Winter 12-15-2017

Influenza A virus Genomic Reassortment and Packaging

Graham Devin Williams
Washington University in St. Louis

WASHINGTON UNIVERSITY IN ST. LOUIS

Division of Biology and Biomedical Sciences
Molecular Microbiology and Microbial Pathogenesis

Dissertation Examination Committee:
Adrianus C. M. Boon, Chair
Gaya Amarasinghe
Michael Diamond
Sebla Kutluay
Deborah Lenschow
David Wang

Influenza A virus Genomic Reassortment and Packaging

by
Graham D. Williams

A dissertation presented to
The Graduate School
of Washington University in
partial fulfillment of the
requirements for the degree
of Doctor of Philosophy

December 2017
St. Louis, Missouri

Table of Contents

List of Figures

List of Tables

Acknowledgments

I would like to thank my mentor, Dr. Adrianus “Jacco” Boon, for support, guidance, and persistent reassurance throughout this project. Dr. Boon has relentlessly worked to better me as a scientist, something, for which I will forever be grateful. Thank you to my thesis committee- all of you have contributed to this project through constructive and challenging feedback, asking inquisitive and provocative questions, and sharing your expertise with me. Over the last six years, the Departments of Molecular Microbiology and Microbial Pathogenesis and Infectious Diseases have become my home. Thank you all for sharing intellectual and material assistance. I am particularly grateful to past and present the members of the Boon Lab including Anshu Gounder, Traci Bricker, Camille Linton, and Brittany Doll for your companionship in lab throughout our years together. Thank you to the Diamond Lab members and alumni for providing support and feedback in all matters. Additionally, to the undergraduate and rotation students I have mentored – thank you for the opportunity to work with you and have a reciprocal teaching opportunity in which we both continued learned.

Thank you to Drs. Sebla Kutluay and Kristine Wylie for enabling our studies of influenza A protein-RNA interactions and computational interpretation of our data. Additionally, thank you to the many collaborators for providing reagents and insight as our studies progressed. Thank you to Jeanne Silvestrine for being our program coordinator and keeping my entire cohort on track. Dr. Sibley, thank you for the opportunity to facilitate the Microbiology Student Journal Club and Dr. Jim Skeath for the opportunity to coordinate the Genetics and Microbiology Student Run Seminar Series. The past and continued support of the Cell and Molecular Biology and Infectious Disease Training Grants as well as the Victoria Fraser Graduate Fellowship facilitated all of this work and must be acknowledged. All of these experiences have enriched my graduate training inside and out of lab.

Outside of lab, I am incredibly thankful to the friendships I have had in St. Louis. Thanks to Carlos, Nick, Melissa, Caitlin, JP, JJ, Shannon, Valerie, Katherine, Sidd, and Greg for your friendship, support, and making graduate school bearable. To the many friends I have made outside of lab including Bill, Elizabeth, Diego, Joey, and Tim you have enhanced nights and weekends. For those who I left to pursue this degree, but have remained intertwined in my life -

Tim, Kayla, Matt, Andy, Evan, Brad, and Brittany – I continue to value our past and present friendship. Finally, to Caitlin, thank you for your unending support and partnership in science and all else.

Thank you to my family for always supporting and loving me. My parents, Thomas and Susan, have been consistently supportive of me, and I would be never had made it here without them. To my brother, Spencer, you are a lightening reprieve from graduate school when it was most necessary. To all of my extended family that has played a part in shaping my path, thank you as well.

Graham D. Williams

Washington University in St. Louis

December 2017

ABSTRACT

Influenza A virus Reassortment and Genome Packaging

by

Graham D. Williams

Doctor of Philosophy in Biology and Biomedical Sciences

Molecular Microbiology and Microbial Pathogenesis

Washington University in St. Louis, 2017

Professor Adrianus Boon, Chair

Influenza A viruses (IAV) are a major human and environmental pathogen. IAV successfully infects a diverse host range and adaptation of new viral strains to humans may cause pandemic events with high morbidity and mortality. As a member of the *Orthomyxoviridae* family, IAV inherently possesses a segmented genome, which enables a process of segment transmission between viruses following cellular co-infection, a process termed reassortment. The high rate of IAV mutation and continued co-circulation of diverse viral strains in divergent host species leads to the persistent prospect for emergence of new IAV with pandemic potential. Therefore, it is of great importance to understand the viral and host factors that restrict and promote the generation of emergent virus strains, their potential for pathogenesis, and discover novel mechanistic countermeasures against IAV, including improved vaccination and targeted therapeutic strategies.

Human and avian IAV co-circulate and occasionally co-infect the same host, leading to the potential for generation of novel genome constellations following reassortment. The specific host and viral molecular determinants that allow replication of reassortant progeny virus are not well defined. Here, I show that the viral genetic context and host cell in which reassortment occurs determine the potential for genetic diversity derived from multiple distantly related strains. Importantly, we identify single gene reassortants between a North American avian strain and the 2009 pandemic H1N1 virus that are capable of causing disease in mammals and replicate in a human cell line as well as induce the production of several pro-inflammatory cytokines linked to severe disease outcomes. Additionally, utilizing a different viral genetic background, I show that the reassortment potential is regulated by species and cell type specific differences in

viral replication due to augmented viral polymerase function dependent on the identity of a single amino acid in the PA protein. Together, these studies provide evidence that context-dependent compatibility between both viral and host factors determine the possibility for generation of novel reassortant genome constellations and regulate their potential for replication and transmission in new host species.

Reassortment between IAV strains is likely dictated by the functional compatibility of vRNA segments bound by IAV nucleoprotein during genome packaging. I hypothesized that nucleoprotein (NP) scaffolds specific RNA elements that are required for genome packaging and interaction between viral RNA (vRNA) genome segments. Therefore, I sought to determine the functional consequences of genome architecture on genome packaging and for the first time determine the nucleotide-resolution landscape of NP-vRNA interactions in infected cells. We utilized Photoactivatable Ribonucleoside-Enhanced Crosslinking and Immunoprecipitation (PAR-CLIP) coupled to next-generation sequencing to determine the specific interaction sites of vRNA bound by NP. We then interrogated the functional importance of regions of vRNA bound or unbound by NP and identified a number of potentially structured RNA features required for efficient genome packaging and virus propagation. These studies provide a framework for understanding the multifactorial restrictions of IAV reassortment and potential for generation of novel genome constellations with pandemic potential. Finally, these studies expand our understanding of how viral and host determinants shape the possible evolutionary trajectories of IAV through reassortment and required genetic elements needed for genome assembly.

CHAPTER 1:
INTRODUCTION

1.1 Influenza A virus as a Cause of Human and Animal Disease

Influenzavirus A viruses are part of the *Orthomyxoviridae* family that also consists of related but distinct members including *Influenzavirus B*, *Influenzavirus C*, *Thogotovirus*, and *Isavirus*[1]. While Influenza A, B, and C genera and *Thogotoviruses* infect humans and a wide variety of other organisms, *Isaviruses* are presently understood as restricted to aquatic vertebrates. Influenza A (IAV) and B (IBV) viruses pose the greatest risk to humans. IAV can infect a highly divergent cohort of birds, humans, domestic pigs, horses, dogs, and aquatic mammals but influenza B virus is restricted to humans.

Influenza infection in humans is generally limited to the respiratory epithelium, although dissemination to other organs has been observed in some highly pathogenic strains[2,3]. Transmission of the virus occurs by respiratory droplet as well as contact with contaminated surfaces. High population density (i.e. schools or workplaces) and direct exposure to infected individuals leads to increased transmission of the disease[4]. As a respiratory pathogen, the mucosal epithelial surface and innate immune system play large roles in host protection[5]. In dry air, mucous in the airway is thinned and leads to greater likelihood of infection[6]. Influenza pathogenesis is mediated by immune infiltration and inflammation in the airway as well as enhanced inflammatory cytokine release leading to edema at the sites of infection and replication[7]. Symptoms consist of coughing, fever, rhinitis, and malaise occurring within 2 days of exposure that general resolve within 7-10 days. Viral shedding and contagiousness may occur before the onset of symptoms making disease control and spread a large public health issue.

1.2 Influenza A Ecology and Zoonosis

IAV is a genetically diverse viral species, presently including 17 HA subtypes and 10 NA subtypes, classified by both differential immune recognition and sequence composition[8]. H1N1, H3N2, and H2N2 subtypes have circulated widely in the human population, while other subtypes are largely restricted to other mammals or the natural reservoir of migratory and shore birds[9]. Avian IAV including subtypes H5, H7, and H9 cause sporadic, sometimes fatal disease in humans and are major agricultural pathogens[10]. Zoonosis of these viruses or derivative lineages formed by reassortment with human transmissible strains may lead to the emergence of viruses with pandemic potential[11–13]. Therefore, it is of great importance to determine the molecular basis of replication and pathogenesis of contemporary avian viruses in the mammalian host. IAV's containing subtype H7 Hemagglutanin may pose the greatest threat to human health beyond seasonal influenzas of subtypes H1 and H3[14]. In the past decade there have been multiple outbreaks of H7N1, H7N3, and most recent H7N9 viruses in avian and human populations[15–17]. These outbreaks are geographically distinct and are not the result of movement of viruses across geographic regions.

Encounters with H7 viruses are likely common in human populations, though substantial adaptation may have to occur before they can cause fulminant respiratory disease. Seroprevalence studies in Italy, China, and North America reveal the presence of antibodies to H7 (at least H7 cross-reactive) antibodies in many people[18]. Additional human cases in North America have been documented. A small 2004 outbreak in British Columbia resulted in the infection of two poultry workers[19].

In addition to these low-pathogenic strains (LPAIs), the H7 subtype may harbor a multi-basic cleavage site (MBCS) found in highly-pathogenic strains (HPAIs)[20]. Interestingly, these strains are able to acquire this cleavage signal likely through recombination with host rRNA.

During 2012-2013 a highly pathogenic H7N3 strain in Mexico acquired an MBCS through inclusion of a 11 amino acid sequence immediately proximal to the H0-H1 cleavage site. The resulting virus is more pathogenic in Galliformes and can cause disease in mammals as evidenced by multiple infections as a result of occupational exposure.

North American H7 viruses isolated from aquatic birds have receptor binding affinity to human-like sialic acids [19]. Additionally, most humans are not seroreactive to H7-HA and this population-level naivety, and lack of neutralizing antibodies, may allow rapid spread of a greater outbreak than potentially a moderately, antigenically distinct H1 or H3. Both H7- and H1-subtype viruses are capable of infecting mammalian hosts. As both H7 and H1 subtype viruses co-circulate geographically, there is the possibility that reassortment a process described below, between the viruses may result in a strain with altered properties[11].

1.3 Genome Reassortment Produces Pandemic Influenza Viruses

IAV possess a genome composed of eight single-stranded negative-sense RNA segments encapsulated in a nucleoprotein coat and polymerase complex packaged in a host-cell derived envelope[21]. In the event of cellular coinfection, viral ribonucleoprotein complexes (vRNPs) from multiple strains of the virus are able to simultaneously utilize the host cell for replication and genome packaging. If a mixed genome is successfully packaged, resulting viruses with segments from both parental strains result - giving rise to a new genome constellation[9]. This process, termed reassortment, is one of the evolutionary strategies employed by IAV to produce new viral strains that may have altered replication, pathogenesis, and transmission profiles when compared to previous isolates. For reassortment to be successful new genomic constellations must be

replication competent and when there is little difference between constituent segments of two viruses reassortment may occur at a high frequency.

Influenza reassortment has led to the generation of multiple viruses capable of generating pandemics[22–25]. Requisite for pandemic formation is the ability of the virus to break animal-to-human transmission barriers, replicate and release efficiently in the human host, and transmit form human-to-human[26]. The profound effect of reassortment may be illustrated by inspecting the phylogeny of pandemic IAVs from 1918 to 2009 and the recently emergent H7N9[27,28]. In each case an antigenically distinct HA was acquired by way of reassortment with a virus capable of transmitting to and replicating in humans[29,30]. Humoral immunity to HA is the primary basis of current vaccination strategies and antigenic shift of HA by reassortment alters recognition by antibodies. The 1918 H1N1 "Spanish Flu" was likely created by reassortment of at least one avian virus and a mammalian virus that circulated in porcine hosts before introduction to humans[31,32]. Between 20-50 million people died as a result of primary influenza infection and subsequent secondary pneumonia. Contemporary segments with high homology to the 1918 H1N1 strain persist in both avian and swine populations[27]. Experimental creation of an H1N1 strain derived from currently circulating avian genome segments was also shown to cause severe disease in mammals and can transmit between ferrets. Moreover, reassortment between human and avian IAV, likely in a porcine vessel, resulted in the generation of H2N2 and H3N2 strains that caused pandemics in 1957 and 1968, respectively. H3N2 viruses have continued to circulate following the pandemic and are, along with H1N1, the primary cause of seasonal epidemics.

Prior to each pandemic, antigenic shift of HA or NA occurred in addition to transfer of internal gene segments. The 2009 pandemic H1N1 emerged following at least three independent reassortment events involving human, porcine, and avian viruses[32]. The compatibility of genome

segments from this virus with both high and low pathogenic contemporary avian influenza viruses in experimental settings is concerning. Mammal-to-mammal transmission of non-adapted single gene reassortants has been experimentally demonstrated[4,11,13,33,34]. Reassortment of IAV viruses occurs readily in nature and may be simulated in the lab using either co-infection or forced reverse genetic approaches.

Many recent studies have assessed the effect of reassortment between 2009 pandemic H1N1 and other virus strains. In many cases reassortment resulted in increased pathogenesis of porcine viruses. Additionally, many cases of reassortment between 2009 pandemic H1N1 and endemic porcine viruses have been documented. Paired with the potential for reassortment with wild aquatic birds, shorebirds and mallards included, investigation of the potential for reassortment and altered pathogenesis in LPAI H7 viruses with the 2009 pandemic H1N1 virus are imperative. For instance, does the acquisition of structural or non-structural genes from human viruses lead to increased replication, pathogenesis, or transmission of environmentally obtained North American avian influenza viruses? This question will be partially addressed by Chapter 2 of this thesis.

1.4 IAV Genome Structure and Virally Encoded Proteins

Upon infection of the host cell, negative-sense genome segments are trafficked to the nucleus and generate a positive-sense RNA intermediate that acts as a template for genome replication and may be enzymatically modified to serve as message for translation of viral proteins[35]. Influenza viruses possess an RNA-dependent RNA Polymerase (RdRP) complex that also facilitates an endonucleolytic cap-snatching mechanism for co-opting the host transcriptional machinery to produce viral proteins.[36] Individual IAV particles are thought to

contain one copy of each of the eight genomic segments[37,38]. The proportion of infectious to defective particles created during infection remains controversial, potentially a consequence of inefficient genome packaging[39]. Encoded by these eight segments are at least 12 functionally distinct proteins: PB2, PB1, PB1-F2, PA, PA-X, HA, NP, NA, M1, M2, NS1, and NEP (NS2) (**Table 1.1**)[1]. PB2, PB1, and PA along with NP compose the viral replication machinery. HA and NA are surface glycoproteins associated with host cell attachment and budding respectively[34]. The nonstructural proteins NS1 and NEP act, respectively, as regulators of the host immune response and in export of vRNPs from the nucleus[40,41]. All of these proteins intimately interact with host cellular components during the course of entry, endosomal fusion, replication, and budding leading to dissemination of the virus and pathogenic features of infection (**Figure 1.1**)[42].

Segment (length, nt)	Protein	Primary Function
1 (2341)	PB2	Cap-binding domain of RdRP*
2 (2341)	PB1	Catalytic subunit of RdRP
	PB1-F2	Pro-apoptotic, immune modulatory
3 (2233)	PA	Cap-snatching endonuclease in RdRP
	PA-X	Host cell shut-off, "non-specific" endonuclease, immune modulation
4 (1775)	HA	Surface glycoprotein, host cell attachment, membrane fusion
5 (1565)	NP	Nucleoprotein, encapsidation of viral RNAs
6 (1413)	NA	Surface glycoprotein, Neuraminidase activity, receptor destruction and particle egress from host cell
7 (1027)	M1	Matrix protein, particle morphology, vRNP export
	M2	Surface cation channel, protects HA conformation
8 (890)	NS1	Regulation of RdRP, Interferon antagonist, enhancement of viral mRNA translation
	NEP (NS2)	Nuclear export factor

Table 1.1. Proteins encoded by each IAV segment and ascribed functions.

*RdRP: RNA-dependent RNA polymerase

1.5 IAV Receptor Binding, Entry, Endosomal Fusion and Uncoating

IAV has a class I fusion protein, Hemagglutinin (HA), that adorns the viral envelope[43]. HA binds sialic acid moieties on the host cell surface and is internalized by a variety of pathways[44,45]. Spherical viruses generally enter through clathrin-dependent, receptor-mediated endocytosis of the virus followed by passage through the endocytic pathway concluding with fusion in the late endosome at low-pH[46]. Filamentous influenza A viruses enter the cell via a clathrin-independent, dynamin-independent pathway[44,47]. HA induces membrane fusion between viral and host membrane and M2 activation occurs and viral acidification induces separation of M1 and vRNPs allowing their export to the cytoplasm and subsequently the nucleus[48].

1.6 vRNP transport to the Host Cell Nucleus

Following fusion and uncoating, vRNPs must traverse the cytoplasm prior to nuclear entry[49]. In the cytoplasm vRNPs are individually bound by importin-alpha proteins that recognized nuclear localization signals within viral proteins. After importin-alpha loading, an active transport process involving host Ran-GTPases allows traversal of the nuclear pore complex[50]. The recognition by importin-alpha likely relies on a non-classical NLS in the solvent exposed N-terminus of NP present at high copy number within vRNPs[35]. Additionally, PB2 is known to interact with host importins as well, though at only one copy per vRNP, the contribution of these interactions to nuclear import remain unclear.

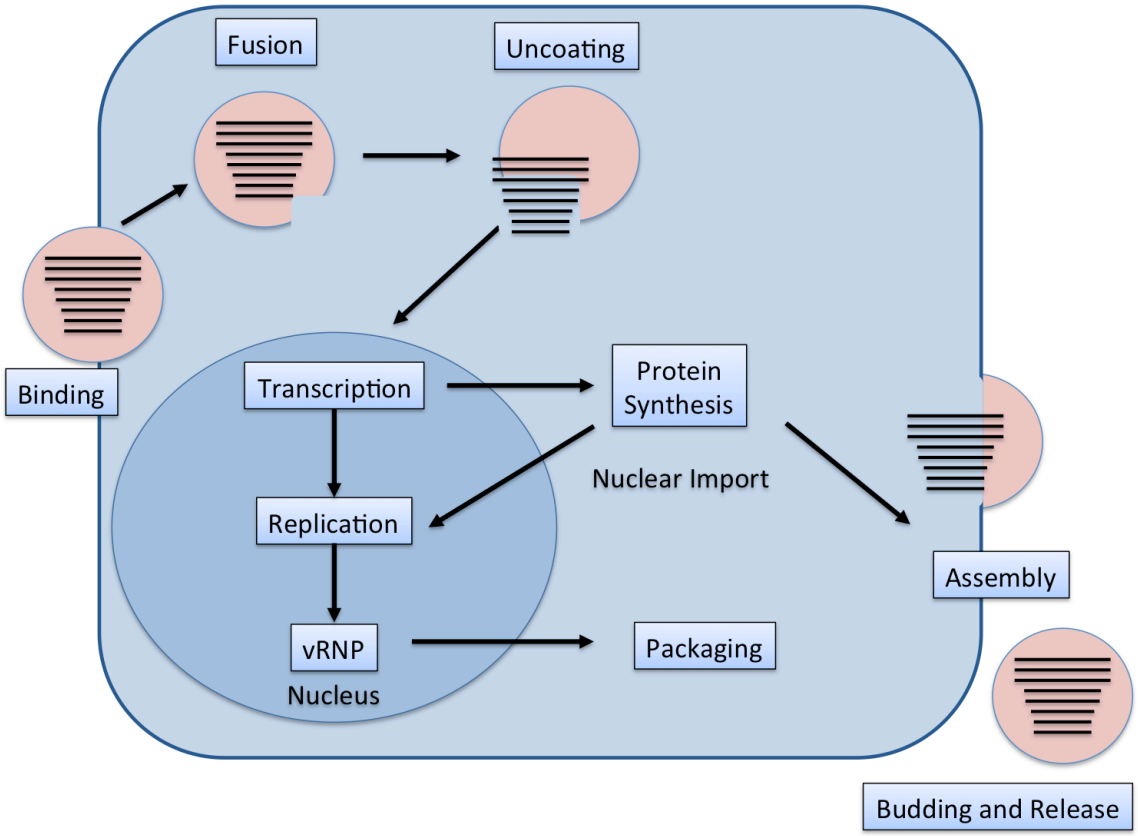


Figure 1.1. Schematic of the influenza A virus infection cycle. Adapted from Shi, et al. Nature Reviews Microbiology. 2013[1]. IAV particles bind host-derived sialic acids via HA and induce virus uptake at the cell membrane. IAV is trafficked to the late endosome prior to pH-dependent HA fusion in the endolysosome. Particle uncoating occurs as M2 is activated and M1 detaches from vRNPs, allowing their export to the host cytoplasm. vRNPs are transported to the nucleus via host proteins and cross the nuclear pore complex and pioneer transcription and protein production proceed. Following accumulation of proteins involved in genome replication, a positive-sense replication intermediate (cRNP) is generated from vRNA template, and then transcription of nascent vRNPs proceeds and amplifies viral RNA content in the nucleus. vRNPs

are actively transported from the nucleus to the cytoplasm where they assemble as multi-segmented foci at Rab11a-positive vesicles and are trafficked en masse to the interior of the plasma membrane at sites of viral budding. Progeny virus bud from the cell membrane and via M2-dependent scission thereby enabling release from the cell.

1.7 Viral Protein synthesis, Cap-snatching, and Genome Replication

Following nuclear import, the initial round of primary transcription from vRNA to viral mRNA must occur. Primary transcription is initiated by trans-acting vRNPs to facilitate (+)-strand synthesis aided by many cellular factors including RNA polymerase II[51]. For viral mRNA to be produced and translated, a host-derived 5'leader sequence must be appended to the end of (+)-strand viral RNA. This process, termed cap-snatching, occurs via the endonucleitic cleavage of host mRNAs, via PA, following cap recognition by PB2 in the context of the heterotrimeric polymerase complex[36]. The first 10-15 nucleotides of host RNA are appended to the viral RNA and used as a template for translation following export from the nucleus[52]. A 3'-poly-A tail is added by the host protein SFPQ and splicing of certain viral mRNAs occurs[53]. In the case of viral transcripts lacking splice variants, these molecules are transported via the canonical host export pathways including NXF1[54]. For transcripts that possess retained introns (M1 and NS1), NS1 protein is required for efficient export due to occlusion of binding by host transcript quality control proteins and the proofreading machinery[54,55]. Once exported the host ribosome translates viral proteins. Each newly synthesized polymerase complex component must then be imported to the nucleus for continued replication[41]. Multiple host factors function in concert with the viral polymerase complex to transcribe and replication genomes[56,57]. After buildup of viral proteins, cRNA production begins and continues processively with the help of the cellular minichromosome maintenance complex, a helicase and clamp complex which allows extended

unwinding and synthesis along full-length transcripts[58]. Once unmodified cRNA is generated, it is used as a copy-back template to synthesize vRNA to be packaged into progeny virions. In both cases the synthesis of cRNA and vRNA likely occurs via RNP non-resident polymerase complexes[59]. As RNA synthesis via the heterotrimeric polymerase component, PB1 proceeds rapidly, it is possible nearly co-transcriptional coating of vRNA by NP may occur[35]. As NP oligomerization along vRNA may happen slower than synthesis, local RNA structures may form. These structures may facilitate potential inter-segment interactions to be discussed later in this text.

1.8 vRNP Export and Transit to Budding Virion

Following vRNP synthesis, genome segments must be actively exported from the nucleus to the sites of virus particle assembly. M1 participates in the transport of viral RNPs (vRNPs) across the nuclear compartment by bridging the vRNP and NEP (NS2)[40,60,61]. This export process is CRM1 and nuclear pore dependent. Following export of the CRM1-NEP-M1-vRNP complex from the nucleus, PB2 in the vRNP associates with a virus-induced Rab11a-positive vesicle[41,62,63]. Docking of vRNPs at this structure has shown to be dependent on vRNP-associated PB2 binding host Rab11a. At this point, or before, higher order clusters of vRNPs form in the cytoplasm[64]. The basis for molecular clustering remains undetermined but this portion of genome assembly provides a platform for competition between genome segments[65]. As subgenomic packets of vRNPs can be found directly following nuclear export, it seems likely that once vRNPs are loaded onto these transport vesicles via PB2, the local concentration of vRNPs is great enough for assembly of higher order, complete genomes. These interactions may occur via vRNA-vRNA interactions, in trans, between vRNPs and be potentiated by proximity[66–68].

