COMPUTATIONAL APPROACHES TO STUDY HOST CELL-MEDIATED IMMUNITY DYNAMICS FROM INFECTION/VACCINATION OF INFLUENZA A VIRUS

by

SWAN TAN

(Under the Direction of Anne De Groot and Justin Bahl)

ABSTRACT

Vaccination is one of the most cost-effective preventive measures for many infectious diseases, including influenza. However, the constant-evolving nature of influenza A virus (IAV) leads to annual revision of seasonal influenza vaccine components and at times, antigenic mismatch can potentially occur with the vaccine target, hemagglutinin (HA). In the case of swine, vaccines in use may not protect against the viruses that spread in pigs as there is no formal vaccine strain recommendation system and variation in swine influenza strains may arise. Human and swine influenza studies have shown that when vaccine strain and circulating strains are poorly matched, highly conserved T cell epitopes may limit disease spread in the absence of cross-reactive antibodies. Despite these findings, most influenza vaccine studies focus on humoral immune mechanisms and means of measuring the correlates of protection for T cell epitopes are still lacking. This dissertation addresses knowledge gaps in the vaccine development from the cellular-mediated immunity perspective. The goal of this dissertation is to evaluate T cell epitope conservation in influenza vaccines against circulating IAV viruses in swine and humans using T cell epitope prediction algorithms and phylogenetic analysis tools. Research aim 1 focuses on assessing a conserved T cell epitope-based prototype vaccine and determining the persistence of

T cell epitope conservation over a 5-year period. Aim 2 concentrates on identifying cross-conserved T cell epitope in HA sequences of human and swine influenza vaccines against emergent H1N1 G4 swine IAV (G4) to evaluate the potential for the G4 strain to impact swine and human populations. The work in aim 3 is about defining the human T cell immune landscape of H3N2 IAV using HA sequence data to estimate potential T cell epitopes and to examine how antigenic drift affects the diversity of T cell epitopes presented by the viral population over time. Collectively, my research findings present rationale for the use of computational means to analyze high dimensional data in the study of host immunity related to infection or vaccination by influenza virus. The studies outcomes also provide useful insights that may enhance influenza vaccine strategies as well as influenza surveillance efforts.

INDEX WORDS: Influenza A virus, swine influenza, human influenza, vaccine, immunoinformatics, T cell epitope, phylogenetic analysis

COMPUTATIONAL APPROACHES TO STUDY HOST CELL-MEDIATED IMMUNITY DYNAMICS FROM INFECTION/VACCINATION OF INFLUENZA A VIRUS

by

SWAN TAN

BS, National University of Malaysia, Malaysia, 2012

MS, Perdana University, Malaysia, 2017

A Dissertation Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2023

COMPUTATIONAL APPROACHES TO STUDY HOST CELL-MEDIATED IMMUNITY DYNAMICS FROM INFECTION/VACCINATION OF INFLUENZA A VIRUS

by

SWAN TAN

Major Professor:	Justin Bahl
Co-major Professor:	Anne S. De Groot
Committee:	Leonard Moise
	Liliana C. M. Salvador
	Stephen M. Tompkins

Electronic Version Approved:

Ron Walcott
Vice Provost for Graduate Education and Dean of the Graduate School
The University of Georgia
May 2023

DEDICATION

To my dearest parents and husband, mother-in-law, sister and brother-in-law, for their unconditional love and patience in my dream-chasing journey. They have been my pillars of support through storms and sunshine, and I cannot imagine these years without them.

ACKNOWLEDGEMENTS

This PhD dissertation would not have been possible without the generous support and help of many individuals who contributed in various ways. Hailing from Malaysia to the United States, and from Rhode Island to Georgia, I extend my sincere appreciate to my beloved family and everyone I came across for making this long journey a reality.

First and foremost, I want to thank Dr. Annie De Groot and Dr. Lenny Moise for offering me the opportunity to pursue my graduate studies. I am grateful for their constant supervision, guidance, and training during my two years in the Institute for Immunology and Informatics (iCubed), University of Rhode Island (URI). Both Dr. De Groot and Dr. Moise continued to mentor me remotely even after my transfer to University of Georgia (UGA), and I am truly appreciative of their dedication and unwavering support. I am also thankful for Dr. Justin Bahl, my major professor at UGA and another mastermind in my PhD career. Dr. Bahl is very resourceful and his passion in this field is inspiring. I am fortunate to be able to work with him, without which part of this work would not have been possible. Dr. Bahl is particularly flexible and considerate, and I certainly enjoy the lab environment in which I am surrounded by a bunch of encouraging and warmhearted peers. This is noteworthy to me as the Covid-19 pandemic hit shortly after I joined the lab.

My thanks and appreciation are also extended to the rest of my committee members, Dr. Liliana Salvador and Dr. Mark Tompkins. Dr. Salvador is caring and provided a lot of constructive advice that was motivating, especially when I hit a brick wall. Dr. Tompkins is very kind that he instantaneously agreed to be onboard when there was a major change to my dissertation committee.

I would like to thank the former committee member, Dr. Ted Ross. I am grateful that I had the chance to join Ross's lab meetings and share my research to others, which helped me gather different ideas and perspectives. Likewise, I have also extended my network and learned from people apart from my field.

Special thanks to all my friends from Rhode Island, Andres Gutiérrez who provided me consultation on immunoinformatics related skills and thought; Darshika Uduwatte who willing shared her experiences as a senior graduate student; Sundos Khan, Kai Yuan, Jiaying Liu, and Rongrong Li who played supportive roles as colleagues and good friends. I would like to also express my heartfelt gratitude to all my friends in Athens: Jiani Chen who has been a great collaborator to work with and we often share our thoughts about research and life with each other; Lambodhar Damodaran and Cody Dailey are another two key collaborators whom I worked very closely with, I could not have survived without them helping my project/writeup/presentations and sometimes cracking scientific jokes that keep my mental health afloat. I enjoyed being in the same working group with them; other members from the Bahl Lab and the Salvador Lab as well as the Ross Lab (Hyesun Jang – my marathon running buddy, Vanessa Silva Moraes and Ying Huang); Jia Hwei Zhang, a Malaysian graduate student at UGA who share the same culture as me and we have each other's moral support to keep us both sane.

Last but not least, I cannot thank enough my parents, my husband, my mother-in-law, my sister and brother-in-law, and all my friends for their boundless care, help and company. Their faith in me motivated me and helped me persevere throughout this journey.

TABLE OF CONTENTS

Page

LIST OF TABLES

Page

LIST OF FIGURES

Page

CHAPTER 1

INTRODUCTION

Influenza infection poses a significant threat to human health due to annual epidemics and occasional pandemics. In the U.S., the Centers for Disease Control and Prevention (CDC) estimates that during the 2019-2020 influenza season, influenza was associated with 38 million illnesses, 18 million medical visits, 405,000 hospitalizations, and 22,000 deaths [1]. Compared to the 2017-2018 season, the influenza burden was higher in young children (0-4 years), than in adults (18-49 years) in 2019-2020, providing evidence to support how severe seasonal influenza can be at any age. Influenza is also linked with significant economic losses in livestock. Influenza A virus (IAV) is among the most devastating pathogen for swine and poultry productions [2]. Several genetically and antigenically diverse IAV strains are endemic in swine worldwide and continue to cause significant losses to the swine industry.

Vaccination remains the most effective public health intervention for combatting influenza infections. However, the virus undergoes rapid evolution in every epidemic season, which makes vaccine strain selection particularly challenging. In addition, host and environmental factors add to the complexity in predicting which strains will dominate in future seasons. Viral genetic evolution and host adaptive immune profiles play essential roles, and these factors are relevant to be considered in the process of vaccine development and public health control.

The immune system has memory of highly specific protective antibodies formed in response to natural infection and vaccination. Human and swine studies have shown that conserved T cells resultant from previous exposure to influenza infection could cross-react to a novel IAV

strain [3], and that cellular immune responses are important for protection against IAV infection [4]. T cell-mediated immune responses are essential in reducing disease morbidity when flu vaccines and circulating IAV strains are poorly matched [6, 7]. While most influenza vaccine studies would focus on humoral immune mechanisms, my research will largely address the knowledge gap from a cellular-mediated immunity perspective.

This research strategy will use available vaccine and circulating IAV sequences to predict T cell epitope binding likelihood in swine and humans and analyze predicted T cell epitope content to discern adaptive immune response changes shaped by influenza vaccination. The proposed approach also presents an opportunity to combine with phylogenetic analysis to better interpret viral evolution and immune selection pressure. Moreover, another advantage of the study allows analysis of large amounts of data spanning a wide range of years and/or analysis of IAV whole proteomes.

In this dissertation, I aim to develop new analysis pipeline by combining a series of T cell epitope prediction tools (immunoinformatic) with phylogenetic approaches, to assess T cell epitope content in existing influenza vaccines and/or experimental vaccines against circulating and/or emergent IAV viruses in swine and humans. The goal of the proposed study outcome is to improve understanding of host adaptive immune profiles against IAV and may be relevant for developing a universal influenza vaccine in which cellular immunity contribution should also be addressed.

To achieve my research goal, the following three specific studies will be carried out:

Specific Aim 1: Assessment of a prototype epitope-based swine IAV vaccine. To evaluate the contribution of cell-mediated immunity to protect against severe disease even in the absence of

antibody response, I will quantify T cell epitope content based on highly conserved T cell epitopes between the prototype vaccine and circulating swine IAV strains spanning a five-year period. Another objective in this aim is to establish an immunology-based approach to estimate T cell epitopes conservation.

Specific Aim 2: Epidemic-risk identification of IAV circulating strains using immunology-based computational approaches. When influenza vaccines and emerging IAV strains are poorly matched, T cell-mediated immunity is crucial if there are no cross-reactive antibodies. This study will focus on an emerging swine IAV strain, G4, bearing Eurasian avian-like origin, which has been shown to have pandemic potential. I will compare the T cell epitope content of G4 strains and determine whether the T cell epitope content matches the existing and/or experimental vaccine strains as well as circulating strains in human and swine to further understand the potential threat of G4 virus in naive populations.

Specific Aim 3: Sequence-based approach to characterize host cell-mediated immunity selection on viral diversity. Although serological data are often used for viral antigenic characterization and vaccine efficacy evaluation, I will consider sequenced-based surveillance approaches to study human seasonal IAV evolution and host T cell immunity dynamic. To evaluate the dynamic of T cell epitope drift and replacement, I will analyze human seasonal influenza H3N2 vaccines and circulating strains to define human IAV evolutionary landscape in terms of immune selection pressure and viral diversification.

CHAPTER 2

LITERATURE REVIEW

Epidemiology

Influenza is an infectious disease of birds and some mammals, including humans, and has been present for more than four centuries since its first documented description in 1580 [7]. The contagious viral disease has caused occasional pandemics and countless seasonal epidemics. To date, influenza continues to raise global health concerns from time to time.

Influenza pandemics occur when an antigenically novel strain of the influenza virus results from the switching of virus gene segments, emerges in an immunity-naïve human population, and transmits efficiently among humans. There are four historic pandemics of influenza that occurred in the 20th century, the great H1N1 pandemic of 1918 (Spanish flu) with an estimated 21 million deaths worldwide, the H2N2 Asian flu of 1957, the H3N2 Hong Kong flu of 1968, and H1N1 Russian flu of 1977 [8], [9]. The H1N1 swine-origin flu pandemic of 2009 is the only influenza pandemic that took place in the 21st century [9]. Most influenza pandemics originated from non-human reservoirs, particularly among aquatic wildfowl, which serves as a natural reservoir [10].

While influenza pandemics pose a significant threat, seasonal influenza is responsible for the annual disease burden of influenza. In the United States (US), the Centers for Disease Control and Prevention (CDC) estimates that during 2019-2020 influenza season, influenza was associated with 38 million illnesses, 18 million medical visits, 405,000 hospitalizations, and 22,000 deaths [11]. Annual influenza epidemics peak during the winter season in temperate regions. Given that there are differences in the timing of winter, the influenza season in the Northern Hemisphere (NH)

usually falls between November and April, and for the Southern Hemisphere (SH), influenza season takes place between May and October. In tropical and subtropical regions, the epidemiology of influenza is more divergent throughout the year and the influenza period remains ambiguous [12].

In contrast to influenza pandemics, seasonal epidemics continue to recur yearly as influenza viruses evolve through point mutations (antigenic drift) and appear to be distinct from previously circulating influenza viruses [13]. This constant antigenic drift enables the virus to escape the host immunity that an individual gained from prior infections or vaccinations, as well as to be transmissible efficiently from human-to-human via respiratory droplets.

Influenza structure and virological features

The human respiratory pathogen that causes influenza is the influenza virus, which belongs to the Orthomyxoviridae family of ribonucleic acid (RNA) viruses [14]. The genome structure of influenza virus is composed of negative-sense, single-stranded segmented RNA. The negative-sense of the genome means it must be transcribed into a positive-sense RNA by RNA-dependent RNA polymerase (RdRp) prior to translation. This positive-sense RNA strand acts as a messenger RNA (mRNA). There are four types of influenza virus, namely influenza virus A, B, C, and D [14]. Aquatic birds are the primary reservoir of influenza A virus (IAV), which is also widespread in various mammals, including humans and pigs. The other members of the family include influenza B virus (IBV), a genus that infects only humans and seals [15]; influenza C virus (ICV), a rare type that is known to have infected human and pigs [16]; and a more recently detected species, influenza D virus (IDV), known to infect pigs and cattle, however, no human infections have been observed.

Scientists have confirmed through sequencing that these four genera of influenza viruses share a common genetic ancestry, however, they are genetically diverse [17]. These viruses are capable of reassorting and exchanging their viral RNA segments within the same genus, or type, but not across types. IAV and IBV have eight genome segments while ICV and IDV have seven genome segments which encode 10 and nine major proteins, respectively [14], [17], [18]. These genome segments are numbered in the order of decreasing length.

The first three segments are the largest genes in the virus genome. Segments one and two encode the polymerase basic two (PB2) and polymerase basic one (PB1) protein, respectively. PB2 is involved in 5' cap recognition and PB1 acts as a transcriptase. Additionally, segment two encodes a small protein, PB1-F2 via a second open reading frame that acts as a pro-apoptotic factor. Segment three encodes for two polymerase acidic proteins, namely PA (known as P3 for ICV and IDV), which acts as an endonuclease, and PA-X. Both PB1-F2 and PA-X are expressed through alternative open reading frames. These accessory proteins are important in host defense suppression, virulence, and pathogenicity. Three of the polymerase proteins (PB2, PB1 and PA), together with the nucleocapsid protein, NP, which are encoded by the fifth segment, form the viral ribonucleoprotein (vRNP) complex. vRNP complex is required for replication and transcription of the viral RNA (vRNA) [14], [17], [19].

For IAV and IBV, segments four and six encode the viral surface glycoproteins, hemagglutinin (HA) and neuraminidase (NA). HA mediates virus binding and internalization, whereas NA aids in release of the virus from the cell surface. ICV and IDV, encode a hemagglutinin-esterase fusion (HEF) protein on one segment that merges the functions of HA and NA. Segment seven (unspliced mRNA) and eight (spliced mRNA) encode the two viral matrix proteins (M1 and M2) and non-structural proteins (NS1 and NS2), respectively [20]–[23]. The

proteins M1, NS1 and NS2 mediate nuclear export of the newly synthesized vRNP complexes (composed of PB2, PB1, PA and NP). The vRNPs, together with M1 (the major viral structural protein), are assembled into virions that detach from the cellular membrane. M2 is also involved in the assembly process which acts as an ion channel protein that mediates function early or late in the infection. Furthermore, the protein NS1 functions as an interferon (IFN) antagonist [14], [17], [24]. A summary of the various functions of the proteins is provided in Table 2.1.

Among the four, IAV, which will be the focus of this review, has the most rapid mutation rate, driven by the error-prone nature of the RNA replication and as a result, IAV exhibits significant genetic diversity [14], [17]. IAVs are further characterized by the subtype of their surface glycoproteins, the HA and the NA. Influenza viruses have a standard nomenclature that includes as follows: the virus type; the species from which it was isolated (if non-human host); the location at which it was isolated; an isolation identifier; the isolate year; and, for IAVs only, numbering for the HA and NA subtype. Thus, A/Hong Kong/1/1968 (H3N2) was a human IAV isolated from Hong Kong in 1968 with isolation number of 1, and it has an HA subtype 3 and an NA subtype 2. Another example, A/Swine/Netherlands/3/1980 (H1N1) was a swine IAV isolated from Netherlands in 1980 with isolation number of 3, and it has an HA subtype 1 and an NA subtype 1. To date, there are 18 different known HA subtypes (H1 to H18) and 11 different known NA subtypes (N1 to N11) [25]. While many genetically distinct subtypes have been discovered in aquatic waterfowl and shorebirds, only three HA (H1, H2, and H3) and two NA (N1 and N2) subtypes have caused human epidemics, attributed to sustained, widespread, person-to-person transmission.

Viral life cycle from cell entry to virion release

The target organ of infection in humans is the respiratory tract [26]. When influenza virus is introduced into the respiratory tract, either by aerosol or by contact with saliva or other respiratory secretions from an infected individual, it attaches to and replicates in epithelial cells. The virus replicates in cells of both the upper and lower respiratory tract. Generally, there are four main stages in an influenza virus replication cycle [17]. It begins with (1) virus attachment and entry, followed by (2) viral RNA replication, (3) viral protein synthesis, and finally, (4) virion assembly and release (Figure 2.1). Each protein plays a vital role as the virus functional driving force [17].

Viral entry and endocytosis

The viral life cycle begins when the virus invades the primary physical barrier by binding to the respiratory mucus layer. Mucus in the respiratory tract contains sialylated glycoproteins, and the nine-carbon acidic monosaccharides, sialic acids (SA), are the target of recognition for influenza virus [14], [25]. The NA is a mushroom-shaped tetramer, anchored to the viral envelope by a transmembrane domain [27]. It possesses receptor-destroying activity, cleaving terminal sialic acid residues from cell-surface glycoproteins and gangliosides to release progeny virus from the host cell. Prior to binding to respiratory endothelial cells, the NA proteins on the viral envelope promote access to the target cells by degrading mucous, which helps to remove extracellular decoy receptors that would impede access to target cells. NA cleaves sialic acids and disables the inhibitory functions of the mucus, thus allowing penetration into the mucus layer [25], [27]. Binding to the target cell is then mediated by the HA proteins that are located on the surface of the viral envelope which enable fusion of the virions to host cell surface membranes that contain SA

receptors. This viral entry process is also known as receptor-mediated endocytosis.

Following endocytosis, the virus is internalized into the cell and the endosome is acidified by cellular lysosomes. The low pH activates the M2 ion channel and triggers a large conformational change in HA that exposes the fusion peptide and releases the matrix protein-coated vRNPs into the host cytosol [25]. The M1 protein shell surrounding the RNPs is also degraded, fully uncoating RNPs into the cytosol. Shortly after the uncoating process, released vRNPs are then translocated into the host nucleus via nuclear pore complexes, where they transcribe and replicate vRNA [25].

Viral RNA replication

The released eight segments of the influenza genome that are covered with NP and attached to RdRp complex next enter the host nucleus via nuclear pores. Given that the influenza RNA segments are negative strands, and hence they are all noncoding strands, each of the eight segments needs to be copied into two complementary (+) strands by the viral RNA polymerase. Once that occurs, the strand can be used for: (1) translating into viral proteins and (2) synthesizing complementary RNA (cRNA) intermediates from which the RNA polymerase subsequently transcribes more copies of negative-sense genomic vRNA. Therefore, the vRNPs (RdRp complex and NP) are responsible for both transcription and replication of the viral RNA genome [17], [25].

To make a mRNA for transcription by host ribosomes, the influenza RdRp complex which consists of PB1, PB2, and PA utilize a mechanism called “cap-snatching” for transcription. “Cap-snatching” is known as a process where the RdRp complex cuts off the 5’ cap from one of the host cell’s mRNA molecules and uses the cap to start transcription of vRNA. PB2 recognizes the 5′ cap of host pre-mRNA, and PA cleaves host mRNA to generate 5′-capped RNA fragments that are

then used as primers to initiate viral mRNA transcription. PB1 is responsible for carrying out viral mRNA synthesis. At the same time, RdRp also synthesizes cRNA which are then used as templates by viral polymerases to synthesize copies of the negative-sense genome [17], [25], [27].

Viral protein synthesis

Once viral mRNA is transcribed, it is exported out of the nucleus and translated by host ribosomes in a cap-dependent manner to synthesize viral proteins [28]. Newly synthesized viral RdRp subunits and NP proteins are imported to the nucleus to further increase the rate of viral replication and form RNPs. The viral surface proteins, HA and NA, and M2 are synthesized from mRNA of viral origin into the endoplasmic reticulum, where they are folded and transported to the Golgi apparatus for post-translational modification and are signaled to the cell membrane for virion assembly. M1 is responsible for bringing the RNP–NS2 complex [29] into contact with the envelope-bound HA, NA, and M2 proteins for packaging at the host cell membrane [17], [30]. Viral non-structural proteins including NS1 and accessory proteins, PB1-F2, and PA-X regulate host cellular processes to disable antiviral responses [14], [29]. PB1-F2 also interacts with PB1 to keep polymerases in the nucleus longer. M1 and NS2 proteins localize to the nucleus during the later stages of infection, bind to viral RNPs and mediate their export to the cytoplasm where they migrate to the cell membrane with the aid of recycled endosomes and are bundled into the segments of the viral genome [14].

Virion assembly and release

Influenza virus is not fully infectious unless its virions contain a full genome of eight segments. Progeny viruses leave the cell by budding from the cell membrane, initiated by an

accumulation of M1 matrix protein at the cytoplasmic side of the lipid bilayer [17]. When budding is complete, HA spikes continue to bind the virions to the sialic acid on the cell surface until virus particles are actively released by the sialidase activity of the NA protein. The NA also removes sialic acid residues from the virus envelope itself, which helps prevent newly assembled viruses from aggregating near the cell surface and improving infectivity [17], [31]. In viruses with inactive or absent NA, or in the presence of neuraminidase inhibitors, virus particles clump at the cell surface and infectivity is consequently reduced [17].

Antigenic drift and shift

Influenza virus and other RNA viruses have a remarkably rapid mutation rate. The segmented nature of the influenza virus genome allows for constant genetic change, mainly through two evolutionary mechanisms, such as antigenic drift (point mutations in the virus gene) and/or antigenic shift (reassortment or exchange of gene segment with another subtype of influenza virus). The RdRp complex encoded by the virus lacks a fidelity checking mechanism, thus the viral replication is error-prone, with a point mutation rate of ~1/104 bases per replication cycle. The point mutations are a major contributor to virus genetic variation via antigenic drift [32]. Due to these mutation events, IAVs can survive and infect a wide range of host species and occasionally, they are capable of transmitting and spreading among the same population, thereby, causing an outbreak.

Antigenic drift involves the accumulation of point mutations in viral genome, especially at the essential sites such as host recognition and antibody binding sites which are located at virus surface proteins. Antigenic drift is especially common for the HA protein. This can occur in response to evolutionary pressure exerted by the host immune response, resulting in the production

of novel strains that can evade pre-existing antibody-mediated and T cell-mediated immunity [33], [34]. Antigenic drift causes continual evolution of seasonal influenza virus [35], therefore contributing to the need for an annual update of influenza vaccine components.

Additionally, the extent of antigenic drift can alter the host receptor binding site (RBS) and specificity of HA protein, a major determinant of host tropism, for preference from avian to human sialic acid linkages [36]. For example, substitution of amino acid glutamine (Q) to leucine (L) at position 226 (Q226L9; H3 numbering) and from glycine (G) to serine (S) at position 228 (G228S) in HA have been described to be key mutations enabling change of binding preference to human SA receptor. The internal proteins NP, PA and PB2 have also been reported to harbor mutation sites, mostly located in the functional domains related to RNP-RNP interactions, which are crucial for viral replication [37]. In PB2, glutamic acid residue (E) mutations are commonly seen in avian viruses, restricting viral growth in humans and monkeys. A change from E to lysine (K) can restore and enhance virus replication in mammalian cells [38]. E627K was also able to convert a nonlethal H5N1 IAV isolated from a human to a lethal virus in mice [39]. Studies have also revealed that with just five mutations influenza A (H5N1) avian virus can become transmissible by air between ferrets (a good model to study human transmission).

The second evolutionary mechanism, antigenic shift, takes place when reassortment/recombination happens. For example, in a cell co-infected by viruses of two different subtypes, the HA segment (or others) of one strain is acquired by the other. Given that influenza viruses have segmented genomes, they are capable of reassortment. Genome segment reassortment and replacement can happen when cells are infected by viruses from a mixture of host species, such as human and animal viruses, which can result in hybrid progeny that is a novel and highly pathogenic strain, with no preexisting immunity in the human population. This is a sudden and

drastic change taking place in an influenza virus's antigen, usually HA. Antigenic shift, however, only occurs among influenza viruses of the same genus [20] and most commonly occurs among IAVs. This evolutionary process of IAVs creates a large diversity of influenza viruses in birds, though it is uncommon in human, equine, and canine lineages [40]. Pigs, bats, and quails have SA receptors for both mammalian and avian IAVs, so they are potential "mixing vessels" for reassortment [41], [42]. Pandemic influenza generally occurs when antigenic shift produces a virus strain to which the human population is immunologically, susceptible, and the virus is capable of sustaining human-to-human transmission [17].

Swine influenza

Besides morbidity and mortality in humans, influenza is also linked with significant economic losses in livestock, and influenza affecting one mammalian species (human, for example) can easily transmit to swine and vice versa [43]–[45]. Swine influenza causes respiratory disease in pigs, and it is one of the most concerning diseases in the pig industry. The common swine influenza viruses circulating in pig populations are H1N1, H1N2 and H3N2 subtypes [46]. The most prevalent circulating swine strains in the US are classical H1N1 swine virus, triple reassortant H3N2, avian-origin swine H1N1 and pandemic 2009 H1N1 virus (H1N1/pdm09) [47]. In contrast, the predominant swine viruses in Europe are H1avN1 (Europe-avian-like lineage), H1huN2 (human-like lineage), human-like reassortant swine H3N2, and H1N1/pdm09 [48], [49]. And since China is the leading pork producer worldwide, multiple lineages of swine viruses from both the US and Europe have been identified in Chinese pig herds. These swine virus strains include classical swine H1N1, Eurasian avian-like (EA) H1N1, H1N1/pdm09, triple-reassortant, and H3N2 lineages [50], [51]. Additionally, a recent emerging swine H1N1 G4 genotype has

drawn the attention of swine influenza experts as it is dominating other swine virus strains in Chinese swine populations, as reported by Sun *et al* [52].

The coexistence of diverse swine viruses within individual pigs promotes gene reassortment and emergence of new strains. The HA of most avian influenza viruses bind specifically to avian α-2,3-galactose sialic acid receptor, which is abundant in epithelial cells of avian trachea. Human influenza viruses, on the other hand, prefer α-2,6-galactose sialic acid linkage [53]. The different binding preferences of HA to the host sialic acid receptor in lung epithelial cells is the primary determinant of species barrier that differentiates avian and mammal viruses. Avian influenza virus are usually not able to infect a human host directly and vice versa [54]. Remarkably, the lung epithelial cells of pig contains both α-2,3- and α-2,6-galactose sialic acid linkages and therefore, pigs can be susceptible to both avian and human influenza viruses. Scientists have hypothesized that pigs act as a "mixing vessel" for swine, avian and human influenza viruses generating novel reassortant influenza viruses [55].

Reassortant swine influenza viruses can circulate amid regional pig farms and potentially be transmissible to humans. Pig farmers who are continuously exposed to reassortant swine viruses are at increased risk of zoonotic infection, and thus establish a vulnerable population of human hosts in which zoonosis and reassortment can develop at the same time [56]. The best-described example is the swine-origin H1N1/pdm09 influenza virus which resulted from viral genes reassortment that previously had not been identified. The H1N1/pdm09 influenza virus bears a combination of gene segments from three different influenza strains circulating in pigs (classical H1N1 swine virus), birds (EA H1N1) and humans (triple reassortant) [57].

Immune responses to infection

During IAV infections, both innate and adaptive immune responses are essential in protecting the host against infections to achieve viral clearance. At the site of infection, infected cells produce cytokines such as type I interferons (IFNs), interleukin (IL)-6, IL-8, tumor necrosis factor α (TNF-α), and chemokines, for example, CCL2 (MCP-1), RANTES, and MIP-1α that recruit immune cells including natural killer (NK) cells, neutrophils, macrophages, and dendritic cells (DCs) to initiate the innate immune response [58]–[60]. These innate and adaptive immune cells contribute to the immunopathology of infection, while also augmenting protection following influenza virus infection.

Innate immune response

Host innate immunity is the crucial first line of defense barrier that prevents viruses from replicating and propagating further in new hosts. As IAV causes infection of the upper and lower respiratory tract, alveolar and bronchial epithelial cells are the primary targets for influenza virus infection and replication [61]. IAV replicates most effectively in respiratory epithelium cells where the HA molecule is cleaved [62]. The virus can infect other cell types as well, including immune cells like tissue-resident macrophages and dendritic cells, however, these innate immune cells seem to be able to limit internal viral production [60], [61].

Cell infection and necrosis/apoptosis of infected cells activates a series of immune responses with production of proinflammatory cytokines such as TNF-α, IL-1, and other chemokines [63], [64]. Meanwhile, two type I IFN families, namely IFN-α and IFN-β and type III interferon, IFN-λ, can inhibit viral replication in epithelial cells [60]. Once the infected lung epithelium begins to release these inflammatory mediators, it leads to the recruitment of additional

innate cells, such as natural killer cells, neutrophils and proinflammatory monocytes to the lungs to kill infected cells and control viral infection [14], [58]. These host responses to infection are critical for viral clearance and initiation of adaptive immune responses. However, in some cases, when virus-infected epithelial cells, leukocytes and tissue-resident macrophages continue to induce high levels production of inflammatory cytokines and chemokines, the inflammatory response at the site of infection can be intensified [65], [66]. This severe immune phenomenon is regarded as cytokine storm and it can lead to immunopathology, lung damage and severe disease.

Adaptive immune response

Adaptive immune response, also known as acquired immune response, provides highly specific protection and usually takes time to develop before becoming fully functional. Even though innate immunity provides immediate viral control in the early stage, adaptive immune responses are necessary for eliminating infected cells, disease recovery, and protection from reinfection via memory cells that have developed over the course of infection. The two main components of adaptive immunity are cell-mediated immunity and antibody-mediated immunity (also known as humoral immunity). Essentially, the cell-mediated arm consists of T lymphocytes, namely helper T cells (CD4+) and cytotoxic T lymphocytes (CTL/ CD8+), whereas the antibody-mediated arm consists of B lymphocytes (plasma cells) and antibodies (immunoglobulins).

Both cell-mediated and antibody-mediated responses can be described by three important features: (1) they exhibit remarkable diversity as they are capable of responding to millions of different antigens; (2) they are able to last long even after an infection has been cleared and react quickly upon an immune recall because memory T cells and memory B cells are produced; (3) they show great specificity as their actions are specially directed against the antigen that initiated

the response. Typically, these highly specific processes are triggered by the initial recognition of the viral fragments or foreign substances that stimulate B or T cell activation to produce a specific response, for instance, antibody generation or T cell expansion. These responses will then precisely target and destroy the infected cells.

Humoral-mediated immunity/ B cell immunity

B cell immunity is essential to defend against viral invaders and can be acquired through either IAV infection or through influenza vaccination. Activation of B cells leads to antibody (Ab) production. Abs bind to pathogens or to foreign substances and neutralize them. Antibody responses against IAV infection are usually robust and long lasting in naive individuals. The B cell immune response begins approximately three days post infection and by day 7, immunoglobulin G (IgG) is secreted [67]. Using murine models, scientists found that most antibody-producing B cells secrete IgG and IgM in the mouse lung, while antibody-producing B cells secrete IgA, primarily in the mouse upper respiratory tract [68].

B cell immunity mainly targets virus external proteins HA, by inhibiting or 'neutralizing' viral cell entry and blocking virus attachment to the cell surface, and NA, by inhibiting virus exit and preventing virus adhesion to the receptors on the epithelial cells [69]–[71]. Neutralizing Abs often correlate with protection and contribute directly to eliminating infection [72], [73]. Typically, neutralizing Abs against IAV are directed to the conformational epitopes on the globular domain of HA. There is also a minor Ab population identified as non-neutralizing Abs, which are specific to other viral epitopes. Non-neutralizing Abs can also be protective by other mechanisms, such as activating complement or promoting antibody-dependent cellular cytotoxicity [74].

As neutralizing Abs against IAV generally recognize the globular head region of HA, this

head domain has been extensively characterized using monoclonal Abs (mAbs) technology, and is regarded as the immunodominant region [69], [70], [75]. There are five well-known discrete (non-overlapping) antibody binding sites: Sb and Sa are located at the top of the globular domain of HA, while Ca1, Ca2 and Cb are located at the bottom of the head. These sites were described as the major regions recognized by neutralizing Abs and showed inhibitory activity in the hemagglutination inhibition (HI) assay [76]. Essentially, an HI titer of 40 or greater often corresponds with protection. Despite persisting issues with its accuracy, the HI assay remains the gold standard of laboratory assays for classifying and subtyping IAVs.

In contrast with the globular head domain of HA, the stem/stalk domain of HA is relatively more conserved. The stem domain has been reported to induce broadly neutralizing Abs (bnAbs) that are capable of neutralizing IAV of different subtypes. An early study found that bnAbs that target the HA stem region in mice had no HI activity, but they were capable of neutralizing H1 and H2 viruses [77]. Another group of researchers, Guthmiller *et al,* identified a novel class of bnAbs that target stem-specific epitope of pre- and post-pandemic H1N1 viruses as well as a swine-origin H1-expressing virus. This recent finding demonstrates the potential for cross-reactivity of pandemic-subtype neutralizing antibodies that target stalk-binding epitopes [78]. Still, studies have also revealed that the viral mutants are capable of escaping bnAbs. New vaccine approaches should aim to boost the generation of Abs against the stem domain in order to get maximal protection against seasonal and pandemic IAVs [33], [78].

Apart from HA-specific Abs, NA-specific Abs can limit viral load by interfering with the exit of virions and anti-NA Abs have shown to have protective potential against influenza infection [71], [79]. There is evidence that anti-NA mAbs and NA vaccination protect against IAV challenge in animal models [79]–[81]. Unfortunately, anti-NA Abs have not been prioritized and NA is not

usually included in influenza vaccines, due to the central role that HA has played in influenza research. The current vaccine designs are mostly HA-focused [82].

IAV infection also induces Abs against internal proteins, such as NP and matrix proteins [83], [84]. M2 protein is the third most abundant protein on the IAV viral surface following HA and NA. The binding of M2-specific mAb to its extracellular N-terminal domain was shown to be able to obstruct IAV strains replication [85]. Although few studies in mice have shown that anti-NP Abs can help to clear influenza infection [86], [87], the protective role of Abs elicited by internal proteins is yet to be determined. Usually, the antibody responses induced by conserved epitopes of the internal proteins are weak and therefore, their potential contributions to protection in the general population is neglected [88].

Besides focusing on host pathogenesis associated with IAV infection, prevention of IAV transmission between hosts reduces the propagation of the virus in the population. Since the mucosal surfaces of the respiratory tract are the primary entrance for IAVs, secreted Abs elicited through mucosal immune responses have been found to be important in restricting IAV transmission. Seibert *et al* proved that only mucosal immunity (including IgA) but not systemic immune responses (IgG) can efficiently block IAV transmission in the guinea pig model [89].

Cell-mediated immunity/ T cell immunity

T cell immunity along with B cell immunity, contributes to efficient and ultimate viral clearance which involves a series of naïve immune cells activation, rapid proliferation, recruitment, and expression of effector activities. During an influenza infection, naïve CD4+ T cells recognize the viral antigens presented via the class II major histocompatibility complex (MHC) of antigen presenting cells (APCs). Upon antigen recognition, naïve CD4+ T cells

differentiate into two main subsets of helper T cells, namely Th1 and Th2, in response to distinct cytokines [90], [91]. Activated CD4+ T cells subsequently support class-switching of antibodies and promote optimal CD8+ T cells responses [92], [93]. In mice, CD4+ T cell responses in the lung reach a maximum 10 days post influenza infection [94]. Influenza virus-specific effector CD4+ T cells isolated from mouse draining lymph nodes and lung can confer protection, in the absence of antibody, to naïve recipient mice following IAV challenge [94].

The cytokine milieu generated during influenza virus infection can lead to polarization of helper T cells to further support the production of either Th1 or Th2 cells [90]. Cytokine IL-12 triggers Th1 immune response, which is also a proinflammatory response, that is targeted at the destruction of cells infected with intracellular parasites (including bacteria and viruses) and boosts autoimmune responses [91]. On the other hand, cytokines IL-4 and IL-2 trigger Th2 immunity that produces anti-inflammatory responses to aid in the development of antibody responses and other immune mechanisms to kill large, extracellular parasites [95].

Th1 cells secrete mainly IFN-γ, IL-2 and TNF-α. These cytokines activate macrophages, as well as CD8+ T cells, promote IgG B cells and mediate cellular immune responses [90], [91]. Th2 cells secrete IL-4, IL-5, IL-6, IL-10 and IL-13 and prompt B cells isotype switching to produce antibodies such as IgE [90], [91]. Th1 immune response is associated to facilitate virus clearance and mediate heterosubtypic immunity by the generation of CTLs compared with Th2 cells. Interestingly, Th2 responses can be converted to Th1 responses in a Th1-driven cellular environment and achieve heterosubtypic immunity [96].

Apart from CD4+ T cells, another pivotal component of cell-mediated immunity is CD8+ T cells. Effector CD8+ T cells/ CTLs are necessary for optimal influenza virus clearance and host protection [97], [98]. Naïve CD8+ T cells are activated through the recognition of peptide-MHC

class I molecules which presents viral peptides of 8-12 amino acids in length. CD8+ T cell responses peak at day 8 post influenza infection in draining lung lymph node and at day 10 post infection in bronchoalveolar lavage fluid [99].

The role of CTL effector T cells is to eliminate infected cells by a process known as cytolysis. Activated CD8+ T cells release effector cytokines and cytolytic molecules once they interact with antigen bearing target cells, including CD45+ APCs and IAV infected epithelial cells via T cell receptor (TCR) signaling [100], [101]. Generally, effector CD8+ T cells utilize the death receptor ligand, Fas ligand (FasL), and perforin-dependent pathways to clear influenza virus infection [98], [102]. CTL effectors expressing FasL binds to Fas on target cells and the Fas-FasL interaction directly kills virus infected cells. Additionally, CTL produce perforin to permeabilize the membranes of infected host cells and secrete cytolytic granules such as granzyme and perforin into cells to induce apoptosis [102].

Studies have suggested that CTL effectors may use other mechanisms to resolve influenza infection in the lung. CTL effectors stimulate antiviral responses by expressing antiviral effector cytokines which include IFN-, TNF- and IL-17, and eliciting multiple chemokines such as CCL3, CCL4 and CCL5 to recruit other immune cells to the infected sites, contributing to viral defense mechanisms [98], [102]. Notably, CTL effectors also produce the regulatory cytokine, IL-10, to control excessive pulmonary inflammation resulted from the immune response to influenza virus infections [98], [102].

Animal and human challenge studies suggest that T cell immunity confers protection against disease, and scientists have hypothesized that cellular immunity may limit symptomatic illness in the absence of antibodies [3], [5], [103], [104]. The rationale behind this hypothesis is that T cell responses targeting highly conserved internal influenza antigens could provide cross-

protective immunity against different IAV subtypes. Notably, T cell responses have been shown to be involved in limiting disease severity in the absence of cross-reactive neutralizing Abs, an activity that is strongly associated with highly conserved CD8+ T cell epitopes derived from IAV internal antigens including NP, M1 and PB1 [103]. In fact, internally-derived-epitopes are more conserved across a broad range of influenza virus strains and subtypes as they are less subject to antibody-selected antigenic drift [14]. Furthermore, there is obvious evidence to support the idea that established CTL memory directed towards conserved viral peptides presented by diverse HLA-prevalent human populations can induce protective heterosubtypic cellular immunity against novel IAV strains [3], [5], [103]–[105].

Vaccination against influenza generally targets HA protein of the circulating viral strain. As a result, antibodies that confer homosubtypic immunity against predominant circulating strains are less potent when there is possible mismatch between vaccine strain and circulating strains [106], and studies have also shown that antibody titers can wane during an intense season [107], [108]. Although repeated annual immunization can prevent infection by matched homologous strains, it may also prevent the induction and maintenance of heterosubtypic cellular immunity, thereby leaving individuals susceptible to more severe disease with new reassortant viral strains that evade vaccine-induced humoral immunity [109].

Given that the present vaccination approach is unable to provide optimal protection against both homologous and heterologous viral strains, there is a broad interest in potential strategies for cross-protective vaccination by incorporating T cell-induced immunity in vaccines. Immunization using live attenuated viral strains have the potential to induce CD8+ T cell responses, but research data suggests that they do not promote strong CTL memory [110]. There is still a major knowledge gap concerning how to achieve optimal cellular immunity before universal influenza vaccines can

be developed. Further studies of T cell immunity in influenza virus infections will advance the overall understanding of cross-protective viral immunity.

Computational approaches to studying host immunity to influenza virus infection

Our adaptive immune system can establish a state of immunological memory following natural infection or through vaccination which is crucial in response rapidly and effectively to the pathogen that has been encountered before. In cell-mediated immunity, T cells exhibit the ability to detect peptides derived from "self" and "non-self", serving as a starting point for prior activation and clonal expansion. T cells execute a precise recognition process that depends on the interaction between the peptide-MHC complex presented on the surface of an APC and the TCR expressed on the T cell surface. The T cell epitope that displayed on the host MHC receptor is critical for the initiation of TCR signaling. The MHC-binding epitope is derived from a fragment of viral antigen, comprised of at least nine amino acid residues, and presented in a linear form. Predicting this important string of amino acid residues from sequence data is achievable using algorithms that search for patterns in viral sequence data. The information can be used to define putative T cell epitopes that are later validated *in vitro* or *in vivo*.

With the advancement of recombinant DNA technology over time and sequencing data becoming easily accessible on public websites, sequence-based analysis for the purpose of understanding pathogens and disease has gained in popularity not only in the field of the Omics (genomics, transcriptomics, proteomics, or metabolomics) and molecular biology but also in the immunology research [111]. The development of sequencing techniques and the improvement of knowledge on the host immune response and virology along with the expanding usage of bioinformatics tools motivated a new generation of vaccine design that can be achieved without

the need to invest in expensive laboratory reagents and human effort [112], [113]. A wide range of computational tools for epitope prediction have been developed for exploring the adaptive immune responses of humans and other vertebrates such as swine and canines, among others. The term, "immunoinformatics", represents the application of computational tools to define and analyze immunological data [114]. To achieve these goals, a thorough understanding of the disease agent or precisely, the critical antigens/epitopes to induce the appropriate immunological reaction and correlates of immunity, is required. These tools are used in a sub-discipline of immunoinformatics called "computational vaccinology". In this context, immunoinformatics tools can be integrated into the "ten steps framework" for vaccine development that has been described by De Groot *et al* (Table 2.2).

T cell epitope prediction tool

Immunoinformatics tools, in particular T cell epitope prediction algorithms, drive the selection of potent T effector epitopes that are likely to be recognized within an individual or in a broader population, while also removing epitopes that may drive immunopathogenic or immune tolerogenic responses by the notion of cross-reactivity with host sequences [113], [114]. T cell epitope prediction and screening can also guide downstream experimental vaccine analysis studies.

The standard first step in the prediction workflow is to evaluate immunogenicity of a given protein sequence. Experiment data has shown that the T cell epitopes interact and attach to the MHC class I and class II binding groove through binding of the R group side chains into pockets in the MHC [115], [116]. Given the fact that T cell epitopes are bound in a linear fashion in the binding groove of MHC molecules, the binding likelihood between sequence-derived amino acid residues and the "binding pockets" of the MHC receptor can be calculated, based on the unique

properties of amino acid R groups, and the interface between ligands and TCR can be modeled with accuracy [114], [117]. Based on these findings, numerous T cell epitope mapping algorithms have been established and integrated into web servers to rapidly identify putative T cell epitopes [118]–[120].

Different MHC class I and class II binding tools have been developed, using matrices based on sequence motifs, position specific scoring matrices (PSSM), quantitative matrices (QM), artificial neural networks (ANN), support vector machines (SVM), quantitative structure activity relationship (QSAR) and molecular docking simulations [113]. For instance, EpiMatrix and PigMatrix, both developed by EpiVax, use PSSM search for human and swine epitopes, respectively. The input sequence is parsed into overlapping frames of nine amino acids, which are equivalent to the minimal length of an MHC-binding peptide. Scoring starts at the beginning of the input sequence and the PSSM is iterated over the sequence, shifting the analysis frame by one residue at a time, until the end of the sequence is reached.

Predicted binding scores can be computed by EpiMatrix for the likelihood of each 9-mer in an antigen to bind to a panel of class I alleles (A*01:01, A*02:01, A*03:02, A*24:02, B*07:02, and B*44:03) and class II (DRB1*01:01, DRB1*03:01, DRB1*04:01, DRB1*07:01, DRB1*08:01, DRB1*09:01, DRB1*11:01, DRB1*13:01, and DRB1*15:01) human leukocyte antigen (HLA) alleles. These are HLA allele supertypes (alleles sharing common binding preferences) that cover the genetic diversity of more than 95% of human populations globally [114], [121], [122]. Similarly, EpiMatrix can be used with swine leukocyte antigen (SLA)-specific PSSM to assign predicted binding scores for each 9-mer from swine IAV sequences. SLA alleles are used in the prediction model include class I (SLA-1*01:01, 1*04:01, 1*08:01, 1*12:01, 1*13:01, 2*01:01, 2*04:01, 2*05:01, 2*12:01, 3*04:01, 3*05:01, 3*06:01, 3*07:01) and class II

(SLA-DRB1*01:01, 02:01, 04:01, 04:02, 06:01, 06:02, 07:01, and 10:01) SLA molecules. The SLA alleles selection for swine is designed to reflect dominant SLA types, each of which has binding pocket preferences that are shared with several other common SLA alleles, based on a study by Gutierrez *et al* [123], [124].

For each 9-mer, each individual allele of a set of MHC alleles, PigMatrix or EpiMatrix raw binding scores, are normalized to Z scores using the average and the standard deviation of scores calculated for 100,000 random 9-mers [114], [125]. To screen the potential 9-mer binders, a binding threshold is defined as 9-mers with Z scores greater or equal to 1.64, which comprise the top 5% in the normalized set of scores for each SLA or HLA allele of sequences. Sequences with scores above 1.64 are predicted to have significant SLA or HLA binding potential. Higher Z-scores associate with higher MHC binding probability [124], [126].

Cross-conservation of T cell epitope

Cross-reactivity of TCR has been extensively explored in recent decades [127], [128]. The fact that TCR can potentially interact with a range of T cell epitopes presents an opportunity to further investigate the connection and significance between conservation with self and non-self, T cell epitope phenotypes (effector or tolerant), and immunodominance. To facilitate research in this growing area of investigation, new tools have been developed. For example, the JanusMatrix tool was designed to expand on EpiMatrix output and search for the potential cross-reactivity between TCR and T cell epitopes derived from the human genome, the human microbiome, and human pathogens [114], [129].

Although T cells recognize linear peptides that are displayed in the form of peptide-MHC complex by human (HLA) or other species' MHC molecules, there are two dimensions to consider,

i.e., the amino acid side chains of T cell epitopes that project out of the MHC binding cleft that binds to the TCR of T cells, and the amino acid side chains of T cell epitope embedded into the groove of the MHC binding cleft. The whole linear peptide or the T cell epitope side chains that attach to TCR is commonly known as the "epitope", while the T cell epitope side chains that stick into MHC-binding "pocket" is referred to as the "agretope".

