EXPLORING POPULATION IMMUNITY TOWARDS POTENTIAL PANDEMIC VIRUSES

Jai Samuel Bolton

Keble College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

HILARY 2024

1. Declaration

I declare that this thesis was composed by myself and that the work contained herein is my own except where explicitly stated in the text. The work has not been submitted for any degree or professional qualification except as specified.

Jai Samuel Bolton, HILARY 2023

2. Acknowledgement

I wish to thank my supervisors, Dr Anita Milicic, Dr Alexandra Spencer, and Dr Craig Thompson for their consistent support and feedback. Special thanks go to Dr Thompson for his tireless kindness, guidance, and mentorship, all of which greatly contributed to my academic and professional development.

I want to thank the Oxford Interdisciplinary Biosciences Doctoral Training Programme and the administration of Keble College for their support. Perhaps most of all, I wish to thank the DTP for introducing me to my wonderful partner Chloé.

I wish to thank the fantastic friends I have made from my DPhil course and college. Special mention goes to my housemates Jacques, Jack, Jeremie, Jochem, Andrew, and Nic. The lounge will always hold a place in my heart.

I wish to thank everyone I have worked with during my time as a DPhil student. Included in these is my Mum, Sandra Bolton, who spent time proofreading various manuscripts.

Although challenging, my time working in the Peter Medawar Building as part of the COVID-19 response will always be a treasured memory, enhanced by the people I worked with.

3. Publications and Contributions

The following publications arose from my time spent as a DPhil Student.

[1] H. Harvala *et al.,* 'Convalescent plasma therapy for the treatment of patients with COVID-19: assessment of methods available for antibody detection and their correlation with neutralising antibody levels', *Transfusion Medicine*, vol. 31, no. 3, pp. 167–175, 2021.
- Work not presented in this thesis.

[2] S. F. Lumley *et al.,* 'SARS-CoV-2 antibody prevalence, titres and neutralising activity in an antenatal cohort, United Kingdom, 14 April to 15 June 2020', *Eurosurveillance*, vol. 25, no. 42, p. 2001721, 2020.
- Work not presented in this thesis.

[3] C. P. Thompson *et al.,* 'Detection of neutralising antibodies to SARS-CoV-2 to determine population exposure in Scottish blood donors between March and May 2020', *Eurosurveillance*, vol. 25, no. 42, p. 2000685, 2020.
- Chapter 9 contains neutralisation data that was produced by a 20-person team, which I was a part of, with permission from the corresponding author.

[4] J. W. Bouvier *et al.,* 'Rubisco adaptation is more limited by phylogenetic constraint than by catalytic trade-off', *Molecular Biology and Evolution*, vol. 38, no. 7, pp. 2880–2896, 2021.
- Work not presented in this thesis.

[5] W. Dejnirattisai *et al.,* 'The antigenic anatomy of SARS-CoV-2 receptor binding domain', *Cell*, vol. 184, no. 8, pp. 2183–2200, 2021.
- Work not presented in this thesis

[6] A. L. McNaughton *et al.,* 'Fatal COVID-19 outcomes are associated with an antibody response targeting epitopes shared with endemic coronaviruses', *JCI insight*, vol. 7, no. 13, 2022.
- Work not presented in this thesis.

[7] K. A. Bates, J. S. Bolton, and K. C. King, 'A globally ubiquitous symbiont can drive experimental host evolution', *Molecular Ecology*, vol. 30, no. 15, pp. 3882–3892, 2021.
- Work not presented in this thesis.

[8] J. S. Bolton, H. Klim, J. Wellens, M. Edmans, U. Obolski, and C. P. Thompson, 'An antigenic thrift-based approach to influenza vaccine design', *Vaccines*, vol. 9, no. 6, p. 657, 2021.
- Introduction 7.11 and 7.13 is loosely based on the concept of this review, with permission from the corresponding author.

4. Abstract

4.1. Introduction

The emergence of SARS coronavirus 2 (SARS-CoV-2) in 2019 prompted unprecedented public health responses globally. Concurrently, H5N1 and H1N1 influenza viruses continue to pose significant threats to public health: with H5N1 viruses posing a concerning pandemic risk. In this study, I conducted a comprehensive serosurvey of human and avian anti-influenza virus immunity. Bioinformatics-informed site-directed mutagenesis techniques (SDM) uncovered HA-head-directed immunity likely present in this serosurvey. Additionally, I assessed the Scottish population immunity towards SARS-CoV-2 through a serosurvey of blood donors during the first pandemic wave.

4.2. Methods

I employed a lentivirus-based pseudotyped virus system to test for neutralising antibodies against the influenza Haemagglutinin (HA) and SARS-CoV-2 Spike (S) proteins in sera samples. I employed structural bioinformatics and SDM to characterise putative antibody binding sites (ABSs) and epitopes on H5N1 and H1N1 hemagglutinins (HAs). The seroprevalence of SARS-CoV-2 antibodies was modelled using logistic regression.

4.3. Results

I found that 479/520 people neutralised the HA from A/bar-headed goose/Qinghai/3/2005/H5N1, a Highly Pathogenic Avian Influenza (HPAI) virus that did not infect people in Scotland. I also found varying levels of population immunity towards pandemic and seasonal H1N1 influenza HAs. Structural bioinformatics approaches identified epitopes of limited variability (ELVs) shared between H5N1 and H1N1 subtypes. Site-directed mutagenesis (SDM) showed that mutating HA head residues in A/bar-

headed goose/Qinghai/3/2005/H5N1 and A/USSR/90/1977/H1N1 significantly modulates the human neutralising response. The SARS-CoV-2 serosurvey revealed negligible population immunity amongst the Scottish population towards the end of the first COVID-19 pandemic wave.

4.4. Discussion

This study reveals potential cross-reactive epitopes of limited variability (ELVs) shared between H5N1 and H1N1 influenza subtypes, challenging conventional notions of subtype-specific immunity. These findings suggest that the immune system recognises conserved regions within the HA head domain, potentially impacting vaccine design and pandemic preparedness efforts. Further validation and exploration are warranted to understand the implications fully. Moreover, my results support the concept of limited drift/ antigenic thrift, wherein conserved immunologically relevant regions drive influenza antigenic evolution. The identification of shared epitopes has profound implications for universal influenza vaccine development. By targeting conserved head ELVs, vaccines may confer broad cross-protection against multiple influenza strains, mitigating the need for frequent updates and enhancing pandemic preparedness.

Moreover, the SARS-CoV-2 serosurvey revealed that another COVID-19 pandemic wave was likely and provided impetus for the UK government to pursue further public health interventions.

5. Abbreviations

Abbreviation	Definition
ABS	Antibody Binding Site
ACE2	Angiotensin-Converting Enzyme 2
ARDS	Acute Respiratory Distress Syndrome
BH	Benjamini-Hochberg
bnAB	Broadly Neutralising Antibody
CDC	Centres for Disease Control and Prevention
DMS	Deep Mutational Scanning
ELV	Epitope of Limited Variability
FDA	Food and Drug Administration
GAM	Generalised Additive Model
GC	Germinal Centre
GP	Glycoprotein
H/ LPAI	Highly Pathogenic/ Low Pathogenic Avian Influenza
HAI	Haemagglutination Inhibition Assay
HA	Haemagglutinin
IC50	Half-Maximal Inhibitory Concentration
M2	Matrix Protein 2
NA	Neuraminidase
OD	Optical Density
OAS	Original Antigenic Sin
PHE	Public Health England
PMN	Pseudotype Microneutralisation Assay
PSO	Post Symptom Onset
RBD/ S	Receptor Binding Domain/ Site
RLU	Relative Light Unit
SA	Sialic Acid
S	Spike Protein
VE	Vestigial Esterase
WHO	World Health Organisation
WOAH	World Organisation for Animal Health

6. Table of contents

7. Introduction

Seasonal influenza causes an estimated three to five million cases of severe disease and between 290,000 to 650,000 deaths per year [1]. Influenza pandemics have also occurred four times since 1918, causing over 50 million deaths [2]. Since 2019, SARS-CoV-2 spread around the world causing over 600 million cases and up to 6 million deaths [3]. In order to combat the threat of emergent viruses, we need to determine which strains have pandemic potential and design vaccines that protect against future pandemics and seasonal outbreaks.

7.1. Virion structure

Figure 1 shows a representation of an influenza virion. Influenza comprises up to seventeen proteins. Three of these are found on the exterior of the virion, the surface glycoproteins Haemagglutinin (HA), Neuraminidase (NA), and Matrix 2 (M2). The virion is enclosed in a host-derived lipid membrane. The surface glycoproteins are anchored to the viral envelope. The most abundant of these glycoproteins is HA, which mediates viral attachment and entry into host cells by binding to sialic acid receptors on the cell surface [4]. HA is estimated to be the target of 60% of anti-influenza antibodies [5]. Consequently, HA is the primary antigen used for most influenza vaccines [5]–[7]. Therefore, this project focuses on immune responses towards HA. NA cleaves sialic acid residues from host cell receptors, thereby facilitating viral egress [8]. The M2 protein provides structural support to the viral envelope and functions as a proton channel, enabling the acidification necessary for viral entry into host cells [9]–[12]. Inside the viral envelope are viral Ribonucleoproteins (vRNPs). These vRNPs consist of the viral genome bound to viral Nucleoproteins (NPs). This is how the negative sense, single-stranded RNA gene segments are stored and protected. Also within the vRNP complexes are the three sub-unit proteins that form the polymerase complex: Polymerase Basic Protein 1 (PB1), Polymerase Basic Protein 2 (PB2), and Polymerase Acidic Protein (PA). TheRNA-dependent RNA polymerase

complex catalysing transcription and replication of the viral genome. The morphology of influenza virions is varied; laboratory virions tend to be spherical, whilst clinical isolates appear to form long filaments [13]–[17].

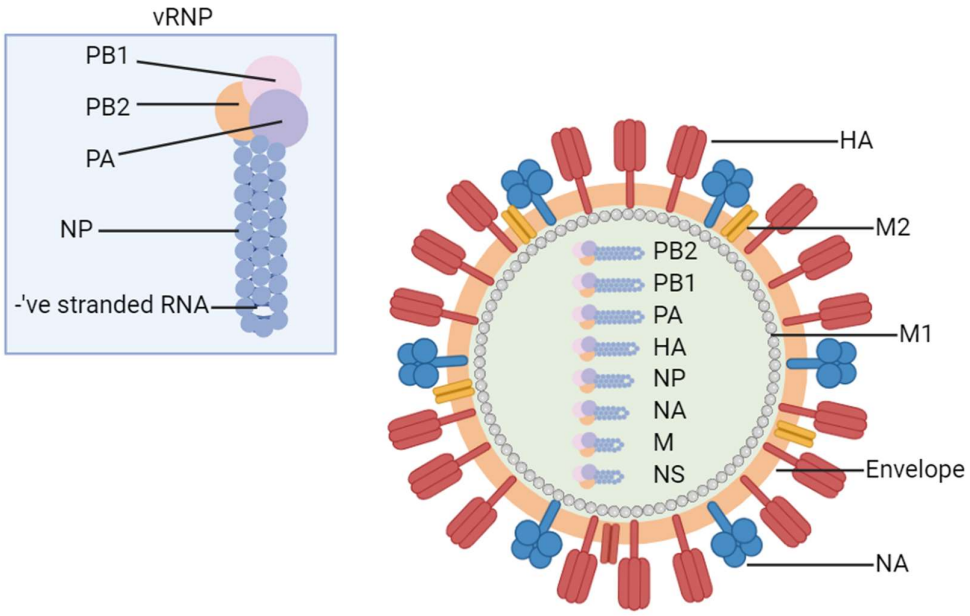


Figure 1. A representation of an influenza virion showing the distribution of the HA, NA and M2 proteins on the influenza virion's surface, as well as the structure of vRNPs. Created using Biorender.com (accessed on 11 April 2024).

7.2. Influenza Subtypes & Lineages

Influenza A and B currently circulate in the human population. There are two influenza A subtypes circulating; H1N1 influenza A and H3N2 influenza A. Before 2020, there were two lineages of Influenza B; Victoria and Yamatagata. Since 2020, the Yamagata lineage has dropped out of circulation [18], [19]. Influenza A subtypes are denoted by HA (1–18) and NA (NA, 1–11) [20]. The subtypes H1N1, H2N2, H3N1, and H5N1 are all influenza A [20]. Multiple strains comprise each subtype showing divergent evolutionary and antigenic properties.

In humans, multiple subtypes have circulated since the last century [5]. H1N1 influenza initially circulated between 1918 and 1957. In 1957, H1N1 was replaced by H2N2 influenza A in the 'Asian flu'.

H2N2 then circulated between 1957 and 1968 when it was replaced by H3N2 influenza A in the 'Hong Kong flu'. In 1977, an H1N1 strain closely related to one previously circulating in 1947 [21] appeared and began to co-circulate with H3N2, the 'Russian flu'. In 2009, a novel strain of H1N1 replaced the strain that appeared in 1977 in the 'Swine flu' pandemic. A timeline of this is given by Figure 2.

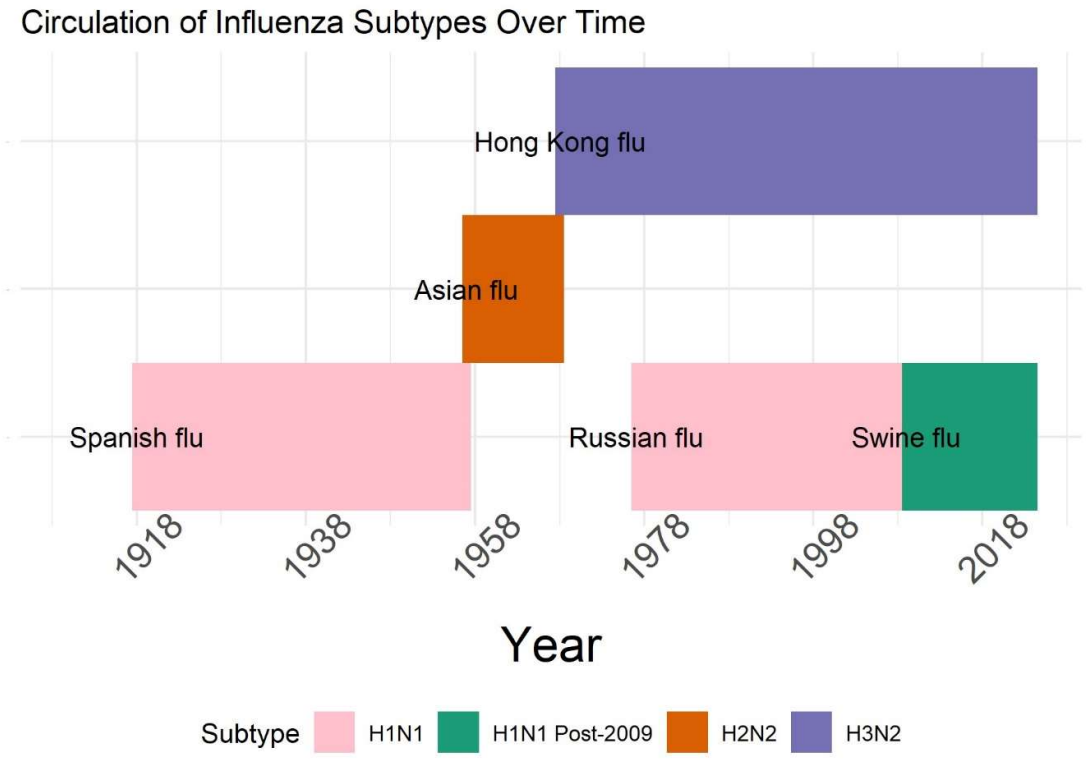


Figure 2. The circulation of human influenza subtypes over time. Notable pandemics that led to subtype introductions are labelled.

In addition to circulating in humans, influenza also circulates in avian populations. H5N1 strains are classified as Low Pathogenic Avian Influenza (LPAI) and Highly Pathogenic Avian Influenza (HPAI) HPAI is associated with severe clinical manifestations and high mortality rates in poultry and occasionally in humans [22]–[24]. LPAI strains cause mild or asymptomatic infections in birds [25]–[29]. LPAI H5N1 strains spread routinely in avian populations, with most genetic diversity situated in water birds [22]–

[24]. HPAI evolves from LPAI[25]–[29]. In poultry, H5 influenza A strains cause large-scale outbreaks, commonly resulting in 70-80% mortality [30].

H5N1 influenza viruses are grouped into 10 phylogenetic clades based on HA gene sequences (Group 0-9) [31], [32], [33]. Avian-to-human H5N1 spillover was first documented in Hong Kong in 1997, killing 6/18 patients [34]. This was defined as clade 0. Between 2004 and 2023, it is estimated that there have been over 874 human H5N1 cases in 23 countries, reviewed recently [35]. Of these, approximately 458 people have died, causing a total case mortality rate of 52.4%. It is likely that this is an overestimate due to a large number of undetected cases. H5N1 remains topical due to a large-scale avian pandemic. In 2022, a strain from H5N1 subclade 2.3.4.4b originated in the Eurasian Steppe spread across the world and has caused spillover into multiple mammals including mink, seals, and dolphins. This has caused the worst avian death toll of any recorded H5N1 pandemic, with over 50 million birds succumbing or culled. Human spillover events have also occurred. To date, H5N1 strains have not displayed efficient human-to-human transmission, however, this remains a possibility.

7.3. HA structure

The pre-fusion structure of HA is well-characterised [36]–[41]. After transcription and translation of the HA gene segment, the naïve HA0 protein is produced (Figure 3 d). This consists of three identical monomers forming a trimer. HA0 contains a proteolytic cleavage site situated in the HA1 subunit, to the 3’ side of the C-Terminal F” subdomain (Figure 3 d). After proteolytic cleavage in the Golgi apparatus, the HA1 and HA2 subunits are held together by a disulfide bridge (Figure 3 b and 3). This disulfide bridge connects the N-terminal F” subdomain to the F subdomain.

The structure of HA can be further divided into the head and stem domains. The head domain consists of the receptor binding domain (RBD) and the vestigial esterase (VE). The stem domain contains the HA1 subunit which consists of the signal peptide and cleavage site, N-terminal F” subdomain and C-terminal F” subdomain.

The stem domain contains the HA2 subunit, consisting of the fusion peptide and F subdomain (which together form the HA2 subunit ectodomain), the transmembrane domain (TMD), and the cytoplasmic tail (Figure 3 d). The head domain is held together by a disulfide bridge (Figure 3 b and d). This holds both sections of the VE together.

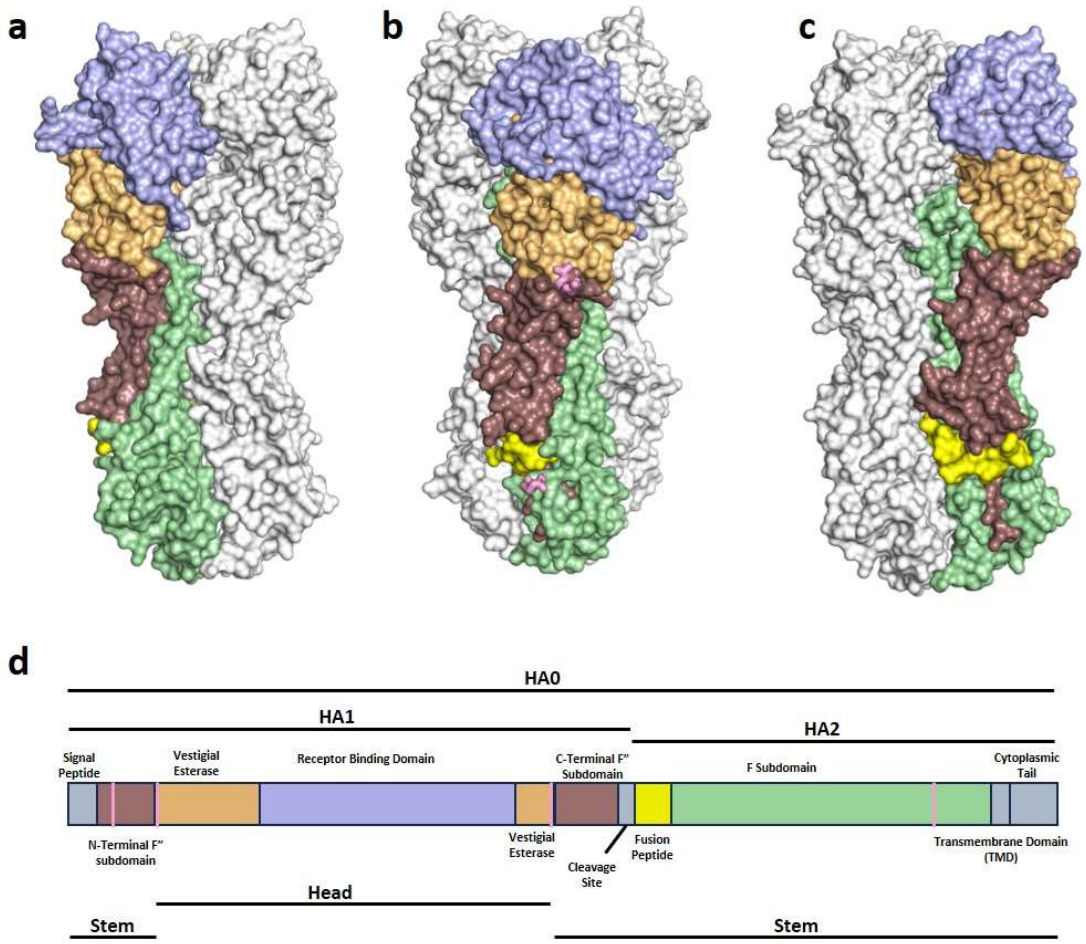


Figure 3. The structure of HA is shown on the '2fko' crystal structure. Figures a-c show the left, middle and right view of the HA trimer. Figure d shows a representation of the HA gene sequence, split into the HA0, HA1, HA2, head, and stem domains. Colouring is consistent between the crystal structures and HA sequence. Pink sticks on b denote two disulfide bridges, the locations of these are denoted in Figure d. The top disulfide bridge in Figure b connects the head domain, denoted by the middle two pink vertical lines on Figure d. The bottom pink stick on Figure b denotes the disulfide bridge joining the N-Terminal F" subdomain of HA1 to the F" subdomain of HA2, denoted by the outer two vertical lines on Figure d.

7.4. Influenza lifecycle and the role of HA

The general lifecycle of influenza is well understood, reviewed in [4], [42], [43], (Figure 4). The influenza lifecycle has six main constituent parts: (1) host cell entry, (2) entry into the host cell nucleus, (3) genome replication, (4) nuclear export, (5) translation, and (6) viral assembly and budding.

The two main functions of HA are (i) enabling influenza virions entry to host cells, and (ii) release of vRNPs into the host cell cytosol [4], [8], [42]–[46]. Sialic Acid (SA) receptors are cell receptors for influenza. During cell entry, influenza virions contact the cell surface membrane of host. NA proteins cleave glycosidic sialic acid-HA bonds that do not mediate cell entry. This occurs via hydrolysis of glycosidic bonds [8]. HA binds to the cell surface receptors using the RBD of the head domain. Situated within the RBD is the receptor binding site (RBS) of HA. The RBS is comprised of α-helices and a β-barrel [4]. The RBS forms a depression towards the top left head domain. RBS secondary structures involved in SA-HA binding include the 130 loop, 190 helix and 220 loop [47]. Virions are then endocytosed through receptor-mediated endocytosis. Although situated in the head domain, adjacent to the RBD, the VE subdomain's function is so far unclear [36], [37].

During the endosomal pathway, the vesicles containing the virions become more acidic via the M2 protein acting as an ion channel. The M2 protein is activated by the acidifying environment. One of M2's functions is the transportation of protons into the virion's core, further lowering the pH [9]–[12]. The lowering pH causes complex conformational changes in HA that allow for viral entry [4], [8], [42]–[46]. HA2's fusion peptide is released due to the lowering pH [4], [8], [42], [43], [45], [46] (Figure 3 d). The fusion peptide inserts into the host cell membrane. At this point, HA is inserted into the host cell surface membrane and anchored to the viral envelope by HA2's TMD. In the process of hemifusion, the outer parts of the viral envelope and host cell membranes merge. Through further complex processes, the inner sections of both membranes merge and form the fusion pore. This process involves the substantial conformational change of HA. The fusion pore allows the vRNPs to enter the host cell interior.

In the cell interior, vRNPs are trafficked to the host cell nucleus by host cell machinery. The influenza genome is then transcribed in the nucleus by the three subunits that form the polymerase complex. In the nucleus, the genome is transcribed to produce mRNA. mRNA encoding HA, NA and M2 are trafficked towards endoplasmic reticulum (ER) associated ribosomes.

ER-associated ribosomes synthesise the nascent hemagglutinin (HA) polypeptides, which are initially produced in their precursor form, HA0 [43], [48] (Figure 3 d). The signal peptide (SP) of the newly produced HA performs important functions [4], [48]. The SP signals target the nascent HA to the ER and binds the signal recognition particle (SRP) which facilitates ER entry. This process happens co-translationally, whilst the ribosome is still translating. The SP is then cleaved from the amino acid sequence. After folding and trimerising in the ER, the nascent HA0 proteins are transported to the Golgi apparatus for post-translational modification (PTM).

In the Golgi apparatus [48], host proteases [4], [8], [43], [48], specifically type II transmembrane serine proteases (TTSPs) [49], cleave the HA0 precursor into two parts: the HA1 and HA2 subunits, exposing the fusion peptide. These are held together by a disulfide bridge (Figure 3 b and d) [36]–[41]. The HA trimers are now in their active form and are transported to the host cell membrane for assembly into virions and subsequent budding.

The rest of the gene segments are trafficked to cytosolic ribosomes, translated into proteins, and then localised back into the nucleus. These proteins take part in replication and are assembled into vRNPs, which are subsequently trafficked to the cell surface. At the cell surface, vRNPs and the cell surface proteins assemble to form influenza virions. These bud out of the host cell by pinching the cell surface membrane to form the virion's envelope. Virion-SA bonds are cleaved by NA and the virions are released from the host cell.

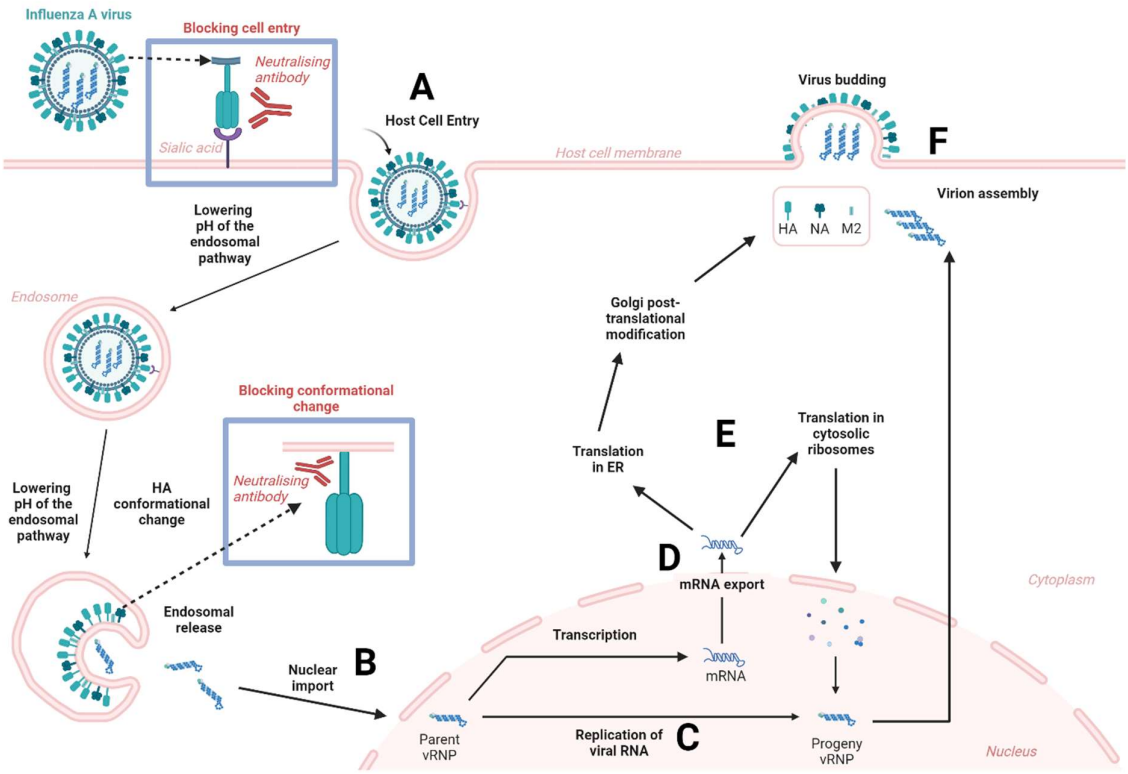


Figure 4. A schematic of the influenza lifecycle. The influenza lifecycle consists of six main stages: (A) host cell entry, (B) entry into the host cell nucleus, (C) genome replication, (D) nuclear export, (E) translation, and (F) viral assembly and budding. The two blue boxes show the two stages of the lifecycle targeted by neutralising antibodies, which is the focus of this thesis. Created using Biorender.com (accessed on 5 November 2024).

7.5. HA's cleavage site as a pathogenicity determinant

LPAI H5N1 HAs typically have a monobasic cleavage site [25]–[29]. These are cleaved by proteases found predominantly in the respiratory and gastrointestinal tracts of birds. HPAI H5N1 HAs typically possess a polybasic cleavage site [25]–[29]. H5's polybasic cleavage site allows cleavage of HA with a greater variety of host proteases, and in particular, furin [25]–[27], which is ubiquitously expressed. This allows the systemic replication of the virus and contributes to increased virulence.

7.6. Anti-HA antibodies

Anti-influenza antibodies attenuate the virus by interrupting parts of the lifecycle. The most common anti-influenza antibodies target HA [5]–[7]. Broadly reactive anti-HA antibodies can be divided into two groups: neutralising and non-neutralising antibodies. Neutralising antibodies stop influenza virions from entering the host cell [50]–[55]. Non-neutralising antibodies have an inhibiting impact on viral function but do not stop virions from entering host cells.

Most anti-HA neutralising antibodies block HA-SA binding and block virion entry [50], [54], [55] (Figure 4 A). These antibodies bind the head domain of HA [50], [54], [55] (Figure 4 A). Binding interference can occur through direct RBS binding [50], [54] or steric hindrance of the interaction [50]. Neutralising antibodies can also block HA-endosome fusion by stabilising HA's structure [50], [54], [56] (Figure 4 B). This occurs by blocking the low-pH-initiated conformational change of HA and the subsequent process of host cell membrane fusion. This primarily occurs for anti-stem antibodies but has also been reported for anti-head antibodies [50], [56] (Figure 4 B). Neutralisation assays such as the pseudotype microneutralisation assay measure the neutralising action of antibodies; the degree to which antibodies stop viral entry. This technique quantifies the action of anti-HA neutralising antibodies [50], [57], [58].

Non-neutralising antibodies have been shown to (i) block virion release by steric hindrance of NA [50], [59] and, (ii) bind the stem domain of HA0 and block proteolytic cleavage, rendering HA inactive [50], [60]. Moreover, influenza antibodies can function in a cell-dependent manner. Antibodies can be split [61]–[64] into the Fab, which binds antigens and the Fc region, which binds host immune cells. Some anti-HA antibodies activate the immune system to attack infected cells when their Fc region binds [50], [51], [65], [66]. This can occur by immune effector cells killing infected host cells through antibody-dependent cellular cytotoxicity (ADCC), the phagogocytosis of infected host cells through antibody-dependent cellular phagocytosis (ADCP), or activation of the complement system [50], [51], [65], [66].

The antigen-binding portion of antibodies, known as the 'Fab' region, is formed from the light chain and part of the heavy chain [61]–[64]. Antibodies are divided into isotypes depending on the type of heavy chain they comprise [61]–[64]. There are five mammalian antibody classes: IgA, IgD, IgE, IgG, and IgM. IgA exists as Y-shaped monomers intracellularly and as dimers when secreted [50], [61]. IgA comprises approximately 15% of serum antibodies and is divided into two subclasses, IgA1 and IgA2. IgD and IgE both form Y-shaped monomers and are present at low levels in serum, 0.5% and 0.01% respectively [50], [61]. IgG is the most common human blood sera antibody, comprising around 75% of antibodies [50], [61]. IgG forms Y-shaped monomers and is split into 4 subclasses, IgG1, IgG2, IgG3, and IgG4. IgM forms pentamers of Y-shaped monomers and comprises approximately 10% of serum antibody [50], [61]. Most human anti-influenza antibodies are of the IgG1 and a lower proportion of IgG3 isotype. IgG performs direct neutralisation and can bind to Fc receptors [63]. It is thought that IgA plays a predominant role in the upper, whereas IgG plays a predominant role in the lower respiratory tract [63]. IgM is the first antibody produced during an influenza infection. It interacts with complement and can be secreted onto mucosal surfaces in a similar manner to IgA [63].

The Fab region of antibodies is split into four broad domains; the heavy and light chain constant regions, and the heavy and light chain variable regions (Fv). The basic structures of these are well-characterised [64]. At the top end of each Fv region is a hypervariable region. The complementarity-determining region (CDR) of each hypervariable region is formed from 3 beta sheets [64]. The CDRs are responsible for binding antigens. CDRs recognise and bind to groups of amino acids on the surface of HA antigens. These regions are defined as 'epitopes' [67]. Inversely, the group of residues comprising the CDRs that can bind an HA epitope is called the 'paratope.'

7.7. The repertoire of anti-influenza immune responses

Antibodies are produced by B cells [68]–[71]. A single B cell clone produces one B cell receptor (BCR), and if activated, releases a corresponding antibody. Multiple B cell genotypes give rise to a polyclonal

immune response containing mAbs with targeting epitopes. For antibody production, B cells must first be activated. The primary immune response is elicited first [68]–[71]. In secondary lymphoid tissue, B cells are primed by BCR-viral antigen binding [68], [69]. These primed, naïve B cells are then activated by binding T_{FH} CD4+ T cells displaying a viral antigen on Major Histocompatibility Complex 2 (MHC2). In B cell follicles, activated B cells replicate by B cell proliferation [68], [69]. They then undergo differentiation into different cell types. These can differentiate into (i) plasmablasts, which rapidly produce IgM and then under 'class switching' to produce IgA/IgG, (ii) germinal centre (GC) independent memory B cells, or (iii) GC-B cells which migrate to the GC of the lymphoid tissue [68], [69].

GC B cells undergo further differentiation and clonal expansion. Over the course of an immune response, GC B cells produce antibodies with progressively higher affinity for viral epitopes [72], [73]. This process is known as 'affinity maturation.' This is thought to occur by a combination of two mechanisms: somatic hypermutation and clonal selection. During the immune response, mutations build up in BCR DNA [74], conferring mutation build-up in the structure of the antibodies produced. This is the process of somatic hypermutation. Somatic hypermutation can lead to changes in the affinity-enhancing CDRs of the antibodies which build up over time. Some of these mutations improve the binding strength and stability of antibodies [72]. In a process known as 'clonal expansion', proliferating GC B cells are continually exposed to antigens, and T_{FH} CD4+ T cells displaying these antigens on MHC2, and become activated [73]. B cells with favourable mutations can be selected for with help from T_{FH} CD4+ T cells. These T cells can favour the survival potential of superior B cells over inferior B cells [73].

Activated GC cells then can differentiate into plasma cells or long-lived memory B cells [68], [69]. Plasma cells commonly take up residence in bone marrow [68], [69]. Plasma cells are long-lived and produce antibodies for potentially a lifetime [50]. Early-activated plasmablasts can also differentiate into plasma cells which are short-lived. Memory B cells do not produce antibody. Memory B cells take

part in immune surveillance and patrol the periphery [68], [69]. In the secondary immune response, when activated, memory B cells can differentiate into long-lived plasma cells or GC B cells [68], [69].

Antibody cross-reactivity refers to the ability of an antibody to bind multiple epitopes that share structural similarities, rather than being specific to a single target [60], [75], [84]–[88], [76]–[83]. Broadly neutralising antibodies (bnAbs) antibodies can effectively neutralise a wide range of different influenza strains. More conserved regions of HA have been typically shown to elicit bnAbs. Multiple cross-reactive antibodies targeting the stem domain have been characterised [60], [77], [86]–[88]. Antibodies against conserved stem epitopes have been shown to cross-react with strains from different subtypes such as H1, H2, and H5. Anti-stem bnAbs typically block conformational changes involved in membrane fusion by binding HA0 [86], [87] or HA1 [60], [88]. The head domain is also targeted by bnAbs. Multiple bnAbs binding to the conserved RBS have been isolated. These antibodies typically inhibit HA-SA binding [84], [85] but have also been reported to inhibit membrane fusion, potentially by crosslinking HAs [75], [82], [83]. The VE subdomain of the head is more conserved than the RBD. It is not clear what function the VE undertakes in influenza A [36], [37]. Recently, multiple anti-VE bnABs have been isolated. They have been reported to interfere with HA-membrane fusion [78], [79] or elicit cell-dependent mechanisms [80], [81]. Finally, bnABs have been characterised that recognise an occluded HA1 epitope that is not external before membrane fusion [76].

7.8. Antigenic Imprinting or “Original Antigenic Sin” (OAS)

Since the 1960s, immune responses towards novel influenza strains have been known to be impacted by previous exposure. It was noted in Francis’ 1950s work that exposure to novel influenza strains tends to generate a stronger immune response to previously encountered, related strains, (Francis, T, 1960). The phenomenon was named “Original Antigenic Sin.” This often leads to a sub-optimal antibody response to the novel strain, as antibodies produced when meeting the novel strain are specific towards slightly different epitopes contained within the original strain [90], [91]. Another

observed phenomenon is the process of 'back boosting' where exposure to a previously encountered strain boosts the immune response to a subsequent, different (historical) strain as well as the original [50], [92]. OAS and back boosting can collectively be referred to as 'antigenic imprinting' [91]. Antigenic imprinting is seen in Influenza [50], [92] and SARS-CoV-2 [91].

Multiple mechanisms have been proposed to explain antigenic imprinting. One theory is that the response is based on the action of memory B cells, produced in prior infections. This suggests that these B cells are able to outcompete the induction of naïve B cells in the GC. This occurs by preferentially interacting with T_{FH} CD4+ T cells required for T-cell dependent activation of B cells, reviewed in [90]. This immune dominance results in a hierarchy of responses being mounted and often leads to a better or worse protective immune response against the same or closely related viruses.

7.9. Population immunity

Population immunity, also known as herd immunity, is a critical concept in epidemiology and public health that describes, in essence, the level of immunity within a population [93], [94]. Population immunity is thought to provide some level of protection against the spread of viruses. Population immunity arises from a combination of prior infections and vaccinations. Population immunity is thought to reduce the susceptible pool of individuals and potentially impedes the transmission of the virus through the population.

A.W. Hedrich proposed the term "herd immunity" to describe the observation that, when a sufficient proportion of the population became infected with measles, the virus would stop spreading before infecting the whole population [95]. The basic reproduction number denoted as R_o, was derived by using this concept as groundwork by MacDonald [96]. R_o is a key parameter in epidemiology used to estimate the average number of secondary infections produced by a single infectious individual in a completely susceptible population. It serves as a mathematical indicator of a pathogen's potential for spread. If R_o is greater than 1, each infected person is, on average, transmitting the virus to more than

one other person, indicating the potential for an epidemic. Conversely, if R_o is less than 1, the virus struggles to sustain transmission within the population.

The proportion of individuals in a population that needs to be immunised to achieve herd immunity is classically calculated using the formula: $P = 1 - \frac{1}{R_0}$ [93], [94]. This formula represents the threshold at which the virus is no longer able to sustain itself within the population. This has important use in public health. P can be used to estimate the number of people in a population that should be targeted in vaccination campaigns. It can also be used to determine how much risk a virus has of spreading in a certain population. It is important to note that the concept of P was generated after studying measles as a case study and may not apply to all respiratory pathogens.

7.10. Antigenic evolution of influenza – Antigenic Drift

Anti-HA antibodies that inhibit the influenza virus lifecycle reduce the virus' evolutionary fitness. Due to influenza's error-prone polymerase, HA mutations are produced during viral replication [97]. To increase evolutionary fitness, mutations allowing for immune escape are favoured [98]–[102]. In the presence of anti-HA antibodies, HA generates escape mutants. These mutations often accumulate in regions of the HA which are bound by antibodies, otherwise known as 'epitopes' [67]. The buildup of escape mutations in influenza is called antigenic drift [98]–[102]. Antigenic drift reduces vaccine efficacy and requires the seasonal vaccine to be updated [92], [103], [104].

There is a large amount of experimental evidence supporting antigenic drift. Strong evidence for drift comes from (i) antibody escape mutants [56], [98], [109]–[117], [99]–[102], [105]–[108], and (ii) haemagglutinin inhibition (HAI) assays showing a progressive divergence in the antigenicity of influenza strains over time [100], [103], [110], [111], [118], [119].

The experimental generation of escape mutants in the laboratory provides evidence that influenza can antigenically drift. Epitope mapping is a technique that elucidates the presence of epitopes on HA's surface [56], [105], [114]–[117], [106]–[113]. Epitope identification guides vaccine design by

pinpointing specific areas where the responses are focussed. This typically involves intentionally introducing mutations, usually by incubation with various mAbs, into HA to identify specific epitopes. This is done by allowing these mutated viruses to replicate and observing the emergence of escape mutations that evade the immune response mounted by the mAbs. Groups of escape mutants become identified as specific epitopes when researchers observe recurring mutations in specific regions of HA across isolates. In an analogous method, structural characterisation of multiple mAbs binding HAs can, when grouped, identify surface epitopes [56], [105], [111], [112]. Epitope mapping can involve computationally identifying sites under selection by grouping structural and genetic data together [109], [114].

Epitope mapping has been applied to identify H1N1 HA epitopes multiple times. Two classically cited studies that focus on HA's head domain are [106], [107]. Epitopes identified in these papers are used widely in the influenza field. These epitopes consist of Sa, Sb, Ca1, Ca2, and Cb (Figure 5 a). These cover the breadth of HA's head domain. Much influenza evolution is thought to occur through escape mutations in these antigenic sites. Since then, multiple stem antigenic sites have also been characterised [60], [77], [86]–[88].

The H5N1 ten phylogenetic clades (Group 0-9) are further split into subclades [31], [32], [33]. Clades have been shown to group together by serological studies. Cluster 1 comprises groups 0, 1, 3, 4, 5, 6, 7.1, and 9; Cluster 2 comprises groups 2.2.1, 2.1.3.2, 2.3.4, 2.4, 2.5 and 8; and Cluster 3 comprises groups 2.3.21 and 7.2 [120], [121]. Differences in antigenicity between H5N1 phylogenetic groups make the task of epitope mapping more challenging than within H1N1 [121]. This is because escape mutations commonly impact strains belonging to different antigenic clusters in diverging ways. For example, certain RBS proximal mutations cause large antigenic impacts in subclade 2.1, but less for other subclades [115]. The classical H5N1 HA head antigenic sites A, B, C, D, and E, were defined in 1989 using the clustering of escape mutations mapped onto the H3N2 HA crystal structure (Figure 5 b) [116], [117]. These sites span the HA head, comprising the RBD and VE domains. Since then, multiple epitope mapping [108], [110]–[117] and bioinformatics [109], [114] studies have identified additional

H5N1 escape mutant-containing sites. In addition to classical sites A-E, escape mutations have been found across the stem-containing HA1 and HA2 domains [108], [109], [111], [121]. Common amongst many studies are mutations proximal, or including, the RBS [108]–[117]. It is thought that RBS-containing/ proximal sites are likely cross-reactive between H5N1 antigenic clusters [121].

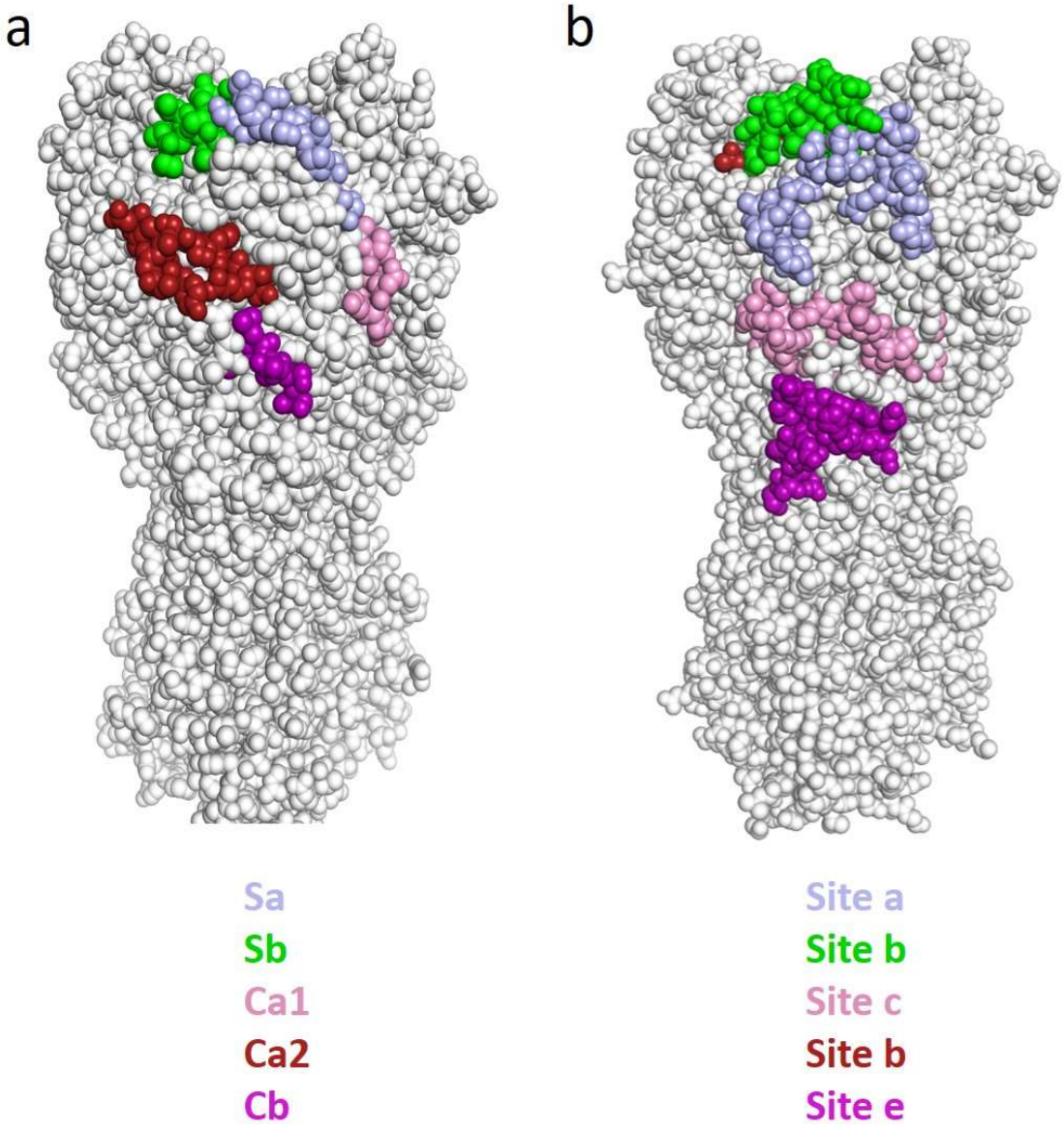


Figure 5. The five classical antigenic sites of the H1N1 [106], [107] and H5N1 [114] HAs. Figure a denotes the H1N1 HA and b denotes the H5N1 HA, crystal structures 1ruz and 5huf respectively.

HAI studies involving vaccinating naïve ferrets or mice with historical influenza strains also provide evidence for antigenic drift. Sera from vaccinated ferrets has been used in HA inhibition (HAI) assays to determine if it could neutralise historical strains. These studies tended to show that sera raised against a single virus only neutralised chronologically similar viruses [5]. Some studies have gone further and analysed large sets of HAI assays using antigenic cartography, finding that the evolutionary trajectory of influenza occurs in a linear manner, suggesting that antigenic distance increases over time, a key feature of antigenic drift [102].

However, Smith's seminal antigenic cartography study [103] was conducted using HAI data generated by fourteen-day post-immunisation antisera. This antisera obtained from primary infected ferrets may not reflect the natural condition because humans usually come into contact with many strains over their lives. Therefore, the accepted experimental method may not capture the original antigenic sin and affinity maturation characteristics (which occur after fourteen days) seen in the real world [122].

Indeed, Carter *et al.,* (2013) found that Haemagglutinin Inhibition (HI) assays performed on ferret antisera taken fourteen days after immunisation with a historical influenza strain does not cross-react with modern strains [123]. However, antisera taken 81 days after immunisation, enough time for affinity maturation, does show cross-reactivity with modern strains. Schmidt *et al.,* (2015) and Raymond *et al.,* (2018) show that human anti-HA antibodies can immunise against strains not previously seen after sequential exposure to divergent influenza strains [124], [125]. Ferrets were serially infected with different strains and immune responses were produced for novel strains [126]. Moreover, Linderman & Hensley, (2016) showed that mice infected with a certain influenza strain (strain one) could produce antibodies specific to strain two, a strain that the mouse was previously exposed to, with little specificity to strain one [127].

HA is attached to the virion's envelope by the stem domain's TMD [36]–[41]. The head domain is therefore more exposed to the host immune system [128]. Consequently, due to the greater selective pressure, mutations build up in the head domain faster than the stem domain, and as a result, the stem is less variable than the head domain [39]. Moreover, several studies used high-throughput PCR-based methods known as deep mutational scanning (DMS) to mutate every residue in the H1N1 HA to every other possible residue. They concluded that the stem domain of H1N1 also has an intrinsically lower capacity to vary than the head domain [129].

Despite being perceived as highly variable due to its rapid mutation rate and ability to generate escape mutants [98]–[102], influenza exhibits substantial genetic and antigenic diversity restrictions, with seasons often dominated by a single or limited number of strains [101], [102], [130]. This is highlighted by the 'ladder-like' morphology shown in Figure 6. Seasonal vaccines targeting specific strains offer considerable protection, highlighting a paradox within the influenza field [131]–[133]. The discrepancy lies in understanding why, despite its perceived variability, influenza seasons consistently witness limited strain dominance and successful vaccination efforts [134].

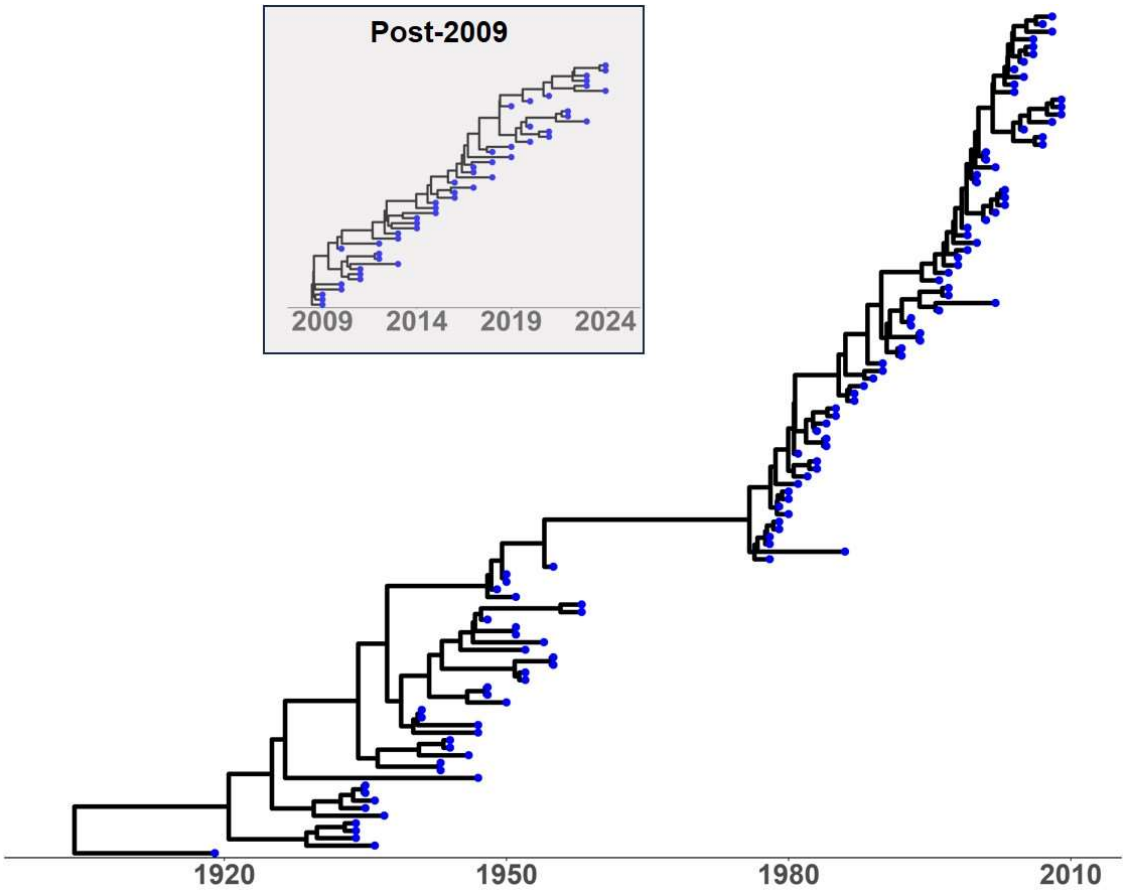


Figure 6. Maximum clade credibility trees of human-circulating H1N1 influenza virus HAs. H1N1 strains were sampled between the years 1918 and 2020. Three human-circulating strains per sampling date were, where possible, randomly sampled for each subtype.

Multiple studies have sought to account for the limited diversity of influenza within the context of antigenic drift. Rambaut *et al.,* (2008) proposed a concept of selective sweeps, where strains exported from regions with weak seasonality—like the tropics—into those with strong seasonality face heightened selective pressure, curtailing diversity [135]. Ferguson *et al.,* (2003) and Tria *et al.,* (2005) introduced mathematical models suggesting the necessity of non-strain-specific immunity for constraining strain diversity under antigenic drift [136], [137]. Koelle *et al.,* (2006) and Bedford *et al.,* (2012) explored the idea of clustered genotypes and genetic bottlenecks, respectively, as factors shaping influenza's evolutionary trajectory [138], [139]. Yuan and Koelle *et al.,* (2013) and Bush *et al.,*

(1999) highlighted the role of receptor avidity and specific genetic codons in influencing strain dominance and diversity [140], [141]. Morris *et al.,* (2020) suggested that within-host selection can substantially reduce influenza's diversity [142]. Despite these insights, consensus on the precise mechanisms governing influenza's antigenic evolution remains elusive, posing challenges for vaccine design.

7.11. Antigenic evolution of influenza – Antigenic recycling

In Francis' seminal work, he posited that influenza antigens were recycled over time [143]. Some justification for this was the presence of antibodies specific to the 1957 H2N2 pandemic strain in older people. The theory was that antigens present in the 1957 pandemic strain had spread before, and immunised older people when they were younger, through the mechanism of OAS. Some researchers have experimented with evolutionary theories of influenza that are similar to the recycling idea. The reason for this is to address some phenomena potentially unexplained using antigenic drift theory.

The Antigenic Thrift theory was proposed to explain the prevalence of single or limited strain dominance in influenza, reconciling this with phylogenetic data [144]–[146]. The theory posits that population immunity targets epitopes of limited variability (ELVs) rather than highly variable epitopes, constraining the diversity of influenza strains. As influenza strains spread and population immunity wanes, strains containing similar ELVs reappear (Figure 7). Experimental evidence supporting this theory has emerged, including studies isolating broadly neutralising antibodies from humans and serological studies in ferrets and mice.

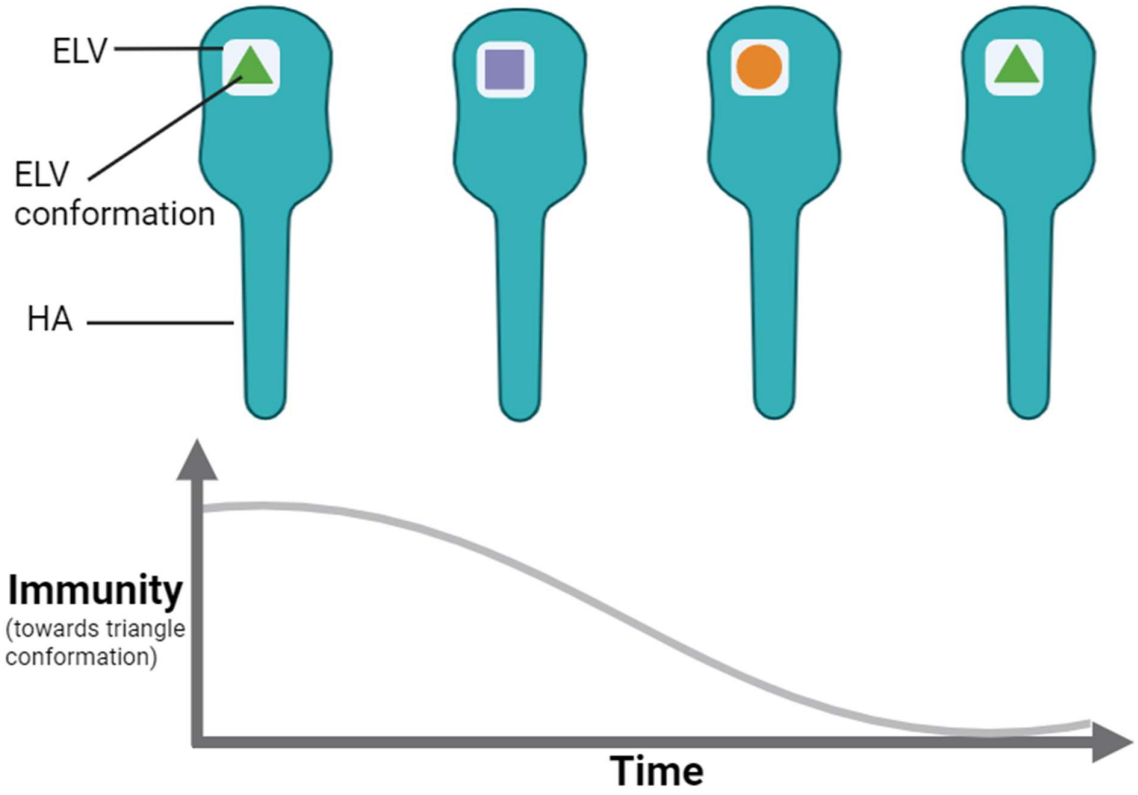


Figure 7. A representation of the antigenic thrift theory. The antigenic thrift theory attempts to explain how influenza viruses change over time to evade population immunity, which is directed towards a limited set of epitopes of limited variability (ELVs). This figure illustrates how one specific theoretical ELV, evolves. Individual conformations this single ELV takes are given by shapes. Initially, the population develops strong immunity against the triangle conformation. As a result, influenza strains carrying the triangle conformation become less common because they are easily neutralised by the immune system. However, over time, the ELV changes to other conformations than triangle to allow escape from the immune response. As a result, strains with different ELV conformations become more prevalent. As people die or move, immunity against the original triangle conformation decreases. Eventually, there's less protection against the triangle shape, allowing strains with this shape to reappear and become common again. This cycle continues as the virus and the population's immunity interact and evolve over time. Created using Biorender.com (accessed on 11 April 2024).

Thompson *et al.,* (2018) provided a significant contribution to this theory by presenting a bioinformatic thrift-based approach to identify less variable epitopes in the H1 HA head domain [147]. They collected sera from children aged 7–11 years old during 2006–2007, who had only been exposed to a limited number of strains. The sera were then used to show that epitopes identified in silico mediated immunity to historical strains. Finally, mice were vaccinated with chimeric HA constructs to elicit

antibodies specifically targeting the many variants of this epitope of interest. The reactivity seen in the sera from children was recapitulated in mice and demonstrated via pseudotyped virus microneutralisation assays and influenza virus challenge. In addition, alternative versions of the epitope were shown to react to a complementary subset of chronologically distinct influenza strains. These findings highlight the presence of epitopes in the HA head domain that cycle through a limited number of conformations, supporting the concept of antigenic thrift and its role in shaping influenza antigenic diversity.

Other studies have shown increased antibody repertoires targeting conserved regions of the HA, indicating the presence of antigenic thrift-consistent ELVs. Andrews *et al.,* (2015) identified cross-reactive antibodies targeting conserved regions of the HA, suggesting their potential for broad reactivity against multiple strains [148]. These findings underscore the role of ELVs in shaping influenza antigenic diversity and offer insights for vaccine design.

Some of the earlier discussed bnAbs exhibit characteristics potentially consistent with antigenic thrift [141], [159], [160]. Whittle *et al.,* (2011) identified a monoclonal antibody binding to the HA RBS, situated in the variable head domain [149]. This effectively neutralised 30 out of 36 distinct H1 strains. Similarly, Nogales *et al.,* (2018) isolated an antibody capable of neutralising both historical and contemporary H1 strains in various assays, including mouse challenge studies [163]. If broadly reactive antibodies akin to those in Whittle *et al.,* (2011) and Nogales *et al.,* (2018) are prevalent in the human population, they could target epitopes of limited variability proposed in the antigenic thrift theory. Interestingly, the antibody described in Whittle *et al.,* (2011) exhibits binding to a location similar to that identified in Thompson *et al.,* (2018) [147], [149].

7.12. Seasonal influenza vaccines

Understanding the antigenic evolution of influenza is pivotal to the design of effective vaccines. Presently, seasonal influenza vaccines endorsed by the WHO's Global Influenza Programme, include

one H1N1, one H3N2 influenza A, and one influenza B (IBV) Victoria strain [150]–[152]. The vaccines are available in multiple formats [150]–[152].

The Inactivated Influenza Vaccine (IIV) contains viruses that have been killed with chemicals. The virus is then broken up into smaller pieces using detergents (split-virus vaccines) or purified to contain just the hemagglutinin HA and NA proteins (subunit vaccines). The inactivated virus or its components cannot replicate. Immune responses are then elicited by viral proteins. Live Attenuated Influenza Vaccines (LAIV) are live viruses that contain mutations causing temperature sensitivity. These mutations replicate poorly at the higher temperatures found in the lower respiratory tract but replicate well in the cooler temperatures of the upper respiratory tract. Immune responses more similar to natural influenza infection than IIVs can be elicited with LAIVs.

The strains for the vaccine are selected by the World Health Organisation (WHO) approximately six months before the onset of the Northern or Southern Hemisphere influenza seasons [153]. Despite efforts, predicting the dominant strain for the upcoming season remains challenging, often leading to mismatches between the vaccine and the circulating strain [6], [132], [154]. The efficacy of seasonal influenza vaccines varies, influenced by factors such as the degree of mismatch and the choice of vector [132]. For instance, during the 2008–2009 flu season, IIVs displayed an average efficacy (VE) of 70%, while the LAIV had a VE of 38%, with the efficacy of IIV decreasing notably over time [155]. In the 2015–2016 flu season, LAIV efficacy dropped to a historic low of 3% among children aged 2–17 years in the United States, prompting the US Centres for Disease Control and Prevention (CDC) to advise against LAIV administration the following season [156]. The variable vaccine efficacy provides an impetus for the design of ‘universal’ influenza vaccines.

7.13. Attempts to develop “Universal” Influenza vaccines

Current seasonal influenza vaccines (1) necessitate the selection of vaccine strains approximately six months ahead of the influenza season, (2) exhibit variable efficacy, and (3) require regular updates (3)

[134] . In response, various strategies have been pursued to develop a universal influenza vaccine[6], [7]. The Universal influenza vaccine is a mostly theoretical concept. Proposed Universal influenza vaccines can be broadly categorised based on the target scope: pan-seasonal vaccines aim to provide protection across multiple influenza seasons; pan-subtype vaccines target multiple strains within a specific subtype (e.g., H1 or H3); pan-group vaccines cover all strains within a major group of influenza A (e.g., Group 1 or Group 2); and pan-influenza vaccines are designed to protect against all influenza A and B viruses.

One approach involves inducing stem-targeted antibodies through chimeric HA constructs or nanoparticle vaccines, as seen in studies by Nachbagauer *et al.,*[157]–[159] and Amitai *et al.,* (2020) [160]. Another avenue involves targeting internal viral antigens, such as nucleoprotein (NP) and matrix 1 (M1), to induce CD8+ [161]–[163] and CD4+ [164] T cell responses. Additionally, the NA [165]–[170] and M2 proteins have been targeted for vaccine development [171]–[179]. Despite promising results in preclinical studies, further research is needed to develop broadly protective influenza vaccines.