1.9 Influenza A Genome Segment and RNP Architecture

Each vRNA genome segment contains at least one open reading frame (ORF), and two distal untranslated regions[69–72]. The 5' and 3' UTR, along with an overlapping portion of the ORF are responsible for packaging of the segment into progeny viruses. Segment termini all contain a universal 12 and 13 nucleotides, 5' and 3' respectively, that are nearly complementary and form a corkscrew-like promoter implicated in priming of cRNA and vRNA synthesis by the polymerase complex composed of PB2, PB1, and PA[49,73,74].

IAV RNP complexes are comprised of the heterotrimeric polymerase complex as well as viral nucleoprotein and RNA. Using a number of methods, including mass-spectrometry and cryo-electron microscopy (EM), nucleoprotein has been hypothesized to bind approximately 24-32 bases per monomer by calculating the number of NP monomers per polymerase in purified virus or by dividing the length of a segment by the number of NP molecules present[59,69,75,76]. Conversely, NP spontaneously assembles into vRNP-like helical structures in the presence of 11-13 nucleotides, and this discrepancy has not been fully reconciled[77]. Modeling of NP-RNA interactions additionally predicts the phosphate backbone of a 12 nucleotide model RNA may be accommodates in the positively charged RNA-binding groove of monomeric NP[78]. Reconstruction of vRNPs via cryo-EM describe these macromolecules as extended helical structures in which the opposing termini of each segment are coordinated by the polymerase complex. Nucleoprotein coils through interactions of a dimer interface along the segment length save for a structurally required loop in the segment end distal to the polymerase complex. NP binds to the phosphate backbone of vRNA in a manner that leaves bases exposed to solvent and

potentially able to form RNA structures in the context of the vRNP, suggesting NP coating of vRNA may be incomplete[72,79]. The structural approaches utilized to date have not quantified the specific nucleotide interactions of vRNA with NP in the context of infected cells or purified virus since the resolution limit of prior studies was not great enough to assign nucleotide identities to vRNA density in class-averaged cryo-EM structures[59,69]. Additionally, single particle cryo-EM studies demonstrate the potential for non-uniform helical torsion within vRNPs that generates heterogeneous structural topology[71]. Therefore, nucleotide resolution analysis of the NP-RNA interaction may elucidate previously unappreciated vRNP features required in IAV biology.

1.10 Influenza A RNA Features Required for Coordinated Genome Packaging

Segment specific packaging signals have been identified for each segment utilizing A/PR/08/1934 (H1N1) and A/WSN/1933 (H1N1) strains[80–84]. A/PR/08/1934 is the same strain of virus we have used in many studies and is also used as one of the genomic backgrounds for human vaccine development. Initial studies focused on internal truncations of large segments, primarily the polymerase coding segments that are able to compete with wild type segments for incorporation into the genome; these truncated segments have since been termed "Defective Interfering RNAs" (DI RNA)[85]. Importantly, these truncated segments were shown to act at the level of packaging, as they are able to compete with wild-type segments for incorporation into progeny viruses and result in replication incompetent virus. Additionally, though many reverse genetic experiments have examined the minimum requirements for DI-like RNA incorporation, most have found the inclusion of terminal coding regions in addition to conserved untranslated regions (UTR) confers a greater rate of incorporation relative to wild-type segments[86]. Similarly,

the inclusion of terminal coding regions promoted more efficient packaging than the UTR alone. More recently, reverse genetic studies employing artificial reporter segments have yielded valuable information about the minimal requirements for efficient packaging. All segments have demonstrated a requirement for a bipartite packaging signal derived from both segment termini[87]. Interestingly, in many cases truncation of one segment leads to a significant decrease in the packaging efficiency of other segments when measured on a segment-specific basis[88]. These results suggest the segments may function as a multipartite, cooperatively packaged entity rather than alone. This finding implies a bundling mechanism within the genome potentiated by segment- segment interactions.

Information regarding structural features within the packaging signals and across entire segments is lacking and may provide insights into how such seemingly minute changes in sequence may have dramatic effects on replicative capacity[89]. Exposed vRNA may fold into conserved structures allowing between-segment interactions[90]. Indeed vRNA is moderately susceptible to RNase cleavage despite RNP formation, suggesting at minimum dynamic protein-RNA interactions and potentially highly exposed, short RNA structures[89]. In support of this hypothesis complementary oligos targeting packaging regions are able to disrupt genome packaging and inhibit viral replication. Complex RNA structures may form within IAV vRNPs following co-transcriptional coating of RNA by nucleoprotein[84]. Consequently, the genome structure of IAV may be attributed to interactions between segments or an undiscovered RNA-binding protein that is able to oligomerize genome segments following export from the nucleus. One hypothesis suggests that the packaging signals may contain the bundling and recognition motifs themselves. In this case RNA-RNA interactions between segments would be required for efficient packaging and accounting within a newly forming genome. Electron micrographs of

both IAV and Influenza B virus depict large regions of electron density that appear to extend between genome segments[37]. More recently, studies have implicated non-canonical packaging elements as critical for determining the relative selection rates of competing genome segments. Additional studies show a single-interaction network between *in vitro* transcribed vRNAs[66,67,90]. Interestingly, the interactions identified by this method differ by viral strain[68]. Further experiments have established that genome packaging is dependent on the genome constellation present at the time of packaging[91,33,92,93]. Experimental coinfection *in vitro* or by classical reassortment following coinfection and antibody selection, has demonstrated co-segregation of genome segments. For instance, during vaccine creation the seasonal NA and PB1 are found together at high frequency[93]. Most importantly, the description of vRNA interactions within this reports lie outside of the canonical packaging signals necessary for efficient incorporation of reporter genomic segments[94]. These studies and our preliminary data described above provide the basis for the following hypothesis: Interactions within and between vRNA segments drive packaging in a competitive context and therefore likely contribute greatly to the potential for reassortment during coinfection.

1.11 Exploring viral and host determinants of reassortment potential

The generation and continued replication of genetically diverse IAV in a wide range of host species presents and ongoing challenge for human and animal health[10,26,95,96]. Within the following chapters, three independent but interconnected studies describing in part the genetic features of IAV that enable, as well as restrict, viral propagation and the rise of new IAV strains are described. In chapter 2, I present a study on the potential for reassortment of a recent North American avian H7N3 isolate with a panel of human and avian virus strains[91]. The consequences of single-gene reassortment on viral replication pathogenesis in a mammalian model are

discussed. Subsequently, in chapter 3, studies examining the reassortment potential of a H1N1 isolate demonstrate that the genetic context in which a gene segment resides may determine the fitness of a newly derived reassortant virus. In chapter 4, the interaction landscape of IAV NP and the viral genome are exported at nucleotide resolution. We find that the NP RNA scaffold likely allows formation of defined and required RNA structures that enable coordinated genome packaging. Finally, in chapter 5, the implications of our findings and recommendations for potential future studies are discussed in depth.

1.12 References

1. Shi, Y., Wu, Y., Zhang, W., Qi, J. & Gao, G. F. Enabling the 'host jump': structural determinants of receptor-binding specificity in influenza A viruses. *Nat. Rev. Microbiol.* **12,** 822–31 (2014).

2. Fukuyama, S. & Kawaoka, Y. The pathogenesis of influenza virus infections : the contributions of virus and host factors. *Curr. Opin. Immunol.* **23,** 481–486 (2011).

3. Heaton, N. S. *et al*. Long-term survival of influenza virus infected club cells drives immunopathology. *J. Exp. Med.* **211,** jem.20140488- (2014).

4. Herfst, S. *et al*. Airborne Transmission of Influenza. **1534,** (2012).

5. Ramos, I. & Fernandez-Sesma, A. Cell receptors for influenza a viruses and the innate immune response. *Front. Microbiol.* **3,** 117 (2012).

6. Sorrell, E. M. *et al*. Predicting 'airborne' influenza viruses: (trans-) mission impossible? *Curr. Opin. Virol.* **1,** 635–42 (2011).

7. Boon, A. C., Williams, R. W., Sinasac, D. S. & Webby, R. J. A novel genetic locus linked to pro-inflammatory cytokines after virulent H5N1 virus infection in mice. *BMC Genomics* **15,** 1017 (2014).

8. Lukarska, M. *et al*. Structural basis of an essential interaction between influenza polymerase and Pol II CTD. *Nature* **541,** 1–17 (2016).

9. Dugan, V. G. *et al*. The evolutionary genetics and emergence of avian influenza viruses in

wild birds. *PLoS Pathog*. **4,** e1000076 (2008).

10. Capua, I. & Munoz, O. Emergence of influenza viruses with zoonotic potential: Open issues which need to be addressed. A review. *Vet. Microbiol*. **165,** 7–12 (2013).

11. Sutton, T. C. *et al*. Airborne Transmission of Highly Pathogenic H7N1 Influenza Virus in Ferrets. *J. Virol*. **88,** 6623–6635 (2014).

12. Li, C. *et al*. Reassortment between avian H5N1 and human H3N2 influenza viruses creates hybrid viruses with substantial virulence. *Proc. Natl. Acad. Sci. U. S. A*. **107,** 4687–92 (2010).

13. Zhang, Y. *et al*. H5N1 Hybrid Viruses Bearing 2009/H1N1 Virus Genes Transmit in Guinea Pigs by Respiratory Droplet. *Science (80-.)*. **340,** 1459–1463 (2013).

14. Webster, R. G. Predicting the Next Influenza Virus. *Science (80-.)*. (2012).

15. Belser, J. A. *et al*. Pathogenesis and transmission of avian influenza A (H7N9) virus in ferrets and mice. *Nature* **501,** 556–9 (2013).

16. Pandemicity, H., Morens, D. M., Taubenberger, J. K. & Anthony, S. H7N9 Avian Influenza A Virus and the Perpetual Challenge of. (2013). doi:10.1128/mBio.00445-13.Editor

17. Belser, J. A. *et al*. Pathogenesis, Transmissibility, and Ocular Tropism of a Highly Pathogenic Avian Influenza A (H7N3) Virus Associated with Human Conjunctivitis. *J. Virol*. **87,** 5746–5754 (2013).

18. Lin, Y. P. *et al*. Population seroprevalence of antibody to influenza A(H7N9) virus,

Guangzhou, China. *BMC Infect. Dis.* **16,** 632 (2016).

19. Belser, J. a *et al.* Contemporary North American influenza H7 viruses possess human receptor specificity: Implications for virus transmissibility. *Proc. Natl. Acad. Sci. U. S. A.* **105,** 7558–63 (2008).

20. Kapczynski, D. R. *et al.* Characterization of the 2012 highly pathogenic avian influenza H7N3 virus isolated from poultry in an outbreak in Mexico: pathobiology and vaccine protection. *J. Virol.* **87,** 9086–96 (2013).

21. Rossman, J. S. & Lamb, R. a. Influenza virus assembly and budding. *Virology* **411,** 229–36 (2011).

22. Lakdawala, S. S. *et al.* Eurasian-origin gene segments contribute to the transmissibility, aerosol release, and morphology of the 2009 pandemic H1N1 influenza virus. *PLoS Pathog.* **7,** (2011).

23. Stincarelli, M. *et al.* Reassortment ability of the 2009 pandemic H1N1 influenza virus with circulating human and avian influenza viruses: public health risk implications. *Virus Res.* **175,** 151–4 (2013).

24. Greenbaum, B. D., Li, O. T. W., Poon, L. L. M., Levine, A. J. & Rabadan, R. Viral reassortment as an information exchange between viral segments. *Proc. Natl. Acad. Sci. U. S. A.* **109,** 3341–6 (2012).

25. Nelson, M. I. *et al.* Multiple reassortment events in the evolutionary history of H1N1 influenza A virus since 1918. *PLoS Pathog.* **4,** e1000012 (2008).

26. Romero-Tejeda, A. & Capua, I. Virus-specific factors associated with zoonotic and pandemic potential. *Influenza Other Respi. Viruses* **7 Suppl 2,** 4–14 (2013).

27. Watanabe, T. *et al*. Circulating avian influenza viruses closely related to the 1918 virus have pandemic potential. *Cell Host Microbe* **15,** 692–705 (2014).

28. Wu, A. *et al*. Sequential reassortments underlie diverse influenza H7N9 genotypes in China. *Cell Host Microbe* **14,** 446–452 (2013).

29. Xu, R. *et al*. Functional balance of the hemagglutinin and neuraminidase activities accompanies the emergence of the 2009 H1N1 influenza pandemic. *J. Virol*. **86,** 9221–32 (2012).

30. Ducatez, M. F. *et al*. Both influenza hemagglutinin and polymerase acidic genes are important for delayed pandemic 2009 H1N1 virus clearance in the ferret model. *Virology* **432,** 389–393 (2012).

31. Neumann, G., Noda, T. & Kawaoka, Y. Emergence and pandemic potential of swine-origin H1N1 influenza virus. *Nature* **459,** 931–9 (2009).

32. Smith, G. J. D. *et al*. Origins and evolutionary genomics of the 2009 swine-origin H1N1 influenza A epidemic. *Nature* **459,** 1122–5 (2009).

33. Kimble, J. B. *et al*. Alternative reassortment events leading to transmissible H9N1 influenza viruses in the ferret model. *J. Virol*. **88,** 66–71 (2014).

34. Kimble, J. B., Sorrell, E., Shao, H., Martin, P. L. & Perez, D. R. Compatibility of H9N2 avian influenza surface genes and 2009 pandemic H1N1 internal genes for transmission in

the ferret model. *Proc. Natl. Acad. Sci. U. S. A.* **108,** 12084–8 (2011).

35. Resa-Infante, P., Jorba, N., Coloma, R. & Ortin, J. The influenza virus RNA synthesis machine: Advances in its structure and function. *RNA Biol.* **8,** 207–215 (2011).

36. Dias, A. *et al.* The cap-snatching endonuclease of influenza virus polymerase resides in the PA subunit. *Nature* **458,** 914–8 (2009).

37. Nakatsu, S. Complete and Incomplete Genome Packaging of Influenza A and B Viruses. *MBio* **7,** 1–7 (2016).

38. Noda, T. *et al.* Three-dimensional analysis of ribonucleoprotein complexes in influenza A virus. *Nat. Commun.* **3,** 639 (2012).

39. Brooke, C. B. *et al.* Most Influenza A Virions Fail to Express At Least One Essential Viral Protein. *J. Virol.* (2013). doi:10.1128/JVI.02284-12

40. Huang, S. *et al.* A second CRM1-dependent nuclear export signal in the influenza A virus NS2 protein (NEP) contributes to the nuclear export of viral ribonucleoproteins. *J. Virol.* (2012). doi:10.1128/JVI.06519-11

41. Hutchinson, E. C. & Fodor, E. Transport of the influenza virus genome from nucleus to nucleus. *Viruses* **5,** 2424–46 (2013).

42. Stertz, S. & Shaw, M. L. Uncovering the global host cell requirements for influenza virus replication via RNAi screening. *Microbes Infect.* **13,** 516–25 (2011).

43. Imai, M. & Kawaoka, Y. The role of receptor binding specificity in interspecies transmission of influenza viruses. *Curr. Opin. Virol.* **2,** 160–7 (2012).

44. Rossman, J. S., Leser, G. P. & Lamb, R. a. Filamentous Influenza Virus Enters Cells via Macropinocytosis. *J. Virol.* **86,** 10950–60 (2012).

45. Roberts, P. C. & Compans, R. W. Host cell dependence of viral morphology. *Proc. Natl. Acad. Sci. U. S. A.* **95,** 5746–51 (1998).

46. de Vries, E. *et al*. Dissection of the influenza A virus endocytic routes reveals macropinocytosis as an alternative entry pathway. *PLoS Pathog.* **7,** e1001329 (2011).

47. de Vries, E. *et al*. Dissection of the influenza A virus endocytic routes reveals macropinocytosis as an alternative entry pathway. *PLoS Pathog.* **7,** e1001329 (2011).

48. Su, W.-C. *et al*. Pooled RNAi screen identifies ubiquitin ligase Itch as crucial for influenza A virus release from the endosome during virus entry. *Proc. Natl. Acad. Sci. U. S. A.* **110,** 17516–21 (2013).

49. Eisfeld, A. J., Neumann, G. & Kawaoka, Y. At the centre: influenza A virus ribonucleoproteins. *Nat. Rev. Microbiol.* **13,** 28–41 (2014).

50. Hudjetz, B. & Gabriel, G. Human-like PB2 627K influenza virus polymerase activity is regulated by importin-α1 and -α7. *PLoS Pathog.* **8,** e1002488 (2012).

51. Engelhardt, O., Smith, M. & Fodor, E. Association of the influenza A virus RNA-dependent RNA polymerase with cellular RNA polymerase II. *J. Virol.* **79,** 5812–5818 (2005).

52. Gu, W. *et al*. Influenza A virus preferentially snatches noncoding RNA caps. *RNA* **21,** 2067–75 (2015).

53. Landeras-Bueno, S., Jorba, N., Pérez-Cidoncha, M. & Ortín, J. The splicing factor proline-glutamine rich (SFPQ/PSF) is involved in influenza virus transcription. *PLoS Pathog*. **7,** (2011).

54. Pereira, C. F., Read, E. K. C., Wise, H. M., Amorim, M. J. & Digard, P. The influenza A virus NS1 protein promotes efficient nuclear export of unspliced viral M1 mRNA. *J. Virol*. JVI.00528-17 (2017). doi:10.1128/JVI.00528-17

55. Huang, X. *et al*. An NS-segment exonic splicing enhancer regulates influenza A virus replication in mammalian cells. *Nat. Commun*. **8,** 14751 (2017).

56. Perez, J. T. *et al*. A small-RNA enhancer of viral polymerase activity. *J. Virol*. **86,** 13475–85 (2012).

57. König, R. *et al*. Human host factors required for influenza virus replication. *Nature* **463,** 813–7 (2010).

58. Kawaguchi, A. & Nagata, K. De novo replication of the influenza virus RNA genome is regulated by DNA replicative helicase, MCM. *Embo J* **26,** 4566–4575 (2007).

59. Moeller, A., Kirchdoerfer, R. N., Potter, C. S., Carragher, B. & Wilson, I. a. Organization of the Influenza Virus Replication Machinery. *Science (80-.).* **338,** 1631–1634 (2012).

60. Elton, D. *et al*. Interaction of the Influenza Virus Nucleoprotein with the Cellular CRM1-Mediated Nuclear Export Pathway Interaction of the Influenza Virus Nucleoprotein with the Cellular CRM1-Mediated Nuclear Export Pathway. (2001). doi:10.1128/JVI.75.1.408

61. Chase, G. P. *et al*. Influenza virus ribonucleoprotein complexes gain preferential access to

cellular export machinery through chromatin targeting. *PLoS Pathog*. **7,** e1002187 (2011).

62. Chou, Y. *et al*. Colocalization of different influenza viral RNA segments in the cytoplasm before viral budding as shown by single-molecule sensitivity FISH analysis. *PLoS Pathog*. **9,** e1003358 (2013).

63. Nturibi, E., Bhagwat, A. R., Coburn, S., Myerburg, M. M. & Lakdawala, S. S. Intracellular Colocalization of Influenza Viral RNA and Rab11A is Dependent upon Microtubule Filaments. *J. Virol*. JVI.01179-17 (2017). doi:10.1128/JVI.01179-17

64. Lakdawala, S. S. *et al*. Influenza a virus assembly intermediates fuse in the cytoplasm. *PLoS Pathog*. **10,** e1003971 (2014).

65. Giese, S., Bolte, H. & Schwemmle, M. The feat of packaging eight unique genome segments. *Viruses* **8,** 1–11 (2016).

66. Fournier, E. *et al*. A supramolecular assembly formed by influenza A virus genomic RNA segments. *Nucleic Acids Res*. **40,** 2197–209 (2012).

67. Fournier, E. *et al*. Interaction network linking the human H3N2 influenza A virus genomic RNA segments. *Vaccine* **30,** 7359–67 (2012).

68. Essere, B. *et al*. Critical role of segment-specific packaging signals in genetic reassortment of influenza A viruses. *Proc. Natl. Acad. Sci. U. S. A.* **110,** E3840-8 (2013).

69. Arranz, R. *et al*. The Structure of Native Influenza Virion Ribonucleoproteins. *Science (80-.).* **338,** 1634–1637 (2012).

70. Moeller, A., Kirchdoerfer, R. N. & Potter, C. S. Arne Moeller, * Robert N. Kirchdoerfer,

* Clinton S. Potter, Bridget. **338,** 1631–1634 (2012).

71. Gallagher, J. R., Torian, U., McCraw, D. M. & Harris, A. K. Structural studies of influenza virus RNPs by electron microscopy indicate molecular contortions within NP supra-structures. *J. Struct. Biol.* (2016). doi:10.1016/j.jsb.2016.12.007

72. Baudin, F., Bach, C., Cusack, S. & Ruigrok, R. W. Structure of influenza virus RNP. I. Influenza virus nucleoprotein melts secondary structure in panhandle RNA and exposes the bases to the solvent. *EMBO J.* **13,** 3158–65 (1994).

73. Flick, R. & Hobom, G. Interaction of influenza virus polymerase with viral RNA in the 'corkscrew' conformation. *J. Gen. Virol.* **80 (Pt 10,** 2565–72 (1999).

74. Tomescu, A. I., Robb, N. C., Hengrung, N., Fodor, E. & Kapanidis, A. N. Single-molecule FRET reveals a corkscrew RNA structure for the polymerase-bound influenza virus promoter. *Proc. Natl. Acad. Sci. U. S. A.* 1–8 (2014). doi:10.1073/pnas.1406056111

75. Hutchinson, E. C. *et al.* Conserved and host-specific features of influenza virion architecture. *Nat Commun* **5,** 4816 (2014).

76. Calder, L. J., Wasilewski, S., Berriman, J. a & Rosenthal, P. B. Structural organization of a filamentous influenza A virus. *Proc. Natl. Acad. Sci. U. S. A.* **107,** 10685–90 (2010).

77. Labaronne, A. *et al.* Binding of RNA by the nucleoproteins of influenza viruses A and B. *Viruses* **8,** 1–14 (2016).

78. Liu, C. L. *et al.* Using mutagenesis to explore conserved residues in the RNA-binding groove of influenza A virus nucleoprotein for antiviral drug development. *Sci Rep* **6,**

21662 (2016).

79. York, A., Hengrung, N., Vreede, F. T., Huiskonen, J. T. & Fodor, E. Isolation and characterization of the positive-sense replicative intermediate of a negative-strand RNA virus. *Proc. Natl. Acad. Sci. U. S. A.* **110,** E4238-45 (2013).

80. Liang, Y., Hong, Y. & Parslow, T. G. cis -Acting Packaging Signals in the Influenza Virus PB1 , PB2 , and PA Genomic RNA Segments cis -Acting Packaging Signals in the Influenza Virus PB1 , PB2 , and PA Genomic RNA Segments. **79,** (2005).

81. Hutchinson, E. C., Kirchbach, J. C. Von, Gog, J. R. & Digard, P. Genome packaging in influenza A virus. 313–328 (2010). doi:10.1099/vir.0.017608-0

82. Marsh, G. a, Rabadán, R., Levine, A. J. & Palese, P. Highly conserved regions of influenza a virus polymerase gene segments are critical for efficient viral RNA packaging. *J. Virol.* **82,** 2295–304 (2008).

83. Gog, J. R. *et al.* Codon conservation in the influenza A virus genome defines RNA packaging signals. **35,** 1897–1907 (2007).

84. Giannecchini, S. *et al.* Packaging signals in the 5'-ends of influenza virus PA, PB1, and PB2 genes as potential targets to develop nucleic-acid based antiviral molecules. *Antiviral Res.* **92,** 64–72 (2011).

85. Duhaut, S. D. & Dimmock, N. J. Defective segment 1 RNAs that interfere with production of infectious influenza A virus require at least 150 nucleotides of 5' sequence: evidence from a plasmid-driven system. *J. Gen. Virol.* **83,** 403–11 (2002).

86. Goto, H., Muramoto, Y., Noda, T. & Kawaoka, Y. The genome-packaging signal of the influenza A virus genome comprises a genome incorporation signal and a genome-bundling signal. *J. Virol.* **87,** 11316–22 (2013).

87. Gerber, M., Isel, C., Moules, V. & Marquet, R. Selective packaging of the influenza A genome and consequences for genetic reassortment. *Trends Microbiol.* **22,** 446–455 (2014).

88. Hutchinson, E. C., Wise, H. M., Kudryavtseva, K., Curran, M. D. & Digard, P. Characterisation of influenza A viruses with mutations in segment 5 packaging signals. **27,** 6270–6275 (2009).

89. Lee, N. *et al.* Genome-wide analysis of influenza viral RNA and nucleoprotein association. *Nucleic Acids Res.* **5,** e1000491 (2017).

90. Gavazzi, C. *et al.* A functional sequence-specific interaction between influenza A virus genomic RNA segments. *Proc. Natl. Acad. Sci. U. S. A.* **110,** 16604–9 (2013).

91. Williams, G. D., Pinto, A. K., Doll, B. & Boon, A. C. M. A North American H7N3 influenza virus supports reassortment with 2009 pandemic H1N1 and induces disease in mice without prior adaptation. *J. Virol.* **90,** JVI.02761-15 (2016).

92. Ince, W. L., Gueye-Mbaye, A., Bennink, J. R. & Yewdell, J. W. Reassortment complements spontaneous mutation in influenza A virus NP and M1 genes to accelerate adaptation to a new host. *J. Virol.* **87,** 4330–8 (2013).

