Deep mutational scanning to understand the evolution of SARS-CoV-2 spike.

Ruthiran Kugathasan

Submitted to Imperial College London
in fulfilment of the requirements for the degree of

DOCTOR OF PHILOSOPHY (PhD)

January 2023

Imperial College London
Department of Infectious Diseases

DECLARATION OF ORIGINALITY

I, Ruthiran Kugathasan, declare that the work presented in this thesis is entirely my own except where appropriately referenced.

COPYRIGHT DECLARATION

ACKNOWLEDGMENTS

I would like to thank my supervisor, Wendy Barclay for letting into the lab and giving me the opportunity to pursue academic research, and not curtailing my numerous digressive ideas. Thank you for providing a nurturing environment from which I could flourish and grow as a scientist.

Thank you to my lab mates, both past and present for their advice and creating a friendly environment to return to daily. Particular mentions goes to Jie Zhou, who facilitated my buying of an inordinate number of primers including spontaneous midnight orders, the other members of the ferreteers (Rebecca and Laury), the Monica Tensions (Ksenia and Maya) for helping support my work when I returned to the clinical world and Lee Coulson, for maintaining a never-ending supply of agar plate. Lastly, I would like to thank the air conditioner in tissue culture for helping to maintain a chill 17oC for the efficient conduction of science.

Thank you to my family for their continued support over the last three and a bit years, especially to Soph, who put up with my many weekend forays to the lab. This PhD is dedicated to you.

And finally thank you to the Wellcome Trust for funding the PhD and making this possible.

ABSTRACT

Deep mutational scanning to understand the evolution of the SARS-CoV-2 spike.

SARS-CoV-2 emerged as a zoonosis in 2019, causing a pandemic that led to significant global mortality, and devastating economic impact. SARS-CoV-2's pathogenicity was related to its novelty in an immunologically naïve population. With increasing population immunity through vaccination and/or natural infection attenuating the disease, societies have been able to return to a semblance of normality, however SARS-CoV-2 has persisted to become endemic. With endemicity SARS-CoV-2 has continued to adapt and evolve, initially to optimise transmission and latterly to escape immune responses. To predict the future evolution, a deep mutagenesis scanning platform was developed. Deep mutagenesis scanning allows the phenotypic effects of thousands of mutations to be explored in a high-throughput manner. Using whole trimeric Alpha spike displayed on mammalian cells provided a physiologically relevant model and allowed the identification of mutations that increase ACE2 binding (the receptor for SARS-CoV-2 spike) and immune escape, which subsequently appeared in the Omicron lineages. Using this novel deep mutagenesis platform, the evolutionary trajectory of the SARS-CoV-2 receptor-binding domain can be seen to be restricted by epistasis. Vaccine induced immune responses against the receptor binding-domain are found to be remarkably focused on one or two residues despite being polyclonal and these residues have been repeatedly selected for in a variety of variants. From work exploring the antigenic effects of receptor-binding domain mutations, it becomes apparent the N-terminal domain contributes significantly to the immune escape seen with Delta and BA.1. This effect of the N-terminal domain does not appear to be mediated by escape from N-terminal

domain directed antibodies, but by making the receptor-binding domain more difficult to neutralise. The plasticity of and focused immune response on the receptor-binding domain make further SARS-CoV-2 antigenic drift inevitable. Work described here suggests the most dramatic changes in antigenicity requires changes in both the N-terminal domain and receptor-binding domain.

TABLE OF CONTENTS

Page

ACKNOWLEDGMENTS vii

ABSTRACT viii

LIST OF TABLES xiii

LIST OF FIGURES xiv

1. Introduction 17

1.1 Life cycle 20
1.2 Evolution of SARS-CoV-2 spike 29
1.2.1 Evolution from a bat reservoir 30
1.2.2 Early SARS-CoV-2 spike evolution in an immunologically naïve population 31
1.2.3 The current phase of SARS-CoV-2 evolution, increasing antigenic distance from an immune population 33
1.2.4 The emergence of the Omicron lineages 35
1.3 Forecasting future evolution by predicting the phenotypic effects of mutations 36
1.3.1 Replicating methods 38
1.3.1.1 Serial passaging SARS-CoV-2 38
1.3.1.2 Chronic infections in immunocompromised hosts 40
1.3.1.3 Serial passaging of a replicating pseudotyped virus 42
1.3.2 Non-replicating methods 45
1.3.2.1 Deep mutational scanning (DMS) 45
1.3.2.2 Phage display 47
1.3.2.3 Yeast display 48
1.3.2.4 Mammalian cell display 49
1.3.2.5 Pseudotypes with spike DMS 51
1.3.3 Iterative methods of mutagenesis 52
1.3.3.1 Error-prone polymerase chain reaction (PCR) 52
1.3.3.2 Somatic hypermutation 52
1.3.4 An ideal approach? 54
1.4 Thesis aims and objectives 55

2 Developing a mammalian cell display platform for library screening 58

2.1 Introduction 58
2.2 Optimising plasmid mixtures for transfection 59
2.3 Can the platform select SARS-CoV-2 spike variants that escape from monoclonal antibodies (mAb)? 61
2.4 Mini discussion 64

2.5 Further optimisation of mammalian cell display 65
2.6 Introducing a fluorescent tag to spike for flow cytometry 66
2.7 Using ACE2 binding for selection of desired populations 68
2.8 sACE2(v2.4)-GFP 70
2.9 sACE2-Fc(IgG) 70
2.10 Sequencing of the sorted cells 72
2.11 Discussion 73
3 Creating the library 74
3.1 Introduction 74
3.2 Error prone PCR versus overlap extension PCR with degenerate primers 75
3.3 Optimising the PCR protocol 76
3.4 Next generation sequencing of the Alpha spike RBD plasmid mutagenesis library 82
3.5 Discussion 84
4 Screening for mutatations in the Alpha spike RBD that increase ACE2 binding 85
4.1 Introduction 85
4.2 Tolerance of the RBD library for ACE2 binding 85
4.3 sACE2-Fc-mScarlet titration 88
4.4 A screen for RBD mutations that increase ACE2 binding in Alpha spike 89
4.5 Discussion 97
5 Predicting monoclonal antibody escape 99
5.1 Introduction 99
5.2 Screening for escape variants from REGN 10933, REGN 10987 and LY-CoV016 101
5.3 Discussion 112
6 Predicting escape from vaccine sera 113
6.1 Introduction 113
6.2 Are SARS-CoV-2 vaccines more like the measles vaccine or the seasonal influenza vaccines? 115
6.3 Immune focusing on the RBD 120
6.4 Discussion 128
7 The importance of RBD mutations in BA.1 for immune escape from vaccine sera 130
7.1 Introduction 130
7.2 Which mutations in the BA.1 RBD contribute to vaccine escape? 131
7.3 Discussion 134

8 The role of domains outside of the RBD in immune escape 136

8.1 Introduction 136
8.2 The contribution of the Delta NTD versus the Delta RBD in immune escape 139
8.3 BA.1 domain swaps 142
8.4 The effect of NTD domain swaps on neutralisation by RBD mAbs 144
8.5 The effect of the NTD on ACE2 binding 146
8.6 Discussion 149
9 Discussion 152

9.1 Future SARS-CoV-2 spike evolution 152
9.2 Immune imprinting and SARS-CoV-2 antigenic evolution 156
9.3 Other applications for the mammalian display platform 158
9.4 Final perspectives 161
10 Materials and Methods 163

10.1 Materials 163
10.1.1 Cell lines 163
10.1.2 Antibodies 163
10.1.3 Plasmids 165
10.1.4 Primers 170
10.1.5 Buffers and media 173
10.2 Methods 174
10.2.1 Polymerase chain reaction (PCR) 174
10.2.2 Gel electrophoresis and extraction 174
10.2.3 DNA assembly 175
10.2.4 RBD library construction 175
10.2.5 Bacterial transformation 177
10.2.6 Minipreps 178
10.2.7 Maxipreps 178
10.2.8 Sanger sequencing 178
10.2.9 RNA extraction 178
10.2.10 cDNA synthesis 179
10.2.11 Next generation sequencing (NGS) 179
10.2.12 Cell culture 179
10.2.13 ACE2-Fc(IgG)-mScarlet production 180
10.2.14 Generation of a non-coding plasmid 180
10.2.15 FACS 180
10.2.16 ACE2 binding measurements 181
10.2.17 Pseudovirus generation 182
10.2.18 Pseudovirus neutralisation assays 182
10.2.19 Western blots 182
BIBLIOGRAPHY 184

LIST OF TABLES

Table Page

Table 1-1: Table of human coronaviruses… 19

Table 6-1: Participant demographics... 119

Table 10-1: Protocol for the fragment PCR used to create the deep mutagenesis libraries...176

Table 10-2: Protocol for the joining PCR used to create the deep mutagenesis.....177

LIST OF FIGURES

Figure Page

Figure 1-1: Structural representations of SARS-CoV-2 spike.....21

Figure 1-2: Structural representation of the “down” and “up” conformations adopted by the RBD of SARS-CoV-2 spike.. 22

Figure 1-3: SARS-CoV-2 life cycle.....25

Figure 1-4: Schematic of SARS-CoV-2 replication and transcription. 27

Figure 1-5: Antigenic cartography of SARS-CoV-2 variants. 35

Figure 1-6: Schematic of deep mutagenesis scanning selection. 47

Figure 2-1: Optimising transfection conditions to reduce the occurrence of multiple plasmids entering a single cell to increase the linkage between genotype and phenotype. 60

Figure 2-2: Schematic of selecting cells expressing SARS-CoV-2 spike variants that escape monoclonal antibody binding.. 62

Figure 2-3: Selecting for spike expressing cells that can escape mAb binding. 63

Figure 2-4: The ER retention signal of SARS-CoV-2 spike was replaced with a flexible linker tagged to a green fluorescent protein, mGreenLantern. 67

Figure 2-5: Increasing ACE2 binding by sACE2 through mutagenesis. 69

Figure 2-6: Tagging sACE2-Fc with mScarlet does not affect binding to SARS-CoV-2 spike. 71

Figure 3-1: Schematic of PCR protocol used to produce the mutagenesis library using degenerate codon primers and overlap extension PCR.....78

Figure 3-2: Frequency distribution of the number of RBD mutations using different PCR cycle numbers. 79

Figure 3-3: Distribution of the number of mutations per RBD in the library. 81

Figure 3-4: The frequency of mutations occurring at positions in the RBD from a sample of plasmids from the plasmid library. 82

Figure 3-5: Correlation between 2 independent NGS sequencing runs of the plasmid library. 83

Figure 4-1: Binding of sACE2-Fc-mScarlet by Alpha spike and the Alpha RBD library. 87

Figure 4-2: sACE2-Fc-mScarlet titrations... 88

Figure 4-3: FACS gating strategy for high ACE2 binding. 90

Figure 4-4: Heatmap of point mutations showing their enrichment score for ACE2 binding.. 91

Figure 4-5: HEK-293T cells expressing spike were incubated with sACE2-Fc-mScarlet overnight, and their binding measured using flow cytometry. 92

Figure 4-6: Relative ACE2 binding for the mutations Q498R and Q498H on Alpha, Wuhan+N501Y(D614G) and Wuhan(D614G) trimeric spikes. 94

Figure 4-7: RBD positions 484(yellow), 498(blue), 501(orange) are directly involved in the interaction with hACE2(red).....96

Figure 5-1: Dilutions of monoclonal antibodies used for sorts to screen for escape mutations. 102

Figure 5-2: Monoclonal antibody escape heatmaps.....104

Figure 5-3: Effect of single mutations in the Alpha RBD on mAb neutralisation. 106

Figure 5-4: Amino acid positions in the RBD of Alpha SARS-CoV-2 spike where monoclonal antibody escape variants were selected from the RBD library for Ly-CoV016, REGN 10987, REGN 10933 monoclonal antibodies. 107

Figure 5-5: Structures of mAbs binding to the SARS-CoV-2 RBD. 109

Figure 5-6: 477 is a position of escape for REGN 10933. 111

Figure 6-1: Evolving immune escape in a virus with multiple co-dominant neutralising epitopes versus a virus with a single dominant neutralising epitope. 114

Figure 6-2: Dilutions of vaccine sera used for the vaccine escape sorts with the Alpha RBD spike library. 118

Figure 6-3: Double BNT162b2 vaccine sera is heterogenous and is dominated by antibody responses against a single site. 122

Figure 6-4: Escape maps for each of the individual double dose BNT162b2 vaccine sera tested. 126

Figure 6-5: Cumulative escape histogram and heatmap from 8 double vaccinated individuals with BNT162b2. 127

Figure 7-1: Alignment of RBD differences between the major circulating SARS-CoV-2 variants. S.....130

Figure 7-2: The cumulative double dose BNT162b2 vaccine escape map filtered to the positions mutated in BA.1 relative to Alpha to show the contributions of each mutation in BA.1 to vaccine escape.....131

Figure 7-3: Single mutations in the RBD of BA.1 are responsible for the escape seen by BA.1's RBD.....132

Figure 8-1 Schematic explaining construction of domain swap chimeric spike proteins.....138

Figure 8-2: Alignment of NTD (top) and RBD (bottom) of major circulating SARS-CoV-2 variants.....139

Figure 8-3: The Delta NTD plays an important in the escape seen from BNT162b2 vaccine sera. 141

Figure 8-4: The BA.1 NTD reduces RBD neutralisation by vaccine sera.....143

Figure 8-5: Pseudovirus neutralisation assays using chimeric spikes featuring the Wuhan-RBD with different neighbouring domains against mAbs targeting the SARS-CoV-2-RBD......145

Figure 8-6: The effect of the BA.1 NTD on ACE2 binding by whole trimeric spike displayed on HEK-293T cells. 147

Figure 8-7:Differences in SARS-CoV-2 spike incorporation into pseudoviruses or spike processivity do not account for the phenotypic differences seen in the chimeric spike pseudoviruses. 148

Figure 9-1: Schematic showing the balance between ACE2 binding and immune evasion on the overall transmissibility of SARS-CoV-2. 156

CHAPTER 1

1. INTRODUCTION

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) was first identified in December 2019 as the causative agent of an unrecognised respiratory syndrome causing an increase in hospitalisations in the Hubei province of Wuhan, China[1, 2]. The respiratory disease would later go on to be renamed coronavirus infectious disease 2019 (COVID19).

Early cases of COVID19 had epidemiological ties to the fresh markets in Huanan, where a variety of live wild animals are sold[3, 4]. Subsequent molecular and phylodynamic studies point to at least two independent introductions of SARS-CoV-2 into humans, these early SARS-CoV-2 viruses were called lineage A and lineage B[3, 4]. Lineage A and lineage B differ by a pair of nucleotides allowing for the realisation there were multiple introductions of SARS-CoV-2 into humans. Lineage B had stronger and more direct epidemiological ties to the Huanan markets and accounted for most cases early in the pandemic[3, 4]. Environmental swabs detected the presence of SARS-CoV-2 RNA in the fresh market. Cages, storage sites and drainage water all tested positive for SARS-CoV-2 RNA providing evidence for a zoonotic origin[3, 5].

SARS-CoV-2 related viruses have been found in Rhinolophus bats from the Indochinese region, however no direct precursor virus has been identified[6]. Multiple SARS-CoV-2 related viruses have been found co-circulating in bat populations and recombination between these viruses is thought to be the likely explanation for the origin of SARS-CoV-

2[7, 8]. SARS-CoV-2 has a mosaic genome sharing highly similar regions with many of the related SARS-CoV-2 like viruses including BANAL-52 (the most closely related to SARS-CoV-2), RaTG13, RmYN02 and RpYN06[6]. The region of the SARS-CoV-2 genome that has yet to be found in SARS-CoV-2 related viruses is the furin cleavage site, however furin cleavage sites have emerged naturally in other coronaviruses[7, 9]. Recombination is also thought to be the explanation underlying the origin of SARS-CoV1, which similarly emerged from bats[10].

SARS-CoV-2 is a positive sense, enveloped virus from the *Coronaviridae* family and is the seventh coronavirus to cross the species barrier into humans[11]. The other coronaviruses to be associated with human disease in chronological order of identification are HCOV-229E, HCOV-OC43, SARS-CoV1, HCOV-NL63, HCOV-HKU-1, MERS-CoV, and SARS-CoV-2[11]. The *Coronaviridae* family is divided into four genera, *Alphacoronavirus, Betacoronavirus, Deltacoronavirus*, and *Gammacoronavirus*[12]. The alphacoronaviruses and betacoronaviruses are pathogens of mammals, while delta- and gammacoronaviruses predominantly infect birds, but some are capable of infection of mammals as well[12]. Coronaviruses causing severe disease in humans of recent are all betacoronaviruses (SARS-CoV1, MERS-CoV and SARS-CoV-2), while alphacoronaviruses include the seasonal coronaviruses that cause a predominantly mild upper respiratory tract infection (HCoV-229E and HCoV-NL63). Crossing the species barrier is a particular concern with the coronavirus family with HCoV-229E, HCoV-NL63, SARS-CoV1, MERS-CoV and SARS-CoV-2 all making the leap from bats into humans, while HCoV-OC43 and HCoV-HKU1 are thought to have

crossed from rodents into humans (Table 1-1)[11]. The major determinant to crossing species barrier is coronavirus spike binding to an appropriate receptor. SARS-CoV1, SARS-CoV-2 and HCoV-NL63 all make use of the ACE2 receptor, MERS-CoV binds to DPP-4, HCoV-229E binds to APN, while HCoV-HKU1 uses sialic acids for cell entry (Table 1-1)[13]. The sarbecovirus family, which includes SARS-CoV1 and SARS-CoV-2 appears to have diversified from an ancestral virus that was capable of binding and using bat ACE2 for cell entry[14]. Single mutations in the receptor binding motif of the sarbecovirus spike can significantly increase ACE2 binding to a variety of mammalian ACE2 providing a broad species range sarbecoviruses can cross over into with minimal mutations[14].

Coronavirus	Genera	Identification	Receptor	Zoonotic origin	Intermediate host	Furin cleavage site
HCoV-229E	Alphacoronavirus	1966	APN	Bat	Camelids?	N
HCoV-OC43	Betacoronavirus	1967	Sialic acid	Rodent	Bovine	Y
SARS-CoV	Betacoronavirus	2002	ACE2	Bat	Palm civet	N
HCoV-NL63	Alphacoronavirus	2004	ACE2	Bat	?	N
HCoV-HKU1	Betacoronavirus	2005	Sialic acid	Rodent	?	Y
MERS-CoV	Betacoronavirus	2012	DPP4	Bat	Dromedary camels	Y
SARS-CoV2	Betacoronavirus	2019	ACE2	Bat	Pangolin?	Y

Table 1-1: Table of human coronaviruses. Adapted from Z. W. Ye et al. Zoonotic origins of human coronaviruses. Int J Biol Sci 2020 Vol. 16 Issue 10 Pages 1686-1697 and Millet et al. Molecular diversity of coronavirus host cell entry receptors. FEMS Microbiology Reviews 2021 Vol. 45 Issue 3 Pages fuaa057**.**

The seasonal coronaviruses (HCoV-229E, HCoV-OC43, HCoV-NL63 and HCoV-HKU1) cause recurrent infections in humans with re-infections occurring 12 months after an infection[15]. There are two possible reasons for reinfection, waning immunity, and antigenic drift. Immunity against coronaviruses wanes with time and re-infection has been shown with the same strain of HCoV-229E following experimental infection of volunteers one year after initial infection[16]. Antigenic drift refers to the process of

genetic change that over time leads to a change in antigenicity that allows escape from recognition by prior antibody responses. Evidence for antigenic drift has been seen in HCoV-OC43 and HCoV-229E with the positive selection of mutations in spike that would be in keeping the selection of mutations that escape antibodies[17-19]. Furthermore, it has been shown that future HCOV-229E viruses are less neutralised by sera from decades prior to the emergence of the respective HCOV-229E virus providing direct evidence of antigenic drift[20]. The change in antigenicity seen with the emergence of novel variants of SARS-CoV-2 has to date occurred more frequently and to a greater magnitude than seen in seasonal coronaviruses. Re-infections with SARS-CoV-2 have necessitated the use of booster and vaccine updates[21-23].

1.1 Life cycle

SARS-CoV-2 uses its spike protein to bind to the host cell receptor ACE2[24, 25]. Spike is a class I fusion protein consisting of approximately 1280 amino acids[26]. Spike in its native state on the virion surface exists as a homotrimer. Spike can be divided into a S1 and S2 subunit. The S1 subunit interacts with the host cell receptor while S2 is responsible for membrane fusion (figure 1-1)[13].

S1 contains the NTD (N-terminal domain) and the CTD, consisting of the RBD (receptor binding domain) and SD1 (subdomain 1) and SD2 (figure 1-1). From an evolutionary perspective the ancestral virus of the coronavirus family is thought to have had only an NTD[27]. The NTD domain has galectin folds, and it is suspected the origin of the NTD may have been from a host galectin gene. Galectin proteins recognise and bind to carbohydrates. Evidence for galectin folds have been found in the CTDs of

alphacoronaviruses suggesting the CTD may have arisen by duplication on the NTD with subsequent divergence leading to the recognition of different host receptors[27]. The CTD may have been under a higher selection pressure due to being the most exposed domain of spike and the major target of antibody responses[27].

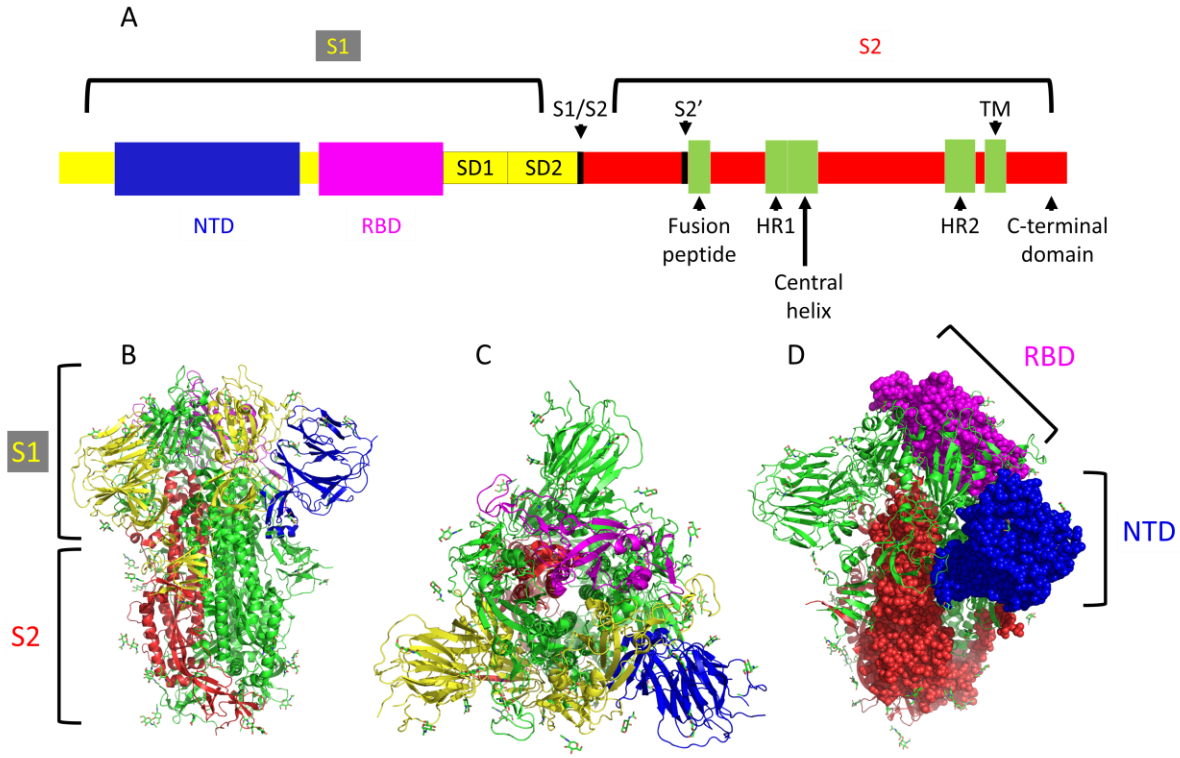


Figure 1-1: Structural representations of SARS-CoV-2 spike. A) Schematic of SARS-CoV-2 spike. B) side on view of SARS-CoV-2 spike to highlight S1 and S2 subunits in yellow and orange respectively. C) view of SARS-CoV-2 spike from above. D) positions of RBD and NTD from the same protomer are shown in magenta and blue respectively. PDB: 6ZP2. Figures created using PyMOL (The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC.)

The S1 subunit uses the RBD to directly interact with ACE2. Spike is a dynamic protein and the RBDs can exist in a "up" or open conformation or "down" or closed conformation (figure 1-2)[26, 28]. The RBD binds better to ACE2 when in the "up"

position. Mutations that increase the ability of the RBD to adopt the "up" position have been selected for in nature, the best example being D614G[29-31]. The precise role of the NTD remains uncertain, however mutations in the NTD have been shown to influence cleavage of spike by proteases and the efficiency of cell entry[32, 33].

The S2 subunit contains the fusion peptide, HR1 (heptad repeat 1), HR2 (heptad repeat 2), the transmembrane domain and a cytoplasmic C terminal tail[26, 28]. Prior to fusion, spike has to be cleaved at the S1-S2 junction and at the S2' cleavage site. SARS-CoV-2 spike contains a furin cleavage site at the S1-S2 junction, this allows cleavage of spike by endogenous furin as it is being processed in the Golgi.

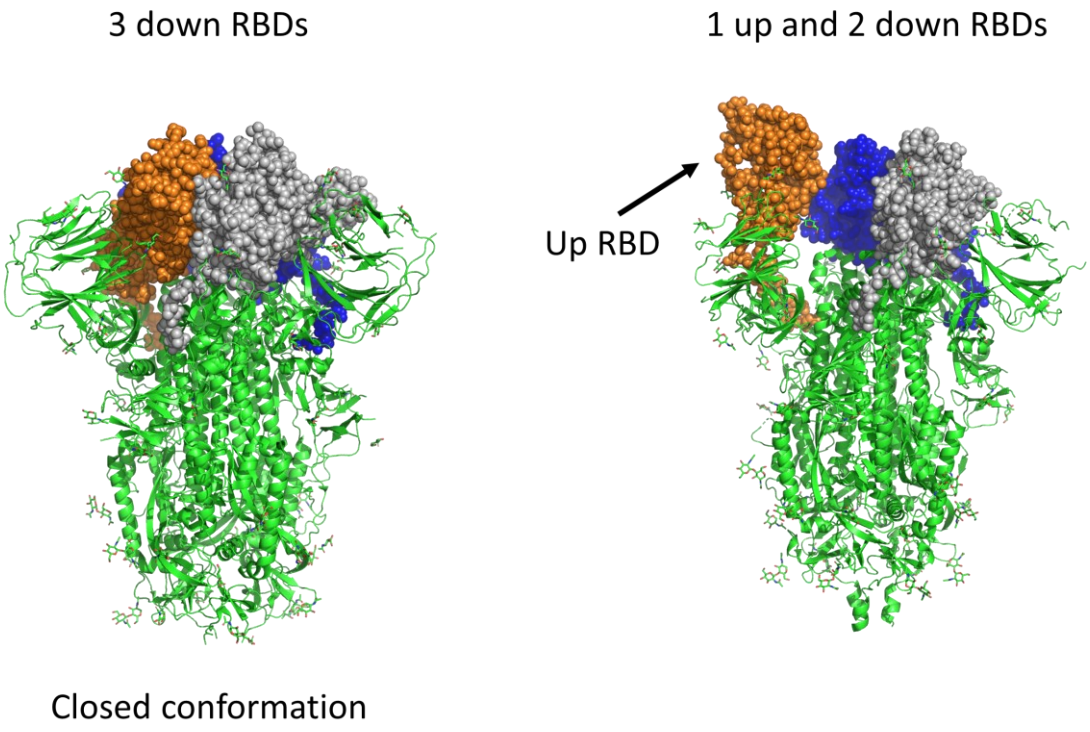


Figure 1-2: Structural representation of the "down" and "up" conformations adopted by the RBD of SARS-CoV-2 spike. PDB: 6ZP2, 7KD1. Figures created using PyMOL (The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC.)

This initial cleavage of spike before SARS-CoV-2 virions leave a host cell primes the SARS-CoV-2 spike for cell entry, as only one cleavage at S2' is required before spike can initiate membrane fusion, increasing the efficiency of this process and reducing the reliance on the endosomal pathway and possible restrictions from pH and host IFITM responses[26, 28]. Following S1-S2 cleavage spike exists in a prefusion metastable state with S1 dissociated from S2, being held only by non-covalent forces. This metastable state makes the subsequent conformational changes S2 undergoes more energetically favourable[26, 28].

ACE2 binding brings S2 into close proximity with proteases on the cell membrane (TMPRSS2) or endosomal membrane (cathepsins) for cleavage of S2, avoiding passive diffusion to find a suitable protease and cell membrane to fuse with, which would significantly reduce the probability of a virion infecting a host cell[27]. From an evolutionary perspective, coronavirus spike protein may have started being only a S2 subunit, as the S2 subunit of spike is the only part of spike necessary for the membrane fusion and cell entry[27]. The acquisition of the S1 subunit increases membrane localisation of spike. Further support for this idea came from the identification of a mouse coronavirus, neurotropic MHV (mouse hepatitis virus) that is able to infect cells independently of receptor binding[27].

Following ACE2 binding and cleavage of the S2' site, the S1 can then dissociate, and the heptad repeats rearrange to form 6 extended helical bundles with the fusion peptide now

exposed and used to insert into the host cell membrane and trigger membrane fusion[26, 28].

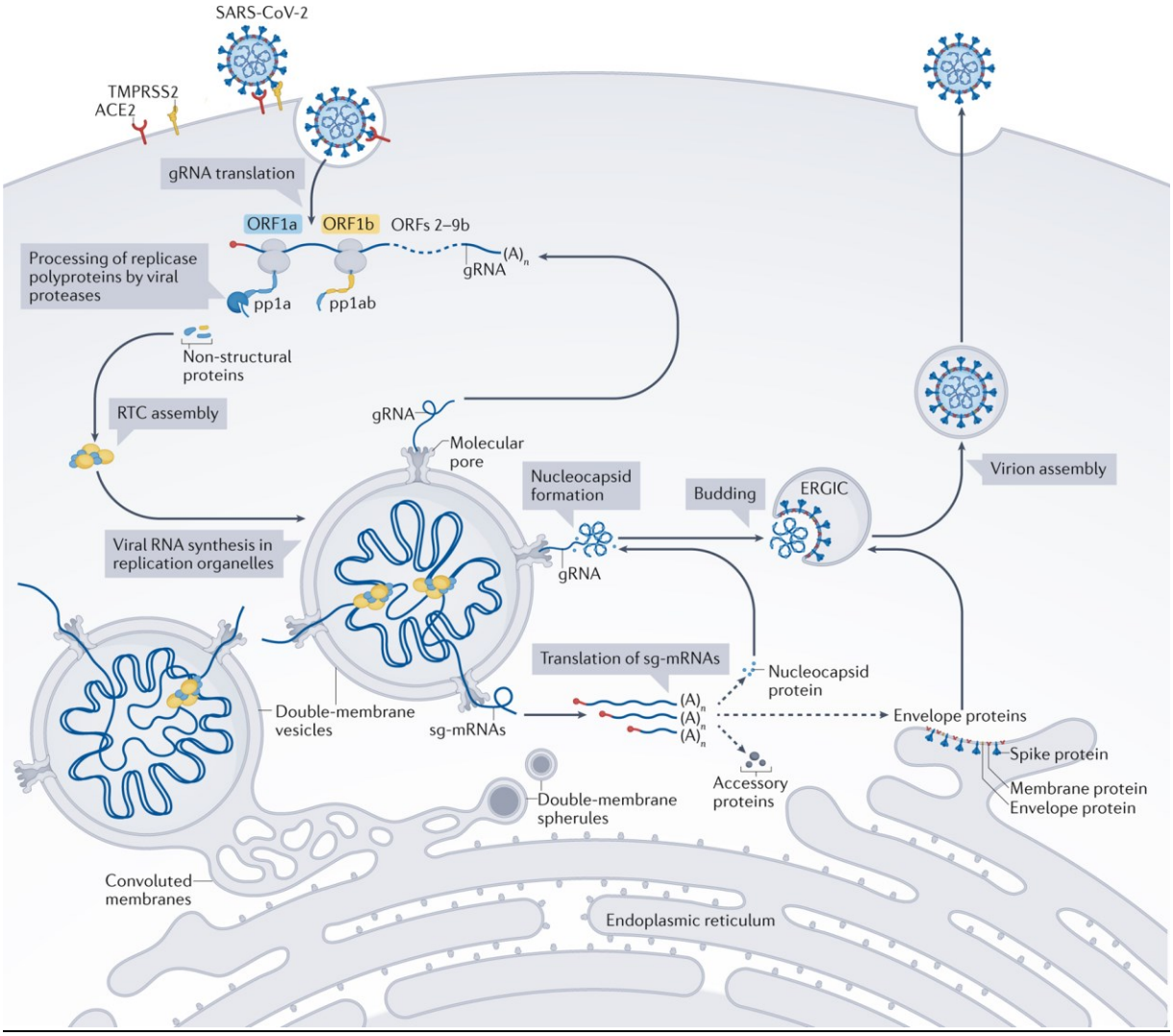


Figure 1-3: SARS-CoV-2 life cycle. Schematic of SARS-CoV-2 replication in a host cell. From Malone et al. Nature Reviews Molecular Cell Biology 2022 Vol. 23 Issue 1 Pages 21-39
131

132 Following membrane fusion either at the cell surface or the endosome, the genome enters
133 the cytoplasm. Coronavirus replication, like many other RNA viruses occurs in the
134 cytoplasm. The genome of coronaviruses is positive sense and on entry to the cytoplasm
135 makes contact with host ribosomes and is directly translated into two polypeptides, pp1a
136 and pp1b (figure 1-3)[12, 34]. The two polypeptides are translated from two reading

frames, ORF1a and ORF1b. ORF1b is produced by continuing translation beyond the stop codon of ORF1a. This is possible through a programmed ribosomal frame shift of -1 nucleotide, the new reading frame does not have a stop codon allowing translation to continue. Secondary RNA structure and a "slippery sequence" (UUUAAAC) cause the translating ribosomal to slow down and slip 40-75% of occasions, which dictates the stoichiometry of pp1a and pp1b. pp1a contains non-structural proteins (NSP)1-11, while pp1b contains NSP1-10 and NSP12-16[12, 34]. The NSPs contribute to the replication-transcription complex of coronaviruses and host cell transcription shut down (figure 1-4). The polypeptides are polycistronic and are cleaved into individual proteins by one of two proteases, NSP3 (PLpro) and NSP5 (Mpro). NSP1 is the first protein to be released from the polypeptides and leads to shut down of host transcription by blocking ribosome entry sites, while still allowing SARS-CoV-2 RNA to be translated[12, 34].

Coronavirus replication occurs using continuous and discontinuous replication. Continuous replication leads to a negative sense copy of the coronavirus genomic RNA that can be used as a template for the production of positive sense genomic RNA that can be packaged into new virions or used as a template for transcription or the production of more negative sense genomic RNA templates (figure 1-4)[12, 34].

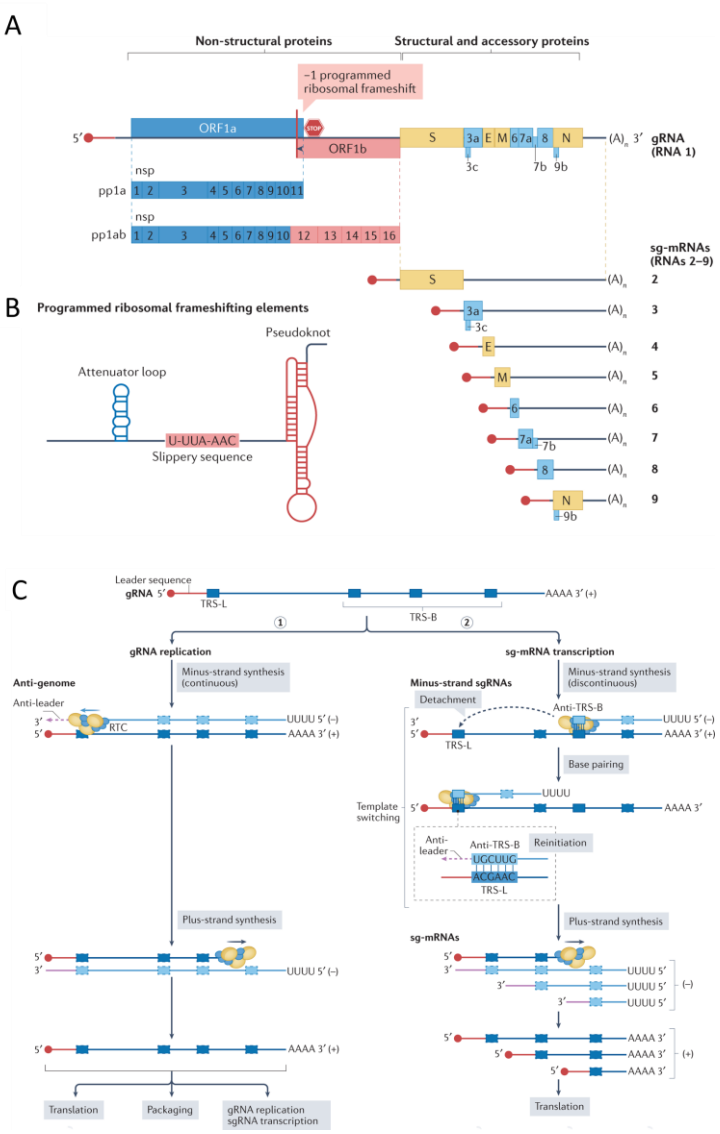


Figure 1-4: Schematic of SARS-CoV-2 replication and transcription. A) Schematic showing the organisation of the SARS-CoV-2 genome. Non-structural proteins (NSP) are located at the 5' end of the genome, while the structural and accessory proteins are located at the 3' end of the genome. B) Mechanisms dictating differential gene expression. Direct translation from genomic RNA and programmed ribosomal frameshifts are used to express the NSPs. Secondary structures in RNA upstream of the ORF1a stop codon are used to pause the ribosome on a "slippery sequence" to allow the ribosome to shift back 1 nucleotide. This occurs 40-75% of the time. C) Differential outcomes of continuous and discontinuous replication of the SARS-CoV-2 genome are used control expression of SARS-CoV-2 proteins. Continuous replication leads to the production of new SARS-CoV-2 genomes that can be packaged into virions, used directly for translation or used as template for the production of further genomes or sub-genomic RNAs. Discontinuous replication is used to produce a set of nested sub-genomic mRNAs that are used to produce the structural proteins and accessory proteins. Adapted from Malone et al. Nature Reviews Molecular Cell Biology 2022 Vol. 23 Issue 1 Pages 21-39

Discontinuous replication is a feature of coronavirus replication and contributes to the frequent recombination seen in coronaviruses. Discontinuous replication produces short subgenomic negative sense RNAs that are used as templates to produce subgenomic mRNAs. These subgenomic mRNAs are translated to produce the structural proteins (nucleocapsid, membrane, envelope and spike) and accessory proteins (figure 1-4)[12, 34].

The SARS-CoV-2 holo-RNA dependent RNA polymerase (RdRp) consists of NSP12 (RdRp) and two co-factors, NSP 7 and NSP 8. Other NSPs and host proteins in addition to the holo-RdRp contribute to the formation of the replication-transcription complex (RTC)[34]. During discontinuous replication as the RTC is transcribing from the 3' end of the genome, the RTC pauses and can switch template leaving the 3' part of the genomic template and rejoining at the 5' end of the template, restarting transcription at this point creating a recombined subgenomic RNA with a large deletion of the middle of the genome. The regions of template switching are not random and occur at transcription regulatory sequences (TRS). TRS-B (body) sequences are located upstream of each of the ORFs in the 3' end of the coronavirus genome. The RTC rejoins the 5'end of the template at the TRS-L (leader) sequence, found upstream of the first ORF of the coronavirus genome[34]. The TRS-L is complementary to the anti-TRS-B of the nascent RNA, helping to localise the RTC to the 5' TRS-L sequence. Thus, all RNA produced by the RTC has the same 5' sequence making it easier for ribosomes to identify ORFs and providing a mechanism to overcome NSP1 blockade of ribosomes[34]. Each TRS-B

serves as a decision point for the RTC, whether to continue replication to produce a negative sense antigenome or to stop and detach at one of the TRS-Bs to produce a subgenomic RNA (figure 1-4)[34].

During coronavirus replication the membranes of the endoplasmic reticulum (ER) are re-organised to form double membrane vesicles (DMV)[12, 34]. Replication and transcription occur within these DMVs, which serve to shield double stranded RNA intermediates from the host innate immune system. The nascent RNAs exit the DMVs through pores in the membrane created by NSP3 into the cytosol where they are encapsidated by nucleocapsid[34]. The encapsidated genomes can acquire a membrane with spike on the surface by budding from the ER-Golgi system before leaving the cell using the lysosomal secretory pathway (figure 1-3)[12, 34].

Coronaviruses have the largest RNA genomes of any RNA virus, at around 30kb. Like all RdRp, the coronavirus polymerase is error prone, but error catastrophe is avoided by the NSP14 gene product having proof reading activity. This led some to assume that coronaviruses have a lower evolution rate than other RNA viruses, and indeed at the start of the COVID pandemic, rapid evolution of the novel virus was thought to be unlikely.

1.2 Evolution of SARS-CoV-2 spike

The evolution of SARS-CoV-2 spike can be thought of in three phases, the zoonotic jump into humans, the early phase of evolution in an immunologically naïve population and the current phase, the continued evolution of spike in an immune population.

1.2.1 Evolution from a bat reservoir

SARS-CoV-2 evolved from bat sarbecoviruses to use human ACE2 efficiently as a cell entry receptor[6, 8, 14]. The initial barrier to species crossover is the ability to bind and enter cells of the new species[35]. The spike proteins of bat sarbecoviruses are often one- or two-point mutations away from being able to bind to other mammal ACE2 proteins, creating a reservoir for potential zoonotic spillover[6, 14]. The ability to use ACE2 as an entry receptor offers sarbecoviruses an effective route of transmission, ACE2 is found extensively in the respiratory and gastrointestinal tracts of mammals[36, 37]. SARS-CoV1 emerged in 2003 from a bat reservoir with the ability to bind to human ACE2 and cause respiratory disease[38]. SARS-CoV1 caused just over 8000 cases and spread to 29 countries and the outbreak was declared over in a few months[39]. SARS-CoV1 transmitted less efficiently than SARS-CoV-2, which contributed to their differing outcomes[40]. Some of this difference can be attributed to the presence of a furin cleavage site in the spike protein of SARS-CoV-2[41].

Prior to fusion with the cell membrane and entry, the SARS-CoV-2 spike protein must be cleaved to allow the S1 head to dissociate, freeing the S2 fusion machinery to carry out its role. The presence of a furin cleavage site allows cleavage to occur within the cell during spike biogenesis by the ubiquitous enzyme furin at the S1-S2 junction. This first cleavage is not necessary for S1-S2 dissociation, but does facilitate cleavage at the S2' site, which is required for cell membrane fusion and entry[42]. This increases the chances the second cleavage can occur at the cell surface by TMPRSS, rather than depending on endosomal entry and cathepsins[24, 43]. The importance of the furin cleavage site in

virulence and pathogenesis has been shown in hamster transmission and ferret experiments using matched reverse genetic SARS-CoV-2 viruses with and without a furin cleavage site[41, 44, 45].

1.2.2 Early SARS-CoV-2 spike evolution in an immunologically naïve population

Early SARS-CoV-2 (Wuhan) had two features relative to other bat sarbecoviruses that allowed infection and transmission within a human population: improved human ACE2 (hACE2) binding and a furin cleavage site. These features have evidence of selective pressure during the ongoing evolution of SARS-CoV-2 spike[45-47].

The early evolution of SARS-CoV-2 spike can be considered as fine-tuning the interaction with hACE2, in keeping with its recent zoonotic origin[3, 4, 8]. The first major mutation to become fixed in spike was D614G, which emerged in March 2020[48]. D614G structurally increased the proportion of spike protomers in the "up" configuration, which increased binding to ACE2[29, 49]. This single mutation increased cell entry across a range of cell types and increased transmission in animal models and epidemiologically in the global population[31, 50]. This mutation is found today in over 99% of sequenced cases worldwide.

N501Y has had less penetrance initially than D614G but has emerged recurrently in different clades. N501Y increases SARS-CoV-2 spike affinity to hACE2 by approximately ten times[51, 52], and this has been shown to be associated with increased transmissibility[53]. N501Y appeared in Alpha, which swept to almost global dominance in early 2021[54, 55], Beta, which dominated regions of Southern Africa[56] and

Gamma, which was the predominant strain in South America in early 2021[57]. Most recently, N501Y emerged in the Omicron lineage [58], which has rapidly spread around the world and continues to dominate through its descendants.

L452R is another mutation in the RBD of SARS-CoV-2 that leads to an increase in ACE2 binding[59]. L452R first arose in a short-lived cluster as part of the Epsilon VOC[60], before really rising to prominence as part of the B.1.617 lineages, which included Delta[61].

Mutations in furin cleavage site have been found in Alpha (P681H) and Delta (P681R) and these lead to increased spike cleavage[45, 46], with Delta showing greater spike cleavage than Alpha[62, 63]. Delta has been shown to have a competitive advantage in replication over Alpha in competition assays in vitro and this has been attributed to the furin cleavage mutations[63]. A host of other mutations have recurrently appeared that have been postulated to modulate furin cleavage including H655Y[64] and N679K[65]. Mutations in furin cleavage sites certainly affect cell entry, however whether increasing the furin cleavage site activity is associated with increased transmissibility is less clear cut, using RG viruses differing at P681H showed no advantaged in a hamster transmission experiment[53]. In addition, Omicron shows evidence of reduced spike cleavage[66, 67], yet is more transmissible than Delta and Alpha[58], although this may be confounded by Omicron's greater escape from neutralisation[68, 69]. A household transmission study showed a transmission advantage for Omicron against Delta in only the vaccinated members of a household, in the unvaccinated household transmission of

Omicron was similar to Delta, suggesting a major contributory reason to Omicron replacing Delta was immune evasiveness[70].

