in these animals. In parallel, the number of group-specific SNPs detected on viral HA, NA, and NP are moderate, with no mutations persisting on PB1. Group-specific mutations on HA in naïve-mock vaccinated animals most often occurred outside of the RBD. Therefore, it seems that unrestricted viral replication certainly leads to the emergence of several viral protein variants that may have occurred randomly and can persist for various reasons [255], but the lack of humoral or cellular immune memory in these animals serves to limit selection of group-specific mutations during the study timeframe. It is important to note that pre-existing adaptive immunity from previous infection and vaccination are absolutely essential to preventing severe disease in humans. Nonetheless, there is perhaps a tradeoff between highly specific immune responses and induction of more viral variants.

5.3 Antigenic analysis of the SARS-CoV-2 spike protein reflects existing vaccination data against the variants

Viral mutations do not always translate to antigenic differences [44,255]. When formulating vaccines to emerging viruses, or when considering vaccine effectiveness toward viruses that continue to evolve, it is essential to translate amino acid changes to possible changes in antigenicity. Therefore, to complement my study of host immune pressures driving antigenic drift, I developed an antigenic analysis pipeline to determine whether specific amino acid substitutions change the predicted antigenicity of a protein with respect to B cell epitopes, T cell epitopes, and glycosylation sites. I then applied this pipeline to SARS-CoV-2 variant spike proteins.

When examining each amino acid substitution across the nine SARS-CoV-2 variants for predicted structural change on the spike protein using Missense 3D, six substitutions had potential damage: these were L18F and D138Y in the N-terminal domain of Gamma variant spike; D80A

in the N-terminal domain of Beta; G142D in the N-terminal domain of Kappa and Delta; E484Q in the receptor-binding domain of Kappa and B.1.617.3; and P681H/R in Alpha and B.1.617 sub-lineages. Several amino acid substitutions across the variants have already been functionally investigated and found to affect the transmissibility and antigenicity of the virus. The well-studied mutation D614G, which quickly became the dominant form of the virus early in the pandemic, was shown to enhance not only the transmissibility of SARS-CoV-2, but also receptor-binding domain-specific antibody susceptibility and trimer stability [132–134,284], although we did not predict major structural impacts from D614G in our Missense3D analysis. The N501Y substitution found in the receptor-binding domain of Alpha, Beta, and Gamma spike has an increased ACE2 receptor-binding affinity with or without the help of the deletion at 69-70 [43]. Additionally, SARS-CoV-2 containing the E484K substitution (present in Beta and Gamma variants) has reduced antibody neutralization ability [150]. The mutation K417T present in the RBD, along with E484K and N501Y, is described as part of a trio of mutations that likely increase transmissibility of the Gamma variant [153]. The L452R mutation found in Epsilon and B.1.617 sub-lineages is said to confer resistance to binding by certain monoclonal antibodies [160,285] along with possibly increasing infectivity [286]. Other spike mutations are shown to impact viral fitness directly. The L18F mutation, for which we identified structural damage in our analysis of the Gamma variant, is said to confer a replicative advantage [287].

Once the variants were identified, essential studies were initiated to determine if antibodies elicited by vaccination or by previous infection with the original SARS-CoV-2 virus could neutralize the new variants. In general, vaccine-induced immunity provides more robust heterotypic immunity than natural infection to emerging SARS-CoV-2 variants of concern [288]. Neutralization assays performed across some variants with vaccine serum have indicated modest decreases in titers against the Alpha variant, with more pronounced losses in titer for variants which contain the E484K substitution (Beta and Gamma) [147]. In general, Alpha is shown to be

susceptible to neutralizing antibodies elicited by ancestral spike vaccines [144–147]. Vaccine clinical trials held in regions with high dominance of a particular variant have demonstrated similar findings, leading to the conclusion that vaccines have greater cross-protective immunity against Alpha compared to other variants, although further analysis is still needed due to small numbers in study participants [145,289]. My study points to additional B cell epitope changes that may account for these results. In my analysis for the Alpha variant, I noted a moderate decrease in the B cell epitope prediction score for the spike sequence spanning residues 180-185 in the N-terminal domain, as well as an additional antigenic site spanning residues 565-568 in the RBD, which is proximal to the A470D mutation. This small antigenic difference in the Alpha RBD compared to the ancestral SARS-CoV-2 spike sequence may give insight into the finding that some RBD-specific antibody titers are decreased in Alpha [144,145]. However, our B cell epitope sequence predictions were mostly conserved between Alpha and D614G spike, suggesting that most antibodies elicited against the vaccine strain will likely cross-react with Alpha.

The Beta variant was the first variant to have identified acquisition of the E484K mutation in the spike RBD. The E484K mutation of Beta is thought to contribute to evasion of host neutralizing antibodies [150] as studies have confirmed reduced susceptibility of Beta against antibodies elicited after natural infection [151], vaccination [147], and treatment with monoclonal antibodies bamlanivimab, estesevimab, and imdevimab [152]. From these data, I hypothesized that the E484K mutation would lead to detectable changes in predicted B cell epitopes for the Beta variant. However, my analysis of B cell epitopes showed no difference in the Beta RBD compared to our reference spike sequence proximal to the E484K substitution, although this analysis only indicates the presence of an epitope and not if the epitopes are the same. I noted antigenic changes in the N-terminal domain spanning residues 248-254 (most proximal to D215G), but it is unclear which mutations contribute to this, or whether it contributes to differential

recognition by host humoral responses. It was not revealed in my analysis how E484K may contribute to antibody evasion.

I also noted several less prominent predicted B cell epitope changes for the Epsilon variant that may contribute to differences in antigenicity. Two antigenic regions in the N-terminal domain, specifically at residues 72 and 146, both proximal to known spike glycosylation sites, showed lower B cell epitope prediction scores in the Epsilon spike compared to the reference spike. Also, an additional antigenic region in the RBD at residues 455-480 was detected in Epsilon and B.1.617 lineages, but not in the other variants under study. Current evidence suggests moderate decreases in antibody titer against Epsilon in those vaccinated with a Wuhan-1 isolate–based messenger RNA vaccine, or in serum from convalescent individuals [290].

Currently, the B.1.617 lineage that emerged in India in 2020 is of particular concern. Of the three sub-lineages described, the Delta variant has accounted for the greatest number of reported cases worldwide and is suggested to have an increased secondary attack rate [291]. Reduction in neutralization by antibodies from vaccinated and previously infected individuals has been identified for the Delta variant, however this reduction is not as large as what was observed for the Beta variant [152,292]. In my analysis of predicted B cell epitopes, I observed minimal differences in antigenicity between the Delta variant and the other B.1.617 sub-lineages. Although the Delta variant contains the T478K substitution that is not present in the other sub-lineages, in addition to lacking the E484Q mutation, these changes did not lead to differences in predicted B cell epitopes or glycosylation sites. However, I noted that the B.1.617 lineage contained B cell epitope sequences not predicted in the original Wuhan lineage or the other variants of concern, in particular the region spanning residues 455-480 in the spike RBD. These unique B cell epitope sequences may account for the moderate resistance to monoclonal antibodies directed against spike RBD that has been reported for the Delta variant [292], but it is currently unclear whether this reduction is shared amongst the other B.1.617 sub-lineages. Additionally, I noted that the

predicted glycosylation site at position 17, which was found in all other variants including the Kappa variant, was not found in the Delta or B.1.617.3. Although unclear how, the lack of a glycosylation site at this position may impact the specific antigenicity of the Delta variant compared to other SARS-CoV-2 spike proteins.

These observed changes in the variants may affect immunodominance, leading to immune skewing during sequential infection-reinfections or the generation of immunodominant B cell clonotypes, which may be unfavorable for vaccine-induced protection.

5.4 Prediction of SARS-CoV-2 spike glycosylation sites may explain differences in vaccine effectiveness against the Gamma variant

Under the hypothesis that spike protein glycosylation may impact protein folding, stability, and evasion of host immune responses, I aimed to compare N-glycosylation sites along the nine SARS-CoV-2 variants in my study. Here I found 22 predicted glycosylation sites that were conserved across the variants and original SARS-CoV-2 viruses. These conserved sites have been confirmed biologically [293]. Specifically, Watanabe and colleagues found the same 22 glycans that were predicted in my analysis, however, with my predictions I identified three additional glycosylation sites that were unique to the Gamma variant at sites 20, 188, and 657. Given the consistency between my computational site predictions and existing biological characterization, support exists that the novel potential sites identified are actual glycosylation patterns on the Gamma variant spike. Further analysis is required to validate the N-glycosylated sites found exclusively in Gamma variant at sequons N20 and N188, since those sites have not yet been characterized.

Increases in protein glycosylation of a virus after introduction to humans from an animal reservoir has been previously observed for other viruses. For example, influenza viruses that initially spill over into humans from animals have a relatively low number of glycosylated residues

within the hemagglutinin (HA) glycoprotein [294,295]. Interestingly, as the virus becomes more humanized and less like the virus as it existed in its original animal reservoir host, the virus gains mutations that facilitate glycosylation of amino acids [294,295]. The specific glycans might influence both innate and adaptive immune responses by affecting the spike protein binding through pattern recognition receptors, like collectins and other lectins, and via changing the accessible HLA antigens, thereby providing a fitness advantage for the virus. The glycans can also shield certain epitopes from antibody neutralization [293], as observed in other viruses including SARS-CoV-1 spike, HIV-1 envelope, influenza HA, and LASV glycoprotein [195,293]. I identified additional glycosylation sites in the Gamma variant only, suggesting that gaining glycosylation modification may not be a general strategy that is advantageous for the SARS-CoV-2 virus as it adapts to its human host. Although I did not identify a generalized increase in glycosylation in the SARS-CoV-2 spike proteins of variant viruses, it will be essential to continue to monitor the SARS-CoV-2 spike protein for increases in glycosylation sites as it circulates in humans as the viruses still might utilize this strategy for host evasion. Additionally, changes in B cell epitopes distal to glycosylation sites may also have synergistic effects on antigenicity [296] which should be further explored. Identifying conserved B cell epitopes that may be impeded by novel glycosylation sites is essential to optimizing epitopes for vaccine formulations.

5.5 MHC class I and II binding data suggests broad conservation of T cell epitopes across SARS-CoV-2 variants

My analysis of T cell epitopes suggests that the major CD8+ and CD4+ T cell epitopes were conserved across SARS-CoV-2 spike variants. This data supports a similar analysis by Tarke and colleagues, which suggested a negligible impact of SARS-CoV-2 variants with respect to CD8+ T cell epitopes [297]. Additionally, their computational analysis was supported by experimental evidence illustrating similar CD4+ and CD8+ T cell simulation by SARS-CoV-2

variant peptide pools. In my study, I noted that high-affinity peptides that form the antigens for CD4+ and CD8+ T cell receptors are highly conserved, with MHC-I-binding peptides 93-99% conserved, and MHC-II peptides 87-99% conserved. The conserved peptides spanned the S1 and S2 subunits of spike in relatively equal proportions. However, lineage-specific peptides were detected in each variant which may contribute to unique antigenic signatures and differentially stimulate host T cells. In particular, the Gamma variant had several lineage-specific epitopes that had the capacity to strongly bind several common MHC-I and MHC-II alleles in humans. Differential recognition by host T cells at infection, or altered immunodominance of specific peptides, may contribute to ineffective immune responses in patients previously infected with SARS-CoV-2 [298]. Immunodominance will be an important next step in the analysis of the adaptive immune response to SARS-CoV-2 variants.