93. Cobbin, J. C. a *et al.* Influenza Virus PB1 and NA Gene Segments can Co-segregate during Vaccine Reassortment Driven by Interactions in the PB1 Coding Region. *J. Virol.*

(2014). doi:10.1128/JVI.01022-14

94. Gilbertson, B. *et al*. Influenza NA and PB1 gene segments interact during the formation of viral progeny: Localization of the binding region within the PB1 gene. *Viruses* **8,** 1–17 (2016).

95. Schrauwen, E. J. & Fouchier, R. A. Host adaptation and transmission of influenza A viruses in mammals. *Emerg. Microbes Infect.* **3,** e9 (2014).

96. Driskell, E. a *et al*. Low pathogenic avian influenza isolates from wild birds replicate and transmit via contact in ferrets without prior adaptation. *PLoS One* **7,** e38067 (2012).

Chapter 2:

A North American H7N3 influenza virus supports reassortment with 2009 pandemic H1N1 and induces disease in mice without prior adaptation

This chapter was published in full as a research article in the Journal of Virology, a Journal of the American Society of Microbiology on 2 March 2016.

2.1 Abstract

Reassortment between H5 or H9 subtype avian and mammalian influenza A viruses (IAV) can generate a novel virus that causes disease and transmits between mammals. This information is currently not available for H7 subtype viruses. We evaluated the ability of a low pathogenic North American avian H7N3 virus (A/shorebird/Delaware/22/2006) to reassort with mammalian or avian viruses using a plasmid-based competition assay. In addition to genome segments derived from an avian H7N9 virus, the H7N3 virus reassorted efficiently with the PB2, NA and M segments from the 2009 pandemic H1N1 ($_{P}$H1N1) virus. *In vitro* and *in vivo* evaluation of the H7N3:$_{P}$H1N1 7+1 reassortant viruses revealed that the PB2, NA, or M segments from $_{P}$H1N1 largely do not attenuate the H7N3 virus, whereas the PB1, PA, NP, or NS genome segments from $_{P}$H1N1 do. Additionally, we assessed the functionality of the H7N3:$_{P}$H1N1 7+1 reassortant viruses by measuring the inflammatory response *in vivo*. We found that infection with wild-type H7N3 resulted in increased inflammatory cytokine production relative to the $_{P}$H1N1, which was further exacerbated by substitution of $_{P}$H1N1 PB2 but not NA or M. Finally, we assessed if any adaptations occurred in the individually substituted segments after *in vivo* inoculation and found no mutations suggesting $_{P}$H1N1 PB2, NA, and M are genetically stable in background of this H7N3 virus. Taken together, we demonstrate that a North American avian H7N3 IAV is genetically and functionally compatible with multiple gene segments from the 2009 pandemic influenza virus strain without prior adaptation.

2.2 Importance

The 2009 pandemic H1N1 continues to circulate and reassort with other influenza viruses creating novel viruses with increased replication and transmission potential in humans. Previous studies have found that this virus can also reassort with H5N1 and H9N2 avian influenza viruses. We now show that several genome segments of the 2009 H1N1 virus are also highly compatible with a low pathogenic avian H7N3 virus and that these reassortant viruses are stable and not attenuated in an animal model. These results highlight the potential for reassortment of H1N1 viruses with avian influenza virus and emphasize the need for continued surveillance of influenza viruses in areas of co-circulation between avian, human and swine viruses.

2.3 Introduction

Reassortment of influenza A viruses (IAV) produces diversity and antigenic novelty within circulating strains, sometimes leading to the emergence of pandemic viruses that cause widespread disease in humans. Avian IAV subtypes, including H5, H7, and H9, have caused sporadic but sometimes fatal disease in humans [1,2]. Zoonosis of these viruses or derivative lineages formed via reassortment with strains capable of human-to-human transmission may lead to the emergence of novel viruses with pandemic potential [3]. Indeed multiple avian-origin viruses, most notably H5N1 and H9N2 strains were able to cause disease in mammals and had limited, but enhanced transmission potential following experimental reassortment with the 2009 pandemic H1N1 ($_{p}$H1N1) [2,4]. Genetically diverse IAV may therefore gain the ability to induce disease and transmit between mammals if an appropriate genetic constellation is assembled through reassortment.

H7-subtype viruses intermittently infect and cause disease in humans following contact with infected birds [1,5]. Outbreaks of H7 viruses in humans have occurred in geographically distinct areas including The Netherlands (2003), Canada (2004), Mexico (2012), and China (2013) [6]. Sporadic infections such as these, that sometimes result in severe disease, suggests that introduction of a H7-subtype virus capable of sustained transmission between humans has the potential to initiate a significant outbreak. Recently, a reassortant, low pathogenic H7N9 virus emerged in China with a case-fatality rate of approximately 25%, making this the most severe and sustained incursion of H7-subtype viruses into the human population [7]. Although person-to-person transmission has not been consistently demonstrated, the virus is capable of limited transmission in guinea pig and ferret models without prior adaptation, suggesting the potential

acquisition of transmission-related adaptations through mutation or reassortment [7–10]. The sustained and ongoing geographic co-circulation of H7-bearing viruses with the $_{P}$H1N1 poses a risk for reassortment that may create H7-bearing viruses containing one or more $_{P}$H1N1-origin gene segments [11].

Reassortment of IAV genes, including those endemic in animal reservoirs has given rise to pandemic IAVs, most recently $_{P}$H1N1 in 2009 [12]. In each case, antigenic shift of external proteins occurred as well as transfer of gene segments encoding internal and nonstructural proteins. $_{P}$H1N1 resulted from three independent reassortment events involving genes from swine, human, and avian viruses. This triple reassortant swine-origin IAV, first identified in Mexico, consists of PB2, PB1, PA, HA, NP, and NS derived from a North American swine virus isolate and the NA and M segment from an Eurasian lineage swine influenza virus [12]. Further reassortment events between $_{P}$H1N1 and swine IAV have resulted in the creation of variant H1N1v, H2N1v, and H3N2v viruses capable of causing disease in humans. All variant viruses contain segment 7 (M) of $_{P}$H1N1, which has been shown to confer aerosol transmissibility to previously non-transmissible viruses [13,14]. The widespread distribution of $_{P}$H1N1 coupled with its high rate of reassortment with environmental viruses suggests the potential for the emergence of novel $_{P}$H1N1 segment-containing IAV genome constellations with increased virulence or transmissibility in humans.

North American H7N3 IAVs co-circulate with $_{P}$H1N1; however, the ability of these avian viruses to reassort with $_{P}$H1N1, modify virologic traits, and induce disease in mammals is not known. This study was initially designed to identify genome segments derived from diverse avian- and human-origin IAV that were most compatible with an avian-origin LPAI (low pathogenic avian influenza) H7N3 virus. We devised a competitive reverse genetics strategy to

examine selection of a single genome segment when multiple versions of that segment are present in a cell, mimicking cellular co-infection. We found $_{p}$H1N1 PB2, NA, and M genome segments are capable of outcompeting the parental H7N3 strain and additional heterologous segments from diverse origins in competitive reverse genetic experiments. As a result of the competitive reverse genetic studies, we generated and characterized the parental H7N3 and $_{p}$H1N1 viruses, and seven H7N3:$_{p}$H1N1 7+1 reassortant viruses. Characterization of single segment reassortants demonstrated multiple genome segments from $_{p}$H1N1 either enhanced (PB2) or maintained (NA or M) H7N3 replication and virulence in a mouse model of infection. Finally, adaptation of substituted $_{p}$H1N1 segments was not required to function within the H7N3 backbone, suggesting these newly formed genomic constellations are genetically and functionally stable.

2.4 Materials and methods

Biosafety and Ethical Considerations

All experiments were carried out under enhanced BSL2 containment and approved by the Washington University in Saint Louis Institutional Biosafety Committee. The H7N3 virus A/shorebird/Delaware/22/2006 used in this study is neuraminidase inhibitor sensitive as are all the reassortant H7N3 viruses that were generated and used in this study. Additionally, the parental H7N3 virus has a genetic signature indicative of α-2,3-sialic acid (avian receptor) binding preference and does not bear a multibasic cleavage site, therefore we did not reasonably anticipate these experiments, in which this H7 HA was maintained, would result in increased host range or transmission to a new (mammalian) host. Finally, the NA gene of A/Memphis/03/2008 (H1N1) was excluded from this study due to a $H_{274}Y$ mutation that confers resistant to neuraminidase inhibitors (16).

Viruses

Viruses used in this study: A/shorebird/Delaware/22/2006 (H7N3), A/California/04/2009 ($_{P}$H1N1), A/mallard/Alberta/177/2004 (H7N9), A/Memphis/03/2008 ($_{S}$H1N1), and A/Puerto Rico/08/1934 ($_{PR8}$H1N1). For $_{S}$H1N1, H7N3, and H7N9, cDNA for all gene segments was cloned into the bi-directional pHW2000 plasmid and used to generate influenza viruses as previously described (17). Dr. Richard Webby at St. Jude Children's Research Hospital kindly provided the reverse genetic plasmids for $_{P}$H1N1 and $_{PR8}$H1N1. Plasmid-derived A/shorebird/Delaware/22/2006 (H7N3), A/California/04/2009 ($_{P}$H1N1), and all H7N3:$_{P}$H1N1 7+1 single reassortant viruses, were generated using the 293T-Madin-Darby canine kidney cell (MDCK) co-culture system and supernatant was injected into 10-day old embryonated chicken

eggs for 48 hours at 35°C (Cackle Hatcheries, IA, USA). Allantoic fluid containing the infectious virus was harvested and stored at -80°C. The viral titer (tissue culture infectious dose 50, $TCID_{50}$) was determined using MDCK cells and viral stock used in this study was titrated at least twice independently. Single reassortant viruses on the background of H7N3 possessing a single genome segment from $_{P}$H1N1 used in this study have been named with the following convention – seven segments from the H7N3 virus plus the substituted $_{P}$H1N1 segment in subscript (ex. $H7N3_{NA}$).

Cells

MDCK cells were maintained in Minimal essential medium (MEM) with 5% Fetal bovine serum (FBS), vitamins, L-glutamine (Invitrogen), penicillin, and streptomycin. 293T cells were maintained in Opti-MEM with 10% FBS, L-glutamine, penicillin and streptomycin. A549 cells were maintained in Dulbecco's Minimal essential medium (DMEM) with 10% FBS vitamins, L-glutamine, penicillin, streptomycin, 25mm HEPES, and non-essential amino acids (NEAA).

Competitive Reverse Genetic Assay

A competitive reverse genetic assay was developed to evaluate the ability of avian and human IAV genome segments to reassort in the context of an avian H7N3 viral background (Figure 1A). 293T-MDCK co-cultures were transfected with pHW2000 plasmids (1 μg per segment) containing seven genome segments of H7N3 virus plus four (for NA segment due to exclusion of NA from sH1N1) or five plasmids encoding a single genome segment (e.g. PB2) from different avian and human influenza viruses. All transfections included the wild-type H7N3 segment in addition to those from divergent strains. We did not evaluate the HA genome

segment because of biosafety considerations. The total amount of co-transfected plasmid DNA for the 8th segment (1 μg) was divided equally between the four or five studied viruses including the parental strain. The DNA was mixed at a 1:2 ratio with Trans-IT LT1 (Mirus) in Opti-MEM for 20 minutes at room temperature and added to the culture medium. Following an overnight incubation, the cell culture medium was removed and 1ml of fresh Opti-MEM supplemented with penicillin, streptomycin, L-glutamine was added. Twenty-four hours later an additional 1ml of Opti-Mem with 1μg/ml TPCK-trypsin (Worthington) was added. Forty-eight hours later supernatants were collected, cell debris removed by centrifugation at 1,200xg, and frozen at -80°C until further analysis. Clonal viral populations were isolated by limiting dilution assay on MDCK cells, then viral RNA extracted, reverse-transcribed using a vRNA-specific primer, and amplified with segment-specific PCR primers. Genome segment amplicons were genotyped by either restriction fragment length polymorphism analysis (RFLP) or Sanger sequencing. RFLP analysis was performed with one to four different restriction enzymes yielding a unique fragment length pattern for each segment and strain: PB2 (EcoRI, HindIII, BamHI); PB1 (HindIII); PA (BamHI, HindIII, XbaI); NP (BamHI, BglII); NA (BamHI, BglII, EcoRI, BsmBI); M (BamHI, HindIII, XcmI, PvuII). All restriction enzymes were obtained from New England Biolabs and used according to their instructions. The NS gene segment was genotyped by Sanger sequencing. Competitive reverse genetic assays were completed at least twice independently for each genome segment. On average we tested 48 clonal viruses per genome segment with a minimum of 24 (M segment) and maximum of 71 (PA segment) viruses.

Virus Genome Sequence Analysis

After expansion of virus in eggs, the nucleotide identity of the singly substituted genome segment in each H7N3:$_{P}$H1N1 7+1 single reassortant virus was verified by Sanger sequencing.

Viral genomic RNA was extracted, and reverse transcribed as described above. Segment-specific PCR was conducted and the resulting amplicon was purified by Agarose gel electrophoresis, and then submitted to for sequencing with an overlapping panel of sequencing primers derived from the parental $_{P}$H1N1 virus consensus sequence. We then assembled contiguous sequences for each segment and aligned them to the parental segment for analysis. The same sample preparation and analysis methods were used for materials derived from lung homogenate.

Multistep Growth Curves of Influenza A virus

MDCK or A549 cells ($2x10^5$) were seeded in 24-well plates and inoculated the next day with 50 $TCID_{50}$ (MDCK) or 10^5 $TCID_{50}$ (A549) of IAV. MDCK cells were washed once with PBS before adding the inoculum in MEM containing penicillin, streptomycin, L-glutamine, and vitamins plus 0.1% bovine serum albumin (M0.1B) for one hour at 37°C. After the one hour, the cells were washed once with PBS and 1.0 ml of M0.1B with 1μg/ml TPCK-trypsin was added to each well. A549 cells were washed once in DMEM containing penicillin, streptomycin, L-glutamine, and vitamins plus 0.1% bovine serum albumin and NEAA (D0.1B) prior to inoculation with virus diluted in D0.1B. After one hour at 37°C, the A549 cells were washed once with D0.1B and 1.0 ml of D0.1B with 0.5μg/ml TPCK-trypsin was added to each well. Culture supernatants from either cell type were collected at 24 and 48 hours post-infection (hpi) and the amount of infectious virus was quantified by titration on MDCK cells. The results are the average of two to three experiments, each performed in duplicate.

Phylogenetic analysis of IAV segments

Phylogenetic trees were generated for complete genome segments using the ClustalW algorithm in Lasergene MegAlign (version 11.1.0, DNASTAR, INC), bootstrapped 1000x, and used to constructed neighbor-joining phylogenetic trees.

$TCID_{50}$ assay

Confluent monolayers of MDCK cells were grown overnight in 96-well plates. The next day, the cells were washed with PBS and inoculated with ten-fold serial dilutions (10^{-1} to 10^{-8}) of culture supernatant, allantoic fluid or lung homogenate for one hour in M0.1B at 37°C and 5% CO_2. After one hour the inoculum was removed and replaced with M0.1B supplemented with 1μg/ml TPCK-trypsin and incubated for 72 hours. Presence of virus was determined by hemagglutination assay using 0.5% turkey red blood cells. $TCID_{50}$ was determined by the Reed-Muench method [15].

Intranasal inoculation of mice with influenza viruses

Six- to eight-week-old male C57BL/6J mice were bred in-house in a barrier facility at Washington University School of Medicine, St Louis, MO, USA. The mice received food and water *ad libitum* and all experiments were conducted in accordance with rules of the Institutional Animal Care and Use Committee. Mice were inoculated with 10^3 or 10^4 $TCID_{50}$ units of IAV intranasally in 30 μl of sterile PBS after sedation with Avertin (2,2,2-tribromoethanol, Sigma-Aldrich, MO, USA). Morbidity and mortality following intranasal inoculation were monitored for 7 days (10^4 $TCID_{50}$) or 14 days (10^3 $TCID_{50}$). To assess lung viral titers, mice inoculated with 10^4 $TCID_{50}$ were sacrificed on days 3 or 7 post-inoculation, the entire lung was collected, homogenized in 1.0 ml of infection media, cleared by centrifugation at 1000xg for 5 minutes and stored in aliquots at –80°C. Experiments in which 10^4 $TCID_{50}$ inoculum dose was used were

terminated at 7 days post-inoculation as all mice survived to the this point, at which viral titration and chemokine cytokine samples were harvested. Viral titers from lung homogenates were determined by $TCID_{50}$ assay. The results from the lung titrations are the average of at least two independent experiments.

Cytokine Array

Cytokine and chemokine production in lung homogenates was measured using a 23-plex cytokine array (Bio-Plex Pro™ Mouse Cytokine 23-plex Assay, Bio-Rad) according to manufacturer's protocol. The cytokine screen included IL-1α, IL-1β, IL-2, IL-3 IL-4, IL-5, IL-6, IL-9, IL-10, IL-12p40, IL-12p70, IL-13, IL-17, Eotaxin, G-CSF, GM-CSF, IFN-γ, KC, MCP-1 MIP-1α, MIP-1β, RANTES, and TNF-α. As many samples fell below the limit of detection we did not include the cytokines (IL-2, IL-3, IL-4, IL-9, and IL-17) in our analysis. Results from the cytokine array are the average of at least two independent experiments (Day 3: $H7N3_{WT}$,n=6; $H7N3_{PB2}$, n=5; $H7N3_{NA}$, n=6, $H7N3_{M}$, n=5; $_{P}H1N1_{WT}$, n=7; Day 7: $H7N3_{WT}$, n=5; $H7N3_{PB2}$, n=7; $H7N3_{NA}$, n=7; $H7N3_{M}$, n=4; $_{P}H1N1_{WT}$, n=10).

Influenza A virus mini-genome reporter assay

The PB2, PB1, PA and NP genes from H7N3 and $_{P}$H1N1 were cloned into the pcDNA3.1+ (Invitrogen) mammalian expression vector from the corresponding pHW2000 plasmid. The PA of pH1N1 containing the $P_{295}L$ mutation was cloned from the reassortant $H7N3_{PA}$ virus generated in this study. The pLuci plasmid was kindly provided by Dr. Yen (Hong Kong University, Hong Kong, China) and contains the firefly luciferase gene flanked by the non-coding regions of NP gene segment in the negative orientation under the control of a human RNA polymerase I promoter. Cells were maintained at 37°C for the duration of the experiment.

A Renilla luciferase expression plasmid was included for normalization. 293T cells were seeded into 24-well plates and transfected with PB2, PB1, PA and NP expression plasmids (83ng/plasmid) along with the two luciferase containing plasmids (total of 500 ng DNA) using TransIT LT1 per well. The following day the media was changed and cells were incubated for 48 hours, harvested and lysed for analysis of luciferase activities (Promega). Each combination of polymerase proteins (set of plasmids) was examined in duplicate and repeated independently in three separate experiments. The relative light units (RLU) of firefly luciferase activity were normalized to the RLU for Renilla luciferase activity within the same sample to account differences in transfection efficiency between wells and experiments. Polymerase activity was normalized to that of the parental H7N3 polymerase proteins.

Statistical analysis

All statistical analyses were performed using GraphPad Prism 6.0 software. Mann-Whitney U-Test was used to determine statistical significance between lung virus titers. One-Way ANOVA with Dunnett's Multiple Comparisons Test was used to determine statistical significance in cytokine and chemokine production in lung homogenates, with all comparisons made to $H7N3_{WT}$ for the indicated day. 2-Way ANOVA was used in the analysis of weight loss differences following influenza A virus infections in mice. Student's T-test was used to assess statistical differences between conditions in the mini-genome reporter assay with all comparisons made to the wild-type combination of H7N3 proteins.

2.5 Results

A North American H7 virus preferentially selects avian and 2009 pandemic H1N1 genome segments under competitive conditions.

Previous studies examining reassortment potential between H5- and H9- subtype avian IAV and human-derived IAV report high compatibility between diverse avian and human origin genome segments [2,4,16,17]. However, it remains unclear if H7-subtype viruses can undergo similar reassortment events with other avian and human IAV. The recent emergence of H7N9 in China demonstrates that H7-subtype viruses can infect humans and, if they acquire the ability to transmit between humans through reassortment, may pose a significant public health threat. Therefore, we sought to determine if North American H7-subtype viruses are capable of extensive reassortment with human or mammalian viruses.

We developed a competitive reverse genetics assay to evaluate the ability of heterologous genome segments from human and avian viruses to reassort efficiently with a North American low pathogenic H7N3 virus (**Figure 2.1 A**). Seven genome segments of A/shorebird/Delaware/22/2006 (H7N3) were maintained as a constant background while multiple versions of the 8th segment from H7N3, A/California/04/2009 ($_{P}$H1N1), A/Puerto Rico/08/1934 ($_{PR8}$H1N1), A/mallard/Alberta/144/2007 (H7N9), or A/Memphis/3/2008 ($_{S}$H1N1) were included in equimolar ratios. Phylogenetic trees were constructed for each genome segment to illustrate the genetic relationship to the H7N3 segment (**Figure 2.1 B**). Pools of infectious viruses were rescued, subjected to limiting dilution assays to culture clonal viral populations, and identify the segment of interest by via RFLP or Sanger sequencing.

A total of 334 viruses were analyzed for 7 genome segments in the context of H7N3 virus (HA was excluded). We observed a high frequency of isolated viruses containing avian H7N3 (29%) or H7N9 (34%) derived genome segments (**Figure 2.1 B**). In particular, the PB1, PA, NP and NS reassorted genome segments were predominantly (81%) of avian virus origin. In contrast, when we examined the identity of competitively selected segments the PB2, NA, and M genome segments were derived mostly from mammalian IAV, with a clear bias towards the $_{P}$H1N1 virus. The NA gene of $_{P}$H1N1 was identified in 100% (68/68) of the NA gene-reassorted H7 viruses, while the M and PB2 gene segment constituted 55% (13/24), and 35% (14/37) of the examined H7 viruses respectively. Combined, 32% of all isolated H7 viruses contained genome segments derived from $_{P}$H1N1, suggesting a high degree of compatibility between H7N3 and $_{P}$H1N1 virus. Genome segments from $_{PR8}$H1N1 were identified sporadically, 5% (15/334), among the H7N3 viruses. Of these, 8 contained the PB2 gene of $_{PR8}$H1N1. Finally, genome segments from $_{S}$H1N1 were never identified, 0% (0/266), among the H7 viruses, suggesting a low level of genetic compatibility between human H1N1 viruses circulating prior to 2009 and avian H7N3 virus.

Most H7N3:$_{P}$H1N1 7+1 viruses replicate efficiently in MDCK cells.

We observed a high degree of genetic compatibility between genome segments of $_{P}$H1N1 and H7N3 viruses in our competitive reverse genetics assay and sought to determine the replicative ability of each single reassortant virus bearing individual $_{P}$H1N1 segments in the background of H7N3. All 7+1 reassortant viruses were generated independently by reverse genetics and amplified in chicken eggs. We inoculated MDCK cells with 50 $TCID_{50}$ and collected supernatant at 24 and 48 hpi (**Figure 2.2 A and B**, respectively). H7N3 virus replicates to high titers at 24 ($10^{7.8}$/ml) and 48 hpi ($10^{7.9}$/ml). In contrast, $_{P}$H1N1 virus replicates to

significantly lower titers at 24 ($10^{5.8}$/ml) and 48 hpi ($10^{6.4}$/ml) ($P<0.01$ for both time-points). H7 viruses containing the NA ($H7N3_{NA}$) or M ($H7N3_{M}$) gene segment of $_{P}$H1N1 virus grow to similar titers ($P>0.2$) at 24 hpi ($10^{7.6}$/ml for $H7N3_{NA}$ and $10^{7.3}$/ml for $H7N3_{M}$) and 48 hpi ($10^{8.2}$/ml for $H7N3_{NA}$ and $10^{8.1}$/ml for $H7N3_{M}$). The titers of H7N3 reassortant viruses containing the PB2 ($H7N3_{PB2}$, $10^{6.8}$/ml), PA ($H7N3_{PA}$, $10^{5.9}$/ml), NP ($H7N3_{NP}$, $10^{5.7}$/ml) or NS ($H7N3_{NS}$, $10^{6.3}$/ml) gene segment of $_{P}$H1N1 were significantly ($P<0.001$) lower at 24 hpi, but not 48 hpi ($P>0.15$), compared to H7N3 parental virus. Finally, the virus titer of $H7N3_{PB1}$ was significantly ($P<0.001$) lower at 24 and 48 hpi ($10^{4.4}$/ml and$10^{4.6}$/ml) compared to parental H7N3. Overall, our results indicate that most H7N3 viruses possessing a single genome segment of $_{P}$H1N1 are able to replicate efficiently in a MDCK cell culture system and the NA and M gene of $_{P}$H1N1 did not attenuate H7N3 virus *in vitro*.

The majority of H7N3:$_{P}$H1N1 7+1 viruses are attenuated in A549 cells.