The antigenic thrift theory postulates that anti-influenza population immunity is directed towards a restricted set of epitopes of limited variability (ELVs). If accurate, this allows for the possibility of designing vaccines that target limited variability epitopes shared between multiple strains, therefore protecting against a large number of strains. Thompson *et al.,* (2018) identified epitope variants using bioinformatics and serology, targeting all possible variants of limited variability to protect against entire subtypes [147]. Similarly, the Ross group developed computationally optimised antigens (COBRA) for H1, H3, and H5 HAs, showing promising results in mice and ferrets [180]–[184]. DIOSynVax utilises bioinformatics tools to identify conserved immunogenic regions within HA [185].

7.14. Pandemics

The WHO lays out a six-phase framework for the generation of influenza pandemics, using the 2009 “swine flu” pandemic as a case study [186]. In phase, one animal influenzas circulate routinely

throughout their animal reservoirs. In phase two, zoonotic transmission of a virus into the human population occurs, causing concern for potential further spillover events. Whilst in phase three, multiple clusters of human cases with no sustained human-to-human transmission occur. Some limited human-to-human transmission may occur due to close contact between patients and caregivers. Phase four consists of "community level" outbreaks demonstrating sustained human-to-human transmission. At this stage, the pandemic risk is high. In phase five, two or more countries are affected by the pathogen. Stage six: the pandemic is underway; community level spread in two counties in different regions. The final stage, phase six, is commonly characterised by increasing population immunity and declining case numbers.

The concept of waves in pandemics refers to distinct phases or periods of increased disease activity followed by periods of reduced activity [187]. The "Spanish flu" pandemic occurred in three distinct waves [187]. This was similar to the characteristics of the COVID-19 pandemic [187]. Waves of infection may be influenced by ongoing viral evolution, stochastic effects, public health interventions, and novel therapeutics.

The initial zoonosis usually occurs due to the generation of a novel subtype. New subtypes emerge via 'antigenic shift', where genetic reassortment of the eight genomic influenza fragments occurs [188]. When antigenic shift occurs, there may be little population immunity to the rearranged subtype [188]. The low population immunity allows the virus to evade the immune response and efficiently replicate.

In order to seed a pandemic, a virus must have the ability to spread efficiently between humans [189]. H1N1 HAs bind human-type receptors characterised by 2,6-linked sialic acids prevalent in the human respiratory tract [190]–[195]. This specificity facilitates efficient human-to-human transmission. Possessing this phenotype is important for pandemic strains. H5N1 HAs exhibit a preferential binding affinity to avian-type receptors, characterised by 2,3-linked sialic acids found in avian hosts [190]–[195]. Concerningly, HPAI strains are capable of mutations allowing the specificity for both 2,3, and

2,6-linked sialic acids [25]–[29], which could indicate adaptation congruent with pre-pandemic changes.

Antigenic shift may occur in swine, which are susceptible to both avian and human influenza, as swine express both 2’6 and 2’3 sialic acid [196], [197]. In this instance, if more than one subtype were to infect a cell at the same time, it is possible vRNA gene segments may reassort, for example, an avian H5N1 strain gaining a human SA-targeting H1N1 HA [196], [197].

Multiple other factors also play a part in the generation of pandemics, including: (i) global connectivity and the ability for humans to travel long distances, (ii) population density, R0 is known to increase in densely populated areas, (iii) ecological changes such as urbanisation bringing humans closer to animal reservoirs; a risk factor in spillover events, (iv) healthcare infrastructure and its ability to treat, identify, and quarantine infected persons, (v) public health interventions such as the availability of vaccines and the implementation of lockdowns, and (vi) virology: the ability of virions to persist in the environment without degradation in more extreme climates [198]–[200].

Pre-existing neutralising immunity may also decrease the potential of a shifted virus to cause a pandemic in a certain population [52]. Therefore, determining the population immunity to a virus in a certain population may allow us to infer whether that population is close to Hedrich’s concept of “Herd Immunity;” and how likely that virus is to cause a pandemic. However, it should be noted that this is just one factor determining how likely a virus is to become pandemic.

With the development of vaccines with greater efficacy for H1N1 and H3N2 comes a heightened risk for immune escape if full immunity to all possible influenza variants is not conferred. Increased vaccine efficacy against entire subtypes could remove the H1N1 and H3N2 strains from circulation, potentially creating the right conditions for a significant zoonotic transfer leading to a pandemic [202]. This threat underscores the importance of working towards a universal vaccine, from which immune evasion would be less likely and also highlights that caution needs to be applied when deploying such vaccines [203].

7.15. Demographic patterns in influenza pandemics

Seasonal influenza mostly impacts the very young and very elderly, presumably due to weakened immune systems. In contrast, influenza pandemics have exhibited unusual age-based severity, mortality, and incidence patterns. “Spanish influenza” was characterised by a “W” shaped mortality curve: most mortality was distributed in the very young and early middle ages; less was seen in teenagers and older adults [204]. During the “Swine flu” pandemic, younger people were more affected than older people [205]. This resembled a “U-shaped” curve. Moreover, during the “Russian flu,” most mortality was seen in those younger than 26 [206].

Francis’ theory of antigenic recycling and the evolutionary models of antigenic thrift/ limited drift could partially explain this phenomenon. Potentially, a combination of childhood immune imprinting and affinity maturation on cyclical epitopes could cause unusual age-based characteristics. The serological studies from Carter *et al.,* (2013), Schmidt *et al.,* (2015), and Linderman & Hensley, (2016) showing that previous exposure to certain strains modulated the response against novel, divergent strains may support this hypothesis [123]–[125], [127]. Potentially, this could occur due to conserved epitopes shared between these strains, such as H1N1 residue 147 identified and Thompson *et al.,* (2018) [126] ,[147].

7.16. SARS-CoV-2 and the COVID-19 pandemic

COVID-19, is caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). The virus primarily targets the respiratory system, akin to influenza, leading to a range of symptoms from mild respiratory distress to severe pneumonia [207], [208]. SARS-CoV-2 was first identified in the Hubei province of China in December 2019 [209], [210]. SARS-CoV-2 is an enveloped virus belonging to the β-coronavirus family which comprises five currently known viruses, [211]. Three of these, MERS-CoV, SARS-CoV-1 and SARS-CoV-2 are zoonotic pathogens which have ‘jumped’ species, passing from an

animal host to humans [211]. The Spike (S) surface glycoprotein is the predominant surface protein on the virion and the primary antibody target (Figure 8).

The life cycle and structure of SARS-CoV-2 has been well-studied [212]–[215]. The structure of the SARS-CoV-2 virion, given in Figure 8, is composed of a lipid envelope derived from the host cell membrane, which encases the viral components. Within this envelope resides the nucleocapsid, formed by the nucleocapsid (N) protein tightly binding to the positive-sense single-stranded RNA genome, providing stability and protection to the viral RNA. The envelope is studded with spike (S) proteins that protrude from its surface, facilitating attachment to the host cell's ACE2 receptor and subsequent entry into the cell. Additionally, the envelope contains the membrane (M) protein, which is essential for virion assembly and maintaining structural integrity, and the envelope (E) protein, which facilitates the budding process by promoting the release of new virions and regulating the envelope's curvature.

The lifecycle, shown in Figure 8, begins with the virus's entry into a host cell, a process that can occur via two principal pathways: direct fusion and endocytosis (Figure 8 a and b). In the direct fusion pathway, the spike (S) protein of the virus binds to the angiotensin-converting enzyme 2 (ACE2) receptor on the host cell surface. This binding induces a conformational change in the S protein, facilitating the fusion of the viral envelope with the host cell membrane, thereby allowing the viral RNA to enter the cytoplasm. Conversely, in the endocytic pathway, the virus is internalised into the host cell through endocytosis, resulting in the formation of an endosome. Within the endosome, the acidic environment activates host proteases, such as cathepsin L, which cleaves the S protein, inducing a conformational change that leads to the fusion of the viral and endosomal membranes and the subsequent release of the viral RNA into the cytoplasm.

Following entry into the host cell, the viral RNA undergoes uncoating, which involves the separation of the viral RNA from the nucleocapsid, allowing the RNA to be accessible for translation. The viral RNA is then translated by host ribosomes to produce viral polyproteins. These polyproteins undergo

co-translational cleavage, a process mediated by viral proteases such as the papain-like protease (PLpro) and the main protease (Mpro), also known as 3CLpro. This cleavage results in the generation of non-structural proteins (nsps).

The nsps subsequently assemble to form the replication-transcription complex (RTC). The RTC operates in association with double-membrane vesicles (DMVs) derived from the endoplasmic reticulum (ER). The RTC synthesises a complementary negative-sense RNA template from the positive-sense viral RNA. This negative-sense RNA template is then utilised to produce new positive-sense genomic RNA. Additionally, the RTC generates sub-genomic RNAs (sgRNAs) that serve as templates for the synthesis of mRNAs encoding structural and accessory proteins.

The structural proteins, including spike (S), envelope (E), and membrane (M) proteins, are translated by ribosomes bound to the rough ER. These proteins undergo post-translational modifications and are transported through the ER-Golgi pathway to the ER-Golgi intermediate compartment (ERGIC). The nucleocapsid (N) protein, translated by free ribosomes in the cytoplasm, binds to the newly synthesised genomic RNA to form the nucleocapsid. The structural proteins and nucleocapsids then form virions are formed by budding together into the ERGIC lumen. These virions are subsequently transported to the cell surface in vesicles and released from the cell via exocytosis.

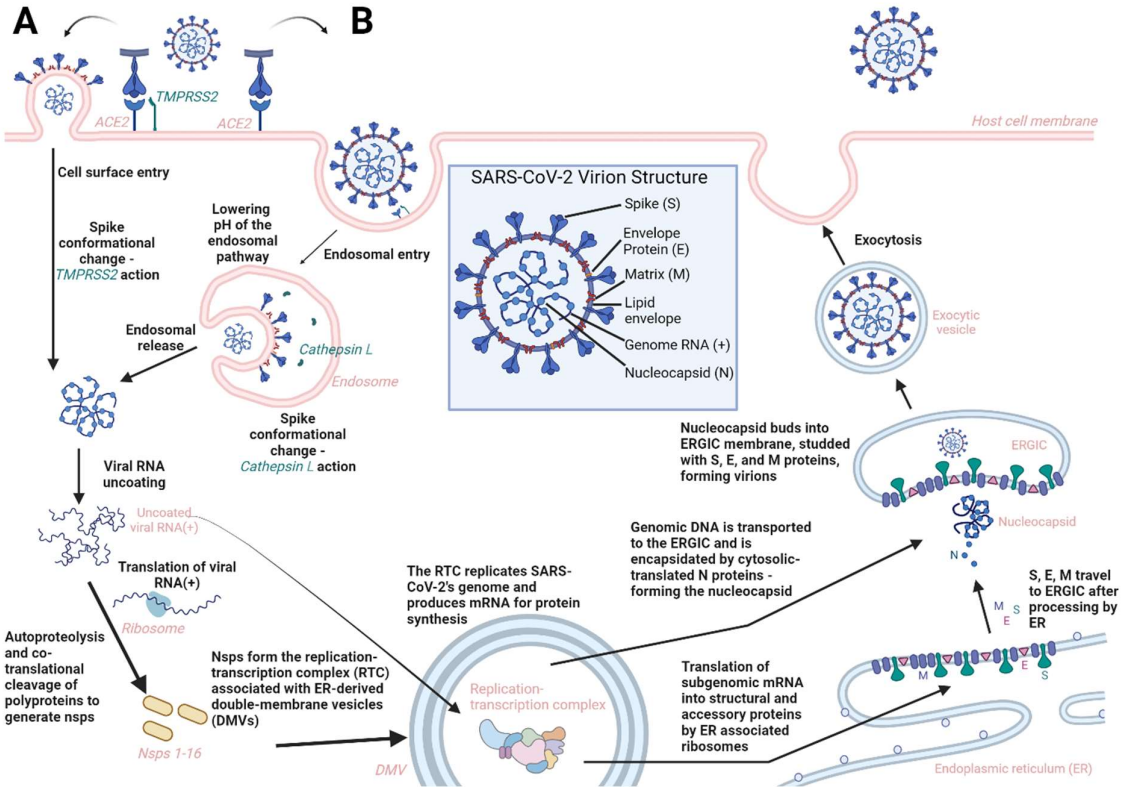


Figure 8. Schematic of the SARS-CoV-2 lifecycle, highlighting key stages of viral entry, replication, translation, assembly, and release. The branching pathways labelled A and B denote two different routes of viral entry: the cell surface (A) and endosomal (B) routes. The box in the centre represents the structure of the SARS-CoV-2 virion, showing the viral components within the lipid envelope. Created using Biorender.com (accessed on 5 November 2024).

Structurally, HA and the S protein share key features that underpin their roles in viral entry [216]. Both are glycoproteins anchored to the viral envelope, containing domains responsible for binding to host cell receptors and mediating membrane fusion. The S protein is trimeric, formed from three identical monomers (Figure 9 a). The Spike protein functions in an analogous manner to HA. The S protein of SARS-CoV-2 contains a cleavage site that is critical for viral entry (Figure 9 c). The S protein is cleaved into two subunits, S1 and S2, during the infection process. The S1 subunit harbours the receptor-binding domain (RBD), responsible for binding to the host cell receptor angiotensin-converting enzyme 2 (ACE2), while the S2 subunit mediates the fusion of the viral and host cell membranes (Figure 9). The fusion peptide mediates the fusion of the viral envelope with the host cell membrane, a

fundamental step in the initiation of infection (Figure 9). The activation of the S protein through proteolytic cleavage is a key step in both the cell surface and endosomal viral entry pathways and is essential for the successful initiation of infection [212]–[215].

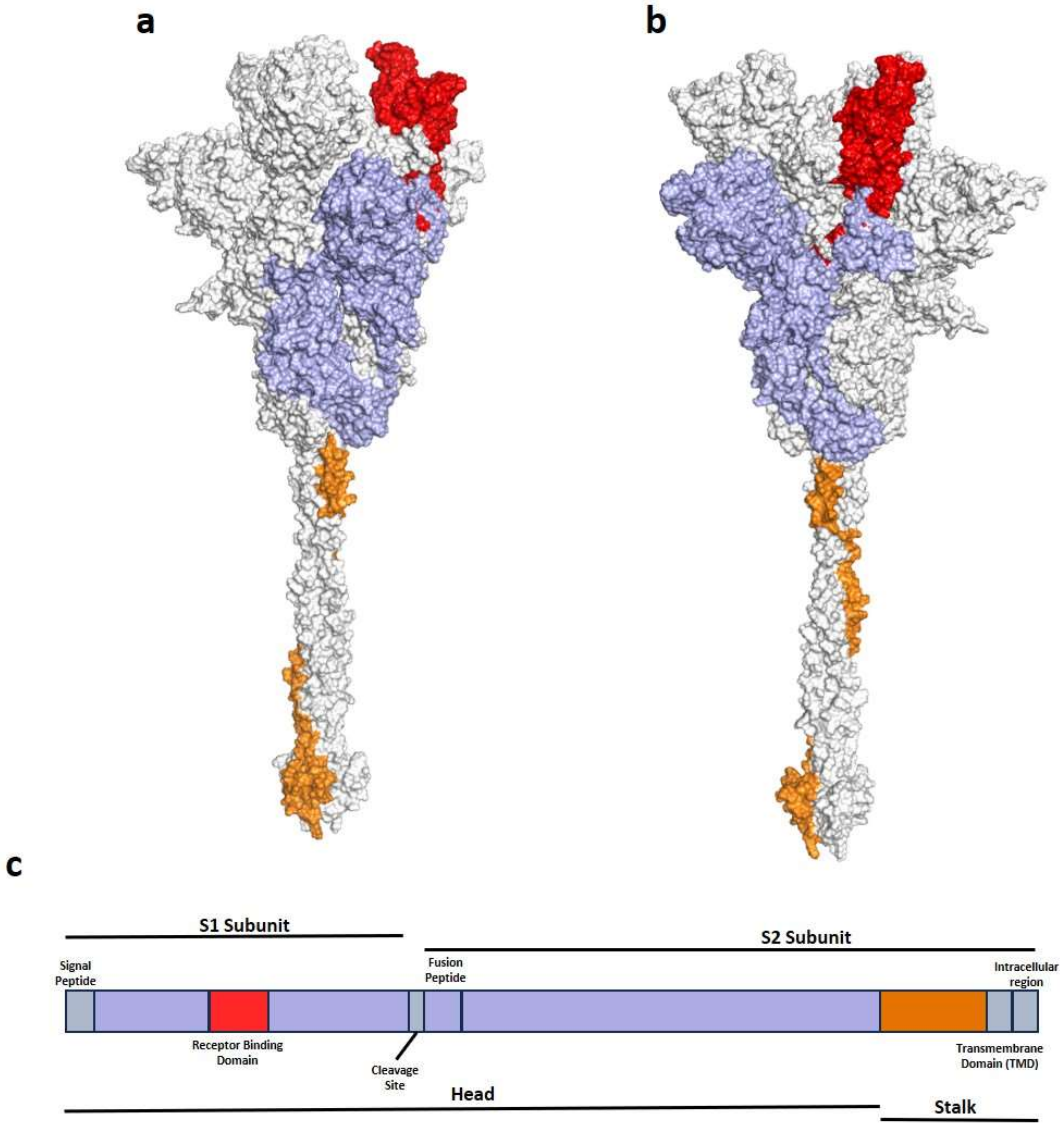


Figure 9. The structure of Spike is shown on the 6vsb_1_1_1_S309 model. Figures a-b show the left-right view of the Spike trimer. Figure 9 c shows a representation of the Spike gene sequence, split into the S1 and S2 subunits as well as the head and stem domains. Colouring is consistent between the crystal structures and amino acid sequence. One monomer is coloured.

Similarly to influenza's HA, neutralising anti-Spike antibodies interfere with the host-cell entry process [217]–[219]. Antibodies have been shown to bind epitopes within the RBD to stop Spike-ACE2 binding. Moreover, antibodies may bind epitopes in the S2 subunit, blocking conformational change or binding

the fusion peptide, interfering with membrane fusion. SARS-CoV-2 is known to undergo antigenic drift, with multiple mutations building up in the RBD.

Pseudotypes containing the Spike can also be produced and used in pseudotype neutralisation assays [220], [221]. Large-scale serosurveys could be conducted using this method to determine population immunity during the pandemic [220].

7.17. The pseudotype microneutralisation assay and other serological assays

The primary serological assay employed in this project is the Pseudotype Microneutralisation (pMN) assay. Pseudotyped viruses allow the study of viral entry, neutralisation, and vaccine efficacy without the need for handling pathogenic viruses. I used a three-plasmid system (Figure 10) to produce replication-incompetent pseudotyped viruses. This system was reviewed comprehensively in [222].

The p8.91 plasmid is a modified HIV-1 vector that lacks the packaging signal (Ψ) and several envelope genes (env, vif, nef, vpu, and vpr) (Figure 10). It utilises the cytomegalovirus (CMV) promoter for high-level expression. The p8.91 plasmid provides the structural proteins (Gag) and enzymes (Pol), including integrase and reverse transcriptase, necessary for forming the pseudoviral core.

The pCSFLW plasmid contains the firefly luciferase gene, which acts as a reporter. Like the p8.91 plasmid, it is also under the control of the CMV promoter. The pCSFLW plasmid supplies the reporter gene and packaging signals, including the long terminal repeats (LTRs), Rev response element (RRE), and packaging signal (Ψ) sequences, which were originally from HIV (Figure 10). These elements are essential for packaging the luciferase gene into the pseudovirus, ensuring that the pseudoviruses carry the luciferase gene for detection.

The pCDNA3.1 plasmid contains the HA gene from influenza, which is incorporated into the viral envelope (Figure 10). This plasmid provides the HA protein that allows the pseudovirus to bind and enter host cells.

The design of this pseudotype virus system incorporates multiple safety features to ensure that the produced viruses are replication-incompetent and safe for laboratory use. The p8.91 plasmid lacks the packaging signal (Ψ), preventing the encapsidation of the viral genome and rendering the pseudoviruses unable to replicate. The removal of env, vif, nef, vpu, and vpr genes from the p8.91 plasmid further ensures that the pseudoviruses cannot produce the proteins necessary for new viral particles.

The cytomegalovirus (CMV) promoter is used in both p8.91 and pCSFLW plasmids for high-level expression. The CMV promoter replaces the original HIV LTR promoter, adding a layer of safety by removing the HIV-specific promoter elements. The pCSFLW plasmid is designed as a self-inactivating (SIN) vector, which includes a deletion in the 3' LTR (U3 promoter region). This modification ensures that the integrated reporter gene cannot be reactivated by cellular promoters, preventing unintended expression.

The production of pseudotype viruses involves the co-transfection of HEK293T cells with the three plasmids (p8.91, pCSFLW, and pCDNA3.1) using polyethyleneimine (PEI). This process initiates the production of pseudoviruses as follows: The plasmids are introduced into HEK293T cells, where they travel to the host cell nucleus and are expressed. The Gag and Pol proteins from p8.91 form the viral core, while the luciferase mRNA from pCSFLW is packaged into the pseudovirions due to the Ψ sequence. The HA protein from pCDNA3.1 is incorporated into the viral envelope. The assembled pseudoviruses bud off from the cell membrane, acquiring the HA protein in their envelope. Exogenous NA is then added which cleaves HA bonds between the pseudotyped viruses and the host cell membrane.

When these pseudoviruses infect a target host cell during an assay, the HA protein on the pseudovirus envelope binds to receptors on the host cell surface, facilitating viral entry. Inside the host cell, the viral core releases the luciferase mRNA, which is reverse-transcribed and integrated into the host genome by the packaged integrase enzyme. The CMV promoter drives the expression of the luciferase gene, resulting in the synthesis of luciferase enzyme. The presence of luciferase is detected using a luminometer after adding the Bright-Glo reagent, which produces a luminescent signal proportional to the level of infection.

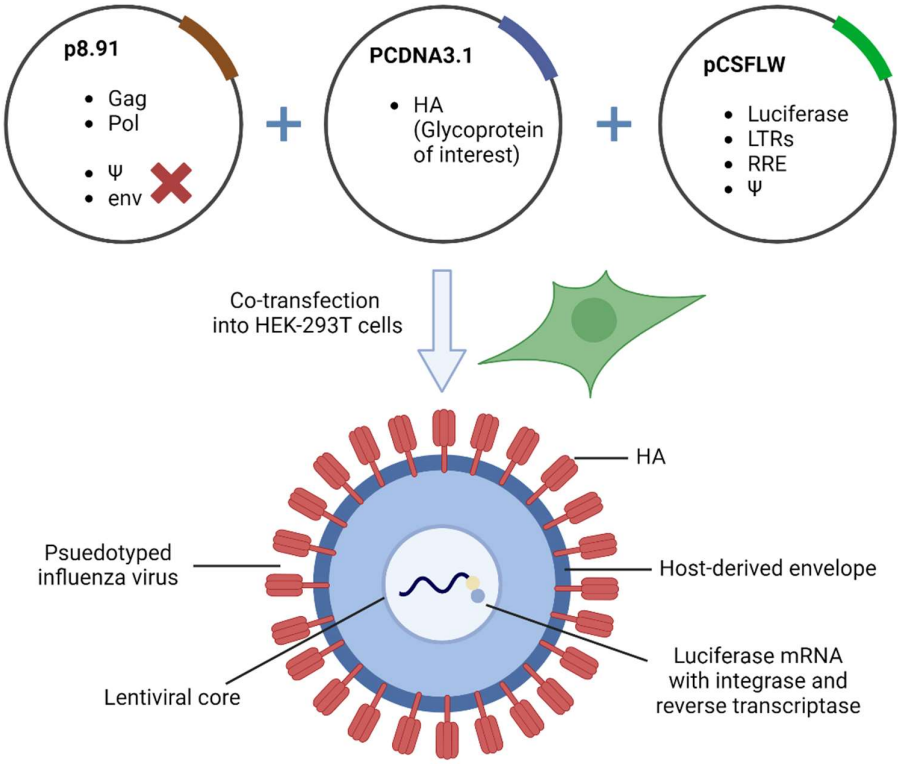


Figure 10. A schematic of the three plasmid pseudotyped virus production system. The top circles detail the three plasmids used to produce pseudotyped viruses: P8.91, pCSFLW and PCDNA3.1. Labels inside the plasmids denote features of note. Below is a representation of the structure of a pseudotyped influenza virus displaying the HA glycoprotein. Created using Biorender.com (accessed on 5 November 2024).

Multiple assays can be employed to measure the presence, concentration, and functional capabilities of antibodies to influenza or SARS-CoV-2. These assays range from traditional techniques

like Microneutralisation (MN) and Western Blotting to more advanced methods such as Surface Plasmon Resonance (SPR). Each assay type offers distinct advantages and focuses on different aspects of the immune response, utilising specific reagents, measurement units, and equipment. Table 1 provides a comparison of these serological assays, highlighting their unique methodologies and applications.

Assay Type	Measurement Focus	Reagents/Components	Measurement Units	Equipment
Microneutralisation (MN) [223]	Ability of serum antibodies to neutralise live virus, assessing viral replication through cytopathic effects or viral RNA quantification.	Virus strain, serum samples, target cell line (e.g., MDCK cells for influenza, Vero E6 for SARS-CoV-2)	IC50 (µg/ml or dilution)	Luminometer, cell culture plates
Pseudotype Neutralisation (pMN) [222]	Assessment of serum neutralisation against pseudotyped viruses expressing viral proteins (HA for influenza, spike for SARS-CoV-2).	Pseudovirus, serum, target cells (e.g., HEK293T), luciferase reporter system	IC_{50} (µg/ml or dilution)	Luminometer, cell culture plates
Western Blotting [224]	Detection of specific antibodies against viral proteins.	Viral proteins, serum, secondary antibodies conjugated to detection enzymes or fluorophores	Visual	Gel electrophoresis system, membrane transfer equipment
Lateral Flow Assays (LFAs) [225]	Rapid detection of viral antigens or antibodies, antigen-recognising antibodies bind to and hold antigen in place.	Detection antibodies, viral antigens or serum, gold nanoparticles or coloured latex beads	Visual	None (visual readout)
Chemiluminescent Immunoassays (CLIAs) [226]	Enhanced sensitivity and dynamic range compared to ELISAs, using chemiluminescent substrates to quantify target antigens or antibodies.	Antigens, serum, enzyme-conjugated secondary antibodies, chemiluminescent substrates	Arbitrary Units (AU)	Chemiluminescence imager
Enzyme-Linked Immunosorbent Assay (ELISA) [227]	Quantitative measurement of the binding strength and presence of specific antibodies against viral proteins.	Antigens (e.g., recombinant HA proteins for influenza, spike proteins for SARS-CoV-2), serum, enzyme-conjugated secondary antibodies, substrates (e.g., TMB)	Optical density (OD)	Microplate reader

Surface Plasmon Resonance (SPR) [228]	Real-time analysis of biomolecular binding interactions without labels.	HA proteins for influenza, spike proteins for SARS-CoV-2, purified antibodies, buffer solutions	Resonance units (RU)	SPR instrument
Flow Cytometry [229]	Multi-parameter analysis of individual cells, assessing specific antibodies on cell surfaces and characterising cell-mediated immune responses.	Cells expressing HA proteins for influenza or spike proteins for SARS-CoV-2, fluorescently labelled antibodies	Mean fluorescence intensity (MFI)	Flow cytometer
Single Radial Haemolysis (SRH) [230]	Measurement of haemolytic zone area indicating the presence and concentration of influenza-specific antibodies.	Haemolysis buffer, sheep red blood cells (SRBCs), influenza antigen, guinea pig complement, antisera	mm^2 (square millimetres)	SRH plates, incubator, light box or viewer, callipers or image analysis software

Antibody Pulldown [165]	Assessment of specific antibody-antigen interactions.	HA2 stalk proteins for influenza, spike proteins for SARS-CoV-2, specific antibodies, magnetic beads or agarose beads	Relative abundance	Magnetic separation system

Table 1. Comparison of Serological Assays for Influenza and SARS-CoV-2. This table summarises various serological assays used to evaluate antibody responses against influenza and SARS-CoV-2. Each assay type is detailed with its measurement focus, reagents/components, measurement units, and required equipment.

When compared to the various serological assays available, Pseudotype Neutralisation (pMN) assays provide a unique balance of safety, sensitivity, and specificity in assessing the neutralising antibody response against viral infections. While Microneutralisation (MN) assays directly measure the ability of serum antibodies to neutralise live viruses, they present significant biosafety concerns due to the need for handling pathogenic strains. In contrast, pMN assays mitigate these risks by employing pseudotyped viruses, allowing researchers to evaluate neutralisation without direct exposure to live pathogens. Moreover, they allow the direct measurement of the function of a glycoprotein of interest, without confounding effects deriving from other proteins in a live virus. Enzyme-Linked Immunosorbent Assays (ELISA) and Chemiluminescent Immunoassays (CLIAs) provide valuable quantitative data on the presence and concentration of antibodies but do not directly assess neutralisation capabilities, limiting their utility in understanding functional immune responses. Surface Plasmon Resonance (SPR) offers detailed kinetic and affinity information on antibody interactions but requires specialised equipment and technical expertise, which may not be readily accessible in all laboratories. Flow cytometry enables multi-parameter analysis of individual cells,

providing insights into cellular immune responses, yet it is more complex and resource-intensive compared to the pMN assay. Single Radial Haemolysis (SRH) is limited in its application, as it specifically measures the presence of influenza-specific antibodies, relying on the haemolytic properties of hemagglutinin (HA) that do not translate to the spike protein of SARS-CoV-2. Overall, while each assay contributes to a comprehensive understanding of the immune response to viral infections, pMN stands out for its balance of safety and specificity, enabling informed decisions regarding public health and vaccination strategies.

To specifically assay the serological response towards the stem domain of HA, pMN assays are not applicable, as viable pseudotypes do not form without including the head domain. Instead, antibody pulldown assays can be employed to isolate antibodies specific to the HA stem domain by immobilising the stem domain on magnetic beads. Following this, the bound antibodies can be eluted and characterised using various downstream assays, including pMN. Western blotting can be adapted to specifically identify antibodies against the HA stem domain by utilising the recombinant stem domain as the antigen in the assay. Furthermore, Enzyme-Linked Immunosorbent Assays (ELISAs) can be adapted to quantify antibodies specifically against the stem domain by coating microplate wells with recombinant stem domain protein, enabling the measurement of antibody levels in serum samples. Immunofluorescence assays (IFAs) can visualise the binding of antibodies to the stem domain expressed on infected cells, offering qualitative insights into the antibody response. Additionally, Lateral Flow Assays (LFAs) can be developed for rapid testing of stem domain-specific antibodies; however, they may exhibit limited sensitivity compared to more sophisticated assays. Surface Plasmon Resonance (SPR) can be utilised to measure the binding kinetics and affinities of antibodies targeting the stem domain, providing detailed information on the strength of these interactions.

7.18. Conclusion

Determining the level of population immunity towards SARS-CoV-2 was essential after and during the first pandemic wave. The level of population immunity was not known. If many undetected asymptomatic cases had already occurred, the population of the UK may have been near herd immunity. This would have profound implications, potentially rendering public health policies such as lockdowns or vaccine roll-outs unnecessary.

Some H5N1 influenza viruses have been highlighted by the WHO as having pandemic potential [231] [232]. It is therefore important to assess the population immunity to H5N1 strains in order to assess their pandemic potential and prepare for future pandemics. Likewise, previous H1N1 strains have caused pandemics and continue to circulate, so it is important to determine how the seasonal circulation of H1N1 viruses impacts the population immunity to current and historical H1N1 strains.

The H5N1 and H1N1 HA are phylogenetically (Figure 11) and structurally (Figures 3 and 11) similar. Because of this, I hypothesise that there could be structurally conserved epitopes shared between the H5N1 and H1N1 subtypes. If this is possible, it may enable us to identify specific epitopes that induce a cross-reactive immune response contained within the H5N1 and H1N1 HAs in order to design vaccines theoretically capable of protecting against future pandemics and be used as a broadly protective or 'universal' influenza vaccine.

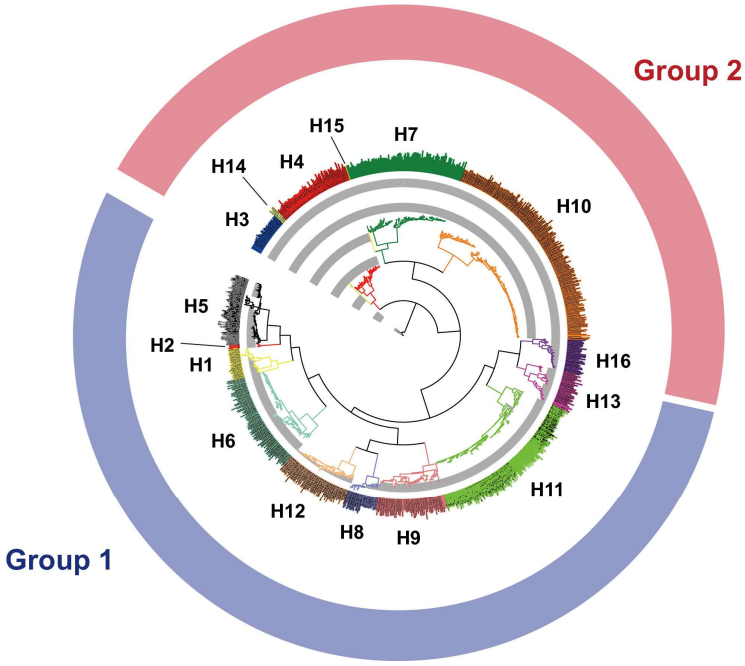


Figure 11. Adapted from [133]. A circular phylogenetic tree depicting the evolutionary relationships between different HA subtypes. Group 1 and 2 HAs are denoted by purple and pink semicircles respectively.

The purpose of this project is twofold: (i) to understand population immunity to SARS-CoV-2, highly pathogenic avian H5N1, and historical H1N1 viruses; and (ii) to identify and characterise ELVs present in avian H5N1 strains that may also be found in human H1N1 strains.

7.19. Aims

1. Determine the extent of population immunity to the novel SARS-CoV-2 virus in Scotland during and after the first COVID-19 wave of infection.

2. Perform serological studies to determine if population immunity exists towards potential pandemic avian H5N1 and historical H1N1 strains using a representative cross-section of the adult human as well as the wild avian population.

3. To identify potential shared epitopes of limited variability in the H5 and H1 HAs using bioinformatics.

4. Mutate potential epitopes of limited variability to confirm their existence using serological assays in H5N1/ H1N1-reactive cohorts of humans identified in 1.

8. Methods

8.1. Immunological techniques

8.1.1 Pseudotyped virus production

Virus pseudotyping refers to the generation of replication-deficient viruses encoding a viral protein of interest [147], [233], [234]. A lentivirus-based pseudotyped virus system was used to display the HA protein on the pseudotyped virus surface. Pseudotyped viruses were generated by transfecting HEK 293 T/17 cells (ATCC) with 1.0 µg of HA protein (in plasmid pcDNA3.1/ pi.18), 1.0 µg of gag/pol (plasmid p8.91) and 1.5 µg of a luciferase reporter construct (plasmid pCSFLW) as part of a plasmid-OptiMEM/PEI solution. Transfections were performed in 10 ml of media DMEM 10% FCS, 1% penicillin–streptomycin, 20% L-glutamine and left for 24hrs at 37°C 5% CO2. Fresh media was added to the cells before leaving them at 37°C 5% CO2 for 48hrs. Supernatant was then harvested and stored at -80°C.

The same lentivirus-based system was used for generating SARS-CoV-2 pseudotypes. Two days before transfection, HEK 293 T/17 cells (ATCC) were transfected with 0.5µg ACE2. This was done using a synthetic codon optimised SARS-COV-2 Spike glycoprotein expression construct on its surface. The methodology has previously been used to produce pseudotyped viruses for Ebola, SARS-CoV-1 and MERS-CoV [235] [233]; [236].

8.1.2 Titration of pseudotyped virus

Titration is essential for calculating the optimal amount of pseudotyped viruses to add when conducting neutralisation assays. This process ensures that the inoculum added is within a defined range of 10^5 relative light units (RLU).

Supernatant is extracted from transfected HEK293T cell flasks and passed through a 0.45-micron cellulose acetate filter to remove cellular debris. An 8-fold serial dilution (1:2) is then prepared in duplicate across a 96-well plate using DMEM media as the dilution reagent.

The pseudotype microneutralisation assay, as described in Methods 8.1.3, is then set up to assess the RLU intensity. Pseudotyped viruses are diluted to a concentration that ensures the resulting measurement is no less than 10^5 RLU.

8.1.3 Pseudotyped virus microneutralisation assay

Pseudotyped viruses that successfully enter target cells integrate their luciferase-encoding genetic material into the host genome using the integrase and reverse transcriptase enzymes as outlined in Introduction 7.17 and [222]. The host cell then produces the luciferase enzyme, and quantifying the luciferase present in the host cells allows us to determine the amount of pseudotyped virus that has entered.

5µl of sera was added to 45µl DMEM media (a 1:20 dilution) containing 1% penicillin-streptomycin, 1% L-glutamine and 10% foetal calf serum before being mixed with 50μl of 10^5 relative light units (RLU) of pseudotyped virus. This was incubated at 37°C for two hours and then mixed with 10^4 HEK 293T ACE2-transfected cells per well. Plates were incubated for 72 hours at 37°C.

To measure luciferase activity, the supernatant from the 96-well plates is removed, and 50µL of Bright-Glo reagent (Promega, UK) is added to each well. The reaction is incubated for 5 minutes at room temperature and pressure (RTP). The Bright-Glo reagent lyses the cells, releasing the luciferase enzyme. Bright-Glo contains luciferin, the substrate for luciferase, and ATP, which is necessary for the subsequent enzymatic reaction. The luciferase enzyme catalyses the oxidation of luciferin, resulting in the production of a quantifiable light signal. This light signal is measured in relative light units (RLU) using a luminometer (Promega, UK).

8.1.4 Serial Dilution

To assess the neutralisation capacity of sera against pseudotyped virus, serial dilution assays were conducted. Serial dilutions were prepared for each serum sample. The dilutions were labelled by their logarithmic base 10 values to facilitate subsequent analysis. For each ID, this was done in two replicates, split over 4 different 96 well plates in order to better control for plate-specific confounding variation.

Serial dilutions were either performed using the Thompson Lab's automated Robotic platform utilising an Opentrons liquid handling robot or manually. When performed using the liquid handling robot, 6-fold dilutions of each sample are spread over 4 plates, in duplicate. When performed manually, as in the Site-Directed Mutagenesis work (Chapter 12), 8-fold serial dilutions were spread over 2 plates, in duplicate.

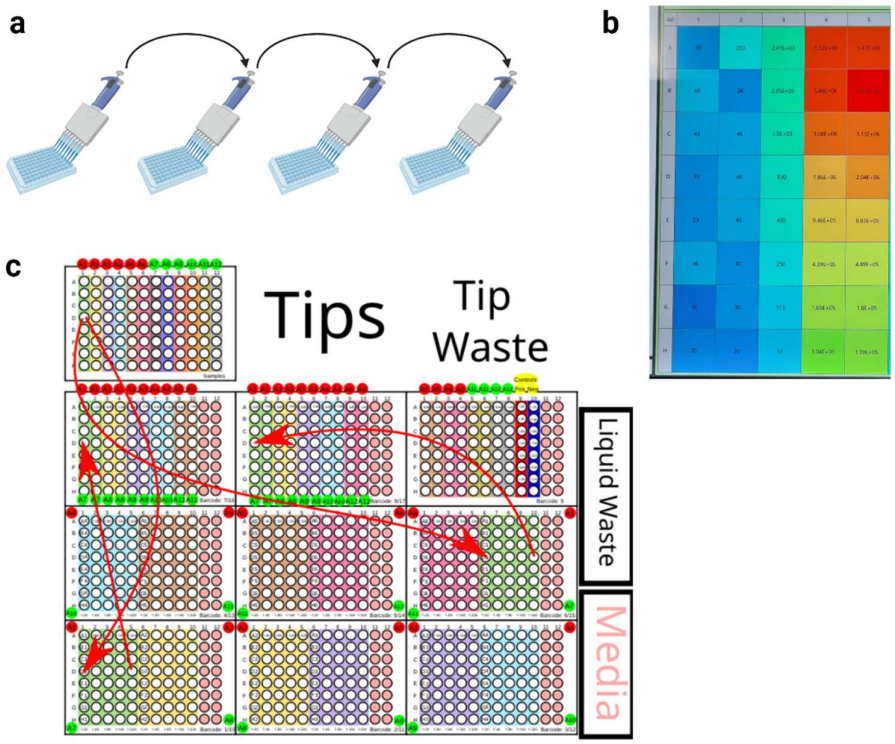


Figure 12. A is a representation of manual two-fold serial dilution using a multi-channel pipette. Figure 12 b is the output of a manually-titrated pseudotyped virus displaying the HA of A/bar-headed goose/Qinghai/3/2005/H5N1. This is a photograph of the luminometer GUI (Promega, UK), (RLU). Columns 4 and 5 were seeded with 50µl pseudotyped virus and serially diluted 1:2 down the plate with DMEM media. Figure 12 c is a schematic of the movement of the robotic platform when setting up a pseudotyped virus microneutralisation assay. Donor samples are loaded, one per well, into a Nunc-Immuno 96 well sample plate. The platform then adds media to each plate and distributes the samples into test plates, denoted by the arrows emanating from the sample plate i.e. setting up the serial dilutions. The robot is able to pipette fully, 34 plates, totalling 98 samples in approximately 1.5 hours. This schematic was provided by Dr Nicolas Grayson.

8.1.5 Standardisation

This method for determining the IC_{50} and subsequent error calculations has been peer-reviewed and has been included in multiple publications [91], [220]. The raw data of each measurement is expressed as Relative Light Units (RLU) which measures the level of fluorescence. The RLU values obtained from the serial dilution assays were standardised to express the neutralisation responses as percentages. Standardisation was done using two plate-wide controls: (i) virus and cell-only control, "Plate Mean Technical Positive", a plate-wide mean of eight virus-only control wells to assess the amount of pseudotyped virus luminescence with the absence of sera neutralisation, (ii) cell only control, "Plate Mean Technical Negative", a plate-wide mean of eight virus only control wells to control for the amount of cell only background present in the assay. This standardisation process was carried out using the following equation:

$$Standardised = 1 - \frac{Value - \text{ Plate Mean Technical Negative}}{\text{Plate Mean Technical Positive} - \text{ Plate Mean Technical Negative}}$$

8.1.6 Determining neutralisation strength

Neutralisation strength is expressed as the Half-Maximal Inhibitory Concentration (IC_{50}) value of sera samples. IC_{50} values are calculated by fitting a logarithmic curve to the standardised data. For each sample ID, IC_{50} measures the dilution factor at which the RLU is reduced by half. The logistic model used to fit the RLU-dilution factor data is described by the following equation:

$$Standardised = \frac{1}{1 + e\ (\frac{xmid - \log_{10} Dilution\ Factor}{Scal})}$$

$xmid$ denotes the $\log_{10}$ Dilution Factor in which the logistic curve reaches its midpoint.

$Scal$ describes how quickly or gradually the curve rises from its minimum to maximum values. The smaller the $Scal$, the gentler the slope will be.

The logistic curves were fitted using the R package "nls.multstart", calling the function "nls_multstart". The function applies the Levenberg-Marquardt nonlinear least-squares optimisation algorithm in order to generate a curve that best fits the data. The function iteratively adjusts both $xmid$ and $Scal$ for 100 replicates until convergence: when more iterations are not likely to further minimise the model's error. Curve selection consists of minimisation of the objective function, simply, the curve selected is the one displaying the minimum sum of squared residuals between the observed and predicted values. The specific objective function is referred to as the sum of squared residuals. This process selects the model with the minimum distance between the observed and predicted values. When expressed as an equation, the objective function can be defined as:

$$f(\theta) = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

θ refers to a model parameter, such as $xmid$

y_i denotes $Standardised$, the response variable

$\hat{y}_i$ denotes the predicted model value

n is the total number of datapoints

i denotes the vector index

The IC_{50} value is given as the $\log_{10}$ dilution factor at which the RLU is reduced by half, denoted in the equation by $xmid$. If no curve with an adequate error can be fit, then the sample does not receive and IC_{50} value; no IC_{50} value is computed.

The standard error of the IC_{50} value is calculated as the square root of the variance of the IC_{50} estimate.

$$Standard\ Error(IC_{50}) = \sqrt{Var(IC_{50})}$$

The variance of the IC_{50} estimate is obtained from the diagonal element of the variance-covariance matrix, which is derived from the fitted logistic curve model.

$$Var(IC_{50}) = V[IC_{50}, IC_{50}]$$

Specifically, the variance-covariance matrix is calculated as the inverse of the Hessian matrix (H).

$$V = (H^{-1})$$

The Hessian matrix is computed from the second-order partial derivatives of the objective function with respect to the parameters of the curve. The Hessian matrix captures the curvature of the objective function surface with respect to changes in the parameters of the curve. Essentially, this allows us to determine how the rate of change of the objective function varies after simultaneously varying parameters. In the following equation, $xmid$ would be one of the parameters included when calculating the variance of IC_{50}:

$$H_{ij} = \frac{\partial^2 f}{\partial \theta_i \, \partial \theta_j}$$

θ_i and θ_j denote a pair of parameters in the logistic regression model.

8.1.7 Robotic automation of pseudotype neutralisation assay

Population level pseudotype microneutralisation assays are carried out using a robotic platform, Figure 12 c. The Opentrons robot pipettes out media and carries out 6-fold serial dilutions using a

Python script obtained from Dr Nicolas Grayson. Samples are divided into two repeats and spread out to four separate plates in order to minimise single-plate confounding effects.

8.1.8 Enzyme-Linked Immunosorbent Assay

ELISA was employed to detect antibodies against the trimeric spike protein. MAXISORP immunoplates (442404; NUNC; Merck, Darmstadt, Germany) were initially coated with StrepMAB-Classic (2–1507–001; IBA Life Sciences, Göttingen, Germany). Following this, the plates underwent blocking with 2% skimmed milk in phosphate buffered saline (PBS) for 1 hour, then incubated with either 0.125 µg of soluble SARS-CoV-2 trimeric spike protein or 2% skimmed milk in PBS. After another hour, plasma samples were added at a 1:50 dilution, followed by alkaline phosphatase (AP)-conjugated anti-human IgG (A9544; Merck, Darmstadt, Germany) at 1:10,000 dilution or AP-conjugated anti-human IgM (A9794; Merck, Darmstadt, Germany) at 1:5,000 dilution. The reaction was initiated by the addition of p-nitrophenyl phosphate (PNPP, Merck, Darmstadt, Germany) substrate and halted with NaOH. Optical density (OD) units were measured at 405 nm after 1 hour. This method was developed using [237] as a reference. Background OD (calculated using milk) was subtracted from the sample's OD readings. As this was performed before much characterisation on how OD relates to the concentration of analytes, we did not relate the OD readings to a standard curve.

8.2. Bioinformatics

8.2.1 Mutational entropy

The concept of Shannon entropy provides a quantitative measure of uncertainty [238], or variability, in a sequence alignment. Higher Shannon entropy values indicate greater mutational diversity, with many different amino acids observed at that position [239]. Lower entropy values suggest conservation, denoting fewer amino acid mutations. Shannon entropy is calculated using the following formula for every amino acid position:

$$H = -\sum_{i=1}^{n} p_i \, log_2(p_i)$$

p_i denotes the probability of observing a specific amino acid

n denotes the total number of distinct amino acids observed at that position

i denotes a vector index

8.2.2 Structural Bioinformatics Pipeline

A pipeline written in R version 4.2.2 enables the variability of putative antibody binding sites (ABS) on the surface of a given trimeric protein to be determined. The pipeline takes two data inputs: i) an amino acid alignment and, ii) a crystal structure. This is based on the structural bioinformatic pipeline used in [147].

Swiss-PdbViewer V 1.4 is used to remove all internal amino acids from the structure. Then, in an automated pipeline utilising R version 4.2.2, Python version 3.10.1 and Pymol version 4.6.0, amino acid crystal structure coordinates are matched with the positions of the amino acid alignment.

The programme iteratively takes each position in the alignment and uses it as the central position in a putative ABS. It then produces a list of amino acids within a distance of an input area, measured

in Å². Sites measuring 400 Å², 600 Å², 800 Å² and 1000 Å² are generated based on empirical antibody binding sites identified from[240]. This can be described using the following equation:

$$abs = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2 + (z_i - z_j)^2} \leq input\_area$$

i, j refers to positions i and j, in which i is taken as a central residue, and i is every other amino acid compared to i

x, y, z refer to x, y, and z coordinates derived from the crystal structure .pdb file

For each putative abs arrived at using the previous function, the script computes the amino acid variability through the metric of mean Hamming distance. Mean Hamming distance is the total number of differences between the number of sequences, divided by the total number of sequences. The calculation is given by this equation:

$$MeanHam_i = \frac{1}{N_i} \sum_{k=1}^{N_i} d(resmat_i[k])$$

$resmat_i$ defines a matrix containing sequences derived from the amino acid alignments at the positions given by abs_i

N_i refers to the total number of sequences in $resmat$

k denotes the index of each sequence within $resmat$

$d(resmat_i[k])$ denotes the Hamming distance between k and all other sequences in $resmat$

i denotes an index of abs

I also applied the structural bioinformatic pipeline to deep mutational scanning data from [241]. This allows me to determine the mean mutational tolerance, expressed as Shannon diversity bits, for every putative abs. The calculation is given by the following equation:

$$MeanPreference = \frac{1}{N}\sum_{i=1}^{N} Bits_i$$

N refers to the number of amino acid positions in abs_i

$Bits$ denotes mutational tolerance; Shannon entropy measured in bits

i denotes an index of abs

8.2.3 Protein structure modelling

Crystal structure pdb files were computed from HA amino acid sequences using the MultiFOLD server, which provides results of comparable quality to AlphaFold2 [242], [243].

8.2.4 Structural alignments

PyMOL version 4.6.0 was utilised to generate structural alignments between HAs. Using PyMOL's alignment tools, structural superposition was performed to align the HA based on their atomic coordinates

8.2.5 Computing pairwise root-mean-squared deviation (RMSD)

A Python script was written to facilitate the structural analysis of protein mutations using PyMOL. The script computes the pairwise root-mean-square deviation (RMSD) between corresponding residues of two protein structures, enabling the comparison of their structural differences. This calculation is expressed by the following equation:

$$RMSD = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - y_i)^2}$$

N denotes the total number of atoms in a residue

x_i and y_i refer to the Cartesian coordinates of the atom i in the template and target respectively

8.2.6 Adding scale bars to differentially coloured crystal structures

I used the "spectrumbar.py" [244] script to add scale bars.

8.2.7 Generating HA sequence databases

I generated large, comprehensive, databases of HA amino acid sequences for the influenza subtypes H5Nx and H1N1. These databases contain 16,145 and 9,340 HA amino acid sequences, and contain 16,222 and 18,503 nucleotide sequences, respectively. I obtained HA nucleotide sequences from Gisaid, Genbank, NCBI, and fludb.org. I also obtained amino acid sequences in the same manner. I employed a reference-based alignment strategy using the lightweight version of the MAFFT online server to align the nucleotide sequences. I did this by utilising a high-confidence 50 HA sequence aligned using MAFFT Version 7.0 [245] as a reference. These sequences were then translated using the R package "Biostrings." All amino acid sequences were then compiled. Duplicated sequences were removed using "Biostrings." I then employed the reference-based lightweight version of the MAFFT online server to align the large amino acid datasets. Sequences were aligned to a reference generated using a structural alignment of 50 divergent amino acid sequences. I then utilised Regex tools from R version 4.2.2 to extract metadata including, for example, host, subtype, evolutionary clade, date of isolation, country, and region.

8.2.8 Specifying amino acid electrostatic properties

Amino acid alignments were converted to comma-separated variables (.csv) files. A script written in R version 4.2.2 converted each amino acid into a classification of its physical property, according to [246]. The properties of each amino acid are outlined in Table 2.

Amino Acid Symbol	Amino acid	Physical Property
A, I, L, M, V	Alanine, Isoleucine, Leucine, Methionine, Valine	Hydrophobic Side Chain – Aliphatic
F, W, Y	Phenylalanine, Tryptophan, Tyrosine	Hydrophobic Side Chain – Aromatic
N, C, Q, S, T	Asparagine, Cystine, Glutamine, Glutamine, Serine, Threonine	Polar Neutral Side Chain)
D, E	Aspartic Acid, Glutamic Acid	Acidic
R, H, K	Arginine, Histidine, Lysine	Basic
G	Glycine	Unique
P	Proline	Unique

Table 2. Details of the electrostatic properties of amino acids.

8.2.9 Phylogenetics

Random samples of sequences from the HA sequence databases, outlined in Methods 8.4.1 were generated using R version 4.2.2. To ensure optimal alignment, nucleotide sequences were ungapped, aligned using mafft version 7, and then codon aligned using translatorX [247]. Conserved blocks were

recovered using trimAl v1.2 [248]. Nucleotide substitution models including specific conversion rates, gamma category count, gamma shape, and the proportion invariant were all selected using ModelTest-NG [249], [250]. Time-calibrated or non-time-calibrated phylogenetic trees were constructed using the Bayesian Markov Chain Monte Carlo (MCMC) approach utilised in BEAST 2.5 [251] [252]. H1N1 and H5N1 subtype trees were generated using the Bayesian Skyline Coalescent along with a relaxed molecular clock [253]–[255]. Trace files were examined, and appropriate burnin percentages were selected using Tracer v1.7.1 [256]. Maximum Clade Credibility (MCC) trees were compiled using TreeAnnotator [251] [252]. Trees were visualised and annotated using the R package "ggtree" [257]–[259].

8.3. Statistical analysis

8.3.1 Modelling the seroprevalence of SARS-CoV-2 antibodies in Scotland

A longitudinal analysis of antibody titres conducted in early 2020 suggested that sample seropositivity would not wane during our study period [220]. Using this assumption, I was able to model population seropositivity using binomial logistic regression. Statistical analysis of the Scottish data set was undertaken using R version 4.2.2. I conducted this analysis using multiple combinations of variables including health board, week, day, postcode, sex, county, and multiple. The final model chosen was arrived at through model validation techniques such as the comparison of AIC scores, residual analysis and Hosmer-Lemeshow goodness of fit tests.

Donors were defined as seropositive or seronegative using the threshold arrived at in Results 9.3.1. A database was constructed consisting of every sample tested. Each sample was assigned a win/ loss value of 1 if seropositive and 0 if seronegative, as defined by pseudotype microneutralisation assay. These binary data were modelled in the logistic regression using the base R function "glm." The week was given by the week of sample collection following the 17th of March. Health Board (Scottish region) describes the NHS health board that the sample was collected from. The model is given by the following equation:

$$logit(p_i) = \beta_0 + \beta_1 Week_i + \beta_2 Health\_Board$$

β_0 represents the intercept (in the unit of log odds)

β_1 and β_2 denote regression coefficients estimating the effect of each predictor variable on the log odds of the response variable

p_i denotes the probability of success (a sample testing positive for neutralising antibodies) at index i. The mathematical definition of p_i is given by the formula:

$$p_i = \frac{1}{1 + e^{-(\beta_0 + \beta_1 Age_i + \beta_2 Hea \quad \_Board_i)}}$$

The $logit$ link function is given by the equation:

$$logit(p_i) = \log\left(\frac{p_i}{1 - p_i}\right)$$

The objective function of the binomial logistic regression is called the negative-log likelihood function. This is derived from the binomial distribution. Simply put, the objective function minimises the discrepancy between the predicted probabilities and the actual outcomes given by the data. This can be given by the equation:

$$Objective\ Function = -\sum_{i=1}^{n}(y_i \log(p_i) + (1 - y_i)\log(1 - p_i))$$

y_i is the outcome given by the data

p_i is the predicted probability

n denotes the number of observations

i is the vector index

Determining whether the predictor variable has a significant impact on the observed data is done through the calculation of the p-value. The "glm" function calculates the p-value using the Wald statistic. The Wald statistic Under the null hypothesis, which states there is no significant relationship, the Wald statistic follows a chi-squared distribution with 1 degree of freedom. The p-value determines the likelihood of the observed Wald statistic arising through chance. It is calculated using the equation:

$$Wald\ Statistic = \frac{\hat{\beta}}{SE(\hat{\beta})}$$

$\hat{\beta}$ here refers to the estimated gradient of the model

The logistic regression's McFadden Pseudo-R^2 describes how well the model fits the data; it calculates the proportion of deviance explained by the model relative to the deviance of the null model. The statistic was computed using the formula:

$$McFadden\ pseudoR^2 = 1 - \frac{D}{D_0}$$

D denotes the deviance of the fitted model

D_0 denotes the deviance of the null model, describing no significant impact of the predictor variable on the gradient The final model's McFadden Pseudo-R2 was 0.59; this can be roughly considered as explaining 59% of the variation.

I used the base R function 'predict.glm' to generate a vector of predicted variables based on the input data. The standard error of these predicted probabilities is calculated using the Delta method. The calculation can be expressed using the following formula:

$$SE(p_i) = \sqrt{X_i^T(V_i)X_i}$$

X_i denotes the vector of predictor variables

V_i denotes the variance-covariance matrix of the estimated coefficients from the model

X_i^T denotes the transposition of X_i

i denotes the vector index

8.3.2 Analysing relationships between neutralisation and age

I determined whether neutralisation strength and percentage neutralisation significantly varied with age. I tested for both linear and non-linear relationships. When modelling neutralisation strength, I examined $\log_{10}(IC_{50})$ instead of IC_{50} for multiple reasons. IC_{50} values commonly vary by multiple magnitudes within the same experiment. This means that the distribution of IC_{50} values can be skewed, with extreme values exerting disproportionate influence when modelling. Taking the logarithm means that the values are compressed, likely leading to a more symmetrical distribution. This approach also allows for a more intuitive inference of model statistics.

Relationships between neutralisation strength $\log_{10}(IC_{50})$ can be modelled using Gaussian statistics. However, relationships between the percentage of donors neutralising per age group cannot be analysed through Gaussian statistics. Percentage neutralisation describes the number of successes (e.g., positive outcomes), and failures (e.g., negative outcomes) for a given age. Gaussian methods assume normality and constant variance in the data; these assumptions are violated by binomially distributed data. Percentage seropositive was modelled as the chance of seropositivity.

8.3.3 Linear regression

I assessed whether any statistically significant linear relationships were present. I did this by fitting linear models using R's base function "lm." The linear model can be expressed mathematically as:

$$y_i = \beta_0 + \beta_1 \, . \, x_i + \varepsilon_i$$

The "lm" function employs the ordinary least squares (OLS) method as its objective function. That is, OLS minimises the difference between the observed and predicted values for the model. This can be expressed mathematically as:

$$\sum_{i=1}^{n} (y_i - (\beta_0 + \beta_1 \, . \, x_i))^2$$

y_i denotes the $\log_{10}(IC_{50})$

x_i denotes the age

β_0 denotes the intercept

β_1 denotes the gradient

ε_i denotes the error term, calculated for every value using the OLS objective function

i denotes the vector index

I repeated this process using weighted and unweighted models. I weighted the linear models using the standard error $\log_{10}(IC_{50})$ associated with each $\log_{10}(IC_{50})$ value. This approach, whilst still assuming a normal distribution, allows for the possibility of residuals displaying different variances. This can be expressed mathematically as:

For weighting method one:

$$\varepsilon_i \sim N(0, \frac{1}{SE_{y_i}})$$

For weighting method two:

$$\varepsilon_i \sim N(0, \frac{1}{SE_{y_i}^{\ 2}})$$

I then performed model validation including assessing residual plots and AIC criterion scores to determine the most appropriate model to use.

In order to determine the statistical significance of the impact of age on the linear model, I computed a p-value for the gradient coefficient β_1. Here, the null hypothesis assumes that age has no statistically significant impact on β_1. To do this, the lm function calculates the t statistic for β_1. This can be expressed mathematically as:

$$t = \frac{\hat{\beta}_j}{SE(\hat{\beta}_j)}$$

Where $\hat{\beta}_j$ denotes the estimated coefficient for the jth predictor variable.

Using the cumulative distribution function of the t-distribution, the p-value associated with the t-statistic is calculated. Simply, the p-value presents the probability of observing the t-value seen.

In order to assess how much variance is captured by the linear model, I computed the R^2 statistic. Simply, this describes how well the model describes the data. The "lm" function calculates this using the equation:

$$R^2 = 1 - \frac{SSresidual}{SStotal}$$

SS denotes the sum of squared residuals, which quantifies the differences between the observed and predicted values. This is calculated using the OLS objective function.

I used the base R function 'predict' to generate a vector of predicted variables based on the input data. The estimated residual variance $\hat{\sigma}^2$ is calculated as the residual sum of squares divided by the degrees of freedom (df). The standard error of the output (the "fit") is calculated by the square root of the estimated residual variance. This can be expressed mathematically by:

$$SE(\hat{y}) = \sqrt{\hat{\sigma}^2}$$

$\hat{y}$ is the predicted value

$\hat{\sigma}^2$ denotes the variance of residuals

8.3.4 Logistic regression

Logistic regression was employed to model associations between age and neutralisation chance with the assumption that the relationship may be approximately linear. I used base R's "glm" function to fit binomial logistic regressions to the data. We can express the regression mathematically using the following formula:

$$logit(p_i) = \beta_0 + \beta_1 \times Age_i$$

β_0 is the intercept

β_1 is the regression coefficient which estimates the effect of each predictor variable on the log odds of the response variable

i is the vector index

p_i denotes the probability of success (a sample testing positive for neutralising antibodies) at index i. The mathematical definition of p_i is given by the formula:

$$p_i = \frac{1}{1 + e^{-(\beta_0 + \beta_1 \times Age_i)}}$$

The mathematical definitions of the: logit link function, unweighted objective function, calculation of Mcfadden pseudo R^2, "predict.glm" function, and regression coefficient p-value are the same as those detailed in Methods 8.3.1. The model validation methods used are also detailed in Methods 8.3.1.

8.3.5 Generalised Additive Modelling

To assess the statistical significance of non-linear age-neutralisation strength relationships, I employed generalised additive models (GAMs) from the R package "mgcv" [260], [261]. GAMs allow for the modelling of flexible, non-linear relationships between neutralisation strength and age. I used model evaluation techniques to assess the appropriateness of the models generated. These included: (i) the examination of residuals-fitted value plots, Q-Q plots, and Partial Dependence Plots (PDPs), (ii) comparisons of the Root Mean Squared Error (RMSE), adjusted R^2, and AIC scores, (iii) sensitivity analysis by varying model parameters, and (iv) only including ages with adequate numbers of donors i.e. not including ages with only 1 donor, exhibiting extreme values. We can illustrate the GAM structure using this equation:

$$log_{10}(IC_{50}) = f(Age) + \varepsilon$$

$f(Age)$ denotes the smooth function applied to the continuous Age variable. This is specified by the REML algorithm, which selects appropriate smooth terms, as discussed below.

ε represents the error term, inferred from the differences between the observed values and values fitted by the model. ε is used to estimate the parameters of the model including the selected smooth term(s)

Smooth terms capture potentially complex non-linear relationships between predictor variables and the response variable. The 'mgcv' package utilises penalised splines to do this. Cubic splines are fitted to the data and penalise the smoothness to prevent overfitting.

The objective function for GAMs aims to maximise the log-likelihood function while adjusting for the estimation of fixed effects and removing bias associated with using the likelihood function alone. This involves maximizing the logarithm of the likelihood of the model parameters given the data while penalising overly complex models to prevent overfitting. Mathematically, the objective function can be represented as:

$$Objective\ Function(\theta \mid data) = \ \mathrm{logL}(\theta \mid \mathrm{data}) - \frac{1}{2}\log \mid \mathrm{Var(y)} \mid$$

θ denotes the model parameters

$\mathrm{logL}(\theta \mid \mathrm{data})$ denotes the log-likelihood function

$\mathrm{Var(y)}$ denotes the variance-covariance matrix of the response variable

The mgcv package in R employs the restricted maximum likelihood estimation (REML) algorithm to fit GAMs. REML selects appropriate smooth terms for the model. It optimises the objective function by removing bias associated with estimating fixed effects, resulting in reliable model estimates. In summary, REML ensures that the GAM captures the underlying patterns in the data while avoiding overfitting.