One published study showed that the SARS-CoV-2 spike protein is poorly immunogenic in respect to CD8+ T cells compared to other coronaviruses, since its reactivity was less than that of other spike proteins from the previous coronaviruses SARS-CoV-1 and MERS-CoV. Despite the interpretation of the authors, my work, together with the evidence that the current vaccine platforms can elicit strong T cell responses [35,288,297], suggests the importance of developing vaccines that are able to elicit T cell responses as a countermeasure for RNA viruses with a propensity to mutate. Additionally, a set of T cell epitopes that are highly conserved among human and animal coronaviruses may represent targets for the development of next-generation vaccines that are cross-protective against divergent coronaviruses [299]. My analysis supports the S2 subunit of as a source of cross-reactive T cell epitopes that can be used to target SARS-CoV-2 variants. In my analysis, the majority of variant-specific MHC-I and MHC-II-binding peptides we detected were derived from the S1 subunit, with greater conservation of S2 sequences. A study by Stoddard and colleagues demonstrated the importance of the S2 subunit as a cross-reactive target, where convalescent serum from SARS-CoV-2 patients had the greatest cross-reactivity

with endemic human coronaviruses when stimulated with peptides derived from the S2 subunit [300]. Moreover, S2-derived epitopes may have the greatest resilience against coronavirus spillover from animal reservoirs. My analysis found that the mink-related variant had the greatest proportion of lineage-specific MHC-I peptides occurring in the RBD compared to the other variants. Unexpectedly, the mink-related spike was the only variant in my investigation to have lineage-specific peptides in the RBD that were predicted to bind MHC-II alleles, with 9/15 lineage-specific peptides spanning the RBD. Although dissemination of the mink-related variant in humans has been limited, these data suggest that variability in the S1 subunit may impact the potential of future spillover into humans, either positively or negatively. Nonetheless, more work is needed to isolate the T cell response against SARS-CoV-2 to determine if T cells alone can provide cross-protection against SARS-CoV-2 variant infection or severe COVID-19 disease.

5.6 Toward the future of vaccine design and pandemic preparedness

The knowledge that the capacity for viral evolution is associated with the hosts' vaccination and infection history can be applied to the development and refinement of future vaccines and vaccination regimes. First, my results suggest that naïve populations, or those unvaccinated or uninfected previously, experience the least restricted viral replication at challenge and contribute the highest frequency of variants. Although naïve hosts do not yet have specific adaptive immune mechanisms that can apply selective pressure, higher viral titer can allow novel mutations to appear that may later be selected when transferred to a vaccinated individual, for example. This scenario underlines the importance of early vaccination in the human population to limit viral replication on a population level and therefore limit the emergence of novel variants as well as protecting from severe disease. Furthermore, evidence supporting unrestricted viral replication as a driver of emerging viral variants in humans has been shown during the SARS-CoV-2 pandemic.

Several case studies have reported increased variability in SARS-CoV-2 gene sequences in immunosuppressed COVID-19 patients during chronic infection and unrestricted viral replication [301]. These novel variants which emerged from the immunosuppressed often had reduced sensitivity to neutralizing antibodies, showing that high viral replication coupled with prolonged viral shedding can be a source of viruses with antigenic differences.

Second, several of the mutations in my influenza virus study appeared in the stalk domain of HA, which is generally conserved and is a primary target for universal influenza vaccination (reviewed here [302]). My results suggest that sequential infection with heterologous influenza viruses can select mutations across the stalk region, rendering it another target on HA that can influence viral escape from host immunity. This may not be solved by simply formulating seasonal influenza virus vaccines that are well-matched to circulating strains. Given the original antigenic sin hypothesis, which contends that memory of the first influenza virus infection forms the baseline of future immune responses against influenza viruses, contemporary infections will continue to stimulate immune memory from early influenza virus exposures, which may be divergent from the imprinting strain [303], thus stimulating antibodies against the stalk and possibly selecting mutations. As suggested by my analysis of SARS-CoV-2 spike variant epitopes, it seems that the key to conserving immunity toward influenza viruses is in stimulating cross-reactive T cell responses as opposed to strain-specific antibodies.

Third, we should consider that the immune memory of the host may also play a role in coronavirus infections. Where human coronaviruses hCoV-NL63, -229E, -OC34 and -HKU1 circulate seasonally and cause milder respiratory tract infections, it is possible that adaptive immune responses generated toward these viruses have cross-reactivity with novel coronavirus strains, such as the pandemic coronavirus SARS-CoV-2. Studies have shown cross-reactivity between seasonal coronaviruses and SARS-CoV that may prime the immune response and lead to early detection of antibodies [304]. Current evidence already suggests some T cell cross-

reactivity between seasonal viruses and the more pathogenic SARS-CoV-2, including cross-reactivity toward spike proteins [107]. It is currently unclear what role, if any, preimmunity plays in coronavirus evolution; however, the importance of preimmunity as a driver of viral mutation may be relevant beyond influenza viruses.

Finally, computational analyses of viral proteins can be an exceptionally useful in formulating vaccines and in managing rapidly progressing pandemics such as SARS-CoV-2. Mutations do not always lead to changes in antigenicity [255]. It is therefore important to harness the predictive power that the decades of B and T cell receptor binding data have provided. Antigenic analysis strategies, such as the pipeline I developed in my study of influenza virus and SARS-CoV-2 variants, can allow us to predict the degree to which certain mutations will impact viral epitopes, as well as which arms of the immune response may be affected. Although computational strategies are not perfect and do not completely capture the dynamics of host-pathogen interactions, they give us valuable leverage against emerging viruses by facilitating a faster response during epidemics and pandemics. Moreover, as our knowledge of viral evolution advances, computational methods such as machine learning can be employed to predict mutational patterns based on previous data [305].

5.7 Future directions

The work conducted in this thesis constitutes the foundation for future studies involving viral evolution and prediction of viral antigenicity. There are several different avenues that can be explored as an immediate follow-up to my studies, with respect to both evolutionary selection pressures and the predictive power of antigenic analysis tools.

Having developed a preimmune-vaccinated mouse model, we can investigate other variations of the sequential infection and vaccination study that will allow us to better elucidate

the host factors contributing to mutation. First is how the degree of similarity between sequential exposure events impacts the immune response and viral mutations; for example, we might use an H1N1 virus as the preimmune strain, but an H3N2 virus as the challenge strain. My results also imply that unrestricted viral replication, together with highly specific immune responses, facilitates the greatest degree of viral mutation. These factors may be investigated separately, using host-to-host transmission as a bottleneck that would provide insight into mutation acquisition in a real-world setting. Moreover, since immune memory as well as immune capacity seem to impact viral mutation, the age of the host may play a role in antigenic drift. The cross-section between age-related immune capacity and accumulated immune memory toward viruses could be investigated by repeating this study with young and aged mouse models, or by transferring viruses from naïve mice to preimmune-vaccinated mice. This would allow us to model the relative contribution of infants and young children (naïve) compared to older individuals with accumulated immune histories (preimmune).

My transcriptome analysis of mouse immune responses at viral challenge showed strong activation of B cell-mediated immunity. However, the greatest gene enrichment in preimmune animals was of immune pathways involving activation of cytotoxic T cells. The C57Bl/6J mouse model used in this study generally shows balanced Th1/Th2 responses. To conduct a more detailed analysis of the humoral response at heterologous challenge, such as sequencing of the B cell receptor repertoire, it may be useful to employ a mouse model skewed toward Th2 responses, such as BALB/c mice.

For analysis of SARS-CoV-2 variants, future studies should apply the computational predictions to biological studies, with a specific focus on immunodominance. My analysis identified several conserved peptides that can cross-react with many MHC-I and MHC-II alleles. These peptides could be used to stimulate T cells from convalescent patient serum to evaluate reactivity toward conserved spike epitopes, possibly providing antigenic targets for vaccines

directed toward stimulating T cell immunity. Moreover, the peptide libraries specific to each variant can be used in cell-based assays, such as enzyme-linked immune absorbent spot (ELISpot) assays, to investigate cross-reactivity between SARS-CoV-2 vaccines and to emerging variants.

5.8 Conclusion

In conclusion, my thesis has highlighted the importance of the specific host immune background as a driver of influenza virus evolution. Strain-specific antibody responses elicited by vaccination, in conjunction with only moderate restriction of viral replication, were associated with mutation of both external (HA and NA) and internal (NP and PB1) influenza virus proteins. I propose that unrestricted replication and specific immune targeting are associated with increased viral mutations and changes in antigenicity, as seen in the experimental groups of my influenza virus study as well as in the unrestricted viral replication of SARS-CoV-2 as it spread through the human population. In contrast, broader antibody repertoires consisting of antibodies toward antigenically similar and divergent influenza viruses, combined with cross-reactive T cell responses, serve to limit viral mutation. I also showed that existing B cell and T cell epitope prediction software can be leveraged during the SARS-CoV-2 pandemic to evaluate the antigenicity of rapidly emerging spike variants. Taken together, the results of my experimental model and that from viral mutations in people will contribute to our overall understanding of viral evolution, which can be used to better predict viral mutational patterns and design more effective vaccines going forward.

REFERENCES

1. McGinnis J, Laplante J, Shudt M, George KS. Next generation sequencing for whole genome analysis and surveillance of influenza A viruses. J Clin Virol Off Publ Pan Am Soc Clin Virol. 2016;79:44–50.

2. Iuliano AD, Roguski KM, Chang HH, Muscatello DJ, Palekar R, Tempia S, et al. Estimates of global seasonal influenza-associated respiratory mortality: a modelling study. Lancet Lond Engl. 2018 31;391(10127):1285–300.

3. Neumann G, Noda T, Kawaoka Y. Emergence and pandemic potential of swine-origin H1N1 influenza virus. Nature. 2009 Jun;459(7249):931–9.

4. Kalil AC, Thomas PG. Influenza virus-related critical illness: pathophysiology and epidemiology. Crit Care. 2019 Dec;23(1):258.

5. Bouvier NM, Palese P. The biology of influenza viruses. Vaccine. 2008 Sep 12;26 Suppl 4:D49-53.

6. CDC. Types of Influenza Viruses [Internet]. Centers for Disease Control and Prevention. 2019 [cited 2021 Mar 1]. Available from: https://www.cdc.gov/flu/about/viruses/types.htm

7. Jang J, Bae S-E. Comparative Co-Evolution Analysis Between the HA and NA Genes of Influenza A Virus. Virol Res Treat [Internet]. 2018 Jul 19 [cited 2019 Nov 12];9. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6053862/

8. Kanegae Y, Sugita S, Endo A, Ishida M, Senya S, Osako K, et al. Evolutionary pattern of the hemagglutinin gene of influenza B viruses isolated in Japan: cocirculating lineages in the same epidemic season. J Virol. 1990 Jun 1;64(6):2860–5.

9. Dou D, Revol R, Östbye H, Wang H, Daniels R. Influenza A Virus Cell Entry, Replication, Virion Assembly and Movement. Front Immunol [Internet]. 2018 Jul 20 [cited 2019 Nov 12];9. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6062596/

10. Samji T. Influenza A: understanding the viral life cycle. Yale J Biol Med. 2009 Dec;82(4):153–9.

11. Noda T, Kawaoka Y. Structure of influenza virus ribonucleoprotein complexes and their packaging into virions. Rev Med Virol. 2010 Nov;20(6):380–91.

12. Grohskopf LA, Alyanak E, Broder KR, Walter EB, Fry AM, Jernigan DB. Prevention and Control of Seasonal Influenza with Vaccines: Recommendations of the Advisory Committee on Immunization Practices - United States, 2019-20 Influenza Season. MMWR Recomm Rep. 2019 Aug 23;68(3):1–21.

13. Del Giudice G, Rappuoli R. Inactivated and Adjuvanted Influenza Vaccines. In: Oldstone MBA, Compans RW, editors. Influenza Pathogenesis and Control - Volume II [Internet]. Cham: Springer International Publishing; 2015 [cited 2019 Nov 29]. p. 151–80. (Current

Topics in Microbiology and Immunology). Available from: https://doi.org/10.1007/82_2014_406

14. Rajão DS, Pérez DR. Universal Vaccines and Vaccine Platforms to Protect against Influenza Viruses in Humans and Agriculture. Front Microbiol [Internet]. 2018 [cited 2019 Nov 29];9. Available from: https://www.frontiersin.org/articles/10.3389/fmicb.2018.00123/full

15. Peck KM, Burch CL, Heise MT, Baric RS. Coronavirus Host Range Expansion and Middle East Respiratory Syndrome Coronavirus Emergence: Biochemical Mechanisms and Evolutionary Perspectives. Annu Rev Virol. 2015;2(1):95–117.