Next, we measured replication of the 7+1 reassortant viruses in a human lung epithelial cell line (A549). H7N3 virus replicated to $10^{3.5}$ $TCID_{50}$/ml at 24 hpi and $10^{4.0}$ $TCID_{50}$/ml at 48 hpi. $_{P}$H1N1 virus replicated to equivalent titers at 24 hpi ($10^{3.7}$ $TCID_{50}$/ml) and significantly higher titers at 48 hpi ($10^{5.2}$ $TCID_{50}$/ml, $P<0.05$) compared to H7N3 virus (**Figure 2.2 C and D**). H7N3 virus containing the M-segment of pH1N1 ($H7N3_{M}$) was able to grow in A549 cells, albeit less efficient compared to the wild type H7N3 virus at 48 hpi ($10^{3.5}$ $TCID_{50}$/ml, $P<0.05$). The 7+1 reassortant virus containing the PB2 genome segment of $_{P}$H1N1 replicated to equivalently titers at 24 hpi ($10^{3.7}$/ml), and higher titers at 48 hpi ($10^{4.7}$/ml, $P<0.05$), compared to H7N3 virus. Finally, we did not detect replicating virus from the $H7N3_{PB1,}$ $H7N3_{PA,}$ $H7N3_{NP,}$ $H7N3_{NA}$, or $H7N3_{NS}$ reassortant viruses in culture supernatant at either time point post-inoculation. The data

suggest that the majority of the gene segments of $_{P}$H1N1 attenuate H7N3 virus growth in human cells.

PB2, PA, and NP segments of $_{P}$H1N1 increase H7N3 polymerase activity in human cells.

We employed a mini-genome reporter assay to assess the polymerase activity of different viral polymerase complexes in human 293T cells. We found that the full complement of $_{P}$H1N1 proteins generated greater reporter activity than did the H7N3 polymerase proteins (1.8-fold increase, $P<0.01$, **Figure 2.3**). Substitution of the PB2 or NP protein with those from $_{P}$H1N1 increased polymerase activity by 1.8-fold ($P<0.01$) and 1.4-fold ($P<0.001$) respectively. The PA of $_{P}$H1N1 increased the polymerase activity more than 20-fold ($P<0.001$). We also evaluated a PA protein of $_{P}$H1N1 containing a Proline to Leucine substitution at position 295 ($P_{295}L$). This mutation was selected for in H7N3$_{PA}$ virus cultured in 10-day old embryonated chicken eggs. The polymerase activity of the mutant PA remained more than 10-fold ($P<0.001$, **Figure 2.3**) greater compared to H7N3 protein complex. Interestingly, substitution of $_{P}$H1N1 PB1 into a majority H7N3 polymerase complex resulted in a 25-fold reduction in polymerase activity ($P<0.001$). To examine if this was due to strain-dependent differences in PB1 protein expression we analyzed protein expression by flow cytometry (mean fluorescence intensity (MFI)) using His-tagged versions of the PB1 of H7N3 and $_{P}$H1N1 virus. We did not observe a difference in MFI between these two proteins (data not shown), suggesting that the 25-fold difference in reporter activity is due to inherent differences in compatibility between the polymerase proteins. Overall, the H7N3 virus appears well adapted to mammalian cells and the introduction of the PB1 of PH1N1 severely attenuates the H7N3 virus.

H7N3:$_{P}$H1N1 7+1 reassortant viruses induce distinct morbidities *in vivo*.

Given the high degree of functional compatibility between $_{P}$H1N1 and H7N3 virus *in vitro*, we next determined the relative fitness of these H7N3 viruses containing individual gene-segments of $_{P}$H1N1 *in vivo*. C57BL/6J mice were inoculated intranasally with either parental (H7N3 or $_{P}$H1N1) or single segment reassortant viruses and weighed at regular intervals for one week (10^4 dose) or two weeks (10^3 dose) depending on the inoculum size. Mice inoculated with 10^4 $TCID_{50}$ of H7N3 demonstrated substantial weight loss beginning at 3 dpi (10%) that continued until the experiment was ended at 7 dpi (21%). Inoculation with 10^4 $TCID_{50}$ of $_{P}$H1N1 induced up to 30% weight loss relative to the starting weight within 7 days (**Figure 2.4 A**) demonstrating that the $_{P}$H1N1 is more pathogenic in mice compared to the avian H7N3 virus. H7N3 viruses containing single gene segments for $_{P}$H1N1 varied in their ability to cause weight loss. Inoculation with 10^3 or 10^4 $TCID_{50}$ of H7N3 virus containing the PB2 segment of $_{P}$H1N1 induced significantly ($P<0.05$) more weight loss compared to H7N3, including rapid weight loss (>20%) within the first three days following inoculation (**Figure 2.4 A**). Importantly, the $H7N3_{PB2}$ virus was significantly ($P<0.005$, days 8 and 10; $P<0.01$, days 12, and 14) less virulent compared to the $_{P}$H1N1 virus at the lower inoculum (**Figure 2.4 D**). Inoculation with 10^4 $H7N3_{NA}$ or $H7N3_{M}$ resulted in equivalent weight loss at all times compared to H7N3 (**Figure 2.4 B**, $P>0.1$). Reassortant H7N3 viruses possessing the PB1, PA, NP, or NS of $_{P}$H1N1 did not induce appreciable weight loss at any day post inoculation, suggesting that these gene-segments from $_{P}$H1N1 attenuate H7N3 virulence *in vivo*.

H7N3 reassortant viruses possessing $_{P}$H1N1 PB2, NA, or M do not attenuate replication *in vivo*.

To further evaluate the compatibility between H7N3 and $_{P}$H1N1 viruses we measured virus titer in lung tissue 3 and 7 dpi with 10^4 $TCID_{50}$ of all parental and single gene reassortants.

Lung viral titers 3 days after $_{P}$H1N1 virus inoculation ($10^{6.4}$/ml) were significantly higher (P<0.005) compared to H7N3 virus ($10^{5.9}$/ml, **Figure 2.5 A**). However, this difference in virus load disappeared by day 7; H7N3 ($10^{5.3}$/ml) and $_{P}$H1N1 ($10^{5.7}$/ml) (P>0.15, **Figure 2.5 B**). The virus load in lungs of H7N3$_{PB2}$ infected animals was significantly higher ($10^{6.5}$/ml, P<0.05) at 3 dpi, but not 7 dpi ($10^{5.0}$/ml, P>0.15), compared to H7N3 virus. Importantly, the viral load in H7N3$_{PB2}$ infected lungs was similar to that of $_{P}$H1N1 virus at day 3. Inoculation with H7N3$_{NA}$ or H7N3$_{M}$ resulted in similar lung virus titer compared to H7N3 at both 3 ($10^{5.7}$/ml for H7N3$_{NA}$ and $10^{5.5}$/ml for H7N3$_{M}$) and 7 dpi ($10^{5.2}$/ml for H7N3$_{NA}$ and $10^{4.3}$/ml for H7N3$_{M}$), (P>0.05 for all comparisons). H7N3 reassortant viruses bearing PA or NS gene segment of $_{P}$H1N1 replicated *in vivo*, however the virus titers were significantly lower at 3 ($10^{4.0}$/ml for H7N3$_{PA}$ and$10^{2.2}$/ml for H7N3$_{NS}$, P<0.005 and P<0.01, respectively) and 7 dpi ($10^{2.0}$/ml for H7N3$_{PA}$ and$10^{2.0}$/ml for H7N3$_{NS}$, P<0.005 and P<0.005, respectively) compared to H7N3 virus (**Figure 2.5 A and B**). Finally, H7N3$_{NP}$ and H7N3$_{PB1}$ viruses were never detected at 3 or 7 dpi. These results suggest that the PB2, NA and M genes of $_{P}$H1N1 are genetically and functionally compatible with the remaining segments of an avian H7 IAV.

$_{P}$H1N1 PB2, NA, or M show no signs of genetic adaptation in the context of a predominantly H7N3 genome constellation.

We analyzed the nucleotide sequence of substituted $_{P}$H1N1 gene segments in the context of the H7N3:$_{P}$H1N1 7+1 viruses after passage in eggs and replication in mice. Six of the seven $_{P}$H1N1 gene-segments contained no nucleotide changes upon culture in eggs. A single point mutation was found in the PA gene of $_{P}$H1N1 ($P_{295}L$) in H7N3$_{PA}$ grown 48 hours in eggs. This same mutation was identified a second time following an independent attempt to rescue virus with genetic sequence identical to the original $_{P}$H1N1 PA gene. To analyze genetic stability *in*

vivo, we isolated RNA from lungs of mice infected 7 days with H7N3$_{PB2}$ (n = 3), H7N3$_{NA}$ (n = 3), and H7N3$_{M}$ (n = 3). No nucleotide changes were found in any of the PB2, NA or M genes of $_{P}$H1N1 in the context of the H7N3 virus. Taken together, these data suggest that the $_{P}$H1N1 gene segments are genetically stable and highly functional in the context of an avian H7N3 virus.

H7N3 and reassortant viruses that contain $_{P}$H1N1 PB2, NA, or M viruses induce robust inflammatory cytokine and chemokine responses *in vivo*.

The virulence and fitness of an influenza virus in the mouse model is often associated with elevated production of pro-inflammatory cytokines and chemokines [18–20]. To further evaluate the functional compatibility between $_{P}$H1N1 and H7N3 genes we quantified the inflammatory response in lung tissue 3 and 7 days post inoculation with the three most virulent H7N3:$_{P}$H1N1 7+1 reassortant viruses (H7N3$_{PB2}$, H7N3$_{NA}$, and H7N3$_{M}$) and compared the response to that of parental H7N3 and $_{P}$H1N1 virus. The parental H7N3 virus induced significantly (P<0.05) higher concentrations of inflammatory mediators, such as IL-12(p40) early following inoculation (day 3) relative to $_{P}$H1N1 (**Table 2.1**). Infection with H7N3$_{PB2}$ induced higher concentrations of key inflammatory mediators including CCL2, CCL3, and CCL5 relative to H7N3 (P<0.05). H7N3$_{PA}$ replication was attenuated *in vivo* and this virus correspondingly induced significantly lower levels of many inflammatory cytokines relative to H7N3 (P<0.05). By 7 days post infection, $_{P}$H1N1 infected lungs had higher concentrations of cytokines relative to parental H7N3 while H7N3$_{PB2}$ produced even more exacerbated responses in many mediators, including IL-10, CCL2, CCL3, CCL5, and TNF-α (P<0.05-P<0.001 relative to H7N3, **Table 2.2**). H7N3$_{NA}$ and H7N3$_{M}$ elicited inflammatory responses that more closely resemble the profile of H7N3 than $_{P}$H1N1 at both time points following infection.

Table 2.1. Inflammatory cytokine and chemokine levels in total lung homogenates three days after intranasal inoculation with parental H7N3, $_{P}$H1N1, and reassortant viruses.

	Concentration (pg/ml) (mean ± S.E.M.) in mouse lung				
	Virus				
Cytokine	**H7N3$_{WT}$**	**H7N3$_{PB2}$**	**H7N3$_{NA}$**	**H7N3$_{M}$**	**$_{P}$H1N1$_{WT}$**
IL-1α	267 ± 5	32 ± 4	23 ± 6	33 ± 7	12 ± 5*
IL-1β	360 ±	436 ± 57	494 ± 73	498 ± 66	169 ± 22
IL-5	294 ± 95	529 ± 56	423 ± 45	356 ± 58	29 ± 13*
IL-6	228 ± 69	363 ± 78	451 ± 49*	288 ± 57	70 ± 30
IL-10	21 ± 3	20 ± 2	16± 2	20 ± 2	13 ± 4
IL-12p40	253 ± 63	421 ± 65	244 ± 36	383 ± 51	85 ± 24*
IL-12p70	100 ± 27	48 ± 7	44 ± 8	55 ± 12	63 ± 36
IL-13	52 ± 3	64 ± 5	45 ± 5$^{\#}$	48 ±5	26 ± 3$^{\&}$
Eotaxin	462 ± 97	613 ± 24	732 ± 60*	570 ± 107	175 ± 25
G-CSF	268 ± 78	579 ± 113*	336 ± 92	263 ± 22	249 ± 91
GM-CSF	67 ± 11	65 ± 5	57 ± 6	50 ± 6	46 ± 13
Interferon-γ	9 ± 1	20 ± 3$^{\&}$	7 ± 1	9 ± 3	5 ± 1*
KC	370 ± 88	435 ± 119	456 ± 95	533 ± 52	151 ± 56
CCL2	586 ±	3145 ±	1115 ± 84$^{\&}$	758 ± 99	835 ± 292
CCL3	52 ± 16	125 ± 18*	77 ± 24	65 ± 21	22 ± 8
CCL4	72 ± 20	128 ± 34	84 ± 17	68 ± 6	44 ± 8
CCL5	89 ± 13	148 ± 27*	125 ± 25	105 ± 31	66 ± 24
TNF-α	172 ± 49	123 ± 12	126 ± 21	100 ± 9	262 ± 52

(*, $P < 0.05$; #, $P < 0.01$; &, $P < 0.005$, compared to H7N3 values).

Table 2.2. Inflammatory cytokine and chemokine levels in total lung homogenates seven days after intranasal inoculation with parental H7N3, $_{P}$H1N1, and reassortant viruses.

	Concentration (pg/ml) (mean ± S.E.M.) in mouse lung				
	Virus				
Cytokine	**H7N3$_{WT}$**	**H7N3$_{PB2}$**	**H7N3$_{NA}$**	**H7N3$_{M}$**	**$_{P}$H1N1$_{WT}$**
IL-1α	13 ± 1	48 ± 4$^{\&}$	22 ± 8	32 ± 2$^{\&}$	31 ± 2$^{\&}$
IL-1β	313 ± 80	669 ± 77	337 ± 99	840 ± 51*	519 ± 112
IL-5	131 ± 35	120 ± 23	313 ± 95	178 ± 36	235 ± 35
IL-6	198 ± 39	148 ± 35	355 ± 105	183 ± 45	790 ± 85$^{\&}$
IL-10	62 ± 6	247 ± 63*	76 ± 27	109 ± 16	123 ± 12$^{\#}$
IL-12p40	328 ± 78	462 ± 66	218 ± 52	713 ± 122*	313 ± 39
IL-12p70	44 ± 9	104 ± 24*	63 ± 10	37 ± 3	82 ± 14*
IL-13	92 ± 14	248 ± 19$^{\&}$	92 ± 16	193 ± 10$^{\&}$	123 ± 44
Eotaxin	1678 ±	465 ± 89	464 ± 94	432 ± 88	393 ± 55
G-CSF	288 ± 53	289 ± 87	333 ± 127	490 ± 141	1362 ± 105$^{\&}$
GM-CSF	46 ± 6	80 ± 7$^{\#}$	37 ± 3	105 ± 4$^{\&}$	152 ± 85
Interferon-γ	46 ± 12	63 ± 27	24 ± 18	51 ± 1	276 ± 96*
KC	196 ± 65	351 ± 54	284 ± 130	153 ± 59	338 ± 80
CCL2	750 ± 25	2580 ±610*	402 ± 198	201 ± 114*	2273 ± 288$^{\&}$
CCL3	172 ± 18	2544 ± 610$^{\#}$	96 ± 35	737 ± 243	692 ± 171*
CCL4	196 ± 41	156 ± 35	99 ± 63	69 ± 13*	110 ± 20
CCL5	692 ± 78	1672 ±333*	271 ± 193	593 ± 116	294 ± 65$^{\#}$
TNF-α	161 ± 27	478 ± 99*	221 ± 44	167	307 ± 35$^{\#}$

(*, $P < 0.05$; #, $P < 0.01$; &, $P < 0.005$, compared to H7N3 values).

2.6 Discussion

The ongoing reassortment of avian and human IAV threatens human populations worldwide as evidenced by numerous recent zoonotic events [21]. Along with H5- and- H9 subtypes, H7-subtype viruses are thought to be candidates for generating a pandemic IAV. We sought to determine the potential for reassortment between a low-pathogenic North American avian H7N3 virus and diverse avian and mammalian virus isolates. Our studies identified a clear bias towards reassortment of H7N3 with gene-segments derived from the 2009 pandemic H1N1 ($_{P}$H1N1) virus, but not other mammalian virus isolates. We found that single reassortant H7N3 viruses containing PB2, NA, or M segments of $_{P}$H1N1 do not attenuate the H7N3 virus *in vitro* or *in vivo*. Our results highlight the potential for environmental reassortment that can lead to antigenically novel viruses capable of inducing disease in mammals and may inform future surveillance efforts. Moreover, these findings support the potential for emergence of future pre-pandemic viruses in avian populations without prior mammalian adaptation.

We used a novel plasmid-based genetic screen to assess the compatibility of human and avian genome segments in the context of a low-pathogenic North American H7N3 virus. We found that $_{P}$H1N1 genome segments were selected most frequently among mammalian derived segments, suggesting a high degree of functional compatibility of these segments with the H7N3 virus when compared to pre-pandemic human isolates. We found no instances of seasonal H1N1 ($_{S}$H1N1) gene segments reassorting with the avian virus and only seldom observed inclusion of genes from the $_{PR8}$H1N1 strain. This striking difference suggests that the current H1N1 virus ($_{P}$H1N1) is more likely to reassort with other viruses and create novel pandemic viruses. While the mechanism underlying this preference is unknown, it is plausible that the extended adaptation

of gene-segments of $_{S}$H1N1 to the human host and other segments of $_{S}$H1N1 has decreased its ability to reassort with avian or other human viruses. In contrast, gene-segments of $_{P}$H1N1 are found in many different virus isolates validating the ability to reassort rapidly with other strains of influenza virus.

We observed a high frequency of H7N3 and H7N9 segment selection in instances when these segments were closely related (**Figure 2.1 B**). However, in not all instances did high nucleotide similarity to the parental H7N3 segment result in high rates of selection, for example the M and NA segments of $_{P}$H1N1 virus. This discrepancy perhaps indicates, that within this system, compatibility between proteins, not genome segments, drives viral replication and therefore selection.

Historically, reassortment of influenza viruses was studied using co-infection models. These experiments provided important information about the natural evolution of IAV; however, they are less controlled and potentially more hazardous if one or both IAVs have previously been shown to transmit between mammals. We provide an alternative method to study influenza reassortment and argue that these experiments are not only safer, but also address a different question and that is; what are the molecular requirements of a single gene-segment to compete with and reassort into an existing virus genome? Knowing the key features in a genome segment that promotes or inhibits reassortment will result in a better understanding of the process as a whole and therefore improve our predictions on the outcome of experimental and natural co-infections.

These studies identified two pandemic IAV gene segments, NA and M, which in the context of H7N3 virus do not impact viral replication in MDCK cells or pathogenesis in mice.

The compatibility of these genes with H7N3 is of interest given their prior association with increased transmission of the 2009 pandemic H1N1 when compared with putative pandemic progenitors [22]. Multiple independent reassortment events between human and swine influenza viruses resulted in the creation of novel genome constellations (H1N1v, H2N1v, and H3N2v) capable of aerosol transmission from pig to human due to acquisition of $_{P}$H1N1 M. Previous investigations of potential reassortant viruses between a highly pathogenic avian H5N1 influenza virus and $_{P}$H1N1 demonstrated high compatibility between H5N1 and the NA and M segments of $_{P}$H1N1 [23]. The resulting reassortant viruses similarly did not have altered *in vitro* replication relative to wild-type H5N1 virus. To further experimentally corroborate the importance of this segment in $_{P}$H1N1 transmission, multiple studies engineered reassortant viruses that either gained or lost the ability to transmit between mammals following manipulation of M [22,24]. In combination with M, NA was capable of enhancing the transmission of $_{P}$H1N1 through altered receptor destroying activity (NA) and filamentous particle morphology (M) that changes virus release from the respiratory tract of infected animals compared to diverse contemporary swine isolates. Although our studies did not examine virion morphology or respiratory shedding, additional studies in a model more suitable for examining such phenotypes may provide insights into the activity of these genes in the genetic background of an avian influenza virus.

The potential for transmission of H7-bearing viruses has been demonstrated experimentally [25]. Since other North American H7-HA are able to bind both avian and mammalian receptors, additional adaptations or acquisition of other gene segments via reassortment may be critical to achieve sustained mammalian transmission [26–28]. Other avian viral subtypes (i.e. H9N2 and H5N1) can reassort with $_{P}$H1N1 and achieve contact or respiratory droplet transmission [2,4,16]. Additionally, reassortant viruses possessing $_{P}$H1N1 M are capable of

directly infecting humans from pigs [29]. Our studies show H7N3 viruses possessing PB2, NA, or M from $_{P}$H1N1 are not attenuated *in vivo* and induce disease comparable to wild type H7N3 without adaptation. Further studies of these viruses and other similar single reassortant avian viruses in mammalian transmission models are required to understand the constellation of reassortants that are potential sources of new zoonoses.

PB2 is the only $_{P}$H1N1 gene segment that increased the virulence and replication *in vivo* relative to wild type H7N3 despite attenuation *in vitro*. Previous studies have identified several amino-acid substitutions that are associated with efficient replication in mammalian cells, including $G_{590}S$, $Q_{591}R$, $E_{627}K$, and $D_{701}N$ [30–32]. The PB2 of H7N3 virus contains all avian residues, while the PB2 of $_{P}$H1N1 has the $G_{590}S$ and $Q_{591}R$ substitution. Based on this information, we anticipated that the PB2 of $_{P}$H1N1 would increase replication of the H7N3 7+1 virus in mammalian cells. As increased viral replication may exacerbate initial inflammatory responses and lead to subsequent morbidity [18,20], we assessed the abundance of cytokines and chemokines present in whole lung homogenate at two points following infection (**Tables 2.1 and 2.2**). Mammalian adapted $_{P}$H1N1 PB2 increased induction of pro-inflammatory cytokines IL-6, CCL2 and CCL3 early during infection (day 3, **Table 2.1**). CCL2 levels also correlated well with morbidity when assessed later during infection at 7 dpi, likely due to increased viral replication early following inoculation (**Table 2.2**). Importantly, while morbidity and viral titer were increased relative to parental H7N3, substitution of $_{P}$H1N1 PB2 did not induce morbidity equivalent to parental $_{P}$H1N1 at low inoculum dose.

While most H7N3 7+1 reassortant viruses did not contain mutations in the $_{P}$H1N1 gene-segment, we repeatedly recovered $H7N3_{PA}$ virus containing an amino acid substitution at position 295 ($P_{295}L$). In a unique isolate from the 2009 pandemic – A/Tennessee/560/09 (H1N1)- mouse

adaptation introduced the inverse mutation PA $L_{295}P$ [33]. Upon further examination, creation of a virus with only this mutation led to increased viral replication in normal human bronchial epithelial (NHBE) cells. In a different study, the same group described $L_{295}P$ as contributing to increased and prolonged viral shedding and disease in both donor and contact ferrets[34]. Additionally, the same polymerase assay that we utilized, A/Tennessee/560/09 PA $L_{295}P$ increased reporter activity in the context of other A/Tennessee/560/09 proteins and the co-expressed level of PB2 but not of NP [33]. The importance of this residue had not to our knowledge been examined in the context of A/California/04/09 PA before this report, though it may be surmised that the $P_{295}L$ mutation is detrimental to the PA function in the mammalian host and perhaps was selected for due to interaction with the avian host during growth in chicken eggs or with a component of the H7N3 polymerase complex.

In the context of the H7N3 polymerase complex, this mutation reduced the polymerase activity 2-fold (**Figure 2.3**). At this time we cannot address if the observed differences in polymerase activity are due to altered protein half-life, expression level, or inherent enzymatic activity of $_{p}$H1N1 PA in either H7N3 or $_{p}$H1N1 complexes. It is not known if this mutation affected the virulence of $H7N3_{PA}$ in mice, but based on published data we expect that an H7N3 virus containing the wild-type PA of $_{p}$H1N1 is even more attenuated.

Taking the data together, we establish that the 2009 pandemic H1N1 has a high degree of compatibility with a North American LPAI H7N3 virus. Further, the genome segment reassortant viruses containing the $_{p}$H1N1 segments most frequently selected (PB2, NA, or M) establish productive infection in the mouse respiratory tract, induce morbidity and disease, and persist for at least one week, potentially allowing dissemination to others by contact or respiratory transmission. Future studies may examine the potential for transmission of these H7 viruses

containing single genes from 2009 pandemic H1N1. These studies emphasize the need for continued surveillance of the avian IAV reservoir and the critical importance of identifying reassortant viruses that contain genomic signatures associated with mammalian disease and transmission.

2.7 Acknowledgements

G.D.W. was supported in part by training grant GM: 007067, the Victoria J. Fraser, M.D. Fellowship for Graduate Studies in Infectious Diseases, and the Division of Biology and Biomedical Sciences at Washington University School of Medicine. Finally, we are thankful for the technical assistance of Traci Bricker.