As with linear modelling, I constructed unweighted as well as weighted GAMS. The weightings I used can be expressed mathematically by:

$$Objective\ Function(\theta \mid data) = \ \mathrm{logL}(\theta \mid \mathrm{data, w}) - \frac{1}{2}\log \mid \mathrm{Var(y, w)} \mid$$

For weighting method one:

$$w = \left(\frac{1}{SE}\right)$$

For weighting method two:

$$w = \left(\frac{1}{SE^2}\right)$$

GAMs can be used to analyse binomially distributed data, such as percentage neutralisation vs age, by using a logit link function. I did this to assess non-linear neutralisation percentage-age relationships, utilising the REML algorithm. The objective function of the binomial GAM model aims to maximise the log-likelihood function while accounting for the binomial nature of the response variable. This model allows us to analyse non-linear relationships in binomially distributed data. The objective function can be captured by the same equation as seen above for the Gaussian GAM. However, the variance-covariance matrix $\mathrm{Var}(\mathrm{y})$ reflects the binomial nature of the response variable.

The linear predictor of the GAM represents the linear combination of predictor variables and model parameters in the model's equation. It includes both linear terms (such as main effects) and smooth terms (such as smooth functions). Mathematically, we can represent the linear predictor as:

$$\eta = f(Age) + \epsilon$$

η represents the linear predictor

However, when adding a logit link function, explained in Methods 8.3.1, the linear predictor is transformed:

$$\eta = \mathrm{logit}(p_i)$$

p denotes the probability of success

$\mathrm{logit}(p)$ denotes the log odds of success

For all GAM models, the adjusted R^2 values and p-values of the smooth term were calculated using the summary.gam() function from the mgcv package. The adjusted R^2 captures the proportion of variance explained by the model, where both the original variance and residual variance are estimated using unbiased estimators. This method takes into account the number of parameters in the model and can be negative. It can be expressed mathematically by:

$$R^2 = 1 - \frac{\mathrm{Var}(y - \hat{y})}{Var(y)} \times \frac{n-1}{n-k-1}$$

$Var(y)$ denotes the total variance of the response variable

$\mathrm{Var}(y - \hat{y})$ denotes the residual variance, estimated using unbiased estimators

n denotes the number of observations

k denotes the number of model parameters, excluding the intercept

The p-value represents the probability of observing a test statistic at least as extreme as the one calculated from the data, under the null hypothesis that the coefficient is equal to zero (i.e., no effect of the smooth term on the model). The p-value of the smooth term is calculated using a Wald-type test, excluding uncertainty in smoothing parameter estimates. This can be expressed mathematically as:

$$p - value = P(|Z| > \frac{|\hat{\beta}|}{SE(\hat{\beta})}$$

Z is a standard normal random variable

$\hat{\beta}$ denotes the estimated coefficient of the smooth term

$SE(\hat{\beta})$ is the standard error of the estimated coefficient

In order to generate plots, I used the 'predict.gam' function from the mgcv package to generate a vector of predicted variables based on the input data. The fit of the predicted values has an associated

standard error, calculated from the square root of the estimated variance of the prediction. The parameter variances are calculated from the variance-covariance matrix generated by the mgcv package. Mathematically, the standard error of the predicted value for observation be expressed using the equation:

$$SE(\hat{y}_i) = \sqrt{Var(\hat{y}_i)}$$

$\hat{y}_i$ is the predicted value

$Var(\hat{y}_i)$ denotes the estimated variance of the predicted value

8.3.6 General statistical inference

All statistical inference was computed at the 95% confidence, 0.05 significance level.

8.3.7 Differences between sample means

The statistical significance of differences in sample means was calculated using the "t_test" function from the R package "rstatix" [262]. This can be expressed by the following equation:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\frac{S_1}{\sqrt{n_1}} + \frac{S_2}{\sqrt{n_2}}}$$

$\bar{x}$ denotes the sample mean

S denotes the standard deviation

n denotes the sample size

$1,2$ denotes group 1 and 2 respectively

P-values represent the probability of observing a t-value as extreme or more extreme than calculated, assuming that the null hypothesis is true. This is determined by whether the t-statistic is greater than

the t-critical value ($t_{\alpha/2}$). $t_{\alpha/2}$ is specified by the significance level (α). This is done under examination of the t-distribution. For a given degrees of freedom (df), calculated by $n-1$, the $t_{\alpha/2}$ value is the value at which the probability density of obtaining that t value is $\leq \frac{\alpha}{2}$.

If the p-value is less than or equal to 0.05, we can say that the difference in sample means is greater than that expected by chance. When performing multiple comparisons, p-values were modulated using the Benjamini-Hochberg (BH) multiple comparison correction in order to control for Type 1 error:

$$\hat{p} = \min\left(1, \frac{mp}{r}\right)$$

$\hat{p}$ denotes the adjusted p-value

p denotes the original p-value

m denotes the number of comparisons in comparison with all ps

r denotes the rank of p

8.3.8 Confidence intervals

For Gaussian distributed data, I constructed 95% confidence intervals to show the possible values the sample mean could take at the 0.05 significance level. I calculated these using the following equation using R version 4.2.2:

$$CI = \bar{x} \pm t_{\alpha/2} \times \left(\frac{s}{\sqrt{n}}\right)$$

$t_{\alpha/2}$ is the t-critical value specified above

$\bar{x}$ denotes the sample mean

s denotes the standard deviation

n denotes the sample size

For binomially distributed wins/ losses data, I calculated Bayesian confidence intervals using the R function “binom.confint” from the “binom” package [263]. This is done using the Markov chain Monte Carlo (MCMC) sampling approach to calculate 95% confidence intervals.

8.3.9 Pearson’s correlation

The Pearson correlation test quantifies the strength and direction of the linear relationship between two continuous variables. This was calculated using the base R function “cor.” The calculation is given by the equation:

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}}$$

r denotes the Pearson’s correlation statistic, the closer the value is to 1, the more positive the relationship between the two variables

x_i and y_i refer to the two variables

$\bar{x}$ and $\bar{y}$ refer to the means of the two variables

n denotes the sample size

As with t-tests, the p-value of r is calculated using the $t_{\alpha/2}$ value, derived from the t-distribution with the df of $n - 2$. The t-value is calculated using the equation using symbols defined above:

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$$

8.4. Cohorts and datasets

8.4.1 Genetic data

All available nucleotide and amino acid HA sequences were obtained from fludb.org, gisaid, NCBI, and GenBank for influenza subtypes H5Nx and H1N1. These databases contain 16,145 and 9,340 HA amino acid and contain 16,222 and 18,503 nucleotide sequences respectively. H5 sequences of all N subtypes were compiled and treated as pan-H5 in analysis unless specified.

8.4.2 Pseudotyped surface glycoproteins

Table 3 includes the names of all sequences used in this project.

Sequence name	Accession number	Subtype	Year collected	Host
A/bar-headed goose/Qinghai/3/2005/H5N1	HM172454	H5N1	2005	Avian
A/chicken/Scotland/1959/H5N1 (codon optimised)	CY015081	H5N1	1959	Avian
A/South Carolina/1/1918/H1N1	AAC57065	H1N1	1918	Human
A/PR/8/1934/H1N1	CCH23213	H1N1	1934	Human
A/USSR/90/1977/H1N1 (codon optimised)	AAA43240	H1N1	1977	Human
A/Solomon Islands/3/2006/H1N1 (codon optimised)	ABU99109	H1N1	2006	Human
A/California/07/2009/H1N1	AEE69009	H1N1	2009	Human
A/Viet-Nam/1203/2004/H5N1	AAW80717	H5N1	2004	Human
A/goose/Guiyang/337/2006/H5N1	ABJ96698	H5N1	2006	Avian
A/Chicken/Deli_Derdang/BBPVI/2005/H5N1	EU124091	H5N1	2005	Avian

SARS-COV-2 Spike glycoprotein (codon optimised)	YP_009724390	SARS-CoV-2	2019	Human

Table 3. All sequences used to produce pseudotyped influenza viruses. HA sequences are not codon optimised unless stated otherwise.

Crystal structures were obtained from rsc.org and charmm-gui.org. Structures used: 5YKC.pdb – A/chicken/Taiwan/0502/2012, 5E2Y.pdb – A/duck/Egypt/10185SS/2010, 6E7G.pdb – A/Viet Nam/1203/2004, 4CQV. – A/turkey/Turkey/1/2005, 6ONA – A/Hickox/1940, 6N41 – A/Netherlands/002P1/1951, 4GXX – A/Brevig Mission/1/1918, 4M4Y – A/California/04/2009, 2FK0.pdb - A/Viet Nam/1203/2004, 1RUZ.pdb – A/South Carolina/1/1918, 6vsb_1_1_1_S309 computational structure linking 6vsb to S309 [264] – SARS-CoV-2.

8.4.3 Scottish Blood Donor Cohort

Serum samples were obtained from the Scottish National Blood and Transfusion Service (SNBTS) [265]. These were sampled from six consecutive timepoints: 17th March n=514, 21st-23rd March n=500, 4th-6th April n=500, 18th-20th April n=558, 2nd-8th May n=1171, 16th -80th May n=512. Samples were heat-inactivated prior to serological testing by incubation at 56°C for 30 minutes. Donor selection criteria are described in Table 4. Ethical permission has been provided for the "virological analysis" of the samples by the SNBTS Research and Sample Governance Committee (IRAS project number 18005).

Anonymous archive and blood samples. SNBTS blood samples are collected annually, with donor consent, for anonymous epidemiological surveillance studies. Six batches of samples, were collected with standard donor consent. These were collected with the following donor deferral criteria: Reactive in standard HCV, HIV-1, HBV, HEV, HTLV-I or syphilis screening

They have felt unwell in the previous 14 days It is less than 28 (17th March, 2020) or 14 days (21st – 23rd March 2020) since resolution of COVID-19 symptoms It is less than 28 (17th March, 2020) or 14 days (21st – 23rd March 2020) from first day of self-isolation During this period if they are over 70 For samples collected on 17th March travel deferrals to countries with high levels of COVID-19 applied For samples collected on and after the 24th March 2020, travel deferral has been removed due to the increased community spread within the UK and to ensure sufficiency of supply. Focus on the health of donors on day of donation/no evidence for transfusion–transmission of SARS/MERS and that blood viraemia is infrequent and seems to only occur in small proportion of individuals with severe symptoms.

Table 4. Details of sample collection for the Scottish sera sample cohort.

8.4.4 Avian Sera cohort

200 vials of wild swan sera were obtained from the Abbotsbury Swannery. Each sample contains metadata including swan age, date, health characteristics and location, GMRA risk assessment 426. The population has been serologically and epidemiologically categorised [266].

Historical data indicate that the swan population at Abbotsbury Swannery has experienced multiple exposures to HPAI H5, with at least two to three documented instances of infection prior to the collection of the samples (the hatch-dates of the samples range from 2008 to 2017). Specifically, the outbreaks experienced by this cohort were caused by: H5N1 in 2007, H5N8 in 2016 and H5N6 in 2017

[266]. Swans are frequently found in densely packed groups which likely facilitates the transmission of influenza. There is also a high infant mortality rate in this population which is likely due to HPAI circulation [266]. Stable seroconversion towards H5 influenza viruses was found in 75% of birds sampled twice at random between June and November 2017, using HI assays [266].

This strongly suggests that most of these birds have been exposed to H5 influenza viruses. Given the historical exposure to HPAIV H5 and the serological evidence, the swan sera samples can be considered as positive controls for H5 influenza viruses in the pseudotype microneutralisation assay.

8.4.5 Stem-targeted Monoclonal Antibodies (mAbs)

When undertaking site-directed mutagenesis (SDM), I used two stem-targeted anti-HA monoclonal antibodies (mAbs), Cr6261 and C179, to normalise for HA head-domain independent pseudovirus changes. Both mAbs have been shown to recognise H5N1 and H1N1 influenza strains [267]–[269]. Concentrations were optimised during titration with pseudoviruses. The Cr6261 mAb was purchased under the name Diridavumab from Life Technologies Ltd (MA542027). The C179 mAb was purchased from Takara Bioscience (SD3145).

9. Detection of neutralising antibodies to SARS coronavirus 2 to determine population exposure in Scottish blood donors between March and May 2020

9.1. Collaborative work statement

I ran neutralisation assays throughout the serosurvey as part of a team of 20 members to generate the data contained within this chapter. I undertook the data analysis contained within this chapter without assistance during the serosurvey and subsequently to that. A complete analysis of the serosurvey was published in [220]. I have permission from the corresponding author to use this data.

9.2. Background

COVID-19 is the disease caused by the novel coronavirus, SARS coronavirus 2 (SARS-CoV-2) [270]. SARS-CoV-2 was first identified in the Hubei province of China in December 2019 [209]; [210]. Since 2019 SARS-CoV-2 spread around the world causing over 600 million cases and up to 6 million deaths [3].

Symptomatic individuals exhibit fever, cough, and shortness of breath 2-14 days after infection [211]. Those who suffer more severe illness can experience pneumonia and acute respiratory distress syndrome (ARDS) [207]. However, one study proposed that around 30%, of individuals experience no symptoms [208]. This figure was arrived at after nucleic acid screenings were performed on close contacts of COVID-19-confirmed patients, of which 24 asymptomatic individuals were characterised. Antibody responses are mounted during the first few days of infection [271], [272] and can last up to eight months after infection [273].

SARS-CoV-2 is an enveloped virus belonging to the β-coronavirus family which comprises five currently known viruses [211]. Three of these MERS-CoV, SARS-CoV-1 and SARS-CoV-2 are pathogens which recently have 'jumped' species, passing from an animal host to humans [270]. By way of comparison, other beta-coronaviruses such as HCoV-OC43 and HCoV-HKU are endemic viruses which often circulate in human populations causing mild disease [274]. Newly emergent coronaviruses that cause human diseases are typically more virulent than the endemic coronaviruses [275].

During 2020, it had been suggested that SARS-CoV-2 'jumped' species via a seafood market in Wuhan, Hubei, China. Evidence for this was epidemiological; the seafood market and surrounding areas comprised the initial outbreak epicentre, with 2/3 of the first COVID-19 patients having had contact with the market [276].

However, some of these cases did not have contact with the market [209]. Phylogenetic analyses published in [277], [278] suggest importation to the Huanan market from elsewhere, and that more than one strain was already in existence at the time of detection [278]. This has led researchers to hypothesise that the Hubei market could have been an early 'superspreader event', where a large number of transmissions took place [279].

Some suggested that the SARS-CoV-2 virus could have emerged as early as Summer 2019. Nsoesie *et al.,* (2020) cite uncharacteristic car park topology and search engine results to point to a cover-up by authorities [280]. This was widely denounced by the scientific community. The timing of global spread is also a subject of debate. There are studies showing the presence of the virus in France in 2019 [281]. The virus was, for example, thought to have caused the first UK index case on the 23rd January 2020, [282].

Genomic analysis suggests that is highly likely that an ancestral bat coronavirus jumped to an intermediary species subsequent to humans [283]. There is also phylogenetic evidence that an

intermediary host species was the pangolin [284]. Ultimately, there is no scientific consensus on how and when SARS-CoV-2 emerged and spread to humans, reviewed in [279].

The COVID-19 pandemic prompted many governments to enact 'lockdown' policies – consisting of banning social gatherings and closing workplaces to try to reduce transmission of the virus. On March 23rd 2020, the UK government implemented this policy [285]. Our study started before the pandemic had established itself in the UK. The first samples we tested came from the 17th of March 2020, a week before the first lockdown was implemented. At this time, the UK had experienced approximately 2,626 cases, with 99 recorded deaths [286], [287]. Although, COVID-19 testing was not widespread during this time so this number may be an underestimate [287].

Our study concluded in the middle of May 2020. The Effective Reproduction number (RE), defined as the number of hosts to which an infected host will transmit a pathogen, was at that time decreasing. It remained likely that the first wave of the epidemic was on a downward trend [288]. The Office for National Statistics (ONS) estimate that case numbers peaked around the start of April 2020 [289]. Lockdown restrictions were eased gradually, at different times by the devolved parliaments of the UK. For example, in Scotland, lessening restrictions were announced on the 28th of May, with greater relaxations announced on the 18th of June 2020 [285].

We theorised that whether further spread of SARS-CoV-2 would or would not occur was to be influenced by the level of population immunity towards SARS-CoV-2. At the end-point of our study, the population immunity of Scotland towards SARS-CoV-2 had not been ascertained. We, therefore, conducted a time-course serosurvey of Scottish blood donors to detect neutralising antibodies directed to the SARS-CoV-2 spike protein between the 17th of March and 18th of May to determine the extent of population immunity to the novel SARS-CoV-2 virus in Scotland.

9.3. Results

The aims of this study were three-fold: (i) to determine if blood donors can be used as a sentinel population to ascertain the population immunity towards novel viruses, (ii) to describe how the SARS-CoV-2 pandemic progressed in Scotland using neutralising antibodies as a marker for infected individuals, and (iii) to determine if there was population immunity to SARS-CoV-2.

I attempted to quantify a time series of the seroprevalence of Scotland between the 17th of March and the 18th of May by repeated sampling of blood donor samples and seroprevalence modelling.

9.3.1 Sensitivity and specificity analysis

We obtained 3,617 blood sera samples from the Scottish National Blood and Transfusion Service (SNBTS) also known as 'Blood Scotland' [265]. In order to determine which of these samples were positive for anti-SARS-CoV-2 neutralising antibodies using our pseudotype neutralisation assay (Methods 8.1.3), we undertook a sensitivity and specificity analysis. We obtained an additional 100 pre-pandemic blood donor samples from the SNBTS anonymous archive to act as negative controls. To serve as positive controls, we obtained 17 samples from asymptomatic contact-traced individuals who were PCR-confirmed as SARS-CoV-2 infected. All the individuals from whom the positive control sera samples were taken had asymptomatic SARS-CoV-2 infections.

I determined the $\log_{10}(IC_{50})$ values for these 17 positive controls and 100 pre-pandemic negative control sera samples using the approach outlined in Methods 8.1. I was able to fit IC_{50} curves for 30/100 negative controls and 17/17 positive controls. There was a significant difference between the means of the positive and negative controls as determined by t-test, p-value < 2.32e-14 (Figure 13 a). 16/17 positive controls yielded greater $\log_{10}(IC_{50})$ values than negative control sera. This gives our assay a 100% (100/100) specificity and 94% (16/17) sensitivity rate based on the sample cohorts assessed.

To bolster the neutralisation assay results, we ran Enzyme-Linked Immunoabsorbant Assays (ELISAs), described in Methods 8.1.8, on selected sera samples that yielded apparent neutralisation. The neutralisation $\log_{10}(IC_{50})$ and ELISA Optical Density measurements correlated positively: Pearson's correlation 0.864 $p<0.001$ and linear model R^2 0.748 (Figure 13 b).

I selected neutralisation assay thresholds based on the results of the sensitivity and specificity analysis. $Log_{10}(IC_{50})$ value and $\log_{10}(\sigma)$ cut-offs of 1.84 (the greatest $\log_{10}(IC_{50})$ of any negative control) and -0.8 (the highest standard error of any positive control) were decided on after comparison of the level of neutralisation with the primary and negative positive controls (Figure 13 c). These samples were analysed with a SARS-CoV-2 S protein ELISA in tandem to bolster and confirm the results if the sample contained SARS-CoV-2 antibodies.

Of the 3,617 post-February samples we tested, 112 contained anti-SARS-CoV-2 spike-neutralising antibodies using the IC_{50} and standard error-based thresholds described in the sensitivity and specificity analysis (Figure 13 c). IC_{50} curves were fit for 310/3,617 samples. 198 samples to which curves were fit were not defined as positive due to them either not exhibiting $\log_{10}(IC_{50})$ values greater than the negative threshold or standard errors greater than the positive threshold.

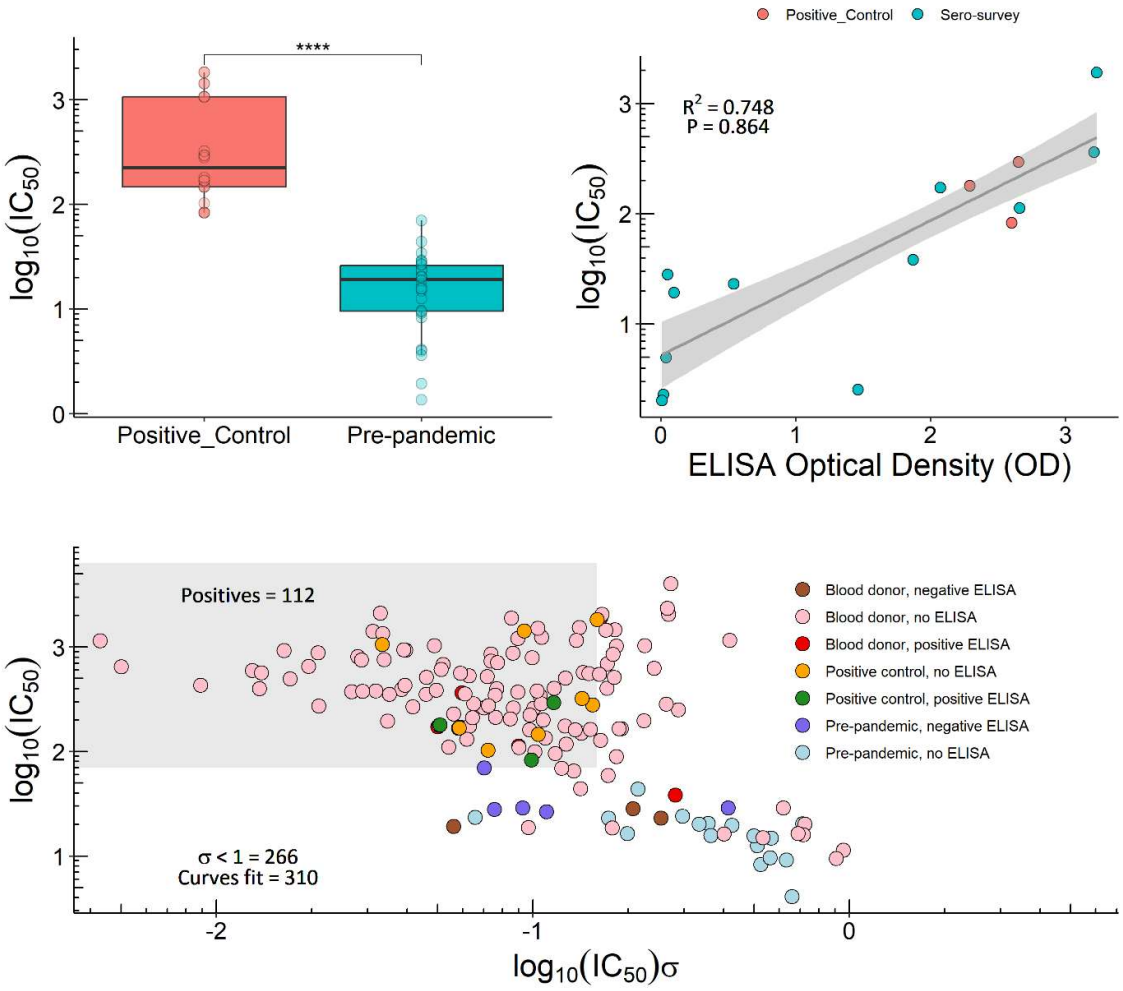


Figure 13. Sensitivity and specificity analysis for our bespoke SARS-CoV-2 Spike pseudotype microneutralisation assay. Figure a is a boxplot comparing the $\log_{10}(IC_{50})$ values of the positive and negative control sera, n=47. The x-axis denotes whether the sample is a positive or negative control and the y-axis denotes neutralisation strength $\log_{10}(IC_{50})$. The median $\log_{10}(IC_{50})$ value is given by the middle lines in each box plot. There is a significant difference between the $\log_{10}(IC_{50})$ values of positive control and pre-pandemic donor samples after the application of a student's t-test: p-value = 2.32e-14. Figure b shows the correlation between the $\log_{10}(IC_{50})$ (y-axis) of our neutralisation assay and the OD of our ELISA (x-axis) for the same sera samples. A linear model (grey line) was fitted with an associated standard error (grey shaded area) to the data. The adjusted R^2 and Pearson's Correlation (P) are labelled. Red points show positive control sera and blue points show serosurvey samples that were defined as positive. Figure c: Selection criteria for classifying a sample as SARS-CoV-2-neutralising, Scotland, March-May 2020 (n = 3,617). Points show the $\log_{10}(\sigma)$ (x-axis) and $\log_{10}(IC_{50})$ measurement (y-axis) for all samples in which a positive $\log_{10}(IC_{50})$ value with a standard error less than 1 was found (n = 266). Colours show each sample's neutralisation assay and ELISA results. All samples situated within the grey rectangle were defined as positive.

9.3.2 Modelling seroprevalence of neutralising anti-SARS-CoV-2 Spike antibodies

We tested 3,617 post-February sera samples with our novel pseudotype neutralisation assay as described in Methods 8.1. These samples were sampled from six consecutive time points which were used in this study: 17th March n=514, 21st-23rd March n=500, 4th-6th April n=500, 18th-20th April n=558, 2nd-8th May n=1171, 16th -80th May n=512. These are expressed as the week of the year: 11, 12, 14, 16, 18, 19, and 20. Scotland is divided into Health Boards by NHS Scotland. Health Boards are responsible for implementing health services for a geographic location. As I was able to obtain the Health Boards of the samples as metadata, I used Health Boards to geolocate clusters of infections.

Longitudinal analysis of antibody titres suggested that sample seropositivity did not decrease during the time analysed within our study period [220]. With this assumption, I was able to model the number of seropositive donors in our sample set using logistic regression as described in Methods 8.1. This allowed me to estimate the true number of seropositive donors within different Scottish regions over time (Figure 14).

Time had a significant positive impact on the numbers of seroprevalence, with a gradient p-value p <2e-16 based on the logistic regression model. This means that time had a statistically positive impact significant on the model's gradient, suggesting that the chance of seropositivity of donors increased over time. The Health Board the samples were sourced from, added as a factor in the logistic regression model, also had a significant impact on the modelled seroprevalence p <2e-16. This means that the seroprevalence modelling suggests that different Health Boards experienced different levels of seropositivity, with different growth rates.

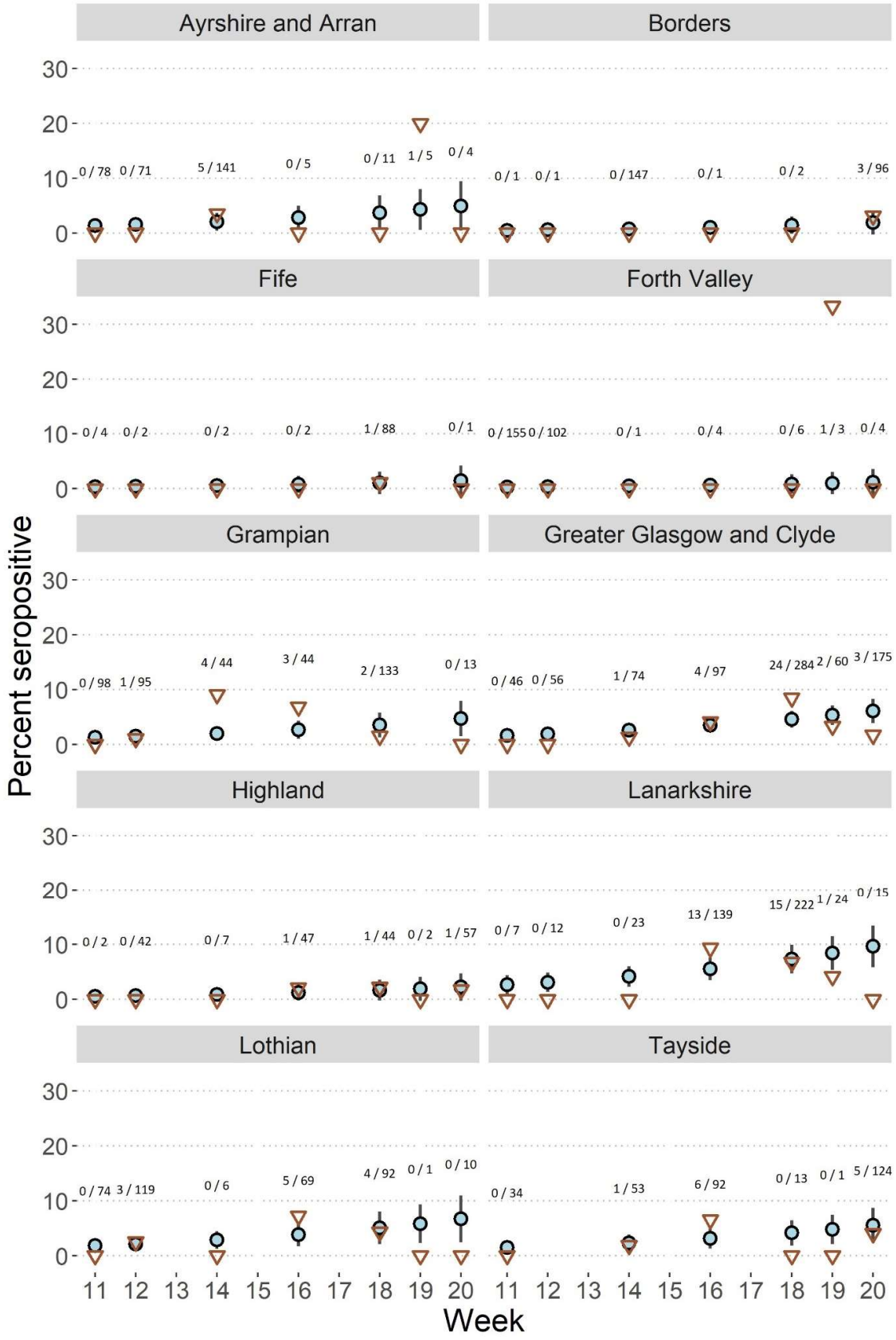

Ayrshire and Arran
Borders
Fife
Forth Valley
Grampian
Greater Glasgow and Clyde
Highland
Lanarkshire
Lothian
Tayside
Percent seropositive
Week
0 / 78
0 / 71
5 / 141
0 / 5
0 / 11
1 / 5
0 / 4
0 / 1
0 / 1
0 / 147
0 / 1
0 / 2
3 / 96
0 / 4
0 / 2
0 / 2
0 / 2
1 / 88
0 / 1
0 / 155
0 / 102
0 / 1
0 / 4
0 / 6
1 / 3
0 / 4
0 / 98
1 / 95
4 / 44
3 / 44
2 / 133
0 / 13
0 / 46
0 / 56
1 / 74
4 / 97
24 / 284
2 / 60
3 / 175
0 / 2
0 / 42
0 / 7
1 / 47
1 / 44
0 / 2
1 / 57
0 / 7
0 / 12
0 / 23
13 / 139
15 / 222
1 / 24
0 / 15
0 / 74
3 / 119
0 / 6
5 / 69
4 / 92
0 / 1
0 / 10
0 / 34
1 / 53
6 / 92
0 / 13
0 / 1
5 / 124
0
10
20
30
11
12
13
14
15
16
17
18
19
20

Figure 14. Seroprevalence of Scottish Health Boards per week following the 17th of March: estimated using the binomial logistic regression model outlined in Methods 8.3.1. Light blue dots denote the estimated seroprevalence per Health Board per week. Red triangles denote observed percentage seroprevalence. Error bars denote 95% confidence intervals (CIs) of estimated seroprevalence. Text labels denote the number of positive samples out of the total number of samples per Health Board per week. Facets denote the Health Board.

Estimates of seroprevalence within different Scottish Health Boards were taken from the final sampling point of the model (Figures 14 to 16). Lanarkshire was predicted to contain the highest seroprevalence percentage of all regions, 9.7% (with a 95% CI of 5.9%-13.5%). Greater Glasgow & Clyde seropositivity was predicted with the highest confidence, 6.10% (95% CI 3.9%-8.3%). Lanarkshire, Lothian, Greater Glasgow & Clyde, Tayside, Ayrshire & Arran, and Grampian were all predicted to have an end-point seropositivity percentage greater than zero. All end-point estimates were characterised by a 95% CI that, when subtracted from the estimate, did not cross zero. Highland, Fife, Borders and Forth Valley were not predicted to have a seroprevalence percentage significantly greater than zero despite all containing some seropositive donors. Forth Valley was characterised by only one seropositive donor.

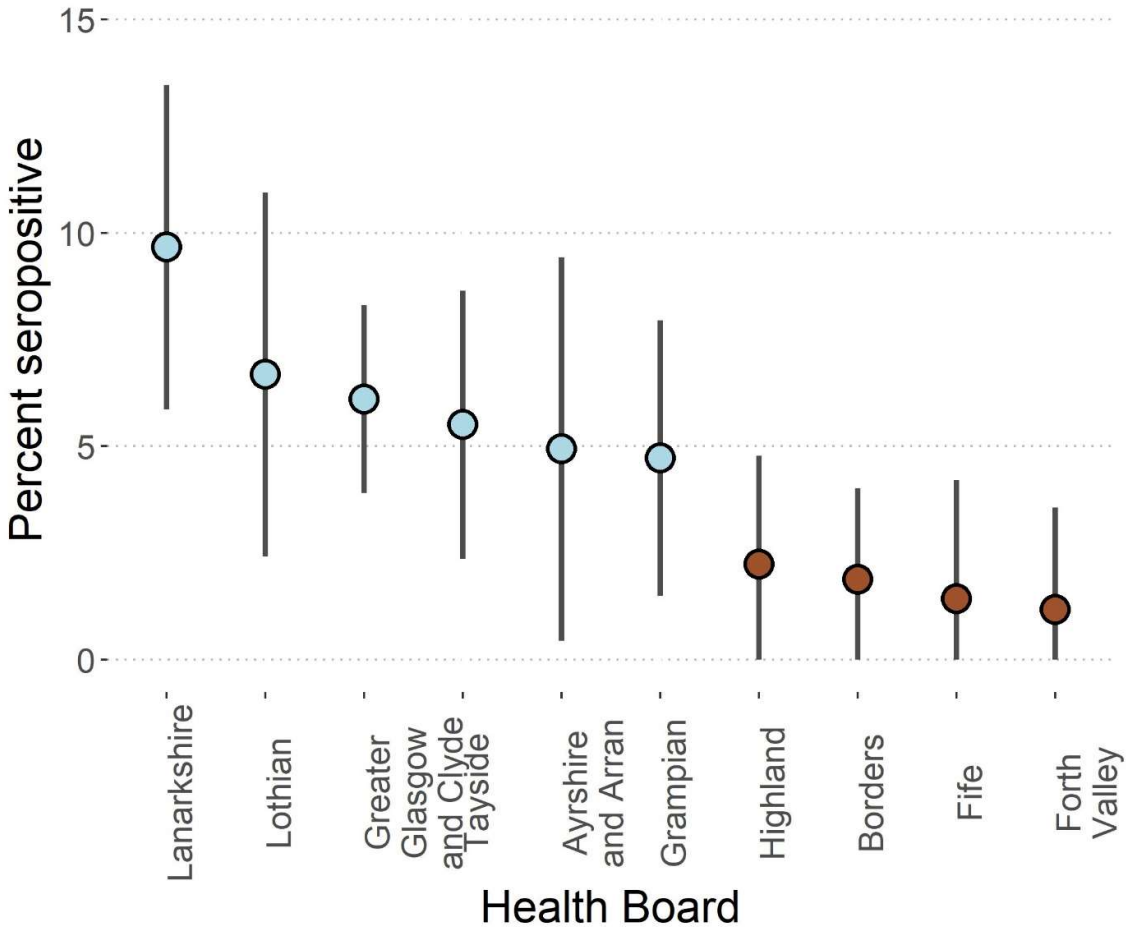


Figure 15. End-point logistic regression seroprevalence estimates of Scottish Health Boards for the final sampling point, week 10. The x-axis denotes the Health Board, the y-axis denotes the percentage chance of a sample being positive. Points show the output of the logistic regression seroprevalence model. Error bars denote the 95% CI of the seroprevalence logistic regression estimates. Light blue points denote samples with estimated CI lower bounds not crossing zero. Sienna-coloured points denote samples with estimated CI lower bounds crossing zero.

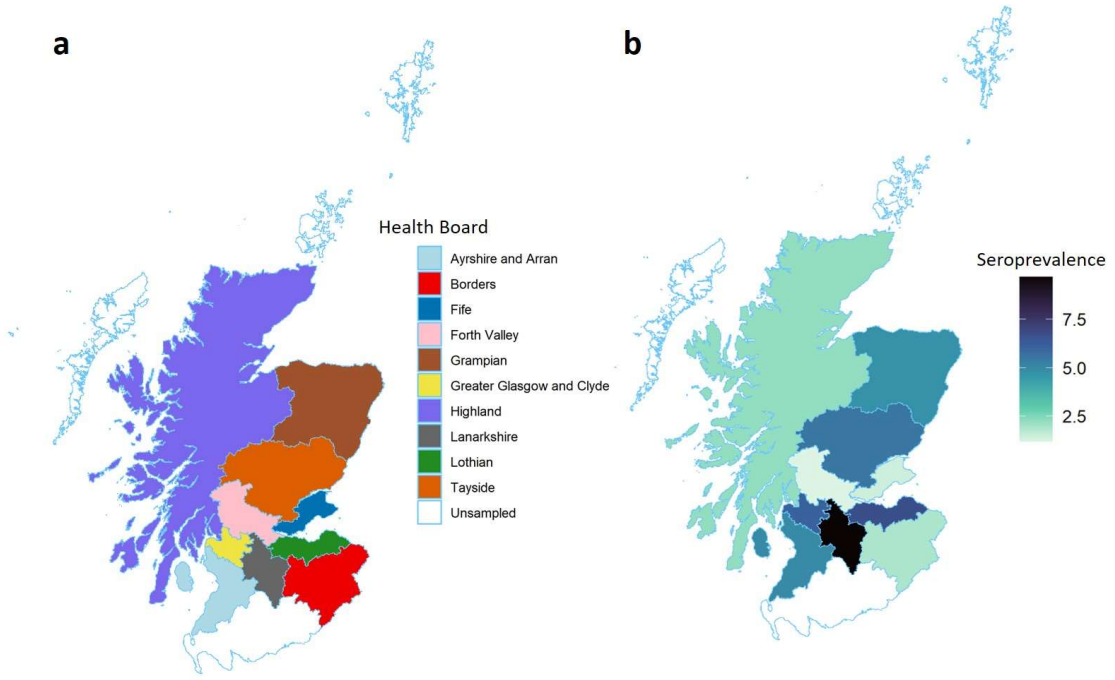


Figure 16. End-point seroprevalence estimates of Scottish Health Boards for the final sampling point, week 10, displayed on a map of Scotland. The Health Boards not filled with colour were unsampled. Figures a and b show the Health Boards of Scotland and their estimated logistic regression end-point seroprevalence respectively.

I was able to obtain the postcodes of where serum donors resided at the time of blood donation for those living in the Greater Glasgow Area. I found that 1/5, 7/191, 1/12, 15/197, and 14/85 donors exhibited neutralising anti-Spike antibodies in the postcodes Falkirk (FK), Glasgow (G), Motherwell (ML), Kilmarnock (KA), and Paisley (PA) respectively (Figure 17).

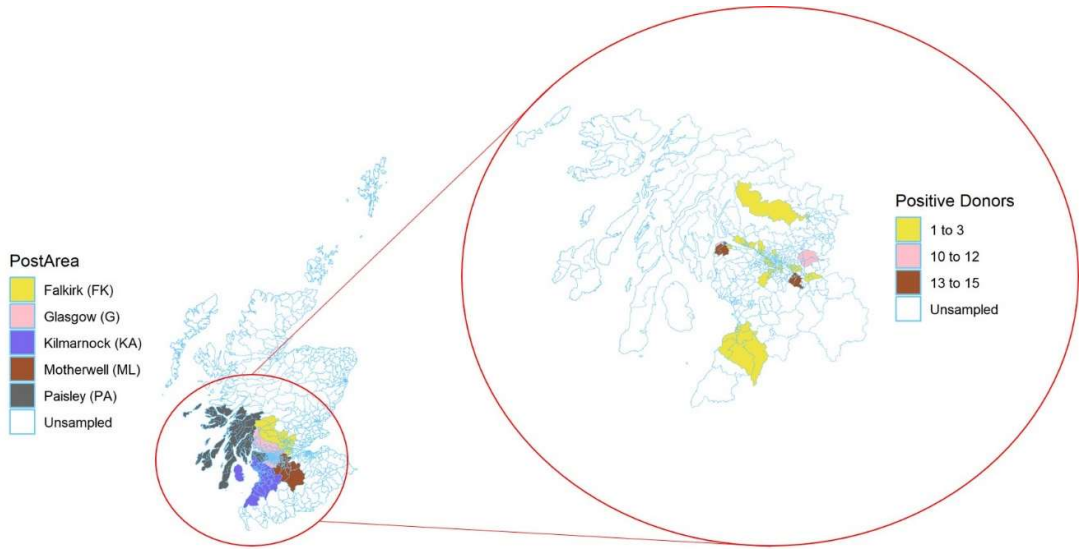


Figure 17. Total number of positive donors found in individual Post Areas of the Greater Glasgow Area between March and May 2020. The fill of the whole Scotland map denotes the number of positive donors found in each postcode of the Greater Glasgow area. The fill of the magnified portion of Greater Glasgow denotes the number of seropositive donors in each Greater Glasgow postcode. Coloured regions denote the individual post areas in which samples were collected, along with the respective numbers of seropositive donors.

9.3.3 Scottish blood donor demographics and representativeness

Figure 18 a shows the sex breakdown of seropositive donors. The mean $\log_{10}(IC_{50})$ for males and females were highly similar, with 2.637 and 2.636 respectively. Donors had an age range from 17 to 80, with no donors of age 72 or 79. The sample set consisted of more female (1869) than male (1631) samples. There was no significant difference between the means when applying a t-test, p-value of 0.9982. No differences were found between the means of neutralisation titre $\log_{10}(IC_{50})$ between male and female Scottish blood donor samples, as analysed by t-test (Figure b). Multiple Benjamini-Hochberg (BH) corrected T-tests found no differences between the means of $\log_{10}(IC_{50})$ values sampled from individual weeks (Figure c). Age did not have a significant impact on the neutralisation titre of positive samples under analysis with linear models (Figure d).

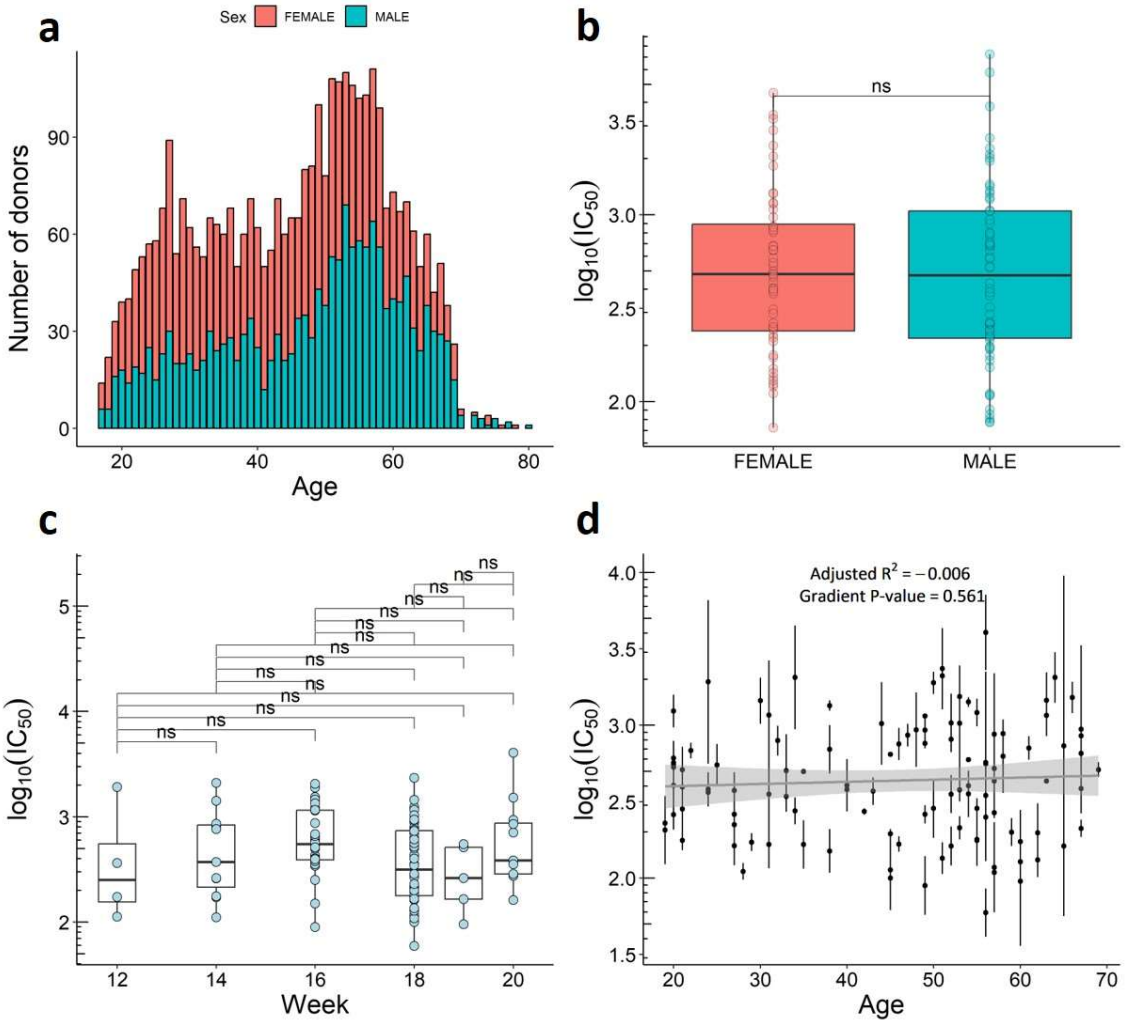


Figure 18. The demographics of seropositive Scottish blood sera samples. Figure a is a histogram consisting of number of donors on the y and donor age on the x-axis. The columns are split up into male and female. Male and female columns are stacked meaning to read female data, take the count from the top of the male bar to the end of the female bar. Figure b compares the neutralisation strength of females and males. The x-axis denotes sex and the y-axis denotes neutralisation strength. The points show individual neutralisation values. The middle lines of the boxes denote the median neutralisation strength. There is no significant difference between male and female $\log_{10}(IC_{50})$ values when tested using a student's t-test (P=0.9982). Figure c comprises boxplots showing the neutralisation strength, denoted by $\log_{10}(IC_{50})$ per sampling week of all positive donors (n = 112). The x-axis denotes the sampling week and y axis denotes the neutralisation strength. Black points show individual neutralisation values. The middle lines of the boxes denote the median neutralisation strength. The significance of multiple BH-corrected T-tests between the neutralisation titre of sampling weeks is denoted by vertical lines. All were non-significant (ns). Figure d explores the relationship between neutralisation strength and donor age of all positive donors (n = 112). The x-axis denotes the age of the donor and the y-axis denotes the neutralisation strength. The vertical lines denote the $\log_{10}(IC_{50})$ standard error for each point. The grey line denotes the output of a linear regression with the shaded area denoting the standard errors of the model's prediction. The adjusted R^2 and p-value of the linear model's gradient are labelled.

9.4. Discussion

9.4.1 Summary

Many studies have shown that anti-SARS-CoV-2 Spike protein antibodies are generated following infection with SARS-CoV-2 [270], [271], [296]–[301], [272], [273], [290]–[295]. Neutralising antibodies, antibodies which prevent viral entry, have also been demonstrated to be generated after SARS-CoV-2 [301]. Our study found anti-SARS-CoV-2 neutralising antibodies in 112 out of 3,653 Scottish blood donor samples. My logistic regression seroprevalence model's results did not find large standing population immunity to SARS-CoV-2 (Figures 14 to 16).

Seroprevalence, determined by positive neutralisation values defined by the thresholds presented in Results 9.3.1, increased in our blood donor cohort over our study period. By the end of our study period, after ten weeks, six health boards: Lothian, Tayside, Greater Glasgow & Clyde, Ayrshire & Arran and Grampian were predicted to have a non-zero seroprevalence using this methodology. Four Health Boards: Highland, Fife, Borders and Forth Valley were not predicted to have a significantly non-zero seroprevalence using this methodology. Lanarkshire was predicted to have the greatest seroprevalence by the end of our study period of 9.7% (95% CI of 5.9%-13.5%). The average seroprevalence, calculated by generating a mean of the end-point seroprevalence estimates across Health Boards, found in our study was 4.4%. Even the highest value from Lanarkshire is far lower than what was predicted to be needed to confer population immunity, which has been estimated to be 80% or above [302]. Therefore, the study suggested that (i) SARS-CoV-2 was likely to infect most of the population due to an absence of pre-existing immunity, and (ii) public health interventions such as vaccinations were necessary.

9.4.2 Antibody kinetics

The kinetics of the serum humoral immune response towards SARS-CoV-2 have been well-characterised [270], [271], [296]–[300], [272], [273], [290]–[295]. Antibodies are produced promptly after infection, with extrafollicular (EF) plasma cells producing antibody within a few days after infection [271], [272]. Early-produced EF-plasma cell antibodies have not undergone much affinity maturation [271], [295]. Seroconversion to the S protein happens within one to three weeks after infection with the virus, these antibodies are primarily IgG and IgM and many are neutralising [294], [298]–[300]. Peak neutralising antibody titres are commonly seen after the first few weeks, ranging from three to five weeks post-symptom onset (PSO) [296], [297], [303], [304]. The repertoire of anti-spike antibodies increases over time in a similar manner to influenza antibodies undergoing affinity maturation in the Germinal Centre (GC) of B-cell follicles [271], [305].

Neutralising antibodies towards SARS-CoV-2 have been shown to decay over time [296], [297], [303], [304], however, the temporal characteristics of this are varied and likely to vary based on the assay, the antigen used in the assay, and the viral load of infection. Anti-Spike IgG antibodies have been reported up to eight months PSO [273]. However, other studies have estimated that titres are only stable for three to four months PSO [292], [296], [297]. Titres of IgM and IgA have been shown to decay faster than IgG [271], [273], [297], [303]. Our study period spanned three months. If antibody titres, especially concerning IgM and IgA, decayed by a substantial amount during our study period, then this could have led to our study underestimating seroprevalence levels. This remains unlikely when considering the previously mentioned antibody kinetics.

9.4.3 Clinical severity and the influence on antibody responses

The antibody titre that patients produce after infection has been shown to be variable [103], [270], [296]–[301], [271]–[273], [290]–[294]. There is a strong correlation between disease severity and titre, with severe cases producing higher titres of IgM and IgG than less severe cases [273], [306], [307].

Older people are more likely to experience more severe clinical manifestations, correlating with increased antibody titres [308]. Using Gaussian statistics, I found no statistically significant effect of age or sex on the neutralisation strength of the donor's sera.

9.4.4 Seroprevalence data and its quality

Pre-existing immunity is known as a correlate of protection for infection with viruses [309], [310]. Pre-existing neutralising immunity may decrease the potential of novel viruses to cause outbreaks in certain populations [50]. Therefore, determining the seroprevalence of a population towards a virus is essential for long-term communicable disease policy generation. During the time of our study, multiple SARS-CoV-2 serosurveys had been conducted, with varying quality [288], [311]–[314]. We conducted bespoke analyses to calibrate our assay thresholds with the equipment and reagents we were using. Many SARS-COV-2 serosurveys have used commercially available assays to test for anti-SARS-COV-2 antibodies and have relied on manufacturer's sensitivity/ specificity data instead of undertaking bespoke analysis.

Serosurveys were conducted by Stringhini *et al.,* (2020) and *PHE.*, (2020) focussing on Geneva, Switzerland and England, UK respectively [288], [314]. Seropositivity in Geneva was found to change due to age, with 8.5% seropositive in 20–49-year-olds and 3.7% in those aged 50+. Public Health England (PHE) found that seroprevalence ranged from 17.5% to 4.2% in London and the North East respectively in May 2020. This serosurvey did not sample every region at all time points. Both used the Euroimmun ELISA to inform the studies. Although sensitivity analyses, [315]; [316] have concluded high reliability of the assay for IgG detection, the FDA (US Food and Drug Administration) cast doubt on its effectiveness, [317].

One of the first surveys undertaken estimated that the SARS-COV-2 seroprevalence of Santa Clara County, California was around 2.8% in March 2019, [311]. This was widely criticised due to the subject recruitment practises. Participant blood donors were reached using targeted Facebook

advertisements. This bodes the question of how representative the sample set was of the wider population.

Bendavid *et al.,* (2020) used lateral flow assays (LFAs) from Premier Biotech sourced from China that were not approved by the Chinese National Medical Products Administration (NMPA) [318]. Another Californian study using Premier Biotech assays surveyed Los Angeles County and found a seroprevalence of 2.5-7%, [313]. However, a more recent study found that Premier's LFAs perform adequately [319].

Doi *et al.,* (2020) used an immunochromatographic assay from KURABO Industries Ltd. to suggest that around 3.3% of the population of Kobe, Japan was seropositive in April 2019 [312]. The study was done using specificity thresholds suggested by the manufacturers and was criticised as inaccurate due to a high false positive rate [320].

In our study, specificity is the chance of identifying an antibody specific to a related coronavirus and sensitivity is the chance of missing antibodies to SARS-CoV-2.

The sensitivity and specificity analysis allowed for sensible $\log_{10}(IC_{50})$ and $\log_{10}(\sigma)$ cut-off thresholds, of 1.84 and -0.8 to be generated. This gave our assay a 100% specificity and 94% sensitivity rate. Moreover, corroboration by an anti-Spike ELISA showed a positive correlation with our neutralisation assay's results (Figure 13 b). Therefore, I contend that the assay produces accurate results.

Due to the: (i) large sample set, (ii) bespoke sensitivity and specificity analysis, and (iii) relatively unbiased sample set, we contend that our study provided one of the most scientifically rigorous COVID-19 seroprevalence datasets that were available at the time of conducting the study.

9.4.5 Sample representativeness

We assayed samples obtained from the SNBTS. Scotland has a low population density (67.2/km^2) but has several large cities including Glasgow, which has a population density of 3,400/km^2 [321]. The

seropositive samples obtained were from 18/56 county areas in Scotland which suggests a wide geographical area was covered by our study. The highest numbers of positives were obtained from primarily urban areas in the Greater Glasgow area (Figures 16 and 17). However, the manner in which the SNBTS obtained the samples may add confounding variation to our study. The methodology the service uses is to obtain samples from one area at one time and collect from elsewhere subsequently, not to randomly select from different areas at the same time point as would be beneficial for representativeness.

Some features of blood donors differ relative to the general population. Those at risk for blood-borne viruses (HIV-1, HCV, HBV) and syphilis are not represented. Children and those who have returned from infection hotspots (i.e. China in March) <4 weeks before sampling are also not included in the study set. Donors were deferred if they had a history of a laboratory-confirmed SARS-CoV-2 diagnosis, or respiratory symptoms consistent with SARS-CoV-2 infection within the previous 28 or 14 days (outlined in Table 4).

9.4.6 Extrapolation to whole populations

This study attempted to model an estimated seroprevalence of the entire population from our sample set. The rationale is that whichever samples are studied, will not be entirely representative of the whole population. Models are constructed so as to better embody the demographic structure of the population at large. Different methods are used in different studies and the accuracy of such are largely unquantified. I used a logistic regression approach to estimate the seroprevalence of our sample set and found that 6/10 Scottish regions studied had a significantly positive seropositivity rate.

9.4.7 Spectrum bias

Studies that include bespoke sensitivity/ specificity analysis may still be underestimating the seroprevalence of SARS-CoV-2. As previously stated, antibody titres among COVID-19 patients are highly variable. Some of this variance is explained by disease severity; higher antibody titres correlate with more severe disease. Positives used in these sensitivity analyses commonly contain high antibody titres after isolation from qRT-PCR-positive patients who have presented with obvious symptoms. This is likely to skew the results of sensitivity analyses, setting the thresholds too high meaning some lower-titre positives are missed, such as those that experienced asymptomatic infections [322]. This is the phenomenon is known as 'spectrum bias' and could potentially influence our results.

9.4.8 Time-course analysis

Surveying the population at different times throughout a pandemic can provide valuable information about pathogens and how policies such as enforced lockdowns have influenced such. Repeated measurements theoretically could give a better idea of representative surveillance than single time point studies. If done using bespoke assay thresholds and representative sample sets then this could provide a gold-standard way of estimating how many individuals have contracted COVID-19 in a population.

Our raw data (without modelling) show the same broad pattern as other time series studies Stringhini *et al.,* (2020) and (*PHE*, 2020) focussing on Geneva, Switzerland and England, UK respectively [288], [314]. Seroprevalence initially increases, then plateaus, and then starts decreasing (Figure 14'sred triangles). However, in order to extrapolate to whole-population seroprevalence, I employed a logistic regression approach which found that seroprevalence did not, in fact, decrease during the sampling period.

9.4.9 Possible implications for public health and policy

Understanding how many people are immune to COVID-19 is vital for the design of public health policy for the long term. Using experimentally influenced models calibrated with data from other coronaviruses, Rosado *et al.,* (2020) concluded that anti-SARS-CoV-2 Spike antibodies may have circulated for over a year before the pandemic [323]. If this was the case, it was possible that, if the seroprevalence was higher than known, this number of susceptible in the population could have been lower than currently thought. Studies estimated the R0 of SARS-CoV-2 to be from 2-3, [324]. This would have given a herd immunity threshold of around 50-67%. If the seroprevalence was higher than thought, population centres may have been closer to herd immunity than thought at that time. A controversial study claimed that in populations comprising people with large variations in susceptibility, the Pc could be as low as 20% [325].

Our study was vitally important because it gave strong evidence that there did not exist substantial population immunity to SARS-CoV-2 conferred by cross-reactivity to seasonal coronaviruses. The standing population immunity was, if at all present, close to 1%; not enough to protect against the spread of SARS-CoV-2. However, the study was limited by the biases surrounding blood donor sampling (age and health for example). The data generated by this serosurvey, as well as informing policy, provided a valuable source for understanding the temporal dynamics of epidemics and the effect of public health measures.

10. A large-scale serosurvey to assess the population immunity towards endemic and potential influenza viruses in representative human and avian populations

10.1. Background

A pandemic occurs when a pathogen emerges that (i) spreads from human to human, and (ii) limited or no protective immunity exists towards it in the general population [189]. Emergence typically occurs via zoonoses: non-human animal pathogens "spill over" [326] by infecting and circulating in the human population. SARS-CoV-2 fulfilled both these conditions. We showed that towards the end of the first wave of the pandemic, neutralising population immunity towards SARS-CoV-2's spike protein was low in Scottish blood donor sera [220]. The highest prevalence observed during the timeframe of our study was 9.7%. Assessing the same donor population for potential pandemic influenza viruses can help determine if protective immunity to this influenza exists and define current immunity to seasonal influenza in the human population.

Pandemic influenza has killed 50 to 100 million people since the start of the 20th century [2]. Of the five influenza pandemics that occurred during the last 106 years, two were caused by the H1N1 subtype. There is also a risk of novel strains of H1N1 causing pandemics in the future.

Highly Pathogenic Avian Influenza (HPAI) of the H5 subtype is regarded by the World Health Organisation (WHO) to be at high risk of generating a high-mortality human pandemic [327]. Many influenza strains currently circulating in non-human animal populations, specifically swine and birds, present a pandemic risk not just to humans but also to agricultural and wild animal populations [328]. Avian influenza pandemics routinely lead to the culling of tens of millions of poultry and cause losses of hundreds of millions of pounds [30], [329].

To this end, I used the same Scottish blood donor samples tested during the SARS-CoV-2 pandemic to test for neutralising population immunity to a panel of pandemic-risk and historical influenza strains. I also tested a panel of pandemic-risk H5N1 viruses against a wild avian population.

10.2. Results

10.2.1 Calibration of the pseudotype microneutralisation assay

I chose the pseudotyped virus ebola to act as a negative control for the pseudotyped virus microneutralisation (pMN) assay. This was in order to define a threshold at which I could determine if a sample was positive for neutralisation detected in our pseudotyped influenza virus neutralisation assays.

We obtained 3,617 blood sera samples from the Scottish National Blood and Transfusion Service (SNBTS) also known as 'Blood Scotland' (Methods 8.4.4) [265]. We also received 200 vials of H5 influenza virus-exposed wild swan sera from the Abbotsbury Swannery (Methods 8.4.4). I was able to use these sample sets to test a panel of pseudotyped viruses with our pMN assay. In total, I ran pseudotyped influenza virus HAs on human blood donor samples for 2,357 runs and on wild swan samples for 376 runs. The estimated IC_{50} values and standard errors for the influenza serosurvey are displayed in Figure 19.

Eliminating background signals from the pseudotyped virus neutralisation assay was key to enable me to examine neutralising population immunity in the respective blood donor and avian cohorts analysed in this study. Ebola is a Filovirus native to the African continent. There has been no documented human or avian Filovirus outbreak in the United Kingdom [330]. The ebola surface glycoprotein (GP) pseudotyped virus is also a commonly used and well-studied pseudotyped virus. The pseudotype is used in biotechnological research producing robust neutralisation outputs [331]. Therefore, the ebola GP was selected to be expressed on a pseudotyped virus, to be used as an a priori negative control pseudotyped virus.

I generated pseudotyped viruses displaying the ebola GP and tested them on the Scottish blood donor and wild avian sample sets. I generated an age-representative random sample of both sample sets. Age representativeness was essential to eliminate confounding variation introduced by age-neutralisation trends.

To define the lower limit of the assay, I ran 96 human samples on the ebola pseudotyped virus microneutralisation assay. $Log_{10}(IC_{50})$ values with a non-zero lower bound were generated for 61 of the samples tested. The 95th percentile of the $log_{10}(IC_{50})$ values generated was defined as the threshold for the negative control which was 2.27. The 95^{th} quartile was chosen to minimise the chance of false negatives introduced by misidentified strong ebola GP pseudotype responders. Any influenza $log_{10}(IC_{50})$ values that fall below this threshold are defined as being assay background and not a positive neutralising immune response. I also ran 40 swan sera samples through the microneutralisation assay. Curves were fit for 2 samples. The highest $log_{10}(IC_{50})$ of these was 1.18. This was defined as the negative threshold for the avian samples. In Figure 19, the grey boxes show the donors defined as positive by examination of the ebola GP pseudotypes.