16. Edridge AWD, Kaczorowska J, Hoste ACR, Bakker M, Klein M, Loens K, et al. Seasonal coronavirus protective immunity is short-lasting. Nat Med. 2020 Nov;26(11):1691–3.

17. Schenten D, Bhattacharya D. Immunology of SARS-CoV-2 infections and vaccines. Adv Immunol. 2021;151:49–97.

18. Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. The Lancet. 2020 Feb 15;395(10223):497–506.

19. COVID-19 Map [Internet]. Johns Hopkins Coronavirus Resource Center. [cited 2021 Aug 5]. Available from: https://coronavirus.jhu.edu/map.html

20. Alanagreh L, Alzoughool F, Atoum M. The Human Coronavirus Disease COVID-19: Its Origin, Characteristics, and Insights into Potential Drugs and Its Mechanisms. Pathogens. 2020 May;9(5):331.

21. Chen Y, Liu Q, Guo D. Emerging coronaviruses: Genome structure, replication, and pathogenesis. J Med Virol. 2020;92(4):418–23.

22. Beniac DR, Andonov A, Grudeski E, Booth TF. Architecture of the SARS coronavirus prefusion spike. Nat Struct Mol Biol. 2006 Aug;13(8):751–2.

23. Delmas B, Laude H. Assembly of coronavirus spike protein into trimers and its role in epitope expression. J Virol. 1990 Nov 1;64(11):5367–75.

24. DeDiego ML, Álvarez E, Almazán F, Rejas MT, Lamirande E, Roberts A, et al. A Severe Acute Respiratory Syndrome Coronavirus That Lacks the E Gene Is Attenuated In Vitro and In Vivo. J Virol. 2007 Feb 15;81(4):1701–13.

25. Nieto-Torres JL, DeDiego ML, Verdiá-Báguena C, Jimenez-Guardeño JM, Regla-Nava JA, Fernandez-Delgado R, et al. Severe Acute Respiratory Syndrome Coronavirus Envelope Protein Ion Channel Activity Promotes Virus Fitness and Pathogenesis. PLOS Pathog. 2014 May 1;10(5):e1004077.

26. Nal B, Chan C, Kien F, Siu L, Tse J, Chu K, et al. Differential maturation and subcellular localization of severe acute respiratory syndrome coronavirus surface proteins S, M and E. J Gen Virol. 86(5):1423–34.

27. Neuman BW, Kiss G, Kunding AH, Bhella D, Baksh MF, Connelly S, et al. A structural analysis of M protein in coronavirus assembly and morphology. J Struct Biol. 2011 Apr 1;174(1):11–22.

28. Cui L, Wang H, Ji Y, Yang J, Xu S, Huang X, et al. The Nucleocapsid Protein of Coronaviruses Acts as a Viral Suppressor of RNA Silencing in Mammalian Cells. J Virol. 89(17):9029–43.

29. Walls AC, Park Y-J, Tortorici MA, Wall A, McGuire AT, Veesler D. Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. Cell. 2020 Apr 16;181(2):281-292.e6.

30. Tumban E. Lead SARS-CoV-2 Candidate Vaccines: Expectations from Phase III Trials and Recommendations Post-Vaccine Approval. Viruses [Internet]. 2020 Dec 31 [cited 2021 Mar 15];13(1). Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7824305/

31. Krammer F. SARS-CoV-2 vaccines in development. Nature. 2020 Oct;586(7830):516–27.

32. Dong Y, Dai T, Wei Y, Zhang L, Zheng M, Zhou F. A systematic review of SARS-CoV-2 vaccine candidates. Signal Transduct Target Ther. 2020 Oct 13;5(1):1–14.

33. Commissioner O of the. COVID-19 Vaccines. FDA [Internet]. 2021 Nov 3 [cited 2021 Mar 15]; Available from: https://www.fda.gov/emergency-preparedness-and-response/coronavirus-disease-2019-covid-19/covid-19-vaccines

34. Canada H. COVID-19 Vaccines: Authorized vaccines [Internet]. aem. 2020 [cited 2021 Mar 15]. Available from: https://www.canada.ca/en/health-canada/services/drugs-health-products/covid19-industry/drugs-vaccines-treatments/vaccines.html

35. Sahin U, Muik A, Derhovanessian E, Vogler I, Kranz LM, Vormehr M, et al. COVID-19 vaccine BNT162b1 elicits human antibody and TH1 T cell responses. Nature. 2020 Oct 22;586(7830):594–9.

36. Sadoff J, Le Gars M, Shukarev G, Heerwegh D, Truyers C, de Groot AM, et al. Interim Results of a Phase 1–2a Trial of Ad26.COV2.S Covid-19 Vaccine. N Engl J Med. 2021 Jan 13;NEJMoa2034201.

37. Hamilton BS, Whittaker GR, Daniel S. Influenza virus-mediated membrane fusion: determinants of hemagglutinin fusogenic activity and experimental approaches for assessing virus fusion. Viruses. 2012;4(7):1144–68.

38. Te Velthuis AJW, Fodor E. Influenza virus RNA polymerase: insights into the mechanisms of viral RNA synthesis. Nat Rev Microbiol. 2016;14(8):479–93.

39. Zheng W, Tao YJ. Structure and assembly of the influenza A virus ribonucleoprotein complex. FEBS Lett. 2013 Apr 17;587(8):1206–14.

40. Ramos-Nascimento A, Kellen B, Ferreira F, Alenquer M, Vale-Costa S, Raposo G, et al. KIF13A mediates trafficking of influenza A virus ribonucleoproteins. J Cell Sci. 2017 Dec 1;130(23):4038–50.

41. Nakatsu S, Sagara H, Sakai-Tagawa Y, Sugaya N, Noda T, Kawaoka Y. Complete and Incomplete Genome Packaging of Influenza A and B Viruses. MBio. 2016 Sep 6;7(5).

42. Jane M, Vidal MJ, Soldevila N, Romero A, Martinez A, Torner N, et al. Epidemiological and clinical characteristics of children hospitalized due to influenza A and B in the south of Europe, 2010-2016. Sci Rep. 2019 Sep 6;9(1):12853.

43. Zheng Z, Paul SS, Mo X, Yuan Y-RA, Tan Y-J. The Vestigial Esterase Domain of Haemagglutinin of H5N1 Avian Influenza A Virus: Antigenicity and Contribution to Viral Pathogenesis. Vaccines. 2018 Sep;6(3):53.

44. Shih AC-C, Hsiao T-C, Ho M-S, Li W-H. Simultaneous amino acid substitutions at antigenic sites drive influenza A hemagglutinin evolution. Proc Natl Acad Sci U S A. 2007 Apr 10;104(15):6283–8.

45. Wilson IA, Skehel JJ, Wiley DC. Structure of the haemagglutinin membrane glycoprotein of influenza virus at 3 Å resolution. Nature. 1981 Jan;289(5796):366–73.

46. Stevens J, Corper AL, Basler CF, Taubenberger JK, Palese P, Wilson IA. Structure of the uncleaved human H1 hemagglutinin from the extinct 1918 influenza virus. Science. 2004 Mar 19;303(5665):1866–70.

47. Palese P, Tobita K, Ueda M, Compans RW. Characterization of temperature sensitive influenza virus mutants defective in neuraminidase. Virology. 1974 Oct 1;61(2):397–410.

48. Varghese JN, Laver WG, Colman PM. Structure of the influenza virus glycoprotein antigen neuraminidase at 2.9 Å resolution. Nature. 1983 May;303(5912):35–40.

49. Harris A, Cardone G, Winkler DC, Heymann JB, Brecher M, White JM, et al. Influenza virus pleiomorphy characterized by cryoelectron tomography. Proc Natl Acad Sci. 2006 Dec 12;103(50):19123–7.

50. Blok J, Air GM. Variation in the membrane-insertion and “stalk” sequences in eight subtypes of influenza type A virus neuraminidase. Biochemistry. 1982 Aug 17;21(17):4001–7.

51. Air GM. Influenza neuraminidase. Influenza Other Respir Viruses. 2012 Jul;6(4):245–56.

52. Sultana I, Yang K, Getie-Kebtie M, Couzens L, Markoff L, Alterman M, et al. Stability of neuraminidase in inactivated influenza vaccines. Vaccine. 2014 Apr 17;32(19):2225–30.

53. V’kovski P, Kratzel A, Steiner S, Stalder H, Thiel V. Coronavirus biology and replication: implications for SARS-CoV-2. Nat Rev Microbiol. 2021 Mar;19(3):155–70.

54. Ziegler CGK, Allon SJ, Nyquist SK, Mbano IM, Miao VN, Tzouanas CN, et al. SARS-CoV-2 Receptor ACE2 Is an Interferon-Stimulated Gene in Human Airway Epithelial Cells and Is Detected in Specific Cell Subsets across Tissues. Cell. 2020 May 28;181(5):1016-1035.e19.

55. Wölfel R, Corman VM, Guggemos W, Seilmaier M, Zange S, Müller MA, et al. Virological assessment of hospitalized patients with COVID-2019. Nature. 2020 May 28;581(7809):465–9.

56. Hamming I, Timens W, Bulthuis M, Lely A, Navis G, van Goor H. Tissue distribution of ACE2 protein, the functional receptor for SARS coronavirus. A first step in understanding SARS pathogenesis. J Pathol. 2004;203(2):631–7.

57. Shieh W-J, Hsiao C-H, Paddock CD, Guarner J, Goldsmith CS, Tatti K, et al. Immunohistochemical, in situ hybridization, and ultrastructural localization of SARS-associated coronavirus in lung of a fatal case of severe acute respiratory syndrome in Taiwan. Hum Pathol. 2005 Mar 1;36(3):303–9.

58. Li W, Moore MJ, Vasilieva N, Sui J, Wong SK, Berne MA, et al. Angiotensin-converting enzyme 2 is a functional receptor for the SARS coronavirus. Nature. 2003 Nov;426(6965):450–4.

59. Zhou P, Yang X-L, Wang X-G, Hu B, Zhang L, Zhang W, et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature. 2020 Mar;579(7798):270–3.

60. Qian Z, Travanty EA, Oko L, Edeen K, Berglund A, Wang J, et al. Innate Immune Response of Human Alveolar Type II Cells Infected with Severe Acute Respiratory Syndrome–Coronavirus. Am J Respir Cell Mol Biol. 2013 Jun 1;48(6):742–8.

61. Gierer S, Bertram S, Kaup F, Wrensch F, Heurich A, Krämer-Kühl A, et al. The Spike Protein of the Emerging Betacoronavirus EMC Uses a Novel Coronavirus Receptor for Entry, Can Be Activated by TMPRSS2, and Is Targeted by Neutralizing Antibodies. J Virol. 2013 May 15;87(10):5502–11.

62. Ulasli M, Verheije MH, de Haan CAM, Reggiori F. Qualitative and quantitative ultrastructural analysis of the membrane rearrangements induced by coronavirus. Cell Microbiol. 2010;12(6):844–61.

63. Wolff G, Limpens RWAL, Zevenhoven-Dobbe JC, Laugks U, Zheng S, de Jong AWM, et al. A molecular pore spans the double membrane of the coronavirus replication organelle. Science. 2020 Sep 11;369(6509):1395–8.

64. Klein S, Cortese M, Winter SL, Wachsmuth-Melm M, Neufeldt CJ, Cerikan B, et al. SARS-CoV-2 structure and replication characterized by in situ cryo-electron tomography. Nat Commun. 2020 Nov 18;11(1):5885.

65. Siu YL, Teoh KT, Lo J, Chan CM, Kien F, Escriou N, et al. The M, E, and N Structural Proteins of the Severe Acute Respiratory Syndrome Coronavirus Are Required for Efficient Assembly, Trafficking, and Release of Virus-Like Particles. J Virol. 2008 Nov 15;82(22):11318–30.

66. Iwasaki A, Medzhitov R. Control of adaptive immunity by the innate immune system. Nat Immunol. 2015 Apr;16(4):343–53.

