High-throughput analysis of the antigenic effects of mutations to influenza hemagglutinin

Frances Welsh

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington
2024

Reading Committee:
Jesse D. Bloom, Chair
Frederick Matsen
Neil King

Program Authorized to Offer Degree:
Molecular and Cellular Biology

University of Washington

Abstract

High-throughput analysis of the antigenic effects of mutations to influenza hemagglutinin

Frances Welsh

Chair of the supervisory committee:
Jesse D. Bloom
Fred Hutchinson Cancer Center

Influenza virus rapidly evolves to escape neutralization by polyclonal antibodies. However, we have a limited understanding of how the effects of viral mutations on antibody neutralization vary across the human population, and how this heterogeneity might affect virus evolution. In my dissertation, I address this question by mapping the antigenic effects of influenza mutations against sera from defined age cohorts.

First, I describe the development of an improved deep mutational scanning system to measure how mutations in hemagglutinin (HA) affect neutralization by human sera. I engineer a chimeric, barcoded HA construct, which both improves sample throughput while also allowing for absolute quantification of both escape and sensitizing mutations. I show that the resulting barcoded libraries can be used to map the HA epitopes targeted by monoclonal antibodies, antibody cocktails, and polyclonal sera, and that these measurements are consistent with results from traditional neutralization assays.

In the remainder of my thesis, I use a barcoded library in the background of the A/Hong Kong/45/2019 H3 HA protein to analyze heterogeneity in serum antibody targeting across individuals and age cohorts. I find that the effects of HA mutations on

serum neutralization differ across age groups, and that these differences can be partially rationalized in terms of exposure histories. For instance, mutations that revert to amino acids found in the HAs of older viral strains often increase neutralization sensitivity in older individuals, but not young children. I incorporate data from similar experiments using the earlier, non-barcoded A/Perth/16/2009 H3 HA library, which also found substantial differences between child and adult escape maps. Natural mutations that fixed in influenza variants after 2020 cause the greatest escape from sera from children and teenagers, suggesting that antigenic pressure from younger age groups play a more prominent role in driving viral evolution.

Overall, my graduate research demonstrates that influenza faces distinct antigenic selection regimes from different age groups, and that this heterogeneity may have a substantial impact on viral evolution. More rigorous characterization of this immune heterogeneity has the potential to improve both evolutionary forecasting and vaccine effectiveness.

TABLE OF CONTENTS

ACKNOWLEDGMENTS

Thank you to my thesis advisor, Jesse Bloom, for your constant guidance and support. Thank you for encouraging me to stick it out, despite the challenges I faced in joining a virology lab at the peak of a global pandemic. I joined Jesse's lab because of the science, but I stayed because of how deeply he cares about the people he mentors. His ability and willingness to adapt to each person's needs is admirable, and I will never understand how he finds the time to be so readily available despite his packed schedule. Thank you for helping me work my way through 50 years of influenza literature, plan experiments during the age of social distancing, and stay motivated through the inevitable twists and turns of research. And most importantly, thank you for your kindness, understanding, and patience. I could not have chosen a better lab and advisor.

Thank you to the Bloom lab, past and present, for creating such a welcoming and supportive place to do science. Thank you to Sarah, my first mentor in the lab, for teaching me my two most important skills (coding and ping pong). Thank you to Andrea for working alongside me on the flu system, answering a never-ending stream of questions, and somehow keeping the lab running on top of everything else. I'm especially grateful to Tyler, Allie, and Caelan for both your scientific support and reminders not to take everything so seriously. David, Rachel, Keara, Will, Andrew, Tim, and many more, thank you for keeping the energy up and making the lab such a fun place to work.

I'm immensely grateful to my mentors, past and present. I firmly believe that I would not have pursued a PhD if it hadn't been for Leslie Saucedo - thank you for inspiring me and taking me under your wing 10 years ago. Harmit Malik sparked my passion for evolutionary biology, and has been incredibly supportive throughout my academic career. Nasser Rusan, Brian Galletta, and Ryan O'Neill taught me to be an independent, thorough, and incredibly efficient scientist. Even though I did a 180 from

microscopy to genomics, graduate school has been a much smoother ride thanks to the training I received in the Rusan lab.

Thank you to my committee - Erick Matsen, Neil King, Michael Emerman, and Maitreya Dunham - for your support and feedback. Even though it's mostly been over Zoom, I've really enjoyed our discussions, and your genuine interest and encouragement has kept me motivated throughout this project.

I am very fortunate to have an incredible support network in Seattle and beyond. To the friends I've made in graduate school, especially Ellen, Carrie, Violet, Caroline, and Kat - living together was chaotic but I wouldn't have it any other way. Karen, Hannah, Cole, and Emily, my favorite dinner club, thank you for bearing with me through all the highs and lows. To my family, whose belief in me has been a source of strength and motivation, and kept me grounded throughout this journey. And to my brother John, who has been my rock for the past 5 years, and my community in San Francisco. I couldn't have done it without any of you.

CHAPTER 1: Introduction

1.1 The human immune response to influenza

1.1.1 Mechanisms of infection

Seasonal influenza viruses are endemic to the human population, and cause approximately 3-5 million severe cases of respiratory infection each year (WHO, 2018). Infection is mediated by two surface glycoproteins: hemagglutinin (HA) and neuraminidase (NA). Hemagglutinin binds to sialic acid receptors on the target cell surface, triggering endocytosis of the virion, membrane fusion, and subsequent release of the viral genome into the infected cell. Neuraminidase plays the opposing role of cleaving terminal sialic acid residues. This allows for release of nascent viral particles, which initially stick to the cell surface due to HA-receptor binding (Krammer, 2019). Blocking the activity of HA or NA at any step in this process can prevent either cell entry or release of infectious virions.

1.1.2 Antibody generation and mechanisms of action

Upon infection or vaccination, a fraction of activated B cells differentiate into short-lived plasmablasts to produce virus-specific antibodies. Another fraction become memory B cells, which act as a first-line response to future antigen exposures by rapidly differentiating into antibody-generating plasmablasts (Ellebedy et al., 2016; Lau et al., 2017). The remaining B cells migrate to germinal centers and undergo affinity maturation. This process results in long-lived, antibody-secreting plasma cells that migrate to the bone marrow and maintain the steady-state serum antibody levels (Krammer, 2019; Lam and Baumgarth, 2019). These serum antibodies are one of the first lines of defense against infection, as it takes approximately 7-10 days for plasmablasts to generate *de novo* high-affinity antibodies (Kitphati et al., 2009; Wrammert et al., 2008), and only 24-72 hours for virus titers to peak (Baccam et al., 2006).

HA is the immunodominant influenza protein, i.e. the main target of antibodies, followed by NA (Chen et al., 2018). Some antibodies are generated against other viral proteins, but these are generally less accessible to the immune system than the surface glycoproteins (Krammer, 2019). There are two main types of antibodies, as defined by their mechanism of protection. **Neutralizing antibodies** prevent infection. Most neutralizing antibodies target the HA head domain and thereby prevent binding to the sialic acid receptor. However, some neutralizing antibodies function by locking HA into its pre-fusion conformation, preventing membrane fusion and release of the viral genome into the cell (Brandenburg et al., 2013). **Non-neutralizing antibodies** limit virus replication and spread. This includes HA antibodies that inhibit viral egress, antibodies that prevent proteolytic cleavage necessary for HA activation, and antibodies that inhibit NA activity (Ekiert et al., 2009; Tan et al., 2014; Wu and Wilson, 2017). Because of the dominant protective role of HA antibodies, and the relative ease of assaying virus neutralization rather than replication, most work thus far has focused on characterizing neutralizing antibodies as correlates of protection from influenza (Monto et al., 2015).

1.2 Influenza evolution

1.2.1 Antigenic shift

Influenza is a multi-segmented virus, and each segment is independently packaged into progeny virions. Co-infection of a cell with different strains can generate novel reassortant viruses (Ghedin et al., 2009; Nelson and Holmes, 2007). Influenza A viruses are subclassified based on HA (H1-H18) and NA (N1-N11) subtype, due to their antigenic significance, and a reassortment event that generates a novel HA and NA combination is known as **antigenic shift** (Han et al., 2023). These shifted variants typically emerge in the wild bird reservoir, where most subtypes are in constant circulation, or in intermediary domestic poultry and swine populations. Zoonotic transmission of a novel subtype to humans results in an influenza pandemic, followed by endemic seasonal circulation (Neumann and Kawaoka, 2015).

Influenza subtypes A/H1N1, A/H2N2, and A/H3N2 have circulated in the human population at various intervals since the 1918 pandemic. A/H3N2 and A/H1N1 have been circulating since 1968 and 1977, respectively, although A/H1N1 was replaced by the antigenically distinct A/H1N1pdm09 in 2009. Circulation of other subtypes in domestic poultry are closely monitored due to their heightened pandemic potential, especially H5, H7, and H9 strains. Many surveillance efforts are currently focused on the ongoing A/H5N1 pandemic in wild birds and domestic poultry, due to documented spillover into mammalian scavenger species and some isolated human infections (Puryear et al., 2023).

1.2.2 Antigenic drift

HA and NA are the main targets of the adaptive immune response, and are therefore under selective pressure to gain mutations that confer antibody escape. **Antigenic drift** refers to the gradual accumulation of antigenically significant mutations over time. As the immunodominant surface protein, HA evolves more rapidly than NA, although antigenic drift in both proteins contributes to immune escape (Sandbulte et al., 2011; Wan et al., 2019; Westgeest et al., 2012). Consequently, individuals are reinfected approximately every five years, despite retaining potent neutralizing antibodies from prior exposures (Kucharski et al., 2015; Ranjeva et al., 2019). Vaccines must also be continually updated to keep pace with antigenic advancement (Kilbourne et al., 2002).

Antigenically distinct strains emerge every few years, resulting in frequent turnover of the virus population (Bedford et al., 2014; Fitch et al., 1997; Strelkowa and Lässig, 2012). This turnover is reflected by influenza's distinctive ladderlike phylogeny, and differs between subtypes. H3N2 evolves rapidly, resulting in a spindly tree, whereas H1N1 shows coexistence of strains for longer periods of time (Bedford et al., 2014). Older children and adults are therefore more susceptible to reinfection by H3N2 than H1N1 (Bedford et al., 2015; Cobey and Hensley, 2017). Although these phylogenies shows gradual genetic change, antigenic evolution proceeds in punctuated bursts (Plotkin et al., 2002; Smith et al., 2004), with recurrent selective sweeps driven by just three to four beneficial mutations (on average) (Strelkowa and Lässig, 2012).

1.2.3 Characterization of major HA epitopes

Much effort has been focused on identifying the amino acid residues likely to drive antigenic change. Early experiments used selections with monoclonal antibodies (Gerhard et al., 1981) and structural analysis of genetic variation in recent strains (Wiley et al., 1981) to roughly define antigenic regions A through E on H3 HA. Incubating virus with a single antibody consistently selected for escape variants with single point mutations. However, no virus could escape neutralization by two or more antibodies targeting distinct antigenic regions (Caton et al., 1982; Laver et al., 1979; Yewdell et al., 1979), raising the following paradox: how does influenza evolve to escape the polyclonal antibodies in human serum?

More recent work has proposed two explanations. Hensley et al. (2009) found that passaging virus in immunized mice selected for mutations that increase receptor binding avidity, suggesting that serum escape is mediated by avidity, rather than antigenicity. The second explanation is that the immune response, despite being polyclonal, is actually highly focused on a single epitope. Analysis of human serum shows that the HA-binding antibody repertoire is dominated by just a few clonotypes, which persist over multiple years and vaccinations (J. Lee et al., 2019; Lee et al., 2016). Increased antigenicity of recent H1N1 and H3N2 strains have also been linked to mutations in a single HA epitope (Davis et al., 2018; Gouma et al., 2020; Huang et al., 2015; J. M. Lee et al., 2019; Linderman et al., 2014; Li et al., 2013; Zost et al., 2017), and studies with ferret sera suggest that just 7 residues near the receptor binding site are sufficient to determine antigenic change in H3 HA (Koel et al., 2013).

Identification of major HA epitopes can improve predictive models of influenza evolution. However, the fitness landscape of circulating strains and the immune landscape of the population are constantly changing, meaning that these antigenically significant sites are only useful if they are frequently updated by experimental data (Huddleston et al., 2020).

1.2.4 Influenza surveillance

Antigenic drift of influenza necessitates regular updates to the annual vaccine strains. To select the strain likely to offer the best protection in the upcoming season, surveillance laboratories must characterize the relative antigenicity of thousands of circulating strains (Barr et al., 2014). The antigenic distance between strains is mainly evaluated using antigenic cartography (Smith et al., 2004). This technique involves infecting ferrets with a strain of interest, harvesting the resulting serum, then measuring how well the antiserum can neutralize different strains via hemagglutination-inhibition (HI) assays. Strains are then positioned on a map based on measured antigenic distance.

There are two main limitations to this approach. First, HI assays measure how well an antibody binds to the sialic acid receptors on red blood cells. Although HI assays have been thoroughly validated as a correlate of protection for influenza infection, binding to sialic acid receptors is not a perfect proxy for actual cell entry. Neutralizing antibodies that prevent membrane fusion, for example, will not be detected by HI. Recent antigenic change in H3 HA also prevented binding to red blood cells used in HI assays, thus rendering the assay ineffective for modern H3 HA variants.

The second limitation is the use of singly-infected ferret antiserum, which does not account for the impact of exposure history on the human antibody response. Previous work has shown that ferret antisera and sera from naive children generate similar antigenic maps (Fonville et al., 2016), while adults with complex exposure histories often have highly divergent antibody responses (Fonville et al., 2014; J. M. Lee et al., 2019; Linderman et al., 2014; Li et al., 2013). Relying on ferret antiserum can obscure antigenically significant changes in emerging strains, sometimes leading to especially low vaccine effectiveness (Linderman et al., 2014). Better understanding how exposure history impacts individual antibody responses, and systematically assessing this heterogeneity in the population, has the potential to significantly improve surveillance and vaccine selection.

1.3 Effects of exposure history on individual antibody responses

The immune response to influenza infection or vaccination is shaped by an individual’s early exposure history, a phenomenon initially described as ‘original antigenic sin’ (Davenport et al., 1953; Davenport and Hennessy, 1956; Francis, 1960). Subsequent research has developed the more nuanced model of ‘antigenic seniority’, where static neutralizing titers are highest against strains circulating early in one’s lifetime, then steadily decline with age at time of exposure (Fonville et al., 2014; Kucharski et al., 2015; Lessler et al., 2012; Yang et al., 2020). Although the exact underlying mechanisms are still unclear, infection and vaccination appear to induce transient back-boosting of antibodies generated by prior exposures to antigenically similar strains, likely due to memory cell stimulation and antibody recall (Fonville et al., 2014). This boosting also interferes with the generation of new antibodies specific to the infecting strain (Arevalo et al., 2020; Kucharski et al., 2015; Lessler et al., 2012). Repeated back-boosting of cross-reactive antibodies over the course of one's lifetime would thereby result in an immune hierarchy favoring older strains.

The phenomenon of antigenic seniority implies that immune specificity varies, at least in part, by birth cohort. In a recent paper analyzing serum neutralization titers to different HA or NA clades, Kim et al. (2023) found that individuals within the same age group have significantly correlated titers, and that average susceptibility to a given strain varies between cohorts. This population-level heterogeneity has been proposed as a factor shaping the antigenic evolution of influenza (Nakajima et al., 2000; Oidtman et al., 2021; Sato et al., 2004; Worby et al., 2015). As previously discussed, human serum that is highly focused on one immunodominant epitope can select for new escape variants.
In a heterogeneous population, smaller cohorts that target similar epitopes may be responsible for driving antigenic drift.

The degree of serum focusing can also vary dramatically between individuals. Deep mutational scanning has shown that single mutations in HA can sometimes lead to large drops in neutralization by human serum, but the magnitude is variable even among just four adults (J. M. Lee et al., 2019). Because narrowly focused serum can more efficiently select escape mutants than broadly neutralizing serum (Davis et al., 2018), this heterogeneity has important consequences for antigenic drift (Nakajima et al., 2000). Prior work suggests that serum from children may be more targeted, due to their limited exposure history (Meade et al., 2020; Natali et al., 1981; Ranjeva et al., 2019; Sato et al., 2004), but thus far there has been no systematic analysis of how viral mutations affect neutralization by serum from different age cohorts. In the following chapters, I will describe our use of deep mutational scanning to address this gap and deepen our understanding of the selective pressures that shape influenza evolution.

1.4 High-throughput methods to characterize antibody targeting of HA

1.4.1 Binding-based assays

One general approach to mapping sites targeted by serum antibodies is to analyze binding to the antigen of interest. A recently developed method, EM polyclonal epitope mapping (EMPEM), uses high-throughput imaging of antibodies in complex with HA (Han et al., 2021). EMPEM can identify precise antibody binding sites on a structure, thus providing a more complete understanding of the serum antibody repertoire. Other high-throughput approaches use deep mutational scanning, which involves generating libraries of variants carrying mutations to a protein of interest. By incubating the library with a serum of interest, then selecting for and sequencing variants that were bound or unbound by antibodies, researchers can map mutations that confer escape from antibody binding. Both yeast display (Greaney et al., 2021; Starr et al., 2022, 2021, 2020) and cell surface display (Salamango et al., 2016; Steichen et al., 2016; Tan et al., 2023) libraries have been used to analyze antibody targeting of other virus glycoproteins.

However, antibodies that successfully bind HA do not necessarily have high neutralizing activity. As previously discussed, antigenic drift of influenza suggests that the human serum response to influenza is immunodominant, where a small subset of antibodies confer most neutralizing activity. Although non-neutralizing antibodies are also correlates of protection from infection and disease, neutralizing antibodies directed towards potent HA head epitopes exert stronger selective pressure on influenza

(Krammer, 2019). Assays that specifically map the neutralizing antibody response are therefore essential for better understanding and predicting antigenic drift.

1.4.2 Neutralization-based assays

Deep mutational scanning (DMS) has previously been used to map both functional effects (Doud and Bloom, 2016; Lee et al., 2018) and antigenic effects (Doud et al., 2018, 2017; J. M. Lee et al., 2019) of all mutations to influenza HA. These studies used replication-competent libraries of influenza, grown using the non-HA segments from the lab-adapted strain A/WSN/1933, and randomly mutagenized HA from a recent strain of interest, such that virion expresses a different HA variant. While yeast and cell surface display libraries express an antigen outside of its native context, these influenza libraries are functionally intact and can therefore be used to assay neutralization. After serum or antibody incubation, cells are infected with the antibody-bound library, and escape mutations can then be identified by sequencing viral RNA from infected cells.

Prior influenza DMS experiments demonstrated that human serum, despite being polyclonal, is often highly targeted (J. M. Lee et al., 2019). Mutation effects varied dramatically even among four individuals, where a mutation causing almost complete escape from one sera might have no effect on another. Many of the sites under strong immune selection also gained mutations in subsequent years. DMS of HA has thereby deepened our understanding of the polyclonal serum antibody repertoire, and appears to somewhat predict influenza evolution. However, due to technical limitations, this study was limited to assaying single-mutant variants, could only resolve escape and not neutralization-sensitizing mutations, and mapped escape from just four sera.

In **Chapter 2**, I will describe the development of an improved mutagenic library that incorporates barcodes into the HA construct. This barcoding advancement overcomes many of the technical limitations of the original HA DMS library, and improves sample throughput while also allowing for absolute quantification of both escape and sensitizing mutations. Then, in **Chapter 3**, I use this library to map neutralizing antibody specificity against 30 sera from three defined age cohorts. The resulting escape profiles demonstrate that the antigenic effects of mutations differ between age cohorts, and suggest ways in which these heterogeneous selective pressures may shape influenza evolution.

CHAPTER 2: Development of a barcoded library of influenza variants with multiple mutations to hemagglutinin

This chapter includes some results and figures selected from a submitted manuscript:

> Welsh FC, Eguia RT, Lee JM, Haddox HK, Galloway J, Van Vinh Chau N, Loes AN, Huddleston J, Yu TC, Le MQ, Nhat NTD, Le Thanh NT, Greninger AL, Chu HY, Englund JA, Bedford T, Matsen FA, Boni MF, Bloom JD, **Age-dependent heterogeneity in the antigenic effects of mutations to influenza hemagglutinin**, *bioRxiv*, 2023.

Prior work has described mutational antigenic profiling of influenza libraries to study viral escape from polyclonal sera (J. M. Lee et al., 2019). Due to technical limitations, this approach was limited to single-mutant libraries, could not resolve sites that make the virus more sensitive to neutralization, and could not reliably compare selection magnitude between separate experiments. Sample preparation was also highly effort-intensive, as the full HA gene had to be amplified as a collection of short, tiled fragments for sequencing (Doud and Bloom, 2016). We have addressed these limitations by developing an updated influenza library system that incorporates barcodes for each variant gene. This advance also allows us to incorporate an external, non-neutralized standard, which can be used to correct for variation in viral growth and generate absolute escape measurements.

Here, we generate a library of H3 HA variants with multiple mutations in the background of A/Hong Kong/45/2019. We validate our library escape measurements using monoclonal antibodies and antibody cocktails, then map serum escape from a child and adult. The antigenic effects of mutations differ between sera, and are recapitulated by traditional neutralization assays. These improved deep mutational scanning libraries allow for precise analysis of serum antibody targeting at a much larger scale than previously possible, making them a powerful tool for exploring population-level heterogeneity in the immune response to influenza.

2.1 Results

2.1.1 Engineering of a barcode H3 HA sequence without disrupting the packaging signal

To systematically map how mutations to hemagglutinin affect neutralization by human sera, we developed barcoded deep mutational scanning libraries in the HA of A/HongKong/45/2019, which was the H3N2 component of the 2020-2021 influenza vaccine. Influenza is multi-segmented, and proper packaging of vRNA into the virion relies on segment-specific RNA sequences called “packaging signals.” These span both coding and noncoding regions at the 3’ and 5’ ends of the vRNA segment (Li et al., 2021). To insert a barcode without disrupting vRNA packaging, we designed a chimeric construct with a duplicated 5’ packaging signal (negative sense), based on chimeric influenza genes designed by (Gao and Palese, 2009)). This chimeric construct uses the packaging signals from the lab-adapted strain A/WSN/1933 HA. The other 7 gene segments used to generate this virus are also from this lab-adapted strain, and prior work suggests that matched packaging signals improve virus rescue efficiency.

The native 3’ packaging signal (negative sense) of H3 HA is replaced by the corresponding sequence from A/WSN/1933 HA (**Figure 2.1A-B**). This is followed by the unmodified A/Hong Kong/45/2019 ectodomain and C-terminal coding region. Although prior barcoded constructs have also replaced the C-terminal region with the corresponding WSN sequence (Wu et al., 2017), we found that maintaining the native H3 HA sequence was essential for virus growth (**Figure 2.1C**). The 90-nucleotide native packaging sequence in the 5’ coding region is synonymously recoded to avoid interference between duplicated sequences. We place the full 150-nucleotide 5’ packaging signal from WSN, spanning both coding and noncoding sequence, after the stop codon of the HA gene **(Figure 2.1B)**. This construct design allows for insertion of a barcode and priming sequence after the stop codon, while leaving the duplicated packaging sequence fully intact.

An extra stop codon was added to the end of the coding sequence to minimize translational readthrough. This is followed by a random 16-nucleotide barcode, then a priming sequence to amplify barcodes for Illumina sequencing, and finally the full-length WSN-HA packaging signal. Homology between the native and duplicated packaging signal could lead to barcode loss; we therefore included a stop codon in the duplicated packaging signal that will be in-frame if it replaces the native packaging signal. This ensures that loss of the barcode will generate a truncated, non-functional HA protein (**Figure 2.1B**).

2.1.2 Incorporation of non-neutralized standards allows for generation of quantitative escape measurements

The generation of an HA construct with barcodes linked to specific variants allows for incorporation of a barcoded growth standard. This is a construct that will not be neutralized by human serum, and can therefore be used to correct for stochastic variation in virus replication across different conditions and experiments. We can then calculate absolute, quantitative escape scores, rather than relative magnitudes used in prior work (J. M. Lee et al., 2019). The barcoded standards are spiked into the library at a low level (approximately 1%), and used to normalize the barcode frequency between mock selections and serum selections. As more of the library is neutralized at increasingly potent serum concentrations, the proportion of library barcodes to standard barcodes will decrease in a consistent manner (**Figure 2.2A-B**).

We included standards at two stages: infection with the virus library, and harvest of RNA from infected cells. The virus standard is identical to the WSN-flanked H3 HA library construct, but uses the ectodomain and 5' coding region from A/Turkey/Mass/1975 H6 HA, a low-pathogenicity avian HA strain (Sandbulte et al., 2009) **(Figure 2.2A)**. Note that although the HA of this viral standard is from a low pathogenicity avian-strain, the virus itself is still only biosafety-level 2, because the seven non-HA gene segments and the H6 HA packaging signals are derived from the lab-adapted WSN strain. This standard is expected to have similar infection dynamics as the library H3 HA variants. We tested neutralization activity of several sera against the WSN-H6 HA standard, and found that any serum used at a dilution of approximately 1:100 or higher begins to neutralize WSN-H6 HA.

Therefore, to allow for analysis of low-potency sera that must be at high concentration for selections, we also included an RNA spike-in standard. This is an RNA fragment that is identical to the WSN-flanked H3 HA sequence from the 5' coding region to the 5' UTR, but replaces the H3 ectodomain with GFP, in order to keep the RNA spike-in standard a similar length to the HA RNA (**Figure 2.2A**). The fragment is generated with a pool of approximately 100 random barcodes. We add a standardized amount of RNA spike-in to the lysis buffer used for harvest of cellular RNA, so that it's equally distributed between wells in each experiment. We tested a range of concentrations, and found that the RNA spike-in behaves similarly to the WSN-H6 HA standard when 0.05ng of RNA spike-in is added per well (**Figure 2.2C**).

2.1.3 Designing H3 HA libraries to enrich for functional, antigenically significant variants

In a library of HA variants with combinations of mutations, the rate of non-functional variants will be much higher than a single-mutant library, as a variant carrying *any* deleterious mutation will likely be non-functional. We therefore maximized the number of functional variants in the library by excluding mutations with highly deleterious functional effects. Using H3 amino acid preferences calculated from previous deep mutational scanning of A/Perth/16/2009 H3 HA (Lee et al., 2018), we retained mutations with functional effects more positive than the 0.75 quantile of stop codon effects (**Figure 2.3A**).

We then identified excluded mutations that were tolerated in other sequence contexts by analyzing natural H3 sequences circulating from 1968 to 2021. After basic quality control and exclusion of egg-passaged strains, we aligned 7,597 H3 variants, and identified 44 excluded mutations that were present at a frequency of at least 0.5% in natural sequences. These were added back into the library, for a total of 2,505 targeted mutations (**Figure 2.3C**).

To ensure that antigenically significant mutations were highly represented in multiple backgrounds, we biased mutagenesis towards 22 selected epitope sites (**Figure 2.3B**). These were defined as sites that had been mutated in at least one major clade from 2015-2021. These sites would be randomly mutated, regardless of functional effects, at a higher rate than the general targeted mutations.

We independently mutagenized two separate plasmid pools using primers targeting the selected sites, generating two fully independent libraries with an average of 2-3 mutations per variant (**Figure 2.3D**). The resulting HA libraries were then incorporated into replication-competent influenza virions, with the other viral genes derived from the lab-adapted WSN strain for biosafety reasons. Virus libraries were passaged at a low MOI such that each cell would be infected by one virion at most. This passaging creates a genotype-phenotype link, where the HA protein expressed on the surface of the virion matches its packaged HA gene. The final passaged libraries each contained approximately 30,000 barcoded H3 HA variants, due to loss of non-functional HA variants during these growth steps.

2.1.4 Validation of library approach using monoclonal antibodies and cocktails

Using these deep mutational scanning libraries, we were able to map mutations across HA that increase either resistance or sensitivity of the virus to neutralization. We began by mapping escape from two previously characterized monoclonal antibodies, 1C04 and 5G04 (Zost et al., 2019), to confirm that our approach can accurately identify known antigenic mutations (**Figure 2.4A**). The mutation "escape scores" in these maps are roughly proportional to the log fold change in neutralization titer (Yu et al., 2022). Positive escape scores indicate mutations that confer resistance to neutralization, while negative scores indicate mutations that increase neutralization.

The escape map for antibody 1C04 shows precise targeting of a few proximal sites in antigenic region E (**Figure 2.4B**). The strongest escape mutations are at site 82 and 83, which is consistent with prior results showing that the mutation K82E strongly disrupts 1C04 binding (Zost et al., 2019). There is some background signal at additional sites, but the effects are minimal when compared to the dominant escape sites, and likely reflect either noise from the assay or changes in binding avidity (Hensley et al., 2009).

Antibody 5G04 is even more highly targeted, with almost all escape mutations at site 227 (**Figure 2.4C**). In contrast to 1C04, we also see negative signal at site 160, indicating that these mutations increase neutralization sensitivity. Mutations at site 160 eliminate an N-linked glycan at site 158, which was introduced in 2014 and masks antigenic region B (Gouma et al., 2020; Zost et al., 2017). Previous work mapped 5G04 binding to an earlier H3 HA strain, A/Victoria/210/2009, and found that mutating a number of sites in the receptor binding region strongly disrupted 5G04 binding. Notably, introducing the 158 glycan with the mutation K160T also reduced antibody binding (Zost et al., 2019). These results suggest that 5G04 typically has a larger binding footprint, but is sterically impeded by the 158 glycan, and that removing this glycan increases the potency of 5G04 against A/Hong Kong/45/2019. By mapping both escape *and* sensitizing mutations, we thus gain a more nuanced understanding of antibody targeting, especially how it may be affected by virus evolution.

We also mapped escape from an antibody cocktail, to test whether we can resolve multiple targeted epitopes using this library approach. Antibody selections using a 1:1 mixture of 1C04 and 5G04 successfully generated the expected escape map, with signal at sites 82 (1C04) and 160/216/227 (5G04) (**Figure 2.4D**). The magnitude of scores from 1C04 and 5G04 are similar, which reflects that the antibodies were present at equal potency. However, when the antibodies were mixed at a 3:1 ratio (based on neutralization potency), only the dominant antibody can be identified on the resulting escape map. It's important to note that a cocktail of highly targeted antibodies, targeting spatially distant epitopes, may be an extreme representation of polyclonal serum. Regardless, these results demonstrate that we can reliably identify sites targeted by multiple potent antibodies, but cannot always resolve antibodies with very minor contributions to overall serum activity.

2.1.5 Serum escape analysis is consistent with likely exposure history for a child and adult

After validating the barcoded deep mutational scanning results using known monoclonal antibodies, we then mapped two sera collected in 2020 in Seattle, one from a 4-year-old child and one from a 42-year-old adult (**Figure 2.5A-B**). We validated that the escape scores measured by deep mutational scanning correlated strongly with the results of traditional neutralization assays (**Figure 2.5C-D**).

The escape map for the child’s serum (serum 3944) is dominated by positive escape at just a few sites, most prominently site 189 (H3 numbering), followed by sites 159 and 160 (**Figure 2.5A**). There is one region (sites 221-228) where mutations modestly increase sensitivity to neutralization. This is likely by decreasing receptor avidity rather than directly affecting antibody binding (Hensley et al., 2009), as mutations at these sites are known to alter receptor binding (Lin et al., 2012; Matrosovich et al., 2000) and have consistent neutralization-sensitizing effects across all mapped sera (see Chapter 3). At some sites (e.g. site 189), virtually any amino acid mutation causes escape, while at other sites only specific mutations cause escape (e.g. S145K) (**Figure 2.5A**).

In contrast, the escape map for the adult’s serum (serum 197C) reveals that some mutations increase neutralization sensitivity while others cause neutralization escape. Specifically, mutations at sites 121-124 and 159-160 increase neutralization by this serum (reflected by negative escape scores), whereas mutations at site 189 cause escape (**Figure 2.5B**). The sites with strongest sensitizing effects all correspond to regions where H3N2 evolution has previously introduced N-linked glycans that mask antibody epitopes. In 1997, H3N2 influenza acquired an N-linked glycan at site 122 that masked a key antigenic site. Mutations at sites 122 and 124 that eliminate this glycan greatly increase sensitivity to neutralization, suggesting that antibodies targeting this now-inaccessible site are still present at high levels in the serum. Similarly, the 158 glycan was introduced in 2014 and masked an immunodominant antigenic region (Gouma et al., 2020a; Zost et al., 2017). Increased neutralization sensitivity from mutations that remove glycosylation motifs introduced many years ago is consistent with the well-described concept of antigenic seniority, where individuals retain neutralizing antibodies against strains encountered early in life (Lessler et al., 2012).