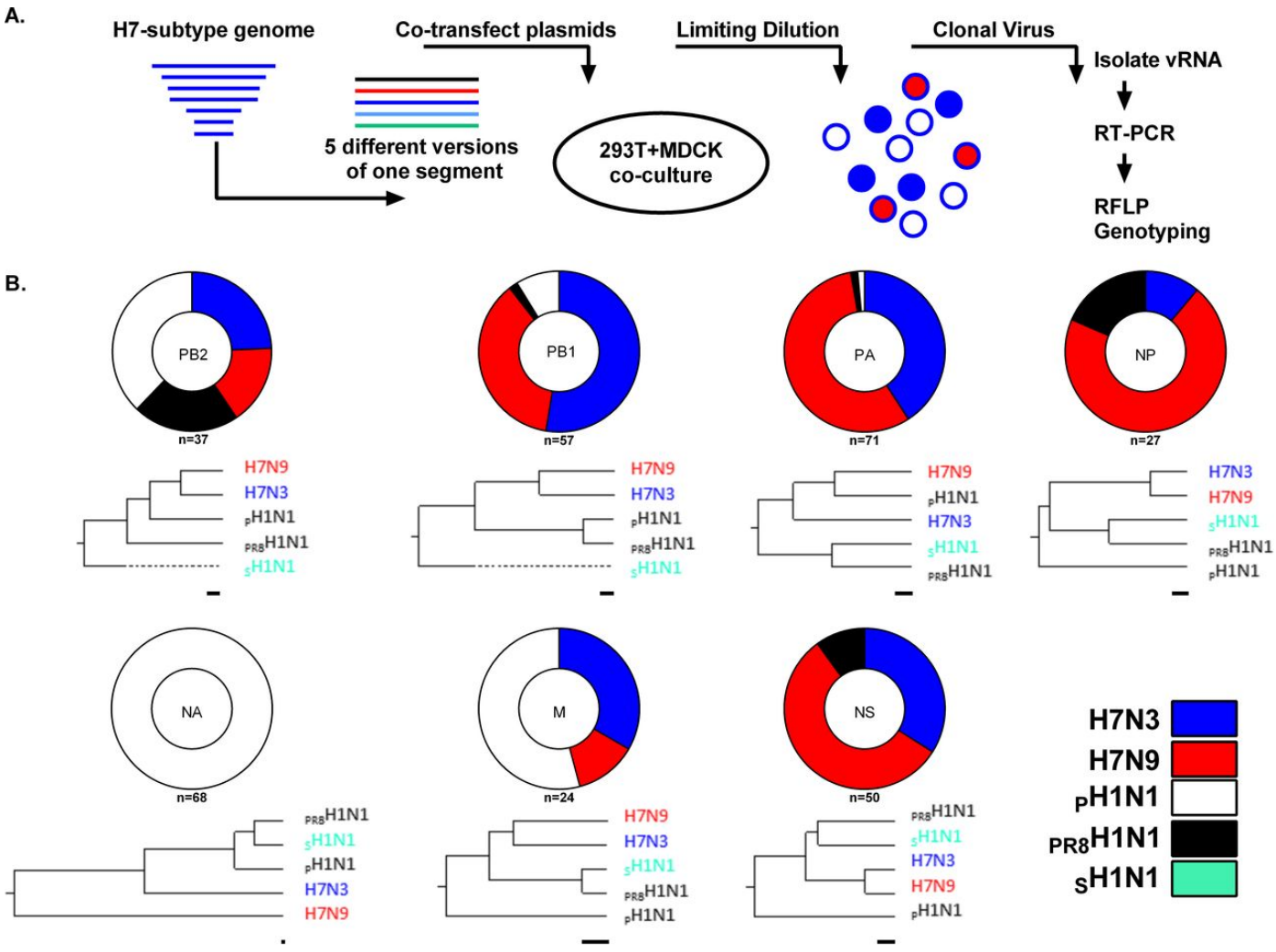

A.
H7-subtype genome
Co-transfect plasmids
Limiting Dilution
Clonal Virus
Isolate vRNA
RT-PCR
RFLP Genotyping
5 different versions of one segment
293T+MDCK co-culture
B.
PB2
n=37
PB1
n=57
PA
n=71
NP
n=27
NA
n=68
M
n=24
NS
n=50
H7N3
H7N9
pH1N1
PR8H1N1
sH1N1

Figure 2.1. A North American H7N3 virus preferentially selects avian and 2009 pandemic H1N1 genome segments following competitive transfection.

A. Seven genome segments derived from A/shorebird/Delaware/22/2006 (H7N3) were co-transfected with four (NA only) or five variants of the eighth genome segment. These variants included the homologous gene-segment from A/shorebird/Delaware/22/2006 (H7N3) as well as that of A/mallard/Alberta/177/2004 (H7N9), A/Memphis/03/2008 ($_{S}$H1N1), A/California/04/2009 ($_{P}$H1N1), and A/Puerto Rico/08/1934 ($_{PR8}$H1N1) virus. Following transfection, individual virus particles were cultured in MDCK cells using a limiting dilution assay. Virus positive wells were identified by HA assay and the remaining supernatant was used to purify viral RNA. cDNA was generated using a universal influenza primer and the segment of interest was amplified using segment-specific PCR primers and analyzed by restriction length fragment polymorphism (RFLP). **B.** Pie charts depicting the relative distribution of four or five gene-segments that were identified among the competitive reverse genetics derived H7 viruses. The results are the cumulative distribution of two or more independently repeated experiments; 334 viruses in total sampled. The NA gene of A/Memphis/03/2008 (NA-inhibitor insensitive) was excluded from these analyses. Phylogenetic trees for each corresponding segment illustrating the genetic relationship to between segments assayed (bootstrapped 1000x). The line beneath each phylogenetic tree represents 1 nucleotide substitution per 100 nucleotides.

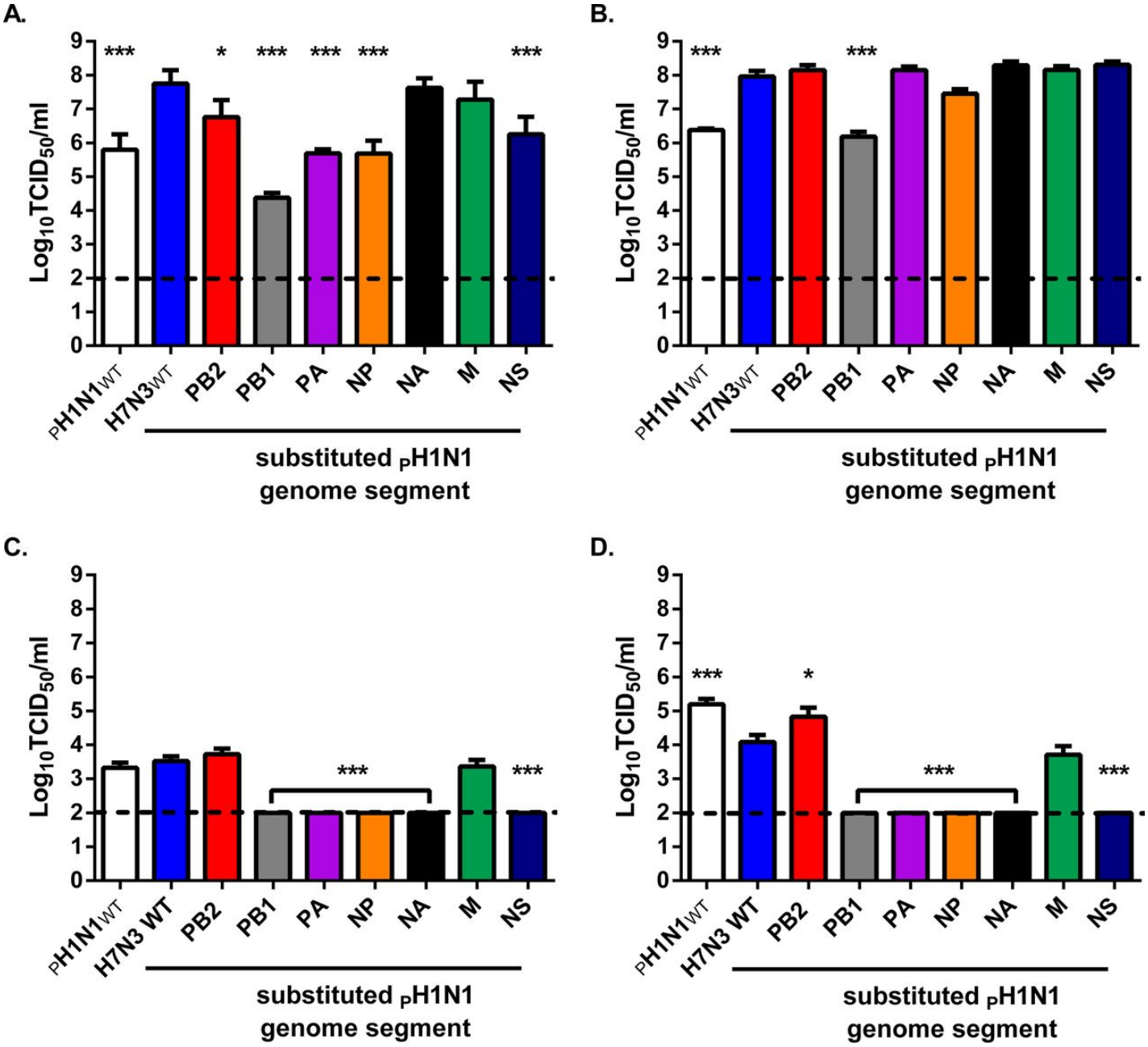

A.
B.
C.
D.
Log10TCID50/ml
pH1N1WT
H7N3WT
H7N3 WT
PB2
PB1
PA
NP
NA
M
NS
substituted pH1N1
genome segment
***
*

Figure 2.2. Several H7-single gene reassortant viruses possessing a $_{P}$H1N1 segment replicate to efficiently in mammalian tissue culture.

MDCK cells were inoculated with 50 $TCID_{50}$ of the indicated virus and virus titer ($TCID_{50}$/ml) in the supernatant was determined at 24 (**A**) or 48 (**B**) hours post- inoculation by virus titration assay on MDCK cells. A549 cells were inoculated with 10^5 $TCID_{50}$ of the indicated virus and virus titer ($TCID_{50}$/ml) in the supernatant was determined at 24 (**C**) or 48 (**D**) hours post-inoculation by virus titration assay on MDCK cells. The limit of detection was 100 $TCID_{50}$ (dotted line). The data represent the geometric mean + the standard error of the mean of four samples derived from two or three independent experiments performed in duplicate (*, $P<0.05$; **, $P<0.01$; ***, $P<0.005$, compared to H7N3 values).

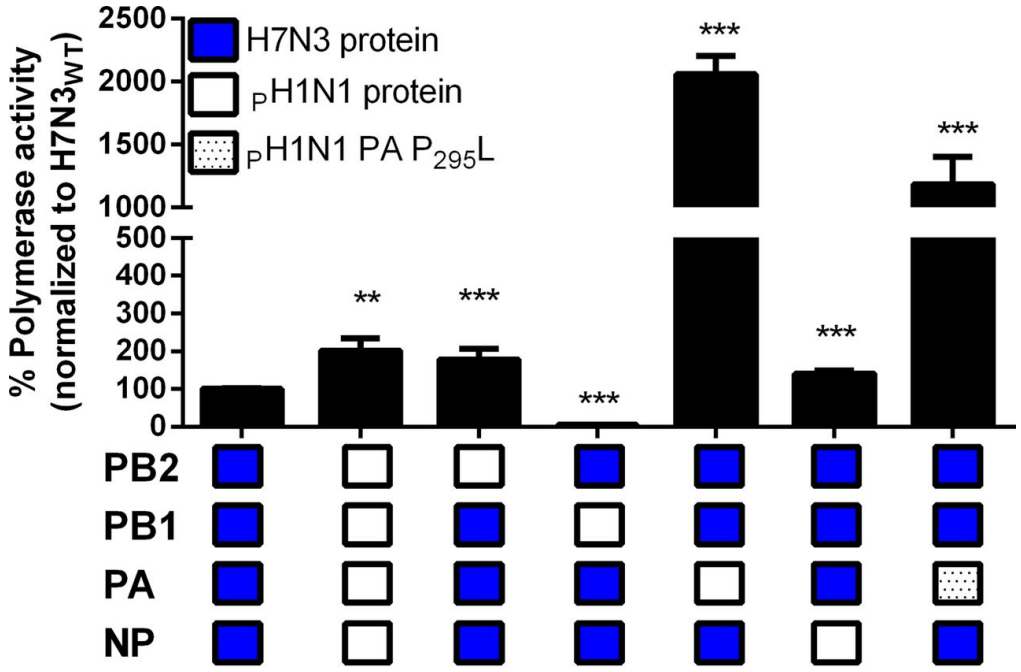

% Polymerase activity
(normalized to H7N3WT)
2500
2000
1500
1000
500
400
300
200
100
0
H7N3 protein
PH1N1 protein
PH1N1 PA P295L
**
***
***
***
***
***
PB2
PB1
PA
NP

Figure 2.3. PA protein from pandemic H1N1 virus increases polymerase activity of H7N3 in human 293T cells.

A. A reporter assay was used to measure the relative polymerase activity of different polymerase protein complex combinations between H7N3 and $_{P}$H1N1 viruses, including $_{P}$H1N1 PA $P_{295}L$ as described in the Materials and Methods. The relative polymerase activity was first normalized to the Renilla activity and compared to that of a full set of H7N3 polymerase proteins. Blue boxes represent H7N3 proteins, white boxes represent pH1N1 proteins, and spotted white boxes represent pH1N1 PA-$P_{295}L$. Data represent the average normalized value from three or five experiments performed in duplicate (*, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.005$, compared to H7N3 values).

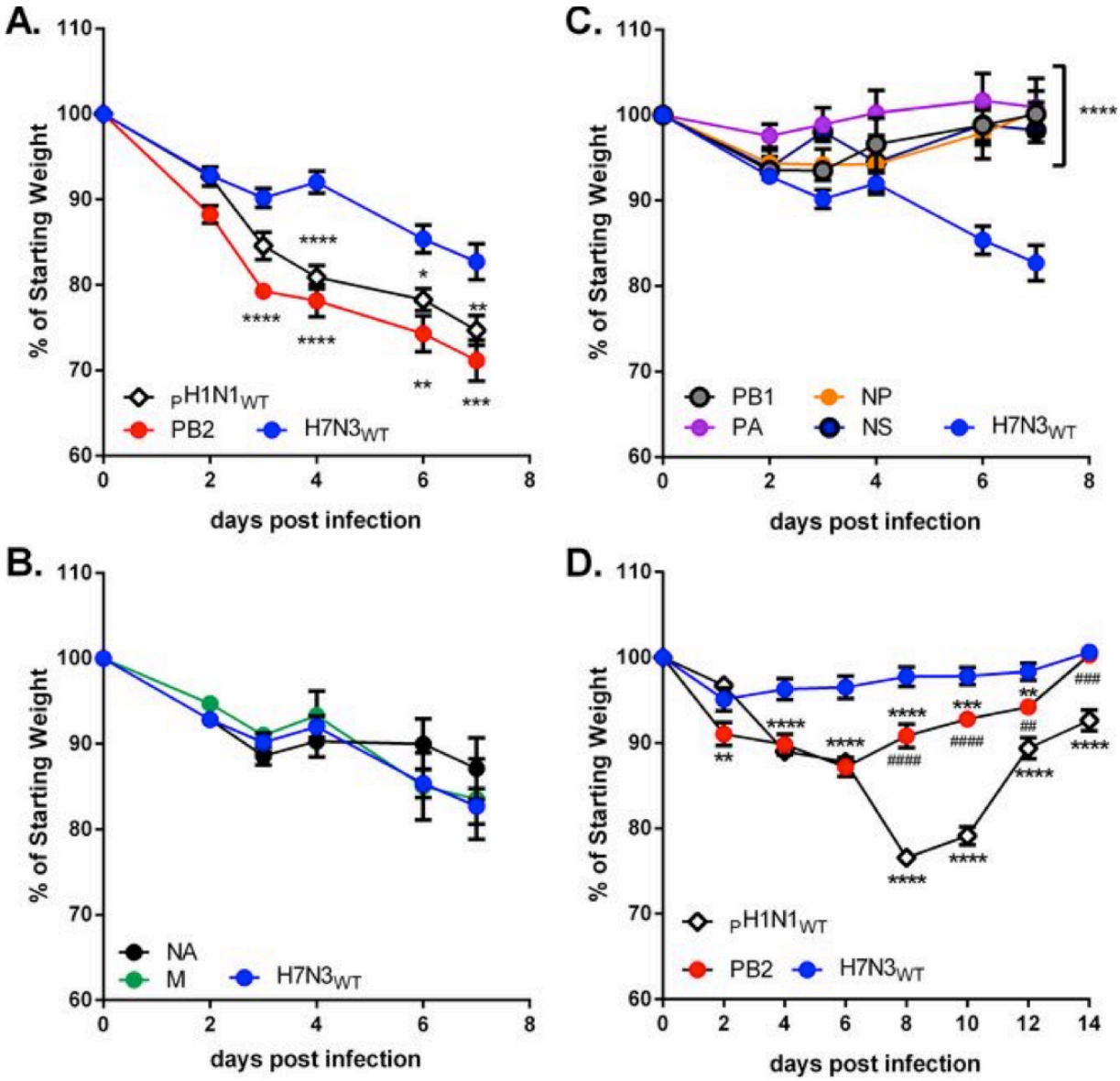
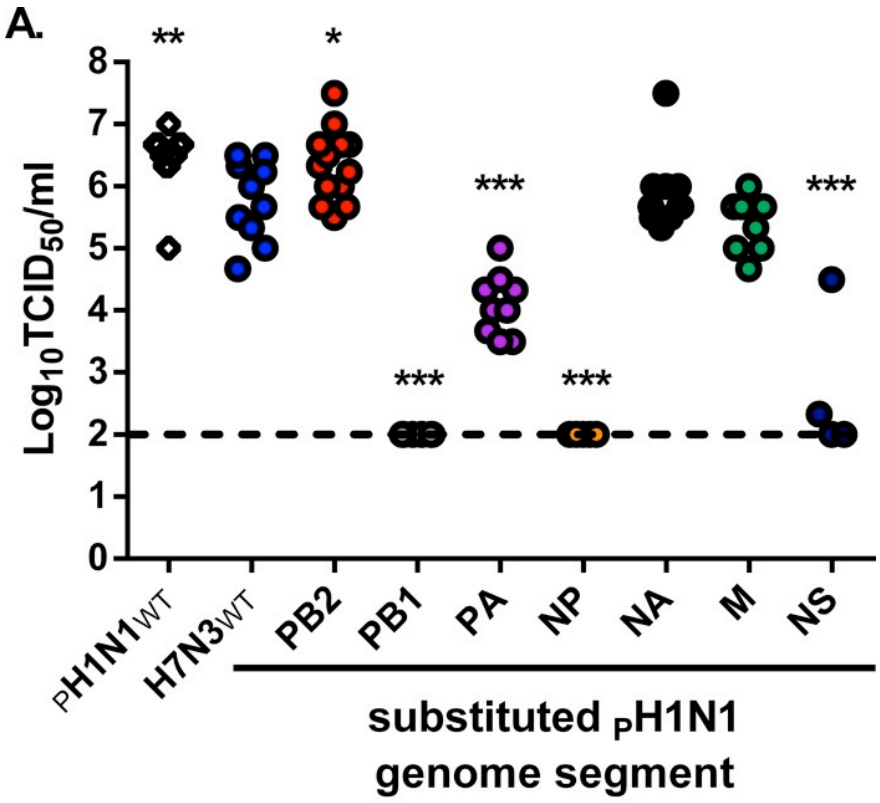

A.
% of Starting Weight
days post infection
pH1N1WT
PB2
H7N3WT
B.
% of Starting Weight
days post infection
NA
M
H7N3WT
C.
% of Starting Weight
days post infection
PB1
PA
NP
NS
H7N3WT
D.
% of Starting Weight
days post infection
pH1N1WT
PB2
H7N3WT

Figure 2.4. Reassortant H7 viruses containing $_{P}$H1N1 PB2, NA, or M induce considerable morbidity in mice

(**A-C**). Groups of male C57BL/6J were inoculated intranasally with 10^4 $TCID_{50}$ of the indicated viruses and weight loss was measured for 7 days. Morbidity is displayed as percent of original body weight remaining for mice inoculated with the indicated virus that showed significantly greater (**A**), equal (**B**), or significantly less (**C**) weight loss than the parental H7N3 virus. (**D**) Male C57BL/6J mice were inoculated intranasally with 10^3 $TCID_{50}$ of H7N3, $_{P}$H1N1 or H7N3 containing the PB2 of $_{P}$H1N1 ($H7N3_{PB2}$) and weight loss was measured for 14 days. The data represents the average and standard error of the mean of at least six animals from two or more independently performed experiments per virus. (*, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.005$, compared to H7N3 values; #, $P < 0.05$; ##, $P < 0.01$; ###, $P < 0.005$; ####, $P < 0.001$, compared to $_{P}$H1N1 values).


A.
**
*
***
***
***
***
Log10TCID50/ml
0
1
2
3
4
5
6
7
8
PH1N1WT
H7N3WT
PB2
PB1
PA
NP
NA
M
NS
substituted PH1N1
genome segment

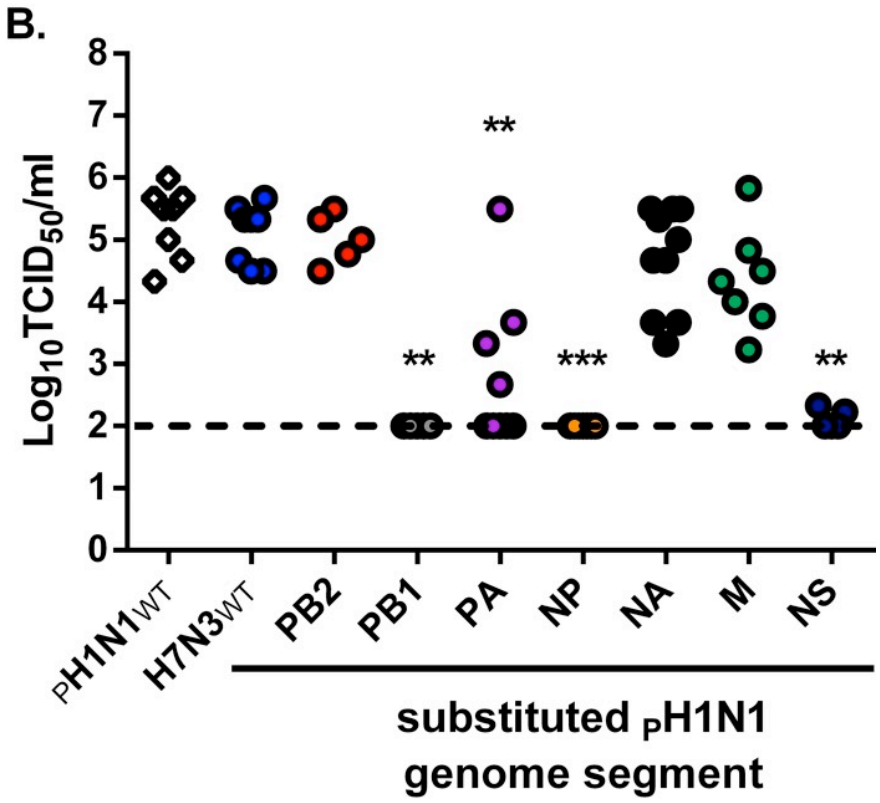

B.
**
**
***
**
Log10TCID50/ml
0
1
2
3
4
5
6
7
8
PH1N1WT
H7N3WT
PB2
PB1
PA
NP
NA
M
NS
substituted PH1N1
genome segment

Figure 2.5. H7N3 replicates extensively in the lungs of C57BL/6J mice and $_{P}$H1N1 PB2, NA, or M do not attenuate H7N3.

Male C57BL/6J mice were inoculated with 10^4 $TCID_{50}$ of H7N3, $_{P}$H1N1, or the indicated reassortant virus and lung viral titers were quantified 3 and 7 days post inoculation. Virus titer present in whole lung homogenate at day 3 (A) or day 7 (B) was determined by virus titration assay on MDCK cells and expressed as $TCID_{50}$/ml. Viral titers were derived from five or more mice from two or more independently performed experiments. The limit of detection was 100 $TCID_{50}$ (dotted line). (*, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.005$, compared to H7N3 values.)

2.8 References

1. Yu, H. *et al.* Human infection with avian influenza A H7N9 virus: An assessment of clinical severity. *Lancet* **382,** 138–145 (2013).

2. Kimble, J. B. *et al.* Alternative reassortment events leading to transmissible H9N1 influenza viruses in the ferret model. *J. Virol.* **88,** 66–71 (2014).

3. Itoh, Y. *et al.* In vitro and in vivo characterization of new swine-origin H1N1 influenza viruses. *Nature* **460,** 1021–1025 (2009).

4. Zhang, Y. *et al.* H5N1 Hybrid Viruses Bearing 2009/H1N1 Virus Genes Transmit in Guinea Pigs by Respiratory Droplet. *Science (80-.).* **340,** 1459–1463 (2013).

5. Wu, A. *et al.* Sequential reassortments underlie diverse influenza H7N9 genotypes in China. *Cell Host Microbe* **14,** 446–452 (2013).

6. Kapczynski, D. R. *et al.* Characterization of the 2012 highly pathogenic avian influenza H7N3 virus isolated from poultry in an outbreak in Mexico: pathobiology and vaccine protection. *J. Virol.* **87,** 9086–96 (2013).