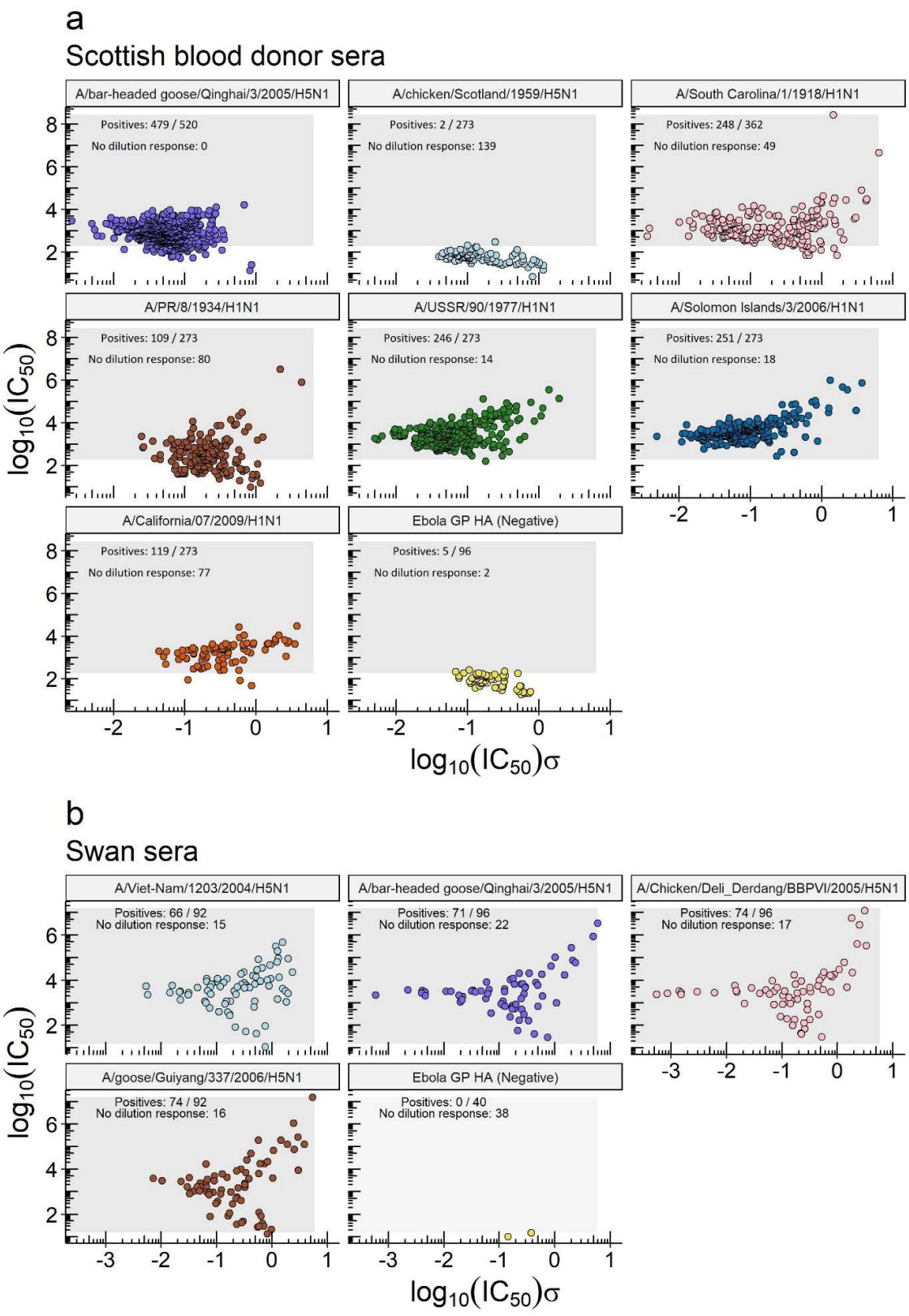

a
Scottish blood donor sera
A/bar-headed goose/Qinghai/3/2005/H5N1
Positives: 479 / 520
No dilution response: 0
A/chicken/Scotland/1959/H5N1
Positives: 2 / 273
No dilution response: 139
A/South Carolina/1/1918/H1N1
Positives: 248 / 362
No dilution response: 49
A/PR/8/1934/H1N1
Positives: 109 / 273
No dilution response: 80
A/USSR/90/1977/H1N1
Positives: 246 / 273
No dilution response: 14
A/Solomon Islands/3/2006/H1N1
Positives: 251 / 273
No dilution response: 18
A/California/07/2009/H1N1
Positives: 119 / 273
No dilution response: 77
Ebola GP HA (Negative)
Positives: 5 / 96
No dilution response: 2
log10(IC50)
log10(IC50)σ
b
Swan sera
A/Viet-Nam/1203/2004/H5N1
Positives: 66 / 92
No dilution response: 15
A/bar-headed goose/Qinghai/3/2005/H5N1
Positives: 71 / 96
No dilution response: 22
A/Chicken/Deli_Derdang/BBPVI/2005/H5N1
Positives: 74 / 96
No dilution response: 17
A/goose/Guiyang/337/2006/H5N1
Positives: 74 / 92
No dilution response: 16
Ebola GP HA (Negative)
Positives: 0 / 40
No dilution response: 38
log10(IC50)
log10(IC50)σ

Figure 19. Selection criteria for defining positive neutralisation responses to pseudotyped influenza viruses. To be counted as having neutralising activity, samples are required to elicit a $log_{10}(IC_{50})$ neutralisation higher than the 95th percentile of Ebola GP pseudotyped virus responses in humans; and higher than the highest Ebola GP pseudotyped virus response in birds. The grey boxes denote these thresholds. Any $log_{10}(IC_{50})$ falling below the box is defined as negative. The y-axis denotes $log_{10}(IC_{50})$ and the x-axis denotes the log_{10} standard error of each sera sample. Text labels denote the number of positives generated by the total number of donor samples run. If a $log_{10}(IC_{50})$ value was not fitted then the donor is defined as not generating an immune response.

10.2.2 Sera from Scottish blood donors collected in 2020 neutralise some potential pandemic H5 pseudotyped viruses

A/bar-headed goose/Qinghai/3/2005/H5N1 was first identified in Qinghai, China in 2005 [332]. The virus caused a large avian pandemic resulting in multiple lineages [333]. Strains derived from A/bar-headed goose/Qinghai/3/2005/H5N1 are referred to as "Qinghai-like" [334]. Qinghai-like strains have caused many human spillover infections ranging from Nigeria to Mongolia [333]. The strain was shown to spread between mammals under laboratory conditions [335]. This led multiple research groups, including scientists from the World Organisation for Animal Health (WOAH), to express concerns about the strain's pandemic potential [231].

Serum from a likely unexposed Scottish blood donor cohort of 520 individuals was tested for the ability to neutralise a pseudotyped virus displaying the HA from A/bar-headed goose/Qinghai/3/2005. 479 of 520 (92%), donors elicited a positive human immune response (Figure 20 a).

The H5N1 strain A/chicken/Scotland/1959/H5N1 caused the first recorded outbreak in the world [336]. The outbreak was confined to a Scottish chicken farm. This strain shares an 89.96% amino acid identity with A/bar-headed goose/Qinghai/3/2005. Out of the 273 donors tested, two produced a positive neutralisation response based on the aforementioned criteria (Figure 20 a). This gives a mean response of 0.73% positive.

10.2.3 Sera from Scottish blood donors collected in 2020 neutralises historical H1 pseudotyped viruses

The oldest blood donor in the Scottish human blood donor cohort has a birth year of 1940. No donor in the sample set has experienced infection with A/South Carolina/1/1918/H1N1 or A/PR/8/1934/H1N1. A/South Carolina/1/1918/H1N1 is the influenza virus that caused the "Spanish Influenza" pandemic of 1918 [337]. The Spanish influenza pandemic was the deadliest of modern times, with a death toll of potentially 50 million people [2]. Of the Scottish donors, 246/362 (68%), exhibited a positive neutralisation response (Figure 20 a). A/PR/8/1934/H1N1 was extracted from a patient in Puerto Rico in 1934 [338]. This is one of the most well-studied influenza strains currently in use. 109/273 (40%), exhibited a positive neutralisation response (Figure 20 a).

10.2.4 Individuals neutralise viruses with which they have potentially previously been infected

In 1977, H1N1 was reintroduced into human circulation after 20 years of extinction. This occurred with the 1977 "Russian flu" epidemic. The epidemic started in Northern Asia and lasted until 1979. Potentially 700,000 people died during this epidemic [2]. A/USSR/90/1977/H1N1 was isolated from the USSR in 1977 and was the causative agent of this epidemic [339]. 93% of Scottish donors (255/273), exhibited a neutralising immune response (Figure 20 a). This is the highest percentage of positives of any pseudotyped influenza virus tested on the Scottish blood donor cohort.

A/Solomon Islands/3/2006/H1N1 (synthesised as a GeneArt String) was isolated from the Solomon Islands in 2006 and was the recommended H1N1 WHO vaccine strain for the 2007 to 2008 vaccine [131]. A/Solomon Islands/3/2006/H1N1 is a seasonal epidemic strain. 251/273 (92%) Scottish donors exhibited a positive immune response. The Bayesian 95% confidence interval (lower bound of 91%) crosses with that of A/USSR/90/1977/H1N1 (Figure 20 a).

The causative agent of the 2009 "Swine flu" influenza pandemic is thought to have emerged due to a "triple reassortment" of avian, swine and human H1N1 influenza viruses [340]. Estimates of mortality

range from ~150,000 to 500,000 people [341], [342]. However, due to many potential asymptomatic cases, the mortality rate of some countries could have been 0.3/100,000 people [341]. I found that 119/277 (43%) donors exhibited a positive neutralisation response to pseudotypes made from A/California/07/2009/H1N1 (Figure 20 a). This was the WHO-recommended H1N1 strain between 2009 and 2010 for the influenza vaccine [343].

10.2.5 The magnitude of immune responses in positive donors is not uniform across influenza pseudotypes

Figure 20 b shows all $\log_{10}(IC_{50})$ values for Scottish blood donor neutralising positive samples. Seropositivity was defined by the negative threshold shown in Figure 19. $\text{Log}_{10}(IC_{50})$ values are not uniform, with means varying from 3 (A/bar-headed goose/Qinghai/3/2005/H5N1) to 3.7 (A/Solomon Islands/3/2006/H1N1).

I applied multiple t-tests to determine if there are significant differences between the neutralisation magnitude of pseudotyped influenza viruses. I used the Benjamini-Hochberg multiple-comparison correction to correct p-values. I excluded A/chicken/Scotland/1959/H5N1 due to the low number of positive donors, two. The results of this analysis are displayed in Table 5. Significant differences exist between all virus pairs apart from two. There was no significant difference in means between A/PR/8/1934/H1N1 and A/bar-headed goose/Qinghai/3/2005/H5N1 as well as between A/South Carolina/1/1918/H1N1 and A/California/07/2009/H1N1 (Table 5).

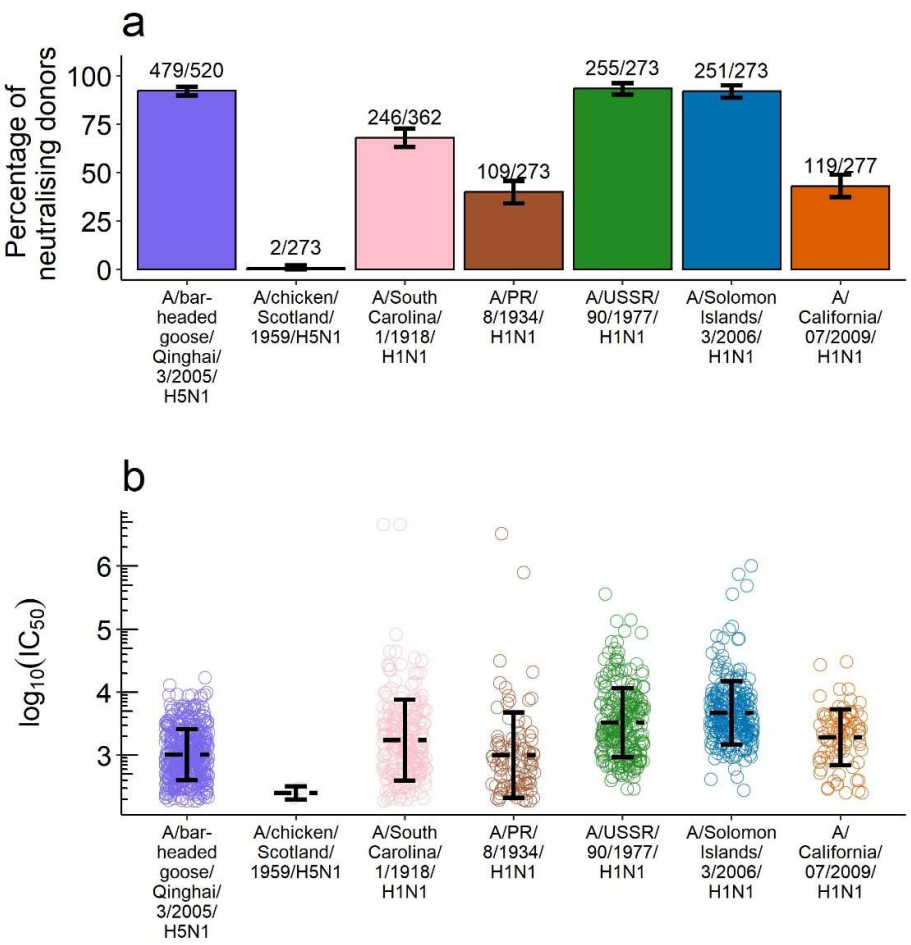


Figure 20. Percentage of neutralising donors and neutralisation strength for each pseudotyped virus run on the Scottish blood donor sample set. Figure a shows the percentage neutralisation. Error bars denote the Bayesian 95% confidence interval of seropositivity chance. The number of positive donors per cohort is labelled above columns. Figure b shows the magnitude of neutralisation. Error bars denote the 95% confidence interval of neutralisation strength; the horizontal dashed lines denote the mean. The y-axis denotes $\log_{10}(IC_{50})$.

.y.	Virus 1	Virus 2	p	p.adj	p.adj.signif
log10_IC50	A/bar-headed goose/Qinghai/3/2005/H5N1	A/California/07/2009/H1N1	5.5e-07	1.2e-06	****
log10_IC50	A/bar-headed goose/Qinghai/3/2005/H5N1	A/PR/8/1934/H1N1	0.93	0.93	ns
log10_IC50	A/bar-headed goose/Qinghai/3/2005/H5N1	A/Solomon Islands/3/2006/H1N1	3.4e-54	5.1e-53	****
log10_IC50	A/bar-headed goose/Qinghai/3/2005/H5N1	A/South Carolina/1/1918/H1N1	2.3e-06	3.8e-06	****
log10_IC50	A/bar-headed goose/Qinghai/3/2005/H5N1	A/USSR/90/1977/H1N1	3.7e-32	2.8e-31	****
log10_IC50	A/California/07/2009/H1N1	A/PR/8/1934/H1N1	0.00091	0.0012	**
log10_IC50	A/California/07/2009/H1N1	A/Solomon Islands/3/2006/H1N1	3.5e-10	8.8e-10	****
log10_IC50	A/California/07/2009/H1N1	A/South Carolina/1/1918/H1N1	0.47	0.5	ns
log10_IC50	A/California/07/2009/H1N1	A/USSR/90/1977/H1N1	0.00015	0.00022	***
log10_IC50	A/PR/8/1934/H1N1	A/Solomon Islands/3/2006/H1N1	2e-15	1e-14	****
log10_IC50	A/PR/8/1934/H1N1	A/South Carolina/1/1918/H1N1	0.004	0.0046	**
log10_IC50	A/PR/8/1934/H1N1	A/USSR/90/1977/H1N1	3.4e-10	8.8e-10	****
log10_IC50	A/Solomon Islands/3/2006/H1N1	A/South Carolina/1/1918/H1N1	9e-15	3.4e-14	****
log10_IC50	A/Solomon Islands/3/2006/H1N1	A/USSR/90/1977/H1N1	0.00093	0.0012	**
log10_IC50	A/South Carolina/1/1918/H1N1	A/USSR/90/1977/H1N1	8.4e-07	1.6e-06	****

Table 5. A comparison between the neutralisation magnitude of pseudotypes run on the Scottish sample set. P.ajd denotes the BH-corrected p-value arising from multiple t-tests. P.adj.signif denotes the significance level of the difference between means at the 95% significance level.

10.2.6 A non-linear relationship between pseudotyped influenza virus neutralisation and age

The “Spanish flu,” “Russian flu” and “Swine flu” outbreaks exhibited unusual age-based severity, mortality, and incidence patterns. This was discussed in Introduction 7.15. “Spanish influenza” was characterised by a “W” shaped mortality curve: most mortality was distributed in the very young and early middle ages; less was seen in teenagers and older adults [204]. A similar pattern was seen during the “Swine flu” pandemic [205]. Moreover, during the “Russian flu,” most mortality was seen in those younger than 26 [206]. Consequently, non-linear patterns of seropositivity chance and neutralisation magnitude may be present in this influenza serosurvey.

I conducted analyses to determine whether age impacts the probability of donors in our Scottish blood donor sample set being seropositive. I modelled seropositivity (percentage of donors eliciting a positive neutralisation response) as a function of age. I did this using multiple methods and selected the most appropriate based on the outcomes.

I first constructed logistic regressions to test for linear, or near-linear, seropositivity-age relationships for each virus (Supplementary Figure 1). Age was significant in models constructed for A/bar-headed goose/Qinghai/3/2005/H5N1 and A/USSR/90/1977/H1N1. Both showed a positive relationship between seropositivity and age, with McFadden pseudo-R^2 values of 0.13 and 0.12, respectively. McFadden pseudo-R^2 can be loosely thought of as describing how much variance is explained by the model.

H1N1 influenza strains of differing antigenicities typically emerge every two to five years [344]. Consequently, I examined a 3-year rolling average of seropositivity. Figure 21 a displays the relationship between seropositivity and age using a 3-year rolling average. Next, I constructed Generalised Additive Models (GAMs) for each virus (Figure 21 b). This approach allowed me to examine non-linear seropositivity-age relationships. Age was a significant component in GAMs constructed for A/Solomon Islands/3/2006/H1N1, A/PR/8/1934/H1N1 and A/bar-headed goose/Qinghai/3/2005/H5N1. Adjusted R^2 values were moderate for A/Solomon Islands/3/2006/H1N1 and A/PR/8/1934/H1N1: 0.4 and 0.28 respectively. The GAM suggests that blood donors responding to A/Solomon_Islands/3/2006/H1N1 have a 100% probability of seropositivity when young and old. This decreases to an 80% probability around the age of 40 (Figure 21 a). From ages 20 to 60, the maximum for A/PR/8/1934/H1N1 is situated around the age of 40. After the age of 40, responses to A/PR/8/1934/H1N1 show another minimum around the age of 60; then an increase until 70 (Figure 21 a).

The seropositivity-age relationships displayed by A/Solomon Islands/3/2006/H1N1, A/PR/8/1934/H1N1, goose/Qinghai/3/2005/H5N1 and A/USSR/90/1977/H1N1 are similar as those modelled by GAM, and by logistic regression for the latter two viruses. Rolling averages for A/South Carolina/1/1918/H1N1 and A/California/07/2009/H1N1 display many peaks and troughs. These relationships are seen in Figure 21.

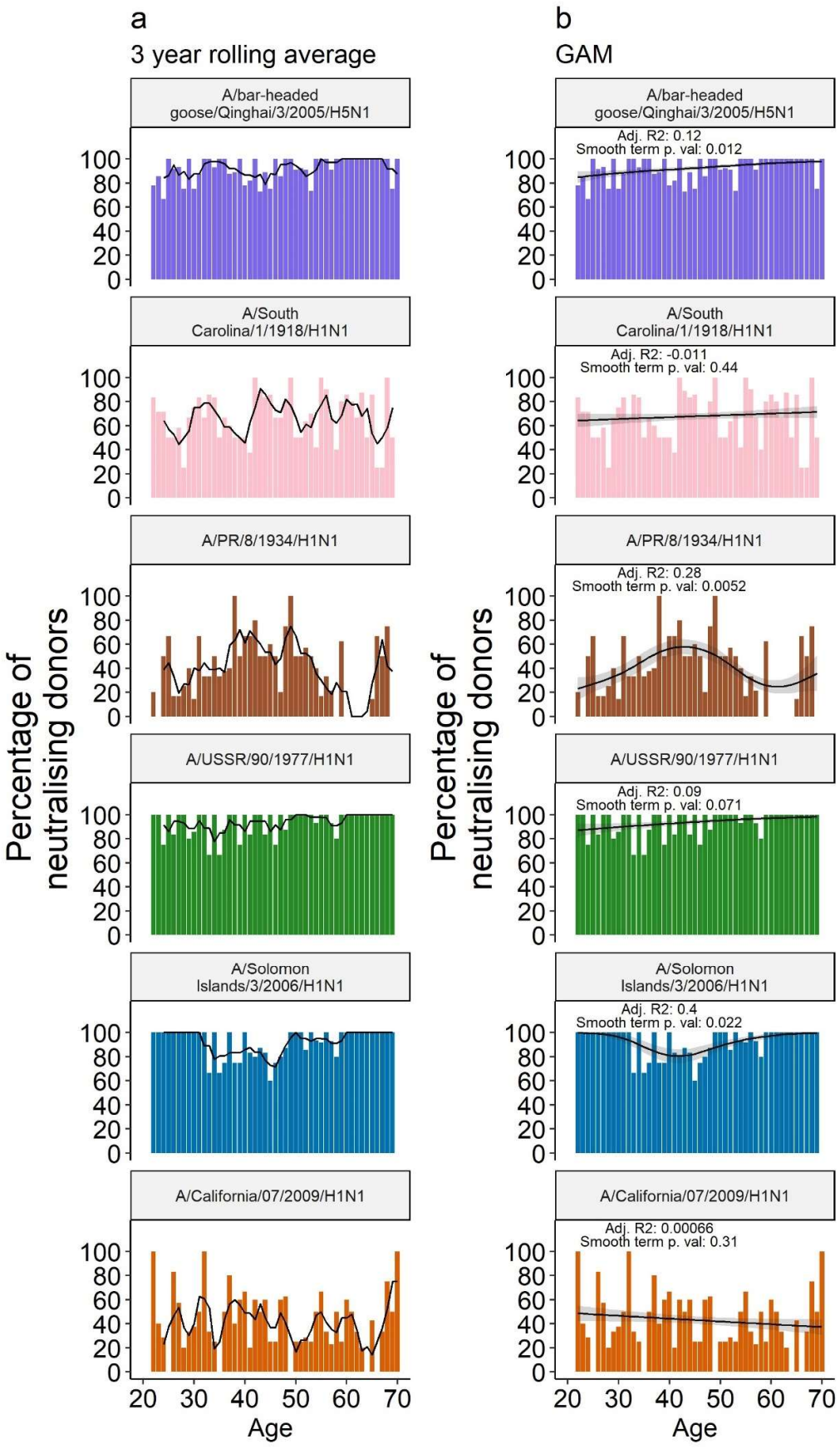

a
3 year rolling average
b
GAM
A/bar-headed goose/Qinghai/3/2005/H5N1
A/South Carolina/1/1918/H1N1
A/PR/8/1934/H1N1
A/USSR/90/1977/H1N1
A/Solomon Islands/3/2006/H1N1
A/California/07/2009/H1N1
Adj. R2: 0.12
Smooth term p. val: 0.012
Adj. R2: -0.011
Smooth term p. val: 0.44
Adj. R2: 0.28
Smooth term p. val: 0.0052
Adj. R2: 0.09
Smooth term p. val: 0.071
Adj. R2: 0.4
Smooth term p. val: 0.022
Adj. R2: 0.00066
Smooth term p. val: 0.31
Percentage of neutralising donors
Age

Figure 21. The relationship between age and the percentage of donors seropositive for pseudotypes run on the Scottish sample set. Bars denote the percentage of seropositive donors per age. Colour denotes individual pseudotypes with a 3-year rolling average and GAMs. Black lines in Figure b denote a 3-year rolling average of seropositivity. The black lines on Figure b show the output of logit-linked GAM models. Shaded areas around the line denote the standard error of the model. Adj. R2 denotes the adjusted R^2 of the model. The smooth term p.val denotes the significance of age in the model.

In addition to examining seropositivity, I also attempted to determine whether naturalisation strength changes as a function of age. Every IC_{50} value has an associated standard error (Figure 19). Standard errors can be used to weight parametric models, thus biasing any patterns towards higher confidence values. When analysing neutralisation strength, I constructed unweighted and weighted models. Through model validation (Methods 8.3.5), the best models were selected to present. All models constructed are shown in Supplementary Figures 2 to 3.

I first constructed weighted and unweighted linear models (Supplementary Figure 2) to find linear relationships between neutralisation strength and age. Age was significant for A/bar-headed goose/Qinghai/3/2005/H5N1 and A/USSR/90/1977/H1N1 (Supplementary Figure 2). For both viruses, the relationship between age and neutralisation strength is positive: the older the donor, the higher the neutralisation strength.

I then examined non-linear relationships between age and neutralisation strength. I constructed both weighted and unweighted GAMs (Supplementary Figure 3). The adjusted R^2 is greater for non-linear models than linear models for both viruses (Figure 22 b and Supplementary Figure 3). Age is significant for A/bar-headed goose/Qinghai/3/2005/H5N1 and A/USSR/90/1977/H1N1 (Figure 22 b). Both show general positive relationships: the older the donor, the greater the neutralisation strength. However, this increase is non-linear. Notably, the increase starts around age 38 for A/USSR/90/1977/H1N1 and decreases after 50 years of age. In addition, there is a trough around age 43 for A/bar-headed goose/Qinghai/3/2005/H5N1.

In addition, I repeated the rolling average analysis with neutralisation strength (Figure 22 a). For both A/bar-headed goose/Qinghai/3/2005/H5N1 and A/USSR/90/1977/H1N1, the relationship follows the trend shown by non-linear modelling (Figure 22). Pronounced peaks and troughs can be seen for all other viruses.

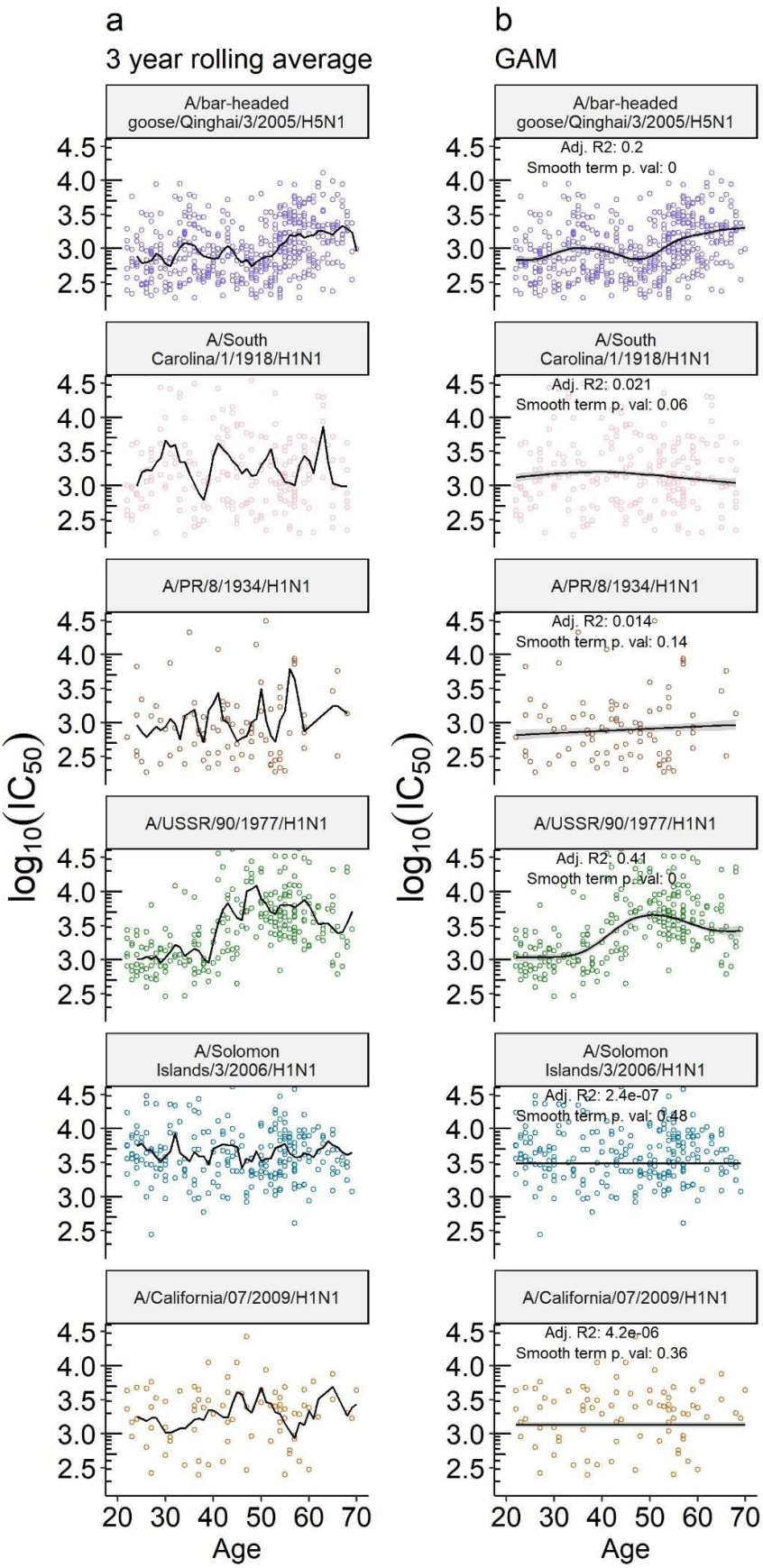

a
3 year rolling average
b
GAM
A/bar-headed goose/Qinghai/3/2005/H5N1
A/South Carolina/1/1918/H1N1
A/PR/8/1934/H1N1
A/USSR/90/1977/H1N1
A/Solomon Islands/3/2006/H1N1
A/California/07/2009/H1N1
Adj. R2: 0.2
Smooth term p. val: 0
Adj. R2: 0.021
Smooth term p. val: 0.06
Adj. R2: 0.014
Smooth term p. val: 0.14
Adj. R2: 0.41
Smooth term p. val: 0
Adj. R2: 2.4e-07
Smooth term p. val: 0.48
Adj. R2: 4.2e-06
Smooth term p. val: 0.36
log10(IC50)
Age

Figure 22. The relationship between age and neutralisation strength for pseudotypes run on the Scottish sample set. Points denote neutralisation strength per age. Colours denote individual pseudotypes. Black lines in Figure a denote a 3-year rolling average of neutralisation strength. Black lines on Figure b show the output of GAM models weighted by $\frac{1}{\delta}$. Shaded areas around the line denotes the standard error of the model. Adj. R2 denotes the adjusted R^2 of the model. The smooth term p.val denotes the significance of age in the model.

10.2.7 Sera from wild swans collected between 2008 and 2017 neutralises potential pandemic H5 pseudotyped avian influenza viruses

Avian-to-human H5N1 spillover was first documented in Hong Kong in 1997, killing six of 18 individuals identified as infected [34]. This was defined as clade 0. Clade 1 originated in Guangdong, China, in 2002. Derived strains spread throughout South East Asia [345]. In Vietnam, these strains spilled over into humans [346], [33], including A/Viet-Nam/1203/2004/H5N1. This virus was the causative agent behind an outbreak of 10 people in January 2004; eight of these died [347]. No patient had direct contact with poultry before the outbreak. Strains in clade 1 routinely infected avian hosts. 71/96 (77%) of swan samples neutralised A/Viet-Nam/1203/2004/H5N1, with a mean $\log_{10}(IC_{50})$ of 3.66 (Figure 23).

Pseudotyped viruses displaying the HA from A/bar-headed goose/Qinghai/3/2005/H5N1 have been run in microneutralisation assays against the Scottish blood donor human sample set. Out of 520, 479 samples were found to contain neutralising antibodies. During the large avian pandemic it caused, infection with the virus produced 80% mortality in water bird species, such as swans, and 100% mortality in chickens [332]. A/bar-headed goose/Qinghai/3/2005/H5N1 is in phylogenetic clade 2.2 [345]. 73/96 (74%) of swan samples neutralised A/bar-headed goose/Qinghai/3/2005/H5N1, with a mean $\log_{10}(IC_{50})$ of 3.46 (Figure 23).

H5N1 clade 2.1.2 spread throughout Western Indonesia. It primarily affected birds but also infected a cluster of humans [345]. A/Chicken/Deli_Derdang/BBPVI/2005/H5N1 was isolated from a chicken in 2005 and is a member of this clade [31]. 68/92 (74%) of swan samples neutralised A/Chicken/Deli_Derdang/BBPVI/2005/H5N1, with a mean $\log_{10}(IC_{50})$ of 3.54 (Figure 23).

A/goose/Guiyang/337/2006/H5N1 was isolated in 2006 from a Southern Chinese market [334]. It belongs to H5N1 clade 4 [32], a clade that exclusively infects birds [345]. 74/92 (72%) of swan samples neutralised A/goose/Guiyang/337/2006/H5N1, with a mean $\log_{10}(IC_{50})$ of 3.35 (Figure 23).

10.2.8 Neutralising immune responses towards avian influenza viruses in a wild swan cohort did not differ between pseudotyped influenza viruses

There were no significant differences observed between the neutralisation response of any pseudotype tested in the microneutralisation assay using the avian sample set. Bayesian 95% confidence intervals for seropositivity chance crossed each other for all pseudotyped viruses (Figure 23 a). Moreover, I applied multiple Bonferroni-corrected t-tests to compare the neutralisation strength of each pseudotyped virus. There were no significant differences between means of any pseudotyped influenza virus (Figure 23 and Table 6).

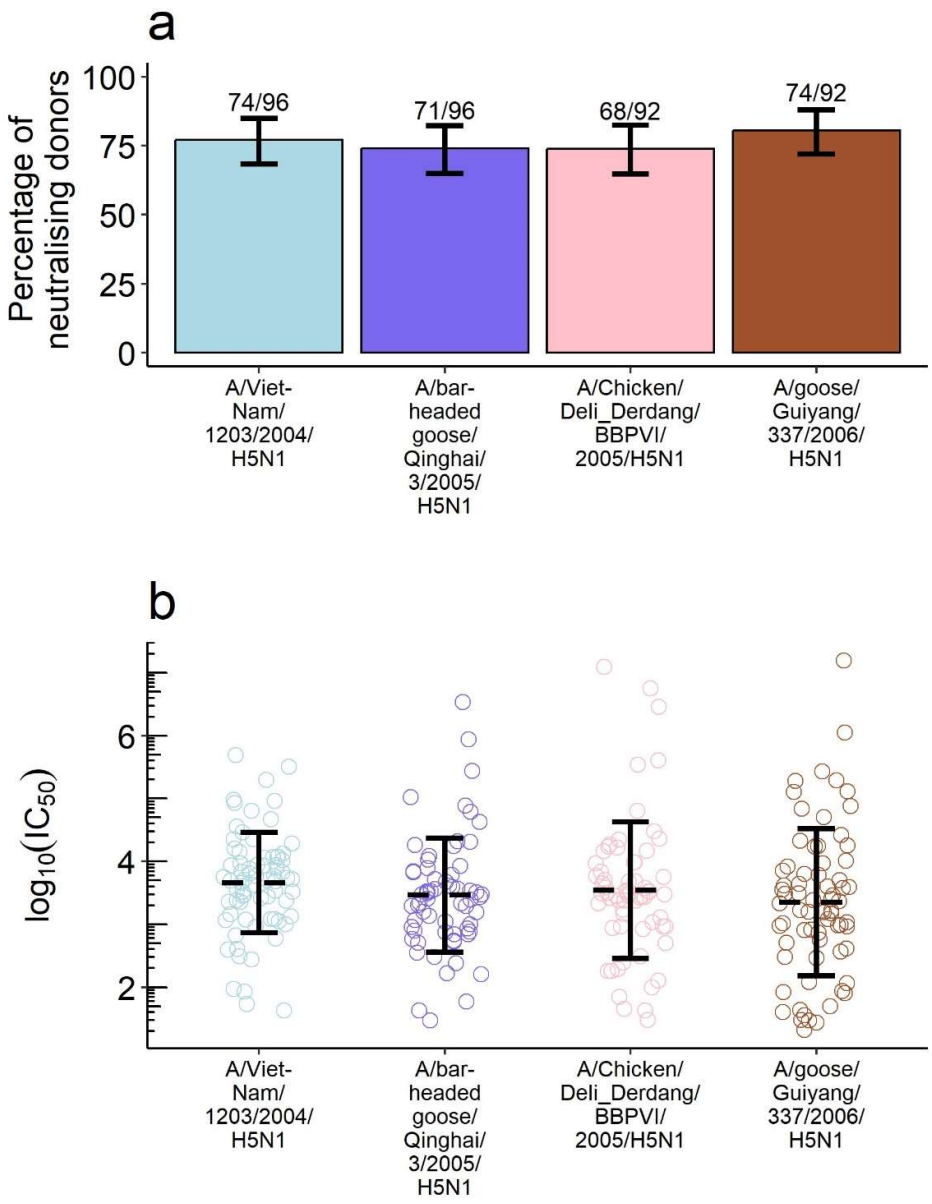


Figure 23. Percentage of neutralising donors and neutralisation strength for each pseudotyped influenza virus run on the wild avian sample set. Figure a shows the percentage neutralisation. Error bars denote the Bayesian 95% confidence interval of seropositivity. The number of positive donors out of the number of samples run is labelled above columns. Figure b. shows the magnitude of neutralisation. Error bars denote the 95% confidence interval of neutralisation strength; the horizontal dashed lines denote the mean. The y-axis denotes $\log_{10}(IC_{50})$.

.y.	Virus 1	Virus 2	p	p.adj	p.adj.signif
log10_IC50	A/bar-headed goose/Qinghai/3/2005/H5N1	A/Chicken/Deli_Derdang/BBPVI/2005/H5N1	0.66	0.66	ns
log10_IC50	A/bar-headed goose/Qinghai/3/2005/H5N1	A/goose/Guiyang/337/2006/H5N1	0.52	0.63	ns
log10_IC50	A/bar-headed goose/Qinghai/3/2005/H5N1	A/Viet-Nam/1203/2004/H5N1	0.17	0.52	ns
log10_IC50	A/Chicken/Deli_Derdang/BBPVI/2005/H5N1	A/goose/Guiyang/337/2006/H5N1	0.33	0.63	ns
log10_IC50	A/Chicken/Deli_Derdang/BBPVI/2005/H5N1	A/Viet-Nam/1203/2004/H5N1	0.46	0.63	ns
log10_IC50	A/goose/Guiyang/337/2006/H5N1	A/Viet-Nam/1203/2004/H5N1	0.065	0.39	ns

Table 6. A comparison between the neutralisation magnitude of pseudotypes run on the wild avian sample set. P.ajd denotes the BH corrected p-value arising from multiple t-tests. P.adj.signif denotes the significance level of the difference between means at the 95% significance level.

10.2.9 Hatch date does not influence neutralisation of pseudotyped influenza viruses displaying avian HAs

As with the previous serosurvey using sera from Scottish blood donors (Chapter 9), I tested to see whether the neutralisation is impacted by the age of the birds. The birds whose samples were used in this study had sera taken during their birth year. Each sample has an associated hatch date except for 13 samples. These were excluded from this analysis.

Most birds were hatched between 2012 and 2017. Five birds had samples taken in 2008. Excluding the 2008 samples, which represent a small minority compared to the larger sample sizes from 2012 and 2017, enhances the robustness and generalisability of the analysis. By focusing on the majority group of samples, potential biases or anomalies associated with the smaller subset from 2008 can be mitigated, resulting in more reliable and representative modelling outcomes. Models were created with both the 2008 samples included and excluded (see Supplementary Figures 4 to 6). For presentation in the main text, I chose to exclude the 2008 samples from the models.

There were too few hatch years to enable us to test for non-linear relationships between age and neutralisation with the GAM approach taken for the human serosurvey. Therefore, I did not test for non-linear relationships between age and neutralisation responses in the avian serosurvey. Moreover,

there were too few hatch years to make conclusions using the rolling average method. I therefore tested for linear, or near-linear, relationships.

I first used linear models to test for linear relationships between hatch year and neutralisation strength. I constructed weighted and unweighted models. Hatch year was significant for A/goose/Guiyang/337/2006/H5N1 (Figure 24 a). The relationship is negative: the later the hatch year, the smaller the neutralisation response. The adjusted R^2 of this model is small: 0.066.

I also used logistic regression to test for linear or near-linear relationships between hatch year and seropositivity chance. Age was significant for A/Viet-Nam/1203/2004/H5N1 (Figure 24 b). The relationship is negative as with A/goose/Guiyang/337/2006/H5N1 with a McFadden pseudo-R^2 of 0.69.

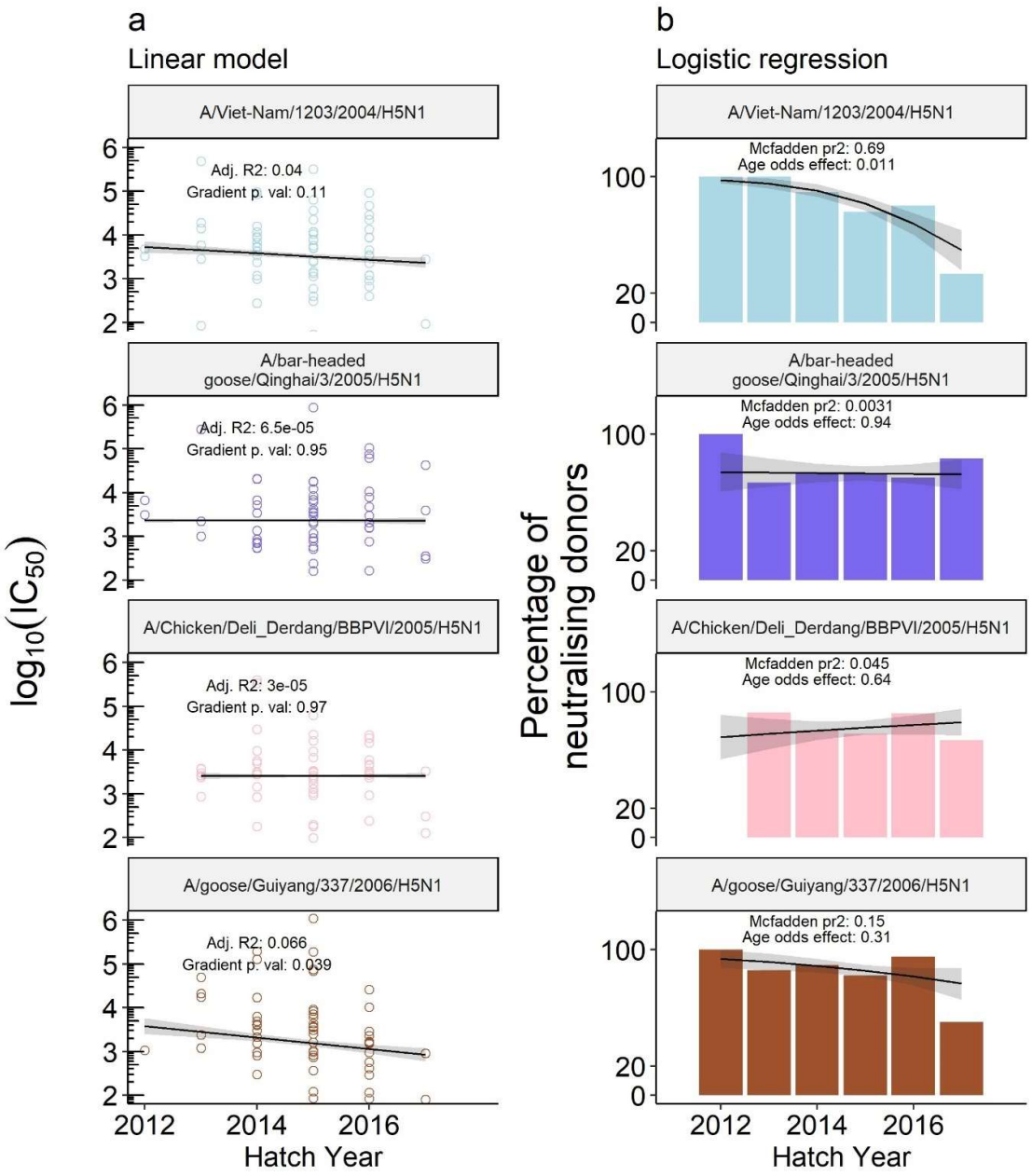


Figure 24. The relationship between hatch year and neutralisation of pseudotypes run on the wild avian sample set. Facets denote individual pseudotypes. Points in Figure a denote neutralisation. Black lines show the output of unweighted linear models. Shaded areas around the line denote the standard error of the model. Adj. R2 denotes the adjusted R^2 of the model. Gradient p.val denotes the significance of age in the model. Bars in Figure b denote the percentage of seropositive donors by hatch year. Black lines on b show the output of logistic regression models. Shaded areas around the line denote the standard error of the model. McFadden pr2 denotes the McFadden pseudo-R2 of the model. Age odds effect denotes the p-value of age's significance in the model.

10.2.10 Comparison between human and H5 virus-exposed avian neutralisation responses towards A/bar-headed goose/Qinghai/3/2005/H5N1 pseudotypes

Pseudotypes expressing the HA of A/bar-headed goose/Qinghai/3/2005/H5N1 were tested with both Scottish blood donors and likely H5-exposed (Methods 8.4.4), swan sera samples. The percentage of donors neutralising pseudotypes expressing A/bar-headed goose/Qinghai/3/2005/H5N1 was greater in humans than in swans. Bayesian 95% confidence intervals do not cross, Figure 25 a. However, the mean magnitude of neutralising responses was greater in swans than in humans, t-test p-value < 0.0005 (Figure 25 b).

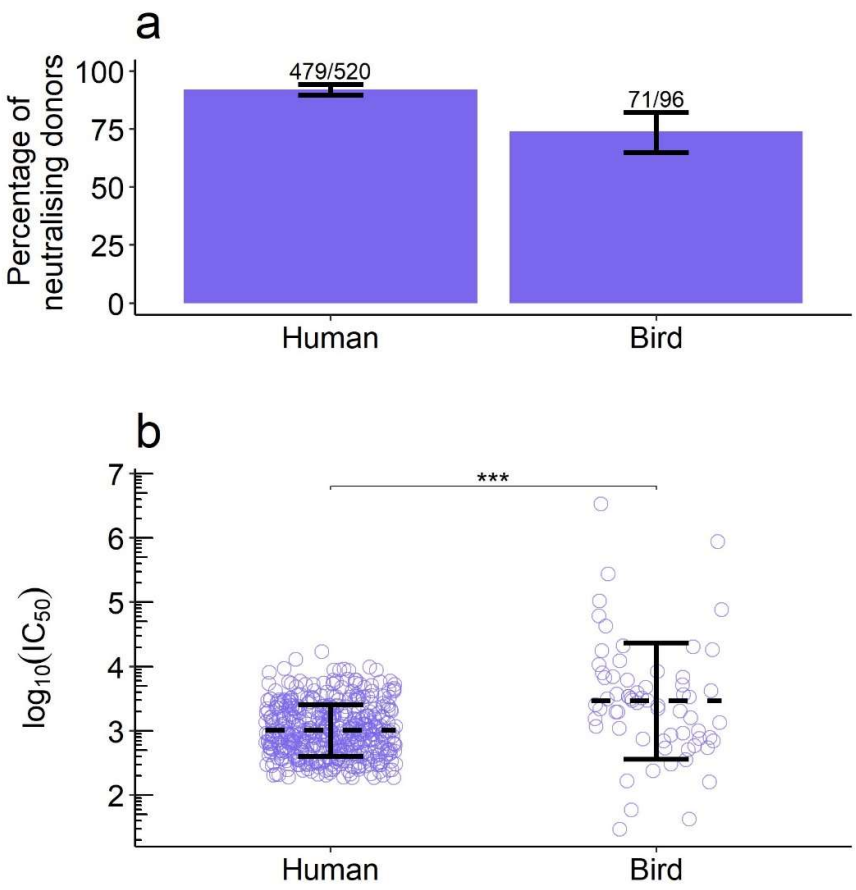


Figure 25. A comparison between the human and avian neutralisation response to A/bar-headed goose/Qinghai/3/2005/H5N1 virus. Figure a shows the percentage neutralisation. Error bars denote the Bayesian 95% confidence interval of seropositivity. The number of donors run is labelled on top of the bars. Figure b shows the magnitude of neutralisation. Error bars denote the 95% confidence interval of neutralisation strength; the horizontal dashed lines denote the mean. The y-axis denotes $\log_{10}(IC_{50})$. Differences between means of human and avian responses were determined by student's t-test. Significance stars: * p-value < 0.05, ** p-value < 0.005, *** p-value < 0.0005.

10.2.11 Scottish blood donor and swan sera sample set demographics

Age and sex-controlled random samples of the Scottish blood sera sample sets were produced using the R package "dplyr." A/bar-headed goose/Qinghai/3/2005/H5N1 pseudotypes were tested on human sample one. A/South Carolina/1/1918/H1N1 pseudotypes were tested on human sample two. All other pseudotypes were tested on human sample two. Human sample two consists of all samples in human sample three but with 89 extra age and sex random samples. Human sample one consists of all samples in human sample two but with 158 extra age and sex random samples. Figure 26 a details histograms showing these distributions.

Age and hatch date controlled random samples of the swan sera sample set were produced using the R package "dplyr." A/Viet-Nam/1203/2004/H5N1 and A/bar-headed goose/Qinghai/3/2005/H5N1 were tested on avian sample one.

A/Chicken/Deli_Derdang/BBPVI/2005/H5N1 and A/goose/Guiyang/337/2006/H5N1 pseudotypes were tested on avian sample two. Avian sample one consists of four more, age and hatch-date-controlled random samples of swan sera than sample two. Figure 26 b details histograms showing these distributions.

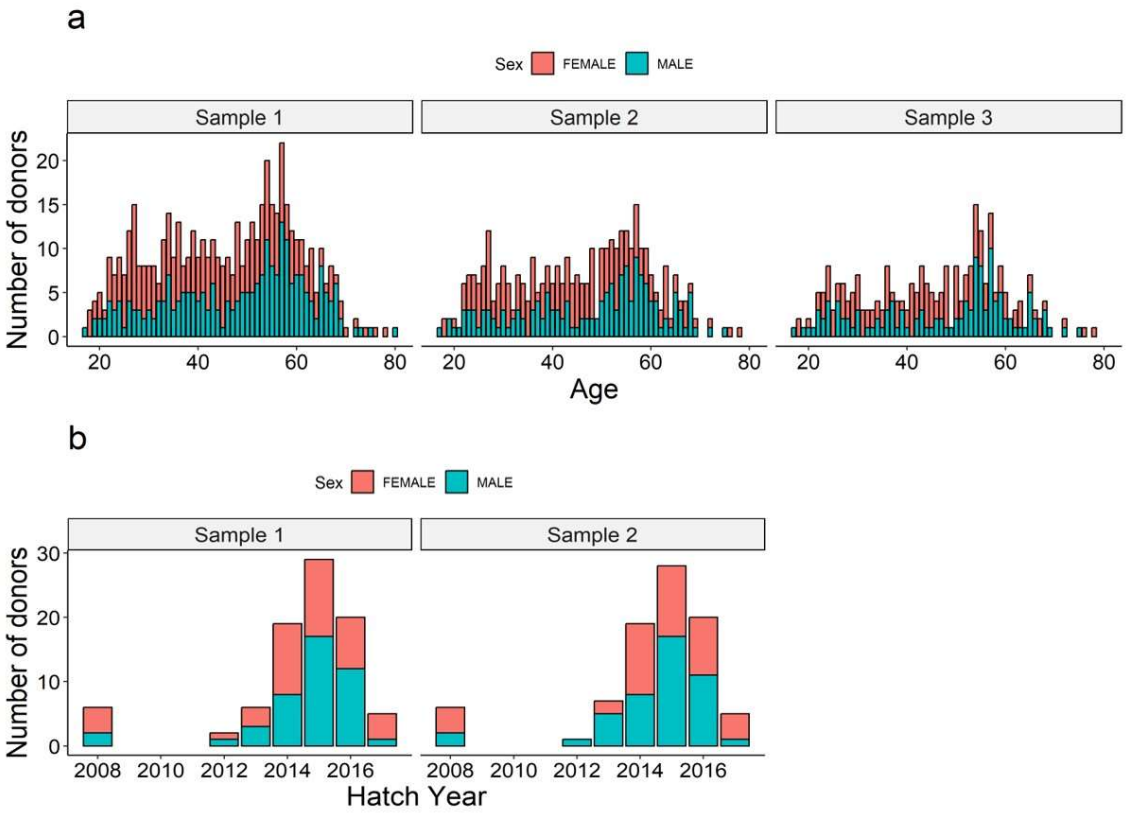


Figure 26. A histogram detailing the demographics of the Scottish blood donor and avian sera samples sets. Figure a details the Scottish blood sera sample set and Figure b details the avian sera sample set. Columns are split up into male and female. Male and female columns are stacked meaning to read female data, take the count from the top of the male bar to the end of the female bar. The y axis denotes the number of samples and the x-axis denotes age in Figure a and hatch year in Figure b. Facets denote the sample used.

10.3. Discussion

10.3.1 Summary

I conducted two large serosurveys comprising 2,357 human blood donors and 376 wild swan sera runs. We developed an in-house pseudotyped influenza microneutralisation assay run on an automated robotic pipeline (Methods 8.1.7). Pseudotype viruses expressing ebola's GP were used to set the limits of detection, allowing me to determine which samples contain neutralising antibodies.

I found substantial population immunity towards the HA of an avian H5N1 virus of pandemic concern in a likely unexposed Scottish blood donor cohort. I found population immunity of differing levels towards historical H1 pseudotyped influenza viruses not previously seen by any members of the cohort. In contrast, I found no population immunity to a historical avian H5N1 virus collected in 1959: the first isolated H5N1 virus. I also found substantial population immunity to four HPAI H5N1 strains in a British wild swan sample cohort.

Age has a significant, non-linear impact on seropositivity chance and neutralisation strength towards certain influenza pseudotypes (Figures 21 and 22). Age significantly impacts the seropositivity probability for A/PR/8/1934/H1N1 and A/Solomon Islands/3/2006/H1N1. The age-neutralisation probability relationships between A/PR/8/1934/H1N1 and A/Solomon Islands/3/2006/H1N1 appear to take opposite shapes between the ages of 20 and 60, (Figure 21 b). Age also significantly impacts the neutralisation strength of A/bar-headed goose/Qinghai/3/2005/H5N1 and A/USSR/90/1977/H1N1 in GAM models. Both show a positive relationship but with peaks and troughs. No clear age relationships were found for: A/South Carolina/1/1918/H1N1, A/California/07/2009/H1N1 or within the avian serosurvey (Figures 21 and 22).

10.3.2 Antigenic imprinting and cross-reactivity between viruses

The high level of anti-H5N1 HA neutralising immunity in the Scottish blood donor sample set (Figure 20) is unlikely to be generated from natural infection with A/bar-headed goose/Qinghai/3/2005/H5N1

or any other H5N1 strain as H5 HA strains do not circulate in the human population. There have been no documented human H5N1 infections in Scotland. It is likely that this immunity is due to cross-reactive antibodies generated by similar HAs. H1N1, H2N2 and H5N1 viruses are phylogenetic related and collectively referred to as Group 1 HAs [348]. The similarities between H1 and H5 HAs are discussed further in Results 8 and 9. H1N1 circulates annually in Scotland. It is possible that human H5 cross-reactivity may be derived from previous infections with H1N1 strains.

Gostic *et al.,* (2016) suggested that incidence and mortality in South East Asian H5N1 cases showed differing patterns depending on birth date [349]. People born before 1968 had a lower incidence/ mortality than those born after 1968. A potential partial explanation for this could be due to antigenic imprinting (Introduction 7.8). H1N1 became extinct from circulation in 1957. H1N1 was replaced by H2N2 (another group 1 HA) which circulated until 1968 when it was replaced by H3N2 [350] [5]. H1N1 started circulating again, succeeding the "Russian flu" epidemic [350]. Post-1977, H1N1 (group 1 HA) has been co-circulating with H3N2 (group 2 HA [348]) (Figure 2), causing around half the cases that H3N2 does [103]. Gostic *et al.,* (2016) propose that children born before 1968 primarily imprinted on group 1 HA viruses (H1N1 and H2N2) [349]. However, from 1968 to 1977, this would not have been possible as H1N1 was not circulating, and therefore children born in this timeframe would be primarily imprinting on H3N2 or H2N2. Moreover, from 1977 to the present day, children are less likely to experience H1N1 than H3N2 annually. In summary, the hypothesis suggests that people born before 1968 developed preferential immune responses towards group 1 HAs (including H5N1) and people born after 1968 developed preferential immune responses towards group 2 HAs.

In this study, donors with an age of 52 were born in 1968 whilst donors with an age of 43 were born in 1977. Intriguingly, the pattern describing the relationship between neutralisation strength and age for A/bar-headed goose/Qinghai/3/2005/H5N1 and A/USSR/90/1977/H1N1 both contain peaks/ troughs around the ages of 52 and 43, respectively (Figure 22). In addition to this, some of the post-60-year-old decline in strength could be explained by immune-senescence.

The highest neutralisation $\log_{10}(IC_{50})$ values were found for the A/USSR/90/1977/H1N1 HA pseudotyped virus from individuals around age 52 (Figure 22). From the late 50s, neutralisation begins to decline until age 70. However, it only declines to half the level of the peak relative to the dip at age 20. Moreover, the A/bar-headed goose/Qinghai/3/2005/H5N1 neutralisation strength-age relationship dips at age 43, lasting towards nearly 50 years old (Figure 22). Neutralisation strength then increases with ages older than 50.

A potential explanation for the A/USSR/90/1977 H1N1 response trend could be that donors of age 52 and older imprinted predominantly on H1N1/ H2N2 epitopes shared with A/USSR/90/1977/H1N1 when young. This would hypothetically cause them to mount a stronger immune response to A/USSR/90/1977/H1N1 than those born before 52. Whereas donors 43 years and older imprinted on H1N1 epitopes shared with A/bar-headed goose/Qinghai/3/2005/H5N1. If no epitopes are shared between H2N2 and A/bar-headed goose/Qinghai/3/2005/H5N1 (which circulated between 1958 and 1968), only between H1N1 and A/bar-headed goose/Qinghai/3/2005/H5N1, this could explain why donors older than 50 mount a stronger immune response to A/bar-headed goose/Qinghai/3/2005/H5N1.

Another explanation for these trends could be that immune responses increase in strength due to sequential exposure to influenza viruses over the course of an individual's life. Sequential exposure to divergent influenza strains has been shown to increase broadly-neutralising antibody immunity (Introduction 7.7). Furthermore, cross-reactivity could be directed towards the conserved stem region or the immunodominant globular head.

In addition to neutralisation strength, the probability of neutralising antibodies also varied with age for A/PR/8/1934/H1N1 and A/Solomon Islands/3/2006/H1N1. Strangely these patterns appear to be inverse from the ages 20 to 60 (Figure 21): with the detected neutralising activity to the A/PR/8/1934/H1N1 peaking between 20-60 and the detected neutralising activity to A/Solomon Islands/3/2006/H1N1 declining. Interestingly, the maximum for A/PR/8/1934/H1N1 and minimum for

A/Solomon Islands/3/2006/H1N1 are situated around age 43, when H1N1 was reintroduced into circulation [350].

Neither A/South Carolina/1/1918/H1N1 or A/California/07/2009/H1N1 exhibited a significant age trend. The 3-year rolling average for seropositivity chance and neutralisation strength contains many peaks and troughs (Figures 21 and 22). This is potentially due to natural variance around zero.

No clear trend between hatch year and neutralisation strength or seropositivity was elucidated in the avian serosurvey. However, a significant negative trend between hatch date and neutralisation strength was seen for A/goose/Guiyang/337/2006/H5N1 (Figure 24). However, this trend has a very low R^2 of 0.066. Moreover, seropositivity chance and hatch date showed a significantly negative trend in A/Viet-Nam/1203/2004/H5N1 (Figure 24). This trend has a high McFadden pseudo-R^2 of 0.69. However, when 2008 is included in the analysis, age becomes insignificant (Supplementary Figure 4). In addition, there were no significant differences between the immune responses between the four H5N1 pseudotypes tested on the swan sera. This could be due to all these pseudotyped HA strains being antigenically similar. The strains were only sampled between 2004 and 2006; all in Asia. Therefore, being antigenically similar is likely.

10.3.3 Suitability of comparisons between the neutralisation of different pseudotyped influenza viruses

Pseudotyped viruses are constructed by displaying a viral glycoprotein (HA in this study) of interest on the membrane of a virus with a lentiviral core, containing an internal luciferase reporter. Differences in glycoprotein structure impact how antisera bind to the pseudotype and thus, entry into HEK293T cells [351]. Therefore, by using multiple pseudotypes, I can compare the neutralisation response against different glycoproteins. However, other factors than fusion glycoprotein structure may impact the neutralisation response. Differences in glycoproteins may change: (i) how much fusion

glycoprotein is displayed on the pseudotype surface and, (ii) the distribution of glycoprotein on the surface of the pseudotype [222]. Both these factors may impact how antisera bind to pseudotypes.

Normalising responses to a monoclonal antibody (mAb) can reduce some of this variation and allow for better comparisons. However, any mAb used should bind the exact same epitope or binding site for each virus. I could not find any mAb that would bind the same conserved site for all HAs I used in our serosurvey.

10.3.4 Antigenic characterisation of A/South Carolina/1/1918/H1N1 and A/California/07/2009/H1N1 have similar antigenicity profiles

Thompson *et al.,* (2018) found that children's sera exposed to A/South Carolina/1/1918/H1N1 and A/California/07/2009/H1N1 reacted similarly [147]. Moreover, both the "Spanish flu" and "Swine flu" showed similar age patterns [352], [353]. I have also shown a similar trend in our serosurvey. There is no significant difference between the neutralisation strengths of the two pseudotyped influenza viruses (Figure 20 and Table 5). Neither virus showed any significant age trend (Figures 21 and 22).

10.3.5 Avian neutralising immunity did not differ between pseudotyped influenza virus strains despite being from diverse phylogenetic clades

I tested four HPAI H5N1 pseudotyped influenza virus strains on a wild avian sera sample set in microneutralisation assays. These strains were collected between 2004 and 2006. Despite being from similar years, all were from different phylogenetic clades as defined by the WHO's unified HPAI nomenclature [31], [32], [33]. Wild avian neutralising immunity showed no differences between the strains used (Figure 23 and Table 6). This occurred for both seropositivity chance and neutralisation strength. This could suggest that avian neutralising immune responses are primarily targeted towards conserved regions shared between strains occupying diverse phylogenetic clades. The detection of conserved regions of neutralising cross-reactivity was investigated further in Chapters 10 and 11.

However, this could also be because, from the years 2004 to 2006, HAs did not have sufficient time to evolve diverging antigenicity profiles.

10.3.6 Definition of Positive Samples

Responses to pseudotyped viruses expressing ebola's GP were used as a negative control. IC_{50} values were used to set thresholds for seropositivity, Figure 19. Ebola has never been documented to infect humans or swans in the United Kingdom [330]. It is possible, if highly unlikely, that people and/or birds in the United Kingdom have been in contact with pathogens that confer cross-reactivity to ebola. Therefore, I hypothesise that any detectable neutralisation responses occurring for the ebola pseudotyped virus may be due to non-specific factors. Non-specific factors may include activation of the complement system, which could lead to the opsonisation and subsequent neutralisation of pseudotypes independent of specific antibody-antigen interactions. Additionally, the presence of natural antibodies or pre-existing anti-lentiviral factors, such as those targeting viral vectors or conserved epitopes shared by unrelated viruses, could also contribute to the observed neutralisation. These non-specific immune responses are important to consider, as they can lead to false-positive results in serological assays, thereby complicating the interpretation of neutralisation data.

10.3.7 Assessing pandemic risk

A pandemic occurs when a pathogen emerges that (i) spreads from human to human, and (ii) limited or no immunity exists towards it in the general population [189]. Neutralising antibodies towards viruses is known to be a correlate of protection [309], [310]. Our serosurvey detecting neutralising antibodies towards SARS-CoV-2's Spike protein for the Scottish National Blood and Transfusion Service (SNBTS) found that population immunity was low during the first COVID-19 pandemic wave. From this, I predicted that further spread of the SARS-CoV-2 virus was likely (Chapter 9).

I found that 479/520 Scottish blood donor samples neutralised A/bar-headed goose/Qinghai/3/2005/H5N1. In contrast, 2/273 Scottish sera samples neutralised A/chicken/Scotland/1959/H5N1. Our sample set contains donors primarily between the ages of 18-65 years old. This result may indicate that there is some standing protection towards the A/bar-headed goose/Qinghai/3/2005/H5N1 HA in the Scottish population. This could reduce the likelihood of this strain spreading in this population.

However, other conditions may also play a part in pandemic production. For example: public health measures taken during the initial outbreak; remoteness of the initial outbreak and opportunities for people to travel; climate at the epicentre and preexisting cellular-based immunity all may play a part in the production and maintenance of pandemics. This serosurvey does not provide data that addresses these points.

Moreover, I cannot assess whether the neutralising immunity I measured will lead to changes in clinical outcomes after an infection. For example, our group showed that a higher degree of cross-reactive humoural immunity between beta coronaviruses and SARS-CoV-2 is associated with negative clinical outcomes in ICU patients [91].

Finally, I have tested the neutralisation response to influenzas' HA glycoprotein. This is justified because this is influenza's primary antigenic determinant [5]. However, I have not assessed immunity towards other influenza proteins.

11. Identification of epitopes of limited variability in the HAs of H5 and H1 influenza viruses using bioinformatics

11.1. Background

HA is influenza's most abundant surface protein and plays a key function in viral entry to target cells, it is estimated to be the target of 60% of anti-influenza antibodies [5]. Consequently, HA is the primary antigen used for most influenza vaccines [5]–[7].