67. Iwasaki A, Pillai PS. Innate immunity to influenza virus infection. Nat Rev Immunol. 2014 May;14(5):315–28.

68. Horisberger MA, Staeheli P, Haller O. Interferon induces a unique protein in mouse cells bearing a gene for resistance to influenza virus. Proc Natl Acad Sci. 1983 Apr 1;80(7):1910–4.

69. Schoggins JW. Interferon-Stimulated Genes: What Do They All Do? Annu Rev Virol. 2019;6(1):567–84.

70. Gazit R, Gruda R, Elboim M, Arnon TI, Katz G, Achdout H, et al. Lethal influenza infection in the absence of the natural killer cell receptor gene Ncr1. Nat Immunol. 2006 May;7(5):517–23.

71. Hashimoto Y, Moki T, Takizawa T, Shiratsuchi A, Nakanishi Y. Evidence for Phagocytosis of Influenza Virus-Infected, Apoptotic Cells by Neutrophils and Macrophages in Mice. J Immunol. 2007 Feb 15;178(4):2448–57.

72. Chaplin DD. Overview of the Immune Response. J Allergy Clin Immunol. 2010 Feb;125(2 Suppl 2):S3-23.

73. Wong P, Pamer EG. CD8 T Cell Responses to Infectious Pathogens. Annu Rev Immunol. 2003;21(1):29–70.

74. Alberts B, Johnson A, Lewis J, Raff M, Roberts K, Walter P. Helper T Cells and Lymphocyte Activation. Mol Biol Cell 4th Ed [Internet]. 2002 [cited 2021 Nov 13]; Available from: https://www.ncbi.nlm.nih.gov/books/NBK26827/

75. Tesmer LA, Lundy SK, Sarkar S, Fox DA. Th17 cells in human disease. Immunol Rev. 2008 Jun;223:87–113.

76. Crotty S. T Follicular Helper Cell Biology: A Decade of Discovery and Diseases. Immunity. 2019 May 21;50(5):1132–48.

77. Cruz-Tapias P, Castiblanco J, Anaya J-M. Major histocompatibility complex: Antigen processing and presentation [Internet]. Autoimmunity: From Bench to Bedside [Internet]. El Rosario University Press; 2013 [cited 2021 Oct 26]. Available from: https://www.ncbi.nlm.nih.gov/books/NBK459467/

78. IPD-IMGT/HLA Database [Internet]. [cited 2021 Oct 26]. Available from: https://www.ebi.ac.uk/ipd/imgt/hla/

79. Mack SJ, Cano P, Hollenbach JA, He J, Hurley CK, Middleton D, et al. Common and Well-Documented HLA Alleles: 2012 Update to the CWD Catalogue. Tissue Antigens. 2013 Apr;81(4):194–203.

80. Cano P, Klitz W, Mack SJ, Maiers M, Marsh SGE, Noreen H, et al. Common and well-documented HLA alleles: report of the Ad-Hoc committee of the american society for histocompatiblity and immunogenetics. Hum Immunol. 2007 May;68(5):392–417.

81. Greenbaum J, Sidney J, Chung J, Brander C, Peters B, Sette A. Functional classification of class II human leukocyte antigen (HLA) molecules reveals seven different supertypes and a surprising degree of repertoire sharing across supertypes. Immunogenetics. 2011 Jun;63(6):325–35.

82. Lund O, Nielsen M, Kesmir C, Petersen AG, Lundegaard C, Worning P, et al. Definition of supertypes for HLA molecules using clustering of specificity matrices. Immunogenetics. 2004 Mar;55(12):797–810.

83. Wang M, Claesson MH. Classification of human leukocyte antigen (HLA) supertypes. Methods Mol Biol Clifton NJ. 2014;1184:309–17.

84. Reth M. Antigen Receptors on B Lymphocytes. Annu Rev Immunol. 1992;10(1):97–121.

85. Finney J, Yeh C-H, Kelsoe G, Kuraoka M. Germinal center responses to complex antigens. Immunol Rev. 2018;284(1):42–50.

86. Turner JS, Zhou JQ, Han J, Schmitz AJ, Rizk AA, Alsoussi WB, et al. Human germinal centres engage memory and naive B cells after influenza vaccination. Nature. 2020 Oct;586(7827):127–32.

87. Oslund KL, Baumgarth N. Influenza-induced innate immunity: regulators of viral replication, respiratory tract pathology & adaptive immunity. Future Virol. 2011 Aug;6(8):951–62.

88. Ramos I, Bernal-Rubio D, Durham N, Belicha-Villanueva A, Lowen AC, Steel J, et al. Effects of Receptor Binding Specificity of Avian Influenza Virus on the Human Innate Immune Response. J Virol. 2011 May;85(9):4421–31.

89. Goffic RL, Balloy V, Lagranderie M, Alexopoulou L, Escriou N, Flavell R, et al. Detrimental Contribution of the Toll-Like Receptor (TLR)3 to Influenza A Virus–Induced Acute Pneumonia. PLOS Pathog. 2006 Jun 9;2(6):e53.

90. Kato H, Takahasi K, Fujita T. RIG-I-like receptors: cytoplasmic sensors for non-self RNA. Immunol Rev. 2011 Sep;243(1):91–8.

91. Pichlmair A, Schulz O, Tan CP, Näslund TI, Liljeström P, Weber F, et al. RIG-I-Mediated Antiviral Responses to Single-Stranded RNA Bearing 5'-Phosphates. Science. 2006 Nov 10;314(5801):997–1001.

92. Swanson KV, Deng M, Ting JP-Y. The NLRP3 inflammasome: molecular activation and regulation to therapeutics. Nat Rev Immunol. 2019 Aug;19(8):477–89.

93. Kanneganti T-D. Central roles of NLRs and inflammasomes in viral infection. Nat Rev Immunol. 2010 Oct;10(10):688–98.

94. Heer AK, Shamshiev A, Donda A, Uematsu S, Akira S, Kopf M, et al. TLR signaling fine-tunes anti-influenza B cell responses without regulating effector T cell responses. J Immunol Baltim Md 1950. 2007 Feb 15;178(4):2182–91.

95. Chan MCW, Cheung CY, Chui WH, Tsao SW, Nicholls JM, Chan YO, et al. Proinflammatory cytokine responses induced by influenza A (H5N1) viruses in primary human alveolar and bronchial epithelial cells. Respir Res. 2005 Nov 11;6:135.

96. Xing Z, Harper R, Anunciacion J, Yang Z, Gao W, Qu B, et al. Host immune and apoptotic responses to avian influenza virus H9N2 in human tracheobronchial epithelial cells. Am J Respir Cell Mol Biol. 2011 Jan;44(1):24–33.

97. Herold S, von Wulffen W, Steinmueller M, Pleschka S, Kuziel WA, Mack M, et al. Alveolar epithelial cells direct monocyte transepithelial migration upon influenza virus infection: impact of chemokines and adhesion molecules. J Immunol Baltim Md 1950. 2006 Aug 1;177(3):1817–24.

98. Yarilina A, Park-Min K-H, Antoniv T, Hu X, Ivashkiv LB. TNF activates an IRF1-dependent autocrine loop leading to sustained expression of chemokines and STAT1-dependent type I interferon–response genes. Nat Immunol. 2008 Apr;9(4):378–87.

99. Kohlmeier JE, Cookenham T, Roberts AD, Miller SC, Woodland DL. Type I interferons regulate cytolytic activity of memory CD8(+) T cells in the lung airways during respiratory virus challenge. Immunity. 2010 Jul 23;33(1):96–105.

100. Coro ES, Chang WLW, Baumgarth N. Type I IFN receptor signals directly stimulate local B cells early following influenza virus infection. J Immunol Baltim Md 1950. 2006 Apr 1;176(7):4343–51.

101. Ichinohe T, Lee HK, Ogura Y, Flavell R, Iwasaki A. Inflammasome recognition of influenza virus is essential for adaptive immune responses. J Exp Med. 2009 Jan 16;206(1):79–87.

102. Wrammert J, Smith K, Miller J, Langley WA, Kokko K, Larsen C, et al. Rapid cloning of high-affinity human monoclonal antibodies against influenza virus. Nature. 2008 May;453(7195):667–71.

103. Reber AJ, Music N, Kim JH, Gansebom S, Chen J, York I. Extensive T cell cross-reactivity between diverse seasonal influenza strains in the ferret model. Sci Rep. 2018 Apr 17;8(1):6112.

104. Grant E, Wu C, Chan K-F, Eckle S, Bharadwaj M, Zou QM, et al. Nucleoprotein of influenza A virus is a major target of immunodominant CD8+ T-cell responses. Immunol Cell Biol. 2013 Feb;91(2):184–94.

105. Lee LY-H, Ha DLA, Simmons C, de Jong MD, Chau NVV, Schumacher R, et al. Memory T cells established by seasonal human influenza A infection cross-react with avian influenza A (H5N1) in healthy individuals. J Clin Invest. 2008 Oct;118(10):3478–90.

106. Blanco-Melo D, Nilsson-Payant BE, Liu W-C, Uhl S, Hoagland D, Møller R, et al. Imbalanced Host Response to SARS-CoV-2 Drives Development of COVID-19. Cell. 2020 May 28;181(5):1036-1045.e9.

107. Grifoni A, Weiskopf D, Ramirez SI, Mateus J, Dan JM, Moderbacher CR, et al. Targets of T Cell Responses to SARS-CoV-2 Coronavirus in Humans with COVID-19 Disease and Unexposed Individuals. Cell. 2020 Jun 25;181(7):1489-1501.e15.

108. Geers D, Shamier MC, Bogers S, den Hartog G, Gommers L, Nieuwkoop NN, et al. SARS-CoV-2 variants of concern partially escape humoral but not T cell responses in COVID-19 convalescent donors and vaccine recipients. Sci Immunol. 2021 May 25;6(59):eabj1750.

109. Rydyznski Moderbacher C, Ramirez SI, Dan JM, Grifoni A, Hastie KM, Weiskopf D, et al. Antigen-Specific Adaptive Immunity to SARS-CoV-2 in Acute COVID-19 and Associations with Age and Disease Severity. Cell. 2020 Nov 12;183(4):996-1012.e19.

110. Ju B, Zhang Q, Ge J, Wang R, Sun J, Ge X, et al. Human neutralizing antibodies elicited by SARS-CoV-2 infection. Nature. 2020 Aug;584(7819):115–9.

111. Rogers TF, Zhao F, Huang D, Beutler N, Burns A, He W, et al. Isolation of potent SARS-CoV-2 neutralizing antibodies and protection from disease in a small animal model. Science. 2020 Aug 21;369(6506):956–63.

112. Subbarao K. SARS-CoV-2: A New Song Recalls an Old Melody. Cell Host Microbe. 2020 May 13;27(5):692–4.

113. Robbiani DF, Gaebler C, Muecksch F, Lorenzi JCC, Wang Z, Cho A, et al. Convergent antibody responses to SARS-CoV-2 in convalescent individuals. Nature. 2020 Aug;584(7821):437–42.

114. Francis ME, McNeil M, Dawe NJ, Foley MK, King ML, Ross TM, et al. Historical H1N1 Influenza Virus Imprinting Increases Vaccine Protection by Influencing the Activity and Sustained Production of Antibodies Elicited at Vaccination in Ferrets. Vaccines. 2019 Sep 28;7(4):133.

115. Krammer F, Smith GJD, Fouchier RAM, Peiris M, Kedzierska K, Doherty PC, et al. Influenza. Nat Rev Dis Primer. 2018 Dec;4(1):3.

116. Belser JA, Pulit-Penaloza JA, Maines TR. Ferreting Out Influenza Virus Pathogenicity and Transmissibility: Past and Future Risk Assessments in the Ferret Model. Cold Spring Harb Perspect Med. 2020 Jan 7;10(7):a038323.

117. Francis ME, King ML, Kelvin AA. Back to the Future for Influenza Preimmunity-Looking Back at Influenza Virus History to Infer the Outcome of Future Infections. Viruses. 2019 30;11(2).

118. Both GW, Sleigh MJ. Conservation and variation in the hemagglutinins of Hong Kong subtype influenza viruses during antigenic drift. J Virol. 1981 Sep;39(3):663–72.