7. Watanabe, T. *et al.* Characterization of H7N9 influenza A viruses isolated from humans. *Nature* **501,** 551–5 (2013).

8. Gabbard, J. D. *et al.* Novel H7N9 Influenza Virus Shows Low Infectious Dose, High Growth Rate, and Efficient Contact Transmission in the Guinea Pig Model. *J. Virol.* **88,** 1502–1512 (2014).

9. Belser, J. A. *et al.* Pathogenesis and transmission of avian influenza A (H7N9) virus in

ferrets and mice. *Nature* **501,** 556–9 (2013).

10. Richard, M. *et al.* Limited airborne transmission of H7N9 influenza A virus between ferrets. *Nature* **501,** 560–3 (2013).

11. Bahl, J. *et al.* Influenza A Virus Migration and Persistence in North American Wild Birds. *PLoS Pathog.* **9,** e1003570 (2013).

12. Smith, G. J. D. *et al.* Origins and evolutionary genomics of the 2009 swine-origin H1N1 influenza A epidemic. *Nature* **459,** 1122–5 (2009).

13. Nelson, M. I., Vincent, A. L., Kitikoon, P., Holmes, E. C. & Gramer, M. R. Evolution of novel reassortant A/H3N2 influenza viruses in North American swine and humans, 2009-2011. *J. Virol.* **86,** 8872–8 (2012).

14. Pascua, P. N. Q. *et al.* Virulence and transmissibility of H1N2 influenza virus in ferrets imply the continuing threat of triple-reassortant swine viruses. *Proc. Natl. Acad. Sci.* **109,** 15900–15905 (2012).

15. Reed, L. J. & Muench, H. A Simple Method of Estimating Fifty Per Cent Endpoints. *Am. J. Hygeine* **27,** 285–297 (1938).

16. Kimble, J. B., Sorrell, E., Shao, H., Martin, P. L. & Perez, D. R. Compatibility of H9N2 avian influenza surface genes and 2009 pandemic H1N1 internal genes for transmission in the ferret model. *Proc. Natl. Acad. Sci. U. S. A.* **108,** 12084–8 (2011).

17. Li, C. *et al.* Reassortment between avian H5N1 and human H3N2 influenza viruses creates hybrid viruses with substantial virulence. *Proc. Natl. Acad. Sci. U. S. A.* **107,**

4687–92 (2010).

18. Boon, A. C., Williams, R. W., Sinasac, D. S. & Webby, R. J. A novel genetic locus linked to pro-inflammatory cytokines after virulent H5N1 virus infection imice. *BMC Genomics* **15,** 1017 (2014).

19. Walsh, K. B. *et al.* Suppression of cytokine storm with a sphingosine analog provides protection against pathogenic influenza virus. *Proc. Natl. Acad. Sci. U. S. A.* **108,** 12018–23 (2011).

20. Shoemaker, J. E. *et al.* An Ultrasensitive Mechanism Regulates Influenza Virus-Induced Inflammation. *PLOS Pathog.* **11,** e1004856 (2015).

21. Capua, I. & Munoz, O. Emergence of influenza viruses with zoonotic potential: Open issues which need to be addressed. A review. *Vet. Microbiol.* **165,** 7–12 (2013).

22. Lakdawala, S. S. *et al.* Eurasian-origin gene segments contribute to the transmissibility, aerosol release, and morphology of the 2009 pandemic H1N1 influenza virus. *PLoS Pathog.* **7,** (2011).

23. Schrauwen, E. J. a *et al.* Reassortment between Avian H5N1 and human influenza viruses is mainly restricted to the matrix and neuraminidase gene segments. *PLoS One* **8,** e59889 (2013).

24. Chou, Y. *et al.* The M segment of the 2009 new pandemic H1N1 influenza virus is critical for its high transmission efficiency in the guinea pig model. *J. Virol.* **85,** 11235–41 (2011).

25. Sutton, T. C. *et al.* Airborne Transmission of Highly Pathogenic H7N1 Influenza Virus in Ferrets. *J. Virol.* **88,** 6623–6635 (2014).

26. Schrauwen, E. J. & Fouchier, R. A. Host adaptation and transmission of influenza A viruses in mammals. *Emerg. Microbes Infect.* **3,** e9 (2014).

27. Belser, J. A. *et al.* Pathogenesis, Transmissibility, and Ocular Tropism of a Highly Pathogenic Avian Influenza A (H7N3) Virus Associated with Human Conjunctivitis. *J. Virol.* **87,** 5746–5754 (2013).

28. Belser, J. a *et al.* Contemporary North American influenza H7 viruses possess human receptor specificity: Implications for virus transmissibility. *Proc. Natl. Acad. Sci. U. S. A.* **105,** 7558–63 (2008).

29. Kitikoon, P. *et al.* Pathogenicity and transmission in pigs of the novel A(H3N2)v influenza virus isolated from humans and characterization of swine H3N2 viruses isolated in 2010-2011. *J. Virol.* **86,** 6804–14 (2012).

30. Mehle, A. & Doudna, J. a. Adaptive strategies of the influenza virus polymerase for replication in humans. *Proc. Natl. Acad. Sci. U. S. A.* **106,** 21312–6 (2009).

31. Shinya, Y. *et al.* Biological and structural characterization of a host-adapting amino acid in influenza virus. *PLoS Pathog.* **6,** 15–16 (2010).

32. Gabriel, G. *et al.* The viral polymerase mediates adaptation of an avian influenza virus to a mammalian host. *Proc. Natl. Acad. Sci. U. S. A.* **102,** 18590–5 (2005).

33. Ilyushina, N. a *et al.* Adaptation of pandemic H1N1 influenza viruses in mice. *J. Virol.*

84, 8607–8616 (2010).

34. Ducatez, M. F. *et al.* Both influenza hemagglutinin and polymerase acidic genes are important for delayed pandemic 2009 H1N1 virus clearance in the ferret model. *Virology* **432,** 389–393 (2012).

Chapter 3:

Amino Acid Identity Affects Competitive Incorporation of influenza A Genome Segments and Limits the Potential for Reassortment

3.1 Abstract

Influenza A virus (IAV) reassortment leads to the generation of new viral strains including those with pandemic potential. The segmented nature of the viral genome allows co-packaging of viral genomic RNA from multiple parental strains upon cellular co-infection. Previous studies have identified so-called packaging signals at the distal ends of each vRNA segment as determinants of selective packaging of reporter RNAs into viral particles. More recently vRNA bundling signals and other structural sequence elements that lie outside of these canonical packaging elements were implicated in the selection of divergent RNA species. Here we utilize a recently developed competitive reverse genetics strategy to comprehensively assess the potential for reassortment between genetically diverse virus segments in the context of an isogenic background commonly used in laboratory and vaccine development settings. Initially, we show that competition occurs *in vitro* in a genome context-dependent manner. We then determine the n-terminal region outside of known packaging elements of segment 3 (PA) is a determinant of selection. These experiments simultaneously reveal inherent flexibility of the viral genome while illuminating sequences that are important for selective packaging during co-infection. Finally, we examine the nucleotide and amino acid identities within segment 3 that drive the preferential selection of one strain's segment over another. Taken together we establish the amino acid identity of PA-184, as an important human-specific determinant of polymerase activity and competitive selection in this system.

3.2 Introduction

Influenza A viruses (IAV) belong to the *orthomyxovirus* family. These viruses possess a genome composed of eight single-stranded negative-sense RNA segments encapsulated in a nucleoprotein coat and polymerase complex packaged in a host-cell derived envelope[1]. In the event of cellular coinfection, viral ribonucleoprotein complexes (vRNPs) from multiple strains of the virus are able to simultaneously utilize the host cell for replication and genome packaging[2]. If a mixed genome is successfully packaged, resulting viruses with segments from both parental strains result - giving rise to a new genome constellation[3]. This process, termed reassortment, is one of the evolutionary strategies employed by IAV to produce new viral strains that may have altered replication, pathogenesis, and transmission profiles when compared to previous isolates. For reassortment to be successful new genomic constellations must be replication competent and when there is little difference between constituent segments of two viruses reassortment may occur at a high frequency[2]. Importantly, reassortment may result in progeny viruses with mixed genomes that possess a combination of virologic traits that induce shifts in host adaptation, transmission potential, and pathogenesis[4].

Influenza reassortment has lead to the generation of multiple viruses capable of generating pandemics. Requisite for pandemic formation is the ability of the virus to break animal-to-human transmission barriers, replicate and release efficiently in the human host, and transmit form human-to-human. The profound effect of reassortment may be illustrated by inspecting the phylogeny of pandemic IAVs from 1918 to 2009 and the recently emergent H7N9[5]. In each case an antigenically distinct HA was acquired by way of reassortment with a virus capable of transmitting to and replicating in humans. Humoral immunity to HA is the

primary basis of current vaccination strategies and antigenic shift of HA by reassortment alters recognition by antibodies[6]. The 1918 H1N1 "Spanish Flu" was likely created by reassortment of at least one avian virus and a mammalian virus that circulated in porcine hosts before introduction to humans[7]. Between 20-50 million people died as a result of primary influenza infection and subsequent secondary pneumonia. Contemporary segments with high homology to the 1918 H1N1 strain persist in both avian and swine populations[8]. Experimental creation of an H1N1 strain derived from currently circulating avian genome segments was also shown to cause severe disease in mammals and can transmit between ferrets[9]. Moreover, reassortment between human and avian influenza A viruses, likely in a porcine vessel, resulted in the generation of H2N2 and H3N2 strains that caused pandemics in 1957 and 1968, respectively. H3N2 viruses have continued to circulate following the pandemic and are, along with H1N1, the primary cause of seasonal epidemics[7,10].

Prior to each pandemic, antigenic shift of external proteins occurred in addition to transfer of internal gene segments. The 2009 pandemic H1N1 emerged following at least three independent reassortment events involving human, porcine, and avian viruses. In each recent human pandemic the PB2 and PA segments co-segregated in part due to a requirement for interacting viral protein interfaces. Co-segregation of H3N2 and H1N1 PB2 and PA segments has been demonstrated experimentally using high-content reassortment assays with replicating viruses[11]. Additionally, the interdependence of PB2 and PA residues has been interrogated extensively *in vitro* utilizing co-transfection mini-genome assays to assess the amino acid residues responsible for polymerase complex activity in the context of reassortant protein complexes. Utilizing phylogenetic data from human and avian strains, interdependence on amino acid residues 184 and 383 in the PA subunit were required for an H1N1 PA to be able to

replicate genome in the context of a polymerase complex otherwise derived from an H3N2 isolate[12].

Here, we utilize a competitive reverse genetics system to identify differences genetic restriction of closely related avian gene segments in the context of a H1N1 virus in human cells[13]. We identify specific polymerase complex combinations in the context of replicating virus and map this activity to a specific genetic locus in PA[14]. Prior studies have focused on the restriction of reassortment between H3N2 and H1N1 viruses, whereas we explore why certain otherwise functional avian genes are non-permissively excluded from H1N1 viruses during competition. We find, that the identity of amino acid 184 dictates the ability of an avian H7N3 PA genome segment to be incorporated into a H1N1 isolate descended from the 1918 pandemic H1N1. Additionally, the selection bias demonstrated during reverse genetic competition assays may be alleviated by addition of the heterologous H7N3 PB2 segment, indicating a specific interplay between these two elements. Manipulation of the H1N1 PA-184 residue attenuated virus replication in human, but not canine cells and this can be attributed to diminished polymerase activity. Restoration of the H1N1-like PA-184 residue increases the polymerase activity of H7N3 PA and relative competitive incorporation rate when compared to the H1N1 virus, indicating that this amino acid is a species-specific and context-dependent regulator of reassortment potential.

3.3 Materials and methods

Biosafety Considerations

All reverse genetic experiments were carried out in the BSL2+ setting, within a class II biosafety hood, utilizing appropriate personal protective equipment. All experiments were discussed with Washington University in Saint Louis's Institutional Biosafety Committee (IBC) and approved by the biosafety officer.

Viruses

Viruses used in this study: A/shorebird/Delaware/22/2006 (H7N3), A/California/04/2009 ($_{P}$H1N1), A/mallard/Alberta/144/2007 (H7N9), A/Memphis/03/2008 ($_{S}$H1N1), and A/Puerto Rico/08/1934 ($_{PR8}$H1N1). For $_{S}$H1N1, H7N3, and H7N9, cDNA for all gene segments was cloned into the bi-directional pHW2000 plasmid and used to generate influenza viruses as previously described (see Chapter 2). Dr. Richard Webby at St. Jude Children's Research Hospital kindly provided the reverse genetic plasmids for $_{P}$H1N1 and $_{PR8}$H1N1. Plasmid-derived A/Puerto Rico/08/1934 (H1N1), A/shorebird/Delaware/22/2006 (H7N3), and all $_{PR8}$H1N1 7+1 single reassortant viruses, were generated using the 293T-Madin-Darby canine kidney cell (MDCK) co-culture system and supernatant was injected into 10-day old embryonated chicken eggs for 48 hours at 35°C (Cackle Hatcheries, IA, USA). Allantoic fluid containing the infectious virus was harvested and stored at -80°C. The viral titer (tissue culture infectious dose 50, $TCID_{50}$) was determined using MDCK cells and viral stock used in this study was titrated at least twice independently. Single reassortant viruses on the background of $_{PR8}$H1N1 possessing a single genome segment from H7N3 used in this study have been named with the following convention

– seven segments from the $_{PR8}$H1N1virus plus the substituted H7N3 segment in subscript (ex. PR8:PA_{H7N3}).

Cells

MDCK cells were maintained in Minimal essential medium (MEM) with 5% Fetal bovine serum (FBS), vitamins, L-glutamine (Invitrogen), penicillin, and streptomycin. 293T cells were maintained in Opti-MEM with 10% FBS, L-glutamine, penicillin and streptomycin. A549 cells were maintained in Dulbecco's Minimal essential medium (DMEM) with 10% FBS vitamins, L-glutamine, penicillin, streptomycin, 25mm HEPES, and non-essential amino acids (NEAA).

Competitive Reverse Genetic Assay

We utilized a previously developed competitive reverse genetic assay to evaluate the ability of avian and human IAV genome segments to reassort in the context of a $_{PR8}$H1N1 viral background (Figure 1A). 293T-MDCK co-cultures were transfected with pHW2000 plasmids (1 μg per segment) containing seven genome segments of $_{PR8}$H1N1 virus plus four (for NA segment due to exclusion of NA from sH1N1) or five plasmids encoding a single genome segment (e.g. PB2) from different avian and human influenza viruses. All transfections included the wild-type $_{PR8}$H1N1 segment in addition to those from divergent strains. We did not evaluate the HA genome segment because of biosafety considerations. The total amount of co-transfected plasmid DNA for the 8th segment (1 μg) was divided equally between the four or five studied viruses including the parental strain. The DNA was mixed at a 1:2 ratio with Trans-IT LT1 (Mirus) in Opti-MEM for 20 minutes at room temperature and added to the culture medium. Following an overnight incubation, the cell culture medium was removed and 1ml of fresh Opti-MEM

supplemented with penicillin, streptomycin, L-glutamine was added. Twenty-four hours later an additional 1ml of Opti-Mem with 1μg/ml TPCK-trypsin (Worthington) was added. Forty-eight hours later supernatants were collected, cell debris removed by centrifugation at 1,200xg, and frozen at -80°C until further analysis. Clonal viral populations were isolated by limiting dilution assay on MDCK cells, then viral RNA extracted, reverse-transcribed using a vRNA-specific primer, and amplified with segment-specific PCR primers. Genome segment amplicons were genotyped by either restriction fragment length polymorphism analysis (RFLP) or Sanger sequencing. RFLP analysis was performed with one to four different restriction enzymes yielding a unique fragment length pattern for each segment and strain: PB2 (EcoRI, HindIII, BamHI); PB1 (HindIII); PA (BamHI, HindIII, XbaI); NP (BamHI, BglII); NA (BamHI, BglII, EcoRI, BsmBI); M (BamHI, HindIII, XcmI, PvuII). All restriction enzymes were obtained from New England Biolabs and used according to their instructions. The NS gene segment was genotyped by Sanger sequencing. Competitive reverse genetic assays were completed at least twice independently for each genome segment.

Multistep Growth Curves of Influenza A virus

$2x10^5$ MDCK cells were seeded in 24-well plates and infected the next day at MOI 0.001. Briefly, MDCK cells were washed once with PBS before adding the inoculum in Minimal Essential Medium (MEM) containing penicillin, streptomycin, glutamine, and vitamins plus 0.1% bovine serum albumin (M0.1B) for one hour at 37°C. After the one hour, the cells were washed with PBS and 1.0 ml of M0.1B with 1μg/ml TPCK-trypsin (Worthington Technologies) was added to each well. Culture supernatants were collected at 24 and 48 post infection and the amount of infectious virus was quantified by titration on MDCK cells. All growth curve experiments were performed at least twice in duplicate. Infection of 293T cells was performed in

24-well plates coated with poly-D-lysine to enhance cellular adherence with each well containing $2x10^5$ cells. Infection of cells was performed at MOI 0.1 and samples collected from supernatant at 24 hours. One hour post-infection, inoculum was removed, cells washed once with PBS, and fresh M0.1B media containing 0.4 μg/ml TPCK-trypsin was added to the culture.

Generation of mutants

Chimeric segments were generated by creation of fragments that carried BsmBI (New England Biosciences) restriction sites to allow scar-less integration of unique pieces. All single-mutant pHW2000 plasmids were generated by use of the Agilent Quick Change II system according to the manufacturer's protocol. Sequence identity of all newly generated chimeric and mutant plasmids was verified by sequencing. All viruses were additionally examined for rescue capacity in the $_{PR8}$H1N1 background and successfully generated.

Mini-genome luciferase assay with polymerase genes of influenza A virus

The PB2, PB1, PA and NP genes from $_{PR8}$H1N1 and H7N3 were previously cloned into the pHW2000 vector. The pLuci plasmid was kindly provided by Dr. Yen (Hong Kong University, Hong Kong, China) and contains the firefly luciferase gene flanked by the non-coding regions of NP gene segment in the negative orientation. The transcription of this influenza A virus-like gene segment is under the control of a human polymerase I promoter. A Renilla luciferase gene containing expression plasmid is used to normalize for transfection efficiency. 293T cells are transfected with plasmids containing the PB2, PB1, PA and NP gene segments of IAV plus the two luciferase containing plasmids (500ng total DNA) using TransIT LT1 (Mirus Bio LLC). The next day the media was changed and the assay was incubated for 48 hours from the time of transfection before the cells were harvested, lysed and used to analyze Firefly and Renilla

luciferase activity (Promega). Each condition (set of plasmids) was done in duplicate and repeated independently in separate experiments three times. The relative light units (RLU) of firefly are normalized to the RLU for Renilla and the activity of varying polymerase combinations are then normalized to that of the full complement of $_{PR8}$H1N1 IAV segments within each experiment.

$TCID_{50}$ assay

Confluent monolayers of MDCK cells were grown overnight in 96-well plates. The next day, the cells were washed with PBS and inoculated with ten-fold serial dilutions (10^{-1} to 10^{-8}) of culture supernatant, allantoic fluid or lung homogenate for one hour in M0.1B at 37°C and 5% CO_2. After one hour the inoculum was removed and replaced with M0.1B supplemented with 1μg/ml TPCK-trypsin and incubated for 72 hours. Presence of virus was determined by hemagglutination assay using 0.5% turkey red blood cells. $TCID_{50}$ was determined by the Reed-Muench method [15].

Inoculation of embryonated eggs with influenza A virus

Embryonated hen eggs from Cackle Hatcheries, Iowa, USA, were inoculated with the indicated dose of virus in 100uL sterile PBS and 1X egg antibiotics (I need to include these specifically probably) by needle injection of the allantois cavity. Eggs were incubated at 35°C for 48 hours, and then chilled overnight. The following day allantois fluid was harvested, cleared of debris by centrifugation at 1,200 RPM for 5 minutes, stored at -80°C for analysis.

RT-qPCR analysis

For all experiments 100uL of viral supernatant was harvested in 275uL lysis buffer and stored at -80C until use. RNA was extracted using the Omega E.Z.N.A. Total RNA I Kit according to the manufacturer's instructions and eluted in 50uL RNase-Free Water. 2 μL of total RNA was reverse transcribed with random hexamers using SuperScript III reverse transcriptase (Invitrogen) according to manufacturer's instructions. After diluting cDNA 1:10 in H_2O, 4 μL was added to a mixture of the indicated primer-probe (IDT), H_2O and 10 μL of 2x qPCR MasterMix (TaqMan® Fast Universal PCR Master Mix (2x) Applied Biosystems by Life Technologies). All qPCR assays were performed using the 7500 Fast system. Cycle-threshold values were uniformly set to 0.4.

Statistical analysis

Statistical analyses were performed using GraphPad Prism 6.0 software. All statistical comparisons were performed using One-Way ANOVA with Dunnet's Correction for multiple comparisons or Two-Way ANOVA with Sidak Correction for multiple comparisons as indicated in figure legends.

3.4 Results

Competitive reverse genetics reveals inherent biases during conditions mimicking reassortment.

Influenza A virus possesses a segmented RNA genome; as such the coordinated packaging of each segment must be maintained for progeny virus formation. Each genome segment must be packaged into newly forming virions to allow replicative success following egress from the host cell. To assess the potential for reassortment between viruses and interrogate viral features that dictate selection of a given segment in a competitive context, we utilized an established in vitro reassortment system. In this system, utilizing an isogenic viral background, we have identified inherent biases in the creation of reverse genetic derived IAV. We chose to examine competitive incorporation in the context a common laboratory strain A/Puerto Rico/08/1934 (H1N1) ($_{PR8}$H1N1) and compared the results to those we determined previously for an avian isolate A/shorebird/DE/22/2006 (H7N3) - to examine if the determinants of reassortment are strain specific. Additionally, we included segments from another avian isolate – A/mallard/Alberta/144/2007 (H7N9) –and two human strains – A/Memphis/03/2008 (H1N1) ($_{S}$H1N1) and A/California/04/2009 (H1N1) ($_{P}$H1N1). In our competitive reverse genetics system, seven segments are held constant as a stable backbone while five options of the eighth segment are provided in equal amounts (**Figure 3.1 A**). The same technique is utilized to generate wild type and reassortant viruses utilized in later studies. Following a period of infection the supernatant is harvested, titered, and subjected to a limiting dilution assay to isolate clonal viruses. Next, individual viruses are isolated, RNA extracted, reverse transcribed, amplified with segment specific primers and identified by restriction fragment length

polymorphisms or Sanger Sequencing (**Figure 3.1 B**). We have assessed the incorporation rate of Segments 1-3 and 5-8 in at least two independent experiments. Segment 4 (HA) was excluded from this study due to biosafety concerns.

$_{PR8}$H1N1 preferentially packages human-derived genome segments

We found repeatedly that the $_{PR8}$H1N1 background preferentially incorporated segments derived from human isolates compared to avian origin segments in all cases. More heterogeneity in resulting viral pools was observed when PA, M, and NS were assessed. We additionally chose to determine the rescue efficiency of all possible single gene reassortants that may be created under these conditions. We were able to efficiently recover 22 of 28 viruses to levels within 10-fold of the parental $_{PR8}$H1N1 background (**Figure 3.1 B**). Only six single reassortant viruses were not generated after multiple attempts, interestingly including the avian-origin PB2 segments, and intriguingly the human origin PA segments. Additionally, although we did not assess the competitive outcomes of Segment 4 (HA) we were able to recover virus representing each of these as well. We then chose to focus on the genes that comprise the vRNP and polymerase complexes, as compatibility of proteins derived from these segments have been identified as primary host restriction factors following reassortment in previous studies.

Preferential Incorporation of Segment 3 (PA) is due replication kinetics in human cells.

We chose to focus on the selection differences of Segment 3 (PA) because of the genetic interaction of PA and PB2 observed in nature[11,12]. Prior studies on co-functionality of these segments identified residues required for function of a H1N1 PA in the context of a H3N2 polymerase complex[12]. We wondered if similar genetic features determined the differential incorporation of closely related avian Segment 3 segments in a H1N1 context (**Figure 3.1 A**). In

head-to-head competitive transfections in the $_{PR8}$H1N1 background between the H7N3 and $_{PR8}$H1N1 segments, only the $_{PR8}$H1N1 segment was selected (**Figure 3.2 A**). Conversely, in the H7N3 context, only H7N3 PA was ever recovered when assessed by limiting dilution assay. We generated egg-derived stocks of $_{PR8}$H1N1 and $_{PR8}$H1N1 7+1 H7N3$_{PA}$ viruses and performed multi-cycle replication studies at low multiplicity of infection and surprisingly found no difference in MDCK cells (**Figure 3.2 B**). Next, we hypothesized there may be differences in coordinated genome packaging efficiency of these two viruses and utilized a particle-normalized single-step growth curve to assess the replication efficiency of these viruses in multiple cell types. Specifically, we included two human cell lines commonly utilized to study influenza viruses. A549 cells are a human adenocarcinoma cell line that is capable of supporting multi-cycle replication of IAV. 293T cells are utilized for transfection-based assays, including generation of reverse genetically-derived IAV, and assessment of IAV polymerase function. As such, we assessed the virus released from cells infected with equal particle number at 24 hours, in a single-round infection. As in the low MOI infection in MDCK cells we saw no replication defect of the $_{PR8}$H1N1 7+1 H7N3$_{PA}$ relative to $_{PR8}$H1N1 (**Figure 3.2 C**). Surprisingly, we saw a multi-log difference in replication in both human cell lines for $_{PR8}$H1N1 7+1 H7N3$_{PA}$ relative to $_{PR8}$H1N1, indicating a human-specific blockade of replication of a virus containing only one avian-origin genome segment. Similarly, we saw a decrease in the production of viral antigen in human, but not canine cells for H7N3$_{PA}$ relative to $_{PR8}$H1N1 (**Figure 3.2 D**).