An effective T cell immunity is achieved when T cells can generate an immune response to antigenic epitopes that have not been encountered before. However, it is estimated that there are exceedingly high number of potential epitopes (more than 10^{12}) whereas the immune system must cope with an estimated number of less than 10^{8} available TCRs in humans [129]. It stands to reason that TCRs may recognize more than one epitope and that theory has been validated by Wooldridge *et al*, demonstrating that each TCR has the potential to recognize as many as one million peptides [129], [130]. This potential for cross-reactivity may seem high, however, given that the number of potential T cell epitopes represented by modifying each of the TCR-facing positions using the 20 amino acids, suggests that the minimum cross-specificity may be 1:100,000 peptides [130]. However, by altering MHC binding residues, the affinity of the peptide can be modified and the shape of the TCR face can be different. Since the TCR has shown the ability to adapt to minor changes, a single TCR can bind to more than one epitope containing the same TCR-facing but different MHC-facing residues [131], [132].

Extending from this observation, immunoinformatics tools were developed to evaluate cross-conservation between host and pathogen T cell epitopes. Initially, the idea was to explain cross-reactivity between pathogen epitopes, however, it was soon discovered that many pathogen epitopes were also conserved with human epitopes. Building on prior research related to T cell epitopes that were highly conserved within the human genome and found to be associated with

regulatory T cell responses, the team of Moise *et al* began to explore the hypothesis that tolerogenic T cell epitopes (Ttol) might bear TCR-facing residues that are cross-conserved in human (self) proteins, whereas the effector T cell epitopes (Teff) might bear TCR-facing residues that are less cross-conserved with human genome proteins [129]. Using this concept as a basis for identifying potentially tolerogenic epitopes, tolerogenic epitopes were discovered in pathogens infecting humans, ranging from *Brugia malayi* (the cause of lymphatic filariasis), to *Mycobacterium tuberculosis* (the causative agent of tuberculosis), to human immuno-deficiency virus (HIV), and Hepatitis C [133]. These discoveries led to the concept of "immune-engineering" pathogens, the rationale being that putatively tolerogenic and cross-conserved epitopes might be important to remove from vaccines as they might induce regulatory T cell responses, reducing the immune response to other T cell epitopes and B cell responses. The JanusMatrix tool has been developed to detect the TCR-facing amino acids that are conserved between pathogen epitopes and human 9-mer peptides by searching through various large databases built into the tool. For instance, JanusMatrix can search for homologies within the human genome and human microbiome databases [134].

In addition, human studies have shown that cross-conserved cellular immune responses are important [94]. For example, conserved influenza-specific T cells resulting from previous influenza infections can also cross-react to similar epitopes found in novel IAV strains [5], [6], [104]. When cross-reactive antibodies are absent due to antigenic mismatch, cross-conserved influenza-specific CD4+ immune responses can reduce morbidity [5]. Therefore, methods for studying potential cross-conservation of T cell epitopes at the residues that interface with TCRs are important. Essentially, this pathogen-to-pathogen comparison is an extended application of TCR cross-reactivity using JanusMatrix since the TCR does not interrogate side chain that are

buried in the MHC binding groove. A T cell epitope can be cross-reactive despite not being identical in sequence to another epitope. In other words, two or more peptide-major histocompatibility complex ligands from different influenza viruses can be recognized by the same T cell's TCR if the TCR-facing amino acids in the two sequences are identical [135].

Taken together, sequence-based approaches are gaining popularity in studying virus evolution and antigenicity. This dissertation documented the application of immunoinformatics to predict and quantify T cell epitope conservation of vaccine strains against circulating strains for cross-protection by conserved T cell epitopes, along with phylogenetic analysis to track conserved T cell epitope changes over time in two different hosts.

Table 2.1. Example genomic information of influenza A/California/07/2009 (H1N1) virus and its encoded proteins.

Segment	Encoded protein(s)	Coding sequence (CDS) length in nucleotides	Protein length in amino acids (aa)	Protein function
1	Polymerase basic 2 (PB2)	2280	759	Polymerase subunit; mRNA cap recognition
2	Polymerase basic 1 (PB1)	2274	757	Polymerase subunit; RNA elongation, endonuclease activity
3	Polymerase acidic (PA)	2151	716	Polymerase subunit; protease activity
	Polymerase acidic accessory protein (PA-X)	700	232	Aids in inhibiting host antiviral and immune response and contribute to viral growth
4	Hemagglutinin (HA)	1701	566	Surface glycoprotein; major antigen, receptor binding and fusion activities
5	Nucleoprotein (NP)	1497	498	RNA binding protein; nuclear import regulation
6	Neuraminidase (NA)	1410	469	Surface glycoprotein: sialidase activity, facilitate virus release process
7	Matrix protein 1 (M1)	759	252	Matrix protein; vRNP interaction, RNA nuclear export regulation, viral budding
	Matrix protein 2 (M2)	294	97	Tetrameric membrane ion channel; virus uncoating and assembly
8	Non-structural protein 1 (NS1)	660	219	Inhibits host antiviral immune responses by suppressing the expression of host mRNAs that enable IFN-induced antiviral phenotypes
	Nuclear export protein (NEP/NS2)	286	121	Transports newly synthesized RNPs out of the nucleus after amplification. Also important for efficient influenza virion formation and budding.

Table 2.2. The ten steps framework to vaccine design.

Step	Description	Details
1	Define disease	Understanding of disease epidemiology associated with a specific pathogen
2	Define pathogen	Identification of pathogen and classification or comparison to existing pathogens
3	Is there immunity?	"Natural" evidence of post-infection immunity to follow-on infections
4	Correlates of immunity	Defining the relevant immune responses to measure during vaccine studies
5	Critical antigens	Identification of key targets of protective response
6	Animal model	Selection of appropriate animal model for pre-clinical testing
7	Prototype vaccine	Selection of a vaccine prototype and experimental proof of concept
8	Safety, toxicity, and stability studies	Preclinical steps prior human subject testing
9	Clinical trials	Phase I, II and III of human subject testing
10	Distribution, acceptance, and access	Access to the vaccine, distribution to, and acceptance by at risk populations

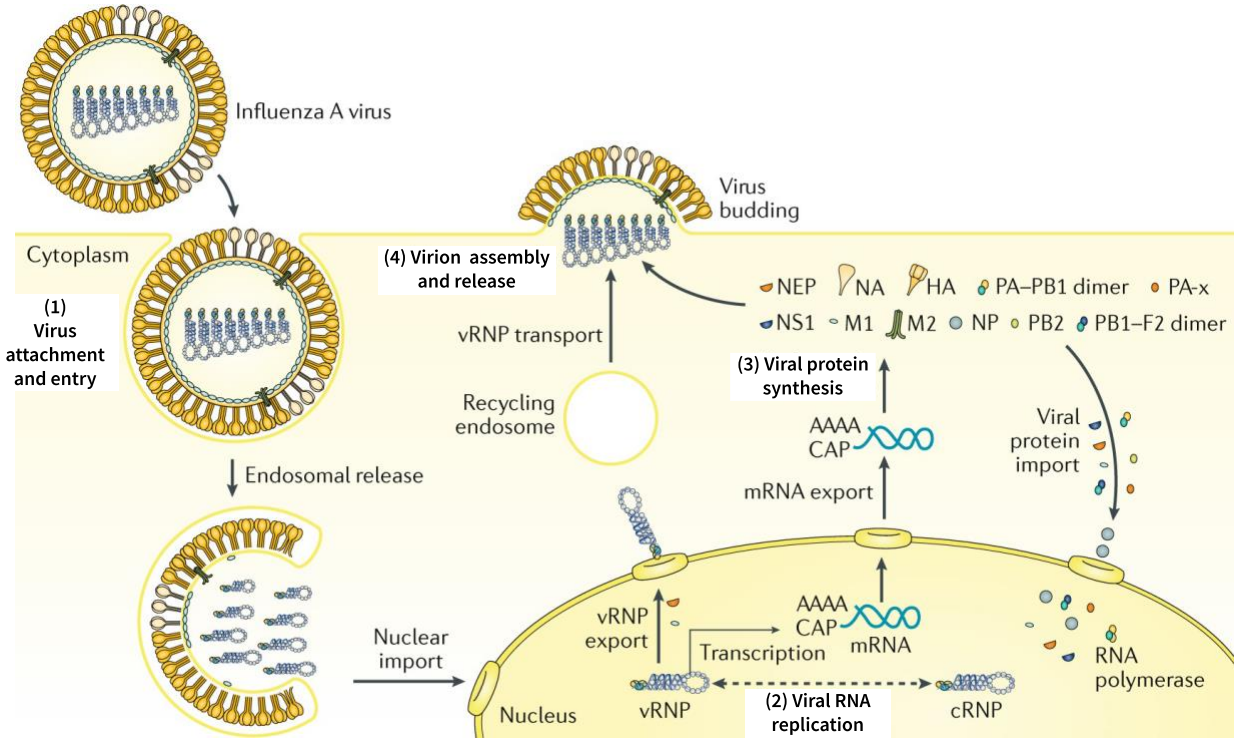


Figure 2.1. The viral life cycle of influenza A virus (IAV). The virus (1) attaches to host cell surface membrane that contains sialic acids (SA) receptor and enters the host respiratory mucus layer via receptor-mediated endocytosis. Acidification of the endosome activates the M2 ion channel and induces conformational change of the hemagglutinin (HA) protein causing the virus to fuse with the endosomal membrane and release viral ribonucleoproteins (vRNPs) into the cytoplasm. The genetic material is imported to the nucleus for transcription of mRNA and replication (2) through a mechanism called "cap-snatching" where the RNA-dependent RNA polymerase (RdRp) complex cuts off the 5' cap from one of the host cell's mRNA molecules and uses the cap to start transcription of viral RNA. At the same time, RdRp also synthesizes complementary ribonucleoprotein (cRNP) which are then used as templates by viral polymerases to synthesize copies of the negative-sense genome. Viral mRNA is exported out of the nucleus and translated into viral proteins by host ribosomes in the cytoplasm (3). Newly synthesized viral RdRp subunits and NP proteins are imported to the nucleus to further increase the rate of viral replication

and form RNPs. HA, NA, and M2 are synthesized from mRNA of viral origin into the endoplasmic reticulum, where they are folded and transported to the Golgi apparatus for post-translational modification and are signaled to the cell membrane for virion assembly. The newly synthesized vRNPs undergo post-translational modification and are signaled to the cell membrane for virion assembly (4). Progeny viruses leave the cell by budding from the cell membrane, initiated by an accumulation of M1 matrix protein at the cytoplasmic side of the lipid bilayer. Figure adapted from Krammer *et al* [14].

CHAPTER 3

QUANTIFYING THE PERSISTENCE OF VACCINE-RELATED T CELL EPITOPES IN CIRCULATING SWINE INFLUENZA A STRAINS FROM 2013–2017 [1]

[1]Tan, S.; Gutiérrez, A.H.; Gauger, P.C.; Opriessnig, T.; Bahl, J.; Moise, L.; De Groot, A.S. Quantifying the Persistence of Vaccine-Related T Cell Epitopes in Circulating Swine Influenza A Strains from 2013–2017. *Vaccines 2021, 9, 468*. https://doi.org/10.3390/vaccines9050468. Reprinted here with permission of the publisher.

Abstract

When swine flu vaccines and circulating influenza A virus (IAV) strains are poorly matched, vaccine-induced antibodies may not protect from infection. Highly conserved T cell epitopes may, however, have a disease-mitigating effect. The degree of T cell epitope conservation among circulating strains and vaccine strains can vary, which may also explain differences in vaccine efficacy. Here, we evaluate a previously developed conserved T cell epitope-based vaccine and determine the persistence of T cell epitope conservation over time. We used a pair-wise homology score to define the conservation between the vaccine's swine leukocyte antigen (SLA) class I and II-restricted epitopes and T cell epitopes found in 1272 swine IAV strains sequenced between 2013 and 2017. Twenty-four of the 48 total T cell epitopes included in the epitope-based vaccine were highly conserved and found in >1000 circulating swine IAV strains over the 5-year period. In contrast, commercial swine IAV vaccines developed in 2013 exhibited a declining conservation with the circulating IAV strains over the same 5-year period. Conserved T cell epitope vaccines may be a useful adjunct for commercial swine flu vaccines and to improve protection against influenza when antibodies are not cross-reactive.

Introduction

When a new strain of pathogen emerges, the first question asked is often whether existing vaccines might be effective against it. In the past, experts have relied on examining the humoral immune response by using antibody assays to determine the potential of existing vaccines to cross-protect [136]. It is now well established that cell-mediated immunity (CMI) contributes to the protection against severe disease even in the absence of antibody response [5], [6], [137], [138]. CMI involves cytotoxic T lymphocytes (CTL) and T helper (Th) lymphocytes, which are triggered to respond when their T cell receptors (TCR) recognize T cell epitopes presented by class I or class II major histocompatibility complex (MHC) molecules on the surface of antigen presenting cells or infected cells [139].

In humans, immune responses to conserved T cell epitopes may result in reduced morbidity, despite the lack of cross-reactive antibody to the new strain [140]–[142]. This is supported by a case-controlled study that investigated the association of the pandemic IAV H1N1 2009 infections with 2008–2009 seasonal trivalent inactivated flu vaccination [143]. Previous seasonal vaccination protected against pandemic H1N1, despite the lack of antibody protection. Using the immunoinformatic tools available to us at the time, we defined T cell epitopes that were present in the newly emergent strain (pH1N1 A/California/04/2009; GenBank accession numbers ACP41105 for hemagglutinin or HA and ACP41107 for the neuraminidase or NA), and highly conserved in the existing seasonal influenza vaccine (containing H1N1 A/Brisbane/59/2007; GenBank accession numbers ACA28844 for HA and ACA28847 for NA) [144]. The *in silico* analysis demonstrated that despite the lack of antibody cross-reactivity, more than 50% of T cell epitopes in the novel pH1N1 virus were also present in the seasonal vaccine, supporting the

concept that pre-existing T cell response due to vaccination or exposure may have protected in the absence of protective antibody response.

Swine experimental studies have also shown that pigs were protected from heterologous infection and challenge between 'avian-like' H1N1 and 2009 pandemic H1N1 lineages in the absence of cross-reactive antibodies, establishing the role of cross-reacting T cells [145], [146]. In concurrence, researchers were able to identify cross-reacting CD8 T cell epitopes in pigs and nucleoprotein (NP)-specific CD8 T cells were induced following immunization by aerosol [147], [148]. Prospective animal studies also confirmed that seasonal H1N1 vaccines that did not induce cross-reactive antibody responses, but induced cross-reactive T cell responses, did not protect against the pandemic pH1N1 infection, but greatly reduced morbidity, mortality, virus replication, and viral shedding [149]. Thus, T cell epitopes can be conserved between both human and swine vaccines and emerging influenza strains and have been shown to contribute to protection.

Antigenic shift and drift are significant challenges not only to human seasonal vaccination but also to effective swine flu vaccination over time. The segmented IAV genome allows for the antigenic shift by reassortment of RNA segments from different viral strains, generating novel viruses [46]. The antigenic drift that is due to the gradual accumulation of mutations in the HA and NA surface antigens over time also contributes to the remarkable diversity of IAVs co-circulating among swine populations. This sequence-level diversity can impact the T cell response since even single amino acid modifications to T cell epitopes can reduce human leukocyte antigen (HLA) binding or T cell recognition, leading to viral escape and viral camouflage [150], [151] that contribute to a lower vaccine efficacy. An additional problem facing swine influenza vaccine developers is the high diversity of circulating IAV genotypes impacting individual pork farms each year [17], making it difficult to know whether a given commercial vaccine will be protective.

To address these challenges in pigs, we applied a previously developed computational method for estimating the degree of epitope conservation between vaccines and outbreak strains. Rather than focus on sequence identity, this algorithm identifies individual epitopes, and searches for epitope pairs that share MHC-binding properties and have identical TCR-facing residues, while allowing for amino acid variability at the T cell agretope (the HLA-binding-pocket facing amino acid residues). For swine IAV, the first step is to use SLA prediction matrices (PigMatrix) [152], and once an SLA-binding T cell epitope is predicted, JanusMatrix is applied to isolate the TCR-facing residues for comparison with similar SLA-binding epitopes in other circulating influenza strains [114], [129]. The third step is to use the Epitope Content Comparison (EpiCC) tool, which compiles the similarities and differences between the T cell epitopes in the vaccine and the circulating strain, assigning a score that reflects the degree of conservation [124].

Here, we compare a computationally designed swine flu vaccine based on conserved T cell epitopes called multi-epitope vaccine (MEpiV) with commercially available inactivated full strain vaccines using the computational algorithms described above. The MEpiV is composed of immunoinformatic-identified conserved SLA class I and class II epitopes assembled head-to-tail as class I and class II poly-epitope genes and formulated for delivery in a DNA vaccine vector [123]. The MEpiV was previously shown to be protective in a heterologous prime-boost vaccination and challenge study when combined with the whole-inactivated vaccine [153]. In this study, we determine if the conserved T cell epitope-based vaccine would maintain conservation with circulating strain T cell epitopes over time.

To evaluate the T cell epitope conservation for the MEpiV vaccine and to compare the conservation of the epitopes selected in 2013 to circulating strains for subsequent years, we used the HA sequence of seasonal inactivated swine flu vaccines as a benchmark for comparison. Then,

we applied EpiCC and demonstrated that the MEpiV vaccine designed using computational tools in 2013 maintains >50% conservation with circulating strains over a 5-year period. As can be expected, EpiCC also indicated that T cell epitopes in commercial seasonal vaccines are less well conserved over the same time period. This evaluation of vaccines using EpiCC shows the approach to understanding T cell epitope conservation and the utility of the tool for comparing vaccines against emerging influenza strains. The analysis also reinforces the utility of designing influenza vaccines based on highly conserved epitopes from circulating viral strains, as these epitopes may be conserved over time.

Materials and Methods

Datasets

The sequences of all available H1N1, H1N2, and H3N2 swine IAV genomes circulating in the United States during the 5-year period 2013–2017 were obtained from the NIAID Influenza Research Database (https://www.fludb.org/, accessed on 29 April 2021) [154]. All of the genome sequences were downloaded and pre-processed to remove partial and duplicated sequences. The final set of 1272 whole genome sequences were translated into protein sequences and were compared to the epitope-based DNA vaccine, MEpiV (Table 3.1) using an immunoinformatic approach as described below. In order to further evaluate the conservation of MEpiV, the sequence from two standard inactivated swine IAV vaccine antigens (FluSureXP 2016) were included for comparison. HA sequences of inactivated swine IAV vaccine comprised of one H1N1, one H1N2, and two H3N2 strains were provided by Zoetis to facilitate the comparison of T cell epitope conservation in the epitope-based and inactivated virus vaccine, applying the same immunoinformatic analysis pipeline. The level of conservation for each of the vaccines (HA from

H1N1 to H3N2, the MEpiV as compared to a representative set of IAV strains for each year was measured relative to 2013 to obtain relative changes in the number of T cell epitopes that were conserved.

Immunoinformatic Tools

Three separate algorithms were used to evaluate the conservation of vaccine epitopes contained in the target vaccine against the complete set of swine IAV sequences: (1) PigMatrix, which defines T cell epitopes for swine class I and class II epitopes, (2) JanusMatrix (JMX), a tool for identifying epitopes that can be compared between strains by looking for epitopes that bind to the same allele and have conserved TCR facing residues which can be used to compare strains, and (3) EpiCC, the T cell epitope content comparison algorithm utilizes results generated from PigMatrix and JMX and produces an overall score for class I and/or class II epitopes on a whole antigen level to enable pairwise comparisons between circulating IAV and vaccine strains (See Figure 3.1). A total of 1272 pairwise comparisons were performed, comparing each 9-mer sequence for a possible conservation of SLA binding and TCR face, between MEpiV vaccine and circulating strains. EpiCC examines all of the epitopes in a vaccine against all of the epitopes in a given strain and produces an overall score for all class I or class II epitopes for each strain sequence. In addition to EpiCC, we used JanusMatrix to perform the same comparison on an epitope-by-epitope basis for the 28 class I and 20 class II epitopes in the vaccine.

T Cell Epitope Prediction Using PigMatrix

Using the pocket profile method and well-defined EpiMatrix binding preferences for human MHC pockets, we developed PigMatrix prediction matrices as previously described [114],

[152]. Matrices were designed based on the binding preferences of the best-matched human leukocyte antigen (HLA) pocket for each SLA pocket. The contact residues involved in the binding pockets were defined from crystal structures of SLA or HLA supertype alleles for class I and II, respectively. The allele selection was based on prior data indicating their prevalence in outbred swine populations [155], [156], and frequencies determined using low-resolution haplotyping in a small number of pigs [123]. For low-resolution SLA-typing results where haplotype associations were not possible, XX01 alleles were selected. Matrices were constructed for SLA alleles with HLA binding pocket similarities above 85% to predict T cell epitope binding to class I (SLA-1*01:01, 1*04:01, 1*08:01, 1*12:01, 1*13:01, 2*01:01, 2*04:01, 2*05:01, 2*12:01, 3*04:01, 3*05:01, 3*06:01, 3*07:01) and class II (SLA-DRB1*01:01, 02:01, 04:01, 04:02, 06:01, 06:02, 07:01, and 10:01) SLA alleles. PigMatrix raw scores were standardized to Z-scores to compare potential epitopes across multiple SLA alleles. Peptides with Z-scores ≥1.64 (the top 5% of any given sample of 9-mers) were identified as likely to be SLA ligands.

Identification of Conserved Vaccine Epitopes in Different Circulating Swine IAV Subtypes

JanusMatrix (JMX) is another immunoinformatic algorithm, which was incorporated to prospectively identify conserved vaccine epitopes among prevalent swine IAV [129]. JMX is used to facilitate the epitope to an epitope-based comparison between swine IAV protein sequences and the vaccine strain. Conserved peptides at the TCR-face were searched against all the circulating strains and hence the presence of these peptides can be identified when there are matches in each individual strain.

T Cell Epitope Content Comparison (EpiCC) Analysis

In order to determine the conservation of the T cell epitopes in the MEpiV among the three-circulating swine IAV subtypes, we applied EpiCC to facilitate the pairwise comparison of protein sequences [124]. This method of comparison is based on an immunological property expressed in terms of T cell epitope content which incorporated JMX computation, rather than sequence identity. Shared (conserved) T cell epitopes between the vaccine target and the circulating swine IAV strains were evaluated. The assumption was based on the fact that given epitopes *i* and *j* from different strains (the circulating strain, *s* and the vaccine strain, *v*), cross-reactive memory T cells can be activated by epitopes with identical TCR-facing residues (TCR*f*) that bind to the same alleles. The potential cross-reactive of class I epitope is calculated by considering identical residues at positions 4, 5, 6, 7, and 8 and for class II, the calculation is taken into account by identical residues at positions 2, 3, 5, 7, and 8. Therefore, the probability to induce the cross T-cell immunity is computed based on the following equation and p stands for the probability for epitope binding to the SLA allele:

$$S(i,j)_a = p(i)_a \cdot p(j)_a$$

By applying the above equation, we further computed the shared T cell epitope content (conserved) between two strains *s* and *v*. The sum of shared epitope scores of each *i, j* was normalized by the total number of compared pairs, p, and by the number of SLA alleles in A. This is to account for different epitope densities and for comparison of values of E determined using different numbers of SLA alleles. Therefore, the normalized shared EpiCC score, (termed as EpiCC score), can be computed by applying the following equation:

$$E(shared)_A = \frac{1}{|p| \cdot |A|} \sum_{i \in s, j \in v} \sum_{a \in A} S(i,j)_a$$

Area under the Curve (AUC) Computation

Given the fact that the complexity of multiple comparisons was done according to years and subtypes, the AUC calculation is applied to represent the T cell epitope conservation of a subtype in a year. EpiCC scores that were calculated for MEpiV and swine IAV sequences were plotted in a radar form (a line plot that is on circular orientation). The area under the radar curve (a numerical integral) was computed by combining spline interpolation and integration with the formula shown below:

$$AUC = \int_{a}^{b} f(x)dx$$

The higher the AUC value, the more T cell epitopes against MEpiV were conserved. Normalization of AUC values with respect to the baseline score of MEpiV vaccine was performed as the number of sequences varied across the years. This enabled a direct comparison of the epitope content conservation across the years.

Phylogenetic Analysis

The T cell epitope conservation was mapped onto a phylogenetic tree to correlate the T cell epitope conservation with a genetic evolution of swine IAV. Phylogenetic trees inferred from the maximum likelihood (ML) were constructed based on the HA protein (H1 and H3 subtypes) of circulating swine IAV strains with RAxML.v8 using the GTR-GAMMA nucleotide substitution model. Both phylogeny trees were rooted with midpoint. MEpiV vaccine epitopes were evaluated against H1 and H3 tree tips using the ggtree package version 2.2.4 in R [157].

Results

Swine IAV Dataset from 2013 to 2017

The goal of this study was to determine whether a vaccine designed in 2013 may continue to provide CMI boosting as was illustrated in 2019 [144]. The MEpiV vaccine contains 28 class I and 20 class II T cell epitopes and was produced as a plasmid DNA vaccine and tested in 2015 [141]. Circulating swine IAV whole genome sequences of three major subtypes (H1N1, H1N2, and H3N2) from 2013 to 2017 were computationally screened in the same stepwise process to evaluate their T cell epitope content in an epitope-to-epitope comparison to circulating strains (Figure 3.2). A total of 1272 whole genome swine influenza A sequences were analyzed, comprising 409 (32.2%) H1N1, 388 (30.5%) H1N2, and 475 (37.3%) H3N2 sequences. The highest number of sequences available was for 2016 (407 sequences; 32.0% of the total), while the lowest number was for 2014 (133 sequences; 10.5% of the total).

T Cell Epitope Content Comparison (EpiCC) of Swine MEpiV Vaccine against H1N1, H1N2, and H3N2 Circulating Swine IAV

In order to determine the conservation of MEpiV vaccine epitopes among the three circulating swine IAV subtypes, we applied EpiCC to facilitate a pairwise comparison of protein sequences. This sequence comparison method is based on an immunological property, potential T cell immunogenicity, rather than sequence identity. Shared (conserved) T cell epitopes between the vaccine target and the circulating swine IAV strains were assessed.

Higher EpiCC scores are thought to be associated with greater protection by vaccines against challenge strains [124]. For MEpiV vaccine class I epitopes, the highest EpiCC score is found for H1N1 swine IAVs (EpiCC score of 0.0256 with 98.5% conservation when normalized

to the MEpiV baseline), and the lowest for H3N2 (EpiCC score of 0.0100 with 38.5% conservation when normalized to the MEpiV baseline). Interestingly, on average, EpiCC scores of MEpiV vaccine class II epitopes for all subtypes is 14.3% higher than scores of class I epitopes. The average range difference (in percentage) of class II EpiCC scores for all three subtypes is 25.6%, while for class I EpiCC scores it is 28.6%. The range difference of class II EpiCC scores is 11.7% smaller than the range difference of class I, indicating that the conservation of MEpiV class II epitopes was consistent in all of the circulating swine IAV subtypes that were analyzed. Detailed information for each of the circulating strains and their respective EpiCC scores are tabulated in Supplemental Table A-1.

While detailed lists of EpiCC scores are informative, we also used radar plots to visualize the EpiCC scores. Radar plots were constructed to describe the degree of conservation of MEpiV vaccine class I and II T cell epitopes in the three prevalent swine flu subtypes (Figure 3.3) and the area under the curve for the EpiCC scores (AUC, outlined in color in Figure 3.3) was used to quantify and compare the T cell epitope conservation between the vaccine and circulating swine IAV each year. As shown in Figure 3.3 and Supplemental Figure A-1, the AUC described by the EpiCC scores is greater for the MEpiV vaccine against H1N1, than the AUC for H3N2 and H1N2 circulating strains. Thus, the vaccine is predicted to be effective against all circulating H1N1 strains in 2013 to 2017. The MEpiV vaccine is predicted to drive a broad CD4 immune response based on data published by Gutierrez *et al* [123] and Hewitt *et al* [153].

Computing the AUC facilitates the qualitative comparison of the vaccine against circulating strains over time. As expected, when considering the MEpiV computer-designed vaccine epitopes, the overall EpiCC scores, compared to circulating viral strains, did not change very much over time. The overall conservation was maintained for all three viral subtypes,

although the total EpiCC scores were lower for H1N2 and H3N2 strains. Class II T cell epitopes were 80.8% more conserved on average, as compared to class I in all subtypes (Supplemental Figure A-1). We visualized these data on the individual antigen level in biaxial plots with the x-axis representing time and the y-axis representing the AUC for vaccine against circulating strains for that year (Figure 3.4). For the HA antigen, there was 79.5% conservation of MEpiV vaccine (both class I and II HA epitopes) in H1N1 over multiple years, whereas the HA epitope conservation in H1N2 and H3N2 were 51.7% and 8.6%, respectively.

The overall conservation of NA class I and II epitopes in H1N1 was 45.0%, while the conservation of vaccine epitopes in H1N2 and H3N2 strains was lower at 10.3% and 9.0%, respectively. The conservation in H1N2 and H3N2 for surface antigens was relatively low compared to H1N1, due to the complete lack of conservation (AUC of zero) for H3 and N2 epitopes in the MEpiV vaccine. Internal antigen epitopes were also well conserved across all subtypes (Supplemental Figure A-2), suggesting that internal proteins might contribute to vaccine efficacy. While the original MEpiV epitopes were selected from seven representative swine influenza strains, this finding suggests that vaccine epitopes that are highly conserved in one set of sequences for a given year may still be relevant and provide cross-protective immunity in the years that follow.

T Cell Epitope Conservation Analysis of Individual Epitopes Using JanusMatrix (JMX)

The EpiCC tool gives an overall score for the combined epitope content, rather than assessing and reporting on each epitope in a vaccine. Since the MEpiV is composed of distinct T cell epitopes, we wished to determine the conservation of each epitope over time, and therefore we performed an additional epitope-by-epitope analysis using JMX comparing the vaccine

epitopes with their homolog in circulating strains. In this case, JMX searches the circulating swine IAV strains for 9-mers with the same TCR-facing amino acids as those of the input class I and II MEpiV vaccine epitopes [129]. The JMX homology score was calculated for every input MEpiV vaccine epitope that appears "homologous" to a given TCR, even though there may be minor variations in the MHC binding residues, as long as the peptide would still be predicted to bind to the same MHC.

While performing the JMX analysis to compare vaccine epitopes to circulating strain epitopes was matched for binding to the same MHC and identical at the TCR-face, we were able to identify the specific TCR-homologous 9-mers in circulating swine IAV strains. We applied JMX homology scores to further examine the level of conservation of individual T cell epitopes in every swine IAV subtype and quantify the overall conservation (Table 3.2).

Doing so, we were able to identify the most highly conserved T cell epitopes. Among 28 class I peptides, 16 of the peptides were more than 80% conserved in the three-circulating swine IAV subtypes throughout the 5-year period (Table 3.2A). Only two surface epitopes from NA were conserved and were N1-specific. Most of the highly conserved peptides were from internal antigens: PB2 (GTEKLTITY), PB1 (VSDGGPNLY, DTVNRTHQY), PA (QVSRPMFLY), NP (AFDERRNKY, CTELKLSDY, ASQGTKRSY, KSCINRCFY, DTVHDRTPY), and M1 (SLLTEVETY, LTEVETYVL, DLLENLQAY, LASCMGLIY, LASCMGLIY, NTDLEALME). The two most conserved peptides were SLLTEVETY and LTEVETYVL from the M1 protein. These peptides were found in all of the 1272 IAV strains. Interestingly, the least conserved peptides (GAKEVALSY and NMDKAVKLY) are also from M1, with conservation less than 3% in all of the subtypes and only being observed in 2013.

In addition, 10 out of 20 class II peptides were highly conserved (>80%) in circulating swine IAV strains (Table 3.2B). None of these highly conserved peptides were found in HA and NA, rather they were found in internal antigens such as PB1 (MMGMFNMLSTVLGVSI, YRYGFVANFSMELPSFGVSG), PA (EVHIYYLEKANKIKSEKTHIHIF, RSKFLLMDALKLSIEDP), NP (IEDLIFLARSALILRGSVAHKSCLP), M1 (TRQMVHAMRTIGTHPSSSA, TYVLSIIPSGPLKAEIAQRLESV, SCMGLIYNRMGTVTTEAAFGLVC), and NS2 (FEQITFMQALQLLLEVE, FQDILMRMSKMQLGSSSE). This suggests that epitopes from the internal antigens are well-conserved across strains and over time may contribute to vaccine efficacy.

Then, we used this epitope matching information generated from the JMX analysis jointly with HA phylogeny trees to visualize the distribution of MEpiV vaccine class I and II epitopes (Figure 3.5). Epitopes from both classes were well conserved in most internal proteins, as indicated by the presence of small bars adjacent to the tips of the respective HA phylogeny tree. Epitopes in the external proteins such as HA and NA are subtype-specific, demonstrating that the MEpiV vaccine consists of H1, N1, and N2-specific epitopes. A big blank under HA for the H3 phylogeny tree shows almost an absence of H3 epitopes in the MEpiV vaccine. Interestingly, although there are subtype-specific epitopes, we would expect that H1N1 and H1N2 IAV strains have a shared conservation in HA epitopes, however, H1-specific epitopes are only found conserved in 47 H1N2 swine IAV strains that are of the same clade as H1N1 IAV strains. Six out of eight class I and half of four class II HA epitopes were absent in the H1N2 swine IAV subtype.

Strains Identification for Conserved Peptides

The epitope to epitope-based comparison can also be used to identify strains that have the most or the least conserved T cell epitopes (Supplemental Table A-2), which may be important when selecting strains for a recombinant or inactivated whole antigen vaccine. Forty-four IAV sequences were shown to be highly conserved against the MEpiV prototype vaccine, with conservation at 75%. The majority of the sequences (42/44) belong to the H1N1 subtype, while two belong to the H1N2 subtype. In contrast, four swine IAV sequences had very few epitopes conserved with the prototype vaccine (46.4%); all of these strains were H3N2 subtypes. This is expected as most of the T cell epitopes included in MEpiV were HA H1-specific, and conservation across subtypes is not optimal, indicating that truly universal vaccines must include epitopes from more than one subtype.

This study demonstrates how EpiCC and JMX can be applied in complement for surveillance and analysis of epitope evolution and/or escape. One of the direct applications of the EpiCC program is to enable the selection of challenge IAV strains for vaccine studies. Furthermore, this work also serves a retrospective analysis that provides a baseline strain coverage estimate for MEpiV but can easily be applied to other (new or old) vaccines against large numbers of new viruses.

Comparison of MEpiV and Commercial Swine Flu Vaccine

Immunity induced by inactivated virus vaccines usually wanes over time when it is no longer a close match to the circulating strains. To further investigate whether the T cell epitope conservation in a vaccine that was computationally designed to contain such epitopes was advantageous as compared to commercial swine flu vaccines, we compared the AUC computed

from the EpiCC analysis for HA antigens of the MEpiV vaccine and the HA found in a commercial vaccine which comprises four HA vaccine strains of the major swine IAV seasonal subtypes, one H1N1, one H1N2, and two H3N2. The H1 components of these commercial vaccines were included since 2011, while the H3 components were introduced in 2016.

A year-to-year comparison was made relative to 2013 for HA antigens of all the vaccine strains except for the H3 components of the commercial vaccine strains that were introduced in 2016. Changes in the conservation of the vaccines against the baseline year were calculated as a ratio, meaning that a score of 1.00 would indicate no change in the T cell epitope conservation (in AUC values), greater than 1.00 indicates an increasing T cell epitope conservation relative to 2013, and a ratio less than 1.00 implies a loss of T cell epitope conservation. The ratio of T cell epitope content (class I and II) for MEpiV over time, remains consistent or increases (except for H1N2) (Figure 3.6). Specifically, the H3 HA class I epitopes in the MEpiV vaccine showed a gradual increase of conservation in circulating swine IAV strains. In contrast, the ratio of conservation for the H3 conventional vaccine strains (FSXP.NC and FSXP.MN) decline over time. The same trend for FSXP.NC and FSXP.MN were observed in class II, however, there was no change in the class II epitope conservation for the MEpiV vaccine, as there were no H3-specific class II epitopes selected for the MEpiV vaccine. This result is consistent with the EpiCC and JMX analyses shown above.

Discussion

In general, vaccine efficacy assessment methods are lacking for swine IAV. More specifically, in lieu of challenge studies, there is no method available for evaluating new vaccines against circulating strains for cross-protection by T cell epitopes. Here, we used the EpiCC tool to

approximate the potential T cell epitope cross-protection between MEpiV and circulating strains. In previously published studies, we established a threshold of cross-conservative epitope protection, using EpiCC to compare one vaccine against IAV strains circulating in 1 year [124]. We have also demonstrated the utility of EpiCC tool applied for another pathogen, Porcine circovirus 2 (PCV2) in a study evaluating multiple vaccines against circulating PCV2 strains [158]. In this study, we demonstrate how EpiCC can be used for the longitudinal analysis against evolving strains circulating in swine populations.

The current analysis applies the EpiCC tool to a computationally designed T cell epitope vaccine and compares the vaccine with circulating strains over a 5-year period. Having established the longitudinal conservation of the H1N1 T cell epitopes in the subunit vaccine, we then compared the 5-year trajectory of the epitope vaccine with that of a typical commercial swine IAV vaccine. The MEpiV retained conservation of T cell epitope content over time. This was especially true for seven T cell epitopes that were previously confirmed as immunogenic in a previous study [123]. In contrast with MEpiV, the antigenic 'drift' was evident for the commercial vaccine, resulting in lower EpiCC scores for the epitopes contained in the HA antigen over time, as expected. Consistency of the area under the curve (AUC) over years (for the MEpiV) suggests that the T cell epitopes in the prototype vaccine could reliably drive robust immune responses in swine regardless of the drift, and that a conserved epitope-driven vaccine may be a valuable adjunct to vaccination with whole, inactivated seasonal vaccine, as was shown by Hewitt *et al* [153].

Comparing the T cell epitope conservation can contribute to assessing the projected efficacy of a vaccine. This study illustrates how EpiCC might be applied to evaluate several different vaccines, and to select the best vaccine strain (based on the T cell epitope conservation)

for any given year. This is as relevant for IAV as it may be relevant for other emerging viruses such as COVID-19.

The analysis also demonstrates the use of JMX, a novel tool that searches for conserved T cell epitopes using TCR facing residues. JMX may make more accurate comparisons between T cell epitopes contained in vaccines as compared to circulating strains over time. By quantifying the conservation using JMX, we are also able to examine which T cell epitopes are conserved in which strains of IAV. This type of analysis may be useful for the selection of challenge strains in vaccine studies. Not surprisingly, epitopes from M1 and PB1 proteins were better conserved with circulating strain epitopes over the 5-year period studied in this example, and as expected, epitopes from HA and NA protein were much less conserved.

Compared to the commercial whole antigen killed vaccine, MEpiV T cell epitopes were highly conserved over time. This finding is particularly relevant for influenza, since cross-reactive antibodies may not be present when influenza strains shift, rather than drift [159]. Experts in the field have advocated for the development of 'universal influenza vaccines' that can boost immune responses in the absence of antibody cross-reactivity for this reason. The fact that lower levels of conservation were observed for H1N2 and H3N2 over time suggests that conserved epitope-based vaccines should be designed for each IAV subtype. We have explored the use of MEpiV-type vaccines given by the heterologous prime-boost with a commercial swine influenza vaccine (which contains a whole HA antigen) and found increased immunogenicity by priming with the MEpiV vaccine over the homologous commercial vaccine prime-boost, an equivalent body temperature control 1 day after the pH1N1 challenge, and reduced lung lesions and influenza antigen, as illustrated by Hewitt *et al* [153]. Reducing the overall viral burden and increasing the average daily gain, distributed across large populations of swine, may prove cost-effective for pork producers.

One application of ‘universal’ T cell epitope-based vaccines being explored in humans is to combine them with seasonal vaccines, a topic which might also be of interest to the animal health community [160]–[162].

Moreover, we note that the SLA alleles selected for this study were reported as prevalent in outbred swine populations [155], [156] and on low-resolution haplotyping results in a small number of pigs [123]. We considered this set of alleles a first proxy for commonly expressed alleles. However, these alleles might not represent the complete SLA diversity or the most prevalent alleles in the US swine outbreed population. While EpiCC scores might be different, T cell epitope predictions for highly prevalent haplotypes that represent a large percentage of the US swine population will likely produce more relevant results. Systematic studies to investigate the distribution of SLA haplotypes in outbred populations of pigs in the US will have a significant impact on our ability to develop prediction models for a more comprehensive set of SLA alleles.

Table 3.1. MEpiV vaccine class I and II peptides.

Antigen	Class I peptide	Class II peptide
PB2	GTEKLTITY	-
PB1	VSDGGPNLY	MMGMFNMLSTVLGVSI
	DTVNRTHQY	YRYGFVANFSMELPSFGVSG
PA	QVSRPMFLY	EVHIYYLEKANKIKSEKTHIHIF
		RSKFLLMDALKLSIEDP
HA	GMVDGWYGY	YEELREQLSSVSSFER
	GMIDGWYGY	STRIYQILAIYSTVASSLVLV
	SVKNGTYDY	GDKITFEATGNLVVPRY
	RIYQILAIY	VPRYAFAMERNAGSGIIIS
	NADTLCIGY	
	TSADQQSLY	
	LSTASSWSY	
	ITIGKCPKY	
NP	AFDERRNKY	IEDLIFLARSALILRGSVAHKSCLP
	CTELKLSDY	TRGVQIASNENVETMDSNTLELR
	ASQGTKRSY	IDPFKLLQNSQVVSLMRP
NA	KSCINRCFY	CRTFFLTQGALLNDKH
	DTVHDRTPY	SVVSVKLAGNSSLCPV
	GTIKDRSPY	NQTYVNISNTNFAAGQSVVSVKL
	EMNAPNYHY	MANLILQIGNIISIWISHS
	ELDAPNYHY	
	EICPKLAEY	
M1	SLLTEVETY	TRQMVHAMRTIGTHPSSSA
	LTEVETYVL	SCMGLIYNRMGTVTTEAAFGLVC
	DLLENLQAY	TYVLSIIPSGPLKAEIAQRLESV
	LASCMGLIY	
	NTDLEALME	
	NMDKAVKLY	
	GAKEVALSY	
NS2	-	FEQITFMQALQLLLEVE
		FQDILMRMSKMQLGSSSE

Table 3.2. Total class I (A) and class II (B) peptides found in circulating IAV strains over a 5-year period, sorted from the greatest to the lowest conservation. Antigens are sorted according to viral surface antigens (HA and NA), followed by internal antigens. Shaded rows represent MEpiV epitopes that show the conservation equal to or greater than 80% in swine circulating IAV strains.

(A)

No.	Antigen	Class I Epitopes	JMX Homology Score (% of Conservation)			Average Conservation (%)
			H1N1	H1N2	H3N2	
1	HA	GMVDGWYGY	401.7 (98.2)	384.8 (99.2)	240.7 (50.7)	79.0
2	HA	GMIDGWYGY	401.7 (98.2)	384.8 (99.2)	240.7 (50.7)	79.0
3	HA	SVKNGTYDY	402.8 (98.5)	308.0 (79.4)	0.5 (0.1)	9.2
4	HA	RIYQILAIY	392.8 (96.0)	57.0 (14.7)	0	37.6
5	HA	NADTLCIGY	375.0 (91.7)	22.0 (5.7)	0	22.9
6	HA	TSADQQSLY	352.0 (86.1)	17.0 (4.4)	0	19.5
7	HA	LSTASSWSY	306.5 (74.9)	16.5 (4.3)	0	17.9
8	HA	ITIGKCPKY	58.8 (14.4)	3.5 (0.9)	0	3.6
9	NA	KSCINRCFY	0	384.0 (99.0)	474.0 (99.8)	99.4
10	NA	DTVHDRTPY	0	371.3 (95.7)	468.0 (98.2)	96.9
11	NA	GTIKDRSPY	322.25 (78.8)	0	0	78.8
12	NA	EMNAPNYHY	337.29 (82.5)	0	0	82.5
13	NA	ELDAPNYHY	381.86 (93.4)	0	0	93.4
14	NA	EICPKLAEY	0	98.4 (25.4)	120.6 (25.4)	25.4
15	PB2	GTEKLTITY	405.7 (99.2)	379.7 (97.9)	454.7 (95.7)	97.6
16	PB1	VSDGGPNLY	408.4 (99.9)	386.2 (99.5)	472.0 (99.4)	99.6
17	PB1	DTVNRTHQY	409.0 (100.0)	386.7 (99.7)	468.0 (98.5)	99.4
18	PA	QVSRPMFLY	400.6 (97.9)	379.6 (97.8)	433.0 (91.2)	95.7
19	NP	AFDERRNKY	407.8 (99.7)	386.5 (99.6)	471.8 (99.3)	99.5
20	NP	CTELKLSDY	406.0 (99.3)	383.5 (98.8)	472.0 (99.4)	99.2
21	NP	ASQGTKRSY	400.0 (97.8)	369.0 (95.1)	464.0 (97.7)	96.9
22	M1	SLLTEVETY	409 (100.0)	388 (100.0)	475.0 (100.0)	100.0

23	M1	LTEVETYVL	409 (100.0)	388 (100.0)	475.0 (100.0)	100.0
24	M1	DLLENLQAY	407 (99.5)	387 (99.7)	468.0 (98.5)	99.2
25	M1	LASCMGLIY	399 (97.6)	388 (100.0)	473.0 (99.6)	99.1
26	M1	NTDLEALME	399 (97.6)	366 (94.3)	463.0 (97.5)	96.5
27	M1	NMDKAVKLY	11 (2.7)	10 (2.6)	16.0 (3.4)	2.9
28	M1	GAKEVALSY	12 (2.9)	9 (2.3)	13.0 (2.7)	2.6

(B)

No.	Antigen	Class II Epitopes	JMX Homology Score (% of Conservation)			Average Conservation (%)
			H1N1	H1N2	H3N2	
1	HA	YEELREQLSSVSSFER	392.6 (96.0)	365.3 (94.1)	0	63.4
2	HA	STRIYQILAIYSTVASSLVLV	393.1 (96.1)	253.4 (65.3)	0	53.8
3	HA	GDKITFEATGNLVVPRY	348.2 (85.1)	56.4 (14.5)	0	33.2
4	HA	VPRYAFAMERNAGSGIIIS	13.0 (3.2)	1.1 (0.3)	0	1.2
5	NA	CRTFFLTQGALLNDKH	408.4 (99.9)	0	0	33.3
6	NA	SVVSVKLAGNSSLCPV	102.9 (25.2)	0	0	8.4
7	NA	NQTYVNISNTNFAAGQSVVSVKL	66.0 (16.1)	0	0	5.4
8	NA	MANLILQIGNIISIWISHS	62.1 (15.2)	0	0	5.1
9	PB1	MMGMFNMLSTVLGVSI	409.0 (100.0)	387.4 (99.8)	474.9 (100.0)	99.9
10	PB1	YRYGFVANFSMELPSFGVSG	409.0 (100.0)	388.0 (100.0)	474.5 (100.0)	100.0
11	PA	EVHIYYLEKANKIKSEKTHIHIF	406.3 (99.3)	386.1 (99.5)	472.9 (99.6)	99.5
12	PA	RSKFLLMDALKLSIEDP	405.9 (99.2)	381.0 (98.2)	474.7 (99.9)	99.1
13	NP	IEDLIFLARSALILRGSVAHKSCLP	400.3 (97.9)	311.3 (80.2)	454.3 (95.6)	91.2
14	NP	TRGVQIASNENVETMDSNTLELR	346.5 (84.7)	243.8 (62.8)	268.5 (56.5)	68.0
15	NP	IDPFKLLQNSQVVSLMRP	343.4 (84.0)	270.6 (69.7)	296.9 (62.5)	72.1
16	M1	TRQMVHAMRTIGTHPSSSA	398.6 (97.5)	380.8 (98.1)	462.1 (97.3)	97.6
17	M1	SCMGLIYNRMGTVTTEAAFGLVC	399.3 (97.6)	382.0 (98.5)	462.7 (97.4)	97.8
18	M1	TYVLSIIPSGPLKAEIAQRLESV	395.3 (96.7)	367.2 (94.7)	465.5 (98.0)	96.5
19	NS2	FEQITFMQALQLLLEVE	407.6 (99.7)	384.9 (99.2)	466.6 (98.2)	99.0
20	NS2	FQDILMRMSKMQLGSSSE	364.9 (89.2)	327.2 (84.3)	358.9 (75.6)	83.0

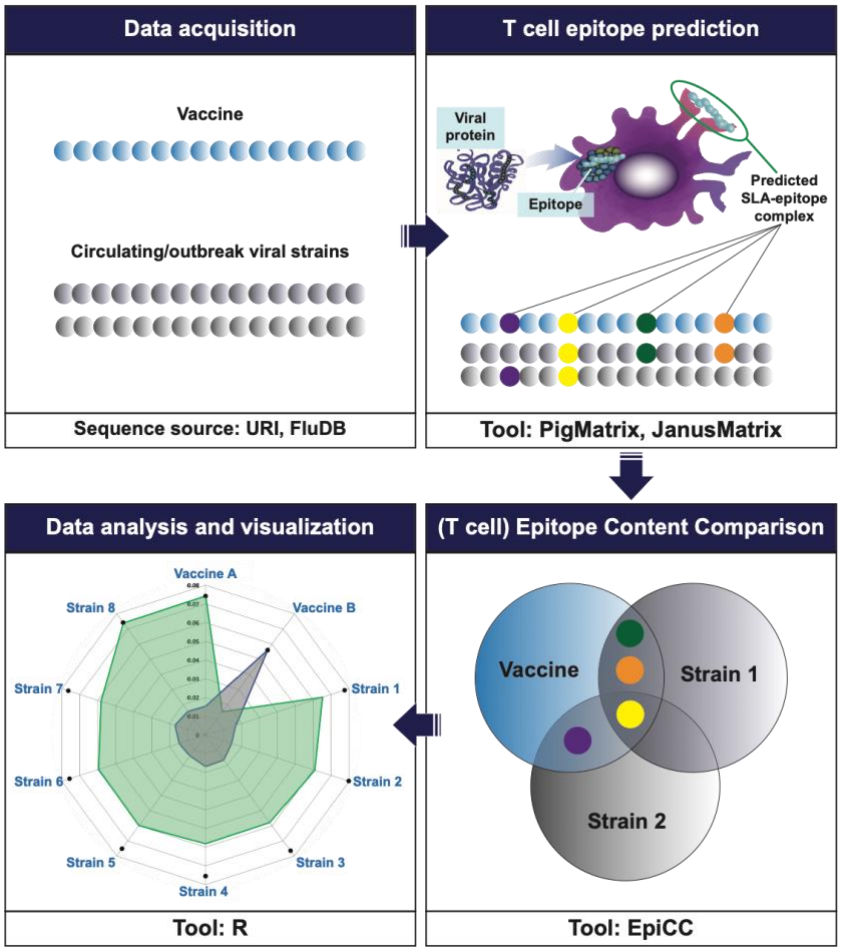


Figure 3.1. Workflow for the typical EpiCC analysis. The vaccine sequence of interest (here, MEpiV) and circulating pathogen strains (swine IAV, in this example) are retrieved and pre-processed prior to performing the EpiCC analysis. T cell epitopes are identified in the vaccine and circulating strains (colored beads) using EpiMatrix (for HLA restricted human T cell epitopes) or Pig-Matrix (for SLA-restricted T cell epitopes). Once all the epitopes are identified, a comparison and quantification of the T cell epitopes is performed using EpiCC. An overall EpiCC score (area under the curve) is calculated for the combined class I and II epitopes for each vaccine/strain compari-son. Greater AUC scores indicate higher numbers of conserved T cell epitopes. EpiCC scores can be compared and contrasted for the selected vaccines (here, MEpiV versus seasonal whole inactivated swine IAV vaccines).