119. Both GW, Sleigh MJ, Cox NJ, Kendal AP. Antigenic drift in influenza virus H3 hemagglutinin from 1968 to 1980: multiple evolutionary pathways and sequential amino acid changes at key antigenic sites. J Virol. 1983 Oct;48(1):52–60.

120. Popova L, Smith K, West AH, Wilson PC, James JA, Thompson LF, et al. Immunodominance of Antigenic Site B over Site A of Hemagglutinin of Recent H3N2 Influenza Viruses. PLOS ONE. 2012 Jul 25;7(7):e41895.

121. Gerhard W, Yewdell J, Frankel ME, Webster R. Antigenic structure of influenza virus haemagglutinin defined by hybridoma antibodies. Nature. 1981 Apr;290(5808):713–7.

122. Treanor J. Influenza Vaccine — Outmaneuvering Antigenic Shift and Drift. N Engl J Med. 2004 Jan 15;350(3):218–20.

123. Hay AJ, Gregory V, Douglas AR, Lin YP. The evolution of human influenza viruses. Philos Trans R Soc Lond B Biol Sci. 2001 Dec 29;356(1416):1861–70.

124. CDC. How Flu Viruses Can Change [Internet]. Centers for Disease Control and Prevention. 2019 [cited 2020 Sep 17]. Available from: https://www.cdc.gov/flu/about/viruses/change.htm

125. Kim H, Webster RG, Webby RJ. Influenza Virus: Dealing with a Drifting and Shifting Pathogen. Viral Immunol. 2018;31(2):174–83.

126. Kelvin AA, Zambon M. Influenza imprinting in childhood and the influence on vaccine response later in life. Eurosurveillance. 2019 Nov 28;24(48):1900720.

127. Glinsky GV. Genomic analysis of pandemic (H1N1) 2009 reveals association of increasing disease severity with emergence of novel hemagglutinin mutations. Cell Cycle. 2010 Mar 1;9(5):958–70.

128. Kilander A, Rykkvin R, Dudman SG, Hungnes O. Observed association between the HA1 mutation D222G in the 2009 pandemic influenza A(H1N1) virus and severe clinical outcome, Norway 2009-2010. Eurosurveillance. 2010 Mar 4;15(9):19498.

129. Chong Y, Ikematsu H. Effect of seasonal vaccination on the selection of influenza A/H3N2 epidemic variants. Vaccine. 2017 Jan 5;35(2):255–63.

130. Parisi A, Lopes JS, Nunes A, Gomes MGM. Heterogeneity in antibody range and the antigenic drift of influenza A viruses. Ecol Complex. 2013 Jun 1;14:157–65.

131. Ralph R, Lew J, Zeng T, Francis ME, Bei X, Rioux M, et al. 2019-nCoV (Wuhan virus), A Novel Coronavirus, linked to Chinese Pneumonia Cases: Human-to-Human Transmission, Travel-Related Cases and Vaccine Readiness. J Infect Dev Ctries. 2020;

132. Weissman D, Alameh M-G, de Silva T, Collini P, Hornsby H, Brown R, et al. D614G Spike Mutation Increases SARS CoV-2 Susceptibility to Neutralization. Cell Host Microbe. 2021 Jan 13;29(1):23-31.e4.

133. Zhang L, Jackson CB, Mou H, Ojha A, Peng H, Quinlan BD, et al. SARS-CoV-2 spike-protein D614G mutation increases virion spike density and infectivity. Nat Commun. 2020 Nov 26;11(1):6013.

134. Zhang J, Cai Y, Xiao T, Lu J, Peng H, Sterling SM, et al. Structural impact on SARS-CoV-2 spike protein by D614G substitution. Science [Internet]. 2021 Mar 16 [cited 2021 Mar 20]; Available from: https://science.sciencemag.org/content/early/2021/03/16/science.abf2303

135. Callaway E. Could new COVID variants undermine vaccines? Labs scramble to find out. Nature. 2021 Jan 7;589(7841):177–8.

136. Tracking SARS-CoV-2 variants [Internet]. [cited 2021 Jul 25]. Available from: https://www.who.int/activities/tracking-SARS-CoV-2-variants

137. Investigation of novel SARS-CoV-2 variant: Variant of Concern 202012/01. :19.

138. Davies NG, Abbott S, Barnard RC, Jarvis CI, Kucharski AJ, Munday JD, et al. Estimated transmissibility and impact of SARS-CoV-2 lineage B.1.1.7 in England. Science [Internet]. 2021 Mar 3 [cited 2021 Mar 15]; Available from: https://science.sciencemag.org/content/early/2021/03/03/science.abg3055

139. Campbell F, Archer B, Laurenson-Schafer H, Jinnai Y, Konings F, Batra N, et al. Increased transmissibility and global spread of SARS-CoV-2 variants of concern as at June 2021. Eurosurveillance [Internet]. 2021 Jun 17 [cited 2021 Jul 26];26(24). Available from: https://www.eurosurveillance.org/content/10.2807/1560-7917.ES.2021.26.24.2100509

140. Horby P, Bell I, Breuer J, Cevik M, Challen R, Davies N, et al. Update note on B.1.1.7 severity. :14.

141. Davies NG, Jarvis CI, Edmunds WJ, Jewell NP, Diaz-Ordaz K, Keogh RH. Increased mortality in community-tested cases of SARS-CoV-2 lineage B.1.1.7. Nature. 2021 May;593(7858):270–4.

142. Here's what we know about the new variant of coronavirus | Sharon Peacock [Internet]. the Guardian. 2020 [cited 2021 Jul 25]. Available from: http://www.theguardian.com/commentisfree/2020/dec/22/new-variant-coronavirus-genomic-sars-cov-2-pandemic

143. Preliminary genomic characterisation of an emergent SARS-CoV-2 lineage in the UK defined by a novel set of spike mutations - SARS-CoV-2 coronavirus / nCoV-2019 Genomic Epidemiology [Internet]. Virological. 2020 [cited 2021 Mar 15]. Available from: https://virological.org/t/preliminary-genomic-characterisation-of-an-emergent-sars-cov-2-lineage-in-the-uk-defined-by-a-novel-set-of-spike-mutations/563

144. Shen X, Tang H, McDanal C, Wagh K, Fischer W, Theiler J, et al. SARS-CoV-2 variant B.1.1.7 is susceptible to neutralizing antibodies elicited by ancestral spike vaccines. Cell Host Microbe. 2021 Mar;S1931312821001025.

145. Supasa P, Zhou D, Dejnirattisai W, Liu C, Mentzer AJ, Ginn HM, et al. Reduced neutralization of SARS-CoV-2 B.1.1.7 variant by convalescent and vaccine sera. Cell [Internet]. 2021 Feb 18 [cited 2021 Mar 20];0(0). Available from: https://www.cell.com/cell/abstract/S0092-8674(21)00222-1

146. Muik A, Wallisch A-K, Sänger B, Swanson KA, Mühl J, Chen W, et al. Neutralization of SARS-CoV-2 lineage B.1.1.7 pseudovirus by BNT162b2 vaccine–elicited human sera. Science. 2021 Mar 12;371(6534):1152–3.

147. Wang P, Liu L, Iketani S, Luo Y, Guo Y, Wang M, et al. Increased Resistance of SARS-CoV-2 Variants B.1.351 and B.1.1.7 to Antibody Neutralization. bioRxiv. 2021 Jan 26;2021.01.25.428137.

148. Tegally H, Wilkinson E, Giovanetti M, Iranzadeh A, Fonseca V, Giandhari J, et al. Emergence and rapid spread of a new severe acute respiratory syndrome-related coronavirus 2 (SARS-CoV-2) lineage with multiple spike mutations in South Africa. medRxiv. 2020 Dec 22;2020.12.21.20248640.

149. Tang JW, Toovey OTR, Harvey KN, Hui DDS. Introduction of the South African SARS-CoV-2 variant 501Y.V2 into the UK. J Infect [Internet]. 2021 Jan 17 [cited 2021 Mar 15];0(0). Available from: https://www.journalofinfection.com/article/S0163-4453(21)00030-X/abstract

150. Wu L, Peng C, Xu Z, Zhu W. Predicting the Potential Effect of E484K Mutation on the Binding of 28 Antibodies to the Spike Protein of SARS-CoV-2 by Molecular Dynamics Simulation and Free Energy Calculation. 2021 Feb 11 [cited 2021 Mar 20]; Available from: /articles/preprint/Predicting_the_Potential_Effect_of_E484K_Mutation_on_the_Binding_of_28_Antibodies_to_the_Spike_Protein_of_SARS-CoV-2_by_Molecular_Dynamics_Simulation_and_Free_Energy_Calculation/13897091/1

151. Weekly epidemiological update on COVID-19 - 20 July 2021 [Internet]. [cited 2021 Jul 26]. Available from: https://www.who.int/publications/m/item/weekly-epidemiological-update-on-covid-19---20-july-2021

152. Hoffmann M, Hofmann-Winkler H, Krüger N, Kempf A, Nehlmeier I, Graichen L, et al. SARS-CoV-2 variant B.1.617 is resistant to bamlanivimab and evades antibodies induced by infection and vaccination. Cell Rep. 2021 Jul;36(3):109415.

153. Genomic characterisation of an emergent SARS-CoV-2 lineage in Manaus: preliminary findings - SARS-CoV-2 coronavirus / nCoV-2019 Genomic Epidemiology [Internet]. Virological. 2021 [cited 2021 Mar 20]. Available from: https://virological.org/t/genomic-characterisation-of-an-emergent-sars-cov-2-lineage-in-manaus-preliminary-findings/586

154. Freitas ARR, Lemos DRQ, Beckedorff OA, Cavalcanti LP de G, Siqueira AM, Mello RCS de, et al. The increase in the risk of severity and fatality rate of covid-19 in southern Brazil after the emergence of the Variant of Concern (VOC) SARS-CoV-2 P.1 was greater among young adults without pre-existing risk conditions. medRxiv. 2021 Apr 19;2021.04.13.21255281.

155. Weekly epidemiological update on COVID-19 - 11 May 2021 [Internet]. [cited 2021 Jun 4]. Available from: https://www.who.int/publications/m/item/weekly-epidemiological-update-on-covid-19---11-may-2021

156. Weekly epidemiological update on COVID-19 - 27 April 2021 [Internet]. [cited 2021 Jun 4]. Available from: https://www.who.int/publications/m/item/weekly-epidemiological-update-on-covid-19---27-april-2021

157. Fisman DN, Tuite AR. Progressive Increase in Virulence of Novel SARS-CoV-2 Variants in Ontario, Canada. medRxiv. 2021 Jul 12;2021.07.05.21260050.

158. Investigation of SARS-CoV-2 variants of concern: variant risk assessments [Internet]. GOV.UK. [cited 2021 Jul 26]. Available from: https://www.gov.uk/government/publications/investigation-of-sars-cov-2-variants-of-concern-variant-risk-assessments

159. “Double mutant”: What are the risks of India’s new Covid-19 variant. BBC News [Internet]. 2021 Mar 25 [cited 2021 Jul 6]; Available from: https://www.bbc.com/news/world-asia-india-56517495

160. Zhang W, Davis BD, Chen SS, Martinez JMS, Plummer JT, Vail E. Emergence of a novel SARS-CoV-2 strain in Southern California, USA. medRxiv. 2021 Jan 20;2021.01.18.21249786.

161. Oude Munnink BB, Sikkema RS, Nieuwenhuijse DF, Molenaar RJ, Munger E, Molenkamp R, et al. Transmission of SARS-CoV-2 on mink farms between humans and mink and back to humans. Science. 2021 Jan 8;371(6525):172–7.

162. Larsen HD, Fonager J, Lomholt FK, Dalby T, Benedetti G, Kristensen B, et al. Preliminary report of an outbreak of SARS-CoV-2 in mink and mink farmers associated with community spread, Denmark, June to November 2020. Eurosurveillance. 2021 Feb 4;26(5):2100009.

163. Schrörs B, Gudimella R, Bukur T, Rösler T, Löwer M, Sahin U. Large-scale analysis of SARS-CoV-2 spike-glycoprotein mutants demonstrates the need for continuous screening of virus isolates. bioRxiv. 2021 Feb 4;2021.02.04.429765.