Identification of Segment 3's N-terminal region, independent of packaging signals, as a determinant of selection.

We designed a strain-specific RT-qPCR assay to determine the relative abundance of viruses in supernatant following transfection and a single blind passage in MDCK cells to

amplify signal, since there are no replication differences in these cells. We found a greater than 25-fold increase in $H1N1_{PA}$ relative to $H7N3_{PA}$ in the $_{PR8}$H1N1 context and the opposite in the H7N3 genomic context (**Figure 3.3 A**).

The terminal regions of vRNA have been identified as determinants of segment packaging. Nearly a decade ago the identity of numerous synonymous mutations in these UTR and at the termini of the open reading frames (ORFs) of each segment were shown to impact the selection of artificial reporter segments encoding the ORF of fluorescent proteins[16,17]. Increasing recent evidence points to the central region of genome segments that were not assessed in early studies as additional critical determinants of selection during competitive conditions between divergent viruses[18,19]. To interrogate the contributions of naturally occurring nucleotide variation across the entire length of the segment we constructed chimeric segments that encode the packaging region of either $_{PR8}$H1N1 and the central region of H7N3 or vice versa (**Figure 3.3 B**). Additionally, we constructed segments that mimic the boundaries of defective interfering RNAs that are able to be packaged and propagate during high multiplicity infection, but lack that ability to produce protein and replicate independently[20]. We find that in competition with a wild-type segment bearing the same packaging signals or with the extended DI boundary constructs, that the central region of the segment is a predominant determinant of segment 3 selection (**Figure 3.3 A**).

Having excluded traditional packaging signals, we explored whether or not the function of the polymerase complex in 293T cells was correlated with discrepant genome selection. We find that in the context of the $_{PR8}$H1N1 polymerase complex $H7N3_{PA}$ is attenuated (**Figure 3.3 C**). Additionally, the central region of the genome segment is responsible for this defect, as the packaging signals of DI-like constructs show no difference from the segment in which the central

region was derived (**Figure 3.3 C**). To further probe the region necessary for altered polymerase activity created additional constructs encoding the N- or C-terminus of either strain fused to the N- or C-terminus of the other strain and a majority H7N3 construct containing only nucleotides 400-700 of $_{PR8}$H1N1 (**Figure 3.3 B)**. We found that the N-terminal region of PR8 (8N3C construct), located between nucleotides 400-700 ($8_{400}3_{700}8$ construct) is sufficient to increase H7N3$_{PA}$ polymerase function to greater than that of WT-$_{PR8}$H1N1$_{PA}$ in the context of the remaining $_{PR8}$H1N1 genome segments (**Figure 3.3 C)**. Together these findings indicate the central region of Segment 3, between the boundaries of terminal packaging signals, but within the N-terminal half of the segment likely determines competitive selection.

PA amino acid 184, but not 190, is a context-dependent determinant of differential segment selection

Having established the central region of segment 3 as the primary segment-intrinsic driver of selection in this given background, and nucleotides 400-700 as the determinant of polymerase activity in human cells, we sought to utilize sequence analysis between the three strains that propagated equivalently but showed divergent selection properties ($_{PR8}$H1N1, H7N3, and H7N9). PA of H7N3 and H7N9 contain seven coding changes[14]. We have previously investigated the contribution of differences between these viruses as critical to regulation of virus replication and pathogenesis. H7N3 segment 3 was selected at a much lower rate than that of H7N9 in the context of a $_{PR8}$H1N1 genome. As such we generated mutant segments bearing single amino acid changes at three positions that differed most in their biochemical properties and assessed the contribution of these amino acids to the polymerase activity and competitive selection of $_{PR8}$H1N1 and H7N3. We generated H7N3$_{PA}$ mutant segments that have the $_{PR8}$H1N1 or H7N9 codons at amino acids 184 and 190, since these are the variable positions in the central

region of PA we find determines differential polymerase activity (**Figure 3.4 A**). We generated each of these viruses as well as a H7N3 PA that has an alternative asparagine codon as a control. Similarly, we generated a $_{PR8}$H1N1$_{PA}$ construct bearing the amino acids of H7N3$_{PA}$ at position 184. All viruses were able to be rescued in two independent attempts and none of the PA-184-variant viruses were attenuated relative to WT-$_{PR8}$H1N1 (**Figure 3.4 B**). Both H7N3 PA-P190S mutants were attenuated in the $_{PR8}$H1N1 genetic background (**Figure 3.4 B**). We then assessed the ability of single mutant PA-184-variant segments to function in the polymerase activity assay. Interestingly, the mutation $_{PR8}$H1N1$_{PA}$ S184N reduced polymerase function by 50% and H7N3$_{PA}$ N184S had the opposite effect of generating greater than WT-$_{PR8}$H1N1 activity (Figure 3. 4 C). A silent mutation of H7N3$_{PA}$ N184N did not dramatically alter polymerase activity, and H7N3$_{PA}$ P190S had no activity restoration effect (**Figure 3.4 C**).

We then used mutant segments bearing H7N3$_{PA}$ 184-variants in the head-to-head competition system and assessed relative incorporation compared to the wild-type PR8-H1N1 segment. Interestingly, the N184S mutants drastically increased the H7N3 segment's incorporation while P190S and Q400P in the c-terminal region had little effect on selection. Therefore, we conclude selection of homologous PA segments in competitive reverse genetics assays in human cells may be due to restriction of PA-184 residues and attenuated polymerase activity.

Genome constellation present during segment selection impacts the outcome of competitive experiments.

IAV genome segments have previously been hypothesized to interact with one and other during assembly although the process through which this occurs remains unclear. Additionally, myriad viral-viral and viral-host protein interactions may have profound effects on properties of

each newly generated reassortant virus. Examining the selection outcomes of Segment 3 in the H7N3 background we found that H7N3 and H7N9 segments were the predominantly selected strains while $_{PR8}$H1N1 and other human-origin segments were selected infrequently[13]. Given the selection disparities between each background we decided to identify the genetic components of the inherently different selection patterns in the context of whole, replicating virus. To assess which segments of the genome most directly impact Segment 3's selection we again utilized a head-to-head competition approach where in one segment of H7N3 was substituted for the -$_{PR8}$H1N1 segment and the other six segments were maintained. In this scheme we identified all segments, save HA and M as capable of altering the selection ratio of $_{PR8}$H1N1$_{PA}$:H7N3$_{PA}$ (**Figure 3.4 C**). PB2 and NA had the most drastic impact on selection, resulting in nearly equivalent levels of the two segments, though these viruses appears attenuated as the CT-values for each were lower than in a WT-$_{PR8}$H1N1 background. These results suggest that the selection preference of $_{PR8}$H1N1 PA in the $_{PR8}$H1N1 context is likely driven by a required interaction between PA and one or more viral genes. Additionally, while possible, the co-selection of avian PA and PB2 seems unlikely given the replication defect observed in this assay.

3.5 Discussion

Here, we uncovered a previously unappreciated genetic switch that allows certain avian-origin polymerase genes to be incorporated into H1N1 genomes, but prevents others from reassortment. The selection of a single amino acid variant allows increased incorporation of H7N3 PA and this is due in part to increased polymerase activity. If segment selection were unbiased, in the presence of five options, an equal distribution of segment selection would be expected. The partial conservation of protein function between divergent strains likely allows rescue of individual reassortant viruses within a reasonable range activity. However, once past a threshold of packaging efficiency or minimal fitness the resulting reassortant virus will attenuated or non-functional. Alternatively, differences in protein-protein interactions may lead to the altered growth of some viruses following reassortment and unequally distributed progeny virus populations[21]. We found that in the case of H7N3 PA, this restriction was not due to previously characterized genome packaging signals, nor was it due to inherent viral replication defects in the highly permissive MDCK cell line. Rather, we noted that in the context of a $_{PR8}$H1N1 genetic background, a single amino acid was sufficient to increase polymerase activity and the likelihood of H7N3 PA segment incorporation into new virions.

The residue we identified, PA-184, has previously been described as an epistatic regulator of reassortant genome polymerase activity in human cells. In the context of human H3N2 polymerase, insertion of H1N1 PA yields a functionally dead complex, likely explaining the recent history of co-assorting of PB2 and PA in human isolates[11,12,22]. Interestingly, this PA residue in the context of H3N2 depends on an additional host-restriction residue in PB2 – the widely studied PB2 627 residue as well as the identity of PA-383 in the same protein[12]. In a

crystal structure of the polymerase complex these amino acids reside together spatially within the vRNP complex to enhance association with a host factor required for replication or efficient transcription initiation[23,24]. Additionally, the PA-184 position is required for interaction with the host factor MCM2, important in processivity of transcription[25,26]. Finally, PA-184 resides near the second required nuclear localization signal of PA (186-247)[27]. Perhaps alteration of the biochemical properties of PA-184 in the context of some, but not other polymerase complexes alters the ability of either vRNPs or PA protein itself to be translocated to the nucleus. Together, our results indicate that the interaction of viral proteins with both host and viral components determines the outcome of reassortment following co-infection and polymerase activity, along with similar tropism and efficient genome packaging, are required for segment co-segregation in newly formed virus particles.

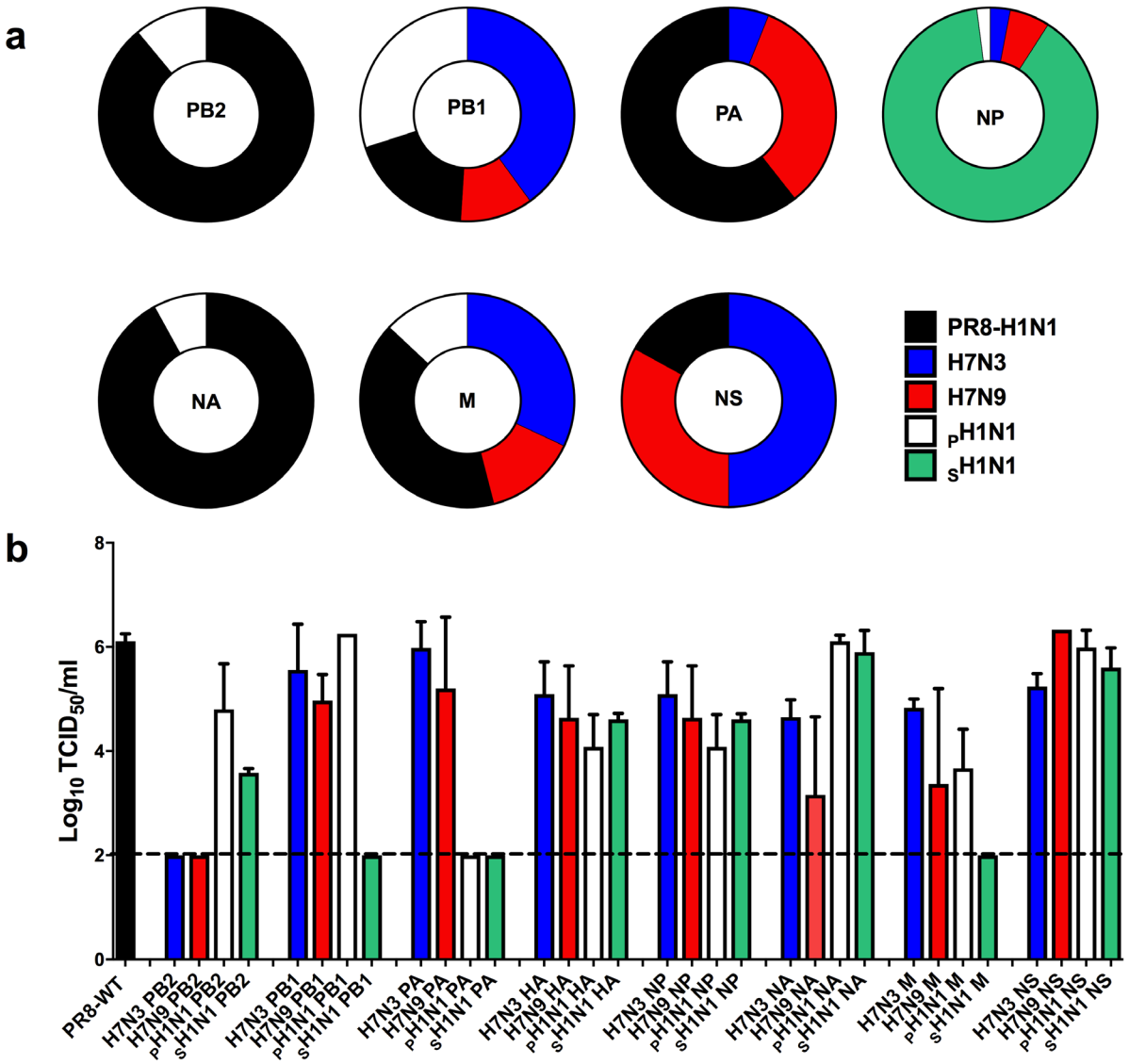


Figure 3.1. Reassortment potential of an H1N1 virus assessed by competitive transfection and individual reverse genetic rescue. A. Percentages of each reassortant virus isolated from culture supernatant of competitive co-transfection assay and limiting dilution assay. Genotyping was conducted by restriction fragment polymorphism length or Sanger Sequencing. B. Rescue efficiency of single gene reassortant viruses. Each reassortant virus was rescued twice independently and titered by $TCID_{50}$ assay (Geometric Mean +/- S.E.M.).

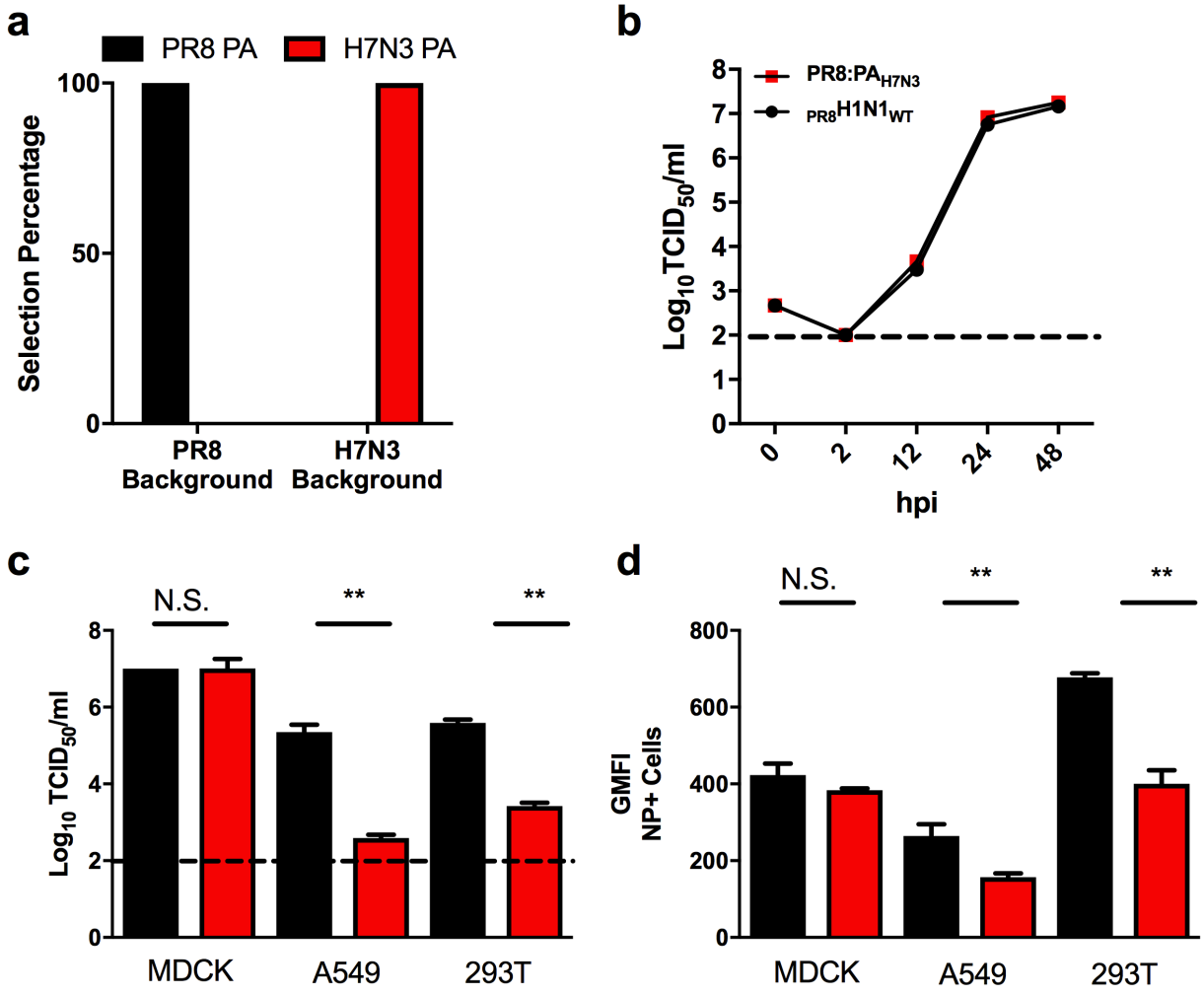

a
PR8 PA
H7N3 PA
Selection Percentage
100
50
0
PR8 Background
H7N3 Background
b
PR8:PA$_{H7N3}$
PR8H1N1$_{WT}$
Log$_{10}$TCID$_{50}$/ml
hpi
c
N.S.
**
**
Log$_{10}$ TCID$_{50}$/ml
MDCK
A549
293T
d
N.S.
**
**
GMFI
NP+ Cells
MDCK
A549
293T

Figure 3.2. Genomic context and host cell determines differential segment 3 selection and virus replication. A. Selection percentage determined by limiting dilution assay of head-to-head selection experiments in the context of either seven PR8 or seven H7N3 genome segments (n=3 experiments). **B.** MDCK infection time course. $2x10^5$ cells were plated in 24-well plates and infected with 250 $TCID_{50}$ of the indicated virus (Geometric mean +/- S.E.M., performed three times in duplicate). **C.** Single cycle replication of PR8 7+1 PA viruses in the indicated cell type. $2x10^5$ cells were plated in 24-well plates and infected at MOI 5 for 24 hours in the absence of TPCK-trypsin to prevent re-infection. Black bars represent WT-PR8; red bars represent PR8 7+1 H7N3 PA. (Geometric mean +/- S.E.M., performed three times in duplicate). **D.** Indicated cells were infected for 8 hours and harvested for flow cytometric analysis of nucleoprotein content. $2x10^5$ cells were plated in 24-well plates and infected at MOI 0.3 for 8 hours. Black bars represent WT-PR8; red bars represent PR8 7+1 H7N3 PA (Mean +/- S.E.M., performed three times in duplicate; **, $P<0.01$).

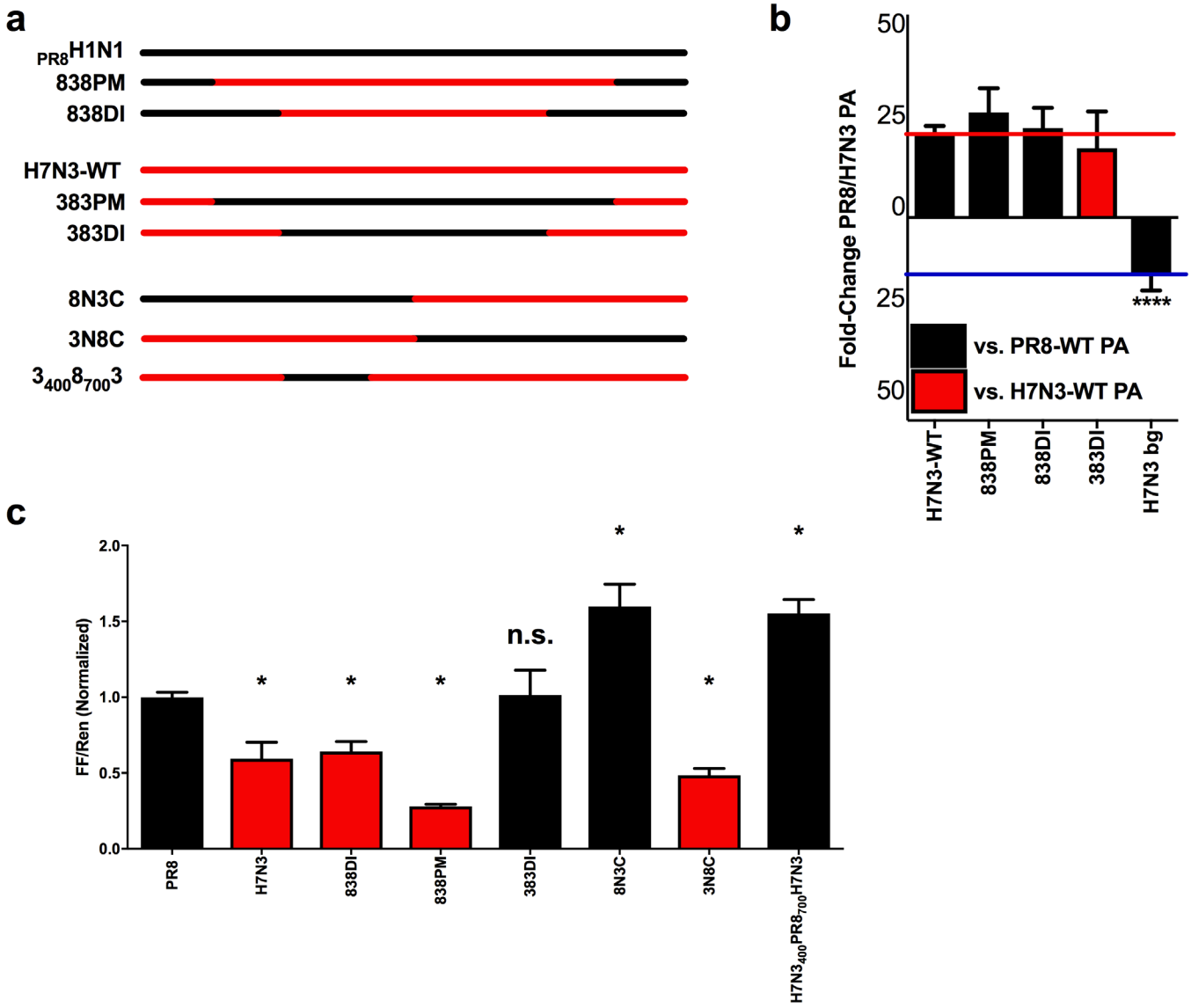

a
PR8H1N1
838PM
838DI
H7N3-WT
383PM
383DI
8N3C
3N8C
3400870033
b
Fold-Change PR8/H7N3 PA
50
25
0
25
50
****
vs. PR8-WT PA
vs. H7N3-WT PA
H7N3-WT
838PM
838DI
383DI
H7N3 bg
c
FF/Ren (Normalized)
2.0
1.5
1.0
0.5
0.0
*
*
*
n.s.
*
*
*
PR8
H7N3
838DI
838PM
383DI
8N3C
3N8C
H7N3400PR8700H7N3

Figure 3.3. The central region of segment 3 determines competitive selection and augments polymerase activity in human cells. A. Schematic of constructs used to assess the location in segment 3 that confers selection advantage of PR8-PA over H7N3-PA. Black regions indicate the region of Segment 3 that is derived from PR8-PA and red indicated H7N3-PA derived regions. **B.** Relative abundance of primarily PR8-PA or H7N3-PA segment in culture supernatant following competitive co-transfection and single passage on MDCK cells. Fold-change was determined with strain-specific TaqMan qPCR probes (N=4-6, mean +/- S.E.M.). **C.** Polymerase activity of reassortant polymerase complexes in 293T cells. All values compared to PR8-PA wild-type segment (N=4-6. mean +/- S.E.M.). All statistical comparisons made by one-way ANOVA corrected for multiple comparisons relative to PR8-PA WT in the context of the PR8-H1N1 genome constelation (****, $P<0.001$; *, $P<0.05$).