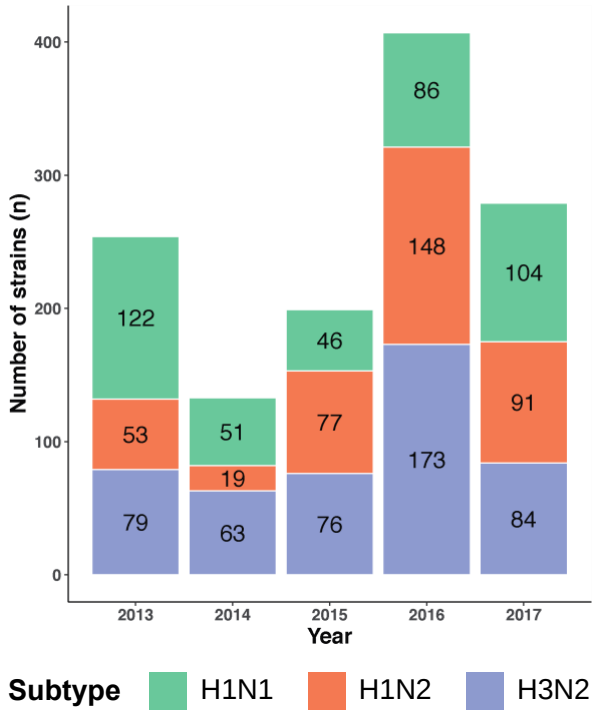


Figure 3.2. Swine IAV genome sequences from the H1N1, H1N2, and H3N2 subtypes from 2013–2017 included in this analysis. The color-coded stacked bar chart represents the three subtypes, each stacked component shows the number of strains per subtype for that year.

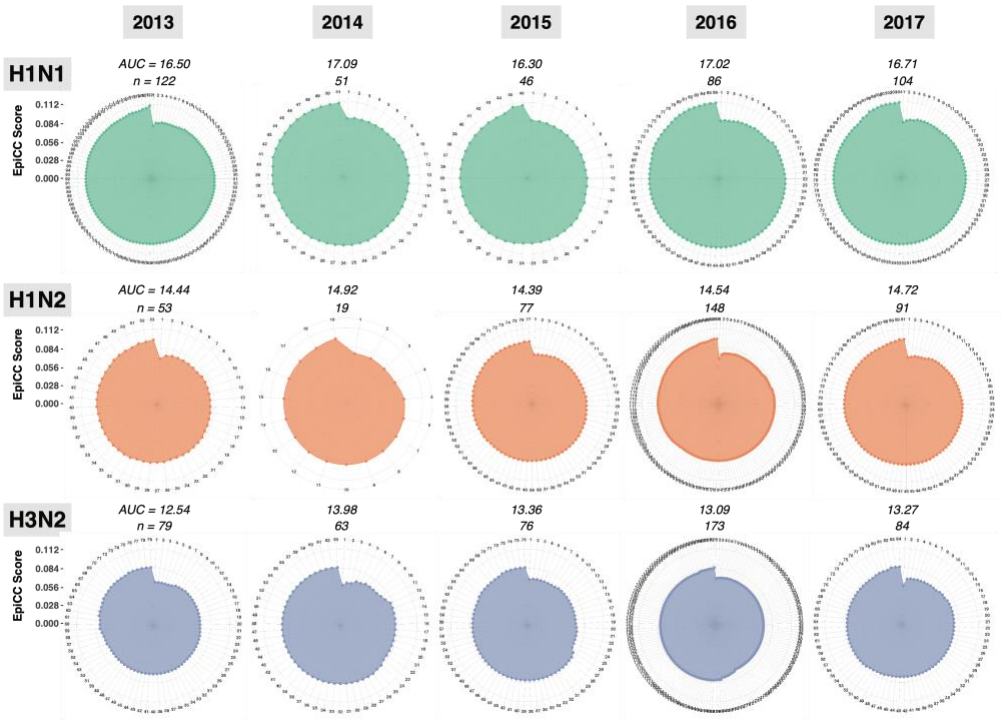
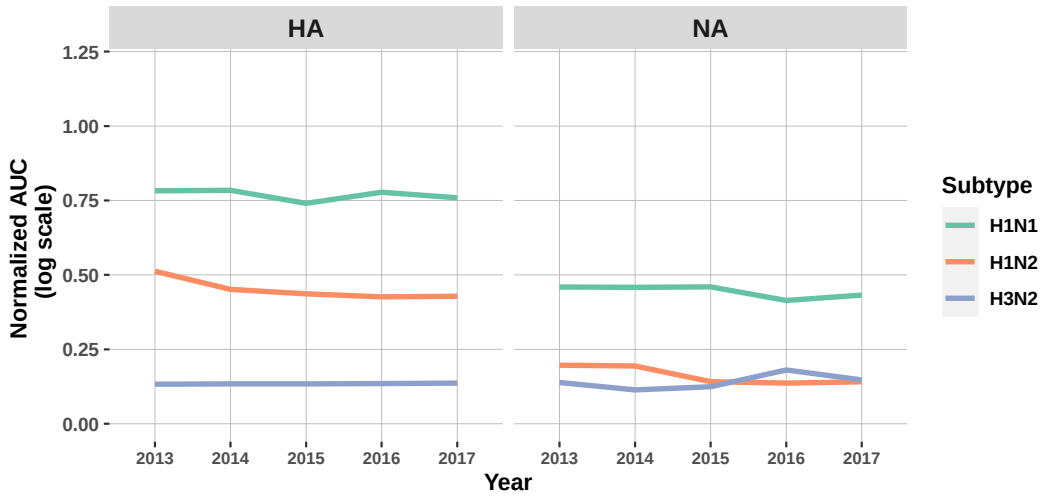
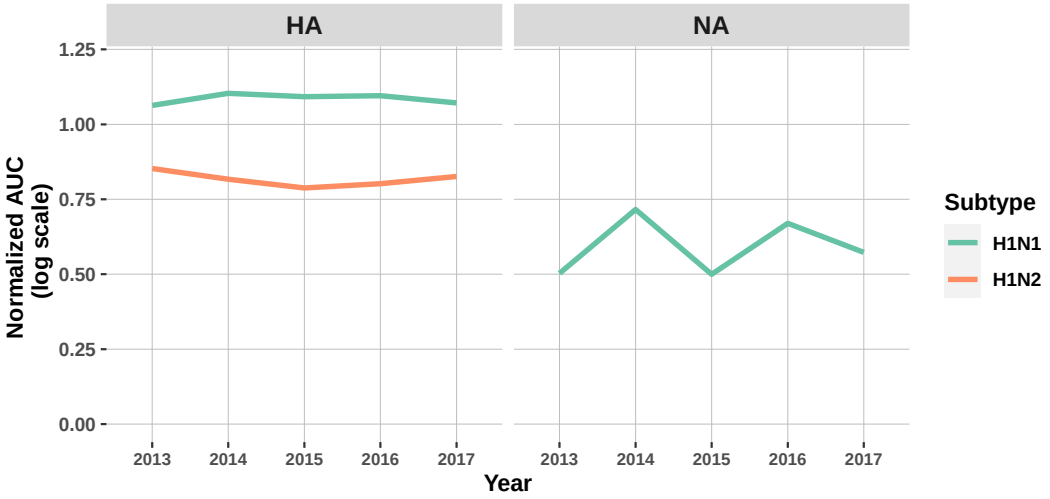


Figure 3.3. Radar plots enable the quantitative analysis of the degree of T cell epitope conservation between the conserved epitopes from all of the IAV proteins contained in a vaccine (here, MEpiV) and the epitopes from all of the IAV proteins contained in whole genome circulating strains for each year. The EpiCC score describing the T cell epitope conservation between the vaccine (MEpiV) against each swine IAV circulating strain is plotted on the radiating axes of radar plot for each year, for a period of 5 years, left to right. Circulating IAV strains were sorted from the lowest to the highest EpiCC scores. Radar plots for class II EpiCC scores are shown here and radar plots for class I are provided in supplemental data (Supplemental Figure A-1).



(A) Class I T cell epitopes of surface antigens HA and NA.



(B) Class II T cell epitopes HA and NA.

Figure 3.4. Line plots showing the normalized AUC for the comparison of MEpiV vaccine epitopes to epitopes found in circulating IAV strains for surface antigens HA and NA, by subtypes and by year for class I SLA (A) and class II SLA (B) epitopes. The AUC is shown on a normalized scale to enable the direct comparison by antigens, subtypes, years, and T cell epitopes classes. The lines that represent H1N2 (NA) and H3N2 (HA and NA) were removed in (B) as they showed no conservation. Similar line plots (different y-axis scaling) for the internal antigens are shown in Supplemental Figure A-2.

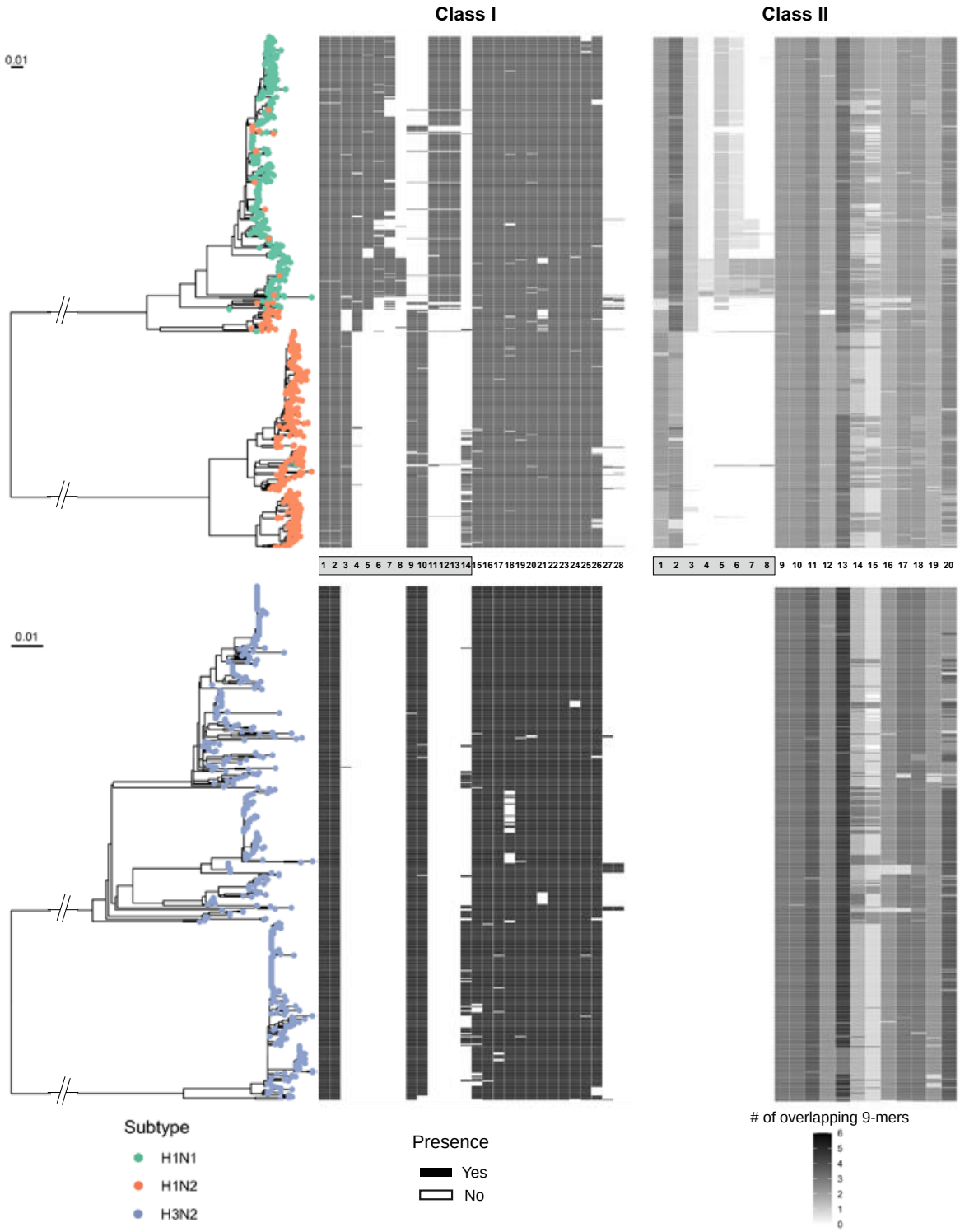


Figure 3.5. Phylogenetic tree of circulating swine H1 and H3 subtype IAV strains with predicted epitopes mapped to the tree tips. Class I and class II of MEpiV vaccine epitopes are shown in the heatmaps aligned with each associated strain. HA subtypes were color-coded. MEpiV

epitopes were listed in the central panel sorted by an external (grey box) and internal proteins arrangement (numberings refer to Table 3.1). The black and white bars mapped adjacent to the phylogeny tree show the presence or absence of respective MEpiV vaccine epitopes in these circulating swine IAV strains.

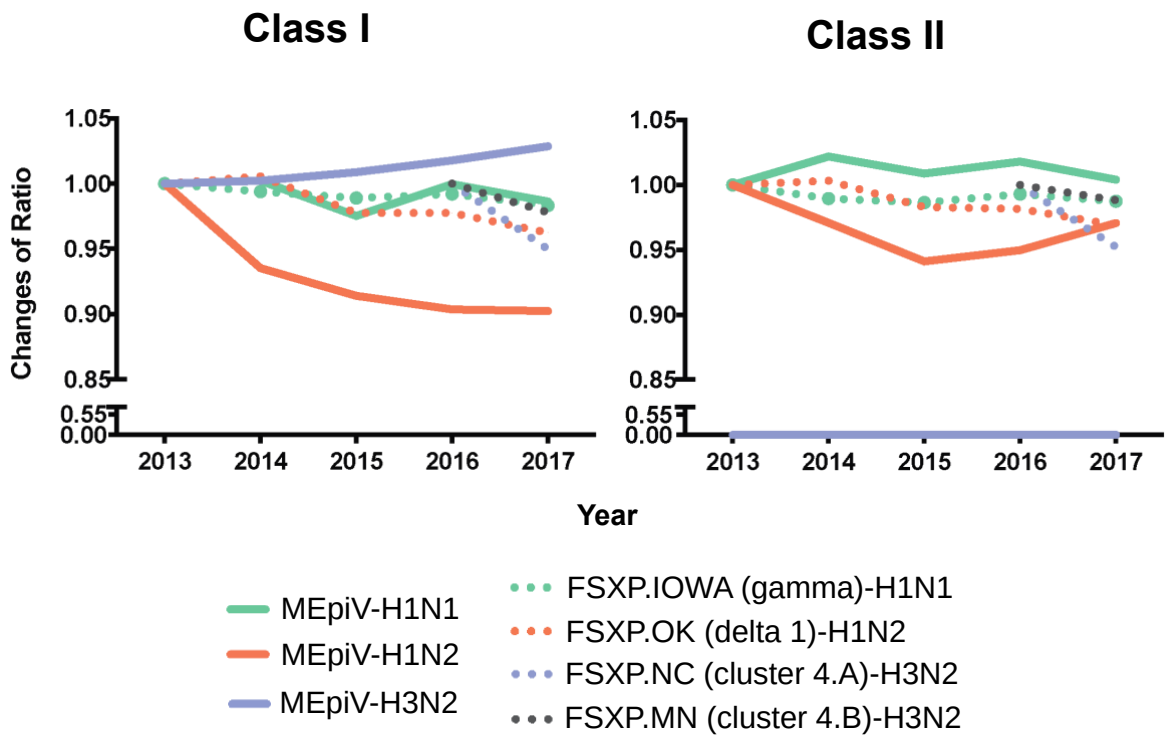


Figure 3.6. MEpiV compared to the commercial seasonal vaccine. The EpiCC analysis for the HA epitopes in MEpiV and HA from commercial (inactivated, whole) vaccines were calculated and then normalized to the EpiCC AUC deter-mined for circulating strains for the vaccine in 2013, to show changes in AUC over time. The solid lines represent the HA antigen of MEpiV from an epitope-based vaccine and the dotted lines illustrate the HA components from a commercial swine vaccine.

CHAPTER 4

H1N1 G4 SWINE INFLUENZA T CELL EPITOPE ANALYSIS IN SWINE AND HUMAN VACCINES AND CIRCULATING STRAINS UNCOVERS POTENTIAL RISK TO SWINE AND HUMANS [2]

[2]Tan, S; Moise, L; Pearce, DS; Kyriakis CS; Gutiérrez, AH; Ross, TM; Bahl, J; De Groot, AS. H1N1 G4 swine influenza T cell epitope analysis in swine and human vaccines and circulating strains uncovers potential risk to swine and humans. *Influenza Other Respi Viruses. 2022; 1- 15*. https://doi.org/10.1111/irv.13058. Reprinted here with permission of the publisher.

Abstract

Pandemic influenza viruses may emerge from animal reservoirs and spread among humans in the absence of cross-reactive antibodies in the human population. Immune response to highly conserved T cell epitopes in vaccines may still reduce morbidity and limit the spread of the new virus even when cross-protective antibody responses are lacking. We used an established epitope content prediction and comparison tool, Epitope Content Comparison (EpiCC), to assess the potential for emergent H1N1 G4 swine influenza A virus (G4) to impact swine and human populations. We identified and computed the total cross-conserved T cell epitope content in HA sequences of human seasonal and experimental influenza vaccines, swine influenza vaccines from Europe and the United States (US) against G4. The overall T cell epitope content of US commercial swine vaccines was poorly conserved with G4, with an average T cell epitope coverage of 35.7%. EpiCC scores for the comparison between current human influenza vaccines and circulating human influenza strains were also very low. In contrast, the T cell epitope coverage of a recent European swine influenza vaccine (HL03) was 65.8% against G4. Poor T cell epitope cross-conservation between emergent G4 and swine and human influenza vaccines in the US may enable G4 to spread in swine and spillover to human populations in the absence of protective antibody response. One European influenza vaccine, HL03, may protect against emergent G4. This study illustrates the use of the EpiCC tool for prospective assessment of existing vaccine strains against emergent viruses in swine and human populations.

Introduction

The annual occurrence of influenza outbreaks causes considerable morbidity and mortality and poses a global public health challenge. The causative agent of these outbreaks is influenza A virus (IAV), although influenza B also contributes to some outbreaks. IAV infects a wide range of hosts including mammals and avian species. Pigs are one of the most important hosts given their susceptibility to a broader range of avian and human influenza viruses, and they are often the source of novel reassortant viruses from avian- and human-origin strains. IAV infection of swine causes significant economic losses for commercial pork producers. In addition, swine IAV poses a threat to human health due to the potential for swine influenza A virus (swIAV) to spill over into the human population, as occurred in 2009 (for H1N1/pdm09) [163], [164].

After the outbreak of H1N1/pdm09 in human in 2009, the virus was reintroduced into pig herds around the world and reassorted with other swine influenza viruses, forming new swIAVs which spread within pig herds in the United States (US), Brazil, Europe (EU), Japan, China, and other countries [50], [165]. Some of these reassortants harboring H1N1/pdm09 internal genes have gradually evolved and replaced previous strains of swine influenza, demonstrating antigenic drift due to genetic selection pressure exerted on H1N1/pdm09 and other strains worldwide [51]. For example, multiple lineages of swIAV have been identified in pig herds in China, including classical swine H1N1, Eurasian avian-like (EA) H1N1, H1N1/pdm09, triple-reassortant (TR), and H3N2 lineages.

Influenza experts have been concerned about a new strain of swIAV, namely the G4 genotype (G4), that is now dominating other strains of IAVs in Chinese swine populations, as reported by Sun et al [52]. Emergent G4 is a reassortment product of EA H1N1 virus, bearing H1N1/pdm09 and TR-derived internal genes. In the reassorted G4, the HA and NA genes are from

the EA H1N1 lineage and in particular, the HA gene falls within the 1C.2.3 lineage. The viral ribonucleoprotein (vRNP) genes and M gene are from the H1N1/pdm09 lineage, and the non-structural (NS) gene is derived from the TR lineage [52]. Given that as many as 20% of pork industry workers in China have been found to be seropositive for G4 antibodies, it appears that G4 has the potential to cross species barriers [52].

Vaccine efficacy evaluation usually involves assessment of cross-reactive influenza-specific antibodies generated by exposure or vaccination. Seasonal vaccination in humans does not generate antibodies that protect against G4 (hemagglutination inhibition, see Figure 1C in Sun et al [52]). In a separate study, monoclonal antibodies isolated from mice immunized with pandemic (A/California/07/09) hemagglutinin (HA) and a novel flu vaccine, computationally optimized broadly reactive antigen (COBRA) P1 HA, generated hemagglutination of G4 virus-like particles in vitro [166]. The relevance of this murine study to swine and human populations remains to be determined.

When cross-reactive antibodies are not present, cross-conserved T cell epitopes in IAV vaccines and strains have been shown to play an important role in reducing morbidity and limiting the spread of IAV, even when vaccines and emergent strains are poorly matched [167]–[169]. There is strong evidence that (1) T cell responses generated by previous influenza exposure cross-reacts with novel IAV strains [3] and (2) T cell responses are critically important for protection against IAV infection in both humans and swine [147], [169], [170]. Thus, even in the absence of cross-reactive antibody to G4, T cell cross-reactivity might be protective.

Here, we apply an immunoinformatics tool to evaluate whether existing vaccines may have the capacity to prevent the spread of G4 in humans and swine. We developed a computational workflow that employs the Epitope Content Comparison (EpiCC) algorithm to measure the degree

of epitope conservation between target vaccines and outbreak strains. In previous studies, we identified an EpiCC score that was correlated with protection in the absence of cross-reactive antibody. We used the same approach to establish thresholds for protective efficacy for vaccines against circulating strains in this study. EpiCC's estimation of T cell epitope conservation between emerging viruses and vaccine strains may be useful as a potential surrogate measure of vaccine efficacy, in conjunction with other methods of pandemic risk assessment.

Methods

Sequences and data processing

Vaccine Strains

The H1 HA sequences of two EU swine influenza vaccine strains and five human seasonal influenza vaccine strains (from seasons 2008-2021) were obtained from the publicly available database, Global Initiative on Sharing Avian Influenza Data (GISAID EpiFlu; http://platform.gisaid.org/epi3/; accessed in August 2020) [171]. Strain information regarding commercial use EU swine influenza vaccine was based on literature review (Table 1A) [172], [173]. Sequences of the US swine influenza vaccines, namely FluSure(FS)XP/IA00, FSXP/OK08 and FSPandemic(FSPDM)/CA09, were provided by Zoetis. Experimental COBRA influenza vaccine sequences including swine (SW1 and SW2), human (X3 and X6) and a hybrid swine/human vaccine, P1, were provided by the Center for Vaccine and Immunology (CVI), University of Georgia (UGA). The COBRA SW1 and SW2 HA antigens were designed to be more cross-protective antigens using HA sequences from swine H1N1 and H1N2 sequences. Human COBRA X3 and X6 were designed using HA sequences from human isolates while COBRA P1 was derived from both swine and human H1 HA sequences [174].

G4 and Circulating Strains

Twenty-nine HA sequences from previously published swine H1N1 G4 genotype strains were used in this study [52]. Given that the 29 HA G4 strains were derived from a shorter time range (2016 – 2018) and that there is great similarity (percentage identity in the range of 95.8 – 100.0%) between these sequences, eight sequences were randomly selected for analysis in this study (Figure 1, Table 1B). All available H1 HA sequences comprising swIAV strains circulating in the US (1939-2020) and EU (1939-2018) and human IAV strains circulating in the US (from 2008-2010 and 2019-2020) were retrieved from GISAID.

Data Curation

Duplicate and partial sequences containing less than 1400 nucleotides were removed using a publicly available python script [175]. Phylogenetic analysis was performed following sequence alignment using MUSCLE 3.8.31 [176]. Maximum-likelihood phylogenetic trees were constructed with RAxML.v8 using the GTR-GAMMA nucleotide substitution model [177]. To ensure computational tractability and to preserve representative clades, Phylogenetic Diversity Analyzer (PDA) was used to subsample 150 sequences each from large dataset that consisted of swIAV strains circulating in the US and EU, respectively, as well as 300 sequences from human IAV strains circulating in the US [178]. The final reduced dataset (Supplementary Table 1A-C) was translated into amino acid sequences and combined with selected G4 strains and respective vaccine strains for three sets of analyses (Figure 1): (1) European circulating swine flu virus and G4 strains; (2) US circulating swine flu and G4 strains and (3) US circulating human influenza virus and G4 strains. To better classify strains according to their respective phylo-clusters, metadata such as H1 strain clade information was acquired using Swine H1 Clade Classification Tool [49].

T cell epitope binding prediction

We initiated our analysis by focusing on HA, given its importance as the critical antigen that is most relevant to protective immunity to influenza. Additional antigens were also evaluated (see section ***Analysis on other IAV antigens***). After compiling IAV sequences as described in Methods and illustrated in Figure 1, the translated HA protein sequences were screened using host-specific T cell epitope identification algorithms developed by EpiVax. Particularly, PigMatrix epitope prediction tools were used for the swine sequences, and EpiMatrix was used for identification of human epitopes. These prediction tools parse sequences into overlapping 9-mer frames to define the relative likelihood of binding to a set of prevalent swine leukocyte antigen (SLA) or supertype human leukocyte antigen (HLA) class I and II alleles [114], [125].

More specifically, using a position-specific scoring matrix, predicted binding scores were computed by PigMatrix for the likelihood of each 9-mer in the HA antigens of IAV to bind to a panel of prevalent class I (SLA-1*01:01, 1*04:01, 1*08:01, 1*12:01, 1*13:01, 2*01:01, 2*04:01, 2*05:01, 2*12:01, 3*04:01, 3*05:01, 3*06:01, 3*07:01) and class II (SLA-DRB1*01:01, 02:01, 04:01, 04:02, 06:01, 06:02, 07:01, and 10:01) SLA alleles. The SLA alleles selection for swine reflect dominant SLA types, each of which has binding pocket preferences that are shared with several other SLA alleles, and was based on a previous study reported by Gutierrez et al [123], [124]. Similarly, predicted binding scores for each 9-mer from human IAV sequences was assigned by EpiMatrix and the HLA alleles that were used in the prediction model include class I (A*01:01, A*02:01, A*03:02, A*24:02, B*07:02, and B*44:03) and class II (DRB1*01:01, DRB1*03:01, DRB1*04:01, DRB1*07:01, DRB1*08:01, DRB1*09:01, DRB1*11:01, DRB1*13:01, and DRB1*15:01) HLA molecules. These are HLA allele supertypes (alleles sharing common binding preferences) that cover the genetic diversity of more than 95% of human populations globally

[114], [121], [122].

For each 9-mer, i in each individual allele a of a set of MHC alleles A, PigMatrix or EpiMatrix raw scores, r, are normalized to Z-scores using the average μ and the standard deviation σ of scores calculated for 100,000 random 9-mers using the formula below [114], [125]. Nine-mers with Z scores greater or equal to 1.64, which comprise the top 5% in the normalized set of scores for each SLA or HLA allele of sequences, are predicted to have significant SLA or HLA binding potential. Higher Z-scores associate with higher MHC binding probability [124], [126].

$$Z(i)_a = \frac{(r - \mu)}{\sigma}$$

Analysis of T cell epitope content comparison (EpiCC)

We applied EpiCC to facilitate the pairwise T cell epitope content comparison of protein sequences (Figure 1). EpiCC enables large scale sequence analysis for conservation of T cell epitopes between swine and human flu vaccines and circulating IAV and G4 strains, focusing only on shared T cell epitopes between the vaccines and the target IAV strains [124], [179]. Once T cell epitope content is defined for each vaccine or strain, the set of conserved T cell epitopes that are shared between two strains can be enumerated.

In mathematical terms, EpiCC assesses the relatedness of T cell epitope, i, contained in a protein sequence of vaccine strain v and T cell epitope, j, contained in a protein sequence of a strain s based on respective PigMatrix SLA binding or EpiMatrix HLA binding score. Since cross-reactive memory T cells can be stimulated by epitopes (i, j) with identical TCR-facing residues (TCRf) that may have different HLA binding pocket residues, as long as they bind to the same alleles, we searched for potentially cross-reactive epitopes that shared TCRf as follows: Cross-conserved class I epitopes were defined by identical residues at positions 4, 5, 6, 7, and 8 and class

II epitopes were defined by identical residues at positions 2, 3, 5, 7, and 8. The score of cross-conserved T cell epitope shared between two strains s and v, was calculated using predicted binding probabilities as follows:

$$S(i,j)_a = p(i)_a \cdot p(j)_a$$

To normalize shared EpiCC score, the sum of shared epitope scores of each i, j was normalized by the total number of compared pairs, p, and by the number of MHC alleles in A. This is to account for different epitope densities and for comparison of values of E determined using different numbers of MHC alleles. Therefore, the normalized shared EpiCC score, (termed as EpiCC score), can be computed by applying the following equation:

$$E(shared)_A = \frac{1}{|p| \cdot |A|} \sum_{i \in s, j \in v} \sum_{a \in A} S(i,j)_a$$

Maximum EpiCC scores were calculated. These scores were derived from shared EpiCC scores computed from the comparison of any sequence to itself. The greater the maximum EpiCC score, the greater the total epitope content of the sequence. Since no sequence can be better matched to another sequence than itself, the maximum value for any comparison between any target sequence and a comparison sequence is always less than or equivalent to their maximum EpiCC scores. Both class I and class II EpiCC analyses were combined by summing class I and class II EpiCC scores (termed as total EpiCC score) for each vaccine-to-strain comparison.

When the shared T cell epitope content of a strain of influenza is highly related or "covered" by a given influenza vaccine sequence, the vaccine-to-strain's EpiCC score approaches the circulating strain's maximum EpiCC score (it approaches the maximum if nearly all the epitopes are identical, as defined above). To determine vaccine-to-strain EpiCC scores coverage, each vaccine-to-strain comparison was divided by that strain's maximum EpiCC score and expressed as a percentage. The greater the T cell epitope coverage (percentage), the better the

vaccine matches or covers the T cell epitope content of the circulating strain sequences (Figure 4.1).

EpiCC scores and EU vaccine efficacy estimation

To identify a threshold of protective efficacy of existing swine vaccines against circulating field strains, we extrapolated from available data, using an approach similar to one that we have already published [124]. For HA sequences, we calculated the EpiCC scores for three H1N1 EU commercial vaccines and three experimental monovalent vaccines against EU circulating swine IAV strains from the same period. Once we had obtained the scores, we identified the minimum EpiCC scores that correlated with protective endpoint results in four published studies that used the commercial and/or experimental vaccine strains. Vaccines were protective if they significantly reduced lung virus titers. The EpiCC score protective threshold was defined as the lowest EpiCC for at which the vaccine strain was shown to be protective. This is the main criterion for evaluating protection in the EU [173]. Scoring was performed independently of and prior to obtaining information about the outcomes of the vaccination and challenge studies.

Statistical analysis

A non-parametric Wilcoxon signed-rank test was used to compare T cell epitope coverage of different groups of vaccines (swine and human) analyzed against G4. P-values (p) less than 0.05 were deemed significant. The analysis was performed using the rstatix package in R version 4.0.3 [180].

Analysis on other IAV antigens

To determine whether other viral antigens might contribute to the protective efficacy of vaccines and circulating strains, we extended and applied similar workflow as described above to additional nine antigens other than HA. The nine antigens included neuraminidase (NA), polymerase basics (PB1 and PB2), polymerase acidic (PA), nucleoprotein (NP), matrix proteins (M1 and M2) and non-structural proteins (NS1 and NS2). The purpose of this analysis was to compare and assess the degree of T cell epitope relatedness between whole-killed vaccines strains, IAV circulating strains and G4 of non-HA antigens. Available whole proteome sequences of EU swine flu vaccines, human seasonal influenza vaccine strains (Table 4.1A) and the eight G4 sequences (Table 4.1B) were obtained from GISAID (accessed in December 2021). Sixteen clade-specific swine H1 IAV strains circulating in EU and five human H1 IAV strains circulating in the US were randomly selected from the previously defined dataset for analysis (Supplemental Table B-1 and B-3).

Results

We set out to evaluate the potential for cross-conserved T cell immune responses to protect against G4 in swine and human populations using immunoinformatics methods. The results of this analysis are divided into two parts due to species-specific MHC binding preferences (swine and human) and the species-specific circulation patterns of influenza strains. The first part of the analysis focused on protection by swine vaccines against G4 in swine, and the second part focused on predicting protection against G4 influenza in case of spillover into human hosts.

For the swine IAV analysis, we evaluated two commercial H1N1 swine influenza strains used in the EU and three US commercial H1N1 vaccine strains against strains from their respective

regions of the world. We also evaluated the experimental (swine) COBRA vaccines that were previously studied for cross-reactive antibodies, for T cell epitope conservation against G4 using swine epitope prediction tools (PigMatrix and EpiCC). In the second section, we evaluated human T cell epitope content relatedness of five seasonal (human) H1N1 IAV vaccine strains and three experimental (human COBRA) IAV vaccine strains in the US human population to determine their potential to control G4 in the event of G4 spillover into the US human population, using EpiMatrix and EpiCC.

As described above, T cell epitope content relatedness was defined as the density of shared T cell epitope content between the vaccine of interest and targeted strain (EpiCC score). A percent coverage was used to normalize score and permit comparison to established protective thresholds. Figure 4.1 illustrates how radar plots are constructed; the results of the swine and human vaccine-to-strain analyses are shown in separate radar plots (Figure 4.2 – 4.4), combining MHC class I and II comparisons (SLA for swine; HLA for human).

In the sections below, we described the results of the EpiCC analysis for HA, as it is the principal target of protective immune responses in influenza infection and is the most variable sequence in the pathogen [181]. Changes to vaccine composition and changes in vaccine efficacy are primarily due to drift and shift in the sequence of HA antigen. In section **T cell epitope conservation among antigens other than HA**, we provided results for additional influenza proteins, to allow for a more comprehensive understanding of their potential to contribute to changes in T cell epitope content.

Swine vaccine-to-strain EpiCC analysis results

EpiCC analysis of US vaccines and strains

EpiCC scores comparing the T cell epitope content of the HA of three commercial swine H1N1 vaccines used in the US (FSXP/IA00, FSXP/OK08 and FSPDM/CA09) to 150 swIAV strains circulating in the US from 1939 to 2020 and eight G4 sequences were calculated (Figure 4.1A). Overall, for these swine vaccine-to-strain comparisons tended to be clade-specific. Higher total EpiCC scores (greater cross-conservation) were observed for FSXP/IA00 when it was compared to US swIAV circulating strains. The average of vaccine T cell epitope coverage was 69.4% for this comparison. In contrast, the average T cell epitope coverage by FSXP/OK08 and FSPDM/CA09 of US swIAV strains was lower, at 43.2% and 59.9%, respectively. FSXP/IA00 had high T cell epitope relatedness to swIAV circulating strains from eight clades (Alpha, Beta, Gamma-PDM-09-like, Gamma2, Gamma2-beta-like and PDM-09) although Delta-like and Delta2 EpiCC scores were lower. FSXP/OK08 (HA from H1N2), only exhibited T cell epitope content shared with swIAV Delta-like and Delta2 strains. The third vaccine strain, FSPDM/CA09, which has been in use since 2009, demonstrated high T cell epitope relatedness to swIAV strains from PDM-09, Gamma-PDM-09-like and Gamma2 lineages.

EpiCC scores that were previously defined using experimental data available for FSXP/IA00 as fully protective or partially protective thresholds in Gutiérrez *et al.* 2017 (in the absence of cross-reactive antibodies) [124] were extrapolated and applied to the swine vaccine-to-strain EpiCC analysis. A total EpiCC score of at least 0.076 was used to define complete protection; a score between 0.065 and 0.076, was used to define partial protection (light and dark grey circles, respectively, in Figure 4.2). Based on these protective thresholds, FSXP/IA00 was predicted to confer protection and partial protection against all swine influenza strains except

Delta2 and G4. A summary of US swine influenza vaccines EpiCC scores and vaccine coverage can be found in Table 4.2. Statistical analysis was also performed to evaluate the potential of protection against G4, and FSXP/IA00 was predicted to have lower T cell epitope coverage and was therefore considered not likely to provide protection ($p < 0.05$; Supplemental Figure B-1). Other US commercial swine influenza vaccines had low T cell epitope coverage as well and were predicted to confer no protection.

COBRA vaccines are computationally optimized antigens designed to provide broad antibody epitope coverage against a wide range of variable sequences, but their T cell epitope content is not currently optimized. We considered the T cell epitope coverage of three experimental COBRA vaccines (COBRA/P1, COBRA/SW1 and COBRA/SW2) [174] for the set of US swIAV strains used in this study. COBRA/SW2 showed the highest T cell epitope relatedness to US field strains, with vaccine T cell epitope coverage of 66.7% (Table 4.2). COBRA/SW1 and COBRA/P1 had lower vaccine T cell epitope coverage of 53.3% and 49.7%, respectively. Again, the vaccines were clade-specific: COBRA/SW2 had greater total shared T cell epitope relatedness to field strains from Alpha, Beta, Gamma2, Gamma2-beta-like and PDM-09 clades (Figure 4.2B). It also had less T cell epitope relatedness to Delta-like and Delta2 field strains. In contrast, COBRA/SW1 was predicted to confer protection only against Delta-like swIAV field strains while COBRA/P1 was predicted to confer only partial protection against Delta-like field strains. These results suggest that similar patterns of T cell epitope relatedness are observed among circulating strains from the same lineage (classical swine influenza versus human seasonal influenza).

EpiCC analysis of EU vaccines and strains

For EU strains, we established thresholds for full and partial protection using retrospective data. The commercial and experimental inactivated virus vaccine strains used in challenge studies were considered protective in our analysis if lung virus titers were significantly lower than the challenge control group and pigs showed no and/or mild clinical signs (low mean temperature or low scores of respiratory diseases) compared to unvaccinated pigs. Partial protection was defined as a significant reduction of lung lesions coupled with a non-significant reduction of lung virus titer when compared to unvaccinated controls.

To define the protective thresholds, we evaluated four published vaccine efficacy studies [182]–[185] which used four different H1N1 challenge strains. Table 4.3 lists all the vaccine strains that conferred protection or partial protection against specific challenge strains and the relevant lung titer data. In six of the eight evaluations, protection was demonstrated by the absence of virus, or by significantly lower lung virus titers than the control group. A shared T cell epitope content EpiCC score was calculated for each vaccine-to-strain comparison (Table 4.3).

Based on these studies, the lowest total EpiCC score that was associated with protective efficacy was 0.0604, defined by comparing T cell epitopes from the swine/Belgium/1/83 challenge strain with the first generation of H1N1 European vaccine strain (NJ76). This threshold is represented by the white area in Figure 4.3. Based on this study and previous studies, swIAV strains that have total EpiCC scores above the threshold are likely to be protective. Similarly, a total EpiCC score of 0.0474 was associated with partial protective efficacy, as defined for the T cell epitope comparison between NJ76 and challenge strain GT/112/07. For this analysis, therefore, we defined total EpiCC score between 0.0474 and 0.0604 as partially protective.

Protective efficacy of EU swine influenza vaccines against circulating and G4

Having established estimated vaccine efficacy thresholds for EU swine influenza vaccines, we could evaluate whether the T cell epitope content of additional swIAV vaccine strains that are commonly used in the EU commercial settings (BK00 and HL03) might protect against 150 swIAV strains circulating in the EU and eight emergent G4 sequences (Figure 4.3).

HL03 had the highest T cell epitope coverage (vaccine-to-strain EpiCC scores) for circulating swIAV strains in the EU and was predicted to provide protection against 76% of the swIAV strains (total EpiCC scores greater than 0.0604). The BK00 strain had lower T cell epitope relatedness against EU swIAV circulating strains (only 8.7% of EU swIAV circulating strains had total EpiCC score of at least 0.0604). EpiCC comparisons showed that T cell epitope cross-conservation was lineage specific: HL03's EpiCC score suggested that it may only confer protection against 1C lineages strains. In contrast, vaccine strain BK00's EpiCC score suggests that it may confer protection to field strains related to 1B and other avian or human lineages but not 1C lineages.

Notably, the HL03 vaccine strain had total EpiCC scores that exceeded the defined protective threshold for emergent G4. The average HL03 vaccine strain T cell epitope coverage (slate blue line in Figure 4.3) for the G4 HA sequences was 65.7%, which is above the protective threshold (Table 4.4). This observation suggests that existing European swine vaccines may have a protective effect against emergent virus G4.

Human vaccine-to-strain EpiCC analysis results

To assess whether existing human vaccines would provide protection against potential spillover of G4 into US human population, we evaluated the T cell epitope relatedness of five

commercial human H1N1 seasonal influenza vaccine strains to G4 using EpiCC (Figure 4.4, Table 4.5).

One seasonal vaccine strain, BR07, had a distinctive pattern with high T cell epitope content relatedness when compared to Delta-like human IAV circulating strains such as strains that were circulating prior to the 2009 pandemic (Figure 4.4A). As expected, CA09, which was introduced in the 2009 pandemic, demonstrates high T cell epitope relatedness when compared to circulating strains of PDM-09 (Figure 4.4A). Newer H1N1 vaccine strains that were introduced after the CA09 pandemic also had high EpiCC scores when compared to PDM-09 circulating strains (approximately 70.0% T cell epitope coverage on average). T cell epitope coverage for human vaccine-to-G4-strains comparison was much lower, at 32.4% on average (Table 4.5).

Unlike the vaccine efficacy studies performed in swine where challenge studies data such as lung virus titers and lung lesion reduction are accessible, human vaccine efficacy is determined via clinical trials [186]. Human vaccine efficacy estimates vary among published efficacy studies [186] and hence, defining an EpiCC protective threshold is not straightforward. Since the protective thresholds were not defined for human vaccines, we used an average vaccine-to-strain EpiCC scores coverage of 38.5% (represented by the dotted line in Supplemental Figure B-1) to evaluate the T cell epitope coverage for human vaccines (Supplemental Figure B-1). The human seasonal influenza vaccines (CA09, MI15, BR18 and GDMN19) showed T cell epitope coverage below the mean and had no significant T cell epitope relatedness to G4 ($p < 0.05$) when compared to G4.

Three novel subunit influenza vaccines (COBRA HA vaccines) that were designed to generate cross-protective B cell epitopes [174], were also compared to both G4, and circulating human IAV strains, to estimate whether they might also generate protective T memory response

against emerging IAV strains. The experimental COBRA vaccine strains only showed T cell epitope content relatedness to pre-PDM strains of human IAV (Figure 4.4B). COBRA vaccine T cell epitopes were poorly conserved with G4, with an average of only 37.3% T cell epitope coverage, as was observed for H1N1 seasonal influenza vaccines (Table 4.5).

T cell epitope conservation among antigens other than HA

Although HA has been the focus of antigenic studies for most influenza vaccines, there is evidence that internal genes such as PB2 and NP may be associated with milder clinical signs and decreased virus shedding [104]. Therefore, we examined the degree of T cell epitope conservation in other viral antigens. Even though conventional inactivated virus vaccines are manufactured as high-producing reassortants containing the HA and NA of target strains and internal genes from the master strain (A/Puerto Rico/8/1934) [187], comparing the T cell epitope content of non-HA proteins of G4, and circulating H1N1 viruses may shed light on the potential for seasonal influenza virus infection to protect against G4.