Delta had two phenotypic features that led to the replacement of Alpha, increased transmissibility[71], and antigenic distance[72-74], signifying a transition point in the evolution of SARS-CoV-2 spike. Population immunity had reached a threshold where it was now limiting transmissibility of SARS-CoV-2.

1.2.3 The current phase of SARS-CoV-2 evolution, increasing antigenic distance from an immune population

The latter phase of SARS-CoV-2 evolution has been driven by the presence of population immunity through natural infection or vaccination. The overall direction of SARS-CoV-2 evolution has been towards increased transmissibility[46]. In the presence of immunity, increasing transmission requires immune evasion, whilst maintaining inherent transmissibility. Neutralising antibodies against SARS-CoV-2 spike are the main barrier to transmission[75]. Spike is the major surface protein of the SARS-CoV-2 virion and the NTD and RBD are co-immunodominant in terms of antibody responses[76]. Despite this, antibodies against the RBD account for around 90% of neutralisation from convalescent and vaccine sera[77, 78]. Neutralising antibodies against the NTD are a minority population in immune sera[76]. The RBD is responsible for ACE2 binding[79], a step required for infection of a host cell. Antibodies directed against the RBD are the most potent at neutralisation[75, 80].

Antigenically distant variants feature mutations in both the RBD and NTD[47]. Mutations in the RBD have appeared recurrently in antigenically distant variants, an example of this is position 484. E484K was present in Beta[56], Gamma[57], and appeared in descendants of Alpha[81], while E484Q emerged in Kappa[61], and E484A most recently in the Omicron lineages[58, 82]. Deep mutational scanning identified this mutation as being an important escape mutation from convalescent sera[77].

Deletions have appeared recurrently in the NTD of SARS-CoV-2 spike and affect a region that is the major target for neutralising NTD antibodies, the antigenic supersite[83, 84]. Mutations affecting this site leads to escape from neutralising NTD monoclonal antibodies (mAb)[84, 85]. NTD deletions exist in the spike proteins of Beta[56], Delta[61] and Omicron[58].

Antigenic cartography utilises neutralisation assay data from a range of different immune sera against different virus variants to mathematically map the position variants occupy in 2 dimensions on a map of antigenicity (figure 1-5). This has been used most extensively in seasonal influenza and data from antigenic cartography is used in decisions on strain selection for the seasonal influenza vaccine by the WHO[86]. Novel SARS-CoV-2 variants continue to emerge and diversify, to be successful they must explore different areas of antigenic space to evade host antibody responses[87, 88]. Early in the evolution of SARS-CoV-2, variants were clustered close together showing relatively small incremental changes in antigenicity[87]. The emergence of BA.1 changed that, BA.1

clustered away from all other variants on an antigenic map and at the time was the most antigenically distant variant known[87] (figure 1-5).

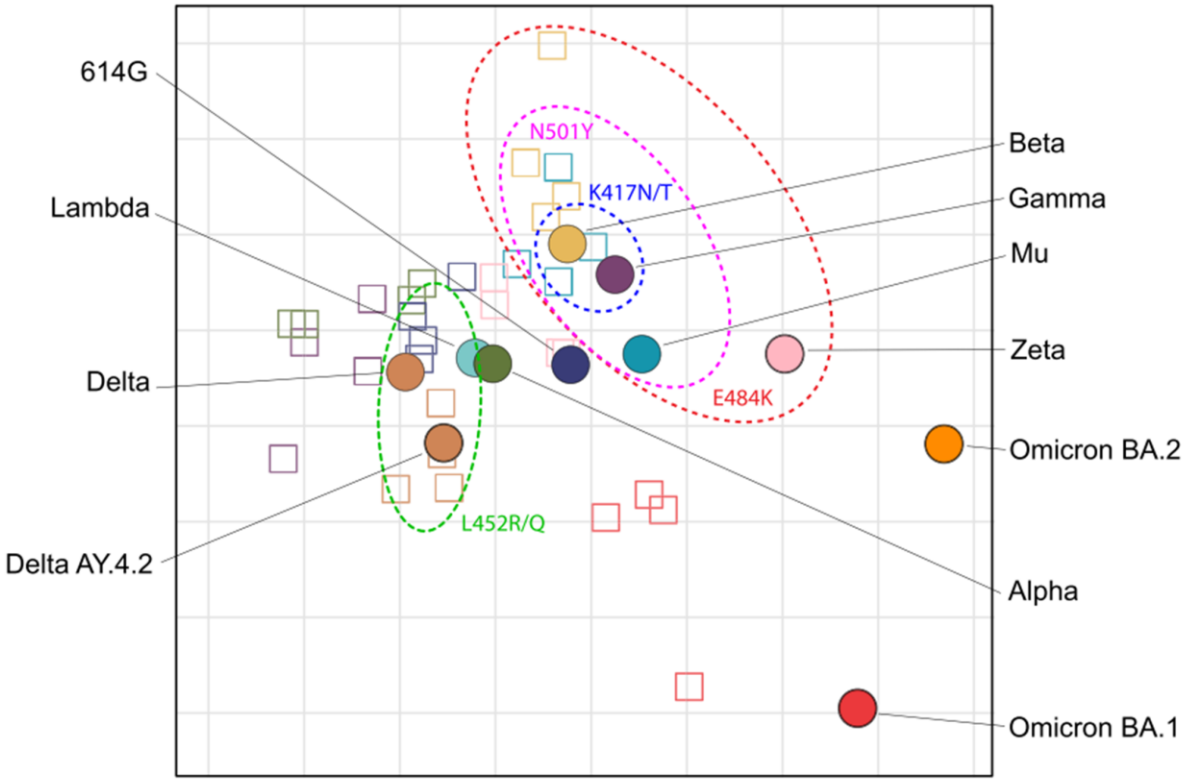


Figure 1-5: Antigenic cartography of SARS-CoV-2 variants. Dotted lines indicate variants with overlapping mutations. Adapted from A. Z. Mykytyn et al. Science Immunology Vol. 7 Issue 75. 10.1126/sciimmunol.abq4450

1.2.4 The emergence of the Omicron lineages

The end of 2021 saw the arrival of the Omicron lineages, the most transmissible and antigenically distant variants to date[58, 68, 69]. BA.1 was the first of the Omicron clade to be identified and was phylogenetically distant to any other known variant[58]. The BA.1 spike had over 30 mutations in spike compared to the ancestral Wuhan spike. To account for this incredible divergence and lack of intermediates, it is thought that Omicron evolved during a chronic infection in an immunocompromised host[89]. Studies

following infections in the immunocompromised have identified the evolution of many similar mutations [89-92].

Omicron differs phenotypically from the variants that have preceded. Early studies suggested the BA.1 spike was less cleaved by furin, showed less fusogenicity and was less pathogenic than previous variants[66, 67]. Structural studies suggest the BA.1 spike stabilises the 3 RBD "down" position of spike in a move away from the previous pattern of evolution that favoured the "up" position, however this maybe be compensated for by an increased propensity to move into the "up" position[93]. Omicron showed a different mechanism of cell entry, favouring the endosomal pathway over the fusion at the cell surface, in contrast to Delta, which is markedly biased towards utilising fusion at the cell surface for entry[62, 66, 67]. The unexpected phenotype of Omicron challenged the existing beliefs on which phenotypes were thought to be advantageous to SARS-CoV-2 transmission.

The success of the Omicron lineage is marked by the continued diversification, expansion, and dominance of its descendants globally[82]. Understanding the phenotypes associated with SARS-CoV-2 transmission, immune evasion and pathogenicity will improve the ability of surveillance to identify variants of concern. Being able to predict the genotypes that create these phenotypes will further improve the sensitivity of surveillance.

1.3 Forecasting future evolution by predicting the phenotypic effects of mutations

The evolution of SARS-CoV-2 spike has focused on increasing ACE2 binding and immune evasion[46]. Assuming that is still the case, then identifying mutations that

confer this phenotype might allow one to forecast the direction of evolution by predicting evolutionary pathways SARS-CoV-2 spike could take to increase ACE2 binding and immune evasiveness. Methods to predict the phenotype of mutations can be arbitrarily divided into those that involve replication and those that do not. Replicating methods will typically involve the serial passaging of the replicating unit in conditions designed to select for the desired phenotype, for instance in the presence of a monoclonal antibody (mAb) to select for escape mutations. This has been done using live SARS-CoV-2 virus[94] and a replicating pseudotyped VSV[95-97]. In replicating methods, the mechanism of mutagenesis is the polymerase of the replicating unit making errors. Reliance on polymerase error limits the scope of mutations that can be explored as single nucleotide mutations are favoured, however this is more reflective of nature.

Non-replicating methods to predict phenotype can be sub-categorised based on the mechanism deployed for mutagenesis, since there is no replicating polymerase to drive mutation. The 2 methods of mutagenesis used commonly are deep mutagenesis scanning and using an error-prone polymerase. These mutagenesis methods are then combined with other experimental platforms to permit selection of the desired phenotype. The platforms used most frequently are a form of display platform (phage, yeast, chicken B cells or mammalian cells) and most recently pseudovirus. Traditionally, more hypothesis driven mutagenesis approaches were used based on heuristics related to structural predictions and amino acid properties, although successful these approaches were biased by existing understanding, typically missing mutations outside of the active site or binding motif exerting long range effects. The increasing availability of deep mutagenesis

scanning, next generation sequencing (NGS) and bioinformatics has shifted scientific approaches towards non-biased mutagenic exploration of the phenotypic landscape.

1.3.1 Replicating methods

1.3.1.1 Serial passaging SARS-CoV-2

The serial passaging of live virus under selection pressure to direct evolution is technically the simplest approach and involves growing the virus under specific conditions until the desired phenotype is seen, then sequencing to identify which mutations are responsible. This has been done historically to identify mutations conferring drug resistance and escape from antibodies, monoclonal[98] and polyclonal[99]. Passaging live virus will require the use of appropriate biological containment procedures usually Cat 2 or 3 depending on the virus, which may not be available to all labs. The product will be a virus with a new phenotype, which may have implications regarding gain of function experiments and the recent controversy surrounding bat coronavirus experiments[100].

In one such experiment, SARS-CoV-2 was passaged in the presence of convalescent sera from a single person on Vero E6 cells. After 13 passages the virus completely escaped neutralisation by the sera. The escape variant had a single mutation in the RBD E484K, a deletion at F140 in the NTD and a 11 amino acid insertion in the NTD between positions 248 and 249[94]. The mechanism of escape from convalescent sera using directed

390 evolution with live virus involved changes in both the RBD and NTD, which are
391 recurrent themes in the evolution of SARS-CoV-2 variants[47, 101].
392
393 Using live virus provides the closest physiological model to evolution in nature when
394 combined with the correct selection parameters. The advantages compared to other
395 directed evolution approaches are that whole trimeric spike is used, and this is presented
396 in a natural conformation and density on the viral envelope. Fitness effects, both positive
397 and negative on spike are accurately predicted, for instance certain mutations that
398 destabilise spike and are deleterious in live virus, actually lead to increased cell entry in
399 spike pseudotyped lentiviruses[102]. Additionally, the effects of mutations outside of
400 spike can be explored when using live virus. Mutations in nucleocapsid have been shown
401 to be associated with a replication advantage in SARS-CoV-2, and these sites have been
402 selected for in current circulating variants[103, 104]. Furthermore, the mutations selected
403 for using live virus are achievable and physiologically relevant in nature, as the evolution
404 is being driven by the same polymerase acting on its native template.
405
406 As highlighted earlier, this serial passaging approach requires Cat 3 containment
407 laboratories and has the potential to evolve a variant that is more transmissible and more
408 pathogenic. Additionally, it could be argued that these experiments are occurring in
409 chronically infected immunocompromised hosts and the same information could be
410 obtained by serial collection and sequencing of virus from these individuals.

1.3.1.2 Chronic infections in immunocompromised hosts

Viral clearance can be delayed in immunocompromised hosts leading to prolonged infections in the presence of sub-neutralising antibody responses. The in-host evolution seen in these cases of chronic infection can provide a portent of future viral evolution, as the virus can try multiple possible evolutionary pathways in the presence of a wide bottleneck. In a series of chronically infected H3N2 influenza patients, deep sequencing at multiple time points revealed dynamic changes in the frequency of mutations, rising and falling over time[105]. Common mutations arose within the viral populations between patients and some of these mutations would become fixed in globally circulating H3N2 influenza virus years later[105].

Chronically infected SARS-CoV-2 infected people have been followed over time to identify evidence of intra-host evolution. The feature that appears most commonly are NTD deletions, deletion of 69/70 and deletions of different lengths at position 141 in the NTD[89, 90, 92, 106]. These NTD deletions have appeared in circulating strains recurrently, most prominently in Alpha and Omicron[83]. Moreover, these NTD mutations tend to occur repeatedly at certain sites or hotspots in the NTD, mirroring what has occurred in nature. Some have suggested these NTD mutational hotspots are the sites of targeting by NTD directed antibodies and we are seeing the selection of escape variants[83], however these NTD changes have also been shown to affect cell entry and the proteolytic cleavage of S1/S2[32, 66].

Point mutations in the RBD in chronically infected people have been described that have become fixed in subsequent circulating SARS-CoV-2 variants, including T478K (found in Delta[61]), E484K (found in Beta[56], Gamma[57]) and N501Y (Alpha, Beta, Gamma, Omicron)[89, 90]. Some of these mutations have also appeared in deep mutational scans looking for mutations that increase ACE2 binding and escape from polyclonal sera[52, 77, 78].

The viral evolution occurring within chronically infected people provides insights into beneficial mutations for the virus from both a replication and immune escape aspect. Chronically infected people have been postulated to be a source from which novel variants emerge[89]. When Alpha and Omicron emerged they were phylogenetically very distant from any possible ancestral strain, suggesting a long period of evolution going undetected, which is most explainable by a chronically infected host[89], although the possibilities of undetected circulation in under surveyed regions of the world or within animal reservoirs remain.

Monitoring in-host evolution in chronically infected individuals has many commonalities to an in-vitro directed evolution experiment, although with no control of the selection parameters. Similar results were observed between the two approaches, however in-host evolution has replication occurring in the respiratory tract compared to cell culture and has interaction with immune cells, albeit with reduced effector function due to the immunosuppression. In addition, mutations outside of spike can be selected that confer replication advantages and escape from host immune responses including both innate and

adaptive. However, neither approach in the way they have been used to date is able to predict the effect on transmission between hosts.

Serial sequencing of chronic SARS-CoV-2 infections offers many of the advantages of in vitro directed evolution using live virus without the experimental setup or the need for a Cat 3 laboratory, however this comes at the cost of a lack of control over the starting virus and the selection forces directing evolution. A safer and more accessible approach uses a replicating pseudotyped virus, such an approach benefits from using whole trimeric spike, allowing for control of selection forces and can be performed at Cat 2.

1.3.1.3 Serial passaging of a replicating pseudotyped virus

Pseudotyping involves replacing the envelope protein of one virus with another (usually more pathogenic) virus. Pseudotyping has most frequently been carried out using non-replicating lentiviruses and vesicular stomatitis virus, which can be replication competent. Pseudotyped viruses are used extensively to examine entry into a range of cells and for neutralisation assays, which can all be carried out at a lower BSL level than the parent virus[107].

For directed evolution experiments, a replicating pseudovirus would have to be used and this is predominantly VSV. The polymerase of VSV has a higher error rate than that of coronavirus, which allows mutations to occur more readily. A replicating VSV pseudotype with SARS-CoV-2 spike was used to identify escape mutations from monoclonal antibodies[95]. Using the information on escape mutations from the VSV pseudotypes with SARS-CoV-2 spikes, two mAbs were chosen to form the REGN-CoV

monoclonal antibody cocktail. The two mAbs chosen had different escape mutations meaning a single mutation could not escape both of the mAbs, creating a higher genetic barrier for resistance to develop[95].

Chimeric VSV/SARS-CoV-2 spike virus was passaged with 27 individual convalescent sera for 6 passages and subsequent plaques, purified and sequenced[96]. Mutations in spike were found in multiple regions, highlighting the polyclonality of convalescent responses[96]. The majority of selected mutations were found in the NTD and the RBD, within these two domains putative escape mutations were focused on specific regions. In the NTD these were concentrated in loops 142-144 and 243-247, which form the antigenic supersite[108] and in the RBD centred on 446, 484 and 493[96]. Mutations at these sites have been occurring naturally in variants[47, 101].

In nature, SARS-CoV-2 variants often feature deletions in the NTD affecting the antigenic supersite[83], in contrast to the point mutations found in chimeric VSV/SARS-CoV-2 spike virus, suggesting deletions and insertions are a property of the SARS-CoV-2 polymerase. The deletions and insertions that are specific to coronavirus polymerases cannot be modelled using a recombinant VSV/SARS-CoV-2 spike system, however regions of antigenic importance can still be identified.

The above experimental techniques all involve the replication of SARS-CoV-2 virus or a pseudotyped virus to predict the direction of evolution. This relies on the mutation rate of the associated polymerase and because mutations are being generated by the polymerase,

they tend to only be a single nucleotide away limiting the depth of mutations and the phenotypic landscape that can be explored. On the one hand, if wanting to predict the most probable next mutation to be fixed, this approach is acceptable. However, exploring the greater mutational landscape beyond a single nucleotide allows inferences to be made on plasticity, regions of structural and functional importance, antibody binding epitopes and a further look into the direction of future evolution. In directed evolution studies to explore beyond single nucleotide changes deep mutational scanning (DMS) is the most frequently used technique. An example of this comes from the exploration of escape mutations from the REGN-CoV mAb cocktail, a combination of two mAbs, REGN 10987 and REGN 10933 that were chosen for their different escape mutations. Using a pseudotyped VSV, no single mutation was identified that could escape from both mAbs[95], however using DMS, the E406W mutation in SARS-CoV-2 spike was identified that could escape from both mAbs. The E406W mutation is not reachable by a single nucleotide change reducing the probability of it occurring, explaining also why it did not appear in the replicating VSV-SARS-CoV-2 spike pseudotype escape screens[109].

1.3.2 Non-replicating methods

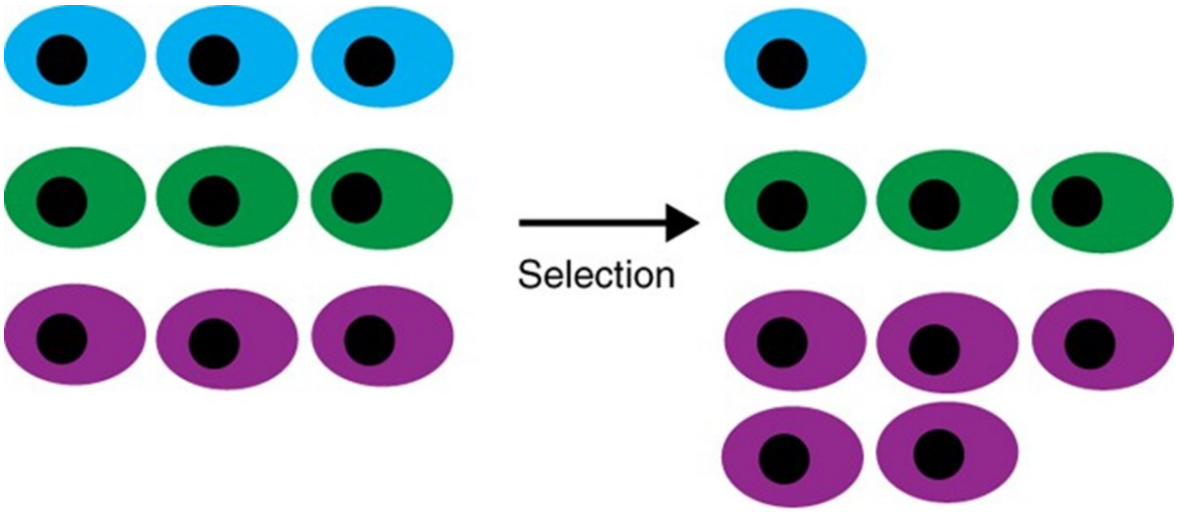
1.3.2.1 Deep mutational scanning (DMS)

The phenotype of a protein is the result of its physicochemical interactions, which are a product of the biochemical properties of its constituent amino acids, while it is possible to make biased predictions on the effect of some amino acids if located near functional domains such as an active site, mutations in other sites can have unexpected and dramatic effects on function and stability. In traditional mutagenesis approaches, it is only feasible to explore a handful of mutations in a targeted approach given the vast number of possibilities. Deep mutational scanning allows the exploration of the functional consequence of every possible amino acid at every site. The data generated can be used to inform on regions important for binding, stability, conserved regions, structure and many other features of a protein[110]. A deep mutational scanning experiment involves the creation of a mutagenesis library, a selection stage to select for mutants with the desired phenotype and subsequent sequencing to identify the selected mutations. The proportion of mutants in the selection can then be compared to the proportion of mutants in the mutagenesis library, mutants with a greater proportion in the selection than in the mutagenesis library will have been positively selected and will have increased function (figure 1-6). The selection stage can take many forms depending on the experimental design, a common strategy exploring proteins-protein interactions such as in antibody binding or receptor binding is to combine cell display of the protein with fluorescence activated cell sorting (FACS). The sorted cells with for instance, higher binding can be sequenced, and the proportion of reads compared to the proportion of reads in the unselected population to produce a phenotypic score for each mutant (figure 1-6). The

associated phenotypic scores can be linked to genotype to create genotype to phenotype maps[110].

DMS was used initially to produce proteins with improved functions such as higher affinity monoclonal antibodies, inhibitor proteins and enzymes with greater activity[110]. The earliest application of DMS for inferring the evolutionary relationships of proteins was for influenza nucleoprotein[111]. In virology, DMS has been most extensively used for influenza. Readily available and easy to use reverse genetics systems that are amenable to mutagenesis has allowed the use of DMS libraries with whole influenza virus. DMS libraries of influenza have been used to scan PB2[112], haemagglutinin[113-116], and nucleoprotein[111, 117]. The use of DMS libraries in whole SARS-CoV-2 is less straight forward due to the potential for producing a gain of function variant and in addition the reverse genetics systems for SARS-CoV-2 are more difficult to use due to the non-segmented and large (~30kB) nature of the genome.

For SARS-CoV-2, the spike protein is amenable to DMS, spike can easily be expressed on cell surfaces and binds to another protein, ACE2, allowing for the creation of a simple functional assay to measure phenotypic effects. DMS approaches to assess phenotypes in spike have been carried out using phage display[118], yeast display systems[52], and mammalian cell displays[119, 120].



Variant	Mutation	Counts (input)	Counts (selected)	Functional score
Blue	A60P	3	1	0.33
Green	WT	3	3	1
Purple	S36T	3	5	1.67

Figure 1-6: Schematic of deep mutagenesis scanning selection. The proportion of a mutant in the post-selection group can be divided by the proportion in the mutagenesis library to create a simple enrichment or functional score, reflecting the effect of the mutation of protein function. Adapted from Fowler et al. Deep mutational scanning: a new style of protein science. Nature Methods 2014 Vol. 11 Issue 8 Pages 801-807. DOI: 10.1038/nmeth.3027

1.3.2.2 Phage display

Phage display involves the presentation of peptides on bacteriophages that have been expressed in bacteria[121]. Bacteriophages can only display small proteins on their surface due to their size and additionally can only display simple proteins due to the absence of endoplasmic reticulum and the molecular chaperones found in eukaryotic cells, which limits the folding of complex proteins. Phage display has been used with

SARS-CoV-2 spike but involved the display of peptide fragments of spike due to the constraints described above[118]. The use of peptide fragments creates linear epitopes and abolishes the secondary, tertiary, and quaternary structures found in spike, consequently antibodies were unable to recognise the RBD peptide fragments[118]. Using phage display highlighted the inter individual differences in antibody profiles and suggested S2 was a major target of antibody responses[118], although how reflective these linear S2 epitopes are of trimeric spike is uncertain. The translatability of phage display of SARS-CoV-2 spike is questionable, given the inability to present a protein that resembles trimeric spike and in addition, the different glycosylation patterns seen in prokaryotic versus mammalian cells[122].

1.3.2.3 Yeast display

Yeast display has been used extensively to study mutations in the RBD of SARS-CoV-2 spike. Yeast cells are eukaryotic and have organelles and chaperone proteins resembling mammalian cells, overcoming the folding constraints with phage display[123]. The N glycosylation pathways of yeast cells do differ with those of mammalian cells, yeast cells predominantly use high mannose N glycosylation in contrast to the more complex glycosylation found in mammalian cells[124]. The spike of SARS-CoV-2 is highly glycosylated[125] and glycosylation is important in spike function, for example loss of the N343 glycosylation site is associated with over a 50% reduction in ACE2 binding[126] and simulations show the pattern of glycosylation e.g. oligomannose, paucimannose etc. will affect spike function[127]. Existing yeast display libraries have only presented the RBD in isolation. The RBD of SARS-CoV-2 spike forms part of the S1 subunit, which in combination with the S2 subunit form a protomer that is natively

found as a trimer in whole virus. Yeast display systems have a limitation on the size of protein they can display and can only display monomers. This will lead to an increased tolerance of mutations at the interface regions between protomers[119]. Additionally, interactions with other domains of spike do affect the function of the RBD, within whole trimeric spike the RBD is mobile, and adopts open or "up" and closed or "down" conformations[30, 128]. The "up" conformation increases ACE2 binding, the importance of which was seen with the global selection and fixation of the D614G mutation that increased the proportion of spike protomers in the "up" conformation resulting in increases in cell entry and transmission[29, 31, 50]. Monomeric RBD is unable to adopt these conformations leading to possible differences in ACE2 binding and epitope presentation. Despite these limitations, yeast display has been used almost exclusively to produce phenotypic maps of ACE2 binding and characterise mutations leading to escape from monoclonal antibodies(mAbs) and polyclonal sera[78, 109, 129, 130].

1.3.2.4 Mammalian cell display

Mammalian cell display was developed to improve the affinity of novel antibodies by experimentally mutagenizing the CDRs on single chain variant fragments (scFv)[131]. A proportion of yeast expressed proteins will misfold in mammalian cells due to differences in post-translational modifications, mammalian cell display was developed to overcome this limitation[132]. The original libraries were relatively small involving only a few sites on the scFv. HEK-293T cells were the mammalian cells chosen, these cells are highly transfectable and express high levels of proteins[132]. Using mammalian cells library sizes of 10^7 are achievable, which is comparable to that of yeast display but less than the 10^{12} possible with phage display[132].

Mammalian cell display has been used to create a modified sACE2 with a higher affinity for SARS-CoV-2 spike to act as a molecular decoy and inhibitor of entry[133]. The same group also used this platform with SARS-CoV-2 spike and DMS of the RBD to show that their modified high affinity sACE2 did not select for naturally occurring SARS-CoV-2 variants[119].

Alanine scanning of the SARS-CoV-2 spike NTD displayed on mammalian cells has been used to identify antibody binding hotspots and regions of escape from NTD directed mAbs[120]. The sites identified correlated with deletional hotspots and the antigenic supersite on the NTD[84, 108]. A major advantage of using mammalian cell display is the ability to use whole trimeric spike, which allows other domains to be scanned and maintains interdomain and interprotomeric interactions, which may alter the phenotype of the domain being investigated. The other is that glycosylation patterns are preserved, spike is heavily glycosylated, and the glycosylation of spike has an important role in its function and masking from antibodies[125].

The major limitation of using mammalian cell display for DMS is the difficulty in linking genotype to phenotype, on transfection multiple plasmids can enter a cell and express different mutants making the selection difficult. Methods to overcome this have included creating a stable cell line using lentiviral transduction and modifying the transfection process.

1.3.2.5 Pseudotypes with spike DMS

Another method that combines whole spike with DMS is using pseudotypes. To date, this has only been done using non-replicating lentiviral pseudotypes[102]. Using pseudotypes allows the effect of mutations to be explored in whole trimeric spike and interdomain and interprotomeric effects to be observed. By using pseudotypes, the effects of mutations on ACE2 binding, cell entry and antibody escape can be measured concurrently allowing fitness costs of mutations to be factored in, which is of particular importance for antibody escape mutations.

The downside of using lentiviral pseudotypes is that spike may not be presented on the lentivirus surface in the same number or density as in SARS-CoV-2, which may have avidity consequences for binding antibodies and mutations that destabilise spike may increase cell entry in a lentiviral pseudotype in contrast to SARS-CoV-2, where such a mutation would be deleterious[102].

The above-described methods of DMS (phage, yeast, mammalian cell display and using pseudotypes) only have a single round of mutagenesis. To do further rounds would involve harvesting the selected mutants, another round of mutagenesis, then re-cloning followed by another round of selection. While combinations of selected single mutants can be explored for additive or synergistic effect, due to epistasis some combinations of beneficial mutations would not be identified using single rounds of mutagenesis where most mutants contain just a single mutation. Using an iterative method of mutagenesis would allow single beneficial mutations to be selected for in the first round and then in

the second-round beneficial mutations that were both additive and/or epistatic could be identified in a stepwise manner.

Two techniques have been used to combine display techniques with iterative mutagenesis, using an error-prone polymerase chain reaction (PCR) and utilising somatic hypermutation.

1.3.3 Iterative methods of mutagenesis

1.3.3.1 Error-prone polymerase chain reaction (PCR)

An error-prone PCR uses conditions that increase the rate of error of the polymerase used in the PCR. This will be described in more detail in a later section. Error-prone PCR has been used in yeast display of monomeric SARS-CoV-2 RBD to iteratively create a higher-affinity ACE2 binding RBD that could be used to competitively block binding sites for SARS-CoV-2[134]. Following successive rounds of mutagenesis and selection, the N501Y mutation, followed by E484K and Q498R were selected. The combination of N501Y and E484K was found in Beta[56] and Gamma[57], while the combination of N501Y and Q498R is currently found in the Omicron lineages[58]. Both are important compensatory mutations that facilitate the emergence of other mutations that may be deleterious to ACE2 binding but have other important functions such as K417N in immune escape[135].

1.3.3.2 Somatic hypermutation

Somatic hypermutation is the mechanism of affinity maturation in B cells. After finding an antigen that is recognised by the B cell receptor, the antigen is endocytosed in the B

cell, broken down in the lysosome and presented on MHC II to complementary T cells. Following T cell activation in a lymph node the naïve B cell undergoes a replicative burst in a germinal centre. The activated B cell turns on an enzyme called activation-induced cytidine deaminase (AID). AID deaminates cytosine to uracil and through the host cell DNA repair machinery, the uracil is removed and can be replaced by any of the four nucleotides depending on the DNA repair system employed[136]. The activity of AID is limited to actively transcribed regions of chromatin and focused on regions of the immunoglobulin genes due to the presence of regulatory regions that recruit AID. AID activity outside of the immunoglobulin regions contributes to the development of B-cell lymphomas[136]. The result of somatic hypermutation mediated by AID are a library of B-cells with mutant immunoglobulins, some of these will bind to antigen better, others the same and some will be worse. This library of B-cells competes with each other for binding of antigen and presentation to T-cells for further activation signals, leading to multiple rounds of selection and division of B cells encoding immunoglobulin mutants with a higher affinity for the antigen[136].

Replacing the immunoglobulin of B-cells with the gene of your protein of interest allows that protein to undergo somatic hypermutation. This approach has been used in the DT-40 cell line, an avian lymphoma cell line[137]. SARS-CoV-2 RBD was cloned into the immunoglobulin locus, where it underwent somatic hypermutation and those mutants that were better able to bind ACE2 were selected. This led to the selection of the previously identified combination of N501Y and Q498R, which work epistatically together to increase ACE2 binding[138]. The advantage of making use of somatic hypermutation is

the simplicity once the gene of interest is cloned in, requiring only the maintenance of the cells to generate mutagenesis libraries. A disadvantage compared to error-prone PCR is that AID is biased to certain locations within a gene, so some regions will have a higher frequency of mutations and cytosine to thymine substitutions are favoured, however this can be overcome to a certain extent by increasing the number of cells that are initially transfected and form the start of the library. The protein of interest is expressed as a fusion protein with immunoglobulin and a PDGF anchor to promote display of the protein on the cell surface[138]. FACS can then be used to screen the libraries for cells expressing the protein with the desired phenotype.

These methods of conducting iterative rounds of mutagenesis with a display platform allow sequential effects of mutations to be explored and are robust to epistasis, however in both cases mutagenesis is largely by single nucleotide changes limiting the mutational space that can be explored compared to DMS.

1.3.4 An ideal approach?

The ideal approach would combine the advantages of the above techniques and might resemble, a DMS library of whole spike in a replicating pseudotyped VSV, thus providing the depth of mutational landscape exploration with DMS, iterative rounds of mutagenesis to explore additive and epistatic mutations following the initial selection using the VSV polymerase and the safety features of not using live SARS-CoV-2. Unfortunately, in the UK VSV is under SAPO classification and our lab does not have a licence for this virus. For this study, a mammalian cell display approach was chosen. Whole trimeric spike will be presented on mammalian cells and DMS used to explore

mutational phenotypes. This thesis serves as a proof of principle for the development of this platform, to that end the DMS is restricted to the RBD of SARS-CoV-2. The RBD of SARS-CoV-2 contains the ACE2 binding interface and mutations of the RBD have been shown to have significant effects on ACE2 binding[25, 52]. The most potent and majority of therapeutic monoclonal antibodies target the RBD. Additionally, the majority of neutralisation from convalescent and vaccine sera are due to RBD-directed antibodies[77, 78]. The NTD of SARS-CoV-2 spike has evolved through combinations of substitutions, deletions, and insertions[47, 101], which are technically more difficult to apply a DMS approach to. The NTD mutations tend to occur repeatedly in certain hotspots, in particular affecting the NTD supersite[84, 108], a biased library of frequently occurring NTD deletions could be incorporated into future designs using this mammalian cell display platform.

1.4 Thesis aims and objectives

SARS-CoV-2 has become endemic and continues to acquire mutations in spike that lead to increased transmissibility and immune escape. The continual antigenic change reduces the efficacy of vaccines, necessitating updates to the vaccines to better match circulating strains of SARS-CoV-2. Vaccines remain our best protection against severe disease from SARS-CoV-2. Developing techniques to forward predict the evolution of spike will allow vaccines updates to be sought pro-actively rather than retrospectively and allow for more informed surveillance and earlier warning signals of VOCs with significant immune escape potential. The overall aim of this thesis is to develop a platform with the potential to forward predict the evolution of the SARS-CoV-2 spike RBD. This will be achieved in the following way:

Chapter 2: To develop and validate a mammalian cell display platform to select cells expressing SARS-CoV-2 spike mutants from a library for the desired phenotype.

Chapter 3: To develop a PCR protocol for creating the RBD mutagenesis library in Alpha SARS-CoV-2 spike.

Chapter 4: Using deep mutagenesis scanning with the mammalian cell display platform, I plan to screen for RBD mutations on whole trimeric Alpha spike that increase ACE2 binding.

Chapter 5: Using deep mutagenesis scanning with the mammalian cell display platform, I plan to screen for RBD mutations on whole trimeric Alpha spike that escape from monoclonal antibodies.

Chapter 6: Using deep mutagenesis scanning with the mammalian cell display platform, I plan to screen for RBD mutations on whole trimeric Alpha spike that escape from vaccine sera and understand the nature of the vaccine antibody response.

Chapter 7: Using data on vaccine escape mutations in the Alpha RBD, I try to understand which mutations in the BA.1 RBD are responsible for the great immune escape seen from convalescent and vaccine sera.

Chapter 8: Using chimeric SARS-CoV-2 spike bearing pseudoviruses, I plan to explore the contributions of spike domains outside of the RBD in immune escape and ACE2 binding.

CHAPTER 2

2 DEVELOPING A MAMMALIAN CELL DISPLAY PLATFORM FOR LIBRARY SCREENING

2.1 Introduction

This chapter describes the construction of the mammalian cell display platform. A major limitation in the delivery of the library to mammalian cells is the efficiency of transfection, in contrast to bacterial or yeast transformation, mammalian cell transfection is so efficient that multiple copies of plasmid will enter cells, making the link between genotype and phenotype harder to discern in a library. One solution to this is to develop the library using lentiviruses and create a stable cell line using a dilution of transducing lentivirus that by probability only leads to the transduction of a cell by a single lentivirus. However, this would need to be followed by rounds of selection and subsequent sequencing to confirm the library is sufficiently represented by the stable cell line taking many weeks to months. Instead, this study sought to alter the standard transfection process in mammalian cells to reduce the number of plasmids entering cells and maintain the linkage between genotype and phenotype.

2.2 Optimising plasmid mixtures for transfection

To create a transfection protocol where the number of plasmids coding for a SARS-CoV-2 spike variant entering a single cell would be reduced, and maintain a tight coupling between genotype to phenotype, the plasmids encoding the library were diluted with a non-coding or empty plasmid. To establish the optimum ratio of coding plasmid to non-coding plasmid to use for library selection, different amounts of an equal mixture of two coding plasmids encoding different fluorescent proteins were mixed with 1500ng of a non-coding plasmid and transfected into 10^6 HEK-293T cells. The proportion of cells expressing a single fluorescent protein and those expressing both was determined using flow cytometry. The higher the total mass of coding plasmid mixture used the more cells expressing both fluorescent proteins were detected. However, below 1ng of total coding plasmid DNA the proportion of cells expressing a single fluorescent protein starts to drop off (figure 2-1). The optimum choice of coding DNA mass is a balance between reducing the number of cells that are multiply transfected, while ensuring a sufficient proportion of cells are transfected with coding plasmid. 1ng was chosen, as at this starting mass of plasmid mixture there are equal amounts of cells expressing a single protein and cells expressing both, while still maintaining a reasonable total transfection percentage of around 10% of cells (figure 2-1). Previous studies [1, 2] have aimed for more stringent coupling, resulting in only 1% of total cells being transfected, however reducing the number of coding plasmids per cell rather than aiming for one plasmid per cell offers the better balance between coupling genotype to phenotype and cells and time required to sort sufficient cells for analysis.

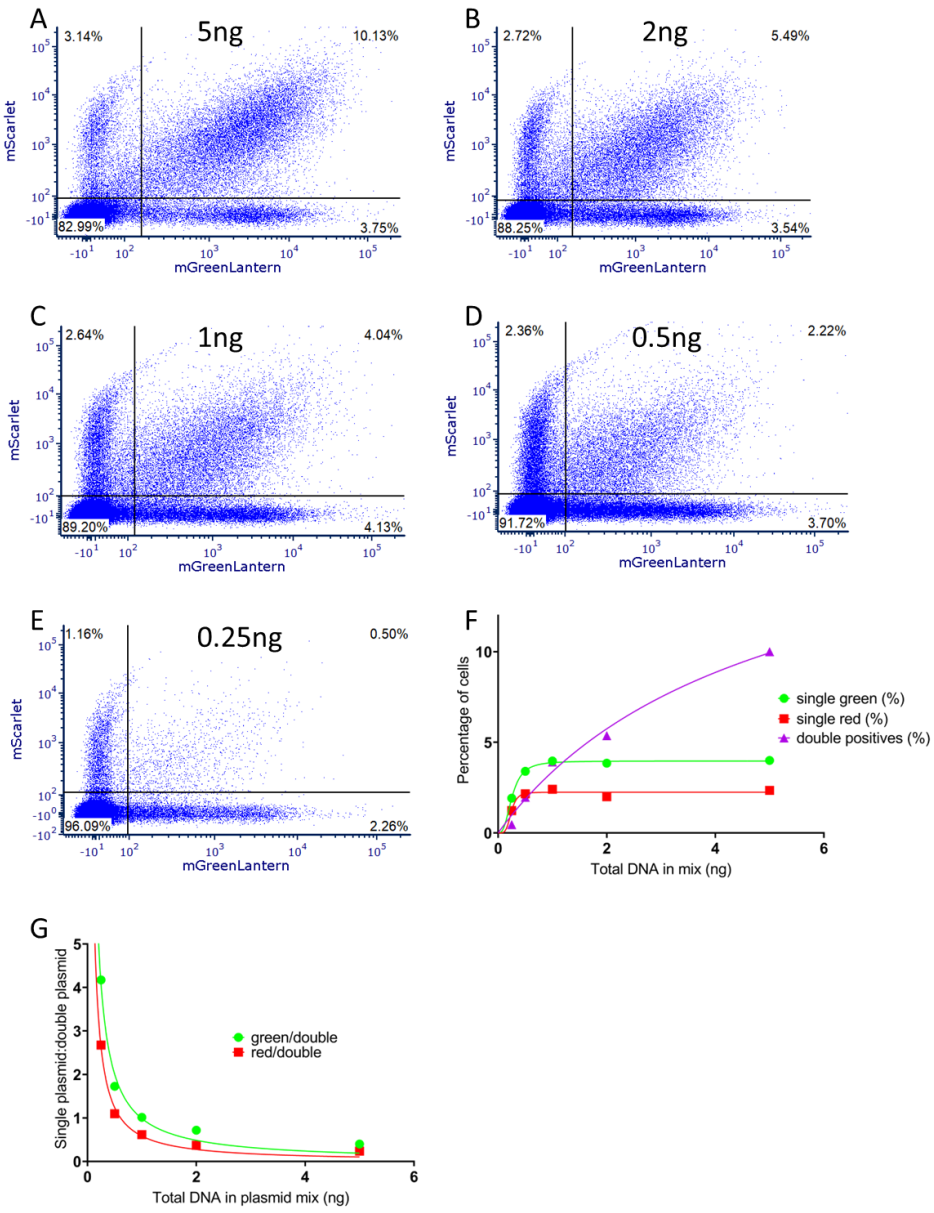


Figure 2-1: Optimising transfection conditions to reduce the occurrence of multiple plasmids entering a single cell to increase the linkage between genotype and phenotype. a) 5ng, b) 2ng, c) 1ng, d) 0.5ng, e) 0.25ng in total of an equal mixture of mScarlet and mGreenLantern plasmid were mixed with 1500ng of non-coding plasmid and transfected into 10^6 HEK-293T cells, dot plots are shown with the Y axis representing expression of mScarlet and the X axis representing mGreenLantern expression. f) The percentage of cells at each quantity of plasmid mixture expressing a single type of plasmid or being positive for both plasmids. g) The ratio of cells expressing a single type of plasmid to the number of cells expressing both types of plasmid at each plasmid mixture.

2.3 Can the platform select SARS-CoV-2 spike variants that escape from monoclonal antibodies (mAb)?

To assess whether the platform could select for antibody escape, the ability to distinguish differences in spike binding to a monoclonal antibody was investigated. A trial sort was devised using LY-CoV016 (also known as CB6), a monoclonal antibody (mAb) designed by Eli Lilly against the RBD of SARS-CoV-2 spike (figure 2-2). Gamma has the following mutations in the RBD K417T, E484K and N501Y[57]. The ancestral Wuhan-D614G spike is neutralised at picomolar concentrations in neutralisation assays by Ly-CoV016[109] and this binding can be detected using flow cytometry (figure 2-3A), whereas Gamma is not neutralised and this can be seen as reduced binding with flow cytometry (figure 2-3B) due to mutations at 417 and 484[57, 109]. Plasmids with the SARS-CoV-2 Wuhan spike in a pcDNA3.1 expression vector were kindly gifted by Paul Mackay. Site directed mutagenesis was used to produce the Wuhan-D614G and Gamma spike plasmids. A 1ng mixture of Wuhan-D614G spike and Gamma spike plasmids in the ratio 10:1 respectively was mixed with 1500ng non-coding plasmid as described above and transfected into cells. Ly-CoV016 was incubated with the HEK-293T cells expressing SARS-CoV-2 spike and stained with a secondary antibody against human IgG-Fc. If the mammalian cell display protocol works and a strong linkage between genotype and phenotype is maintained, it should be possible by sorting to enrich the population for cells expressing Gamma spike.