2.2 Discussion

We have described the development of a new influenza deep mutational scanning system that links each variant in the library to a unique nucleotide barcode. This barcoding advancement dramatically improves assay throughput by simplifying sample preparation and sequencing. It also allows for generation of quantitative escape measurements, and identification of mutations that increase virus neutralization, thus representing a significant advance over prior influenza deep mutational scanning methods. Here, we have demonstrated the utility of this system using both monoclonal antibodies and human serum samples.

The question of how influenza successfully evolves to escape neutralization by polyclonal serum has been a longstanding paradox in the field, as selections with just two antibodies targeting distinct regions prevent the growth of escape variants (Caton et al., 1982; Laver et al., 1979; Yewdell et al., 1979). Previous work has suggested that the serum antibody response to influenza is highly immunodominant (Davis et al., 2018; Huang et al., 2015; J. Lee et al., 2019; Lee et al., 2016; Linderman et al., 2014; Li et al., 2013). Consistent with this hypothesis, deep mutational scanning of influenza H3 showed that single amino acid mutations could lead to significant reduction of human serum neutralization (J. M. Lee et al., 2019). However, it was not clear whether these experiments were powered to resolve multiple targeted epitopes, as the library was limited to single-mutant variants.

We have now generated libraries with an average of 2-3 mutations per variant, and successfully mapped escape from an antibody cocktail targeting two spatially distinct regions. These results demonstrate that our deep mutational scanning approach will identify all sites that contribute substantially to serum neutralization activity. The relative magnitude of escape scores is meaningful: antibodies at equal potency will generate roughly equivalent signal. When one antibody was added to the cocktail at 3-fold lower potency than the other, we were unable to resolve the low-potency antibody over background noise. It's unclear exactly how this translates to serum, as a cocktail of two potently neutralizing antibodies is much more highly targeted than polyclonal serum, but these results suggest that we cannot reliably identify minor epitopes.

We show that the barcoded deep mutational scanning libraries can be used to generate quantitative escape measurements that include neutralization-sensitizing mutations. With this advancement, we gain a more nuanced understanding of serum escape from individuals of different ages. Both the child and adult serum are mainly escaped by mutations at site 189. However, the magnitude of escape scores from the child serum are much higher, which is consistent with previous work showing that children mount a strong, highly specific neutralizing antibody response (Islam et al., 2017; Meade et al., 2020; Nakajima et al., 2000; Ranjeva et al., 2019). The negative signal at sites 121-124 and 160 in the adult serum suggests a broader antibody response targeting 3 distinct regions, though only one of those epitopes is still accessible. The presence of highly sensitizing mutations in the adult serum, but not the child serum, is consistent with their exposure histories, as it's unlikely that the child has had prior exposures to strains antigenically distinct from A/Hong Kong/45/2019.

Although the sample size is small, these results demonstrate the utility of the barcoded deep mutational scanning system for mapping serum targeting of influenza HA. Resolution of sensitizing and escape mutations allows us to identify both sites under positive selection, and sites that are constrained from reverting to prior amino acid identities. The ability to generate more comprehensive, quantitative escape measurements, combined with the relative ease of sequencing, make these barcoded libraries a powerful high-throughput tool for analyzing population-level heterogeneity in the serum antibody response to influenza.

2.3 Figures

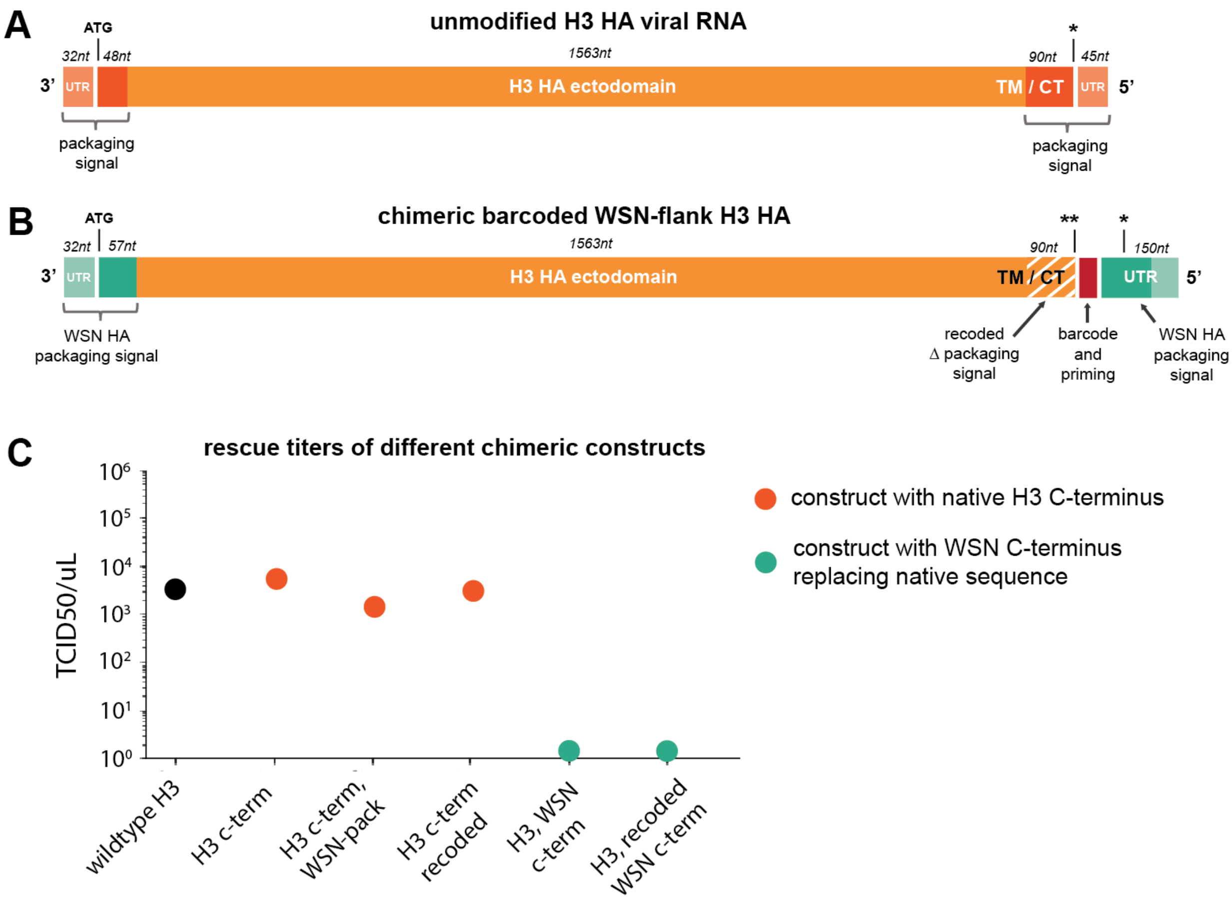


Figure 2.1: Design of chimeric barcoded WSN-flank H3 HA construct.

All segments are shown in the reverse orientation of the negative-sense viral genome, with the 3' to 5' labels indicating ends in the negative-sense viral genome. Stop codons are denoted as asterisks. (A) Schematic of normal unmodified H3 HA. Packaging signals, untranslated regions (UTRs), ectodomain, transmembrane domain (TM), and cytoplasmic tail (CT) are labeled for a normal H3 HA. Corresponding nucleotide length is labeled based on sequence of A/Hong Kong/45/2019. (B) Schematic of chimeric WSN-flank H3 HA used for barcoding. The 3' and duplicated 5' packaging signals are taken from the lab-adapted A/WSN/1933 HA. The native packaging sequence in the 5' coding region is synonymously recoded to avoid interference between duplicated sequences. (C) Rescue titers for different chimeric HA constructs. Virus was generated with the other seven gene segments from A/WSN/1933 HA, rescued after 48 hours, and titered by TCID50 assay. All constructs that retained the wildtype H3 amino acid sequence at the C-terminus grew to similar titers as the wildtype H3 HA control. Constructs where the C-terminus was replaced with the WSN-HA sequence had undetectable titers.

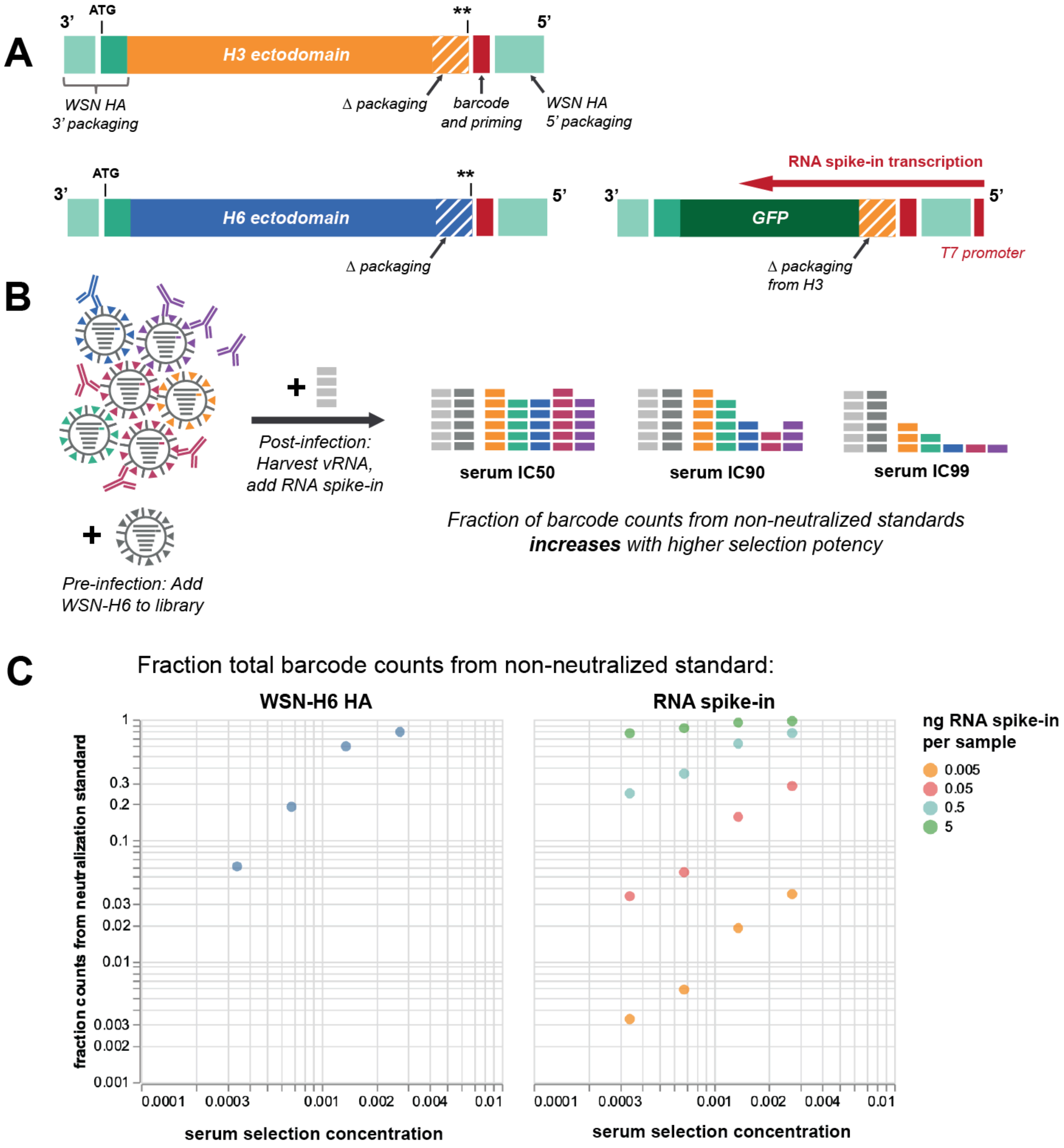


Figure 2.2: Incorporation of non-neutralized standards to generate quantitative escape measurements.

(A) Design of WSN-H6 HA and RNA spike-in neutralization standards. The WSN-H6 HA standard is identical to the H3 HA library construct, but uses the ectodomain and 5' region from A/Turkey/Mass/1975, a low-pathogenicity avian HA strain (GenBank Accession AB296072.1) (Sandbulte et al., 2009). The RNA spike-in is generated by *in vitro* transcription of a construct where the H3 HA ectodomain is replaced by GFP. (B) WSN-H6 HA is spiked into the H3 HA library before incubating with serum and infecting cells. The RNA spike-in standard is added when harvesting cellular RNA at 13 hours post-infection. As neither standard is neutralized by human sera, the number of barcodes from the standards remains constant, while H3 HA barcodes decrease at higher concentrations of neutralizing antibodies. The overall fraction of standard barcodes should therefore increase at a constant rate with increasing serum potency. (C) At a range of serum concentrations used for selections, the fraction of counts from the WSN-H6 HA neutralization standard increases in a consistent manner as the H3 HA variant library is more potently neutralized. Spiking in 0.05ng of RNA standard per sample achieves the same effect.

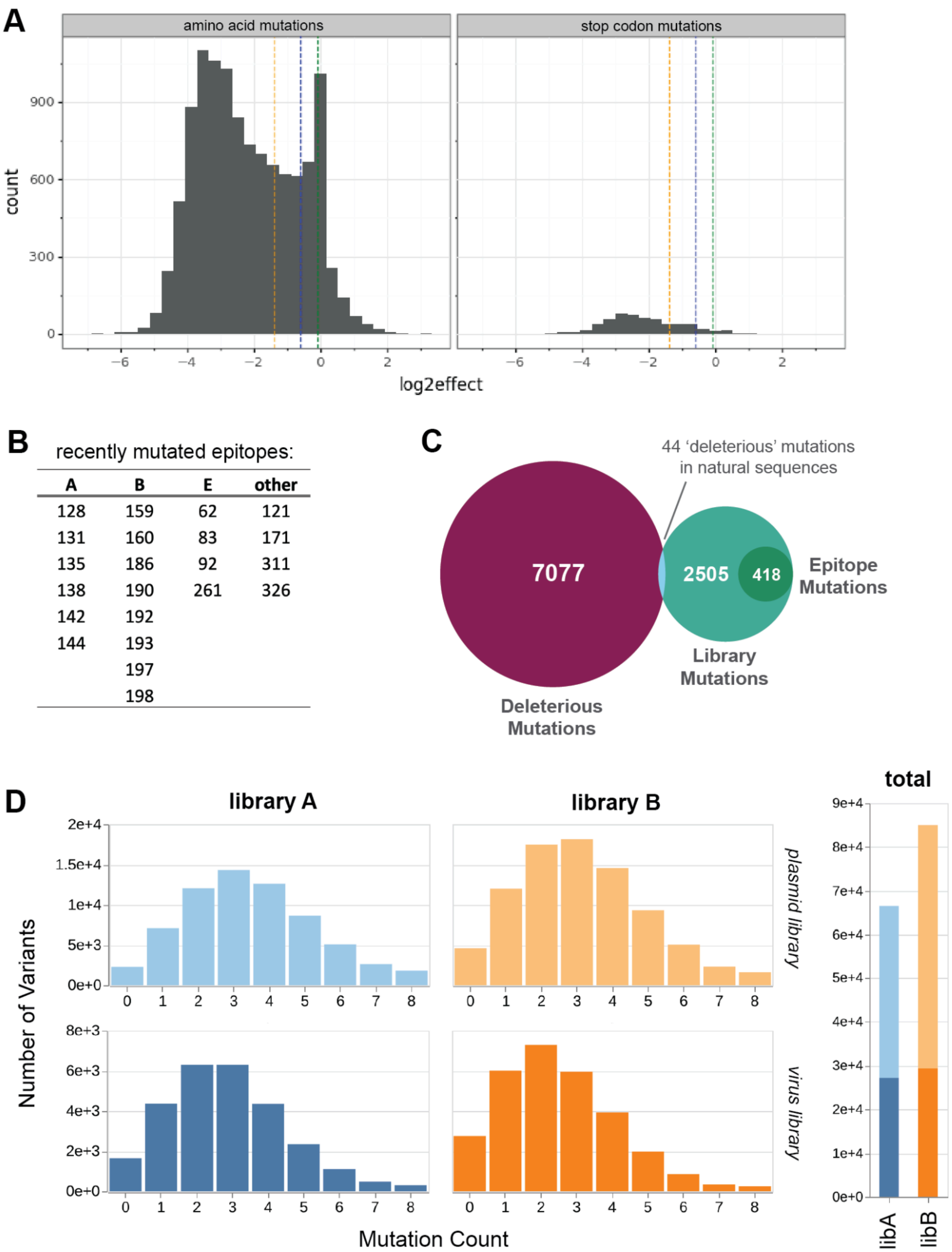


Figure 2.3: Composition of H3 HA libraries.

(A) Functional effects of all single amino acid mutations to A/Perth/16/2009 H3 HA, calculated by (Lee et al., 2018). Vertical lines delimit 75%, 90%, and 95% of the most deleterious stop codons. These quantiles are transposed onto all amino acid mutations, and 7,077 mutants corresponding to the 75th quantile of stop codons were excluded from the current library. (B) List of epitope sites, defined as sites that have been mutated in at least one major H3 HA clade between 2015 and 2021. These sites are randomly mutagenized in the library at a higher rate than the general non-deleterious mutations. (C) Visualization of types of mutations included in the library. In total, 7,077 mutations in the H3 HA ectodomain region were identified as functionally deleterious. 44 of these excluded mutants were identified in H3 HA strains circulating between 1968 and 2021, and added back into the library. This left 2,505 mutations to target in the library. 418 mutations at epitope sites were introduced at a higher frequency than general library mutations. (D) Distribution of mutations per variant in library A and library B, shown for both the starting plasmid libraries and the virus libraries after rescue and passaging. The bar plot on the right shows the total number of variants in the plasmid library (approx. 65,000 and 85,000) and the virus library (approx. 30,000 in each library).

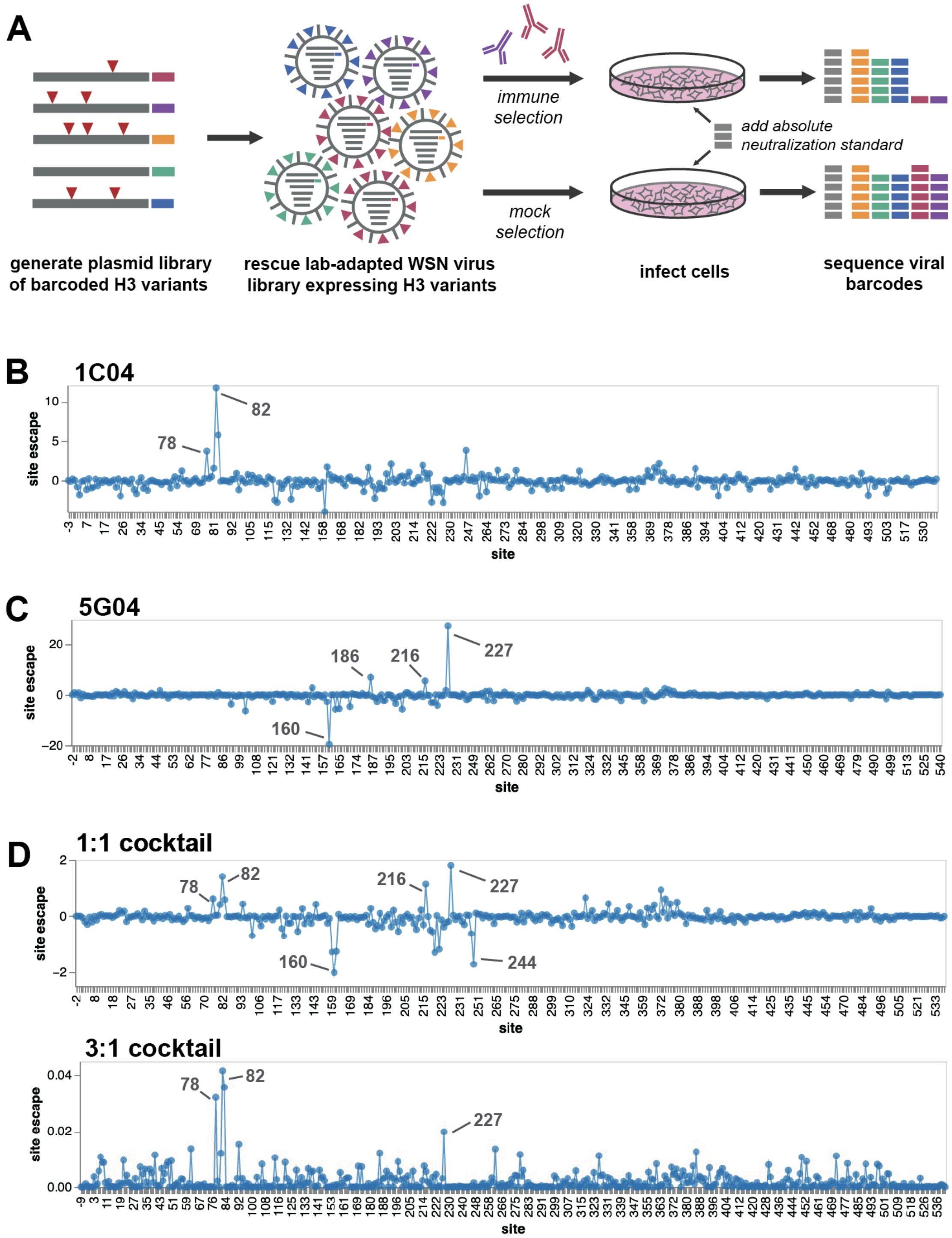


Figure 2.4: Validation of deep mutational scanning using monoclonal antibodies and cocktails.

(A) Experimental workflow for mapping HA serum escape. The virus library is incubated with either the serum of interest, or media alone, then added to cells. At 13 hours post-infection, RNA is harvested from cells, and viral barcodes are reverse-transcribed and amplified by PCR. Escape scores are calculated for all sampled mutations by sequencing barcodes from variants that successfully infected cells in the serum-selection condition, and comparing to the mock-selection. A neutralization standard is used to convert relative sequencing counts into absolute neutralization (**Figure 2**). Escape maps from monoclonal antibody 1C04 (B), 5G04 (C), and 1C04-5G04 cocktails mixed at either 1:1 potency, or with 1C04 at 3-fold higher potency (D). Line plots show the summed effects of all sampled mutations at each site ("site escape", roughly proportional to fold change in neutralization), with positive values indicating neutralization escape and negative values indicating increased neutralization. Escape map from the 3:1 cocktail (D) is floored at 0, to better resolve sites of positive escape.

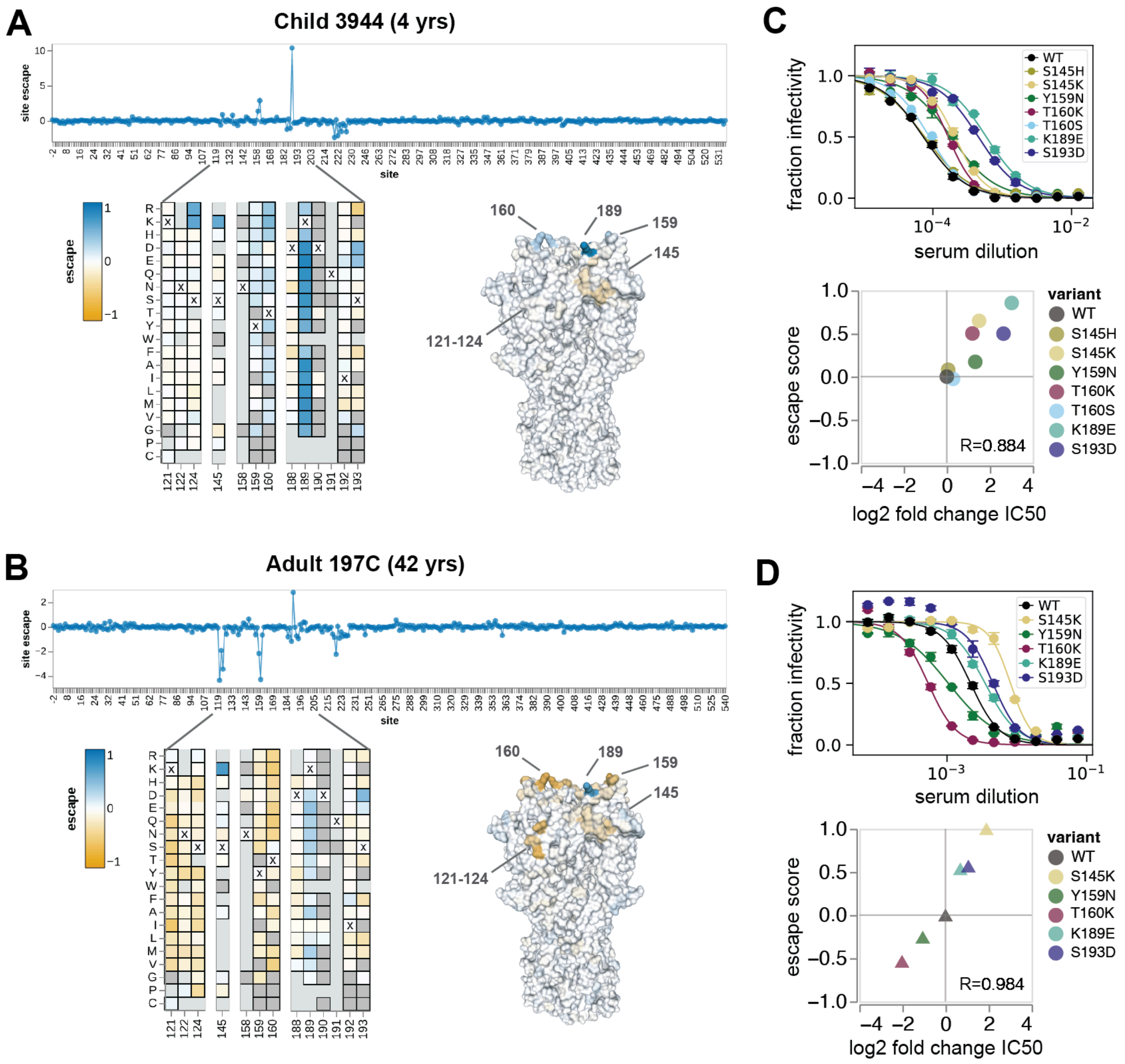


Figure 2.5: HA mutations have heterogeneous effects on neutralization by different sera.

Serum escape maps from a child (A) and an adult (B). The line plot shows the summed effects of all sampled mutations at each site ("site escape", roughly proportional to fold change in neutralization), with positive values indicating neutralization escape and negative values indicating increased neutralization. Heatmaps show the effects of individual mutations for key sites. In the heatmaps, X indicates the wildtype amino acid in A/Hong Kong/45/2019 at each site, dark gray indicates mutations measured to be highly deleterious to HA-mediated infection, and light gray indicates mutations not sampled in the libraries. The structures are colored by the summed site escape scores, and show the HA structure for the A/Victoria/361/2011 H3 HA (PDB 4O5N). (C-D) Validations of escape scores by traditional neutralization assays for selected mutants against the child (C) and adult (D) sera. Correlation plots show the escape scores versus the log2 fold change in IC50 between the wildtype library strain and the mutant of interest. R indicates the Pearson correlation.

2.4 Tables

Primers for amplifying WSN-flanked H3 HA coding sequence	
Forward linearizing primer (primer_088)	gcaaaactactggtcctgttatatgcatttgtagc
Reverse linearizing primer (primer_089)	ctcattatatacagatgttgcatcggatgttgcc
Primers for mutagenizing H3 HA ectodomain	
General mutagenic primer pool (forward and reverse)	https://github.com/dms-vep/flu_h3_hk19_dms/blob/main/library_design/hk19_primers.csv
Epitope primer pool (forward and reverse)	https://github.com/dms-vep/flu_h3_hk19_dms/blob/main/library_design/hk19_single_epitope_primers.csv
Paired epitope primer pool (forward and reverse)	https://github.com/dms-vep/flu_h3_hk19_dms/blob/main/library_design/hk19_paired_epitope_primers.csv
Primers for barcoding WSN-flanked H3 HA sequences	
Forward linearizing primer (primer_088)	gcaaaactactggtcctgttatatgcatttgtagc
Reverse barcoding primer (primer_090)	acactctttccctacacgacgctcttccgatctNNNNNNNNNNNNNNNNctcattatatacagatgttgc
Primers for amplifying H3 HA C-terminus with GFP overlap	
Forward GFP overlap primer (primer_115)	TGGACGAGCTGTACAAGTAATAGgttgagctgaagtcaggatacaaagattggatc
Reverse linearizing primer (primer_089)	ctcattatatacagatgttgcatcggatgttgcc
Primers for generating RNA spike-in template with T7 promoter	
Forward U12-annealing primer (primer_113)	agcaaaagcaggggaaaataaaaacaacc
Reverse T7-appending primer (primer_116)	TTACGATAATACGACTCACTATAGGGagtagaaacaagggtgtttttccttatatttctg
Primers for reverse transcription of viral barcodes	
Viral mRNA-annealing RT primer (primer_110)	ggcaacatccgatgcaacatctgtatataatga
Round one Illumina barcode sequencing preparation primers	
Illumina round 1 forward primer (primer_098)	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTggcaacatccgatgcaacatctgtatataatgag
Illumina round 1 reverse primer (primer_099)	acactctttccctacacgacgctcttccgatct

Round two Illumina barcode sequencing preparation primers	
Unique i5 indexing prime	AATGATACGGCGACCACCGAGATCTACACxxxxxxxxxxACACTCTTTCCCTACACGACGCTCTTCCGATCT Where "xxxxxxxxxx" are NextFlex indices unique to each primer.
Unique i7 indexing primer	CAAGCAGAAGACGGCATACGAGATtatcttcagcGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT Where "xxxxxxxxxx" are NextFlex indices unique to each primer.

Table 2.1: Primers and sequences referenced in methods.

2.5 Methods

2.5.1 Biosafety

All mutant viruses in the deep mutational scanning studies derived their HAs from a recent human H3N2 vaccine strain, A/Hong Kong/45/2019, and the remaining seven genes from the lab-adapted A/WSN/1933 (H1N1) strain. These viruses are therefore classified as biosafety-level 2, and the work was approved by the Fred Hutchinson Cancer Center biosafety committee.

2.5.2 Antibodies and human sera

The monoclonal antibodies in **Figure 2.4** are described in Zost et al. (2019). These antibodies were isolated from peripheral blood mononuclear cells of human donors 7 days post-vaccination with the 2010–2011 influenza vaccine containing the A/Victoria/210/2009 strain as the H3N2 component, using the approach in (Smith et al., 2009). The VH and VL chains were amplified using single-cell RT-PCR, and cloned into human IgG expression vectors. To produce the monoclonal antibodies, 293 T cells were transfected with plasmids encoding the heavy and light chains, and the antibodies were purified using protein A/G affinity purification.

The human sera in **Figure 2.5** were collected in Seattle, Washington in 2020. Serum 3944 from the 4-year-old child is residual sera collected at Seattle Children's Hospital with approval from the Human Subjects Institutional Review Board. Serum 197C from the 42-year-old adult was collected as part of the prospective longitudinal Hospitalized or Ambulatory Adults with Respiratory Viral Infections (HAARVI) cohort study of individuals with SARS-CoV-2 infection. Written informed consent was obtained for each participant.

2.5.3 Generation of plasmid library of barcoded H3 HA variants

See https://github.com/jbloomlab/CodonTilingPrimers for a general description of the PCR mutagenesis strategy we use here. Note that we are using primers that generate targeted amino-acid mutations, as described above.

To mutagenize the A/Hong Kong/45/2019 HA sequence, we first amplified the HA coding sequence from the plasmid containing the barcoded WSN-flanked H3 HA gene under control of the pHH21 promoter. The PCR was performed with the following conditions: PCR mix: 20μL H2O, 1.5μL 10μM forward linearizing primer (primer_088), 1.5μL 10μM reverse linearizing primer (primer_089), 2μL 5ng/μL template plasmid, and 25μL 2x KOD Hot Start Master Mix (Sigma-Aldrich, Cat. No. 71842). Cycling conditions: (1) 95C/2min (2) 95C/20sec (3) 70C/1sec (4) 54C/30sec, cooling at 0.5C/sec (5) 70C/40sec (6) Return to Step 2 x19.

The amplified, linearized chimeric H3 HA sequence was gel purified using the NucleoSpin Gel and PCR Clean-up kit (Takara, Cat. No. 740609.5) and then purified using Ampure XP beads (Beckman Coulter, Cat. No. A63881) at 1:1 sample to bead ratio.