164. Rajao DS, Perez DR. Universal Vaccines and Vaccine Platforms to Protect against Influenza Viruses in Humans and Agriculture. Front Microbiol. 2018;9:123.

165. Vemula SV, Sayedahmed EE, Sambhara S, Mittal SK. Vaccine approaches conferring cross-protection against influenza viruses. Expert Rev Vaccines. 2017 Nov;16(11):1141–54.

166. Tisa V, Barberis I, Faccio V, Paganino C, Trucchi C, Martini M, et al. Quadrivalent influenza vaccine: a new opportunity to reduce the influenza burden. J Prev Med Hyg. 2016;57(1):E28-33.

167. Fiore AE, Shay DK, Broder K, Iskander JK, Uyeki TM, Mootrey G, et al. Prevention and control of seasonal influenza with vaccines: recommendations of the Advisory Committee on Immunization Practices (ACIP), 2009. MMWR Recomm Rep. 2009 Jul 31;58(RR-8):1–52.

168. Vemula SV, Sayedahmed EE, Sambhara S, Mittal SK. Vaccine approaches conferring cross-protection against influenza viruses. Expert Rev Vaccines. 2017 Nov 2;16(11):1141–54.

169. Andrews SF, Huang Y, Kaur K, Popova LI, Ho IY, Pauli NT, et al. Immune history profoundly affects broadly protective B cell responses to influenza. Sci Transl Med. 2015 Dec 2;7(316):316ra192.

170. Fonville JM, Wilks SH, James SL, Fox A, Ventresca M, Aban M, et al. Antibody landscapes after influenza virus infection or vaccination. Science. 2014 Nov 21;346(6212):996–1000.

171. Pan K. Understanding original antigenic sin in influenza with a dynamical system. PLoS One. 2011;6(8):e23910.

172. Dong W, Bhide Y, Sicca F, Meijerhof T, Guilfoyle K, Engelhardt OG, et al. Cross-Protective Immune Responses Induced by Sequential Influenza Virus Infection and by Sequential Vaccination With Inactivated Influenza Vaccines. Front Immunol [Internet]. 2018 [cited 2020 Sep 18];9. Available from: https://www.frontiersin.org/articles/10.3389/fimmu.2018.02312/full

173. Yewdell JW, Bennink JR, Smith GL, Moss B. Influenza A virus nucleoprotein is a major target antigen for cross-reactive anti-influenza A virus cytotoxic T lymphocytes. Proc Natl Acad Sci U S A. 1985 Mar;82(6):1785–9.

174. Tu W, Mao H, Zheng J, Liu Y, Chiu SS, Qin G, et al. Cytotoxic T Lymphocytes Established by Seasonal Human Influenza Cross-React against 2009 Pandemic H1N1 Influenza Virus. J Virol. 2010 Jul 1;84(13):6527–35.

175. Monto AS, Malosh RE, Petrie JG, Martin ET. The Doctrine of Original Antigenic Sin: Separating Good From Evil. J Infect Dis. 2017 Jun 15;215(12):1782–8.

176. Henry C, Palm AE, Krammer F, Wilson PC. From Original Antigenic Sin to the Universal Influenza Virus Vaccine. Trends Immunol. 2018 Jan;39(1):70–9.

177. Ellebedy AH, Ahmed R. Re-engaging cross-reactive memory B cells: the influenza puzzle. Front Immunol. 2012;3:53.

178. Hancock K, Veguilla V, Lu X, Zhong W, Butler EN, Sun H, et al. Cross-reactive antibody responses to the 2009 pandemic H1N1 influenza virus. N Engl J Med. 2009 Nov 12;361(20):1945–52.

179. Gostic KM, Ambrose M, Worobey M, Lloyd-Smith JO. Potent protection against H5N1 and H7N9 influenza via childhood hemagglutinin imprinting. Science. 2016 Nov 11;354(6313):722–6.

180. Linderman SL, Hensley SE. Antibodies with “Original Antigenic Sin” Properties Are Valuable Components of Secondary Immune Responses to Influenza Viruses. PLoS Pathog. 2016 Aug;12(8):e1005806.

181. Ramakrishnan MA. Determination of 50% endpoint titer using a simple formula. World J Virol. 2016 May 12;5(2):85–6.

182. Hoffmann E, Stech J, Guan Y, Webster RG, Perez DR. Universal primer set for the full-length amplification of all influenza A viruses. Arch Virol. 2001 Dec;146(12):2275–89.

183. Green MR, Sambrook J, Sambrook J. Molecular cloning: a laboratory manual. 4th ed. Cold Spring Harbor, N.Y: Cold Spring Harbor Laboratory Press; 2012. 3 p.

184. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, et al. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinforma Oxf Engl. 2012 Jun 15;28(12):1647–9.

185. Stecher G, Tamura K, Kumar S. Molecular Evolutionary Genetics Analysis (MEGA) for macOS. Mol Biol Evol. 2020 Apr 1;37(4):1237–9.

186. Elbe S, Buckland-Merrett G. Data, disease and diplomacy: GISAID's innovative contribution to global health: Data, Disease and Diplomacy. Glob Chall. 2017 Jan;1(1):33–46.

187. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, et al. Clustal W and Clustal X version 2.0. Bioinforma Oxf Engl. 2007 Nov 1;23(21):2947–8.

188. Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJE. The Phyre2 web portal for protein modeling, prediction and analysis. Nat Protoc. 2015 Jun;10(6):845–58.

189. Ittisoponpisan S, Islam SA, Khanna T, Alhuzimi E, David A, Sternberg MJE. Can Predicted Protein 3D Structures Provide Reliable Insights into whether Missense Variants Are Disease Associated? J Mol Biol. 2019 May 17;431(11):2197–212.

190. Waterhouse A, Bertoni M, Bienert S, Studer G, Tauriello G, Gumienny R, et al. SWISS-MODEL: homology modelling of protein structures and complexes. Nucleic Acids Res. 2018 Jul 2;46(Web Server issue):W296–303.

191. Benkert P, Biasini M, Schwede T. Toward the estimation of the absolute quality of individual protein structure models. Bioinforma Oxf Engl. 2011 Feb 1;27(3):343–50.

192. Kringelum JV, Lundegaard C, Lund O, Nielsen M. Reliable B Cell Epitope Predictions: Impacts of Method Development and Improved Benchmarking. PLOS Comput Biol. 2012 Dec 27;8(12):e1002829.

193. Wickham H. ggplot2: Elegant Graphics for Data Analysis. 2nd ed. 2016. Cham: Springer International Publishing : Imprint: Springer; 2016. 1 p. (Use R!).

194. Blom N, Sicheritz-Pontén T, Gupta R, Gammeltoft S, Brunak S. Prediction of post-translational glycosylation and phosphorylation of proteins from the amino acid sequence. Proteomics. 2004 Jun;4(6):1633–49.

195. Grant OC, Montgomery D, Ito K, Woods RJ. Analysis of the SARS-CoV-2 spike protein glycan shield reveals implications for immune recognition. Sci Rep. 2020 Sep 14;10(1):14991.

196. Weiskopf D, Angelo MA, de Azeredo EL, Sidney J, Greenbaum JA, Fernando AN, et al. Comprehensive analysis of dengue virus-specific responses supports an HLA-linked protective role for CD8+ T cells. Proc Natl Acad Sci U S A. 2013 May 28;110(22):E2046-2053.

197. Jensen KK, Andreatta M, Marcatili P, Buus S, Greenbaum JA, Yan Z, et al. Improved methods for predicting peptide binding affinity to MHC class II molecules. Immunology. 2018 Jul;154(3):394–406.

198. Wickham H, Averick M, Bryan J, Chang W, McGowan L, François R, et al. Welcome to the Tidyverse. J Open Source Softw. 2019 Nov 21;4(43):1686.

199. Crooke SN, Ovsyannikova IG, Kennedy RB, Poland GA. Immunoinformatic identification of B cell and T cell epitopes in the SARS-CoV-2 proteome. Sci Rep. 2020 Aug 25;10(1):14179.

200. Past Seasons Estimated Influenza Disease Burden | CDC [Internet]. 2020 [cited 2021 Oct 1]. Available from: https://www.cdc.gov/flu/about/burden/past-seasons.html

201. CDC. Key Facts About Seasonal Flu Vaccine [Internet]. Centers for Disease Control and Prevention. 2021 [cited 2021 Oct 4]. Available from: https://www.cdc.gov/flu/prevent/keyfacts.htm

202. Paules CI, Marston HD, Eisinger RW, Baltimore D, Fauci AS. The Pathway to a Universal Influenza Vaccine. Immunity. 2017 Oct 17;47(4):599–603.

203. Memoli MJ, Han A, Walters K-A, Czajkowski L, Reed S, Athota R, et al. Influenza A Reinfection in Sequential Human Challenge: Implications for Protective Immunity and "Universal" Vaccine Development. Clin Infect Dis Off Publ Infect Dis Soc Am. 2020 Feb 14;70(5):748–53.

204. CDC. How Flu Viruses Can Change [Internet]. Centers for Disease Control and Prevention. 2021 [cited 2021 Oct 1]. Available from: https://www.cdc.gov/flu/about/viruses/change.htm

205. Hensley SE, Das SR, Bailey AL, Schmidt LM, Hickman HD, Jayaraman A, et al. Hemagglutinin Receptor Binding Avidity Drives Influenza A Virus Antigenic Drift. Science. 2009 Oct 30;326(5953):734–6.

206. Wang Q. Influenza antigenic drift: what is the driving force? Cellscience. 2010 Jan;6(3):1–8.

207. O'Donnell CD, Vogel L, Wright A, Das SR, Wrammert J, Li G-M, et al. Antibody pressure by a human monoclonal antibody targeting the 2009 pandemic H1N1 virus hemagglutinin drives the emergence of a virus with increased virulence in mice. mBio. 2012;3(3):e00120-12.

208. Wu NC, Wilson IA. A Perspective on the Structural and Functional Constraints for Immune Evasion: Insights from Influenza Virus. J Mol Biol. 2017 Aug 18;429(17):2694–709.

209. Paquette SG, Banner D, Zhao Z, Fang Y, Huang SSH, León AJ, et al. Interleukin-6 Is a Potential Biomarker for Severe Pandemic H1N1 Influenza A Infection. PLoS ONE [Internet].

2012 Jun 5 [cited 2020 Jan 27];7(6). Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3367995/

210. Rowe T, Banner D, Farooqui A, Ng DCK, Kelvin AA, Rubino S, et al. In vivo ribavirin activity against severe pandemic H1N1 influenza A/Mexico/4108/2009. J Gen Virol. 2010;91(12):2898–906.

211. Kash JC, Qi L, Dugan VG, Jagger BW, Hrabal RJ, Memoli MJ, et al. Prior infection with classical swine H1N1 influenza viruses is associated with protective immunity to the 2009 pandemic H1N1 virus. Influenza Other Respir Viruses. 2010 May;4(3):121–7.

212. Kelvin AA, Degousee N, Banner D, Stefanski E, León AJ, Angoulvant D, et al. Lack of group X secreted phospholipase A2 increases survival following pandemic H1N1 influenza infection. Virology. 2014 Apr 1;454–455:78–92.

213. Influenza Research Database - Strain A/FM/1/47-MA [Internet]. [cited 2021 Oct 1]. Available from: https://www.fludb.org/brc/fluStrainDetails.spg?strainName=A/FM/1/47-MA&decorator=influenza

214. Gamblin SJ, Skehel JJ. Influenza hemagglutinin and neuraminidase membrane glycoproteins. J Biol Chem. 2010 Sep 10;285(37):28403–9.

215. Zheng Z, Paul SS, Mo X, Yuan Y-RA, Tan Y-J. The Vestigial Esterase Domain of Haemagglutinin of H5N1 Avian Influenza A Virus: Antigenicity and Contribution to Viral Pathogenesis. Vaccines. 2018 Sep;6(3):53.

216. Kringelum JV, Nielsen M, Padkjær SB, Lund O. Structural analysis of B-cell epitopes in antibody:protein complexes. Mol Immunol. 2013 Jan 1;53(1):24–34.