HA can be divided into two domains: a head domain (52-277 in H3 numbering) and a stem domain (AA <52, >277 in H3 numbering) [39] (Figure 3). The HA head domain binds to sialic acid, the cell receptor on the surface of target cells. Endocytosis is initiated once bound to the target cell and the virion is enclosed in a vesicle. The pH drops within the vesicle, which in turn causes the stem domain via the fusion peptide to change conformation and initiate fusion of the viral and cell membranes to release the genome into the target cell (discussed in Introduction 7.3 to 7.4).

Influenza evolves through the incremental accumulation of mutations in surface proteins, such as HA. This allows viruses containing these mutations to avoid antibodies produced by the host immune system [5], [99]–[103], [106], [107]. This process is known as 'antigenic drift' [19]–[25]. Strains containing advantageous mutations spread throughout the host population to become the most prevalent seasonal strain (Introduction 7.10 to 7.11).

The stem domain anchors HA to the virion and contains the fusion components. In contrast, the globular head domain is more exposed to the host immune system, and consequently, is under greater selective pressure (Introduction 7.10). The stem is less variable than the head domain [241]. Deep mutational scanning has been performed on the HAs of H1N1 and has shown that the stem domain is intrinsically less variable than the head domain [241].

Despite the intrinsic variability of the influenza virus proposed by the antigenic drift theory, epidemiological and phylogenetic studies of influenza circulating in the human population have shown that the genetic and antigenic diversity of HA must be restricted to some extent [16], [98], [144], [354] as influenza seasons tend to be dominated by a single or limited number of strains [135], [145]. The constrained nature of the antigenic evolution of influenza in humans remains a matter of debate. However, several modelling-based hypotheses exist; each approach constrains the evolution of the virus in some manner[126], [135], [355]–[362], [136]–[142], [146].

Some of these evolutionary mechanisms predict the existence of epitopes of limited variability (ELVs) situated within the more variable head domain [144]–[146]. Studies involving monoclonal antibodies have identified regions of HA that cross-react with multiple strains within a subtype. These antibodies tend to bind to areas of HA which carry out a function and are consequently restricted in their variability. These regions include the i) receptor binding site (RBS) [84], [149], [363], which binds to sialic acid, ii) lateral patch [125], which enables monomers of the HA trimerise, and iii) the fusion peptide of the stem domain [364], which initiates fusion of the virion membrane and the vesicle membrane during infection

A second strand of evidence suggesting that ELVs exist is serological studies. These studies involve running serological assays against a panel of chronologically distinct influenza viruses or antigens [5], [123]. These studies often show that immunity to one strain may confer protection against a number of chronologically distinct strains, this suggests that epitopes are 'recycled' over time as population immunity changes [5], [123]. For example, Thompson *et al.,* (2018), showed that specific conformation of a restricted variability HA head epitope encompassing the RBS induces cross-reactive immunity to H1N1 strains that last circulated in 1934, 1977 and 2006 [147]. The existence of low variability HA head epitopes, which exist as a limited number of conformations, and are shared between multiple strains.

No studies have identified ELVs that are shared between strains from different subtypes, such as between H5N1 and H1N1. I have shown that, within the same Scottish blood donor sera sample set, population immunity exists towards the H5N1 strain A/bar-headed goose/Qinghai/3/2005 and multiple H1N1 influenza strains (Chapter 10).

The HAs of H5N1 and H1N1 are phylogenetically closely related [365]. Moreover, the HAs appear to be structurally similar (Figure 27). Consequently, I hypothesise that the anti-H5N1 neutralising immunity is partially elicited due to cross-reactive H5N1-H1N1 antibodies. I hypothesise that these cross-reactive antibodies recognise ELVs shared between H5N1 and H1N1.

Therefore, I performed structural bioinformatic analyses to identify ELVs that are shared between H5N1 and H1N1. The rationale for this analysis was to produce hypothesis-testing data. Putative ELVs identified during this analysis were then to be tested in the laboratory setting to provide supporting evidence of their existence.

11.2. Results

11.2.1 The head domain of Haemagglutinin is more variable than the stem domain

Previous studies have shown that the HA head is more variable than the stem domain. I first attempted to determine if this is true using the amino acid and crystal structure datasets described in Methods 8.4.1.

Shannon entropy is an information theory concept that describes uncertainty in a dataset. By applying this to an amino acid alignment, I can determine the total amino acid existence uncertainty of a certain residue position. This can be thought of as the likelihood of residues in that position to mutate. Thus, this can be described as mutational entropy [241].

Figure 27 a shows the mutational entropy for every position in the H5N1 and H1N1 alignments. Figure 27 c and d displays this data on H5N1 and H1N1 crystal structures. Examination shows that the head domains of both H5N1 and H1N1 have a greater mutational entropy than the stem domains. This is confirmed by statistical analysis. For both subtypes, there is a significant difference between the mean variability of head and stem when applying a student's t-test (Figure 27 b).

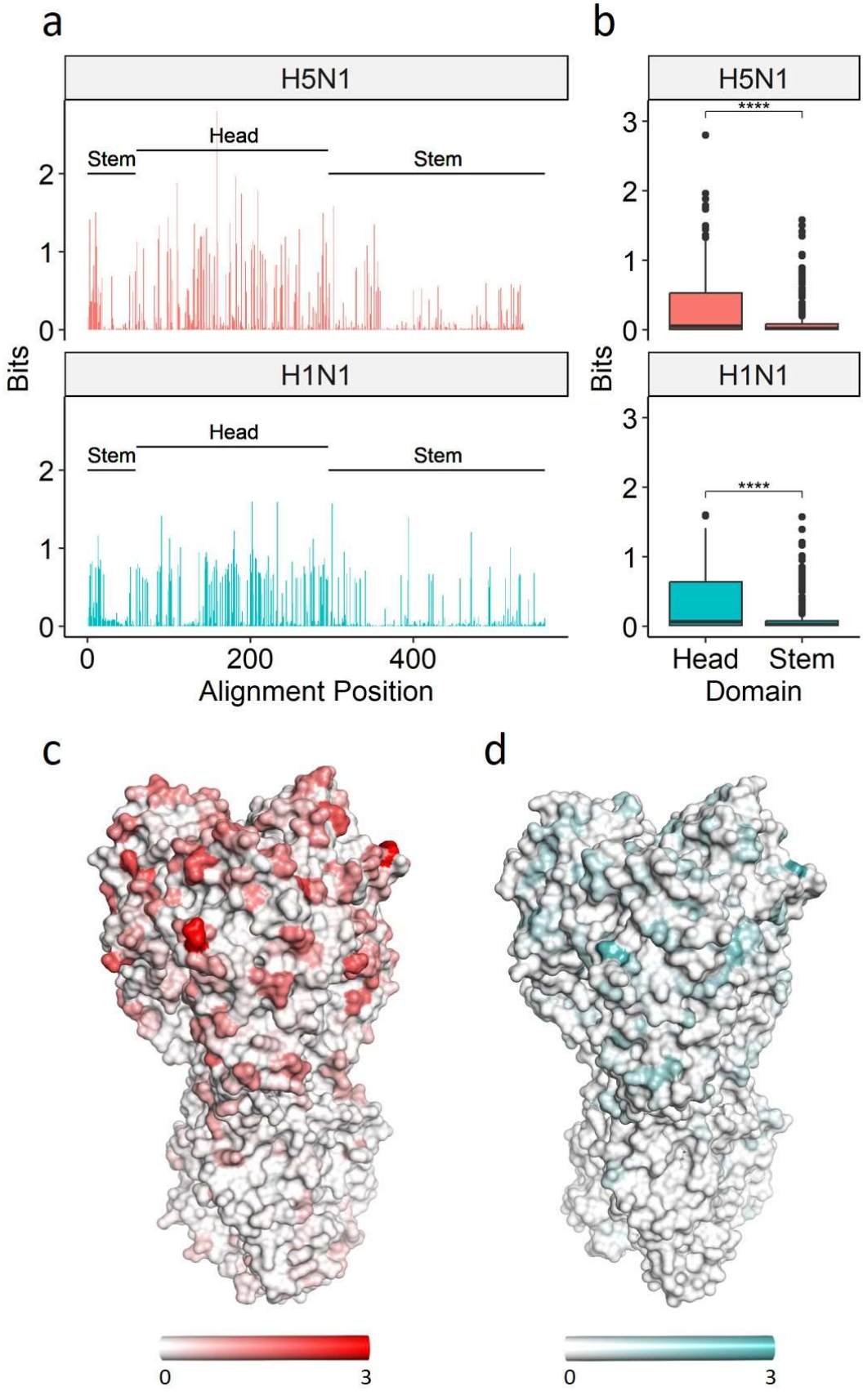

a
H5N1
Head
Stem
Stem
H1N1
Head
Stem
Stem
Bits
0
1
2
0
200
400
Alignment Position
b
H5N1
****
H1N1
****
Bits
0
1
2
3
Head
Stem
Domain
c
d
0
3
0
3

Figure 27. Variability of the head and stem domains of H5NX and H1N1. Mutational entropy is displayed in the unit "bits". Figure a shows the mutational entropy for every position in the H5Nx and H1N1 alignments. Figure b is a boxplot showing the median, interquartile range and outliers for H5Nx and H1N1 amino acid alignments. This is split into head and stem domains. Differences between means of head and stem were determined by student's t-test. Significance stars: * p-value < 0.05, ** p-value < 0.005, *** p-value < 0.0005 and **** p-value <0.00005. Figures c and d show the mutational entropy displayed on the crystal structures 5e2y (H5N1) and 6ona (H1N1). Scale bars denote the mutational entropy extent.

11.2.2 There are regions of limited variability in the H5 HA head domain

I attempted to determine if there are immunogenic regions of limited variability within the H5 HAs. I used the structural bioinformatics program described in Methods 8.2.2 to do this. In brief, the programme divides HA up into theoretical antibody binding sites (ABSs). It does this by iteratively taking every residue in a crystal structure and selecting nearby residues. Nearby residues are determined by selecting an area, in Å², from which to look. With this data, I can map theoretical ABSs onto an amino acid alignment.

ABS sizes of 400 Å², 600 Å², 800 Å², and 1000 Å² were used as they correspond to a range of empirically determined ABS sizes from antibody-HA crystal structures [240]. ABSs are comprised of solvent-accessible residues. Therefore, to generate putative ABSs I used SwissModel to select different thresholds of solvent accessibility to perform the analysis with. I selected the thresholds of 5%, 10%, 20%, and 30% solvent accessibility to perform analysis (Supplementary Figures 7 to 9).

Hamming distance as a concept is the measurement of the number of differences between strings. Mean hamming distance describes the average dissimilarity between multiple strings. By applying this concept to the bioinformatics pipeline, I can determine the average number of differences for ABSs mapped onto an amino acid alignment. I can think of this as measuring the variability of theoretical ABSs.

I applied this concept to determine the variability of putative ABSs for pan H5Nx influenza subtypes. I selected the crystal structure 5e2y (H5N1) for the analysis. I used the amino acid alignments described in Methods 8.4.1 to compute mean Hamming distances. These comprised a pan-H5 amino acid alignment (H5Nx) containing strains isolated from avian hosts. This describes the variability of putative ABSs as found in nature.

Putative ABS variability is greater in the head than in the stem domain (Figure 28). This trend is conserved across all putative ABS sizes. It is also conserved across all solvent accessibilities (Supplementary Figure 7). Interestingly, despite being more variable than the stem domain, the head domain contains dips in variability. Dips situated around alignment numbers 157 and 189 are notably lower variability than many of the putative ABSs in the stem domain.

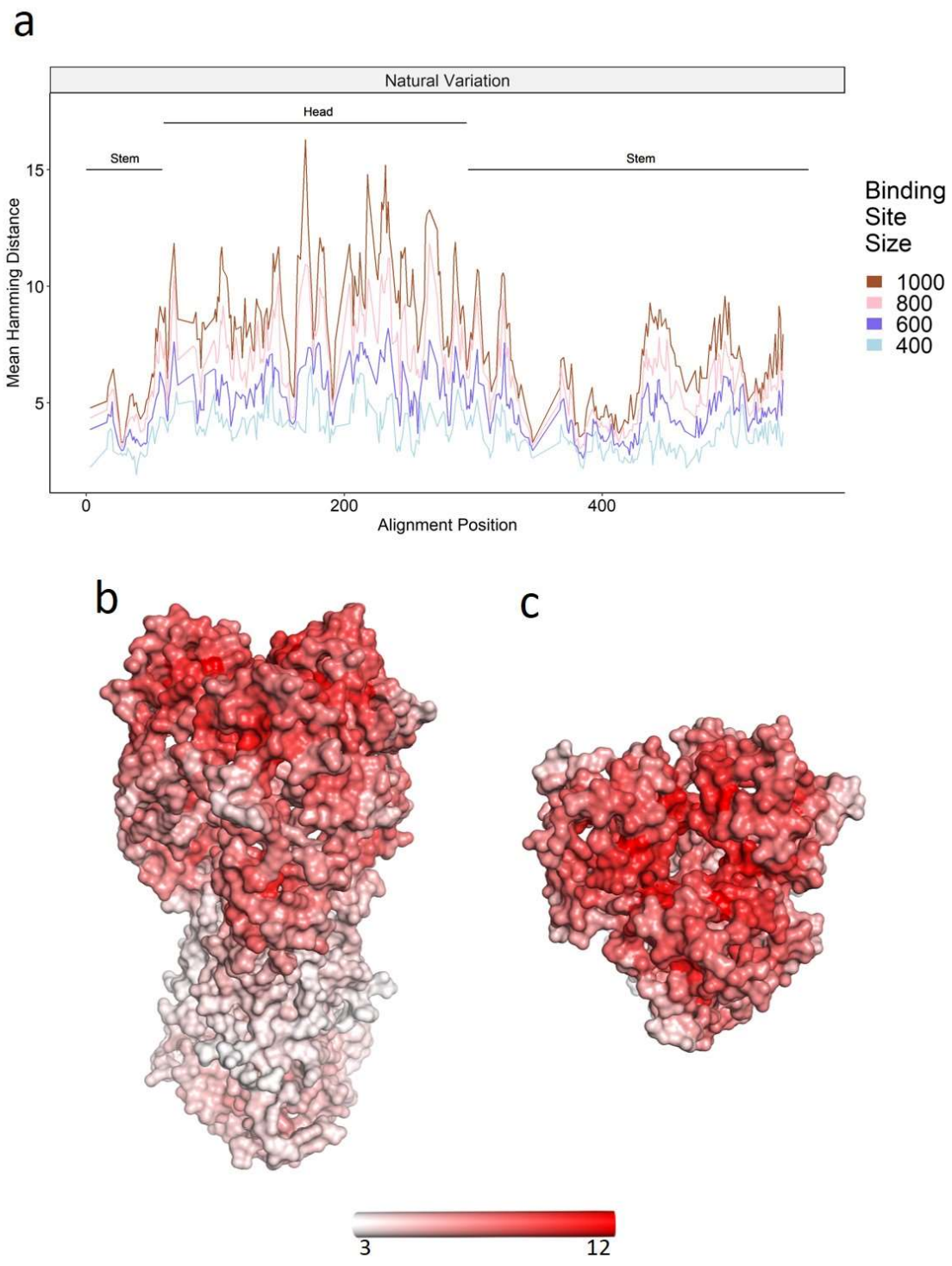

a
Natural Variation
Head
Stem
Stem
Mean Hamming Distance
15
10
5
0
200
400
Alignment Position
Binding Site Size
1000
800
600
400
b
c
3
12

Figure 28. Variability analysis of the H5 HA. Figure a shows the output from the bioinformatic structural variability pipeline for avian H5 HA sequences mapped to the 5e2y (A/duck/Egypt/10185SS/2010) crystal structure. The x-axis is the numbering from a structural alignment between H1N1 and H5Nx. The x-axis denotes the position in the alignment which occupies the central position of the predicted ABS. The HA head and stem are labelled above. The y-axis denotes variability in the units of mean Hamming distance. The four lines on each graph denote putative ABS of size 400 Å², 600 Å², 800 Å² and 1000 Å² by the line colours. Figures b and c show the variability analysis displayed on the 5e2y trimer crystal structure. Figure c shows the top view of the crystal structure. The scale bar denotes the variability.

11.2.3 There are regions of limited variability in the H1 HA head domain

I then ran the same analysis using the H1 HA (crystal structure: 6ona A/Hickox/1940) and an alignment of H1N1 HA amino acid sequences known to have circulated in humans. As with H5Nx, the variability of putative ABSs is higher in the head than in the stem. The same pattern across the alignment is conserved between all binding site sizes. Moreover, there are also notable dips in variability situated in the head domain that are lower than some putative ABSs in the stem domain, Figures 29 a, c and e.

The above analysis was done to assess the variability of putative H1N1 ABSs seen in nature. In addition to this, I performed analyses to assess the theoretical structural variability of these putative ABSs, i.e. to see how plastic these ABSs could be without evolutionary selection pressure. To do this I ran data from a comprehensive deep mutational scanning study performed by Bloom *et al.,* (2016) through the pipeline [241].

Deep mutational scanning involves mutating every residue in the H1N1 HA to all possible amino acids. By determining the frequency of these mutations, the amino acid preferences can be worked out. Amino acid preferences refer to the chance of a residue mutating to a certain amino acid. The Shannon mutational entropy of each position describes the inherent mutational tolerance of that residue. A higher mutational tolerance (measured in bits) describes how plastic each residue in the alignment position is.

I ran the mutational tolerance data for each H1N1 residue through our putative structural bioinformatics pipeline. I determined the mean mutational preference for each putative ABS at 400

Å², 600 Å², 800 Å², and 1000 Å²; and 5%, 10%, 20%, and 30% solvent accessibility (Figure 29 b, d and f).

Mutational tolerance is not uniform across the H1N1 HA crystal structure. As with the natural variation analysis, there are peaks and troughs throughout the structure (Figure 29 b). There is higher mutational tolerance within the head of HA than the stem (Figure 29 b, d, and f).

There are also multiple differences between the mutational tolerance (Figure 29 b) and natural variation outputs (Figure 29 a). In particular, there is a difference between the variability and mutational tolerance of putative stem ABSs. Some stem putative ABSs exhibit a mutational tolerance that is relatively high when compared to the head domain, Figure 29 b. Whereas, predominantly, stem ABSs exhibit Hamming distances that are relatively low when compared to the head domain, Figure 29 a.

Interestingly, the pattern of mutational tolerance in the head seems to be the inverse of the natural variability analysis (Figure 29). Notably, troughs in variability at 157, 189, and 260 are all matched by a peak in mutational tolerance (Figure 29 a and b). This is visible in the crystal structures shown in Figures 29 d to f. For example, the middle head domain is shown to be of a high natural variability but a low mutational tolerance, Figure 29. This is exemplified by the correlation between the two outputs. The Pearson's correlation between the head domain is 0.07 (0.14 to 0.015 95% confidence interval).

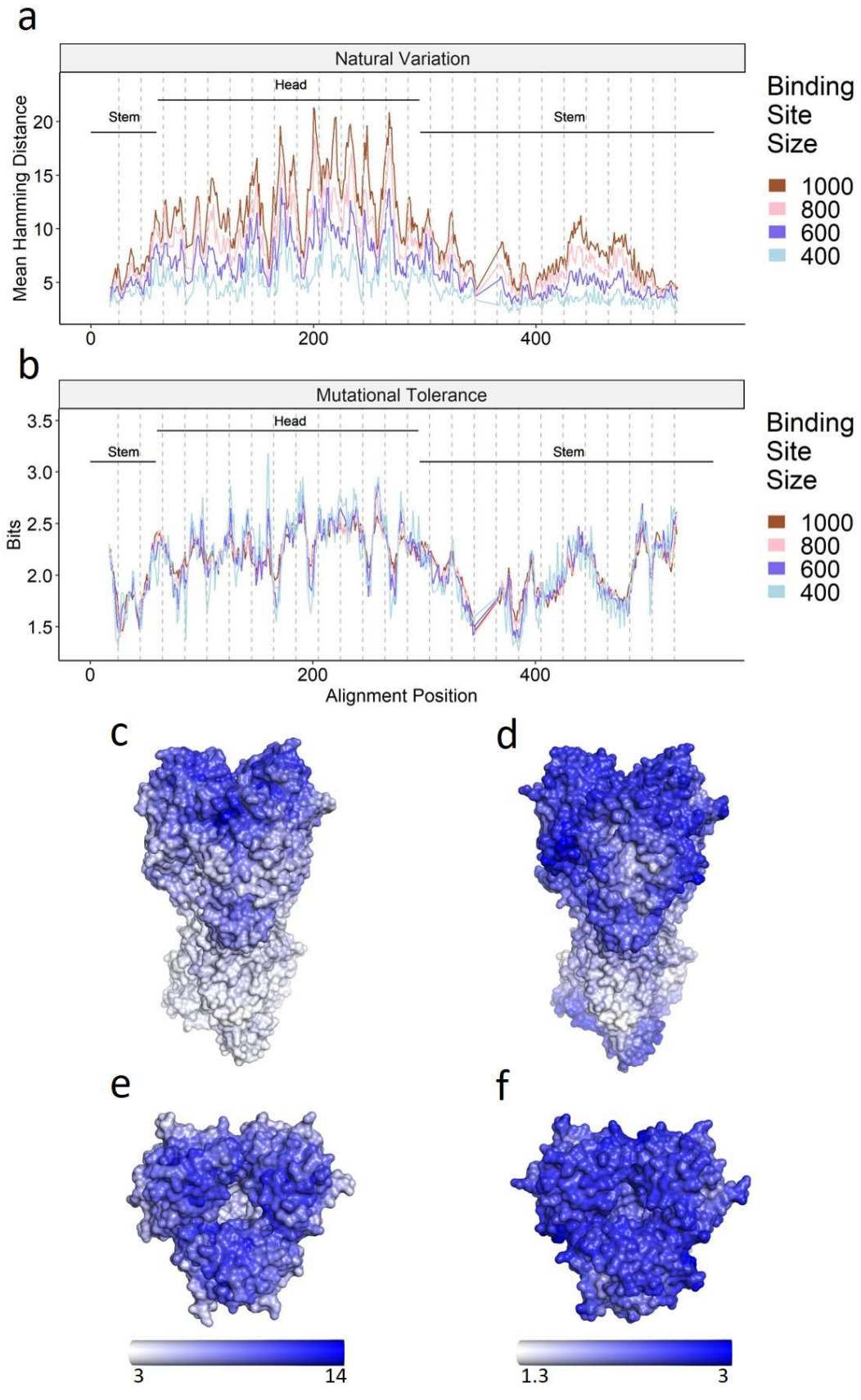

a
Natural Variation
Head
Stem
Stem
Mean Hamming Distance
20
15
10
5
0
200
400
Binding Site Size
1000
800
600
400
b
Mutational Tolerance
Head
Stem
Stem
Bits
3.5
3.0
2.5
2.0
1.5
0
200
400
Alignment Position
Binding Site Size
1000
800
600
400
c
d
e
f
3
14
1.3
3

Figure 29. Natural variability and mutational tolerance analysis for putative ABSs in the H1N1 HA. Figure a shows the output from the bioinformatic structural variability pipeline for human H1N1 HA sequences mapped to the 6ona crystal structure. Figure b shows the output from the structural bioinformatic analysis using mutational tolerance data [241]. The x-axis is the numbering from a structural alignment between H1N1 and H5Nx. The x-axis denotes the position in the alignment which occupies the central position of the predicted ABS. The HA head and stem are labelled above. The y-axis denotes variability in the units of mean Hamming distance and bits respectively. The four lines on each graph denote epitopes of size 400 Å², 600 Å², 800 Å² and 1000 Å² by the line colours. Figures c and d show the variability and mutational tolerance analysis displayed on the 6ona trimer crystal structure. e and f show the top view of the crystal structure. The scale bar denotes the variability/ mutational tolerance.

11.2.4 I identified a putative low variability antibody binding site shared between H5N1 and H1N1

The output from the H5Nx and H1N1 natural variation structural variability analyses showed a high degree of similarity (Figures 28 and 29). The Pearson's correlation between the two outputs shown in Figure 30 a is 0.65 (0.58 to 0.71 95% confidence interval). This indicates that the variability of secondary and tertiary structural elements is consistent between H1N1 and H5Nx.

This method has previously been used by Thompson *et al.,* (2018) to identify the "OREO" ELV [147]. Due to structural similarities, I hypothesise that there may be a structural equivalent of this ELV in the H5N1 HA. The H1N1 ELV 'OREO' was identified by a structural variability dip around residues 157 to 159. Indeed, there is a remarkably similar dip at the same position in the H5Nx analysis (Figure 30 a).

Figures 30 b and c show the putative ABS drawn centred around H5Nx-H1N1 structural alignment position 159 at 600 Å in area. Both are in the same places on the trimer and look strikingly similar. This predicted ABS appears to potentially share sequence similarity between the H1 and H5 Hs.

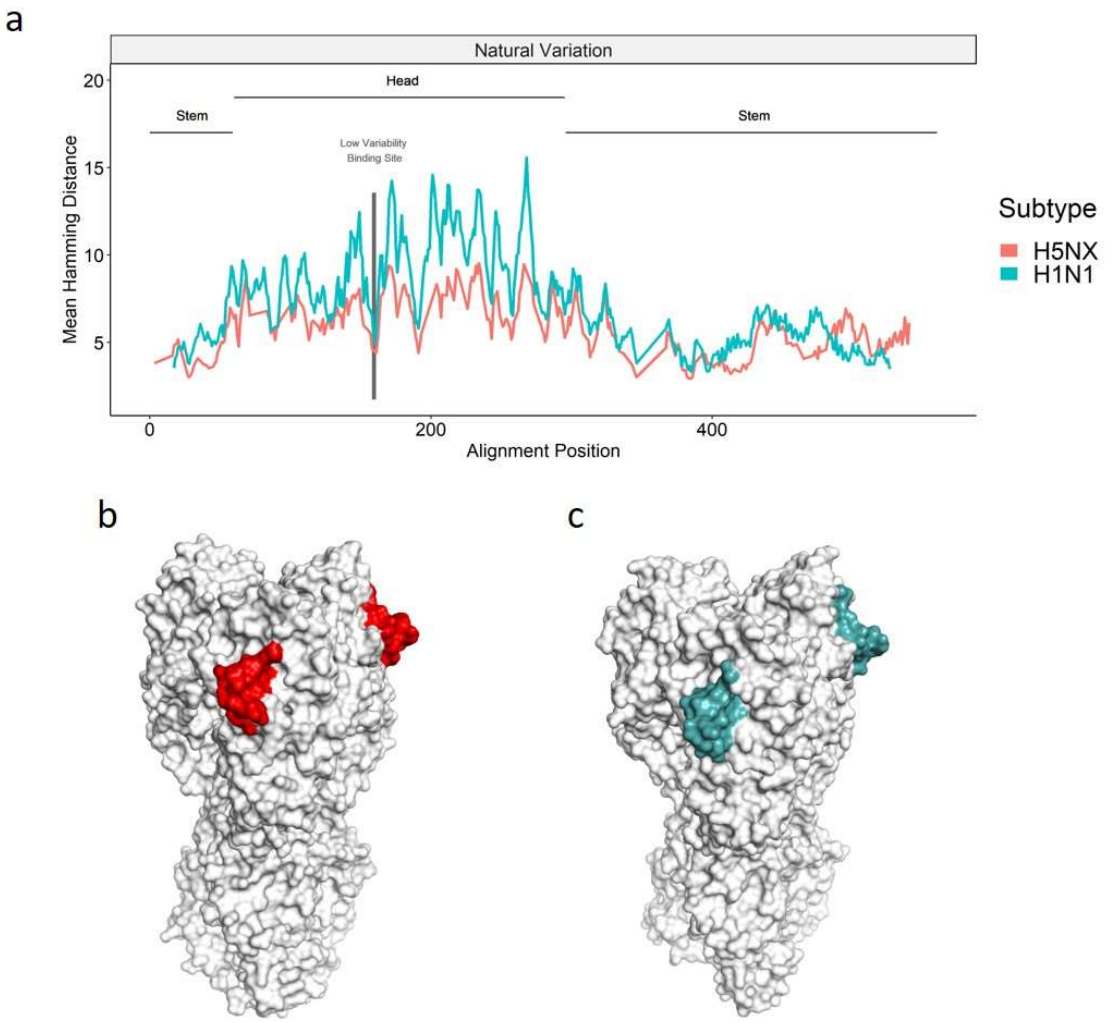


Figure 30. Variability analysis of the H5Nx and H1N1 HAs was used to identify a putative ELV potentially shared between the two subtypes. Figure a shows the output from the bioinformatic structural variability pipeline for avian H5 HA sequences mapped to the 5e2y crystal structure and human H1N1 HA sequences mapped to the 6ona crystal structure. The x-axis is the numbering from a structural alignment between H1N1 and H5Nx. The x-axis denotes the position in the alignment which occupies the central position of the predicted ABS. The HA head and stem are labelled above. The y-axis denotes the mean variability between all epitope sizes in the units of mean Hamming distance. Figures b and c show the putative ELV, of 600 Å in size, centred on structural alignment position 159 for H5Nx and H1N1 respectively.

11.2.5 The putative shared H5Nx-H1N1 ELV is situated in an important location in the HA receptor binding site and displays temporal and evolutionary characteristics of ELVs

During infection, the influenza virus uses the HA head domain to bind to sialic acid, allowing the virus to enter the host cell and replicate itself.. The mechanism of binding is dependent on conserved secondary structure elements: the 130 loop, the 220 loop and the 190 helix (Introduction 7.3). These, along with certain conserved amino acids, make up the HA receptor binding site (RBS). The H1N1 OREO ELV and the putative H5N1 equivalent (herein referred to as "the putative H5 ELV ABS", hereby referred to as H5PUTELV) both encompass part of the HA RBS (figure 31 a and b). Thompson *et al.,* (2018) showed that conformations of the OREO ELV are substantially influenced by the H1N1 residue 147 [147]. By changing the residue at position 147, particularly to amino acids of differing electrostatic properties, they were able to significantly impact the neutralisation of H1N1 viruses. The H5 structural equivalent of H1N1 residue 147 is 145. This amino acid is positioned in H5PUTELV to the right-hand side of the putative ABS, just as 147 is in OREO (Figures 31 a and b).

The electrostatic properties of amino acids can impact the secondary structure of proteins. Amino acid mutations that change amino acid properties can therefore alter regions of proteins [366]. For OREO and H5PUTELV, positions 147 and 145 are commonly made up of amino acids consisting of three distinct electrostatic properties. The majority of H1N1 residues at position 147 are either deleted, basic or polar (Figure 31 c). The majority of H5N1 residues at position 145 are either deleted, hydrophobic or polar (Figure 31 c).

Some evolutionary models predict that ELVs spread in a cyclical manner. This can be expressed by time and/ or evolutionary clade [98], [144], [145]. Laboratory evidence was gathered for this hypothesis by Thompson *et al.,* (2018) [147]. They showed that the spread of OREO confirmations stopped temporarily before resuming again. This pattern is shown by the time-calibrated H1N1 phylogeny of Figure 31 e detailing the electrostatic properties of residue 147. A phylogenetic clade of H1N1 spread during the 1940s that contained strains with a deletion at 147. Subsequently, these were replaced by basic residues of another clade in the 1950s. Deletions began spreading again in the early

2000s by a further different clade, which were then replaced again by basic residues following the 2009 “swine flu” pandemic. From 2022, it appears that the predominant clade contains polar amino acids.

The temporal and phylogenetic pattern of H5N1 residue 145 shows similarities with H1N1 residue 147 (Figure 31 d and e). From 1959 to the early 2000s, residue 145 predominantly expressed polar residues. During this time, multiple clades were formed. One contained hydrophobic residues and one contained polar residues which then changed to deletions around the 2010s. In the current year, most strains are now expressing hydrophobic residues.

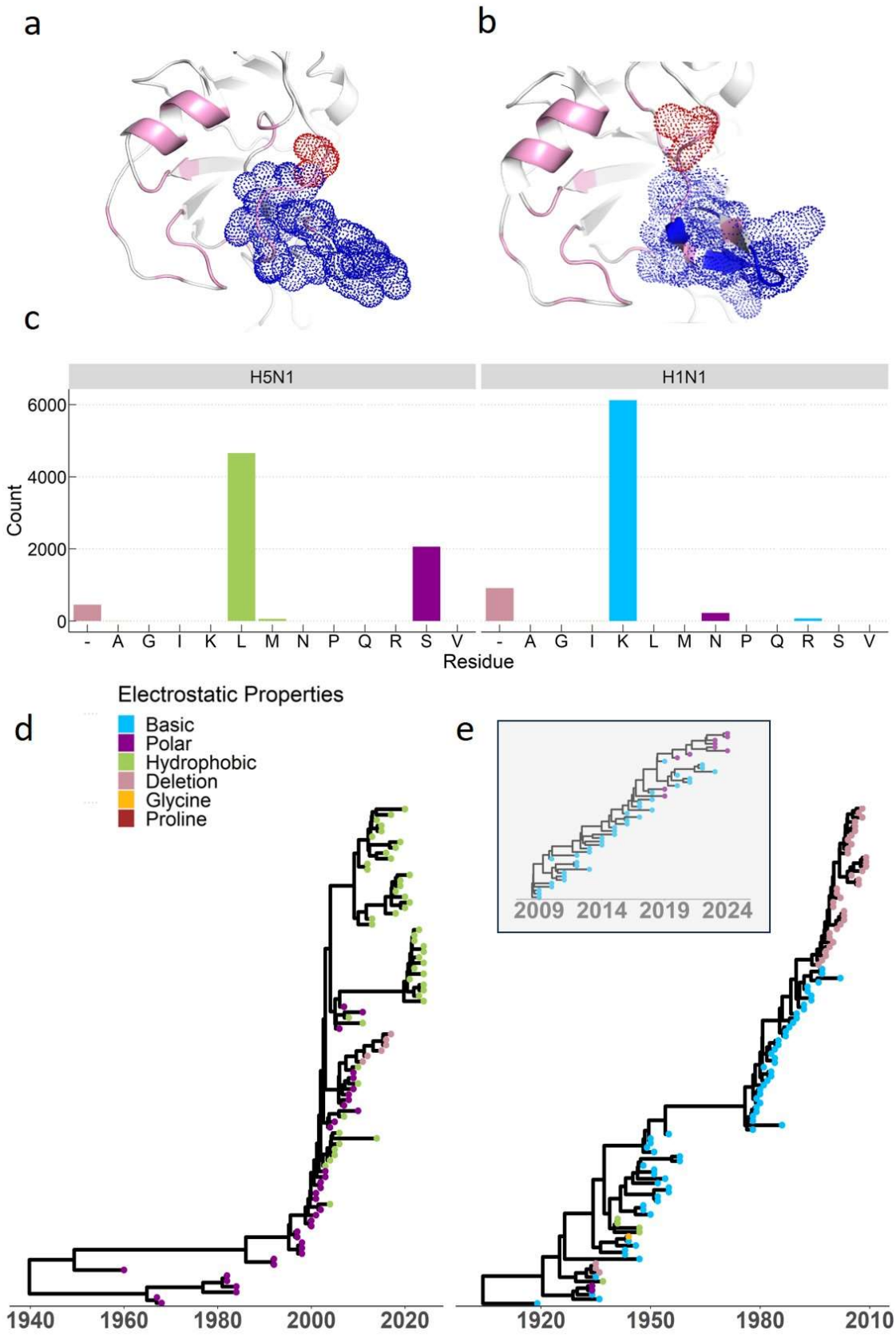

a
b
c
H5N1
H1N1
6000
4000
2000
0
Count
- A G I K L M N P Q R S V
- A G I K L M N P Q R S V
Residue
Electrostatic Properties
Basic
Polar
Hydrophobic
Deletion
Glycine
Proline
d
1940 1960 1980 2000 2020
e
2009 2014 2019 2024
1920 1950 1980 2010

Figure 31. The structural and phylogenetic characteristics of the putative H5 ELV and its structural equivalent in H1N1. Figures a and b show the putative ABS, of 600 Å in size, on the 5e2y (H5N1) and 6ona (H1N1) crystal structures, respectively. Cartoons coloured pink denote the HA receptor binding site (RBS). Blue dots denote the putative ELV. Red dots denote H5N1 residue 145 and H1N1 147 respectively. Figure c denotes the most common residues at H5N1 position 145 and H1N1 position 147. Count denotes the number of sequences in the amino acid database for H5N1 and H1N1 (Methods 8.4.1) and the x-axis denotes the specific amino acid. Facets denote the influenza subtype. Figures d and e show time-calibrated Maximum Clade Credibility (MCC) trees, with tip labels denoting the electrostatic properties of H5N1 residue 145 and H1N1 residue 147. Phylogenies were constructed by randomly sampling 3 sequences per year (where possible) from the H5N1 and H1N1 databases. Bar colours of c and tip label colours of d and e denote electrostatic properties.

11.2.6 Homology modelling identified that mutations at position 145 have structural impacts on an integral part of the HA receptor binding site

To test this in silico, I attempted to study the structural impacts of 145 mutations by homology modelling (Methods 8.2.3). Homology modelling is a technique used to predict the 3D structure of a theoretical protein by comparison to known, similar structures. The most common amino acids at H5N1 position 145 are serine (polar) and leucine (hydrophobic). I therefore generated structures based on H5N1 crystal structure 5e2y with 145 S and L mutations. I then structurally aligned these two structures and examined the differences. Between 145S and 145L, there is a difference in the orientation of the 130 binding loop (Figure 32 a).

I then calculated the Root Mean Square Deviation (RMSD) between every amino acid pair in the structural alignment. RMSD quantifies the difference between the x, y, and z coordinates of two objects. The greater the RMSD, the greater the difference in structure. Most amino acid pairs showed a RMSD of 0. 14 amino acids showed a non-zero RMSD. 8/14 of these are situated in the RBS, 3 of which are in the 130 loop, with the highest total RMSD in the 130 loop (Figure 32 b).

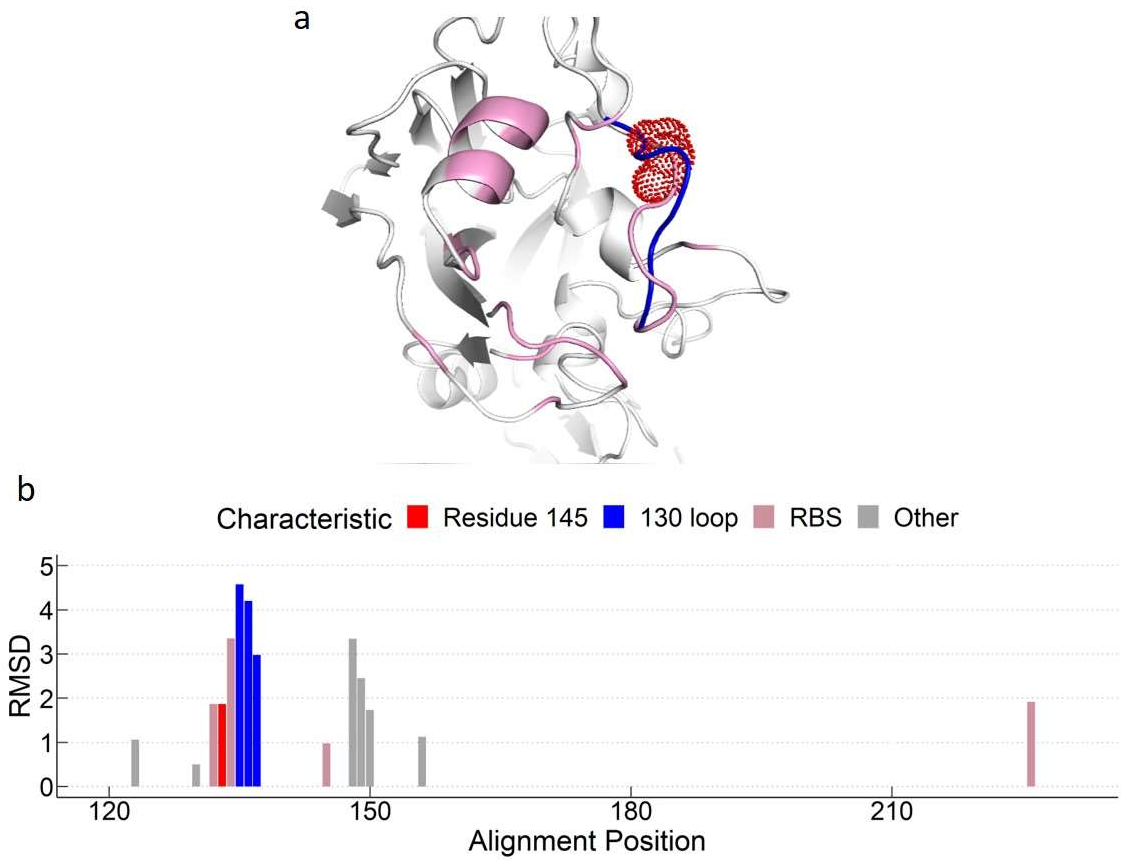


Figure 32. Theoretical impact of mutations in H5N1 residue 145 on HA's RBS. Figure a shows two structurally aligned homology-modelled HA structures, based on crystal structure 5e2y (H5N1). These structures contain 145S and 145L mutations. Cartoons coloured pink comprise HA's RBS. Cartoons coloured blue show a peptide chain that differs in orientation between the two aligned structures. Figure b shows the pairwise Root Mean Square Deviation (RMSD) between every amino acid pair. Bars denote the RMSD between pairs and alignment position denotes position in the structural alignment. Figure b zooms in from residue 120 to 220 as all other positions show a pairwise RMSD of 0.

11.3. Discussion

11.3.1 Summary

My analyses show that the globular head domain of the H5 and H1 subtype HA is more variable than the stem (Figure 27).

I applied a structural bioinformatic pipeline to characterise the variability of putative ABSs on the HAs of H5N1 and H1N1. This was done using a range of ABS sizes and solvent-accessibility thresholds (Figures 28 to 30 and Supplementary Figures 7 to 10). Using my amino acid alignment database, I showed that variability is not uniform across the structure. There exist pronounced peaks and troughs in variability throughout the head domain. Some putative ABSs in the head domain are, in fact, of lower variability than those in the stem domain.

In addition to the natural variation, I applied mutational tolerance data to the structural bioinformatic pipeline. Bloom *et al.,* (2016) determined the capacity of HA to vary by generating every mutation possible in the H1N1 HA [241]. I used this data to determine the theoretical mutational tolerance of the putative ABSs generated earlier (Figure 29). These data were not highly concordant with the natural variation analysis applied to H1N1. The distribution of mutational tolerance showed a near-opposite pattern across the amino acid alignment in the head domain.

By examining the natural variation analysis output for H5N1 and H1N1, I identified a low variability theoretical ABSs that could be shared between the H5 and H1 subtypes (Figure 30), hereby referred to as 'H5PUTELV'. The putative H5 low variability ABS appears to be structurally similar to the H1N1 ELV OREO (Figure 31 a and b). Residue 145 in H5PUTELV is the structural analogue of OREO residue 147. This residue was shown to have an interesting temporal pattern of spread. I showed that a similar phylogenetic and temporal pattern is exhibited by residue 145 found within the low variability H5 HA ABS (Figure 31 e and f).

ELVs have been found in structurally important parts of the HA's globular head domain [84], [149], [363]. H5PUTELV is situated on top of the HA's sialic acid RBS. I, therefore, attempted to determine whether mutations in H5PUTELV residue 145 would structurally impact the RBS. Through homology modelling and structural alignment, I showed that mutations theoretically alter the orientation of the RBS 130 loop (Figure 32).

11.3.2 Development of a theoretical "universal influenza vaccine"

The primary antigenic determinant of influenza is the globular head of HA. Therefore, it is the primary target for vaccines [39]. To date, alternative methodologies, such as vaccines targeting the more conserved HA stem or conserved internal proteins such as NP and M1 have failed to show efficacy in clinical trials [355]. However, the disadvantage of targeting the head domain is that epitopes within it can be more variable than those in the stem. The analysis in Figures 28 to 30 shows that some putative head epitopes have lower variabilities than some stem epitopes in the HAs of H1N1 and H5N1. These putative epitopes, such as H5PUTELV in H5N1 and OREO in H1N1 [147], if shared between subtypes, could be ideal vaccine targets for an H1/H5 subtype influenza vaccine.

11.3.3 Hypothesis Generating data

The rationale for this bioinformatic analysis was to produce hypothesis-generating data. I wanted to attempt to understand why a large proportion of a likely-unexposed Scottish blood donor sera cohort neutralised the A/bar-headed goose/Qinghai/3/2005 H5N1 HA. I hypothesised that this occurs due to cross-reactive immune responses generated by exposure to H1N1 HAs. If this is the case, I hypothesised that H5N1 and H1N1 HAs share conserved epitopes on the globular head. I planned to test this by undertaking site-directed-mutagenesis (SDM) on the residues identified by bioinformatics. These tests would be designed to explain whether the predicted residues: i) explain a significant

portion of the neutralisation strength towards the A/bar-headed goose/Qinghai/3/2005 HA, ii) are shared between H1N1 and H5 strains, and iii) the impact of mutating these residues in order to characterise and isolate the predicted ELV.

11.3.4 The H5 residue 145 is antigenically important and has been associated with HPAI epidemics as well as human spillover risk

I identified H5PUTELV as the structural analogue of H1N1's OREO ELV (Figure 31). Thompson *et al.,* (2018) identified H1N1 residue 147 as integral to the OREO epitope [147]. I identified the H5 structural analogue of H1N1 residue 147 to be residue 145 (Figure 31 a and b).

H5 residue 145 is situated within the H5 antigenic site "Site A", also known as "Group 1" (Figure 5) [113], [114], [118]. Antigenic sites have been identified in epitope mapping studies. These studies test a group of mAbs with different, unknown, binding sites against viruses with known mutations. By determining which mutations impact the binding of which mAbs, and then grouping these together, researchers can define antigenic sites. Rockman *et al.,* (2013) conducted fine epitope mapping, progressively introducing amino acid substitutions to A/Vietnam/1194/04 until it became A/Bar-headed Goose/Qinghai/65/05 [367]. Mutating A/Vietnam/1194/04 to L145S caused significant immune escape to the mAb 153b. L145S has been shown to cause immune escape against chicken sera [114]. Moreover, viruses containing the mutations L145 S/Del were inferred to elicit immune escape when assayed against ferret sera [118]. These studies suggest that mutations at residue 145 can elicit immune escape against a variety of antibodies. This makes 145 an ideal candidate for SDM.

Rockman *et al.,* (2013) performed homology modelling and molecular dynamics simulations involving residue 145 [367]. In this in silico analysis, they predict L145S changes the hydrogen bonding characteristics and electrostatic interactions in adjacent areas. These in turn cause conformational changes in secondary structures which may alter antigenicity by impacting antibody binding. My in-silico analysis also predicts that S145L can impact the secondary structure of H5N1 HA. I predict that

the transition from hydrophobic to polar residues causes the RBS 130 loop to change in orientation (Figure 32).

In addition to immune escape, mutations at residue 145 have been linked with HPAI epidemics. Residue 145 has been shown to be under directional selection during H5 outbreaks [368], and associated with the generation of HPAI outbreaks [369]. This suggests that mutations at this site improve viral fitness and increase the longevity of outbreaks. This could occur due to 145 mutations allowing immune escape in populations becoming immune to the previous residue. Mutations at residue 145 have also been shown to be associated with increased virulence in avian outbreaks [370].

In general, human influenza viruses HAs preferentially bind to α2,6-linked sialic acid situated in the upper respiratory tract. Avian influenza viruses such as H5N1 bind preferentially to α2,3-linked sialic acid and show poor avidity for α2,6-linked sialic acid [371][372]. This means avian influenza viruses have difficulty infecting human hosts. Shifting in preference from α2,3 to α2,6-linked sialic acid causes spillover concern [372].

The double mutation 129Del/I151T (residue 129 is residue 145 in H5 numbering) is associated with increased HA avidity to α2,6-linked sialic acid [373] [374]. The double mutation also showed increased murine virulence [374]. The 145Del mutation has occurred commonly in Egypt [375]. The triple mutation 120, 129Del and I151T (H5 numbering) has been found in many Egyptian cases, with 145S also common [376]. Strains containing 145Del were shown to be statistically over-represented in fatal Egyptian human H5N1 cases [377].

11.3.5 Temporal and phylogenetic signals of ELV existence

Mathematical models of influenza evolution containing low variability epitopes are becoming increasingly common [355]. The model known as “antigenic thrift” stipulates immunity to epitopes of limited variability drives the emergence of new strains as population immunity changes due to birth

and death within the population [144]–[146]. Due to the length of the human lifespan, humans are infected with multiple strains of influenza throughout their lives. A large part of the neutralising immunity to influenza could be directed against a small number of structurally constrained epitopes. If true, due to structural constraints, these epitopes could only have a small number of conformations they could occupy.

For example, conformation A of an ELV may spread for a decade, and people then become increasingly immune to conformation A until a level of population immunity is reached that makes conformation A not evolutionarily favourable to spread, at which point strains with conformation B would spread, Introduction 7.11 (Figure 7). This allows for the reappearance of historical epitopes (conformation A) once immunity against them has decreased due to immune people dying or migrating. A scenario similar to this occurs within H1N1, with immunity directed towards the OREO ELV [147]. This pattern appears to be reflected by the time-calibrated phylogeny shown in Figure 31 f which describes the H1N1 OREO residue 147.

Birds have shorter lifespans than humans [378]. This means that birds are infected with fewer strains throughout their lives than humans. This is predicted to have an impact on the spread of limited-conformation ELVs [145]. Birds may not live long enough to become immune to an ELV conformation and in doing so, cause switching of this conformation. Instead, this pattern may be seen via phylogenetic clades. Figure 31 e shows that electrostatic charges of H5 residue 145 appear to be grouped phylogenetically and temporally. This could be evidence that H5 residue 145 may spread in an analogous manner to the OREO-containing H1N1 residue 147.

11.3.6 Contrast between natural variability and mutational tolerance ABS analysis in H1N1

I captured putative ABS variability that occurs in nature through our structural bioinformatic pipeline. I did this by using a large H1N1 amino acid alignment generated from laboratory-sampled influenza viruses. I also captured the mutational tolerance of putative ABSs by feeding mutational preference

data from [241] through the bioinformatics pipeline. I hypothesised that the two analyses would show similar patterns. In fact, the patterns did not correlate well, with a Pearson's correlation of just 0.07 in the head domain. This may suggest that, even though certain putative ABSs have a high mutational tolerance, they do not vary much in nature. This could potentially happen due to structural constraints imposed by surrounding residues.

12. Identification of epitopes of limited variability in the head domains of H5 and H1 influenza virus HAs using site-directed mutagenesis

12.1. Background

Site-directed mutagenesis (SDM) is a precise genetic engineering technique used to selectively modify DNA sequences to change one amino acid into another. Consequently, SDM enables precise modification of amino acids within a protein [379]. In the context of this chapter, SDM allows us to determine the impact of mutations on the antigenicity of HA. By observing how mutations modify the neutralising immune responses observed in Chapter 10, I can determine whether certain residues form parts of an epitope and responsible for an observed immune response.

In Chapter 10, I identified neutralising immune responses towards A/bar-headed goose/Qinghai/3/2005/H5N1 and multiple historical H1N1 strains using a pseudotyped influenza virus microneutralisation assay and sera from blood donors taken in 2020.

The epitopes recognised could be located in the more variable head domain of HA or the conserved stem. Cross-reactive responses between different subtypes are typically thought to be mediated by stem-targeted antibodies [60], [77], [86]–[88]. However, based on the bioinformatic analysis of the H1 and H5 haemagglutinins in Chapter 11, it appears that cross-reactive epitopes could also exist in the head domain of the H1 and H5 HAs.

Furthermore, the observed antibody response could be (1) polyclonal, in which multiple epitopes are recognised, (2) oligoclonal, in which several epitopes are recognised, or (3) monoclonal, in which a single epitope is recognised by a single class of antibody [380].

In this chapter, I therefore mutated residues 145 in the H5 HA of A/bar-headed goose/Qinghai/3/2005/H5N1, and its H1 HA counterpart, 147 in A/USSR/90/1977/H1N1 to determine if the observed cross-reactivity was due to a shared epitope containing this residue. I used two control antibodies targeting the stem of the H5 and H1 HAs displayed by the pseudotyped viruses.

I show that mutating H5 residue 145 significantly impacts human and avian neutralising immune responses towards A/bar-headed goose/Qinghai/3/2005/H5N1. I show that mutating H5 145's structural analogue, H1 residue 147, significantly impacts the human immune response towards A/USSR/90/1977/H1N1. These results are in contrast to control mutations in other H5 residues which do not confer a significant change in immune response.

12.2. Results

12.2.1 Mutating the head domain of A/bar-headed goose/Qinghai/3/2005 HA impacts neutralisation with monoclonal antibodies

I performed site-directed mutagenesis on residue 145 in the HA of A/bar-headed goose/Qinghai/3/2005. Residue 145 (Figures 33 a, d, and e) was selected as it is the structural analogue of the H1 residue 147. Residue 147 in the H1 HA has been previously shown to be important in the H1 ELV OREO, mediating cross-reactivity to historical strains [147]. Although the residue does vary, it is situated within the putative low-variability H5 ELV (H5PUTELV Results 11.2.5). Mutating it allows us to determine whether the previously observed reactivity to the A/bar-headed goose/Qinghai/3/2005 HA pseudotyped influenza virus HA in (i) the Scottish blood donor sera obtained in 2020, and (ii) the swan sera obtained from the Abbott Swannery between 2008 and 2017 is reduced.

Using this information, I also identified residue 215 (Figures 33 c, f and g) for SDM. Residue 215 is a structural analogue of an H1 residue adjacent to, but not in, the H1N1 epitope Sb. I also selected other H5 residues that do not have H1 analogues situated in defined H1N1 epitopes. The rationale for this was twofold: (i) to test whether the SDM technique would cause statistically significant differences in many residues, which is a-priori unlikely and may suggest the generation of false positives, and (ii) to determine whether any anti-HA head cross-reactivity is widespread or confined to a limited number of residues, as is predicted by mathematical models [144]–[146]. To this aim, I selected H5 residues: 70, 99, 123, 279 and 281 (Figure 33) to act as residues for negative control mutations.

Using PCR-based techniques, I introduced non-synonymous amino acid point mutations into the head domain of the A/bar-headed goose/Qinghai/3/2005/H5N1 HA to test the predictions outlined above. I then displayed these HAs on pseudotyped influenza viruses using the technique described in Methods 8.1.3. Mutations that appear in the amino acid database (Methods 8.4.1) were selected; all mutations have appeared in nature. I designed mutations to change the electrostatic properties of the residue in question.

When exposed to a polyclonal antibody sample, such as blood sera extracted from humans, antibodies are likely to bind to multiple epitopes on the pseudotyped virus surface. I wanted to determine if mutations in the head domain impact the binding of antibodies to these potential epitopes. Any differences between WT and mutant would likely be due to structural changes caused by mutations impacting the shape of epitopes in question. However, mutations may change more than the conformation of head epitopes. Mutations could impact: (i) the amount of HA expressed in the pseudotype, and/ or (ii) the distribution of HA in the pseudotype [222]**.** These changes could impact how mutant pseudotyped influenza viruses are neutralised by mechanisms not related to changes in surface epitope conformation.

Monoclonal antibodies (mAbs) are antibodies that recognise a single epitope. The mAbs Cr6261 and C179 both recognise epitopes on the conserved stem domain of HA [267]–[269]. C179 neutralises the same epitope in H1, H2, H5, and H6 HAs [267], [268]. Cr6261 neutralises the same epitope in and protects mice from H1N1 and H5N1 viruses [269].

All the mutations I introduced target the head domain of HA. The stem domains are unchanged between mutants. By testing these pseudotyped influenza viruses with stem-targeted mAbs, I can determine if the mutations have impacted the pseudotype's neutralisation characteristics irrespective of head epitope structural changes. This is crucial for later work as this data can be used to normalise differences in human and avian sera responses and allow for accurate comparison of mutants. I therefore, tested the WT A/bar-headed goose/Qinghai/3/2005/H5N1 HA expressing pseudotyped influenza virus against both mAbs. I then repeated this with every mutant HA expressing pseudotyped virus (Figure 33).

All mutants were neutralised by both mAbs (Figure 33). No mutants apart from 145k showed statistically different neutralisation characteristics than the WT as determined by 95% confidence intervals. This occurred for Cr6261 mAb but not C179.

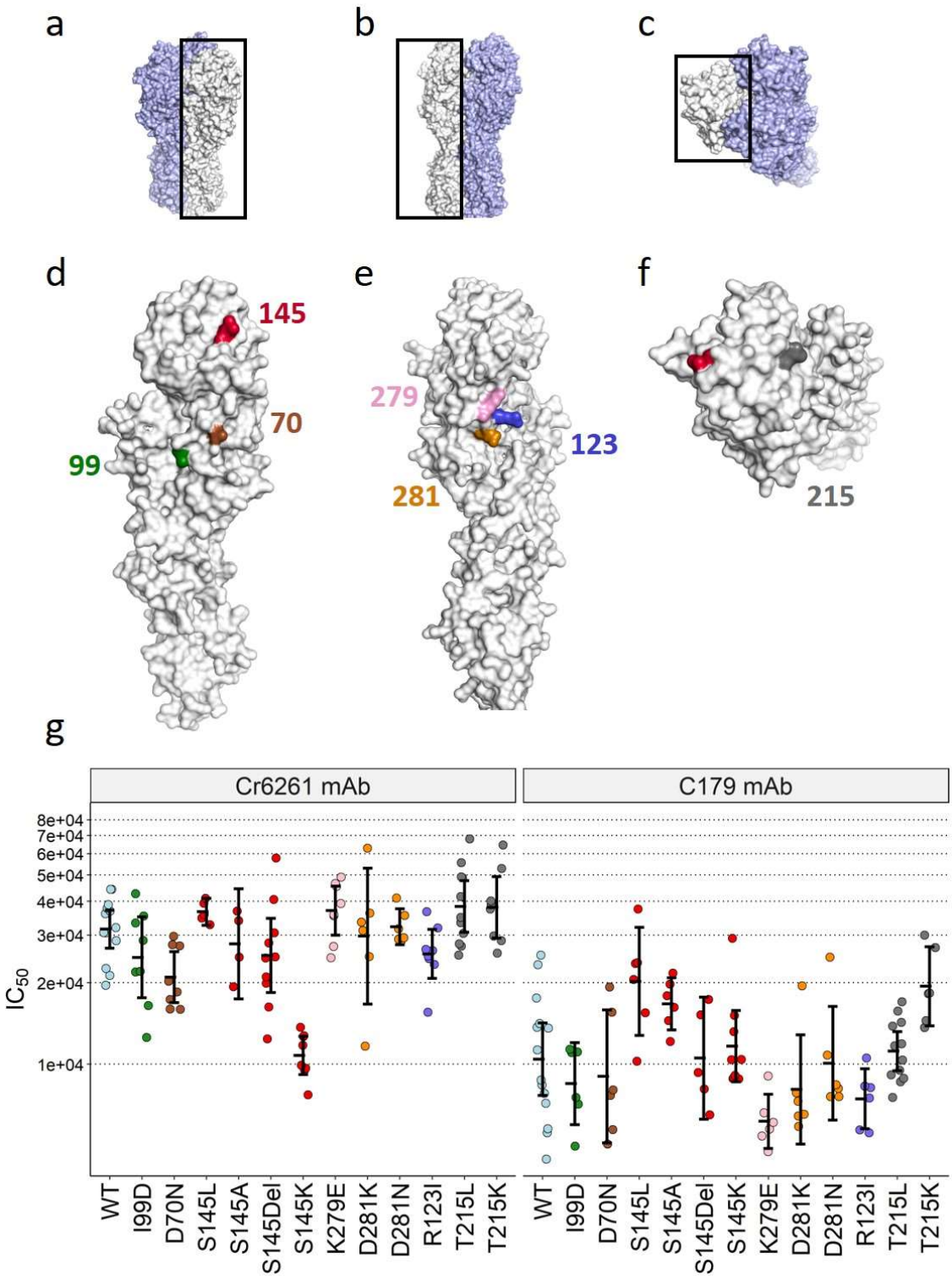

a
b
c
d
145
70
99
e
279
123
281
f
215
g
Cr6261 mAb
C179 mAb
IC50
8e+04
7e+04
6e+04
5e+04
4e+04
3e+04
2e+04
1e+04
WT
I99D
D70N
S145L
S145A
S145Del
S145K
K279E
D281K
D281N
R123I
T215L
T215K

Figure 33. mAb characterisation of pseudotyped influenza viruses displaying the HA of WT and mutant A/bar-headed goose/Qinghai/3/2005/H5N1. Figures a to c show the zoomed-out view of the monomeric modelled H5N1 HA crystal structure 5e2y. Figure c shows the top view of the structure. Chain A is coloured white. The black rectangle highlights the zoomed-in view displayed in Figures d to f. Figures d to f show the monomeric form of the crystal structure from Figures a to c. Residues coloured and labelled are those that were mutated in mutant A/bar-headed goose/Qinghai/3/2005/H5N1. Figure g shows the characterisation of pseudotypes displaying the WT and mutant A/bar-headed goose/Qinghai/3/2005/H5N1 HAs. The y-axis denotes IC_{50} on a $\log_{10}$ scale. The x-axis denotes the WT or mutant pseudotype. Facets denote the mAb. Error bars denote the 95% confidence interval. Horizontal dashed lines show the mean IC_{50}. Statistics were calculated using $\log_{10}(IC_{50})$ values and then back-transformed.

12.2.2 Mutating the head domain of A/USSR/90/1977/H1N1 HA impacts neutralisation with monoclonal antibodies

To determine whether mutating residue 147 in H1N1 would significantly impact the neutralisation response, Thompson *et al.,* (2018) demonstrated that a deleted residue at position 147 was required for cross-reactivity to historical strains [147]. As H5 does not circulate in the human population, the observed reactivity to H5 is likely due to the circulation of H1 in the human population. Consequently, if any reduction in neutralisation occurs when residue 145 is mutated, this is likely due to a pan-subtype ELV. The bioinformatic analysis suggested that mutating the H5 structural analogue of H1 147 (H5 145) may significantly impact the immune response (Chapter 11). If mutations of both these residues significantly impact the neutralisation response with the same sera samples, there is evidence for a shared, pan-subtype, ELV. I therefore mutated H1 residue 147 to amino acids seen in nature (Figure 34).

I also wanted to determine how polyclonal the neutralisation towards A/USSR/90/1977/H1N1 HA is. I, therefore, introduced amino acid point mutations into residues not part of the classic H1 head epitopes (Figure 5) defined in [106], [107]. These residues are 63, 101, 103, 189 and 252 (Figure 34 c and d).

As with H5, I produced pseudotyped influenza viruses containing these mutated HAs. I then tested the WT and mutant pseudotyped influenza viruses using the stem-targeting mAbs Cr6261 and C179. No significant differences were detected between the neutralisation of WT to mutants apart from

K147Del as determined by a 95% confidence interval. A significant difference occurred with the Cr6261 mAb.