The linearized chimeric H3 HA sequence was then mutagenized using a modification of a previously described PCR mutagenesis technique (Bloom, 2014). Forward and reverse primers for mutagenic PCR were pooled separately at a molar ratio of 10:1:1 between epitope : paired epitope : general mutagenic primers. Mutated fragments of the chimeric H3 HA sequence were then PCR-amplified in two separate reactions, using either the forward primer pool and reverse linearizing primer, or the reverse primer pool and forward linearizing primer. We performed two rounds of mutagenesis, with 7 mutagenic PCR cycles the first round and 11 the second round. The mutagenesis PCRs were performed with the following conditions: PCR mix: 20μL H2O, 1.5μL 4.5μM forward or reverse primer pool, 1.5μL 4.5μM reverse (primer_089) or forward (primer_088) linearizing primer, 4µL 3ng/µL linearized template, and 15µL KOD. Cycling conditions: (1) 95C/2min (2) 95C/20sec (3) 70C/1sec (4) 50C/30sec, cooling at 0.5C/sec (5) 70C/55sec (6) Return to Step 2 for the number of cycles described above (7) 70C/2min. Note that mutagenesis was performed in duplicate, to generate two independently mutagenized replicate libraries.

Between each round of mutagenic PCR, the mutagenized partial H3 HA fragments from the forward and reverse reactions were joined in a separate PCR reaction to generate full-length H3 HA sequences. The joining PCR was performed with the following conditions: PCR mix: 4 µL each of 1:4 diluted forward and reverse mutagenesis reactions, 1.5µL 4.5µM forward primer (primer_088), 1.5µL 4.5µM reverse primer (primer_089), 3µL H2O, and 15µL KOD. Cycling conditions: (1) 95C/2min (2) 95C/20sec (3) 70C/1sec (4) 50C/30sec, cooling at 0.5C/sec (5) 70C/40sec (6) Return to Step 2 x19 (7) 70C/2min. Joined PCR mutagenesis products were gel and Ampure bead purified after each joining reaction, as described previously.

We then appended random 16-nucleotide barcodes to the mutagenized H3 HA sequences, followed by a constant overlap sequence for cloning into the backbone plasmid, using a PCR with the following conditions: PCR mix: 11.2µL H2O, 0.9µL 10µM forward linearizing primer (primer_088), 0.9µL 10µM reverse barcoding primer (primer_090), 2µL 15ng/µL mutagenized template, and 15µL KOD. Cycling conditions: (1) 95C/2min (2) 95C/20sec (3) 70C/1sec (4) 50C/30sec, cooling at 0.5C/sec (5) 70C/40sec (6) Return to Step 2 x9 (7) 70C/2min.

The barcoded mutagenized H3 HA sequences were gel and Ampure bead purified as described previously, using a 1:1 sample to bead ratio, then cloned into a plasmid containing the flanking WSN HA packaging signals, 5’ and 3’ NCRs, and pHH21 promoter and terminator. The plasmid contains GFP in place of the H3 HA ectodomain to ensure that virions cannot be generated from untransformed plasmid. The plasmid was digested with BsmBI and NotI, then gel and Ampure bead purified. We then used a 2:1 insert to vector ratio in a 1 hour HiFi assembly reaction using NEBuilder HiFi DNA Assembly kit (NEB, Cat. No. E2621). HiFi assembly products were Ampure bead purified and eluted into 20uL of H2O for higher electroporation efficiency.

We electroporated 20μL of 10-beta electrocompetent E. coli cells (NEB, C3020K) with 1μL of the purified HiFi product. Electroporated cells were plated on 15cm LB+ampicillin plates at an estimated bottleneck of 80,000 to 100,000 colony forming units to limit library size. After approximately 18 hours of outgrowth, colonies were scraped into liquid LB+ampicillin, and grown at 37°C for 2.5 hours in liquid culture prior to plasmid purification using QIAprep Spin Miniprep Kit (Cat. No. 27104). We generated two independently mutagenized plasmid libraries, referred to as libA and libB. The final structure of the mutagenized H3 HA library construct is illustrated in **Figure 2.1B.**

2.5.4 PacBio sequencing of H3 HA plasmid library

We used long-read PacBio sequencing to link variants to their random nucleotide barcodes, and determine the composition of the plasmid library, similar to the approach used in (Starr et al., 2020). PacBio sequencing inserts were prepared from the purified plasmid pools via NotI-HF restriction digest, rather than PCR amplification, which eliminates the possibility of PCR strand exchange scrambling barcodes. Digested inserts were then gel and Ampure bead purified, and eluted into 30uL of EB. Each library was then barcoded for PacBio sequencing using SMRTbell prep kit 3.0, bound to polymerase using Sequel II Binding Kit 3.2, and then sequenced using a PacBio Sequel IIe sequencer with a 20-hour movie collection time. Analysis is described in the section ‘PacBio sequencing data analysis’.

2.5.5 Generation of H3 HA variant library in the context of lab-adapted WSN virus

Mutant virus libraries were generated by reverse genetics (Hoffmann et al., 2000). We used 6-well plates rather than 15cm plates for transfection, to limit the extent to which jackpotting of certain variants in each well could bottleneck library diversity. We transfected 40 wells of 6-well plates for each library. Each well was plated with 4×10^5 HEK-293T cells in D10 media (DMEM, supplemented with 10% heat-inactivated FBS, 2 mM L-glutamine, 100 U/mL penicillin, and 100 μg/mL streptomycin). 21 hours after plating, cells were transfected with 220ng each of H3 HA plasmid library (or wildtype control), the pHW18* series of plasmids (Hoffmann et al., 2000) for all non-HA viral genes, and pHAGE2-EF1aInt-TMPRSS2-IRES-mCherry (Lee et al., 2018). The pHW18* plasmids encode genes from the lab-adapted A/WSN/1933 strain, and so all genes in our viruses derive from the lab-adapted WSN strain except for the HAs. We used Bioland BioT (Cat. No. B01-01) for the transfections, and followed the manufacturer's recommendations for the protocol and DNA / transfection reagent ratios.

At 20 hours post-transfection, we removed the D10 media and added 3×10^5 MDCK-SIAT1-TMPRSS2 cells per well in low-serum IGM (Influenza Growth Media, consisting of Opti-MEM supplemented with 0.01% heat-inactivated FBS, 0.3% BSA, 100 U/mL penicillin, 100 μg/mL streptomycin, and 100 μg/mL calcium chloride). Note that overlay of MDCK-SIAT1-TMPRSS2 cells post-transfection yielded higher virus titers than transfection of co-cultured cells in pilot rescue experiments. At 49 hours post-transfection, the transfection supernatants were harvested, clarified by centrifugation at 2,000x*g* for 5min, aliquoted, frozen at -80°C, and titered in MDCK-SIAT1-TMPRSS2 cells. The titers were 3,162 TCID50/μL for both libA and libB.

To generate a genotype-phenotype link, where the barcoded HA gene carried by each virion matches the HA variant expressed on its surface, we passaged the libraries at a low MOI such that each cell should be infected by one virion at most. For each library, we infected cells in two 5-layer 875 cm^2 flasks (Corning, Cat. No. 353144) at an MOI of 0.01. For each flask, we gently mixed 7.875x10^7 MDCK-SIAT1-TMPRSS2 cells in 150mL of IGM (targeting a density of 9x10^4 cells/cm^2 when plated) with 7.875x10^5 TCID50 of library transfection supernatant in a sterile 250mL bottle. The cell-virus mix was then added directly to the 5-layer flask. Virus supernatant was collected at 45 hours post-infection, clarified by centrifugation at 2,000x*g* for 5min, split into 1mL aliquots, frozen at -80°C, and titered in MDCK-SIAT1-TMPRSS2 cells. The titers were 3.09x10^4 and 1.43x10^4 TCID50/µL for libA and libB, respectively. Assuming a maximum library size of 100,000 variants, libraries were passaged at over 10x coverage of variant diversity.

For future deep mutational scanning studies, we would recommend rescuing the library at 40 hours post-transfection, and reducing passaging to approximately 36 hours. These conditions yielded similar titers, and minimizing library growth times will reduce bottlenecking from variants carrying tissue-culture-adaptive mutations.

2.5.6 Generation of WSN-H6 HA neutralization standard

The WSN-H6 HA standard was generated from the same GFP recipient vector used to clone the WSN-flanked H3 HA library construct. The plasmid was digested with BsmBI and XbaI, then gel and Ampure bead purified. To avoid sequencing errors, we designed four different barcodes for the WSN-H6 HA standard, such that the nucleotide at a given site was different in all four barcodes. The four barcoded H6 HA ectodomain sequences were synthesized from Twist with a 20bp overlap with the digested recipient vector, and cloned into the recipient vector using a 2:1 insert to vector ratio in a 1 hour HiFi assembly reaction. NEB 5α High Efficiency Cells (Cat. No. C2987) were transformed with 2µL of the HiFi assembly, and plated on LB+ampicillin for 18 hours of outgrowth. Single colonies were grown in liquid LB+ampicillin at 37°C for 18 hours, plasmid was purified using QIAprep Spin Miniprep Kit, and sequence was confirmed by Sanger sequencing. The four barcoded plasmids were then combined at equal concentrations.

To generate the virus, we transfected a coculture of $4x10^5$ HEK-293T and $5x10^4$ MDCK-SIAT1 cells with 220ng each of the pooled WSN-H6 HA plasmid, the pHW18* series of plasmids (Hoffmann et al., 2000) for all non-HA viral genes, and pHAGE2-EF1aInt-TMPRSS2-IRES-mCherry. Transfections were performed in D10 media using BioT transfection reagent. Media was changed to IGM at 20 hours post-transfection. At 68 hours post-transfection, transfection supernatants were harvested, clarified by centrifugation at 2000x*g* for 5 min, aliquoted, frozen at -80°C, and titered in MDCK-SIAT1-TMPRSS2 cells. Titers were $3.16x10^5$ TCID50/µL.

2.5.7 Generation of RNA spike-in neutralization standard

The RNA spike-in standard was generated from the GFP recipient vector, with the final 46 amino acids from the library strain appended at the 5' end of the GFP sequence (in negative-sense viral orientation), such that the 5' coding, random barcode region, and priming sequence were identical to the library strain. A T7 promoter sequence was added after the 5' NCR for in vitro transcription of the RNA sequence. The GFP recipient vector was digested with XbaI alone, then gel and Ampure bead purified. The C-terminal insert sequence was PCR-amplified off of the chimeric library plasmid, using a forward primer that appended a 20-bp overlap with the GFP sequence. PCR conditions were as follows: PCR mix: 21µL H2O, 1.5µL 10µM forward GFP overlap primer (primer_115), 1.5µL 10µM reverse linearizing primer (primer_089), 1µL 10ng/µL template plasmid, and 25µL 2x KOD Hot Start Master Mix. Cycling conditions: (1) 95C/2min (2) 95C/20sec (3)

63C/10sec (4) 70C/10sec (5) Return to Step 2 x19 (6) 70C/2min. Amplified insert was gel purified.

We then appended random 16-nucleotide barcodes to the amplified insert, followed by a constant overlap sequence for cloning into the GFP recipient vector, using a PCR with the following conditions: PCR mix: 10.66µL H2O, 0.9µL 10µM forward linearizing primer (primer_088), 0.9µL 10µM reverse barcoding primer (primer_090), 2.54µL 11.8ng/µL insert template, and 15µL 2x KOD Hot Start Master Mix. Cycling conditions: (1) 95C/2min (2) 95C/20sec (3) 70C/1sec (4) 50C/30sec, cooling at 0.5C/sec (5) 70C/20sec (6) Return to Step 2 x9 (7) 70C/2min.

The barcoded insert was gel purified, then cloned into the XbaI-digested GFP recipient vector, using a 3:1 insert to vector ratio in a 1 hour HiFi assembly reaction. NEB 5α High Efficiency Cells were transformed with 2µL of the HiFi assembly, and plated on LB+ampicillin for 18 hours of outgrowth. To generate a diverse pool of barcodes, 20 individual colonies were grown in liquid LB+ampicillin at 37°C for 18 hours, and plasmids were purified using QIAprep Spin Miniprep Kit. Sequence was confirmed by Sanger sequencing. The 16 plasmids with no SNPs were pooled at equal concentrations.

From this randomly-barcoded pool, we amplified the linear DNA template from the U12 to U13 region, and added a T7 promoter sequence in the reverse orientation after U13 (5' terminus of the linear template, negative-sense viral orientation). PCR conditions were as follows: PCR mix: 11.2µL H2O, 0.9µL 10µM forward U12-annealing primer (primer_113), 0.9µL 10µM reverse T7-appending primer (primer_116), 2µL 10ng/µL template plasmid, and 15µL 2x KOD Hot Start Master Mix. Cycling conditions: (1) 95C/2min (2) 95C/20sec (3) 62C/10sec (4) 70C/40sec (5) Return to Step 2 x19 (6) 70C/2min. We ran two replicate 30uL PCR reactions, and gel purified both reactions using the same column to maximize concentration of the linear DNA template.

The RNA spike-in standard was then generated via *in vitro* transcription of the linear DNA template with T7 promoter, using the T7 RiboMAX Express Large Scale RNA Production System (Promega, Cat. No. P1320) with 1µg DNA template input. Reaction was incubated at 37°C for 45min. To remove DNA template after transcription, we added 1µL of RQ1 RNase-Free DNase and incubated for an additional 15 min at 37°C. RNA was then purified using the Monarch RNA Cleanup Kit (NEB, Cat. No. T2040).

Final RNA yield was at a concentration of 2.24µg/µL. This was diluted to 0.025ng/µL, aliquoted, and stored at -80°C. Barcodes were amplified and sequenced on a NextSeq 2000 alongside library selection samples, as described in the section 'Barcode amplification from infected cells for Illumina sequencing'. The final pool of RNA spike-in neutralization standard contained 125 unique barcodes.

2.5.8 Serum preparation

All sera were heat-inactivated and treated with receptor-destroying enzyme (RDE) before use. RDE-treatment removes residual sialic acids in human serum samples that can bind to the virus libraries. We mixed serum with resuspended, lyophilized RDE II (VWR, Cat. No. 370013) at a ratio of 1:3, and incubated in heatblocks at 37°C for 2.5hr. We then heat-inactivated the RDE and serum by incubating at 55°C for 30min, aliquoted, and stored at -80°C. Note that all sera used in this study are at a starting dilution of 1:4 due to RDE treatment, which is accounted for in reported concentrations.

2.5.9 GFP neutralization assays for serum screening and validation

Neutralization assays were performed using influenza viruses carrying GFP in the PB1 segment. These GFP viruses were generated by reverse genetics using the following plasmids: wildtype WSN-flanked A/Hong Kong/45/2019 HA (the library background strain), pHH-PB1flank-eGFP (which encodes GFP on the PB1 segment; (Bloom et al., 2010)), and the pHW18* series of plasmids (Hoffmann et al., 2000) for all viral genes except HA and PB1. The same approach was used to generate GFP virus for validation neutralization assays, with the mutation of interest introduced into the WSN-flanked A/Hong Kong/45/2019 HA plasmid.

We added $4x10^5$ 293T-CMV-PB1 (Bloom et al., 2010) and $5x10^4$ MDCK-SIAT1-CMV-PB1-TMPRSS2 (Lee et al., 2018) cells in D10 media to each well of a 6-well plate. Cells that constitutively express PB1 are necessary because the PB1 segment is replaced by GFP in these viruses. 18 hours after plating, wells were transfected with 220ng of each reverse genetics plasmid and the pHAGE2-EF1aInt-TMPRSS2-IRES-mCherry plasmid, using BioT transfection reagent. At 20 hours post-transfection, we changed the media in each well to 2mL NAM (Neutralization Assay Media, consisting of Medium-199 supplemented with 0.01% heat-inactivated FBS, 0.3% BSA, 100 U/mL penicillin, 100 μg/mL streptomycin, 100 μg/mL calcium chloride, and 25mM HEPES). This media is used because it has low background fluorescence at 488 nm. At approximately 40 hours post-transfection, the transfection supernatants were harvested, clarified by centrifugation at 2,000x*g* for 5min, aliquoted, and frozen at -80°C.

To titer these viruses, we infected $1x10^5$ MDCK-SIAT1-CMV-PB1-TMPRSS2 in a 12-well plate with different dilutions of virus, chose wells that showed 1-10% GFP positivity at 18 hours post-infection, measured the fraction of GFP-positive cells by flow cytometry, and used this fraction to calculate the titer of infectious particles per μL.

For accurate measurement of serum neutralization curves, the correlation between number of infectious particles and change in GFP fluorescence should be linear. Adding too much or too little virus per well will result in inaccurate readings. It's therefore important to generate an MOI curve for each GFP virus, where MDCK-SIAT1-CMV-PB1-TMPRSS2 cells in a 96-well plate are infected with 2-fold dilutions of virus, and GFP signal is measured 18 hours post-infection. The virus concentration used for neutralization assays should correspond to the highest MOI in the linear range of MOI:signal.

Neutralization assays were performed as described previously (Doud et al., 2018; Hooper and Bloom, 2013) (see also https://github.com/jbloomlab/flu_PB1flank-GFP_neut_assay for detailed protocol). Serum was diluted 2-fold in NAM, for final volumes of 40μL per well of a 96-well plate. These dilutions were 2x more concentrated than the target dilutions, to account for the addition of 40μL of virus. Note that calculation of neutralization curves assumes that antibody is in excess of virus, and if this is not the case, the measured IC50 will be artificially high. For some highly potent sera (IC50=$1x10^{-4}$ dilution or lower), serum dilution volumes were therefore increased to 100μL to avoid saturation of serum antibodies. The serum-virus mix was incubated for 1.5 hours at 37°C before adding $4x10^{4}$ MDCK-SIAT1-CMV-PB1-TMPRSS2 cells per well. When screening for neutralization activity against the wildtype library strain, only one replicate was run per serum. Sera selected for escape mapping were then re-run with two replicates per serum to generate more accurate curves. All validation neutralization assays were run with two replicates per serum and virus, and curves represent the mean and standard error of these replicates.

GFP signal was measured 18 hours post-infection. To calculate fraction infectivity, we subtracted the average background signal from uninfected cells from each condition, then divided the signal from each serum dilution by the average signal from no-serum control infections. Fraction infectivities were then used to fit Hill-like neutralization curves using the neutcurve package (https://jbloomlab.github.io/neutcurve/), version 0.5.7.

2.5.10 Antibody and serum selection of library escape variants

Selections were performed at three concentrations for each antibody or serum, targeting a range of concentrations from serum IC90 to IC99.99. Using multiple potent antibody concentrations for selections generates more accurate measurements of enriched escape variants, and improves fitting with our biophysical escape models (Yu et al., 2022).

For each experiment, we plated MDCK-SIAT1 cells in IGM 4 hours before infection. Because large differences in MOI impact the amount of viral mRNA transcribed by each cell (Bacsik et al., 2023), we plated $1.5x10^5$ cells per well for each selection condition, and $1x10^6$ cells per well for the no-antibody controls, to roughly account for the expected difference in infectious particles. Selections used 12-well plates while no-antibody controls used 6-well plates, to keep cell density relatively consistent. We made a master mix of library virus with WSN-H6 HA neutralization standard added at 0.1% (libA) or 0.3% (libB) of the total virus titer. These ratios were calibrated in pilot selections such that approximately 1% of counts in the no-antibody control conditions would come from the neutralization standard. Using a master mix of virus is important as it ensures comparable results between the no-antibody and serum selection conditions. 400µL of virus master mix was mixed with 400µL serum (diluted to 2x the targeted concentration, to account for virus volume), or 400µL media for the mock selection control. All dilutions were in IGM. The virus-serum or virus-media mixes were incubated at 37°C for one hour, then added dropwise to each well. RNA was harvested from infected cells at 13 hours post-infection (described below).

In initial pilot experiments, we added $6x10^5$ TCID50 per infection for each library for approximately 20-fold coverage of variant diversity, and ran selections in minimal volumes to conserve serum. We found that these conditions resulted in serum antibody saturation, where there were more virions than antibodies, and results were skewed by measurement of unbound virions. 800μL was the minimal selection volume that avoided saturation effects for potent sera. We also found that the calculated probability of escape for wildtype variants was consistently higher for libB than libA. We therefore reduced the titer of libB added per infection to $3x10^5$ TCID50, which led to consistent results between libraries for wildtype variants. This discrepancy is likely due to inconsistency between TCID50, measured as virus spread in cell culture, and selection experiment results, measured based on cell entry. We therefore recommend using the probability escape for wildtype variants from a pilot experiment, which should reflect targeted ICXX values (e.g. 10% probability escape at serum IC90), to calibrate the amount of virus added per selection.

To minimize bottlenecking of variant diversity during infections, we infected 3 wells with no-antibody controls for each library, prepped the RNA and viral barcodes separately (described below), and pooled the Illumina sequencing reads into a single no-antibody control sample.

2.5.11 Barcode amplification from infected cells for Illumina sequencing

RNA was harvested from infected cells 13 hours post-infection using the RNeasy Plus Mini Kit (Qiagen, Cat. No. 74134). We made a master mix of RLT lysis buffer with 1% β-mercaptoethanol to inhibit RNases, and 0.14ng/mL RNA spike-in standard (i.e. 2μL of 0.025ng/μL standard for every 350μL in the master mix). This correlates to 0.05ng of standard in 352μL lysis master mix, added to each well. Working with one row of wells at a time, we removed the media, washed with 1mL PBS, and aspirated completely. We then added 352μL of the lysis master mix with RNA standard. Cells were washed from each well and transferred to 1.5mL aliquots, then vortexed for 1min at full speed to homogenize. After lysing and homogenizing cells from all conditions in the experiment, we proceeded with RNA extraction, following the extraction kit protocol. RNA was eluted twice from each column in 30µL RNase-free water to maximize yields. Extraction was typically completed within 1.5 hours from homogenization.

Viral mRNA was transcribed as cDNA, using SuperScript III First Strand Synthesis SuperMix (Invitrogen, Cat. No. 18080400). We normalized amount of input RNA across all samples in the experiment, and input anywhere from 800ng-1µg RNA per sample, depending on the maximum input from the lowest-concentration sample. Reaction was carried out according to manufacturer protocol, using the mRNA-annealing RT primer (primer_110) to amplify off of viral mRNA. We chose to target mRNA as it is more abundant in infected cells than vRNA.

We then used two rounds of PCR to amplify the barcode region and add the necessary Illumina adapters and indices for each sample. The first round of PCR uses a forward primer that anneals to the C-terminus of the HA sequence, and a reverse primer that anneals to the Illumina Truseq Read 1 sequence downstream of the barcode. We use the maximum recommended input of cDNA, 10µL, to minimize bottlenecking of barcode diversity during amplification. Conditions were as follows: PCR mix: 14µL H2O, 0.5µL 5µM Illumina round 1 forward primer (primer_098), 0.5µL 5µM Illumina round 1 reverse primer (primer_099), 10µL cDNA template, 25µL KOD. Cycling conditions: (1) 95C/2min (2) 95C/20sec (3) 70C/1sec (4) 64C/10sec, cooling at 0.5C/sec (5) 70C/20sec (6) Return to Step 2 x19 (7) 70C/2min.

DNA concentration was quantified using a Qubit Fluorometer (ThermoFisher). A second round of PCR was then performed using custom dual indexing primers, which anneal to the Read 1 or Read 2 sites, and append the P5 Illumina adapter and i5 sample index at the Read 1 terminus, or the P7 Illumina adapter and i7 sample index at the Read 2 terminus. Conditions were as follows: PCR mix: 5ng of round 1 PCR product (variable volumes), H2O to 19µL, 3µL 2.5µM unique i5 indexing primer, 3µL 2.5µM unique i7 indexing primer, 25µL KOD. Cycling conditions: (1) 95C/2min (2) 95C/20sec (3) 70C/1sec (4) 55C/20sec (5) 70C/20sec (6) Return to Step 2 x9 (7) 70C/2min.

DNA concentration of each round 2 PCR product was quantified using Qubit. Serum selection samples were pooled at an even ratio, while no-antibody control samples were added at 3x the concentration of each selection sample, to account for greater barcode diversity in the no-antibody controls. Samples were gel purified and Ampure bead cleaned at a 1:2 sample to beads ratio, then sequenced using either P2 or P3 reagent kits on a NextSeq 2000. Analysis is described in the section 'Illumina barcode sequencing data analysis'.

2.5.12 PacBio sequencing data analysis

PacBio CCSs were analyzed using *alignparse* (see https://jbloomlab.github.io/alignparse/ for documentation) (Crawford and Bloom, 2019). We computed the empirical accuracy of the CCSs as the frequency at which CCSs with the same barcode were associated with the same variant sequence. Empirical accuracies for both libraries were approximately 0.99.

We filtered out any barcodes that had less than two CCSs, to ensure confidence in the consensus sequences. We also filtered out barcodes that had minor fractions of substitutions or indels above 0.2, which would indicate that multiple variants share that barcode. The resulting consensus sequences were saved as barcode / variant lookup tables. This file and the associated analysis notebook, which includes quality control plots, can be found on the main HTML page described in 'Computational pipeline overview'.

2.5.13 Illumina barcode sequencing data analysis

We processed the Illumina barcode sequencing using the parser https://jbloomlab.github.io/dms_variants/dms_variants.illuminabarcodeparser.html to determine the counts of each variant in each selection condition. We filtered barcoded mutants that had low counts in the no-antibody control conditions, as these were likely to be randomly bottlenecked during infections and generate noisy escape scores. See https://github.com/dms-vep/flu_h3_hk19_dms/blob/main/config.yaml for the configuration file where these thresholds are specified.

2.5.14 Modeling the effects of mutations on H3 HA function

We modeled the functional effects of mutations on A/Hong Kong/45/2019 H3 HA as previously described (Dadonaite et al., 2023). We computed functional scores based on the frequency of a given mutation in the plasmid library (where there is no functional pressure), compared to the passaged virus library (where non-functional HA variants have been purged). The functional score for a variant v is calculated as $log_2\left(\left[n^{v}_{post} / n^{wt}_{post}\right]/\left[n^{v}_{pre} / n^{wt}_{pre}\right]\right)$, where n^{v}_{post} is the variant count in the no-antibody mock selection condition, n^{v}_{pre} is the variant count in the plasmid library, and n^{wt}_{post} and n^{wt}_{pre} are the counts of all wildtype barcodes in each condition.

Many variants in the library carry multiple mutations. We therefore use global epistasis models (Otwinowski et al., 2018; Sailer and Harms, 2017) to deconvolve this data and estimate the fitness effects of each individual mutation. Under these models, each mutation has an additive effect on a “latent” (unmeasured) phenotype. These additive effects are then transformed by a nonlinear function to generate the “observed” (measured) phenotype. The nonlinear transformation accounts for epistatic interactions between mutations, and thereby generates more accurate estimates for mutation effects.

2.5.15 Modeling the effects of mutations on antibody and serum escape

We modeled the effects of mutations on serum escape as described previously in Dadonaite et al. (2023). Briefly, we computed the fraction of non-neutralized variants at each serum selection concentration, normalized by counts from the relevant neutralization standard (either WSN-H6 HA or the RNA spike-in standard). Fraction of non-neutralized variants corresponds to probability of escape, and should fall between 0 (variant is fully neutralized) and 1 (variant fully escapes neutralization). Values greater than 1 are clipped. We then estimated the effects of each individual mutation on escape using the software package *polyclonal*, version 6.2 (see https://jbloomlab.github.io/polyclonal/ for documentation) (Yu et al., 2022). Fits were constrained to a single epitope to simplify analysis, as most sera were highly targeted.

For each serum, the heat maps and line plots are generated using the mutation-level escape scores calculated by *polyclonal*. All sera had measurements from two replicate libraries, and the final reported values are the average (median) across those replicates. Files containing the escape scores from individual replicates, the averaged scores, analysis notebooks with relevant quality control plots, and interactive plots summarizing the final estimates are available on the main HTML page described in ‘Computational pipeline overview’. Figures in this chapter only include mutations seen in at least three distinct variants (averaged across libraries), with a functional effect of at least -1.38.

CHAPTER 3: Age-dependent heterogeneity in the antigenic effects of mutations to influenza hemagglutinin

A version of this chapter is currently submitted as:

> Welsh FC, Eguia RT, Lee JM, Haddox HK, Galloway J, Van Vinh Chau N, Loes AN, Huddleston J, Yu TC, Le MQ, Nhat NTD, Le Thanh NT, Greninger AL, Chu HY, Englund JA, Bedford T, Matsen FA, Boni MF, Bloom JD, **Age-dependent heterogeneity in the antigenic effects of mutations to influenza hemagglutinin**, *bioRxiv*, 2023.

Seasonal human influenza virus undergoes continual antigenic evolution in response to immune selection from the host population. Although human individuals retain potent neutralizing antibodies to historical strains from prior exposures, they are susceptible to re-infection roughly every five years, due to rapid accumulation of mutations in the hemagglutinin (HA) surface protein that erode antibody neutralization (Kucharski et al., 2015; Ranjeva et al., 2019). Antigenically distinct strains of influenza H3N2 emerge every few years (Smith et al., 2004), resulting in frequent turnover of the virus population (Bedford et al., 2014; Fitch et al., 1997; Strelkowa and Lässig, 2012). Because of this continual antigenic evolution, individuals in the population have diverse exposure histories, which in turn shape their antibody response to subsequent infection or vaccination (Fonville et al., 2014; Francis, 1960; Krammer, 2019; Linderman et al., 2014; Skowronski et al., 2017). Neutralizing antibody titers are typically highest against strains seen early in life, and steadily decline against strains seen later in life (Fonville et al., 2014; Kucharski et al., 2015; Lessler et al., 2012; Ranjeva et al., 2019; Yang et al., 2020), which suggests that immune specificity varies, at least in part, by birth cohort.

Population-level heterogeneity has been proposed as a factor shaping the antigenic evolution of influenza (Nakajima et al., 2000; Oidtman et al., 2021; Sato et al., 2004), but we lack a detailed understanding of how the antigenic impacts of viral mutations differ across individuals in the population. Prior work has characterized this heterogeneity by analyzing human serum neutralization or hemagglutination-inhibition of recently circulating strains (Kim et al., 2023; Nakajima et al., 2000; Sato et al., 2004), or analyzing demographics of confirmed influenza cases over many seasons (Arevalo et al., 2020; Worby et al., 2015). Kim et al. (2023) found that individuals within the same age group have significantly correlated neutralization titers to different HA or NA clades. However, these studies have focused on susceptibility to naturally circulating strains, and cannot fully resolve how specific viral mutations affect neutralization by the polyclonal antibodies of different individuals.

Deep mutational scanning can be used to map the mutations in influenza HA that confer escape from neutralization by human sera, and has shown that single mutations can sometimes lead to large drops in neutralization (Lee et al., 2019). However, prior deep mutational scanning studies have not analyzed the antigenic effects of mutations across many individuals from different age groups. To address this gap, we here use deep mutational scanning to map how mutations to the HAs of two H3N2 vaccine strains (A/Hong Kong/45/2019 and A/Perth/16/2009) affect neutralization by sera collected from individuals of different ages during the timeframe that strain was in the vaccine. We find that the antigenic effects of mutations differ across individuals and especially age cohorts, with single mutations often causing large drops in neutralization. Differences among age cohorts are partially explained by inferred exposure history. We also find some evidence that the mutations that fix during natural influenza evolution in subsequent years often confer substantial neutralization escape from child and especially teenage sera. Overall, our work provides a deeper understanding of the antigenic selection exerted by different subsets of the population, and suggests ways in which this heterogeneity may shape influenza evolution.

3.1 Results

3.1.1 Antigenic effects of HA mutations differ among individuals and age cohorts

To better understand how antibody escape mutations differ among human individuals, we generated escape maps for sera collected in 2020 from 8-10 individuals each in three different age cohorts: children (2-5 years of age), teenagers (15-20 years), and adults (40-45 years) (**Figure 3.1**). Most individuals were vaccinated within the past year (**Tables 3.1-3.2**). We also mapped serum from an unvaccinated 2-month old infant and a 68-year-old adult. Sera were chosen from a larger set, based on having good neutralization activity against the parental A/Hong Kong/45/2019 library strain HA (**Figure 3.6**). To confirm that our escape maps are replicable, we mapped two sets of sera sampled from the same individual on different days, and found that the results were almost identical (**Figure 3.**7).

For all sera, the strongest escape mutations are within a small subset of sites in HA1 that correspond to classically defined antigenic regions (Wiley et al., 1981) (**Figure 3.1A-B**) . The mutations that cause the greatest escape are in antigenic regions A and B, which are near the sialic-acid receptor binding pocket and have alternated in immunodominance throughout H3N2 evolution (Popova et al., 2012). Mutations in antigenic regions C and E, which are lower on the HA head, cause weaker escape from the sera of some individuals. Two outlier sera, child 4584 and teenager 3856, exclusively target regions C and E and have no escape mutations in the immunodominant regions A and B (**Figure 3.8**). Notably, out of all mapped sera, these two sera had the lowest neutralization activity against the parental library strain.