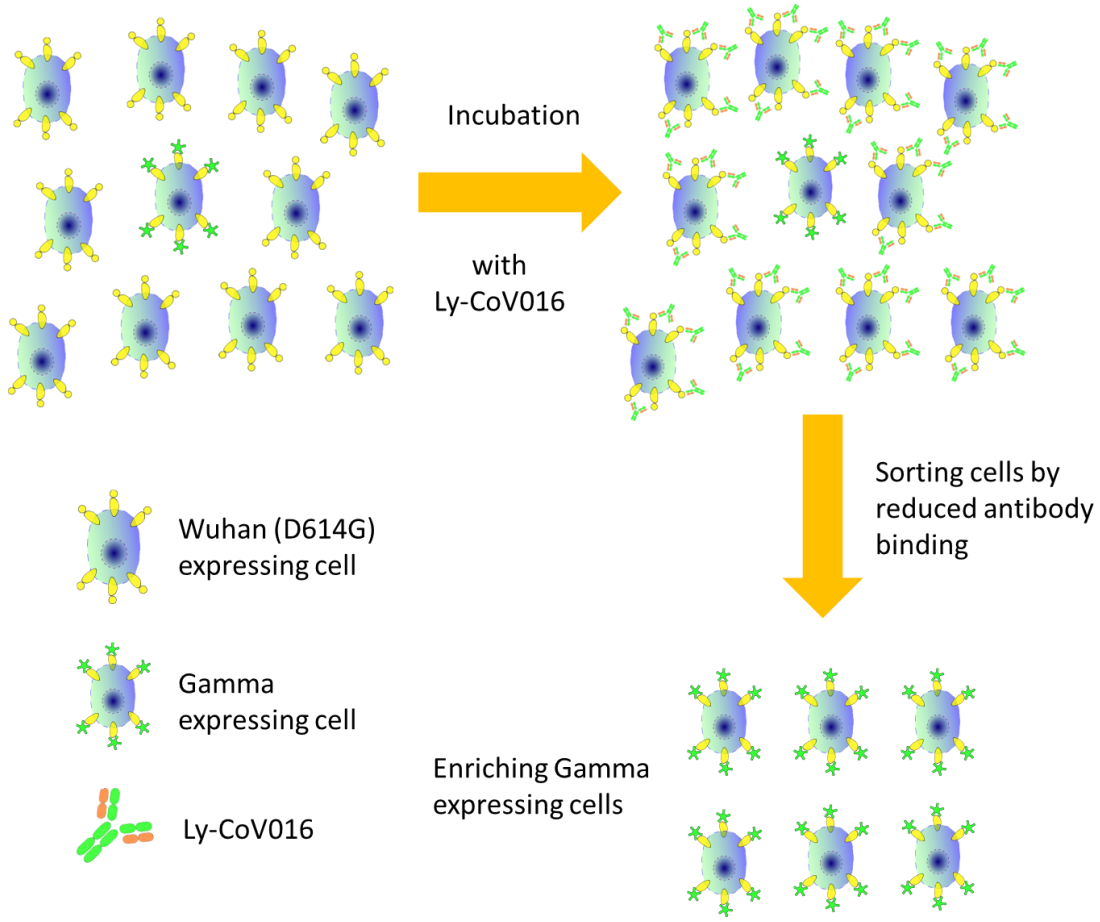
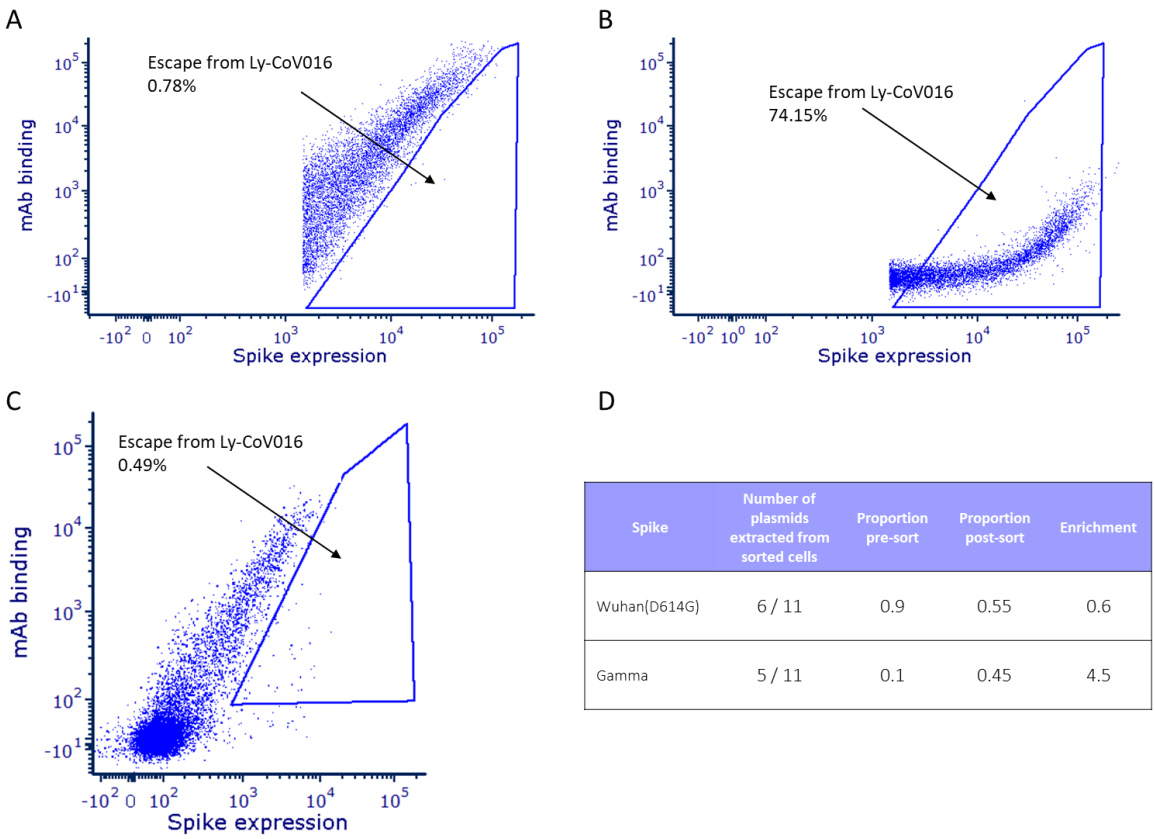


Figure 2-2: Schematic of selecting cells expressing SARS-CoV-2 spike variants that escape monoclonal antibody binding. A mixture of Wuhan (D614G) and Gamma spike expressing plasmids (in a ratio of 10:1) were transfected into HEK-293T cells. They were incubated with the mAb Ly-CoV016. The Gamma spike is a known escape variant of Ly-CoV016. Sorting the mixture of Wuhan (D614G) and Gamma spike expressing cells by the least bound by Ly-CoV016 should lead to selection of cells expressing the Gamma spike variant.

Cells with a lower level of LY-CoV016 binding were sorted by FACS and plasmids collected from the sorted cells using a commercial miniprep kit (figure 2-3D). The plasmids were used to transform bacteria, which were subsequently outgrown, miniprepped and the plasmids sequenced by Sanger sequencing. The results of the Sanger sequencing showed Gamma plasmids were enriched in the sorted populations 4.6 times, whereas the Wuhan-D614G plasmid was selected against.



Spike	Number of plasmids extracted from sorted cells	Proportion pre-sort	Proportion post-sort	Enrichment
Wuhan(D614G)	6 / 11	0.9	0.55	0.6
Gamma	5 / 11	0.1	0.45	4.5

Figure 2-3: Selecting for spike expressing cells that can escape mAb binding. The Gamma SARS-CoV-2 spike variant is a known escape variant of the mAb Ly-CoV016. The ability to sort cells expressing Gamma spike was assessed by flow cytometry by selecting for cells with reduced mAb binding. A mixture of plasmids expressing Wuhan-D614G spike and Gamma spike in a ratio of 90:10 was transfected into HEK-293T cells. 1ng of the plasmid mixture was mixed with 1500ng of a non-coding plasmid per 10^6 cells. Representative flow cytometry plots are shown of a) Wuhan-D614G spike, b) Gamma spike, c) Wuhan-D614G 90%: Gamma 10% spike expressing cells incubated with the mAb LY-CoV016. The gate in figure c was chosen to sort the lowest 5% of LY-CoV016 bound cells expressing spike corrected for levels of spike expression. The plasmids from the sorted cells were subsequently harvested and sequenced using Sanger sequencing, the number of sequences for each type of plasmid and the enrichment ratio is shown in d). Enrichment ratio = proportion in sorted population / proportion in plasmid mix.

848

849

2.4 Mini discussion

Here, using a non-coding plasmid to dilute the coding plasmid it was possible to maintain a tight genotype to phenotype link during transfection of mammalian cells and in a single round of sorting enrich greater than 4-fold for an escape variant from a mAb. Moreover, this served to validate the ability of flow cytometry to identify antibody escape variants and even other phenotypes.

The current experimental design is limited due to using reduced antibody binding to screen for escape variants from antibodies, as this will not distinguish those spike variants that have incurred a significant fitness cost and are no longer able to bind human ACE2 (hACE2). Additionally, using commercial miniprep kits to recover plasmid from mammalian cells, when such a small amount of coding plasmid is transfected is not very efficient. Both limitations will be discussed in further detail in the next sections.

2.5 Further optimisation of mammalian cell display

Having established the mammalian cell display platform can be used to screen for different spike phenotypes, this study next sought to improve upon the design. Normalising for the level of expression of spike on the cell surface is required to exclude phenotypic differences are due to differential expression of spike on the surface of cells. At the time, this was being done using a S2 antibody, which is subsequently stained with a secondary antibody. However, this is approach may become limited with ongoing evolution of the S2, requiring updated S2 antibodies if future SARS-CoV-2 spike variants are to be displayed.

Selection of antibody escape variants in the previous chapter relied on decreased antibody binding, however this does not consider escape mutations that significantly impair ACE2 binding. To correct for this, a method of measuring ACE2 binding during FACS was developed.

Lastly, as mentioned in the previous section the use of a commercial miniprep kit to extract plasmids from sorted cells was inefficient and will miss variants from a diverse library. This chapter will aim to improve on measuring spike expression, ACE2 binding and the subsequent extraction and sequencing of plasmid spike sequence data.

2.6 Introducing a fluorescent tag to spike for flow cytometry

Using mammalian cell display it was possible to select between antigenically different SARS-CoV-2 spike proteins, however further improvements to the system could be made. Using a primary antibody against S2 does not future proof the platform for the display of SARS-CoV-2 spike variants that may be sufficiently antigenically distance to escape S2 antibody binding. Additionally, secondary antibodies against a SARS-CoV-2 spike S2 mouse antibody were being used to normalise for spike expression, the secondary used was not very bright meaning cells with lower levels of spike expression were being missed. Further, large amounts of secondary would have to be used given the number of sorts planned, which would be costly and time consuming. To overcome these issues, the SARS-CoV-2 spike protein was tagged at the C-terminal end by replacing the endoplasmic reticulum (ER) retention signal with a flexible Gly-Ser linker attached to mGreenLantern[139], a bright monomeric fluorescent protein. mGreenLantern is the brightest monomeric green fluorescent protein and has improved folding and stability properties compared to enhanced green and yellow fluorescent proteins[139] (figure 2-4A). The fluorescent protein would be synthesised with the spike protein in HEK-293T cells following transfection, allowing spike expression to be normalised without further processing. Binding to hACE2 using purified recombinant hACE2 protein and a novel construct, hACE2-Fc-mScarlet (will be described later) was compared between the tagged spike protein and untagged spike protein to ensure the fluorescent protein did not adversely affect spike function. Figure 2-4B&C shows there was no difference in binding to hACE2 between the fluorescently tagged spike compared to the untagged spike proteins.

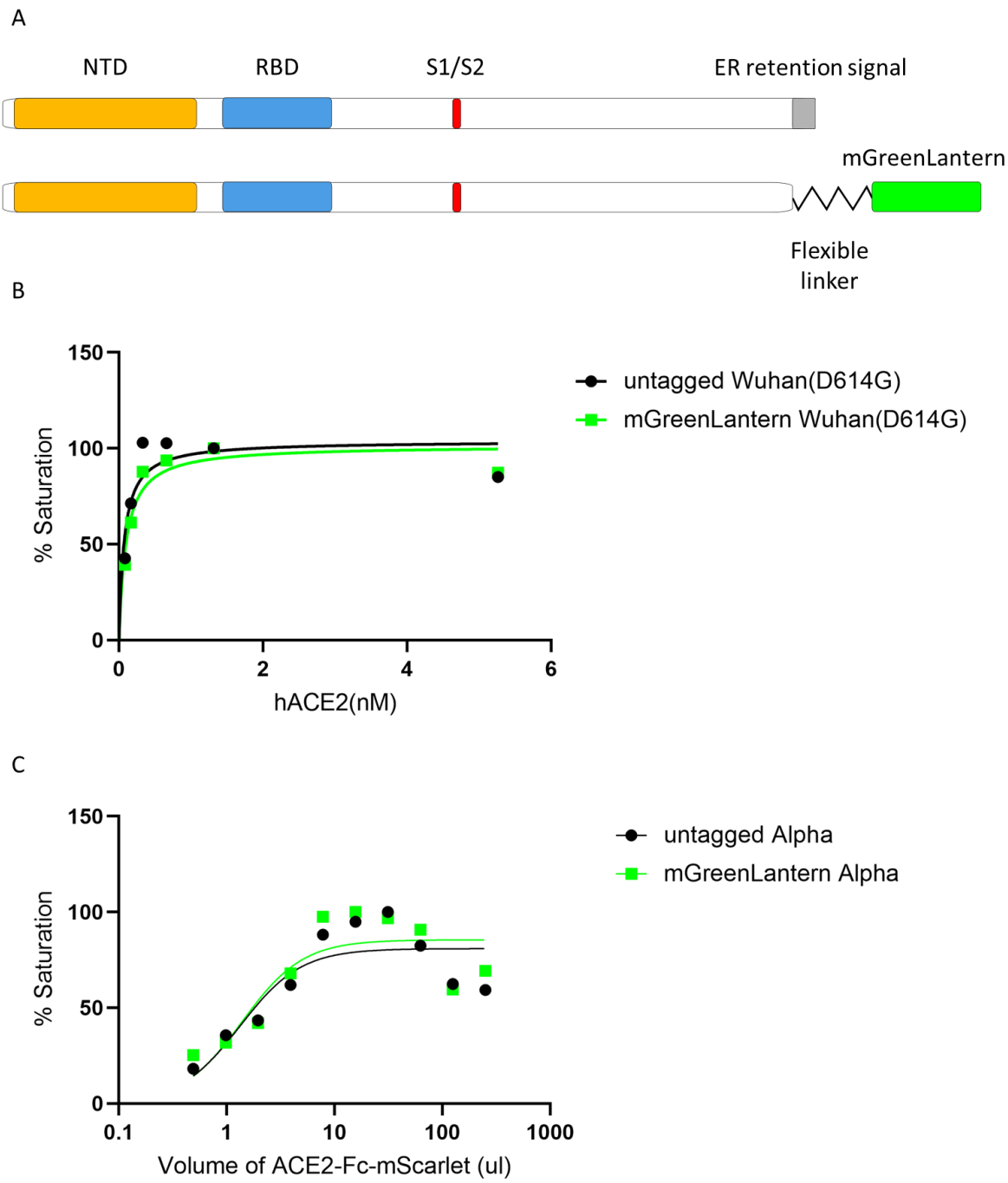


Figure 2-4: The ER retention signal of SARS-CoV-2 spike was replaced with a flexible linker tagged to a green fluorescent protein, mGreenLantern. a) Schematic of SARS-CoV-2 spike protein with mGreenLantern tag. To ensure ACE2 binding of spike was not affected by tagging with mGreenLantern, binding to ACE2 was measured using flow cytometry at a range of concentrations comparing tagged and untagged spikes expressed on 293T cells. b) Flow cytometry plot of Wuhan(D614G) spike tagged and untagged binding to purified hACE2-Fc (Abcam ab273885). c) Flow cytometry plot of Alpha spike tagged and untagged binding to supernatant containing expressed ACE2-Fc-mScarlet

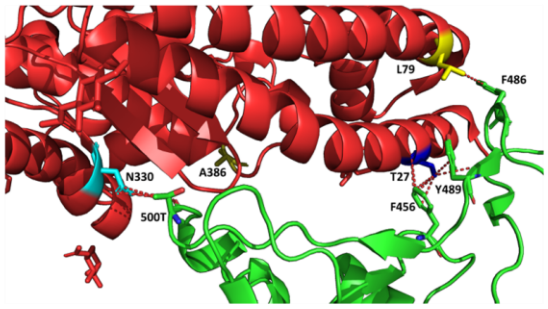
2.7 Using ACE2 binding for selection of desired populations

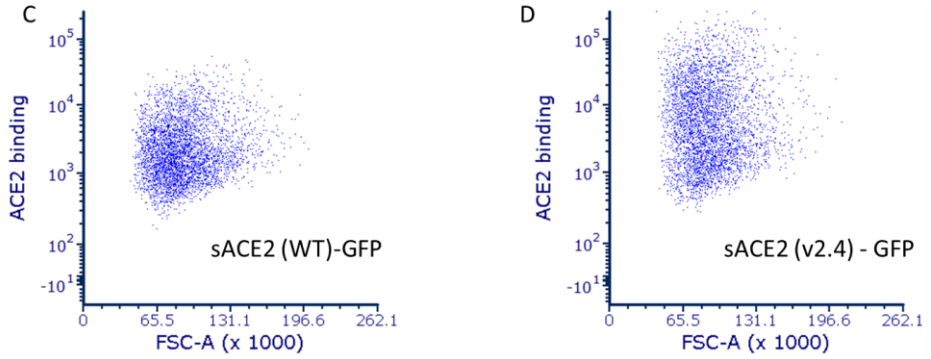
During the selection for LY-CoV016 escape (figure 2-3), Gamma spike expressing cells were selected for based on reduced mAb binding. A library will contain spike mutants that may escape from antibody binding but be unable or have severely impaired binding to hACE2. To ensure mAb escape spike mutants retain the ability to bind hACE2, escape mutants will be selected for by their ability to bind hACE2 in the presence of a neutralising mAb. HEK-293T cells expressing the spike library will be incubated with the mAb at a titration that is sub-saturating, the cells will then be incubated with hACE2, cells expressing spike mutants that escape antibody binding, and have the least cost on ACE2 binding will have a higher hACE2 signal on FACS and greater enrichment in the sorted population.

Soluble ACE2(sACE2) binds to spike poorly due to its high off rate (figure 2-5C), sACE2 can be modified to increase its affinity to spike by mutagenesis to reduce the off rate (figure 2-5) or alternatively sACE2 can be fused to the Fc of IgG, which also reduces the off rate by an avidity effect as each sACE2-Fc(IgG) has two sACE2 domains[133].

A

Position	sACE2 (wt)	sACE2 (v2.4)
27	T	Y
79	L	T
330	N	Y
386	A	L





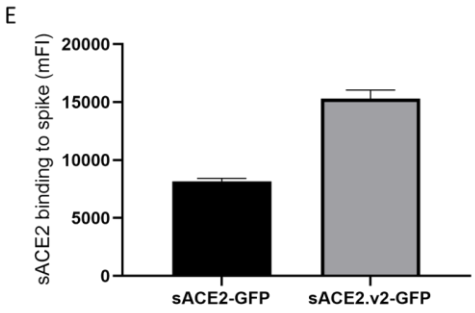


Figure 2-5: Increasing ACE2 binding by sACE2 through mutagenesis. a) sACE2.v2-GFP was created from sACE2-GFP by site directed mutagenesis, introducing the four mutations: T27Y, L79T, N330Y and A386L. b) expanded view of the mutated sites in ACE2(red) interacting with the RBD interface(green). sACE2(wt)-GFP and sACE2.v2-GFP was expressed following transfection of 293T cells, and the supernatant collected after 48 hours. 1ml of supernatant was used to incubate HEK-293T cells expressing Wuhan(D614G) spike for 1 hour before analysis by flow cytometry. Dot plots of spike binding to c) sACE2(wt)-GFP and d) sACE2.v2-GFP. e) binding of sACE2-GFP and sACE2.v2-GFP to spike expressing HEK-293T cells as determined by MFI(median fluorescence intensity)

928
929

2.8 sACE2(v2.4)-GFP

The mutations in ACE2 that increase binding to the SARS-CoV-2 RBD were identified using deep mutagenesis scanning and combinations of the mutations trialled to see the effect on ACE2 binding[133]. The mutations on sACE2 improve ACE2 binding but modifying sACE2 may lead to the selection of different spike mutations that bind better to the modified sACE2(v2.4). Using a sACE2 and modified sACE2(v2.4) mixture to incubate HEK-293T cells expressing whole Wuhan spike with an RBD library, showed rare mutations having differential binding to sACE2(wt) and sACE2(v2.4), with Y449K leading to increased binding to sACE2(v2.4)[119]. Although, a rare occurrence, using sACE2(v2.4) could bias results to mutations that bind to sACE2(v2.4) better than sACE2(wt).

2.9 sACE2-Fc(IgG)

sACE2-Fc(IgG) has the benefits of increased spike binding while maintaining a wild-type sACE2. The Fc tag of sACE2-Fc(IgG (human)) could not be used for detection of ACE2 binding, as the secondary antibody (which would have to be an anti-human IgG) would cross react with monoclonal or polyclonal antibodies that would be used in some of the selections planned to identify escape mutations. To overcome this, the sACE2-Fc(IgG) was tagged at the Fc end with a red fluorescent protein, mScarlet[140]. Tagging the fusion protein with mScarlet did not affect its binding to SARS-CoV-2 spike (figure 2-6).

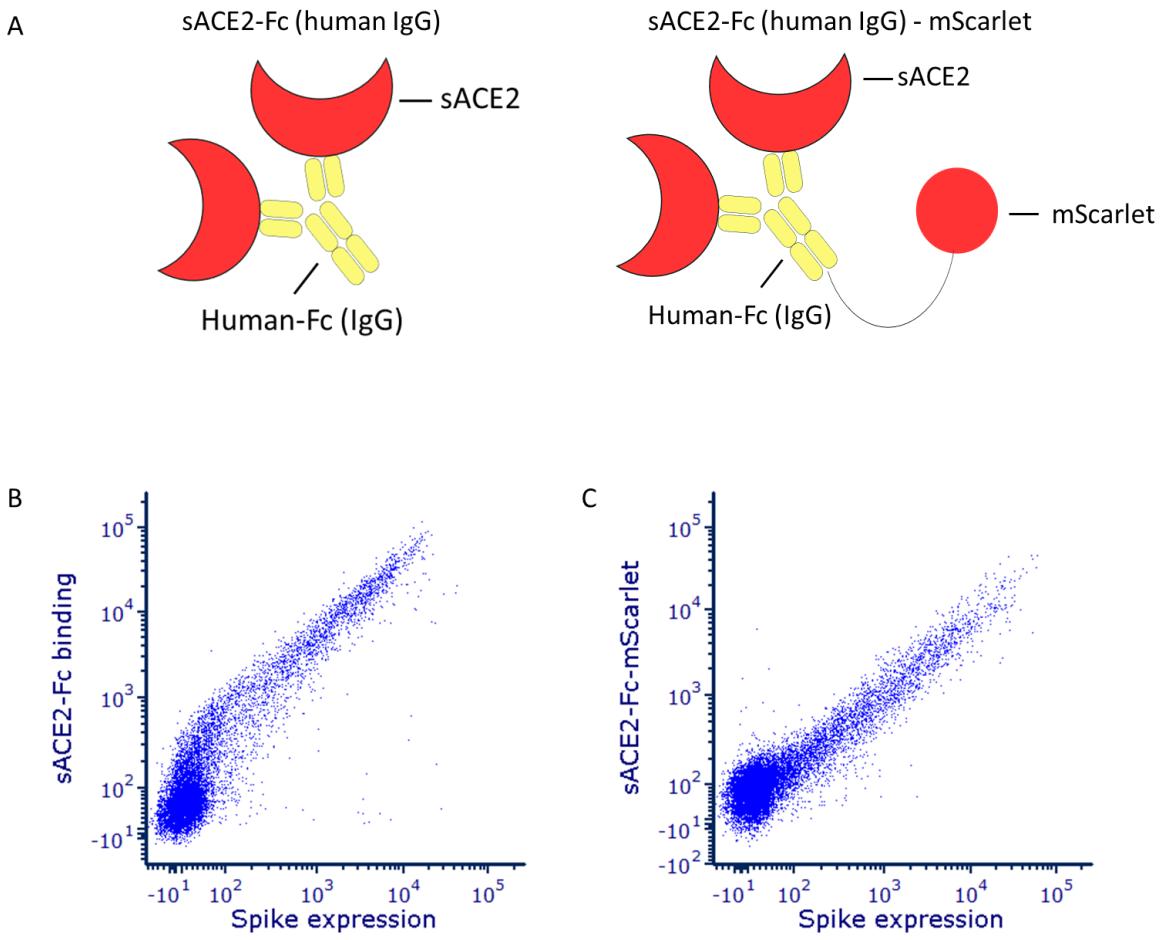


Figure 2-6: Tagging sACE2-Fc with mScarlet does not affect binding to SARS-CoV-2 spike. a) Schematic of sACE2-Fc-mScarlet. HEK-293T cells expressing spike were incubated with sACE2-Fc and sACE2-Fc-mScarlet for 1 hour. Spike expressing HEK-293T cells incubated with sACE2-Fc were secondarily incubated with an anti-IgG(human) for 30 minutes. Dot plots of b) sACE2-Fc and c) sACE2-Fc-mScarlet binding to HEK-293T cells expressing spike.

950

951

2.10 Sequencing of the sorted cells

During the trial sort with Ly-CoV016 and the Wuhan-D614G: Gamma spike mix, plasmids were extracted from sorted cells using a commercial miniprep kit, bacteria transformed with the extracted plasmids, transformed bacteria outgrown and plasmids extracted, and Sanger sequenced. The extraction of plasmids from the sorted HEK-293T cells was inefficient with approximately 210 bacterial colonies being transformed from 3000 sorted cells (data not shown). While this may be sufficient for a low diversity library, for a high diversity library this will not provide adequate data to make reliable conclusions, unless many orders more of cells are sorted.

To increase the efficiency of obtaining sequence information from sorted cells it was decided that extraction of RNA transcribed from the plasmids would be used instead. Assuming single coding plasmids are being transfected into each mammalian cell using the modified approach, for each plasmid multiple RNA copies will be transcribed increasing the probability of successful RNA extraction and sequencing. Thus, total RNA extraction of the sorted cell pellet would be carried out. The extracted RNA is then reverse transcribed to cDNA using a gene specific primer for the RBD. The cDNA is then PCR amplified for next generation sequencing (NGS). Previous studies have aimed for a minimum of 10,000 cells from a selection using FACS[119, 133].

2.11 Discussion

In this chapter, the initial design of the mammalian display platform and sequence extraction strategy has been improved upon by tagging spike with a fluorescent protein to allow automatic measurement of expression at the cell surface, developing a method for selection using ACE2 binding to measure the fitness effects of mutations on binding and improving the efficiency of sequencing extraction by extracting RNA instead of plasmid DNA.

In the next chapter, the library to be expressed and displayed on the mammalian cell display will be constructed.

CHAPTER 3

3 CREATING THE LIBRARY

3.1 Introduction

Deep mutagenesis scanning (DMS) involves the creation of a library of mutations encompassing all the possible amino acid options and then the exploration of their phenotype in a high throughput manner, coupled with sequencing to identify the desired genotypes. In nature, viral evolution occurs mostly through the serial acquisition of one or two mutations, such as with antigenic drift in seasonal influenza[141] and drift in the seasonal coronavirus 229E[20]. Viruses can in addition to this genetic drift, evolve more dramatically through recombination, a common mechanism used by coronaviruses[10] and reassortment, a unique mechanism of evolution to segmented viruses involving the acquisition of segments from other strains, which is the underlying mechanism for antigenic shift in influenza[141]. Recombination and reassortment are more difficult to model using mutagenesis libraries. In this study, it was aimed to create a mutagenesis library consisting of mostly single or double mutations since this mirrors the genetic drift that occurs in evolution more closely.

A DMS library can be made using PCR, there are 2 main ways to create a single site mutagenesis library using PCR; error prone PCR or using degenerate codon primers.

3.2 Error prone PCR versus overlap extension PCR with degenerate primers

The two options for creating a single site mutagenesis library are using an error prone PCR or using primers encoding degenerate codons with overlap extension PCR. The aim of an error prone PCR is to alter conditions to reduce to fidelity of the PCR allowing mutations to occur at a higher rate. This can be done by using a polymerase with a higher error rate, using a higher magnesium concentration, altering the balance of nucleotides, and increasing the extension time[142]. Error prone PCRs have a low rate of introducing new mutations, are biased in the mutations produced and produce libraries with a lower depth of amino acid coverage[142]. The degenerate codon approach has the ability to produce mutagenesis libraries including every amino acid as single mutations at each site in the target sequence using a few rounds of PCR

.

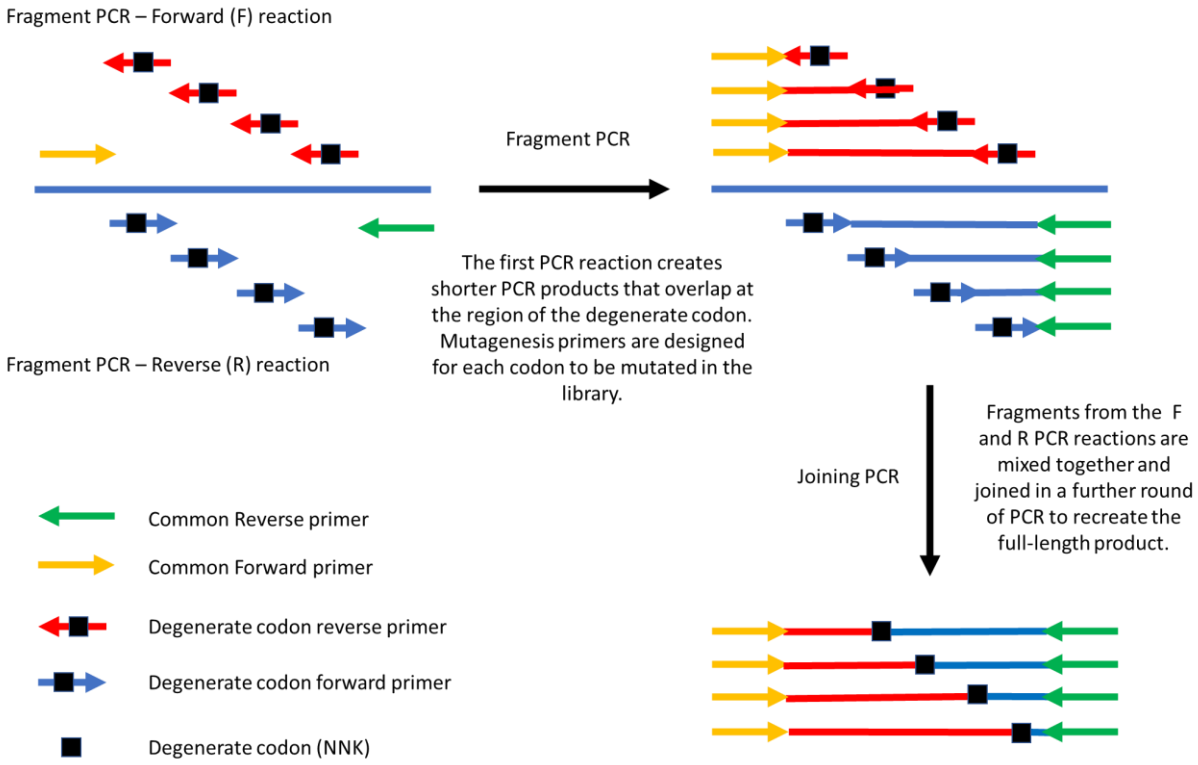
A degenerate codon is made by incorporating a mixture of nucleotides during oligonucleotide synthesis such that the oligonucleotides synthesised contain a mixture of different codons at the site of degeneracy. There are 64 possible codons encoding 20 amino acids and 3 stop codons. A range of degenerate codon options are available that limit the number of codons included. This can be used to reduce the size of the library, making it easier to construct fully representative libraries and screen the effects of these novel amino acid combinations on protein function. The simplest degenerate codon is NNN (where N= A, C, T or G), which includes all 64 possible codons, due to the degeneracy of the codon code, amino acids are not equally represented by codons[143]. For instance, using a NNN approach, methionine and tryptophan are encoded by a single codon, whereas arginine is encoded by 6 codons, so a library produced using NNN

codons would be expected to contain 6 times the number of arginine containing mutants relative to the rarest amino acids. NNK (where K = G or T) is an alternative degenerate codon that only includes 1 stop codon, while still encoding all 20 amino acids, in addition the difference in codon coverage for the rarest encoded and most frequently encoded amino acid is 3 fold[143]. Using NNK for library production reduces the library size required for any depth of coverage compared to using NNN, due to the higher probability of premature stop codon and lower probability of including rarer codons with NNN[143].

3.3 Optimising the PCR protocol

The PCR protocol for generating the mutagenesis libraries was based on Dingens et al[144], which consisted of 3 rounds of overlapping PCR with each round divided into a low cycle number (7 cycles) fragment PCR used to generate fragments of the region to be mutated and then a joining PCR designed to combine the fragments forming the full length product (figure 3-1). Mutagenesis primers were designed with the degenerate codon in the middle of the primer with equal numbers of complementary bases upstream and downstream of the degenerate primer. The length of the primers was determined by the target melting temperature (65ºC) with bases added until the target melting temperature was reached. The mutagenesis primers (forward and reverse) were designed for each amino acid residue of the SARS-CoV-2 RBD (positions 319 - 529) using a python script (https://github.com/jbloomlab/CodonTilingPrimers [144]). The major modification to this protocol was to reduce the number of rounds of overlapping PCR to 1 (figure 3-1). To compensate for the lower number of PCR rounds, the cycle number of the fragment PCR was increased to ensure an adequate frequency of mutation. A further optimisation was to reduce the amount of template DNA used in the fragment PCR to

2ng compared to 20ng in the Dingens protocol[144], this would also be expected to increase the frequency of mutations in the overlapping PCR reaction. Figure 3-2 shows the effect of mutation frequency using different PCR cycle numbers for the fragment PCR reaction.



Dingens et al. mutagenesis PCR for DMS protocol

Fragment PCR – 7 cycles
Joining PCR – 20 cycles
} 3 x

Modified protocol

Fragment PCR – 10 cycles
Joining PCR – 20 cycles
} 1 x

PCR protocol:

Forward		Reverse	
10 x KOD Buffer	2.5	10 x KOD Buffer	2.5
MgS04	1.5	MgS04	1.5
dNTPs	2.5	dNTPs	2.5
F pool	0.75	R pool	0.75
R primer	0.75	F primer	0.75
H20	15.5	H20	15.5
Template (2ng)	1	Template (2ng)	1
KOD	0.5	KOD	0.5
Total	**25**	**Total**	**25**

Thermocycler settings:

	Temp (°C)	Time	
Activation	95	3 mins	
Denature	95	30 s	10 cycles
Anneal	58	45 s	
Extension	70	3 mins	
Last cycle	70	5 mins	
Hold	4		

Joining PCR	
10 x KOD Buffer	5
MgS04	4
dNTPs	5
F fragments (1:4)	6.66
R fragments (1:4)	6.66
5' primer	1.8
3' primer	1.8
H20	18.08
KOD	1
Total	**50**

	Temp (°C)	Time	
Activation	95	3 mins	
Denature	95	30 s	20 cycles
Anneal	58	45 s	
Extension	70	3 mins	
Last cycle	70	5 mins	
Hold	4		

Figure 3-1: Schematic of PCR protocol used to produce the mutagenesis library using degenerate codon primers and overlap extension PCR. The protocol used was modified from Dingens et al.[144] The protocol and thermocycler conditions used for the fragment and joining PCRs are shown at the bottom of the figure.

1054

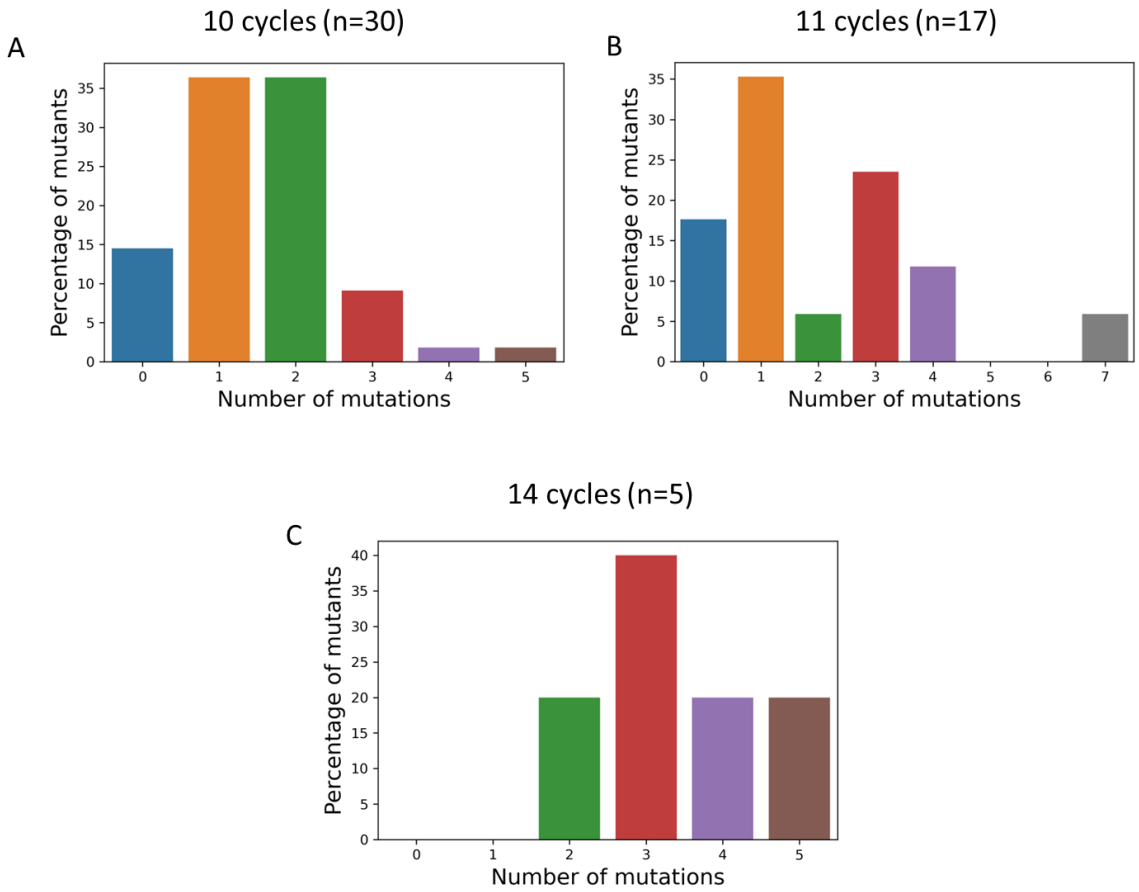


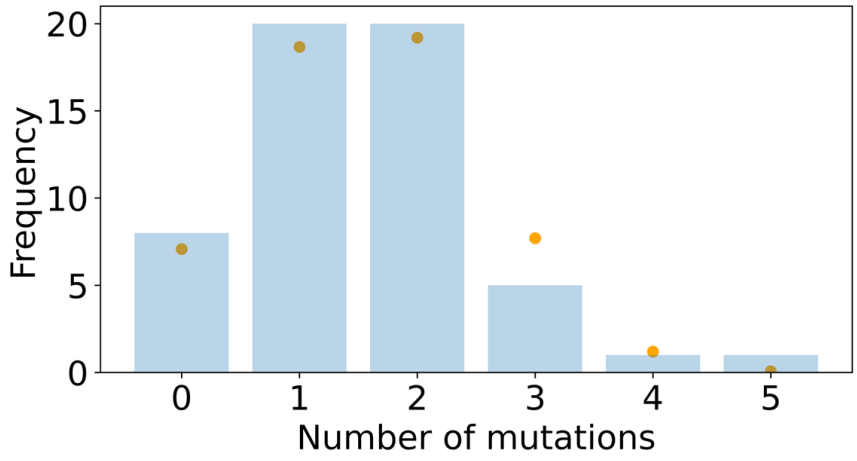
Figure 3-2: Frequency distribution of the number of RBD mutations using different PCR cycle numbers. Plasmids from libraries constructed using the different PCR cycle lengths were used to transform bacteria, a sample of the transformed bacterial colonies had their plasmids extracted and Sanger sequenced. a) 10 cycles (n=30). b) 11 cycles (n=17) and c) 14 cycles (n=5)

10 cycles were chosen, as this offered the best balance between the number of plasmids having single mutations, no mutations and more than one mutation.

To predict the number of clones that would need to be included in the library, 55 clones transformed from the 10 PCR cycle library were Sanger sequenced. The frequency distribution is shown in figure 3-3A, the data appears to be normally distributed. Using a normally distributed model, the probability of having a single mutation in the RBD was

calculated and used to estimate the number colonies that would have to be included in the library to encompass all possible single mutations assuming primer efficiency was equal figure 3-3B. There are 211 amino acids in the RBD of SARS-CoV-2 and 19 other amino acids that each RBD position could be, giving rise to 4009 (19*211) mutations to be included in a complete mutagenesis library. The frequency distribution of RBD mutations in the library constructed using 10 cycles fits most closely a normal distribution. Fitting a normal distribution to the frequency distribution, the probability of having just a single mutation in the RBD is 0,339. Assuming all the mutagenesis primers across the RBD worked evenly, a minimum of 11,816 (= 4009/0.339) colonies would have to be included in the library, however using NNK degenerate primers means the rarest codons are present 3 times less than the most common, consequently 3 x 11,816 colonies = 35,448 colonies would have to be included in the library figure 3-3C.

A



B

Normal distribution	Probability	Expected number of mutations
probability of 0 mutations	**0.129**	**7.1**
probability of 1 mutation	**0.339**	**18.7**
probability of 2 mutations	**0.349**	**19.2**
probability of 3 mutations	**0.140**	**7.7**
probability of 4 mutations	**0.022**	**1.2**
probability of 5 mutations	**0.001**	**0.1**

C

Total number of mutations	**=19 * 211**	**4009**
Number of colonies to include in library for all single mutations	**=4009 / 0.339**	**11816**

Figure 3-3: Distribution of the number of mutations per RBD in the library. a) 55 transformed colonies using plasmids from the library, had their plasmids extracted and Sanger sequenced. The number of plasmids with the specified number of mutations present in the RBD of each cloned sequenced is shown as a histogram. The blue bars show the actual distribution, while the orange circles show the expected numbers from the fitted normal distribution. b) The distribution of the number of mutations per RBD is normally distributed. The probability of a plasmid from the library having a specified number of mutations and the expected numbers in the 55 plasmids sampled are shown. c) Based on the probability of having single mutations, the expected number of colonies that would have to be included in the library, assuming mutations at each position in the RBD are represented evenly is shown.

1076

To explore the efficiencies of the mutagenesis primers used in the construction of the library, the position of 119 mutations from 64 colonies transformed from the mutagenesis library were plotted on a frequency distribution (figure 3-4). The frequency distribution shows most locations had a single mutation, 6 positions had 3 mutations and one position had 6 mutations.

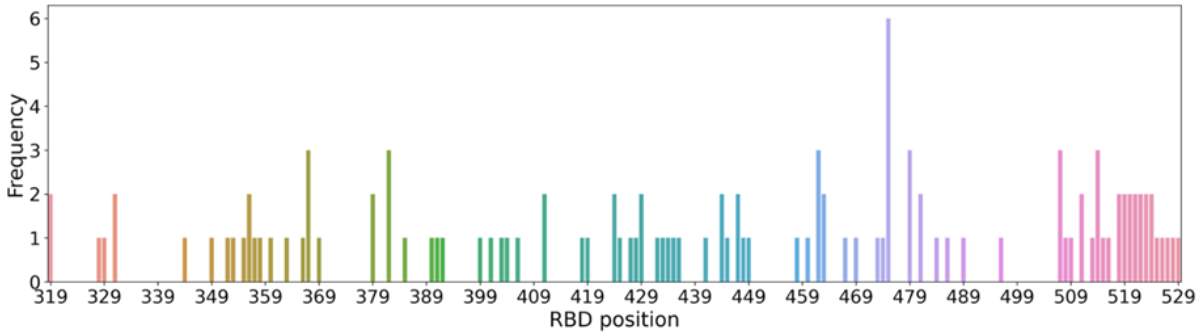


Figure 3-4: The frequency of mutations occurring at positions in the RBD from a sample of plasmids from the plasmid library. 119 mutations were identified from Sanger sequencing of 64 transformed bacterial colonies using the plasmids from the library. The number of mutations occurring at each position in the RBD are shown.

3.4 Next generation sequencing of the Alpha spike RBD plasmid mutagenesis library

To further explore the diversity of the Alpha spike RBD library, the plasmid library was sequenced using next generation sequencing (NGS) in two independent sequencing runs. The RBD was divided into 2 amplicons of 500 nucleotides for NGS sequencing. Due to the strategy of amplicon sequencing it will not be possible to exclude linkage of mutations that occur in the same RBD, but in different amplicons. The proportion of each mutation in the library was determined by dividing the number of reads with a mutation by the total number of reads. Mutations were included in the library if the proportion of the mutation in two independent reads differed by less than 4-fold. The correlation of the

proportions of each mutation included in the library from the two independent sequencing run is shown in figure 3-5. Of the possible 4220 mutations (including wild type), 4183 mutations were present in the library, representing 99.1% coverage. 37 mutations were missing from the library, particularly around positions 500-504 due to primer failure.

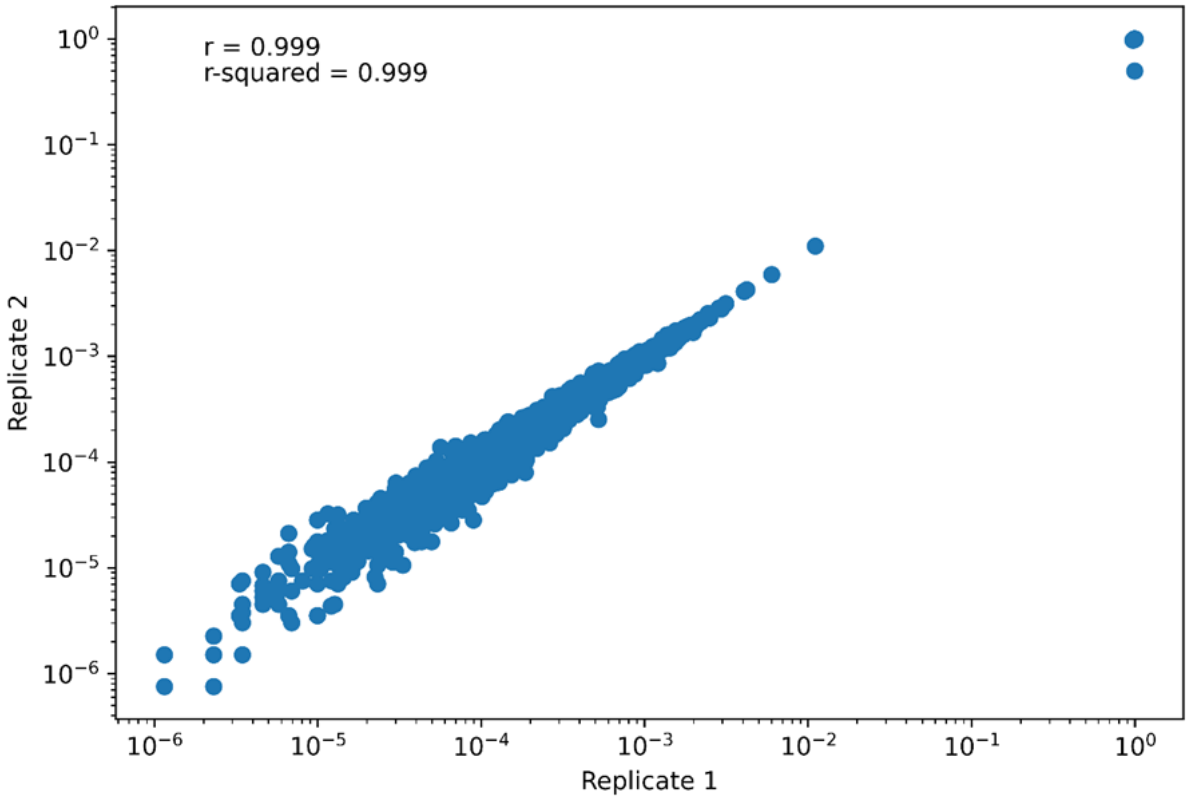


Figure 3-5: Correlation between 2 independent NGS sequencing runs of the plasmid library. The RBD of the Alpha plasmid library was sequenced independently in duplicate by NGS. The proportion of reads containing each point mutation was calculated. The axes represent the proportion of reads for each point mutation in each of the 2 sequencing runs. Point mutations with proportions differing by a greater than 4-fold difference between replicates were excluded from the library. The scatter plot shows the correlation between proportions of point mutations included in the subsequent analyses. R value (Pearson rank coefficient).