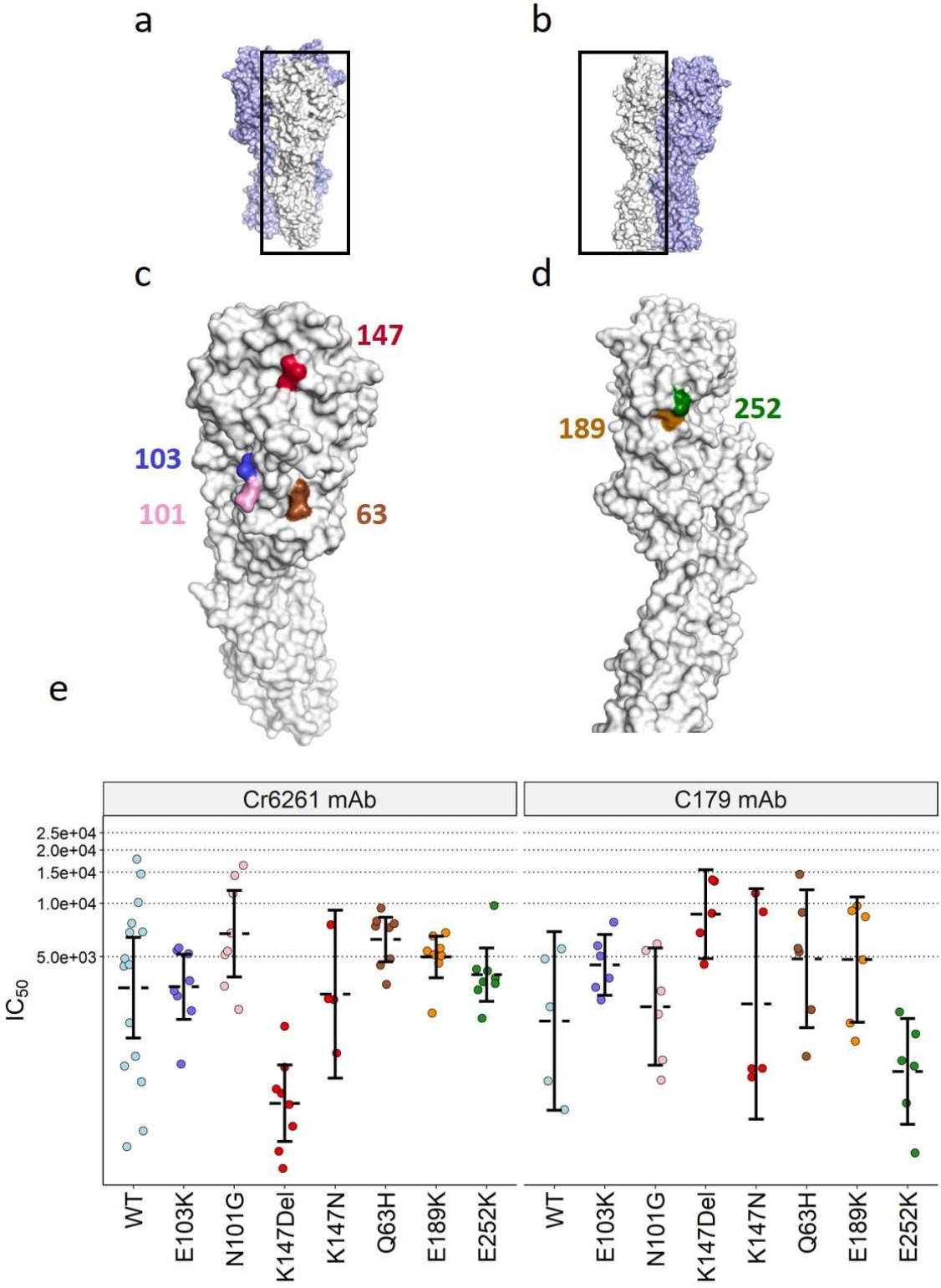


Figure 34. mAb characterisation of pseudotyped influenza viruses displaying the HA of WT and mutant A/USSR/90/1977/H1N1. Figures a to b show the zoomed-out view of the trimeric H1N1 HA crystal structure 6ona. Chain A is coloured white. The black rectangle highlights the zoomed-in view displayed in Figures c to d. Figures c to d show the monomeric form of the crystal structure 6ona from Figures a and b. Residues coloured and labelled are those that were mutated in mutant A/USSR/90/1977/H1N1. Figure e shows the characterisation of pseudotypes displaying the WT and mutant A/USSR/90/1977/H1N1 HAs. The y-axis denotes IC_{50} on a $\log_{10}$ scale. The x-axis denotes the WT or mutant pseudotype. Facets denote the mAb. Error bars denote the 95% confidence interval. Horizontal dashed lines show the mean IC_{50}. Statistics were calculated using $\log_{10}(IC_{50})$ values and then back-transformed.

12.2.3 Mutations to key residues in the head domain of the A/bar-headed goose/Qinghai/3/2005 HA impact the neutralising action of human sera

I defined Scottish sera samples that were positive for neutralisation against A/bar-headed goose/Qinghai/3/2005/H5N1 in Results 10.2.1. I selected an age-representative random sample of 30 of these sera samples. I then tested these sera samples for neutralising immune responses against pseudotyped influenza viruses displaying the WT and mutant HAs of A/bar-headed goose/Qinghai/3/2005/H5N1, (Figure 35).

Mutants differed in neutralising immune responses compared to WT, but none significantly, as determined by 95% confidence interval, (Figure 35). All mutants apart from D281K showed neutralisation responses for some samples below the negative threshold defined in Results 10.2.1.

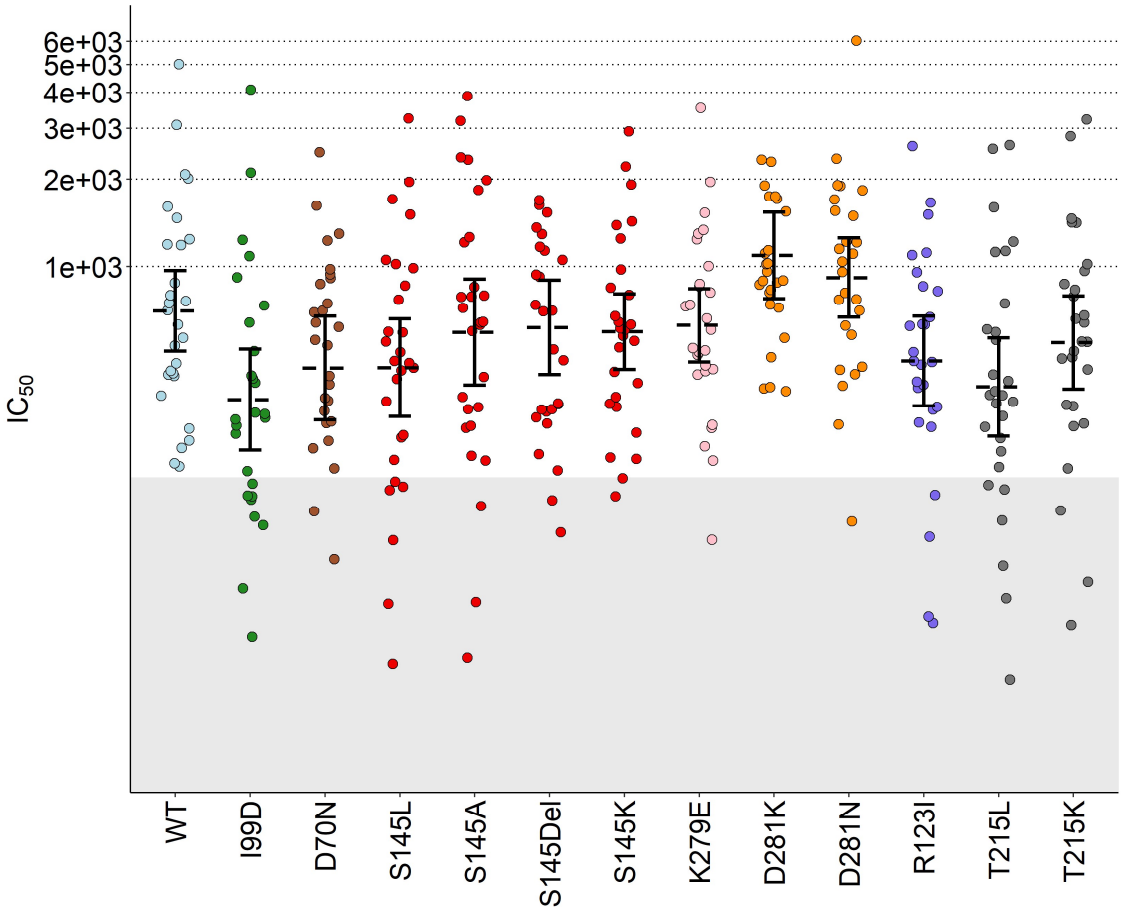


Figure 35. Human neutralising immune responses towards WT and mutant A/bar-headed goose/Qinghai/3/2005/H5N1 HA pseudotypes. The y-axis denotes IC_{50} on a log scale. Error bars denote the 95% confidence interval. Statistics were calculated using $\log_{10}(IC_{50})$ values and then back-transformed. Horizontal dashed lines show the mean IC_{50}. The grey shaded area denotes the negative threshold defined in Results 10.2.1.

12.2.4 Mutations to key residues in the head domain of the A/bar-headed goose/Qinghai/3/2005 HA impact the neutralising action of swan sera

I tested a randomised, hatch-date representative, 30-strong swan sera sample set for neutralising immune responses against pseudotyped influenza viruses displaying the WT and mutant HAs of A/bar-headed goose/Qinghai/3/2005/H5N1 (Figure 36).

Mutants differed significantly in neutralising immune responses compared to W*T*. The mutants D70N and S145A showed significantly different neutralising responses to WT as determined by a 95%

confidence interval, Figure 36. No mutants showed neutralisation responses below the negative threshold defined in Results10.2.1.

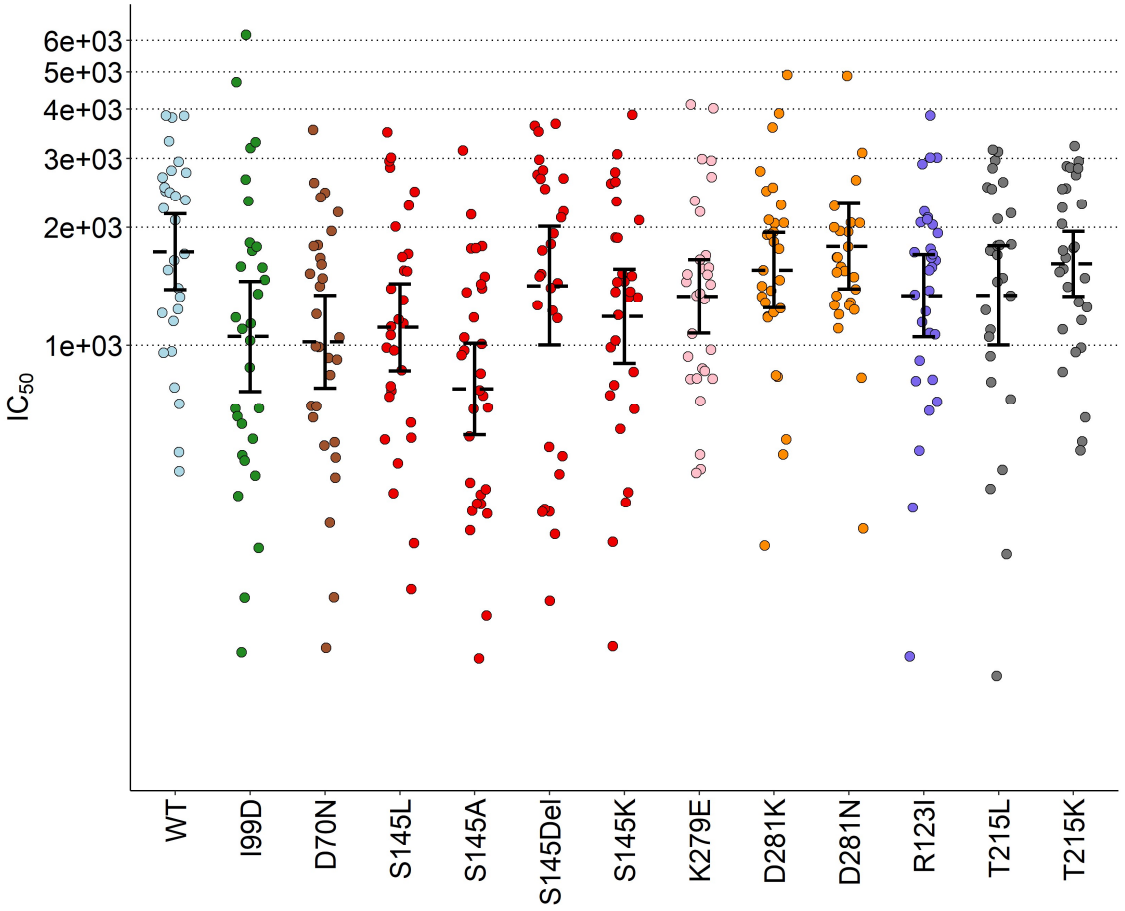


Figure 36. Avian neutralising immune responses towards WT and mutant A/bar-headed goose/Qinghai/3/2005/H5N1 HA pseudotypes. The y-axis denotes IC_{50} on a log scale. Error bars denote the 95% confidence interval. Horizontal dashed lines show the mean IC_{50}. Statistics were calculated using $\log_{10}(IC_{50})$ values and then back-transformed. No samples recorded IC_{50} values below the negative threshold defined in Results 10.2.1 so this is not displayed.

12.2.5 Mutations to key residues in the head domain of the A/USSR/90/1977/H1N1 HA impact the neutralising action of human sera

I used the same age-representative random sample of 30 Scottish sera samples tested against A/bar-headed goose/Qinghai/3/2005/H5N1 mutants to test for neutralising immune responses against pseudotyped influenza viruses displaying the WT and mutant HAs of A/USSR/90/1977/H1N1, Figure 37.

Mutants differed in neutralising immune responses compared to WT. All differed significantly apart from E189K as determined by a 95% confidence interval (Figure 37). All mutants apart from E252K showed neutralisation responses for some samples below the negative threshold defined in Results 10.2.1.

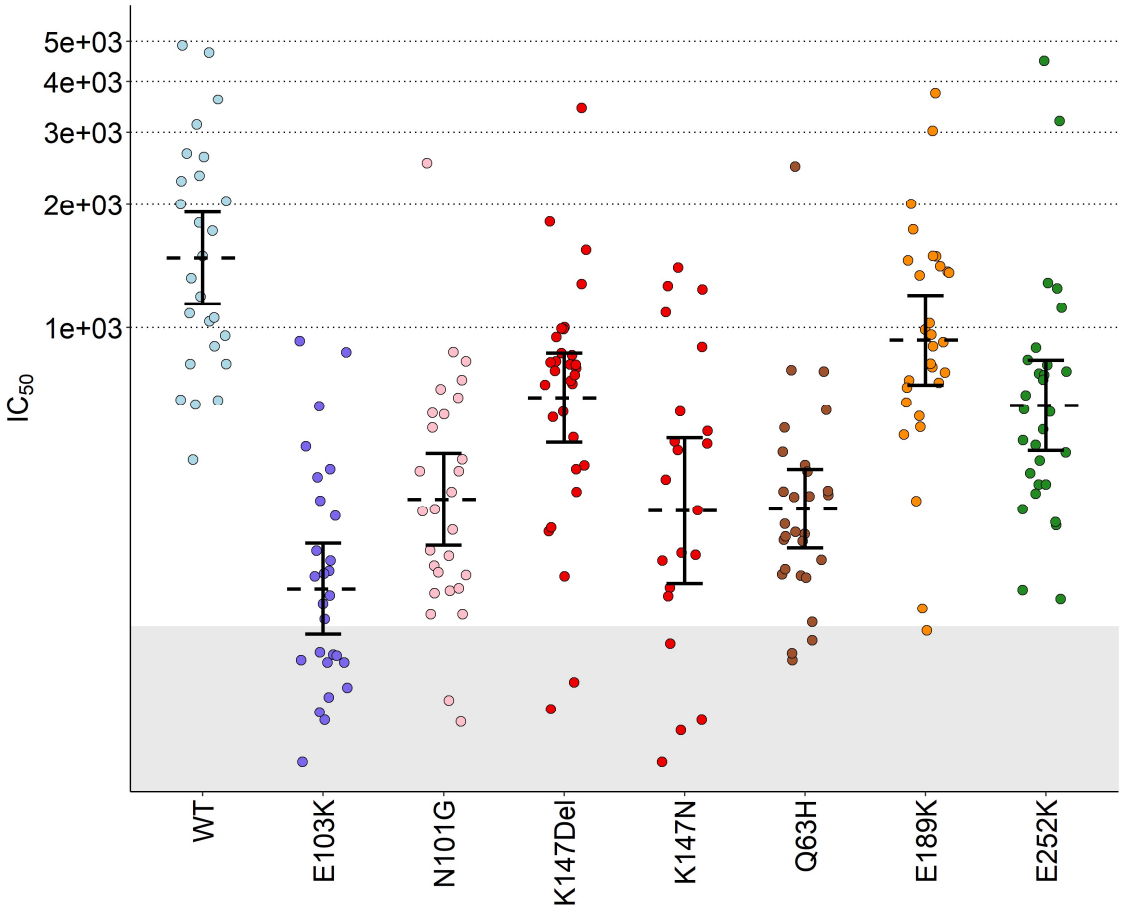


Figure 37. Human neutralising immune responses towards WT and mutant A/USSR/90/1977/H1N1 HA pseudotypes. The y-axis denotes IC_{50} on a log scale. Error bars denote the 95% confidence interval. Horizontal dashed lines show the mean IC_{50}. Statistics were calculated using $\log_{10}(IC_{50})$ values and then back-transformed. The grey shaded area denotes the negative threshold defined in Results 10.2.1.

12.2.6 Mutations to key residues in the head domain of the A/bar-headed goose/Qinghai/3/2005 HA confer a significant change to the human and avian neutralisation response

Amino acid mutations introduced to the head domain of HAs could alter: (i) the amount of HA expressed in the pseudotype, and/ or (ii) the distribution of HA in the pseudotype [222]. This could impact the neutralising immune response of sera samples when exposed to pseudotyped influenza virus mutant HAs. This study aims to assess the impact of HA head mutations on antibody binding to HA head epitopes. Therefore, I conducted analyses that would control for confounding variation that could be introduced by non-head epitope-related changes in mutant pseudotype influenza viruses.

I calculated the percentage change in neutralisation between *WT* HA and mutant A/bar-headed goose/Qinghai/3/2005/H5N1 HA pseudotyped influenza viruses through the equation $\left(\left(\frac{WT\ \mathrm{IC}_{50}}{\mathrm{Mutant\ IC}_{50}}\right)-1\right)\times 100$ [147]. I did this for both the stem-targeted mAb and sera sample experiments. This analysis was conducted for $\log_{10}$ transformed IC_{50} values to eliminate bias from skewed distributions, misleading mean values, and unfulfilled statistical assumptions. I then compared the changes in mAb to the changes in sera sample neutralisation responses with student's t-tests. A significant difference between mAb and sera samples infers a difference in antibody-head epitope binding caused by the mutation.

I did this for both the Scottish blood donor sera samples and wild-avian sera samples. Significant differences were seen between both stem-targeted mAbs and human sera for the mutations: I99D, S145L, T215L, and T215K (Figure 38). No significant differences were seen between both mAbs and human sera for mutations: D70N, S145Del, K279E, D281K, D281N, and R123I (Figure 38). In contrast, there was a significant difference between stem-targeted mAb Cr6261 and human sera for S145A (Figure 38) as well as a significant difference between human sera and mAb Cr6261 for S145K (Figure 38).

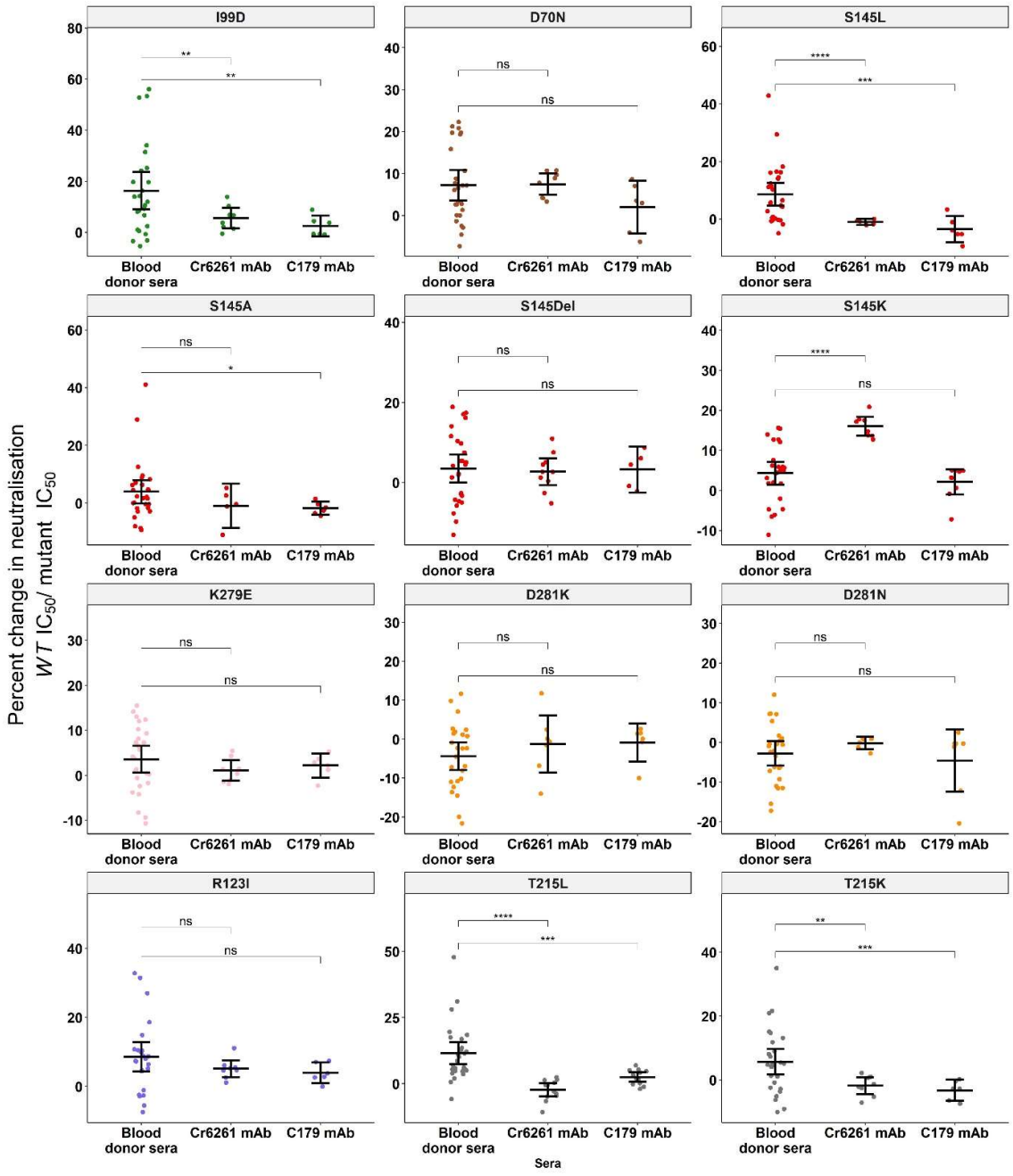


Figure 38. Differences in the human neutralising immune response against mutant and WT A/bar-headed goose/Qinghai/3/2005/H5N1 HA pseudotyped influenza viruses. The y axis denotes the percentage change in neutralisation between *WT* HA and mutant A/bar-headed goose/Qinghai/3/2005/H5N1 HA pseudotyped influenza. The x-axis denotes sera or mAb. Facets denote different mutations. Error bars denote the 95% confidence interval. Statistics were calculated using $\log_{10}(IC_{50})$ values. Differences between means of sera and stem-targeted mAbs were determined by student's t-test. Significance stars: * p-value < 0.05, ** p-value < 0.005, *** p-value < 0.0005 and **** p-value <0.00005.

Significant differences were seen between both stem-targeted mAbs and avian sera for the mutations S145L and S145A (Figure 39). No significant differences were seen between both mAbs and avian sera for mutations: D70N, S145Del, K279E, D281K, D281N, and R123I (Figure 39). There was a weak significant difference between stem-targeted mAb C179 and avian sera for I99D and T125K, and a stronger significant difference for T215L (Figure 39). There was a significant difference between mAb Cr6261 and avian sera for S145K (Figure 39).

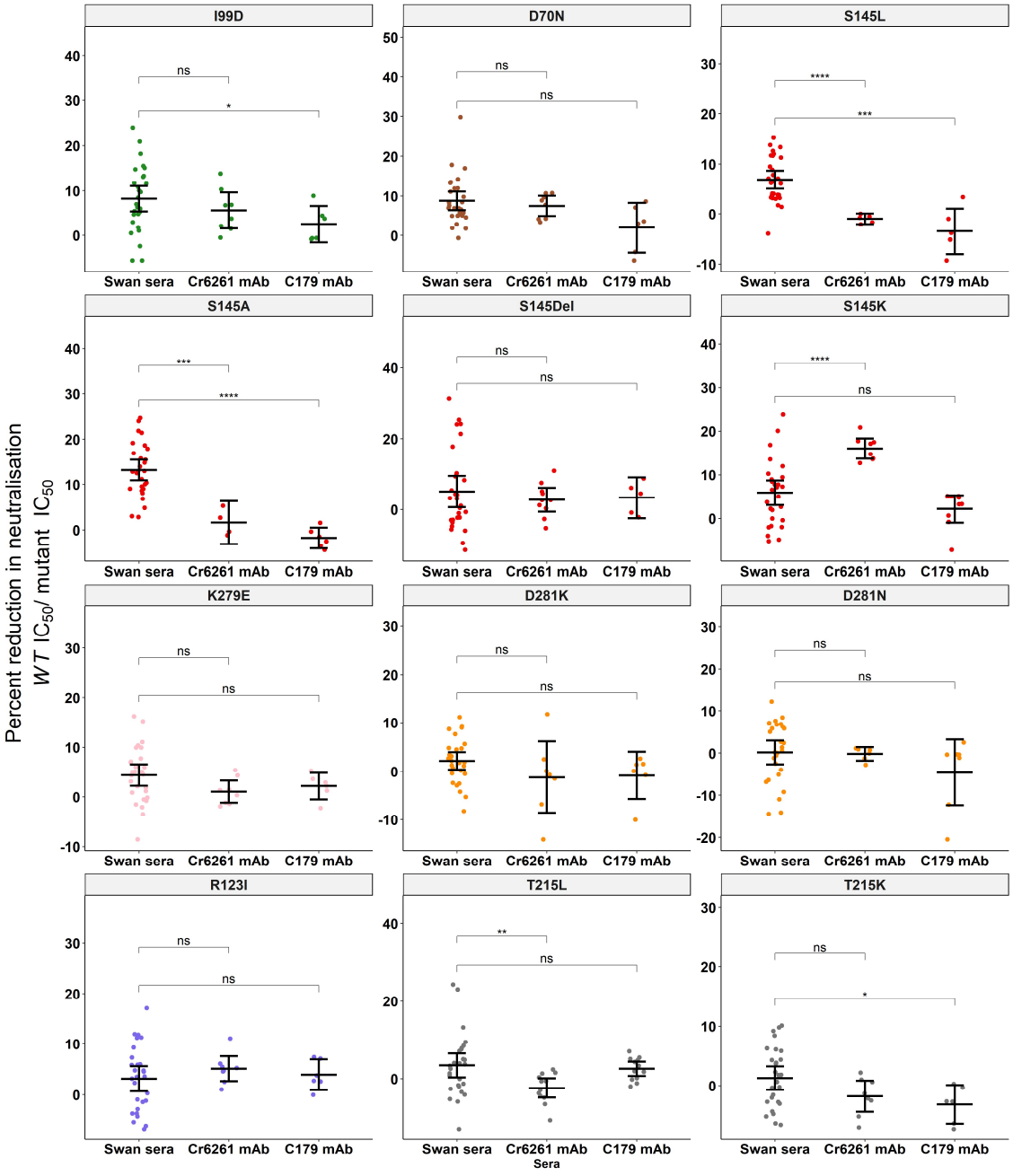


Figure 39. Differences in the avian neutralising immune response against mutant and WT A/bar-headed goose/Qinghai/3/2005/H5N1 HA pseudotyped influenza viruses. The y-axis denotes the percentage change in neutralisation between WT HA and mutant A/bar-headed goose/Qinghai/3/2005/H5N1 HA pseudotyped influenza viruses. The x-axis denotes sera or mAb. Facets denote different mutations. Error bars denote the 95% confidence interval. Statistics were calculated using $\log_{10}(IC_{50})$ values. Differences between means of sera and stem-targeted mAbs were determined by student's t-test. Significance stars: * p-value < 0.05, ** p-value < 0.005, *** p-value < 0.0005 and **** p-value <0.00005.

12.2.7 Mutations to key residues in the head domain of the A/USSR/90/1977/H1N1 HA confer a significant change to the human neutralisation response

I calculated the percentage change in immune response between human sera and stem-targeted mAbs for *WT* and mutant A/USSR/90/1977/H1N1 HA pseudotyped influenza viruses. There were significant differences between human sera and both mAbs for all A/USSR/90/1977/H1N1 mutants determined by student's t-tests (Figure 40).

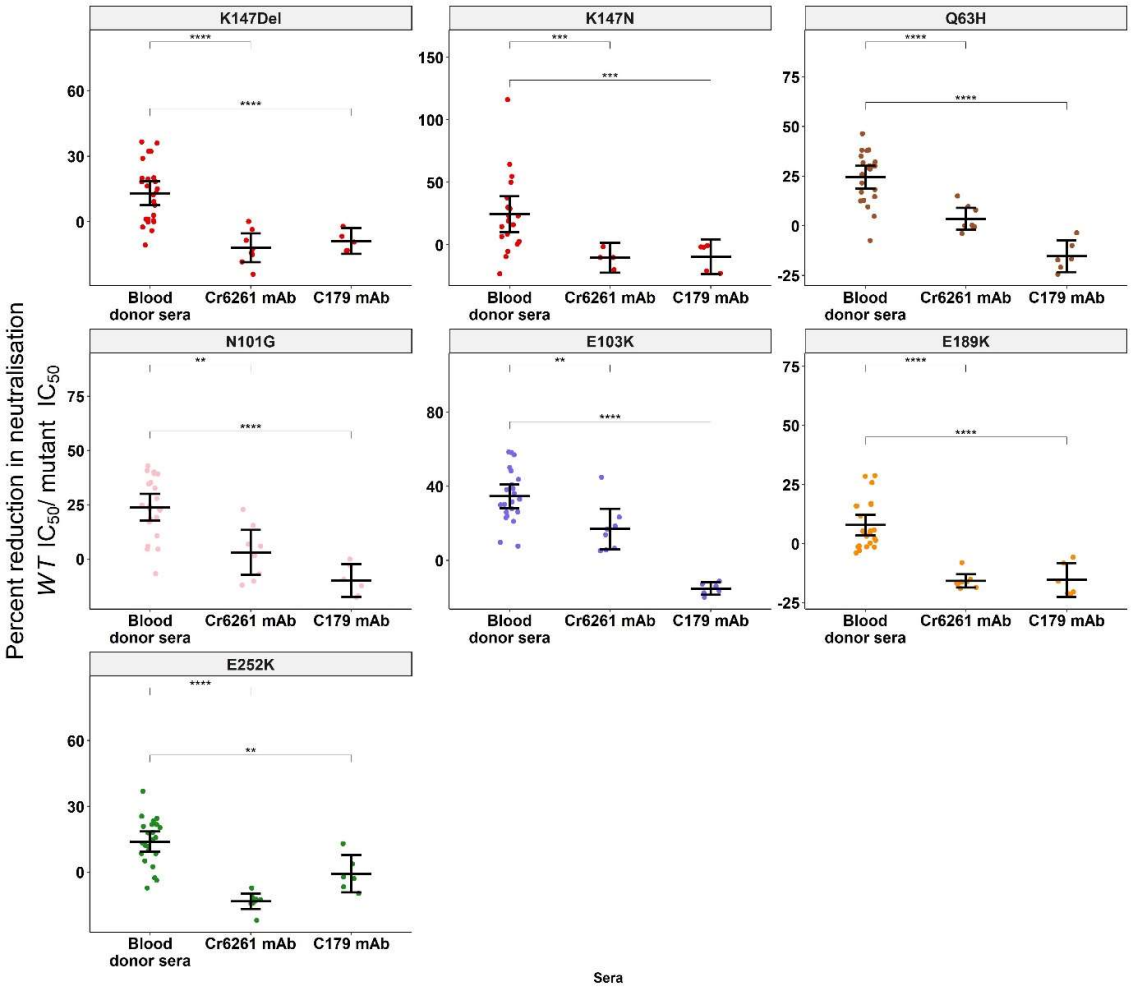


Figure 40. A comparison between the change in human neutralising immune response for mutant and *WT* A/USSR/90/1977/H1N1 HA pseudotyped influenza viruses. The y-axis denotes the percentage change in neutralisation between *WT* HA and mutant A/USSR/90/1977/H1N1 HA pseudotyped influenza viruses. The x-axis denotes sera or mAb. Facets denote different mutations. Error bars denote the 95% confidence interval. Statistics were calculated using $\log_{10}(IC_{50})$ values. The difference between means of sera and stem-targeted mAbs was determined by student's T-test. Significance stars: * p-value < 0.05, ** p-value < 0.005, *** p-value < 0.0005 and **** p-value <0.00005.

12.3. Discussion

12.3.1 Summary

I introduced amino acid point mutations to the HA head domains of A/bar-headed goose/Qinghai/3/2005/H5N1 and A/USSR/90/1977/H1N1. Mutation selection was informed by the results of the bioinformatic analysis (Results 11.2.5). I pseudotyped these mutant HAs and normalised the neutralisation using two stem-targeted mAbs with known epitopes (Figure 33 and 34).

I tested the mutant A/bar-headed goose/Qinghai/3/2005/H5N1 pseudotyped influenza viruses on an age-randomised selection of 30 Scottish and 30 wild avian sera samples (Figures 35 and 36). I tested the mutant A/USSR/90/1977/H1N1 pseudotyped influenza viruses on the same Scottish blood donor sera sample as above (Figure 37).

I then accurately determined the impact of the mutations on neutralisation. I did this by controlling for differences in HA amount and distribution on the pseudotyped virus surface that may impact neutralisation. I showed that some A/bar-headed goose/Qinghai/3/2005/H5N1 mutations caused a significant reduction in neutralisation response when testing against the Scottish blood donor sera: I99D, S145L, T215L, and T215K (Figure 38). Other mutations did not show any significant difference: D70N, S145Del, K279E, D281K, D281N, and R123I (Figure 38). The mutations S145K and S145A showed discordant results between mAbs (Figure 33).

Fewer A/bar-headed goose/Qinghai/3/2005/H5N1 mutations caused significant differences when tested against avian sera. S145L and S145A both elicited significant differences. No significant differences were seen for mutations: D70N, S145Del, K279E, D281K, D281N, and R123I (Figure 39). Discordant results between mAbs were seen for I99D, T215L and T215K.

Every A/USSR/90/1977/H1N1 mutation tested caused significant differences between *WT* and mutant pseudotypes when tested against human sera (Figure 40).

12.3.2 Neutralising immunity in sera from blood donors is directed against the H5 HA head domain

Mutations at H5 position 145 caused a significant impact on the human and avian neutralisation response. In addition, mutations at positions 99 and 215 caused a significant impact on the human neutralisation response. All these mutations impacted the HA head domain. The assay was controlled for non-head-specific pseudotype changes. Therefore, my results suggest that single amino acid residue changes in the head domain are sufficient to significantly modulate the human and avian immune response towards A/bar-headed goose/Qinghai/3/2005/H5N1. This is strong evidence that the population immunity seen in Chapter 10 is partly head-domain specific.

12.3.3 There is evidence for a cross-reactive ELV shared between H5N1 and H1N1, this has implications for the design of a "universal" influenza vaccine

In my bioinformatic analysis, I defined a putative ELV containing H5 residue 145, naming it "H5PUTELV" (Results 11.2.5). H5PUTELV is at the same position on the H5N1 HA as the H1 ELV OREO is on H1N1 [147]. OREO was shown to function as an intra-subtype H1 ELV [147]. OREO comprises the H5 145 structurally analogous H1 residue 147. I therefore mutated both the H5/H1 145/147 residues and tested the impact on neutralisation with the same human sera samples.

Mutating H5 residue 145 significantly modulates the human and avian neutralising immune response (Figures 38 and 39). Furthermore, mutating H1 position 147, which is the structural analogue of H5 residue 145, significantly modulates the human immune response (Figure 40). Both H5 and H1 assays were performed with the same Scottish sera samples. Therefore, by definition, the sera comprising the sample set is cross-reactive. The cross-reactive antibodies are modulated by changes in the H5/H1 structural analogue residue 145/147. This suggests that the anti-A/bar-headed goose/Qinghai/3/2005/H5N1 and anti-A/USSR/90/1977/H1N1 population immunity seen in Chapter 10 is partially directed towards an epitope containing H5/H1 residue 145/147. This is strong evidence

that there is an H5/H1 residue 145/147-containing conserved epitope that is shared between the H5N1 and H1N1 subtypes.

The mutations I99D, T215K, and T215K significantly impacted the human but not avian neutralisation response (Figures 38 and 39). S145A significantly impacted the avian but not human immune response (Figures 38 and 39). This suggests that the human and avian anti-A/bar-headed goose/Qinghai/3/2005/H5N1 population immunity seen in Chapter 10 is polyclonal; and directed towards different epitopes. However, changing A/bar-headed goose/Qinghai/3/2005/H5N1 residue 145 from Serine to Leucine significantly impacted both the human and avian neutralisation responses (Figures 38 and 39). This suggests that human and avian anti-A/bar-headed goose/Qinghai/3/2005/H5N1 neutralising immunity targets a conserved epitope containing position 145. This adds credence to the hypothesis that there is an epitope of limited variability (ELV) comprising residue 145 because human and avian immunity are both modulated when mutating residue 145.

Both H5ELV and OREO comprise the HA's sialic acid receptor binding site (RBS), an important and structurally conserved region of the HA (Results 11.2.5). Through homology modelling, I predicted that mutating H5 residue 145 from Serine to Leucine would alter the orientation of the RBS 130 loop, which is important in HA-sialic acid binding (Results 11.2.6). This is likely the reason why mutating residue 145 significantly impacts the immune response. It is likely that the anti-A/bar-headed goose/Qinghai/3/2005/H5N1 partially recognises an epitope with one orientation of the 130 loop but not the orientation after the S145L mutation. Furthermore, this may partially explain why the S145L and L145S mutations have happened repeatedly over time and phylogenetic clade, and comprise the majority of all H5N1 strains (Results 11.2.5): there is a structurally conserved epitope comprising the RBS (H5PUTELV).

The head domain is the most antigenic region of HA [5]–[7]. 60% of anti-HA antibodies are directed towards the head domain [5]. This makes the head domain the primary target for influenza vaccines

[5]–[7]. However, the head domain is highly variable (Results 11.2.1). Due to antigenic drift causing a progressive buildup of immune escape mutations, it is difficult to design cross-reactive influenza vaccines [355] [102], [103], [122], [344], [367], [381], [382]. If H5PUTELV and OREO are cross-reactive H5/H1 subtype ELVs, then a vaccine targeting this area may immunise against both subtypes.

12.3.4 Non-significant HA head mutations act as negative controls

It was conceivable that the process of SDM itself introduced significant differences in immune responses. I show that mutating A/bar-headed goose/Qinghai/3/2005/H5N1 residues 70, 123, 279, and 281 elicit no significant differences between human/ avian sera and two stem-target mAbs (Figures 38 and 39). These mutations act as negative controls for the SDM technique.

12.3.5 The human neutralising immune response towards A/USSR/90/1977/H1N1 is highly polyclonal and directed towards non-classically defined H1N1 HA head epitopes

Apart from H1 residue 147, I mutated residues that are not part of classically defined H1 epitopes (Figure 34) [106], [107]. Residues 63, 101, and, 103 are situated in the N-Terminal Vestigial Esterase (VE) subdomain of the head [36], [37], [41]. The role of VE is not adequately known [36]. The VE is known to be more conserved than the Receptor Binding Domain (RBD).

Residues 189 and 252 are in the RBD [36], [37], [41], a highly immunogenic and variable region [47]. I initially suspected that mutating residues 147, 189 and 252 would cause more significant differences than those in the VE. However, all mutations caused highly significant differences. My results suggest that the anti-A/USSR/90/1977/H1N1 population immunity seen in Chapter 10 is highly polyclonal. As well as residue 147, mutating multiple other residues not in classically defined epitopes is sufficient to significantly modulate the human immune response.

12.3.6 The human and avian anti-A/bar-headed goose/Qinghai/3/2005/H5N1 neutralising immune response is polyclonal

I show that human anti-A/bar-headed goose/Qinghai/3/2005/H5N1 population immunity is partially directed towards epitopes containing amino acid residues: 99, 145, and 215 (Figure 38). I show that avian population immunity is not partially directed towards residues 99 and 215 in the same manner as humans (Figure 39). I also suggest that S145A alters the conformation of a head epitope in a way that impacts avian but not human-neutralising immunity. This evidence suggests that the avian is less polyclonal than human anti-A/bar-headed goose/Qinghai/3/2005/H5N1 population immunity.

I hypothesise that human anti-A/bar-headed goose/Qinghai/3/2005/H5N1 population immunity is elicited due to cross-reactive antibodies raised against seasonally circulating H1N1 strains. I hypothesise that avian anti-A/bar-headed goose/Qinghai/3/2005/H5N1 population immunity is elicited due to routine low pathogenicity avian influenza (LPAI) and sporadically spreading high pathogenicity avian influenza (HPAI) H5 strains.

It is conceivable that H5 residues 99 and 215 comprise epitopes that have H1N1 analogues. Humans develop immune responses to these H1N1 epitopes, and therefore these antibodies recognise the analogous ones in H5. H5 residue 215's H1N1 structural analogue is adjacent to the classically defined antigenic site Sb. Therefore, humans may recognise a Sb analogue in H5N1. Moreover, H5 residue 99 is the structural analogue of H1 residue 100; next to H1 mutation 101. Mutating both these H5/H1 residues significantly modulates the immune response in the same human samples. This could be evidence that multiple shared H5/H1 epitopes exist.

If this is the case, mutations in H5 residues 99 and 215 not causing a significant immune response could be explained. Birds do not experience the same circulation of H1N1 as humans do. If cross-reactive antibodies to shared H5/H1 epitopes is the reason for mutations in H5 residues 99, 145 and, 215 significantly impacting the human neutralisation response, birds are absent from this phenomenon. Potentially, LPAI H5 strains that the swans experience do not contain the same cross-reactive epitopes as H1N1.

13. Discussion

13.1. Summary

13.1.1 SARS-CoV-2 serosurvey

SARS coronavirus 2 (SARS-CoV-2) emerged in 2019 and has since spread around the world causing over 700 million cases and up to 7 million deaths [3]. It prompted many governments to enact 'lockdown' policies – consisting of banning social gatherings and closing workplaces to try to contain the virus until herd immunity could be achieved through vaccination [383]. In order to ascertain whether lockdown needed to be maintained after the first wave (March to May 2020), and determine how close the UK population was to obtaining herd immunity, we conducted a time-course serosurvey of Scottish blood donors between the 17th of March and 18th of May.

We used a lentivirus-based pseudotyped virus system displaying the SARS-CoV-2's Spike (S) protein to detect neutralising antibodies against the S protein. We performed this on 3,500 sera samples from Scottish blood donors provided by the SNBTS. An Enzyme-Linked Immunosorbent Assay (ELISA) was also used to test selected samples. I then used logistic regression to model the seroprevalence of the Scottish population.

I found a total of 112/3500 Scottish blood donor samples contained neutralising anti-SARS-CoV-2 antibodies as defined by a novel pseudotyped virus microneutralisation assay. Seroprevalence increased over our study period. By the end of our study period, after ten weeks, six Health Boards: Lothian, Tayside, Greater Glasgow & Clyde, Ayrshire & Arran, and Grampian were predicted to have a non-zero seroprevalence. four Health Boards: Highland, Fife, Borders and Forth Valley were not predicted to have a significantly non-zero seroprevalence. Lanarkshire was predicted to have the greatest seroprevalence by the end of our study period, 9.7% (with a 95% CI of 5.9%-13.5%).

13.1.2 Influenza serosurvey

I determined the population immunity to a range of human pandemic risk H5N1 and historical H1N1 pseudotyped influenza viruses. I did this using a sample of the Scottish population taken from blood donors aged 17-80. I found that 479/520 people neutralised the HA from A/bar-headed goose/Qinghai/3/2005/H5N1, an HPAI influenza virus that did not infect people in Scotland. This strain was part of a devastating avian pandemic that seeded multiple human spillover events and was shown to spread between mammals. In contrast, only 2/273 people responded to A/chicken/Scotland/1959/H5N1. I showed that 242/362 and 119/273 neutralised A/South Carolina/1/1918/H1N1 (the causative agent of the 'Spanish influenza' pandemic) and A/California/07/2009/H1N1 (the causative agent of the 'Swine flu' pandemic) respectively. Moreover, 109/273, 246/273, and 251/273 people neutralised the A/PR/8/1934/H1N1, A/USSR/90/1977/H1N1, and A/Solomon Islands/3/2006/H1N1 influenza strains respectively.

I also conducted a serosurvey of wild avian samples with a range of pandemic-risk H5N1 strains. H5N1's natural reservoir is water birds, so using wild swan sera allows us to examine a potential sentinel population. I showed that 66/92 swans neutralised the HA from A/VietNam/1203/2004/H5N1, an H5N1 strain that infected people in Vietnam and was deemed to be of pandemic risk. In addition, 71/92, 74/96, and 74/92 swans neutralised the HAs from A/bar-headed goose/Qinghai/3/2005/H5N1, A/Chicken/Deli_Derdang/BBPVI/2005/H5N1, and A/goose/Guiyang/337/2006/H5N1 respectively.

There were significant human neutralisation strength differences between every pseudotyped influenza viruses apart from A/bar-headed goose/Qinghai/3/2005/H5N1 and A/PR/8/1934/H1N1, and A/South Carolina/1/1918/H1N1 and A/California/07/2009/H1N1. No differences were seen in the avian serosurvey.

Statistical modelling showed that there could be significant non-linear relationships between human neutralising responses and age. Binomial generalised additive models (GAMs) showed that the chance

of seropositivity towards A/PR/8/1934/H1N1 and A/Solomon Islands/3/2006/H1N1 is significantly impacted by age (adjusted R^2 of 0.28 and 0.4 respectively). Interestingly, these show directly antagonistic patterns, even though the same samples were used. Moreover, gaussian GAMs showed that the neutralisation strength against A/USSR/90/1977/H1N1 was significantly impacted by age, and to a lesser extent against A/bar-headed goose/Qinghai/3/2005/H5N1 (adjusted R^2 of 0.41 and 0.2 respectively). Hatch-year does not seem to impact the avian neutralisation response. Statistical modelling trends appear to be congruent with an examination by a 3-year rolling average. However, statistical modelling may not reflect reality or show mechanisms.

13.1.3 Bioinformatics

The globular head domain of the H5Nx and H1N1 subtype HA is more variable than the stem domain. I showed this by calculating the Shannon mutational entropy of the large amino acid datasets characterised in Methods 8.4.1.

I applied a structural bioinformatic pipeline to characterise the variability of putative antibody binding sites (ABSs) on the HAs of H5N1 and H1N1. This was done using a range of ABS sizes and solvent-accessibility thresholds in the same manner [147]. I applied the pipeline to the large amino acid database to assess the natural variability of putative ABSs. Variability is not uniform across the HA's structure. There exist pronounced peaks and troughs in variability throughout the head domain. Some troughs were selected to explore further, with the hypothesis that they may be epitopes of limited variability (ELVs).

I also applied different data to the structural bioinformatic pipeline. Bloom et al., (2016) determined the mutational tolerance of the H1N1 HA [241]. They did this by generating every mutation possible and quantifying the frequency observed. I used this data, expressed in the unit "bits," to calculate the theoretical mutational tolerance of the putative ABSs. Interestingly, these data did not match the

pattern across the data as my natural variation analysis. In fact, in the head domain of the H1N1 HA, the mutational showed the opposite pattern.

I identified a putative ABS situated at position 157 on the H5N1 alignment which I named "H5PUTELV". This is situated at a pronounced dip in natural variability computed using the structural bioinformatic pipeline. This minimum is structurally aligned to a corresponding minimum in the H1N1 HA. The corresponding H1N1 minimum was used to identify the ELV 'OREO'. I therefore hypothesised that H5PUTELV may be an ELV shared between the H5N1 and H1N1 subtype.

The H1N1 residue 147 was shown to be structurally key to the functioning of the H1N1 OREO ELV. I showed that the phylogenetic behaviour of H5Nx residue 145 is similar to H1N1 residue 147, the structural analogue of H5Nx residue 145.

H5PUTELV, like H1N1's OREO, is situated on top of the HA's sialic acid RBS. I, therefore, attempted to predict whether mutations in H5PUTELV residue 145 would structurally impact the RBS. Through homology modelling and structural alignment, I showed that mutations theoretically alter the orientation of the RBS 130 loop.

13.1.4 Site-Directed Mutagenesis

I hypothesised that the neutralising antibody response towards the HA of A/bar-headed goose/Qinghai/3/2005/H5N1 may be comprised of antibodies originally elicited by exposure to H1N1 influenza strains. This hypothesis was generated because no one in Scotland has been documented to be infected with A/bar-headed goose/Qinghai/3/2005/H5N1 or other H5N1 strains. It appears that the structure and variability patterns of the H5N1 and H1N1 HAs are similar. Therefore, neutralising antibodies towards the H1N1 HA may also recognise epitopes on the H5N1 HA. I hypothesised that if this is the case, then the antibodies may be partially directed towards shared epitopes in the variable

globular head domain of HA. I identified 'H5PUTELV' which is a putative ABS on the head of the H5N1 HA that appears to be structurally analogous to the H1N1 ELV named OREO.

To test this, I introduced amino acid point mutations to the H5N1 residue 145 and the structurally analogous H1N1 residue 147, which was shown to be key in the functioning of the H1N1 OREO epitope. This is the process of site-directed mutagenesis (SDM). I introduced these mutations into the HAs of A/bar-headed goose/Qinghai/3/2005/H5N1 and A/USSR/90/1977/H1N1. I also introduced various other mutations across the head domains of these viruses to control for type 1 error introduced by SDM.

I pseudotyped these mutants and characterised them with two stem-targeted monoclonal antibodies (mAbs). I tested the mutant A/bar-headed goose/Qinghai/3/2005/H5N1 and A/USSR/90/1977/H1N1 pseudotypes on an age-randomised selection of 30 Scottish sera samples. I also tested the mutant A/bar-headed goose/Qinghai/3/2005/H5N1 HA pseudotyped virus on a hatch-date randomised selection of 30 wild swan sera samples. I then accurately determined the impact of the mutations on neutralisation. I did this by controlling for differences in HA amount and distribution on the pseudotyped virus surface using the mAb data. I showed that some A/bar-headed goose/Qinghai/3/2005/H5N1 mutations caused a significant reduction in neutralisation response when testing against the Scottish sera: I99D, S145L, T215L, and T215K. Other mutants did not show any significant difference: D70N, S145Del, K279E, D281K, D281N, and R123I. The mutations S145K and S145A showed discordant results between mAbs.

Fewer A/bar-headed goose/Qinghai/3/2005/H5N1 mutations caused significant differences when tested against avian sera. S145L and S145A both elicited significant differences. No significant differences were seen for mutations: D70N, S145Del, K279E, D281K, D281N, and R123I (Figure 38). Discordant results between mAbs were seen for I99D, T215L and T215K.

Every A/USSR/90/1977/H1N1 mutation tested caused significant differences between *WT* and mutant pseudotypes when tested against human sera.

13.2. Outlook

13.2.1 Our SARS-CoV-2 serosurvey suggested the likelihood of subsequent COVID-19 waves

These data indicate that serosurveys of blood donor samples can serve as a useful tool for tracking the emergence and progression of an epidemic like the SARS-CoV-2 outbreak. Although sampling is partially non-uniform and biased, these results also stated that only a relatively small proportion of Scotland had anti-SARS-CoV-2 antibodies towards the end of the UK's first COVID-19 pandemic wave. Our study indicated that the seroprevalence of Scotland in May 2020 remained low, suggesting the need for further public health interventions.

13.2.2 People in the UK neutralise the HA from H5N1 avian influenza

The discovery of widespread anti-H5N1 neutralising antibodies within the UK population has potential ramifications for influenza research and public health policy. Historically, H5N1 avian influenza viruses have gained attention for their high pathogenicity and potential to cause severe illness and mortality in humans [25]–[29]. Sporadic cases of H5N1 infection in humans have been repeatedly documented, however, this has not been documented to occur in Scotland. Human cases are typically associated with avian populations or individuals with direct occupational exposure to poultry. Indeed, studies have found seroprevalence of anti-H5N1 antibodies amongst poultry workers in Vietnam [384]. The presence of widespread neutralising immunity to H5N1 among individuals in the UK unveils a challenge towards previous understandings. The observation of human immunity towards H5N1 in the UK could suggest that multiple individuals may have mounted effective immune responses against H5N1 before. However, this is unlikely as H5N1 has not been documented to spread amongst the UK population. This suggests that the anti-H5 neutralising immunity seen is due to cross-reactivity with another structurally similar virus displaying similar HA epitopes. Evidence for this was uncovered in Chapter 12. This may suggest that anti-influenza immunity, which is characterised by long-term, sequential strain exposure, is generally more reactive than commonly thought. This corroborates

[123], [126], [127] that the repertoire of anti-HA responses is increased by sequential exposure, likely due to cross-reactive antibodies produced after affinity maturation in germinal centres.

Our SARS-CoV-2 serosurvey was conducted during the early stages of the COVID-19 pandemic, Chapter 9. At this stage of the pandemic, it was not known whether there would be a second wave. Neutralising antibodies towards viral glycoproteins are known to be a correlate of protection [309], [310]. It is also known that large-scale population immunity can reduce the virus' basic reproduction number (R_0) [93], [94]. Therefore, it is important to understand how extensive anti-SARS-CoV-2 neutralising antibodies were in the Scottish population. Our study showed that the population immunity towards SARS-CoV-2 was low after the first wave; around ten percent in the highest Health Board. This allowed public health services to predict that a second wave was likely and informed public health policy such as the implementation of lockdowns and the necessity to pursue vaccine campaigns.

Drawing parallels with this, the implications of widespread human immunity towards H5N1 may extend beyond individual-level protection to broader considerations of pandemic risk assessment and preparedness. Avian influenza viruses from the H5N1 subtype pose an ever-present threat of human spillover events which risk becoming global pandemics [35]. The detection of human immunity towards H5N1 suggests that certain segments of the UK population may possess a degree of protection against this highly pathogenic influenza virus, potentially influencing the dynamics of H5N1 transmission and the likelihood of a pandemic outbreak. While the exact extent and duration of this immunity remain to be elucidated, the presence of pre-existing immunity could serve as a mitigating factor in the event of an H5N1 pandemic, reducing the severity and impact of disease transmission. Evidence of pre-existing immunity to H5N1 among individuals in the UK may contribute to enabling better pandemic risk assessment, calibrating epidemiological models, and the facilitation of implementation of public health interventions. The recognition of pre-existing immunity to H5N1

within the UK population underscores the need for ongoing surveillance efforts aimed at monitoring influenza viruses with pandemic potential.

Moreover, the observed cross-reactivity between H5 HPAI and H1N1 HAs adds complexity to the understanding of influenza serosurvey results. This cross-reactivity may make it challenging to differentiate between immunity acquired from past H1N1 infections and immunity triggered by exposure to other influenza A strains, such as H5 HPAI. Consequently, it becomes more difficult to accurately assess individual immune histories and the breadth of protection against different influenza A subtypes.

13.2.3 Human anti-H5N1 neutralising immunity is partially directed towards the head domain of HA

Cross-subtype anti-HA influenza antibodies have been identified that recognise both H5N1 and H1N1 strains [267]–[269]. Epitopes that these antibodies recognise have been isolated in the conserved stem domain of HA [60], [77], [86]–[88], [267]–[269] This type of cross-reactive antibody response has been shown to positively correlate with age [348]. The stem subdomain is known to be an immune-subdominant region of the HA [54]. However, I have shown that introducing amino acid point mutations into the head domain of the H5N1 HA significantly modulates the human-neutralising immune response. This was controlled for using stem-targeted mAbs which eliminates non-head domain changes in the assay. This provides strong evidence that some of the antibodies identified in Chapter 10 target the immunodominant HA head domain.

13.2.4 Identification of cross-reactive, pan-subtype epitopes of limited variability

Epitopes of limited variability (ELVs) have been shown to exist within strains of the same subtype [147]. However, HA head-domain ELVs have not been shown to cross-react between different subtypes. I have shown that mutating the same, structurally analogous, residues in H5N1 and H1N1

significantly modulates the immune response. Indeed, the location of H5N1 residue 145 is structurally analogous to the H1N1 residue 147. H1N1 residue 147 was shown to be critical in the immunological behaviour of the H1N1 OREO ELV [147]. I suggest that 'H5PUTELV' is a structurally limited epitope which exists in H5N1 in the same orientation as OREO does in H1N1. These data provide evidence for the existence of a cross-reactive epitope shared between these influenza virus subtypes. This finding suggests that the immune system recognises and responds to a conserved region within HA that is structurally similar between H1N1 and H5N1 strains. This finding represents a significant contribution to the field but must be elaborated on by further experiments to bolster its validity. I should treat this finding as a 'proof of concept,' that cross-subtype ELVs exist but it must be studied further.

13.2.5 The antigenic evolution of influenza in relation to these findings

Many studies have found that immunity towards influenza is predominantly strain-specific. This immunity decreases with antigenic distance, which correlates with time. That is, as influenza antigenically drifts away from the original strain due to escape mutations, antibodies cross-react less and less [98]–[102].

Various mathematical models such as the antigenic thrift model [144]–[146] or the epochal evolution model from Katia Koelle's research group [139] state that the antigenic evolution of influenza is driven by changes in a limited number of key residues or epitopes/alleles of limited variability. I have shown that mutating single amino acids can have a significant impact on the human immune response. This corroborates Doud, Hensley and Bloom (2017) that single, key amino acids are critical in determining the immune response towards influenza viruses. Antigenic thrift/ limited drift predicts that conserved, immunologically important regions on the head of HA are impacted by the aforementioned mutations. In essence, these concepts posit the existence of HA epitopes characterised by limited variability due to functional constraints (ELVs). A growing number of serological and monoclonal antibody (mAb) studies are providing evidence for this [75], [76], [85]–[88], [123], [126], [127], [147], [385], [386],

[77]–[84]. I believe this study provides important evidence for the existence of a restricted ELV that cross-reacts between the H5 and H1 subtypes. The presence of shared epitopes between different influenza virus subtypes challenges the conventional understanding of subtype-specific antigenicity. Therefore, I suggest that these findings provide evidence to support the aforementioned theories regarding the antigenic evolution of influenza.

13.2.6 Implications for the development of 'universal' influenza vaccines

A 'universal' influenza vaccine is a theoretical vaccine that can induce broad and durable immunity against multiple strains and/ or subtypes of influenza [59], [134], [167], [387], [388]. The rationale for this is to provide protection against divergent strains, thus reducing the need for frequent vaccine updates. Moreover, highly broad protection may allow for protection against future pandemic strains that have not yet evolved.

Immunodominance refers to the phenomenon in which certain influenza epitopes are preferentially recognised by the host immune system, eliciting a robust and dominant immune response [54], [148], [160]. HA is comprised of multiple epitopes that can be targeted by antibodies [56], [105], [114]–[117], [106]–[113]. However, some are more targeted than others. One example of this is the dominance of HA's head over the stem domain [54], [148], [160]. Conserved epitopes on the stem domain have been targeted by teams attempting to design 'universal' influenza vaccines [157]–[159], [389], elaborated on in Introduction 7.13. None of these vaccines have so far been successful. This could be due to immunodominance; although present, these conserved epitopes may not produce a strong enough immune response to provide protection. This is the reason why head epitopes are the most sought-after antigens to incorporate into influenza vaccines.

The existence of a shared, head-based epitope between H1N1 and H5N1 has significant implications for influenza vaccine design and pandemic preparedness. If incorporated into vaccines, the findings may allow researchers to overcome the limitations of current influenza vaccines, which classically

provide strain-specific protection and require frequent updates to match circulating strains [92], [103], [104]. I suggest that targeting these conserved head ELVs within the head of HA may confer broad cross-protection against multiple, pan-subtype influenza strains. These theoretical vaccines may reduce the need for annual vaccine updates, eliminating the intensive and difficult process of strain-matching.

Moreover, these results may have significant implications for pandemic preparedness. Influenza pandemics, such as the 2009 "swine flu", highlight the urgent need for vaccines that can provide rapid and effective protection against shifted influenza viruses. Universal vaccines that target conserved epitopes offer a proactive approach to pandemic preparedness, as they may have the potential to provide cross-reactive immunity against novel influenza virus strains with pandemic potential.

13.3. Limitations

13.3.1 Pseudotype neutralisation assay

The human immune response towards influenza is complex and varied. Humans elicit cellular and humoural responses towards all three surface glycoproteins of influenza [50]. I have focused on anti-HA responses as this is influenza's most targeted antigen. However, I have not assessed antibody responses to NA for example. I have also not assayed for cellular responses in our study. Moreover, the antibody response towards human influenza comprises neutralising and non-neutralising antibodies. I have not tested for non-neutralising antibodies in our study.

Live-virus neutralisation assays are not possible to perform for HPAI H5N1 influenza viruses unless utilising biosafety level 4 (BSL4) laboratories. However, using live-virus-neutralising antibodies would paint a more comprehensive picture of the cross-reactivity I identified. Moreover, by using reverse genetics (RG), reassortant viruses could be generated with a PR8 H1H1 (Methods 8.4.2) backbone and H5 HPAI HA/NA, with the multi-basic cleavage site removed. This approach would produce mutants that could be studied in BSL2/3 laboratories, enabling the live-virus use of H5 HPAI.

13.3.2 Population immunity and its public health impact

Neutralising antibodies towards influenza viruses are known to be correlates of protection [309], [310]. I suggest that having a large amount of population immunity towards influenza viruses may contribute to the risk of these viruses spreading in those respective populations. These are the conclusions that I drew during our SARS-CoV-2 serosurvey. However, it is not certain that this is the case. Studying this phenomenon would be extremely difficult.

13.3.3 Comprehensiveness of the Scottish blood donor cohort

The sera sample sets I used for this project are large. They also contain a broad range of ages, which I was able to sample randomly to generate relatively equal age distributions. However, most of these samples were from 20 years of age or older. Testing those younger than 20, including children, would allow a more comprehensive assessment of anti-influenza HA cross-reactivity.

Moreover, the human samples I used were all sampled from similar time points: from March to May 2020. The samples, therefore, represent a snapshot of immunity at one specific time. These conclusions do not capture any potential longitudinal change in the immune response, which could be impacted by sequential seasonal influenza exposure. This is different for our avian sample set, which contains samples ranging over multiple years.

13.3.4 Clinical relevance of cross-reactive epitopes

These findings provide valuable insights into the cross-reactivity of conserved epitopes between H1N1 and H5N1 strains. This has significant potential implications for the design of a ‘universal’ influenza vaccine. The clinical relevance of the H1N1 ELV OREO has been demonstrated in [147] where immunisation with the epitope was able to protect mice against multiple divergent H1N1 strains.

However, although theoretically similar, I have not demonstrated the clinical relevance of H5PUTELV. This must occur before I can determine whether H5PUTELV could be used as a vaccine antigen.

13.4. Areas of future study

13.4.1 Further characterisation of the H5PUTELV epitope

The identification of single residues within theoretical ELVs which modulate immune responses highlights the need for detailed characterisation of entire epitopes. Techniques such as alanine scanning and targeted SDM based on bioinformatics results should be performed to determine which residues surrounding H5 residue 145 also comprise the H5PUTELV epitope. This should also be done for other potential epitopes identified in this study, such as H5 residues 99 and 215, which both show changes in cross-reactivity. It is crucial to understand the different conformations of these epitopes. The H1N1 OREO was shown to take five main conformations [147]. Immunisation with distinct conformations was shown to provide protection to ferrets against multiple H1N1 strains. However, no conformation provided protection to all strains. Therefore, it is important to characterise the different conformations that H5PUTELV can take in order for ultimate 'universal' vaccine production.

Applying linear ELVs to platform constructs could be used to isolate mAbs specific to the theoretical H5PUTELV and others. Targeted SDM could then be used to identify which mutations allow for immune escape to these mAbs. Repeating this process with multiple mAbs would allow high-confidence conclusions as to the antigenic structure of H5PUTELV. This would function in a similar manner to the epitope mapping studies [56], [105], [114]–[117], [106]–[113], however, this would be far more targeted, and hopefully, constructive.

Employing advanced methodologies such as deep mutational scanning, as pioneered by Jesse Bloom's lab, can provide comprehensive insights into the functional significance of individual epitope residues. Using this technique would provide insights into the plasticity of these epitopes. This would help us to determine the distinct conformations that H5PUTELV can take.

13.4.2 Pulldown assays

Pulldown studies are used to study the interaction between proteins. In this assay, a bait protein, in this case a construct displaying the head domain of H5, and then H5PUTELV, is immobilised on a magnetic bead column. A sera sample is then washed through the column. This serum likely contains polyclonal antibodies. Antibodies specific to the bait protein bind and non-specific antibodies are washed through the column. After washing away non-specific antibodies, the captured antibody-bait protein complexes are eluted.