There are clear differences in escape maps among age cohorts, though they share some sites of strong escape. This can best be seen by analyzing both the overall magnitude of escape across sera (**Figure 3.1C**), and the sites that confer the strongest escape for each serum, normalizing by their maximum escape score (**Figure 3.1D**). For most cohorts, the highest overall escape is at site 189 in the immunodominant antigenic region B. Most adult sera are also escaped by mutations in region A, which was immunodominant in the early 1990s (Nobusawa et al., 2012). Teenage sera are more often escaped by mutations at sites 156-159 in region B, which has likely been immunodominant for their entire lifetime (Chambers et al., 2015; Popova et al., 2012) (**Figure 3.1D**). Some child sera target a different site in region B, site 160, and additional sites in region C. These region C sites are distant from the receptor-binding pocket (**Figure 3.1B**), and proximal to the vestigial esterase subdomain, a target of some broadly neutralizing antibodies (Wu and Wilson, 2017). The escape map for the serum from an elderly individual is distinct from other cohorts, with the strongest escape at site 145, and no signal at 189. This individual was born 16 years before the introduction of H3N2, and differences in childhood HA exposure could contribute to their unique escape map.

Sites where mutations strongly increase neutralization sensitivity (indicated by negative values in the escape maps) are more common in older age cohorts (**Figure 3.1C**). There are two regions of strongly sensitizing mutations in adults: sites 121-124 and site 160. For teenagers, only site 160 exhibits strongly sensitizing mutations. Some children have weakly sensitizing mutations at site 160. Sites 158-160 and 122-124 are glycosylation motifs that appeared during the lifetimes of certain cohorts, and individuals with sensitizing mutations at these sites were exposed to prior strains without this glycan. The 122-124 glycan appeared in 1997. Adults exposed to H3N2 prior to 1997 could have generated neutralizing antibodies targeting this region, but the epitope has always been masked in strains seen by younger cohorts. The 158-160 glycan appeared in 2014, during the lifetime of both teenagers and adults, so these cohorts have similar sensitizing mutations at site 160. The weak negative signal in some children may be attributed to vaccination against A/Kansas/14/2017 H3N2 in the previous year, which does not express the 158-160 glycan (Gouma et al., 2020b) (**Table S1**). There are additional sites of weak negative signal, such as 159 and 193 in adults, and 135 in teenagers, that correspond to sites mutated during their lifetimes (**Figure 3.10**).

By visual inspection, it appears that teenagers and adults often have more heterogeneous escape maps within their age cohorts than children (**Figure 3.1D**). The main escape site for all children is site 189, with the exception of one outlier that exclusively targets antigenic region C. Secondary escape sites are limited and highly consistent among individuals. While 189 is also a major target for most teenagers and adults, the magnitude is comparable to escape at other sites for many individuals, and these other sites are variable between sera. This suggests that a naive population with limited, likely common prior exposure to influenza develops a narrow and fairly stereotyped antibody response. As exposure history increases, antibody responses diversify, even within an age cohort.

The escape map for the 2-month-old unvaccinated infant is similar to the adults, with neutralization-sensitizing mutations at sites 121-124 (**Figure 3.1C**). Influenza antibodies are transferred from mother to fetus during pregnancy, and confer passive immunity in newborns for several months (Nunes et al., 2015). We hypothesize that this infant gained antibodies targeting the glycan-masked 121-124 region from their mother, alongside antibodies with neutralization activity at sites 145 and 189.

In short, these escape maps demonstrate that sera from different age cohorts are escaped by mutations at different sites in HA. Children with limited prior exposure tend to have serum that is escaped by a fairly homogenous and focused set of mutations, whereas escape mutations become more heterogeneous with age. The prevalence of neutralization sensitizing mutations in the escape maps for adults indicated that older individuals often retain antibodies that target strains seen earlier in their lives. At a population level, site 189 is the most potent target of neutralizing antibodies, but individual teenagers and adults also target diverse additional sites in several known antigenic regions. This heterogeneity is especially notable given that most individuals were vaccinated against the same strains in the prior season (**Tables 3.1-3.2**).

3.1.2 Validation of the deep mutational scanning in traditional neutralization assays

We validated key mutations using traditional neutralization assays for eight representative sera from different age cohorts. The changes in IC50 measured by traditional neutralization assays strongly correlate with the escape scores measured by deep mutational scanning (**Figure 3.2, Figure 3.9**).

The validation neutralization assays also confirm the serum-to-serum heterogeneity observed in the deep mutational scanning. For example, Y159N confers escape from four sera, but makes the virus more sensitive to neutralization by sera 3856 and 197C (**Figure 3.2**). Despite K189E and S193D being strong escape mutations for almost all other sera, they had neutral or slightly sensitizing effects for sera 3856 and 4584. At some sites, only specific mutations confer escape. We confirmed that S145K and S193D were escape mutants, but S145H and S193Y were indeed neutral. These validations emphasize that antigenic effects of mutations are often highly variable across a population.

Mutations at site 160 are sensitizing for most sera, but confer escape in three children. This is especially notable because 160T is part of a glycosylation motif. It’s unclear whether these children have antibodies targeting the glycan itself, or a glycan-adjacent region, but we validated that T160K is indeed an escape mutant for child 3944. We also confirmed that T160S, which maintains the glycan, is neutral.

3.1.3 Age-dependent heterogeneity in serum escape is also seen in an unvaccinated cohort measured against a different HA strain

The aforementioned trends were all based on escape maps for the HA of the A/Hong Kong/45/2019 (H3N2) strain generated using sera collected in 2020 in Seattle, Washington, USA. To assess whether trends in age-dependent heterogeneity in escape mutations are a general phenomenon, we also analyzed sera from nine children (2-4 years of age) and three adults (30-35 years of age) collected between 2010 and 2011 from Ho Chi Minh City, Vietnam against the HA of the A/Perth/16/2009 (H3N2) strain. In contrast with the high vaccination rates in the Seattle cohort (**Tables 3.1-3.2**), vaccination rates in Vietnam are negligible (Nhat et al., 2017). We also generated escape maps for four ferrets infected with the A/Perth/16/2009 library strain. The A/Perth/16/2009 HA deep mutational scanning used an older approach with previously described libraries (J. M. Lee et al., 2019; Lee et al., 2018) that are not barcoded. This older deep mutational scanning approach measures relative escape scores for each serum, but the magnitude of these escape scores is not directly comparable between sera. Furthermore, the older deep mutational scanning approach can only reliably measure escape mutations, not sensitizing mutations.

The ferret and child sera have similar escape mutations within their respective cohorts, while the adult sera target heterogeneous sites (**Figure 3.3A-D**). All four singly-infected ferrets have almost identical escape maps and target sites 189 and 193. The children have slightly more variable escape maps, but are still quite similar to each other and all target some combination of sites 189, 193, and 159-160. In contrast, each adult serum most strongly targets a different site: 159, 160, 189, or 192. These sites are proximal to one another (**Figure 3.3B**), but only confer escape from certain sera, suggesting that neutralizing activity against HA from polyclonal sera is highly focused on specific residues. We validated the antigenic effects of key mutants in neutralization assays, and found that they correlated very well with the predicted escape scores (**Figure 3.3E, Figure 3.12**). Note that the slope of the correlation line between escape score and measured fold-change in IC50 varies among sera in **Figure 3.3E**, because the magnitude of escape scores for this older deep mutational scanning approach are not directly comparable among sera.

Site 189 has high escape scores for five of the child sera, and is neutral for the other four (**Figure 3.3F**). The difference between these two groups of children is likely due to the arrival of the N189K mutation in H3N2 HA in 2009, which was halfway between their birth and serum collection. Children with positive escape scores at site 189 have neutralizing antibodies that can bind to K189 in the parental library strain, which suggests that they were exposed to a H3N2 virus with K at site 189. In contrast, we speculate that the children who do not have positive escape scores were likely exposed to H3N2 viruses with N189, and so do not have antibodies targeting K189 in the library strain. Note that the children with the strongest escape at sites 159-160 are all unaffected by mutations at site 189 (**Figure 3.11**).

Ferret sera have escape maps similar to children, but not adults (**Figure 3.3A,C,D**), which suggests that exposure history may be more important than species in shaping specificity. Previous work has shown that singly-infected children and ferrets have similar immune specificity, based on HAI titers against different H3N2 strains (Fonville et al., 2016). Our escape maps confirm that children and ferrets generate neutralizing antibodies against similar sites.

3.1.4 Shifts in immune specificity between 2010-2011 and 2020 cohorts reflect H3N2 evolution during this time period

All the 2010-2011 sera mapped against A/Perth/16/2009 are escaped mainly by mutations in antigenic region B, whereas the 2020 sera mapped against A/Hong Kong/45/2019 are escaped by mutations in antigenic regions A, B, C, and E to varying degrees (**Figure 3.4A-B**). This shift in the location of key escape mutations is likely due to the introduction of the 158 glycan in 2014, which masked nearby sites in antigenic region B (**Figure 3.4B**) (Gouma et al., 2020). The addition of this 158 glycan may explain why the dominant escape site in A/Hong Kong/45/2019 is 189 for all age cohorts: site 189 is still in the immunodominant antigenic region B, but is further from the 158 glycan, and has not been substituted in H3N2 evolution since 2009 (**Figure 3.3F, Figure 3.4B**). A/Hong Kong/45/2019 has also lost glycans at sites 133 and 144 relative to A/Perth/16/2009, which may contribute to increased escape in antigenic region A in A/Hong Kong/45/2019 by unmasking this region (**Figure 3.4B**).

Most escape maps from child sera are dominated by sites 159, 160, 189, and 193 in A/Perth/16/2009. Notably, mutations at three of these four sites rose to at least 90% frequency in the next 5-10 years (**Figure 3.4C**). Some children in the 2020 cohort also have sera that are escaped by mutations at site 50. The mutation E50K began to appear in circulating viruses in 2019, and reached 90% frequency in 2023 (**Figure 3.4C**). These evolutionary patterns suggest that the escape sites mapped here, especially from child sera, reflect antigenic pressure on influenza virus.

Site 189 in A/Hong Kong/45/2019 is generally the dominant escape site for the child sera collected in 2020 (**Figure 3.4C**). The differences in escape sites in the 2020 versus 2010-2011 child sera can be explained by mutations introduced between 2010 and 2020. The 158-160 glycan likely hindered the generation of antibodies targeting 159-160. However, mutations at site 160 in A/Hong Kong/45/2019 are either moderate escape or sensitizing mutations against child sera. This indicates that site 160 is still somewhat of a target of neutralizing antibodies for children in the 2020 cohort, but these antibodies overall make a smaller contribution to serum activity than for children in the 2010-2011 cohort.

Similarly, the mutation F193S was introduced in 2019, and children from the 2020 cohort were likely exposed to strains with the prior genotype F193. Reverting to F193 makes A/Hong Kong/45/2019 more sensitive to neutralization by these sera, and 193D remains a strong escape mutation for both cohorts, which indicates that these children have neutralizing antibodies against site 193 in their serum repertoire (**Figure 3.4C**). In short, the neutralizing antibody response of children with limited exposure history is largely directed towards sites 160, 189, and 193 in both cohorts.

3.1.5 H3 HA evolution after 2020 correlates more with immune pressure imposed by children and teenagers than adults

These results show that influenza faces somewhat distinct antigenic selection regimes from different age groups. This raises the possibility that some mutations could be especially beneficial for the virus within subsets of the population (**Figure 3.5A**). In order to see if the actual evolution of the virus strongly favored escape mutations from specific age cohorts, we analyzed serum escape for approximately 1,200 representative H3N2 strains circulating from 2012 to 2023. The overall escape score for a given strain was calculated as the sum of the escape scores for all of its HA amino-acid mutations relative to A/Hong Kong/45/2019. For each serum, strains with negative scores are more potently neutralized than A/Hong Kong/45/2019, while strains with positive scores can better escape neutralization.

Analysis of the average strain escape scores from each cohort shows that the evolutionary trajectory of H3N2 during this timeframe strongly correlates with escape from teenage sera (slope=0.14, Pearson *r*=0.92) (**Figure 3.5B**). Escape from child sera also correlates to a lesser degree (slope=0.06, r=0.70), although this timeframe includes years prior to their birth. In contrast, mutations gained after 2020 did not substantially escape neutralization by serum from the adult cohort. We confirmed that the increase over time in average strain escape scores from teenage sera are significantly higher than adult sera (*p*=0.004) or child sera (*p*=0.022). Although the number of sera in our analysis is modest, these results show how different age cohorts are differentially affected by viral evolution, and suggest that teenagers and children could play an especially prominent role in driving viral evolution.

3.2 Discussion

We have mapped the mutations that escape neutralization by the polyclonal serum antibodies of 40 human individuals from different age cohorts. Sites of neutralization escape for any given serum are often highly focused, with a few specific mutations often leading to 5-10 fold drops in neutralization by polyclonal serum antibodies—a finding consistent with prior work (Davis et al., 2018; Huang et al., 2015; Lee et al., 2019; Linderman et al., 2014). We also identify mutations that increase virus sensitivity to neutralization, typically by restoring epitopes present in older strains, such as removal of recently acquired N-linked glycosylation sites. The existence of these sensitizing mutations likely constrains the virus' evolution by disfavoring reversions to epitopes present in older strains.

The effects of mutations on neutralization escape differ between age cohorts in a manner consistent with likely exposure history. Many teenage sera are most escaped by mutations in antigenic region B, whereas adult sera are more strongly affected by mutations in antigenic region A, which reflects the regions that were likely immunodominant during their respective childhoods (Chambers et al., 2015; Nobusawa et al., 2012; Popova et al., 2012). Heterogeneity between cohorts can be highly specific: although some child and teenage sera are escaped by mutations in antigenic region C, these mutations are at site 50 for children and site 276 for teenagers. The relative number and magnitude of neutralization-sensitizing mutations also increases with age, likely due to older individuals having more antibodies from prior exposures to older strains.

The child sera analyzed in this study are escaped by mutations at a fairly narrow and homogenous set of sites. This homogeneity suggests a relatively stereotyped initial neutralizing antibody response focused on sites near the receptor binding pocket of HA, which is consistent with prior characterization of the antibody response in children (Islam et al., 2017; Meade et al., 2020; Nakajima et al., 2000). Four out of five sites targeted by children have mutated in subsequent years, suggesting that deep mutational scanning can be used to identify HA sites under strong immune pressure in contemporary human influenza strains.

Exposure history may be more important than species in shaping neutralizing specificity, as ferrets infected with A/Perth/16/2009 have similar escape mutations to children from the 2010-2011 cohort. The similarities between ferrets and naive children, but not adults with complex exposure histories, further illustrates the limitations of using primary ferret antisera as a proxy for adult human sera (Fonville et al., 2014; Lee et al., 2019; Linderman et al., 2014; Li et al., 2013). Nevertheless, ferret sera may still be useful for identifying the selective pressures imposed by young children (Fonville et al., 2016).

The introduction of the 158-160 glycan on the HA head in 2014 resulted in major antigenic change, as it hindered antibody binding to the immunodominant antigenic region B, leaving many individuals highly vulnerable to infection (Gouma et al., 2020a; Zost et al., 2017). It was therefore surprising that this glycan was subsequently lost in 2021 with the mutation T160I (Bolton et al., 2022). Our analysis of sera from 2020 identifies three children whose serum neutralization is escaped by mutations at site 160, suggesting that the T160I mutation may have been gained in response to antigenic selection from young children.

Our data illustrate how mutations to HA can have heterogeneous effects on virus neutralization by sera from different individuals in the human population. The heterogeneity in escape mutations is substantially clustered by age cohort, raising the question of whether certain age groups (e.g., children, teenagers, or adults) might play an especially important role in driving viral antigenic evolution (Kim et al., 2023; Nakajima et al., 2000). Previous work has proposed that school-aged children are important drivers of influenza epidemics in communities, based on transmission modeling (Basta et al., 2009; Wallinga et al., 2006), earlier peaks of infection in this age group (Worby et al., 2015), and the impact of school closure on transmission (Cauchemez et al., 2008; Huang et al., 2014). Whether the populations driving epidemics also drive antigenic drift is currently unclear. Although our study does not examine a sufficient number of sera or viruses to draw definitive conclusions, our results suggest that children and teenagers play an important role in shaping viral evolution, since mutations that fixed in HA often cause more escape from child and teenage sera than adults. More extensive characterization of the major escape mutations in different age cohorts and viruses could further test this hypothesis, and any insights into which segments of the population drive antigenic evolution could be used to improve evolutionary forecasting for influenza vaccine-strain selection.

Heterogeneity in the antigenic effects of mutations across the population could also have an effect on the overall rate of evolution of human influenza virus. If a specific mutation confers escape from neutralizing antibodies across all individuals, it will have a greater beneficial effect than if it only confers escape from some individuals—and this heterogeneity could impact evolutionary dynamics (Cobey and Pascual, 2011; Desai and Fisher, 2007; Gupta and Galvani, 1999; Luksza and Lässig, 2014). We therefore suggest that continued characterization of population-level heterogeneity in viral escape mutations, and incorporation of these measurements into models of evolution, could shed further light on the evolutionary dynamics of influenza and other viruses.

3.3 Limitations

Our experiments analyzed sera from a limited number of individuals from each age cohort, and these cohorts were sampled from a single city for each year (Seattle for the 2020 cohort, and Ho Chi Minh City for the 2010-2011 cohort). While this sample size was sufficient to illustrate heterogeneity in serum antibody targeting between age cohorts, larger sample sizes are needed to draw firm conclusions about how the heterogeneity in the antigenic effects of mutations shapes influenza evolution.

A further caveat is that we focused on sera that had high neutralizing titers against the vaccine strains used to build our libraries, in large part to reduce the serum volumes needed to perform the deep mutational scans. Therefore, if the antigenic effects of mutations vary systematically with the overall serum neutralizing titer, then our results might only reflect the subset of the population with high titers.

Our study coincides with the beginning of the SARS-CoV-2 pandemic, which may limit the generalizability of results measured from the 2020 cohort. More specifically, differences in social distancing measures implemented for children and adults may have influenced which age groups were stronger drivers of influenza evolution in subsequent years.

3.4 Figures

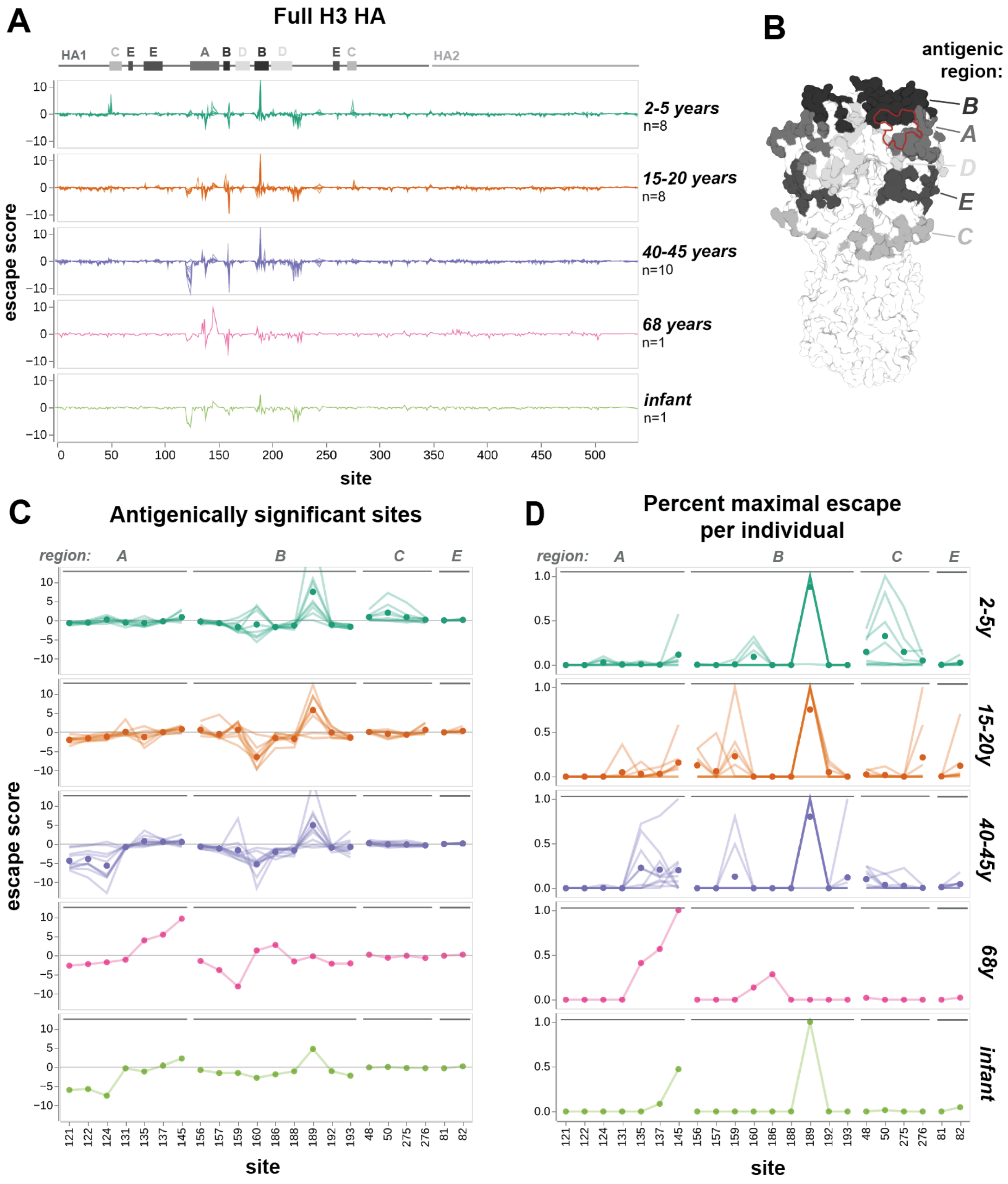


Figure 3.1: Escape maps from different cohorts show age-dependent trends in serum targeting of A/Hong Kong/45/2019 HA.

(A) Serum escape maps for different human individuals in each age cohort. Each line represents an individual, and individuals within the same cohort are overlaid in the same plot. Escape scores are the summed effects of all sampled mutations at each site. Approximate locations of HA1, HA2, and antigenic regions A through E (as defined by (Muñoz and Deem, 2005)) are labeled above the plots. (B) H3 HA trimer colored by antigenic region, with RBS outlined in red. Structure is from A/Victoria/361/2011 H3 HA (PDB 4O5N). (C) Serum escape maps at antigenically significant sites for human individuals in each age cohort. Each line is an individual escape map, and points represent the mean escape score for that age cohort at that site. The y-axis is clipped to 12.5 in both (A) and (C), although one child has an escape score of 23.9 at site 189. (D) Escape maps for key sites with escape floored at zero and normalized to the maximal positive escape score for that individual. Lines represent individuals and points represent the mean normalized escape score for that age cohort.

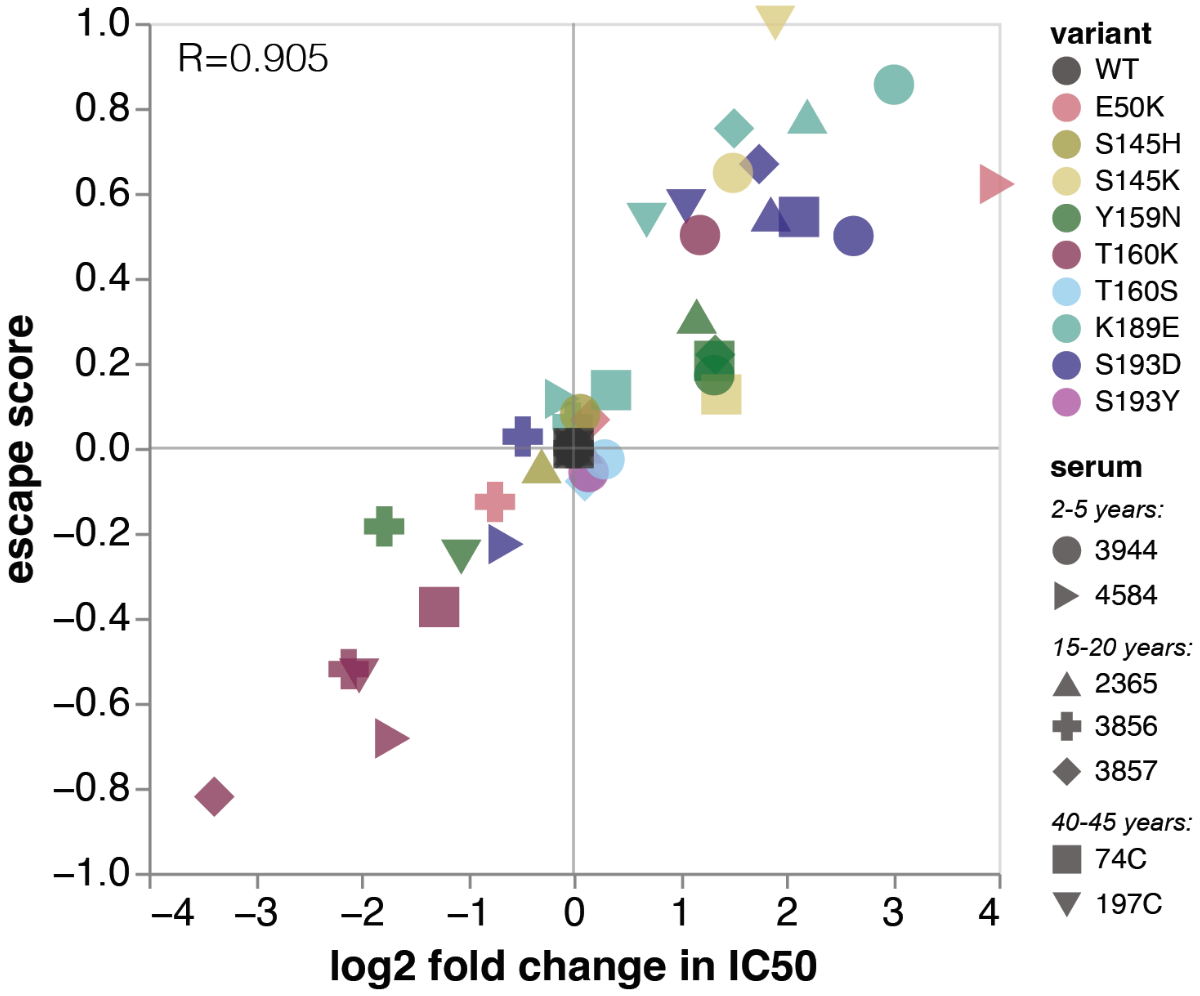


Figure 3.2: Validation of deep mutational scanning results using neutralization assays. The escape score measured by deep mutational scanning for each mutation is plotted against the log2 fold change in IC50 between the wildtype library strain and the mutant of interest. R indicates the Pearson correlation. See **Figure 3.9** for full neutralization curves and serum-level correlation plots.

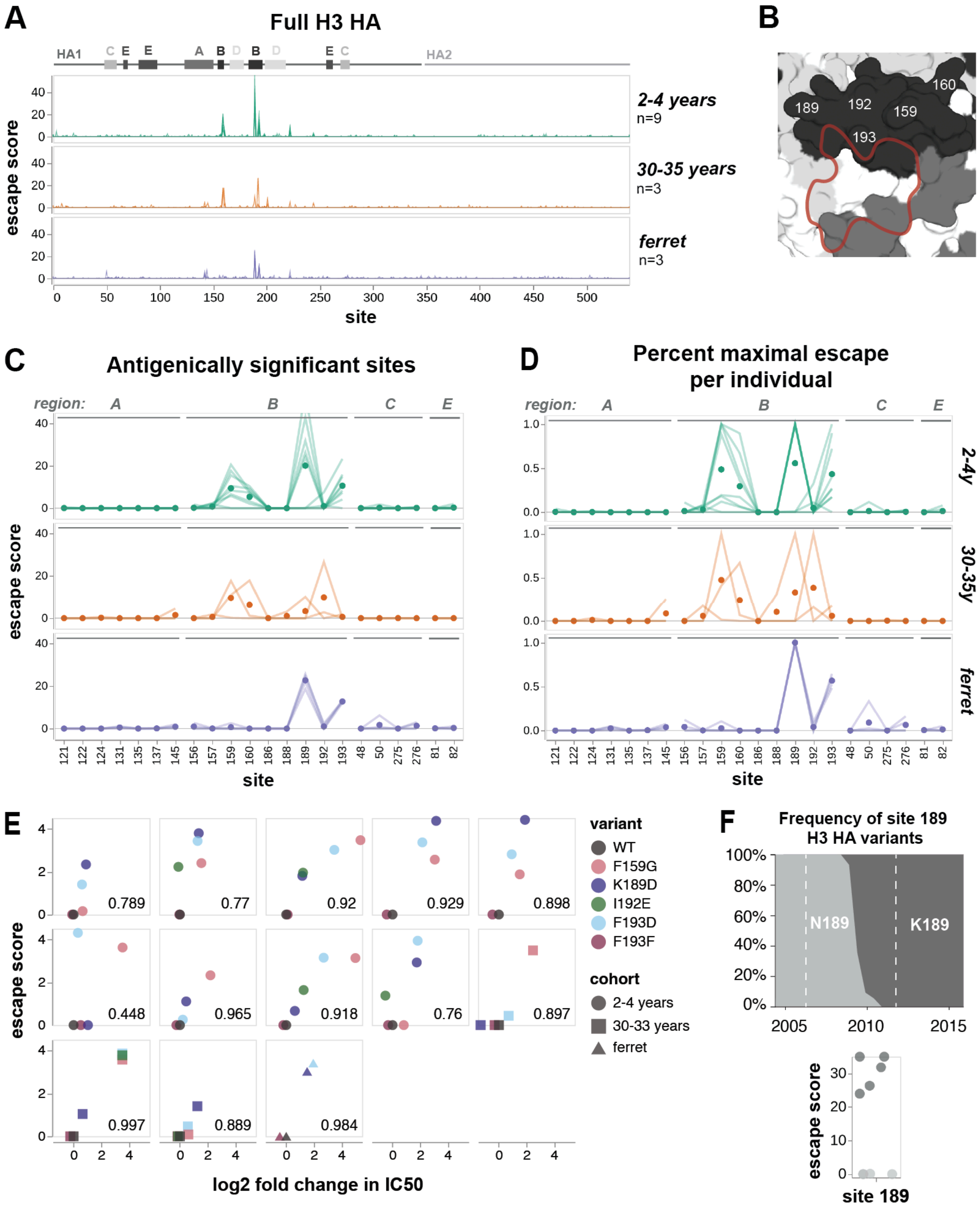


Figure 3.3: Escape maps from A/Perth/16/2009 H3 HA, measured against sera from an unvaccinated cohort and singly-infected ferrets.

(A) Serum escape maps for different individuals in each cohort. Each line represents an individual, and individuals within the same cohort are overlaid in the same plot. Escape scores are the summed effects of all sampled mutations at each site. Approximate locations of HA1, HA2, and antigenic regions A through E are labeled above the plots. (B) Antigenic region B of H3 HA is shown in black, with RBS outlined in red. Structure is from A/Victoria/361/2011 H3 HA (PDB 4O5N). (C and D) Serum escape maps at antigenically significant sites for individuals in each cohort. Each line is an individual escape map, and points represent the mean escape score for the cohort at that site. (C) shows the escape scores, while (D) shows the max normalized escape scores for each individual. (E) Correlation of escape scores measured in deep mutational scanning versus the fold-change in IC50 measured in traditional neutralization assays for each serum. F193F is a synonymous mutant used as a control. The Pearson correlation is indicated for each serum. See **Figure 3.12** for full neutralization curves. (F) Global frequency of site 189 H3 HA variants, compared to escape scores at site 189 from 2-4-year-old children. Dashed lines on frequency plot indicate maximum exposure period for this cohort, from the earliest birth date to latest serum collection date. Frequency plot adapted from the Nextstrain real-time pathogen evolution website (Hadfield et al., 2018; Neher and Bedford, 2015).