217. Couch RB. An overview of serum antibody responses to influenza virus antigens. Dev Biol. 2003 Jan 1;115:25–30.

218. Hancock K, Veguilla V, Lu X, Zhong W, Butler EN, Sun H, et al. Cross-Reactive Antibody Responses to the 2009 Pandemic H1N1 Influenza Virus. N Engl J Med. 2009 Nov 12;361(20):1945–52.

219. Weinfurter JT, Brunner K, Iii SVC, Li C, Broman KW, Kawaoka Y, et al. Cross-Reactive T Cells Are Involved in Rapid Clearance of 2009 Pandemic H1N1 Influenza Virus in Nonhuman Primates. PLOS Pathog. 2011 Nov 10;7(11):e1002381.

220. Ellebedy AH, Ahmed R. Re-Engaging Cross-Reactive Memory B Cells: The Influenza Puzzle. Front Immunol [Internet]. 2012 [cited 2019 Nov 20];3. Available from: http://journal.frontiersin.org/article/10.3389/fimmu.2012.00053/abstract

221. Noda T, Kawaoka Y. Structure of influenza virus ribonucleoprotein complexes and their packaging into virions. Rev Med Virol. 2010 Nov;20(6):380–91.

222. Kerry PS, Willsher N, Fodor E. A cluster of conserved basic amino acids near the C-terminus of the PB1 subunit of the influenza virus RNA polymerase is involved in the regulation of viral transcription. Virology. 2008 Mar 30;373(1):202–10.

223. Groom JR, Luster AD. CXCR3 ligands: redundant, collaborative and antagonistic functions. Immunol Cell Biol. 2011;89(2):207–15.

224. Van Raemdonck K, Van den Steen PE, Liekens S, Van Damme J, Struyf S. CXCR3 ligands in disease and therapy. Cytokine Growth Factor Rev. 2015 Jun 1;26(3):311–27.

225. Zhang Y-L, Li Q, Yang X-M, Fang F, Li J, Wang Y-H, et al. SPON2 Promotes M1-like Macrophage Recruitment and Inhibits Hepatocellular Carcinoma Metastasis by Distinct Integrin–Rho GTPase–Hippo Pathways. Cancer Res. 2018 May 1;78(9):2305–17.

226. Allen IC, Scull MA, Moore CB, Holl EK, McElvania-TeKippe E, Taxman DJ, et al. The NLRP3 Inflammasome Mediates In Vivo Innate Immunity to Influenza A Virus through Recognition of Viral RNA. Immunity. 2009 Apr 17;30(4):556–65.

227. Caignard G, Leiva-Torres GA, Leney-Greene M, Charbonneau B, Dumaine A, Fodil-Cornu N, et al. Genome-Wide Mouse Mutagenesis Reveals CD45-Mediated T Cell Function as Critical in Protective Immunity to HSV-1. PLOS Pathog. 2013 Sep 12;9(9):e1003637.

228. Kazanietz MG, Durando M, Cooke M. CXCL13 and Its Receptor CXCR5 in Cancer: Inflammation, Immune Response, and Beyond. Front Endocrinol. 2019;10:471.

229. Matsuo K, Kitahata K, Kawabata F, Kamei M, Hara Y, Takamura S, et al. A Highly Active Form of XCL1/Lymphotactin Functions as an Effective Adjuvant to Recruit Cross-Presenting Dendritic Cells for Induction of Effector and Memory CD8+ T Cells. Front Immunol. 2018;9:2775.

230. Spolski R, Leonard WJ. Interleukin-21: a double-edged sword with therapeutic potential. Nat Rev Drug Discov. 2014 May;13(5):379–95.

231. Bergmann H, Yabas M, Short A, Miosge L, Barthel N, Teh CE, et al. B cell survival, surface BCR and BAFFR expression, CD74 metabolism, and CD8– dendritic cells require the intramembrane endopeptidase SPPL2A. J Exp Med. 2013 Jan 14;210(1):31–40.

232. Oettgen HC, Geha RS. IgE regulation and roles in asthma pathogenesis. J Allergy Clin Immunol. 2001 Mar 1;107(3):429–41.

233. Chu PG, Arber DA. CD79: A Review. Appl Immunohistochem Mol Morphol. 2001 Jun;9(2):97–106.

234. Correa I, Bix M, Liao NS, Zijlstra M, Jaenisch R, Raulet D. Most gamma delta T cells develop normally in beta 2-microglobulin-deficient mice. Proc Natl Acad Sci. 1992 Jan 15;89(2):653–7.

235. Wang D, Saga Y, Mizukami H, Sato N, Nonaka H, Fujiwara H, et al. Indoleamine-2,3-dioxygenase, an immunosuppressive enzyme that inhibits natural killer cell function, as a useful target for ovarian cancer therapy. Int J Oncol. 2012 Apr 1;40(4):929–34.

236. Pietra G, Vitale M, Moretta L, Mingari MC. How melanoma cells inactivate NK cells. OncoImmunology. 2012 Sep 1;1(6):974–5.

237. Mellor AL, Keskin DB, Johnson T, Chandler P, Munn DH. Cells Expressing Indoleamine 2,3-Dioxygenase Inhibit T Cell Responses. J Immunol. 2002 Apr 15;168(8):3771–6.

238. Godin-Ethier J, Hanafi L-A, Piccirillo CA, Lapointe R. Indoleamine 2,3-Dioxygenase Expression in Human Cancers: Clinical and Immunologic Perspectives. Clin Cancer Res. 2011 Nov 15;17(22):6985–91.

239. Latchman Y, Wood CR, Chernova T, Chaudhary D, Borde M, Chernova I, et al. PD-L2 is a second ligand for PD-1 and inhibits T cell activation. Nat Immunol. 2001 Mar;2(3):261–8.

240. Do J, Kim D, Kim S, Valentin-Torres A, Dvorina N, Jang E, et al. Treg-specific IL-27Rα deletion uncovers a key role for IL-27 in Treg function to control autoimmunity. Proc Natl Acad Sci. 2017 Sep 19;114(38):10190–5.

241. Hubbe M, Altevogt P. Heat-stable antigen/CD24 on mouse T lymphocytes: evidence for a costimulatory function. Eur J Immunol. 1994;24(3):731–7.

242. CDC. Coronavirus Disease 2019 (COVID-19) [Internet]. Centers for Disease Control and Prevention. 2020 [cited 2021 Aug 5]. Available from: https://www.cdc.gov/coronavirus/2019-ncov/variants/variant-info.html

243. Lokman SM, Rasheduzzaman Md, Salauddin A, Barua R, Tanzina AY, Rumi MH, et al. Exploring the genomic and proteomic variations of SARS-CoV-2 spike glycoprotein: A computational biology approach. Infect Genet Evol. 2020 Oct;84:104389.

244. Vigerust DJ, Shepherd VL. Virus glycosylation: role in virulence and immune interactions. Trends Microbiol. 2007 May;15(5):211–8.

245. Marth JD, Grewal PK. Mammalian glycosylation in immunity. Nat Rev Immunol. 2008 Nov;8(11):874–87.

246. Helle F, Duverlie G, Dubuisson J. The Hepatitis C Virus Glycan Shield and Evasion of the Humoral Immune Response. Viruses. 2011 Oct 14;3(10):1909–32.

247. Tate M, Job E, Deng Y-M, Gunalan V, Maurer-Stroh S, Reading P. Playing Hide and Seek: How Glycosylation of the Influenza Virus Hemagglutinin Can Modulate the Immune Response to Infection. Viruses. 2014 Mar 14;6(3):1294–316.

248. Wrapp D, Wang N, Corbett KS, Goldsmith JA, Hsieh C-L, Abiona O, et al. Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. Science. 2020 Mar 13;367(6483):1260–3.

249. Sanda M, Morrison L, Goldman R. N and O glycosylation of the SARS-CoV-2 spike protein. bioRxiv. 2020 Jul 26;2020.07.05.187344.

250. Shajahan A, Supekar NT, Gleinich AS, Azadi P. Deducing the N- and O-glycosylation profile of the spike protein of novel coronavirus SARS-CoV-2. Glycobiology. 2020 Dec 9;30(12):981–8.

251. Amrun SN, Lee CY-P, Lee B, Fong S-W, Young BE, Chee RS-L, et al. Linear B-cell epitopes in the spike and nucleocapsid proteins as markers of SARS-CoV-2 exposure and disease severity. EBioMedicine. 2020 Aug;58:102911.

252. Noorimotlagh Z, Karami C, Mirzaee SA, Kaffashian M, Mami S, Azizi M. Immune and bioinformatics identification of T cell and B cell epitopes in the protein structure of SARS-CoV-2: A systematic review. Int Immunopharmacol. 2020 Sep 1;86:106738.

253. Ismail S, Ahmad S, Azam SS. Immunoinformatics characterization of SARS-CoV-2 spike glycoprotein for prioritization of epitope based multivalent peptide vaccine. J Mol Liq. 2020 Sep 15;314:113612.

254. Grifoni A, Sidney J, Zhang Y, Scheuermann RH, Peters B, Sette A. A Sequence Homology and Bioinformatic Approach Can Predict Candidate Targets for Immune Responses to SARS-CoV-2. Cell Host Microbe. 2020 Apr 8;27(4):671-680.e2.

255. Sanjuán R, Domingo-Calap P. Mechanisms of viral mutation. Cell Mol Life Sci. 2016;73(23):4433–48.

256. Li WH, Wu CI, Luo CC. A new method for estimating synonymous and nonsynonymous rates of nucleotide substitution considering the relative likelihood of nucleotide and codon changes. Mol Biol Evol. 1985 Mar;2(2):150–74.

257. Rambaut A, Pybus OG, Nelson MI, Viboud C, Taubenberger JK, Holmes EC. The genomic and epidemiological dynamics of human influenza A virus. Nature. 2008 May 29;453(7195):615–9.

258. León AJ, Banner D, Xu L, Ran L, Peng Z, Yi K, et al. Sequencing, Annotation, and Characterization of the Influenza Ferret Infectome. J Virol. 2013 Feb 15;87(4):1957–66.

259. Padilla-Quirarte HO, Lopez-Guerrero DV, Gutierrez-Xicotencatl L, Esquivel-Guadarrama F. Protective Antibodies Against Influenza Proteins. Front Immunol. 2019;10:1677.

260. Gamblin SJ, Skehel JJ. Influenza Hemagglutinin and Neuraminidase Membrane Glycoproteins. J Biol Chem. 2010 Oct 9;285(37):28403–9.

261. Caton AJ, Brownlee GG, Yewdell JW, Gerhard W. The antigenic structure of the influenza virus A/PR/8/34 hemagglutinin (H1 subtype). Cell. 1982 Dec 1;31(2, Part 1):417–27.

262. Lubeck MD, Gerhard W. Topological mapping of antigenic sites on the influenza A/PR/8/34 virus hemagglutinin using monoclonal antibodies. Virology. 1981 Aug 1;113(1):64–72.

263. Rosa-Zamboni D de la, Vázquez-Pérez JA, Ávila-Ríos S, Carranco-Arenas AP, Ormsby CE, Cummings CA, et al. Molecular Characterization of the Predominant Influenza A(H1N1)pdm09 Virus in Mexico, December 2011–February 2012. PLOS ONE. 2012 Nov 29;7(11):e50116.

264. Ives JAL, Carr JA, Mendel DB, Tai CY, Lambkin R, Kelly L, et al. The H274Y mutation in the influenza A/H1N1 neuraminidase active site following oseltamivir phosphate treatment

leave virus severely compromised both in vitro and in vivo. Antiviral Res. 2002 Aug 1;55(2):307–17.

265. Paradis EG, Pinilla LT, Holder BP, Abed Y, Boivin G, Beauchemin CAA. Impact of the H275Y and I223V Mutations in the Neuraminidase of the 2009 Pandemic Influenza Virus In Vitro and Evaluating Experimental Reproducibility. PloS One. 2015;10(5):e0126115.