The SDM study gave strong evidence that human neutralising immunity towards the HA of a pandemic avian H5N1 virus is directed partially towards the head domain. To add confidence to this assertion, HA head pulldowns should be performed with the sera. If the eluted sera still neutralises the H5N1 HA then this is added evidence that the response is partially head-specific. This can be furthered by using a linear epitope of H5PUTELV expressed on a construct for the pulldown assay. This will add confidence to the assertion that some of the anti-HA H5N1 neutralising immunity is directed towards the H5PUTELV epitope.

13.4.3 Further characterisation of cross-reactivity identified in this study

I showed that sera taken from the same people neutralise HAs from H5N1 and various H1N1 antibodies. I furthered this observation by showing that this cross-reactivity likely targets a conserved epitope shared between H5N1 and H1N1. These observations of cross-reactivity should be furthered by expanding: (i) the repertoire of influenza strains tested against the same sera samples; testing viruses from multiple subtypes and more divergent strains to understand the limits of this cross-reactivity, (ii) performing the same SDM on an expanded number of background viruses to understand how widespread these epitopes are, and (iii) the range of sera samples, including those of younger ages, from different geographic locations, and sampled from different times.

These studies should be undertaken with multiple immunological techniques. This project primarily used the pseudotyped influenza virus microneutralisation assay to identify the action of anti-HA neutralising antibodies. To build a more comprehensive picture of the anti-HA antibody cross-reactivity, I should also repeat these observations with live virus neutralisation assays and ELISA. Multiple techniques, if the findings are similar, would provide more support for our findings. I employed an anti-Spike ELISA in our SARS-CoV-2 serosurvey to assess how the pseudotype influenza virus microneutralisation assay compares. I found that our assay and the ELISA correlated strongly, with a Pearson's correlation statistic of 0.84. Due to this, I assume that anti-HA ELISAs will yield a similar result.

13.4.4 Surface plasmon resonance

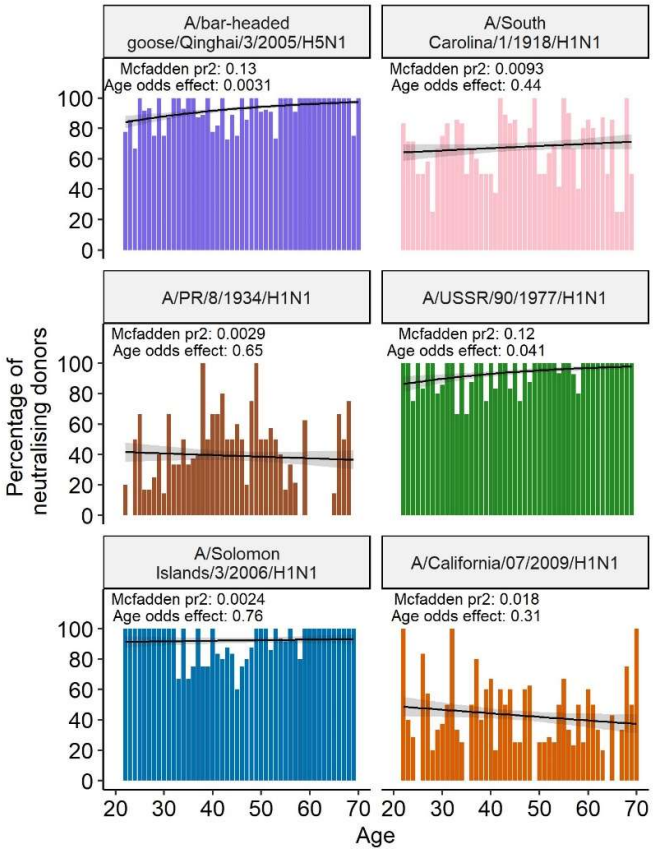
Surface plasmon resonance (SPR) can be employed to characterise the binding affinity and kinetics of neutralising antibodies towards H5PUTELV. Through this, I can characterise the association and dissociation rates of antibodies to epitope candidates, ultimately calculating the KD constant. This will allow us to select variants of H5PUTELV and other epitopes that bind antibodies stronger and for a longer time. These may be better candidates for vaccine antigens. I can also use this to study the specificity of antibodies recognising H5PUTELV, i.e. selecting those that bind only to H5PUTELV or ones that bind more targets.

13.4.5 Incorporation of H5PUTELV into a 'universal' influenza vaccine

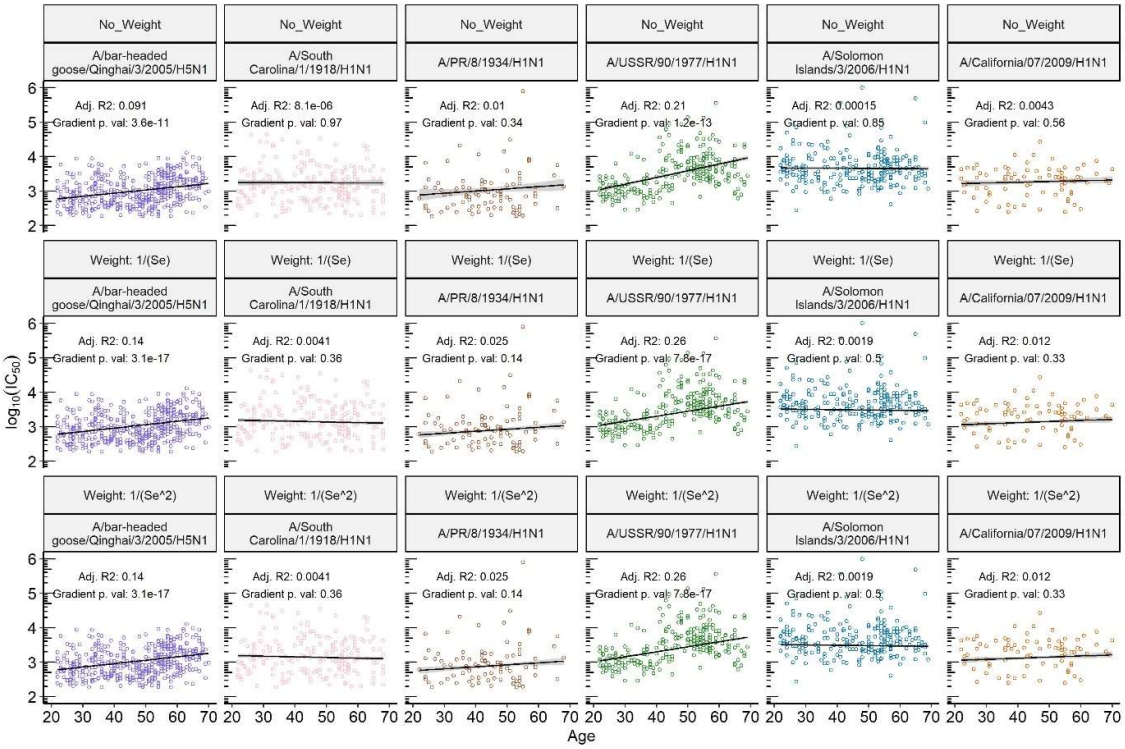
The design and implementation of an epitope-displaying vaccine is the natural progression of this project. The realisation of this will take many steps. First, the linear epitope must be expressed on a vaccine platform. If cross-reactive sera I identified as positive reacts with these constructs, then animal studies must be employed. Performing animal studies will ascertain whether immunisation with

H5PUTELV can elicit protective immunity against H5N1, as well as H1N1 strains. After animal studies, stringent clinical trials would be necessary to ascertain the validity of this approach.

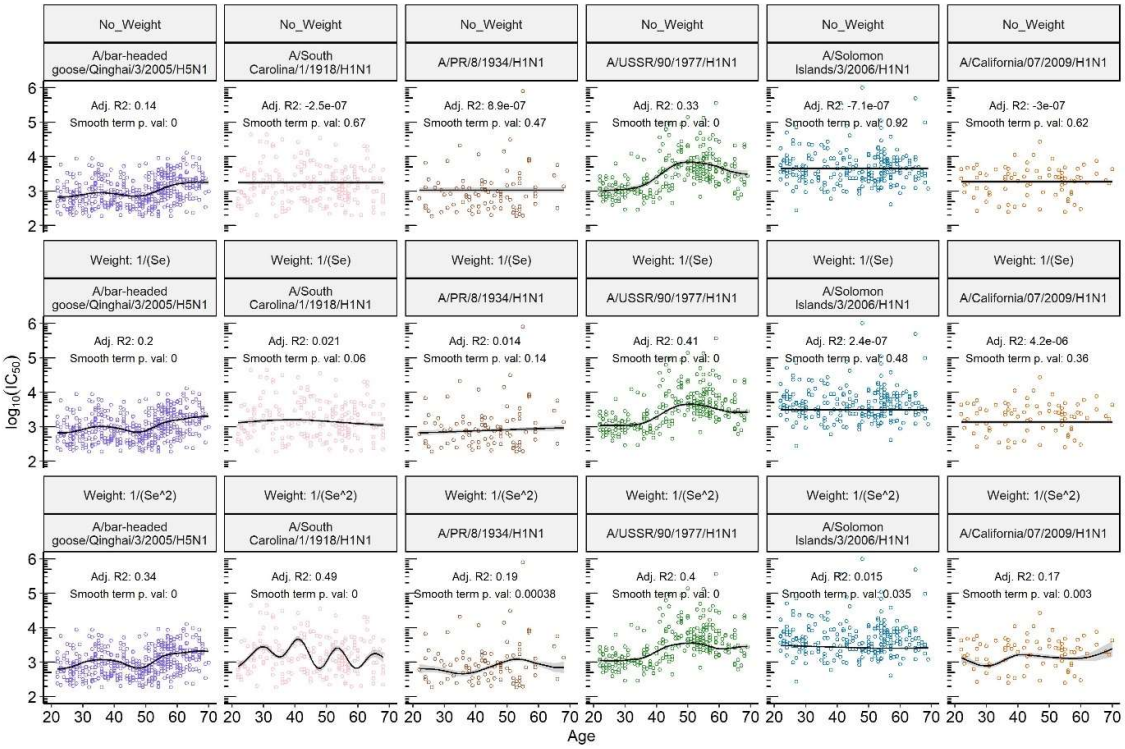
14. Supplementary materials



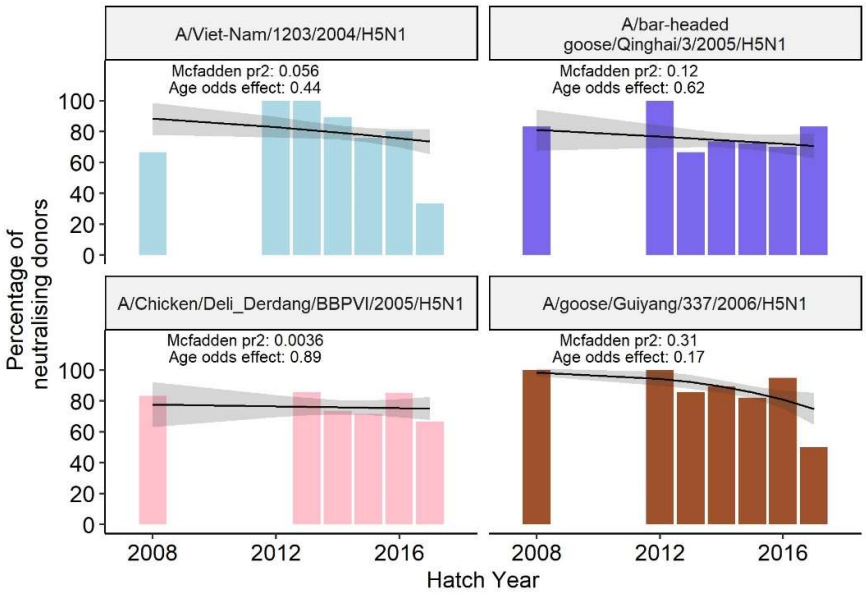
Supplementary Figure 1. The relationship between age and the percentage of donors seropositive for pseudotyped influenza viruses run on the Scottish blood donor cohort. Bars denote the percentage of seropositive donors per age. Facets denote individual pseudotypes. Black lines show the output of logistic regression models. Shaded areas around the line denote the standard error of the model. McFadden pr2 denotes the McFadden pseudo-R^2 of the model. Age odds effect denotes the significance of age in the model.



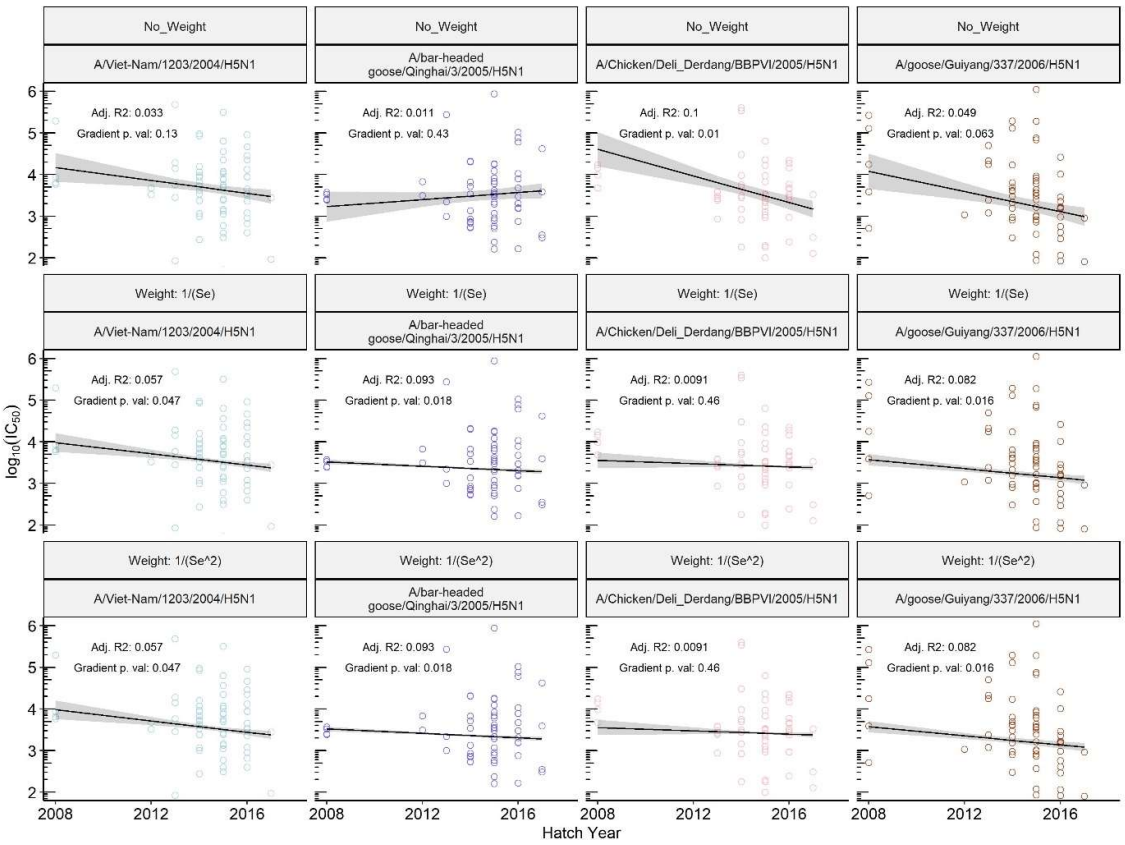
Supplementary Figure 2. The relationship between age and neutralisation strength for pseudotypes run on the Scottish sample set. Points denote neutralisation strength per age. Facets denote individual pseudotypes and weights of the models used. Black lines show the output of linear models. Shaded areas around the line denote the standard error of the model. Adj. R2 denotes the adjusted R^2 of the model. Gradient p.val denotes the significance of age in the model.



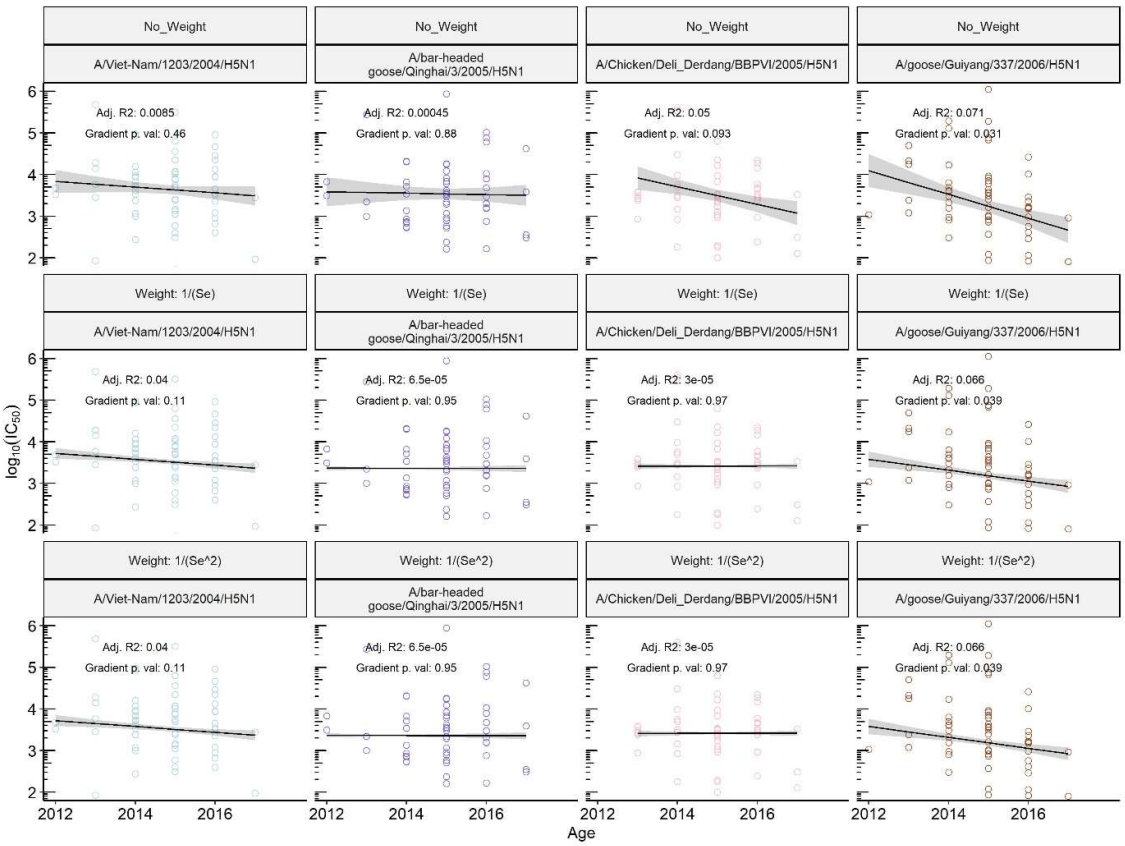
Supplementary Figure 3. The relationship between age and neutralisation strength for pseudotyped influenza viruses run on the Scottish blood donor cohort. Points denote neutralisation strength per age. Facets denote individual pseudotypes and weights of the models used. Black lines show the output of GAM models. Shaded areas around the line denote the standard error of the model. Adj. R2 denotes the adjusted R^2 of the model. The smooth term p.val denotes the significance of age in the model.



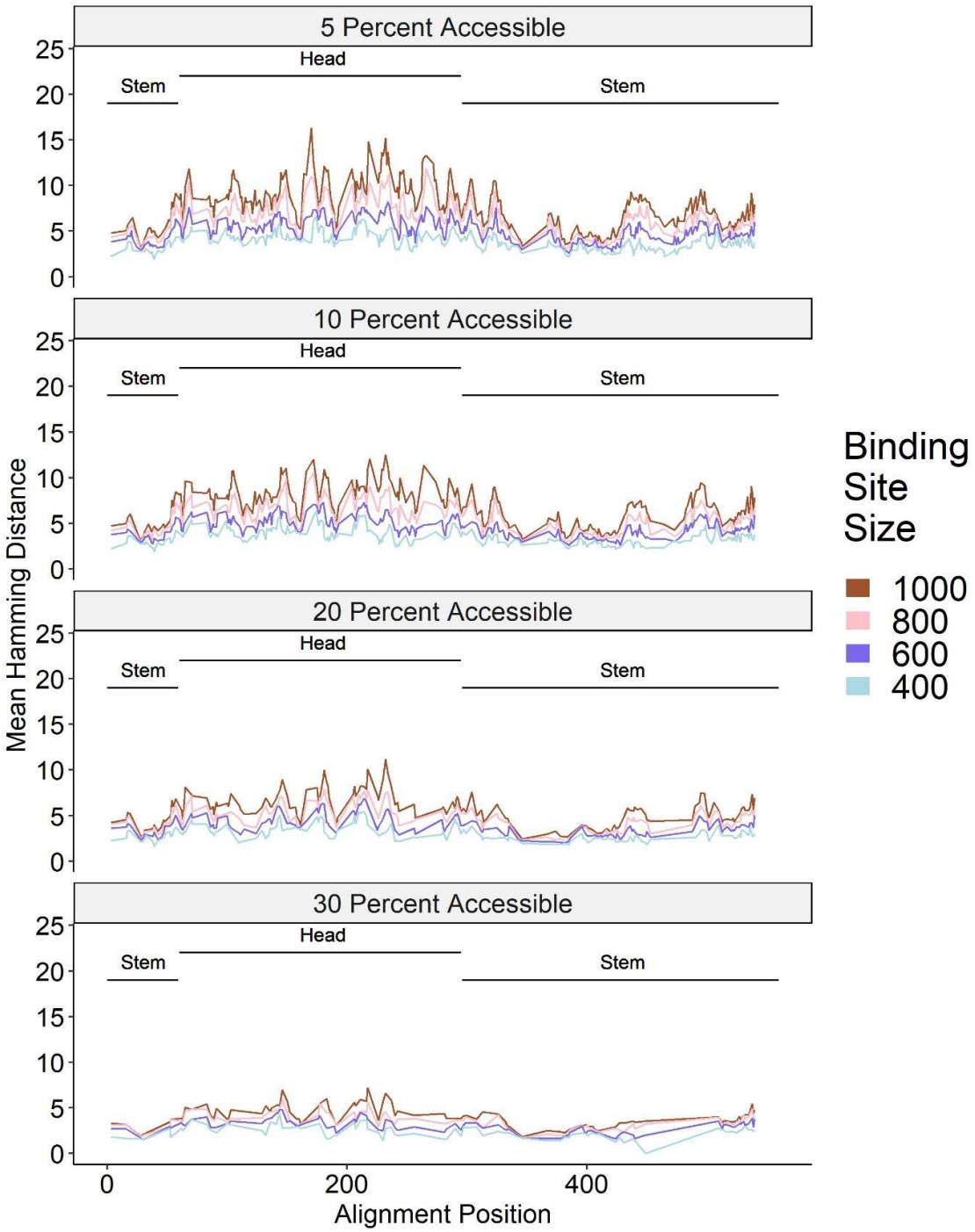
Supplementary Figure 4. The relationship between hatch year and neutralisation strength for pseudotyped influenza viruses run on the wild avian sample set. Samples collected in 2008 are included. Bars denote the seropositivity percentage per hatch year. Facets denote individual pseudotyped viruses. Black lines show the output of logistic regression. Shaded areas around the line denote the standard error of the model. McFadden pr2 denotes the McFadden pseudo-R^2 of the model. Age odds effect denotes the significance of age in the model.



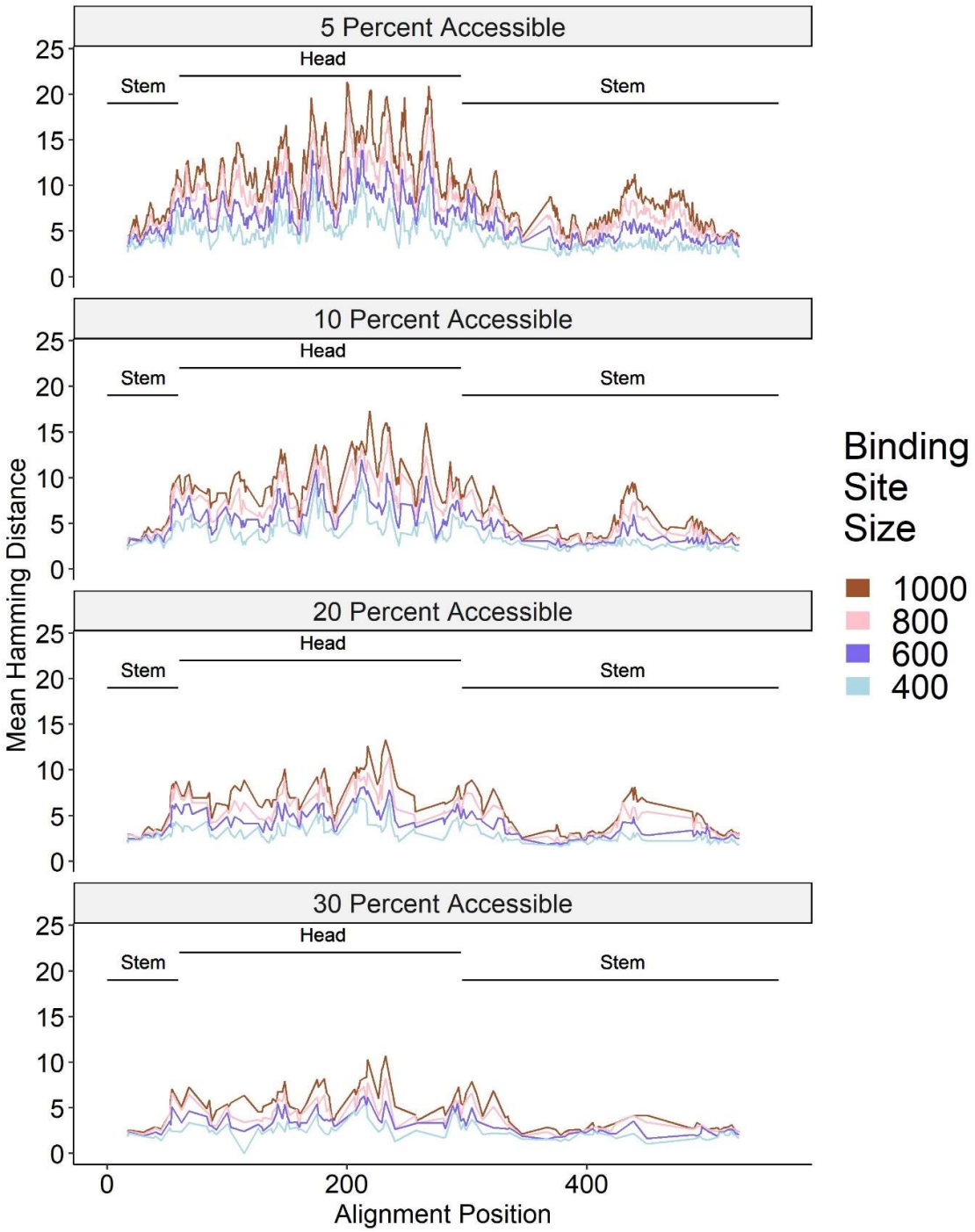
Supplementary Figure 5. The relationship between hatch year and neutralisation strength for pseudotyped influenza virus run on the wild avian sample set. Samples collected in 2008 are included. Points denote neutralisation strength per hatch year. Facets denote individual pseudotypes and weights of the models used. Black lines show the output of linear models. Shaded areas around the line denote the standard error of the model. Adj. R2 denotes the adjusted R^2 of the model. Gradient p.val denotes the significance of age in the model.



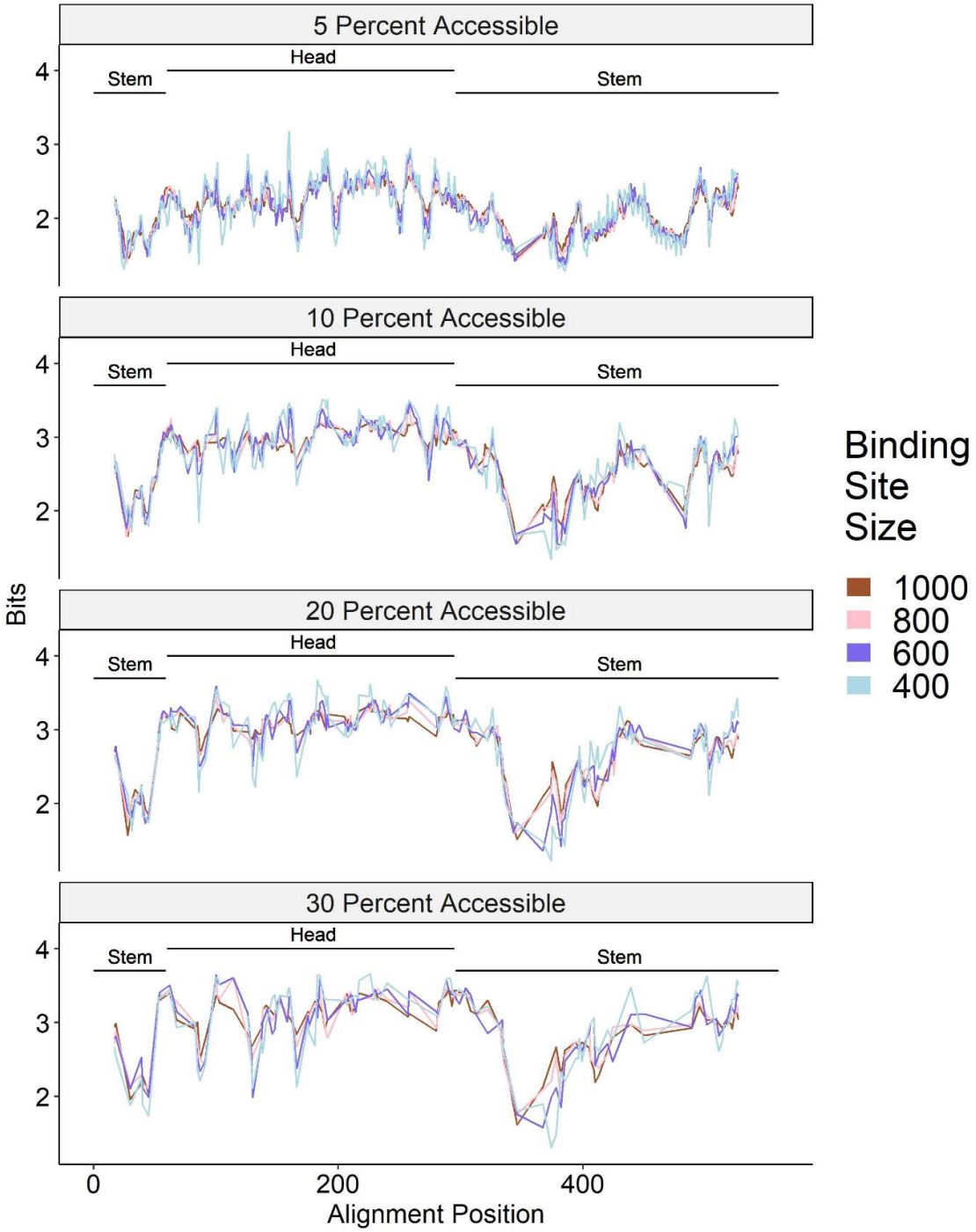
Supplementary Figure 6. The relationship between hatch year and neutralisation strength for pseudotypes run on the wild avian sample set. Samples collected in 2008 are not included. Points denote neutralisation strength per hatch year. Facets denote individual pseudotyped viruses and weights of the models used. Black lines show the output of linear models. Shaded areas around the line denote the standard error of the model. Adj. R2 denotes the adjusted R^2 of the model. Gradient p.val denotes the significance of age in the model.



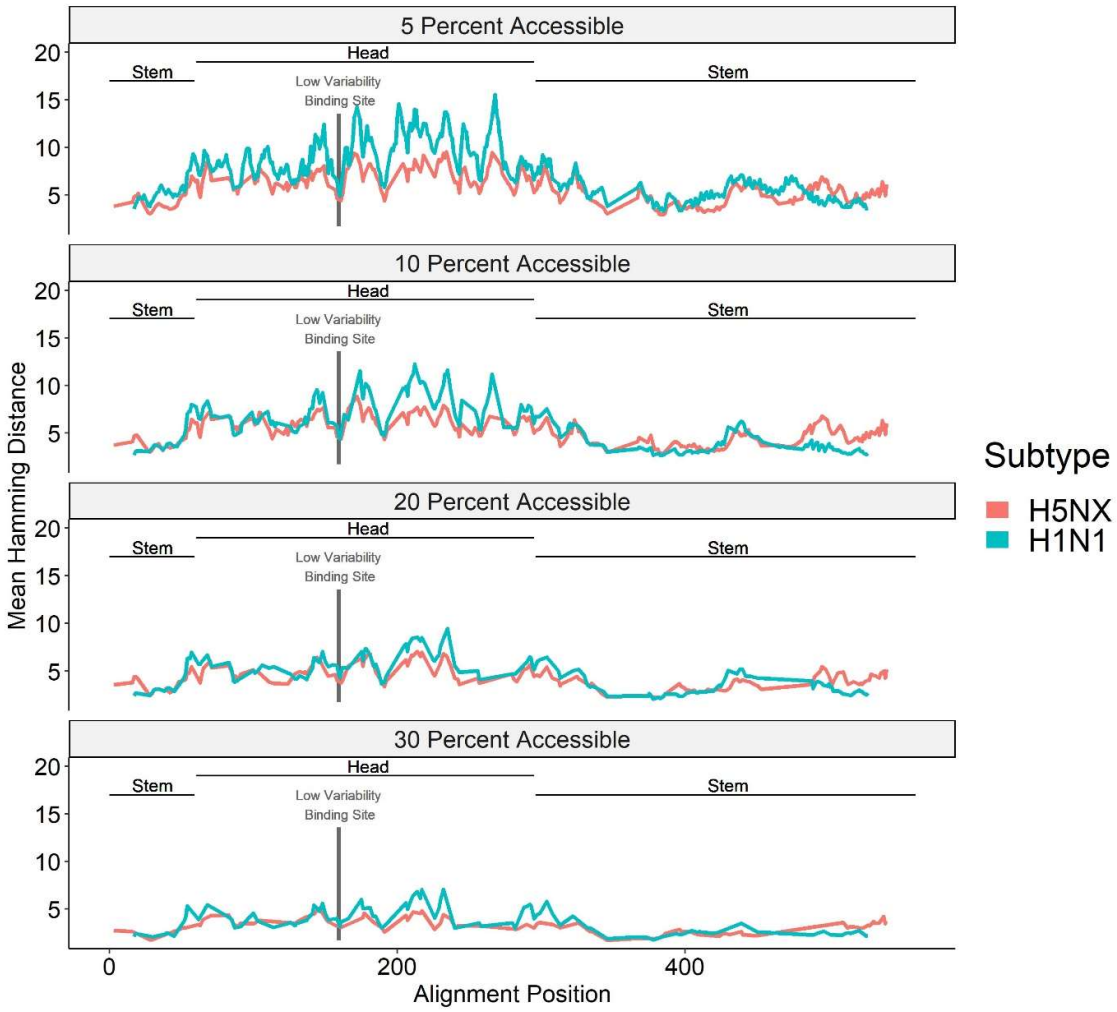
Supplementary Figure 7. Variability analysis of the H5 HA. Supplementary Figure 7 shows the output from the bioinformatic structural variability pipeline for avian H5 HA sequences mapped to the 5e2y crystal structure. The x-axis is the numbering from a structural alignment between H1N1 and H5Nx. The x-axis denotes the position in the alignment which occupies the central position of the predicted ABS. The HA head and stem are labelled above. The y-axis denotes variability in the units of mean Hamming distance. The 4 lines on each graph denote epitopes of size 400 Å², 600 Å², 800 Å² and 1000 Å² by the line colours. Facets denote the threshold of residue solvent accessibility.



Supplementary Figure 8. Variability analysis of the H1N1 HA. Supplementary Figure 8 shows the output from the bioinformatic structural variability pipeline for avian H1N1 HA sequences mapped to the 6ona crystal structure. The x-axis is the numbering from a structural alignment between H1N1 and H5Nx. The x-axis denotes the position in the alignment which occupies the central position of the predicted ABS. The HA head and stem are labelled above. The y-axis denotes variability in the units of mean Hamming distance. The 4 lines on each graph denote epitopes of size 400 Å², 600 Å², 800 Å² and 1000 Å² by the line colours. Facets denote the threshold of residue solvent accessibility.


5 Percent Accessible
10 Percent Accessible
20 Percent Accessible
30 Percent Accessible
Head
Stem
Stem
Bits
Alignment Position
0
200
400
Binding Site Size
1000
800
600
400

Supplementary Figure 9. Mutational tolerance analysis for putative ABSs in the H1N1 HA. This shows the output from the bioinformatic structural variability pipeline for human H1N1 HA using mutational tolerance data from [241]. The x-axis is the numbering from a structural alignment between H1N1 and H5Nx. The x-axis denotes the position in the alignment which occupies the central position of the predicted ABS. The HA head and stem are labelled above. The y-axis denotes variability in the units of mean Hamming distance and bits respectively. The 4 lines on each graph denote epitopes of size 400 Å², 600 Å², 800 Å² and 1000 Å² by the line colours. Facets denote the threshold of residue solvent accessibility.



Supplementary Figure 10. Variability analysis of the H5Nx and H1N1 HAs was used to identify a putative ELV potentially shared between the two subtypes. Figure a shows the output from the bioinformatic structural variability pipeline for avian H5 HA sequences mapped to the 5e2y crystal structure and human H1N1 HA sequences mapped to the 6ona crystal structure. The x-axis is the numbering from a structural alignment between H1N1 and H5Nx. The x-axis denotes the position in the alignment which occupies the central position of the predicted ABS. The HA head and stem are labelled above. The y-axis denotes the mean variability between all epitope sizes in the units of mean Hamming distance. Figures b and c show the putative ELV, of 600 Å in size, centred on structural alignment position 159 for H5Nx and H1N1 respectively. Facets denote the threshold of residue solvent accessibility.

15. Bibliography

[1] A. T. Harding and N. S. Heaton, “Efforts to Improve the Seasonal Influenza Vaccine,” *Vaccines*, vol. 6, no. 2, p. 19, Mar. 2018.

[2] P. Saunders-Hastings and D. Krewski, “Reviewing the History of Pandemic Influenza: Understanding Patterns of Emergence and Transmission,” *Pathogens*, vol. 5, no. 4, p. 66, Dec. 2016.

[3] “COVID-19 deaths | WHO COVID-19 dashboard.” [Online]. Available: https://data.who.int/dashboards/covid19/deaths?n=c. [Accessed: 25-Mar-2024].

[4] M. Luo, “Influenza Virus Entry,” *Viral Mol. Mach.*, vol. 726, p. 201, 2012.

[5] S. S. H. Huang *et al.*, “Immunity toward H1N1 influenza hemagglutinin of historical and contemporary strains suggests protection and vaccine failure,” *Sci. Rep.*, vol. 3, no. 1, p. 1698, Dec. 2013.

[6] E. J. Erbelding *et al.*, “A Universal Influenza Vaccine: The Strategic Plan for the National Institute of Allergy and Infectious Diseases,” *J. Infect. Dis.*, vol. 218, no. 3, pp. 347–354, Jul. 2018.

[7] B. N. Corder, B. L. Bullard, G. A. Poland, and E. A. Weaver, “A Decade in Review: A Systematic Review of Universal Influenza Vaccines in Clinical Trials during the 2010 Decade,” *Viruses*, vol. 12, no. 10, Oct. 2020.

[8] I. Kosik and J. W. Yewdell, “Influenza Hemagglutinin and Neuraminidase: Yin–Yang Proteins Coevolving to Thwart Immunity,” *Viruses 2019, Vol. 11, Page 346*, vol. 11, no. 4, p. 346, Apr. 2019.

[9] T. Leiding, J. Wang, J. Martinsson, W. F. DeGrado, and S. P. Årsköld, “Proton and cation

transport activity of the M2 proton channel from influenza A virus," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 107, no. 35, pp. 15409–15414, Aug. 2010.

[10] M. Schotsaert *et al.*, "Long-Lasting Cross-Protection Against Influenza A by Neuraminidase and M2e-based immunization strategies," *Sci. Rep.*, vol. 6, Apr. 2016.

[11] E. Alvarado-Facundo, Y. Gao, R. M. Ribas-Aparicio, A. Jiménez-Alberto, C. D. Weiss, and W. Wang, "Influenza Virus M2 Protein Ion Channel Activity Helps To Maintain Pandemic 2009 H1N1 Virus Hemagglutinin Fusion Competence during Transport to the Cell Surface," *J. Virol.*, vol. 89, no. 4, pp. 1975–1985, Feb. 2015.

[12] R. M. Pielak and J. J. Chou, "Influenza M2 proton channels," *Biochimica et Biophysica Acta - Biomembranes*, vol. 1808, no. 2. NIH Public Access, pp. 522–529, Feb-2011.

[13] T. Noda, "Native Morphology of Influenza Virions," *Front. Microbiol.*, vol. 2, no. JAN, 2011.

[14] P. W. CHOPPIN, J. S. MURPHY, and I. TAMM, "Studies of two kinds of virus particles which comprise influenza A2 virus strains. III. Morphological characteristics: independence to morphological and functional traits," *J. Exp. Med.*, vol. 112, no. 5, pp. 945–952, 1960.

[15] C. M. Chu, I. M. Dawson, and W. J. Elford, "Filamentous forms associated with newly isolated influenza virus," *Lancet (London, England)*, vol. 1, no. 6554, Apr. 1949.

[16] K. Tanaka, A. Mitsushima, Y. Kashima, T. Nakadera, and H. Osatake, "Application of an ultrahigh-resolution scanning electron microscope (UHS-T1) to biological specimens," *J. Electron Microsc. Tech.*, vol. 12, no. 2, pp. 146–154, 1989.

[17] K. F. Shortridge *et al.*, "Characterization of Avian H5N1 Influenza Viruses from Poultry in Hong Kong," 1998.

[18] M. Koutsakos, A. K. Wheatley, K. Laurie, S. J. Kent, and S. Rockman, "Influenza lineage extinction during the COVID-19 pandemic?," *Nat. Rev. Microbiol.*, vol. 19, no. 12, pp. 741–

742, Dec. 2021.

[19] Z. Vajo and P. Torzsa, “Extinction of the Influenza B Yamagata Line during the COVID Pandemic—Implications for Vaccine Composition,” *Viruses*, vol. 14, no. 8, Aug. 2022.

[20] “Types of Influenza Viruses | CDC.” [Online]. Available: https://www.cdc.gov/flu/about/viruses/types.htm. [Accessed: 05-Jan-2022].

[21] M. Rozo and G. K. Gronvall, “The reemergent 1977 H1N1 strain and the gain-of-function debate,” *MBio*, vol. 6, no. 4, Aug. 2015.

[22] R. G. Webster, W. J. Bean, O. T. Gorman, T. M. Chambers, and Y. Kawaoka, “Evolution and ecology of influenza A viruses.,” *Microbiol. Rev.*, vol. 56, no. 1, p. 152, 1992.

[23] E. S. M. Abdelwhab, J. Veits, and T. C. Mettenleiter, “Genetic changes that accompanied shifts of low pathogenic avian influenza viruses toward higher pathogenicity in poultry,” *Virulence*, vol. 4, no. 6, p. 441, Aug. 2013.

[24] Y. N. Lee *et al.*, “Genetic characteristics and pathogenesis of H5 low pathogenic avian influenza viruses from wild birds and domestic ducks in South Korea,” *Sci. Rep.*, vol. 10, no. 1, Dec. 2020.

[25] A. Stieneke-Gröber *et al.*, “Influenza virus hemagglutinin with multibasic cleavage site is activated by furin, a subtilisin-like endoprotease.,” *EMBO J.*, vol. 11, no. 7, p. 2407, Jul. 1992.

[26] J. Bogs *et al.*, “Highly Pathogenic H5N1 Influenza Viruses Carry Virulence Determinants beyond the Polybasic Hemagglutinin Cleavage Site,” *PLoS One*, vol. 5, no. 7, 2010.

[27] W. Garten and H. D. Klenk, “Understanding influenza virus pathogenicity,” *Trends Microbiol.*, vol. 7, no. 3, pp. 99–100, Mar. 1999.

[28] Y. Liang, “Pathogenicity and virulence of influenza,” *Virulence*, vol. 14, no. 1, 2023.

[29] B. S. Hamilton, D. W. J. Gludish, and G. R. Whittaker, “Cleavage Activation of the Human-

Adapted Influenza Virus Subtypes by Matriptase Reveals both Subtype and Strain Specificities," *J. Virol.*, vol. 86, no. 19, p. 10579, Oct. 2012.

[30] L. Clark and J. Hall, "Avian Influenza in Wild Birds: Status as Reservoirs, and Risks to Humans and Agriculture," *Ornithol. Monogr.*, no. 60, pp. 3–29, Jan. 2006.

[31] GISRS, "Updated unified nomenclature system for the highly pathogenic H5N1 avian influenza viruses Background and Objectives," 2011.

[32] R. O. Donis *et al.*, "Toward a Unified Nomenclature System for Highly Pathogenic Avian Influenza Virus (H5N1)," *Emerg. Infect. Dis.*, vol. 14, no. 7, p. e1, Jul. 2008.

[33] Q. M. Le *et al.*, "Pathogenicity of highly pathogenic avian H5N1 influenza A viruses isolated from humans between 2003 and 2008 in northern Vietnam," *J. Gen. Virol.*, vol. 91, no. Pt 10, p. 2485, Oct. 2010.

[34] P. K. S. Chan, "Outbreak of avian influenza A(H5N1) virus infection in Hong Kong in 1997," *Clin. Infect. Dis.*, vol. 34 Suppl 2, no. SUPPL. 2, May 2002.

[35] P. Venkatesan, "Avian influenza spillover into mammals," *The Lancet Microbe*, vol. 4, no. 7, p. e492, Jul. 2023.

[36] S. M. Ghafoori *et al.*, "Structural characterisation of hemagglutinin from seven Influenza A H1N1 strains reveal diversity in the C05 antibody recognition site," *Sci. Reports 2023 131*, vol. 13, no. 1, pp. 1–9, Apr. 2023.

[37] Z. Zheng, S. S. Paul, X. Mo, Y. R. A. Yuan, and Y. J. Tan, "The Vestigial Esterase Domain of Haemagglutinin of H5N1 Avian Influenza A Virus: Antigenicity and Contribution to Viral Pathogenesis," *Vaccines*, vol. 6, no. 3, Sep. 2018.

[38] R. Hai *et al.*, "Influenza Viruses Expressing Chimeric Hemagglutinins: Globular Head and Stalk Domains Derived from Different Subtypes," *J. Virol.*, vol. 86, no. 10, p. 5774, May 2012.

[39] E. Kirkpatrick, X. Qiu, P. C. Wilson, J. Bahl, and F. Krammer, “The influenza virus hemagglutinin head evolves faster than the stalk domain,” *Sci. Rep.*, vol. 8, no. 1, Dec. 2018.

[40] J. Steel *et al.*, “Influenza Virus Vaccine Based on the Conserved Hemagglutinin Stalk Domain,” *MBio*, vol. 1, no. 1, 2010.

[41] Y. Lu, J. P. Welsh, and J. R. Swartz, “Production and stabilization of the trimeric influenza hemagglutinin stem domain for potentially broadly protective influenza vaccines,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 111, no. 1, pp. 125–130, Jan. 2014.

[42] T. Samji, “Influenza A: Understanding the Viral Life Cycle,” *Yale J. Biol. Med.*, vol. 82, no. 4, p. 153, Dec. 2009.

[43] D. Dou, R. Revol, H. Östbye, H. Wang, and R. Daniels, “Influenza A Virus Cell Entry, Replication, Virion Assembly and Movement,” *Front. Immunol.*, vol. 9, no. JUL, Jul. 2018.

[44] D. J. Benton, S. J. Gamblin, P. B. Rosenthal, and J. J. Skehel, “Structural transitions in influenza haemagglutinin at membrane fusion pH,” *Nat. 2020 5837814*, vol. 583, no. 7814, pp. 150–153, May 2020.

[45] D. J. Benton *et al.*, “Influenza hemagglutinin membrane anchor,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 115, no. 40, pp. 10112–10117, Oct. 2018.

[46] D. C. Wiley and J. J. Skehel, “The structure and function of the hemagglutinin membrane glycoprotein of influenza virus,” *Annu. Rev. Biochem.*, vol. 56, pp. 365–394, 1987.

[47] N. Sriwilaijaroen and Y. Suzuki, “Molecular basis of the structure and function of H1 hemagglutinin of influenza virus.,” *Proc. Jpn. Acad. Ser. B. Phys. Biol. Sci.*, vol. 88, no. 6, pp. 226–49, 2012.

[48] C. S. Copeland, R. W. Doms, E. M. Bolzau, R. G. Webster, and A. Helenius, “Assembly of influenza hemagglutinin trimers and its role in intracellular transport,” *J. Cell Biol.*, vol. 103,

no. 4, pp. 1179–1191, 1986.

[49] T. H. Bugge, T. M. Antalis, and Q. Wu, “Type II Transmembrane Serine Proteases,” *J. Biol. Chem.*, vol. 284, no. 35, p. 23177, Aug. 2009.

[50] F. Krammer, “The human antibody response to influenza A virus infection and vaccination,” *Nature Reviews Immunology*, vol. 19, no. 6. Nature Publishing Group, pp. 383–397, 01-Jun-2019.

[51] I. Zanella, M. Degli Antoni, V. Marchese, F. Castelli, and E. Quiros-Roldan, “Non-neutralizing antibodies: Deleterious or propitious during SARS-CoV-2 infection?,” *Int. Immunopharmacol.*, vol. 110, p. 108943, Sep. 2022.

[52] K. Miyauchi *et al.*, “Influenza virus infection expands the breadth of antibody responses through IL-4 signalling in B cells,” *Nat. Commun. 2021 121*, vol. 12, no. 1, pp. 1–15, Jun. 2021.

[53] Z. N. Li *et al.*, “IgM, IgG, and IgA Antibody Responses to Influenza A(H1N1)pdm09 Hemagglutinin in Infected Persons during the First Wave of the 2009 Pandemic in the United States,” *Clin. Vaccine Immunol.*, vol. 21, no. 8, p. 1054, 2014.

[54] S. J. Zost, N. C. Wu, S. E. Hensley, and I. A. Wilson, “Immunodominance and Antigenic Variation of Influenza Virus Hemagglutinin: Implications for Design of Universal Vaccine Immunogens,” *J. Infect. Dis.*, vol. 219, no. Suppl 1, p. S38, Apr. 2019.

[55] C. C. Lee *et al.*, “An Effective Neutralizing Antibody Against Influenza Virus H1N1 from Human B Cells,” *Sci. Reports 2019 91*, vol. 9, no. 1, pp. 1–11, Mar. 2019.

[56] N. Chai *et al.*, “Two Escape Mechanisms of Influenza A Virus to a Broadly Neutralizing Stalk-Binding Antibody,” *PLoS Pathog.*, vol. 12, no. 6, Jun. 2016.

[57] M. Baz *et al.*, “Seroconversion to Seasonal Influenza Viruses after A(H1N1)pdm09 Virus Infection, Quebec, Canada,” *Emerg. Infect. Dis.*, vol. 18, no. 7, p. 1132, Jul. 2012.

[58] R. Nachbagauer *et al.*, "Defining the antibody cross-reactome against the influenza virus surface glycoproteins," *Nat. Immunol.*, vol. 18, no. 4, p. 464, Mar. 2017.

[59] T. J. Wohlbold and F. Krammer, "In the Shadow of Hemagglutinin: A Growing Interest in Influenza Viral Neuraminidase and Its Role as a Vaccine Antigen," *Viruses*, vol. 6, no. 6, p. 2465, Jun. 2014.

[60] D. C. Ekiert *et al.*, "Antibody recognition of a highly conserved influenza virus epitope," *Science*, vol. 324, no. 5924, pp. 246–251, Apr. 2009.

[61] A. Solomon and D. T. Weiss, "Structural and functional properties of human lambda-light-chain variable-region subgroups.," *Clin. Diagn. Lab. Immunol.*, vol. 2, no. 4, p. 387, 1995.

[62] J. Charles A Janeway, P. Travers, M. Walport, and M. J. Shlomchik, "The structure of a typical antibody molecule," 2001.

[63] G. Vidarsson, G. Dekkers, and T. Rispens, "IgG Subclasses and Allotypes: From Structure to Effector Functions," *Front. Immunol.*, vol. 5, no. OCT, 2014.

[64] M. L. Chiu, D. R. Goulet, A. Teplyakov, and G. L. Gilliland, "Antibody Structure and Function: The Basis for Engineering Therapeutics," *Antibodies*, vol. 8, no. 4, Dec. 2019.

[65] M. Terajima *et al.*, "Complement-dependent lysis of influenza a virus-infected cells by broadly cross-reactive human monoclonal antibodies," *J. Virol.*, vol. 85, no. 24, pp. 13463–13467, Dec. 2011.

[66] D. J. DiLillo, P. Palese, P. C. Wilson, and J. V. Ravetch, "Broadly neutralizing anti-influenza antibodies require Fc receptor engagement for in vivo protection," *J. Clin. Invest.*, vol. 126, no. 2, p. 605, Feb. 2016.

[67] M. H. Van Regenmortel, "The concept and operational definition of protein epitopes," *Philos. Trans. R. Soc. Lond. B. Biol. Sci.*, vol. 323, no. 1217, pp. 451–466, 1989.

[68] T. Kurosaki, K. Kometani, and W. Ise, “Memory B cells,” *Nat. Rev. Immunol. 2015 153*, vol. 15, no. 3, pp. 149–159, Feb. 2015.

[69] M. Akkaya, K. Kwak, and S. K. Pierce, “B cell memory: building two walls of protection against pathogens,” *Nat. Rev. Immunol. 2019 204*, vol. 20, no. 4, pp. 229–238, Dec. 2019.

[70] R. R. Ramiscal and C. G. Vinuesa, “T-cell subsets in the germinal center,” *Immunol. Rev.*, vol. 252, no. 1, pp. 146–155, Mar. 2013.

[71] S. L. Nutt, P. D. Hodgkin, D. M. Tarlinton, and L. M. Corcoran, “The generation of antibody-secreting plasma cells,” *Nat. Rev. Immunol. 2015 153*, vol. 15, no. 3, pp. 160–171, Feb. 2015.

[72] M. C. Julian, L. Li, S. Garde, R. Wilen, and P. M. Tessier, “Efficient affinity maturation of antibody variable domains requires co-selection of compensatory mutations to maintain thermodynamic stability.,” *Sci. Rep.*, vol. 7, p. 45259, 2017.

[73] N. Fazilleau, L. Mark, L. J. McHeyzer-Williams, and M. G. McHeyzer-Williams, “Follicular helper T cells: lineage and location.,” *Immunity*, vol. 30, no. 3, pp. 324–35, Mar. 2009.

[74] D. McKean, K. Huppi, M. Bell, L. Staudt, W. Gerhard, and M. Weigert, “Generation of antibody diversity in the immune response of BALB/c mice to influenza virus hemagglutinin,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 81, no. 10, pp. 3180–3184, 1984.

[75] C. Jiao, B. Wang, P. Chen, Y. Jiang, and J. Liu, “Analysis of the conserved protective epitopes of hemagglutinin on influenza A viruses,” *Front. Immunol.*, vol. 14, p. 1086297, Feb. 2023.

[76] Y. Adachi *et al.*, “Exposure of an occluded hemagglutinin epitope drives selection of a class of cross-protective influenza antibodies,” *Nat. Commun. 2019 101*, vol. 10, no. 1, pp. 1–13, Aug. 2019.

[77] J. J. Guthmiller *et al.*, “Broadly neutralizing antibodies target a haemagglutinin anchor epitope,” *Nat. 2021 6027896*, vol. 602, no. 7896, pp. 314–320, Dec. 2021.

[78] L. Du *et al.*, "Identification and structural characterization of a broadly neutralizing antibody targeting a novel conserved epitope on the influenza virus H5N1 hemagglutinin," *J. Virol.*, vol. 87, no. 4, pp. 2215–2225, Feb. 2013.

[79] Y. Iba *et al.*, "Conserved Neutralizing Epitope at Globular Head of Hemagglutinin in H3N2 Influenza Viruses," *J. Virol.*, vol. 88, no. 13, p. 7130, Jul. 2014.

[80] G. S. Tan *et al.*, "Broadly-Reactive Neutralizing and Non-neutralizing Antibodies Directed against the H7 Influenza Virus Hemagglutinin Reveal Divergent Mechanisms of Protection," *PLoS Pathog.*, vol. 12, no. 4, Apr. 2016.

[81] S. Bangaru *et al.*, "A Site of Vulnerability on the Influenza Virus Hemagglutinin Head Domain Trimer Interface," *Cell*, vol. 177, no. 5, pp. 1136-1152.e18, May 2019.

[82] C. Barbey-Martin *et al.*, "An antibody that prevents the hemagglutinin low pH fusogenic transition," *Virology*, vol. 294, no. 1, pp. 70–74, 2002.

[83] P. S. Daniels *et al.*, "The receptor-binding and membrane-fusion properties of influenza virus variants selected using anti-haemagglutinin monoclonal antibodies," *EMBO J.*, vol. 6, no. 5, pp. 1459–1465, 1987.

[84] D. C. Ekiert *et al.*, "Cross-neutralization of influenza A viruses mediated by a single antibody loop," *Nature*, vol. 489, no. 7417, pp. 526–532, Sep. 2012.

[85] R. Yoshida *et al.*, "Cross-Protective Potential of a Novel Monoclonal Antibody Directed against Antigenic Site B of the Hemagglutinin of Influenza A Viruses," *PLoS Pathog.*, vol. 5, no. 3, p. 1000350, Mar. 2009.

[86] Y. Fu *et al.*, "A broadly neutralizing anti-influenza antibody reveals ongoing capacity of haemagglutinin-specific memory B cells to evolve," *Nat. Commun. 2016 71*, vol. 7, no. 1, pp. 1–13, Sep. 2016.

[87] D. Corti *et al.*, “A neutralizing antibody selected from plasma cells that binds to group 1 and group 2 influenza A hemagglutinins,” *Science*, vol. 333, no. 6044, pp. 850–856, Aug. 2011.

[88] C. Dreyfus *et al.*, “Highly conserved protective epitopes on influenza B viruses,” *Science*, vol. 337, no. 6100, pp. 1343–1348, Sep. 2012.

[89] “On the Doctrine of Original Antigenic Sin on JSTOR.” [Online]. Available: https://www.jstor.org/stable/985534?seq=1#metadata_info_tab_contents. [Accessed: 05-Jan-2022].

[90] A. S. Monto, R. E. Malosh, J. G. Petrie, and E. T. Martin, “The Doctrine of Original Antigenic Sin: Separating Good From Evil,” *J. Infect. Dis.*, vol. 215, no. 12, p. 1782, Jun. 2017.

[91] A. L. McNaughton *et al.*, “Fatal COVID-19 outcomes are associated with an antibody response targeting epitopes shared with endemic coronaviruses,” *JCI Insight*, vol. 7, no. 13, Jul. 2022.

[92] J. M. Fonville *et al.*, “Antibody landscapes after influenza virus infection or vaccination,” *Science*, vol. 346, no. 6212, pp. 996–1000, Nov. 2014.

[93] P. Fine, K. Eames, and D. L. Heymann, “‘Herd Immunity’: A Rough Guide,” *Clin. Infect. Dis.*, vol. 52, no. 7, pp. 911–916, Apr. 2011.

[94] T. J. John and R. Samuel, “Herd immunity and herd effect: new insights and definitions,” *Eur. J. Epidemiol.*, vol. 16, no. 7, pp. 601–606, 2000.

[95] A. W. Hedrich, “Monthly estimates of the child population ‘susceptible’’ to measles, 1900-1931, Baltimore, MD,’” *Am. J. Epidemiol.*, vol. 17, no. 3, pp. 613–636, May 1933.

[96] G. MACDONALD, “The analysis of equilibrium in malaria.,” *Trop. Dis. Bull.*, vol. 49, no. 9, pp. 813–829, Sep. 1952.

[97] S. Boivin, S. Cusack, R. W. H. Ruigrok, and D. J. Hart, “Influenza A Virus Polymerase: Structural Insights into Replication and Host Adaptation Mechanisms,” *J. Biol. Chem.*, vol. 285, no. 37, p.

28411, Sep. 2010.

[98] P. S. Wikramaratna, O. G. Pybus, and S. Gupta, “Contact between bird species of different lifespans can promote the emergence of highly pathogenic avian influenza strains,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 111, no. 29, pp. 10767–10772, Jul. 2014.

[99] M. F. Boni, “Vaccination and antigenic drift in influenza.,” *Vaccine*, vol. 26 Suppl 3, no. Suppl 3, pp. C8-14, Jul. 2008.

[100] J. M. Fonville *et al.*, “Antigenic Maps of Influenza A(H3N2) Produced With Human Antisera Obtained After Primary Infection,” *J. Infect. Dis.*, vol. 213, no. 1, pp. 31–38, Jan. 2016.

[101] J. M. Fonville *et al.*, “Antibody landscapes after influenza virus infection or vaccination,” *Science (80-.).*, vol. 346, no. 6212, pp. 996–1000, Nov. 2014.

[102] M. B. Doud, S. E. Hensley, and J. D. Bloom, “Complete mapping of viral escape from neutralizing antibodies,” *PLOS Pathog.*, vol. 13, no. 3, p. e1006271, Mar. 2017.

[103] D. J. Smith *et al.*, “Mapping the antigenic and genetic evolution of influenza virus.,” *Science*, vol. 305, no. 5682, pp. 371–6, Jul. 2004.

[104] E. Ghedin *et al.*, “Large-scale sequencing of human influenza reveals the dynamic nature of viral genome evolution,” *Nature*, vol. 437, no. 7062, pp. 1162–1166, Oct. 2005.

[105] M. B. Doud, 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 91*, vol. 9, no. 1, pp. 1–12, Apr. 2018.

[106] Y. Matsuzaki *et al.*, “Epitope mapping of the hemagglutinin molecule of A/(H1N1)pdm09 influenza virus by using monoclonal antibody escape mutants,” *J. Virol.*, vol. 88, no. 21, pp. 12364–12373, Nov. 2014.

[107] A. J. Caton, G. G. Brownlee, J. W. Yewdell, and W. Gerhard, “The antigenic structure of the

influenza virus A/PR/8/34 hemagglutinin (H1 subtype)," *Cell*, vol. 31, no. 2 PART 1, pp. 417–427, Dec. 1982.

[108] D. Höper, D. Kalthoff, B. Hoffmann, and M. Beer, "Highly Pathogenic Avian Influenza Virus Subtype H5N1 Escaping Neutralization: More than HA Variation," *J. Virol.*, vol. 86, no. 3, p. 1394, Feb. 2012.

[109] V. R. S. K. Duvvuri, B. Duvvuri, W. R. Cuff, G. E. Wu, and J. Wu, "Role of Positive Selection Pressure on the Evolution of H5N1 Hemagglutinin," *Genomics. Proteomics Bioinformatics*, vol. 7, no. 1–2, p. 47, Jun. 2009.

[110] I. Sitaras, D. Kalthoff, M. Beer, B. Peeters, and M. C. M. De Jong, "Immune Escape Mutants of Highly Pathogenic Avian Influenza H5N1 Selected Using Polyclonal Sera: Identification of Key Amino Acids in the HA Protein," *PLoS One*, vol. 9, no. 2, p. 84628, Feb. 2014.

[111] S. Khurana *et al.*, "Antigenic Fingerprinting of H5N1 Avian Influenza Using Convalescent Sera and Monoclonal Antibodies Reveals Potential Vaccine and Diagnostic Targets," *PLOS Med.*, vol. 6, no. 4, p. e1000049, 2009.

[112] T. Ohkura *et al.*, "Epitope mapping of neutralizing monoclonal antibody in avian influenza A H5N1 virus hemagglutinin," *Biochem. Biophys. Res. Commun.*, vol. 418, no. 1, pp. 38–43, Feb. 2012.

[113] G. Cattoli *et al.*, "Antigenic Drift in H5N1 Avian Influenza Virus in Poultry Is Driven by Mutations in Major Antigenic Sites of the Hemagglutinin Molecule Analogous to Those for Human Influenza Virus," *J. Virol.*, vol. 85, no. 17, p. 8718, Sep. 2011.

[114] Y. Zhang *et al.*, "Key Amino Acid Residues That Determine the Antigenic Properties of Highly Pathogenic H5 Influenza Viruses Bearing the Clade 2.3.4.4 Hemagglutinin Gene," *Viruses 2023, Vol. 15, Page 2249*, vol. 15, no. 11, p. 2249, Nov. 2023.