All publicly available non-HA protein sequence data was retrieved for seasonal vaccine strains, G4 and circulating IAV strains. Given that there is no complete proteome sequence for the US swine influenza vaccine strains, this full proteome analysis was only performed for EU swine and US human IAV strains. There were variations in terms of the degree of T cell epitope conservation among viral antigens (Supplemental Figure B-2A - B, Supplemental Table B-4 - B5). For the EU swIAV internal protein analysis, T cell epitopes were found to be conserved between G4 and EU field strains for the PB2, PB1, PA and NP proteins. T cell epitope conservation was particularly high between HL03 and G4 for the M1 protein. Lower T cell epitope conservation was observed between EU vaccine strains and G4 for the NA protein and other internal antigens such

as M2, NS1 and NS2 (average vaccine coverage less than 65%, Supplemental Figure B-2A; Supplemental Table B-4).

As compared to EU swine vaccine strains, the internal antigens of the human vaccine strains for the US, particularly CA09 and MI15, had higher T cell epitope conservation with human circulating strains and G4 (Supplemental Figure B-2B, Supplemental Table B-5). Internal antigens PB2, PB1, PA, NP, M1 and M2 had total EpiCC scores of at least 0.097, i.e., at least 85% T cell epitope coverage (Supplemental Figure B-2B, Supplemental Table B-5). In contrast, internal proteins from the vaccine strain BR07 had lower T cell epitope conservation with G4, with T cell epitope coverage ranging between 12.1%-75.7%. This suggests that the T cell epitopes from the internal antigens of CA09 and MI15 (H1N1/pdm09 lineage) are more highly conserved with T cell epitopes from the internal antigens of G4.

Discussion

Vaccination remains the most effective public health intervention for combatting influenza infections in both swine and humans. However, the influenza virus is constantly undergoing drift and shift events, making it difficult for some vaccine strains to provide adequate protection. This is particularly worrisome when influenza viruses with reported pandemic potential, such as G4, begin to emerge in the established host population or are shown to possess critical adaptations that allow infection and possible transmission to a new host. While antibodies are usually considered to be the major correlate of protection following influenza vaccination, influenza vaccines containing highly cross-conserved T cell epitopes have also been shown to reduce morbidity and limit spread in the absence of antibodies, even when there is a mismatch between vaccines and emergent strains [3], [169], [170].

To assess the potential epidemic risk posed by the emergent G4 in swine and human populations that are naïve to this virus, we compared the HA antigen T cell epitopes contained in three US commercial swIAV vaccine strains, two commercial-use EU swIAV vaccine strains, five seasonal H1N1 human IAV vaccine strains, and five experimental COBRA vaccines (two strains designed for each swine and human host and one hybrid strains) to the T cell epitopes contained in emergent G4 HA and to circulating IAV strains from each respective region and host.

Approaches for influenza vaccination differ between swine and humans. While human vaccines rely on the World Health Organization (WHO) to make annual vaccine strain recommendations, and no standardized guidelines have been established for swine vaccine strains and dosages [173]. Moreover, the strains of influenza used in vaccines for swine differ in Europe and the US due to differences in the requirements for vaccine approval by regulatory agencies. Further variation in vaccine strains may occur, as some US-based pork producers have been applying for an exemption to the United States Department of Agriculture (USDA) rules to use "autologous" influenza strains [173].

Current commercial swine influenza vaccines used in the US are polyvalent and contain vaccines targeting distinct circulating H1 and/or H3 strains. For this comparison, we examined the H1 components of the vaccines, consisting of γ-cluster and δ-like cluster H1N1 vaccine strains, and δ1-cluster H1N2 vaccine strain. We determined that neither US commercial nor experimental swine influenza COBRA vaccines were predicted to have high T cell epitope relatedness against G4. The average T cell epitope coverage for G4 was lower than the threshold established for protective T cell activity in previously published studies. This analysis suggests that the US swine population may be susceptible to the emergent G4, even if vaccinated with current commercial vaccines.

The commercial swIAV vaccine strains used in Europe are slightly older, containing H1avN1 (Europe-avian-like lineage), H1huN2 (human-like lineage), and H3N2 subtypes, which are more related to G4. Using published vaccine efficacy data from Europe, we defined putative cross-protective thresholds for the EU influenza vaccines using a similar approach established by Gutierrez et al [124]. Lung virus titers are used as the primary measure for defining protective thresholds for Europe swine influenza vaccines rather than reduction in lung lesions. The criteria used in this analysis differed slightly from Gutierrez et al, as lung virus titers data were not available for all challenge strains in their study [124]. EpiCC analysis demonstrated that vaccine strain HL03 had the highest EpiCC scores against G4, and the potential protection predicted for HL03 was significantly different when compared to other vaccines. G4 contains EA surface proteins (HA and NA) [52]. The EpiCC analysis suggests that only one vaccine strain, HL03, from EA lineage (1C clade) may protect the European swine population against the emergent virus.

While we confirmed that there was high T cell epitope relatedness between seasonal human influenza vaccines with circulating human IAV strains (as expected), there was very low T cell epitope relatedness between human influenza vaccine HA and G4 HA. This suggests vaccination with seasonal influenza vaccine HA antigen would not induce cross-protective T cell memory against G4. We also evaluated novel influenza vaccines, known as COBRAs, and found low conservation of the HA antigen T cell epitopes with G4 HA T cell epitopes.

In addition to analyzing the HA protein of swine and human influenza for T cell epitope conservation with the HA of G4, we conducted a more comprehensive protein analysis to compare G4 conservation with other different viral antigens. Phylogenetic analysis and genotype characterization have determined that G4 carries HA and NA from the EA IAV lineage and a mix of internal genes from H1N1/pdm09 and TR lineages [52]. We hypothesized that vaccine strains

that carried H1N1/pdm09 internal genes might have high T cell epitope relatedness to internal antigens of G4. Indeed, we found that internal antigens of human vaccine strains CA09 and MI15 that are from H1N1/pdm09 lineage (all internal antigens except NS antigen) have greater T cell epitope conservation to G4. Even though the HA of G4 is divergent, cross-conservation of internal protein epitopes between current circulating influenza strains and emergent G4 virus may provide some cross-protective T cell response to swine and humans.

In conclusion, we estimated the risk of pandemic emergence of the G4 lineage by comparing the T cell epitope content of G4 strains and determining whether the T cell epitope profile matches circulating strains in both human and swine using immunoinformatics approaches. Poor T cell epitope cross-conservation between G4 and human influenza vaccines may indicate that there is a greater spillover risk to the human population than existed when pH1N1 emerged in 2009. Steps should be taken to prepare for the potential spread of G4 strains. In the absence of G4 vaccines, it may be useful to test available European swine influenza vaccines (HL03) for efficacy against G4. This study also suggests that the emergent G4 may be a greater threat to the US pork industry than to the EU industry, due to the lack of commercial vaccines that could provide cross-protective immunity to G4. Improving vaccination systems by updating vaccine strains used in pork farms and transitioning to include G4 or EA lineage should be prioritized.

There are limitations in this study that could be addressed in future research. First, while the data subsampling strategy is applied to deal with a large sequence dataset and to avoid overrepresentation of data for certain years or geographical areas, having more data subsampling replicates would better ensure results consistency. Second, the putative cross-protective thresholds determined for the EU influenza vaccines were based on commercial NJ76 vaccine strains and applied for BK00 and HL03 analyses. While protection thresholds for different vaccine strains

may vary, having more experimental data available for BK00 or HL03 would help refine the current thresholds used.

As previously mentioned, T cell epitope conservation between circulating virus strains and seasonal vaccines may contribute to the efficacy of existing (human and swine) influenza vaccines. While this study does not absolutely confirm the relevance of the EpiCC tool for the prediction of human and swine influenza vaccine efficacy, a relationship between EpiCC scores and vaccine efficacy is observed and could be used to establish a threshold for vaccine efficacy in the context of European vaccine strains. In a separate study, EpiCC correctly predicted the efficacy of a novel porcine circovirus type 2 (PCV2) viral vaccine against circulating strains of PCV2 in swine [158], [188]. That prospective study and this retrospective analysis of G4 influenza serve to illustrate the utility of EpiCC analysis for additional prospective studies of existing vaccine strains against emergent strains.

Table 4.1A. Source of H1 HA sequences used in the analyses.

Region	Host	Strain Name	Category	Label	Accession No.
EU	Swine	A/swine/Bakum/1832/2000	Commercial	BK00	EU053148
		A/swine/Haseluenne/IDT2617/2003		HL03	GQ161119
US	Swine	A/swine/Iowa/110600/2000	Commercial	FSXP/IA00	Not available
		A/swine/Oklahoma/0726H/2008		FSXP/OK08	
		A/California/04/2009		FSPDM/CA09	EPI_ISL_393964
	Human and swine hybrid	COBRA/SW1	Experimental	COBRA/SW1	Not available
		COBRA/SW2		COBRA/SW2	
		COBRA/P1		COBRA/P1	Not available
		COBRA/X3		COBRA/X3	
		COBRA/X6		COBRA/X6	
Global	Human	A/Brisbane/59/2007	Seasonal	BR07	KF009550
		A/California/07/2009		CA09	CY121680
		A/Michigan/45/2015		MI15	KU933493
		A/Brisbane/02/2018		BR18	EPI1692062
		A/Guangdong-Maonan/SWL1536/2019		GDMN19	EPI1719956

Table 4.1B. H1N1 G4 swIAV strains included in analyses.

No.	Strain Name	Accession No.
1	A/swine/Heilongjiang/1214/2016	MN416609
2	A/swine/Jilin/21/2016	MN416627
3	A/swine/Shandong/1207/2016	MN416643
4	A/swine/Hebei/0113/2017	MN416596
5	A/swine/Anhui/0202/2018	MN416586
6	A/swine/Beijing/0301/2018	MN416589
7	A/swine/Henan/SN11/2018	MN416620
8	A/swine/Jiangsu/J006/2018	MN416626

Table 4.2. Average EpiCC score, and vaccine T cell epitope coverage of US swine influenza vaccine strains compare to US circulating swIAV and G4 strains.

Vaccine			FSXP/IA00	FSXP/OK08	FSPDM/CA09	COBRA/SW1	COBRA/SW2	COBRA/P1[†]
Type			Commercial			Experimental		
US Circulating IAV Strains (n = 150)	Average EpiCC Score ($x10^{-2}$)	Class I	3.49	1.98	2.98	2.48	3.27	2.35
		Class II	4.35	2.90	3.78	3.54	4.26	3.26
		Total	7.84	4.88	6.76	6.02	7.53	5.61
	Average Maximum EpiCC Score (sd)[‡]		11.29 (0.25)					
	Vaccine Coverage (%)[§]		69.4	43.2	59.9	53.3	66.7	49.7
	Predicted Vaccine Efficacy[¶]		Protective	Not protective	Partial protective	Not protective	Protective	Not protective
G4 (n = 8)	Average EpiCC Score ($x10^{-2}$)	Class I	1.69	1.69	1.69	1.77	1.97	2.06
		Class II	2.20	2.50	2.12	2.81	2.59	2.93
		Total	3.89	4.19	3.81	4.58	4.56	4.99

	Average Maximum EpiCC Score (sd) ‡	11.66 (0.15)					
	Vaccine Coverage (%)§	33.4	35.9	32.7	39.3	39.1	42.8
	Predicted Vaccine Efficacy¶	Not protective	Not protective	Not protective	Not protective	Not protective	Not protective

†COBRA/P1 is an experimental vaccine that derived from both swine and human HA sequences (termed as 'hybrid' in this study).

‡Average maximum EpiCC score (and standard deviation) of full-length swIAV strains, expressed in $x10^{-2}$.

§Vaccine T cell epitope coverage range: >65%: protective, 56.6%–65%: partial protective, and <56.6%: non-protective.

¶Vaccine efficacy threshold defined in Gutiérrez *et al* 2017 using experimental data that served as a proxy for this analysis.

Table 4.3. Commercial and experimental vaccination and H1N1 challenge studies in EU.

No.	Vaccine		H1N1 challenge virus†	Mean HI antibody titers prior to challenge		Virus titer in lungs	Clinical outcome	EpiCC score (x10^{-2})			Ref.
	Product name	H1N1 strain		Against homologous vaccine strain	Against challenge virus			Class I	Class II	Total	
1	Gripovac®	NJ76	BE/1/83	80-320‡	40-160‡	Negative	Protection	3.50	2.54	**6.04**§	[182]
2	Gripovac®	NJ76	GT/112/07	86	5	Significantly lower (only in the left lung half)	Partial protection	2.73	2.01	**4.74**¶	[183]
	Suvaxyn® Flu	NL80		80	10	Significantly lower	Protection	4.02	2.87	6.89	
3	Experimental	NJ76	BE/1/98	305	16	No significant difference	Protection	2.73	2.08	4.81	[184]
		BE83		235	91	Significantly lower	Protection	4.26	3.33	7.60	
		BE98		197	197	Significantly lower	Protection	6.44	5.29	11.73	

	Gripovac®	NJ76		610	197	Significantly lower	Protection	2.73	2.08	4.81	
4	Respiporc® Flu3	HL03	PD/15/1981	> 256	Not detected	Significantly lower	Protection but less effective	3.80	2.84	6.64	[185]

Note: This analysis focused on the H1N1 component of previously available swine influenza vaccines in EU or experimental swIAV strains. The primary criterion for assessing vaccine protection in EU vaccine challenge studies was measurement of lung virus titers [14], [15]. Abbreviation: EpiCC, Epitope Content Comparison.

[†]All H1N1 challenge viruses were all swine strains, with abbreviations BE: Belgium; GT: Gent; PD: Potsdam.

[‡]Original data reported HI antibody titers prior to challenge against homologous vaccine strain/challenge virus as range instead of mean.

[§]Minimum EpiCC score that predicted to confer protection.

[¶]Minimum EpiCC score that predicted to confer partial protection.

Table 4.4. Average EpiCC score, and vaccine T cell epitope coverage of EU swine influenza vaccine strains compare to EU circulating and G4 strains.

Vaccine			BK00	HL03
Type			**Commercial**	
EU Circulating IAV Strains (n = 150)	Average EpiCC Score ($x10^{-2}$)	Class I	2.07	2.93
		Class II	2.86	3.73
		Total	4.93	6.66
	Average Maximum EpiCC Score (sd)†		11.55 (0.28)	
	Vaccine Coverage (%)‡		42.7	57.7
	Predicted Vaccine Efficacy§		Partially protective	Likely protective
G4 (n = 8)	Average EpiCC Score ($x10^{-2}$)	Class I	1.80	3.46
		Class II	2.52	4.20
		Total	4.32	7.66
	Average Maximum EpiCC Score (sd)†		11.66 (0.15)	
	Vaccine Coverage (%)‡		37.0	65.7
	Predicted Vaccine Efficacy§		Not protective	Likely protective

†Average maximum EpiCC score (and standard deviation) of full-length swIAV strains, expressed in $x10^{-2}$.

‡ Vaccine T cell epitope coverage range for EU swine influenza vaccines: >51.8%: protective, 40.6%–51.8%: partial protective, <40.6%: non-protective.

§Vaccine efficacy threshold as defined in Table 4.3 that served as a proxy for this analysis.

Table 4.5. Average EpiCC score, and vaccine T cell epitope coverage of US human influenza vaccine strains compare to US human circulating IAV and G4 strains.

Vaccine			BR07	CA09	MI15	BR18	GDMN19	COBRA/X3	COBRA/X6	COBRA/P1†
Type			Seasonal H1N1					Experimental		
US Human Circulating IAV Strains (n = 300)	Average EpiCC Score ($x10^{-2}$)	Class I	3.57	4.27	4.56	4.69	4.74	3.20	3.57	3.03
		Class II	3.95	4.56	4.72	4.64	4.77	3.63	3.88	3.34
		Total	7.52	8.83	9.28	9.33	9.51	6.83	7.45	6.37
	Average Maximum EpiCC Score (sd)‡		12.76							
	Vaccine Coverage (%)		58.9	69.2	72.7	73.1	74.5	53.5	58.4	49.9
G4 (n = 8)	Average EpiCC Score ($x10^{-2}$)	Class I	2.02	2.08	2.00	2.06	2.00	2.06	2.12	2.41
		Class II	2.66	1.96	1.92	2.12	1.96	2.50	2.55	2.71
		Total	4.68	4.04	3.92	4.18	3.96	4.56	4.67	5.12
	Average Maximum EpiCC Score (sd)‡		12.84							
	Vaccine Coverage (%)		36.4	31.5	30.5	32.6	30.8	35.5	36.4	39.9

†COBRA/P1 is an experimental vaccine that derived from both swine and human HA sequences (termed as 'hybrid' in this study).

‡Average maximum EpiCC score (and standard deviation) of full-length human IAV strains, expressed in $x10^{-2}$.

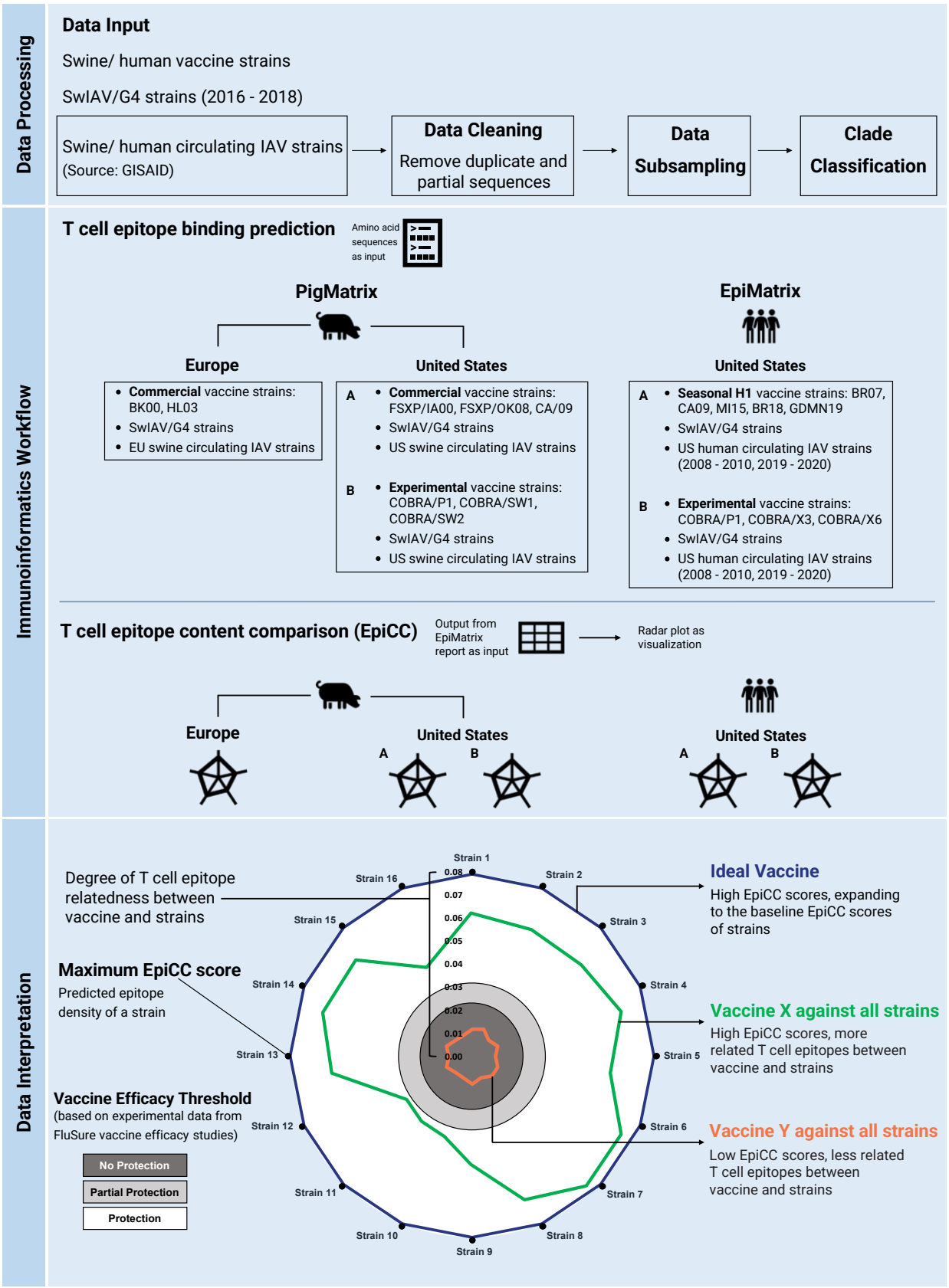

Data Processing
Data Input
Swine/ human vaccine strains
SwIAV/G4 strains (2016 - 2018)
Swine/ human circulating IAV strains (Source: GISAID)
Data Cleaning
Remove duplicate and partial sequences
Data Subsampling
Clade Classification
Immunoinformatics Workflow
T cell epitope binding prediction
Amino acid sequences as input
PigMatrix
EpiMatrix
Europe
United States
Commercial vaccine strains: BK00, HL03
SwIAV/G4 strains
EU swine circulating IAV strains
A
Commercial vaccine strains: FSXP/IA00, FSXP/OK08, CA/09
SwIAV/G4 strains
US swine circulating IAV strains
B
Experimental vaccine strains: COBRA/P1, COBRA/SW1, COBRA/SW2
SwIAV/G4 strains
US swine circulating IAV strains
United States
A
Seasonal H1 vaccine strains: BR07, CA09, MI15, BR18, GDMN19
SwIAV/G4 strains
US human circulating IAV strains (2008 - 2010, 2019 - 2020)
B
Experimental vaccine strains: COBRA/P1, COBRA/X3, COBRA/X6
SwIAV/G4 strains
US human circulating IAV strains (2008 - 2010, 2019 - 2020)
T cell epitope content comparison (EpiCC)
Output from EpiMatrix report as input
Radar plot as visualization
Europe
United States
A
B
United States
A
B
Data Interpretation
Degree of T cell epitope relatedness between vaccine and strains
Maximum EpiCC score
Predicted epitope density of a strain
Vaccine Efficacy Threshold
(based on experimental data from FluSure vaccine efficacy studies)
No Protection
Partial Protection
Protection
Strain 1
Strain 2
Strain 3
Strain 4
Strain 5
Strain 6
Strain 7
Strain 8
Strain 9
Strain 10
Strain 11
Strain 12
Strain 13
Strain 14
Strain 15
Strain 16
0.08
0.07
0.06
0.05
0.04
0.03
0.02
0.01
0.00
Ideal Vaccine
High EpiCC scores, expanding to the baseline EpiCC scores of strains
Vaccine X against all strains
High EpiCC scores, more related T cell epitopes between vaccine and strains
Vaccine Y against all strains
Low EpiCC scores, less related T cell epitopes between vaccine and strains

Figure 4.1. Analysis workflow to quantify the degree of epitope conservation between target vaccines, G4 and circulating IAV strains. The key steps include Data Processing, Immunoinformatics Workflow and Data Interpretation.

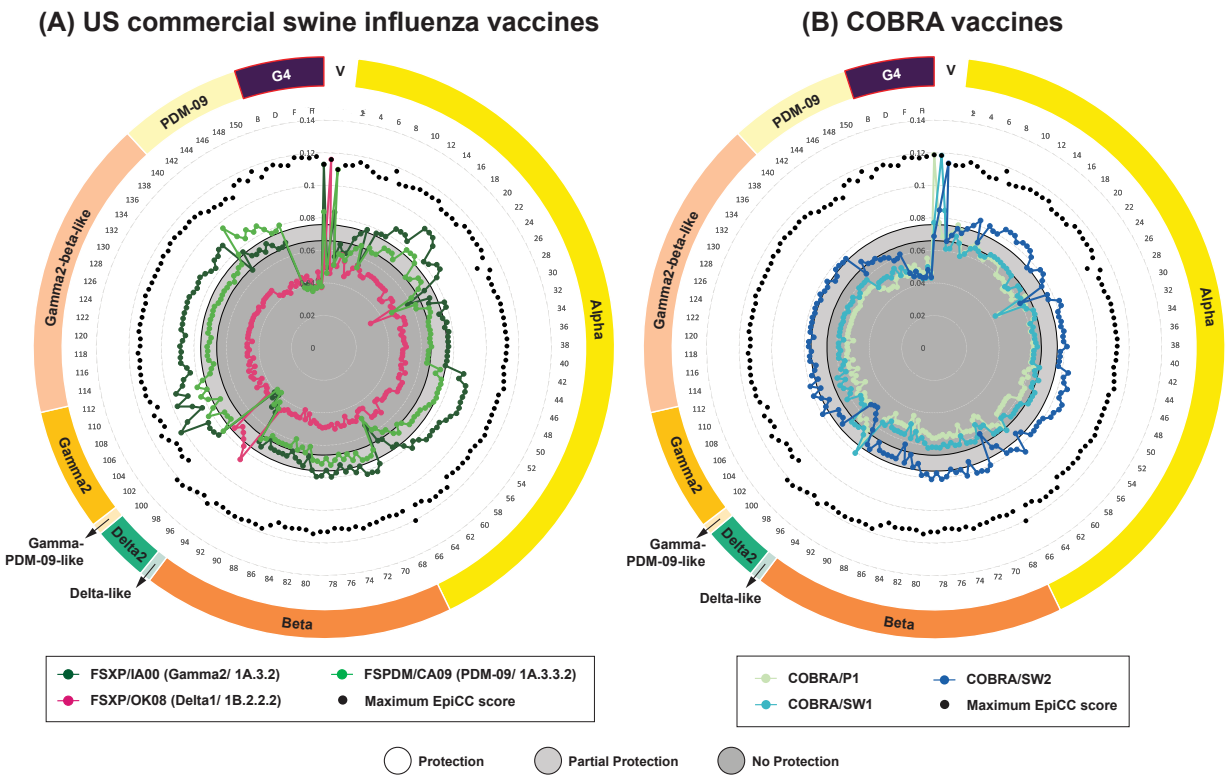


Figure 4.2. This radar plot of the EpiCC analysis enables the comparison of T cell epitope relatedness of the US swine influenza vaccines (A: Commercial; B: COBRA) to circulating US swIAV and G4 strains. EpiCC scores are plotted in a radial fashion in order of chronological time while also grouping flu variants into strain families, and color-coded lines represent each of the vaccine strains (V) compared to each of the circulating strains. Each of the vaccine strain labels is shown in legend. The ring surrounding the radar plot identifies the swIAV sequences metadata using the US clade naming system, which includes alpha, beta, delta, gamma and pandemic 2009 lineage (PDM-09). The two shaded grey circles near the center of the radius define vaccine efficacy thresholds as reported in Gutiérrez et al 2017 [124]. Refer to Figure 4.1 – Data Interpretation for more details. As indicated in both A and B, the G4 strain has few cross-conserved T cell epitopes with vaccine strains (all vaccines EpiCC scores fall below the protective thresholds).

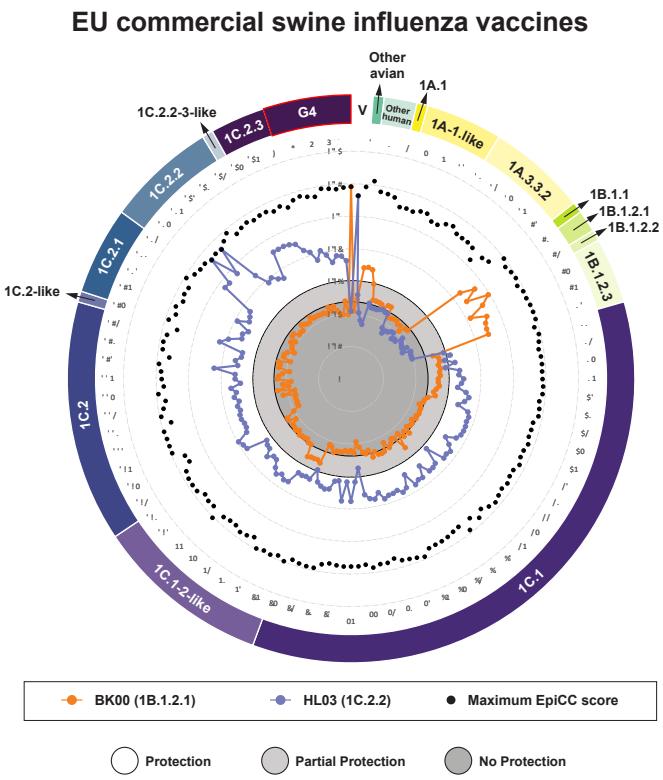


Figure 4.3. Comparison of total (class I and class II) shared T cell epitope relatedness of European swine influenza vaccines against EU circulating swIAV and G4 strains as measured using the EpiCC algorithm. Representative individual strain sequences are numbered, and their clade designation is labeled in the outer ring surrounding the radar plot. The letter 'V' designates the EpiCC scores for the European vaccine strains. The experimentally defined protection thresholds for vaccine efficacy are shown by the shaded grey circles. Note that the observed EpiCC scores indicate that the European vaccine strain, HL03, may provide protective immunity against G4.

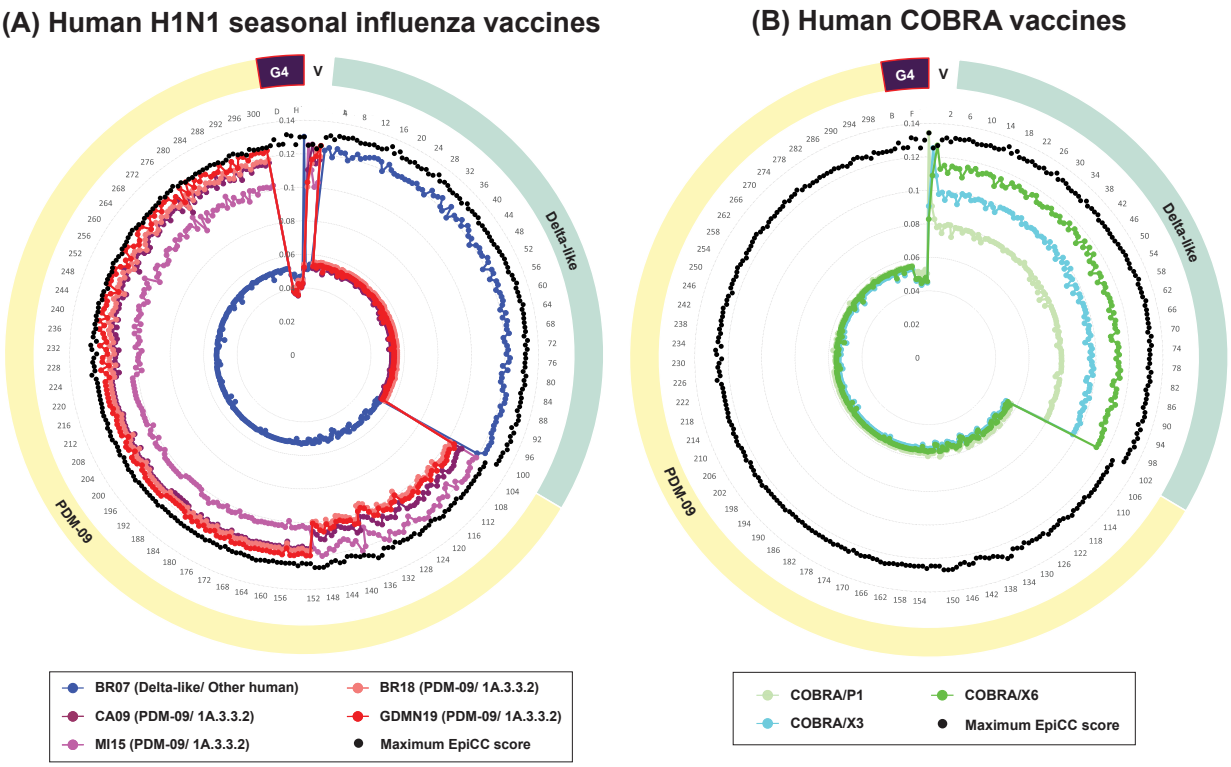


Figure 4.4. EpiCC analysis of human influenza vaccines against US circulating human IAV and G4 strains (A: human H1N1 seasonal influenza vaccines; B: human COBRA influenza vaccines). The outer ring surrounding the radar plot shows sequence metadata. The symbol 'V' represents vaccine strains, and each vaccine strain is color-coded as indicated in the legend. A protective vaccine efficacy T cell epitope conservation (EpiCC) threshold for human influenza vaccine efficacy has not been defined, as most influenza vaccines generate cross-reactive antibody. Note that EpiCC scores for G4 are extremely low compared to scores for other strains.

CHAPTER 5

SEQUENCE-BASED EVALUATION OF CD4+ T CELL IMMUNE LANDSCAPE OF H3N2 SEASONAL INFLUENZA A VIRUS [3]

[3]Tan, S; Damodaran, L; Dailey, CA; Chen J; Kondor, R; Moise, L; De Groot, AS; Bahl, J. Sequence-based evaluation of CD4+ T cell immune landscape of H3N2 seasonal influenza A virus. To be submitted to *Frontiers in Immunology*, 4/2023.

Abstract

The most effective public health intervention to fight against seasonal influenza infection is through immunization. However, the effectiveness of each seasonal influenza vaccine varies, and in some cases, vaccines are antigenically mismatched with the circulating virus hemagglutinin (HA) surface protein. To improve vaccine antigen selection, it may be necessary to anticipate viral evolution and identify key antigenic changes that may decreasing vaccine efficacy. Here, we aim to define the human T cell immune landscape of H3N2 influenza A virus and to measure drifts in the sequences of T cell epitopes with viral evolution. We use HA sequence data to predict potential T cell epitopes to examine how antigenic drift correlates with the diversity of T cell epitopes presented by the viral population over time. Our findings show that conserved CD4+ effector T cell epitopes (Teff) decline in subsequent viral populations, after a new vaccine virus is introduced, which may be linked to immune escape from the imprinted T cell repertoire. Multidimensional scaling and *k*-means clustering analyses demonstrate that Teff can be categorized into clusters based on six T cell attributes. At least four distinct clusters are detected prior to 1990, however, clusters overlap thereafter, indicating that there is an increase in the diversity of T cell epitopes over time. Our results also show that H3 HA T cell epitope evolution is linear with HA genetic drift. T cell epitope landscape changes over time appears to be driven by immune pressure, resulting in the loss of conserved T cell epitopes as well as new T cell epitope introductions. The observed decrease in T cell epitope conservation with each vaccine virus after it is introduced suggests that changes are due to immune pressure. New influenza vaccination strategies may be warranted.

Introduction

Seasonal influenza has resulted in substantial morbidity and mortality as well as direct socioeconomic impact on global populations. The World Health Organization (WHO) estimates that seasonal influenza contributes to one billion illnesses annually, of which 3 - 5 million are severe. There are 290,000 – 650,000 influenza-related respiratory deaths each year [189], [190]. The viruses that are primarily responsible for this episodic disease are influenza A viruses (IAVs), including the H1N1 and H3N2 subtypes, and influenza B viruses (IBVs), of the Victoria and Yamagata lineages. The H3N2 subtype viruses tend to be associated with more severe influenza seasons [191].

Six of the last ten influenza seasons in the northern hemisphere were associated with the influenza A H3N2 subtype (A/H3N2). While all influenza viruses undergo antigenic drift, A/H3N2 viruses have exhibited more frequent genetic changes since their emergence in the human population in 1968 as compared to the influenza A H1N1 subtype (A/H1N1) and IBVs [191]–[193]. The rapid pace of A/H3N2 evolution has frustrated vaccination efforts, due to genetic and/or antigenic disparity between the H3 components of the influenza vaccine and circulating influenza viruses. Mismatches between the vaccine virus and the circulating virus hemagglutinin (HA) protein contribute to lower H3 vaccine effectiveness [194]. Other factors such as egg-adapted mutations that are more likely present in A/H3N2 viruses during influenza vaccine production process can also reduce their potential effectiveness against circulating influenza viruses. The Centers for Disease Control and Prevention (CDC) reports confirm that vaccine effectiveness is more often achieved against IBVs and A/H1N1 than for A/H3N2 viruses [193], [195].

Even though the selection of the seasonal influenza vaccine virus appears to be difficult, especially for A/H3N2 vaccines, annual vaccination remains the most preferred preventive

measure for combatting seasonal influenza infection. It is likely that A/H3N2 viruses are more genetically and antigenically diverse due to accelerated evolution to escape host immune pressures. Thus, understanding A/H3N2 viral evolution and the impact on host immune selection pressures may be useful for optimizing vaccine design strategies.

The immune system develops highly specific protective antibodies and memory T cells in response to natural infection and vaccination. Thus, two types of immune pressure may contribute to immune escape and viral evolution. If antibody-mediated immunity is the only pressure that causes viral evolution, T cell epitopes should remain unchanged in the evolving viral sequence or exhibit no greater change than would be observed by random chance. However, changes in T cell epitope content over time would seem to indicate that T cell mediated immunity also contributes to viral evolution.

Previous studies of influenza vaccines have primarily focused on the contributions of humoral immunity to antigenic drift. This study will address an existing knowledge gap related to the impact of cellular-mediated immunity, specifically the impact of CD4+ T cell immunity on viral evolution. Since CD4+ T cells provide essential help to B cells and CD8+ T cells, they play a critical role in defense against influenza virus infections. Their role in protection from infection through antibody production and viral clearance has been confirmed [92], [93], [170], [196]. Human studies have shown that cellular immune responses are important to protection [94] and that conserved influenza-specific T cells resulting from previous influenza infections can cross-react with similar epitopes found in novel IAV strains [5], [6], [104]. In addition, when cross-reactive antibodies are absent due to antigenic mismatch, influenza-specific CD4+ immune responses reduce disease morbidity [5].

The hemagglutinin inhibition (HI) assay that has been used to determine the antigenicity of the HA has become less reliable over the last decade due to changes in glycosylation [197], [198]. Sequence-based approach is a newer strategy that may help characterize antigenicity and inform research on viral diversity [199], [200]. In this study, we used T cell epitope mapping tools along with a new method for studying potential cross-conservation of T cell epitopes at the residues that interface with T cell receptors (TCRs). Essentially, a T cell epitope can be cross-reactive despite not being identical in sequence to another epitope, if the TCR-facing (TCR*f*) amino acids in the two sequences are identical [129], [201]. In other words, two or more peptide-major histocompatibility complex (p-MHC) ligands from different influenza viruses can be recognized by the same T cell's TCR [135]. Here, we apply the T cell epitope sequence-based approach to identify potential CD4+ T cell immune pressure on A/H3N2 virus sequences in viral evolution. Instead of focusing on sequence identity as a means of understanding viral evolution, we analyze changes in the T cell epitopes by predicting class II T cell epitopes with potential binding affinity to human leukocyte antigen (HLA) alleles and analyzing how TCR*f* residues contribute to epitope cross-conservation and variation between vaccine strains and circulating viruses [124]. We also categorize T cell epitopes as being T effector or potentially tolerated (non-T effector) by examining the TCR*f* residues that are conserved with the human genome.

Materials and Methods

Data collection

To study the human T cell immune landscape of A/H3N2 virus, we analyzed two sets of sequences that spanned across different time frames: dataset A, 1968 - 2004 and dataset B, 2010 - 2020 (Figure 5.1). Dataset A contained 269 HA nucleotide sequences used by Smith *et al* [202].

This well-characterized dataset consisted of 253 circulating and 16 vaccine viruses (Supplemental Table C-1A) with corresponding HI data; (2) We used the same analysis approach for dataset B which contained 1020 sequences isolated between 2010 and 2020. We retrieved these more recent sequences from a publicly available database, Global Initiative on Sharing Avian Influenza Data (GISAID EpiFlu; http://platform.gisaid.org/epi3/; accessed in July 2022).

Data curation and subsampling of modern (2010 to 2020) sequences

For dataset B, we first compiled a large set of HA sequences that have associated HI data provided by the Centers for Disease Control and Prevention (CDC). Prior to finalizing the set of sequences, we removed sequences that have a missing isolation date. We also removed sequences that had discrepancies in labeling. The processed dataset resulted in a total of 9654 A/H3N2 HA sequences, and we then used IQ-TREE with substitution model [203] to construct a phylogenetic tree prior to data subsampling. To efficiently examine A/H3N2 phylogenetic diversity observed over the last ten influenza seasons, the Phylogenetic Diversity Analyzer (PDA) [178] was used to subsample 1020 HA nucleotide sequences worldwide ranging from 2010 to 2020. Of the 1020 sub-sampled A/H3N2 strains, 1007 sequences had associated HI data, including seven vaccine viruses (Supplemental Table C-1B). The remaining 13 sequences did not have available HI data of which 11 were egg-based vaccine viruses and two were cell-based vaccine viruses. The egg-based and cell-based vaccine viruses were identified by referring to the World Health Organization (WHO) recommendations for influenza vaccine composition [204].

To evaluate the consistency of our results for the modern sequences, we added two random subsampling replicates from the total of 9654 A/H3N2 sequences and analyzed these subsamples

in addition to the phylogenetic-based subsampling dataset B. The two random subsampling replicates consisted of 1036 and 1034 A/H3N2 HA sequences, respectively.

Control group for dataset B

To confirm the results, we established a control group using canine IAV sequences was obtained from GISAID. This set of control group sequences consisted of 252 canine A/H3N2 HA sequences isolated between 2009 and 2019. Changes in the T cell epitope content of these sequences was considered less likely to have been due to human immune pressure.

T cell epitope prediction

The first step in the immunoinformatics analysis for this study was to identify potential T cell epitopes that are highly likely to form a peptide-major histocompatibility complex (p-MHC) for T cell receptor (TCR) recognition and signaling in a broad human population. To evaluate the immunogenicity of each protein sequence, the HA protein sequences were translated using Sequence Manipulation Suite [205] and were screened for potential binding to human HLA using HLA-restricted T cell epitope prediction tool called EpiMatrix (EMX), a T cell epitope prediction algorithm developed by EpiVax (Figure 5.1). This tool is available for through academic collaborations.

Given the relevance of class II T cell epitope for stimulating CD4+ T cell immunity [114] and the relevance of CD4+ T cell help to the generation of effective antibody responses [92], [93], [170], [196], this study focused on class II HLA-restricted T cell epitopes. Amino acid sequences were first parsed into overlapping 9-mers to capture all possible peptides of all input sequences. Next, using a position-specific scoring matrix, each 9-mer peptide was assessed for class II binding

likelihood across nine HLA supertype alleles, i.e., DRB1*01:01, DRB1*03:01, DRB1*04:01, DRB1*07:01, DRB1*08:01, DRB1*09:01, DRB1*11:01, DRB1*13:01, and DRB1*15:01. These class II HLA allele supertypes used in the algorithm cover the genetic diversity of more than 95% of the global human populations [114], [121], [122].

For each 9-mer peptide (*i*) in each individual allele (*a*) of a set of HLA alleles (A), EMX raw scores, *r*, are normalized to Z-scores using the average μ and the standard deviation σ of scores calculated for 100,000 random 9-mers [114]. Peptides with Z-scores ≥ 1.64, which correspond to the top 5% of any given 9-mers sample, are assessed as having higher HLA binding potential as compared to those with lower scores [114], [125]. Higher Z-scores are associated with higher HLA binding probability and considered to be a potential epitope or "hit" [124], [126]. In addition, any 9-mer peptide that is identified as having high scores or "hits" for at least four out of nine HLA supertype alleles is highly likely to be recognized by a broad range of individuals in any given population. This type of pattern defines a 9-mer peptide as being a potential 'promiscuous' T cell epitope and of greater immunological importance for immune responses at the population level. Peptides that have more than four hits in one 9-mer frame are said to have high epitope density and are prioritized for vaccine design. In general, higher scoring peptides that have higher T cell epitope content are more likely to be validated or to have been validated for T cell response. In the context of this analysis, epitope predictions were compared to experimentally verified and published influenza epitopes reported in the Immune Epitope Database (IEDB) as a means of external validation.

Cross-conservation analysis

JanusMatrix (JMX) identifies T cell epitopes that are restricted by the same HLA that may be different in terms of HLA binding amino acid residues (agretope) but have the same TCR-facing (TCR*f*) residues (epitope). JMX expands on EMX output, using EMX-identified epitopes to search for the potential cross-reactivity between the MHC-binding epitope TCR*f* and T cell epitopes derived from the human genome or human pathogens [114], [129].

Cross-conservation of T cell epitopes at the TCR face is highly relevant for two reasons. First, cross-conservation at the TCR face with epitopes that have been encountered in the context of prior infections may drive pre-existing immune response [206]–[208]. In addition, TCR can interact with a range of T cell epitopes that presented in the form of p-MHC by HLA. If these T cell epitopes are derived from self-antigens, it is highly likely that the T cell with the cognate TCR has been trained to be non-reactive or eliminate in the thymus (tolerant) or, it may be actively tolerogenic [129]. We have determined that peptides that contain TCR*f* residues that are highly cross-conserved with self-epitopes restricted by the same HLA can be tolerogenic *in vitro* and *in vivo* [201]. Thus, in addition to using JMX to find potential cross-reactive T cell epitopes in different viruses, we use the JMX tool to categorize T cell epitope phenotypes (effector or non-effector) based on the similarity at the TCR face with self and non-self-epitopes (Figure 5.1).

In this study, T cell cross-reactivity of the A/H3N2 strains from both datasets were evaluated by defining the binding likelihood of every epitope from input sequences that has identical TCR*f* residues to epitopes found in other sequences of A/H3N2 HA, given that the epitopes bind to the same HLA allele. Two epitope peptides are assumed to be potentially cross-reactive if they have identical TCR*f* at positions 2, 3, 5, 7, 8 (from 9-mer class II epitopes binding core) regardless of differences on their HLA-facing amino acids. In addition, JMX is used to

compare HA epitopes with human genome epitopes to separate epitopes by putative phenotype into effector (immunogenic) T cell epitopes (Teff) and non-effector T cell epitopes (non-Teff). This division is based on a threshold of cross-conservation with self-epitopes that has been validated in retrospective and prospective studies [209]–[211]. JMX identifies potential human-like T cell epitopes in each IAV by assessing the amino acid residues between IAV epitopes and human epitopes at TCR*f*. An R script (available at https://github.com/swantan) was developed to compute a ratio of self-reactive hits to pathogen reactive hits, termed JMX Human Homology score, for each A/H3N2 HA peptide by applying the following equation, where a hit was defined as the putative epitope with EMX Z-scores greater than or equal to 1.64:

$$\text{JMX Human Homology score} = \frac{number\ of\ cross-reactive\ hits\ of\ human\ epitope\ of\ the\ same\ allele}{number\ of\ IAV\ HA\ putative\ epitope\ allele\ hits}$$

With the computation of JMX Human Homology score, we can categorize putative T cell epitopes identified by EMX in A/H3N2 HA sequences as being of the Teff or non-Teff based on their degree of homology with epitopes in the human genome. As Teff is the focus of our study, if a predicted T cell epitope of HA has a JMX Human Homology score less than or equal to 5, it is considered to be more likely to induce effector T cell response. If it has a JMX Human Homology score greater than 5, it is considered to be non-Teff.

We then consider T cell cross-reactivity between IAV sequences as mentioned earlier and perform pairwise comparison to detect cross-conservation of T cell epitopes between IAV strains using custom R scripts (available at https://github.com/swantan). A pair of epitopes from two different A/H3N2 viruses that have identical TCR*f* are regarded as shared T cell epitopes and are

assumed to have a high likelihood to induce cross-reactive T cell response. In contrast, if a pair of epitopes have non-identical TCR*f,* they are considered to be unique epitopes (non-cross-conserved) to the compared A/H3N2 virus sequence. The number of shared and unique epitopes for Teff are calculated and two *n* x *n* matrices are generated (shared Teff and unique Teff) respectively.

Antigenic cartography and clade designation

To compare T cell epitope changes to antigenic evolution of A/H3N2, we followed the method of Smith *et al* [202] using the published antigenic cartography method to construct antigenic map of all available HI assay data. H3 antigenic maps of the two datasets were constructed using the *racmacs* package in R [212].

We used the antigenic group information of the well-characterized dataset identified from the Smith *et al* study. For the contemporary sequence data collected between 2010-2020, we used the Nextclade program to assign clade as a discrete trait for each taxon [213].