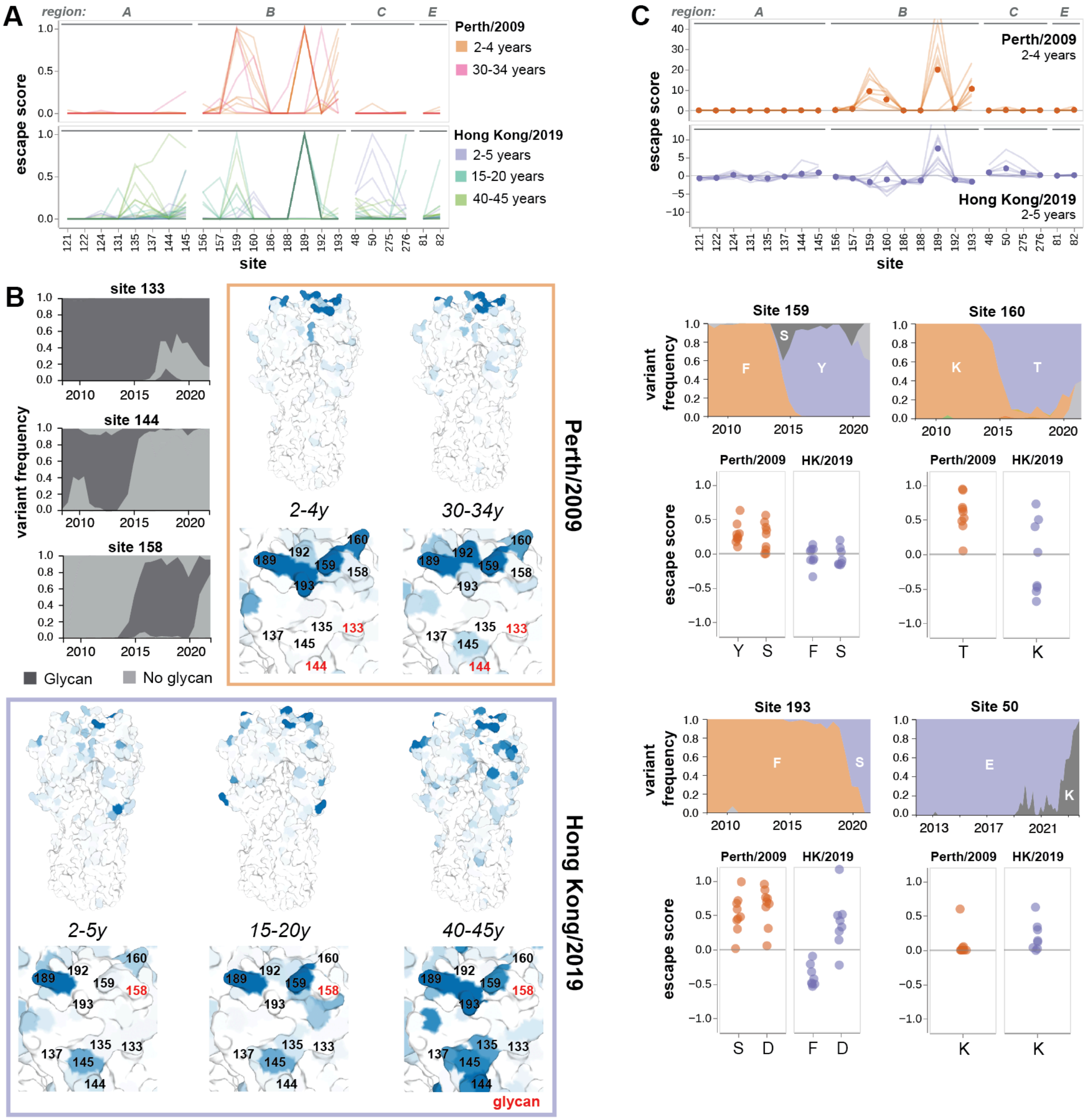


Figure 3.4: Comparison of serum escape maps between 2010-2011 and 2020 cohorts.

(A) Normalized serum escape maps at antigenically significant sites for different individuals in each cohort, colored by age group. Each line is an individual escape map, normalized by the maximum escape for that individual. (B) Maximal normalized escape score for each age cohort visualized on the H3 HA structure. Zoomed panels show key sites in antigenic regions A and B. Glycosylated sites (site 133 and 144 for A/Perth/16/2009, and site 158 for A/Hong Kong/45/2019) are indicated with red text, and the global frequency of variants with these glycans over time are plotted to the left. Structures are from A/Victoria/361/2011 H3 HA (PDB 4O5N). (C) Changes in escape maps from young children between 2010 and 2020, compared to the emergence of new variants at those sites. Global frequency plots of H3 HA variants at sites 159, 160, and 193 are colored by the wildtype amino acid in A/Perth/16/2009 (F159, K160, F193) or A/Hong Kong/45/2019 (Y159, T160, S193). At site 50, both library strains have the same genotype (E50). Escape scores for mutations to amino acids circulating from 2010-2020 are plotted below. Points represent the mutation escape score for each individual in the cohort. For the 2010-2011 cohort, escape scores were normalized to one at each site, to facilitate comparison to escape scores from the 2020 cohort. Frequency plots in (B) and (C) were adapted from the Nextstrain real-time pathogen evolution website (Hadfield et al., 2018; Neher and Bedford, 2015).

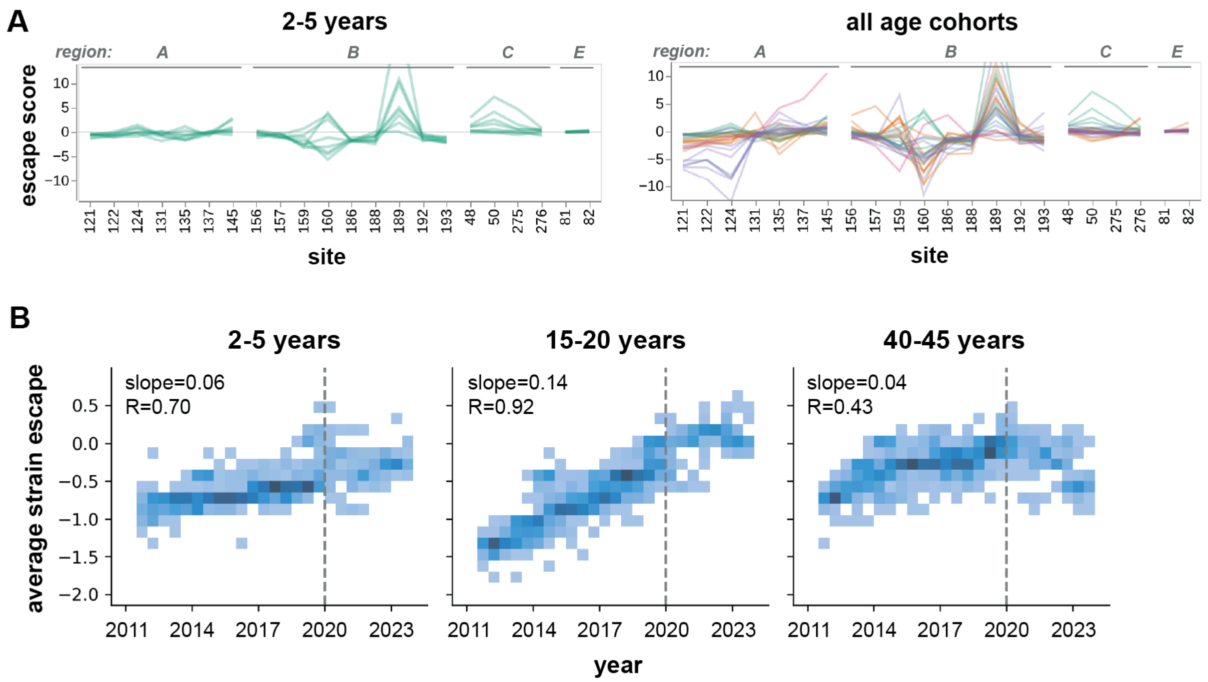


Figure 3.5: Relationship between escape mapped in deep mutational scanning and actual H3N2 HA evolution in humans.

(A) Serum escape maps at key antigenic sites for individuals in either the 2-5-year-old cohort, or all age cohorts, tested against the A/Hong Kong/45/2019 HA. Each line is an individual escape map. (B) Escape scores for H3N2 influenza strains circulating from 2012 to 2023, averaged across sera within each cohort. Strain escape is calculated based on the sum of the cohort-average escape scores for all mutations in that strain, relative to the parental library strain A/Hong Kong/45/2019, which has an escape score of zero. Plots show a 2D histogram of escape scores for approximately 1,200 H3N2 HA variants sampled during this timeframe. Dashed line indicates the year of serum collection (2020). The slope of the best-fit line relating year to average strain escape (line not shown) is reported for each cohort, along with the Pearson correlation. The slope is significantly higher for the teenage cohort compared to the child (*p*=0.022) and adult (*p*=0.004) cohorts.

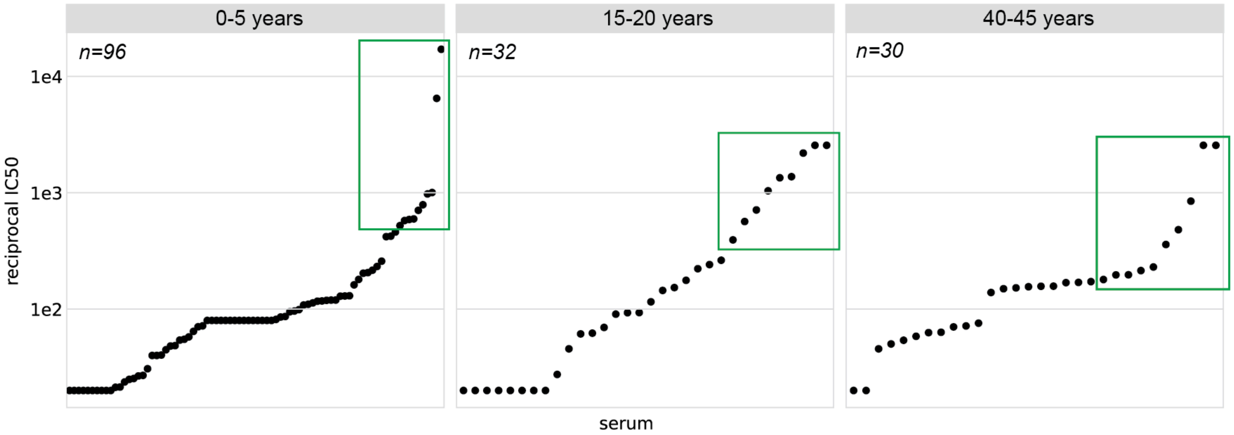


Figure 3.6: Serum neutralizing titers against the A/Hong Kong/45/2019 HA barcoded library strain.

Neutralization assays were performed using influenza carrying GFP in the PB1 segment, as described previously (Doud et al., 2018; Hooper and Bloom, 2013); see https://github.com/jbloomlab/flu_PB1flank-GFP_neut_assay for detailed protocol). Only a single replicate was run for each serum. IC50 values were calculated by fitting Hill-like curves using the neutcurve Python package (https://jbloomlab.github.io/neutcurve/). The ten sera from each age group with the highest neutralization activity against the parental library strain, boxed in green, were selected for serum escape mapping.

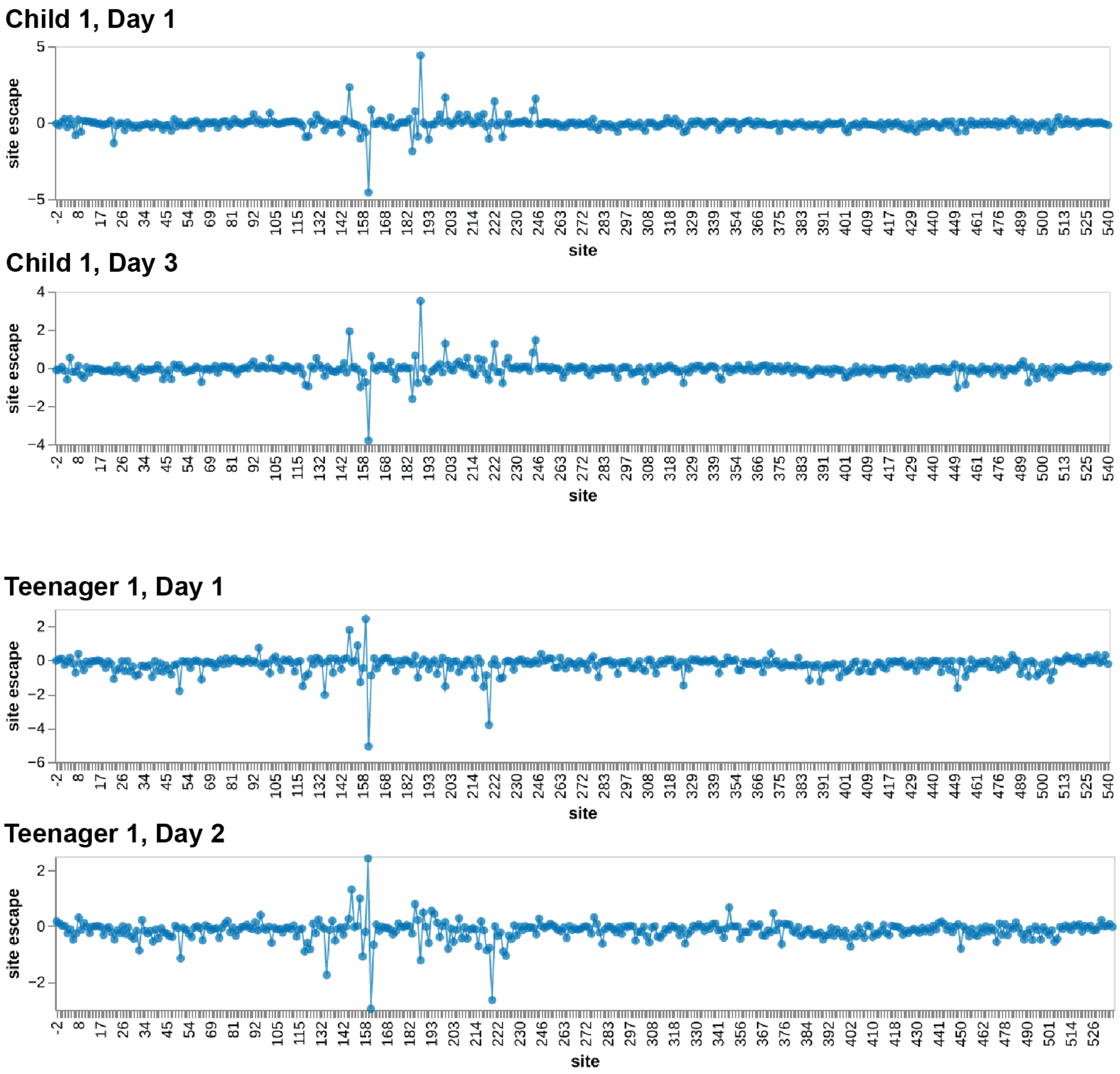
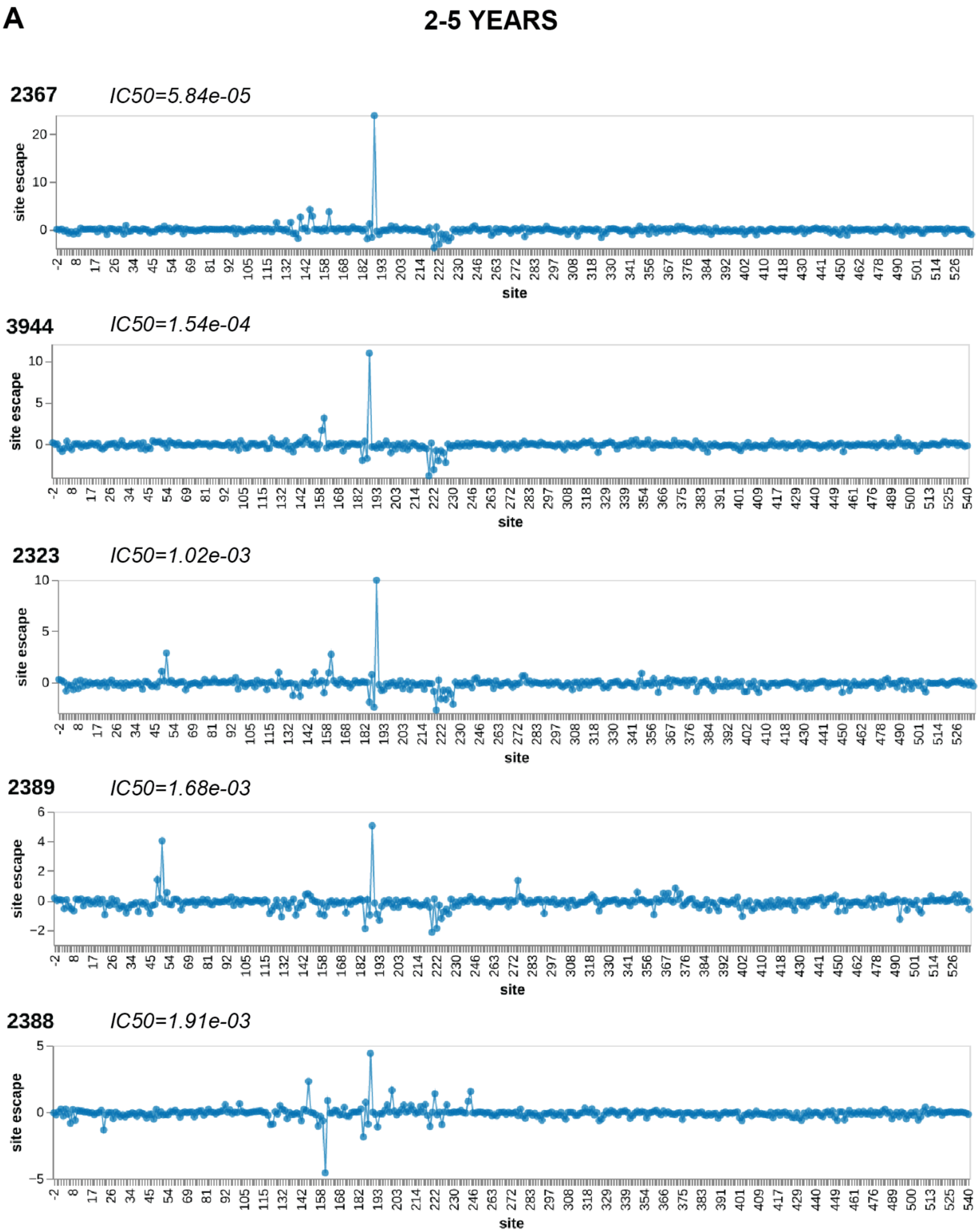
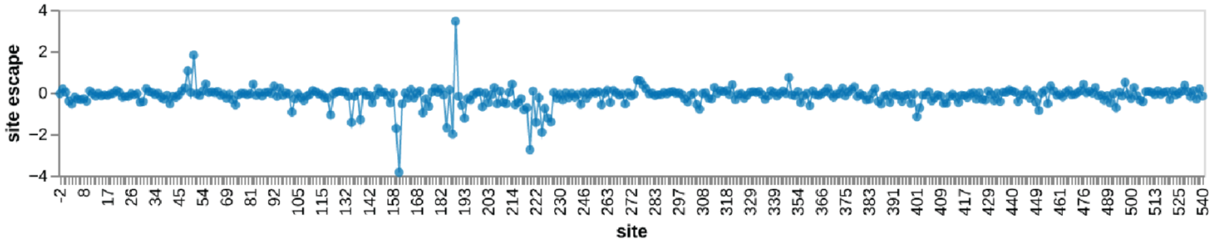


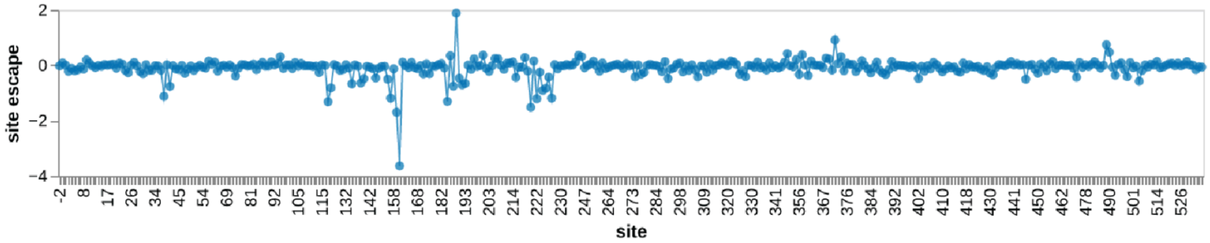
Figure 3.7: Escape maps for sera collected on different days from the same individual. One set of repeated samples comes from a child (sample 2388), and the other from a teenager (sample 3862). Line plots show summed escape scores of each sampled mutation at that site.


A
2-5 YEARS
2367
IC50=5.84e-05
site escape
20
10
0
-2 8 17 26 34 45 54 69 81 92 105 115 132 142 158 168 182 193 203 214 222 230 246 263 272 283 297 308 318 330 341 356 367 376 384 392 402 410 418 430 441 450 462 478 490 501 514 526
site
3944
IC50=1.54e-04
site escape
10
5
0
-2 8 17 26 34 45 54 69 81 92 105 115 132 142 158 168 182 193 203 214 222 230 246 263 272 283 297 308 318 329 339 354 366 375 383 391 401 409 417 429 440 449 461 476 489 500 513 525 540
site
2323
IC50=1.02e-03
site escape
10
5
0
-2 8 17 26 34 45 54 69 81 92 105 115 132 142 158 168 182 193 203 214 222 230 246 263 272 283 297 308 318 330 341 356 367 376 384 392 402 410 418 430 441 450 462 478 490 501 514 526
site
2389
IC50=1.68e-03
site escape
6
4
2
0
−2
-2 8 17 26 34 45 54 69 81 92 105 115 132 142 158 168 182 193 203 214 222 230 246 263 272 283 297 308 318 330 341 356 367 376 384 392 402 410 418 430 441 450 462 478 490 501 514 526
site
2388
IC50=1.91e-03
site escape
5
0
−5
-2 8 17 26 34 45 54 69 81 92 105 115 132 142 158 168 182 193 203 214 222 230 246 263 272 283 297 308 318 329 339 354 366 375 383 391 401 409 417 429 440 449 461 476 489 500 513 525 540
site

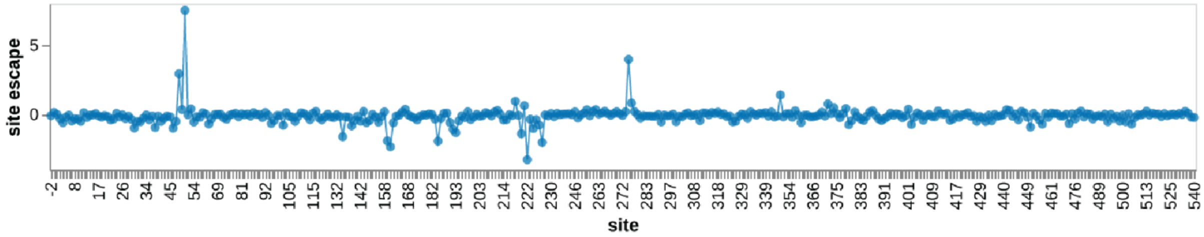
(2-5 YEARS)



4299 *IC50=3.87e-03*



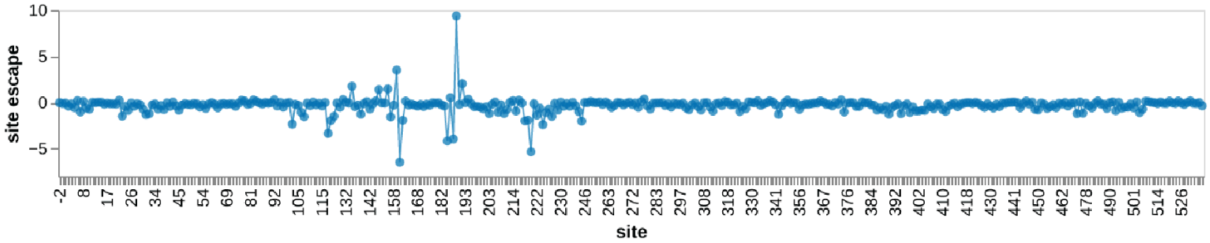
3973 *IC50=4.30e-03*



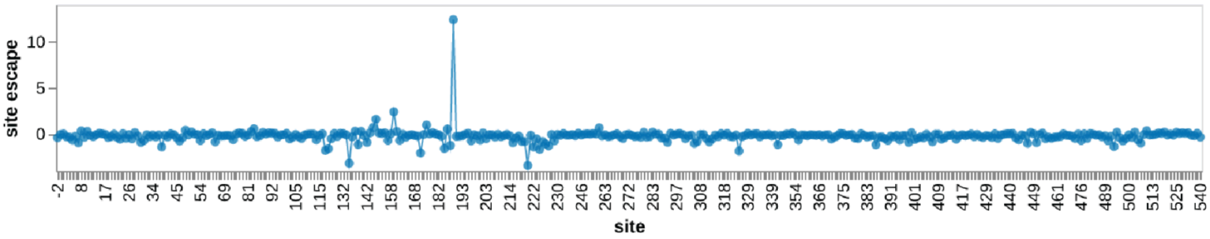
4584 *IC50=4.63e-03*

B

15-20 YEARS



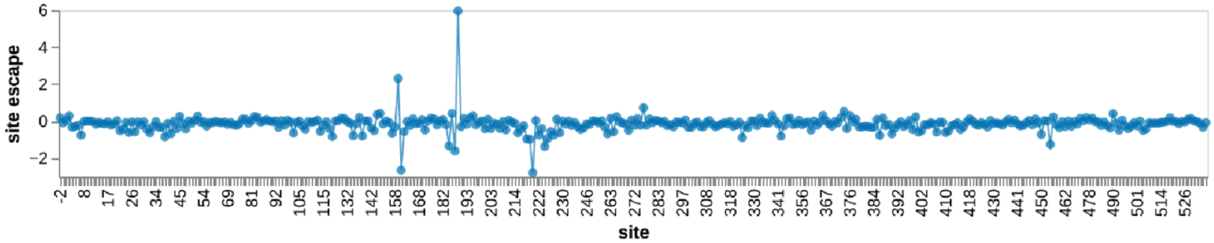
2350 *IC50=3.91e-04*



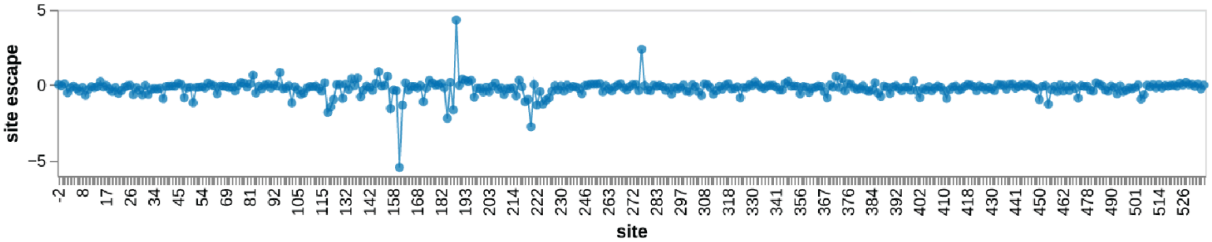
3866 *IC50=3.91e-04*

(15-20 YEARS)

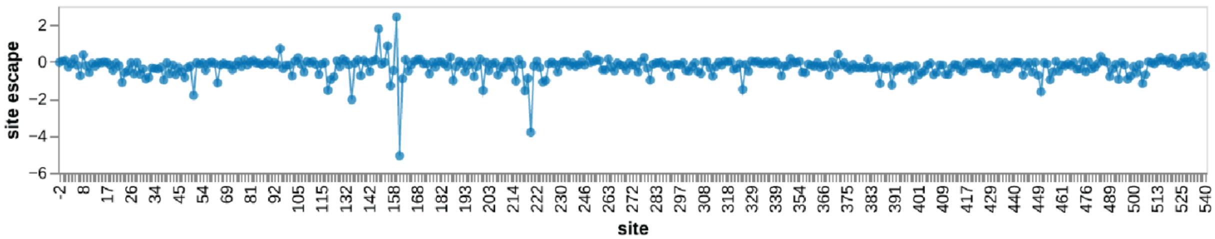
2365 *IC50=4.56e-04*



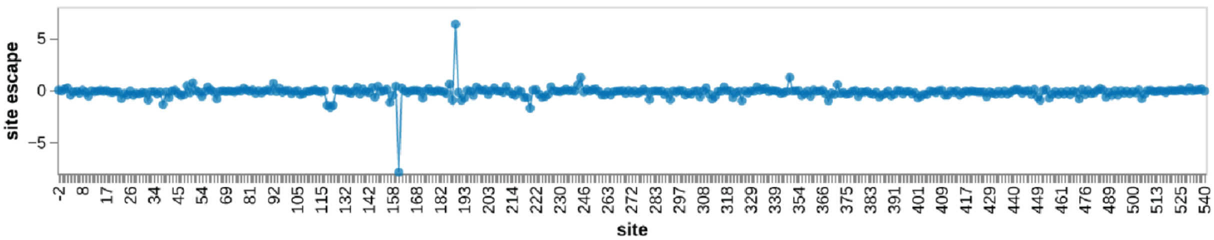
2382 *IC50=7.43e-04*



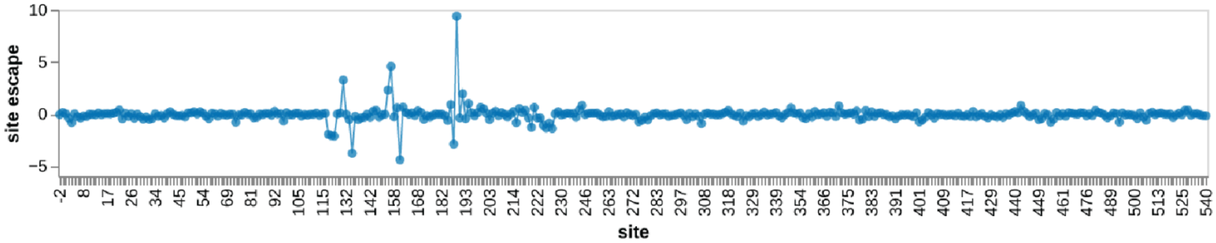
3862 *IC50=9.61e-04*



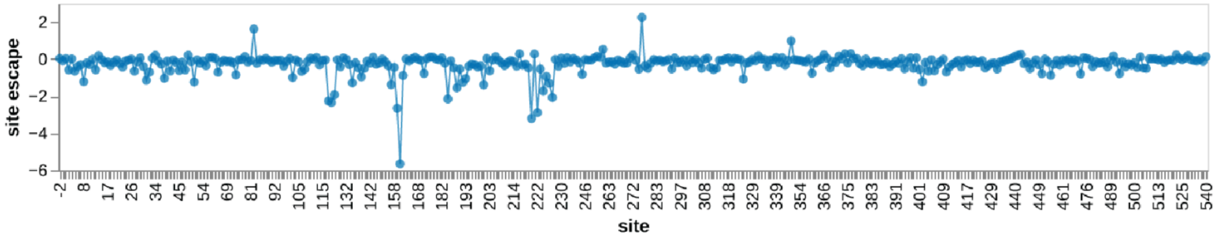
3857 *IC50=2.54e-03*



2380 *IC50=3.80e-03*



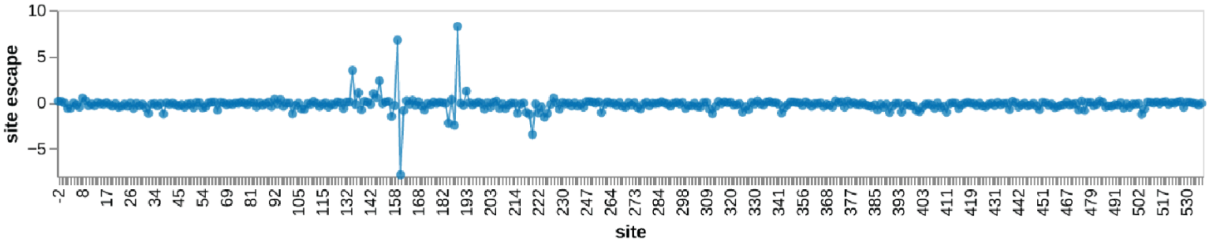
3856 *IC50=4.49e-03* **(15-20 YEARS)**



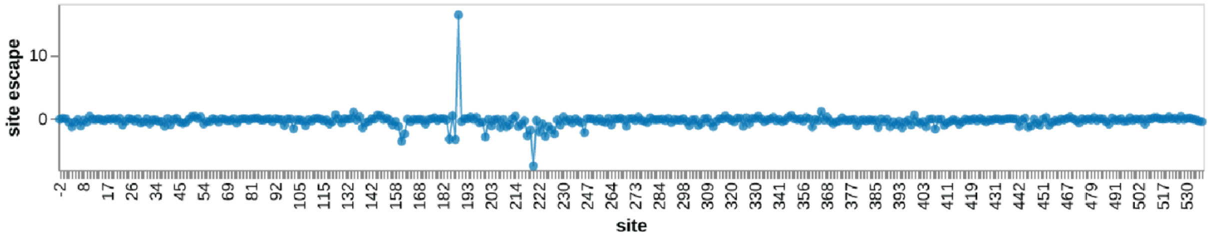
C

40-45 YEARS

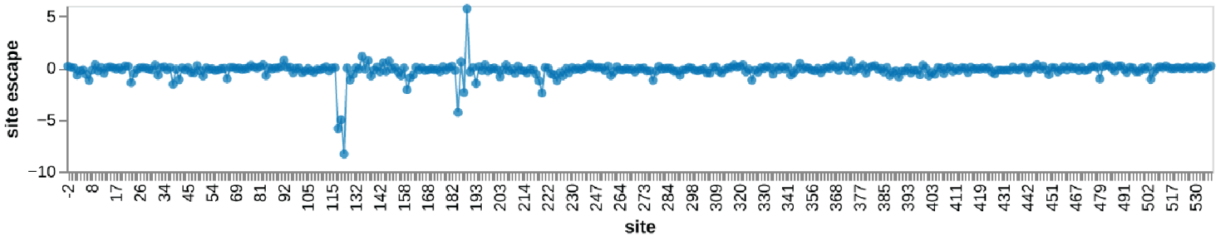
215C *IC50=1.98e-04*



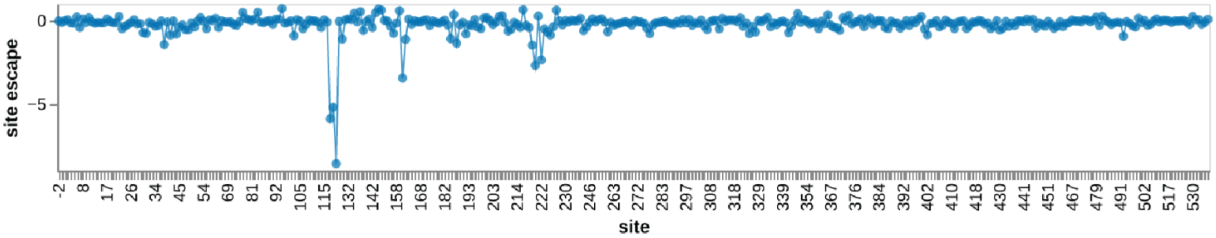
34C *IC50=2.79e-04*



33C *IC50=1.31e-03*

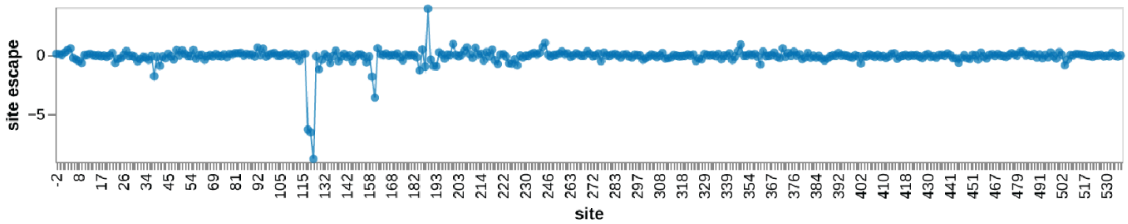


74C *IC50=2.08e-03*

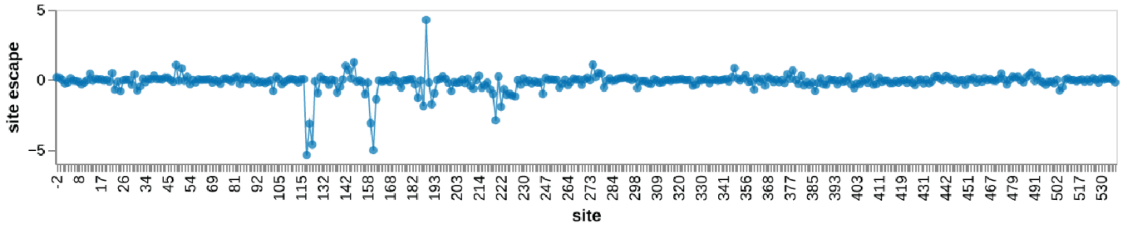


(40-45 YEARS)