[115] B. F. Koel *et al.*, "Antigenic Variation of Clade 2.1 H5N1 Virus Is Determined by a Few Amino

Acid Substitutions Immediately Adjacent to the Receptor Binding Site," *MBio*, vol. 5, no. 3, Jun. 2014.

[116] M. Philpott, B. C. Easterday, and V. S. Hinshaw, "Neutralizing epitopes of the H5 hemagglutinin from a virulent avian influenza virus and their relationship to pathogenicity," *J. Virol.*, vol. 63, no. 8, pp. 3453–3458, Aug. 1989.

[117] M. Philpott, C. Hioe, M. Sheerar, and V. S. Hinshaw, "Hemagglutinin mutations related to attenuation and altered cell tropism of a virulent avian influenza A virus," *J. Virol.*, vol. 64, no. 6, pp. 2941–2947, Jun. 1990.

[118] H. Kong *et al.*, "Plasticity of the influenza virus H5 HA protein," *MBio*, vol. 12, no. 1, pp. 1–14, 2021.

[119] T. Peacock *et al.*, "Antigenic mapping of an H9N2 avian influenza virus reveals two discrete antigenic sites and a novel mechanism of immune escape," *Sci. Rep.*, vol. 6, Jan. 2016.

[120] F. Zhou *et al.*, "A Triclade DNA Vaccine Designed on the Basis of a Comprehensive Serologic Study Elicits Neutralizing Antibody Responses against All Clades and Subclades of Highly Pathogenic Avian Influenza H5N1 Viruses," *J. Virol.*, vol. 86, no. 12, p. 6970, Jun. 2012.

[121] T. Velkov *et al.*, "The antigenic architecture of the hemagglutinin of influenza H5N1 viruses," *Mol. Immunol.*, vol. 56, no. 4, pp. 705–719, Dec. 2013.

[122] S. L. Linderman *et al.*, "Potential antigenic explanation for atypical H1N1 infections among middle-aged adults during the 2013-2014 influenza season.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 111, no. 44, pp. 15798–803, Nov. 2014.

[123] D. M. Carter *et al.*, "Sequential Seasonal H1N1 Influenza Virus Infections Protect Ferrets against Novel 2009 H1N1 Influenza Virus," *J. Virol.*, vol. 87, no. 3, pp. 1400–1410, Feb. 2013.

[124] A. G. Schmidt *et al.*, "Immunogenic Stimulus for Germline Precursors of Antibodies that

Engage the Influenza Hemagglutinin Receptor-Binding Site," *Cell Rep.*, vol. 13, no. 12, pp. 2842–2850, Dec. 2015.

[125] D. D. Raymond *et al.*, "Conserved epitope on influenza-virus hemagglutinin head defined by a vaccine-induced antibody," *Proc. Natl. Acad. Sci.*, vol. 115, no. 1, pp. 168–173, Jan. 2018.

[126] Y. Li *et al.*, "Immune history shapes specificity of pandemic H1N1 influenza antibody responses," *J. Exp. Med.*, vol. 210, no. 8, pp. 1493–1500, 2013.

[127] S. L. Linderman and S. E. Hensley, "Antibodies with 'Original Antigenic Sin' Properties Are Valuable Components of Secondary Immune Responses to Influenza Viruses," *PLOS Pathog.*, vol. 12, no. 8, p. e1005806, Aug. 2016.

[128] A. K. Harris *et al.*, "Structure and accessibility of HA trimers on intact 2009 H1N1 pandemic influenza virus to stem region-specific neutralizing antibodies," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 110, no. 12, pp. 4592–4597, Mar. 2013.

[129] J. M. Lee *et al.*, "Deep mutational scanning of hemagglutinin helps predict evolutionary fates of human H3N2 influenza variants," *Proc. Natl. Acad. Sci.*, 2018.

[130] K. Y. A. Huang *et al.*, "Focused antibody response to influenza linked to antigenic drift," *J. Clin. Invest.*, vol. 125, no. 7, pp. 2631–2645, Jul. 2015.

[131] K. Suwannakarn *et al.*, "Molecular Evolution of Human H1N1 and H3N2 Influenza A Virus in Thailand, 2006–2009," *PLoS One*, vol. 5, no. 3, 2010.

[132] A. C. Tricco *et al.*, "Comparing influenza vaccine efficacy against mismatched and matched strains: A systematic review and meta-analysis," *BMC Med.*, vol. 11, no. 1, p. 153, Jun. 2013.

[133] H. Fukuyama, R. Shinnakasu, and T. Kurosaki, "Influenza vaccination strategies targeting the hemagglutinin stem region," *Immunol. Rev.*, vol. 296, no. 1, pp. 132–141, Jul. 2020.

[134] C. P. Thompson and U. Obolski, "Influenza vaccination and the 'diversity paradox,'" *Hum.*

Vaccin. Immunother., vol. 14, no. 12, p. 3005, Dec. 2018.

[135] A. Rambaut, O. G. Pybus, M. I. Nelson, C. Viboud, J. K. Taubenberger, and E. C. Holmes, "The genomic and epidemiological dynamics of human influenza A virus," *Nat. 2008 4537195*, vol. 453, no. 7195, pp. 615–619, Apr. 2008.

[136] N. M. Ferguson, A. P. Galvanl, and R. M. Bush, "Ecological and immunological determinants of influenza evolution," *Nature*, vol. 422, no. 6930, pp. 428–433, Mar. 2003.

[137] F. Tria, M. Lässig, L. Peliti, and S. Franz, "A minimal stochastic model for influenza evolution," *J. Stat. Mech. Theory Exp.*, vol. 2005, no. 07, p. P07008, Jul. 2005.

[138] T. Bedford, A. Rambaut, and M. Pascual, "Canalization of the evolutionary trajectory of the human influenza virus," *BMC Biol.*, vol. 10, no. 1, pp. 1–12, Apr. 2012.

[139] K. Koelle, S. Cobey, B. Grenfell, and M. Pascual, "Epochal evolution shapes the phylodynamics of interpandemic influenza A (H3N2) in humans," *Science*, vol. 314, no. 5807, pp. 1898–1903, Dec. 2006.

[140] H. Y. Yuan and K. Koelle, "The evolutionary dynamics of receptor binding avidity in influenza A: a mathematical model for a new antigenic drift hypothesis," *Philos. Trans. R. Soc. B Biol. Sci.*, vol. 368, no. 1614, Mar. 2013.

[141] R. M. Bush, C. A. Bender, K. Subbarao, N. J. Cox, and W. M. Fitch, "Predicting the evolution of human influenza A," *Science*, vol. 286, no. 5446, pp. 1921–1925, Dec. 1999.

[142] D. H. Morris *et al.*, "Asynchrony between virus diversity and antibody selection limits influenza virus evolution," *Elife*, vol. 9, pp. 1–62, Oct. 2020.

[143] T. FRANCIS, "Influenza: the new acquayantance," *Ann. Intern. Med.*, vol. 39, no. 2, pp. 203–221, Aug. 1953.

[144] M. Recker, O. G. Pybus, S. Nee, and S. Gupta, "The generation of influenza outbreaks by a

network of host immune responses against a limited set of antigenic types," *Proc. Natl. Acad. Sci.*, vol. 104, no. 18, pp. 7711–7716, May 2007.

[145] D. Zinder, T. Bedford, S. Gupta, and M. Pascual, "The Roles of Competition and Mutation in Shaping Antigenic and Genetic Diversity in Influenza," *PLoS Pathog.*, vol. 9, no. 1, p. e1003104, Jan. 2013.

[146] P. S. Wikramaratna, M. Sandeman, M. Recker, and S. Gupta, "The antigenic evolution of influenza: drift or thrift?," *Philos. Trans. R. Soc. B Biol. Sci.*, vol. 368, no. 1614, pp. 20120200–20120200, Feb. 2013.

[147] C. P. Thompson *et al.*, "A naturally protective epitope of limited variability as an influenza vaccine target," *Nat. Commun.*, vol. 9, no. 1, p. 3859, Dec. 2018.

[148] S. F. Andrews *et al.*, "Immune history profoundly affects broadly protective B cell responses to influenza," *Sci. Transl. Med.*, vol. 7, no. 316, Dec. 2015.

[149] J. R. R. Whittle *et al.*, "Broadly neutralizing human antibody that recognizes the receptor-binding pocket of influenza virus hemagglutinin," *Proc. Natl. Acad. Sci. USA*, vol. 108, no. (34); 08, 2011, pp. 14216–14221, Sep. 2011.

[150] "Global Influenza Programme." [Online]. Available: https://www.who.int/teams/global-influenza-programme/vaccines/who-recommendations. [Accessed: 09-Mar-2024].

[151] W. E. O'Gorman *et al.*, "The Split Virus Influenza Vaccine rapidly activates immune cells through Fcγ receptors," *Vaccine*, vol. 32, no. 45, pp. 5989–5997, Oct. 2014.

[152] L. Martínez-Sobrido, O. Peersen, and A. Nogales, "Temperature Sensitive Mutations in Influenza A Viral Ribonucleoprotein Complex Responsible for the Attenuation of the Live Attenuated Influenza Vaccine," *Viruses*, vol. 10, no. 10, Oct. 2018.

[153] S. Yamayoshi and Y. Kawaoka, "Current and future influenza vaccines," *Nature Medicine*, vol.

25, no. 2. Nature Publishing Group, pp. 212–220, 01-Feb-2019.

[154] M. T. Osterholm, N. S. Kelley, A. Sommer, and E. A. Belongia, “Efficacy and effectiveness of influenza vaccines: a systematic review and meta-analysis,” *Lancet. Infect. Dis.*, vol. 12, no. 1, pp. 36–44, Jan. 2012.

[155] J. G. Petrie *et al.*, “Modest Waning of Influenza Vaccine Efficacy and Antibody Titers During the 2007–2008 Influenza Season,” *J. Infect. Dis.*, vol. 214, no. 8, pp. 1142–1149, Oct. 2016.

[156] “ACIP votes down use of LAIV for 2016-2017 flu season | CDC Online Newsroom | CDC.” [Online]. Available: https://www.cdc.gov/media/releases/2016/s0622-laiv-flu.html. [Accessed: 11-Feb-2021].

[157] R. Nachbagauer *et al.*, “A universal influenza virus vaccine candidate confers protection against pandemic H1N1 infection in preclinical ferret studies,” *npj Vaccines 2017 21*, vol. 2, no. 1, pp. 1–13, Sep. 2017.

[158] R. Nachbagauer *et al.*, “A chimeric hemagglutinin-based universal influenza virus vaccine approach induces broad and long-lasting immunity in a randomized, placebo-controlled phase I trial,” *Nat. Med. 2020 271*, vol. 27, no. 1, pp. 106–114, Dec. 2020.

[159] R. Nachbagauer *et al.*, “A chimeric haemagglutinin-based influenza split virion vaccine adjuvanted with AS03 induces protective stalk-reactive antibodies in mice,” *npj Vaccines 2016 11*, vol. 1, no. 1, pp. 1–10, Sep. 2016.

[160] A. Amitai *et al.*, “Defining and Manipulating B Cell Immunodominance Hierarchies to Elicit Broadly Neutralizing Antibody Responses against Influenza Virus,” *Cell Syst.*, vol. 11, no. 6, pp. 573-588.e9, Dec. 2020.

[161] S. Sridhar *et al.*, “Cellular immune correlates of protection against symptomatic pandemic influenza,” *Nat. Med.*, vol. 19, no. 10, pp. 1305–1312, Oct. 2013.

[162] A. C. Hayward *et al.*, "Natural T Cell-mediated Protection against Seasonal and Pandemic Influenza. Results of the Flu Watch Cohort Study," *Am. J. Respir. Crit. Care Med.*, vol. 191, no. 12, pp. 1422–1431, Jun. 2015.

[163] Z. Wang *et al.*, "Recovery from severe H7N9 disease is associated with diverse response mechanisms dominated by $CD8^+$ T cells," *Nat. Commun.*, vol. 6, May 2015.

[164] T. M. Wilkinson *et al.*, "Preexisting influenza-specific CD4+ T cells correlate with disease protection against influenza challenge in humans," *Nat. Med.*, vol. 18, no. 2, pp. 274–280, Feb. 2012.

[165] Y. Q. Chen *et al.*, "Influenza Infection in Humans Induces Broadly Cross-Reactive and Protective Neuraminidase-Reactive Antibodies," *Cell*, vol. 173, no. 2, pp. 417-429.e10, Apr. 2018.

[166] M. C. Eichelberger and H. Wan, "Influenza neuraminidase as a vaccine antigen," *Curr. Top. Microbiol. Immunol.*, vol. 386, pp. 275–299, Jun. 2015.

[167] F. Krammer *et al.*, "NAction! how can neuraminidase-based immunity contribute to better influenza virus vaccines?," *MBio*, vol. 9, no. 2, Mar. 2018.

[168] S. B. Morgan *et al.*, "Aerosol Delivery of a Candidate Universal Influenza Vaccine Reduces Viral Load in Pigs Challenged with Pandemic H1N1 Virus," *J. Immunol.*, vol. 196, no. 12, pp. 5014–5023, Jun. 2016.

[169] T. M. Doyle *et al.*, "Universal anti-neuraminidase antibody inhibiting all influenza A subtypes," *Antiviral Res.*, vol. 100, no. 2, pp. 567–574, 2013.

[170] K. H. Kim *et al.*, "Neuraminidase expressing virus-like particle vaccine provides effective cross protection against influenza virus," *Virology*, vol. 535, pp. 179–188, Sep. 2019.

[171] S. L. Zebedee, C. D. Richardson, and R. A. Lamb, "Characterization of the influenza virus M2

integral membrane protein and expression at the infected-cell surface from cloned cDNA.," *J. Virol.*, vol. 56, no. 2, pp. 502–511, 1985.

[172] S. L. Zebedee and R. A. Lamb, "Nucleotide sequences of influenza A virus RNA segment 7: A comparison of five isolates," *Nucleic Acids Research*, vol. 17, no. 7. Oxford University Press, p. 2870, 11-Apr-1989.

[173] T. Ito, O. T. Gorman, Y. Kawaoka, W. J. Bean, and R. G. Webster, "Evolutionary analysis of the influenza A virus M gene with comparison of the M1 and M2 proteins.," *J. Virol.*, vol. 65, no. 10, pp. 5491–5498, 1991.

[174] K. El Bakkouri *et al.*, "Universal Vaccine Based on Ectodomain of Matrix Protein 2 of Influenza A: Fc Receptors and Alveolar Macrophages Mediate Protection," *J. Immunol.*, vol. 186, no. 2, pp. 1022–1031, Jan. 2011.

[175] W. Fiers *et al.*, "M2e-based universal influenza A vaccine," *Vaccine*, vol. 27, no. 45, pp. 6280–6283, Oct. 2009.

[176] D. Mezhenskaya, I. Isakova-Sivak, and L. Rudenko, "M2e-based universal influenza vaccines: A historical overview and new approaches to development," *Journal of Biomedical Science*, vol. 26, no. 1. BioMed Central Ltd., pp. 1–15, 19-Oct-2019.

[177] J. W. Huleatt *et al.*, "Potent immunogenicity and efficacy of a universal influenza vaccine candidate comprising a recombinant fusion protein linking influenza M2e to the TLR5 ligand flagellin," *Vaccine*, vol. 26, no. 2, pp. 201–214, Jan. 2008.

[178] L. M. Tsybalova *et al.*, "Development of a candidate influenza vaccine based on virus-like particles displaying influenza M2e peptide into the immunodominant region of hepatitis B core antigen: Broad protective efficacy of particles carrying four copies of M2e," *Vaccine*, vol. 33, no. 29, pp. 3398–3406, Jun. 2015.

[179] S. Neirynck, T. Deroo, X. Saelens, P. Vanlandschoot, W. M. Jou, and W. Fiers, "A universal

influenza A vaccine based on the extracellular domain of the M2 protein," *Nat. Med.*, vol. 5, no. 10, pp. 1157–1163, Oct. 1999.

[180] D. M. Carter *et al.*, "Design and Characterization of a Computationally Optimized Broadly Reactive Hemagglutinin Vaccine for H1N1 Influenza Viruses," *J. Virol.*, vol. 90, no. 9, pp. 4720–4734, May 2016.

[181] T. M. Wong *et al.*, "Computationally Optimized Broadly Reactive Hemagglutinin Elicits Hemagglutination Inhibition Antibodies against a Panel of H3N2 Influenza Virus Cocirculating Variants," *J. Virol.*, vol. 91, no. 24, Dec. 2017.

[182] C. J. Crevar, D. M. Carter, K. Y. J. Lee, and T. M. Ross, "Cocktail of H5N1 COBRA HA vaccines elicit protective antibodies against H5N1 viruses from multiple clades," *Hum. Vaccin. Immunother.*, vol. 11, no. 3, pp. 572–583, 2015.

[183] J. D. Allen, S. Ray, and T. M. Ross, "Split inactivated COBRA vaccine elicits protective antibodies against H1N1 and H3N2 influenza viruses," *PLoS One*, vol. 13, no. 9, Sep. 2018.

[184] T. M. Ross, L. hairong, B. S. Chia, E. Hill, H. Weirback, and S. Zimmer, "Prevalence of antibodies against seasonal influenza A and B viruses during the 2009-2010 and 2010-2011 influenza seasons in residents of Pittsburgh, PA, USA.," *PLoS Curr.*, vol. 3, 2011.

[185] "Vaccine Antigen Design | DIOSynVax | Cambridge." [Online]. Available: https://www.diosvax.com/. [Accessed: 09-Mar-2024].

[186] "THE WHO PANDEMIC PHASES - Pandemic Influenza Preparedness and Response - NCBI Bookshelf." [Online]. Available: https://www.ncbi.nlm.nih.gov/books/NBK143061/. [Accessed: 08-Mar-2024].

[187] G. Cacciapaglia, C. Cot, and F. Sannino, "Multiwave pandemic dynamics explained: how to tame the next wave of infectious diseases," *Sci. Reports 2021 111*, vol. 11, no. 1, pp. 1–8, Mar. 2021.

[188] N. M. Bouvier and P. Palese, “THE BIOLOGY OF INFLUENZA VIRUSES,” *Vaccine*, vol. 26, no. Suppl 4, p. D49, Sep. 2008.

[189] G. Neumann, T. Noda, and Y. Kawaoka, “Emergence and pandemic potential of swine-origin H1N1 influenza virus,” *Nature*, vol. 459, no. 7249. Nature Publishing Group, pp. 931–939, 18-Jun-2009.

[190] Y. Suzuki, “Sialobiology of Influenza: Molecular Mechanism of Host Range Variation of Influenza Viruses,” *Biol. Pharm. Bull.*, vol. 28, no. 3, pp. 399–408, Mar. 2005.

[191] M. Imai *et al.*, “Experimental adaptation of an influenza H5 HA confers respiratory droplet transmission to a reassortant H5 HA/H1N1 virus in ferrets,” *Nat. 2012 4867403*, vol. 486, no. 7403, pp. 420–428, May 2012.

[192] J. C. Paulson and R. P. de Vries, “H5N1 receptor specificity as a factor in pandemic risk,” *Virus Res.*, vol. 178, no. 1, pp. 99–113, Dec. 2013.

[193] M. R. A. Welkers *et al.*, “Genetic diversity and host adaptation of avian H5N1 influenza viruses during human infection,” *Emerg. Microbes Infect.*, vol. 8, no. 1, p. 262, Jan. 2019.

[194] A. Gambaryan, A. Tuzikov, G. Pazynina, N. Bovin, A. Balish, and A. Klimov, “Evolution of the receptor binding phenotype of influenza A (H5) viruses,” *Virology*, vol. 344, no. 2, pp. 432–438, Jan. 2006.

[195] P. Auewarakul *et al.*, “An Avian Influenza H5N1 Virus That Binds to a Human-Type Receptor,” *J. Virol.*, vol. 81, no. 18, p. 9950, Sep. 2007.

[196] W. Ma, R. E. Kahn, and J. A. Richt, “The pig as a mixing vessel for influenza viruses: Human and veterinary implications,” *J. Mol. Genet. Med.*, vol. 3, no. 1, p. 158, 2009.

[197] G. A. Landolt and C. W. Olsen, “Up to new tricks - a review of cross-species transmission of influenza A viruses,” *Anim. Heal. Res. Rev.*, vol. 8, no. 1, pp. 1–21, 2007.

[198] N. Madhav, B. Oppenheim, M. Gallivan, P. Mulembakani, E. Rubin, and N. Wolfe, “Pandemics: Risks, Impacts, and Mitigation,” *Dis. Control Priorities, Third Ed. (Volume 9) Improv. Heal. Reducing Poverty*, pp. 315–345, Nov. 2017.

[199] G. E. C. Charnley, I. Kelman, K. A. M. Gaythorpe, and K. A. Murray, “Traits and risk factors of post-disaster infectious disease outbreaks: a systematic review,” *Sci. Reports 2021 111*, vol. 11, no. 1, pp. 1–14, Mar. 2021.

[200] H. J. Kim, H. Hwang, H. Hong, J. J. Yim, and J. Lee, “A systematic review and meta-analysis of regional risk factors for critical outcomes of COVID-19 during early phase of the pandemic,” *Sci. Reports 2021 111*, vol. 11, no. 1, pp. 1–13, May 2021.

[201] F. Krammer, “The human antibody response to influenza A virus infection and vaccination,” *Nat. Rev. Immunol. 2019 196*, vol. 19, no. 6, pp. 383–397, Mar. 2019.

[202] X. S. Zhang, R. Pebody, D. De Angelis, P. J. White, A. Charlett, and J. W. McCauley, “The Possible Impact of Vaccination for Seasonal Influenza on Emergence of Pandemic Influenza via Reassortment,” *PLoS One*, vol. 9, no. 12, p. e114637, Dec. 2014.

[203] R. Subramanian, A. L. Graham, B. T. Grenfell, and N. Arinaminpathy, “Universal or Specific? A Modeling-Based Comparison of Broad-Spectrum Influenza Vaccines against Conventional, Strain-Matched Vaccines,” *PLOS Comput. Biol.*, vol. 12, no. 12, p. e1005204, Dec. 2016.

[204] N. P. A. S. Johnson, “Measuring a pandemic: Mortality, demography and geography,” *Pop. e Stor.*, vol. 4, no. 2, pp. 31–51, 2003.

[205] L. J. Donaldson *et al.*, “Mortality from pandemic A/H1N1 2009 influenza in England: public health surveillance study,” *BMJ*, vol. 339, no. 7737, p. 82, Dec. 2009.

[206] M. B. Gregg, A. R. Hinman, and R. B. Craven, “The Russian Flu: Its History and Implications for This Year’s Influenza Season,” *JAMA*, vol. 240, no. 21, pp. 2260–2263, Nov. 1978.

[207] Z. Xu *et al.*, "Pathological findings of COVID-19 associated with acute respiratory distress syndrome," *Lancet Respir. Med.*, vol. 8, no. 4, pp. 420–422, Apr. 2020.

[208] Z. Hu *et al.*, "Clinical characteristics of 24 asymptomatic infections with COVID-19 screened among close contacts in Nanjing, China," *Sci. China Life Sci.*, vol. 63, no. 5, pp. 706–711, May 2020.

[209] Q. Li *et al.*, "Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus–Infected Pneumonia," *N. Engl. J. Med.*, vol. 382, no. 13, pp. 1199–1207, Mar. 2020.

[210] P. Zhou *et al.*, "A pneumonia outbreak associated with a new coronavirus of probable bat origin," *Nature*, vol. 579, no. 7798, pp. 270–273, Mar. 2020.

[211] Y. Jin *et al.*, "Virology, Epidemiology, Pathogenesis, and Control of COVID-19," *Viruses*, vol. 12, no. 4, p. 372, Mar. 2020.

[212] C. B. Jackson, M. Farzan, B. Chen, and H. Choe, "Mechanisms of SARS-CoV-2 entry into cells," *Nat. Rev. Mol. Cell Biol. 2021 231*, vol. 23, no. 1, pp. 3–20, Oct. 2021.

[213] P. V'kovski, A. Kratzel, S. Steiner, H. Stalder, and V. Thiel, "Coronavirus biology and replication: implications for SARS-CoV-2," *Nat. Rev. Microbiol. 2020 193*, vol. 19, no. 3, pp. 155–170, Oct. 2020.

[214] M. Pizzato *et al.*, "SARS-CoV-2 and the Host Cell: A Tale of Interactions," *Front. Virol.*, vol. 1, p. 815388, Jan. 2021.

[215] I. P. Trougakos *et al.*, "Insights to SARS-CoV-2 life cycle, pathophysiology, and rationalized treatments that target COVID-19 clinical complications," *J. Biomed. Sci. 2021 281*, vol. 28, no. 1, pp. 1–18, Jan. 2021.

[216] Y. Huang, C. Yang, X. feng Xu, W. Xu, and S. wen Liu, "Structural and functional properties of SARS-CoV-2 spike protein: potential antivirus drug development for COVID-19," *Acta*

Pharmacol. Sin. 2020 419, vol. 41, no. 9, pp. 1141–1149, Aug. 2020.

[217] W. Dejnirattisai *et al.*, "The antigenic anatomy of SARS-CoV-2 receptor binding domain," *Cell*, vol. 184, no. 8, pp. 2183-2200.e22, Apr. 2021.

[218] C. O. Barnes *et al.*, "Structures of human antibodies bound to SARS-CoV-2 spike reveal common epitopes and recurrent features of antibodies," *bioRxiv*, May 2020.

[219] A. T. Huang *et al.*, "A systematic review of antibody mediated immunity to coronaviruses: antibody kinetics, correlates of protection, and association of antibody responses with severity of disease," *medRxiv*, p. 2020.04.14.20065771, Apr. 2020.

[220] C. P. Thompson *et al.*, "Detection of neutralising antibodies to SARS-CoV-2 to determine population exposure in Scottish blood donors between March and May 2020," *Eurosurveillance*, vol. 25, no. 42, Oct. 2020.

[221] S. F. Lumley *et al.*, "SARS-CoV-2 antibody prevalence, titres and neutralising activity in an antenatal cohort, United Kingdom, 14 April to 15 June 2020," *Eurosurveillance*, vol. 25, no. 42, Oct. 2020.

[222] G. W. Carnell, F. Ferrara, K. Grehan, C. P. Thompson, and N. J. Temperton, "Pseudotype-based neutralization assays for influenza: A systematic analysis," *Front. Immunol.*, vol. 6, no. MAR, p. 140476, Apr. 2015.

[223] Y. Sun, W. Huang, H. Xiang, and J. Nie, "SARS-CoV-2 Neutralization Assays Used in Clinical Trials: A Narrative Review," *Vaccines 2024, Vol. 12, Page 554*, vol. 12, no. 5, p. 554, May 2024.

[224] T. Mahmood and P. C. Yang, "Western Blot: Technique, Theory, and Trouble Shooting," *N. Am. J. Med. Sci.*, vol. 4, no. 9, p. 429, Sep. 2012.

[225] J. Budd *et al.*, "Lateral flow test engineering and lessons learned from COVID-19," *Nat. Rev. Bioeng. 2023 11*, vol. 1, no. 1, pp. 13–31, Jan. 2023.

[226] A. N. Grossberg *et al.*, "A multiplex chemiluminescent immunoassay for serological profiling of COVID-19-positive symptomatic and asymptomatic patients," *Nat. Commun. 2021 121*, vol. 12, no. 1, pp. 1–11, Feb. 2021.

[227] M. Gdoura, F. Ben Ghaloum, M. Ben Hamida, W. Chamsa, H. Triki, and C. Bahloul, "Development of an in-house quantitative ELISA for the evaluation of different Covid-19 vaccines in humans," *Sci. Reports 2022 121*, vol. 12, no. 1, pp. 1–9, Jul. 2022.

[228] H. H. Nguyen, J. Park, S. Kang, and M. Kim, "Surface Plasmon Resonance: A Versatile Technique for Biosensor Applications," *Sensors (Basel).*, vol. 15, no. 5, p. 10481, May 2015.

[229] J. L. R. Zamora and H. C. Aguilar, "Flow virometry as a tool to study viruses," *Methods*, vol. 134–135, pp. 87–97, Feb. 2018.

[230] T. L. Smith, R. Jennings, and C. W. Potter, "Use of single radial haemolysis for assessing antibody response to influenza virus vaccines in animals," *Med. Microbiol. Immunol.*, vol. 176, no. 6, pp. 329–339, Oct. 1987.

[231] H. Chen, "Annual reports of OIE Reference Laboratories and Collaborating Centres - Avian Influenza," 2009.

[232] WHO, "Rapid Risk Assessment Assessment of risk associated with recent influenza A(H5N1) clade 2.3.4.4b viruses Background."

[233] K. Grehan, F. Ferrara, and N. Temperton, "An optimised method for the production of MERS-CoV spike expressing viral pseudotypes," *MethodsX*, vol. 2, pp. 379–384, 2015.

[234] F. Ferrara and N. J. Temperton, "Pseudotype Neutralization Assays: From Laboratory Bench to Data Analysis," *Methods Protoc.*, vol. 1, no. 1, p. 8, Jan. 2018.

[235] N. J. Temperton *et al.*, "Longitudinally profiling neutralizing antibody response to SARS coronavirus with pseudotypes," *Emerg. Infect. Dis.*, vol. 11, no. 3, pp. 411–416, 2005.

[236] K. Ewer *et al.*, “A Monovalent Chimpanzee Adenovirus Ebola Vaccine Boosted with MVA,” *N. Engl. J. Med.*, vol. 374, no. 17, pp. 1635–1646, Apr. 2016.

[237] D. W. Crook *et al.*, “Antibody testing for COVID-19: A report from the National COVID Scientific Advisory Panel,” *Wellcome Open Res.*, vol. 5, 2020.

[238] C. E. Shannon, “A Mathematical Theory of Communication,” *Bell Syst. Tech. J.*, vol. 27, no. 3, pp. 379–423, Jul. 1948.

[239] K. Pan and M. W. Deem, “Quantifying selection and diversity in viruses by entropy methods, with application to the haemagglutinin of H3N2 influenza,” *J. R. Soc. Interface*, vol. 8, no. 64, p. 1644, Nov. 2011.

[240] J. V. Forrester, A. D. Dick, P. G. McMenamin, F. Roberts, and E. Pearlman, “Immunology,” *Eye*, pp. 370-461.e2, Jan. 2016.

[241] M. B. Doud and J. D. Bloom, “Accurate Measurement of the Effects of All Amino-Acid Mutations on Influenza Hemagglutinin,” *Viruses*, vol. 8, no. 6, Jun. 2016.

[242] R. Adiyaman, N. S. Edmunds, A. G. Genc, S. M. A. Alharbi, and L. J. McGuffin, “Improvement of protein tertiary and quaternary structure predictions using the ReFOLD refinement method and the AlphaFold2 recycling process,” *Bioinforma. Adv.*, vol. 3, no. 1, Jan. 2023.

[243] L. J. McGuffin, N. S. Edmunds, A. G. Genc, S. M. A. Alharbi, B. R. Salehe, and R. Adiyaman, “Prediction of protein structures, functions and interactions using the IntFOLD7, MultiFOLD and ModFOLDdock servers,” *Nucleic Acids Res.*, vol. 51, no. W1, pp. W274–W280, Jul. 2023.

[244] “Pymol-script-repo/spectrumany.py at master · Pymol-Scripts/Pymol-script-repo · GitHub.” [Online]. Available: https://github.com/Pymol-Scripts/Pymol-script-repo/blob/master/spectrumany.py. [Accessed: 17-Mar-2024].

[245] K. Katoh and D. M. Standley, “MAFFT Multiple Sequence Alignment Software Version 7:

Improvements in Performance and Usability," *Mol. Biol. Evol.*, vol. 30, no. 4, pp. 772–780, Apr. 2013.

[246] "Amino Acids Reference Chart." [Online]. Available: https://www.sigmaaldrich.com/GB/en/technical-documents/technical-article/protein-biology/protein-structural-analysis/amino-acid-reference-chart. [Accessed: 20-Mar-2022].

[247] F. Abascal, R. Zardoya, and M. J. Telford, "TranslatorX: multiple alignment of nucleotide sequences guided by amino acid translations," *Nucleic Acids Res.*, vol. 38, no. Web Server issue, Apr. 2010.

[248] S. Capella-Gutierrez, J. M. Silla-Martinez, and T. Gabaldon, "trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses," *Bioinformatics*, vol. 25, no. 15, pp. 1972–1973, Aug. 2009.

[249] T. Flouri *et al.*, "The phylogenetic likelihood library," *Syst. Biol.*, vol. 64, no. 2, pp. 356–362, Mar. 2015.

[250] D. Darriba, D. Posada, A. M. Kozlov, A. Stamatakis, B. Morel, and T. Flouri, "ModelTest-NG: A New and Scalable Tool for the Selection of DNA and Protein Evolutionary Models," *Mol. Biol. Evol.*, vol. 37, no. 1, pp. 291–294, Jan. 2020.

[251] R. Bouckaert *et al.*, "BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis," *PLoS Comput. Biol.*, vol. 15, no. 4, p. e1006650, 2019.

[252] D. Gasbarra, M. Pirinen, M. J. Sillanpää, E. Salmela, and E. Arjas, "Estimating genealogies from unlinked marker data: A Bayesian approach," *Theor. Popul. Biol.*, vol. 72, no. 3, pp. 305–322, Nov. 2007.

[253] A. J. Drummond, "Bayesian Coalescent Inference of Past Population Dynamics from Molecular Sequences," *Mol. Biol. Evol.*, vol. 22, no. 5, pp. 1185–1192, Feb. 2005.

[254] A. J. Drummond, S. Y. W. Ho, M. J. Phillips, and A. Rambaut, “Relaxed phylogenetics and dating with confidence,” *PLoS Biol.*, vol. 4, no. 5, pp. 699–710, May 2006.

[255] T. Stadler *et al.*, “How well can the exponential-growth coalescent approximate constant-rate birth–death population dynamics?,” *Proc. R. Soc. B Biol. Sci.*, vol. 282, no. 1806, 2015.

[256] A. Rambaut, A. J. Drummond, D. Xie, G. Baele, and M. A. Suchard, “Posterior summarization in Bayesian phylogenetics using Tracer 1.7,” *Syst. Biol.*, vol. 67, no. 5, pp. 901–904, Sep. 2018.

[257] G. Yu, “Using ggtree to Visualize Data on Tree-Like Structures,” *Curr. Protoc. Bioinforma.*, vol. 69, no. 1, Mar. 2020.

[258] G. Yu, T. T. Y. Lam, H. Zhu, and Y. Guan, “Two Methods for Mapping and Visualizing Associated Data on Phylogeny Using Ggtree,” *Mol. Biol. Evol.*, vol. 35, no. 12, pp. 3041–3043, Dec. 2018.

[259] G. Yu, D. K. Smith, H. Zhu, Y. Guan, and T. T. Y. Lam, “ggtree: an r package for visualization and annotation of phylogenetic trees with their covariates and other associated data,” *Methods Ecol. Evol.*, vol. 8, no. 1, pp. 28–36, Jan. 2017.

[260] S. N. Wood, “Generalized additive models: An introduction with R, second edition,” *Gen. Addit. Model. An Introd. with R, Second Ed.*, pp. 1–476, Jan. 2017.

[261] “CRAN - Package mgcv.” [Online]. Available: https://cran.r-project.org/web/packages/mgcv/index.html. [Accessed: 17-Mar-2024].

[262] A. Kassambara, “Pipe-Friendly Framework for Basic Statistical Tests [R package rstatix version 0.7.2],” Feb. 2023.

[263] “CRAN - Package binom.” [Online]. Available: https://cran.r-project.org/web/packages/binom/index.html. [Accessed: 17-Mar-2024].

[264] H. Woo *et al.*, “Developing a fully glycosylated full-length SARS-COV-2 spike protein model in

a viral membrane," *J. Phys. Chem. B*, vol. 124, no. 33, pp. 7128–7137, Aug. 2020.

[265] "Scotblood | Homepage." [Online]. Available: https://www.scotblood.co.uk/. [Accessed: 29-Mar-2024].

[266] S. C. Hill *et al.*, "Comparative micro-epidemiology of pathogenic avian influenza virus outbreaks in a wild bird population," *Philos. Trans. R. Soc. B*, vol. 374, no. 1775, 2019.

[267] C. Dreyfus, D. C. Ekiert, and I. A. Wilson, "Structure of a Classical Broadly Neutralizing Stem Antibody in Complex with a Pandemic H2 Influenza Virus Hemagglutinin," *J. Virol.*, vol. 87, no. 12, p. 7149, Jun. 2013.

[268] S. Sakabe *et al.*, "A cross-reactive neutralizing monoclonal antibody protects mice from H5N1 and pandemic (H1N1) 2009 virus infection," *Antiviral Res.*, vol. 88, no. 3, p. 249, Dec. 2010.

[269] T. C. Sutton *et al.*, "In Vitro Neutralization Is Not Predictive of Prophylactic Efficacy of Broadly Neutralizing Monoclonal Antibodies CR6261 and CR9114 against Lethal H2 Influenza Virus Challenge in Mice," *J. Virol.*, vol. 91, no. 24, Dec. 2017.

[270] A. E. Gorbalenya *et al.*, "The species Severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2," *Nat. Microbiol. 2020 54*, vol. 5, no. 4, pp. 536–544, Mar. 2020.

[271] H. Qi, B. Liu, X. Wang, and L. Zhang, "The humoral response and antibodies against SARS-CoV-2 infection," *Nat. Immunol. 2022 237*, vol. 23, no. 7, pp. 1008–1020, Jun. 2022.

[272] R. A. Elsner and M. J. Shlomchik, "Germinal Center and Extrafollicular B Cell Responses in Vaccination, Immunity, and Autoimmunity," *Immunity*, vol. 53, no. 6, pp. 1136–1150, Dec. 2020.

[273] J. M. Dan *et al.*, "Immunological memory to SARS-CoV-2 assessed for up to 8 months after infection," *Science*, vol. 371, no. 6529, Feb. 2021.

[274] H. Wang *et al.*, “The genetic sequence, origin, and diagnosis of SARS-CoV-2,” 2020.

[275] S. K. P. Lau *et al.*, “Coronavirus HKU1 and other coronavirus infections in Hong Kong,” *J. Clin. Microbiol.*, vol. 44, no. 6, pp. 2063–2071, Jun. 2006.

[276] M. Worobey *et al.*, “The Huanan Seafood Wholesale Market in Wuhan was the early epicenter of the COVID-19 pandemic,” *Science (80-.).*, vol. 377, no. 6609, pp. 951–959, Aug. 2022.

[277] W. Bin Yu, G. Da Tang, L. Zhang, and R. T. Corlett, “Decoding the evolution and transmissions of the novel pneumonia coronavirus (SARS-CoV-2 / HCoV-19) using whole genomic data,” *Zool. Res.*, vol. 41, no. 3, pp. 247–257, May 2020.

[278] C. Xiong, L. Jiang, Y. Chen, and Q. Jiang, “Evolution and variation of 2019-novel coronavirus,” *bioRxiv*, p. 2020.01.30.926477, Jan. 2020.

[279] J. Li, S. Lai, G. F. Gao, and W. Shi, “The emergence, genomic diversity and global spread of SARS-CoV-2,” *Nat. 2021 6007889*, vol. 600, no. 7889, pp. 408–418, Dec. 2021.

[280] E. O. Nsoesie, B. Rader, Y. L. Barnoon, L. Goodwin, and J. Brownstein, “Analysis of hospital traffic and search engine data in Wuhan China indicates early disease activity in the Fall of 2019,” 2020.

[281] A. Deslandes *et al.*, “SARS-CoV-2 was already spreading in France in late December 2019,” *Int. J. Antimicrob. Agents*, vol. 55, no. 6, p. 106006, Jun. 2020.

[282] P. J. Lillie *et al.*, “Novel coronavirus disease (Covid-19): The first two patients in the UK with person to person transmission,” *Journal of Infection*, vol. 80, no. 5. W.B. Saunders Ltd, pp. 578–606, 01-May-2020.

[283] R. Lu *et al.*, “Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding,” *Lancet*, vol. 395, no. 10224, pp. 565–574,

Feb. 2020.

[284] T. T.-Y. Lam *et al.*, “Identifying SARS-CoV-2 related coronaviruses in Malayan pangolins,” *Nature*, pp. 1–6, Mar. 2020.

[285] S. Priddy, “The coronavirus timeline: Measures taken by the House of Commons.” .

[286] “Coronavirus: 32 more COVID-19 patients die in England - taking UK total to 104 | UK News | Sky News.” [Online]. Available: https://news.sky.com/story/coronavirus-32-more-covid-19-patients-die-in-england-taking-uk-total-to-104-11959809. [Accessed: 29-Mar-2024].

[287] “COVID-19 confirmed deaths in England (to 31 December 2020): report - GOV.UK.” [Online]. Available: https://www.gov.uk/government/publications/covid-19-reported-sars-cov-2-deaths-in-england/covid-19-confirmed-deaths-in-england-to-31-december-2020-report. [Accessed: 29-Mar-2024].

[288] PHE, “Confirmed cases in England,” 2020.

[289] “Coronavirus (COVID-19) Infection Survey technical article - Office for National Statistics.” [Online]. Available: https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/articles/coronaviruscovid19infectionsurveytechnicalarticle/wavesandlagsofcovid19inenglandjune2021. [Accessed: 29-Mar-2024].

[290] L. Grandjean *et al.*, “Long-Term Persistence of Spike Protein Antibody and Predictive Modeling of Antibody Dynamics After Infection With Severe Acute Respiratory Syndrome Coronavirus 2,” *Clin. Infect. Dis. An Off. Publ. Infect. Dis. Soc. Am.*, vol. 74, no. 7, p. 1220, Apr. 2022.

[291] A. Bläckberg, N. Fernström, E. Sarbrant, M. Rasmussen, and T. Sunnerhagen, “Antibody kinetics and clinical course of COVID-19 a prospective observational study,” *PLoS One*, vol. 16, no. 3, p. e0248918, Mar. 2021.

[292] N. Ortega *et al.*, "Seven-month kinetics of SARS-CoV-2 antibodies and role of pre-existing antibodies to human coronaviruses," *Nat. Commun. 2021 121*, vol. 12, no. 1, pp. 1–10, Aug. 2021.

[293] "Longitudinal Monitoring of SARS-CoV-2 IgM and IgG Seropositivity to Detect COVID-19," *J. Appl. Lab. Med.*, vol. 6, no. 2, p. 565, Mar. 2021.

[294] Q. X. Long *et al.*, "Antibody responses to SARS-CoV-2 in patients with COVID-19," *Nat. Med. 2020 266*, vol. 26, no. 6, pp. 845–848, Apr. 2020.

[295] J. H. Lam, F. L. Smith, and N. Baumgarth, "B Cell Activation and Response Regulation During Viral Infections," *Viral Immunol.*, vol. 33, no. 4, pp. 294–306, 2020.

[296] J. Seow *et al.*, "Longitudinal observation and decline of neutralizing antibody responses in the three months following SARS-CoV-2 infection in humans," *Nat. Microbiol. 2020 512*, vol. 5, no. 12, pp. 1598–1607, Oct. 2020.

[297] B. Isho *et al.*, "Persistence of serum and saliva antibody responses to SARS-CoV-2 spike antigens in COVID-19 patients," *Sci. Immunol.*, vol. 5, no. 52, Oct. 2020.

[298] L. Premkumar *et al.*, "The receptor binding domain of the viral spike protein is an immunodominant and highly specific target of antibodies in SARS-CoV-2 patients," *Sci. Immunol.*, vol. 5, no. 48, Jun. 2020.

[299] M. S. Suthar *et al.*, "Rapid Generation of Neutralizing Antibody Responses in COVID-19 Patients," *Cell Reports Med.*, vol. 1, no. 3, p. 100040, Jun. 2020.

[300] L. Piccoli *et al.*, "Mapping Neutralizing and Immunodominant Sites on the SARS-CoV-2 Spike Receptor-Binding Domain by Structure-Guided High-Resolution Serology," *Cell*, vol. 183, no. 4, p. 1024, Nov. 2020.

[301] F. Wu *et al.*, "Neutralizing antibody responses to SARS-CoV-2 in a COVID-19 recovered patient

cohort and their implications," *medRxiv*, 2020.

[302] S. Bolotin, S. Wilson, and M. Murti, "Achieving and sustaining herd immunity to SARS-CoV-2," *C. Can. Med. Assoc. J.*, vol. 193, no. 28, p. E1089, Jul. 2021.

[303] A. Wajnberg *et al.*, "Robust neutralizing antibodies to SARS-CoV-2 infection persist for months," *Science*, vol. 370, no. 6521, pp. 1227–1230, Dec. 2020.

[304] K. Wang *et al.*, "Longitudinal Dynamics of the Neutralizing Antibody Response to Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) Infection," *Clin. Infect. Dis.*, vol. 73, no. 3, pp. E531–E539, Aug. 2021.

[305] B. J. Laidlaw and A. H. Ellebedy, "The germinal centre B cell response to SARS-CoV-2," *Nat. Rev. Immunol. 2021 221*, vol. 22, no. 1, pp. 7–18, Dec. 2021.

[306] Y. Chen *et al.*, "Quick COVID-19 Healers Sustain Anti-SARS-CoV-2 Antibody Production," *Cell*, vol. 183, no. 6, p. 1496, Dec. 2020.

[307] L. Kuri-Cervantes *et al.*, "Comprehensive mapping of immune perturbations associated with severe COVID-19," *Sci. Immunol.*, vol. 5, no. 49, Jul. 2020.

[308] C. Cervia *et al.*, "Systemic and mucosal antibody secretion specific to SARS-CoV-2 during mild versus severe COVID-19," *bioRxiv*, p. 2020.05.21.108308, May 2020.

[309] D. Hobson, R. L. Curry, A. S. Beare, and A. Ward-Gardner, "The role of serum haemagglutination-inhibiting antibody in protection against challenge infection with influenza A2 and B viruses," *J. Hyg. (Lond).*, vol. 70, no. 4, p. 767, 1972.

[310] W. Smith, C. H. Andrewes, and P. P. Laidlaw, "A VIRUS OBTAINED FROM INFLUENZA PATIENTS," *Lancet*, vol. 222, no. 5732, pp. 66–68, Jul. 1933.

[311] E. Bendavid *et al.*, "COVID-19 Antibody Seroprevalence in Santa Clara County, California," *medRxiv*, p. 2020.04.14.20062463, Apr. 2020.

[312] A. Doi *et al.*, “Estimation of seroprevalence of novel coronavirus disease (COVID-19) using preserved serum at an outpatient setting in Kobe, Japan: A cross-sectional study.,” *medRxiv*, p. 2020.04.26.20079822, May 2020.

[313] N. Sood *et al.*, “Seroprevalence of SARS-CoV-2–Specific Antibodies Among Adults in Los Angeles County, California, on April 10-11, 2020,” *JAMA*, May 2020.

[314] S. Stringhini *et al.*, “Repeated seroprevalence of anti-SARS-CoV-2 IgG antibodies in a population-based sample from Geneva, Switzerland,” Cold Spring Harbor Laboratory Press, May 2020.

[315] B. Meyer, “Validation of a commercially available SARS-CoV-2 serological Immunoassay,” Cold Spring Harbor Laboratory Press, May 2020.

[316] K. G. Beavis *et al.*, “Evaluation of the EUROIMMUN Anti-SARS-CoV-2 ELISA Assay for detection of IgA and IgG antibodies,” *J. Clin. Virol.*, vol. 129, p. 104468, Aug. 2020.

[317] FDA, “Anti-SARS-CoV-2 ELISA (IgG) Instruction for use.”

[318] “Unapproved Chinese coronavirus antibody tests being used in at least 2 states.” [Online]. Available: https://www.nbcnews.com/health/health-news/unapproved-chinese-coronavirus-antibody-tests-being-used-least-2-states-n1185131. [Accessed: 29-Mar-2024].

[319] S. E. Conklin *et al.*, “Evaluation of Serological SARS-CoV-2 Lateral Flow Assays for Rapid Point-of-Care Testing,” *J. Clin. Microbiol.*, vol. 59, no. 2, Feb. 2021.

[320] A. Doi *et al.*, “A cross-sectional follow up study to estimate seroprevalence of coronavirus disease 2019 in Kobe, Japan,” *Medicine (Baltimore).*, vol. 100, no. 48, Dec. 2021.

[321] “Glasgow Population 2024.” [Online]. Available: https://worldpopulationreview.com/world-cities/glasgow-population. [Accessed: 29-Mar-2024].

[322] I. R.-B. Saki Takahashi, Bryan Greenhouse, “Are SARS-CoV-2 seroprevalence estimates

biased?," *OSF Prepr.*, vol. 1, no. 1, pp. 1–12, 2020.

[323] J. Rosado *et al.*, "Serological signatures of SARS-CoV-2 infection: Implications for antibody-based diagnostics," Cold Spring Harbor Laboratory Press, May 2020.

[324] Y. Liu, A. A. Gayle, A. Wilder-Smith, and J. Rocklöv, "The reproductive number of COVID-19 is higher compared to SARS coronavirus," *J. Travel Med.*, vol. 2020, pp. 1–4, 2020.

[325] M. Gabriela *et al.*, "Individual variation in susceptibility or exposure to SARS-CoV-2 lowers the 2 herd immunity threshold."

[326] S. S. Morse *et al.*, "Prediction and prevention of the next pandemic zoonosis," *The Lancet*, vol. 380, no. 9857. Elsevier, pp. 1956–1965, 01-Dec-2012.

[327] C. Adlhoch *et al.*, "Avian influenza overview December 2022 – March 2023," *EFSA J.*, vol. 21, no. 3, Mar. 2023.

[328] R. J. Webby and R. G. Webster, "Are We Ready for Pandemic Influenza?," *Science*, vol. 302, no. 5650. American Association for the Advancement of Science, pp. 1519–1522, 28-Nov-2003.

[329] T. C. Sutton, "The pandemic threat of emerging H5 and H7 avian influenza viruses," *Viruses*, vol. 10, no. 9. MDPI AG, 01-Sep-2018.

[330] "Ebola: overview, history, origins and transmission - GOV.UK." [Online]. Available: https://www.gov.uk/government/publications/ebola-origins-reservoirs-transmission-and-guidelines/ebola-overview-history-origins-and-transmission. [Accessed: 12-Jan-2024].

[331] K. Steeds *et al.*, "Pseudotyping of VSV with Ebola virus glycoprotein is superior to HIV-1 for the assessment of neutralising antibodies," *Sci. Reports 2022 101*, vol. 10, no. 1, pp. 1–10, Aug. 2020.

[332] J. Y. Zhou *et al.*, "Characterization of a highly pathogenic H5N1 influenza virus derived from

bar-headed geese in China," *J. Gen. Virol.*, vol. 87, no. 7, pp. 1823–1833, Jul. 2006.

[333] S. L. Salzberg *et al.*, "Genome Analysis Linking Recent European and African Influenza (H5N1) Viruses," *Emerg. Infect. Dis.*, vol. 13, no. 5, p. 713, 2007.

[334] G. J. D. Smith *et al.*, "Emergence and predominance of an H5N1 influenza variant in China," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 103, no. 45, pp. 16936–16941, Nov. 2006.

[335] Y. Gao *et al.*, "Identification of Amino Acids in HA and PB2 Critical for the Transmission of H5N1 Avian Influenza Viruses in a Mammalian Host," *PLoS Pathog.*, vol. 5, no. 12, p. 1000709, Dec. 2009.

[336] H. G. Pereira, B. Tůmová, and V. G. Law, "Avian influenza A viruses," *Bull. World Health Organ.*, vol. 32, no. 6, p. 855, 1965.

[337] A. H. Reid, T. G. Fanning, J. V Hultin, and J. K. Taubenberger, "Origin and evolution of the 1918 '"Spanish"' influenza virus hemagglutinin gene," vol. 96, pp. 1651–1656, 1999.

[338] T. Francis and T. P. Magill, "IMMUNOLOGICAL STUDIES WITH THE VIRUS OF INFLUENZA," *J. Exp. Med.*, vol. 62, no. 4, p. 505, Sep. 1935.

[339] P. Concannon, I. W. Cummings, and W. A. Salser1', "Nucleotide sequence of the influenza virus A/USSR/90/77 hemagglutinin gene.," *J. Virol.*, vol. 49, no. 1, pp. 276–278, Jan. 1984.

[340] V. Shinde *et al.*, "Triple-Reassortant Swine Influenza A (H1) in Humans in the United States, 2005–2009," *N. Engl. J. Med.*, vol. 360, no. 25, pp. 2616–2625, Jun. 2009.

[341] L. Simonsen *et al.*, "Global Mortality Estimates for the 2009 Influenza Pandemic from the GLaMOR Project: A Modeling Study," *PLoS Med.*, vol. 10, no. 11, Nov. 2013.

[342] F. S. Dawood *et al.*, "Estimated global mortality associated with the first 12 months of 2009 pandemic influenza A H1N1 virus circulation: A modelling study," *Lancet Infect. Dis.*, vol. 12, no. 9, pp. 687–695, Sep. 2012.

[343] S. Marchi *et al.*, “Serologically-Based Evaluation of Cross-Protection Antibody Responses among Different A(H1N1) Influenza Strains,” *Vaccines 2020, Vol. 8, Page 656*, vol. 8, no. 4, p. 656, Nov. 2020.

[344] V. N. Petrova and C. A. Russell, “The evolution of seasonal influenza viruses,” *Nat. Publ. Gr.*, vol. 16, 2017.

[345] G. J. D. Smith and R. O. Donis, “Continued evolution of highly pathogenic avian influenza A (H5N1): updated nomenclature,” *Influenza Other Respi. Viruses*, vol. 6, no. 1, p. 1, Jan. 2012.

[346] N. T. Liem *et al.*, “Clinical features of human influenza A (H5N1) infection in Vietnam: 2004-2006,” *Clin. Infect. Dis.*, vol. 48, no. 12, pp. 1639–1646, Jun. 2009.

[347] T. T. Hien *et al.*, “Avian Influenza A (H5N1) in 10 Patients in Vietnam,” *n engl j med*, vol. 12, pp. 1179–88, 2004.

[348] L. Sánchez-de Prada *et al.*, “Group 1 and group 2 hemagglutinin stalk antibody response according to age,” *Front. Immunol.*, vol. 14, 2023.

[349] K. M. Gostic, M. Ambrose, M. Worobey, and J. O. Lloyd-Smith, “Potent protection against H5N1 and H7N9 influenza via childhood hemagglutinin imprinting,” *Science (80-.).*, vol. 354, no. 6313, pp. 722–726, Nov. 2016.

[350] C. Scholtissek, W. Rohde, V. Von Hoyningen, and R. Rott, “On the origin of the human influenza virus subtypes H2N2 and H3N2,” *Virology*, vol. 87, no. 1, pp. 13–20, Jun. 1978.

[351] D. Cantoni, M. Mayora-Neto, and N. Temperton, “The role of pseudotype neutralization assays in understanding SARS CoV-2,” *Oxford Open Immunol.*, vol. 2, no. 1, 2021.

[352] N. Eshima *et al.*, “Sex- and Age-Related Differences in Morbidity Rates of 2009 Pandemic Influenza A H1N1 Virus of Swine Origin in Japan,” *PLoS One*, vol. 6, no. 4, 2011.

[353] C. Viboud, J. Eisenstein, A. H. Reid, T. A. Janczewski, D. M. Morens, and J. K. Taubenberger,

"Age- and Sex-Specific Mortality Associated With the 1918–1919 Influenza Pandemic in Kentucky," *J. Infect. Dis.*, vol. 207, no. 5, pp. 721–729, Mar. 2013.

[354] A. C. McHardy and B. Adams, "The Role of Genomics in Tracking the Evolution of Influenza A Virus," *PLoS Pathog.*, vol. 5, no. 10, Oct. 2009.

[355] J. S. Bolton, H. Klim, J. Wellens, M. Edmans, U. Obolski, and C. P. Thompson, "An Antigenic Thrift-Based Approach to Influenza Vaccine Design," *Vaccines*, vol. 9, no. 6, Jun. 2021.

[356] U. Obolski, Y. Ram, and L. Hadany, "Key issues review: evolution on rugged adaptive landscapes," *Rep. Prog. Phys.*, vol. 81, no. 1, Jan. 2018.

[357] K. Koelle and D. A. Rasmussen, "The effects of a deleterious mutation load on patterns of influenza A/H3N2's antigenic evolution in humans," *Elife*, vol. 4, no. September, Sep. 2015.

[358] S. Luo, M. Reed, J. C. Mattingly, and K. Koelle, "The impact of host immune status on the within-host and population dynamics of antigenic immune escape," *J. R. Soc. Interface*, vol. 9, no. 75, pp. 2603–2613, Oct. 2012.

[359] I. Volkov, K. M. Pepin, J. O. Lloyd-Smith, J. R. Banavar, and B. T. Grenfell, "Synthesizing within-host and population-level selective pressures on viral populations: the impact of adaptive immunity on viral immune escape," *J. R. Soc. Interface*, vol. 7, no. 50, pp. 1311–1318, Sep. 2010.

[360] V. N. Petrova and C. A. Russell, "The evolution of seasonal influenza viruses," *Nat. Rev. Microbiol. 2017 161*, vol. 16, no. 1, pp. 47–60, Oct. 2017.

[361] A. X. Han, S. Maurer-Stroh, and C. A. Russell, "Individual immune selection pressure has limited impact on seasonal influenza virus evolution," *Nat. Ecol. Evol. 2018 32*, vol. 3, no. 2, pp. 302–311, Dec. 2018.

[362] C. K. Lumby, L. Zhao, J. Breuer, and C. J. R. Illingworth, "A large effective population size for

established within-host influenza virus infection," *Elife*, vol. 9, pp. 1–17, Aug. 2020.

[363] J. C. Krause *et al.*, "A Broadly Neutralizing Human Monoclonal Antibody That Recognizes a Conserved, Novel Epitope on the Globular Head of the Influenza H1N1 Virus Hemagglutinin," *J. Virol.*, vol. 85, no. 20, p. 10905, Oct. 2011.

[364] G. S. Tan *et al.*, "Characterization of a Broadly Neutralizing Monoclonal Antibody That Targets the Fusion Domain of Group 2 Influenza A Virus Hemagglutinin," *J. Virol.*, vol. 88, no. 23, pp. 13580–13592, Dec. 2014.

[365] R. J. Russell *et al.*, "Structure of influenza hemagglutinin in complex with an inhibitor of membrane fusion," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 105, no. 46, p. 17736, Nov. 2008.

[366] C. M. Saad-Roy, N. Arinaminpathy, N. S. Wingreen, S. A. Levin, J. M. Akey, and B. T. Grenfell, "Implications of localized charge for human influenza A H1N1 hemagglutinin evolution: Insights from deep mutational scans," *PLoS Comput. Biol.*, vol. 16, no. 6, Jun. 2020.

[367] S. Rockman *et al.*, "Reverse engineering the antigenic architecture of the haemagglutinin from influenza H5N1 clade 1 and 2.2 viruses with fine epitope mapping using monoclonal antibodies," *Mol. Immunol.*, vol. 53, no. 4, pp. 435–442, Apr. 2013.

[368] S. L. Kosakovsky Pond, A. F. Y. Poon, A. J. Leigh Brown, and S. D. W. Frost, "A maximum likelihood method for detecting directional evolution in protein sequences and its application to influenza A virus," *Mol. Biol. Evol.*, vol. 25, no. 9, pp. 1809–1824, 2008.

[369] M. Escalera-Zamudio *et al.*, "Parallel evolution in the emergence of highly pathogenic avian influenza A viruses," *Nat. Commun. 2020 111*, vol. 11, no. 1, pp. 1–11, Nov. 2020.

[370] M. L. Perdue and D. L. Suarez, "Structural features of the avian influenza virus hemagglutinin that influence virulence," *Vet. Microbiol.*, vol. 74, no. 1–2, pp. 77–86, May 2000.

[371] H. S. Y. Leung, O. T. W. Li, R. W. Y. Chan, M. C. W. Chan, J. M. Nicholls, and L. L. M. Poon,

"Entry of Influenza A Virus with a α2,6-Linked Sialic Acid Binding Preference Requires Host Fibronectin," *J. Virol.*, vol. 86, no. 19, p. 10704, Oct. 2012.

[372] M. Imai and Y. Kawaoka, "The role of receptor binding specificity in interspecies transmission of influenza viruses," *Curr. Opin. Virol.*, vol. 2, no. 2, p. 160, 2012.

[373] X. Xiong *et al.*, "Enhanced human receptor binding by H5 haemagglutinins," *Virology*, vol. 456–457, no. 100, pp. 179–187, 2014.

[374] Y. Watanabe *et al.*, "Acquisition of Human-Type Receptor Binding Specificity by New H5N1 Influenza Virus Sublineages during Their Emergence in Birds in Egypt," *PLOS Pathog.*, vol. 7, no. 5, p. e1002068, May 2011.

[375] E.-S. M. Abdelwhab, H. M. Hafez, M. M. Aly, C. Grund, and T. C. Harder, "Increasing Prevalence of Unique Mutation Patterns in H5N1 Avian Influenza Virus HA and NA Glycoproteins from Human Infections in Egypt," *Sequencing*, vol. 2010, pp. 1–3, Jul. 2010.

[376] A. Arafa *et al.*, "Phylodynamics of avian influenza clade 2.2.1 H5N1 viruses in Egypt," *Virol. J.*, vol. 13, no. 1, 2016.

[377] V. R. Perovic *et al.*, "Novel Phylogenetic Algorithm to Monitor Human Tropism in Egyptian H5N1-HPAIV Reveals Evolution toward Efficient Human-to-Human Transmission," *PLoS One*, vol. 8, no. 4, Apr. 2013.

[378] K. Healy *et al.*, "Ecology and mode-of-life explain lifespan variation in birds and mammals," *Proceedings. Biol. Sci.*, vol. 281, no. 1784, Jun. 2014.

[379] M. J. Gething, R. W. Doms, D. York, and J. White, "Studies on the mechanism of membrane fusion: site-specific mutagenesis of the hemagglutinin of influenza virus," *J. Cell Biol.*, vol. 102, no. 1, p. 11, Jan. 1986.

[380] D. M. Hinke *et al.*, "Applying valency-based immuno-selection to generate broadly cross-

reactive antibodies against influenza hemagglutinins," *Nat. Commun. 2024 151*, vol. 15, no. 1, pp. 1–17, Feb. 2024.

[381] D. Baxter, "Evaluating the case for trivalent or quadrivalent influenza vaccines," *Human Vaccines and Immunotherapeutics*, vol. 12, no. 10. Taylor and Francis Inc., pp. 2712–2717, 02-Oct-2016.

[382] K. E. Kistler and T. Bedford, "Evidence for adaptive evolution in the receptor-binding domain of seasonal coronaviruses OC43 and 229E."

[383] T. Hale *et al.*, "A global panel database of pandemic policies (Oxford COVID-19 Government Response Tracker)," *Nat. Hum. Behav.*, vol. 5, no. 4, pp. 529–538, Apr. 2021.

[384] C. Schultsz *et al.*, "Prevalence of Antibodies against Avian Influenza A (H5N1) Virus among Cullers and Poultry Workers in Ho Chi Minh City, 2005," *PLoS One*, vol. 4, no. 11, Nov. 2009.

[385] A. Nogales *et al.*, "A Highly Potent and Broadly Neutralizing H1 Influenza-Specific Human Monoclonal Antibody," *Sci. Reports 2018 81*, vol. 8, no. 1, pp. 1–15, Mar. 2018.

[386] D. I. Bernstein *et al.*, "Immunogenicity of chimeric haemagglutinin-based, universal influenza virus vaccine candidates: interim results of a randomised, placebo-controlled, phase 1 clinical trial," *Lancet Infect. Dis.*, vol. 20, no. 1, pp. 80–91, Jan. 2020.

[387] R. Nachbagauer *et al.*, "Induction of Broadly Reactive Anti-Hemagglutinin Stalk Antibodies by an H5N1 Vaccine in Humans," *J. Virol.*, vol. 88, no. 22, p. 13260, Nov. 2014.

[388] T. C. Sutton *et al.*, "Protective efficacy of influenza group 2 hemagglutinin stem-fragment immunogen vaccines," *npj Vaccines*, vol. 2, no. 1, p. 35, Dec. 2017.

[389] A. Choi *et al.*, "Chimeric Hemagglutinin-Based Influenza Virus Vaccines Induce Protective Stalk-Specific Humoral Immunity and Cellular Responses in Mice," *ImmunoHorizons*, vol. 3, no. 4, pp. 133–148, Apr. 2019.