3.5 Discussion

A PCR protocol was established to create a library using degenerate codons (NNK), sequencing showed 99% of possible mutations were included in the library. It is not possible to say if these are all single mutations as amplicon sequencing was used, however Sanger sequencing of a sample of the library suggests the library consists mostly of Alpha spike plasmids containing one or two mutations per RBD.

Now that the Alpha RBD spike library is established, the mammalian cell display platform will be used to identify which mutations confer the following phenotypes increased ACE2 binding, escape from monoclonal antibodies and escape from vaccine sera.

CHAPTER 4

4 SCREENING FOR MUTATATIONS IN THE ALPHA SPIKE RBD THAT INCREASE ACE2 BINDING

4.1 Introduction

The early phase of SARS-CoV-2 spike evolution has been focused on increasing ACE2 binding, highlighted by the successive appearance and fixation of D614G[48, 50, 145], N501Y[53], and L452R[60] in variants that became globally dominant. To ask what further mutations might arise, the first DMS screen sought to identify mutations in the Alpha spike RBD that further increased ACE2 binding.

4.2 Tolerance of the RBD library for ACE2 binding

With the library established, it was important to next see how well the library is expressed as spike on the cell surface and how it binds to sACE2-Fc-mScarlet, which will be used for selection in all experimental sorts. The sACE2-Fc-mScarlet was produced by transfecting the sACE2-Fc-mScarlet plasmid into HEK-293T cells, collecting and filtering the supernatant after 72 hours and freezing a batch into small aliquots for storage at -20^{o}C. The library was transfected into HEK-293T cells using 1ng of plasmid expressing Alpha spike with the RBD library tagged at the C-terminal end with mGreenLantern diluted with a non-coding plasmid, as previously described such that most cells that express spike would have only received a single coding plasmid. Figure 4-1 shows dot plots of ACE2 binding against spike expression comparing WT Alpha spike compared to the Alpha spike RBD library. The dot plot showing WT Alpha spike binding to ACE2 is a straight line with correlation between the level of spike expression per cell

and the amount of ACE2 bound with 67% (100*1.71 / (1.71+0.83)) of spike expressing cells being able to bind ACE2. By comparison the Alpha spike library binding to ACE2 shows 36% (100*0.82 / (0.82+1.43)) of spike library expressing cells were able to bind ACE2.

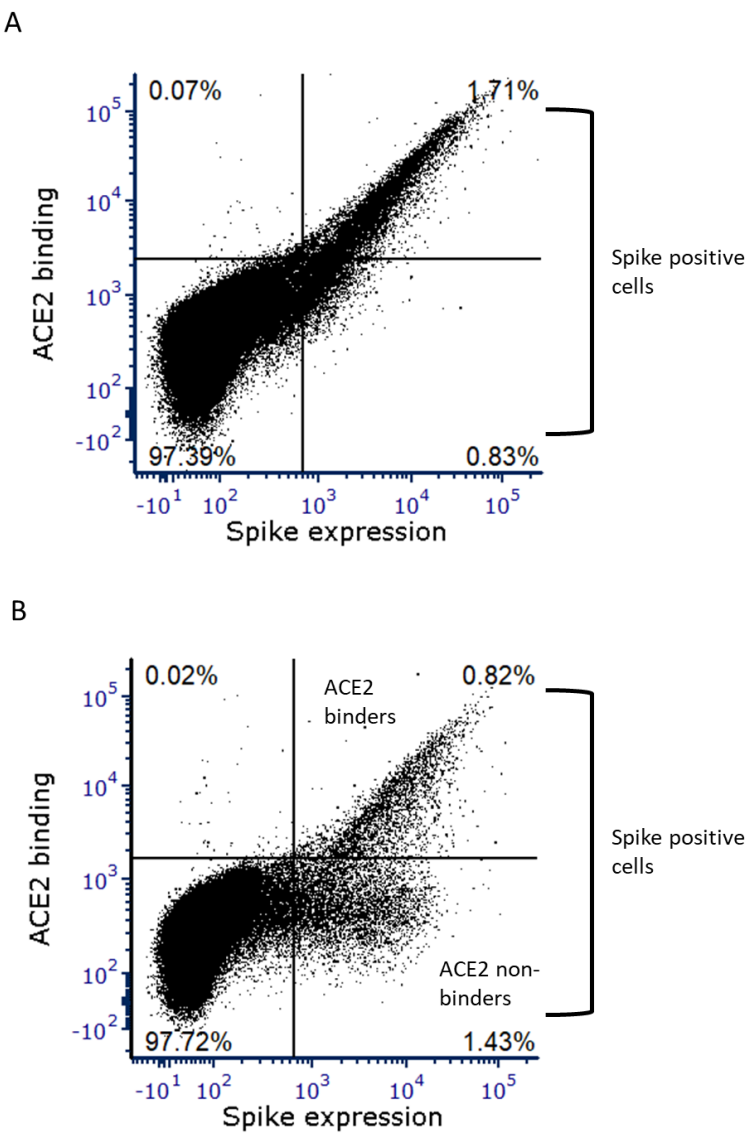


Figure 4-1: Binding of sACE2-Fc-mScarlet by Alpha spike and the Alpha RBD library. HEK-293T cells were transfected with 1ng of spike expressing plasmid tagged at the C-terminus with mGreenLantern mixed with 1500ng of non-coding plasmid per 10^6 cells. 24 hours later cells were dissociated and incubated with sACE2-Fc-mScarlet for 30 minutes before analysis on the flow cytometer. a) WT Alpha spike, b) Alpha spike RBD library

1141

1142

4.3 sACE2-Fc-mScarlet titration

Prior to sorting, the sACE2-Fc-mScarlet was titrated for binding against WT Alpha spike expressed on HEK-293T cells using flow cytometry (figure 4-2). A volume of ACE2-Fc-mScarlet was chosen that was sub-saturating to allow mutations that increase ACE2 binding to be more apparent. For this reason, 31.25ul per ~5*10^5 transfected cells was chosen for sorts screening for mutations that increase ACE2 binding.

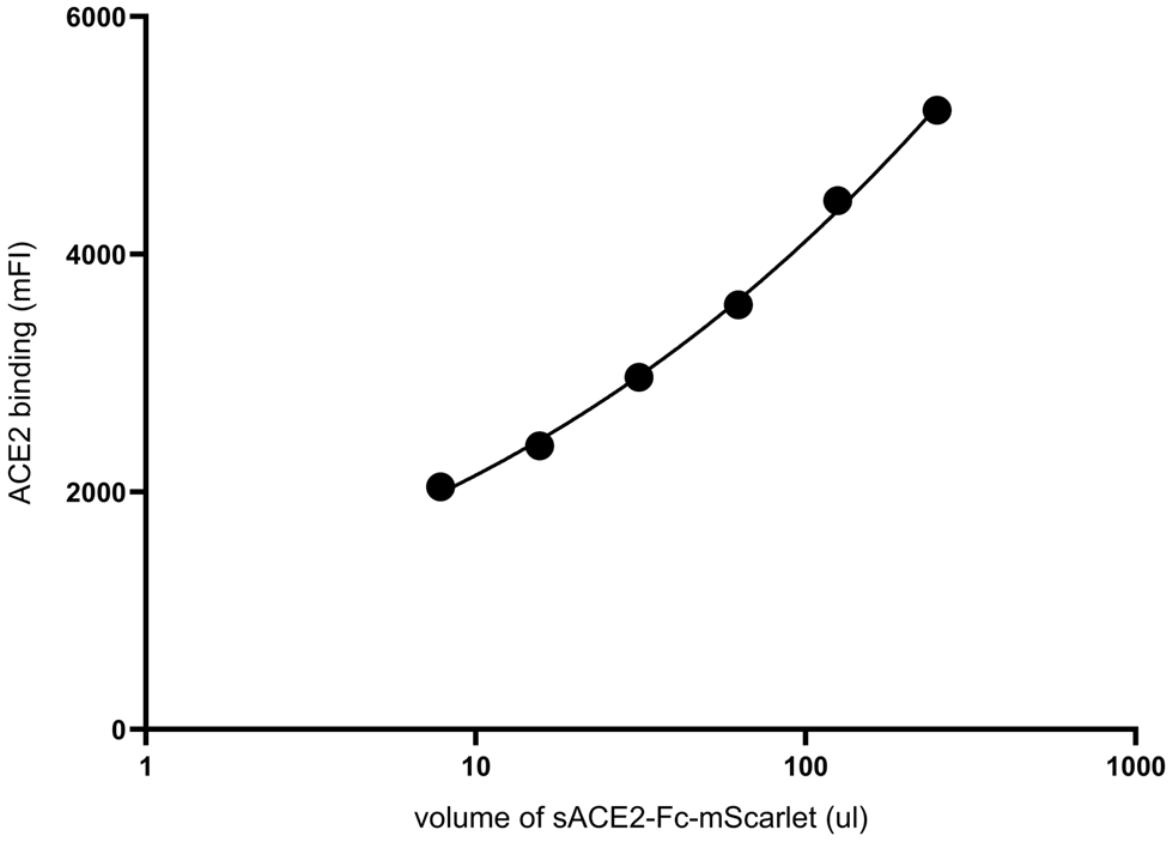


Figure 4-2: sACE2-Fc-mScarlet titrations. sACE2-Fc-mScarlet was transfected into HEK-293Ts cells and the supernatant filtered and harvested after 72 hours. A range of volumes of the sACE2-Fc-mScarlet supernatant was used to incubate HEK-293Ts cells expressing spike tagged with mGreenLantern. Binding of sACE2-Fc-mScarlet was assessed using flow cytometry, the median fluorescence intensity is shown.

4.4 A screen for RBD mutations that increase ACE2 binding in Alpha spike

22 x 10^6 293T cells were transfected with the Alpha spike RBD library tagged with mGreenLantern mixed with a non-coding plasmid at a ratio of 1:1500 to reduce the number of coding plasmids entering the same cells and obfuscating the link between genotype and phenotype, 1ng of coding DNA was transfected per 10^6 cells. This led to less than 10% of cells expressing spike. 24 hours after the transfection, cells were dissociated and incubated with the sACE2-Fc-mScarlet at a sub-saturating level for 30 minutes before being sorted by FACS.

The gates for sorting can be seen in figure 4-3, DAPI was used as a live/dead dye to exclude dead cells. The top 5% of ACE2 bound cells that express spike were sorted. It was noted the ACE2 signal would drop during the sort with time, requiring monitoring and resetting of the gates as required to maintain the top 5% of ACE2 bound cells. The drop in signal was likely due to dissociation of ACE2 from spike or of S1 from S2 and has previously been recognised[133]. Due to the dropping in ACE2 signal, the sort was batched with a new set of cells for sorting after 2 hours. This continued until at least 10,000 cells had been collected, where sorts had to be paused overnight, cells were pelleted and stored at -80^{o}C.

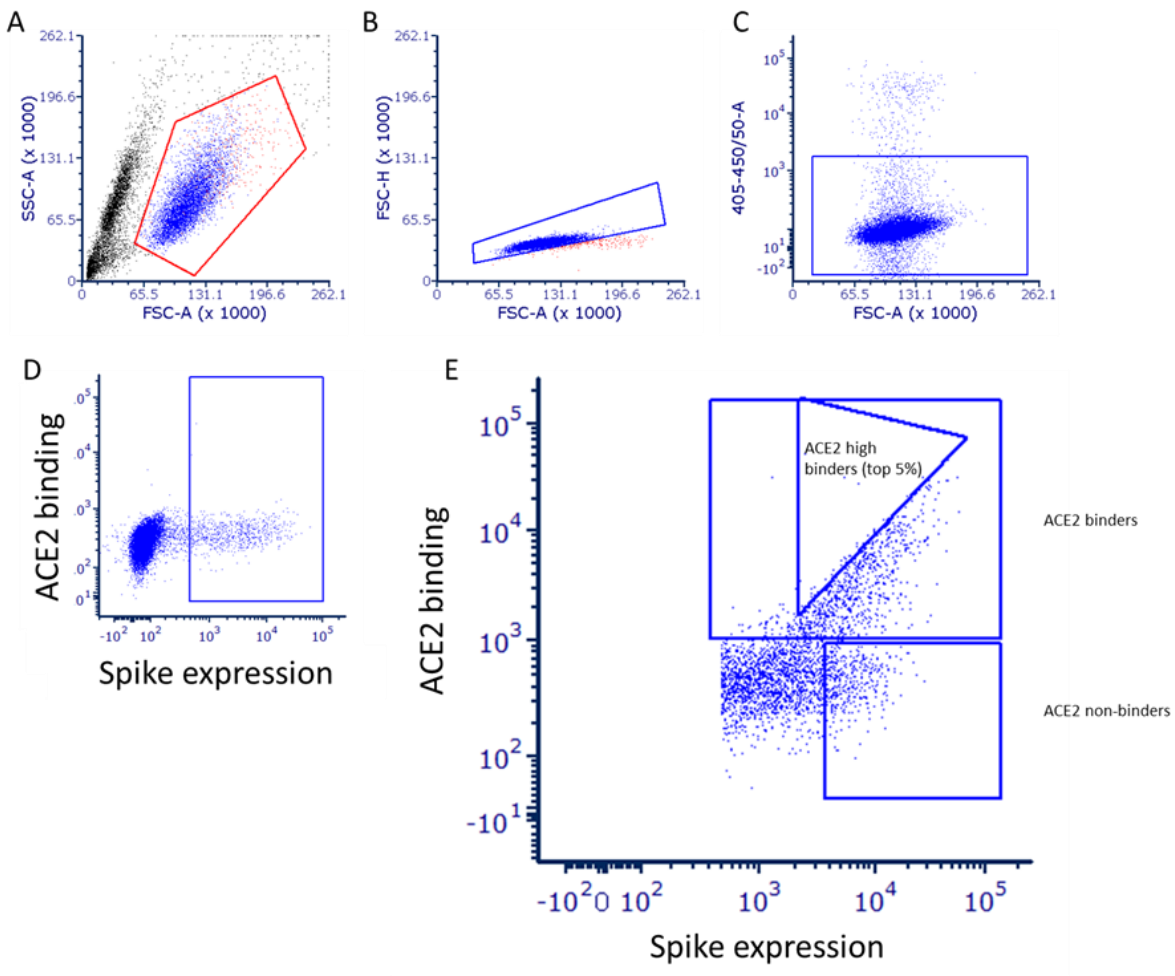


Figure 4-3: FACS gating strategy for high ACE2 binding. a) FSC-A against SSC-A to remove dead cells and debris. b) FSC-A against FSC-H to remove doublets. c) FSC-A against 405-450/50-A to remove dead cells by staining with DAPI. d) 488-525/50-A (spike expression) against 561-582/15 –A (ACE2 binding) to select spike expressing cells. These cells were not incubated with ACE2-Fc-mScarlet to serve as a control. e) 488-525/50-A against 561-582/15-A, the top 5% of ACE2 binding cells were sorted from an ACE2 positive, spike expressing population. The spike protein used were all tagged at the C-terminal end with mGreenLantern and the sACE2-IgG Fc tagged with mScarlet.

Total RNA was extracted from the pelleted cells and reverse transcribed using a gene specific primer for the RBD. The RBD cDNA was PCR amplified as 2 amplicons and adapters for Illumina sequencing added. Next generation sequencing (NGS) and analysis of the sequencing data was conducted as described in the methods. The proportion of the

variants in the sorted population was divided by the proportion of the variant in the starting plasmid library to produce the enrichment score. Variants that were selected for in the sorted population would have scores greater than 1.

The heatmap (figure 4-4) has been filtered to show positions in the RBD that are involved directly in ACE2 binding[52] and those positions mutated in BA.1[58] (the dominant variant at the time of writing) relative to Wuhan. From the heatmap, it can be seen that E484K and Q498R have the largest effect on increasing ACE2 binding by Alpha spike. Interestingly, Q498R emerged in BA.1 and is conserved in the Omicron descendants in keeping with the findings from the screen, which would suggest Q498R has an important role in increasing ACE2 binding (figure 4-4).

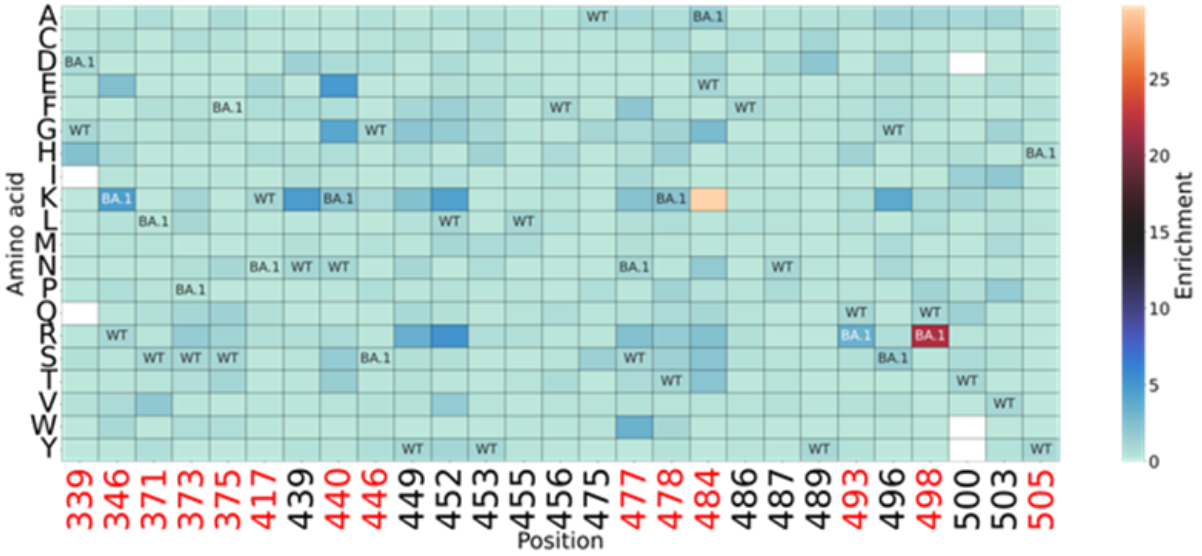


Figure 4-4: Heatmap of point mutations showing their enrichment score for ACE2 binding. The RBD positions shown have been filtered to show those involved in the ACE2 binding interface[52] and mutations that occurred in BA.1. Red numbers on the x axis, represent positions mutated in BA.1. WT = wild type amino acid. BA.1 = mutation found in BA.1. Blank squares represent point mutations not covered by the library.

E484K and Q498R are the two mutations that standout from the heatmap as increasing ACE2 binding (figure 4-4). To confirm the results of the screen, the individual point mutations were engineered to Alpha spike and the effect on hACE2 binding determined using flow cytometry (figure 4-5 and 4-6). Spike expressing cells were incubated with a fixed volume of sACE2-Fc-mScarlet and the median fluorescence intensity of ACE2 binding measured for a fixed level of spike expression to obtain measures of ACE2 binding. The fold difference in ACE2 binding compared to the parent Alpha spike is shown in figure 4-5.

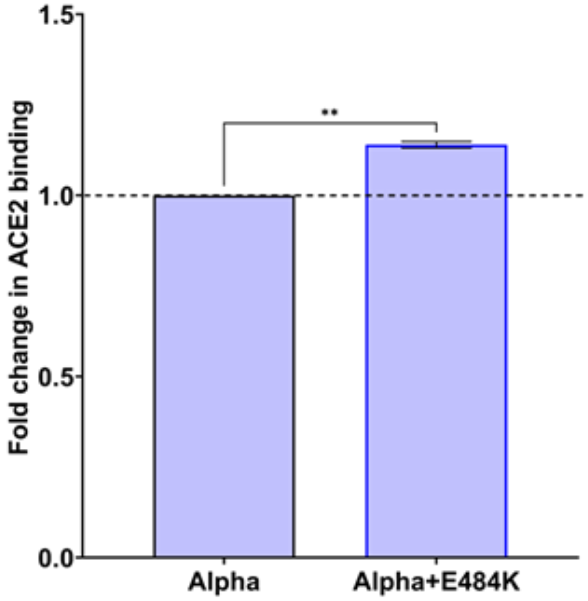


Figure 4-5: HEK-293T cells expressing spike were incubated with sACE2-Fc-mScarlet overnight, and their binding measured using flow cytometry. The data presented is the fold difference in the median ACE2 binding for each mutant relative to Alpha spike normalised for expression. N=2, error bars represent the range. ** p value < 0.01, one-way Anova.

The addition of E484K led to a 20% increase in ACE2 binding by Alpha spike (figure 4-5). Position 498 was selected for in a yeast DMS screen using a Wuhan RBD for

increased ACE2 binding, however it was Q498H that was selected for and the mutation Q498R was predicted to reduce ACE2 binding. To explore if the differential predictions for Q498R were due to epistasis or related to using trimeric spike displayed on mammalian cells, the Q498R and Q498H mutations were added to Alpha, Wuhan+N501Y and Wuhan spike and ACE2 binding measured (figure 4-6).

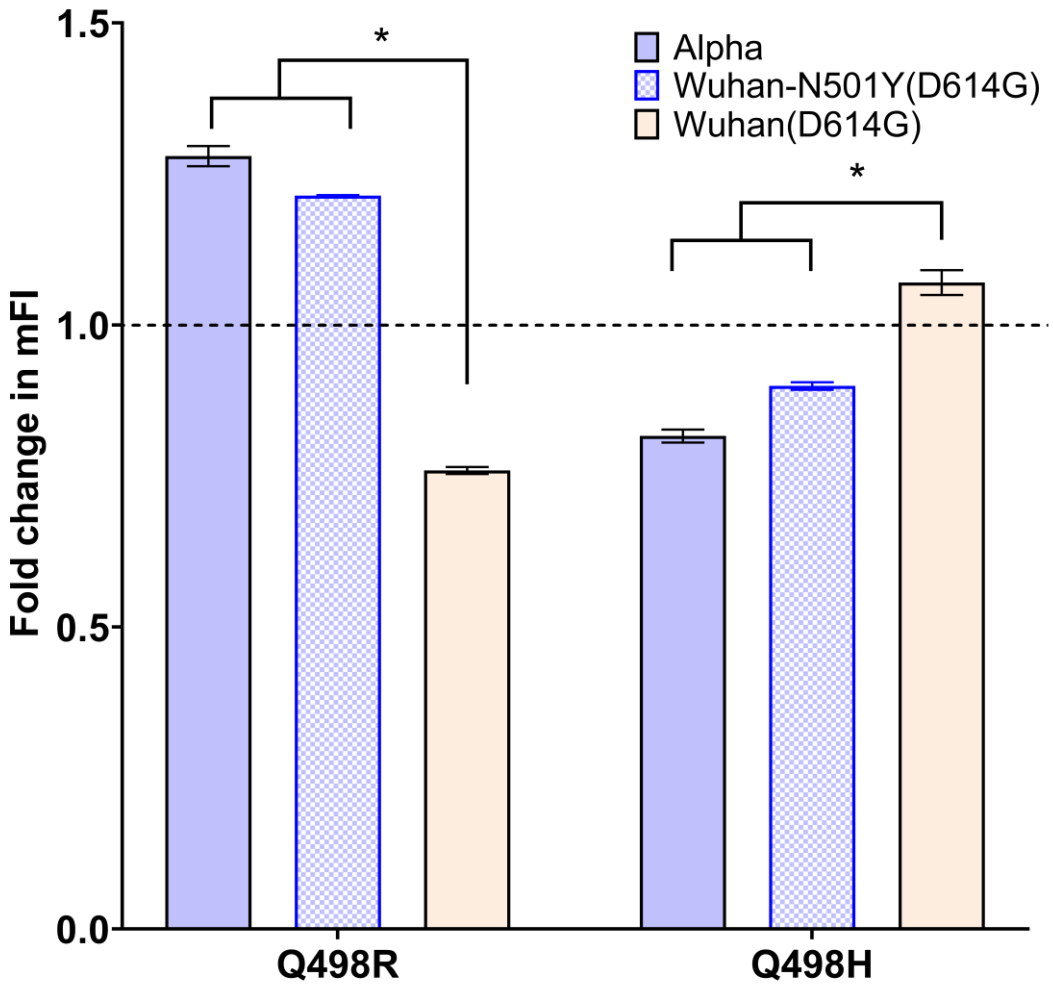


Figure 4-6: Relative ACE2 binding for the mutations Q498R and Q498H on Alpha, Wuhan+N501Y(D614G) and Wuhan(D614G) trimeric spikes. HEK-293T cells expressing spike tagged at the C terminal end with mGreenLantern were incubated with sACE2-Fc-mScarlet supernatant overnight. ACE2 binding was measured as the median fluorescence intensity for a fixed level of spike expression using flow cytometry. Data presented is the fold difference in median ACE2 binding for each mutant relative to their WT parent, normalised for expression. N=2, error bars represent the range. * $p<0.05$ paired t-test comparing difference in ACE2 binding between a 501Y bearing RBD (Wuhan-N501Y and Alpha) compared to a 501N bearing RBD (Wuhan(D614G)) following the introduction of Q498R and Q498H.

1204

1205

The effect of the Q498R mutation on ACE2 binding was dependent on the nature of the residue at position 501. Where 501 is a tyrosine (Y), Q498R increased ACE2 binding, by contrast in the presence of the asparagine (N) at 501, Q498R has a deleterious effect. The reverse appears to be true for Q498H, which only increased ACE2 binding with 501N present in the RBD.

A possible reason for the phenotypic interdependence of residues 498 and 501 can be illustrated structurally in Figure 4-7. Residues 498 and 501 are located in close proximity to each other, tyrosine and histidine are bulky residues containing an aromatic ring, while arginine has no such rings. Steric hindrance between the aromatic rings of histidine and tyrosine may account for the negative effects on ACE2 binding seen with Q498H with N501Y.

Further support for the results from the ACE2 binding screen can be seen in figure 4-7, residues 484 and 498 each makes close contact with ACE2, mutations at these residues would therefore be expected to have effects on ACE2 binding.

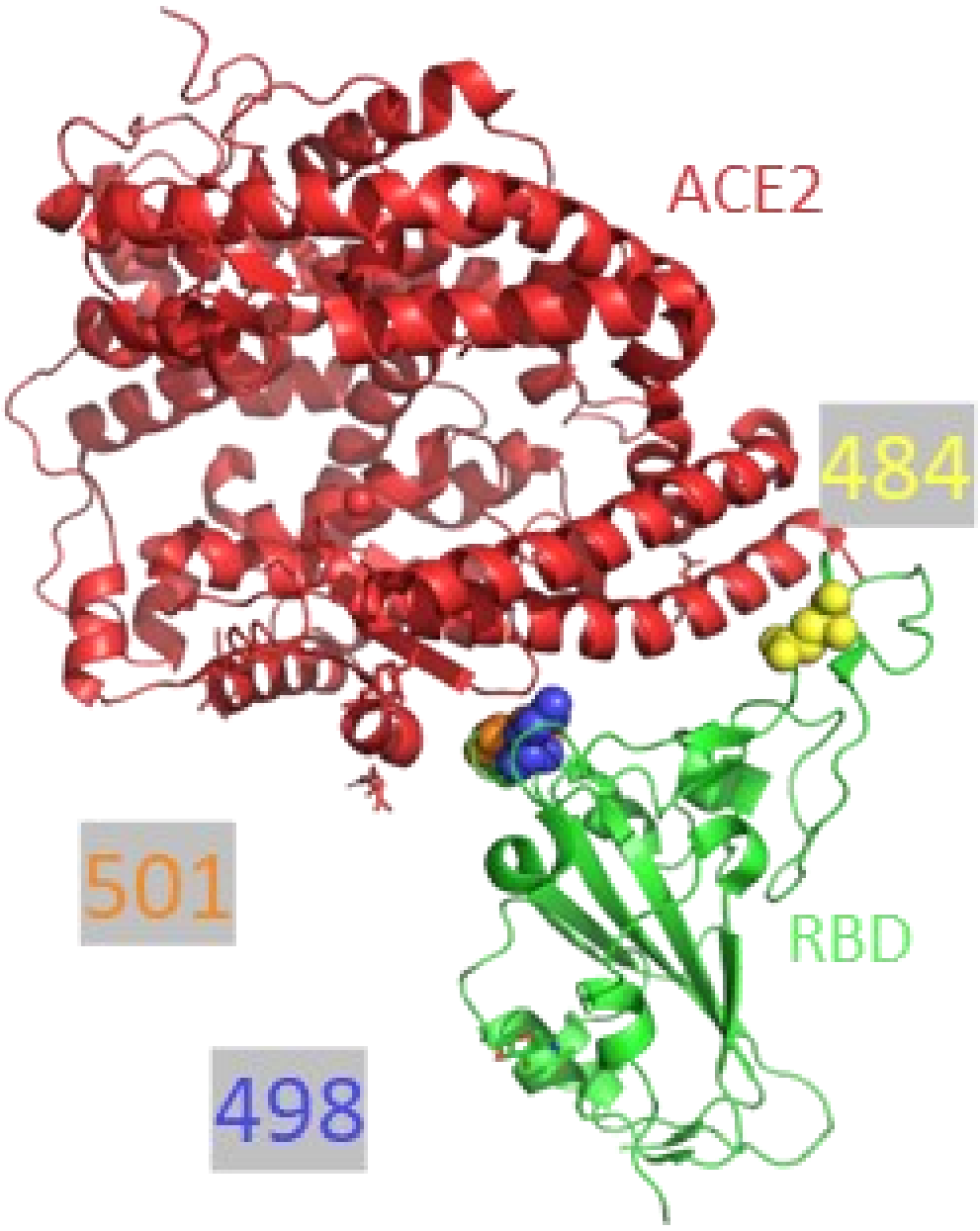


Figure 4-7: RBD positions 484(yellow), 498(blue), 501(orange) are directly involved in the interaction with hACE2(red). Figure created using PyMOL (The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC.) PDB: 6M0J (Lan, Ge et al. 2020)

1222

1223

4.5 Discussion

Here, we have used an Alpha spike RBD library displayed on mammalian cells to screen for mutations that increase human ACE2 binding. 36% of cells expressing the Alpha spike library were capable of binding ACE2 at a sub-saturating concentration compared to 67% of cells expressing the WT Alpha spike library, suggesting 54% (100* 36/67) of mutations in the spike library maintained the ability to bind human ACE2. This is in keeping with the results from yeast[52] and mammalian DMS[119] using a Wuhan RBD and denotes a remarkable degree of mutational tolerance and plasticity in the RBD of SARS-CoV-2 spike.

An Alpha spike was chosen for the DMS screens, as at the time of conception Alpha was the dominant variant and N501Y was present in multiple circulating variants, Alpha, Beta and Gamma. Two mutations were enriched in the sort for increased ACE2 binding, E484K and Q498R. Early screens for RBD mutations that increased ACE2 binding based on a Wuhan RBD identified Q498H as increasing ACE2 binding, and Q498R as decreasing ACE2 binding[52]. In a yeast RBD display forward evolution study using error prone PCR selecting for increased ACE2 binding from a Wuhan RBD, N501Y appeared first followed by the selection of Q498R and E484K[134]. The selection of Q498R as a mutation that increases ACE2 binding only occurred following the acquisition of N501Y[134]. In the presence of a RBD with N501Y, Q498H does not increase ACE2 binding, whereas Q498R does lead to an increase. This finding has been further supported by more recent DMS work using yeast[146] and chicken cell display platforms[138].

This has been reflected in nature with the emergence of the Omicron lineages[58, 82] that all bear the combination of N501Y and Q498R in the RBD and are the most transmissible variants to date.

This phenomenon, where the effect of a mutation is dependent on the surrounding residues is referred to as epistasis, as SARS-CoV-2 continues to diversify, predictions of point mutation phenotype need to be considered in the RBD background in which they might arise.

The plasticity and mutational tolerance seen in the Alpha RBD library and Wuhan RBD DMS ACE2 binding studies shows SARS-CoV-2 has a large mutational field to explore and with epistasis point mutations can have multiple phenotypes further expanding the phenotypic landscape SARS-CoV-2 can evolve into. This has implications for crossing novel species barriers, altering virulence and tropism, increasing transmissibility and immune escape.

In the next chapter, we explore how this mutational tolerance impacts escape from therapeutic monoclonal antibodies.

CHAPTER 5

5 PREDICTING MONOCLONAL ANTIBODY ESCAPE

5.1 Introduction

Monoclonal antibodies were considered as a feasible therapy for a respiratory infection for the first time during the COVID pandemic. The effectiveness of monoclonal antibodies in treating COVID varies with the stage of the infection. They are most effective when used as prophylaxis[147-149] or early in the disease in non-hospitalised mild to moderate cases[150, 151], but have less impact on those requiring hospitalisation due to COVID[152], who are typically in the late phase of the disease, which is driven by aberrant immune responses rather than ongoing viral replication.

The danger with relying on monoclonal therapies is the ease at which a virus can evolve resistance, only a single point mutation is required to significantly reduce binding and evade neutralisation. To this end, the early monoclonal antibodies clinically approved for the treatment of COVID were used in combination with the exception of Sotrovimab[153]. The idea being that by having two monoclonal antibodies with different epitopes, at least two different mutations would be required to escape from both the mAbs presenting a higher genetic barrier to resistance.

Monoclonal antibodies against SARS-CoV-2 operate in a balance between how potently they neutralise and the ease of antibody escape evolving. The most potent monoclonal antibodies with picomolar IC50s all target the ACE2 binding interface of the RBD[154, 155], however this region is very tolerant of mutations and naturally occurring variation

in the RBD is a frequent event. Sotrovimab on the other hand neutralises without blocking ACE2 binding and has a lower IC50 (10-20 fold) than potent ACE2 blocking monoclonals. However, it has a higher genetic barrier for the evolution of resistance[156, 157]. An ideal monoclonal would be one that could target a conserved region of the ACE2 binding interface of the RBD and thus have the dual properties of high potency and high barrier to resistance.

The effectiveness of any given monoclonal antibody will be dependent on the presence of escape mutations in the circulating strains of virus. In the absence of rapid, universally available viral genotyping, monoclonals would be used largely empirically, so knowing which mutations escape from monoclonals will inform decisions on their empirical use based on circulating variants. Using a DMS approach for screening of monoclonal escape mutations offers the advantage that mutations that have not yet arisen, or are a greater genetic distance away (i.e. more than a single nucleotide away) can be explored, These may not be accessible to the virus at the time, but with future mutations that genetic distance can become reachable, allowing future mutations to be characterised and more detailed maps of antibody escape to be developed. The other main advantage is a rapid or high throughput method for highlighting the relative importance a position in the RBD has on escape.

5.2 Screening for escape variants from REGN 10933, REGN 10987 and LY-CoV016

To screen for monoclonal antibody escape mutations using the Alpha spike RBD mutagenesis library, the following monoclonals were chosen: REGN 10933, REGN 10987 and LY-CoV016. The mAbs were a kind gift from Paul Kellam and Anne Palser. The mAbs were produced from expression plasmids using publicly available sequences for each of the mAbs. The mAbs were then purified and concentrated. The REGN monoclonals together form the cocktail REGN-COV, which prior to the emergence of Omicron had been shown to be effective in the prophylaxis and treatment of non-hospitalised patients infected with SARS-CoV-2[147, 150]. LY-CoV016, had similarly shown benefit when used in combination with LY-CoV555 in an outpatient setting[151].

The monoclonal antibodies were titrated by incubating serial dilutions with HEK-293Ts cells expressing the Alpha spike RBD library and choosing a dilution of monoclonal that did not block 100% of ACE2 binding by the library (figure 5-1). The amount of blockade was determined by incubating the cells expressing the spike library with sACE2-Fc-mScarlet and measuring the decrease in ACE2 signal compared to spike expressing cells not incubated with mAbs. The advantage of using the ability to bind ACE2 to positively select escape variants is that mutations that lead to escape from antibody binding but are also deleterious to ACE2 binding are selected against. Opting for a sub-blocking titre rather than a stricter titration that would block 100% of WT spike expressing cells from binding ACE2 reduces the number of cells that need to be sorted to obtain sufficient cells with escape variants. Moreover this strategy, allows greater nuance by allowing the identification of mutations that may partially decrease monoclonal binding, rather than

completely abolishing, which would be of importance clinically where monoclonal concentrations in physiologically relevant areas such as the lung are lower and risk of breakthrough higher[158].

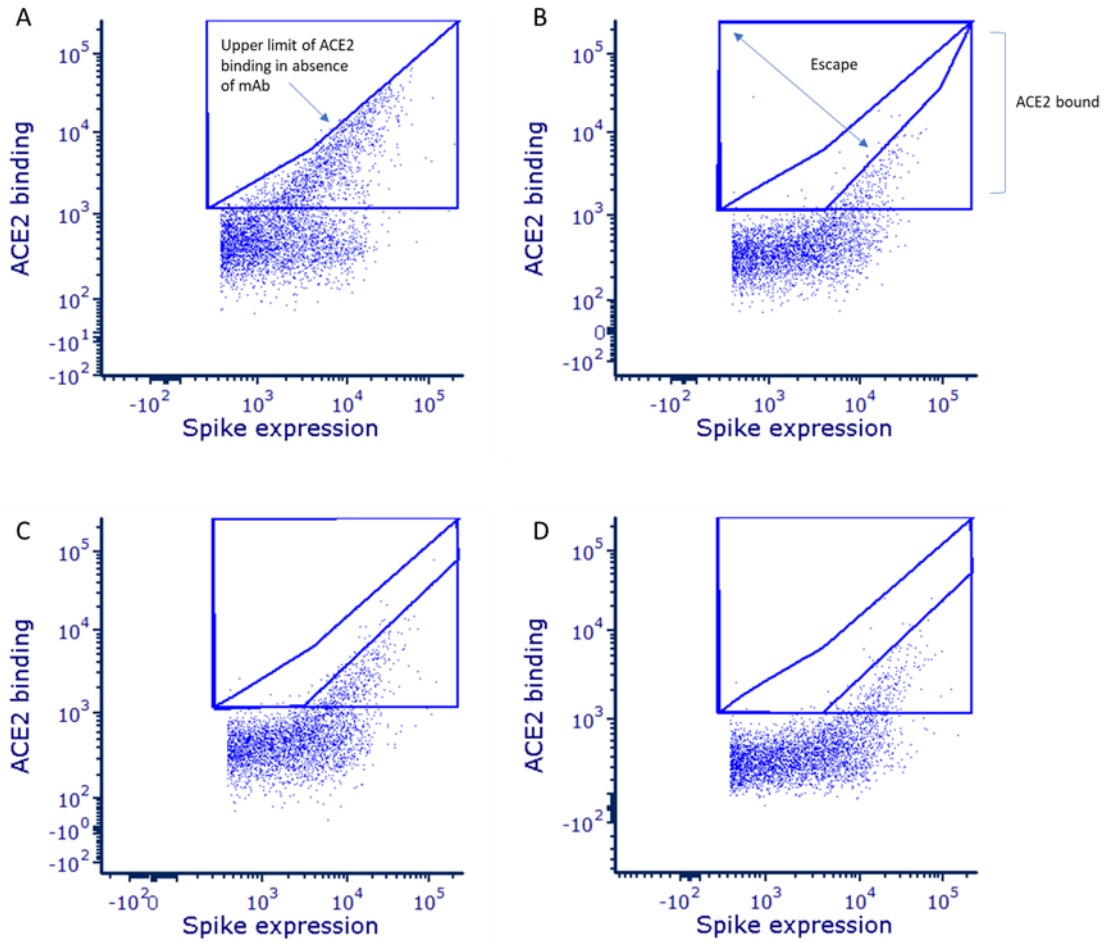


Figure 5-1: Dilutions of monoclonal antibodies used for sorts to screen for escape mutations. A dilution was chosen that did not completely block ACE2 binding by cells expressing the Alpha RBD spike library. HEK-293T cells were transfected with the Alpha RBD spike mutagenesis library tagged with mGreenLantern. 1ng of the Alpha RBD spike mutagenesis library plasmid was diluted in 1500ng of non-coding plasmid per 10^6 cells. 24 hours later spike expressing cells were incubated with the monoclonal antibody for 30 minutes, followed by incubation with sACE2-Fc-mScarlet for 30 minutes before analysis by flow cytometry. X axis represents spike expression, Y axis represents ACE2 binding. a) no monoclonal antibody, b) Ly-CoV-016 (400 ng/mL), c) REGN 10933 (80 ng/mL), d) REGN 10987 (160 ng/mL)

Sorts for each of the monoclonals were conducted using the concentration of antibody determined above (figure 5-1). The top 10% of ACE2 bound cells that express spike were sorted until a minimum of 10,000 cells were collected. Due to the fall in ACE2 signal with time as previously mentioned, gates were monitored throughout the sort and reset as required and the sort was batched with new samples after 2 hours. The cell pellets had their total RNA extracted, spike RNA was reversed transcribed using a gene specific primer, and the cDNA PCR amplified for NGS. Enrichment scores were calculated by dividing the proportion of the variant in the sort population by the proportion in the unselected plasmid library. The enrichment score was adjusted to give more weight to positions of the RBD having many amino acids escaping by multiplying the enrichment score by the fraction of other amino acids that are also enriched in the escape population (i.e. the highest fraction being 1, which would represent all other 19 amino acids being enriched to escape at that position in the RBD) to produce the adjusted enrichment score.

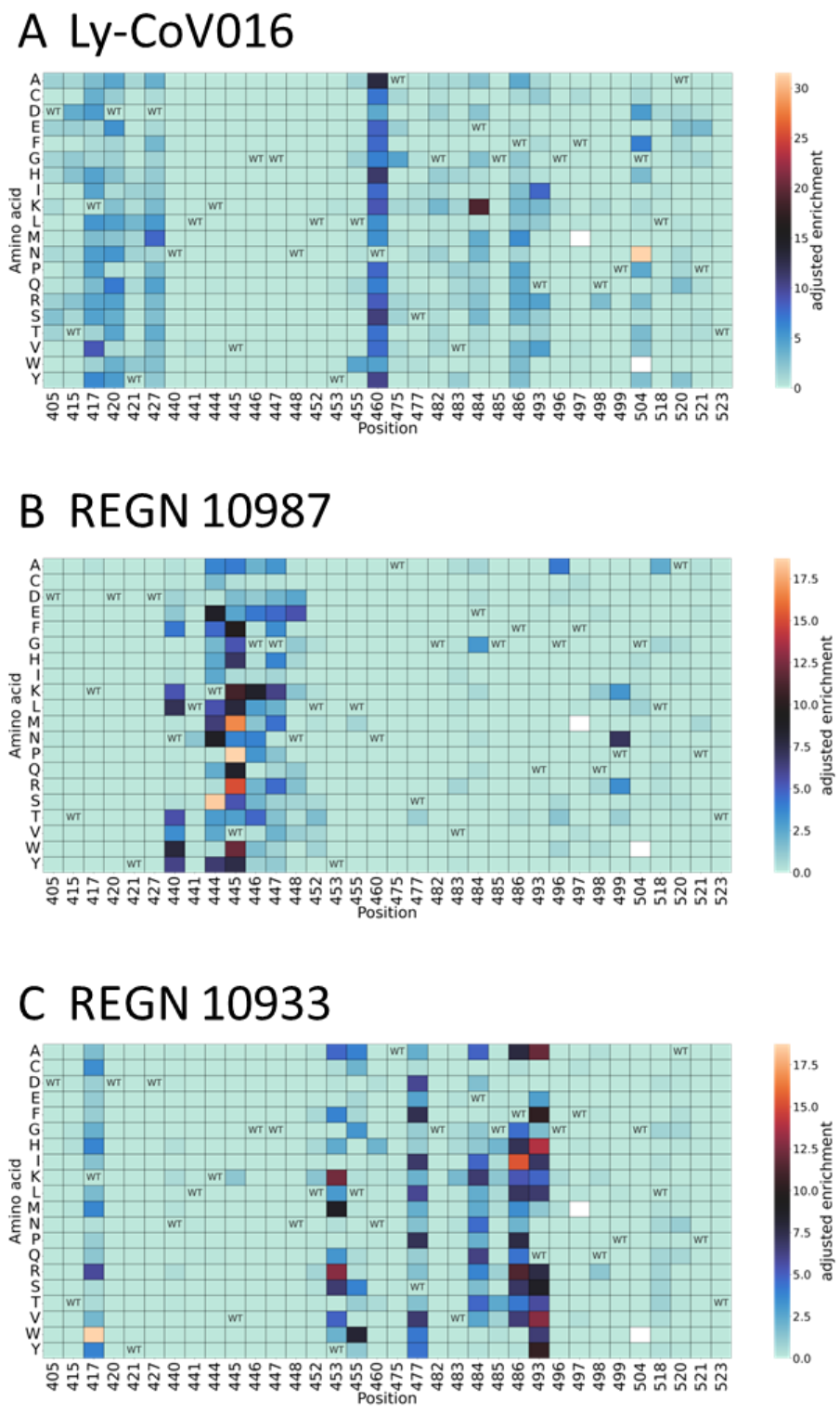


Figure 5-2: Monoclonal antibody escape heatmaps. Heatmaps showing the adjusted enrichment scores for mAb escape of RBD mutations in Alpha. RBD positions in the heatmap have been filtered to those with high enrichment scores. a) Ly-CoV016, b) REGN 10987, c) REGN 10933.

Figure 5-2 shows heatmaps of escape mutations for each monoclonal antibody. To validate the results of the screen, point mutations predicted to lead to escape from each of the monoclonals were chosen and added to a WT Alpha spike for use in pseudovirus neutralisation assays. The same panel of pseudoviruses were tested for neutralisation by all three monoclonal antibodies to show the specificity of the screen results. Most point mutations predicted by the screen to lead to escape caused complete escape from the respective mAb (figure 5-3), emphasising the relevance of these phenotypic maps based on libraries with single mutations, given single point mutations can have dramatic effects. The maps are specific, mutations not predicted to escape a particular mAb do not cause increases in the IC50. Even mutations at the same position are correctly predicted, for instance the escape heatmap for REGN 10933 shows K417W would be predicted to escape, but K417N would not. Pseudoviruses bearing K417W cause a greater than 100-fold increase in the IC50, whereas pseudoviruses bearing K417N do not show an increase in the IC50 (figure 5-3).