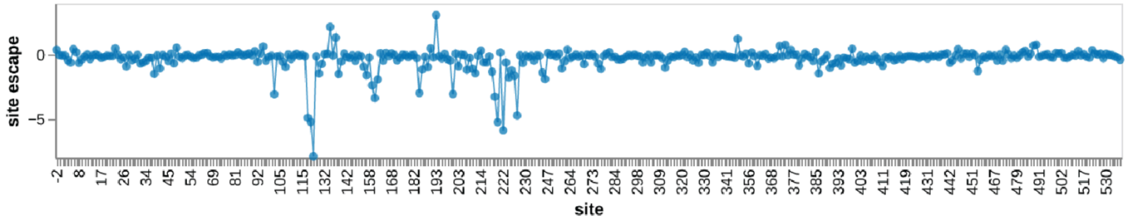
18C *IC50=2.78e-03*



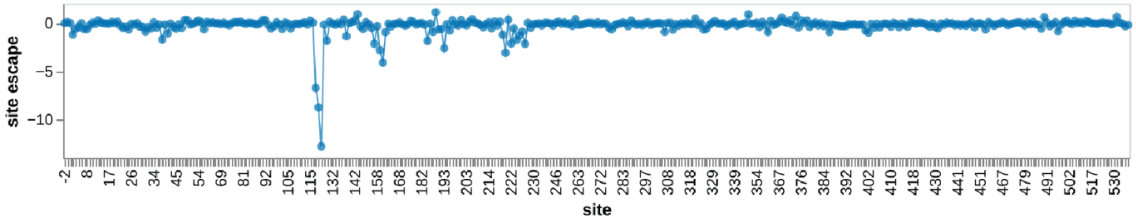
197C *IC50=3.3e-03*



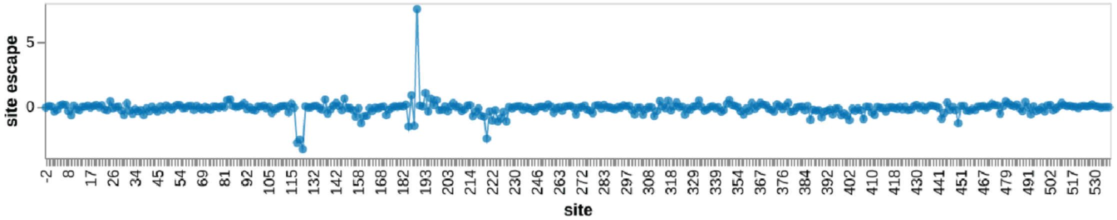
199C *IC50=5.08e-03*



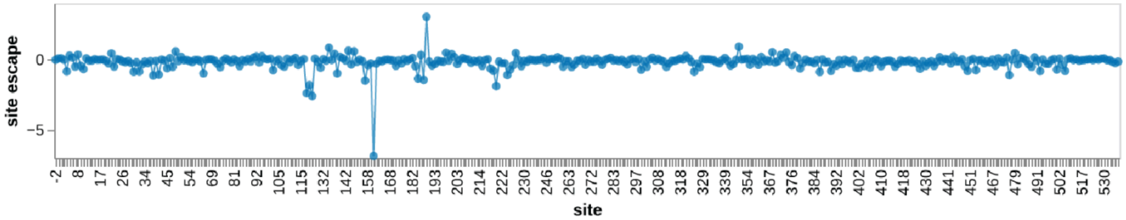
210C *IC50=5.80e-03*



150C *IC50=6.42e-03*



68C *IC50=7.19e-03*



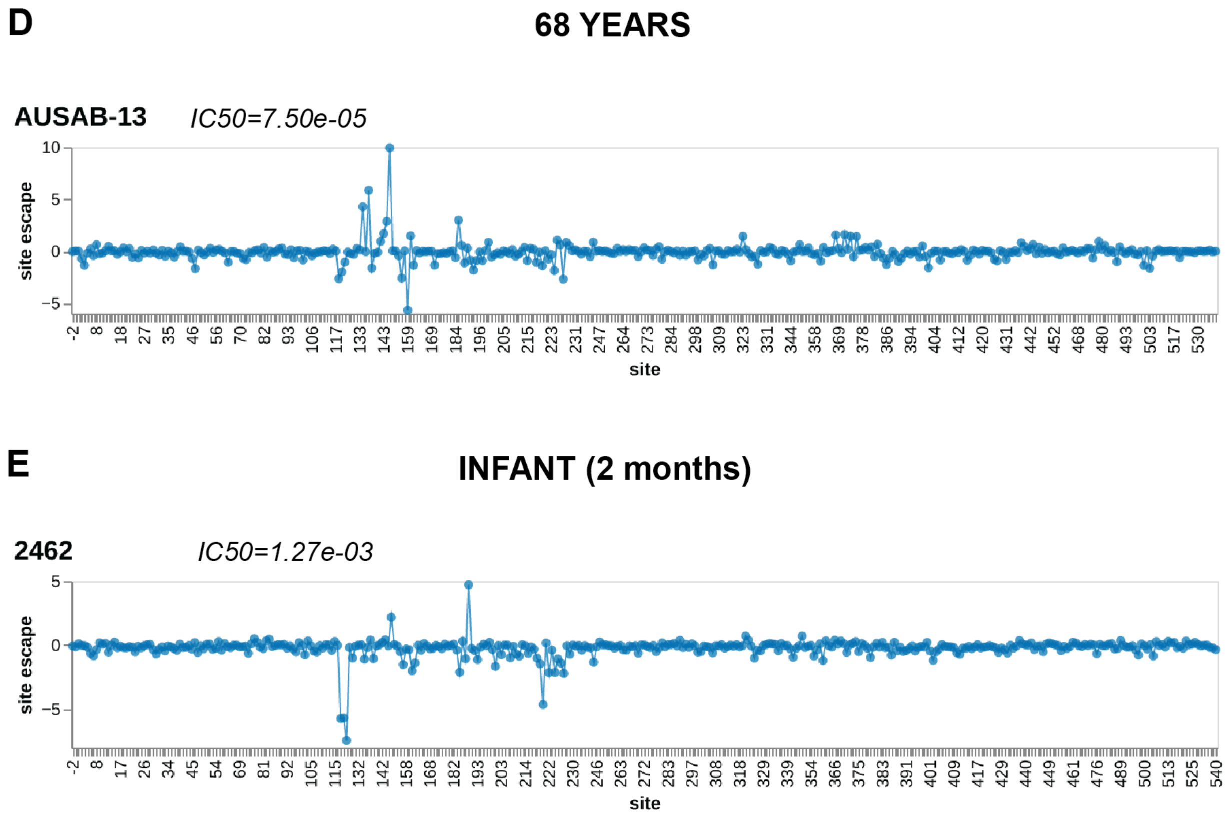


Figure 3.8: Individual escape maps for each serum in the 2020 cohort, analyzed against the A/Hong Kong/45/2019 library.

Line plots show summed escape scores of each sampled mutation at that site. Sera are ordered by decreasing potency within each age group, with IC50s labeled for each serum. (A) Child sera, (B) teenage sera, (C) adult sera, (D) elderly serum, and (E) infant serum.

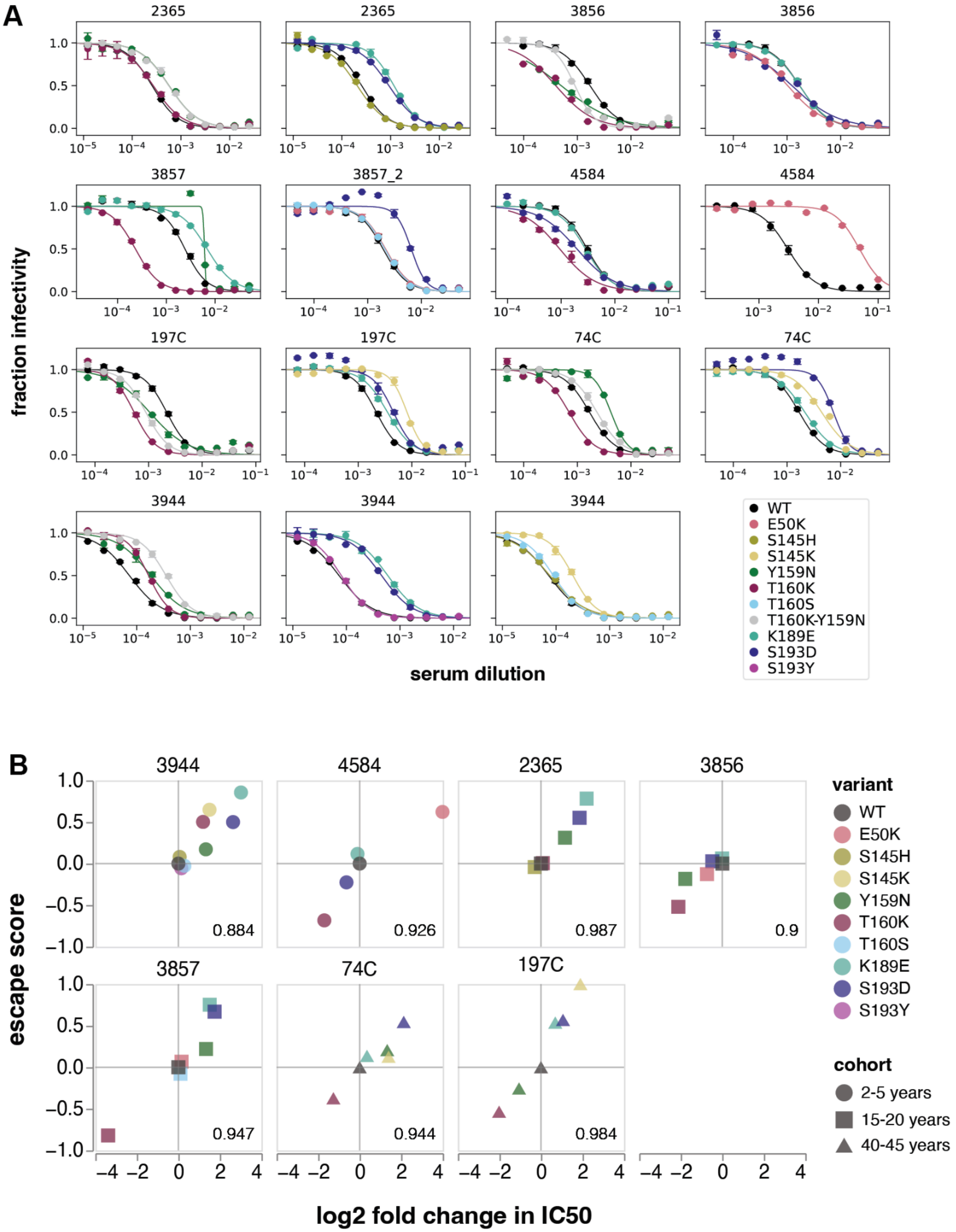

A
2365
2365
3856
3856
3857
3857_2
4584
4584
197C
197C
74C
74C
3944
3944
3944
fraction infectivity
1.0
0.5
0.0
10^{-5}
10^{-4}
10^{-3}
10^{-2}
10^{-1}
serum dilution
WT
E50K
S145H
S145K
Y159N
T160K
T160S
T160K-Y159N
K189E
S193D
S193Y
B
3944
4584
2365
3856
3857
74C
197C
escape score
1.0
0.5
0.0
−0.5
−1.0
0.884
0.926
0.987
0.9
0.947
0.944
0.984
−4
−2
0
2
4
log2 fold change in IC50
variant
WT
E50K
S145H
S145K
Y159N
T160K
T160S
K189E
S193D
S193Y
cohort
2-5 years
15-20 years
40-45 years

Figure 3.9: Neutralization assay results for validation of deep mutational scanning measurements against the A/Hong Kong/45/2019 library.

(A) Neutralization curves for unmutated A/Hong Kong/45/2019 H3 HA (‘WT’) and selected mutants against seven representative sera from different age cohorts. (B) Correlation between escape score and the log2 fold change in serum IC50 for each variant, plotted for each serum independently. Pearson R correlation is noted on each plot. The IC50 for each variant was compared to the IC50 for the wildtype strain run in the same experiment, to control for potential variation in serum concentration and cell number between independent experiments.

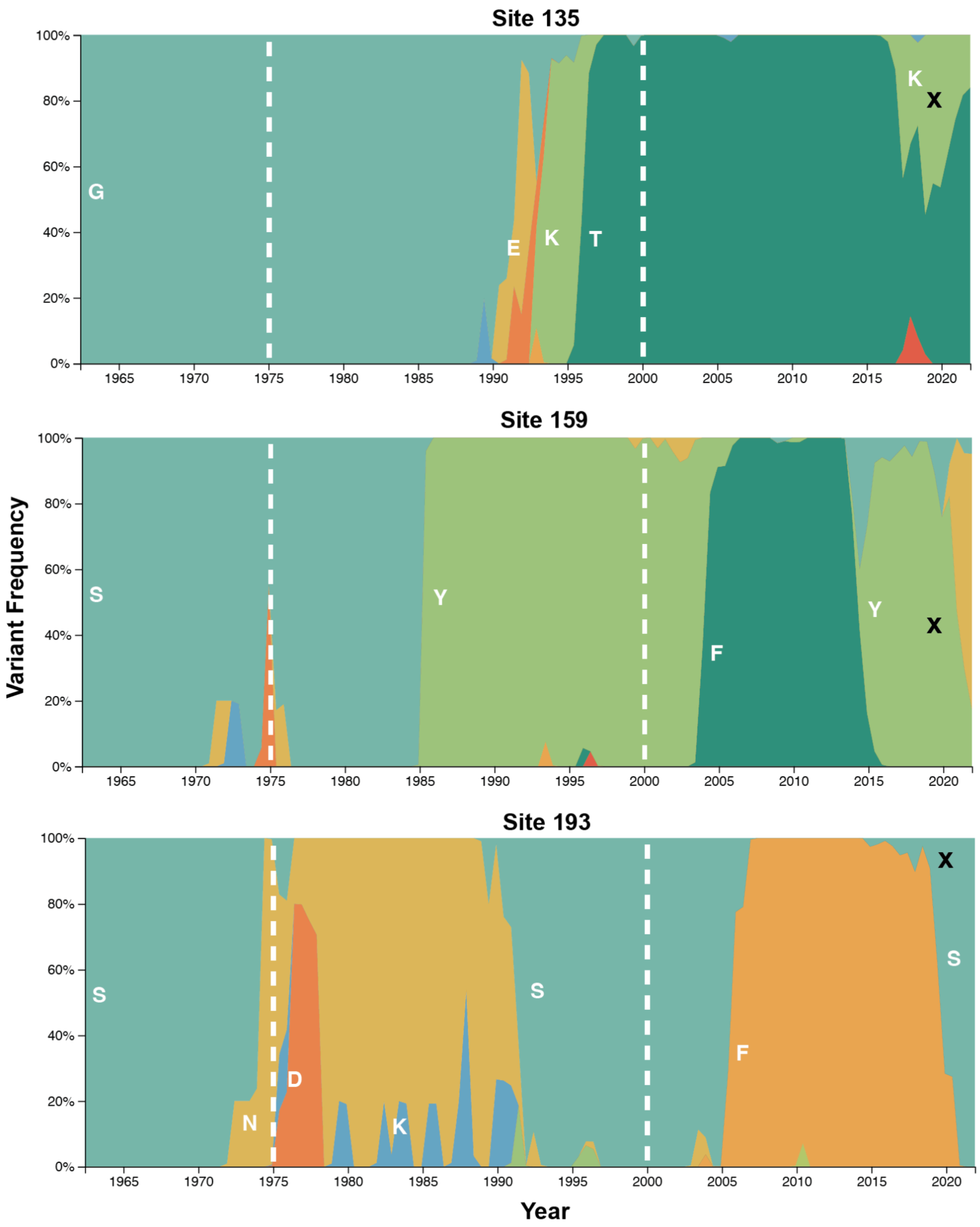
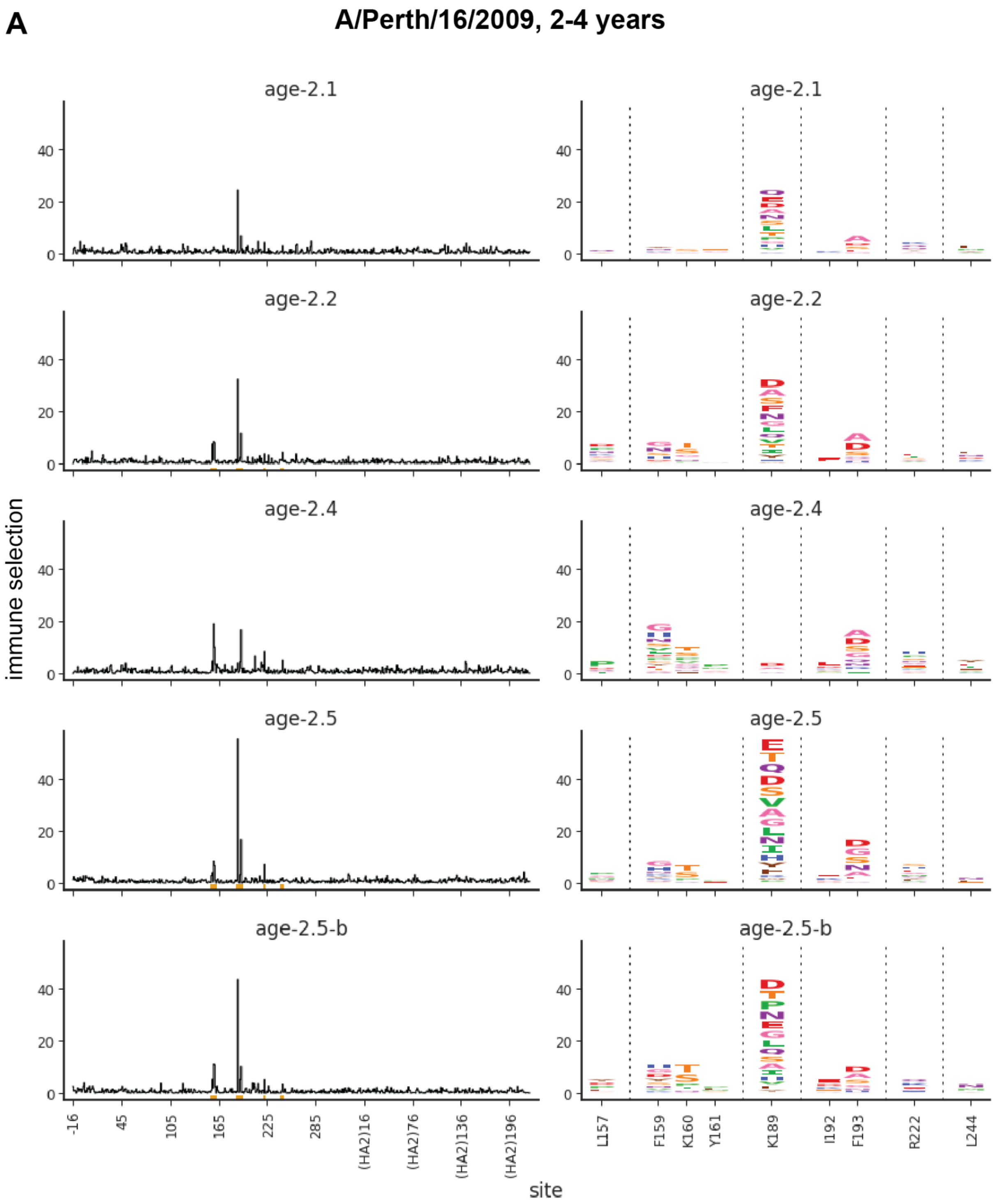
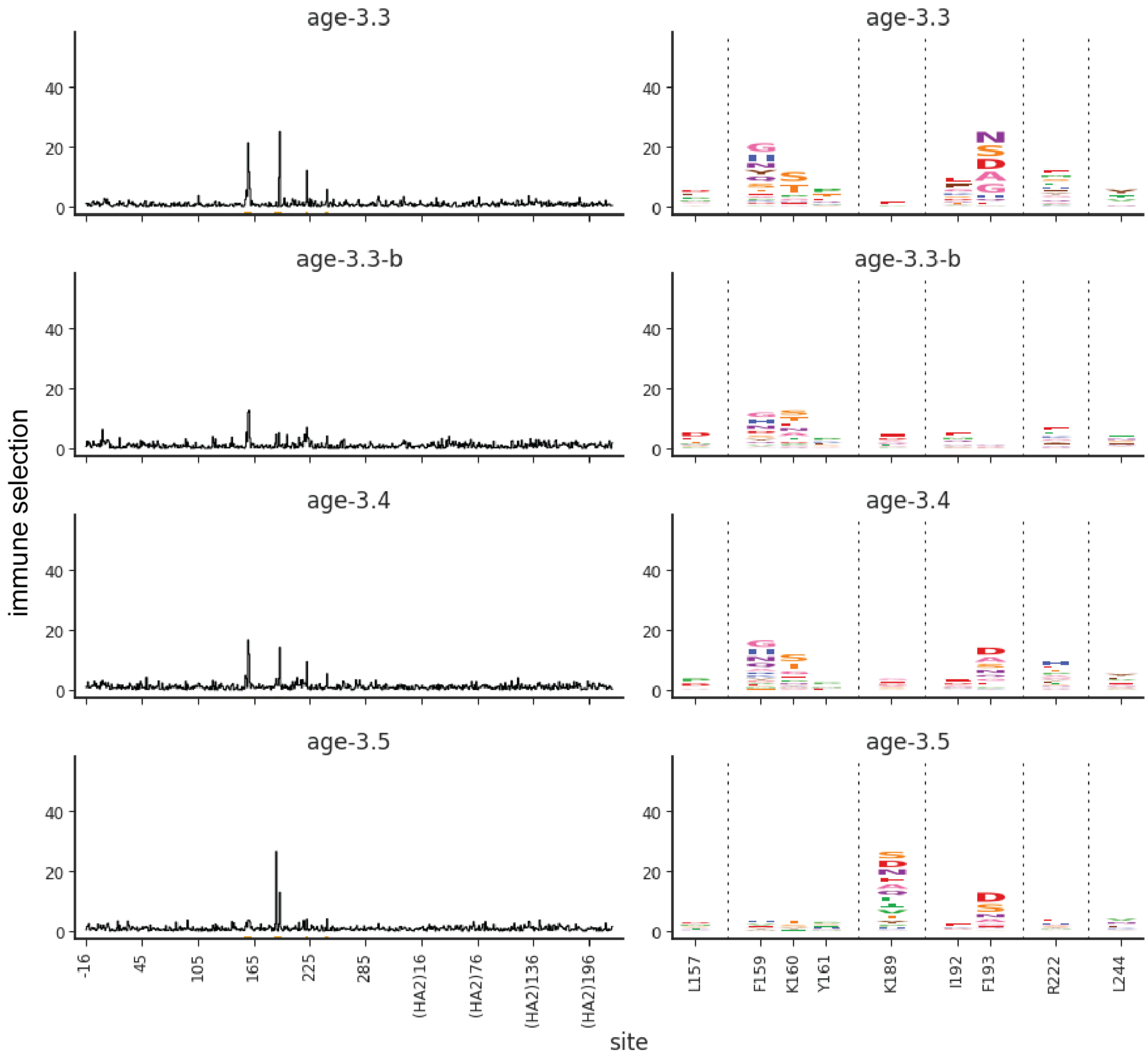


Figure 3.10: Global frequency of variants at sites 135, 159, and 193.

Dashed lines indicate the earliest birthdate of individuals in the adult cohort (1975) and teenage cohort (2000). Black 'X' indicates the amino acid identity of the wildtype library strain, A/Hong Kong/45/2019. Frequency plot adapted from the Nextstrain real-time pathogen evolution website (Hadfield et al., 2018; Neher and Bedford, 2015).


A
A/Perth/16/2009, 2-4 years
age-2.1
age-2.1
age-2.2
age-2.2
age-2.4
age-2.4
age-2.5
age-2.5
age-2.5-b
age-2.5-b
immune selection
0
20
40
-16
45
105
165
225
285
(HA2)16
(HA2)76
(HA2)136
(HA2)196
site
L157
F159
K160
Y161
K189
I192
F193
R222
L244

A/Perth/16/2009, 2-4 years (cont.)



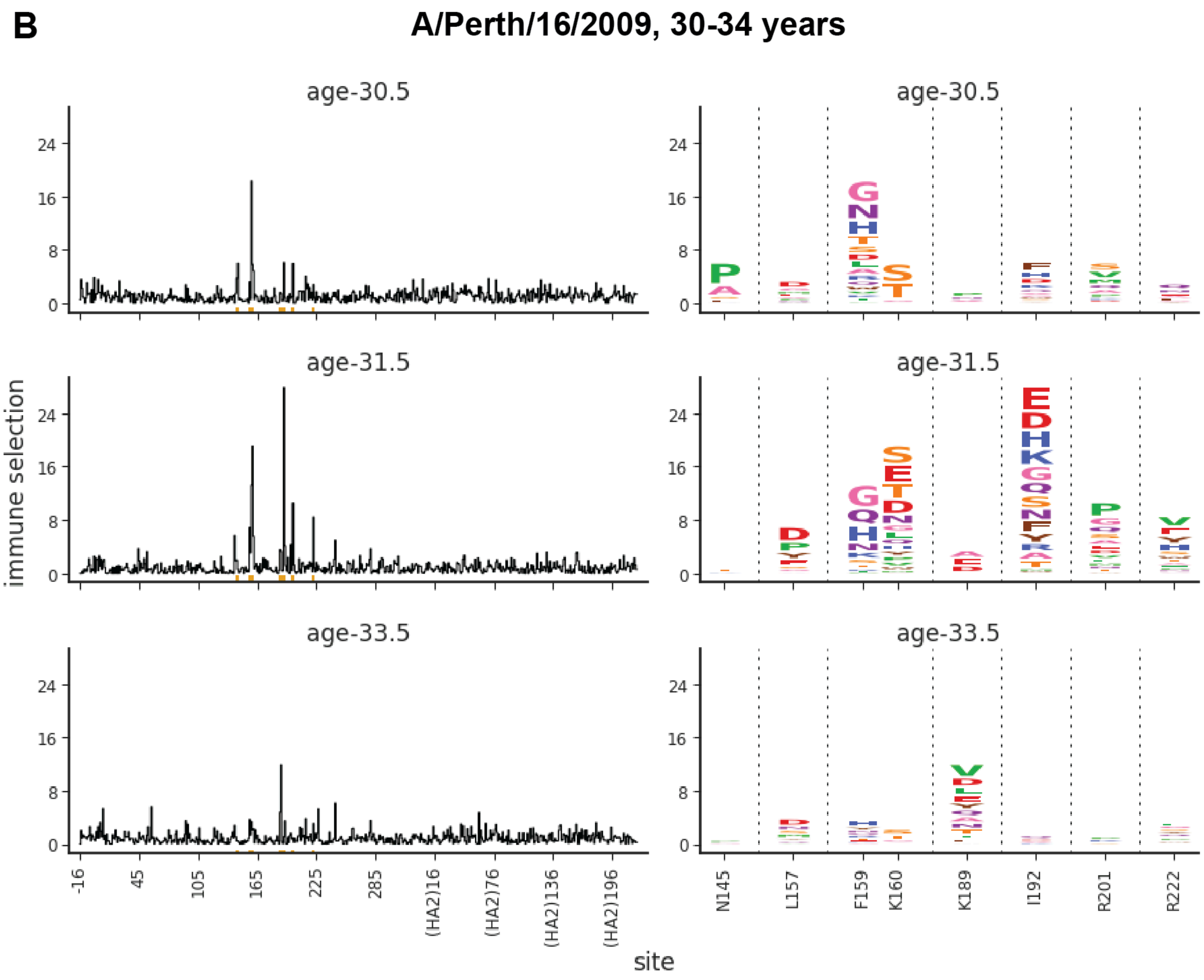


Figure 3.11: Individual escape maps for each serum in the 2010-2011 cohort, analyzed against the A/Perth/16/2009 library.

Line plots show summed escape scores of each sampled mutation at that site. Logo plots show mutation-level escape at key sites, where the height of each letter corresponds to the escape score for that amino acid substitution. (A) Child sera, (B) adult sera.