266. Slaine PD, MacRae C, Kleer M, Lamoureux E, McAlpine S, Warhuus M, et al. Adaptive Mutations in Influenza A/California/07/2009 Enhance Polymerase Activity and Infectious Virion Production. Viruses. 2018 18;10(5).

267. Ilyushina NA, Khalenkov AM, Seiler JP, Forrest HL, Bovin NV, Marjuki H, et al. Adaptation of Pandemic H1N1 Influenza Viruses in Mice. J Virol. 2010 Sep 1;84(17):8607–16.

268. Krause JC, Tsibane T, Tumpey TM, Huffman CJ, Basler CF, Crowe JE. A Broadly Neutralizing Human Monoclonal Antibody That Recognizes a Conserved, Novel Epitope on the Globular Head of the Influenza H1N1 Virus Hemagglutinin. J Virol. 2011 Oct 15;85(20):10905–8.

269. Whittle JRR, Zhang R, Khurana S, King LR, Manischewitz J, Golding H, et al. Broadly neutralizing human antibody that recognizes the receptor-binding pocket of influenza virus hemagglutinin. Proc Natl Acad Sci. 2011 Aug 23;108(34):14216–21.

270. Hong M, Lee PS, Hoffman RMB, Zhu X, Krause JC, Laursen NS, et al. Antibody Recognition of the Pandemic H1N1 Influenza Virus Hemagglutinin Receptor Binding Site. J Virol. 2013 Nov 15;87(22):12471–80.

271. Lee PS, Ohshima N, Stanfield RL, Yu W, Iba Y, Okuno Y, et al. Receptor mimicry by antibody F045–092 facilitates universal binding to the H3 subtype of influenza virus. Nat Commun. 2014 Apr 10;5(1):3614.

272. Fonville JM, Wilks SH, James SL, Fox A, Ventresca M, Aban M, et al. Antibody landscapes after influenza virus infection or vaccination. Science. 2014 Nov 21;346(6212):996–1000.

273. Research C for BE and. Influenza neuraminidase antigenicity and efficacy in vaccines. FDA [Internet]. 2021 Jul 27 [cited 2021 Oct 19]; Available from: https://www.fda.gov/vaccines-blood-biologics/biologics-research-projects/influenza-neuraminidase-antigenicity-and-efficacy-vaccines

274. Sylte MJ, Suarez DL. Influenza neuraminidase as a vaccine antigen. Curr Top Microbiol Immunol. 2009;333:227–41.

275. Ilyushina NA, Komatsu TE, Ince WL, Donaldson EF, Lee N, O'Rear JJ, et al. Influenza A virus hemagglutinin mutations associated with use of neuraminidase inhibitors correlate with decreased inhibition by anti-influenza antibodies. Virol J [Internet]. 2019 Nov 29 [cited 2020 Sep 30];16. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6884823/

276. Kirchenbaum GA, Carter DM, Ross TM. Sequential Infection in Ferrets with Antigenically Distinct Seasonal H1N1 Influenza Viruses Boosts Hemagglutinin Stalk-Specific Antibodies. García-Sastre A, editor. J Virol. 2016 Jan 15;90(2):1116–28.

277. Pan C, Cheung B, Tan S, Li C, Li L, Liu S, et al. Genomic Signature and Mutation Trend Analysis of Pandemic (H1N1) 2009 Influenza A Virus. PLoS ONE [Internet]. 2010 Mar 8 [cited 2020 Jun 17];5(3). Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2833199/

278. Binh NT, Wakai C, Kawaguchi A, Nagata K. The N-terminal region of influenza virus polymerase PB1 adjacent to the PA binding site is involved in replication but not transcription of the viral genome. Front Microbiol [Internet]. 2013 Dec 18 [cited 2020 Jun 18];4. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3866587/

279. Kerry PS, Willsher N, Fodor E. A cluster of conserved basic amino acids near the C-terminus of the PB1 subunit of the influenza virus RNA polymerase is involved in the regulation of viral transcription. Virology. 2008 Mar 30;373(1):202–10.

280. Sugiyama K, Obayashi E, Kawaguchi A, Suzuki Y, Tame JRH, Nagata K, et al. Structural insight into the essential PB1–PB2 subunit contact of the influenza virus RNA polymerase. EMBO J. 2009 Jun 17;28(12):1803–11.

281. Pizzolla A, Nguyen THO, Smith JM, Brooks AG, Kedzierska K, Heath WR, et al. Resident memory CD8+ T cells in the upper respiratory tract prevent pulmonary influenza virus infection. Sci Immunol [Internet]. 2017 Jun 2 [cited 2020 Sep 24];2(12). Available from: https://immunology.sciencemag.org/content/2/12/eaam6970

282. Yewdell JW, Bennink JR, Smith GL, Moss B. Influenza A virus nucleoprotein is a major target antigen for cross-reactive anti-influenza A virus cytotoxic T lymphocytes. Proc Natl Acad Sci. 1985 Mar 1;82(6):1785–9.

283. Andrews SF, Huang Y, Kaur K, Popova LI, Ho IY, Pauli NT, et al. Immune history profoundly affects broadly protective B cell responses to influenza. Sci Transl Med. 2015 Dec 2;7(316):316ra192-316ra192.

284. Plante JA, Liu Y, Liu J, Xia H, Johnson BA, Lokugamage KG, et al. Spike mutation D614G alters SARS-CoV-2 fitness. Nature. 2020 Oct 26;1–6.

285. Li Q, Wu J, Nie J, Zhang L, Hao H, Liu S, et al. The Impact of Mutations in SARS-CoV-2 Spike on Viral Infectivity and Antigenicity. Cell. 2020 Sep;182(5):1284-1294.e9.

286. Tchesnokova V, Kulakesara H, Larson L, Bowers V, Rechkina E, Kisiela D, et al. Acquisition of the L452R mutation in the ACE2-binding interface of Spike protein triggers recent massive expansion of SARS-Cov-2 variants. bioRxiv. 2021 Feb 22;2021.02.22.432189.

287. Grabowski F, Kochańczyk M, Lipniacki T. L18F substrain of SARS-CoV-2 VOC-202012/01 is rapidly spreading in England. medRxiv. 2021 Feb 9;2021.02.07.21251262.

288. Skelly DT, Harding AC, Gilbert-Jaramillo J, Knight ML, Longet S, Brown A, et al. Natural and vaccine-induced antibody and cellular responses against emerging SARS-CoV-2 variants of concern [Internet]. In Review; 2021 Feb [cited 2021 Mar 20]. Available from: https://www.researchsquare.com/article/rs-224655/v1

289. Emary KRW, Golubchik T, Aley PK, Ariani CV, Angus BJ, Bibi S, et al. Efficacy of ChAdOx1 nCoV-19 (AZD1222) Vaccine Against SARS-CoV-2 VOC 202012/01 (B.1.1.7) [Internet]. Rochester, NY: Social Science Research Network; 2021 Feb [cited 2021 Mar 20]. Report No.: ID 3779160. Available from: https://papers.ssrn.com/abstract=3779160

290. McCallum M, Bassi J, Marco AD, Chen A, Walls AC, Iulio JD, et al. SARS-CoV-2 immune evasion by the B.1.427/B.1.429 variant of concern. Science. 2021 Aug 6;373(6555):648–54.

291. Weekly epidemiological update on COVID-19 - 22 June 2021 [Internet]. [cited 2021 Jul 13]. Available from: https://www.who.int/publications/m/item/weekly-epidemiological-update-on-covid-19---22-june-2021

292. Planas D, Veyer D, Baidaliuk A, Staropoli I, Guivel-Benhassine F, Rajah MM, et al. Reduced sensitivity of infectious SARS-CoV-2 variant B.1.617.2 to monoclonal antibodies and sera from convalescent and vaccinated individuals. bioRxiv. 2021 May 27;2021.05.26.445838.

293. Watanabe Y, Allen JD, Wrapp D, McLellan JS, Crispin M. Site-specific glycan analysis of the SARS-CoV-2 spike. Science. 2020 Jul 17;369(6501):330–3.

294. Kim P, Jang YH, Kwon SB, Lee CM, Han G, Seong BL. Glycosylation of Hemagglutinin and Neuraminidase of Influenza A Virus as Signature for Ecological Spillover and Adaptation among Influenza Reservoirs. Viruses. 2018 Apr;10(4):183.

295. Altman MO, Angel M, Košík I, Trovão NS, Zost SJ, Gibbs JS, et al. Human Influenza A Virus Hemagglutinin Glycan Evolution Follows a Temporal Pattern to a Glycan Limit. Griffin DE, editor. mBio. 2019 Apr 2;10(2):e00204-19, /mbio/10/2/mBio.00204-19.atom.

296. Skowronski DM, Sabaiduc S, Leir S, Rose C, Zou M, Murti M, et al. Paradoxical clade- and age-specific vaccine effectiveness during the 2018/19 influenza A(H3N2) epidemic in Canada: potential imprint-regulated effect of vaccine (I-REV). Euro Surveill Bull Eur Sur Mal Transm Eur Commun Dis Bull. 2019 Nov;24(46).

297. Tarke A, Sidney J, Methot N, Zhang Y, Dan JM, Goodwin B, et al. Negligible impact of SARS-CoV-2 variants on CD4+ and CD8+ T cell reactivity in COVID-19 exposed donors and vaccinees. bioRxiv. 2021 Mar 1;2021.02.27.433180.

298. Faria NR, Mellan TA, Whittaker C, Claro IM, Candido D da S, Mishra S, et al. Genomics and epidemiology of the P.1 SARS-CoV-2 lineage in Manaus, Brazil. Science. 2021 May 21;372(6544):815–21.

299. Prakash S, Srivastava R, Coulon P-G, Dhanushkodi NR, Chentoufi AA, Tifrea DF, et al. Genome-Wide B Cell, CD4+, and CD8+ T Cell Epitopes That Are Highly Conserved between Human and Animal Coronaviruses, Identified from SARS-CoV-2 as Targets for Preemptive Pan-Coronavirus Vaccines. J Immunol Baltim Md 1950. 2021 Jun 1;206(11):2566–82.

300. Stoddard CI, Galloway J, Chu HY, Shipley MM, Sung K, Itell HL, et al. Epitope profiling reveals binding signatures of SARS-CoV-2 immune response in natural infection and cross-reactivity with endemic human CoVs. Cell Rep. 2021 May 25;35(8):109164.

301. Corey L, Beyrer C, Cohen MS, Michael NL, Bedford T, Rolland M. SARS-CoV-2 Variants in Patients with Immunosuppression. N Engl J Med. 2021 Aug 5;385(6):562–6.

302. Fukuyama H, Shinnakasu R, Kurosaki T. Influenza vaccination strategies targeting the hemagglutinin stem region. Immunol Rev. 2020 Jul;296(1):132–41.

303. Henry C, Palm A-KE, Krammer F, Wilson PC. From Original Antigenic Sin to the Universal Influenza Virus Vaccine. Trends Immunol. 2018 Jan;39(1):70–9.

304. Ho M-S, Chen W-J, Chen H-Y, Lin S-F, Wang M-C, Di J, et al. Neutralizing Antibody Response and SARS Severity. Emerg Infect Dis. 2005 Nov;11(11):1730–7.

305. Hayati M, Biller P, Colijn C. Predicting the short-term success of human influenza virus variants with machine learning. Proc Biol Sci. 2020 08;287(1924):20200319.

APPENDIX 1

The following sequence data are available on the NCBI Sequence Read Repository (SRA) under the BioProject Accession PRJNA787976:

- Paired end reads from Illumina MiSeq of influenza A virus at day three post-infection
- Single end reads Illumina NovaSeq of Mus musculus lung at day three post-infection

These files can be accessed by searching the above BioProject Accession or by the following link:

https://www.ncbi.nlm.nih.gov/bioproject?LinkName=biosample_bioproject&from_uid=23894514

The following raw data files are available on DalSpace:

- Spreadsheet containing all influenza virus protein SNPs (.csv)
- DiscoTope output for SARS-CoV-2 variant spike Swiss-Models (.txt files)
- T cell epitope data for MHC I (NetMHCpan EL 4.1) and MHC II (NetMHCII 2.3) (.txt files)