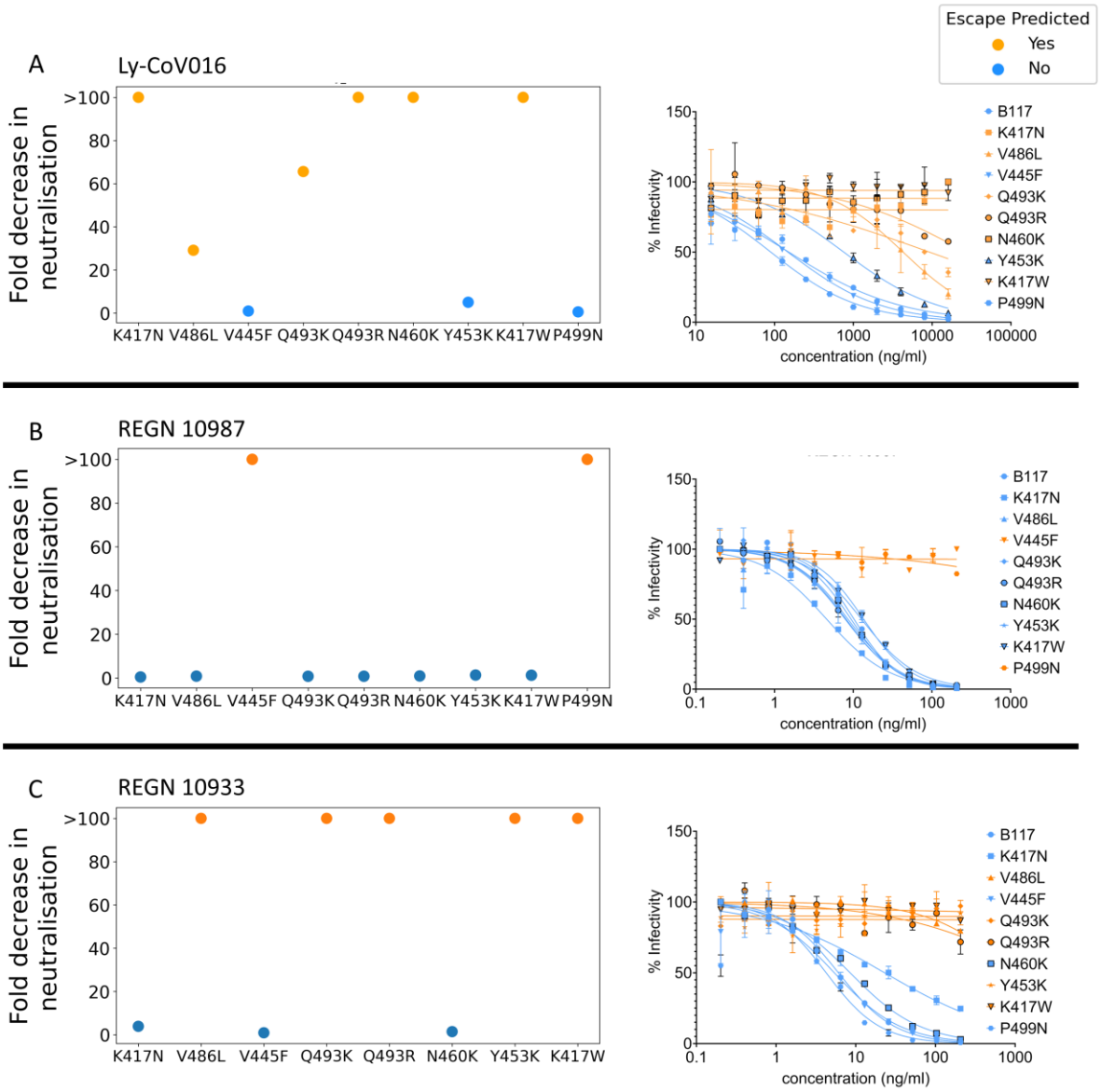


Figure 5-3: Effect of single mutations in the Alpha RBD on mAb neutralisation. The fold changes in neutralisation from WT Alpha pseudovirus are shown on the left, and the pseudovirus neutralisation assays are shown to the right. a) Ly-CoV016, b) REGN 10987, c) REGN 10933. Orange = a predicted escape mutation. Blue = not a predicted escape mutation,

1366

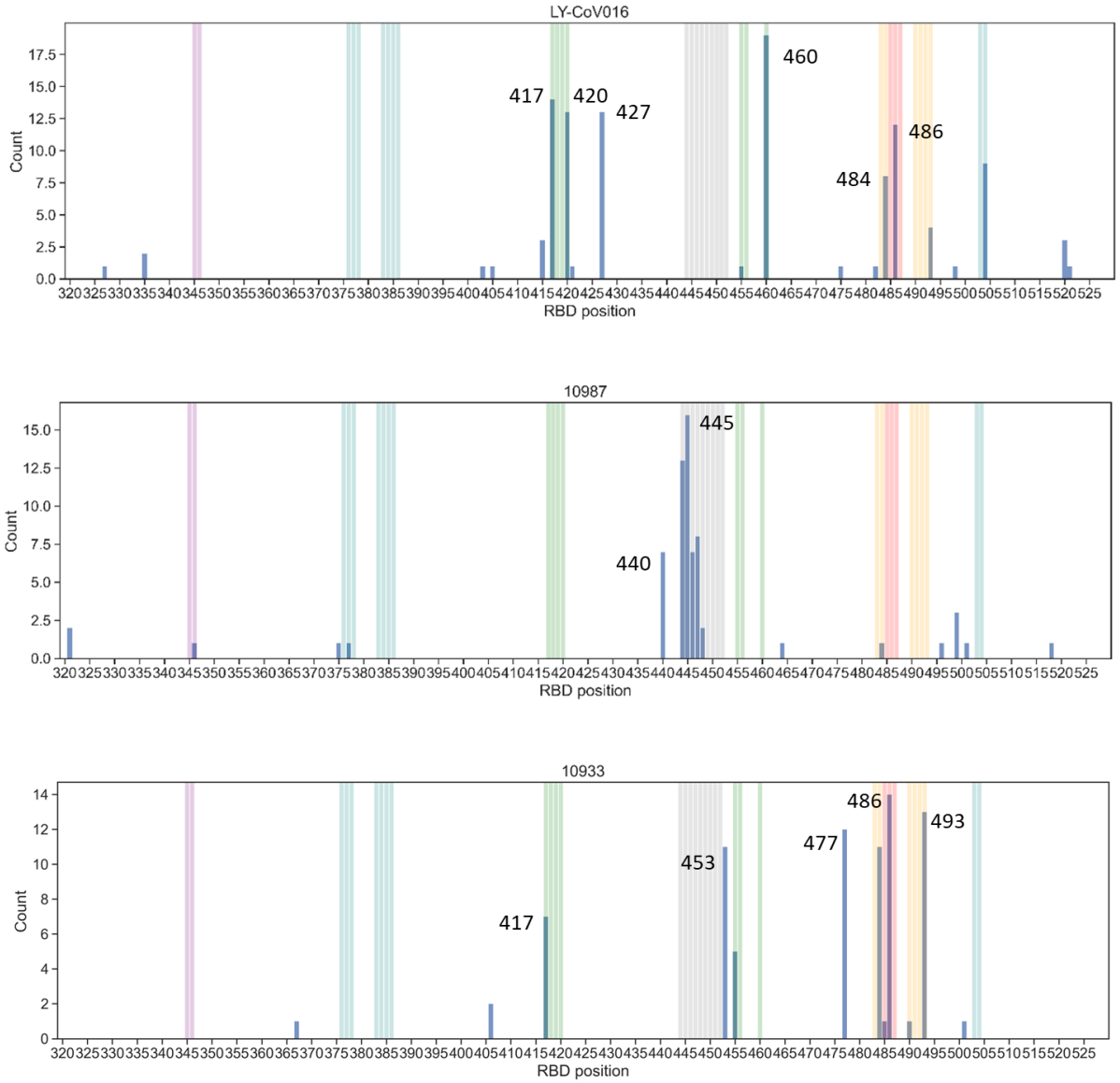


Figure 5-4: Amino acid positions in the RBD of Alpha SARS-CoV-2 spike where monoclonal antibody escape variants were selected from the RBD library for Ly-CoV016, REGN 10987, REGN 10933 monoclonal antibodies. Histogram showing the number of mutations at each position in the RBD that were selected for enrichment in mAb escape DMS experiments. a) Ly-CoV016, b) REGN 10987, c) REGN 10933. The coloured bars represent classes of mAbs (Cao, Wang et al. 2021): Class A = Green, Class B = Red, Class C = Orange, Class D = Gray, Class E = Purple, Class F = Blue.

1367
1368

Figure 5-4 plots the number of different substitutions that lead to escape at each position in the RBD for each respective monoclonal antibody. LY-CoV016 has an antibody binding footprint consisting of positions 417, 420, 427, 460, 484, 486, 493, and 504 in the RBD, corresponding to antigenic class A. REGN 10933 is focused on positions 417, 453, 455, 477, 484, 485, 486, and 493 (monoclonal antibody class B), while REGN 10987 is focused on a narrower footprint consisting of 440, 444, 445, 446, 447, and 499 (mAb class D). REGN 10933 and REGN 10987 were chosen to be part of the REGN-COV cocktail because they do not have overlapping footprints. These correlate with the predicted contact residues from structural maps (figure 5-5).

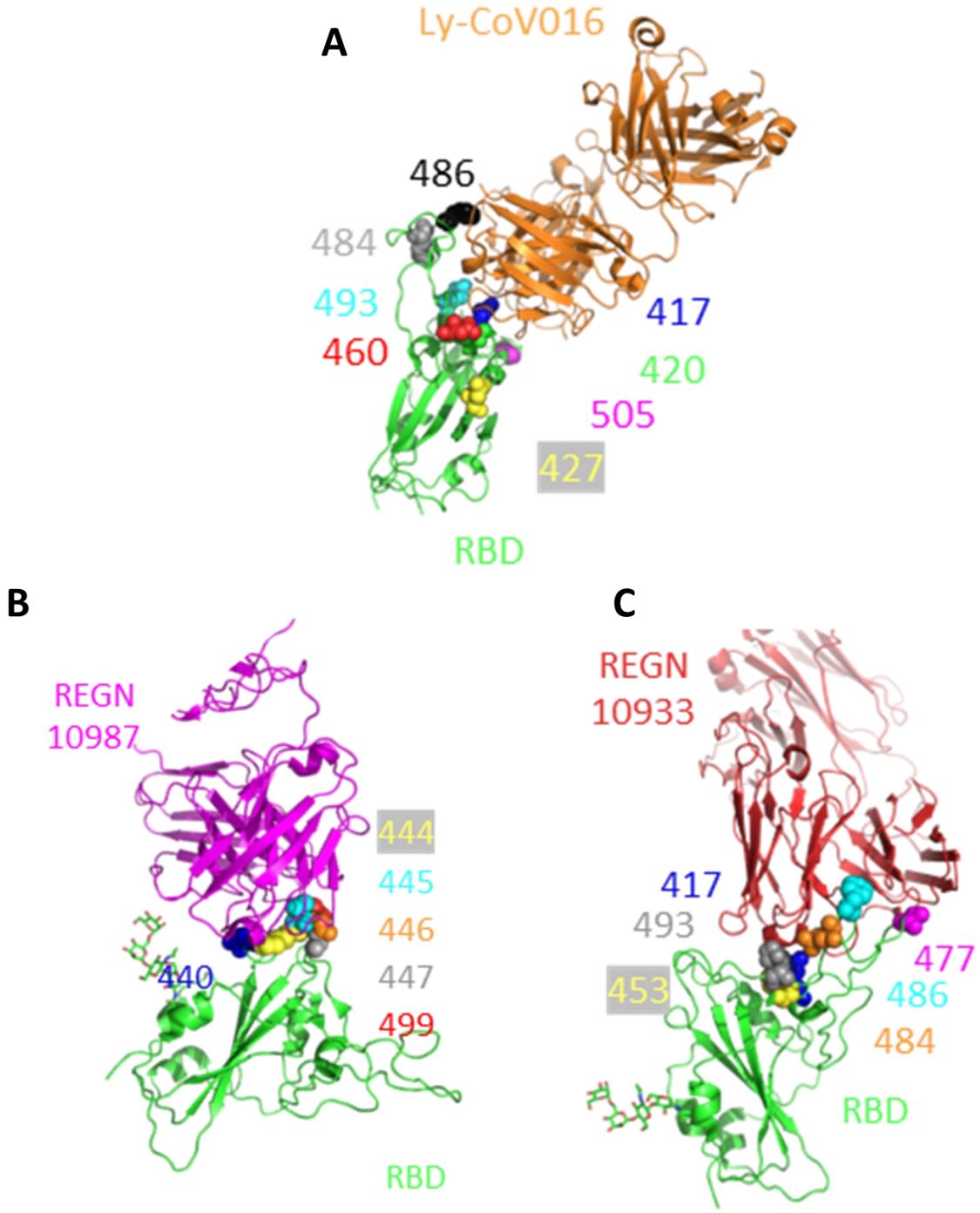


Figure 5-5: Structures of mAbs binding to the SARS-CoV-2 RBD. A) Ly-CoV016, B) REGN 10987, C) REGN 10933. PDB: A- 7C01(Shi, Shan et al. 2020), B, C - 6XDG (Hansen, Baum et al. 2020). Figures created using PyMOL (The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC.)

1379

1380

Monoclonal antibody escape maps have been previously produced using a yeast display of a monomeric Wuhan-RBD only library[109, 159]. A major difference in the escape maps presented here occurs at position 477 for REGN 10933. Using the mammalian cell whole trimeric Alpha spike display, we predicted position 477 to be an escape mutation for REGN 10933 (figure 5-2). In contrast there is no evidence of enrichment for escape at this position in the yeast RBD display platforms. This could be due to epistasis from using an Alpha spike that has already 501Y in the RBD or due to the use of whole spike, which presents the RBD in a different conformation due to allowing the "Up" and "Down" movements of the RBD to occur[29, 128].

To distinguish between epistasis and the effect of using whole spike on RBD antigenic presentation, the point mutations S477P and S477D were engineered to the Wuhan(D614G) or Alpha spikes and used in pseudovirus neutralisation assays. S477P caused at least a 20-fold increase in the IC50 in each spike backbone compared to the parental spike, while S477D caused even greater escape (figure 5-6A). This highlights the importance of using whole trimeric spike in these screens to produce a more physiologically accurate model. The difference in the effect of 477 was not due to epistasis, but rather due to using a whole trimeric spike display platform compared to a monomeric RBD yeast display platform. This may be related to different presentations of position 477 in the open and closed conformations of the RBD in whole spike, which are not recapitulated in monomeric RBD[29, 128].

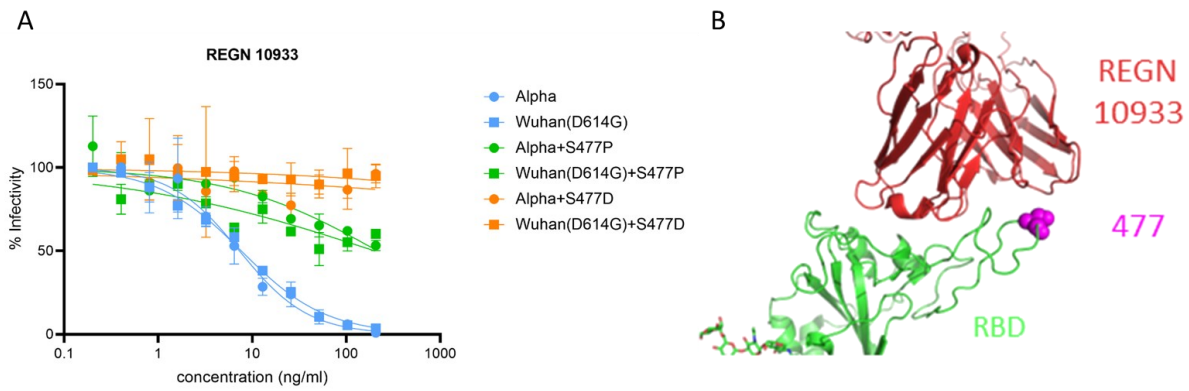


Figure 5-6: 477 is a position of escape for REGN 10933. A) 477D and 477P were added to Alpha and Wuhan(D614G) and pseudovirus neutralisation assays conducted using REGN 10933. B) Structure of REGN 10933 binding to the SARS-CoV-2 RBD, with position 477 highlighted. PDB: 6XDG(Hansen, Baum et al. 2020). Figures created using PyMOL (The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC.)

1402

1403

5.3 Discussion

We have validated the whole trimeric spike mammalian cell display DMS platform for the detection of antibody escape using mAbs and identified antibody binding footprints that correlate with structural data from each respective mAb binding to the RBD. The data produced largely agrees with that from yeast RBD DMS escape studies, except for the identification of position 477 as being important for mediating escape from REGN 10933. We have shown this was not due to epistasis in using an Alpha spike instead of using a Wuhan spike but is due to the differing methodologies. While the yeast RBD DMS for the most part is correct and offers high throughput screening, the platform is not the best physiological model and that should be borne in mind when interpreting data from yeast RBD DMS screens.

Having validated the ability of the platform to identify escape mutations from mAbs, the next chapter will focus on polyclonal sera from vaccinated individuals. Escape mutations from convalescent and vaccine induced immune responses will increasingly shape the direction of evolution of SARS-CoV-2 spike, as seropositivity increases and transmissibility becomes more dependent on immune evasion.

CHAPTER 6

6 PREDICTING ESCAPE FROM VACCINE SERA

6.1 Introduction

The ideal vaccine is one that can produce a lifelong response that is cross-reactive and maintains protection against any variant that should emerge. The measles vaccine (live attenuated) is a good example of a near ideal vaccine. It is given in the UK as part of a combined MMR(measles, mumps, rubella) live attenuated vaccine that requires 2 doses in childhood and has a vaccine efficacy of 96% in preventing measles, additionally the measles strain used in the vaccine has not need to be updated for 60 years[160]. The seasonal influenza vaccine represents the other end of the spectrum with highly variable vaccine efficacy, requiring annual boosters and regular updates to strain choice[161, 162]. Measles and influenza viruses have similar rates of mutation, and while the influenza haemagglutinin is more tolerant of mutations, the degree of tolerance does not account for this difference[163-165]. Measles has 2 co-dominant surface proteins haemagglutinin and fusion protein, in contrast influenza haemagglutinin is immunodominant over neuraminidase the other major surface protein of influenza. Measles haemagglutinin has 8 co-dominant epitopes, and mutations in at least 5 of these epitopes were required to cause a drop in neutralisation activity of measles antisera, which comes at a significant fitness cost (figure 6-1)[160]. Influenza haemagglutinin by contrast elicits a more focused immune response, with many individuals recognizing one or two dominant epitopes, allowing single mutations in these epitopes to have large effects on escape antibody neutralisation without significant fitness costs (figure 6-1)[86, 116, 166]. The focused immune responses raised by influenza haemagglutinin are

compounded by the effect of antigenic sin whereby repeated challenges from vaccination or infection preferentially activate cross reactive memory B cells, which tend to be in regions that are less or not neutralising, as the neutralising immunodominant epitopes evolve to escape antibody recognition[167].

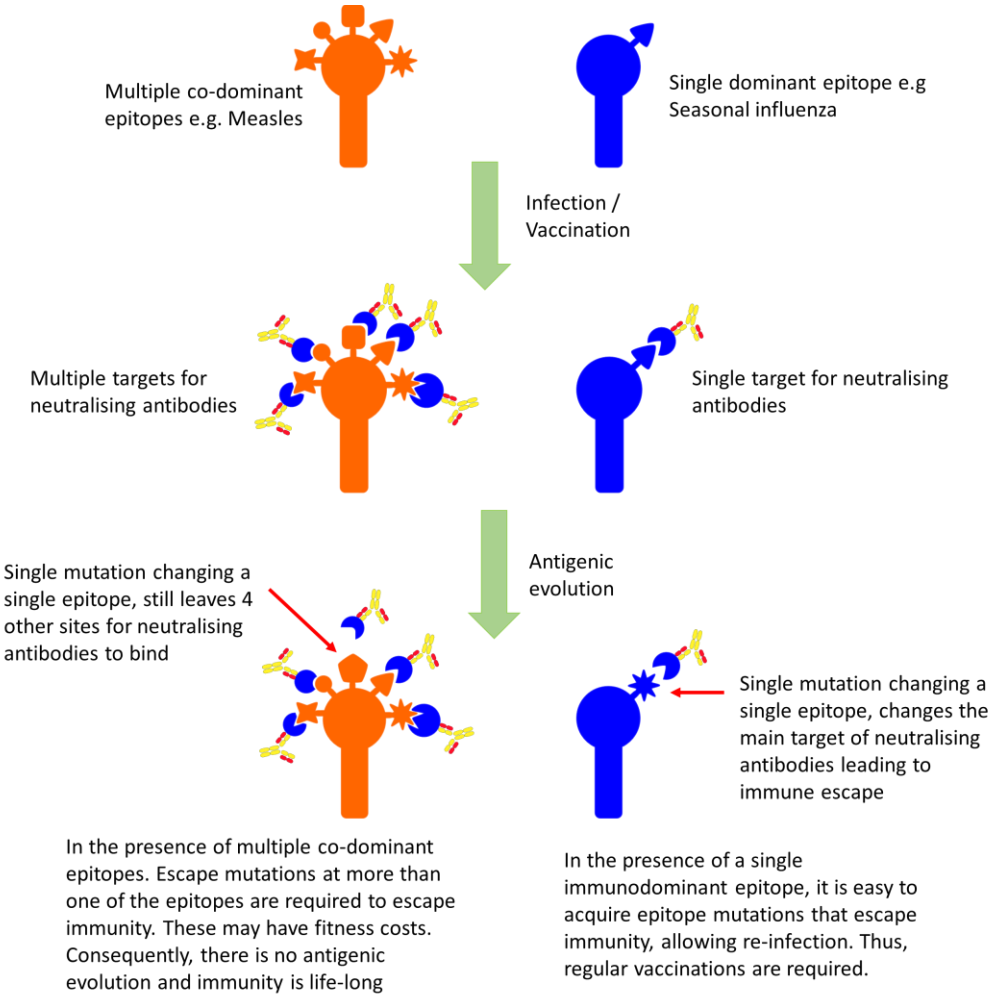
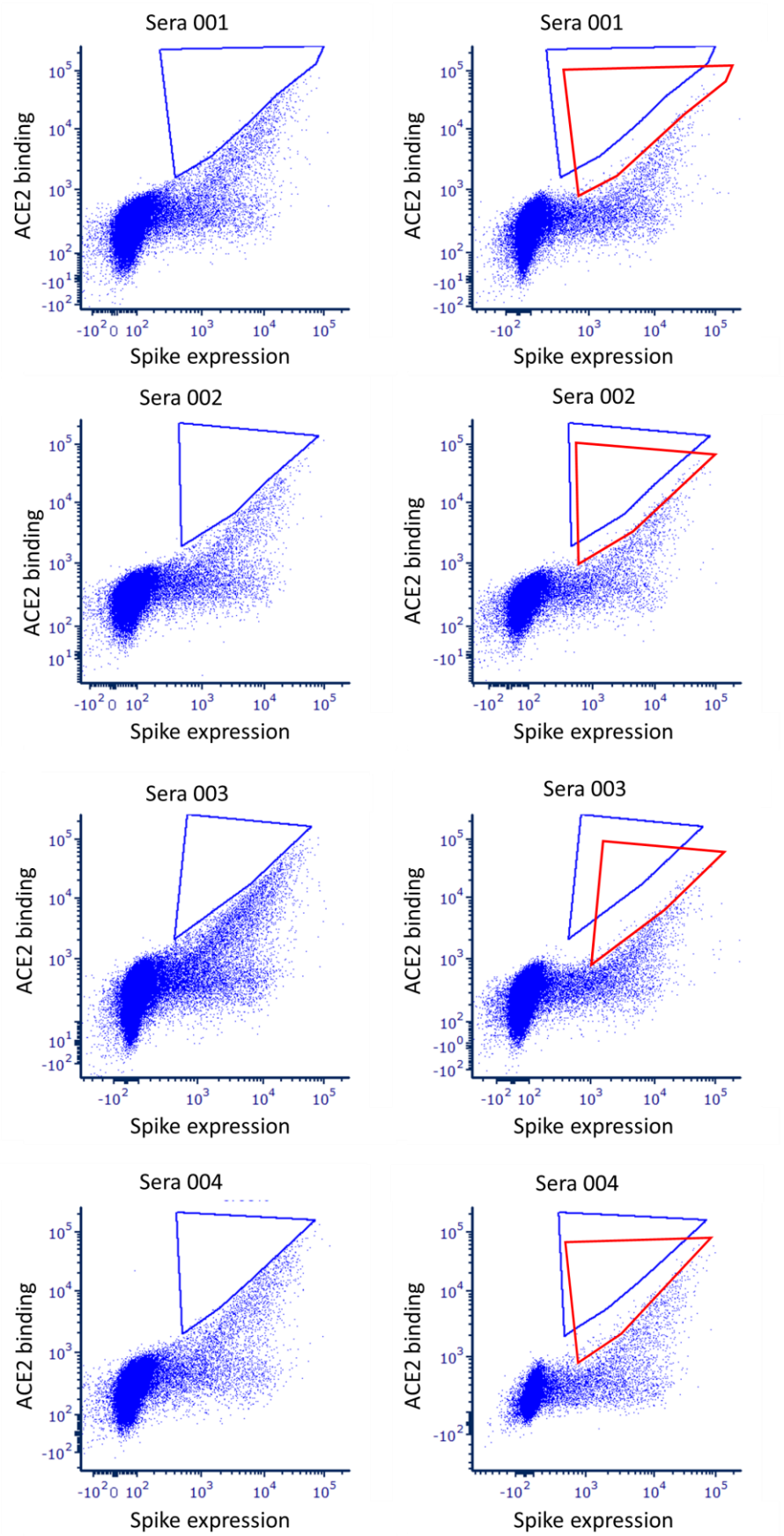


Figure 6-1: Evolving immune escape in a virus with multiple co-dominant neutralising epitopes versus a virus with a single dominant neutralising epitope. The orange virus has 5 co-dominant neutralising epitopes and requires escape mutations in more than one of these epitopes to escape from immunity, making immune escape more difficult to evolve. In contrast, the blue virus with a single immunodominant epitope only needs to evolve a single escape mutation to evade immune responses and cause a re-infection.

6.2 Are SARS-CoV-2 vaccines more like the measles vaccine or the seasonal influenza vaccines?

To explore this question further, sera from 8 double BNT162b2 vaccinated healthy adults was collected within 1-2 months of their second dose and used in a DMS with the trimeric Alpha spike RBD library platform to identify immunodominant epitopes and regions of escape.

Vaccine sera was titrated using flow cytometry to identify a dilution that would reduce the ability of HEK-293T cells expressing the spike library to reduce the sACE2-Fc-mScarlet signal by approximately half (figure 6-2). Details of the individuals who provided sera and the dilutions used for each serum can be found in table 6-1. Vaccine sera was incubated with HEK-293T cells expressing the Alpha spike library for 30 minutes, after washing, the cells were then incubated with sACE2-Fc-mScarlet for 30 mins before sorting by FACS. The top 10% of ACE2 bound cells that expressed spike were sorted, as this population should be enriched for variants with reduced neutralisation of ACE2 binding by the polyclonal vaccine sera.


Sera 001
ACE2 binding
Spike expression
Sera 001
ACE2 binding
Spike expression
Sera 002
ACE2 binding
Spike expression
Sera 002
ACE2 binding
Spike expression
Sera 003
ACE2 binding
Spike expression
Sera 003
ACE2 binding
Spike expression
Sera 004
ACE2 binding
Spike expression
Sera 004
ACE2 binding
Spike expression

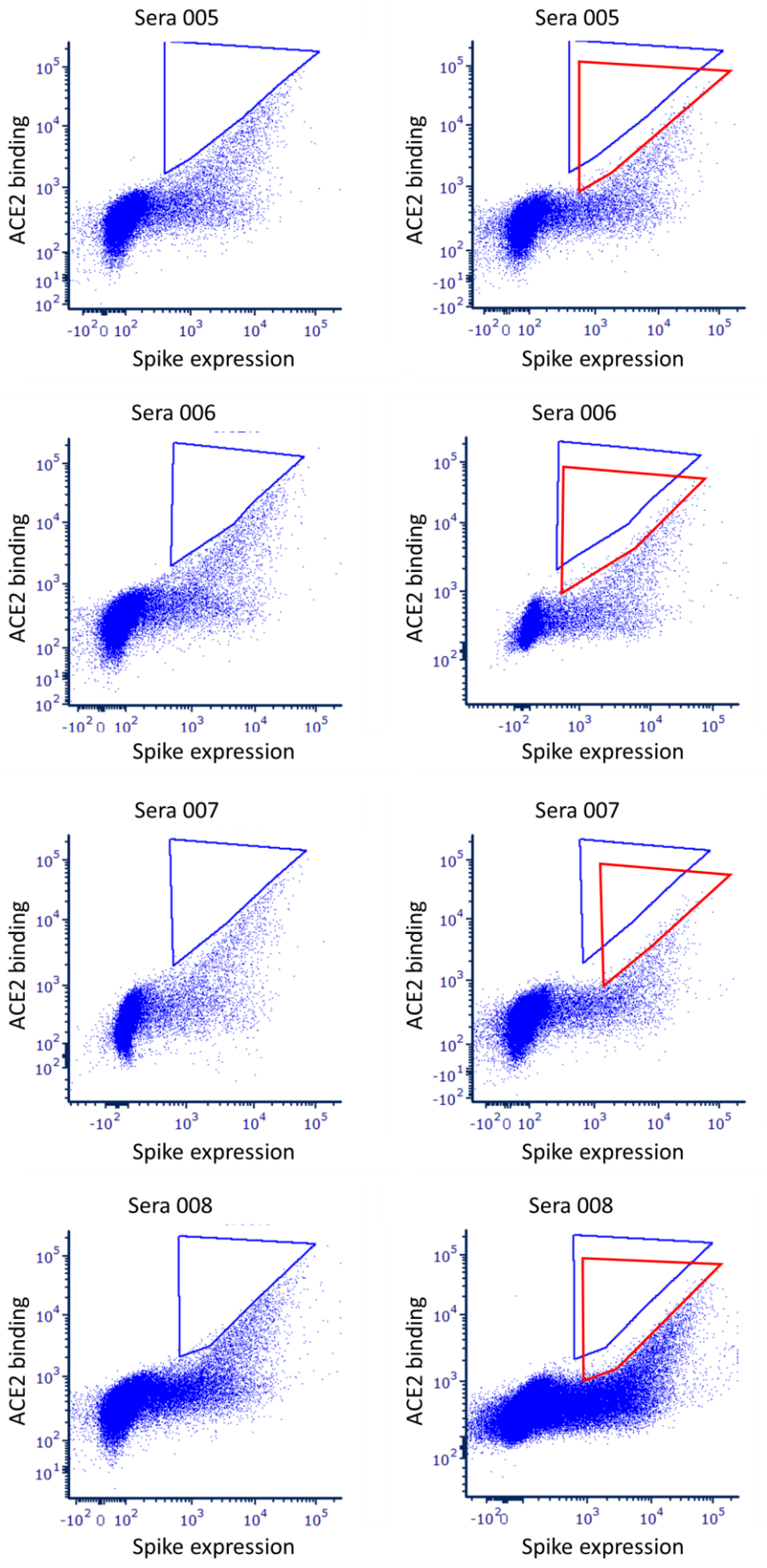

Sera 005
Sera 005
Sera 006
Sera 006
Sera 007
Sera 007
Sera 008
Sera 008
ACE2 binding
Spike expression

Figure 6-2: Dilutions of vaccine sera used for the vaccine escape sorts with the Alpha RBD spike library. HEK-293T cells were transfected with the Alpha spike RBD library, 1ng of Alpha spike RBD library tagged with mGreenLantern plasmid diluted in 1500ng of non-coding plasmid per 10^6 cells. 24 hours later spike expressing cells were incubated with the vaccine sera for 30 minutes, followed by incubation with sACE2-Fc-mScarlet for 30 minutes before analysis by flow cytometry. Sera titrations were chosen that reduce the sACE2-Fc-mScarlet signal by at least 50%. Lower edge of the blue polygon demarcates the upper limit of ACE2 binding by the Alpha RBD spike library in the absence of sera. The dot plots to the left of each pair represent ACE2 binding by HEK-293T cells expressing the Alpha spike library in the absence of sera. The dot plot to the right of each pair represents ACE2 binding by HEK-293T cells expressing the Alpha spike library in the presence of the sera dilution used for the DMS screen looking for escape showing the blockade of ACE2 binding by the respective sera. The blue gate is shown in the dot plot with sera to show the degree of ACE2 blockade by the sera on the spike library expressing HEK-293T cells. For sorts the same gate was moved down until the top 10% of the ACE2 binding population of spike expressing cells were gated, represented by the red gate. X axis represents spike expression measured by mGreenLantern, while the Y axis represents ACE2 binding measured by mScarlet. a) no sera, b) sera 001, c) sera 002, d) sera 003, e) sera 004, f) sera 005, g) sera 006, h) sera 007, i) sera 008

The collected cells were pelleted, and total RNA extracted, the RNA was processed as described before and sequenced by NGS. An adjusted enrichment ratio was calculated as for the monoclonal antibody escape screens (figure 5-2).

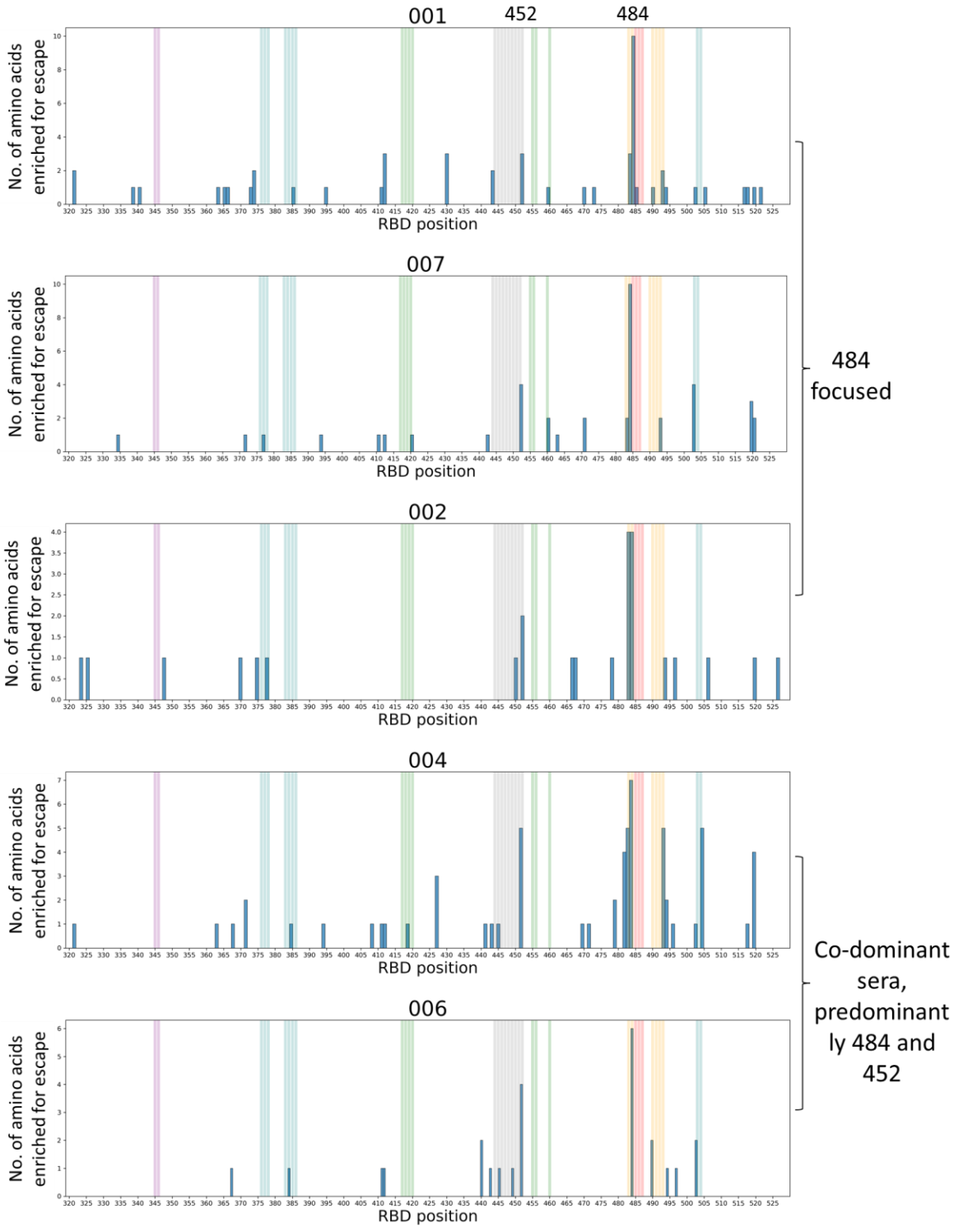
ID	Date bled	1st dose	2nd dose	Age	Gender	Dilution
001	27.04.21	20.1.21	29.3.21	30-40	Female	50
002	27.04.21	14.1.21	07.04.21	30-40	Female	100
003	28.04.21	20.1.21	29.3.21	20-30	Female	33
004	28.04.21	20.1.21	07.4.21	30-40	Male	200
005	04.05.21	20.1.21	07.4.21	30-40	Male	50
006	04.05.21	14.1.21	07.4.21	30-40	Male	294
007	24.05.21	14.1.21	07.4.21	30-40	Male	50
008	24.05.21	01.02.21	07.05.21	30-40	Male	33

Table 6-1: Participant demographics. The table provides dates of vaccinations with BNT162b2, bleeds, ages and gender of the people who donated sera used in this study. The dilution of sera used to incubate with the Alpha spike RBD mutagenesis library is shown.

6.3 Immune focusing on the RBD

To address how focused the antibody responses of vaccine sera are against the epitopes in the RBD of SARS-CoV-2 spike, histograms are shown presenting the number of mutations (maximum number 19, i.e., all amino acids other than WT are enriched for escape at this position) enriched for escape at each position in the RBD for each vaccine sera (figure 6-3).

Individual vaccine responses are heterogenous, with the majority being focused on position 484, sera 005 was predominantly focused on the core RBD, while sera 003 and 008 appeared to have poorly focused immune responses as evidenced by the number of positions with escape mutations (figure 6-3). Sera 001, 002 and 007 appear monospecific being focused on position 484, while sera 004 and 006 have 2 positions of immune dominance primarily 484, and secondarily 452 (figure 6-3). This pattern can be seen in the heatmaps showing the effects of individual mutations on vaccine escape for each serum, the majority of mutations being enriched for escape being found in positions 484 and 452 (figure 6-4).


001
452
484
No. of amino acids enriched for escape
RBD position
007
No. of amino acids enriched for escape
RBD position
002
No. of amino acids enriched for escape
RBD position
484 focused
004
No. of amino acids enriched for escape
RBD position
006
No. of amino acids enriched for escape
RBD position
Co-dominant sera, predominantly 484 and 452

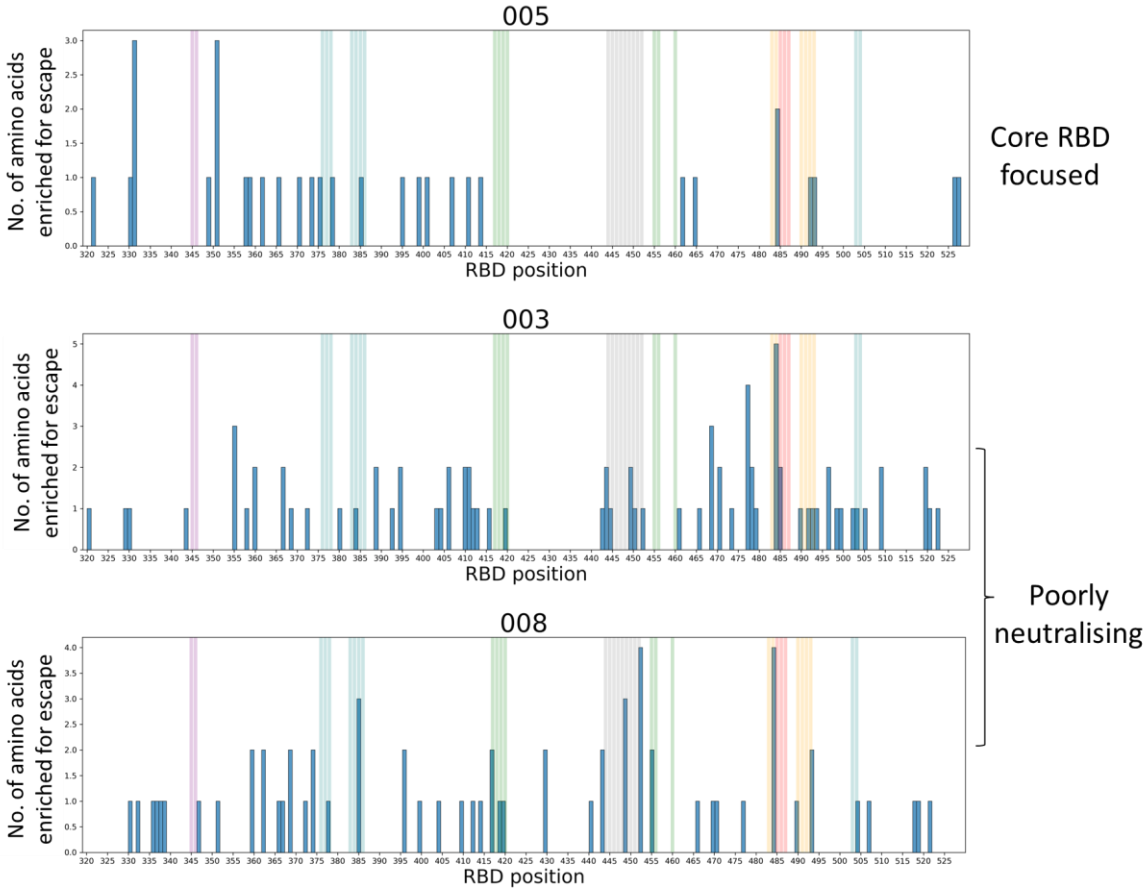
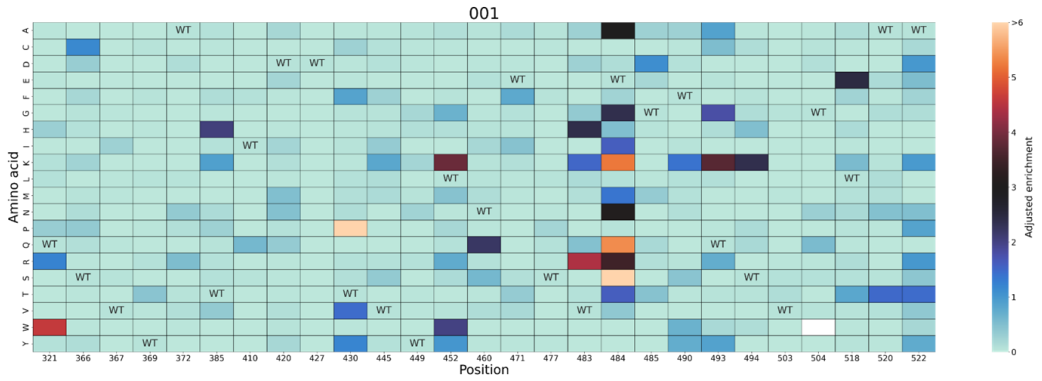
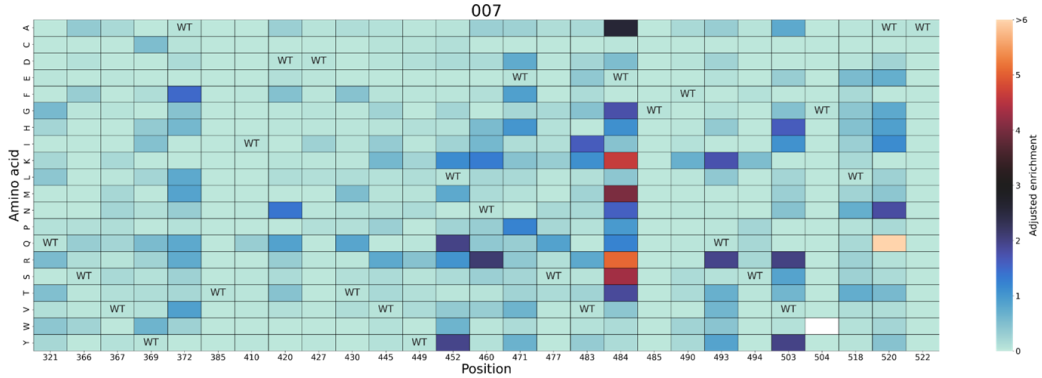


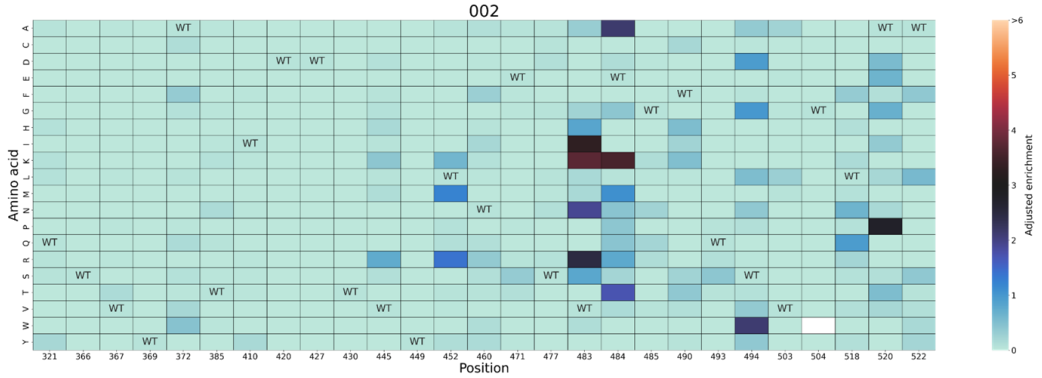
Figure 6-3: Double BNT162b2 vaccine sera is heterogenous and is dominated by antibody responses against a single site. Escape histograms for the 8 double BNT162b2 vaccine sera. The bars represent the number of amino acids at each RBD position that have an adjusted enrichment score greater than 1. Individual histograms are shown for each serum tested. The histograms have been grouped into arbitrary patterns by escape profiles. The coloured bars represent classes of mAbs (Cao, Wang et al. 2021): Class A = Green, Class B = Red, Class C = Orange, Class D = Gray, Class E = Purple, Class F = Blue.