Multidimensional scaling (MDS) and k-means clustering analysis

To better represent the large amount of T cell epitope data generated from T cell epitope prediction and cross-conservation analysis, we used a combination of multidimensional scaling (MDS) approach and the *k*-means clustering algorithm to visualize partitions and to interpret T cell epitope data [214], [215]. Classic multidimensional scaling calculation is applied to transform each high dimensional dataset (containing the matrices of T cell epitope attributes such as binding affinity scores for nine class II supertype alleles, T cell epitope distance, epitope allele hit count, T cell cross-conservation scores/JMX Human Homology scores, and human epitope cross-reactive

hit count) and to reduce it to a low dimensional set of data while retaining most of the information. The retained information is represented by two-dimensional coordinates which illustrate the distance between two viruses based on defined measurable attributes.

Subsequently, clustering of the A/H3N2 viruses was performed based on the computed coordinates of the previously mentioned attributes using the *k*-means algorithm. Given that k-means is an unsupervised machine learning algorithm, Bayesian Information Criterion (BIC) was used to estimate the optimum number of *k*-means clusters [216]. MDS and *k*-means clustering were carried out using the *cmdscale* and *kmeans* package in R [214], [215]. A T cell antigenic map was created for each of the previously mentioned attributes to obtain a better visualization.

Additionally, we conducted a principal component analysis (PCA) for each of the T cell epitope attributes mentioned above. PCA is another statistical method for reducing high-dimensional data by linear transformation. Principal components were summarized in their contribution to total variation in the data and relative loadings from binding affinity scores of nine class II supertype alleles, T cell epitope distance and T cell cross-conservation scores (JMX Human Homology scores). This served as a sensitivity check for MDS, as well as a means of determining the source of variation. PCA was performed using the *prcomp* package in R [217]. We used *screeplot* to plot the variances against the number of principal component (PC) and chose appropriate PCs that best represent the datasets [217]. Subsequently, we sorted out variable loadings/rotation to identify characteristics that significantly vary between isolates.

Genetic hamming distance

Genetic hamming distance is defined as the number of bases by which two nucleotide sequences differ. The distance was calculated by summing the number of different bases between each

nucleotide sequence in the datasets. A python script developed by Chen *et al* was applied to perform the calculation [218].

Phylogenetic inference

The two datasets of 269 and 1020 nucleotide sequences of the HA gene were used to reconstruct the phylogeny of A/H3N2 separately, using the Bayesian phylogenetic method. The Bayesian phylogenetic analysis was performed using a lognormal relaxed molecular clock in a Bayesian statistical framework implemented in BEAST v.1.10.4 [219]. Molecular clock rates are uncorrelated with an initial mean of 0.0033 with a uniform prior ranging from 0.0 to 1.0. A GTR-GAMMA nucleotide substitution model [220] and a GMRF Bayesian Skyride coalescent prior [221] were used. Six independent Markov chain Monte Carlo (MCMC) simulations of 100 million generations were combined using Logcombiner [222] after removing 10% burn-in, with effective sample size (ESS) value greater than 300 in Tracer [223]. TreeAnnotator version 1.10.4 (http://beast.bio.ed.ac.uk/) [222] was used to generate the maximum clade credibility (MCC) phylogenetic tree. Clustering information of all T cell epitope attributes against H3 HA phylogeny tree tips were visualized using *ggtree* package version 2.2.4 in R [224].

Bayesian Stochastic Search Variable Selection (BSSVS) analysis

An empirical tree set of 500 posterior trees was created using the previously described Bayesian phylogenetic reconstruction. This tree set was used to implement a discrete trait diffusion model in BEAST v.1.10.4. The antigenic group (previously described) for isolates was used as the discrete trait for the inference of discrete trait diffusion patterns and to infer co-variate influence on the diffusion process. An asymmetric substation model was utilized using non-reversible

continuous time Markov chain. To reduce the number of rates to the most parsimonious supported network, the Bayesian Stochastic Search Variable Selection (BSSVS) was employed [225]. Three independent runs of 10 million generations were performed for the BSSVS models, log files resulting from the analyses were diagnosed using Tracer v1.7.2 and combined using LogCombiner v1.10.4. The Bayes factor support for the significant non-zero transition rates were calculated using SpreaD3 v0.9.7 [226]. The mean transition rates were calculated from non-zero actual rates of the BSSVS log file using an in-house python script. The level of Bayes factor support indicating the level of support for a given model/rate versus a second model/rate can be described using the criteria established by Kass and Raftery [227].

Results

T cell epitope diversity of the characteristic dataset

Given the relevance of class II T cell epitope to CD4+ T cell immunity, which was the focus of this study, the EMX class II epitope prediction algorithm was employed to predict 9-mer sequence binding affinity to a panel of nine prevalent class II HLA alleles for each of the 9-mer peptides contained in the sequence of the HA antigens of selected A/H3N2 viruses. The 269 full-length HA sequences spanning 35 years were computationally screened for T cell epitopes for each overlapping 9-mer frame (14). On average, each analyzed HA sequence was 566 amino acids in length and as each frame overlaps the previous one by 8 residues, there were roughly 558 9-mer assessments for each strain.

To determine the binding affinity of each 9-mer frame, the EMX Z-score was calculated for each of the nine supertype HLA DR alleles. Any Z-score that is greater than or equal to 1.64, which corresponds to the top 5% of scores for any of the HLA DR alleles, is defined as having

significant binding potential to the respective allele and is considered to be an "allele hit". A 9-mer peptide that has allele hits for at least four out of nine HLA DR supertypes is considered to be a promiscuous epitope with a broader allele coverage of general human populations.

A total of 4,421 potential T cell epitopes were identified. Among these, 282 (6.4%) were predicted to have at least four out of nine HLA allele hits (Table 5.1). 3,548 of the potential 9-mer sequences (80.3%) were predicted to have limited potential to any of the HLA alleles (non-binders) and 591 of the 9-mers (13.4%) had between one to three allele hits.

Cross-conservation analysis of class II HA epitopes and human peptides

Cross-conservation of T cell epitopes can be delineated when the TCR of a given T cell interacts with a range of T cell epitopes that have identical TCR*f* residues even though they may have different on their HLA-facing amino acids, as long as they bind to the same HLA allele. Epitopes that are highly 'human-like' or cross-conserved with multiple human T cell epitopes, as defined by JMX, may limit immune response due to tolerance mechanisms (tolerance to self as opposed to non-self-epitopes). In other words, the introduction of human-like T cell epitopes may suggest that the virus is adapting to the host. We applied JMX to detect cross-conservation between the T cell epitopes predicted from HA and peptides derived from human proteins. Similar to the EMX allele hit definition, a cross-conserved hit is defined as a human peptide with the same TCR*f* residues as the HA T cell epitope and has significant binding potential to the same HLA allele (Z-score greater than or equal to 1.64).

The JMX Human Homology score (T cell epitope cross-conservation score) is derived from the number of cross-reactive hits to human peptides and HA allele hits. In general, the lower the number of cross-reactive human epitope hits to any given HA T cell epitope (JMX Human

Homology score less than or equal to 5), indicates that there is a higher likelihood for the viral epitope to stimulate effector T cell response (due to the absence of tolerance to the peptide). We used the JMX Human Homology score for each HA 9-mer peptide to classify T cell epitope phenotypes into Teff or non-Teff. T effector immune responses by CD4+ T cells would support B cell responses and other immune functions.

Since the binding potential of HA epitopes (HA allele hit) is also a determinant of effective T cell response, we implemented a set of rules that combined the allele hit and JMX Human Homology score to sort HA epitopes into three possible categories: putative T effector epitopes (Teff), potentially tolerated epitopes (non-Teff) and non-binders (Table 5.1). Considering the results for the EMX and JMX analyses for each 9-mer, among the 4,421 putative T cell epitopes that have at least one allele hit, 501 (11.3%) could be categorized as Teff while 90 (2.0%) were categorized as non-Teff. Of the 282 (6.4%) putative T cell epitopes with at least four allele hits, 268 (6.1%) were predicted to be Teff and were promiscuous T effector epitopes and 14 were identified as being potentially tolerated.

Characterization of HA diversity

As cross-reactive memory T cells can be stimulated by epitopes with identical TCR*f* residues despite having different HLA binding residues, as long as the sequences bind to the same alleles, this aspect may be pertinent to characterization of T cell immune dynamics that change over time. To assess the relatedness of T cell epitope between two A/H3N2 viruses (vaccine or circulating virus), we expanded cross-conservation analysis to quantify epitopes and the predicted phenotypes of epitopes that were cross-conserved between two A/H3N2 viruses or unique to individual A/H3N2 virus (non-cross-conserved).

To perform this analysis, a pairwise comparison of T cell epitope between the 269 IAV viruses was carried out for Teff epitope. Each Teff epitope for a given HLA-DR allele was evaluated at the TCR*f* and was denoted as conserved between IAV viruses when the TCR*f* amino acid residues were identical between the two IAV viruses, if not, it was identified as unique (non-cross-conserved). Briefly, we generated two matrices for this data: conserved Teff and unique Teff.

Distribution of Teff epitopes over time

In addition to seasonal influenza virus evolution, the HA component of seasonal influenza vaccines change almost annually. We compared seasonal influenza viruses to vaccines, measuring the presence of conserved Teff epitopes over time, first in reference to the original A/Hong Kong/1/1968 (HK68) strain and then for each of the 15 known A/H3N2 vaccine viruses from 1968 to 2004 (Figure 5.2; Supplemental Figure C-1A). We observed that there was an immediate decline of conserved Teff epitopes after each vaccine introduction and the decrease of conserved epitopes was significant in 12 out of 16 seasons (Figure 5.2A). When all subsequent circulating A/H3N2 viruses were compared to the original HK68 virus 'origin', the number of conserved Teff epitopes dropped continuously, indicating that conserved Teff epitopes continually decrease with time.

In 2000-2004 season, the vaccine virus, A/Moscow/10/1999 (MO99) and A/H3N2 viruses circulating in the same season differed by as many as 20 conserved Teff epitopes ($R = -0.83$; p-value = 5.1×10^{-4}). There were as few as nine (minimum) to as many as 30 (maximum) conserved Teff epitopes changes between the vaccine viruses and the A/H3N2 viruses circulating at a given time (Supplemental Table C-2A). A/Sichuan/02/1987 (SI87) had the smallest total number of changes (Figure 5.2B): nine conserved Teff epitopes were lost in the A/H3N2 viruses circulating in that season. Although there was no consistent pattern in terms of the median changes of

conserved Teff epitopes (Figure 5.2B), we observed that there was no change in median for the two consecutive seasons: TX77 and BK79. The largest increase median of eight conserved Teff epitopes between PH92 and LE86 whereas the largest drop of median of seven conserved Teff epitopes happened between BJ89 and BJ92. Overall, there were more decreasing of conserved Teff epitopes compared to increasing conserved Teff epitopes (Figure 5.2B).

In contrast, the number of unique Teff epitopes exhibited an inverted trend as compared to conserved Teff epitopes. The number of unique Teff epitopes steadily increased over time with respect to the original virus HK68. Decreases in number of conserved Teff epitopes was generally associated with greater total unique Teff epitopes gain, expanding the distance between the vaccine and circulating viruses. A gradual increase of unique Teff epitopes was observed from BJ89 to MO99.

Appearance and disappearance of T cell epitopes

In general, conserved Teff epitopes between seasonal viruses and seasonal vaccines appear to decrease following the introduction of a new vaccine. Beyond changes in total T cell epitope counts, sequence mutations may modify the T cell epitope phenotypes leading to changes in the categorization of the T cell epitope. Since vaccines that contain T cell epitopes that are highly conserved with newly emerging viruses may reduce morbidity and mitigate the spread of the virus, it may be important to track both the number of T cell epitopes as well as dynamic changes in the T cell epitope phenotype, to improve vaccine virus selection.

For this analysis, we confined our study to 47 promiscuous T cell epitopes that were confirmed in the IEDB database (Figure 5.3). We noted that the HA1 domain appears to be more diverse than the HA2 domain; the T cell epitope phenotype appears to switch more frequently in

HA1 as viruses evolve over time. Some positions appear to be partially or fully conserved while others show T cell phenotype transitions. More specifically, the putative Teff epitopes located at three positions (163, 310 and 330) in HA1 appear to have 100% conservation starting from 1968 to 2003. The epitope at position 324 was partially conserved with intermittent change in the putative T cell phenotype due to single mutation at the HLA face or the TCR*f*. The epitope at position 261 was conserved at the HLA-binding face and TCR*f* prior to 1982, and the predicted phenotype remained (Teff) even with a single mutation at the HLA face. More than five T cell epitopes located at positions in HA1 demonstrated inconsistent switching of the T cell phenotypes, while several epitopes in HA1 demonstrated more consistent phenotypic change. Interestingly, limited T cell phenotype switching was observed in antigenic group BJ89 where there is only one vaccine virus update. T cell epitope phenotypes were more diverse in other antigenic groups prior BJ89 as well as between BJ92 and FU02.

The HA2 domain epitopes were relatively more conserved in terms of their phenotype, when comparing T cell epitopes in the original HK68 virus to viruses circulating in 2003 (Figure 5.3). Generally, positions that were predicted to be promiscuous Teff epitopes remained the same throughout the 35 years of virus evolution even though a few mutations did occur at the HLA face or TCR*f*. Epitopes in six out of 24 positions in HA2 (347, 408, 439, 444, 512 and 518) were 100% conserved at both the HLA face and TCR*f* in the 269 A/H3N2 viruses. While single mutation on either HLA face or TCR*f* might change a Teff epitope to a non-Teff epitope or to 9-mer that was predicted to have no binding affinity, mutations also occur that do not influence the phenotype of the epitope. Epitopes at position 397 and 400 had an interesting pattern in which the 9-mer sequences in the older viruses were predicted to be non-binders but switched to binders (Teff epitopes) with no changes at TCR*f*, however, from1994 onwards, we observe modulation between

Teff epitope and non-Teff epitope phenotypes due to mutations on the TCR*f*. There were also five epitope positions in the HA2 (464, 533, 536, 540, 541) that were predicted as non-Teff epitopes.

MDS and clustering analysis

In view of the extensive and complexity of T cell epitope data, we used the MDS approach, and the *k*-means clustering algorithm to visualize and further investigate the T cell epitope landscape. We focused on analyzing all putative T cell epitopes predicted from the coding region. We ran six separate MDS analyses and *k*-means optimization on different T cell epitope properties such as the binding affinity score of the nine HLA alleles, HLA allele hit, human peptide cross-reactive hit, number of human peptide match, JMX Human Homology score and unique Teff results. We found that A/H3N2 viruses clustered into seven groups based on human T cell epitope conservation (JMX Human Homology), and into eight groups when considering T cell epitope cross-conservation between IAV sequences (JMX IAV homology score). There were nine groups based on the analysis of allele hits, cross-reactive hits and unique Teff epitope landscape. The lowest number of clusters that was detected (six clusters) was defined using T cell epitope binding affinity.

To compare and correlate the evolution of A/H3N2 T cell immunity profiles against antigenic and genetic evolution, we constructed two-dimensional map representations following MDS and *k*-means clustering analysis using information on HI titer data, genetic hamming distance and two T cell attributes such as the T cell epitope binding affinity and T cell epitope distance (Figure 5.4). Genetic hamming distance is defined as the number of different bases between two nucleotide sequences of A/H3N2 viruses, whereas T cell epitope distance is defined as the difference of Teff epitope content between pair of A/H3N2 viruses.

We noticed that using the T cell epitope binding affinity and T cell epitope distance variables, the A/H3N2 T cell epitope landscape progresses in a similar pattern as the antigenic transformation and genetic distance in the two-dimensional space, indicating that both T cell epitope binding affinity and T cell epitope content may play an influential role in shaping T cell responses over time. We further decomposed this analysis and plotted each dimension against virus isolation year. Interestingly, T cell binding affinity showed an episodic evolution pattern where distinct clusters were detected prior to 1990 and clusters overlap thereafter, whereas T cell distance undergoes continuous evolution. Both results also demonstrated that A/H3N2 T cell epitope evolution of HA protein is linear with HA genetic drift which was measured by genetic hamming distance (Figure 5.4; Supplemental Figure C-4).

To examine the relationship between T cell epitope diversity and evolution, we mapped the cluster data to the respective HA phylogeny tips (Figure 5.5). We observed that all clusters in antigenic groups prior 1989 have gradual progression whereas A/H3N2 viruses in the antigenic group BJ89 is consistent across HI titer, genetic distance and T cell epitope attributes. Antigenic groups that contained more than two vaccine updates had irregular changes of clusters, especially from 1992 onwards. The binding affinity score of the nine HLA alleles, HLA allele hit, and the number of human peptide match demonstrated similar changes that correlated with T cell epitope cluster transitions. However, other measurements, for instance, human peptide cross-reactive hit and JMX Human Homology score display more distinct patterns that change according to the HI titer data.

T cell epitope immune landscape of A/H3N2 HA from 2010-2020

We applied the same analysis pipeline to sequences isolated from ten influenza seasons to investigate the divergence of A/H3N2 in more recent years. A total of 7,735 unique 9-mer sequences contained in contemporary HA sequences were assessed (Table 5.1). 1,756 (22.7%) of the 9-mer sequences were identified as having a hit for at least one out of nine HLA alleles (top 5%), and after screening these for human homology with JMX, 1,535 9-mers (19.8%) were predicted to have T effector potential (Teff) (Table 5.1). Among these, 476 (6.2%) were identified as promiscuous Teff epitopes because they contained allele hits for at least four out of nine HLA alleles. 5,979 9-mers (77.3%) were predicted to have no binding potential to any of the HLA alleles.

Distribution of putative Teff epitopes

We then examined the pattern of Teff epitopes distributed over the course of ten influenza seasons from 2010 to 2020 (Figure 5.6, Supplemental Figure C-1B). We noticed a similar trend as compared to the older dataset in which the number of conserved Teff epitopes contained in seasonal viruses for each time period was negatively correlated with the vaccine virus, following each vaccine introduction.

There was a significant decrease of conserved Teff epitopes as compared to vaccine virus A/Perth/16/2009 (PE09) was in used (correlation coefficient, $R = -0.25$; p-value $= 2.1 \times 10^{-3}$), however, the decrease in the number of conserved Teff epitopes between 2012-2014 and 2015-2016 seasons was not significant. The number of conserved Teff epitopes continued to decline and exhibited an overall significant negative correlation between vaccine and circulating viruses in four consecutive influenza seasons after 2016. On average, the decline in the number of conserved

Teff epitopes between vaccine and circulating viruses was greater than was observed for the older dataset, suggesting that the diversity of T cell epitopes may have increased in contemporary influenza seasons.

To ensure results consistency, we randomly subsampled two datasets of human contemporary viruses. Subsample 1 consisted of 1036 sequences while subsample 2 had 1034 sequences (Supplemental Figure C-2). The findings appeared to be consistent with the results found for dataset B. To determine whether the decrease in Teff epitopes following vaccine introduction was due to the T cell epitope evolution or random and unrelated to immune pressure, we performed an additional analysis using canine IAV sequences. The method was the same when evaluating changes in HLA-binding T cell epitope content (Supplemental Figure C-3). Over 10 consecutive influenza seasons, the number of conserved Teff epitopes remained constant. This finding suggested that the decrease in human T cell epitope content following vaccine changes is not a random event and more likely due to T cell epitope drifts that responding to immune pressure.

Clade-specific T cell epitope

We also tracked T cell epitopes phenotype change and conservation at TCR*f* to understand the potential influence of immune selection pressure on A/H3N2 strains in 2010-2020 as we observed for the dataset A. The HA1 domain had greater diversity with frequent amino acid substitutions at TCR*f* than the HA2 domain (Figure 5.7). Phylogenetic reconstruction shows that A/H3N2 phylogeny evolved into two major clades in this period, i.e., 3C.2a and 3C.3a lineages. Conserved T cell epitopes were observed at positions 163, 330, 347 and 408 across ten influenza seasons, while other positions appeared to demonstrate lineage-specific changes of T cell epitope phenotypes. For instance, epitopes at seven positions (103, 156, 169, 415, 540, 542) in clade

3C.3a1 showed amino acid substitutions at TCR*f* positions but remained Teff epitopes, T cell epitopes at position 486 transitioned from Teff epitope to non-binders, following a mutation at the HLA facing residue. TCR*f* mutations in epitopes at positions 273 and 276 were clade 3C.2a2-specific.

While amino acid substitutions at either the TCR*f* and/or HLA face were associated with T cell epitope phenotypic change, such as transitioning from Teff epitope to a T cell epitope that had no binding affinity or to a non-Teff epitope, some of the mutations at either the TCR*f* and/or HLA face did not affect T cell epitope phenotypic change. Interestingly, there were four T cell epitope phenotype switches observed at epitope position 325. At epitope position 540, the 3C, 3C.3 and 3C.3a lineages have non-Teff epitopes that are conserved at the TCR*f* residues whereas 3C.3a1 and 3C.2a1b lineages have non-Teff epitopes with mutations at the TCR*f*.

MDS and clustering analysis for contemporary dataset

We analyzed high dimensional T cell epitope data to further characterize T cell immune landscape for the contemporary dataset. As we did with the dataset A, we performed MDS and *k*-means clustering on HI titer, genetic distance and six T cell epitope properties (Figure 5.8 - 5.9). Interestingly, there were only three clusters detected for HI titer and separation between clusters were not as obvious as the older dataset (Figure 5.8). The genetic distance map showed a unique pattern and the dimension 1 of genetic distance extended into two major clades which appeared to reflect the topology of its phylogeny. There were nine distinct clusters detected using predicted T cell epitope binding affinity and T cell distance. When plotting individual dimension against time, both dimension 1 of the T cell epitope binding affinity and T cell epitope distance variables showed a linear progression over time, as vaccine viruses were updated. However, unlike the older dataset,

which showed a gradual change of antigenicity, in the contemporary dataset, there are mixed clusters observed within any given time frame and clusters are sustained for at least two years, suggesting more than one clade may be circulating at the same time (Figure 5.8).

A summary of clustering analysis results for each contemporary A/H3N2 viruses was shown as a heatmap and mapped to the tips of the respective phylogeny (Figure 5.9). For other T cell epitope properties, there were nine clusters detected for T cell epitope allele hit. There were fewer clusters (7-8 clusters) detected for T cell epitope properties that related to human cross-reactivity and these clusters are not well spread out.

Sensitivity check using PCA

We carry out PCA on multiple T cell epitope variables to identify the sources of variation (Supplemental Table C-3). These parameters include the sequence binding affinity score, epitope binding affinity score of each HLA alleles, TCR*f* unique count, JMX Human Homology score for each sequence, and JMX Human Homology score for each epitope. 26 components were analyzed and by plotting the proportion of variances against 26 PCs, we found that the first four PCs had better representation of the data. For the older dataset, we find that PC1 comprises of the binding affinity score of the nine HLA alleles, the overall binding affinity for each virus sequence and T cell epitope distance from MDS dimension 1 (Supplemental Table C-3A). The binding affinity score of the alleles DRB11501, DRB10801, DRB11101, DRB10401, DRB10101, the overall binding affinity for each virus sequence and T cell epitope distance from MDS dimension 2 contribute to PC2. For the contemporary dataset, six HLA alleles (DRB11101, DRB10301, DRB10101, DRB10801, DRB10701, DRB11501) and T cell epitope distance from MDS dimension 1 contribute to PC1 (Supplemental Table C-3B). PC2 consists of two HLA alleles

(DRB10401 and DRB11301) from MDS dimension 1, five alleles (DRB10401, DRB10901, DRB10101, DRB10301, and DRB11101) from MDS, total binding affinity score and T cell epitope distance from MDS dimension 2.

For dataset A, the results of the BSSVS asymmetric discrete trait substitution model showed a temporal pattern of older antigenic groups transitioning to newer antigenic groups (Figure 5.10). The highlighting of only rates with greater than 50% posterior support shows that the transition is only significant for two groups that are occur in direct succession (i.e., an antigenic group 1968 would only have significant transition to antigenic group found in 1972). This result is consistent with the pattern of antigenic drift observed in the phylogeny and accurately represents the temporal structuring of the discrete traits across the tree.

Discussion

While humoral immunity is certainly important for neutralizing viruses and correlates with protection from infection by influenza viruses, T cell-mediated immune response is thought to confer protection, even in the in the absence of influenza-specific neutralizing antibodies [170], [228]. Specific to our study, as vaccine efficacy of A/H3N2 is unpredictable and vaccine viruses that are antigenically different (in terms of B cell immunity) from emerging viruses can fail to protect, T cell epitope cross-conservation may be critical for cross-strain vaccine efficacy [124], [229]. Since the contribution of cell-mediated immunity to viral clearance and limiting severe disease has been recognized in recent years, we considered the possibility that T cell epitopes might also evolve in the course of viral evolution, and also considered the interaction between T cell epitopes and immune pressure due to vaccine introduction to be worth exploration. In this study, we employed a novel sequence-based approach to define the evolution of the CD4+ T cell

epitope landscape in two populations of viruses over time. Specifically, we evaluated the relationship between CD4+ T cell epitopes identified in A/H3N2 virus sequences using immunoinformatics tools and genetic drift to explore the impact of vaccine changes on T cell epitopes in circulating viruses and the relationship between T cell epitopes and HI titers over time.

Loss of T cell epitopes might contribute to a decrease in cross-protective effect between vaccine and circulating viruses. We identified T cell epitopes and assessed epitope similarities by focusing on TCR*f* residues of T cell epitopes. We defined cross-conserved Teff epitopes as T cell epitopes that were predicted to have effector function and that had identical TCR*f* residues when compared to another virus or a vaccine virus. Interestingly, there was a consistent decrease in the number of conserved virus-to-vaccine Teff epitopes after vaccine introduction in both an older dataset of IAV sequences and in contemporary sequences. This is consistent with the hypothesis that there may be immune escape from the virus selected for the seasonal vaccine. An increase in the number of unique Teff epitopes was also observed, which may reflect changes to T cell imprinting repertoire. The magnitude of conserved Teff epitopes changes between vaccine and circulating viruses was greater in recent seasons compared to seasons prior 2004. It is interesting to note that the average number of conserved epitopes that were lost between circulating and vaccine viruses, prior to introduction of a new vaccine, was eight conserved T effector epitopes. Therefore, in addition to evidence of declining HI efficacy, and phylogenetic distance, a drop of between eight to 11 conserved Teff epitopes may dictate an A/H3N2 vaccine update.

In terms of overall T cell epitope content, 9-mers that were predicted to have no binding affinity (non-binders) represented the largest group of epitopes followed by Teff epitopes that have less than four allele hits (non-promiscuous epitopes). Amino acid substitutions in a 9-mer sequence has been noted to contribute to a change in the phenotype of the T cell response. When mutations

increase conservation with human genome T cell epitopes, tolerance to the pathogen epitope may develop. This type of modification has also been observed in neoantigen mutations in cancer [209], [230]. In the context of this analysis of IAV epitopes, we note that phenotypic change of the epitopes was more frequent in HA1 domain of HA which is consistent with the fact that HA1 is the segment that encodes the viral surface of the HA protein and evolves at a higher rate than the HA2 domain [231]. Determining the T cell epitope phenotype in conserved and modified T cell epitopes may aid in identifying lineage-specific epitopes and ideally to better interpreting immune selection.

In addition, we noted that there were important differences in the MDS and clustering analysis results according to different T cell epitope properties. The introduction of T cell epitope analysis may improve the practice of antigenic cartography, which currently only uses HI assays data to represent the distance between antigenically varying virus strains. Our study demonstrated that binding affinity, allele hit, and unique T cell epitope content were important factors that distinguished virus isolates in the two-dimensional space. Viruses that were identified prior also 1990 form distinct clusters compared to strains post 1990 and in contemporary sequences, where mixed clusters were detected, suggesting that greater T cell epitope diversity of A/H3N2 circulating viruses in recent years. When plotting one of the dimensions against time, the progression of T cell epitope binding affinity and unique T cell epitope content is punctuated even though it shows linear relationship with genetic drift. Most importantly, T cell epitope properties such as binding affinity, allele hit, and unique T cell epitope content appeared to contribute to shaping the antigenic landscape.

We also identified some limitations to this study that could be addressed in future research. For instance, this is a retrospective study of published sequences as well as a collection of

sequences deposited in the public domain. The analysis was performed after influenza infections and viral evolution had already taken place. It will be important to perform the same type of study in a prospective manner as this may shed some insight and supplement into influenza virus prediction pipeline. An additional limitation for this study was the focus on HA alone. We limited ourselves to the analysis of HA because we have observed that internal antigens have only minimal variation in T cell epitope content over time. We do note that the surface protein neuraminidase (NA) may be relevant in stimulating CD4+ immune response, and therefore, it should probably be included in further study, for a more comprehensive perspective.

In conclusion, we applied a novel approach to antigenic cartography and influenza phylogeny, while also attempting to determine whether T cell responses contributed to the antigenic changes observed in A/H3N2 viruses. This study may inform future evaluations of new viruses where cross-reactive memory T cells also play a role, reducing the severity of infection and promoting B cell responses. Leveraging this information may lead to the development of vaccines that activate cross-strain protective cellular immunity from T cells as an effective means of protecting against the threat of constantly evolving influenza viruses [229]. All in all, the goal of investigating host CD4+ T cell immunity dynamics is to promote the consideration of cell-mediated immunity as one of the assessment criteria to improve the influenza vaccine selection process.

Table 5.1. A detailed breakdown of putative T cell epitopes predicted for the two datasets used in the study. There are three categories of predicted epitope phenotype. A putative T cell epitope that has a Z score at top 5% of a normal distribution is thought of having significant binding potential and is regarded as an "allele hit". Allele hit count is defined as the binding potential to an HLA allele and the value greater or equal to four is known to be promiscuous T cell epitope.

Putative T cell epitope phenotype	**Janus homology score**	**Allele hit Top 5%**	**Characteristic dataset**		**Contemporary dataset**	
Total			4421		7735	
Non-binder	0	0	3548		5979	
T_{eff}	≤ 5	< 4 ≥ 4	501 268	Total = 769	1059 476	Total = 1535
Non-T_{eff}	> 5	< 4 ≥ 4	90 14	Total = 104	185 3	Total = 221

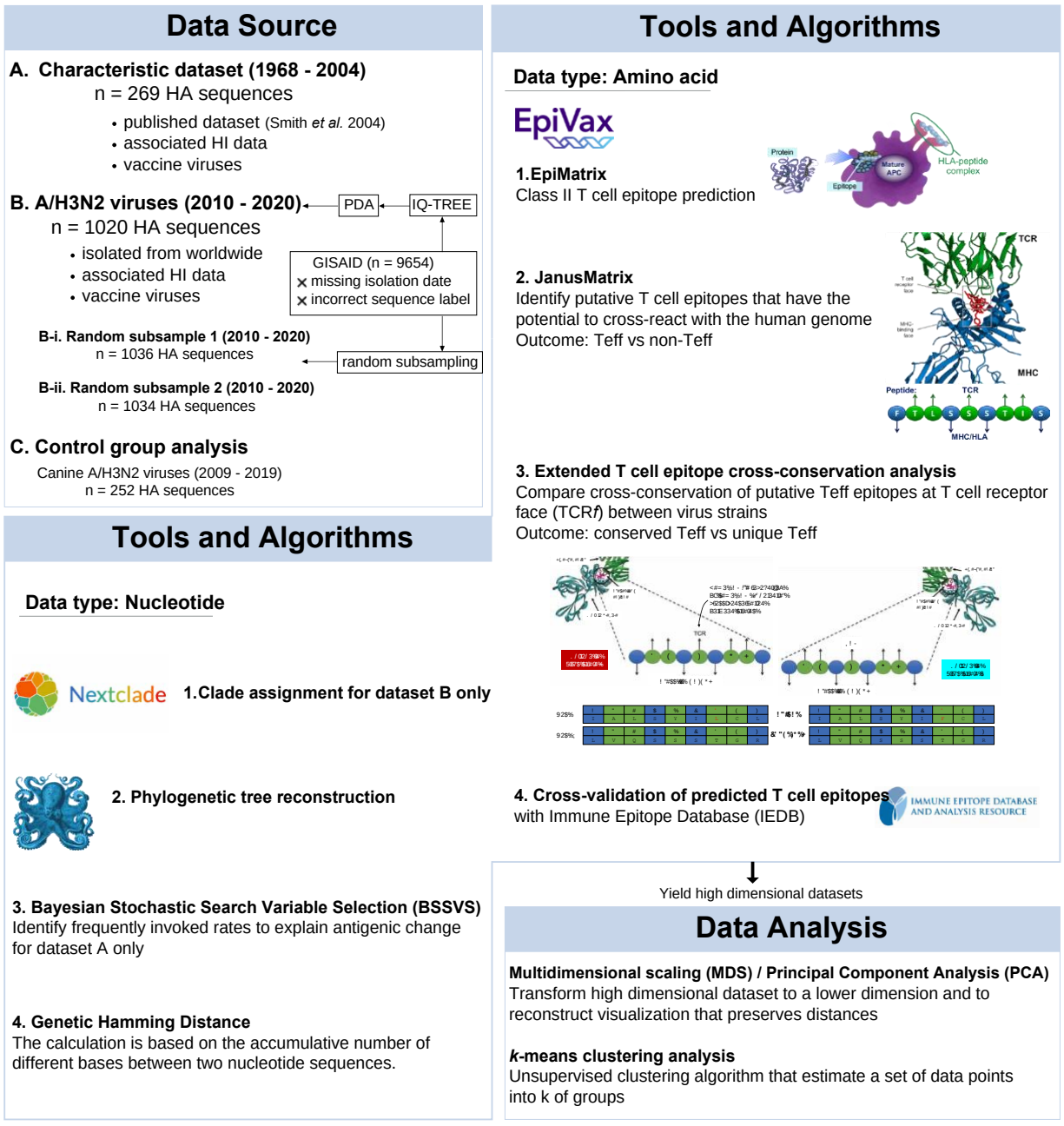


Figure 5.1. Summary of the complete analysis workflow. All mentioned tools and algorithms, as well as data analysis method are applied on data sources A and B independently. For the two random subsamples and the control group analysis, only T cell epitope prediction pipeline is executed.

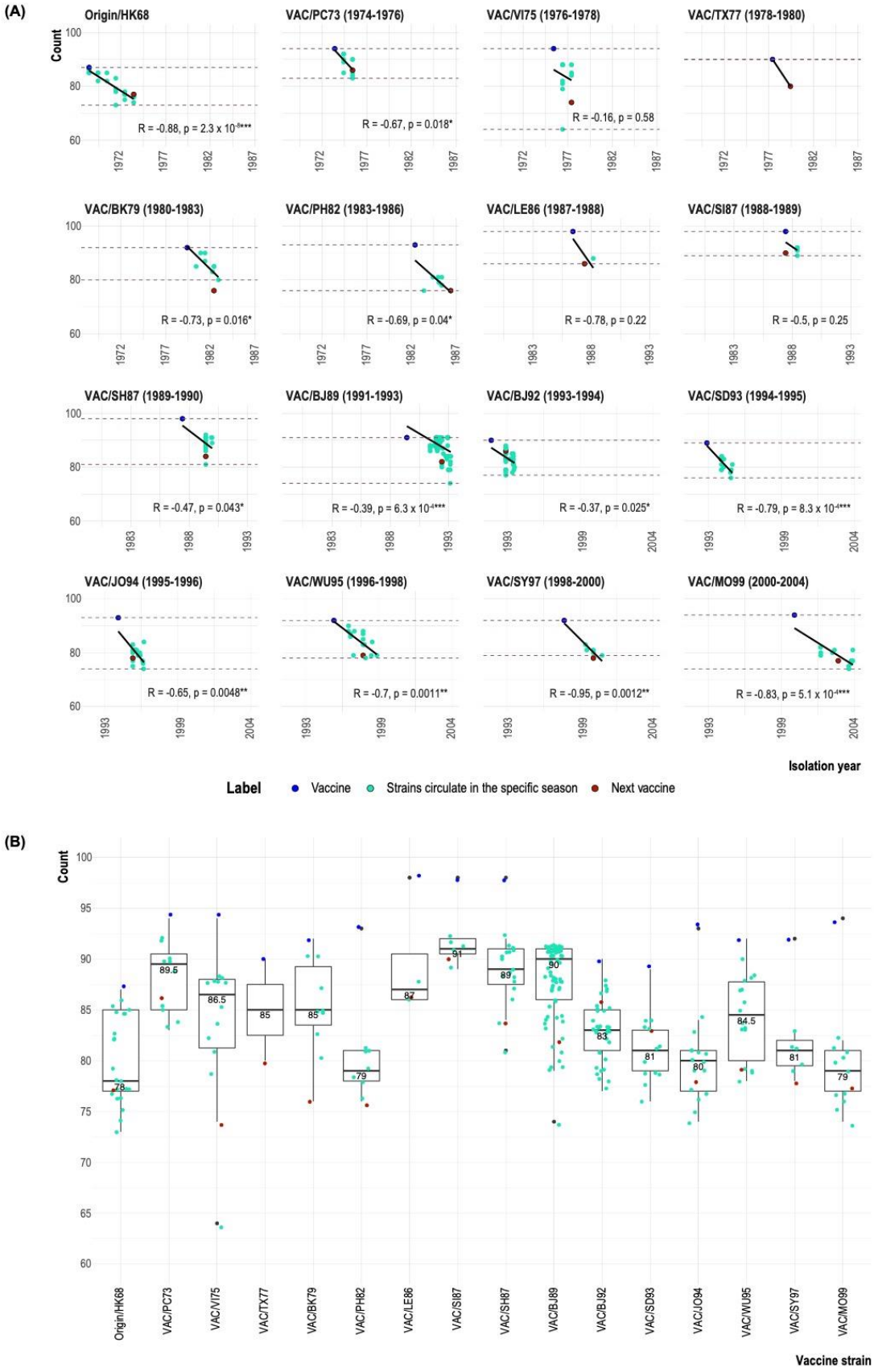

(A)
Origin/HK68
VAC/PC73 (1974-1976)
VAC/VI75 (1976-1978)
VAC/TX77 (1978-1980)
VAC/BK79 (1980-1983)
VAC/PH82 (1983-1986)
VAC/LE86 (1987-1988)
VAC/SI87 (1988-1989)
VAC/SH87 (1989-1990)
VAC/BJ89 (1991-1993)
VAC/BJ92 (1993-1994)
VAC/SD93 (1994-1995)
VAC/JO94 (1995-1996)
VAC/WU95 (1996-1998)
VAC/SY97 (1998-2000)
VAC/MO99 (2000-2004)
Count
R = -0.88, p = 2.3 x 10-5****
R = -0.67, p = 0.018*
R = -0.16, p = 0.58
R = -0.73, p = 0.016*
R = -0.69, p = 0.04*
R = -0.78, p = 0.22
R = -0.5, p = 0.25
R = -0.47, p = 0.043*
R = -0.39, p = 6.3 x 10-4****
R = -0.37, p = 0.025*
R = -0.79, p = 8.3 x 10-4****
R = -0.65, p = 0.0048**
R = -0.7, p = 0.0011**
R = -0.95, p = 0.0012**
R = -0.83, p = 5.1 x 10-4****
Isolation year
Label
Vaccine
Strains circulate in the specific season
Next vaccine
(B)
Count
Origin/HK68
VAC/PC73
VAC/VI75
VAC/TX77
VAC/BK79
VAC/PH82
VAC/LE86
VAC/SI87
VAC/SH87
VAC/BJ89
VAC/BJ92
VAC/SD93
VAC/JO94
VAC/WU95
VAC/SY97
VAC/MO99
Vaccine strain

Figure 5.2. Overview of conserved Teff epitopes over time with reference to H3N2 vaccine virus updates from 1968 to 2004. (A) Top panel highlights the changes of conserved Teff epitopes within the mentioned influenza seasons. The notation R and p are the correlation coefficient and significant value, respectively. The negative R values indicate negative association between T cell epitope content and strain isolation year. (B) The box plot summarizes the distribution of epitope change with reference to each H3N2 vaccine strain update. The solid line represents the median of conserved Teff epitope count.

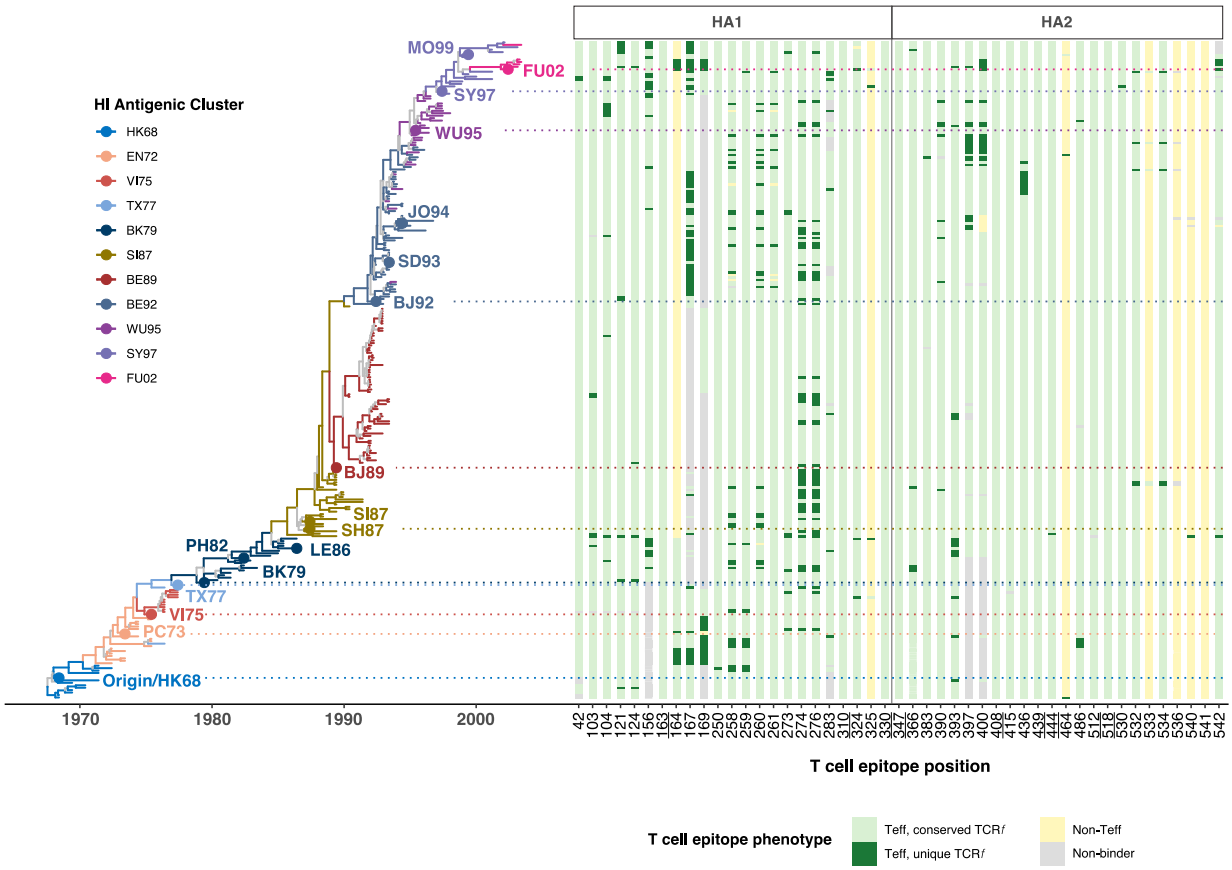


Figure 5.3. Tracking of T cell epitope phenotypes over the course of 30 years. A heatmap is used to demonstrate the transition of T cell epitope phenotypes and each row of observation at selected positions is mapped to the tips of A/H3N2 HA phylogeny.

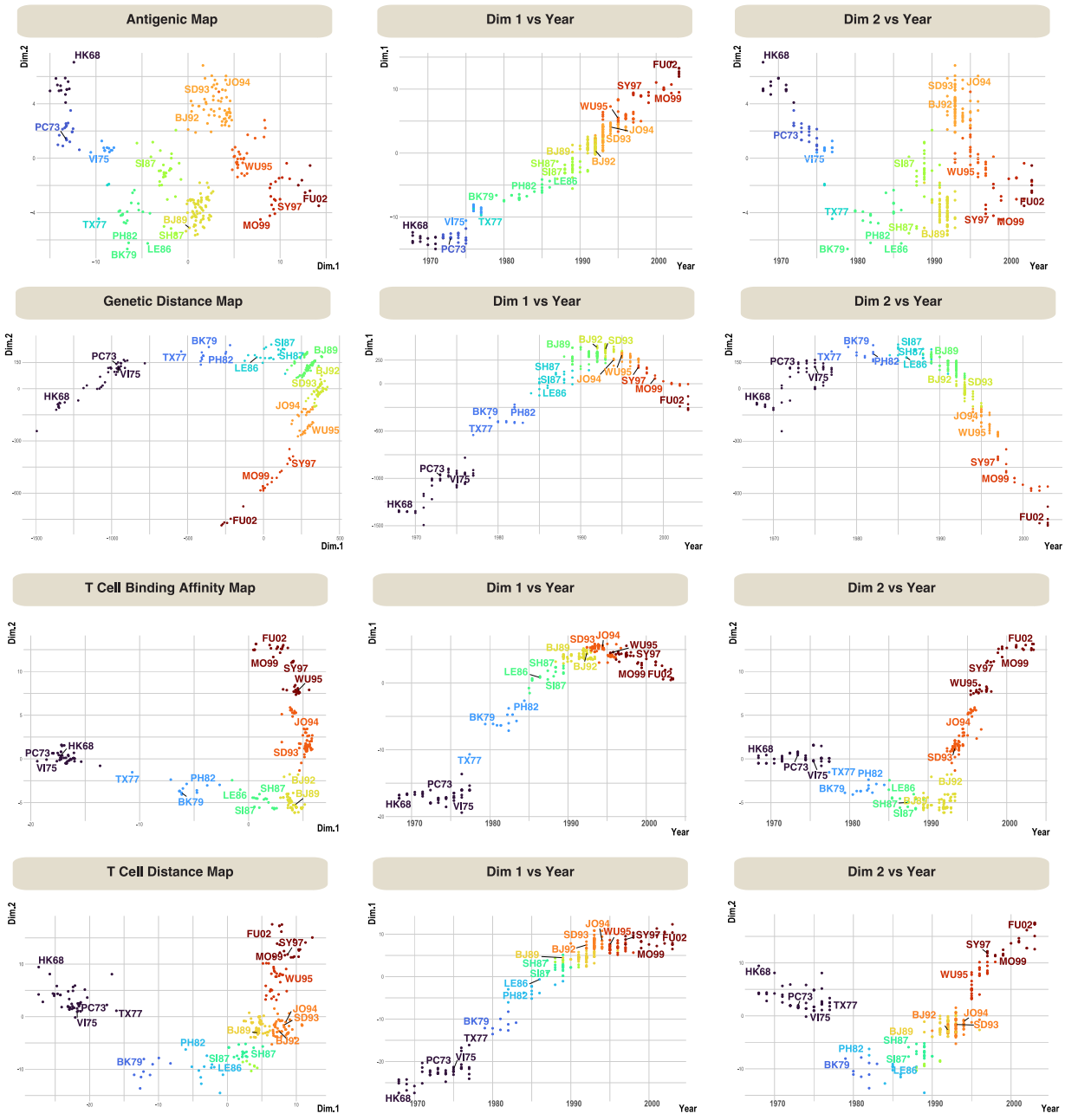


Figure 5.4. Comparison between HI measurement, genetic distance and T cell epitope prediction (the last two rows). The distribution of antigenic map is different from genetic distance and T cell epitope cluster map. The HI-defined antigenic clusters are presented as 11 clusters, as defined in Smith *et. al* [202]. Genetic distance resulted in nine groups, while the T cell epitope

properties (binding affinity and T cell distance and T cell cross-conservation) are grouped into six and nine clusters, respectively. Plotting each of the two dimensions against time, T cell epitope evolution exhibits linear rate, which corresponds with HA antigenic and genetic drift.