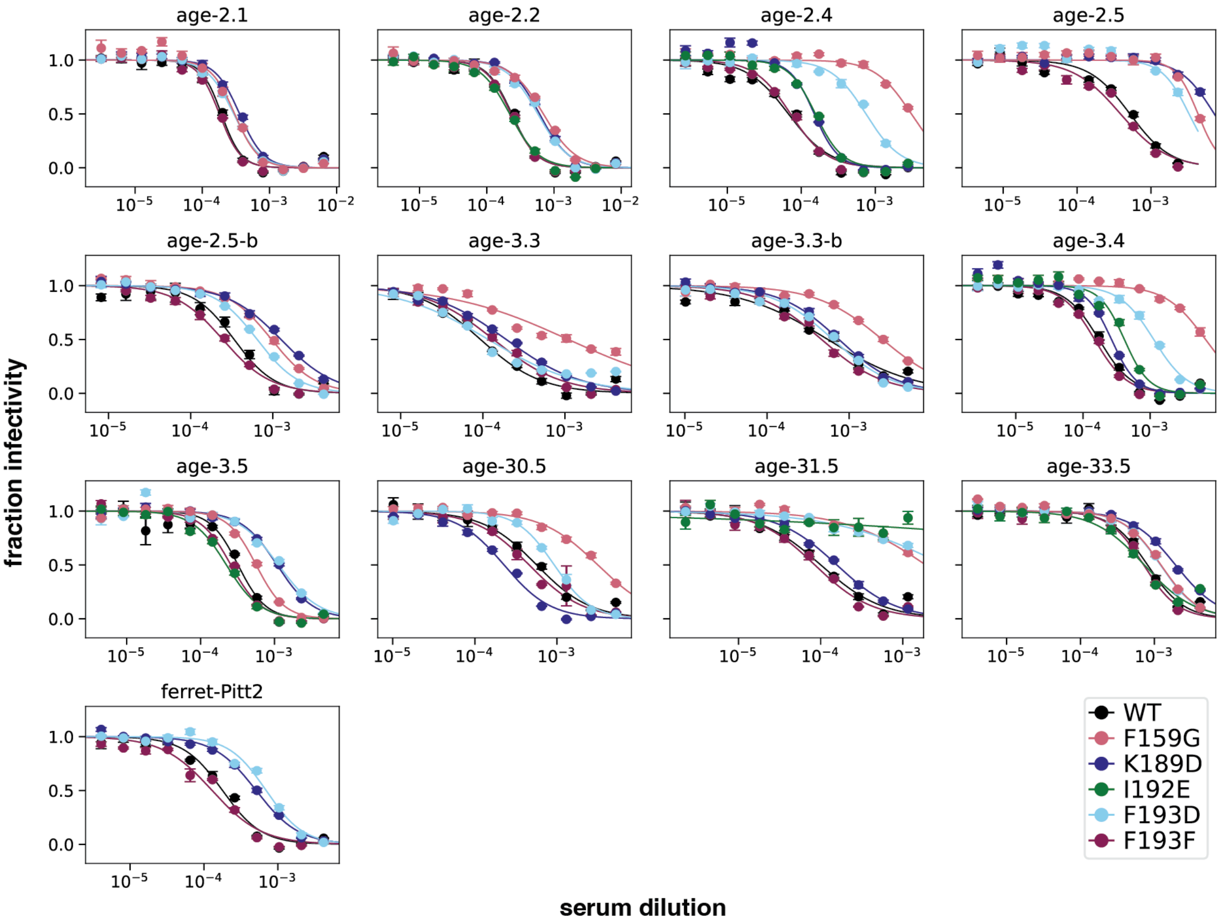
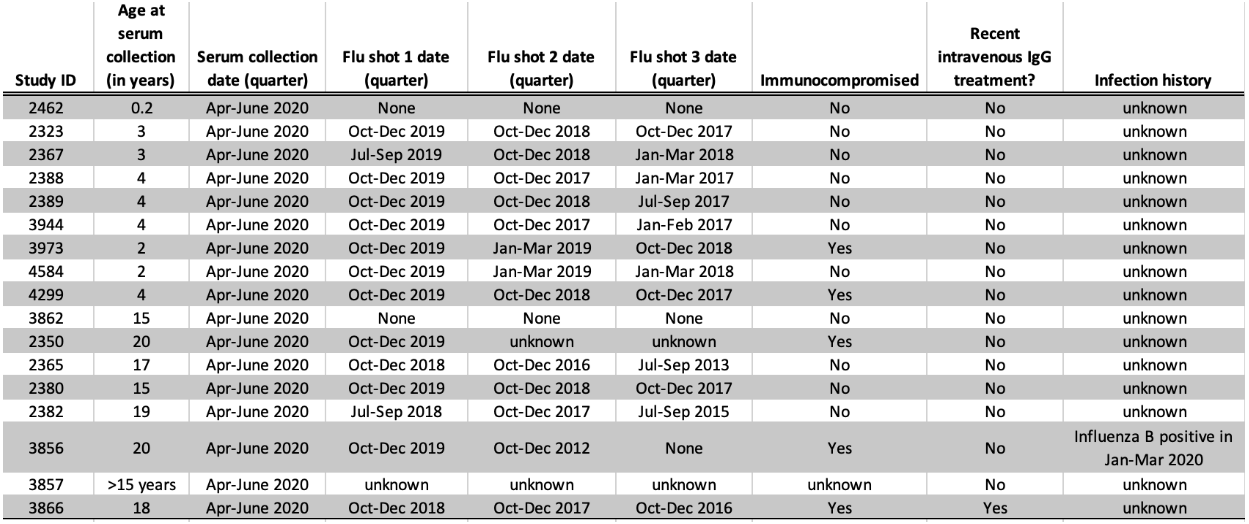


Figure 3.12: Neutralization assay results for validation of deep mutational scanning measurements against the A/Perth/16/2009 library.

Neutralization curves are shown for unmutated A/Perth/16/2009 H3 HA ('WT') and selected mutants against all human sera from the 2010-2011 cohort, plus one representative ferret serum.

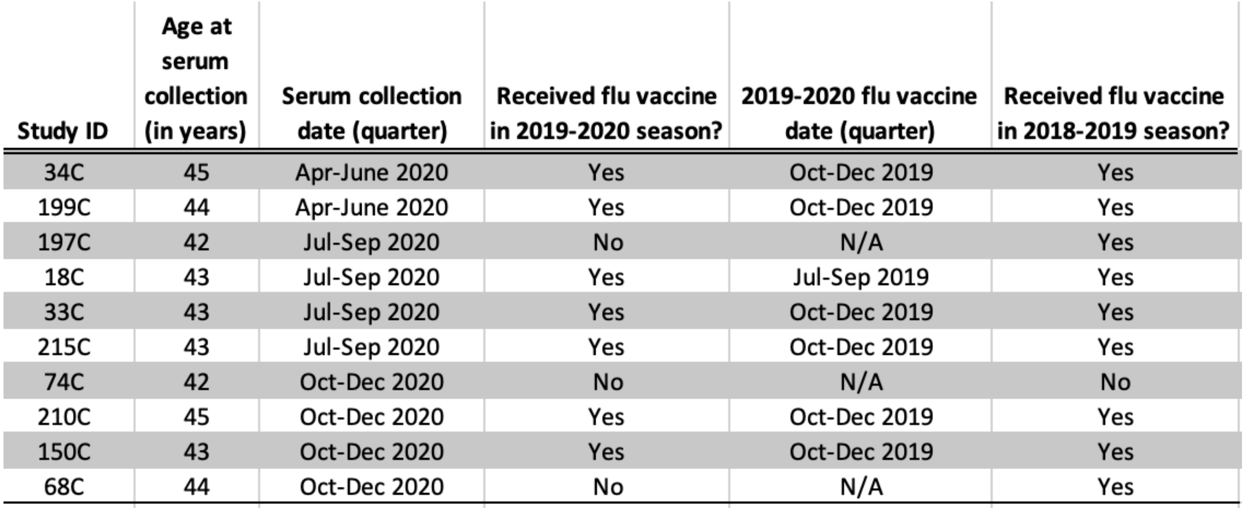
3.5 Tables



Study ID	Age at serum collection (in years)	Serum collection date (quarter)	Flu shot 1 date (quarter)	Flu shot 2 date (quarter)	Flu shot 3 date (quarter)	Immunocompromised	Recent intravenous IgG treatment?	Infection history
2462	0.2	Apr-June 2020	None	None	None	No	No	unknown
2323	3	Apr-June 2020	Oct-Dec 2019	Oct-Dec 2018	Oct-Dec 2017	No	No	unknown
2367	3	Apr-June 2020	Jul-Sep 2019	Oct-Dec 2018	Jan-Mar 2018	No	No	unknown
2388	4	Apr-June 2020	Oct-Dec 2019	Oct-Dec 2017	Jan-Mar 2017	No	No	unknown
2389	4	Apr-June 2020	Oct-Dec 2019	Oct-Dec 2018	Jul-Sep 2017	No	No	unknown
3944	4	Apr-June 2020	Oct-Dec 2019	Oct-Dec 2017	Jan-Feb 2017	No	No	unknown
3973	2	Apr-June 2020	Oct-Dec 2019	Jan-Mar 2019	Oct-Dec 2018	Yes	No	unknown
4584	2	Apr-June 2020	Oct-Dec 2019	Jan-Mar 2019	Jan-Mar 2018	No	No	unknown
4299	4	Apr-June 2020	Oct-Dec 2019	Oct-Dec 2018	Oct-Dec 2017	Yes	No	unknown
3862	15	Apr-June 2020	None	None	None	No	No	unknown
2350	20	Apr-June 2020	Oct-Dec 2019	unknown	unknown	Yes	No	unknown
2365	17	Apr-June 2020	Oct-Dec 2018	Oct-Dec 2016	Jul-Sep 2013	No	No	unknown
2380	15	Apr-June 2020	Oct-Dec 2019	Oct-Dec 2018	Oct-Dec 2017	No	No	unknown
2382	19	Apr-June 2020	Jul-Sep 2018	Oct-Dec 2017	Jul-Sep 2015	No	No	unknown
3856	20	Apr-June 2020	Oct-Dec 2019	Oct-Dec 2012	None	Yes	No	Influenza B positive in Jan-Mar 2020
3857	>15 years	Apr-June 2020	unknown	unknown	unknown	unknown	No	unknown
3866	18	Apr-June 2020	Oct-Dec 2018	Oct-Dec 2017	Oct-Dec 2016	Yes	Yes	unknown

Table 3.1: Serum metadata for pediatric cohorts.

Age, serum collection date, recent vaccination history, and relevant medical history for the unvaccinated infant and individuals in the 2-5 and 15-20 year age cohorts from Seattle, Washington.



Study ID	Age at serum collection (in years)	Serum collection date (quarter)	Received flu vaccine in 2019-2020 season?	2019-2020 flu vaccine date (quarter)	Received flu vaccine in 2018-2019 season?
34C	45	Apr-June 2020	Yes	Oct-Dec 2019	Yes
199C	44	Apr-June 2020	Yes	Oct-Dec 2019	Yes
197C	42	Jul-Sep 2020	No	N/A	Yes
18C	43	Jul-Sep 2020	Yes	Jul-Sep 2019	Yes
33C	43	Jul-Sep 2020	Yes	Oct-Dec 2019	Yes
215C	43	Jul-Sep 2020	Yes	Oct-Dec 2019	Yes
74C	42	Oct-Dec 2020	No	N/A	No
210C	45	Oct-Dec 2020	Yes	Oct-Dec 2019	Yes
150C	43	Oct-Dec 2020	Yes	Oct-Dec 2019	Yes
68C	44	Oct-Dec 2020	No	N/A	Yes

Table 3.2: Serum metadata for adult cohort.

Age, serum collection date, and recent vaccination history for individuals in the 40-45 year age cohort from Seattle, Washington.

3.6 Methods

3.6.1 Biosafety

All mutant viruses in the deep mutational scanning studies derived their HAs from one of two recent human H3N2 vaccine strains (A/Perth/16/2009 or A/Hong Kong/45/2019) and the remaining seven genes from the lab-adapted A/WSN/1933 (H1N1) strain. These viruses are therefore classified as biosafety-level 2, and the work was approved by the Fred Hutchinson Cancer Center biosafety committee.

3.6.2 Human sera

The human sera used for mapping escape from A/Hong Kong/45/2019 were collected in Seattle, Washington in 2020. The sera from the 2-month-old infant, the 2-5-year age cohort, and the 15-20-year age cohort are residual sera collected at Seattle Children's Hospital with approval from the Human Subjects Institutional Review Board. Sera from the 40-45-year age cohort were collected as part of the prospective longitudinal Hospitalized or Ambulatory Adults with Respiratory Viral Infections (HAARVI) cohort study of individuals with SARS-CoV-2 infection. Written informed consent was obtained for each participant. Serum from the 68-year-old individual is deidentified residual serum from the University of Washington Virology Lab that was collected for testing for HBsAb (for which they tested negative), under a study approved by the University of Washington IRB with a consent waiver.

The human sera used for mapping escape from A/Perth/16/2009 were obtained from a serum bank established in 2010 in central and southern Vietnam for the purposes of calculating disease burden, pathogen exposure, and vaccine coverage for a range of pathogens (Berto et al., 2018; Boni et al., 2013; Lam et al., 2019; Nhat et al., 2017; Quan et al., 2018; Thwaites et al., 2023; Vinh et al., 2021). Anonymized samples were collected, as residual serum, from the biochemistry and haematology labs of the Hospital for Tropical Diseases in Ho Chi Minh City, Vietnam, during 2010 and 2011. The study was approved by the Scientific and Ethical Committee of the Hospital for Tropical Diseases and the Oxford Tropical Research Ethics Committee at the University of Oxford.

3.6.3 A/Perth/16/2009 and A/Hong Kong/45/2019 library overview

The analysis in this paper uses two distinct libraries: a non-barcoded library in the background of A/Perth/16/2009, and a barcoded library in the background of A/Hong Kong/45/2019. Because the A/Perth/16/2009 library is non-barcoded, these selection experiments use a barcoded-subamplicon sequencing approach, where the HA gene is amplified as a collection of short, tiled fragments for Illumina sequencing (see (Doud and Bloom, 2016) and https://jbloomlab.github.io/dms_tools2/bcsubamp.html for more detailed descriptions of this approach). There is no linkage information between these fragments, so escape scores are calculated as the enrichment of a given mutation in the serum selection relative to the mock selection. These libraries are therefore designed to be mostly single-mutant variants.

In contrast, the incorporation of barcodes in the A/Hong Kong/45/2019 libraries allows for more straightforward sequencing, and the ability to resolve multiple mutations in a single variant. These libraries were therefore designed to have an average of 2-3 mutations per variant. Escape scores are calculated as the enrichment of a given *variant* in a selection experiment (as measured by sequencing counts for each unique barcode), rather than enrichment of each single mutation.

The A/Perth/16/2009 experiments use the exact library stocks generated in (J. M. Lee et al., 2019). Serum selection experiments and data analysis using the A/Perth/16/2009 library were carried out as described in the methods of (J. M. Lee et al., 2019). The following methods sections all pertain to analyses using the A/Hong Kong/45/2019 barcoded libraries.

3.6.4 Experimental methods

For details on the design and generation of all A/Hong Kong/45/2019 library components, serum preparation and screening, library serum selection assays, sample preparation and sequencing, and GFP neutralization assays, see **chapter 2.5.**

3.6.5 Computational pipeline overview

For analysis of A/Hong Kong/45/2019 deep mutational scanning data, we used the modular pipeline https://github.com/dms-vep/dms-vep-pipeline (version 2.6.5). The repository https://github.com/dms-vep/flu_h3_hk19_dms contains all analyses performed in this paper, including the main *dms-vep-pipeline* and all scripts, notebooks, and settings necessary to recreate the analysis. Full analysis of A/Perth/16/2009 deep mutational scanning data can be found at https://github.com/jbloomlab/map_flu_serum_Vietnam_H3_Perth2009 and does not use *dms-vep-pipeline* for analysis. However, we have imported the finalized escape scores and validation neutralization data to the main https://github.com/dms-vep/flu_h3_hk19_dms repository, and integrated these results into the pipeline. See https://dms-vep.github.io/flu_h3_hk19_dms/ for HTML rendering of key analyses and interactive plots. These pages are the best way to explore the analyses and interactive plots of the results described in this paper.

3.6.6 Estimating escape scores for circulating H3N2 strains

The GitHub repository at https://github.com/matsengrp/seasonal-flu-dmsa has the code we used to perform the analyses in this section, as well as files with key results. We analyzed a set of 1,478 circulating H3N2 strains obtained from a publicly available Nextstrain (Hadfield et al., 2018) phylogenetic tree that samples strains from the past 12 years (https://nextstrain.org/flu/seasonal/h3n2/ha/12y). This tree is regularly updated. For this paper, we analyzed the strains from the tree that was available on October 19th, 2023. For each strain, we obtained the corresponding HA gene nucleotide sequence and associated metadata from the GISAID database (https://gisaid.org/) (Shu and McCauley, 2017). The above repository provides the GISAID accession number for each strain's HA sequence and acknowledges contributing labs.

To estimate an escape score for each strain relative to each serum sample, we ran the above sequences and metadata, including the sequence and metadata for the HA sequence of the A/Hong Kong/45/2019 strain used in the DMS experiment, through a version of the core *seasonal-flu* Nextstrain pipeline (https://github.com/nextstrain/seasonal-flu/) that we modified to compute escape scores. This pipeline generates a multiple-sequence alignment of translated HA protein sequences. For each strain in the alignment, it then makes a list of all amino-acid mutations that are needed to convert the A/Hong Kong/45/2019 sequence to the sequence of the given strain. Then, to compute that strain's escape score for a given serum sample, it simply sums the serum-specific DMS-measured escape scores of all mutations in the above list, such that escape scores for the A/Hong Kong/45/2019 DMS strain always equals zero. We only summed escape scores for mutations that were seen in both libraries and in at least three distinct variants (averaged across libraries) in the DMS data for that serum sample. If a strain's list of mutations included one or more mutations that were not measured by DMS or did not meet the above criteria, we masked that strain in the analysis, unless the mutations in question occurred at sites that we deemed unlikely to affect escape (see the repository for more detail). After the masking step, this process yielded a total of 1,233-1,238 strains with estimated escape scores for each serum sample.

Figure 3.5B shows the distribution of strain-specific escape scores as a function of strain collection date, with scores averaged across all sera in a given cohort. For each cohort, we fit a linear-regression model that predicts a strain's average escape score as a function of its collection date and reported the inferred slope of the resulting best-fit line in the figure caption. Next, we used a randomization approach to determine if these slopes were significantly different between cohorts. In each of 1,000 independent trials, we randomized which serum sample was assigned to which cohort, recomputed average strain-specific escape scores using the randomized cohort assignments, and then fit a linear-regression model to the resulting data, fitting one model per cohort as above. We then compared the slopes obtained by randomization to the slopes observed in the non-randomized data. In **Figure 3.5B**, the slope associated with teenagers is 0.1 units greater than the slope associated with adults (=0.14-0.04). Out of the 1,000 randomization trials, only four trials resulted in a difference in slope (=teenagers-adults) greater than or equal to 0.1, indicating that the slope for teenagers is significantly larger than the slope for adults in the non-randomized data (p=0.004). The randomization test also indicated that the slope for teenagers is significantly larger than the slope for children (p=0.022), though the slope for children was not significantly larger than the slope for adults (p=0.323).

3.7 Acknowledgments

We thank Sarah Cobey for scientific advice. We gratefully acknowledge all data contributors, including the authors and their originating laboratories responsible for obtaining the specimens, and their submitting laboratories for generating the genetic sequence and metadata and sharing via the GISAID Initiative, on which part of this research is based. This work was supported in part by the NIH/NIAID award R01AI165821 to TB and JDB, contract 75N93021C00015 to JDB and TB, and award R01AI146028 to FAM. Serum collection in Vietnam was funded by a Wellcome Trust/Royal Society Sir Henry Dale Fellowship (098511/Z/12/Z) to MFB. TCY was supported by the CMB training grant T32 GM007270 and the NSF graduate research fellowship DGE-2140004. JDB, TB, and FAM are Investigators of the Howard Hughes Medical Institute. This research was supported by the Genomics & Bioinformatics Shared Resource (RRID:SCR_022606) of the Fred Hutch/University of Washington/Seattle Children's Cancer Consortium (P30CA015704), and by Fred Hutch Scientific Computing (NIH grants S10-OD-020069 and S10-OD-028685).

CHAPTER 4: Conclusion

In my dissertation, I developed a barcoded deep mutational scanning system for influenza HA, and demonstrated its utility for high-throughput analysis of serum antibody targeting. By mapping the antigenic effects of HA mutations against sera from different age cohorts, I showed that influenza faces distinct antigenic selection regimes from different subsets of the population, suggesting that certain cohorts may be the primary drivers of influenza evolution. Here, I summarize the main findings of my dissertation and suggest directions for future work.

4.1 Future directions for barcoded influenza libraries

4.1.1 Generate an updated library in the background of a contemporary H3 HA strain

I have described the development of barcoded mutagenic libraries in the background of A/Hong Kong/45/2019, which was the H3N2 component of the 2020-2021 influenza vaccine. Influenza circulation was at historic lows during the first year of the SARS-CoV-2 pandemic, with a 98% decrease in activity starting spring 2020 (Olsen et al., 2020). This did not limit the availability of sera with high titers against H3 HA, as A/Hong Kong/45/2019 was circulating prior to SARS-CoV-2, and we primarily analyzed serum samples collected from April to June 2020. However, differences in social distancing measures implemented for children and adults may have influenced which age groups were stronger drivers of influenza evolution in subsequent years.

Furthermore, the wildtype library strain carries an N-linked glycan at site 158 that sterically shields the immunodominant antigenic region B. This glycan was introduced in 2014, and was somewhat surprisingly lost in 2021 (Bolton et al., 2022). Adult sera collected from the Vietnam cohort in 2010-2011 are much more heterogeneous than the

2020 cohort, with each individual targeting a different site in antigenic region B. This raises the question of whether the relative homogeneity within the 2020 adult cohort is due to the 158 glycan, as 189 is one of the few remaining sites in region B still accessible to neutralizing antibodies. Alternatively, the apparent heterogeneity of the 2010-2011 cohort may be an artifact of the small sample size (3 sera) compared to 2020 (10 sera).

Generating an updated library in the background of a contemporary H3N2 strain that has lost the 158 glycan, and analyzing within-cohort heterogeneity, would therefore be highly informative for understanding how antigenic and structural changes in HA might shape heterogeneity in selective pressures.

4.1.2 Analyze serum antibody responses to vaccination or infection

Prior work has shown that vaccination can either shift the specificity of serum antibodies, or primarily boost existing specificities (Fonville et al., 2014; Henry et al., 2019; J. Lee et al., 2019; Lee et al., 2016; J. M. Lee et al., 2019). An effective flu vaccine would shift serum specificity to induce high antibody titers against contemporary strains, regardless of exposure history, but the best strategy for achieving this goal remains unclear. Lee et al. (2019) showed that the effect of vaccination on immune specificity is heterogeneous between individuals, but this study was limited to four sera from different ages.

The results reported in Chapter 3 illustrate how immune specificity is relatively homogenous within an age cohort with similar exposure histories. If the antibody response to vaccination is mainly a factor of exposure history, then individuals within an age cohort should have relatively homogenous responses. Mapping antibody targeting from longitudinal pre- and post-vaccination samples, collected from defined age cohorts, will shed light on the extent to which exposure history shapes the vaccine response.

The antibody response to natural infection is broader and more long-lived than the antibody response to vaccination, which can sometimes wane even within a season (Belongia et al., 2015; Fisman et al., 2009; Hancock et al., 2009; Kissling et al., 2013; Margine et al., 2013; Petrie et al., 2015; Puig-Barberà et al., 2017; Yu et al., 2008). Better understanding the cause of these different responses is crucial for improving vaccine efficacy. Mapping the impact of infection on immune specificity using longitudinal serum samples, and comparing to the impact of vaccination, could help identify antibody epitopes that contribute to long-lasting protection.

4.1.3 Develop libraries for generation of multiplexed neutralization data

Deep mutational scanning of influenza has thus far been limited to variant libraries assaying mutation effects in a single background strain. The development of a barcoded HA construct makes it possible to assay full-length variants with multiple mutations. Therefore, in addition to mutagenic libraries, we can now also generate libraries carrying many strains of interest. By simply running selections at a wider range of serum or antibody potencies, this deep mutational scanning approach can be used to measure full neutralization curves for each strain in the library.

The multiplexed neutralization assay has been successfully piloted for a library of circulating H1 HA strains (Loes, unpublished). This has significant value for surveillance efforts, as it allows for rapid assessment of the antigenicity of hundreds of circulating strains, and measures neutralization rather than HI titers. We can also use this system to generate a library of historic strains spanning several decades of influenza evolution. While the rate of antigenic drift over time is of great interest to the influenza field, it has only ever been measured using HI assays against primary ferret antisera, which tend to be much more narrowly targeted than human sera (Fonville et al., 2016, 2014; Li et al., 2013; Smith et al., 2004). These studies have also described asymmetry in influenza evolution, where serum exhibits more cross-reactivity with previously circulating strains than strains emerging in the future (Sandbulte et al., 2011; Underwood, 1980). Measuring human serum neutralization against a library of strains would allow for systematic assessment of the rate at which influenza evolution erodes neutralizing titers. By analyzing strains from before and after serum collection, we would also be able to characterize the degree of asymmetry in antigenic drift.

4.2 Assessment of population-level heterogeneity for improved surveillance and evolutionary forecasting

In Chapter 3, I described how different age groups exhibit distinct patterns of serum antibody targeting. Comparison of these selective pressures with subsequent antigenic drift indicate that younger individuals may be stronger drivers of influenza evolution. These results suggest that vaccine strain selection could be significantly improved by targeting well-defined ‘driver’ groups in surveillance and evolutionary forecasting.

Thus far, our analysis has been limited to 8-10 individuals from just three age cohorts. While this was sufficient to demonstrate age-dependent patterns of serum antibody targeting, a broader sample size is needed to draw any firm conclusions about how this heterogeneity impacts influenza evolution. Kim et al. (2023) analyzed neutralizing titers from almost 500 individuals against a panel of 8 reference strains and also identified patterns of age-dependent heterogeneity. Similar to our deep mutational scanning results, they found that the neutralization titers of the 5-17-year-old cohort were highly predictive of the dominant clade in the following season. Simple neutralization assays facilitate high-throughput analysis of serum samples, while deep mutational scanning is lower-throughput, but precisely identifies the epitopes targeted by each serum. I therefore propose a two-pronged approach for more rigorous characterization of population-level immunity, where we measure neutralizing titers for a large set of sera spanning all age cohorts, and select a subset from each cohort for epitope mapping. Using these techniques in conjunction, and comparing the results to subsequent evolution, would more definitively identify the cohorts driving antigenic drift.

The epitopes identified by deep mutational scanning can also be integrated into evolutionary forecasting models to improve predictions. Luksza and Lässig (2014) developed a model based on epitope antigenic novelty, where strain fitness is determined based on mutations at epitope sites and mutational load at non-epitope sites. However, this model is inherently limited by the accuracy of the defined epitopes, as antigenic sites are highly dependent on strain background and population immunity in a given season (Huddleston et al., 2020). Prior epitope definitions relied on computational analyses or aggregated escape data (Luksza and Lässig, 2014; Shih et al., 2007; Wolf et al., 2006), which overfit to historical context.

Here, I have experimentally identified sites that confer antigenic escape from a relatively broad cohort with diverse exposure histories. Initial analysis shows that the number of mutations at putative epitope sites (defined as any site above a minimum threshold of escape for any serum) seem to be well correlated with clade success since 2021 (J Huddleston, personal communication, December 2023). An updated epitope-based model has yet to be formally tested, and it will be informative to determine whether assigning higher weights to epitope sites from child and teenage cohorts improves model performance. It is also uncertain how frequently these epitope sites should be updated to remain informative. This could be roughly determined by generating a new library in the background of a more recent H3N2 strain, defining epitope sites using updated serum antibody mapping data, and analyzing prediction accuracy over time for both sets of epitopes.

If antigenic drift is indeed driven primarily by immune selection from school-aged children, this suggests that population-level immunity may be achieved more efficiently by focusing vaccination efforts on this cohort, rather than elderly individuals. Programs for vaccinating children have previously been implemented in Japan and the UK, with promising results. Japanese schoolchildren were legally required to be vaccinated from 1977 until 1987. A retrospective analysis found that excess death rates during winter influenza seasons declined by over 50% during this period, then rose steeply again after the policy was lifted, most likely due to the effects of vaccine-induced herd immunity (Reichert et al., 2001). In the UK, vaccination was extended from high-risk individuals to also include schoolchildren, based on results from a transmission modeling study (Baguelin et al., 2013). This policy led to a 60% reduction in physician visits by adults and the elderly for influenza-like illnesses (Kassianos et al., 2020). There are substantial ethical considerations to take into account before attempting a similar approach in the US, as it would primarily protect high-risk populations and not necessarily the healthy individuals receiving the vaccines. Nevertheless, these results warrant further studies evaluating how vaccine effectiveness might be improved by shifting target populations.

Our findings demonstrate that effective strategies for controlling influenza transmission must account for the heterogeneous effects of exposure history on antigenicity. More rigorous characterization of these effects may substantially improve both evolutionary forecasting and vaccine effectiveness.

REFERENCES

Arevalo CP, Le Sage V, Bolton MJ, Eilola T, Jones JE, Kormuth KA, Nturibi E, Balmaseda A, Gordon A, Lakdawala SS, Hensley SE. 2020. Original antigenic sin priming of influenza virus hemagglutinin stalk antibodies. *Proc Natl Acad Sci U S A* **117**:17221–17227.

Baccam P, Beauchemin C, Macken CA, Hayden FG, Perelson AS. 2006. Kinetics of influenza A virus infection in humans. *J Virol* **80**:7590–7599.

Bacsik DJ, Dadonaite B, Butler A, Greaney AJ, Heaton NS, Bloom JD. 2023. Influenza virus transcription and progeny production are poorly correlated in single cells. *Elife* **12**. doi:10.7554/eLife.86852

Baguelin M, Flasche S, Camacho A, Demiris N, Miller E, Edmunds WJ. 2013. Assessing optimal target populations for influenza vaccination programmes: an evidence synthesis and modelling study. *PLoS Med* **10**:e1001527.

Barr IG, Russell C, Besselaar TG, Cox NJ, Daniels RS, Donis R, Engelhardt OG, Grohmann G, Itamura S, Kelso A, McCauley J, Odagiri T, Schultz-Cherry S, Shu Y, Smith D, Tashiro M, Wang D, Webby R, Xu X, Ye Z, Zhang W, Writing Committee of the World Health Organization Consultation on Northern Hemisphere Influenza Vaccine Composition for 2013–2014. 2014. WHO recommendations for the viruses used in the 2013-2014 Northern Hemisphere influenza vaccine: Epidemiology, antigenic and genetic characteristics of influenza A(H1N1)pdm09, A(H3N2) and B influenza viruses collected from October 2012 to January 2013. *Vaccine* **32**:4713–4725.

Bedford T, Riley S, Barr IG, Broor S, Chadha M, Cox NJ, Daniels RS, Gunasekaran CP, Hurt AC, Kelso A, Klimov A, Lewis NS, Li X, McCauley JW, Odagiri T, Potdar V, Rambaut A, Shu Y, Skepner E, Smith DJ, Suchard MA, Tashiro M, Wang D, Xu X, Lemey P, Russell CA. 2015. Global circulation patterns of seasonal influenza viruses vary with antigenic drift. *Nature* **523**:217–220.

Bedford T, Suchard MA, Lemey P, Dudas G, Gregory V, Hay AJ, McCauley JW, Russell CA, Smith DJ, Rambaut A. 2014. Integrating influenza antigenic dynamics with molecular evolution. *Elife* **3**:e01914.

Belongia EA, Sundaram ME, McClure DL, Meece JK, Ferdinands J, VanWormer JJ. 2015. Waning vaccine protection against influenza A (H3N2) illness in children and older adults during a single season. *Vaccine* **33**:246–251.

Berto A, Pham HA, Thao TTN, Vy NHT, Caddy SL, Hiraide R, Tue NT, Goodfellow I, Carrique-Mas JJ, Thwaites GE, Baker S, Boni MF, VIZIONS consortium. 2018. Hepatitis E in southern Vietnam: Seroepidemiology in humans and molecular epidemiology in pigs. *Zoonoses Public Health* **65**:43–50.

Bloom JD. 2014. An experimentally determined evolutionary model dramatically improves phylogenetic fit. *Mol Biol Evol* **31**:1956–1978.

Bloom JD, Gong LI, Baltimore D. 2010. Permissive secondary mutations enable the evolution of influenza oseltamivir resistance. *Science* **328**:1272–1275.

Bolton MJ, Ort JT, McBride R, Swanson NJ, Wilson J, Awofolaju M, Furey C, Greenplate AR, Drapeau EM, Pekosz A, Paulson JC, Hensley SE. 2022. Antigenic and virological properties of an H3N2 variant that continues to dominate the 2021-22 Northern Hemisphere influenza season. *Cell Rep* **39**:110897.