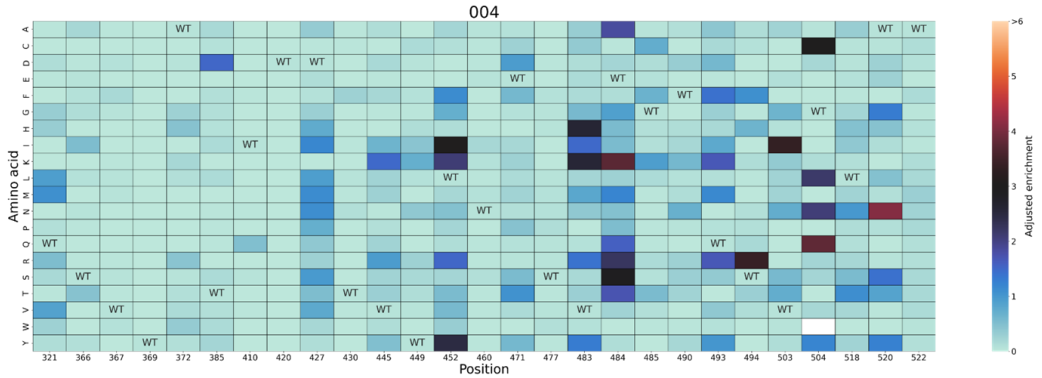
SARS-CoV-2 antigenic evolution occurs at a population level, while individual escape mutations may be initially transmitted from an individual, purifying selection occurs at a population level[168]. For an escape variant to be successful it must escape from many people's antibody responses rather than an individual. To better elucidate commonalities in immunodominance between the sera, the escape histograms were combined and the

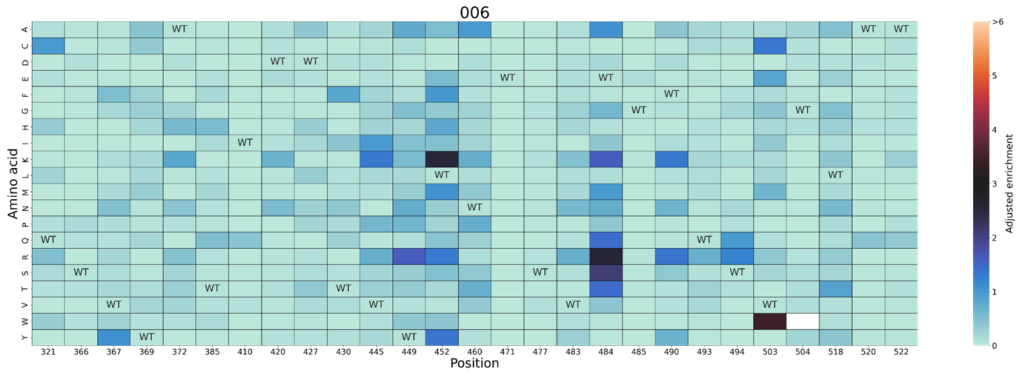
mutations that would be expected to mediate escape across the many the heatmaps were combined (figure 6-5).

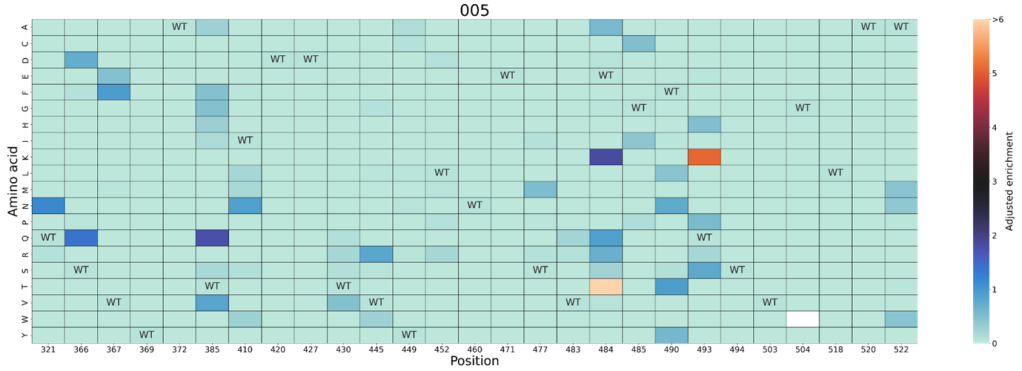

001
Amino acid
Position
Adjusted enrichment

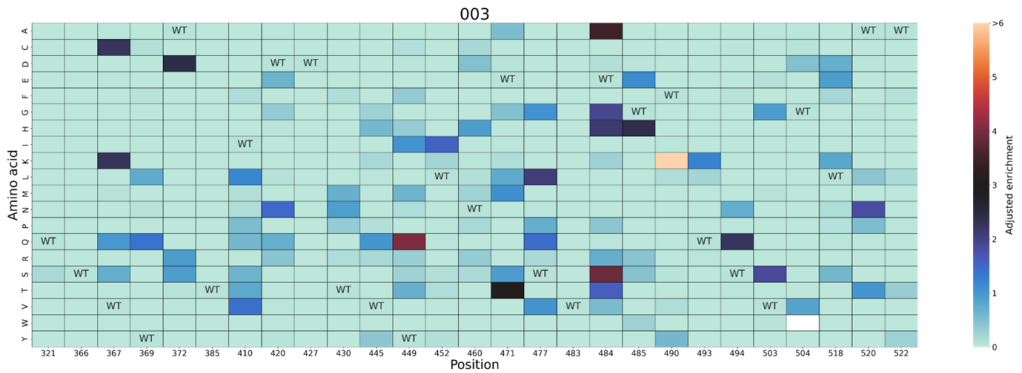

007
Amino acid
Position
Adjusted enrichment


002
Amino acid
Position
Adjusted enrichment


004
Amino acid
Position
Adjusted enrichment


006
Amino acid
Position
Adjusted enrichment


005
Amino acid
Position
Adjusted enrichment


003
Amino acid
Position
Adjusted enrichment

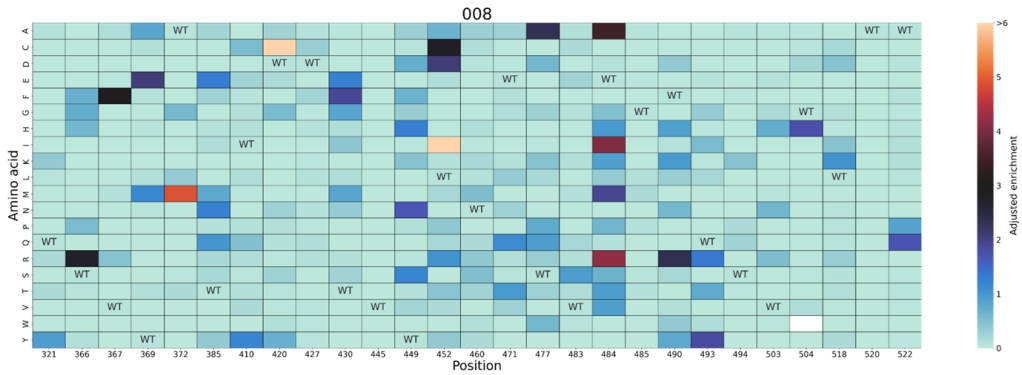

008
Amino acid
Position
Adjusted enrichment

Figure 6-4: Escape maps for each of the individual double dose BNT162b2 vaccine sera tested. The positions in the RBD shown are those having the highest frequency of amino acids across all the sera with adjusted enrichment scores greater than 1. Adjusted enrichment is the enrichment score multiplied by the fraction of amino acids other than WT that would be predicted to escape.

The number in each cell of the heatmap indicates how many of the 8 sera that amino acid was enriched for escape in the selection. The combined histogram shows that the immune responses are focused on 484 with position 452 being the next immunodominant position. The combined heatmaps show that double BNT162b2 vaccination produces an immune response against the RBD that is highly focused on position 484, as most amino acids at this position are enriched for escape and this position has the highest number of amino acid substitutions that are enriched for escape in over half of sera tested (9/19 amino acids). Positions of the RBD found in the ACE2 binding face are over-represented as having substitutions enriched for escape, and at these positions the positively charged amino acids (lysine and arginine) have a greater effect on escape. From the combined heatmap it can be seen positions 452, 484, 490 and 493 have substitutions that lead to escape from most sera (figure 6-5).

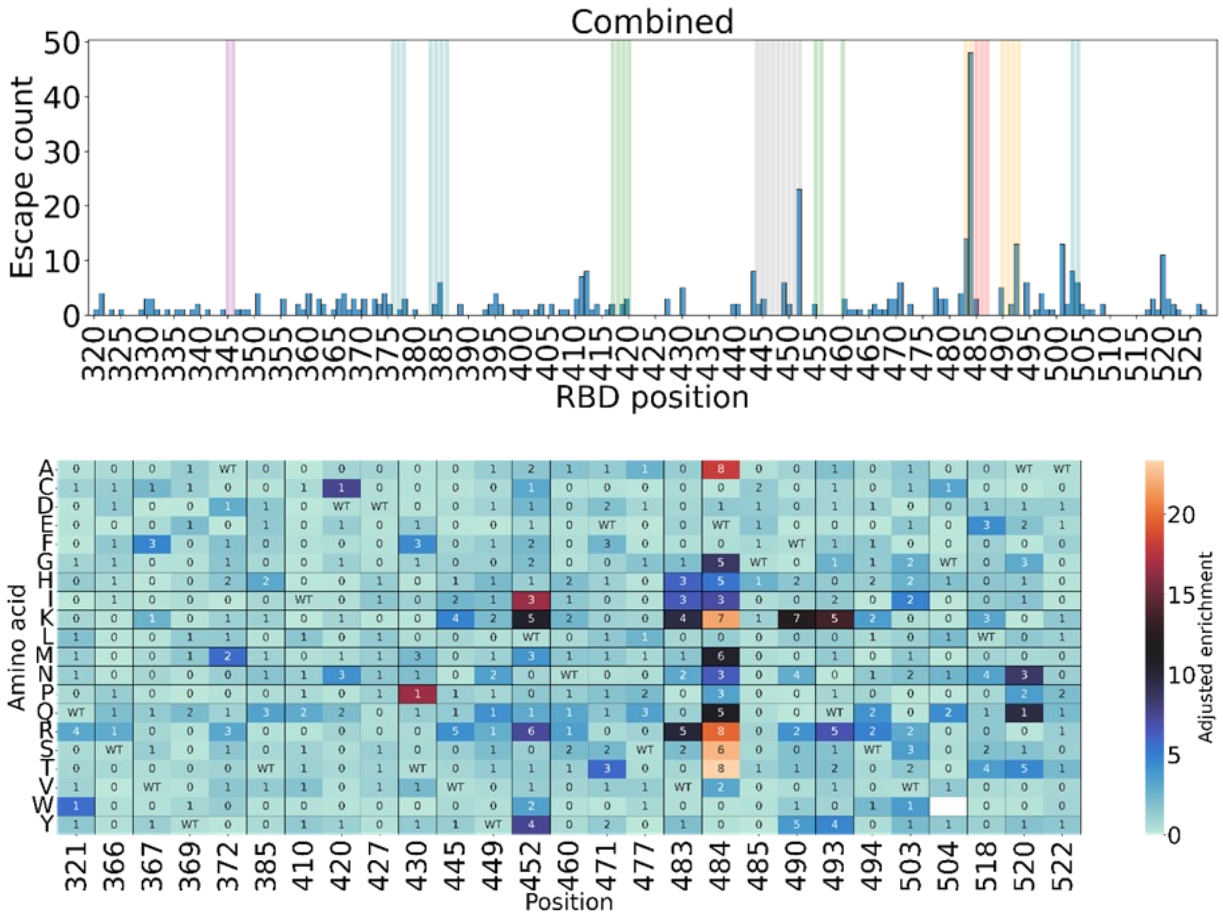


Figure 6-5: Top) Cumulative escape histogram from 8 double vaccinated individuals with BNT162b2. The bars represent the number of amino acids at each RBD position that have an adjusted enrichment score greater than 1, the maximum possible being 19 (19 other amino acids than WT) for an individual sera x 8 (summated across all 8 sera) = 152. The coloured bars represent classes of mAbs (Cao, Wang et al. 2021): Class A = Green, Class B = Red, Class C = Orange , Class D = Gray , Class E = Purple, Class F = Blue
Bottom) Cumulative escape heatmap from summating adjusted enrichment scores from 8 double vaccinated individuals with BNT162b2. Adjusted enrichment is the enrichment score multiplied by the fraction of amino acids other than WT that would be predicted to escape. RBD positions have been filtered to show positions with a higher frequency of amino acids that are predicted to escape. WT = wild type, blank squares = amino acids not represented in the library. The number in each cell represents the number of sera that the mutation had an adjusted enrichment score greater than 0.5.

1514

1515

6.4 Discussion

The RBD of SARS-CoV-2 spike is thought to be immunodominant and the target of most of the neutralising antibody responses[75, 77, 78]. Sera from individuals double vaccinated with BNT162b2 vaccine was mostly focused on one or two immunodominant positions in the RBD. Position 484 was the major immunodominant site in the RBD followed by 452.

Variants with mutations at position 484 have recurrently emerged in nature, being E484K in Beta[56], and Gamma[57], E484Q in Kappa[61] and E484A in the Omicron lineages[58, 82] and have been associated with vaccine escape. Following increasing population immunity from natural infection and vaccination, variants with mutations at 452 started to appear. Delta (B.1.617.2)[61, 169] was the most prevalent variant containing L452R, and this mutation was also retained in descendants of the B.1.617 lineage including Kappa (B.1.617.1) and B.1.617.3[61]. This mutation was key to the success of Epsilon in the US[60], while L452Q in Lambda contributed to the immune escape seen by this variant[170].

The highly focused nature of the vaccine immune response from this small cohort on the RBD suggest SARS-CoV-2 vaccines will behave more like influenza vaccines rather than measles vaccine, necessitating regular boosters rather than providing lifelong protection. Although the cohort used was small, a similar focusing of the immune response was seen in yeast DMS screens[77, 78].

Evidence of waning has been seen leading to wealthier countries instituting a 3rd vaccine booster dose after 6 months of the second dose[21-23] and some countries have already started recommending a fourth dose (second booster)[21, 171]. The rate of waning has been accelerated with the emergence of variants that are capable of escaping antibody neutralisation. The vaccine escape DMS screen uses ACE2 binding to select, the number of mutations enriched across the heatmaps highlights the plasticity of the RBD in maintaining ACE2 binding and the extent of the potential mutational landscape the SARS-CoV-2 RBD can explore.

When BA.1 emerged, it represented the most antigenically distant variant known. Phylogenetically, it was the most distant variant to have emerged having >30 mutations in spike[58]. The BA.1 RBD contained 14 mutations relative to the Alpha RBD[58], in the next chapter the contributions of those 14 mutations to antigenic escape will be explored in detail.

CHAPTER 7

7 THE IMPORTANCE OF RBD MUTATIONS IN BA.1 FOR IMMUNE ESCAPE FROM VACCINE SERA

7.1 Introduction

The RBD of BA.1 has 15 mutations relative to the RBD of Wuhan (figure 7-1). The Omicron lineages are thought to have evolved from chronic infections in immunocompromised hosts, as similar mutations and number of mutations have been found from serial sequencing of these persistently infected populations[89, 90, 92, 106]. Phylogenetically the Omicron lineages are so distant from other circulating variants that the number of mutations seen could not arise in that timespan during typical selection in human-to-human transmission due to the narrowness of the transmission bottleneck[172], by contrast replication within a immunocompromised hosts offers a broad bottleneck allowing many mutational directions to be explored and expanded, accelerating the rate of evolution.

	339	371	373	375	376	405	408	417	440	446	452	477	478	484	486	493	496	498	501	505
Wuhan	G	S	S	S	T	D	R	K	N	G	L	S	T	E	F	Q	G	Q	N	Y
Alpha	G	S	S	S	T	D	R	K	N	G	L	S	T	E	F	Q	G	Q	Y	Y
Delta	G	S	S	S	T	D	R	K	N	G	R	S	K	E	F	Q	G	Q	N	Y
BA.1	D	L	P	F	T	D	R	N	K	S	L	N	K	A	F	R	S	R	Y	H
BA.2	D	F	P	F	A	N	S	N	K	G	L	N	K	A	F	R	G	R	Y	H
BA.2.12.1	D	F	P	F	A	N	S	N	K	G	Q	N	K	A	F	R	G	R	Y	H
BA.4 / BA.5	D	F	P	F	A	N	S	N	K	G	R	N	K	A	V	Q	G	R	Y	H

Figure 7-1: Alignment of RBD differences between the major circulating SARS-CoV-2 variants. Shading of the residues corresponds to the variant that residue was first identified in, Wuhan = no shading, Alpha = grey, Delta = red, BA.1 = orange, BA.2 = green, BA.2.12.1 = blue, BA.4/BA.5 = purple.

7.2 Which mutations in the BA.1 RBD contribute to vaccine escape?

To understand which of the RBD mutations in BA.1 are contributing to the immune escape seen, the data from the combined vaccine escape DMS screens was filtered to the RBD positions found altered in BA.1 (figure 7-2).

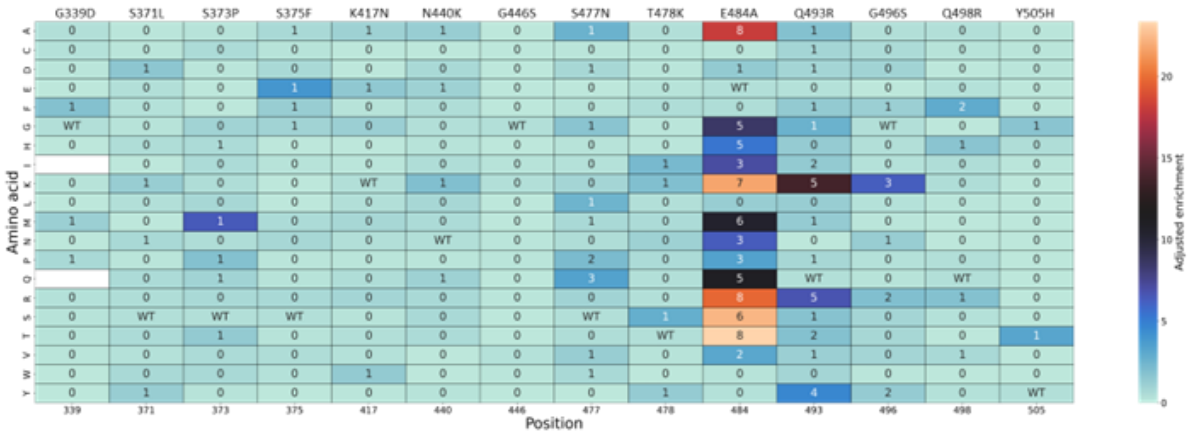


Figure 7-2: The cumulative double dose BNT162b2 vaccine escape map filtered to the positions mutated in BA.1 relative to Alpha to show the contributions of each mutation in BA.1 to vaccine escape. Adjusted enrichment is the enrichment score multiplied by the fraction of amino acids other than WT that would be predicted to escape. RBD positions have been filtered to show positions with a higher frequency of amino acids with adjusted enrichment scores greater than 0.5. The mutations on top of the heatmap represent the residues in BA.1. WT = wild type, blank squares = amino acids not represented in the library. The number in each cell represents the number sera (from 8 sera) that the mutation had an adjusted enrichment score greater than 0.5.

Only 2 of the mutations in the BA.1 RBD appear to be selected for escape in the combined vaccine sera escape heatmap, E484A and Q493R. To assess the contributions of these two mutations to vaccine escape, the RBD of BA.1 was replaced with a RBD containing E484A, Q493R alone and both E484A+Q493R on a Wuhan RBD with N501Y. The N501Y mutation was kept, because all the screens in the present study were done in Alpha spike, which has a N501Y RBD. N501Y has been repeatedly shown not to

lead to significant immune escape[173, 174] but does have a role in epistasis and increasing ACE2 binding[53, 138, 146].

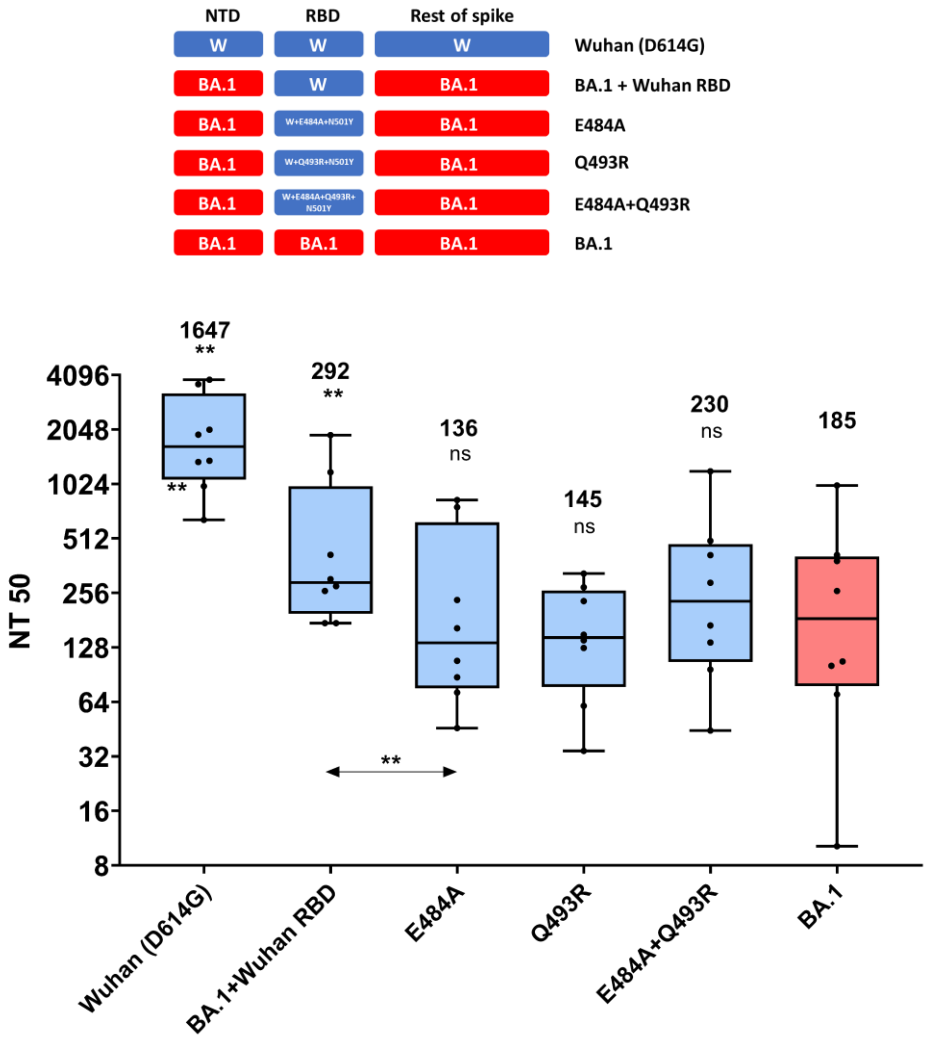


Figure 7-3: Single mutations in the RBD of BA.1 are responsible for the escape seen by BA.1's RBD. Top) Schematic of chimeric spikes used in the pseudovirus neutralisation assays. Bottom) Pseudovirus neutralisation assays using the double dose BNT162b2 vaccine sera against pseudovirus bearing BA.1 spike with a Wuhan RBD and combinations of mutations from the BA.1 RBD to identify their respective contributions to vaccine escape, as shown in the schematic. Median neutralisation titres are shown. *p value <0.05, ** p value <0.01, significantly different from BA.1's neutralisation titre (Wilcoxon matched-pairs sign rank test). ns = non – significant from BA.1's neutralisation titre.

The RBD containing just the two mutations E484A+Q493R in BA.1 was sufficient to cause the same degree of immune escape as BA.1 RBD (figure 7-3). In fact, either mutation alone was able to cause immune escape not significantly different from BA.1 WT. The presence of both mutations (E484A and Q493R) was not additive, suggesting the BA.1 RBD has redundancy in terms of immune escape (figure 7-3). Interestingly, replacing the BA.1 RBD with a Wuhan RBD in BA.1 spike only increased the median neutralisation titre by 1.6 fold (292/185) compared to the 9 fold (1647/1850 difference between Wuhan and BA.1 spikes (figure 7-3), which contrasts with previous studies suggesting that antibodies against the RBD account for 90% of neutralising activity[75, 77, 78].

7.3 Discussion

Despite the 15 mutations found in the RBD of BA.1, only 2 contribute to the immune escape seen from the vaccine sera tested (E484A and Q493R). Position 484 was the immunodominant position identified in the vaccine escape screen, while the heatmap assessing which mutations have the largest effects highlighted both E484A and Q493R as being selected for significant escape. Point mutation data from the DMS is helpful in making predictions regarding phenotype, but they do not account for epistasis. The presence of both E484A and Q493R mutations did not further increase immune escape, suggesting they both led to escape from the same class of antibodies.

The redundancy of E484A and Q493R is supported by events in nature, the Omicron lineages have further diversified and the new lineages BA.4/BA.5 are sweeping into dominance[82]. BA.4 and BA.5 have the revertant mutation R493Q, and the new mutation, L452R in the RBD[82]. The neutralisation experiments above suggest that Q493R could be lost without impact on immune escape and the combined vaccine escape data suggested that position 452 was the 2nd immunodominant site for vaccine sera, with L452R and L452K having the biggest effects on immune escape from this position. In keeping with this data, BA.4/BA.5 show a further 2-3-fold escape from vaccine sera compared to BA.1[175].

Studies suggest that 90% of the neutralising activity of convalescent and vaccine sera is due to antibodies against the RBD[75, 77, 78]. Despite this replacing the BA.1 RBD with a Wuhan RBD only leads to a 50% reduction in escape, suggesting domains outside of

the RBD play a role in determining escape from antibodies. The next chapter will explore the roles of domains outside of the RBD in immune escape.

CHAPTER 8

8 THE ROLE OF DOMAINS OUTSIDE OF THE RBD IN IMMUNE ESCAPE

8.1 Introduction

The S2 region of spike is the most highly conserved region between variants and between betacoronaviruses[176, 177]. Cross reactive immunity from SARS1 and seasonal betacoronaviruses are largely due to antibodies targeting this region[177]. The NTD directly abuts the RBD and is connected by a flexible linker that plays an important role in the transitions to the open and closed conformations of the RBD[30]. The NTD is a major target of the immune system being co-immunodominant with the RBD[76]. Antibodies binding to the NTD can be neutralising despite having no direct involvement in interfering with ACE2 binding[84, 108, 178]. Neutralising antibodies against the NTD target one major site called the antigenic supersite, structural changes to this site leads to escape from the majority of NTD targeting monoclonal antibodies[84, 108]. Support for a role in immune evasion comes from the finding the NTD has recurrently acquired deletions, substitutions and insertions[47, 83] that become fixed in new variants suggesting the actions of positive selection, with immune escape and transmission being the major selective pressures[179]. The NTD exhibits a plasticity that exceeds that seen in the RBD, which has largely evolved through substitutions[47, 101].

The 69/70 deletion in the NTD has featured in Alpha[179], BA.1[58], BA.4 and BA.5[180], del 144 in Alpha[181], del 143/145 in BA.1[58], del 157/158 in Delta[61], del 212 in BA.1[58], and del 241/243 has been found in Beta[56]. These deletions or

similar have also emerged in immunocompromised populations with prolonged infections of SARS-CoV-2[90, 92, 106].

The role of deletions and insertions of the NTD in immune escape from polyclonal sera was first reported by Andreano et al[94], who serially passaged SARS-CoV-2 in convalescent sera, until the virus was no longer neutralised by the sera. Sequencing of the virus revealed the acquisition of a deletion of F140 and the insertion of a 12 amino acid stretch at position 248, as well as the E484K substitution in the RBD. The two NTD changes were required to completely escape from the convalescent sera and each individually contributed to the same amount of escape as the E484K substitution in the RBD[94].

Despite evidence for a role of NTD mutations in immune escape, in convalescent and mRNA vaccine sera, most of the neutralisation from sera is due to antibodies against the RBD. In vaccine induced antibody responses the proportion of the neutralising response that is due to RBD directed antibodies (>90%) is higher than from convalescent sera[77, 78].

To understand the effect the NTD and other domains have on immune escape, pseudoviruses were created with chimeric spike proteins consisting of NTD and S2 domains swapped for ancestral B.1 domains (figure 8-1) and neutralisation assays conducted to assess the change in neutralisation titre.

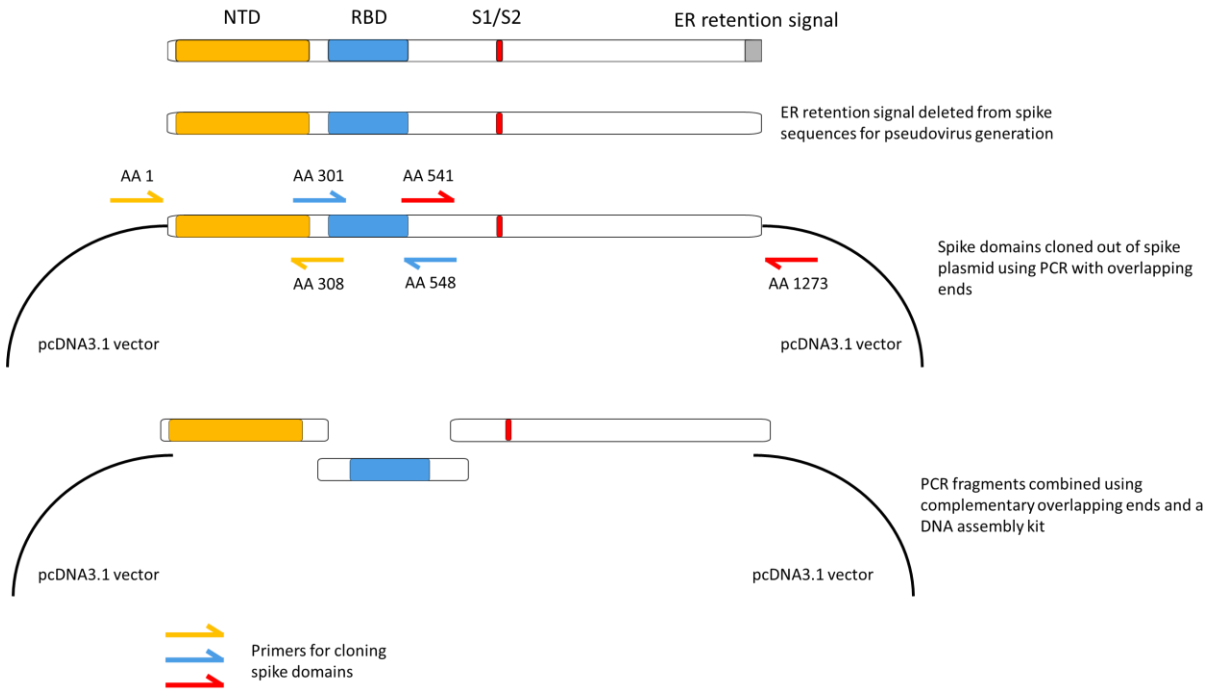


Figure 8-1 Schematic explaining construction of domain swap chimeric spike proteins. The ER retention signal was removed by introduction of a premature stop codon for SARS-CoV-2 spike sequences to be used for pseudovirus generation. PCR was used to clone out the spike domains to be used in the chimeric spike proteins The amino acid (AA) positions of the primers used for cloning are shown in the figure. The PCR fragments of the domains to be joined to create the chimeric spike proteins were designed with overlapping ends to allow assembly and cloning into the pcDNA3.1 vector using a DNA assembly kit.

1663

1664

8.2 The contribution of the Delta NTD versus the Delta RBD in immune escape

In April 2021, a sudden rise in the number of SARS-CoV-2 cases occurred in India[169]. This was caused by a new variant that was more transmissible than the previous Alpha variant[169, 182]. Delta represented the first instance where a variant swept to dominance globally having a phenotype conferring increased transmission (relative to Wuhan) as well as increased immune evasion[74, 183]. Delta was the most successful member of the B.1.617 lineage, the main subclades beings B.1.617.1 (Kappa), B.1.617.2 (Delta) and B.1.617.3[61]. Delta had the following mutations in the RBD relative to Wuhan, L452R and T478K and in the NTD relative to Wuhan, Delta had T19R, G142D, del 144, and del 157/158[61] (figure 8-2). Immune escape studies showed Delta caused significant escape from vaccine and convalescent sera [72-74, 184].

	19	24	25	26	27	67	69	70	95	142	143	144	145	156	157	158	211	212	213	214
Wuhan	T	L	P	P	A	A	H	V	T	G	V	Y	Y	E	F	R	N	L	V	R
Alpha	T	L	P	P	A	A	-	-	T	G	V	-	Y	E	F	R	N	L	V	R
Delta	R	L	P	P	A	A	H	V	T	D	V	Y	Y	G	-	-	N	L	V	R
BA.1	T	L	P	P	A	A	-	-	I	D	-	-	-	E	F	R	I	-	V	INS
BA.2	I	-	-	-	S	V	H	V	T	D	V	Y	Y	E	F	R	N	L	G	R
BA.2.12.1	I	-	-	-	S	V	H	V	T	D	V	Y	Y	E	F	R	N	L	G	R
BA.4 / BA.5	I	-	-	-	S	V	H	V	T	D	V	Y	Y	E	F	R	N	L	G	R

	339	371	373	375	376	405	408	417	440	446	452	477	478	484	486	493	496	498	501	505
Wuhan	G	S	S	S	T	D	R	K	N	G	L	S	T	E	F	Q	G	Q	N	Y
Alpha	G	S	S	S	T	D	R	K	N	G	L	S	T	E	F	Q	G	Q	Y	Y
Delta	G	S	S	S	T	D	R	K	N	G	R	S	K	E	F	Q	G	Q	N	Y
BA.1	D	L	P	F	T	D	R	N	K	S	L	N	K	A	F	R	S	R	Y	H
BA.2	D	F	P	F	A	N	S	N	K	G	L	N	K	A	F	R	G	R	Y	H
BA.2.12.1	D	F	P	F	A	N	S	N	K	G	Q	N	K	A	F	R	G	R	Y	H
BA.4 / BA.5	D	F	P	F	A	N	S	N	K	G	R	N	K	A	V	Q	G	R	Y	H

Figure 8-2: Alignment of NTD (top) and RBD (bottom) of major circulating SARS-CoV-2 variants. Spike residues shown are those that have varied among these variants that caused global sweeps. Shading of the residues corresponds to the variant that residue was first identified in, Wuhan = no shading, Alpha = grey, Delta = red, BA.1 = orange, BA.2 = green, BA.2.12.1 = blue, BA.4/BA.5 = purple.

From the combined vaccine escape heatmaps (figure 6-4), only L452R in the RBD would contribute to immune escape, and only modestly so and certainly not enough to account for the escape seen with Delta. Using chimeric pseudovirus (figure 8-3), the Delta RBD (L452R and T478K relative to Wuhan RBD) in a Wuhan spike, showed no significant change in neutralisation compared to Wuhan itself (figure 8-3). In contrast, the Delta RBD in Delta spike causes a 5.4-fold (1646/308) decrease in neutralisation and even the Wuhan RBD in Delta spike shows increased immune escape, a 2.4-fold (1646/693) decrease in neutralisation relative to Wuhan(D614G) (figure 8-3.).

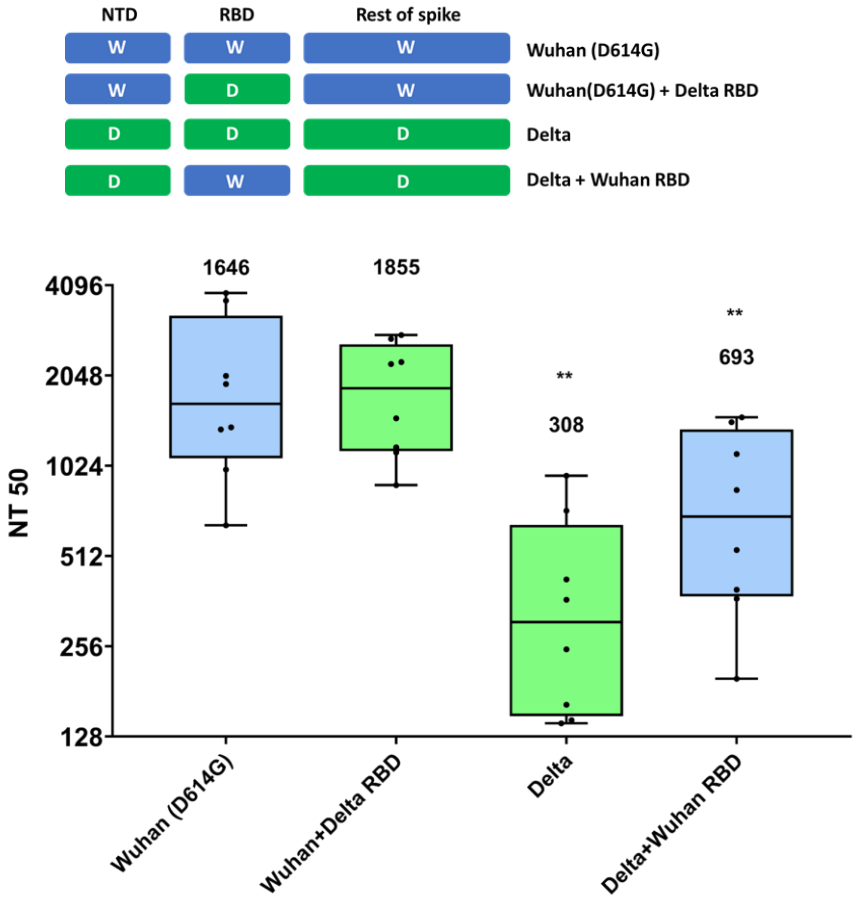


Figure 8-3: The Delta NTD plays an important in the escape seen from BNT162b2 vaccine sera. Top) Schematic of domain swaps used in pseudovirus neutralisation assays. Bottom) Pseudovirus neutralisation assays using the double dose BNT162b2 vaccine sera against pseudoviruses bearing chimeric spikes involving domain swaps between Delta spike and Wuhan(D614G) spike. Shown are the median neutralisation titres. Blue = contains a Wuhan RBD. Green = contains a Delta RBD. ** p value <0.01, significantly different from Wuhan neutralisation titre (Wilcoxon matched-pairs sign rank test). N=2

1687

1688

8.3 BA.1 domain swaps

Having shown that domains outside of the RBD in Delta strongly influenced immune escape, chimeric pseudoviruses were constructed with domain exchanges between BA.1 and Wuhan(D614G) to assess if the same observations were present in a different variant. BA.1 domains were used to replace equivalent domains in a Wuhan(D614G) spike.

The BA.1 RBD in a Wuhan(D614G) spike causes a 4-fold (1647/400) decrease in neutralisation titre of vaccine sera, while BA.1 WT pseudovirus causes an 8.9-fold (1647/185) decrease in titre relative to Wuhan(D614G) (figure 8-4).The Wuhan RBD in BA.1 led to a 5.6-fold (1647/292) decrease in neutralisation titre, while replacing the NTD of Wuhan with BA.1 NTD leads to a similar magnitude decrease in neutralisation titre (5.4-fold (1647/304)), suggesting the change in neutralisation of the Wuhan RBD in the chimera can be localised to the BA.1 NTD. In addition, the replacement of BA.1 NTD with Wuhan NTD in BA.1 spike, increased the neutralisation titre relative to BA.1 (the chimera is 5.6-fold (1031/185) easier to neutralise than WT BA.1), further supporting the role of the NTD in altering the neutralisation of spike (figure 8-4).

1706

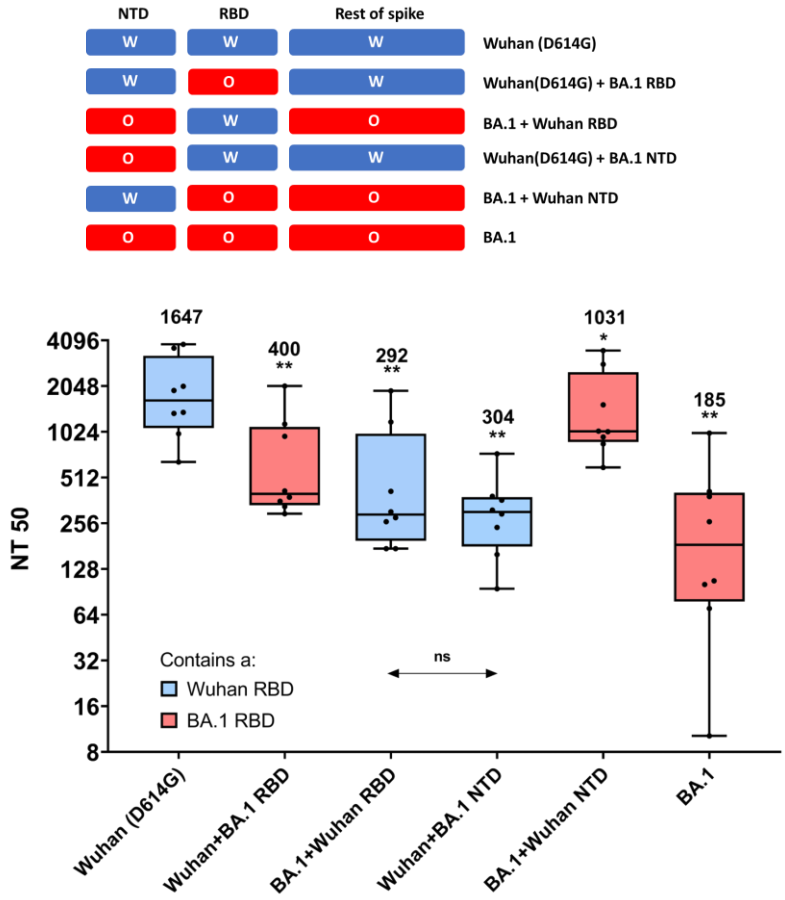


Figure 8-4: The BA.1 NTD reduces RBD neutralisation by vaccine sera. Top) Schematic of domain swaps used in pseudovirus neutralisation assays. Bottom) Pseudovirus neutralisation assays using the double dose BNT162b2 vaccine sera against pseudovirus bearing Wuhan, BA.1, and chimeric domain swaps between the Wuhan and BA.1, as shown in the schematic. Median neutralisation titre is shown. Blue = contains a Wuhan RBD. Red = contains a BA.1 RBD. *p value <0.05, ** p value <0.01, significantly different from Wuhan neutralisation titre (Wilcoxon matched-pairs sign rank test). ns = non – significant.

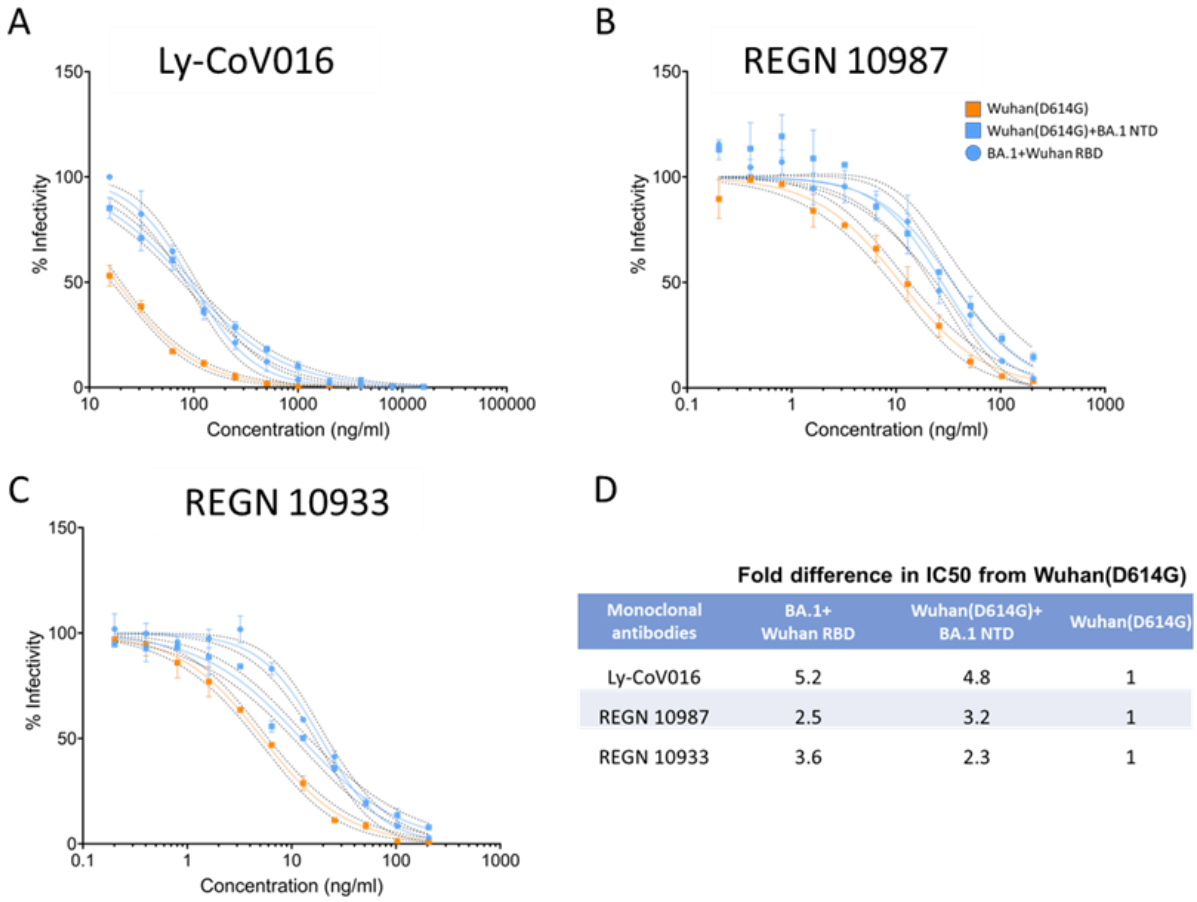
1707

1708

8.4 The effect of NTD domain swaps on neutralisation by RBD mAbs

Domain swaps involving the NTD of BA.1 and Delta appear to have a significant effect on the neutralisation titre of vaccine sera. To exclude the possibility the effects being seen are due to this sample of sera being particularly NTD focused, the chimeric pseudoviruses were used in neutralisation assays against mAbs targeting the RBD. If the effects of the NTD swaps against vaccine sera being seen were due to escape from NTD directed antibodies, then there should be no effect with the RBD directed mAbs.