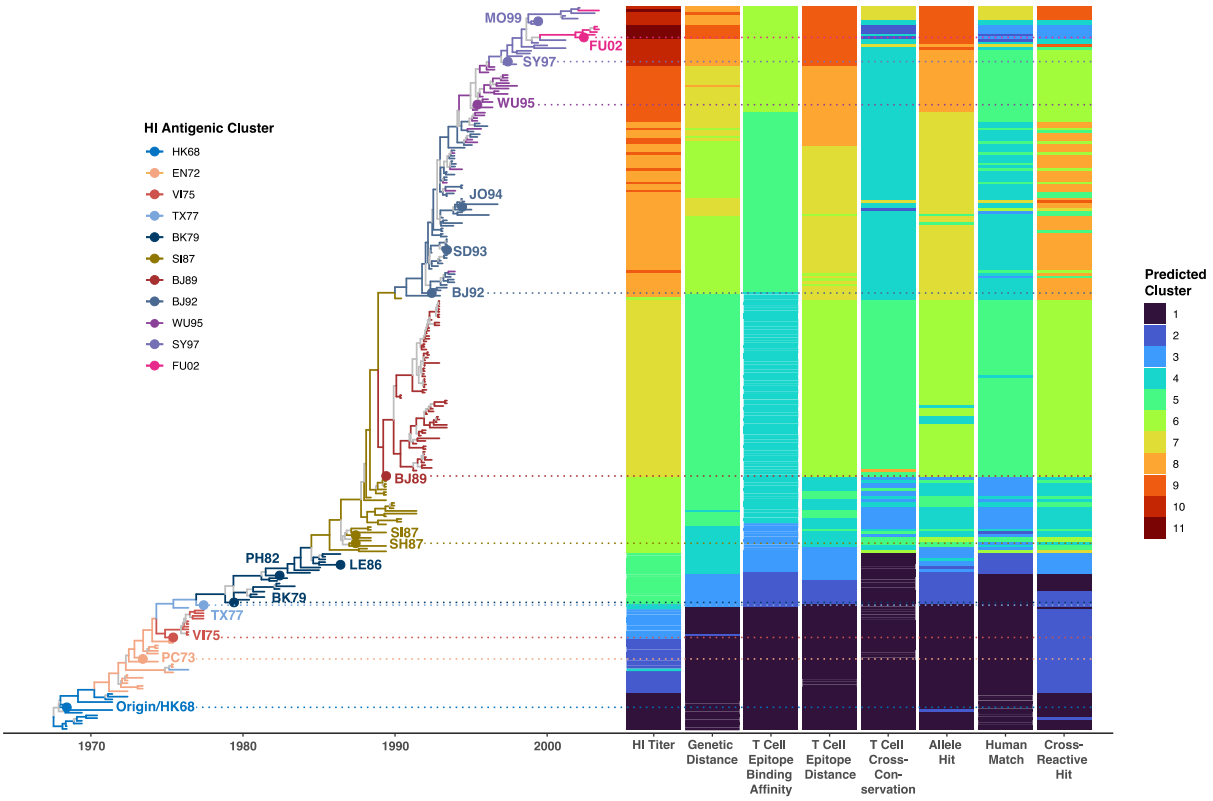


Figure 5.5. Phylogenetic tree reconstruction of the 269 HA nucleotide sequences of seasonal A/H3N2 using BEAST skyride coalescent model. The HI antigenic clusters are labelled together with the predicted clusters of genetic distance and T cell epitope properties are mapped to the tree tips. The corresponding color-coded antigenic group, genetic distance and T cell epitope clusters are shown in the heatmaps aligned with each associated strain.

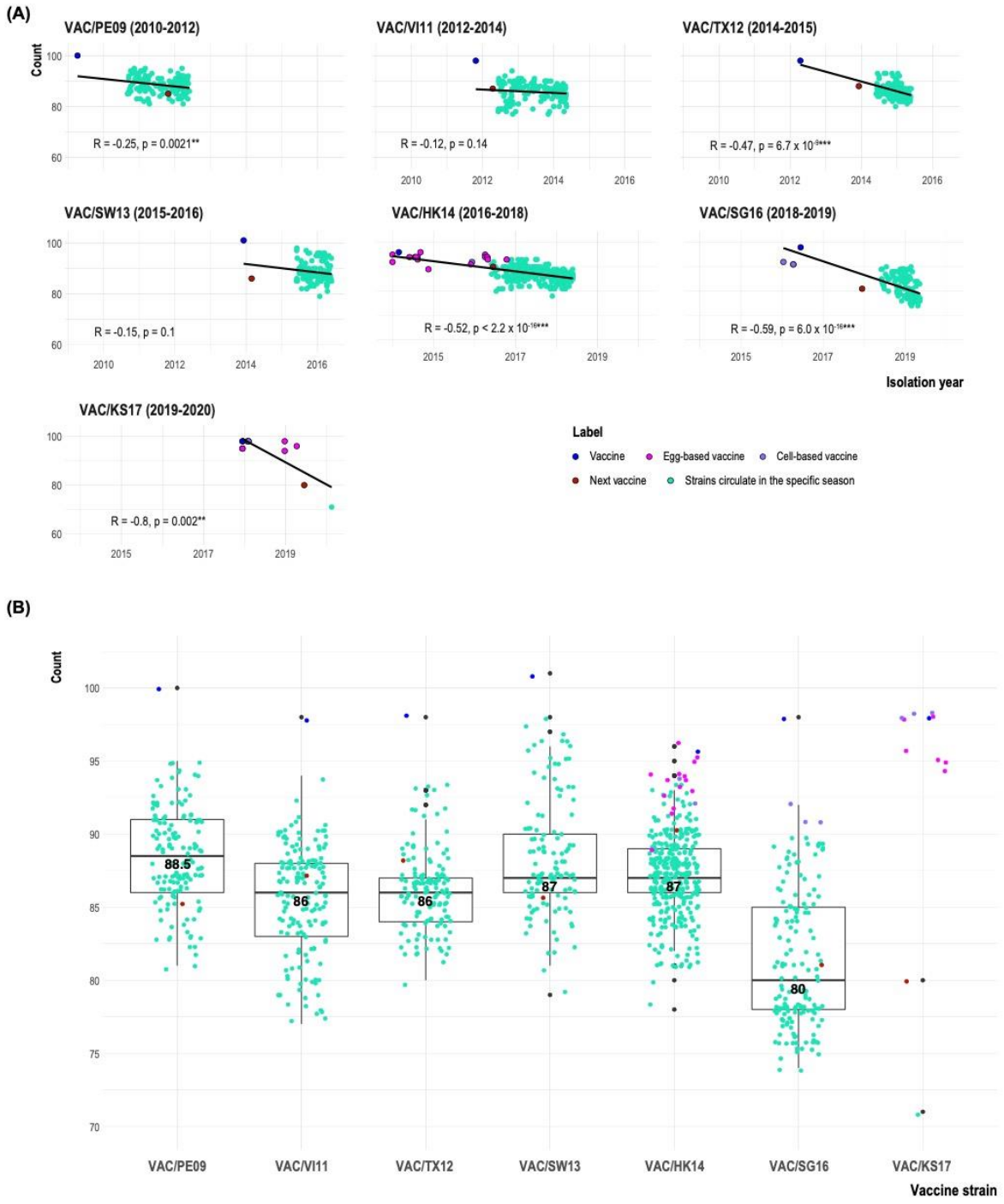


Figure 5.6. Overview of conserved Teff epitopes over time with reference to H3N2 vaccine virus updates from 2010 to 2020. Similar to the analysis performed using dataset A, the top panel (A) highlights the changes of conserved Teff epitopes within the mentioned influenza seasons. The

notation R and p are the correlation coefficient and significant value, respectively. The negative R values indicate negative association between T cell epitope content and strain isolation year. The box plot (B) summarizes the distribution of epitope change with reference to each H3N2 vaccine strain update. The solid line represents the median of conserved Teff epitope count.

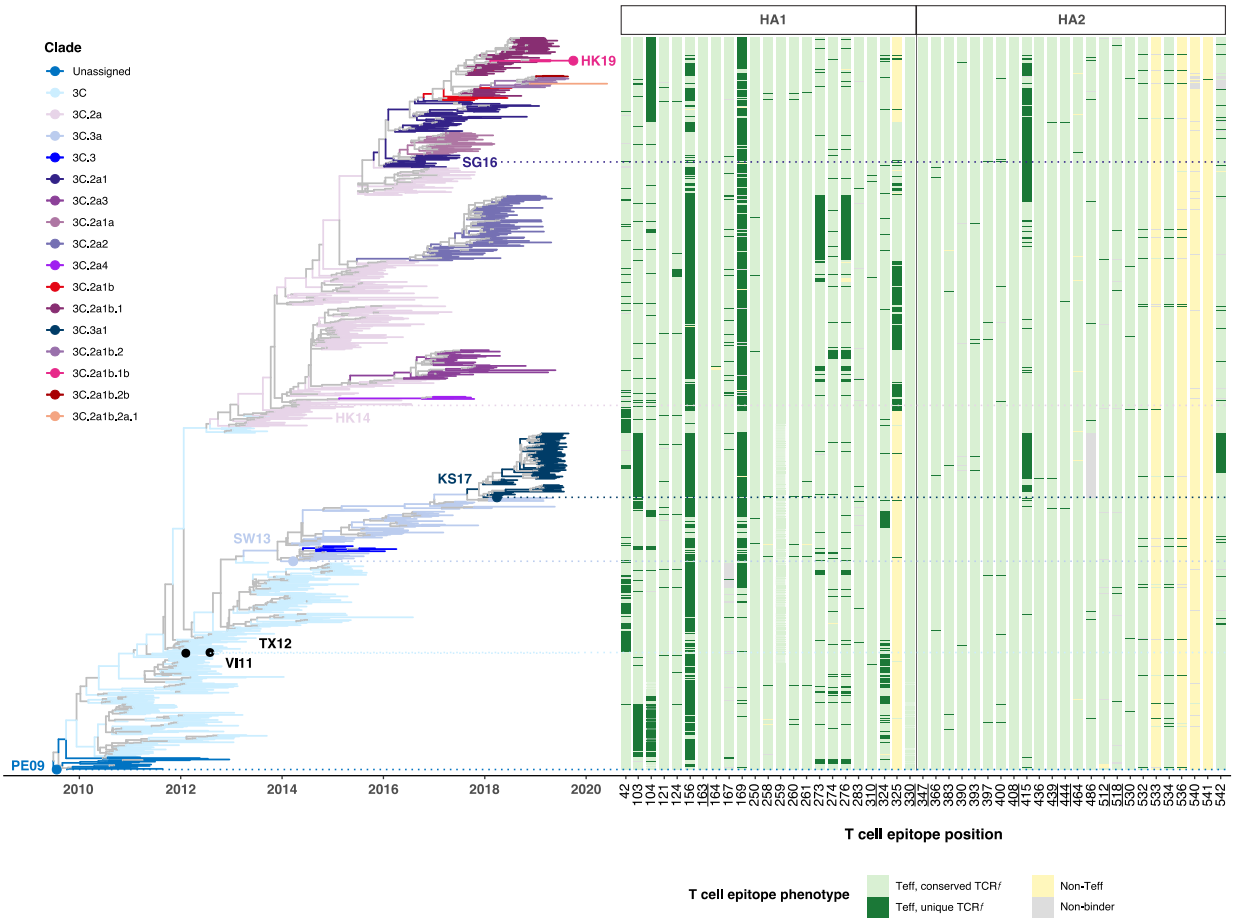


Figure 5.7. T cell epitope phenotypes dynamic of the contemporary dataset. A heatmap that showed T cell epitope phenotype change is used to demonstrate the transition of T cell epitope phenotypes. Each row of observation at selected positions is mapped to the tips of contemporary A/H3N2 HA phylogeny.

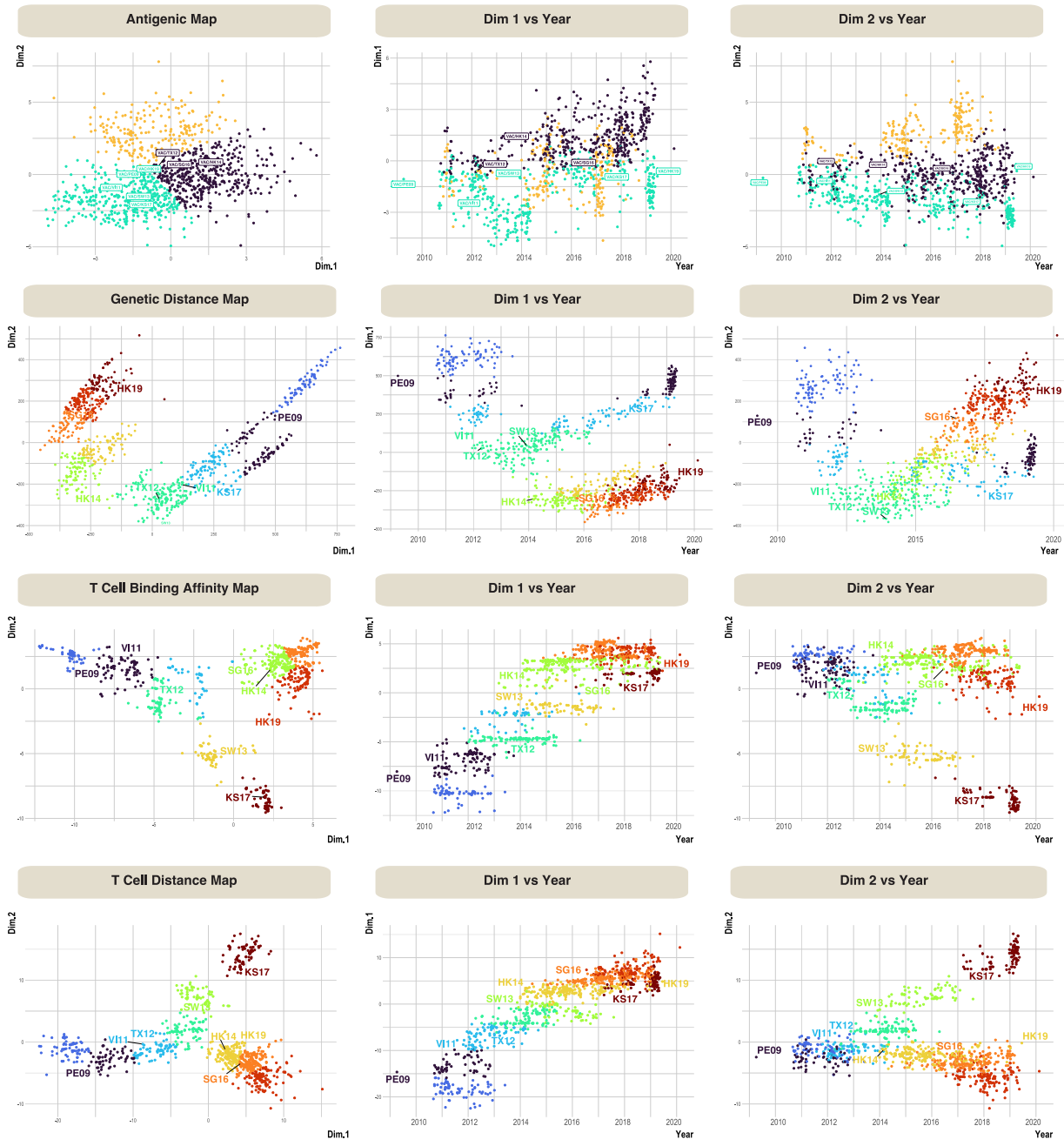


Figure 5.8. Comparison between HI measurement, genetic distance and T cell epitope prediction properties of the contemporary dataset B. The antigenic map of dataset B shows minimal distinction between clusters. The clustering pattern of genetic distance map is unique but T cell epitope cluster maps display similar bifurcation. Decomposing the two dimensions by plotting each of the dimension against time, T cell epitope evolution exhibits a linear distribution.

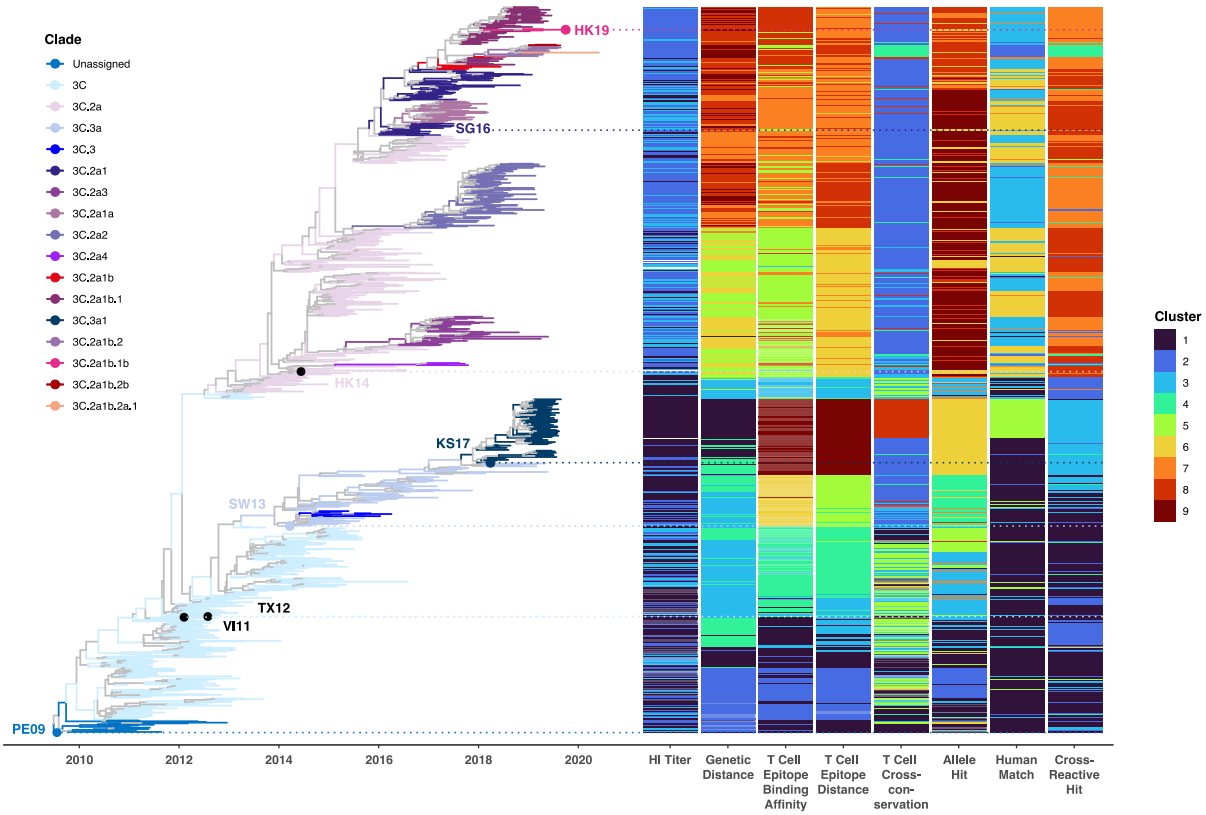


Figure 5.9. *k*-means clustering summary of the contemporary dataset. The contemporary A/H3N2 HA phylogeny is color-coded according to clade assignment. There are two major clades, i.e., 3C2a and 3C3a. The predicted clusters of HI titer, genetic distance and T cell epitope properties are shown in a heatmap and mapped to HA phylogeny tree tips.

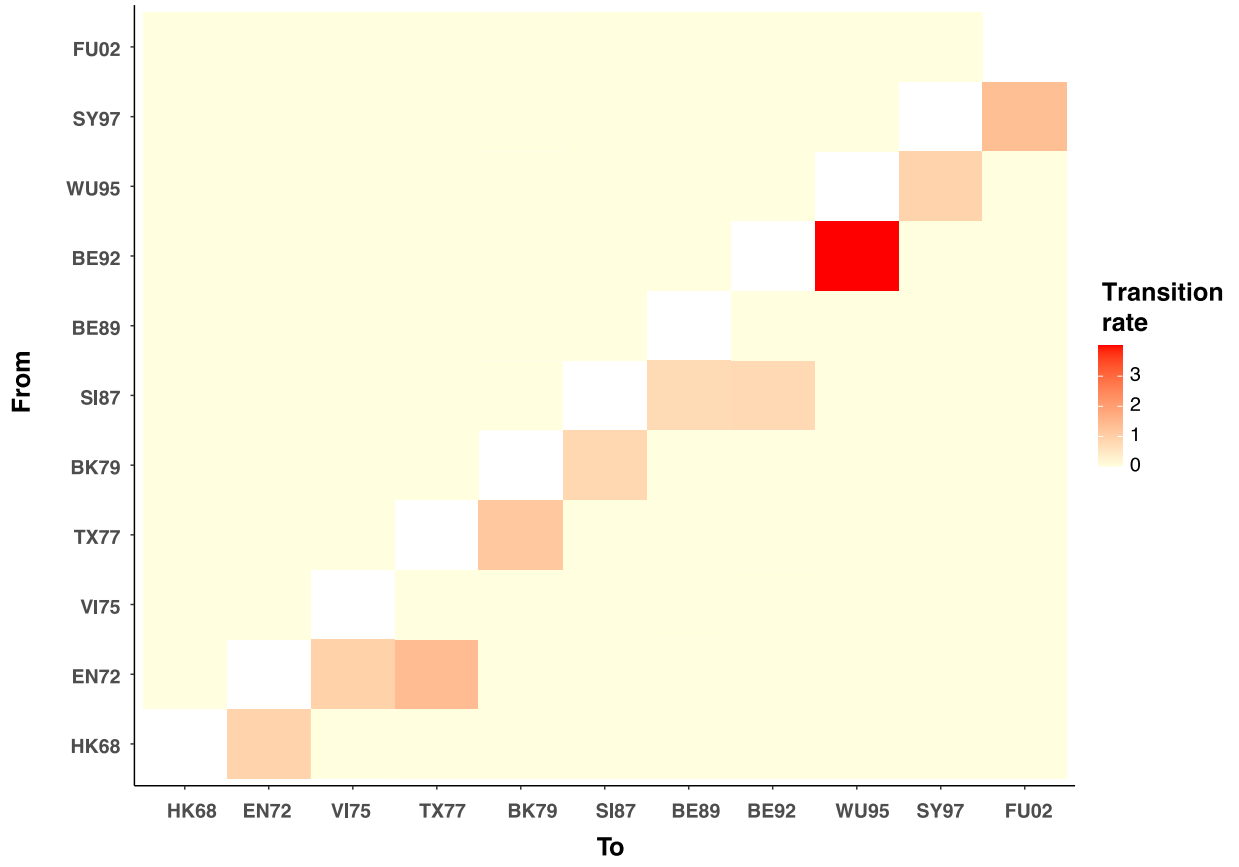


Figure 5.10. BSSVS substitution matrix for characteristic dataset. H3N2 mean transition rates estimated using the BSSVS for isolates collected between 1968 and 2004. Rates with posterior probability greater than 50% are visualized.

CHAPTER 6

SUMMARY AND CONCLUSIONS

The segmented RNA genome structure of influenza virus and its error-prone RNA-dependent RNA polymerases enable the virus to undergo antigenic drift, leading to adaptive immune responses escape in host species, including swine and humans [17]. Despite the effort in preventing and treating influenza for more than a century since the devastating 1918 influenza pandemic [14], influenza vaccination in either animal or humans is a longstanding concern due to the constant-evolving nature and adapting ability of influenza virus in multiple hosts.

Influenza vaccine generally stimulates the production of neutralizing antibody that targets the conformational epitopes on the globular domain of hemagglutinin (HA) which correlates to disease protection. Although neutralizing antibodies secreted by the B cells contribute directly to eliminating infection [72], [73], the role of T cells is equally critical in orchestrating host adaptive immune responses and providing help for optimal influenza virus clearance and host protection [97], [98]. Considering the current influenza vaccination approach, vaccine effectiveness can be suboptimal when a mismatch between vaccine strains and circulating strains occur. Thus, there has been a broad interest in potential strategies for improving cross-protective vaccination by incorporating T cell-induced immunity.

Animal and human challenge studies have shown that cell-mediated immune response contributes to limiting disease spread and reducing disease burden [3], [5], [103], [104]. When influenza vaccines and circulating strains are poorly matched, T cell responses targeting highly conserved influenza antigens could provide cross-protective immunity against different IAV

subtypes in the absence of cross-reactive neutralizing antibodies. Therefore, understanding cell-mediated immunity is equally important to combatting this annual burden. The main objective of this dissertation was to develop novel analysis framework to improve understanding of host adaptive immune profiles against influenza virus and further contribute to better vaccination strategies. In short, I applied a series of T cell epitope prediction tools (immunoinformatic) with phylogenetic approaches to evaluate T cell epitope content between influenza vaccines and circulating influenza strains in swine and human hosts.

Key findings

Highly conserved T cell epitopes are important in the absence of vaccine-induced antibodies when swine influenza vaccines and circulating influenza A virus (IAV) strains are poorly matched. The degree of T cell epitope conservation among circulating strains and vaccine strains can vary, which may explain differences in vaccine efficacy. However, vaccine efficacy assessment methods are lacking for swine IAV and besides conducting challenge studies, there is no method available for evaluating new vaccines against circulating strains for cross-protection by T cell epitopes. Therefore, the immunology-based approach established in aim 1 (Chapter 3) was to estimate T cell epitopes conservation of a prototype epitope-based swine IAV vaccine and determine the persistence of T cell epitope conservation over time. Using T cell epitope prediction algorithm (EpiMatrix) and T cell epitope comparison tool (EpiCC), I quantified highly conserved T cell epitopes based on pairwise homology score between the previously developed epitope-based vaccine and 1272 circulating swine IAV strains isolated from 2013 to 2017. Half of the total T cell epitopes included in the epitope-based vaccine were highly conserved and found in >1000 circulating swine IAV strains spanning the five-year period. In contrast, commercial swine IAV

vaccines developed in 2013 exhibited a declining conservation with the circulating IAV strains over the same 5-year period. Compared to the commercial whole antigen killed vaccine, the T cell epitopes in the prototype vaccine were highly conserved over time. This study was also supported by the experimental results where priming the prototype epitope-based vaccine with a commercial swine influenza vaccine aided in increasing immunogenicity, had equivalent body temperature control post-challenge, and reduced lung lesions and influenza antigen [153]. Taken together, conserved T cell epitope vaccines may be a useful adjunct for commercial swine influenza vaccines and to improve protection against influenza when antibodies are not cross-reactive. This finding is particularly relevant for the idea of 'universal influenza vaccines' that can boost immune responses in the absence of antibody cross-reactivity.

Extending from the notion of T cell-mediated immunity is crucial if there are no cross-reactive antibodies when influenza vaccines and circulating IAV strains are poorly matched, aim 2 (Chapter 4) documented on identifying epidemic-risk of the emergent H1N1 G4 swine influenza A virus (G4) that may impact swine and human populations through assessing T cell epitope conservation between emergent G4 and swine and human influenza vaccines strains as well as circulating strains. The total cross-conserved T cell epitope content in hemagglutinin (HA) sequences of human seasonal and experimental influenza vaccines, swine influenza vaccines from Europe and the United States (US) against G4 was identified and computed using EpiCC tool. From the results, the overall T cell epitope content of US commercial swine vaccines was poorly conserved with G4, with an average T cell epitope coverage of 35.7%. EpiCC scores for the comparison between current human influenza vaccines and circulating human influenza strains were also very low. In contrast, the T cell epitope coverage of a recent European swine influenza vaccine (HL03) was 65.8% against G4. Briefly, the European influenza vaccine, HL03, may

protect against emergent G4, yet poor T cell epitope cross-conservation between emergent G4 and swine and human influenza vaccines in the US may enable G4 to spread in swine and spillover to human populations in the absence of protective antibody response. As shown here, quantifying protein sequences based on immunological property, i.e., T cell epitope content, rather than sequence identity enables a comparison between vaccines and field strains/emergent strains, is useful for identifying whether existing vaccines might have efficacy (at the T cell epitope level) against an emerging infection. In short, this study demonstrated comparative assessment of T cell epitope conservation between existing vaccine strains against emergent viruses can inform epidemic risk in swine and human populations.

As antigenic mismatch with vaccine target would occur, vaccine effectiveness of each seasonal influenza vaccine remains uncertain. Considering the novel findings in the preceding chapters that supported the significance of T cell epitopes found in vaccine strains and conserved in circulating IAV strains, research aim 3 (Chapter 5) sought to characterize human conserved T cell epitopes of H3N2 seasonal IAV so as to determine their impact on influenza virus evolution and the host immune landscape. This chapter features a novel sequence-based approach to predicting potential T cell epitopes in large datasets of HA sequences, to examine how antigenic drift correlates with the diversity of T cell epitopes presented by the viral population over time. In this chapter, I found that conserved T effector T cell epitopes ($T_{CD4+eff}$) decline after new vaccine strain introduction, suggesting that there is potential immune escape from the imprinted T cell repertoire. I used multidimensional scaling and k-means clustering analyses to demonstrate that $T_{CD4+eff}$ can be categorized into clusters based on six T cell attributes. The results showed that at least four distinct clusters are detected prior to 1990, however, clusters overlap thereafter, indicating greater diversity of T cell epitopes. Interestingly, H3 HA T cell epitope evolution is

linear, which corresponds with HA genetic drift. All in all, T cell epitope landscape changes over time may be driven by immune pressure, resulting in conserved T cell epitopes loss and new epitopes' introduction. The decrease in epitope conservation with each vaccine strain after it is introduced suggests that new approaches to influenza vaccination strategies may be warranted.

Challenges

Alleles diversity and selection

The studies apply a T cell epitope prediction tool as the standard first step to identify important putative T cell epitopes that have significant binding affinity to selected sets of major histocompatibility complex (MHC) alleles. Therefore, information about swine leukocyte antigen (SLA) and human leukocyte antigen (HLA) allele diversity in swine and human populations are critically important both to evaluate whether the analysis is comprehensive and to be sure that the analysis focuses on the most prevalent SLA or HLA alleles in a broader population. While it has been determined that a human T cell epitope can leverage the concept of HLA supertypes for selection of few representative HLA alleles from different clusters to cover a high percentage of the HLA diversity in human population [121], [122], SLA allele selection for swine studies (Chapter 3 and 4) is more difficult. Thus, the selected alleles might not represent the complete SLA diversity in the United States (US) swine outbreed population. The SLA alleles included were reported as prevalence in outbred swine populations based on previous data [155], [156] and allele selection frequencies were determined using low-resolution haplotyping in a small number of pigs [123]. Although the selected set of alleles may serve as a first proxy for commonly expressed alleles, systematic studies of swine populations and higher-throughput methods for determination of higher resolution SLA-typing would further improve the ability of researchers to integrate SLA

diversity into epidemiological studies. Additionally, it may be possible to cluster SLA molecules into supertypes if the prevalence and diversity of the US swine SLA are better understood.

Data subsampling

With genomic sequencing techniques becoming more cost-effective and scalable, there are tremendous number of virus sequences deposited in public databases such as Global Initiative on Sharing Avian Influenza Data (GISAID) and Influenza Research Database (IRD), creating significant challenges for downstream analyses and data visualization. To process the large set of data prior sequence analysis, data subsampling is one of the most common ways to deal with the large amount of sequence data and to avoid overrepresentation of data for certain years or geographical areas. A phylogenetic-based subsampling method namely phylogenetic diversity analyzer (PDA) [178] was applied in Chapter 4 and 5 to generate subsample datasets while retaining the genetic diversity of the initial large dataset.

Retrospective study

The core data source of the three studies in this dissertation was based on sequence data made available in public domain, thus, the studies were performed retrospectively using information on events that have taken place in the past, i.e., influenza infection. While retrospective studies can provide meaningful insight and have important impact on evaluation of the disease, there are potential source of bias in retrospective studies. To avoid this confounding error, having more data subsampling replicates can ensure consistency and prospective studies should be performed as well.

Future directions

This work demonstrates the application of immunoinformatics to predict and quantify T cell epitope conservation of vaccine strains against circulating strains for cross-protection by conserved T cell epitopes, along with phylogenetic analysis to track conserved T cell epitope changes over time. As the increasing number of sequence data become available, it is laborious to execute enormous data with the present analysis workflow. Therefore, it would be interesting to automate the existing pipeline for more efficient productivity. A popular workflow management system used in genomic analysis, Snakemake, can be employed to create reproducible and scalable T cell epitope data analyses [232]. Snakemake workflows can set up a description of required software which can be automatically deployed to any execution environment such as server, cluster, grid, and cloud environments.

Another scope of improvement would be having experimental evidence to support prediction and correlate of protection for T cell epitopes. Even though the relationship between the level of conserved T cell epitope and vaccine efficacy was established based on presumptive-defined threshold, it is reasonable to question the accuracy of these tools for the prediction of human and swine influenza vaccine efficacy. As the protection thresholds for different vaccine strains may vary, having more experimental data available would help refine the putative thresholds used.

Concluding remarks

The continuous evasion of host adaptive immune responses is one of the factors that hinders the efforts of making potent and long-lasting influenza vaccines against the disease [17]. Ongoing influenza research focusing on vaccine platforms, antigen design and vaccine adjuvants

and among others are required to identify the best approaches to improve immunogenicity as well as to meet safety guidelines. Clinical studies demonstrate influenza vaccine efficacy are costly and challenging, thus, it is necessary that to utilize computational approaches to improve our understanding of the correlates of host immune protection and evaluation of the existing influenza virus vaccines [14].

In addition to studying antibody response, researchers are starting to pay attention to the contribution of cell-mediated immunity against influenza and leveraging immunoinformatics towards the next generation of influenza vaccine. All in all, this dissertation encompasses 1) the development of novel computational workflows using T cell epitope prediction tools for swine and human to evaluate existing and/or experimental vaccines and 2) the study of cross-conserved T cell epitopes shaped over time in swine and human hosts. The three research aims demonstrate the great potential of immunoinformatics and phylogenetics approaches, to aid in analyzing high dimensional T cell epitope data and to study host immunity related to infection or vaccination. The effort presented herein may guide the way for future advancements and applications in examining host T cell immunity in response to other pathogens.

REFERENCES

[1] Centers for Disease Control and Prevention [CDC], “Estimated Influenza Illnesses, Medical visits, Hospitalizations, and Deaths in the United States — 2019–2020 Influenza Season,” *Centers for Disease Control and Prevention*, 2020. .

[2] S. J. Yoo, T. Kwon, and Y. S. Lyoo, “Challenges of influenza A viruses in humans and animals and current animal vaccines as an effective control measure,” *Clinical and Experimental Vaccine Research*, vol. 7, no. 1. 2018, doi: 10.7774/cevr.2018.7.1.1.

[3] K. Scheible *et al.*, “CD8+ T cell immunity to 2009 pandemic and seasonal H1N1 influenza viruses,” *Vaccine*, 2011, doi: 10.1016/j.vaccine.2010.12.073.

[4] S. B. Morgan *et al.*, “Aerosol Delivery of a Candidate Universal Influenza Vaccine Reduces Viral Load in Pigs Challenged with Pandemic H1N1 Virus,” *J. Immunol.*, 2016, doi: 10.4049/jimmunol.1502632.

[5] T. M. Wilkinson *et al.*, “Preexisting influenza-specific CD4 + T cells correlate with disease protection against influenza challenge in humans,” *Nat. Med.*, vol. 18, no. 2, 2012, doi: 10.1038/nm.2612.

[6] S. Sridhar *et al.*, “Cellular immune correlates of protection against symptomatic pandemic influenza,” *Nat. Med.*, vol. 19, no. 10, pp. 1305–1312, 2013, doi: 10.1038/nm.3350.

[7] D. M. Morens, M. North, and J. K. Taubenberger, “Eyewitness accounts of the 1510 influenza pandemic in Europe,” *The Lancet*, vol. 376, no. 9756. 2010, doi: 10.1016/S0140-6736(10)62204-0.

[8] W. I. Beveridge, “The chronicle of influenza epidemics.,” *Hist. Philos. Life Sci.*, vol. 13,

no. 2, 1991.

[9] Elisha Hall, “Influenza,” *CDC. The Pink Book Home.* https://www.cdc.gov/vaccines/pubs/pinkbook/flu.html (accessed Apr. 27, 2022).

[10] A. Blagodatski *et al.*, “Avian influenza in wild birds and poultry: Dissemination pathways, monitoring methods, and virus ecology,” *Pathogens*, vol. 10, no. 5. MDPI AG, May 01, 2021, doi: 10.3390/pathogens10050630.

[11] Centers for Disease Control and Prevention [CDC], “Estimated Influenza Illnesses, Medical visits, Hospitalizations, and Deaths in the United States — 2019–2020 Influenza Season,” *Centers for Disease Control and Prevention*, 2020. https://www.cdc.gov/flu/about/burden/2019-2020.html#anchor_1601407136591 (accessed Jan. 08, 2021).

[12] C. Viboud, W. J. Alonso, and L. Simonsen, “Influenza in tropical regions,” *PLoS Medicine*, vol. 3, no. 4. 2006, doi: 10.1371/journal.pmed.0030089.

[13] V. N. Petrova and C. A. Russell, “The evolution of seasonal influenza viruses,” *Nature Reviews Microbiology*, vol. 16, no. 1. 2018, doi: 10.1038/nrmicro.2017.118.

[14] F. Krammer *et al.*, “Influenza,” *Nature Reviews Disease Primers*. 2018, doi: 10.1038/s41572-018-0002-y.

[15] A. D. Osterhaus, “Influenza B Virus in Seals,” *Science*, vol. 288. pp. 1051–1053, 2000, doi: 10.1126/science.288.5468.1051.

[16] G. Yuanji, J. Fengen, and W. Ping, “Isolation of influenza C virus from pigs and experimental infection of pigs with influenza C virus,” *J. Gen. Virol.*, vol. 64, pp. 177–182, 1983, doi: 10.1099/0022-1317-64-1-177.

[17] N. M. Bouvier and P. Palese, “The biology of influenza viruses,” *Vaccine*, vol. 26, 2008,

doi: 10.1016/j.vaccine.2008.07.039.

[18] B. K. Sederdahl and J. V. Williams, “Epidemiology and clinical characteristics of influenza C virus,” *Viruses*, vol. 12, no. 1. 2020, doi: 10.3390/v12010089.

[19] W. S. Barclay, A. V Cauldwell, J. S. Long, O. Moncorge, O. Moncorgé, and W. S. Barclay, “Viral determinants of influenza A virus host range.,” *J. Gen. Virol.*, vol. 95, no. Pt 6, pp. 1193–210, Jun. 2014, doi: 10.1099/vir.0.062836-0.

[20] R. Lamb, R. Krug, and D. Knipe, “Orthomyxoviridae: The Viruses and Their Replication,” in *Fields Virology*, 2001, pp. 1487–1531.

[21] T. Watanabe, S. Watanabe, G. Neumann, H. Kida, and Y. Kawaoka, “Immunogenicity and protective efficacy of replication-incompetent influenza virus-like particles.,” *J. Virol.*, vol. 76, pp. 767–773, 2002, doi: 10.1128/JVI.76.2.767-773.2002.

[22] N. J. Cox, G. Neumann, and R. O. Donis, “Orthomyxoviruses: influenza,” *Eras*, vol. 2003, pp. 634–698, 2004, doi: 10.1002/9780470688618.

[23] S. Federhen, “The NCBI Taxonomy database,” *Nucleic Acids Res.*, vol. 40, 2012, doi: 10.1093/nar/gkr1178.

[24] G. Neumann, H. Chen, G. F. Gao, Y. Shu, and Y. Kawaoka, “H5N1 influenza viruses: outbreaks and biological properties,” *Cell Res*, vol. 20, no. 1, pp. 51–61, Nov. 2009, [Online]. Available: http://dx.doi.org/10.1038/cr.2009.124.

[25] D. Dou, R. Revol, H. Östbye, H. Wang, and R. Daniels, “Influenza A virus cell entry, replication, virion assembly and movement,” *Frontiers in Immunology*, vol. 9, no. JUL. 2018, doi: 10.3389/fimmu.2018.01581.

[26] Centers for Disease Control and Prevention [CDC], “Key Facts about Influenza (Flu) & Flu Vaccine | Seasonal Influenza (Flu) | CDC,” *Centers for Disease Control and*

Prevention (CDC), 2015. http://www.cdc.gov/flu/keyfacts.htm (accessed Apr. 04, 2016).

[27] P. M. Colman, J. N. Varghese, and W. G. Laver, "Structure of the catalytic and antigenic sites in influenza virus neuraminidase.," *Nature*, vol. 303, no. 5912, pp. 41–44, May 1983.

[28] S. Rampersad and P. Tennant, "Replication and Expression Strategies of Viruses," in *Viruses: Molecular Biology, Host Interactions, and Applications to Biotechnology*, 2018.

[29] W. Hao, L. Wang, and S. Li, "Roles of the non-structural proteins of influenza a virus," *Pathogens*, vol. 9, no. 10. 2020, doi: 10.3390/pathogens9100812.

[30] P. Palese and M. Shaw, *Orthomyxoviridae: The Viruses and their Replication*, Second. Philadelphia: Lippincott Williams & Wilkins, 2007.

[31] P. Palese and R. W. Compans, "Inhibition of influenza virus replication in tissue culture by 2 deoxy 2,3 dehydro N trifluoroacetylneuraminic acid (FANA): mechanism of action," *J. Gen. Virol.*, vol. 33, no. 1, 1976, doi: 10.1099/0022-1317-33-1-159.

[32] M. C. Zambon, "The pathogenesis of influenza in humans," *Reviews in Medical Virology*, vol. 11, no. 4. pp. 227–241, 2001, doi: 10.1002/rmv.319.

[33] J. K. Park *et al.*, "Pre-existing immunity to influenza virus hemagglutinin stalk might drive selection for antibody-escape mutant viruses in a human challenge model," *Nat. Med.*, vol. 26, no. 8, 2020, doi: 10.1038/s41591-020-0937-x.

[34] R. G. Woolthuis, C. H. Van Dorp, C. Keşmir, R. J. De Boer, and M. Van Boven, "Long-term adaptation of the influenza A virus by escaping cytotoxic T-cell recognition," *Sci. Rep.*, vol. 6, 2016, doi: 10.1038/srep33334.

[35] P. R. Saunders-Hastings and D. Krewski, "Reviewing the history of pandemic influenza: Understanding patterns of emergence and transmission," *Pathogens*, vol. 5, no. 4. 2016, doi: 10.3390/pathogens5040066.

[36] L. Glaser *et al.*, “A single amino acid substitution in 1918 influenza virus hemagglutinin changes receptor binding specificity,” *J Virol*, vol. 79, pp. 11533–11536, 2005, doi: 79/17/11533 [pii]\r10.1128/JVI.79.17.11533-11536.2005.

[37] G. W. Chen *et al.*, “Genomic signatures of human versus avian influenza A viruses,” *Emerg. Infect. Dis.*, vol. 12, no. 9, pp. 1353–1360, 2006, doi: 10.3201/eid1209.060276.

[38] E. K. Subbarao, W. London, and B. R. Murphy, “A single amino acid in the PB2 gene of influenza A virus is a determinant of host range.,” *J. Virol.*, vol. 67, pp. 1761–1764, 1993.

[39] K. Shinya, S. Hamm, M. Hatta, H. Ito, T. Ito, and Y. Kawaoka, “PB2 amino acid at position 627 affects replicative efficiency, but not cell tropism, of Hong Kong H5N1 influenza A viruses in mice,” *Virology*, vol. 320, no. 2, pp. 258–66, Mar. 2004, doi: 10.1016/j.virol.2003.11.030.

[40] S. J. Lycett, F. Duchatel, and P. Digard, “A brief history of bird flu,” *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 374, no. 1775. 2019, doi: 10.1098/rstb.2018.0257.

[41] C. Brown, “Spillover: Animal Infection and the Next Human Pandemic,” *Emerg. Infect. Dis.*, 2013, doi: 10.3201/eid1902.121694.

[42] D. M. Morens and A. S. Fauci, “The 1918 influenza pandemic: insights for the 21st century.,” *J. Infect. Dis.*, vol. 195, pp. 1018–1028, 2007, doi: 10.1086/511989.

[43] G. J. D. Smith *et al.*, “Origins and evolutionary genomics of the 2009 swine-origin H1N1 influenza a epidemic,” *Nature*, 2009, doi: 10.1038/nature08182.

[44] M. I. Nelson, M. R. Gramer, A. L. Vincent, and E. C. Holmes, “Global transmission of influenza viruses from humans to swine,” *J. Gen. Virol.*, vol. 93, no. PART 10, 2012, doi: 10.1099/vir.0.044974-0.

[45] D. S. Rajao, A. L. Vincent, and D. R. Perez, “Adaptation of human influenza viruses to swine,” *Frontiers in Veterinary Science*, vol. 5, no. JAN. 2019, doi: 10.3389/fvets.2018.00347.

[46] A. L. Vincent, K. M. Lager, and T. K. Anderson, “A brief introduction to influenza a virus in swine,” *Methods Mol. Biol.*, 2014, doi: 10.1007/978-1-4939-0758-8_20.

[47] A. S. Bowman, J. D. Workman, J. M. Nolting, S. W. Nelson, and R. D. Slemons, “Exploration of risk factors contributing to the presence of influenza A virus in swine at agricultural fairs,” *Emerg. Microbes Infect.*, vol. 3, 2014, doi: 10.1038/emi.2014.5.

[48] G. Simon *et al.*, “European surveillance network for influenza in pigs: Surveillance programs, diagnostic tools and swine influenza virus subtypes identified in 14 European countries from 2010 to 2013,” *PLoS One*, vol. 9, no. 12, 2014, doi: 10.1371/journal.pone.0115815.

[49] T. K. Anderson *et al.*, “A Phylogeny-Based Global Nomenclature System and Automated Annotation Tool for H1 Hemagglutinin Genes from Swine Influenza A Viruses,” *mSphere*, vol. 1, no. 6, 2016, doi: 10.1128/msphere.00275-16.

[50] M. I. Nelson *et al.*, “Global migration of influenza A viruses in swine,” *Nat. Commun.*, 2015, doi: 10.1038/ncomms7696.

[51] D. Henritzi *et al.*, “Surveillance of European Domestic Pig Populations Identifies an Emerging Reservoir of Potentially Zoonotic Swine Influenza A Viruses,” *Cell Host Microbe*, vol. 28, no. 4, 2020, doi: 10.1016/j.chom.2020.07.006.

[52] H. Sun *et al.*, “Prevalent Eurasian avian-like H1N1 swine influenza virus with 2009 pandemic viral genes facilitating human infection,” *Proc. Natl. Acad. Sci. U. S. A.*, 2020, doi: 10.1073/pnas.1921186117.

[53] L. A. Reperant, T. Kuiken, and A. D. M. E. Osterhaus, “Adaptive pathways of zoonotic influenza viruses: From exposure to establishment in humans,” *Vaccine*, vol. 30, no. 30. pp. 4419–4434, 2012, doi: 10.1016/j.vaccine.2012.04.049.

[54] J. Stevens, A. L. Corper, C. F. Basler, J. K. Taubenberger, P. Palese, and I. A. Wilson, “Structure of the Uncleaved Human H1 Hemagglutinin from the Extinct 1918 Influenza Virus,” *Science (80-.).*, vol. 303, no. 5665, pp. 1866–1870, Mar. 2004, [Online]. Available: http://science.sciencemag.org/content/303/5665/1866.abstract.

[55] T. Ito *et al.*, “Molecular Basis for the Generation in Pigs of Influenza A Viruses with Pandemic Potential,” *J. Virol.*, vol. 72, no. 9, 1998, doi: 10.1128/jvi.72.9.7367-7373.1998.