Boni MF, Chau NVV, Dong N, Todd S, Nhat NTD, de Bruin E, van Beek J, Hien NT, Simmons CP, Farrar J, Koopmans M. 2013. Population-level antibody estimates to novel influenza A/H7N9. *J Infect Dis* **208**:554–558.

Brandenburg B, Koudstaal W, Goudsmit J, Klaren V, Tang C, Bujny MV, Korse HJWM, Kwaks T, Otterstrom JJ, Juraszek J, van Oijen AM, Vogels R, Friesen RHE. 2013. Mechanisms of hemagglutinin targeted influenza virus neutralization. *PLoS One* **8**:e80034.

Caton AJ, Brownlee GG, Yewdell JW, Gerhard W. 1982. The antigenic structure of the influenza virus A/PR/8/34 hemagglutinin (H1 subtype). *Cell* **31**:417–427.

Chen Y-Q, Wohlbold TJ, Zheng N-Y, Huang M, Huang Y, Neu KE, Lee J, Wan H, Rojas KT, Kirkpatrick E, Henry C, Palm A-KE, Stamper CT, Lan LY-L, Topham DJ, Treanor J, Wrammert J, Ahmed R, Eichelberger MC, Georgiou G, Krammer F, Wilson PC. 2018. Influenza Infection in Humans Induces Broadly Cross-Reactive and Protective Neuraminidase-Reactive Antibodies. *Cell* **173**:417–429.e10.

Cobey S, Hensley SE. 2017. Immune history and influenza virus susceptibility. *Curr Opin Virol* **22**:105–111.

Crawford KHD, Bloom JD. 2019. alignparse: A Python package for parsing complex features from high-throughput long-read sequencing. *J Open Source Softw* **4**. doi:10.21105/joss.01915

Dadonaite B, Crawford KHD, Radford CE, Farrell AG, Yu TC, Hannon WW, Zhou P, Andrabi R, Burton DR, Liu L, Ho DD, Chu HY, Neher RA, Bloom JD. 2023. A pseudovirus system enables deep mutational scanning of the full SARS-CoV-2 spike. *Cell* **186**:1263–1278.e20.

Davenport FM, Hennessy AV. 1956. A serologic recapitulation of past experiences with influenza A; antibody response to monovalent vaccine. *J Exp Med* **104**:85–97.

Davenport FM, Hennessy AV, Francis T Jr. 1953. Epidemiologic and immunologic significance of age distribution of antibody to antigenic variants of influenza virus. *J Exp Med* **98**:641–656.

Davis AKF, McCormick K, Gumina ME, Petrie JG, Martin ET, Xue KS, Bloom JD, Monto AS, Bushman FD, Hensley SE. 2018. Sera from Individuals with Narrowly Focused Influenza Virus Antibodies Rapidly Select Viral Escape Mutations In Ovo. *J Virol* **92**. doi:10.1128/JVI.00859-18

Doud MB, Bloom JD. 2016. Accurate Measurement of the Effects of All Amino-Acid Mutations on Influenza Hemagglutinin. *Viruses* **8**. doi:10.3390/v8060155

Doud MB, Hensley SE, Bloom JD. 2017. Complete mapping of viral escape from neutralizing antibodies. *PLoS Pathog* **13**:e1006271.

Doud MB, Lee JM, Bloom JD. 2018. How single mutations affect viral escape from broad and narrow antibodies to H1 influenza hemagglutinin. *Nat Commun* **9**:1386.

Ekiert DC, Bhabha G, Elsliger M-A, Friesen RHE, Jongeneelen M, Throsby M, Goudsmit J, Wilson IA. 2009. Antibody recognition of a highly conserved influenza virus epitope. *Science* **324**:246–251.

Ellebedy AH, Jackson KJL, Kissick HT, Nakaya HI, Davis CW, Roskin KM, McElroy AK, Oshansky CM, Elbein R, Thomas S, Lyon GM, Spiropoulou CF, Mehta AK, Thomas PG, Boyd SD, Ahmed R. 2016. Defining antigen-specific plasmablast and memory B cell subsets in human blood after viral infection or vaccination. *Nat Immunol* **17**:1226–1234.

Fisman DN, Savage R, Gubbay J, Achonu C, Akwar H, Farrell DJ, Crowcroft NS, Jackson P. 2009. Older age and a reduced likelihood of 2009 H1N1 virus infection. *N Engl J Med* **361**:2000–2001.

Fitch WM, Bush RM, Bender CA, Cox NJ. 1997. Long term trends in the evolution of H(3) HA1 human influenza type A. *Proc Natl Acad Sci U S A* **94**:7712–7718.

Fonville JM, Fraaij PLA, de Mutsert G, Wilks SH, van Beek R, Fouchier RAM, Rimmelzwaan GF. 2016. Antigenic Maps of Influenza A(H3N2) Produced With Human Antisera Obtained After Primary Infection. *J Infect Dis* **213**:31–38.

Fonville JM, Wilks SH, James SL, Fox A, Ventresca M, Aban M, Xue L, Jones TC, Le NMH, Pham QT, Tran ND, Wong Y, Mosterin A, Katzelnick LC, Labonte D, Le TT, van der Net G, Skepner E, Russell CA, Kaplan TD, Rimmelzwaan GF, Masurel N, de Jong JC, Palache A, Beyer WEP, Le QM, Nguyen TH, Wertheim HFL, Hurt AC, Osterhaus ADME, Barr IG, Fouchier RAM, Horby PW, Smith DJ. 2014. Antibody landscapes after influenza virus infection or vaccination. *Science* **346**:996–1000.

Francis T. 1960. On the Doctrine of Original Antigenic Sin. *Proc Am Philos Soc* **104**:572–578.

Gao Q, Palese P. 2009. Rewiring the RNAs of influenza virus to prevent reassortment. *Proc Natl Acad Sci U S A* **106**:15891–15896.

Gerhard W, Yewdell J, Frankel ME, Webster R. 1981. Antigenic structure of influenza virus haemagglutinin defined by hybridoma antibodies. *Nature* **290**:713–717.

Ghedin E, Fitch A, Boyne A, Griesemer S, DePasse J, Bera J, Zhang X, Halpin RA, Smit M, Jennings L, St George K, Holmes EC, Spiro DJ. 2009. Mixed infection and the genesis of influenza virus diversity. *J Virol* **83**:8832–8841.

Gouma S, Kim K, Weirick ME, Gumina ME, Branche A, Topham DJ, Martin ET, Monto AS, Cobey S, Hensley SE. 2020. Middle-aged individuals may be in a perpetual state of H3N2 influenza virus susceptibility. *Nat Commun* **11**:4566.

Greaney AJ, Starr TN, Gilchuk P, Zost SJ, Binshtein E, Loes AN, Hilton SK, Huddleston J, Eguia R, Crawford KHD, Dingens AS, Nargi RS, Sutton RE, Suryadevara N, Rothlauf PW, Liu Z, Whelan SPJ, Carnahan RH, Crowe JE Jr, Bloom JD. 2021. Complete Mapping of Mutations to the SARS-CoV-2 Spike Receptor-Binding Domain that Escape Antibody Recognition. *Cell Host Microbe* **29**:44–57.e9.

Hadfield J, Megill C, Bell SM, Huddleston J, Potter B, Callender C, Sagulenko P, Bedford T, Neher RA. 2018. Nextstrain: real-time tracking of pathogen evolution. *Bioinformatics* **34**:4121–4123.

Han AX, de Jong SPJ, Russell CA. 2023. Co-evolution of immunity and seasonal influenza viruses. *Nat Rev Microbiol* **21**:805–817.

Hancock K, Veguilla V, Lu X, Zhong W, Butler EN, Sun H, Liu F, Dong L, DeVos JR, Gargiullo PM, Brammer TL, Cox NJ, Tumpey TM, Katz JM. 2009. Cross-reactive antibody responses to the 2009 pandemic H1N1 influenza virus. *N Engl J Med* **361**:1945–1952.

Han J, Schmitz AJ, Richey ST, Dai Y-N, Turner HL, Mohammed BM, Fremont DH, Ellebedy AH, Ward AB. 2021. Polyclonal epitope mapping reveals temporal dynamics and diversity of human antibody responses to H5N1 vaccination. *Cell Rep* **34**:108682.

Henry C, Zheng N-Y, Huang M, Cabanov A, Rojas KT, Kaur K, Andrews SF, Palm A-KE, Chen Y-Q, Li Y, Hoskova K, Utset HA, Vieira MC, Wrammert J, Ahmed R, Holden-Wiltse J, Topham DJ, Treanor JJ, Ertl HC, Schmader KE, Cobey S, Krammer F, Hensley SE, Greenberg H, He X-S, Wilson PC. 2019. Influenza Virus Vaccination Elicits Poorly Adapted B Cell Responses in Elderly Individuals. *Cell Host Microbe* **25**:357–366.e6.

Hensley SE, Das SR, Bailey AL, Schmidt LM, Hickman HD, Jayaraman A, Viswanathan K, Raman R, Sasisekharan R, Bennink JR, Yewdell JW. 2009. Hemagglutinin receptor binding avidity drives influenza A virus antigenic drift. *Science* **326**:734–736.

Hoffmann E, Neumann G, Kawaoka Y, Hobom G, Webster RG. 2000. A DNA transfection system for generation of influenza A virus from eight plasmids. *Proc Natl Acad Sci U S A* **97**:6108–6113.

Hooper KA, Bloom JD. 2013. A mutant influenza virus that uses an N1 neuraminidase as the receptor-binding protein. *J Virol* **87**:12531–12540.

Huang K-YA, Rijal P, Schimanski L, Powell TJ, Lin T-Y, McCauley JW, Daniels RS, Townsend AR. 2015. Focused antibody response to influenza linked to antigenic drift. *J Clin Invest* **125**:2631–2645.

Huddleston J, Barnes JR, Rowe T, Xu X, Kondor R, Wentworth DE, Whittaker L, Ermetal B, Daniels RS, McCauley JW, Fujisaki S, Nakamura K, Kishida N, Watanabe S, Hasegawa H, Barr I, Subbarao K, Barrat-Charlaix P, Neher RA, Bedford T. 2020. Integrating genotypes and phenotypes improves long-term forecasts of seasonal influenza A/H3N2 evolution. *Elife* **9**. doi:10.7554/eLife.60067

Islam S, Mohn KG-I, Krammer F, Sanne M, Bredholt G, Jul-Larsen Å, Tete SM, Zhou F, Brokstad KA, Cox RJ. 2017. Influenza A haemagglutinin specific IgG responses in children and adults after seasonal trivalent live attenuated influenza vaccination. *Vaccine* **35**:5666–5673.

Kassianos G, MacDonald P, Aloysius I, Reynolds A. 2020. Implementation of the United Kingdom’s childhood influenza national vaccination programme: A review of clinical impact and lessons learned over six influenza seasons. *Vaccine* **38**:5747–5758.

Kilbourne ED, Smith C, Brett I, Pokorny BA, Johansson B, Cox N. 2002. The total influenza vaccine failure of 1947 revisited: major intrasubtypic antigenic change can explain failure of vaccine in a post-World War II epidemic. *Proc Natl Acad Sci U S A* **99**:10748–10752.

Kissling E, Valenciano M, Larrauri A, Oroszi B, Cohen JM, Nunes B, Pitigoi D, Rizzo C, Rebolledo J, Paradowska-Stankiewicz I, Jiménez-Jorge S, Horváth JK, Daviaud I, Guiomar R, Necula G, Bella A, O’Donnell J, Głuchowska M, Ciancio BC, Nicoll A, Moren A. 2013. Low and decreasing vaccine effectiveness against influenza A(H3) in 2011/12 among vaccination target groups in Europe: results from the I-MOVE multicentre case-control study. *Euro Surveill* **18**. doi:10.2807/ese.18.05.20390-en

Kitphati R, Pooruk P, Lerdsamran H, Poosuwan S, Louisirirotchanakul S, Auewarakul P, Chokphaibulkit K, Noisumdaeng P, Sawanpanyalert P, Puthavathana P. 2009. Kinetics and longevity of antibody response to influenza A H5N1 virus infection in humans. *Clin Vaccine Immunol* **16**:978–981.

Koel BF, Burke DF, Bestebroer TM, van der Vliet S, Zondag GCM, Vervaet G, Skepner E, Lewis NS, Spronken MIJ, Russell CA, Eropkin MY, Hurt AC, Barr IG, de Jong JC, Rimmelzwaan GF, Osterhaus ADME, Fouchier RAM, Smith DJ. 2013. Substitutions near the receptor binding site determine major antigenic change during influenza virus evolution. *Science* **342**:976–979.

Krammer F. 2019. The human antibody response to influenza A virus infection and vaccination. *Nat Rev Immunol* **19**:383–397.

Kucharski AJ, Lessler J, Read JM, Zhu H, Jiang CQ, Guan Y, Cummings DAT, Riley S. 2015. Estimating the life course of influenza A(H3N2) antibody responses from cross-sectional data. *PLoS Biol* **13**:e1002082.

Lam HM, Phuong HT, Thao Vy NH, Le Thanh NT, Dung PN, Ngoc Muon TT, Van Vinh Chau N, Rodríguez-Barraquer I, Cummings DAT, Wills BA, Boni MF, Rabaa MA, Clapham HE. 2019. Serological inference of past primary and secondary dengue infection: implications for vaccination. *J R Soc Interface* **16**:20190207.

Lam JH, Baumgarth N. 2019. The Multifaceted B Cell Response to Influenza Virus. *J Immunol* **202**:351–359.

Lau D, Lan LY-L, Andrews SF, Henry C, Rojas KT, Neu KE, Huang M, Huang Y, DeKosky B, Palm A-KE, Ippolito GC, Georgiou G, Wilson PC. 2017. Low CD21 expression defines a population of recent germinal center graduates primed for plasma cell differentiation. *Sci Immunol* **2**. doi:10.1126/sciimmunol.aai8153

Laver WG, Air GM, Webster RG, Gerhard W, Ward CW, Dopheide TA. 1979. Antigenic drift in type A influenza virus: sequence differences in the hemagglutinin of Hong Kong (H3N2) variants selected with monoclonal hybridoma antibodies. *Virology* **98**:226–237.

Lee J, Boutz DR, Chromikova V, Joyce MG, Vollmers C, Leung K, Horton AP, DeKosky BJ, Lee C-H, Lavinder JJ, Murrin EM, Chrysostomou C, Hoi KH, Tsybovsky Y, Thomas PV, Druz A, Zhang B, Zhang Y, Wang L, Kong W-P, Park D, Popova LI, Dekker CL, Davis MM, Carter CE, Ross TM, Ellington AD, Wilson PC, Marcotte EM, Mascola JR, Ippolito GC, Krammer F, Quake SR, Kwong PD, Georgiou G. 2016. Molecular-level analysis of the serum antibody repertoire in young adults before and after seasonal influenza vaccination. *Nat Med* **22**:1456–1464.

Lee JM, Eguia R, Zost SJ, Choudhary S, Wilson PC, Bedford T, Stevens-Ayers T, Boeckh M, Hurt AC, Lakdawala SS, Hensley SE, Bloom JD. 2019. Mapping person-to-person variation in viral mutations that escape polyclonal serum targeting influenza hemagglutinin. *Elife* **8**. doi:10.7554/eLife.49324

Lee JM, Huddleston J, Doud MB, Hooper KA, Wu NC, Bedford T, Bloom JD. 2018. Deep mutational scanning of hemagglutinin helps predict evolutionary fates of human H3N2 influenza variants. *Proc Natl Acad Sci U S A* **115**:E8276–E8285.

Lee J, Paparoditis P, Horton AP, Frühwirth A, McDaniel JR, Jung J, Boutz DR, Hussein DA, Tanno Y, Pappas L, Ippolito GC, Corti D, Lanzavecchia A, Georgiou G. 2019. Persistent Antibody Clonotypes Dominate the Serum Response to Influenza over Multiple Years and Repeated Vaccinations. *Cell Host Microbe* **25**:367–376.e5.

Lessler J, Riley S, Read JM, Wang S, Zhu H, Smith GJD, Guan Y, Jiang CQ, Cummings DAT.

2012. Evidence for antigenic seniority in influenza A (H3N2) antibody responses in southern China. *PLoS Pathog* **8**:e1002802.

Linderman SL, Chambers BS, Zost SJ, Parkhouse K, Li Y, Herrmann C, Ellebedy AH, Carter DM, Andrews SF, Zheng N-Y, Huang M, Huang Y, Strauss D, Shaz BH, Hodinka RL, Reyes-Terán G, Ross TM, Wilson PC, Ahmed R, Bloom JD, Hensley SE. 2014. Potential antigenic explanation for atypical H1N1 infections among middle-aged adults during the 2013-2014 influenza season. *Proc Natl Acad Sci U S A* **111**:15798–15803.

Li X, Gu M, Zheng Q, Gao R, Liu X. 2021. Packaging signal of influenza A virus. *Virol J* **18**:36.

Li Y, Myers JL, Bostick DL, Sullivan CB, Madara J, Linderman SL, Liu Q, Carter DM, Wrammert J, Esposito S, Principi N, Plotkin JB, Ross TM, Ahmed R, Wilson PC, Hensley SE. 2013. Immune history shapes specificity of pandemic H1N1 influenza antibody responses. *J Exp Med* **210**:1493–1500.

Luksza M, Lässig M. 2014. A predictive fitness model for influenza. *Nature* **507**:57–61.

Margine I, Hai R, Albrecht RA, Obermoser G, Harrod AC, Banchereau J, Palucka K, García-Sastre A, Palese P, Treanor JJ, Krammer F. 2013. H3N2 influenza virus infection induces broadly reactive hemagglutinin stalk antibodies in humans and mice. *J Virol* **87**:4728–4737.

Meade P, Kuan G, Strohmeier S, Maier HE, Amanat F, Balmaseda A, Ito K, Kirkpatrick E, Javier A, Gresh L, Nachbagauer R, Gordon A, Krammer F. 2020. Influenza Virus Infection Induces a Narrow Antibody Response in Children but a Broad Recall Response in Adults. *MBio* **11**. doi:10.1128/mBio.03243-19

Monto AS, Petrie JG, Cross RT, Johnson E, Liu M, Zhong W, Levine M, Katz JM, Ohmit SE. 2015. Antibody to Influenza Virus Neuraminidase: An Independent Correlate of Protection. *J Infect Dis* **212**:1191–1199.

Muñoz ET, Deem MW. 2005. Epitope analysis for influenza vaccine design. *Vaccine* **23**:1144–1148.

Nakajima S, Nobusawa E, Nakajima K. 2000. Variation in response among individuals to antigenic sites on the HA protein of human influenza virus may be responsible for the emergence of drift strains in the human population. *Virology* **274**:220–231.

Natali A, Oxford JS, Schild GC. 1981. Frequency of naturally occurring antibody to influenza virus antigenic variants selected in vitro with monoclonal antibody. *J Hyg* **87**:185–190.

Neher RA, Bedford T. 2015. nextflu: real-time tracking of seasonal influenza virus evolution in humans. *Bioinformatics* **31**:3546–3548.

Nelson MI, Holmes EC. 2007. The evolution of epidemic influenza. *Nat Rev Genet* **8**:196–205.

Neumann G, Kawaoka Y. 2015. Transmission of influenza A viruses. *Virology* **479-480**:234–246.

Nhat NTD, Todd S, de Bruin E, Thao TTN, Vy NHT, Quan TM, Vinh DN, van Beek J, Anh PH, Lam HM, Hung NT, Thanh NTL, Huy HLA, Ha VTH, Baker S, Thwaites GE, Lien NTN, Hong TTK, Farrar J, Simmons CP, Chau NVV, Koopmans M, Boni MF. 2017. Structure of general-population antibody titer distributions to influenza A virus. *Sci Rep* **7**:6060.

Oidtman RJ, Arevalo P, Bi Q, McGough L, Russo CJ, Vera Cruz D, Costa Vieira M, Gostic KM. 2021. Influenza immune escape under heterogeneous host immune histories. *Trends Microbiol* **29**:1072–1082.

Olsen SJ, Azziz-Baumgartner E, Budd AP, Brammer L, Sullivan S, Pineda RF, Cohen C, Fry AM. 2020. Decreased influenza activity during the COVID-19 pandemic-United States, Australia, Chile, and South Africa, 2020. *Am J Transplant* **20**:3681–3685.

Otwinowski J, McCandlish DM, Plotkin JB. 2018. Inferring the shape of global epistasis. *Proc Natl Acad Sci U S A* **115**:E7550–E7558.

Petrie JG, Ohmit SE, Johnson E, Truscon R, Monto AS. 2015. Persistence of Antibodies to Influenza Hemagglutinin and Neuraminidase Following One or Two Years of Influenza Vaccination. *J Infect Dis* **212**:1914–1922.

Plotkin JB, Dushoff J, Levin SA. 2002. Hemagglutinin sequence clusters and the antigenic evolution of influenza A virus. *Proc Natl Acad Sci U S A* **99**:6263–6268.

Puig-Barberà J, Mira-Iglesias A, Tortajada-Girbés M, López-Labrador FX, Librero-López J, Díez-Domingo J, Carballido-Fernández M, Carratalá-Munuera C, Correcher-Medina P, Gil-Guillén V, Limón-Ramírez R, Mollar-Maseres J, Otero-Reigada MC, Schwarz H, Valencia Hospital Network for the Study of Influenza and other Respiratory Viruses (VAHNSI, Spain). 2017. Waning protection of influenza vaccination during four influenza seasons, 2011/2012 to 2014/2015. *Vaccine* **35**:5799–5807.

Puryear W, Sawatzki K, Hill N, Foss A, Stone JJ, Doughty L, Walk D, Gilbert K, Murray M, Cox E, Patel P, Mertz Z, Ellis S, Taylor J, Fauquier D, Smith A, DiGiovanni RA Jr, van de Guchte A, Gonzalez-Reiche AS, Khalil Z, van Bakel H, Torchetti MK, Lantz K, Lenoch JB, Runstadler J. 2023. Highly Pathogenic Avian Influenza A(H5N1) Virus Outbreak in New England Seals, United States. *Emerg Infect Dis* **29**:786–791.

Quan TM, Phuong HT, Vy NHT, Thanh NTL, Lien NTN, Hong TTK, Dung PN, Chau NVV, Boni MF, Clapham HE. 2018. Evidence of previous but not current transmission of chikungunya virus in southern and central Vietnam: Results from a systematic review and a seroprevalence study in four locations. *PLoS Negl Trop Dis*

12:e0006246.

Ranjeva S, Subramanian R, Fang VJ, Leung GM, Ip DKM, Perera RAPM, Peiris JSM, Cowling BJ, Cobey S. 2019. Age-specific differences in the dynamics of protective immunity to influenza. *Nat Commun* **10**:1660.

Reichert TA, Sugaya N, Fedson DS, Glezen WP, Simonsen L, Tashiro M. 2001. The Japanese experience with vaccinating schoolchildren against influenza. *N Engl J Med* **344**:889–896.

Sailer ZR, Harms MJ. 2017. Detecting High-Order Epistasis in Nonlinear Genotype-Phenotype Maps. *Genetics* **205**:1079–1088.

Salamango DJ, Alam KK, Burke DH, Johnson MC. 2016. In Vivo Analysis of Infectivity, Fusogenicity, and Incorporation of a Mutagenic Viral Glycoprotein Library Reveals Determinants for Virus Incorporation. *J Virol* **90**:6502–6514.

Sandbulte MR, Gao J, Straight TM, Eichelberger MC. 2009. A miniaturized assay for influenza neuraminidase-inhibiting antibodies utilizing reverse genetics-derived antigens. *Influenza Other Respi Viruses* **3**:233–240.

Sandbulte MR, Westgeest KB, Gao J, Xu X, Klimov AI, Russell CA, Burke DF, Smith DJ, Fouchier RAM, Eichelberger MC. 2011. Discordant antigenic drift of neuraminidase and hemagglutinin in H1N1 and H3N2 influenza viruses. *Proc Natl Acad Sci U S A* **108**:20748–20753.

Sato K, Morishita T, Nobusawa E, Tonegawa K, Sakae K, Nakajima S, Nakajima K. 2004. Amino-acid change on the antigenic region B1 of H3 haemagglutinin may be a trigger for the emergence of drift strain of influenza A virus. *Epidemiol Infect* **132**:399–406.

Shih AC-C, Hsiao T-C, Ho M-S, Li W-H. 2007. Simultaneous amino acid substitutions at antigenic sites drive influenza A hemagglutinin evolution. *Proc Natl Acad Sci U S A* **104**:6283–6288.

Shu Y, McCauley J. 2017. GISAID: Global initiative on sharing all influenza data - from vision to reality. *Euro Surveill* **22**. doi:10.2807/1560-7917.ES.2017.22.13.30494

Smith DJ, Lapedes AS, de Jong JC, Bestebroer TM, Rimmelzwaan GF, Osterhaus ADME, Fouchier RAM. 2004. Mapping the antigenic and genetic evolution of influenza virus. *Science* **305**:371–376.

Smith K, Garman L, Wrammert J, Zheng N-Y, Capra JD, Ahmed R, Wilson PC. 2009. Rapid generation of fully human monoclonal antibodies specific to a vaccinating antigen. *Nat Protoc* **4**:372–384.

Starr TN, Greaney AJ, Addetia A, Hannon WW, Choudhary MC, Dingens AS, Li JZ, Bloom JD. 2021. Prospective mapping of viral mutations that escape antibodies used to treat

COVID-19. *Science* **371**:850–854.

Starr TN, Greaney AJ, Hilton SK, Ellis D, Crawford KHD, Dingens AS, Navarro MJ, Bowen JE, Tortorici MA, Walls AC, King NP, Veesler D, Bloom JD. 2020. Deep Mutational Scanning of SARS-CoV-2 Receptor Binding Domain Reveals Constraints on Folding and ACE2 Binding. *Cell* **182**:1295–1310.e20.

Starr TN, Zepeda SK, Walls AC, Greaney AJ, Alkhovsky S, Veesler D, Bloom JD. 2022. ACE2 binding is an ancestral and evolvable trait of sarbecoviruses. *Nature* **603**:913–918.

Steichen JM, Kulp DW, Tokatlian T, Escolano A, Dosenovic P, Stanfield RL, McCoy LE, Ozorowski G, Hu X, Kalyuzhniy O, Briney B, Schiffner T, Garces F, Freund NT, Gitlin AD, Menis S, Georgeson E, Kubitz M, Adachi Y, Jones M, Mutafyan AA, Yun DS, Mayer CT, Ward AB, Burton DR, Wilson IA, Irvine DJ, Nussenzweig MC, Schief WR. 2016. HIV Vaccine Design to Target Germline Precursors of Glycan-Dependent Broadly Neutralizing Antibodies. *Immunity* **45**:483–496.

Strelkowa N, Lässig M. 2012. Clonal interference in the evolution of influenza. *Genetics* **192**:671–682.

Tan GS, Lee PS, Hoffman RMB, Mazel-Sanchez B, Krammer F, Leon PE, Ward AB, Wilson IA, Palese P. 2014. Characterization of a broadly neutralizing monoclonal antibody that targets the fusion domain of group 2 influenza A virus hemagglutinin. *J Virol* **88**:13580–13592.

Tan TJC, Mou Z, Lei R, Ouyang WO, Yuan M, Song G, Andrabi R, Wilson IA, Kieffer C, Dai X, Matreyek KA, Wu NC. 2023. High-throughput identification of prefusion-stabilizing mutations in SARS-CoV-2 spike. *Nat Commun* **14**:2003.

Thwaites CL, Thanh TT, Ny NTH, Nguyet LA, Nhat NTD, Thuy CT, Thanh NTL, Dung NT, OUCRU Sero-surveillance Network, Campbell J, Thai PQ, Van Tan L, Choisy M, Boni MF. 2023. Seroprotection against tetanus in southern Vietnam. *Vaccine* **41**:2208–2213.

Underwood PA. 1980. Serology and energetics of cross-reactions among the H3 antigens of influenza viruses. *Infect Immun* **27**:397–404.

Vinh DN, Nhat NTD, de Bruin E, Vy NHT, Thao TTN, Phuong HT, Anh PH, Todd S, Quan TM, Thanh NTL, Lien NTN, Ha NTH, Hong TTK, Thai PQ, Choisy M, Nguyen TD, Simmons CP, Thwaites GE, Clapham HE, Chau NVV, Koopmans M, Boni MF. 2021. Age-seroprevalence curves for the multi-strain structure of influenza A virus. *Nat Commun* **12**:6680.

Wan H, Gao J, Yang H, Yang S, Harvey R, Chen Y-Q, Zheng N-Y, Chang J, Carney PJ, Li X, Plant E, Jiang L, Couzens L, Wang C, Strohmeier S, Wu WW, Shen R-F, Krammer F, Cipollo JF, Wilson PC, Stevens J, Wan X-F, Eichelberger MC, Ye Z. 2019. The neuraminidase of A(H3N2) influenza viruses circulating since 2016 is antigenically distinct from the A/Hong Kong/4801/2014 vaccine strain. *Nat Microbiol* **4**:2216–2225.

Westgeest KB, de Graaf M, Fourment M, Bestebroer TM, van Beek R, Spronken MIJ, de Jong JC, Rimmelzwaan GF, Russell CA, Osterhaus ADME, Smith GJD, Smith DJ, Fouchier RAM. 2012. Genetic evolution of the neuraminidase of influenza A (H3N2) viruses from 1968 to 2009 and its correspondence to haemagglutinin evolution. *J Gen Virol* **93**:1996–2007.

Wiley DC, Wilson IA, Skehel JJ. 1981. Structural identification of the antibody-binding sites of Hong Kong influenza haemagglutinin and their involvement in antigenic variation. *Nature* **289**:373–378.

Wolf YI, Viboud C, Holmes EC, Koonin EV, Lipman DJ. 2006. Long intervals of stasis punctuated by bursts of positive selection in the seasonal evolution of influenza A virus. *Biol Direct* **1**:34.

Worby CJ, Chaves SS, Wallinga J, Lipsitch M, Finelli L, Goldstein E. 2015. On the relative role of different age groups in influenza epidemics. *Epidemics* **13**:10–16.

World Health Organization. 2018. Influenza (seasonal). https://www.who.int/en/news-room/fact-sheets/detail/influenza-(seasonal)

Wrammert J, Smith K, Miller J, Langley WA, Kokko K, Larsen C, Zheng N-Y, Mays I, Garman L, Helms C, James J, Air GM, Capra JD, Ahmed R, Wilson PC. 2008. Rapid cloning of high-affinity human monoclonal antibodies against influenza virus. *Nature* **453**:667–671.

Wu NC, Wilson IA. 2017. A Perspective on the Structural and Functional Constraints for Immune Evasion: Insights from Influenza Virus. *J Mol Biol* **429**:2694–2709.

Wu NC, Xie J, Zheng T, Nycholat CM, Grande G, Paulson JC, Lerner RA, Wilson IA. 2017. Diversity of Functionally Permissive Sequences in the Receptor-Binding Site of Influenza Hemagglutinin. *Cell Host Microbe* **21**:742–753.e8.

Yang B, Lessler J, Zhu H, Jiang CQ, Read JM, Hay JA, Kwok KO, Shen R, Guan Y, Riley S, Cummings DAT. 2020. Life course exposures continually shape antibody profiles and risk of seroconversion to influenza. *PLoS Pathog* **16**:e1008635.

Yewdell JW, Webster RG, Gerhard WU. 1979. Antigenic variation in three distinct determinants of an influenza type A haemagglutinin molecule. *Nature* **279**:246–248.

Yu TC, Thornton ZT, Hannon WW, DeWitt WS, Radford CE, Matsen FA 4th, Bloom JD. 2022. A biophysical model of viral escape from polyclonal antibodies. *Virus Evol* **8**:veac110.

Yu X, Tsibane T, McGraw PA, House FS, Keefer CJ, Hicar MD, Tumpey TM, Pappas C, Perrone LA, Martinez O, Stevens J, Wilson IA, Aguilar PV, Altschuler EL, Basler CF, Crowe JE Jr. 2008. Neutralizing antibodies derived from the B cells of 1918 influenza pandemic survivors. *Nature* **455**:532–536.

Zost SJ, Lee J, Gumina ME, Parkhouse K, Henry C, Wu NC, Lee C-CD, Wilson IA, Wilson PC, Bloom JD, Hensley SE. 2019. Identification of Antibodies Targeting the H3N2 Hemagglutinin Receptor Binding Site following Vaccination of Humans. *Cell Rep* **29**:4460–4470.e8.

Zost SJ, Parkhouse K, Gumina ME, Kim K, Diaz Perez S, Wilson PC, Treanor JJ, Sant AJ, Cobey S, Hensley SE. 2017. Contemporary H3N2 influenza viruses have a glycosylation site that alters binding of antibodies elicited by egg-adapted vaccine strains. *Proc Natl Acad Sci U S A* **114**:12578–12583.