The combination of a BA.1 NTD with a Wuhan RBD reduced neutralisation by RBD directed mAbs by 2.3 to 5.2-fold, with the largest difference being seen with LyCoV016 compared with Wuhan(D614G) spike (figure 8-5).



Fold difference in IC50 from Wuhan(D614G)

Monoclonal antibodies	BA.1+ Wuhan RBD	Wuhan(D614G)+ BA.1 NTD	Wuhan(D614G)
Ly-CoV016	5.2	4.8	1
REGN 10987	2.5	3.2	1
REGN 10933	3.6	2.3	1

Figure 8-5: Pseudovirus neutralisation assays using chimeric spikes featuring the Wuhan-RBD with different neighbouring domains against mAbs targeting the SARS-CoV-2-RBD. (A) Ly-CoV016, (B) REGN10987, (C) REGN10933. Dashed lines around each fitted curve represent the 95% confidence interval. (D). Summary of fold differences in IC50 relative to Wuhan(D614G) pseudovirus against the tested mAb. All pseudovirus neutralisation assays were conducted in duplicate.

1720

1721

8.5 The effect of the NTD on ACE2 binding

The NTD affects the neutralisation of the RBD by both mAbs and vaccine sera. ACE2 binding is another function that localises to the RBD. To explore if the NTD domain can influence ACE2 binding, chimeric spikes were expressed on mammalian cells and their ability to bind ACE2 measured by flow cytometry. BA.1 spike was chosen, as whole BA.1 spike shows the greatest binding difference compared to Wuhan (D614G) spike using flow cytometry. Replacing the BA.1 RBD with the Wuhan RBD on BA.1 whole spike leads to decrease in ACE2 binding, surprisingly replacing the NTD of BA.1 with Wuhan NTD, also leads to decrease in ACE2 binding (figure 8-6).

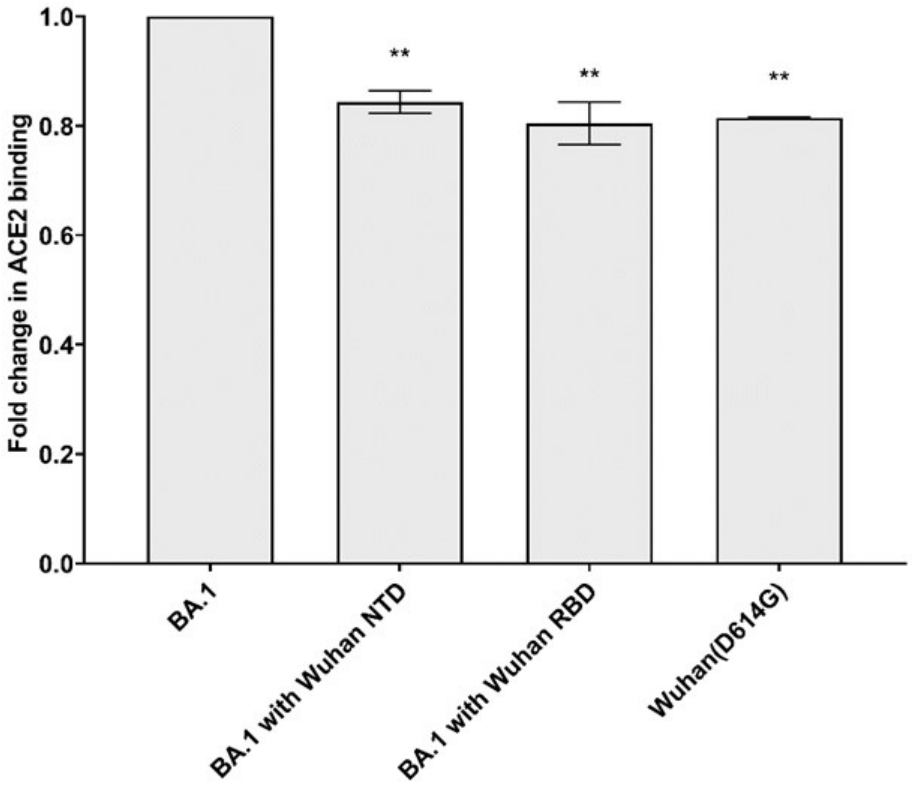


Figure 8-6: The effect of the BA.1 NTD on ACE2 binding by whole trimeric spike displayed on HEK-293T cells. HEK-293T cells were transfected with the respective plasmid, 24 hours later cells were dissociated and incubated with sACE2-Fc-IgG-mScarlet for 1 hour, before measuring the median fluorescence intensity (mfi) on the flow cytometer. The mfi was normalised to spike expression. Shown is the relative difference in mfi to BA.1. n = 2. ** p value < 0.001 using one-way ANOVA relative to BA.1.

The NTD domain has been shown to alter the efficiency of spike cleavage by host proteases[32, 33]. To confirm the phenotypic differences in neutralisation and ACE2 binding seen with the NTD domain swap chimeric SARS-CoV-2 spikes are not due to differences in spike incorporation into pseudovirus or spike processivity, a Western blot was carried out (figure 8-7). From the Western blot there was no difference in SARS-CoV-2 spike cleavage (figure 8-7b) and no difference in spike incorporation into pseudovirus particles as measured by the ratio of total spike to p24 (figure 8-7c).

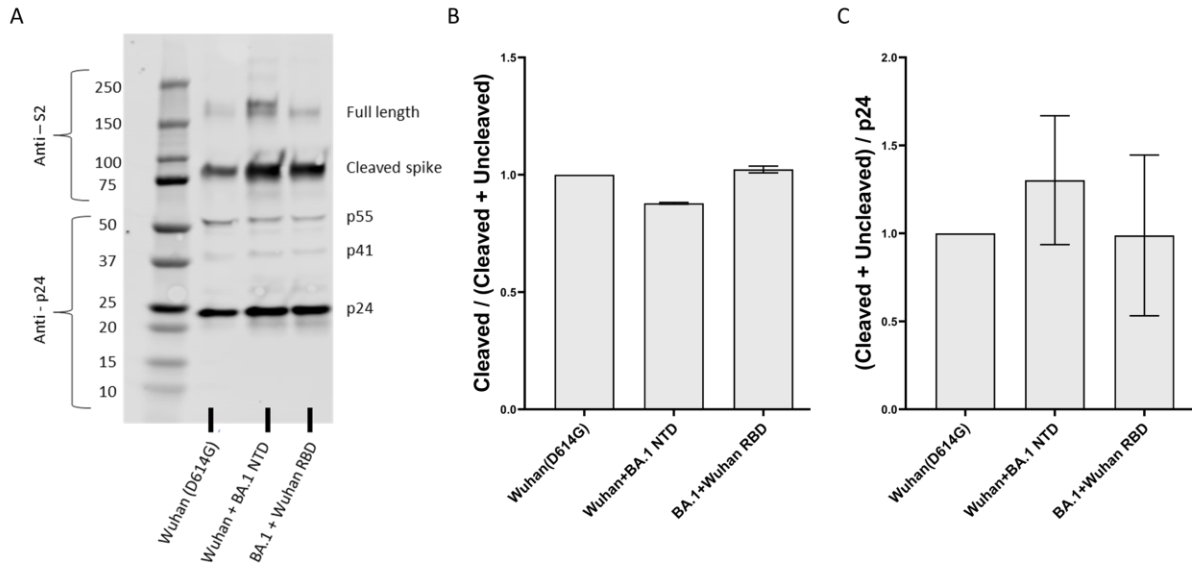


Figure 8-7:Differences in SARS-CoV-2 spike incorporation into pseudoviruses or spike processivity do not account for the phenotypic differences seen in the chimeric spike pseudoviruses. A) Western blot of concentrated pseudoviruses bearing Wuhan (D614G) and Wuhan/BA.1 domain swap spikes. Blots were stained for p24 and SARS-CoV-2 spike S2. B) Densitometry from Western blot in A showing ratio of cleaved to total spike. C) Densitometry from Western blot in C showing ratio of total spike to p24 antigen. Bars represent the range. n=2.

8.6 Discussion

Antibodies against the RBD account for the majority of neutralisation in both convalescent and vaccine sera[75, 77, 78], however domains outside of the RBD influence this neutralisation. In both Delta and BA.1, the NTD the RBD is paired with alters the neutralisation titre in chimeric spike bearing pseudoviruses. The Delta RBD without the Delta NTD shows no escape from vaccine sera, while the Delta NTD decreases neutralisation when associated with a Wuhan RBD and with a Delta RBD. The same pattern is seen in BA.1, except that the BA.1 RBD without the BA.1 NTD is still able to escape vaccine sera, although not to the same extent as when paired with the BA.1 NTD. Similarly, the BA.1 NTD reduces the neutralizability of the Wuhan RBD, when paired together in chimeric pseudoviruses.

There are 2 main hypotheses to explain the effects seen with the NTDs of Delta and BA.1. The first, mutations found in the NTDs directly lead to escape from neutralising antibodies, however it has been shown that the majority of neutralisation is due to antibodies against the RBD[77] and this is particularly the case for vaccine sera[78]. The second hypothesis is that the NTD is able to modulate the neutralizability of the RBD, i.e., inter-domain epistasis between the NTD and RBD.

In domain swap experiments, we have shown that the NTD domain can decrease the neutralisability of the RBD against RBD directed mAbs providing evidence for inter domain epistasis, which is further supported by evidence that the NTD domain can influence ACE2 binding.

The NTD is directly connected to the RBD in the same protomer by a flexible linker and in a trimer directly abuts the neighbouring protomer RBDs[30]. The mutations found in the Delta and BA.1 NTD reduce the neutralizability of the RBD, one mechanism through which this may be possible is by altering how the RBD is presented on whole spike. For instance, if certain epitopes are now less visible, one would expect the neutralisation fall. We have already seen that vaccine sera is predominantly focused on one or two regions on the RBD, alterations in the structural presentation of either region would be expected to have significant effects on neutralisation. Any modification of how the RBD is presented on whole spike might be expected to impact on ACE2 binding and we have seen that this is the case with the BA.1 NTD.

The recurrent emergence of NTD changes in SARS-CoV-2 variants[47, 83] and in chronic infections in immunocompromised hosts[89, 90, 92, 106] has largely been unexplained, suggestions of immune escape from NTD directed antibodies have been proposed, however NTD mediated antibodies contribute to a small proportion of neutralisation and would not be expected to provide sufficient selection pressure. The phenomenon of NTD mediated epistasis on RBD neutralisation and ACE2 binding provides a better explanation for why NTD changes are under such a selection pressure and appear so frequently.

A directed evolution experiment using live SARS-CoV-2 serially passaged in convalescent sera produce a mutant that was able to completely escape the convalescent sera[94]. The escape mutant had only a single RBD mutation (E484K) and required both

a NTD deletion and a NTD insertion, the NTD changes appeared first and accounted for larger drops in the neutralisation titre than the RBD substitution[94].

CHAPTER 9

9 DISCUSSION

9.1 Future SARS-CoV-2 spike evolution

SARS-CoV-2 spike has evolved continuously since its emergence to increase transmissibility[46]. Initially, mutations in spike were selected for that increased human ACE2 binding as the SARS-CoV-2 spike transitioned from a virus with zoonotic origins to an increasingly human adapted virus[8]. The intrinsic transmissibility of SARS-CoV-2 reached a plateau with Delta and the Omicron variants with further increases in transmissibility being limited by antibody responses within a mostly immune population[185, 186] (figure 9-1). Prior to the emergence of the Omicron lineages, novel SARS-CoV-2 variants battled for supremacy by exploring combinations of mutations in spike that offered the ideal balance between ACE2 binding and antibody evasion for that time period[46, 47, 187].

The shift towards immune evasion as being the key factor in increasing transmission is now evident in the remarkable convergence of the RBD mutations that escape antibody responses in the Omicron family[188]. The Omicron family has diversified remarkably into multiple lineages through continuous evolution and recombination. Despite this diversification, mutations at 346, 356, 444, 445, 450, 452, 460, 486 and 490 have been repeatedly and independently acquired by the multiple sub-lineages of the Omicron family[188]. Using mAbs derived from vaccinated individuals with BA.1, BA.2 and BA.5 breakthrough in DMS escape screens, it can be seen that the positions of

convergent evolution represent sites of escape from host antibody responses[188]. Each additional mutation progressively erodes away protective antibody responses.

Antigenic sites on the SARS-CoV-2 spike have been classified into clusters based on the position and nature of monoclonal antibody binding. The earliest attempt at classification, the Barnes classification used neutralising monoclonal antibodies and crystallography data of antibody binding to define 4 classes of antibody by how they bound to SARS-CoV-2 spike[189]. Contact residues for these 4 classes of neutralising monoclonal antibody were used to denote epitopes. Class 1 referred to antibodies that could bind to the SARS-CoV-2 RBD in the open conformation and blocked ACE2 binding. Class 2 referred to antibodies that bound to the RBD in both the open and closed conformations and blocked ACE2 binding. Class 3 referred to antibodies that bind outside of the ACE2 binding site and recognise both the open and closed conformations, and Class 4 antibodies that bind outside of the ACE2 binding site and only recognise the open conformation[189]. The Barnes classification was aimed at classifying different groups of neutralising monoclonal antibodies and early on showed how clustered the locations of neutralising antibody binding was on the RBD. More in depth antigenic mapping by Dejnirattisai et al.[80], identified 5 antigenic sites where neutralising monoclonal antibodies bound to the RBD corresponding to discrete regions of the RBD. In the most detailed antigenic mapping to date, Cao et al.[159] used yeast display and DMS to characterise the escape mutation profile of 247 neutralising monoclonal antibodies and define 6 antigenic sites on the RBD that neutralising monoclonal antibodies bind to.

Evidence has shown convalescent and vaccine sera can be considered as combinations of monoclonal antibodies to understand escape mutations and immune pressure driving antigenic changes[80, 159]. The most strongly selected escape mutations in the SARS-CoV-2 RBD are at position 484. Position 484 is an important escape mutation from Class C mAbs[159]. Of note position 493 also contributes to escape from Class C mAbs, providing an explanation for the lack of additional escape seen with the combination of E484A and Q493R in this study. From this study, two other positions strongly selected for escape from vaccine sera were position 483 and 490. Positions 483, 484, 490 and 493 are all important escape mutations for Class C mAbs[159]. The vaccine sera from our cohort strongly resembles the escape mutational profile for a Class C mAb, suggesting most of the ACE2 blocking from our vaccine sera and by proxy neutralisation activity came from Class C type antibodies. This highlights how focused antibody responses are against SARS-CoV-2[77, 78]. The most potent neutralising mAbs are Classes A-D, as these overlap the ACE2 binding site. Early Omicron lineages BA.1 and BA.2 had less complete escape from these other classes of mAbs. BA.1 contained the mutation K417N[58], which leads to some escape from Class A mAbs, but XBB.1 and BQ1.1 have since additionally gained N460K[190] leading to escape from a greater range of Class A mAbs. Class B mAbs are escaped incompletely by mutations at position 484 and 493 found in BA.1[159], mutations at position 486 are responsible for more complete escape from Class B mAbs[159] and first arose in BA.4/BA.5 and have since been maintained in their descendants[175]. Position 452 was the second immunodominant site identified in the vaccine cohort in this study. Position 452 leads to some escape from Class D mAbs[159], however mutations at positions 444 lead to more complete escape and have

been selected for in BQ.1[159, 190], while XBB.1 is trying to escape Class D mAbs through multiple less effective escape mutations at position 445 and 446[190]. The Omicron sublineages have acquired mutations that allow escape from the most neutralising classes of antibodies (A-D) and are now selecting for mutations leading to escape from the less potent neutralising antibody class E[159], seen with the selection of mutations at position 346 in XBB.1[190]. Selection for escape mutations of Class F antibodies is yet to be seen but may be the next site of antigenic drift as the Omicron lineages have weakened neutralisation by antibodies at other epitope sites.

The continual antigenic diversification of the Omicron linages represents a departure from the previous pattern of sudden and large jumps in antigenic distance and evolution, where the next strain to sweep to dominance would be genotypically distantly related to the previous dominant variant. Instead, there is the co-circulation of several sublineages that genotypically differ from a previous circulating variant by a few mutations[82, 175]. This situation bears the hallmarks of antigenic drift in seasonal influenza, where multiple co-circulating clades are found that are closely related to circulating strains from the previous year[168, 191-193].

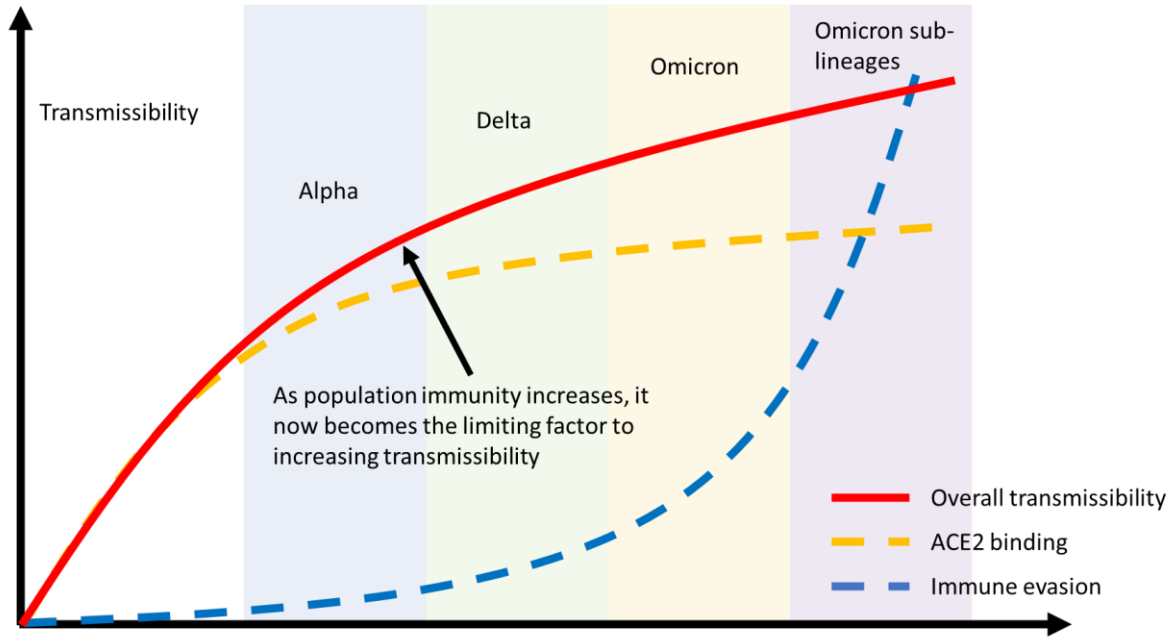


Figure 9-1: Schematic showing the balance between ACE2 binding and immune evasion on the overall transmissibility of SARS-CoV-2. Hypothetical graph showing the relationship between ACE2 binding and immune evasion and transmissibility of SARS-CoV-2. Transmissibility of SARS-CoV-2 shown on the y-axis, ACE2 binding and immune evasion represented on the x-axis. The emergences of Alpha, Delta and Omicron are placed on the graph. Alpha had a significant increase in ACE2 binding relative to Wuhan, but relatively little immune escape. Delta had increased ACE2 binding and immune evasion compared to Wuhan. Omicron had similar innate transmissibility to Delta but had greater transmissibility in an immune population. Subsequent Omicron sub-lineages show continued subtle increases in ACE2 binding, but transmissibility is mainly being driven by increasing antigenic distance.

9.2 Immune imprinting and SARS-CoV-2 antigenic evolution

Immune imprinting is an immunological phenomenon, whereby on repeated challenge with a related pathogen, the immune system preferentially boosts antibody responses against the first version of that pathogen it encountered[194]. It was first recognised in influenza in the 1960s, where it was noted a person's first influenza infection would alter the outcome of the next influenza infection with a different strain[195, 196]. Individuals whose first infection was more likely to be with a HI1N1 had a higher mortality rate

when during seasonal influenza seasons dominated by H3N2[197]. Further, it has been shown serologically that individuals imprinted with an early H3N2 infection possessed non-neutralising antibodies against modern day H3N2s despite individuals on average being infected with seasonal influenza every 4-5 years[168], in contrast younger individuals imprinted on a more contemporary and closely related H3N2 still possessed neutralising antibodies against the current seasonal H3N2s[198].

Immune imprinting makes antigenic drift possible, because it results in a suboptimal immune response to each subsequent virus. Most people in the world today with SARS-CoV-2 immunity will have a Wuhan based imprinted antibody repertoire due to prior infection or vaccination, since all vaccines still retain Wuhan spikes, and first doses are based on the original Wuhan immunogen. This gives good protection against infection with Wuhan and variants that are not antigenically distant such as Alpha[187, 199]. Similarly, imprinting after a primary Beta variant infection gives protection against reinfection with Beta and the antigenically close Gamma[187]. The Omicron lineages are antigenically distant to the antigens that have imprinted most people's immune systems. Breakthrough infections with Omicron lineages will preferentially back boost cross-reactive memory B cells that produce antibodies that recognise both the original imprinting SARS-CoV-2 spike and the new Omicron spike antigen[200-202]. These cross-reactive antibodies and the lack of stimulation of new B cells against the Omicron lineage spike produce a weakly neutralising antibody response. This allows antigenically closely related variants such as with the Omicron sublineages to escape breakthrough immunity from earlier members of the Omicron family and cause re-infection. Immune

imprinting has been well described for seasonal influenza[196, 203] and is being increasingly recognised for SARS-CoV-2[188, 194, 204].

Due to immune imprinting, we are likely to see a prolonged phase of antigenic drift, particularly as the SARS-CoV-2 spike still has mutational space to increase ACE2 binding, which can facilitate a greater range of immune evasive mutations by compensating for mutations that lead to escape but reduce ACE2 binding[146]. For a novel variant to replace the Omicron lineages, a large change in antigenic distance, akin to antigenic shift in influenza would be required. This would necessitate novel mutations in both the RBD and NTD, as well as antigenic distance such a variant could have a different virulence phenotype. Genetic surveillance should continue to monitor for increases in SARS-CoV-2 sequences of spike bearing novel mutations in the NTD and RBD of spike combined with some biological assessment of virulence or replication in different cell types[205].

9.3 Other applications for the mammalian display platform

Epistasis is the phenomenon where the effect of a mutation is dependent on the genetic background in which it arises. Epistasis shapes the evolution of proteins, as the SARS-CoV-2 spike continues to evolve and diversify further from Wuhan, DMS genotype to phenotype maps must be updated to remain relevant. This has been done with yeast DMS libraries using Alpha, Beta, Delta and Eta as the RBD base and highlights the constraints on evolution imposed by the genetic background[146].

The mammalian display platform developed here can be readily updated with libraries generated from novel SARS-CoV-2 spikes as they continue to evolve, providing updated phenotypic maps on the possible directions of spike evolution. A limitation of using DMS to forecast evolution is the inability to model recombination. Recombination in coronaviruses is an important contributor to the evolution of coronaviruses[206]. Both SARS-CoV1 and SARS-CoV-2 arose from recombination from bat coronaviruses[207]. Recombination occurs at a relatively high frequency in coronaviruses, being reported in murine hepatitis virus to be as high as 25%[206]. Evidence for recombination events in SARS-CoV-2 has been found from genetic surveillance[208, 209] and is responsible for the recently emerged VOC XBB.1[210].

The potential to cross species barriers from animals to humans and the reverse can be investigated using the platform. An initial barrier to crossing a species barrier is host receptor binding. The number and nature of mutations required to increase binding to another species' receptor can be explored using DMS[14] and can form part of pandemic risk surveillance. This is particularly pertinent for the many bat coronaviruses currently in circulation but is also of use in understanding the pandemic risk from highly pathogenic avian influenza (HPAI), whose transmission in humans is often limited by a reduced ability to bind α2,6 sialic acids in the upper airways[211, 212].

During the SARS-CoV-2 pandemic we have seen cross-over from humans into other species, namely mink[213] and white-tailed deer[214]. Circulating within these animal populations SARS-CoV-2 acquired adaptive mutations in spike[214-216]. Reverse

zoonosis from mink adapted SARS-CoV-2 led to small clusters of infections[213]. Ongoing spike adaptation within other species has the potential to create antigenically novel variants that have the potential to re-enter human circulation[213, 215]. Using DMS can assess the fitness impact these animal adaptive mutations on spike have on human ACE2 binding and the subsequent potential for reverse zoonosis.

Exploring the number of mutations away from being able to use another species' ACE2 or alternative host receptor can be used to assess the pandemic potential of animal coronaviruses and create informed surveillance systems. The phenotypic effect of mutations on receptor binding can be used to infer and recreate evolutionary models to identify ancestral strains of viruses, as the sequence of mutations acquired will be shaped by fitness costs and epistasis[14].

Using mammalian display with SARS-CoV-2 spikes and DMS, modified spikes can be designed for improved binding to animal ACE2 to design improved animal adapted SARS-CoV-2 viruses. Mouse models are the most extensively animal model used in research and have the most reagents available, however SARS-CoV-2 does not readily infect mice and does not transmit between them. Using mice that overexpressed hACE2 has been one way to overcome this, but the disease caused does not represent that in humans with frequent extra-pulmonary manifestations and death from encephalitis[217]. Other mouse adapted SARS-CoV-2 strains have been developed using serial passaging in mouse tissue and these do cause a more representative spectrum of disease than the hACE2 overexpressing mouse models, however the level of replication in the upper

airways still do not reflect those in human infection[218, 219]. The mouse-adapted strains developed using serial passaging may have been limited in their potential to explore the optimal mutations for increasing mouse ACE2 use by nucleotide distance, since amino acids two or more nucleotides away will not have been present in serially passaged SARS-CoV-2 virus. DMS may highlight amino acid substitutions that improve mouse ACE2 binding further. An improved mouse model would offer advantages in understanding pathogenicity, novel therapeutics, and vaccines.

In the future antigenic evolution using mammalian display with DMS could be explored for other virus entry proteins such as Ebola, where it will be safer to work with isolated proteins than live virus. DMS could help identify those regions that are highly variable and then by default those that are not. These conserved regions between different strains may be good targets for vaccination that may offer cross protective neutralisation and have a higher genetic barrier for antigenic drift.

9.4 Final perspectives

This thesis has been a proof of concept for the development of a platform to screen mutagenesis libraries using whole trimeric spike on mammalian cells. This was used to explore mutations that increase ACE2 binding and led to escape from antibody responses. This will hopefully serve as a foundation for future DMS work using current SARS-CoV-2 spikes and sera to explore potential evolutionary escape pathways and inform surveillance and predictive vaccinology.

In the process of validating the findings from the DMS screens, the interplay between the NTD and RBD was highlighted. NTD-RBD interactions influenced ACE2 binding and antibody escape. Further work will be needed to support the existence of epistasis between the NTD and RBD, however in the meantime the role of the NTD in VOCs should not be forgotten and considered in risk assessments of novel variants.

The highly focused nature of the antibody response against the SARS-CoV-2 RBD, the mutational plasticity of the RBD, and immune imprinting mean continued antigenic drift/evolution and reinfections are the likely course of events for the foreseeable future.

10 MATERIALS AND METHODS

10.1 Materials

10.1.1 Cell lines

Cells	Notes	Source
HEK 293Ts	Human embryonic kidney cells stably expressing the SV40 T antigen	ATCC
ACE2 expressing HEK 293Ts	HEK 293Ts stably transduced to express human ACE2	Barclay lab

10.1.2 Antibodies

Antibody	Application	Source
Recombinant Human ACE2 (mutated H374N + H378N) protein (Fc Chimera)	To measure ACE2 binding by flow cytometry	Abcam (ab273885)
Goat anti-human Fc 650 DyLight® 650. Ex: 654nm, Em: 673nm	Secondary antibody for flow cytometry	Abcam (ab97006)
Goat Anti-Mouse IgG H&L (Alexa Fluor® 488) Ex: 495nm, Em: 519nm	Secondary antibody for flow cytometry	Abcam (ab150113)
Goat Anti-Mouse IgG H&L (Alexa Fluor® 555) preadsorbed Ex: 555nm, Em: 565nm	Secondary antibody for flow cytometry	Abcam (ab150118)
REGN-CoV10933	Monoclonal antibody used to screen for escape variants and in neutralisation assays	Gift from Paul Kellam
REGN-CoV10987	Monoclonal antibody used to screen for escape variants and in neutralisation assays	Gift from Paul Kellam
Ly-CoV016	Monoclonal antibody used to screen for escape	Gift from Paul Kellam

	variants and in neutralisation assays	
S309		
RK001	Human sera from double dose BNT162b2 vaccine sera used to screen escape variants and in neutralisation assays	This study
RK002	Human sera from double dose BNT162b2 vaccine sera used to screen escape variants and in neutralisation assays	This study
RK003	Human sera from double dose BNT162b2 vaccine sera used to screen escape variants and in neutralisation assays	This study
RK004	Human sera from double dose BNT162b2 vaccine sera used to screen escape variants and in neutralisation assays	This study
RK005	Human sera from double dose BNT162b2 vaccine sera used to screen escape variants and in neutralisation assays	This study
RK006	Human sera from double dose BNT162b2 vaccine sera used to screen escape variants and in neutralisation assays	This study
RK007	Human sera from double dose BNT162b2 vaccine sera used to screen escape variants and in neutralisation assays	This study
RK008	Human sera from double dose BNT162b2 vaccine sera used to screen escape variants and in neutralisation assays	This study

2015
2016

10.1.3 Plasmids

Plasmid	Note	Source
pcDNA3-sACE2(WT)-Fc(IgG1)	Expression plasmid encoding sACE2(WT)-Fc(IgG1)	A gift from Erik Procko (Addgene plasmid # 145163 ; http://n2t.net/addgene:145163 ; RRID:Addgene_145163)[133]
pmScarlet_C1	Expression plasmid encoding the red fluorescent protein mScarlet	pmScarlet_C1 was a gift from Dorus Gadella (Addgene plasmid # 85042 ; http://n2t.net/addgene:85042 ; RRID:Addgene_85042)[140]
pcDNA3.1-mGreenLantern	Expression plasmid encoding the green fluorescent protein mGreenLantern	pcDNA3.1-mGreenLantern was a gift from Gregory Petsko (Addgene plasmid # 161912 ; http://n2t.net/addgene:161912 ; RRID:Addgene_161912)[139]
pcDNA3.1- SARS-CoV-2 spike (Wuhan)	Expression plasmid encoding the SARS-CoV-2 Wuhan spike protein	Paul McKay (Shattock lab)
pcDNA3.1- SARS-CoV-2 spike (BA.1)	Expression plasmid encoding the SARS-CoV-2 BA.1 spike protein	Barclay lab
pcDNA3.1- SARS-CoV-2 spike (BA.4)	Expression plasmid encoding the SARS-CoV-2 BA.4 spike protein	Barclay lab
pcDNA3.1- SARS-CoV-2 spike (Alpha)	Expression plasmid encoding the SARS-CoV-2 Alpha spike protein	Barclay lab
pcDNA3.1- SARS-CoV-2 spike (Alpha+L452R)	Expression plasmid encoding the SARS-CoV-2 Alpha with	This study

	L452R, predicted to escape vaccine sera	
pcDNA3.1- SARS-CoV-2 spike (Alpha+E484K)	Expression plasmid encoding the SARS-CoV-2 Alpha with E484K, predicted to escape vaccine sera	This study
pcDNA3.1- SARS-CoV-2 spike (Alpha+Q493K)	Expression plasmid encoding the SARS-CoV-2 Alpha with Q493K, predicted to escape vaccine sera	This study
pcDNA3.1- SARS-CoV-2 spike (Alpha+Q493R)	Expression plasmid encoding the SARS-CoV-2 Alpha with Q493R, predicted to escape vaccine sera	This study
pcDNA3.1- SARS-CoV-2 spike (Alpha+L452R+E484K+Q493K)	Expression plasmid encoding the SARS-CoV-2 Alpha with L452R+E484K+Q493R, predicted to escape vaccine sera	This study
pcDNA3.1- SARS-CoV-2 spike (Wuhan+BA.1 RBD)	Expression plasmid encoding the chimeric spike Wuhan with a BA.1 RBD	This study
pcDNA3.1- SARS-CoV-2 spike (Wuhan+BA.1 NTD)	Expression plasmid encoding the chimeric spike Wuhan with a BA.1 NTD	This study
pcDNA3.1- SARS-CoV-2 spike (BA.1+Wuhan NTD)	Expression plasmid encoding the chimeric spike BA.1 with a Wuhan NTD	This study
pcDNA3.1- SARS-CoV-2 spike (BA.1+Wuhan RBD)	Expression plasmid encoding the chimeric spike BA.1 with a Wuhan RBD	This study
pcDNA3.1- SARS-CoV-2 spike (BA.1+Wuhan RBD+E484A+N501Y)	Expression plasmid encoding the chimeric spike BA.1 with a Wuhan RBD+E484A+N501Y	This study

pcDNA3.1- SARS-CoV-2 spike (BA.1+Wuhan RBD+Q493R+N501Y)	Expression plasmid encoding the chimeric spike BA.1 with a Wuhan RBD+Q493R+N501Y	This study
pcDNA3.1- SARS-CoV-2 spike (BA.1+Wuhan RBD+E484A+Q493R+N501Y)	Expression plasmid encoding the chimeric spike BA.1 with a Wuhan RBD+E484A+Q493R+N501Y	This study
pcDNA3.1- SARS-CoV-2 spike (Wuhan)-tagged with mGreenLantern	Expression plasmid encoding SARS-CoV-2 spike (Wuhan) -tagged with mGreenLantern for flow cytometry	This study
pcDNA3.1- SARS-CoV-2 spike (BA.1)-tagged with mGreenLantern	Expression plasmid encoding SARS-CoV-2 spike (BA.1) -tagged with mGreenLantern for flow cytometry	This study
pcDNA3.1- SARS-CoV-2 spike (Delta)-tagged with mGreenLantern	Expression plasmid encoding SARS-CoV-2 spike (Delta) -tagged with mGreenLantern for flow cytometry	This study
pcDNA3.1- SARS-CoV-2 spike (Alpha+Q498R)- tagged with mGreenLantern	Expression plasmid encoding SARS-CoV-2 spike (Alpha+Q498R) -tagged with mGreenLantern for flow cytometry	This study
pcDNA3.1- SARS-CoV-2 spike (Alpha+Q493H)- tagged with mGreenLantern	Expression plasmid encoding SARS-CoV-2 spike (Alpha+Q498H) -tagged with mGreenLantern for flow cytometry	This study
pcDNA3.1- SARS-CoV-2 spike (Wuhan+Q498R)-tagged with mGreenLantern	Expression plasmid encoding SARS-CoV-2 spike (Wuhan+Q498R) -tagged with mGreenLantern for flow cytometry	This study

pcDNA3.1- SARS-CoV-2 spike (Wuhan+Q498H)-tagged with mGreenLantern	Expression plasmid encoding SARS-CoV-2 spike (Wuhan+Q498H) -tagged with mGreenLantern for flow cytometry	This study
pcDNA3.1- SARS-CoV-2 spike (Wuhan+Q498R+N501Y)-tagged with mGreenLantern	Expression plasmid encoding SARS-CoV-2 spike (Wuhan+Q498R+N501Y) -tagged with mGreenLantern for flow cytometry	This study
pcDNA3.1- SARS-CoV-2 spike (Wuhan+Q498H+N501Y)-tagged with mGreenLantern)	Expression plasmid encoding SARS-CoV-2 spike (Wuhan+Q498H+N501Y) -tagged with mGreenLantern for flow cytometry	This study
pcDNA3.1- SARS-CoV-2 spike (Wuhan+N501Y+Q498R-tagged with mGreenLantern)	Expression plasmid encoding SARS-CoV-2 spike (Wuhan+Q498R+N501Y) -tagged with mGreenLantern for flow cytometry	This study
pcDNA3.1- SARS-CoV-2 spike (Alpha+K417N)	Expression plasmid encoding SARS-CoV-2 spike (Alpha) with K417N, a mutation predicted to escape monoclonal antibodies	This study
pcDNA3.1- SARS-CoV-2 spike (Alpha+V486L)	Expression plasmid encoding SARS-CoV-2 spike (Alpha) with V486L, a mutation predicted to escape monoclonal antibodies	This study
pcDNA3.1- SARS-CoV-2 spike (Alpha+V445F)	Expression plasmid encoding SARS-CoV-2 spike (Alpha) with V445F, a mutation predicted to escape monoclonal antibodies	This study

pcDNA3.1- SARS-CoV-2 spike (Alpha+Q493R)	Expression plasmid encoding SARS-CoV-2 spike (Alpha) with Q493R, a mutation predicted to escape monoclonal antibodies	This study
pcDNA3.1- SARS-CoV-2 spike (Alpha+N460K)	Expression plasmid encoding SARS-CoV-2 spike (Alpha) with N460K, a mutation predicted to escape monoclonal antibodies	This study
pcDNA3.1- SARS-CoV-2 spike (Alpha+Y453K)	Expression plasmid encoding SARS-CoV-2 spike (Alpha) with Y453K, a mutation predicted to escape monoclonal antibodies	This study
pcDNA3.1- SARS-CoV-2 spike (Alpha+K417W)	Expression plasmid encoding SARS-CoV-2 spike (Alpha) with K417W, a mutation predicted to escape monoclonal antibodies	This study
pcDNA3.1- SARS-CoV-2 spike (Alpha+P499N)	Expression plasmid encoding SARS-CoV-2 spike (Alpha) with K499N, a mutation predicted to escape monoclonal antibodies	This study
pcDNA3-sACE2(WT)-Fc(IgG1)-tagged with mScarlet	Expression plasmid encoding sACE2 with a IgG1 Fc tag, tagged with mScarlet	This study
pcDNA3-sACE2-sGFP	Expression plasmid encoding sACE2 tagged with superfolder GFP	pcDNA3-sACE2(WT)-sfGFP was a gift from Erik Procko (Addgene plasmid # 145171 ; http://n2t.net/addgene:145171 ; RRID:Addgene_145171)[133]
pcDNA3-sACE2+T27Y+L79T+N330Y+A386L-sGFP	Expression plasmid encoding sACE2 with the mutations T27Y+L79T+N330Y+A	This study

	386L, shown to increase binding to SARS-CoV-2 spike, tagged with superfolder GFP	
pcDNA3-ΔCMV	Used as a non-coding plasmid during transfections.	This study
pCMV-Δ8.91	Plasmid encoding HIV gag-pol for pseudovirus generation.	Paul McKay (Shattock lab)
pCSFLW	Plasmid encoding lentiviral genome reporter, firefly luciferase.	Paul McKay (Shattock lab)

2018

2019 **10.1.4 Primers**

Primer name	Sequence	Note
pcDNA3.1 F	CGAAATTAATACGACTCACTATAGGG	Sanger sequencing of the pC3.1DNA plasmid
pcDNA3.1 R	GCAACTAGAAGGCACAGTCGAGGC	Sanger sequencing of the pC3.1DNA plasmid
RBD F	GTGCACCCTGAAGTCCTTCACCGTG	Sanger sequencing of the SARS-C0V2 spike plasmid
RBD R	GCCGGTCAGGCCGTTGAAGTTGAAG	Sanger sequencing of the SARS-C0V2 spike plasmid
3’701	CAGTGTCTGAAACCGGGTGAT	Sanger sequencing of the SARS-C0V2 spike plasmid

3’1700	CCCGGCCAAACTGCTGGAATG	Sanger sequencing of the SARS-C0V2 spike plasmid
3’3200	CGTATGTCACGTGCAGAAACA	Sanger sequencing of the SARS-C0V2 spike plasmid
5’500	TTCGAGTACGTGTCCCAGCCT	Sanger sequencing of the SARS-C0V2 spike plasmid
5’1500	ATGGCGTGGGCTATCAGCCCT	Sanger sequencing of the SARS-C0V2 spike plasmid
5’3109	GCAAGAGAGTGGACTTTTGCG	Sanger sequencing of the SARS-C0V2 spike plasmid
NGS RBD F (amp1) + Illumina F	TCGTCGGCAGCGTCAGATGTGTATAAGAG ACAGGAAAAGGGCATCTACCAGACCAGCAAC	For first PCR of amplicon 1 of RBD for NGS library prep
NGS RBD R (amp 1) + illumina R	GTCTCGTGGGCTCGGAGATGTGTATAAGAG ACAGGCCTGATAGATCTCGGTGGAGATG	For first PCR of amplicon 1 of RBD for NGS library prep
NGS RBD F (amp 2) + Illumina F	TCGTCGGCAGCGTCAGATGTGTATAAGAGAC AGGCACCTTCAAGTGCTACGGCG	For first PCR of amplicon 2 of RBD for NGS library prep
NGS RBD R (amp 2) + illumina R	GTCTCGTGGGCTCGGAGATGTGTATAAGAG ACAGGAAGTTCACGCATTTGTTCTTCACGAG	For first PCR of amplicon 2 of RBD for NGS library prep

2020
2021

10.1.5 Buffers and media

Buffer/media	Components	Use
Cell culture media	DMEM 10% Fetal bovine serum 1% Penicillin/Streptomycin 1% Non-essential amino acids	Passaging of HEK-293Ts and ACE2-HEK-293Ts Cells were maintained by splitting three times a week Additionally, 1:1000 puromycin was added to the media for ACE2-HEK-293T once a week
FACS buffer	1% PBS (Phosphate buffered saline) 5% Fetal bovine serum	Washing of cells for flow cytometry/FACS
Cell dissociation buffer, enzyme free	Gibco Cat no. 13151014	Dissociation of cells for FACS and flow cytometry

10.2 Methods

10.2.1 Polymerase chain reaction (PCR)

PCRs were done using Q5 hot start high-fidelity polymerase (NEB). PCRs were done according to manufacturer's instructions. In brief, 10ng of plasmid template or 20ng of DNA template was added to a mix containing 5ul of Q5 buffer, 0.5ul of 10uM dNTPs, 1.25ul of 10uM F primer, 1.25ul of 10uM R primer, 0.25ul of Q5 polymerase and nuclease free water was added to a total volume of 25ul. 5ul of GC enhancer was added to complex templates or for long PCR products. The thermocycling conditions used were according to manufacturer's instructions, the annealing temperature used was primer dependent.

10.2.2 Gel electrophoresis and extraction

Gel electrophoresis was conducted using a 1% agarose gel in TAE buffer. Gel red (Invitrogen) was used to visualise DNA. 6x loading dye (NEB) was used to load DNA samples into the gel and a 100bp or 1kB ladder used to estimate the size of the DNA bands. Gels were run at 100V for a time dependent on the expected length of the PCR product. Gels were then visualised on the Biorad gel doc XR+ imaging system.

DNA bands for extraction were cut from gels under UV light visualisation. DNA was extracted from the excised DNA bands using the Monarch gel extraction kit (NEB) in accordance with manufacturer's instructions.

10.2.3 DNA assembly

Fragments for assembly into plasmid were made using PCR including the vector, which was linearized. The fragments to be assembled were designed to have 15-20 nucleotide overlap with their neighbouring fragments. DNA assembly was then conducted using the NEBuilder HiFi DNA assembly kit in accordance with the manufacturer's instructions.

10.2.4 RBD library construction

Overlapping PCR using degenerate primers (NNK) was used to create the RBD library. Primers containing the degenerate codons were designed using a python script https://github.com/jbloomlab/CodonTilingPrimers[111, 144]. The first mutagenesis PCR used 10 cycles using equal amounts of a F primer and the pooled R (degenerate primers) for the forward fragment reaction and equal amounts of a R primer and the pooled F (degenerate primers) for the reverse fragment reaction (table 10-1). Equal amounts diluted 1:4 from the forward fragment and reverse fragment reaction were then combined in a joining PCR of 20 cycles (table 10-2). The PCR product was gel extracted and used to clone into a spike containing vector tagged at the C-terminal end with mGreenLantern[139].

Forward		Reverse	
10 x KOD Buffer	2.5	10 x KOD Buffer	2.5
MgS04	1.5	MgS04	1.5
dNTPs	2.5	dNTPs	2.5
F pool	0.75	R pool	0.75
R primer	0.75	F primer	0.75
H20	15.5	H20	15.5
Template (2ng)	1	Template (2ng)	1
KOD	0.5	KOD	0.5
Total	**25**	**Total**	**25**

	Temp (°C)	Time	
Activation	95	3 mins	
Denature	95	30 s	10 cycles
Anneal	58	45 s	
Extension	70	3 mins	
Last cycle	70	5 mins	
Hold	4		

Table 10-1: Protocol for the fragment PCR used to create the deep mutagenesis libraries.

Joining PCR	
10 x KOD Buffer	5
MgS04	4
dNTPs	5
F fragments (1:4)	6.66
R fragments (1:4)	6.66
5' primer	1.8
3' primer	1.8
H20	18.08
KOD	1
Total	**50**

	Temp (°C)	Time	
Activation	95	3 mins	
Denature	95	30 s	20 cycles
Anneal	58	45 s	
Extension	70	3 mins	
Last cycle	70	5 mins	
Hold	4		

Table 10-2: Protocol for the joining PCR used to create the deep mutagenesis libraries.

10.2.5 Bacterial transformation

1.5ul of assembly product was added to 25ul of DH5α ultracompetent cells (NEB) and incubated on ice for 30 minutes. The bacterial plasmid mix was heat shocked at 42°C for 30 seconds, then cooled on ice for 2 minutes. Pre-warmed SOC outgrowth media (NEB) was added to each transformation reaction and placed on the shaker at 37°C at 200-225 rpm for 1 hour. 100ul of the transformation reaction was plated onto pre-warmed agar plates impregnated with ampicillin and incubated overnight at 37°C.

10.2.6 Minipreps

Transformed bacteria were grown in 10ml of LB with 100ug/ml of ampicillin at 37ºC overnight. The bacterial suspensions were pelleted, and plasmids extracted and purified using the Monarch miniprep kit (NEB). Plasmids were eluted using 60ul of elution buffer.

10.2.7 Maxipreps

Transformed bacteria were grown in 250ml of LB with 100ug/ml of ampicillin at 37ºC overnight. The bacterial suspensions were pelleted, and plasmids extracted and purified using the Qiagen HiSpeed maxiprep kit (Qiagen). Plasmids were eluted using 1ml of TE buffer.

10.2.8 Sanger sequencing

DNA was Sanger sequenced using either Eurofins or Source Bioscience. Plasmid/DNA and the appropriate sequencing primer was sent according to the sample submission guidelines from the sequencer.