[56] G. C. Gray and G. Kayali, “Facing pandemic influenza threats: The importance of including poultry and swine workers in preparedness plans,” *Poult. Sci.*, vol. 88, no. 4, 2009, doi: 10.3382/ps.2008-00335.

[57] F. J. U. M. Van Der Meer, K. Orsel, and H. W. Barkema, “The new influenza a H1N1 virus: Balancing on the interface of humans and animals,” *Canadian Veterinary Journal*, vol. 51, no. 1. 2010.

[58] K. Van Reeth, “Cytokines in the pathogenesis of influenza,” in *Veterinary Microbiology*, 2000, vol. 74, no. 1–2, doi: 10.1016/S0378-1135(00)00171-1.

[59] I. Julkunen, K. Melén, M. Nyqvist, J. Pirhonen, T. Sareneva, and S. Matikainen, “Inflammatory responses in influenza A virus infection,” *Vaccine*, vol. 19, no. SUPPL. 1. 2000, doi: 10.1016/S0264-410X(00)00275-9.

[60] S. Crotta *et al.*, “Type I and Type III Interferons Drive Redundant Amplification Loops to Induce a Transcriptional Signature in Influenza-Infected Airway Epithelia,” *PLoS Pathog.*, vol. 9, no. 11, 2013, doi: 10.1371/journal.ppat.1003773.

[61] A. Ibricevic *et al.*, “Influenza Virus Receptor Specificity and Cell Tropism in Mouse and Human Airway Epithelial Cells,” *J. Virol.*, vol. 80, no. 15, 2006, doi: 10.1128/jvi.02677-05.

[62] A. C. Kalil and P. G. Thomas, “Influenza virus-related critical illness: Pathophysiology and epidemiology,” *Critical Care*, vol. 23, no. 1. 2019, doi: 10.1186/s13054-019-2539-x.

[63] J. Helft *et al.*, “Cross-presenting CD103+ dendritic cells are protected from influenza virus infection,” in *Journal of Clinical Investigation*, 2012, vol. 122, no. 11, doi: 10.1172/JCI60659.

[64] M. C. W. Chan *et al.*, “Proinflammatory cytokine responses induced by influenza A (H5N1) viruses in primary human alveolar and bronchial epithelial cells,” *Respir. Res.*, vol. 6, 2005, doi: 10.1186/1465-9921-6-135.

[65] A. D. T. Barrett and L. R. Stanberry, *Vaccines for biodefense and emerging and neglected diseases*. 2009.

[66] L. Zhu *et al.*, “High Level of Neutrophil Extracellular Traps Correlates with Poor Prognosis of Severe Influenza A Infection,” *J. Infect. Dis.*, vol. 217, no. 3, 2018, doi: 10.1093/infdis/jix475.

[67] E. E. Waffarn and N. Baumgarth, “Protective B Cell Responses to Flu—No Fluke!,” *J. Immunol.*, vol. 186, no. 7, 2011, doi: 10.4049/jimmunol.1002090.

[68] S. I. Tamura and T. Kurata, “Defense mechanisms against influenza virus infection in the respiratory tract mucosa,” *Japanese Journal of Infectious Diseases*, vol. 57, no. 6. 2004.

[69] M. D. Lubeck and W. Gerhard, “Topological mapping of antigenic sites on the influenza A/PR/8/34 virus hemagglutinin using monoclonal antibodies,” *Virology*, vol. 113, no. 1, 1981, doi: 10.1016/0042-6822(81)90136-7.

[70] A. J. Caton, G. G. Brownlee, J. W. Yewdell, and W. Gerhard, “The antigenic structure of the influenza virus A/PR/8/34 hemagglutinin (H1 subtype),” *Cell*, vol. 31, no. 2 PART 1, 1982, doi: 10.1016/0092-8674(82)90135-0.

[71] H. O. Padilla-Quirarte, D. V. Lopez-Guerrero, L. Gutierrez-Xicotencatl, and F. Esquivel-Guadarrama, “Protective antibodies against influenza proteins,” *Frontiers in Immunology*, vol. 10, no. JULY. 2019, doi: 10.3389/fimmu.2019.01677.

[72] W. Smith, C. H. Andrewes, and P. P. Laidlaw, “A VIRUS OBTAINED FROM INFLUENZA PATIENTS,” *Lancet*, vol. 222, no. 5732, 1933, doi: 10.1016/S0140-6736(00)78541-2.

[73] C. C. Lee *et al.*, “An Effective Neutralizing Antibody Against Influenza Virus H1N1 from Human B Cells,” *Sci. Rep.*, vol. 9, no. 1, 2019, doi: 10.1038/s41598-019-40937-4.

[74] L. Hangartner, R. M. Zinkernagel, and H. Hengartner, “Antiviral antibody responses: The two extremes of a wide spectrum,” *Nature Reviews Immunology*, vol. 6, no. 3. 2006, doi: 10.1038/nri1783.

[75] J. W. Yewdell and W. Gerhard, “Antigenic characterization of viruses by monoclonal antibodies.,” *Annual review of microbiology*, vol. 35. 1981, doi: 10.1146/annurev.mi.35.100181.001153.

[76] D. Angeletti *et al.*, “Defining B cell immunodominance to viruses,” *Nat. Immunol.*, vol. 18, no. 4, 2017, doi: 10.1038/ni.3680.

[77] Y. Okuno, Y. Isegawa, F. Sasao, and S. Ueda, “A common neutralizing epitope conserved between the hemagglutinins of influenza A virus H1 and H2 strains,” *J. Virol.*, vol. 67, no. 5, 1993, doi: 10.1128/jvi.67.5.2552-2558.1993.

[78] J. J. Guthmiller *et al.*, “Broadly neutralizing antibodies target a haemagglutinin anchor

epitope," *Nature*, vol. 602, no. 7896, 2022, doi: 10.1038/s41586-021-04356-8.

[79] D. Stadlbauer *et al.*, "Broadly protective human antibodies that target the active site of influenza virus neuraminidase," *Science (80-.).*, vol. 366, no. 6464, 2019, doi: 10.1126/science.aay0678.

[80] T. J. Wohlbold *et al.*, "Broadly protective murine monoclonal antibodies against influenza B virus target highly conserved neuraminidase epitopes," *Nat. Microbiol.*, vol. 2, no. 10, 2017, doi: 10.1038/s41564-017-0011-8.

[81] Y. Q. Chen *et al.*, "Influenza Infection in Humans Induces Broadly Cross-Reactive and Protective Neuraminidase-Reactive Antibodies," *Cell*, vol. 173, no. 2, 2018, doi: 10.1016/j.cell.2018.03.030.

[82] L. Gérentes, N. Kessler, and M. Aymard, "Difficulties in standardizing the neuraminidase content of influenza vaccines.," *Dev. Biol. Stand.*, vol. 98, 1999.

[83] H. A. Vanderven *et al.*, "What Lies Beneath: Antibody Dependent Natural Killer Cell Activation by Antibodies to Internal Influenza Virus Proteins," *EBioMedicine*, vol. 8, 2016, doi: 10.1016/j.ebiom.2016.04.029.

[84] S. Jegaskanda, M. D. T. Co, J. Cruz, K. Subbarao, F. A. Ennis, and M. Terajima, "Induction of H7N9-cross-reactive antibody-dependent cellular cytotoxicity antibodies by human seasonal influenza a viruses that are directed toward the nucleoprotein," *J. Infect. Dis.*, vol. 215, no. 5, 2017, doi: 10.1093/infdis/jiw629.

[85] S. L. Zebedee and R. A. Lamb, "Influenza A virus M2 protein: monoclonal antibody restriction of virus growth and detection of M2 in virions," *J. Virol.*, vol. 62, no. 8, 1988, doi: 10.1128/jvi.62.8.2762-2772.1988.

[86] D. M. Carragher, D. A. Kaminski, A. Moquin, L. Hartson, and T. D. Randall, "A Novel

Role for Non-Neutralizing Antibodies against Nucleoprotein in Facilitating Resistance to Influenza Virus," *J. Immunol.*, vol. 181, no. 6, 2008, doi: 10.4049/jimmunol.181.6.4168.

[87] M. W. LaMere *et al.*, "Contributions of Antinucleoprotein IgG to Heterosubtypic Immunity against Influenza Virus," *J. Immunol.*, vol. 186, no. 7, 2011, doi: 10.4049/jimmunol.1003057.

[88] F. Krammer, "The human antibody response to influenza A virus infection and vaccination," *Nature Reviews Immunology*, vol. 19, no. 6. 2019, doi: 10.1038/s41577-019-0143-6.

[89] C. W. Seibert *et al.*, "Recombinant IgA Is Sufficient To Prevent Influenza Virus Transmission in Guinea Pigs," *J. Virol.*, vol. 87, no. 14, 2013, doi: 10.1128/jvi.00979-13.

[90] S. L. Swain, "T-Cell Subsets: Who does the polarizing?," *Curr. Biol.*, vol. 5, no. 8, 1995, doi: 10.1016/S0960-9822(95)00170-9.

[91] A. Berger, "Science commentary: Th1 and Th2 responses: What are they?," *British Medical Journal*, vol. 321, no. 7258. 2000, doi: 10.1136/bmj.321.7258.424.

[92] J. C. Sun and M. J. Bevan, "Defective CD8 T cell memory following acute infection without CD4 T cell help," *Science (80-.).*, vol. 300, no. 5617, 2003, doi: 10.1126/science.1083317.

[93] K. K. McKinstry *et al.*, "Memory CD4 + T cells protect against influenza through multiple synergizing mechanisms," *J. Clin. Invest.*, vol. 122, no. 8, 2012, doi: 10.1172/JCI63689.

[94] D. M. Brown, S. Lee, M. de la L. Garcia-Hernandez, and S. L. Swain, "Multifunctional CD4 Cells Expressing Gamma Interferon and Perforin Mediate Protection against Lethal Influenza Virus Infection," *J. Virol.*, vol. 86, no. 12, 2012, doi: 10.1128/jvi.07172-11.

[95] T. R. Mosmann and R. L. Coffman, “TH1 and TH2 cells: Different patterns of lymphokine secretion lead to different functional properties,” *Annual Review of Immunology*, vol. 7. 1989, doi: 10.1146/annurev.iy.07.040189.001045.

[96] T. M. Moran, H. Park, A. Fernandez-Sesma, and J. L. Schulman, “Th2 responses to inactivated influenza virus can be converted to Th1 responses and facilitate recovery from heterosubtypic virus infection,” *J. Infect. Dis.*, vol. 180, no. 3, 1999, doi: 10.1086/314952.

[97] A. J. McMichael, F. M. Gotch, G. R. Noble, and P. A. S. Beare, “Cytotoxic T-Cell Immunity to Influenza,” *N. Engl. J. Med.*, vol. 309, no. 1, 1983, doi: 10.1056/nejm198307073090103.

[98] H. E. Jung and H. K. Lee, “Host protective immune Responses against influenza a virus infection,” *Viruses*, vol. 12, no. 5. 2020, doi: 10.3390/v12050504.

[99] G. T. Belz, W. Xie, J. D. Altman, and P. C. Doherty, “ A Previously Unrecognized H-2D b -Restricted Peptide Prominent in the Primary Influenza A Virus-Specific CD8 + T-Cell Response Is Much Less Apparent following Secondary Challenge ,” *J. Virol.*, vol. 74, no. 8, 2000, doi: 10.1128/jvi.74.8.3486-3493.2000.

[100] M. M. Hufford, T. S. Kim, J. Sun, and T. J. Braciale, “Antiviral CD8+ T cell effector activities in situ are regulated by target cell type,” *J. Exp. Med.*, vol. 208, no. 1, 2011, doi: 10.1084/jem.20101850.

[101] M. M. Hufford *et al.*, “Influenza-Infected Neutrophils within the Infected Lungs Act as Antigen Presenting Cells for Anti-Viral CD8+ T Cells,” *PLoS One*, vol. 7, no. 10, 2012, doi: 10.1371/journal.pone.0046581.

[102] J. Sun and T. J. Braciale, “Role of T cell immunity in recovery from influenza virus infection,” *Current Opinion in Virology*, vol. 3, no. 4. 2013, doi:

10.1016/j.coviro.2013.05.001.

[103] S. Sridhar *et al.*, “Cellular immune correlates of protection against symptomatic pandemic influenza,” *Nat. Med.*, vol. 19, no. 10, pp. 1305–1312, 2013, doi: 10.1038/nm.3350.

[104] A. C. Hayward *et al.*, “Natural T cell-mediated protection against seasonal and pandemic influenza: Results of the flu watch cohort study,” *Am. J. Respir. Crit. Care Med.*, vol. 191, no. 12, 2015, doi: 10.1164/rccm.201411-1988OC.

[105] Z. Wang *et al.*, “Recovery from severe H7N9 disease is associated with diverse response mechanisms dominated by CD8+ T cells,” *Nat. Commun.*, vol. 6, 2015, doi: 10.1038/ncomms7833.

[106] V. Demicheli, T. Jefferson, E. Ferroni, A. Rivetti, and C. Di Pietrantonj, “Vaccines for preventing influenza in healthy adults,” *Cochrane Database of Systematic Reviews*, vol. 2018, no. 2. 2018, doi: 10.1002/14651858.CD001269.pub6.

[107] E. A. Belongia, M. E. Sundaram, D. L. McClure, J. K. Meece, J. Ferdinands, and J. J. VanWormer, “Waning vaccine protection against influenza A (H3N2) illness in children and older adults during a single season,” *Vaccine*, vol. 33, no. 1, 2015, doi: 10.1016/j.vaccine.2014.06.052.

[108] J. M. Ferdinands *et al.*, “Intraseason waning of influenza vaccine protection: Evidence from the US influenza vaccine effectiveness network, 2011-2012 through 2014-2015,” *Clin. Infect. Dis.*, vol. 64, no. 5, 2017, doi: 10.1093/cid/ciw816.

[109] R. Bodewes, J. H. Kreijtz, and G. F. Rimmelzwaan, “Yearly influenza vaccinations: a double-edged sword?,” *The Lancet Infectious Diseases*, vol. 9, no. 12. 2009, doi: 10.1016/S1473-3099(09)70263-4.

[110] S. Sridhar, K. A. Brokstad, and R. J. Cox, “Influenza vaccination strategies: Comparing

inactivated and live attenuated influenza vaccines," *Vaccines*, vol. 3, no. 2. 2015, doi: 10.3390/vaccines3020373.

[111] M. Jackwood, L. Hickle, S. Kapil, and R. Silva, "Vaccine development using recombinant DNA technology," *Counc. Agric. Sci. Technol. Issue Pap.*, vol. 7, no. 38, 2008.

[112] V. Brusic and D. R. Flower, "Bioinformatics tools for identifying T-cell epitopes," *Drug Discovery Today: BIOSILICO*, vol. 2, no. 1. 2004, doi: 10.1016/S1741-8364(04)02374-1.

[113] R. E. Soria-Guerra, R. Nieto-Gomez, D. O. Govea-Alonso, and S. Rosales-Mendoza, "An overview of bioinformatics tools for epitope prediction: Implications on vaccine development," *Journal of Biomedical Informatics*, vol. 53. 2015, doi: 10.1016/j.jbi.2014.11.003.

[114] L. Moise *et al.*, "Ivax: An integrated toolkit for the selection and optimization of antigens and the design of epitope-driven vaccines," *Hum. Vaccines Immunother.*, vol. 11, no. 9, pp. 2312–2321, 2015, doi: 10.1080/21645515.2015.1061159.

[115] K. Falk, O. Rötzschke, S. Stevanović, G. Jung, and H. G. Rammensee, "Allele-specific motifs revealed by sequencing of self-peptides eluted from MHC molecules," *Nature*, vol. 351, no. 6324, 1991, doi: 10.1038/351290a0.

[116] O. Rötzschke, K. Falk, S. Stevanovic, G. Jung, P. Walden, and H. -G Rammensee, "Exact prediction of a natural T cell epitope," *Eur. J. Immunol.*, vol. 21, no. 11, 1991, doi: 10.1002/eji.1830211136.

[117] A. S. De Groot, L. Moise, J. A. McMurry, and W. Martin, "Epitope-Based Immunome-Derived Vaccines: A Strategy for Improved Design and Safety," in *Clinical Applications of Immunomics*, 2009.

[118] J. McMurry, H. Sbai, M. L. Gennaro, E. J. Carter, W. Martin, and A. S. De Groot,

"Analyzing Mycobacterium tuberculosis proteomes for candidate vaccine epitopes," *Tuberculosis*, vol. 85, no. 1-2 SPEC.ISS., 2005, doi: 10.1016/j.tube.2004.09.005.

[119] A. S. De Groot and L. Moise, "New tools, new approaches and new ideas for vaccine development," *Expert Review of Vaccines*, vol. 6, no. 2. 2007, doi: 10.1586/14760584.6.2.125.

[120] Y. Kim, A. Sette, and B. Peters, "Applications for T-cell epitope queries and tools in the Immune Epitope Database and Analysis Resource," *J. Immunol. Methods*, vol. 374, no. 1–2, 2011, doi: 10.1016/j.jim.2010.10.010.

[121] A. Sette and J. Sidney, "Nine major HLA class I supertypes account for the vast preponderance of HLA-A and -B polymorphism," *Immunogenetics*, vol. 50, no. 3–4. 1999, doi: 10.1007/s002510050594.

[122] S. Southwood *et al.*, "Several common HLA-DR types share largely overlapping peptide binding repertoires.," *J. Immunol.*, vol. 160, no. 7, 1998.

[123] A. H. Gutiérrez *et al.*, "In vivo validation of predicted and conserved T cell epitopes in a swine influenza model," *PLoS One*, vol. 11, no. 7, 2016, doi: 10.1371/journal.pone.0159237.

[124] A. H. Gutiérrez *et al.*, "T-cell epitope content comparison (EpiCC) of swine H1 influenza A virus hemagglutinin," *Influenza Other Respi. Viruses*, vol. 11, no. 6, pp. 531–542, 2017, doi: 10.1111/irv.12513.

[125] L. Moise *et al.*, "New Immunoinformatics Tools for Swine: Designing Epitope-Driven Vaccines, Predicting Vaccine Efficacy, and Making Vaccines on Demand," *Frontiers in Immunology*, vol. 11. 2020, doi: 10.3389/fimmu.2020.563362.

[126] A. S. De Groot *et al.*, "Mapping cross-clade HIV-1 vaccine epitopes using a

bioinformatics approach," *Vaccine*, vol. 21, no. 27–30, 2003, doi: 10.1016/S0264-410X(03)00390-6.

[127] L. K. Selin, S. R. Nahill, and R. M. Welsh, "Cross-reactivities in memory cytotoxic t lymphocyte recognition of heterologous viruses," *J. Exp. Med.*, vol. 179, no. 6, 1994, doi: 10.1084/jem.179.6.1933.

[128] E. W. Newell *et al.*, " Structural Basis of Specificity and Cross-Reactivity in T Cell Receptors Specific for Cytochrome c –I-E k ," *J. Immunol.*, vol. 186, no. 10, 2011, doi: 10.4049/jimmunol.1100197.

[129] L. Moise *et al.*, "The two-faced T cell epitope: Examining the host-microbe interface with JanusMatrix," in *Human Vaccines and Immunotherapeutics*, 2013, vol. 9, no. 7, pp. 1577–1586, doi: 10.4161/hv.24615.

[130] L. Wooldridge *et al.*, "A single autoimmune T cell receptor recognizes more than a million different peptides," *J. Biol. Chem.*, vol. 287, no. 2, 2012, doi: 10.1074/jbc.M111.289488.

[131] G. J. Kersh and P. M. Allen, "Structural basis for T cell recognition of altered peptide ligands: A single T cell receptor can productively recognize a large continuum of related ligands," *J. Exp. Med.*, vol. 184, no. 4, 1996, doi: 10.1084/jem.184.4.1259.

[132] D. L. Donermeyer, K. S. Weber, D. M. Kranz, and P. M. Allen, "The Study of High-Affinity TCRs Reveals Duality in T Cell Recognition of Antigen: Specificity and Degeneracy," *J. Immunol.*, vol. 177, no. 10, 2006, doi: 10.4049/jimmunol.177.10.6911.

[133] A. S. De Groot *et al.*, "Immune camouflage: Relevance to vaccines and human immunology," *Human Vaccines and Immunotherapeutics*, vol. 10, no. 12. 2014, doi: 10.4161/hv.36134.

[134] R. Vita *et al.*, "The immune epitope database (IEDB) 3.0," *Nucleic Acids Res.*, vol. 43, no. D1, 2015, doi: 10.1093/nar/gku938.

[135] G. Petrova, A. Ferrante, and J. Gorski, "Cross-reactivity of T cells and its role in the immune system," *Critical Reviews in Immunology*, vol. 32, no. 4. 2012, doi: 10.1615/CritRevImmunol.v32.i4.50.

[136] J. C. Pedersen, "Hemagglutination-inhibition assay for influenza virus subtype identification and the detection and quantitation of serum antibodies to influenza virus," *Methods Mol. Biol.*, vol. 1161, 2014, doi: 10.1007/978-1-4939-0758-8_2.

[137] A. Grifoni *et al.*, "Targets of T Cell Responses to SARS-CoV-2 Coronavirus in Humans with COVID-19 Disease and Unexposed Individuals," *Cell*, vol. 181, no. 7, 2020, doi: 10.1016/j.cell.2020.05.015.

[138] J. Zhao, J. Zhao, and S. Perlman, "T Cell Responses Are Required for Protection from Clinical Disease and for Virus Clearance in Severe Acute Respiratory Syndrome Coronavirus-Infected Mice," *J. Virol.*, vol. 84, no. 18, 2010, doi: 10.1128/jvi.01049-10.

[139] J. L. Sanchez-Trincado, M. Gomez-Perosanz, and P. A. Reche, "Fundamentals and Methods for T- and B-Cell Epitope Prediction," *Journal of Immunology Research*, vol. 2017. 2017, doi: 10.1155/2017/2680160.

[140] J. B. Ulmer *et al.*, "Protective CD4+ and CD8+ T Cells against Influenza Virus Induced by Vaccination with Nucleoprotein DNA," *J. Virol.*, 1998, doi: 10.1128/jvi.72.7.5648-5653.1998.

[141] T. J. Powell *et al.*, "Priming with Cold-Adapted Influenza A Does Not Prevent Infection but Elicits Long-Lived Protection against Supralethal Challenge with Heterosubtypic Virus," *J. Immunol.*, 2007, doi: 10.4049/jimmunol.178.2.1030.

[142] T. Wu *et al.*, "Quantification of epitope abundance reveals the effect of direct and cross-presentation on influenza CTL responses," *Nat. Commun.*, 2019, doi: 10.1038/s41467-019-10661-8.

[143] L. Garcia-Garcia *et al.*, "Partial protection of seasonal trivalent inactivated vaccine against novel pandemic influenza A/H1N1 2009: Case-control study in Mexico City," *BMJ*, vol. 339, no. 7725, p. 847, 2009, doi: 10.1136/bmj.b3928.

[144] A. S. De Groot, M. Ardito, E. M. McClaine, L. Moise, and W. D. Martin, "Immunoinformatic comparison of T-cell epitopes contained in novel swine-origin influenza A (H1N1) virus with epitopes in 2008-2009 conventional influenza vaccine," *Vaccine*, vol. 27, no. 42, pp. 5740–5747, 2009, doi: 10.1016/j.vaccine.2009.07.040.

[145] N. Busquets *et al.*, "Experimental infection with H1N1 European swine influenza virus protects pigs from an infection with the 2009 pandemic H1N1 human influenza virus," *Vet. Res.*, vol. 41, no. 5, 2010, doi: 10.1051/vetres/2010046.

[146] Y. Qiu, K. De Hert, and K. Van Reeth, "Cross-protection against European swine influenza viruses in the context of infection immunity against the 2009 pandemic H1N1 virus: Studies in the pig model of influenza," *Vet. Res.*, vol. 46, no. 1, 2015, doi: 10.1186/s13567-015-0236-6.

[147] M. Baratelli *et al.*, "Identification of cross-reacting T-cell epitopes in structural and non-structural proteins of swine and pandemic H1N1 influenza a virus strains in pigs," *J. Gen. Virol.*, vol. 98, no. 5, 2017, doi: 10.1099/jgv.0.000748.

[148] K. Tungatt *et al.*, "Induction of influenza-specific local CD8 T-cells in the respiratory tract after aerosol delivery of vaccine antigen or virus in the Babraham inbred pig," *PLoS Pathog.*, vol. 14, no. 5, 2018, doi: 10.1371/journal.ppat.1007017.

[149] A. H. Ellebedy *et al.*, “Contemporary seasonal influenza A (H1N1) virus infection primes for a more robust response to split inactivated pandemic influenza A (H1N1) virus vaccination in ferrets,” *Clin. Vaccine Immunol.*, vol. 17, no. 12, pp. 1998–2006, 2010, doi: 10.1128/CVI.00247-10.

[150] L. He, A. S. De Groot, A. H. Gutierrez, W. D. Martin, L. Moise, and C. Bailey-Kellogg, “Integrated assessment of predicted MHC binding and cross-conservation with self reveals patterns of viral camouflage,” *BMC Bioinformatics*, vol. 15, 2014, doi: 10.1186/1471-2105-15-S4-S1.

[151] R. Liu *et al.*, “H7N9 T-cell Epitopes that mimic human sequences are less immunogenic and may induce Treg-mediated tolerance,” *Hum. Vaccines Immunother.*, vol. 11, no. 9, pp. 2241–2252, 2015, doi: 10.1080/21645515.2015.1052197.

[152] A. H. Gutiérrez, W. D. Martin, C. Bailey-Kellogg, F. Terry, L. Moise, and A. S. De Groot, “Development and validation of an epitope prediction tool for swine (PigMatrix) based on the pocket profile method,” *BMC Bioinformatics*, vol. 16, no. 1, 2015, doi: 10.1186/s12859-015-0724-8.

[153] J. S. Hewitt *et al.*, “A prime-boost concept using a T-cell epitope-driven DNA vaccine followed by a whole virus vaccine effectively protected pigs in the pandemic H1N1 pig challenge model,” *Vaccine*, 2019, doi: 10.1016/j.vaccine.2019.06.044.

[154] Y. Zhang *et al.*, “Influenza Research Database: An integrated bioinformatics resource for influenza virus research,” *Nucleic Acids Res.*, vol. 45, no. D1, pp. D466–D474, 2017, doi: 10.1093/nar/gkw857.

[155] C. S. Ho *et al.*, “Molecular characterization of swine leucocyte antigen class i genes in outbred pig populations,” *Anim. Genet.*, vol. 40, no. 4, pp. 468–478, Aug. 2009, doi:

10.1111/j.1365-2052.2009.01860.x.

[156] C. S. Ho *et al.*, “Molecular characterization of swine leucocyte antigen class II genes in outbred pig populations,” *Anim. Genet.*, 2010, doi: 10.1111/j.1365-2052.2010.02019.x.

[157] G. Yu, D. K. Smith, H. Zhu, Y. Guan, and T. T. Y. Lam, “ggtree: an r package for visualization and annotation of phylogenetic trees with their covariates and other associated data,” *Methods Ecol. Evol.*, vol. 8, no. 1, 2017, doi: 10.1111/2041-210X.12628.

[158] M. Bandrick *et al.*, “T cell epitope content comparison (EpiCC) analysis demonstrates a bivalent PCV2 vaccine has greater T cell epitope overlap with field strains than monovalent PCV2 vaccines,” *Vet. Immunol. Immunopathol.*, 2020, doi: 10.1016/j.vetimm.2020.110034.

[159] E. B. Clemens, C. Van de Sandt, S. S. Wong, L. M. Wakim, and S. A. Valkenburg, “Harnessing the power of T cells: The promising hope for a universal influenza vaccine,” *Vaccines*, vol. 6, no. 2. 2018, doi: 10.3390/vaccines6020018.

[160] E. J. Erbelding *et al.*, “A universal influenza vaccine: The strategic plan for the national institute of allergy and infectious diseases,” *J. Infect. Dis.*, vol. 218, no. 3, 2018, doi: 10.1093/infdis/jiy103.

[161] E. Van Doorn *et al.*, “Evaluating the immunogenicity and safety of a BiondVax-developed universal influenza vaccine (Multimeric-001) either as a standalone vaccine or as a primer to H5N1 influenza vaccine,” *Med. (United States)*, vol. 96, no. 11, 2017, doi: 10.1097/MD.0000000000006339.

[162] G. H. Lowell, S. Ziv, S. Bruzil, R. Babecoff, and T. Ben-Yedidia, “Back to the future: Immunization with M-001 prior to trivalent influenza vaccine in 2011/12 enhanced protective immune responses against 2014/15 epidemic strain,” *Vaccine*, vol. 35, no. 5,

2017, doi: 10.1016/j.vaccine.2016.12.063.

[163] R. J. Garten *et al.*, “Antigenic and genetic characteristics of swine-origin 2009 A(H1N1) influenza viruses circulating in humans,” *Science (80-.).*, vol. 325, no. 5937, 2009, doi: 10.1126/science.1176225.

[164] C. Dykhuis Haden, T. Painter, T. Fangman, and D. Holtkamp, “Assessing production parameters and economic impact of swine influenza, PRRS and Mycoplasma hyopneumoniae on finishing pigs in a large production system,” in *Proceedings of the AASV Annual Meeting*, 2012.

[165] C. S. Kyriakis *et al.*, “Molecular epidemiology of swine influenza A viruses in the Southeastern United States, highlights regional differences in circulating strains,” *Vet. Microbiol.*, vol. 211, 2017, doi: 10.1016/j.vetmic.2017.10.016.

[166] G. A. Sautto, J. W. Ecker, and T. M. Ross, “An H1N1 Computationally Optimized Broadly Reactive Antigen Elicits a Neutralizing Antibody Response against an Emerging Human-Infecting Eurasian Avian-Like Swine Influenza Virus,” *J. Virol.*, vol. 95, no. 13, 2021, doi: 10.1128/jvi.02421-20.

[167] V. R. S. K. Duvvuri *et al.*, “Highly conserved cross-reactive CD4+ T-cell HA-epitopes of seasonal and the 2009 pandemic influenza viruses,” *Influenza Other Respi. Viruses*, vol. 4, no. 5, 2010, doi: 10.1111/j.1750-2659.2010.00161.x.

[168] V. R. Duvvuri, B. Duvvuri, V. Jamnik, J. B. Gubbay, J. Wu, and G. E. Wu, “T cell memory to evolutionarily conserved and shared hemagglutinin epitopes of H1N1 viruses: A pilot scale study,” *BMC Infect. Dis.*, vol. 13, no. 1, 2013, doi: 10.1186/1471-2334-13-204.

[169] J. T. Weinfurter *et al.*, “Cross-reactive T cells are involved in rapid clearance of 2009

pandemic H1N1 influenza virus in nonhuman primates," *PLoS Pathog.*, vol. 7, no. 11, 2011, doi: 10.1371/journal.ppat.1002381.

[170] A. F. Altenburg, G. F. Rimmelzwaan, and R. D. de Vries, "Virus-specific T cells as correlate of (cross-)protective immunity against influenza," *Vaccine*. 2015, doi: 10.1016/j.vaccine.2014.11.054.

[171] S. Elbe and G. Buckland-Merrett, "Data, disease and diplomacy: GISAID's innovative contribution to global health," *Glob. Challenges*, vol. 1, no. 1, 2017, doi: 10.1002/gch2.1018.

[172] K. Van Reeth and W. Ma, "Swine influenza virus vaccines: To change or not to change-that's the question," *Curr. Top. Microbiol. Immunol.*, vol. 370, 2013, doi: 10.1007/82-2012-266.

[173] K. Van Reeth, A. L. Vincent, and K. M. Lager, "Vaccines and vaccination for swine influenza: differing situations in Europe and the USA," in *Animal Influenza*, 2016.

[174] A. L. Skarlupka, S. O. Owino, L. P. Suzuki-Williams, C. J. Crevar, D. M. Carter, and T. M. Ross, "Computationally optimized broadly reactive vaccine based upon swine H1N1 influenza hemagglutinin sequences protects against both swine and human isolated viruses," *Hum. Vaccines Immunother.*, vol. 15, no. 9, 2019, doi: 10.1080/21645515.2019.1653743.

[175] P. J. A. Cock *et al.*, "Biopython: Freely available Python tools for computational molecular biology and bioinformatics," *Bioinformatics*, vol. 25, no. 11, 2009, doi: 10.1093/bioinformatics/btp163.

[176] R. C. Edgar, "MUSCLE: Multiple sequence alignment with high accuracy and high throughput," *Nucleic Acids Res.*, vol. 32, no. 5, 2004, doi: 10.1093/nar/gkh340.

[177] A. Stamatakis, “RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies,” *Bioinformatics*, vol. 30, no. 9, 2014, doi: 10.1093/bioinformatics/btu033.

[178] O. Chernomor *et al.*, “Split diversity in constrained conservation prioritization using integer linear programming,” *Methods Ecol. Evol.*, vol. 6, no. 1, 2015, doi: 10.1111/2041-210X.12299.

[179] S. Tan *et al.*, “Quantifying the Persistence of Vaccine-Related T Cell Epitopes in Circulating Swine Influenza A Strains from 2013–2017,” *Vaccines*, vol. 9, no. 5, 2021, doi: 10.3390/vaccines9050468.

[180] A. Kassambara, “rstatix: Pipe-Friendly Framework for Basic Statistical Tests,” *Why We Need the Journal of Interactive Advertising*, vol. 3, no. 1. 1997.

[181] M. Khanna, S. Sharma, B. Kumar, and R. Rajput, “Protective immunity based on the conserved hemagglutinin stalk domain and its prospects for universal influenza vaccine development,” *BioMed Research International*, vol. 2014. 2014, doi: 10.1155/2014/546274.

[182] F. Haesebrouck and M. B. Pensaert, “Effect of intratracheal challenge of fattening pigs previously immunised with an inactivated influenza H1N1 vaccine,” *Vet. Microbiol.*, vol. 11, no. 3, 1986, doi: 10.1016/0378-1135(86)90026-X.

[183] C. S. Kyriakis, M. R. Gramer, F. Barbé, J. Van Doorsselaere, and K. Van Reeth, “Efficacy of commercial swine influenza vaccines against challenge with a recent European H1N1 field isolate,” *Vet. Microbiol.*, vol. 144, no. 1–2, 2010, doi: 10.1016/j.vetmic.2009.12.039.

[184] K. Van Reeth, G. Labarque, S. De Clercq, and M. Pensaert, “Efficacy of vaccination of pigs with different H1N1 swine influenza viruses using a recent challenge strain and

different parameters of protection," *Vaccine*, vol. 19, no. 31, 2001, doi: 10.1016/S0264-410X(01)00206-7.

[185] R. Duerrwald, M. Schlegel, K. Bauer, T. Vissiennon, P. Wutzler, and M. Schmidtke, "Efficacy of Influenza Vaccination and Tamiflu® Treatment - Comparative Studies with Eurasian Swine Influenza Viruses in Pigs," *PLoS One*, vol. 8, no. 4, 2013, doi: 10.1371/journal.pone.0061597.

[186] "How Flu Vaccine Effectiveness and Efficacy are Measured," *Centers for Disease Control and Prevention, National Center for Immunization and Respiratory Diseases (NCIRD)*. https://www.cdc.gov/flu/vaccines-work/effectivenessqa.htm.

[187] S. S. Wong and R. J. Webby, "Traditional and new influenza vaccines," *Clin. Microbiol. Rev.*, vol. 26, no. 3, 2013, doi: 10.1128/CMR.00097-12.

[188] M. Bandrick *et al.*, "A bivalent porcine circovirus type 2 (PCV2), PCV2a-PCV2b, vaccine offers biologically superior protection compared to monovalent PCV2 vaccines," *Vet. Res.*, vol. 53, no. 1, p. 12, 2022, doi: 10.1186/s13567-022-01029-w.

[189] "Seasonal influenza factsheet," *Geneva: World Health Organization (WHO)*, 2018. https://www.who.int/news-room/fact-sheets/detail/influenza-(seasonal) (accessed Jun. 20, 2022).

[190] A. D. Iuliano *et al.*, "Estimates of global seasonal influenza-associated respiratory mortality: a modelling study," *Lancet*, vol. 391, no. 10127, 2018, doi: 10.1016/S0140-6736(17)33293-2.

[191] B. J. Jester, T. M. Uyeki, and D. B. Jernigan, "Fifty years of influenza A(H3N2) following the pandemic of 1968," *Am. J. Public Health*, vol. 110, no. 5, 2020, doi: 10.2105/AJPH.2019.305557.

[192] J. D. Allen and T. M. Ross, “H3N2 influenza viruses in humans: Viral mechanisms, evolution, and evaluation,” *Human Vaccines and Immunotherapeutics*, vol. 14, no. 8. 2018, doi: 10.1080/21645515.2018.1462639.

[193] “Vaccine Effectiveness: How Well Do Flu Vaccines Work?,” *Centers for Disease Control and Prevention (CDC)*, 2021. .

[194] S. Gouma, M. Weirick, and S. E. Hensley, “Potential Antigenic Mismatch of the H3N2 Component of the 2019 Southern Hemisphere Influenza Vaccine,” *Clin. Infect. Dis.*, 2019, doi: 10.1093/cid/ciz723.

[195] E. A. Belongia and H. Q. McLean, “Influenza Vaccine Effectiveness: Defining the H3N2 Problem,” *Clin. Infect. Dis.*, vol. 69, no. 10, 2019, doi: 10.1093/cid/ciz411.

[196] S. Alam and A. J. Sant, “Infection with Seasonal Influenza Virus Elicits CD4 T Cells Specific for Genetically Conserved Epitopes That Can Be Rapidly Mobilized for Protective Immunity to Pandemic H1N1 Influenza Virus,” *J. Virol.*, vol. 85, no. 24, 2011, doi: 10.1128/jvi.05728-11.

[197] J. B. Plotkin, J. Dushoff, and S. A. Levin, “Hemagglutinin sequence clusters and the antigenic evolution of influenza A virus,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 99, no. 9, 2002, doi: 10.1073/pnas.082110799.

[198] P. A. Jorquera *et al.*, “Insights into the antigenic advancement of influenza A(H3N2) viruses, 2011–2018,” *Sci. Rep.*, vol. 9, no. 1, 2019, doi: 10.1038/s41598-019-39276-1.

[199] M. W. Deem and K. Pan, “The epitope regions of HI -subtype influenza A, with application to vaccine efficacy,” *Protein Eng. Des. Sel.*, vol. 22, no. 9, 2009, doi: 10.1093/protein/gzp027.

[200] X. Qiu, V. R. Duvvuri, J. B. Gubbay, R. J. Webby, G. Kayali, and J. Bahl, “Lineage-

specific epitope profiles for HPAI H5 pre-pandemic vaccine selection and evaluation," *Influenza Other Respi. Viruses*, vol. 11, no. 5, 2017, doi: 10.1111/irv.12466.

[201] C. S. Eickhoff *et al.*, "Highly conserved influenza T cell epitopes induce broadly protective immunity," *Vaccine*, vol. 37, no. 36, 2019, doi: 10.1016/j.vaccine.2019.07.033.

[202] D. J. Smith *et al.*, "Mapping the antigenic and genetic evolution of influenza virus," *Science (80-.).*, 2004, doi: 10.1126/science.1097211.

[203] L. T. Nguyen, H. A. Schmidt, A. Von Haeseler, and B. Q. Minh, "IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies," *Mol. Biol. Evol.*, vol. 32, no. 1, 2015, doi: 10.1093/molbev/msu300.

[204] World Health Organization [WHO], "Seasonal influenza candidate vaccine viruses," 2022. https://www.who.int/teams/global-influenza-programme/vaccines/who-recommendations/candidate-vaccine-viruses (accessed Jul. 06, 2022).

[205] P. Stothard, "The sequence manipulation suite: JavaScript programs for analyzing and formatting protein and DNA sequences.," *Biotechniques*, vol. 28, no. 6, 2000, doi: 10.2144/00286ir01.

[206] A. Becerra-Artiles *et al.*, "Broadly recognized, cross-reactive SARS-CoV-2 CD4 T cell epitopes are highly conserved across human coronaviruses and presented by common HLA alleles," *Cell Rep.*, vol. 39, no. 11, 2022, doi: 10.1016/j.celrep.2022.110952.

[207] D. Gfeller *et al.*, "Improved predictions of antigen presentation and TCR recognition with MixMHCpred2.2 and PRIME2.0 reveal potent SARS-CoV-2 CD8(+) T-cell epitopes.," *Cell Syst.*, vol. 14, no. 1, pp. 72-83.e5, Jan. 2023, doi: 10.1016/j.cels.2022.12.002.

[208] L. M. Meyers *et al.*, "Highly conserved, non-human-like, and cross-reactive SARS-CoV-2 T cell epitopes for COVID-19 vaccine design and validation," *npj Vaccines*, vol. 6, no. 1,

2021, doi: 10.1038/s41541-021-00331-6.

[209] A. S. De Groot *et al.*, “Better epitope discovery, precision immune engineering, and accelerated vaccine design using Immunoinformatics tools,” *Front. Immunol.*, vol. 11, 2020, doi: 10.3389/fimmu.2020.00442.

[210] S. Wirsching, M. Fichter, M. L. Cacicedo, K. Landfester, and S. Gehring, “Modification of Regulatory T Cell Epitopes Promotes Effector T Cell Responses to Aspartyl/Asparaginyl β-Hydroxylase.,” *Int. J. Mol. Sci.*, vol. 23, no. 20, Oct. 2022, doi: 10.3390/ijms232012444.

[211] A. S. De Groot *et al.*, “Immune Tolerance-Adjusted Personalized Immunogenicity Prediction for Pompe Disease,” *Front. Immunol.*, vol. 12, 2021, doi: 10.3389/fimmu.2021.636731.

[212] Sam Wilks, “Racmacs: R Antigenic Cartography Macros,” 2022. https://acorg.github.io/Racmacs.

[213] I. Aksamentov, C. Roemer, E. Hodcroft, and R. Neher, “Nextclade: clade assignment, mutation calling and quality control for viral genomes,” *J. Open Source Softw.*, vol. 6, no. 67, 2021, doi: 10.21105/joss.03773.

[214] J. C. Gower, “Some Distance Properties of Latent Root and Vector Methods Used in Multivariate Analysis,” *Biometrika*, vol. 53, no. 3/4, 1966, doi: 10.2307/2333639.

[215] A. Hartigan and M. A. Wong, “A K-Means Clustering Algorithm,” *J. R. Stat. Soc.*, vol. 28, no. 1, 1979.

[216] L. Scrucca, M. Fop, T. B. Murphy, and A. E. Raftery, “Mclust 5: Clustering, classification and density estimation using Gaussian finite mixture models,” *R J.*, vol. 8, no. 1, 2016, doi: 10.32614/rj-2016-021.

[217] W. N. Venables and B. D. Ripley, “Modern applied statistics with S Springer-Verlag,” *New York, New York, USA*, 2002.

[218] J. Chen *et al.*, “Diversity and evolution of computationally predicted T cell epitopes against human respiratory syncytial virus.,” *PLoS Comput. Biol.*, vol. 19, no. 1, p. e1010360, Jan. 2023, doi: 10.1371/journal.pcbi.1010360.

[219] M. A. Suchard, P. Lemey, G. Baele, D. L. Ayres, A. J. Drummond, and A. Rambaut, “Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10,” *Virus Evol.*, vol. 4, no. 1, 2018, doi: 10.1093/ve/vey016.

[220] F. Rodríguez, J. L. Oliver, A. Marín, and J. R. Medina, “The general stochastic model of nucleotide substitution,” *J. Theor. Biol.*, vol. 142, no. 4, 1990, doi: 10.1016/S0022-5193(05)80104-3.

[221] V. N. Minin, E. W. Bloomquist, and M. A. Suchard, “Smooth skyride through a rough skyline: Bayesian coalescent-based inference of population dynamics,” *Mol. Biol. Evol.*, vol. 25, no. 7, 2008, doi: 10.1093/molbev/msn090.

[222] A. J. Drummond and A. Rambaut, “BEAST: Bayesian evolutionary analysis by sampling trees,” *BMC Evol. Biol.*, vol. 7, no. 1, 2007, doi: 10.1186/1471-2148-7-214.

[223] A. Rambaut, A. J. Drummond, D. Xie, G. Baele, and M. A. Suchard, “Posterior summarization in Bayesian phylogenetics using Tracer 1.7,” *Syst. Biol.*, vol. 67, no. 5, 2018, doi: 10.1093/sysbio/syy032.

[224] G. Yu, “ggtree: an R package for visualization of tree and annotation data,” *bioRxiv*, 2020.

[225] P. Lemey, A. Rambaut, A. J. Drummond, and M. A. Suchard, “Bayesian phylogeography finds its roots,” *PLoS Comput. Biol.*, vol. 5, no. 9, 2009, doi: 10.1371/journal.pcbi.1000520.

[226] F. Bielejec, G. Baele, B. Vrancken, M. A. Suchard, A. Rambaut, and P. Lemey, "SpreaD3: Interactive Visualization of Spatiotemporal History and Trait Evolutionary Processes," *Mol. Biol. Evol.*, vol. 33, no. 8, 2016, doi: 10.1093/molbev/msw082.

[227] R. E. Kass and A. E. Raftery, "Bayes factors," *J. Am. Stat. Assoc.*, vol. 90, no. 430, 1995, doi: 10.1080/01621459.1995.10476572.

[228] N. L. La Gruta and S. J. Turner, "T cell mediated immunity to influenza: Mechanisms of viral control," *Trends in Immunology*, vol. 35, no. 8. 2014, doi: 10.1016/j.it.2014.06.004.

[229] L. Hensen *et al.*, "T Cell Epitope Discovery in the Context of Distinct and Unique Indigenous HLA Profiles.," *Front. Immunol.*, vol. 13, p. 812393, 2022, doi: 10.3389/fimmu.2022.812393.

[230] G. Richard *et al.*, "Multi-step screening of neoantigens' HLA- and TCR-interfaces improves prediction of survival," *Sci. Rep.*, vol. 11, no. 1, 2021, doi: 10.1038/s41598-021-89016-7.

[231] E. Kirkpatrick, X. Qiu, P. C. Wilson, J. Bahl, and F. Krammer, "The influenza virus hemagglutinin head evolves faster than the stalk domain," *Sci. Rep.*, vol. 8, no. 1, 2018, doi: 10.1038/s41598-018-28706-1.

[232] J. Köster *et al.*, "Sustainable data analysis with Snakemake," *F1000Research*, vol. 10, 2021, doi: 10.12688/f1000research.29032.2.

APPENDIX A

SUPPLEMENTAL MATERIAL FOR CHAPTER 3

QUANTIFYING THE PERSISTENCE OF VACCINE-RELATED T CELL EPITOPES IN CIRCULATING SWINE INFLUENZA A STRAINS FROM 2013–2017 [1]

Supplemental Table A-1. EpiCC scores of all circulating swine IAV strains. Detailed information can be found online at https://www.mdpi.com/article/10.3390/vaccines9050468/s1 (Table S2).

Supplemental Table A-2. List of circulating swine IAV strains that have the most and the least conserved class I (A) and II (B) epitopes.