10.2.9 RNA extraction

Total cellular RNA was extracted from pelleted cells using the RNAeasy kit (Qiagen) according to manufacturer's instructions. RNA was eluted in 30ul of RNAse free water and stored at -80ºC.

10.2.10 cDNA synthesis

cDNA synthesis was carried out using SuperScript IV (Thermofisher) according to manufacturer's instructions using a gene specific primer. 10ul of RNA extract from the sorted cells was used as template for cDNA synthesis. The cDNA was used as a template for PCR with Q5 polymerase (NEB).

10.2.11 Next generation sequencing (NGS)

Library preparation for Illumina sequencing was done using 2 rounds of PCR. The first round of PCR amplified the region of interest and added complementary overhangs for the Illumina Nextera XT indices. The first round PCR used 25 PCR cycles. The second round of PCR added the Illumina Nextera XT indices for NGS. The second round of PCR used 8 cycles. After each round, PCR products were cleaned up using Ampure XP beads (Becker-Coultman) according to manufacturer's instructions. The second round PCR products were normalised by mass and pooled into a single library. The pool was submitted to the Imperial BRC Genomics Facility for quality check and sequencing on the Miseq.

10.2.12 Cell culture

HEK-293T and ACE2-HEK-293T cells were passaged in cell culture media and maintained at 37ºC and 5% CO_2. Cell lines were split 3 times a week. Puromycin at 1mcg/ml was added to the cell culture of ACE2-HEK-293T cells once a week to maintain expression of ACE2.

10.2.13 ACE2-Fc(IgG)-mScarlet production

HEK-293T cells were transfected with 500ng of the ACE2-Fc(IgG)-mScarlet plasmid per 10^6 cells using Lipofectamine 3000 (Thermofisher) as per the manufacturer's instructions. 48 hours after the transfection, the supernatant was aspirated and filtered using a 0.45µm filter. The supernatant was aliquoted and stored at -20^{o}C for future use. The supernatant was used directly in DMS screens and ACE2 binding studies without concentration. The volume to be used in each assay was determined by making a titration curve of a range of volumes of the ACE2-Fc(IgG)-mScarlet supernatant and measuring the mScarlet signal against spike expressing HEK-293T cells using flow cytometry.

10.2.14 Generation of a non-coding plasmid

A non-coding plasmid was created from pcDNA3.1 by deleting the CMV promoter using site directed mutagenesis (pcDNA3.1ΔCMV). This left a non-expressing plasmid that was still capable of replicating to high numbers in bacterial cells used for transformation and plasmid preparations.

10.2.15 FACS

HEK-293T cells were transfected with 1ng of the library plasmid mixed with 1500ng of a non-coding plasmid using Lipofectamine 3000 (Thermofisher) as per the manufacturer's instructions. 24 hours later transfected cells were dissociated using an enzyme free cell dissociation buffer and incubated with monoclonal antibody or sera for 30 minutes. The concentrations of monoclonal antibodies used were: Ly-CoV-016 (400 ng/mL), REGN 10933 (80 ng/mL, REGN 10987 (160 ng/mL), the dilutions for each sera used can be

found in Table 6-1. The dilution of sera was chosen following titration to identify the dilution that reduces binding of ACE2 by the population of spike library expressing cells by approximately 50%. For ACE2 binding screens, the antibody incubation step was omitted. Cells were then incubated with ACE2-Fc(IgG)-mScarlet for 1 hour. For the monoclonal and vaccine sera escape screens, a saturating volume of ACE2-Fc(IgG)-mScarlet was used, whereas for the ACE2 binding screens a sub-saturating volume of ACE2-Fc(IgG)-mScarlet was used. The volume for saturation and sub-saturation was determined by measuring a titration curve of a range of volumes of ACE2-Fc(IgG)-mScarlet binding to spike expressing HEK-293T cells. Cells were washed with FACS buffer twice before analysis and sorting on the BD Aria III. Cells with the highest ACE2 binding were sorted. For the monoclonal antibody and vaccine sera escape screens, the top 10% of ACE2 bound cells were sorted, whereas the top 5% was used for the ACE2 binding screens. Sorts continued until at least 10,000 cells were sorted. Sorted populations were received in FACs buffer and stored at -80ºC.

10.2.16 ACE2 binding measurements

HEK-293T cells were transfected with SARS-CoV-2 expressing plasmid in a mixture with non-coding plasmid at a ratio of 1:1500 respectively. The transfected cells were dissociated after 24 hours. The transfected cells were incubated with supernatant containing sACE2-Fc-mScarlet overnight. The cells were washed with FACS buffer and binding analysed on the flow cytometer. ACE2 binding was measured as the median fluorescence intensity of mScarlet for a fixed level of spike expression.

10.2.17 Pseudovirus generation

Plasmids encoding spike (pcDNA3.1), gag-pol (pCMV-Δ8.91) and luciferase (pCSFLW), were transfected in the ratio 1:1:1.5 for a total of 2500ng per 10^6 HEK-293T cells using Lipofectamine 3000 in accordance with manufacturer's instructions. 72 hours after transfections supernatant was collected, filtered using a 0.45um filter and stored at -80ºC. Prior to use, pseudovirus was titrated by serially diluting the collected supernatant 1:3 on ACE2-293T cells. 48-72 hours later, the media was removed, and relative luciferase units (RLU) was measured using the Bright-Glo Luciferase Assay System (Promega).

10.2.18 Pseudovirus neutralisation assays

mAbs or sera were serially diluted 1:2 in cell culture media in a 96 well plate. To the antibody dilutions ~10^6 RLU of pseudovirus was added. The antibody pseudovirus mixtures were incubated at 37ºC for 1 hour. 24,000 ACE2-HEK-293T cells in 100ul of cell culture media were added to each well of the 96 well plate. The plates were incubated at 37ºC and 5% CO_2 for 72 hours before being read using the Bright-Glo Luciferase Assay System (Promega).

Neutralisation titres 50% (NT50) and IC50s were calculated by fitting the data to a non-linear regression curve in GraphPad prism (version 9.2.0).

10.2.19 Western blots

Pseudoviruses were concentrated by ultracentrifugation before being reduced in Laemmlli buffer with 10% β-mercaptoethanol. Proteins were transferred to a

nitrocellulose membrane and were probed overnight with a polyclonal rabbit anti-SARS spike protein (NOVUS; NB100-56578) and mouse anti-p24 (abcam; ab9071), followed by 1 hour incubation with the secondary antibodies, IRDye® 680RD Goat anti-mouse (abcam; ab216776) and IRDye® 800CW Goat anti-rabbit (abcam; ab216773). Fluorescence was measured using the Odyssey Imaging System (LI-COR Biosciences).

BIBLIOGRAPHY

1. Zhou, P., et al., *A pneumonia outbreak associated with a new coronavirus of probable bat origin.* Nature, 2020. **579**(7798): p. 270-273.
2. Wu, F., et al., *A new coronavirus associated with human respiratory disease in China.* Nature, 2020. **579**(7798): p. 265-269.
3. Worobey, M., et al., *The Huanan Seafood Wholesale Market in Wuhan was the early epicenter of the COVID-19 pandemic.* Science, 2022. **377**(6609): p. 951-959.
4. Pekar, J.E., et al., *The molecular epidemiology of multiple zoonotic origins of SARS-CoV-2.* Science, 2022. **377**(6609): p. 960-966.
5. Gao, G., et al., *Surveillance of SARS-CoV-2 in the environment and animal samples of the Huanan Seafood Market (preprint).* 2022.
6. Temmam, S., et al., *Bat coronaviruses related to SARS-CoV-2 and infectious for human cells.* Nature, 2022. **604**(7905): p. 330-336.
7. Holmes, E.C., et al., *The origins of SARS-CoV-2: A critical review.* Cell, 2021. **184**(19): p. 4848-4856.
8. MacLean, O.A., et al., *Natural selection in the evolution of SARS-CoV-2 in bats created a generalist virus and highly capable human pathogen.* PLOS Biology, 2021. **19**(3): p. e3001115.
9. Liu, X., Q. Wu, and Z. Zhang, *Global Diversification and Distribution of Coronaviruses With Furin Cleavage Sites.* Front Microbiol, 2021. **12**: p. 649314.
10. Cui, J., F. Li, and Z.-L. Shi, *Origin and evolution of pathogenic coronaviruses.* Nature Reviews Microbiology, 2019. **17**(3): p. 181-192.
11. Ye, Z.-W., et al., *Zoonotic origins of human coronaviruses.* International journal of biological sciences, 2020. **16**(10): p. 1686-1697.
12. V'kovski, P., et al., *Coronavirus biology and replication: implications for SARS-CoV-2.* Nature Reviews Microbiology, 2020.
13. Millet, J.K., J.A. Jaimes, and G.R. Whittaker, *Molecular diversity of coronavirus host cell entry receptors.* FEMS Microbiology Reviews, 2021. **45**(3): p. fuaa057.
14. Starr, T.N., et al., *ACE2 binding is an ancestral and evolvable trait of sarbecoviruses.* Nature, 2022. **603**(7903): p. 913-918.
15. Edridge, A.W.D., et al., *Seasonal coronavirus protective immunity is short-lasting.* Nature Medicine, 2020. **26**(11): p. 1691-1693.
16. Callow, K.A., et al., *The time course of the immune response to experimental coronavirus infection of man.* Epidemiol Infect, 1990. **105**(2): p. 435-46.
17. Chibo, D. and C. Birch, *Analysis of human coronavirus 229E spike and nucleoprotein genes demonstrates genetic drift between chronologically distinct strains.* J Gen Virol, 2006. **87**(Pt 5): p. 1203-1208.
18. Ren, L., et al., *Genetic drift of human coronavirus OC43 spike gene during adaptive evolution.* Scientific Reports, 2015. **5**(1): p. 11451.
19. Wang, C., et al., *Antigenic structure of the human coronavirus OC43 spike reveals exposed and occluded neutralizing epitopes.* Nature Communications, 2022. **13**(1): p. 2921.

20. Eguia, R.T., et al., *A human coronavirus evolves antigenically to escape antibody immunity.* PLOS Pathogens, 2021. **17**(4): p. e1009453.
21. Regev-Yochay, G., et al., *Efficacy of a Fourth Dose of Covid-19 mRNA Vaccine against Omicron.* New England Journal of Medicine, 2022. **386**(14): p. 1377-1380.
22. Barda, N., et al., *Effectiveness of a third dose of the BNT162b2 mRNA COVID-19 vaccine for preventing severe outcomes in Israel: an observational study.* The Lancet, 2021. **398**(10316): p. 2093-2100.
23. Muhsen, K., et al., *Effects of BNT162b2 Covid-19 Vaccine Booster in Long-Term Care Facilities in Israel.* New England Journal of Medicine, 2021. **386**(4): p. 399-401.
24. Shang, J., et al., *Cell entry mechanisms of SARS-CoV-2.* Proceedings of the National Academy of Sciences, 2020. **117**(21): p. 11727-11734.
25. Shang, J., et al., *Structural basis of receptor recognition by SARS-CoV-2.* Nature, 2020. **581**(7807): p. 221-224.
26. Huang, Y., et al., *Structural and functional properties of SARS-CoV-2 spike protein: potential antivirus drug development for COVID-19.* Acta Pharmacologica Sinica, 2020. **41**(9): p. 1141-1149.
27. Li, F., *Structure, Function, and Evolution of Coronavirus Spike Proteins.* Annual Review of Virology, 2016. **3**(1): p. 237-261.
28. Duan, L., et al., *The SARS-CoV-2 Spike Glycoprotein Biosynthesis, Structure, Function, and Antigenicity: Implications for the Design of Spike-Based Vaccine Immunogens.* Front Immunol, 2020. **11**: p. 576622.
29. Benton, D.J., et al., *The effect of the D614G substitution on the structure of the spike glycoprotein of SARS-CoV-2.* Proceedings of the National Academy of Sciences, 2021. **118**(9): p. e2022586118.
30. Gobeil, S.M.C., et al., *D614G mutation alters SARS-CoV-2 spike conformation and enhances protease cleavage at the S1/S2 junction.* Cell Reports.
31. Yurkovetskiy, L., et al., *Structural and Functional Analysis of the D614G SARS-CoV-2 Spike Protein Variant.* Cell, 2020. **183**(3): p. 739-751.e8.
32. Meng, B., et al., *SARS-CoV-2 spike N-terminal domain modulates TMPRSS2-dependent viral entry and fusogenicity.* Cell Rep, 2022. **40**(7): p. 111220.
33. Cantoni, D., et al., *Evolutionary remodelling of N-terminal domain loops fine-tunes SARS-CoV-2 spike.* EMBO reports, 2022. **23**(10): p. e54322.
34. Malone, B., et al., *Structures and functions of coronavirus replication–transcription complexes and their relevance for SARS-CoV-2 drug design.* Nature Reviews Molecular Cell Biology, 2022. **23**(1): p. 21-39.
35. Yan, H., et al., *ACE2 receptor usage reveals variation in susceptibility to SARS-CoV and SARS-CoV-2 infection among bat species.* Nature Ecology & Evolution, 2021. **5**(5): p. 600-608.
36. Lean, F.Z.X., et al., *Differential susceptibility of SARS-CoV-2 in animals: Evidence of ACE2 host receptor distribution in companion animals, livestock and wildlife by immunohistochemical characterisation.* Transbound Emerg Dis, 2022. **69**(4): p. 2275-2286.

37. Hamming, I., et al., *Tissue distribution of ACE2 protein, the functional receptor for SARS coronavirus. A first step in understanding SARS pathogenesis.* J Pathol, 2004. **203**(2): p. 631-7.
38. Wang, L.F. and B.T. Eaton, *Bats, civets and the emergence of SARS.* Curr Top Microbiol Immunol, 2007. **315**: p. 325-44.
39. Prevention, C.f.D.C.a. *CDC SARS Response Timeline*. 2013 April 26, 2013 [cited 2022 14/09/2022]; Available from: https://www.cdc.gov/about/history/sars/timeline.htm.
40. Petersen, E., et al., *Comparing SARS-CoV-2 with SARS-CoV and influenza pandemics.* The Lancet Infectious Diseases, 2020. **20**(9): p. e238-e244.
41. Peacock, T.P., et al., *The furin cleavage site in the SARS-CoV-2 spike protein is required for transmission in ferrets.* Nature Microbiology, 2021. **6**(7): p. 899-909.
42. Papa, G., et al., *Furin cleavage of SARS-CoV-2 Spike promotes but is not essential for infection and cell-cell fusion.* PLOS Pathogens, 2021. **17**(1): p. e1009246.
43. Jackson, C.B., et al., *Mechanisms of SARS-CoV-2 entry into cells.* Nature Reviews Molecular Cell Biology, 2022. **23**(1): p. 3-20.
44. Johnson, B.A., et al., *Loss of furin cleavage site attenuates SARS-CoV-2 pathogenesis.* Nature, 2021. **591**(7849): p. 293-299.
45. Magazine, N., et al., *Mutations and Evolution of the SARS-CoV-2 Spike Protein.* Viruses, 2022. **14**(3).
46. Wrobel, A.G., et al., *Evolution of the SARS-CoV-2 spike protein in the human host.* Nature Communications, 2022. **13**(1): p. 1178.
47. Harvey, W.T., et al., *SARS-CoV-2 variants, spike mutations and immune escape.* Nature Reviews Microbiology, 2021. **19**(7): p. 409-424.
48. Korber, B., et al., *Tracking Changes in SARS-CoV-2 Spike: Evidence that D614G Increases Infectivity of the COVID-19 Virus.* Cell, 2020. **182**(4): p. 812-827.e19.
49. Mansbach, R.A., et al., *The SARS-CoV-2 Spike variant D614G favors an open conformational state.* Science Advances, 2021. **7**(16): p. eabf3671.
50. Plante, J.A., et al., *Spike mutation D614G alters SARS-CoV-2 fitness.* Nature, 2020.
51. Tian, F., et al., *N501Y mutation of spike protein in SARS-CoV-2 strengthens its binding to receptor ACE2.* eLife, 2021. **10**: p. e69091.
52. Starr, T.N., et al., *Deep Mutational Scanning of SARS-CoV-2 Receptor Binding Domain Reveals Constraints on Folding and ACE2 Binding.* Cell, 2020. **182**(5): p. 1295-1310.e20.
53. Liu, Y., et al., *The N501Y spike substitution enhances SARS-CoV-2 infection and transmission.* Nature, 2021.
54. Davies, N.G., et al., *Estimated transmissibility and impact of SARS-CoV-2 lineage B.1.1.7 in England.* Science, 2021: p. eabg3055.
55. Alpert, T., et al., *Early introductions and transmission of SARS-CoV-2 variant B.1.1.7 in the United States.* Cell, 2021.
56. Tegally, H., et al., *Detection of a SARS-CoV-2 variant of concern in South Africa.* Nature, 2021. **592**(7854): p. 438-443.
57. Faria, N.R., et al., *Genomics and epidemiology of the P.1 SARS-CoV-2 lineage in Manaus, Brazil.* Science, 2021: p. eabh2644.

58. Viana, R., et al., *Rapid epidemic expansion of the SARS-CoV-2 Omicron variant in southern Africa.* Nature, 2022. **603**(7902): p. 679-686.
59. Motozono, C., et al., *SARS-CoV-2 spike L452R variant evades cellular immunity and increases infectivity.* Cell Host & Microbe, 2021.
60. Deng, X., et al., *Transmission, infectivity, and neutralization of a spike L452R SARS-CoV-2 variant.* Cell.
61. Cherian, S., et al., *SARS-CoV-2 Spike Mutations, L452R, T478K, E484Q and P681R, in the Second Wave of COVID-19 in Maharashtra, India.* Microorganisms, 2021. **9**(7).
62. Mlcochova, P., et al., *SARS-CoV-2 B.1.617.2 Delta variant replication and immune evasion.* Nature, 2021.
63. Liu, Y., et al., *Delta spike P681R mutation enhances SARS-CoV-2 fitness over Alpha variant.* Cell Reports, 2022. **39**(7): p. 110829.
64. Escalera, A., et al., *Mutations in SARS-CoV-2 variants of concern link to increased spike cleavage and virus transmission.* Cell Host Microbe, 2022. **30**(3): p. 373-387.e7.
65. Pastorio, C., et al., *Determinants of Spike infectivity, processing, and neutralization in SARS-CoV-2 Omicron subvariants BA.1 and BA.2.* Cell Host & Microbe, 2022.
66. Meng, B., et al., *Altered TMPRSS2 usage by SARS-CoV-2 Omicron impacts infectivity and fusogenicity.* Nature, 2022. **603**(7902): p. 706-714.
67. Willett, B.J., et al., *SARS-CoV-2 Omicron is an immune escape variant with an altered cell entry pathway.* Nature Microbiology, 2022. **7**(8): p. 1161-1179.
68. Cele, S., et al., *Omicron extensively but incompletely escapes Pfizer BNT162b2 neutralization.* Nature, 2022. **602**(7898): p. 654-656.
69. Schmidt, F., et al., *Plasma Neutralization of the SARS-CoV-2 Omicron Variant.* New England Journal of Medicine, 2021.
70. Lyngse, F.P., et al., *Household transmission of the SARS-CoV-2 Omicron variant in Denmark.* Nature Communications, 2022. **13**(1): p. 5573.
71. Hart, W.S., et al., *Generation time of the alpha and delta SARS-CoV-2 variants: an epidemiological analysis.* The Lancet Infectious Diseases, 2022. **22**(5): p. 603-610.
72. Wall, E.C., et al., *Neutralising antibody activity against SARS-CoV-2 VOCs B.1.617.2 and B.1.351 by BNT162b2 vaccination.* The Lancet, 2021. **397**(10292): p. 2331-2333.
73. Davis, C., et al., *Reduced neutralisation of the Delta (B.1.617.2) SARS-CoV-2 variant of concern following vaccination.* PLOS Pathogens, 2021. **17**(12): p. e1010022.
74. Liu, C., et al., *Reduced neutralization of SARS-CoV-2 B.1.617 by vaccine and convalescent serum.* Cell.
75. Piccoli, L., et al., *Mapping Neutralizing and Immunodominant Sites on the SARS-CoV-2 Spike Receptor-Binding Domain by Structure-Guided High-Resolution Serology.* Cell, 2020. **183**(4): p. 1024-1042.e21.
76. Amanat, F., et al., *SARS-CoV-2 mRNA vaccination induces functionally diverse antibodies to NTD, RBD, and S2.* Cell, 2021. **184**(15): p. 3936-3948.e10.

77. Greaney, A.J., et al., *Comprehensive mapping of mutations in the SARS-CoV-2 receptor-binding domain that affect recognition by polyclonal human plasma antibodies.* Cell Host & Microbe, 2021. **29**(3): p. 463-476.e6.
78. Greaney, A.J., et al., *Antibodies elicited by mRNA-1273 vaccination bind more broadly to the receptor binding domain than do those from SARS-CoV-2 infection.* Science Translational Medicine, 2021. **13**(600): p. eabi9915.
79. Lan, J., et al., *Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor.* Nature, 2020. **581**(7807): p. 215-220.
80. Dejnirattisai, W., et al., *The antigenic anatomy of SARS-CoV-2 receptor binding domain.* Cell.
81. Yang, W.T., et al., *SARS-CoV-2 E484K Mutation Narrative Review: Epidemiology, Immune Escape, Clinical Implications, and Future Considerations.* Infect Drug Resist, 2022. **15**: p. 373-385.
82. Tegally, H., et al., *Emergence of SARS-CoV-2 Omicron lineages BA.4 and BA.5 in South Africa.* Nature Medicine, 2022.
83. McCarthy, K.R., et al., *Recurrent deletions in the SARS-CoV-2 spike glycoprotein drive antibody escape.* Science, 2021. **371**(6534): p. 1139.
84. Cerutti, G., et al., *Potent SARS-CoV-2 neutralizing antibodies directed against spike N-terminal domain target a single supersite.* Cell Host & Microbe, 2021. **29**(5): p. 819-833.e7.
85. McCallum, M., et al., *SARS-CoV-2 immune evasion by the B.1.427/B.1.429 variant of concern.* Science, 2021. **373**(6555): p. 648-654.
86. Koel, B.F., et al., *Substitutions near the receptor binding site determine major antigenic change during influenza virus evolution.* Science, 2013. **342**(6161): p. 976-9.
87. Mykytyn, A.Z., et al., *Antigenic cartography of SARS-CoV-2 reveals that Omicron BA.1 and BA.2 are antigenically distinct.* Science Immunology. **0**(0): p. eabq4450.
88. van der Straten, K., et al., *Antigenic cartography using sera from sequence-confirmed SARS-CoV-2 variants of concern infections reveals antigenic divergence of Omicron.* Immunity, 2022. **55**(9): p. 1725-1731.e4.
89. Cele, S., et al., *SARS-CoV-2 prolonged infection during advanced HIV disease evolves extensive immune escape.* Cell Host Microbe, 2022. **30**(2): p. 154-162.e5.
90. Choi, B., et al., *Persistence and Evolution of SARS-CoV-2 in an Immunocompromised Host.* New England Journal of Medicine, 2020. **383**(23): p. 2291-2293.
91. Clark, S.A., et al., *SARS-CoV-2 evolution in an immunocompromised host reveals shared neutralization escape mechanisms.* Cell.
92. Kemp, S.A., et al., *SARS-CoV-2 evolution during treatment of chronic infection.* Nature, 2021.
93. Gobeil, S.M.C., et al., *Structural diversity of the SARS-CoV-2 Omicron spike.* Molecular Cell, 2022.
94. Andreano, E., et al., *SARS-CoV-2 escape from a highly neutralizing COVID-19 convalescent plasma.* Proceedings of the National Academy of Sciences, 2021. **118**(36): p. e2103154118.

95. Baum, A., et al., *Antibody cocktail to SARS-CoV-2 spike protein prevents rapid mutational escape seen with individual antibodies.* Science, 2020: p. eabd0831.
96. Weisblum, Y., et al., *Escape from neutralizing antibodies by SARS-CoV-2 spike protein variants.* Elife, 2020. **9**.
97. Schmidt, F., et al., *High genetic barrier to SARS-CoV-2 polyclonal neutralizing antibody escape.* Nature, 2021. **600**(7889): p. 512-516.
98. Gerhard, W., et al., *Antigenic structure of influenza virus haemagglutinin defined by hybridoma antibodies.* Nature, 1981. **290**(5808): p. 713-717.
99. Davis, A.K.F., et al., *Sera from Individuals with Narrowly Focused Influenza Virus Antibodies Rapidly Select Viral Escape Mutations In Ovo.* J Virol, 2018. **92**(19).
100. Shinomiya, N., et al., *Reconsidering the need for gain-of-function research on enhanced potential pandemic pathogens in the post-COVID-19 era.* Front Bioeng Biotechnol, 2022. **10**: p. 966586.
101. Oude Munnink, B.B., et al., *The next phase of SARS-CoV-2 surveillance: real-time molecular epidemiology.* Nature Medicine, 2021. **27**(9): p. 1518-1524.
102. Dadonaite, B., et al., *A pseudovirus system enables deep mutational scanning of the full SARS-CoV-2 spike.* bioRxiv, 2022: p. 2022.10.13.512056.
103. Johnson, B.A., et al., *Nucleocapsid mutations in SARS-CoV-2 augment replication and pathogenesis.* PLoS Pathog, 2022. **18**(6): p. e1010627.
104. Syed Abdullah, M., et al., *Rapid assessment of SARS-CoV-2 evolved variants using virus-like particles.* Science. **0**(0): p. eabl6184.
105. Xue, K.S., et al., *Parallel evolution of influenza across multiple spatiotemporal scales.* Elife, 2017. **6**.
106. Avanzato, V.A., et al., *Case Study: Prolonged Infectious SARS-CoV-2 Shedding from an Asymptomatic Immunocompromised Individual with Cancer.* Cell, 2020.
107. Ferrara, F. and N. Temperton, *Pseudotype Neutralization Assays: From Laboratory Bench to Data Analysis.* Methods and protocols, 2018. **1**(1): p. 8.
108. McCallum, M., et al., *N-terminal domain antigenic mapping reveals a site of vulnerability for SARS-CoV-2.* Cell.
109. Starr, T.N., et al., *Prospective mapping of viral mutations that escape antibodies used to treat COVID-19.* Science, 2021. **371**(6531): p. 850.
110. Fowler, D.M. and S. Fields, *Deep mutational scanning: a new style of protein science.* Nature Methods, 2014. **11**(8): p. 801-807.
111. Bloom, J.D., *An Experimentally Determined Evolutionary Model Dramatically Improves Phylogenetic Fit.* Molecular Biology and Evolution, 2014. **31**(8): p. 1956-1978.
112. Soh, Y.S., et al., *Comprehensive mapping of adaptation of the avian influenza polymerase protein PB2 to humans.* Elife, 2019. **8**.
113. Doud, M.B. and J.D. Bloom, *Accurate Measurement of the Effects of All Amino-Acid Mutations on Influenza Hemagglutinin.* Viruses, 2016. **8**(6).
114. Doud, M.B., S.E. Hensley, and J.D. Bloom, *Complete mapping of viral escape from neutralizing antibodies.* PLoS Pathog, 2017. **13**(3): p. e1006271.
115. Doud, M.B., J.M. Lee, and J.D. Bloom, *How single mutations affect viral escape from broad and narrow antibodies to H1 influenza hemagglutinin.* Nat Commun, 2018. **9**(1): p. 1386.

116. Lee, J.M., et al., *Deep mutational scanning of hemagglutinin helps predict evolutionary fates of human H3N2 influenza variants.* Proceedings of the National Academy of Sciences, 2018. **115**(35): p. E8276-E8285.
117. Ashenberg, O., et al., *Deep mutational scanning identifies sites in influenza nucleoprotein that affect viral inhibition by MxA.* PLoS Pathog, 2017. **13**(3): p. e1006288.
118. Garrett, M.E., et al., *High resolution profiling of pathways of escape for SARS-CoV-2 spike-binding antibodies.* Cell.
119. Chan, K.K., et al., *An engineered decoy receptor for SARS-CoV-2 broadly binds protein S sequence variants.* Science Advances, 2021. **7**(8): p. eabf1738.
120. Javanmardi, K., et al., *Rapid characterization of spike variants via mammalian cell surface display.* Molecular Cell, 2021. **81**(24): p. 5099-5111.e8.
121. Clackson, T., et al., *Making antibody fragments using phage display libraries.* Nature, 1991. **352**(6336): p. 624-628.
122. Dell, A., et al., *Similarities and Differences in the Glycosylation Mechanisms in Prokaryotes and Eukaryotes.* International Journal of Microbiology, 2010. **2010**: p. 148178.
123. Boder, E.T. and K.D. Wittrup, *Yeast surface display for screening combinatorial polypeptide libraries.* Nature Biotechnology, 1997. **15**(6): p. 553-557.
124. Wildt, S. and T.U. Gerngross, *The humanization of N-glycosylation pathways in yeast.* Nature Reviews Microbiology, 2005. **3**(2): p. 119-128.
125. Watanabe, Y., et al., *Site-specific glycan analysis of the SARS-CoV-2 spike.* Science, 2020. **369**(6501): p. 330-333.
126. Sztain, T., et al., *A glycan gate controls opening of the SARS-CoV-2 spike protein.* Nature Chemistry, 2021. **13**(10): p. 963-968.
127. Harbison, A.M., et al., *Fine-tuning the spike: role of the nature and topology of the glycan shield in the structure and dynamics of the SARS-CoV-2 S.* Chemical Science, 2022. **13**(2): p. 386-395.
128. Turoňová, B., et al., *In situ structural analysis of SARS-CoV-2 spike reveals flexibility mediated by three hinges.* Science, 2020. **370**(6513): p. 203.
129. Greaney, A.J., et al., *Complete mapping of mutations to the SARS-CoV-2 spike receptor-binding domain that escape antibody recognition.* Cell Host & Microbe, 2020.
130. Starr, T.N., et al., *Complete map of SARS-CoV-2 RBD mutations that escape the monoclonal antibody LY-CoV555 and its cocktail with LY-CoV016.* Cell Reports Medicine, 2021. **2**(4).
131. Ho, M., S. Nagata, and I. Pastan, *Isolation of anti-CD22 Fv with high affinity by Fv display on human cells.* Proceedings of the National Academy of Sciences, 2006. **103**(25): p. 9637.
132. Ho, M. and I. Pastan, *Display and selection of scFv antibodies on HEK-293T cells.* Methods in molecular biology (Clifton, N.J.), 2009. **562**: p. 99-113.
133. Chan, K.K., et al., *Engineering human ACE2 to optimize binding to the spike protein of SARS coronavirus 2.* Science, 2020. **369**(6508): p. 1261.
134. Zahradník, J., et al., *SARS-CoV-2 variant prediction and antiviral drug design are enabled by RBD in vitro evolution.* Nature Microbiology, 2021. **6**(9): p. 1188-1198.

135. Starr, T.N., et al., *Deep mutational scans for ACE2 binding, RBD expression, and antibody escape in the SARS-CoV-2 Omicron BA.1 and BA.2 receptor-binding domains.* bioRxiv, 2022: p. 2022.09.20.508745.
136. Odegard, V.H. and D.G. Schatz, *Targeting of somatic hypermutation.* Nature Reviews Immunology, 2006. **6**(8): p. 573-583.
137. Brindle, N.P.J., et al., *Directed Evolution of an Angiopoietin-2 Ligand Trap by Somatic Hypermutation and Cell Surface Display *.* Journal of Biological Chemistry, 2013. **288**(46): p. 33205-33212.
138. Bate, N., et al., *In vitro evolution predicts emerging SARS-CoV-2 mutations with high affinity for ACE2 and cross-species binding.* PLoS Pathog, 2022. **18**(7): p. e1010733.
139. Campbell, B.C., et al., *mGreenLantern: a bright monomeric fluorescent protein with rapid expression and cell filling properties for neuronal imaging.* Proceedings of the National Academy of Sciences, 2020. **117**(48): p. 30710.
140. Bindels, D.S., et al., *mScarlet: a bright monomeric red fluorescent protein for cellular imaging.* Nature Methods, 2017. **14**(1): p. 53-56.
141. Wille, M. and E.C. Holmes, *The Ecology and Evolution of Influenza Viruses.* Cold Spring Harbor Perspectives in Medicine, 2019.
142. Wong, T.S., et al., *Sequence saturation mutagenesis (SeSaM): a novel method for directed evolution.* Nucleic acids research, 2004. **32**(3): p. e26-e26.
143. Nov, Y., *When second best is good enough: another probabilistic look at saturation mutagenesis.* Applied and environmental microbiology, 2012. **78**(1): p. 258-262.
144. Dingens, A.S., et al., *Comprehensive Mapping of HIV-1 Escape from a Broadly Neutralizing Antibody.* Cell Host Microbe, 2017. **21**(6): p. 777-787.e4.
145. Volz, E., et al., *Evaluating the Effects of SARS-CoV-2 Spike Mutation D614G on Transmissibility and Pathogenicity.* Cell, 2021. **184**(1): p. 64-75.e11.
146. Starr Tyler, N., et al., *Shifting mutational constraints in the SARS-CoV-2 receptor-binding domain during viral evolution.* Science. **0**(0): p. eabo7896.
147. O'Brien, M.P., et al., *Subcutaneous REGEN-COV Antibody Combination to Prevent Covid-19.* New England Journal of Medicine, 2021.
148. Jones, B.E., et al., *The neutralizing antibody, LY-CoV555, protects against SARS-CoV-2 infection in non-human primates.* Science Translational Medicine, 2021: p. eabf1906.
149. Chen, R.E., et al., *In vivo monoclonal antibody efficacy against SARS-CoV-2 variant strains.* Nature, 2021.
150. Weinreich, D.M., et al., *REGN-COV2, a Neutralizing Antibody Cocktail, in Outpatients with Covid-19.* New England Journal of Medicine, 2020.
151. Dougan, M., et al., *Bamlanivimab plus Etesevimab in Mild or Moderate Covid-19.* New England Journal of Medicine, 2021.
152. Abani, O., et al., *Casirivimab and imdevimab in patients admitted to hospital with COVID-19 (RECOVERY): a randomised, controlled, open-label, platform trial.* The Lancet, 2022. **399**(10325): p. 665-676.
153. Gupta, A., et al., *Early Treatment for Covid-19 with SARS-CoV-2 Neutralizing Antibody Sotrovimab.* New England Journal of Medicine, 2021. **385**(21): p. 1941-1950.

154. Andreano, E., et al., *Extremely potent human monoclonal antibodies from COVID-19 convalescent patients.* Cell, 2021. **184**(7): p. 1821-1835.e16.
155. Brouwer, P.J.M., et al., *Potent neutralizing antibodies from COVID-19 patients define multiple targets of vulnerability.* Science, 2020. **369**(6504): p. 643.
156. Pinto, D., et al., *Cross-neutralization of SARS-CoV-2 by a human monoclonal SARS-CoV antibody.* Nature, 2020. **583**(7815): p. 290-295.
157. Starr, T.N., et al., *SARS-CoV-2 RBD antibodies that maximize breadth and resistance to escape.* Nature, 2021. **597**(7874): p. 97-102.
158. Shah, D.K. and A.M. Betts, *Antibody biodistribution coefficients: inferring tissue concentrations of monoclonal antibodies based on the plasma concentrations in several preclinical species and human.* MAbs, 2013. **5**(2): p. 297-305.
159. Cao, Y., et al., *Omicron escapes the majority of existing SARS-CoV-2 neutralizing antibodies.* Nature, 2022. **602**(7898): p. 657-663.
160. Muñoz-Alía, M.Á., et al., *Serotypic evolution of measles virus is constrained by multiple co-dominant B cell epitopes on its surface glycoproteins.* Cell Reports Medicine, 2021. **2**(4): p. 100225.
161. Pebody, R.G., et al., *Uptake and effectiveness of influenza vaccine in those aged 65 years and older in the United Kingdom, influenza seasons 2010/11 to 2016/17.* Euro Surveill, 2018. **23**(39).
162. Belongia, E.A., et al., *Variable influenza vaccine effectiveness by subtype: a systematic review and meta-analysis of test-negative design studies.* Lancet Infect Dis, 2016. **16**(8): p. 942-51.
163. Greaney, A.J., F.C. Welsh, and J.D. Bloom, *Co-dominant neutralizing epitopes make anti-measles immunity resistant to viral evolution.* Cell Rep Med, 2021. **2**(4): p. 100257.
164. Fulton, B.O., et al., *Mutational Analysis of Measles Virus Suggests Constraints on Antigenic Variation of the Glycoproteins.* Cell Rep, 2015. **11**(9): p. 1331-8.
165. Heaton, N.S., et al., *Genome-wide mutagenesis of influenza virus reveals unique plasticity of the hemagglutinin and NS1 proteins.* Proceedings of the National Academy of Sciences, 2013. **110**(50): p. 20248-20253.
166. Lee, J.M., et al., *Mapping person-to-person variation in viral mutations that escape polyclonal serum targeting influenza hemagglutinin.* eLife, 2019. **8**: p. e49324.
167. Arevalo, C.P., et al., *Original antigenic sin priming of influenza virus hemagglutinin stalk antibodies.* Proceedings of the National Academy of Sciences, 2020. **117**(29): p. 17221-17227.
168. Petrova, V.N. and C.A. Russell, *The evolution of seasonal influenza viruses.* Nature Reviews Microbiology, 2018. **16**(1): p. 47-60.
169. Dhar Mahesh, S., et al., *Genomic characterization and epidemiology of an emerging SARS-CoV-2 variant in Delhi, India.* Science, 2021. **374**(6570): p. 995-999.
170. Kimura, I., et al., *The SARS-CoV-2 Lambda variant exhibits enhanced infectivity and immune resistance.* Cell Reports, 2021: p. 110218.
171. Bar-On, Y.M., et al., *Protection by a Fourth Dose of BNT162b2 against Omicron in Israel.* New England Journal of Medicine, 2022.

172. Lythgoe, K.A., et al., *SARS-CoV-2 within-host diversity and transmission.* Science, 2021. **372**(6539): p. eabg0821.
173. Wong, T.Y., et al., *Evaluating Antibody Mediated Protection against Alpha, Beta, and Delta SARS-CoV-2 Variants of Concern in K18-hACE2 Transgenic Mice.* J Virol, 2022. **96**(6): p. e0218421.
174. Muik, A., et al., *Neutralization of SARS-CoV-2 lineage B.1.1.7 pseudovirus by BNT162b2 vaccine–elicited human sera.* Science, 2021. **371**(6534): p. 1152.
175. Kimura, I., et al., *Virological characteristics of the SARS-CoV-2 Omicron BA.2 subvariants, including BA.4 and BA.5.* Cell, 2022. **185**(21): p. 3992-4007.e16.
176. Walls, A.C., et al., *Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein.* Cell, 2020. **181**(2): p. 281-292.e6.
177. Wang, C., et al., *A conserved immunogenic and vulnerable site on the coronavirus spike protein delineated by cross-reactive monoclonal antibodies.* Nature Communications, 2021. **12**(1): p. 1715.
178. Suryadevara, N., et al., *Neutralizing and protective human monoclonal antibodies recognizing the N-terminal domain of the SARS-CoV-2 spike protein.* Cell.
179. Meng, B., et al., *Recurrent emergence of SARS-CoV-2 spike deletion H69/V70 and its role in the Alpha variant B.1.1.7.* Cell Reports, 2021. **35**(13).
180. Kimura, I., et al., *Virological characteristics of the novel SARS-CoV-2 Omicron variants including BA.2.12.1, BA.4 and BA.5.* bioRxiv, 2022: p. 2022.05.26.493539.
181. Ulrich, L., et al., *Enhanced fitness of SARS-CoV-2 variant of concern Alpha but not Beta.* Nature, 2021.
182. Elliott, P., et al., *Exponential growth, high prevalence of SARS-CoV-2, and vaccine effectiveness associated with the Delta variant.* Science. **374**(6574): p. eabl9551.
183. Lopez Bernal, J., et al., *Effectiveness of Covid-19 Vaccines against the B.1.617.2 (Delta) Variant.* New England Journal of Medicine, 2021.
184. Wall, E.C., et al., *AZD1222-induced neutralising antibody activity against SARS-CoV-2 Delta VOC.* The Lancet.
185. Lyngse, F.P., et al., *Increased transmissibility of SARS-CoV-2 lineage B.1.1.7 by age and viral load.* Nature Communications, 2021. **12**(1): p. 7251.
186. Yuan, S., et al., *Pathogenicity, transmissibility, and fitness of SARS-CoV-2 Omicron in Syrian hamsters.* Science, 2022. **377**(6604): p. 428-433.
187. Dejnirattisai, W., et al., *SARS-CoV-2 Omicron-B.1.1.529 leads to widespread escape from neutralizing antibody responses.* Cell.
188. Cao, Y., et al., *Imprinted SARS-CoV-2 humoral immunity induces convergent Omicron RBD evolution.* Nature, 2022.
189. Barnes, C.O., et al., *SARS-CoV-2 neutralizing antibody structures inform therapeutic strategies.* Nature, 2020.
190. Wang, Q., et al., *Alarming antibody evasion properties of rising SARS-CoV-2 BQ and XBB subvariants.* Cell, 2022.
191. Chan, M.C.W., et al., *Frequent Genetic Mismatch between Vaccine Strains and Circulating Seasonal Influenza Viruses, Hong Kong, China, 1996-2012.* Emerg Infect Dis, 2018. **24**(10): p. 1825-1834.

192. Horwood, P.F., et al., *Circulation and characterization of seasonal influenza viruses in Cambodia, 2012-2015.* Influenza and Other Respiratory Viruses, 2019. **13**(5): p. 465-476.
193. Puzelli, S., et al., *Co-circulation of the two influenza B lineages during 13 consecutive influenza surveillance seasons in Italy, 2004–2017.* BMC Infectious Diseases, 2019. **19**(1): p. 990.
194. Wheatley, A.K., et al., *Immune imprinting and SARS-CoV-2 vaccine design.* Trends Immunol, 2021. **42**(11): p. 956-959.
195. Francis, T., *On the Doctrine of Original Antigenic Sin.* Proceedings of the American Philosophical Society, 1960. **104**(6): p. 572-578.
196. Francis, M.E., M.L. King, and A.A. Kelvin *Back to the Future for Influenza Preimmunity—Looking Back at Influenza Virus History to Infer the Outcome of Future Infections*. Viruses, 2019. **11**, DOI: 10.3390/v11020122.
197. Gostic, K.M., et al., *Childhood immune imprinting to influenza A shapes birth year-specific risk during seasonal H1N1 and H3N2 epidemics.* PLOS Pathogens, 2019. **15**(12): p. e1008109.
198. Gouma, S., et al., *Middle-aged individuals may be in a perpetual state of H3N2 influenza virus susceptibility.* Nature Communications, 2020. **11**(1): p. 4566.
199. Supasa, P., et al., *Reduced neutralization of SARS-CoV-2 B.1.1.7 variant by convalescent and vaccine sera.* Cell, 2021. **184**(8): p. 2201-2211.e7.
200. Kaku, C.I., et al., *Recall of preexisting cross-reactive B cell memory after Omicron BA.1 breakthrough infection.* Science Immunology. **7**(73): p. eabq3511.
201. Muik, A., et al., *Omicron BA.2 breakthrough infection enhances cross-neutralization of BA.2.12.1 and BA.4/BA.5.* Science Immunology. **0**(0): p. eade2283.
202. Quandt, J., et al., *Omicron BA.1 breakthrough infection drives cross-variant neutralization and memory B cell formation against conserved epitopes.* Sci Immunol, 2022. **7**(75): p. eabq2427.
203. Guthmiller, J.J. and P.C. Wilson, *Harnessing immune history to combat influenza viruses.* Current Opinion in Immunology, 2018. **53**: p. 187-195.
204. Röltgen, K., et al., *Immune imprinting, breadth of variant recognition, and germinal center response in human SARS-CoV-2 infection and vaccination.* Cell, 2022. **185**(6): p. 1025-1040.e14.
205. Hui, K.P.Y., et al., *SARS-CoV-2 Omicron variant replication in human bronchus and lung ex vivo.* Nature, 2022. **603**(7902): p. 715-720.
206. Focosi, D. and F. Maggi, *Recombination in Coronaviruses, with a Focus on SARS-CoV-2.* Viruses, 2022. **14**(6).
207. Li, X., et al., *Emergence of SARS-CoV-2 through recombination and strong purifying selection.* Science Advances. **6**(27): p. eabb9153.
208. Turakhia, Y., et al., *Pandemic-scale phylogenomics reveals the SARS-CoV-2 recombination landscape.* Nature, 2022. **609**(7929): p. 994-997.
209. Jackson, B., et al., *Generation and transmission of interlineage recombinants in the SARS-CoV-2 pandemic.* Cell, 2021. **184**(20): p. 5179-5188.e8.
210. Tamura, T., et al., *Virological characteristics of the SARS-CoV-2 XBB variant derived from recombination of two Omicron subvariants.* bioRxiv, 2022: p. 2022.12.27.521986.

211. Kaplan, B.S. and R.J. Webby, *The avian and mammalian host range of highly pathogenic avian H5N1 influenza.* Virus Res, 2013. **178**(1): p. 3-11.
212. Herfst, S., et al., *Airborne Transmission of Influenza A/H5N1 Virus Between Ferrets.* Science, 2012. **336**(6088): p. 1534.
213. Oude Munnink, B.B., et al., *Transmission of SARS-CoV-2 on mink farms between humans and mink and back to humans.* Science, 2020: p. eabe5901.
214. Hale, V.L., et al., *SARS-CoV-2 infection in free-ranging white-tailed deer.* Nature, 2021.
215. Hoffmann, M., et al., *SARS-CoV-2 mutations acquired in mink reduce antibody-mediated neutralization.* Cell Reports, 2021. **35**(3): p. 109017.
216. Zhou, J., et al., *Mutations that adapt SARS-CoV-2 to mink or ferret do not increase fitness in the human airway.* Cell Rep, 2022. **38**(6): p. 110344.
217. Oladunni, F.S., et al., *Lethality of SARS-CoV-2 infection in K18 human angiotensin-converting enzyme 2 transgenic mice.* Nature Communications, 2020. **11**(1): p. 6122.
218. Dinnon, K.H., et al., *A mouse-adapted model of SARS-CoV-2 to test COVID-19 countermeasures.* Nature, 2020. **586**(7830): p. 560-566.
219. Leist, S.R., et al., *A Mouse-Adapted SARS-CoV-2 Induces Acute Lung Injury and Mortality in Standard Laboratory Mice.* Cell, 2020. **183**(4): p. 1070-1085.e12.