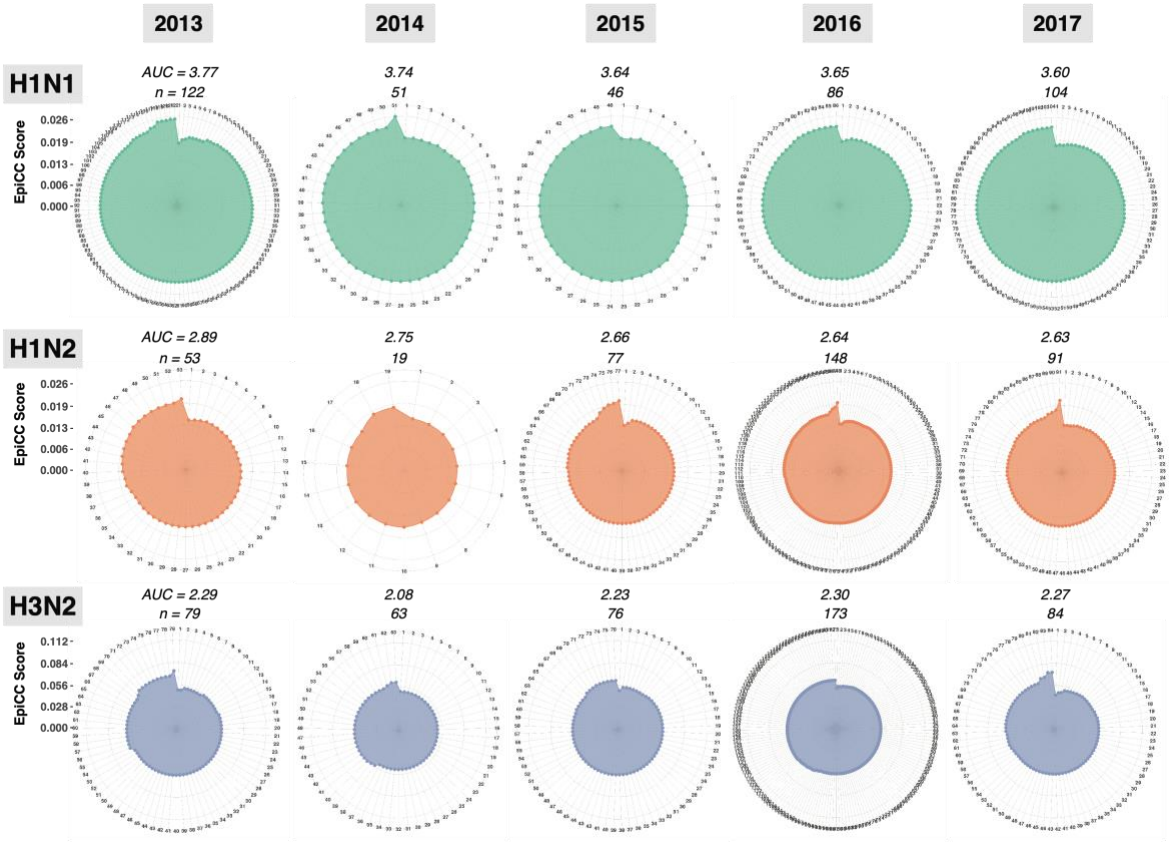
[1]Tan, S.; Gutiérrez, A.H.; Gauger, P.C.; Opriessnig, T.; Bahl, J.; Moise, L.; De Groot, A.S. Quantifying the Persistence of Vaccine-Related T Cell Epitopes in Circulating Swine Influenza A Strains from 2013–2017. *Vaccines 2021, 9, 468*. https://doi.org/10.3390/vaccines9050468. Reprinted here with permission of the publisher.

(A) Class I most conserved strains against the vaccine. Conservation at 75% (21/28)		
No.	**Strain Name**	**Subtype**
1	A/Swine/Arkansas/D0386/2013	
2	A/Swine/Minnesota/A01392911/2013	
3	A/Swine/Minnesota/A01394863/2013	
4	A/Swine/Minnesota/MT1301S79/2013	
5	A/Swine/Ohio/A01349978/2013	
6	A/Swine/Ohio/A01432602/2013	
7	A/Swine/Illinois/A01490609/2014	
8	A/Swine/Illinois/A01492501/2014	
9	A/Swine/Illinois/A01493472/2014	
10	A/Swine/Iowa/A01410472/2014	
11	A/Swine/Kansas/A01410327/2014	
12	A/Swine/Minnesota/A01491447/2014	
13	A/Swine/Minnesota/A01491704/2014	
14	A/Swine/Missouri/A01492887/2014	
15	A/Swine/Nebraska/A01366774/2014	
16	A/Swine/Nebraska/A01491300/2014	
17	A/Swine/Nebraska/A01492657/2014	
18	A/Swine/Nebraska/A01566172/2014	H1N1
19	A/Swine/North Carolina/A01410573/2014	
20	A/Swine/Oklahoma/A01410195/2014	
21	A/Swine/Oklahoma/A01476227/2014	
22	A/Swine/Indiana/A01260972/2015	
23	A/Swine/Illinois/A01729364/2016	
24	A/Swine/Illinois/A01749912/2016	
25	A/Swine/Illinois/A01749913/2016	
26	A/Swine/Illinois/A01749914/2016	
27	A/Swine/Illinois/A01775937/2016	
28	A/Swine/Illinois/A01776206/2016	
29	A/Swine/Illinois/A01777039/2016	
30	A/Swine/Illinois/A01778882/2016	
31	A/Swine/Indiana/16TOSU4933/2016	
32	A/Swine/Indiana/A01671620/2016	
33	A/Swine/Indiana/A01812242/2016	
34	A/Swine/Iowa/A01781047/2016	
35	A/Swine/Iowa/A01782230/2016	

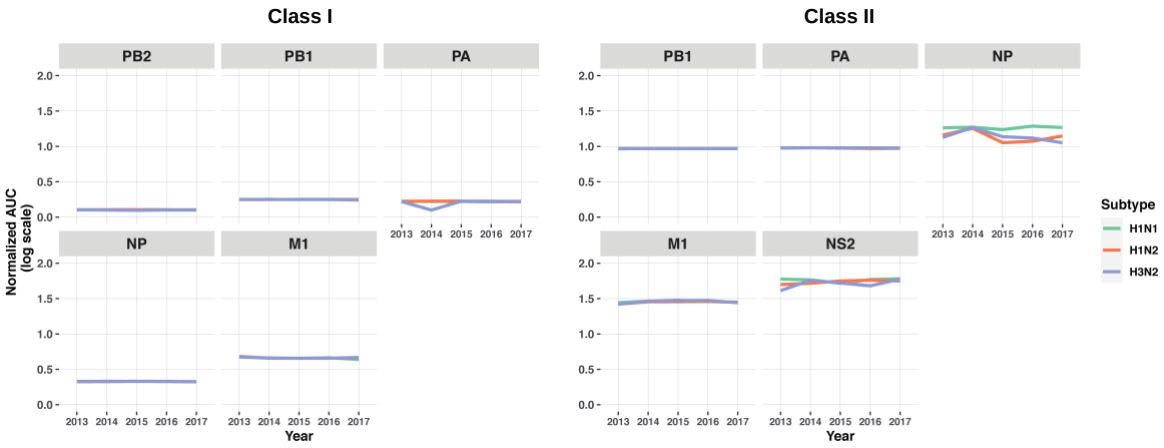
36	A/Swine/Missouri/A01775100/2016	
37	A/Swine/Missouri/A01775109/2016	
38	A/Swine/Nebraska/A01783006/2016	
39	A/Swine/Ohio/A01104092/2016	
40	A/Swine/Tennessee/A01894329/2016	
41	A/Swine/Iowa/A02214835/2017	
42	A/Swine/Kansas/A01378027/2017	
1	A/Swine/Indiana/16TOSU0646/2016	H1N2
2	A/Swine/Michigan/A01259077/2017	
Class I least conserved strains against the vaccine. Conservation at 46.4% (13/28)		
1	A/Swine/Kansas/A01377649/2015	H3N2
2	A/Swine/Missouri/A01840324/2015	
3	A/Swine/Florida/UF1/2017	
4	A/Swine/Missouri/A02136832/2017	

(B) Class II most conserved strains against the vaccine. Conservation at 32.3% (74/229)

No.	Strain Name	Subtype
1	A/Swine/Arkansas/D0386/2013	H1N1
2	A/Swine/Minnesota/A01381276/2013	
3	A/Swine/Ohio/A01349978/2013	
4	A/Swine/Ohio/A01432602/2013	
Class II least conserved strains against the vaccine. Conservation at 15.3% (35/229)		
1	A/Swine/Nebraska/A01493915/2014	H3N2
2	A/Swine/Florida/UF1/2017	



Supplemental Figure A-1. Radar plots showing degree of class I T cell epitope conservation between the conserved epitopes in the MEpiV for the HA antigen and epitopes contained in HA from circulating strains for each year. The radar plots show T cell epitope conservation between the T cell-directed multi-epitope DNA vaccine against each swine IAV circulating strain (axes of radar plot) over five years. The strains were sorted from lowest to highest EpiCC scores. The multi-epitope DNA vaccine is predicted to drive better CD4 immune response based on data published in Gutierrez 2017 and Hewitt 2019.



Supplemental Figure A-2. Line plots showing the degree of conservation by antigens, subtypes, and years. The degree of conservation (AUC) is normalized to ease direct comparison by internal antigens, subtypes, years, and T cell epitopes classes.

APPENDIX B

SUPPLEMENTAL MATERIAL FOR CHAPTER 4

H1N1 G4 SWINE INFLUENZA T CELL EPITOPE ANALYSIS IN SWINE AND HUMAN VACCINES AND CIRCULATING STRAINS UNCOVERS POTENTIAL RISK TO SWINE AND HUMANS [2]

Supplemental Table B-1. H1 HA sequences of EU swIAV vaccine strains, swIAV strains circulating in the EU (from 1939-2018) and swIAV G4 virus strains that are included in the analysis. Detailed information can be found online at https://onlinelibrary.wiley.com/doi/full/10.1111/irv.13058 (Supporting Information: Table S1).

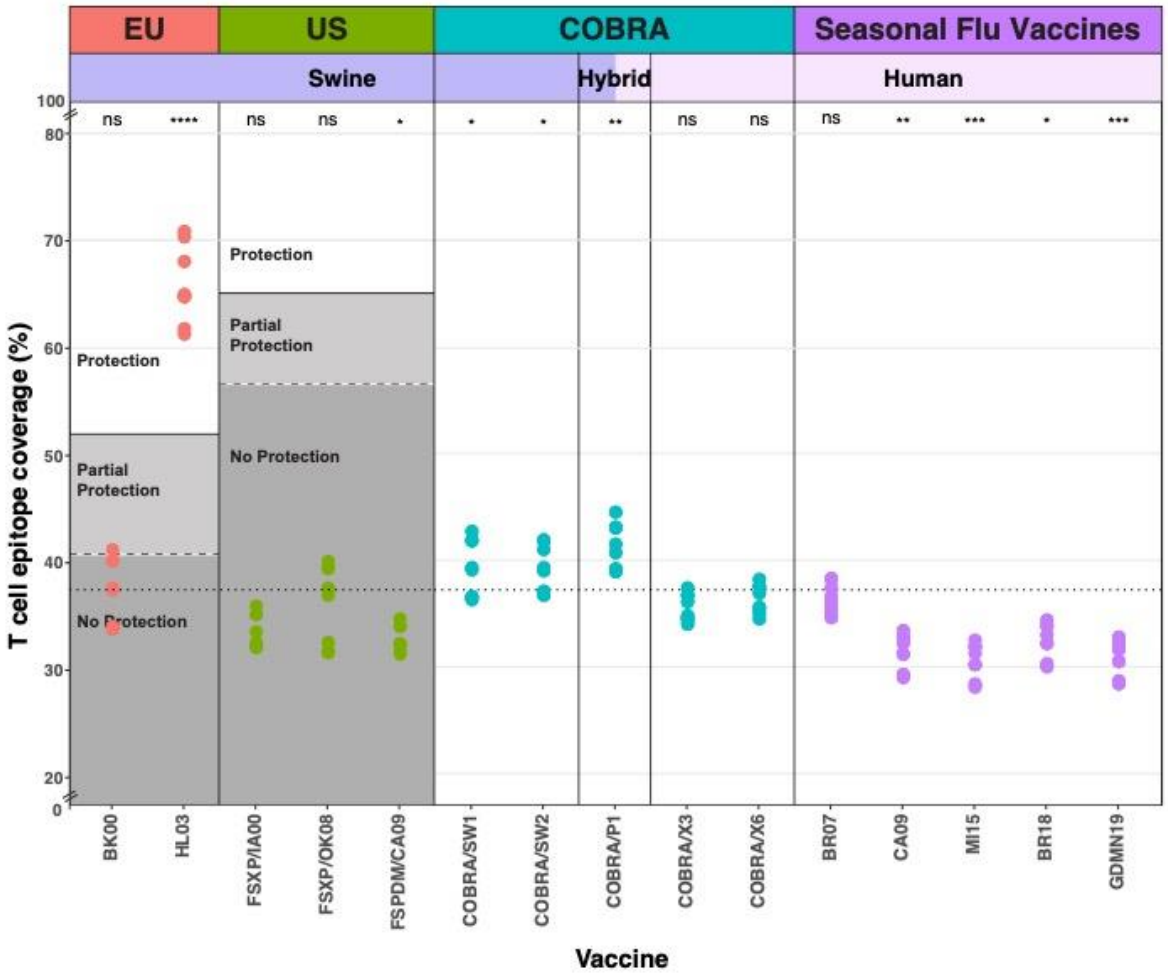
Supplemental Table B-2. H1 HA sequences of US swIAV vaccine strains, swIAV strains circulating in the US (from 1939-2020) and swIAV G4 virus strains that are included in the analysis. Detailed information can be found online at https://onlinelibrary.wiley.com/doi/full/10.1111/irv.13058 (Supporting Information: Table S1).

[2]Tan, S; Moise, L; Pearce, DS; Kyriakis CS; Gutiérrez, AH; Ross, TM; Bahl, J; De Groot, AS. H1N1 G4 swine influenza T cell epitope analysis in swine and human vaccines and circulating strains uncovers potential risk to swine and humans. *Influenza Other Respi Viruses. 2022; 1- 15*. https://doi.org/10.1111/irv.13058. Reprinted here with permission of the publisher.

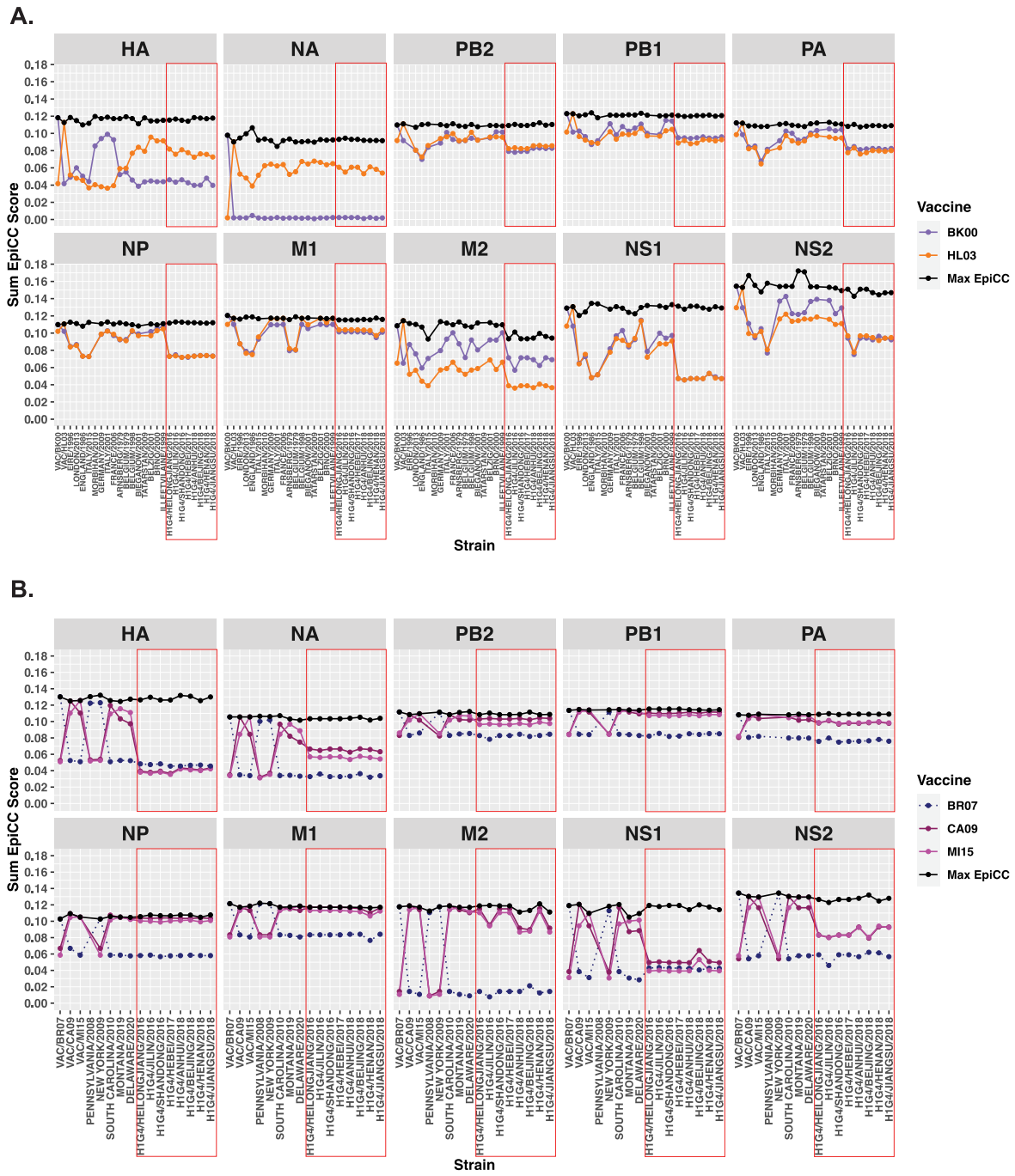
Supplemental Table B-3. H1 HA sequences of human IAV vaccine strains, human IAV strains circulating in the US (from 2008-2010 and 2019-2020) and swIAV G4 virus strains that are included in the analysis. Detailed information can be found online at https://onlinelibrary.wiley.com/doi/full/10.1111/irv.13058 (Supporting Information: Table S1).

Supplemental Table B-4. Average EpiCC score, and vaccine T cell epitope coverage of EU swine influenza vaccine strains compare to EU circulating and G4 strains for the whole proteome analysis. Detailed information can be found online at https://onlinelibrary.wiley.com/doi/full/10.1111/irv.13058 (Supporting Information: Table S2).

Supplemental Table B-5. Average EpiCC score, and vaccine T cell epitope coverage of US human influenza vaccine strains compare to US human circulating IAV and G4 strains for the whole proteome analysis. Detailed information can be found online at https://onlinelibrary.wiley.com/doi/full/10.1111/irv.13058 (Supporting Information: Table S2).



Supplemental Figure B-1. A vaccine-to-vaccine comparison of T cell epitope coverage for each of the vaccine strains against G4. Vaccine strains are color-coded and arranged in order following host and respective region, year, and vaccine type. The Y-axis shows vaccine T cell epitope coverage, shown as a percentage of the maximum score. The horizontal dotted line represents mean of T cell epitope coverage for the strains evaluated in this report. Shaded grey areas the threshold for predicted vaccine efficacy for the US (refer to Table 2 - footnote) and EU (Table 4 - footnote). Only HL03 is predicted to be protective against G4 in this analysis. Statistically significant p-values are labelled for each comparison.



Supplemental Figure B-2. Whole proteome EpiCC analysis of EU swine influenza vaccines (A) and human influenza vaccines (B) against circulating IAV strains in respective regions and G4 strains. Circulating IAV strains were arranged according to clade and year. Red boxes showed H1N1 G4 strains.

APPENDIX C

SUPPLEMENTAL MATERIAL FOR CHAPTER 5

SEQUENCE-BASED EVALUATION OF CD4+ T CELL IMMUNE LANDSCAPE OF H3N2 SEASONAL INFLUENZA A VIRUS [3]

[3]Tan, S; Damodaran, L; Dailey, CA; Chen J; Kondor, R; Moise, L; De Groot, AS; Bahl, J. Sequence-based evaluation of CD4+ T cell immune landscape of H3N2 seasonal influenza A virus. To be submitted to *Frontiers in Immunology*, 4/2023.

Supplemental Table C-1. A list of WHO recommended vaccine component for A/H3N2. (A) listed vaccine viruses that are recommended from 1974 to 2004. (B) detailed the recommendation of A/H3N2 vaccine component from 2010 to 2019. The recommendation included egg-based and cell-based vaccine viruses.

(A)

Seasonal year	WHO recommended H3 composition	Seasonal year	WHO recommended H3 composition
1974-1975	A/Port Chalmers/1/1973	1991-1992	A/Beijing/353/1989
1975-1976		1992-1993	
1976-1977	A/Victoria/3/1975	1993-1994	A/Beijing/32/1992
1977-1978		1994-1995	A/Shandong/9/1993
1978-1979	A/Texas/1/1977	1995-1996	A/Johannesburg/33/1994
1979-1980		1996-1997	A/Wuhan/359/1995
1980-1981	A/Bangkok/01/1979	1997-1998	A/Sydney/5/1997
1981-1982		1998-1999	
1982-1983		1999-2000	
1983-1984	A/Philippines/2/1982	2000-2001	A/Moscow/10/1999
1984-1985		2001-2002	
1985-1986		2002-2003	
1986-1987	A/Christchurch/4/1985	2003-2004	
1987-1988	A/Leningrad/360/1986	2004-2005	A/Fujian/411/2002
1988-1989	A/Sichuan/02/1987		
1989-1990	A/Shanghai/11/1987		
1990-1991	A/Guizhou/54/1989		

(B)

Seasonal year	WHO recommended H3 composition	Egg-based	Cell-based
2010-2011 2011-2012	A/Perth/16/2009		
2012-2013 2013-2014	A/Victoria/361/2011		
2014-2015	A/Texas/50/2012		
2015-2016	A/Switzerland/9715293/2013		
2016-2017	A/Hong Kong/4801/2014	A/Victoria/673/2014 A/New Caledonia/71/2014 A/Hong Kong/7127/2014 A/Hong Kong/4801/2014 (X-263A/B) A/Norway/2178/2014 A/Saitama/103/2014 (CEXP002)	A/Singapore/GP2050/2015 A/Hawaii/47/2014
2017-2018			
2018-2019	A/Singapore/INFIMH-16-0019/2016		A/North Carolina/04/2016 A/Canberra/7/2016
2019-2020	A/Kansas/14/2017	A/Kansas/14/2017 (X-327) A/Indiana/08/2018	A/Kansas/14/2017 (X-327) A/Indiana/08/2018

Supplemental Table C-2. A summary of the number of conserved Teff epitope changes according to influenza season and respective point of reference viruses.

(A) Characteristic dataset (Dataset A)

Influenza season	Point of reference	Conserved Teff epitope (count)						
		Maximum	Minimum	Range	Median	Mean	Average change within season*	Average change between season**
1968-1974	Origin/HK68	87	73	14	78.0	79.96	-7.04	-
1974-1976	PC73	94	83	11	89.5	88.33	-5.67	8.38
1976-1978	VI75	94	64	30	86.5	83.64	-10.36	-4.69
1978-1980	TX77	90	80	10	85.0	85.00	-5.00	1.36
1980-1983	BK79	92	76	16	85.0	85.30	-6.70	0.30
1983-1986	PH82	93	76	17	79.0	80.33	-12.67	-4.97
1987-1988	LE86	98	86	12	87.0	89.50	-8.50	9.17
1988-1989	SI87	98	89	9	91.0	91.86	-6.14	2.36
1989-1990	SH87	98	81	17	89.0	88.89	-9.11	-2.96
1991-1993	BJ89	91	74	17	90.0	87.82	-3.18	-1.07
1993-1994	BJ92	90	77	13	83.0	83.03	-6.97	-4.79
1994-1995	SD93	89	76	13	81.0	81.34	-7.64	-1.67

1995-1996	JO94	93	74	19	80.0	79.94	-13.06	-1.42
1996-1998	WU95	92	78	14	84.5	84.33	-7.67	4.39
1998-2000	SY97	92	78	14	81.0	82.00	-10.00	-2.33
2000-2004	MO99	94	74	20	79.0	79.46	-14.54	-2.54

(B) Contemporary sequences (Dataset B)

Influenza season	**Point of reference**	**Conserved Teff epitope (count)**						
		Maximum	**Minimum**	**Range**	**Median**	**Mean**	**Average change within season**	**Average change between seasons**
2010-2012	VAC/PE09	100	81	19	88.5	88.48	-11.52	-
2012-2014	VAC/VI11	98	77	21	86.0	85.68	-12.32	-2.80
2014-2015	VAC/TX12	98	80	18	86.0	86.27	-11.73	0.59
2015-2016	VAC/SW13	101	79	22	87.0	88.58	-12.43	2.31
2016-2018	VAC/HK14	96	78	18	87.0	87.34	-8.66	-1.23
2018-2019	VAC/SG16	98	74	24	80.0	81.36	-16.64	-5.98
2019-2020	VAC/KS17	98	71	27	97.0	93.25	-4.75	11.89

*The average change within season is calculated by subtracting the mean from the maximum number of conserved Teff epitope. Negative values indicate decrease in conserved Teff epitope.

**The average change between seasons is calculated by taking the difference of mean between two consecutive seasons. Negative values indicate decrease in conserved Teff epitope.

Supplemental Table C-3. The selected principal components (PC1 - 4) that best represent the variances of the analyzed datasets. The variable loadings/rotation are shown for each PCs. The numbers in bold indicate the respective property/characteristic is significantly vary between isolates.

(A) Dataset A

Source of variation	PC1	PC2	PC3	PC4
Total Binding Affinity Dim.1	**-0.3006**	-0.01241	0.05398	-0.008993
DRB10101 Dim.1	**-0.2975**	-0.02809	-0.02573	0.01131
DRB10401 Dim.1	**0.2925**	-0.01833	0.05705	-0.04460
DRB11101 Dim.1	**-0.2874**	-0.06834	0.03655	-0.09573
T Cell Epitope Distance Dim.1	**-0.2856**	-0.07052	-0.06897	0.06838
DRB10901 Dim.1	**-0.2845**	-0.05783	-0.03906	0.1120
DRB10301 Dim.1	**-0.2795**	-0.05191	0.1618	-0.07442
DRB10801 Dim.1	**-0.2733**	0.1023	-0.03222	0.1316
DRB10701 Dim.1	**-0.2717**	0.05788	0.1662	-0.05902
DRB11301 Dim.1	**0.2551**	-0.05564	**-0.2347**	-0.1103
DRB11501 Dim.1	**-0.2513**	0.1957	-0.01092	0.08756
Sequence Immunogenicity Score	-0.1889	0.1677	-0.1335	0.04288
DRB11501 Dim.2	0.1507	**0.3066**	-0.1156	**0.2089**
T Cell Cross-conservation Dim.2	0.1320	0.07239	0.1621	**0.2627**
Sequence JMX Homology Score	-0.1297	-0.1017	0.1651	**0.2637**
DRB11301 Dim.2	-0.1237	-0.1056	**-0.4448**	0.07527
T Cell Cross-conservation Dim.1	-0.07345	0.1830	-0.01266	**-0.4124**
T Cell Epitope Distance Dim.2	-0.06902	**0.3424**	0.07952	-0.1384
DRB10701 Dim.2	-0.06832	-0.004963	**-0.4948**	-0.03249
DRB10801 Dim.2	0.06166	**0.3246**	-0.03034	-0.1650
DRB10301 Dim.2	0.05271	0.1756	**0.4540**	0.06794
DRB11101 Dim.2	-0.04963	**0.3276**	-0.1346	0.03926

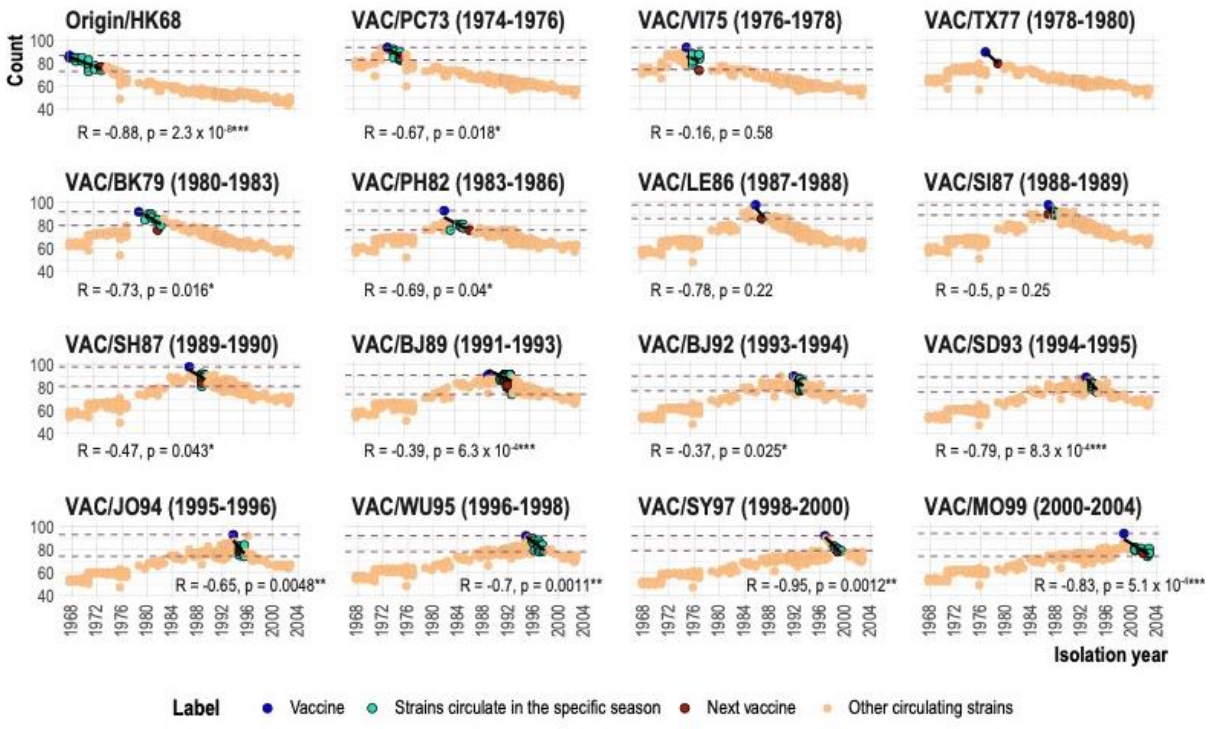
DRB10901 Dim.2	-0.04317	0.009210	-0.03411	**-0.6750**
DRB10401 Dim.2	0.04108	**0.2553**	**-0.3073**	**0.2360**
DRB10101 Dim.2	-0.01584	**0.3973**	-0.05968	-0.0001559
Total Binding Affinity Dim.2	-0.002226	**0.4019**	0.1313	0.02224

(B) Dataset B

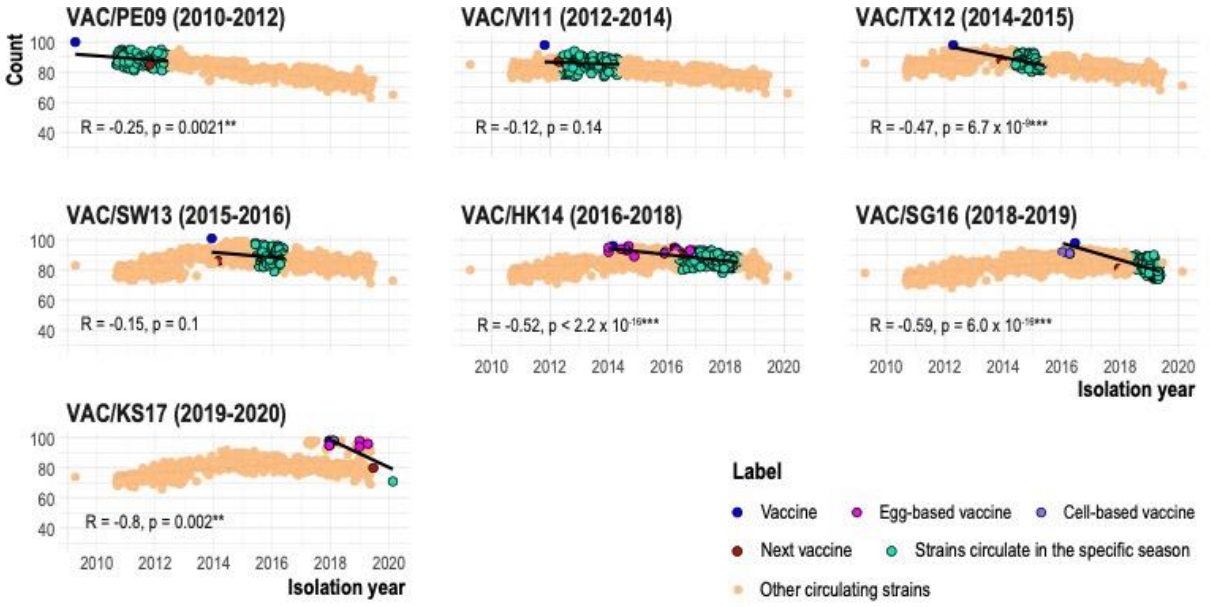
Source of variation	**PC1**	**PC2**	**PC3**	**PC4**
Total Binding Affinity Dim.1	**0.2659**	0.006660	0.04239	0.02734
DRB11101 Dim.1	**0.2655**	0.02769	0.03093	0.009567
DRB10301 Dim.1	**0.2653**	-0.03160	0.02785	0.01105
DRB10101 Dim.1	**0.2648**	0.03398	0.03558	0.02011
DRB10801 Dim.1	**0.2634**	0.04270	0.03943	0.01892
DRB10701 Dim.1	**-0.2575**	-0.02221	-0.06606	-0.04009
DRB11501Dim.1	**-0.2496**	-0.02805	-0.08099	-0.1791
T Cell Epitope Distance Dim.1	**0.2113**	-0.1226	-0.1327	0.1066
DRB10401 Dim.1	0.1918	**-0.2451**	0.03695	-0.002229
DRB10401 Dim.2	0.1670	**0.2457**	0.05843	0.1067
DRB10901 Dim.2	0.1316	**-0.2223**	**-0.2110**	0.1770
DRB11301 Dim.1	-0.1236	**0.3004**	-0.01195	-0.03226
Sequence Immunogenicity Score	-0.1092	-0.09121	**0.2734**	-0.09195
DRB10901 Dim.1	0.09571	-0.1262	**0.3930**	-0.09218
DRB11501 Dim.2	0.06539	0.02152	**-0.2933**	**-0.4132**
Sequence JMX Homology Score	0.02532	0.02340	0.1044	0.1843
T Cell Cross-conservation Dim.1	0.02512	-0.009797	0.1107	**0.2634**
DRB10801 Dim.2	0.02491	-0.006903	**-0.3377**	-0.1593
DRB10701 Dim.2	-0.02441	0.004862	-0.1006	**0.3105**
T Cell Cross-conservation Dim.2	0.02393	0.03514	**-0.2057**	-0.1832
DRB10101 Dim.2	0.02379	**-0.2475**	**-0.2846**	0.1364
DRB10301 Dim.2	0.02282	**0.3524**	-0.02384	0.04005

DRB11101 Dim.2	-0.02216	**0.3481**	-0.0005820	0.01179
T Cell Epitope Distance Dim.2	-0.02097	**-0.2734**	**0.2208**	-0.04261
DRB11301 Dim.2	-0.01255	-0.01659	**-0.3861**	**0.3352**
Total Binding Affinity Dim.2	0.006436	**-0.3597**	0.005273	0.0008344

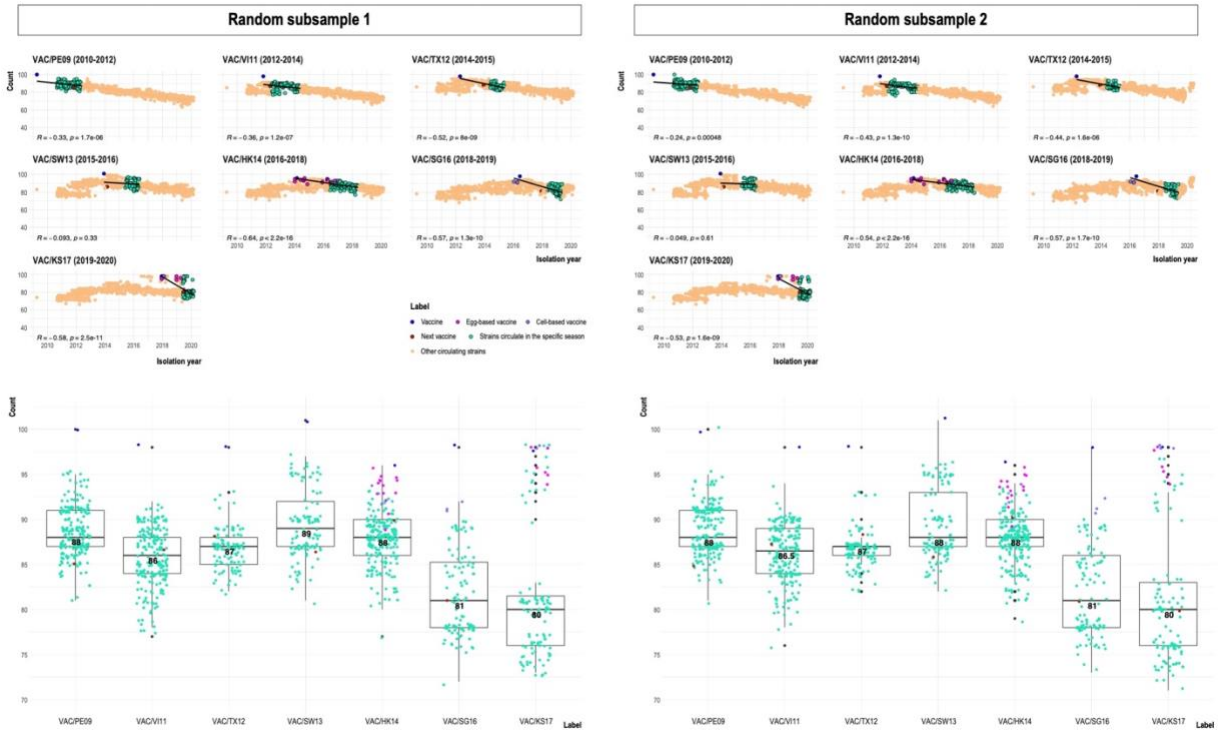
(A)


Origin/HK68
R = -0.88, p = 2.3 x 10-8***
VAC/PC73 (1974-1976)
R = -0.67, p = 0.018*
VAC/VI75 (1976-1978)
R = -0.16, p = 0.58
VAC/TX77 (1978-1980)
VAC/BK79 (1980-1983)
R = -0.73, p = 0.016*
VAC/PH82 (1983-1986)
R = -0.69, p = 0.04*
VAC/LE86 (1987-1988)
R = -0.78, p = 0.22
VAC/SI87 (1988-1989)
R = -0.5, p = 0.25
VAC/SH87 (1989-1990)
R = -0.47, p = 0.043*
VAC/BJ89 (1991-1993)
R = -0.39, p = 6.3 x 10-4***
VAC/BJ92 (1993-1994)
R = -0.37, p = 0.025*
VAC/SD93 (1994-1995)
R = -0.79, p = 8.3 x 10-4***
VAC/JO94 (1995-1996)
R = -0.65, p = 0.0048**
VAC/WU95 (1996-1998)
R = -0.7, p = 0.0011**
VAC/SY97 (1998-2000)
R = -0.95, p = 0.0012**
VAC/MO99 (2000-2004)
R = -0.83, p = 5.1 x 10-4***
Count
Isolation year
Label
Vaccine
Strains circulate in the specific season
Next vaccine
Other circulating strains

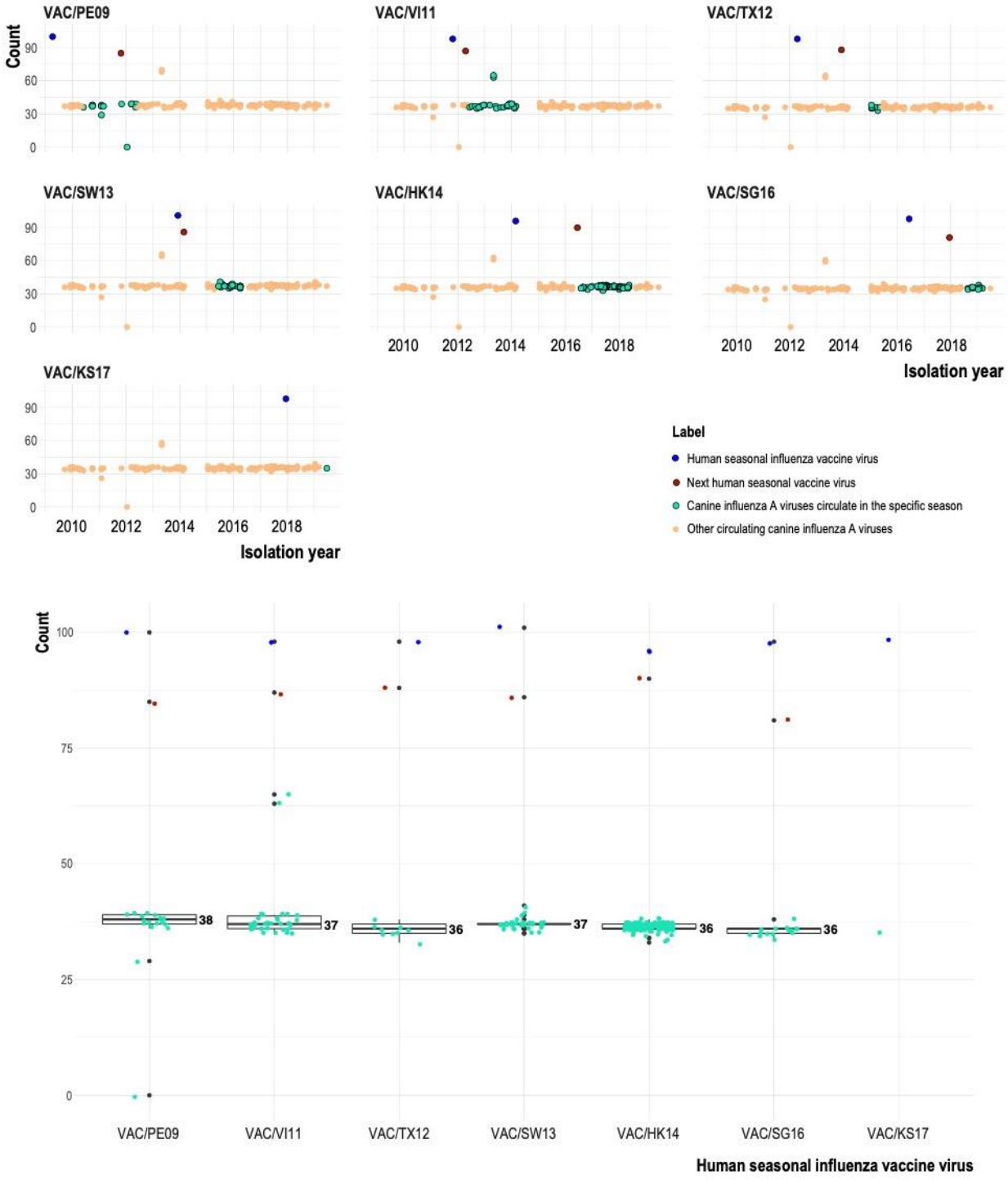
(B)


VAC/PE09 (2010-2012)
R = -0.25, p = 0.0021**
VAC/VI11 (2012-2014)
R = -0.12, p = 0.14
VAC/TX12 (2014-2015)
R = -0.47, p = 6.7 x 10-9***
VAC/SW13 (2015-2016)
R = -0.15, p = 0.1
VAC/HK14 (2016-2018)
R = -0.52, p < 2.2 x 10-16***
VAC/SG16 (2018-2019)
R = -0.59, p = 6.0 x 10-16***
VAC/KS17 (2019-2020)
R = -0.8, p = 0.002**
Count
Isolation year
Label
Vaccine
Egg-based vaccine
Cell-based vaccine
Next vaccine
Strains circulate in the specific season
Other circulating strains

Supplemental Figure C-1. An extended overview of conserved Teff epitopes over time with reference to H3N2 vaccine virus updates from (A) 1968 to 2004 and (B) 2010 to 2020. Refer to Figure 5.2 and 5.6 for more details.

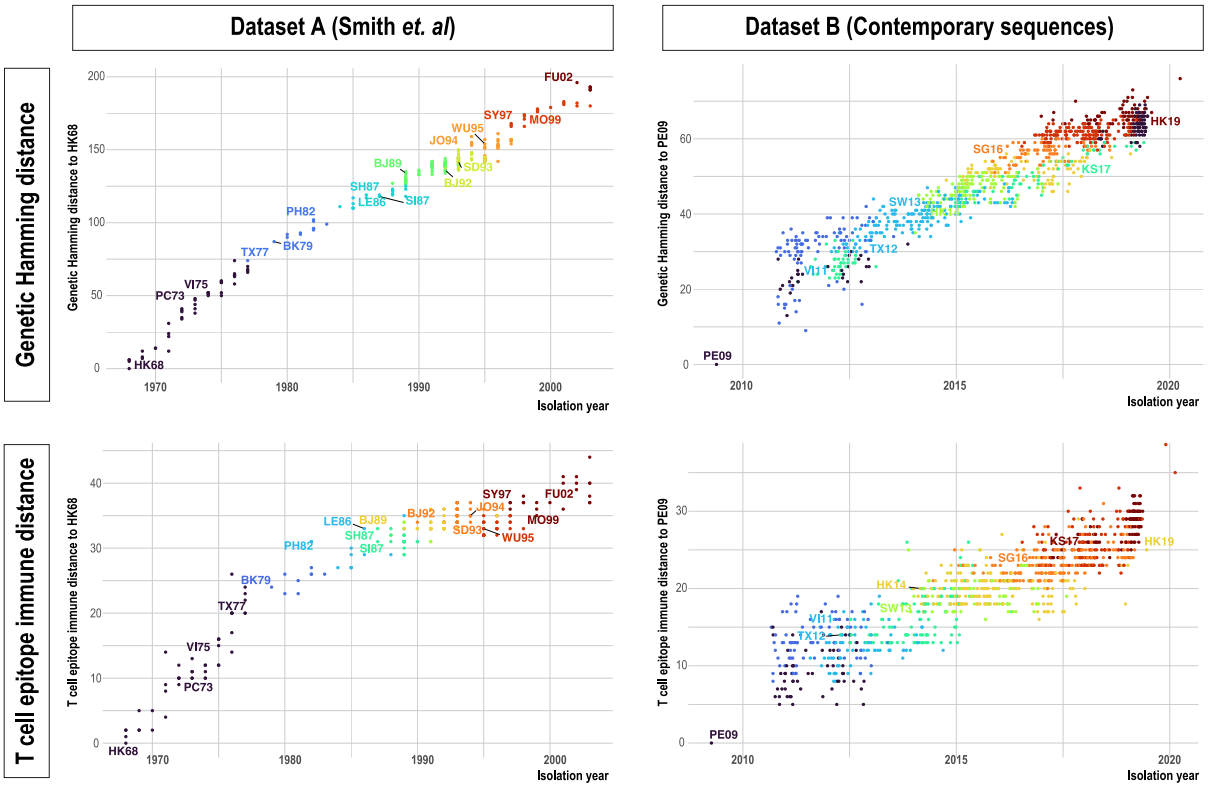


Supplemental Figure C-2. The analysis of the conserved Teff epitopes over time with reference to H3N2 vaccine virus updates from 2010 to 2020 for two replicate random subsamples. The results show outcomes that are consistent with findings in dataset B (Figure 5.6 and Supplemental Figure C-1B).



Supplemental Figure C-3. The control group analysis of the conserved Teff epitopes over time with reference to H3N2 vaccine virus updates from 2009 to 2019 using canine A/H3N2

viruses. The result confirms that the outcome in Figure 5.2, 5.6, Supplemental Figure C-1 and C-2 are not of random event.



Supplemental Figure C-4. The comparison of genetic hamming distance to T cell epitope immune distance for the two datasets with respect to HK68 and PE09, respectively. The progression of both genetic drift and T cell epitope immune pressure appear to be linear over time.