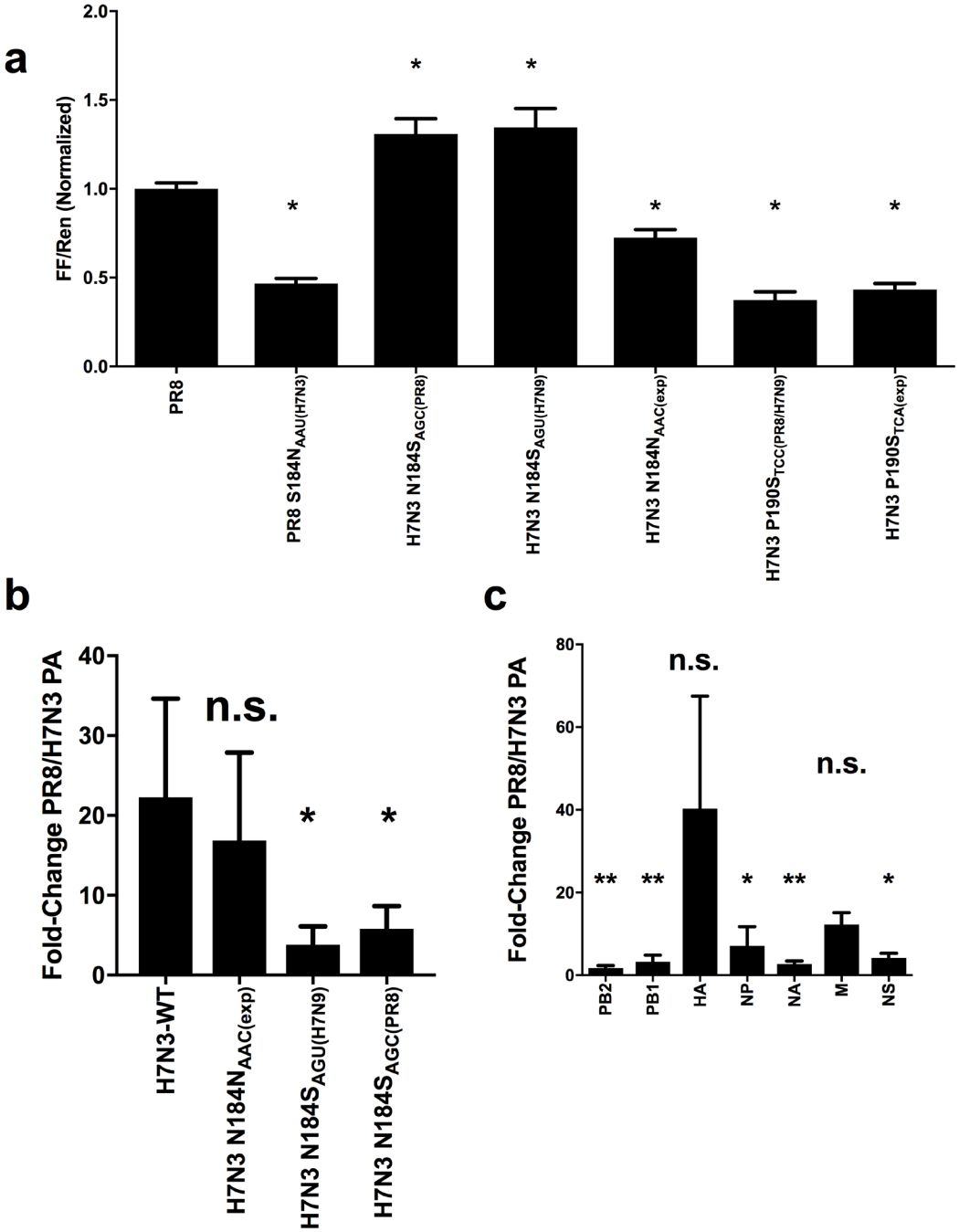

a
FF/Ren (Normalized)
2.0
1.5
1.0
0.5
0.0
*
*
*
*
*
*
PR8
PR8 S184N$_{AAU(H7N3)}$
H7N3 N184S$_{AGC(PR8)}$
H7N3 N184S$_{AGU(H7N9)}$
H7N3 N184N$_{AAC(exp)}$
H7N3 P190S$_{TCC(PR8/H7N9)}$
H7N3 P190S$_{TCA(exp)}$
b
Fold-Change PR8/H7N3 PA
40
30
20
10
0
n.s.
*
*
H7N3-WT
H7N3 N184N$_{AAC(exp)}$
H7N3 N184S$_{AGU(H7N9)}$
H7N3 N184S$_{AGC(PR8)}$
c
Fold-Change PR8/H7N3 PA
80
60
40
20
0
n.s.
n.s.
**
**
*
**
*
PB2
PB1
HA
NP
NA
M
NS

Figure 3.4. Identity of PA amino acid 184 and genomic context determines polymerase function and reassortment potential in human cells. **A** Polymerase activity of single amino acid mutant PA polymerase complexes in 293T cells. All values compared to PR8-PA wild-type segment (N=4-6. mean +/- S.E.M.). All statistical comparisons made by one-way ANOVA corrected for multiple comparisons relative to PR8-PA WT in the context of the PR8-H1N1 genome constelation (**, $P<0.01$; *, $P<0.05$). **B.** Relative abundance of primarily PR8-PA or H7N3-PA mutant segments in culture supernatant following competitive co-transfection and single passage on MDCK cells. Fold-change was determined with strain-specific TaqMan qPCR probes (N=4-6, mean +/- S.E.M.). **C.** Relative abundance of PR8-PA or H7N3-PA segments in culture supernatant following competitive co-transfection and single passage on MDCK cells when a the indicated H7N3 segment was substituted for the same PR8 segment. Fold-change was determined with strain-specific TaqMan qPCR probes (N=4-6, mean +/- S.E.M.; (**, $P<0.01$; *, $P<0.05$).

3.6 References

1. Resa-Infante, P., Jorba, N., Coloma, R. & Ortin, J. The influenza virus RNA synthesis machine: Advances in its structure and function. *RNA Biol.* **8,** 207–215 (2011).

2. Marshall, N., Priyamvada, L., Ende, Z., Steel, J. & Lowen, A. C. Influenza virus reassortment occurs with high frequency in the absence of segment mismatch. *PLoS Pathog.* **9,** e1003421 (2013).

3. Li, C., Hatta, M., Watanabe, S., Neumann, G. & Kawaoka, Y. Compatibility among Polymerase Subunit Proteins Is a Restricting Factor in Reassortment between Equine H7N7 and Human H3N2 Influenza Viruses Compatibility among Polymerase Subunit Proteins Is a Restricting Factor in Reassortment between Equine H7N7 and Hu. (2008). doi:10.1128/JVI.01445-08

4. Gerber, M., Isel, C., Moules, V. & Marquet, R. Selective packaging of the influenza A genome and consequences for genetic reassortment. *Trends Microbiol.* **22,** 446–455 (2014).

5. Wu, A. *et al.* Sequential reassortments underlie diverse influenza H7N9 genotypes in China. *Cell Host Microbe* **14,** 446–452 (2013).

6. van de Sandt, C. E., Kreijtz, J. H. C. M. & Rimmelzwaan, G. F. Evasion of influenza a viruses from innate and adaptive immune responses. *Viruses* **4,** 1438–76 (2012).

7. Nelson, M. I. *et al.* Multiple reassortment events in the evolutionary history of H1N1 influenza A virus since 1918. *PLoS Pathog.* **4,** e1000012 (2008).

8. Watanabe, T. *et al.* Circulating avian influenza viruses closely related to the 1918 virus have pandemic potential. *Cell Host Microbe* **15,** 692–705 (2014).

9. Itoh, Y. *et al.* In vitro and in vivo characterization of new swine-origin H1N1 influenza viruses. *Nature* **460,** 1021–1025 (2009).

10. Nelson, M. I., Vincent, A. L., Kitikoon, P., Holmes, E. C. & Gramer, M. R. Evolution of novel reassortant A/H3N2 influenza viruses in North American swine and humans, 2009-2011. *J. Virol.* **86,** 8872–8 (2012).

11. Phipps, K. L. *et al.* Seasonal H3N2 and 2009 pandemic H1N1 influenza A viruses reassort efficiently but produce attenuated progeny. *J. Virol.* JVI.00830-17 (2017). doi:10.1128/JVI.00830-17

12. Hara, K., Nakazono, Y., Kashiwagi, T., Hamada, N. & Watanabe, H. Co-incorporation of the PB2 and PA polymerase subunits from human H3N2 influenza virus is a critical determinant of the replication of reassortant ribonucleoprotein complexes. *J. Gen. Virol.* **94,** 2406–2416 (2013).

13. Williams, G. D., Pinto, A. K., Doll, B. & Boon, A. C. M. A North American H7N3 influenza virus supports reassortment with 2009 pandemic H1N1 and induces disease in mice without prior adaptation. *J. Virol.* **90,** JVI.02761-15 (2016).

14. DesRochers, B. L. *et al.* Residues in the PB2 and PA genes contribute to the pathogenicity of avian H7N3 influenza A virus in DBA/2 mice. *Virology* **494,** 89–99 (2016).

15. Reed, L. J. & Muench, H. A Simple Method of Estimating Fifty Per Cent Endpoints. *Am.*

J. Hygeine **27,** 285–297 (1938).

16. Marsh, G. a, Hatami, R. & Palese, P. Specific residues of the influenza A virus hemagglutinin viral RNA are important for efficient packaging into budding virions. *J. Virol.* **81,** 9727–36 (2007).

17. Marsh, G. a, Rabadán, R., Levine, A. J. & Palese, P. Highly conserved regions of influenza a virus polymerase gene segments are critical for efficient viral RNA packaging. *J. Virol.* **82,** 2295–304 (2008).

18. Cobbin, J. C. a *et al.* Influenza Virus PB1 and NA Gene Segments can Co-segregate during Vaccine Reassortment Driven by Interactions in the PB1 Coding Region. *J. Virol.* (2014). doi:10.1128/JVI.01022-14

19. Gavazzi, C. *et al.* A functional sequence-specific interaction between influenza A virus genomic RNA segments. *Proc. Natl. Acad. Sci. U. S. A.* **110,** 16604–9 (2013).

20. Saira, K. *et al.* Sequence analysis of in vivo defective interfering-like RNA of influenza A H1N1 pandemic virus. *J. Virol.* **87,** 8064–74 (2013).

21. Hudjetz, B. & Gabriel, G. Human-like PB2 627K influenza virus polymerase activity is regulated by importin-α1 and -α7. *PLoS Pathog.* **8,** e1002488 (2012).

22. Hemerka, J. N. *et al.* Detection and characterization of influenza A virus PA-PB2 interaction through a bimolecular fluorescence complementation assay. *J. Virol.* **83,** 3944–55 (2009).

23. Cusack, S. & Ruigrok, R. W. H. Influenza A Virus Polymerase : Structural Insights into

Replication and Host Adaptation Mechanisms *. **285,** 28411–28417 (2010).

24. Pflug, A., Guilligay, D., Reich, S. & Cusack, S. Structure of influenza A polymerase bound to the viral RNA promoter. *Nature* **516,** 355–60 (2014).

25. Kawaguchi, A. & Nagata, K. De novo replication of the influenza virus RNA genome is regulated by DNA replicative helicase, MCM. *Embo J* **26,** 4566–4575 (2007).

26. Wang, Q. *et al.* Host cell interactome of PA protein of H5N1 influenza A virus in chicken cells. *J. Proteomics* **136,** 48–54 (2016).

27. Liang, Y., Danzy, S., Dao, L. D., Parslow, T. G. & Liang, Y. Mutational analyses of the influenza A virus polymerase subunit PA reveal distinct functions related and unrelated to RNA polymerase activity. *PLoS One* **7,** (2012).

Chapter 4:

Nucleotide resolution mapping of influenza A virus nucleoprotein-RNA interactions reveals the landscape of viral RNA features required for replication

This chapter has been submitted in its entirety as a research manuscript that is currently in revision for publication at Nature Communications.

4.1 Abstract

Influenza A virus nucleoprotein (NP) associates with all eight negative-sense genomic RNA segments during virus replication. Although the positioning of protein components within viral ribonucleoprotein complexes (RNPs) is well defined, the native interaction of NP with the viral RNA (vRNA) and the mechanisms by which a complete multi-segment genome assembles remain elusive. Here, we applied photoactivatable ribonucleoside enhanced crosslinking and immunoprecipitation (PAR-CLIP) to assess the native-state of NP-vRNA interactions in infected human cells. NP bound short fragments of RNA (~12 nucleotides) non-uniformly and without apparent sequence specificity. Moreover, NP binding was reduced at specific locations within the viral genome including regions previously identified as required for viral genome segment packaging. Synonymous, structural mutations in these low-NP binding regions impacted genome packaging and resulted in virus attenuation, whereas mutagenesis of NP-bound regions had no effect. Finally, we demonstrate that the sequence conservation of low-NP binding regions is required in multiple genome segments for propagation of diverse mammalian and avian IAV in host cells.

4.2 Introduction

Influenza A virus (IAV) possesses a segmented, negative-sense RNA genome that is bound by the viral nucleoprotein (NP) throughout replication. The structure of the protein components within the viral ribonucleoprotein complex (RNP) was recently solved, but the structure of viral RNA and its interaction with NP have not been elucidated[1,2]. NP is thought to coat viral RNA (vRNA) uniformly in cells and virus particles; however uniform coating likely would preclude the possibility for RNA structure formation in RNPs. Numerous virus families utilize structured RNA elements for specific biological processes throughout infection, including genome packaging[3,4]. For example, viral RNA elements are required for efficient replication, mRNA splicing, and genome packaging of IAV[5–7]. Structure formation has been demonstrated with *in vitro* folded IAV vRNA and engineered genome segments but the structural constraints imposed by nucleoprotein on vRNA generated during infection is not known[8]. Elucidation of the physiological interaction between NP and viral genomic RNA may provide novel insights into how IAV is capable of coordinating its lifecycle. Thus, we set out to determine the *in vivo* landscape of NP-vRNA interactions.

Infection and complete replication of IAV requires delivery of all eight genome segments into a recipient cell. All IAV segments require packaging signals derived from the termini on each segment[9,10]. Interaction between vRNAs has been demonstrated *in vitro* and disruption of packaging signals or interacting segment regions attenuated virus replication at the stage of genome packaging[11–13]. In many cases, mutation of a single segment leads to a significant decrease in the packaging efficiency of other segments[5,14]. Additionally, viral particles largely package only one copy of each genome segment[15–17]. Together, these results suggest that genome

segments function as a multipartite, cooperatively packaged entity, possibly potentiated by segment-segment interactions, rather than a stochastically generated particle[18,19]. In this study, we set out to determine how IAV NP interacts with vRNA during infection in cells.

4.3 Results

Nucleotide resolution mapping of NP-vRNA interactions

Photoactivatable ribonucleoside enhanced crosslinking and immunoprecipitation (PAR-CLIP) coupled to next generation sequencing was used to resolve the interaction between the negative-sense RNA genome of IAV and NP during infection of human 293T cells[20]. We infected human cells with WT-PR8 virus for 16 h in the presence of 4-thiouridine (4-SU) to enhance crosslinking of NP-RNA complexes and then generated Illumina 1x50 sequencing libraries of the NP-bound RNA (**Figure 4.1 A**). The impact of 4-SU on viral replication was assessed in 293T cells. WT-PR8 replicated to equivalent titers 24 h post-infection (hpi) in mock- or 4-SU-treated (100 μM) cells (**Figure 4.1 B**). Additionally, NP localization after 4-SU treatment was assessed by confocal microscopy at 16 hpi, and no alteration was observed at this time point (**Figure 4.1 B**). These results suggest 4-SU treatment does not substantially impact IAV protein production or replication in human cells.

To determine the sensitivity and specificity of the PAR-CLIP assay, we performed Western blotting analysis for IAV NP and cellular ß-actin on input lysate and immunoprecipitated proteins (**Figure 4.1 C,** bottom). Compared to immunoprecipitations performed without antibody or a control anti-HA antibody, immunoprecipitation with a monoclonal antibody (MAb) against IAV NP produced a specific band. UV-exposure of infected cells in the presence of 4-SU enabled greater recovery of NP-RNA complexes (**Figure 4.1 C**, top). The protein purity in the immunoprecipitate was verified by silver stain or Western blotting with an anti-IAV polyclonal serum (**Figure 4.2 A, B**) or IP with a different MAb to IAV-NP (data not shown).

Influenza A nucleoprotein minimally binds 11-14 nucleotides per monomer at non-uniform intervals

PAR-CLIP identified both human and virus-derived RNAs that interacted with NP. The procedure enriched for IAV RNA sequences relative to RNA-seq libraries (**Fig 4.1 D**), and the majority of the viral RNA sequences were derived from the negative-strand vRNA (**Figure 4.3**). The average length of the NP-bound vRNA was 12 nucleotides (range 11-14, **Figure 4.1 E**) and the distance between high confidence binding events, identified as crosslinking-induced U-to-C transitions in vRNA sequences, was approximately 25 nucleotides (**Figure 4.1 F**).

We then compared PAR-CLIP and RNA-seq libraries to identify contiguous regions of vRNA that are significantly under- and overrepresented among the NP-bound RNAs. Using the criteria of >3-fold difference, $Q<0.01$, and ≥18 nucleotides long, we identified 25 regions in the viral genome that were low in NP binding relative to RNA-seq and 17 regions that met two of these three criteria that we did not investigate further (**Figure 4.4 A-H**, **Table 4.1**). Moreover, four high-NP binding regions that bound RNA were overrepresented relative to the control RNA-seq sets (**Table 4.1**). The low-NP binding regions together represent approximately ten percent of the viral genome and do not differ in base composition from the remainder of the genome (**Figure 4.1 G**). Analysis of low-NP binding regions revealed that RNA secondary structures might form in the absence of NP binding in some of these regions. Thus, NP binding might be affected by local secondary structures in the genome of IAV.

Nucleotide identity of low-NP binding regions in Segment 5 vRNA elements is important for viral replication

To assess the significance of the over- and underrepresented RNA regions, we selected six regions in segment 5 that have variable NP-binding profiles (**Figure 4.5 A)**. The lack of alternative reading frames and splicing and relative conservation of segment 5 compared to other IAV genome segments made this segment more suitable for extensive characterization. Of these, two regions were underrepresented (low-NP binding) in PAR-CLIP data sets (NP_{22-68} and $NP_{1410-1495}$). The $NP_{1410-1495}$ region contains a previously hypothesized vRNA pseudoknot[5]. Additional regions of either NP intermediate binding ($NP_{456-480}$, $NP_{145-175}$, and $NP_{1058-1081}$) or highly bound vRNA ($NP_{584-608}$) also were also included in the following studies. Computational prediction of RNA structures using RNAfold in low-NP binding regions guided further mutational analysis, and these regions were mutagenized to either disrupt ($NP_{22-68:A}$ and $NP_{1410-1495}$) or maintain ($NP_{22-68:B}$) secondary and tertiary structure formation (**Figure 4.9)**. Mutant viruses bearing 2-7 synonymous structural nucleotide changes in these regions were generated and assessed for ability to replicate *in vitro* and *in vivo*.

WT and mutant viruses were subjected to a focus forming assay in MDCK cells to determine replication competence *in vitro* (**Figure 4.5 B**). Mutations disrupting the segment 5 vRNA pseudoknot ($NP_{1410-1495}$) formation and those predicted to destabilize a stem-loop structure ($NP_{22-68:A}$) in the 3' region of the vRNA segment resulted in reduction of focus area, as a measure of multi-cycle replication and spread (**Figure 4.5 B**). Conversely, mutation of intermediate- or highly-bound vRNA regions did not alter the focus area. Synonymous mutations in NP_{22-68}, predicted to maintain the secondary structure ($NP_{22-68:B}$), also did not affect focus size (**Figure 4.5 B**). Multi-cycle replication assays in MDCK cells of select mutant viruses confirmed these results (**Figure 4.5 C**). Finally, mice were inoculated with 10^3 $TCID_{50}$ of each virus, and the presence of infectious virus in the lung was assessed 48 h later. WT-PR8 replicated to high titers

at this time point whereas destabilizing mutations in low-NP binding regions resulted in attenuation (**Figure 4.5 D**). Synonymous structural mutations in intermediate or highly bound regions had no effect on virus infection *in vivo* (**Figure 4.5 D**). These results collectively suggest that structural features of the low-NP binding regions are important for IAV replication.

To determine the cause of attenuated replication of low-NP binding mutant viruses, we evaluated effects on specific stages in the IAV lifecycle. All viruses displayed similar cytoplasmic distribution of NP 8 hpi in MDCK cells when assessed by indirect immunofluorescence and confocal microscopy (**Figure 4.6 A**). The ability of the NP proteins from the mutant viruses to transcribe and replicate a firefly luciferase reporter genome segment was tested in human cells; all viruses displayed equivalent reporter activity (**Figure 4.6 B**). Similarly, infection of MDCK cells with all viruses generated equivalent amounts of viral antigen (NP) 8 hpi when measured by flow cytometry (**Figure 4.6 C**). Therefore, the synonymous mutations introduced within NP segment did not alter NP expression or its activity by all measurable outcomes.

We next assessed the virus particle to infectious unit ratio of WT and mutant PR8 viruses. The infectious titer ($TCID_{50}$/ml) of $NP_{1410\text{-}1495}$ and $NP_{22\text{-}68:A}$ mutant viruses was significantly lower ($P<0.05$) compared to all other viruses when normalized to HA-units, indicating a greater production of non-infectious particles (**Figure 4.6 D**). The ability of WT and mutant viruses to package all eight genome segments was assessed using a population level measure of relative vRNA segment abundance in purified viral particles[21]. RNA from WT and mutant IAVs was subjected to RT-qPCR and the abundance of segments was normalized to segment 7 (**Figure 4.6 E**). Mutation of either the predicted pseudoknot ($NP_{1410\text{-}1495}$) or stem-loop structure ($NP_{22\text{-}68:A}$) resulted in aberrant genome constellation stoichiometry. Consistent with the viral replication

assays, mutation of intermediate- or highly-bound regions or mutations that maintained the predicted 3' stem-loop structure ($NP_{22\text{-}68:B}$) result in unaltered ratios of genome segments in all but one segment-specific instance (**Figure 4.6 and Figure 4.7 A**). These results suggest that RNA structure-destabilizing mutations, in regions of vRNA poorly bound by NP, impact replication at the stage of coordinated genome packaging. Finally, we determined the co-expression of NP and matrix (M) proteins in singly-infected cells (**Figure 4.6 F**, multiplicity of infection (MOI) of 0.05)[22]. In agreement with the RT-qPCR data suggesting a packaging defect, a lower percentage of $NP_{1410\text{-}1495}$ and $NP_{22\text{-}68:A}$ mutant virus infected cells co-expressed both proteins compared to cells infected with WT or other mutant viruses. Notably, high MOI infection with multiple defective particles simultaneously infecting the same cell is able to overcome this defect (**Figure 4.7 B**).

Mutation of low-NP binding regions in additional genome segments results in attenuated viral replication.

Low-NP binding regions in segments 1, 2, and 8 also were evaluated for their contribution to viral replication by mutagenesis designed to disrupt predicted secondary and tertiary RNA structures (**Figure 4.8 A–C**, **Table 4.1**). Nucleotide substitutions that alter the predicted RNA secondary structure but have no impact on the protein, i.e. synonymous structural mutation, in $PB2_{1823\text{-}1944}$, $PB1_{497\text{-}561}$ or $NS_{23\text{-}86}$ reduced the focus area of mutant viruses in MDCK cells (**Figure 4.8 D**). Similar to the attenuation of segment 5 viruses, the $PB2_{1823\text{-}1944}$, $PB1_{497\text{-}561}$ or $NS_{23\text{-}86}$ mutant viruses generated fewer $TCID_{50}$ particles per HA-unit (**Figure 4.8 E**). Segment 1 and 2 mutant viruses did not display altered reporter activity (**4.10 A**). Additionally, virus replication in mice was diminished 48 hpi compared to WT-PR8 (**Figure 4.8 F**). Assessment of vRNA abundance in virus particles revealed that segment 2 ($PB1_{497\text{-}561}$) mutant viruses packaged

reduced levels of segment 6 vRNA (NA) (**Figure 4.8 G**). Manipulation of segment 8 vRNA sequence ($NS_{23\text{-}86}$) resulted in decreased packaging of segments 3, 4, 5, and 7 relative to segment 1 (**Figure 4.8 G**). Unexpectedly, the $PB2_{1823\text{-}1944}$ virus did not show altered segment abundance despite the largest *in vivo* replication defect (**Figure 4.8 F and Figure 4.9 C**) and difference in virus particle to infectious unit ratio. Two additional viruses with mutations in this region ($PB2_{1823\text{-}1944}$) also had reduced focus area despite no differences in segment packaging (**Figure 4.10 B, C**). Together, these data suggest that vRNA sequences that are low in NP binding help to coordinate packaging of a full complement of eight vRNA segments and changes to the predicted RNA structures in these regions results in virus attenuation.

Mutation of IAV-PR8 low-NP binding nucleotides attenuates replication of IAV-H7N3 and IAV-pH1N1

We evaluated the contribution of two low-NP binding regions ($PB2_{1823\text{-}1944}$ and $NP_{1410\text{-}1495}$), identified in WT-PR8, to the replication of a North American avian IAV: A/shorebird/Delaware/22/2006 (IAV-H7N3) and the 2009 pandemic H1N1 strain A/California/04/2009 (H1N1) (IAV-pH1N1). The PB2 and NP gene-segment of IAV-PR8 and IAV-H7N3 are divergent and representative of mammalian and avian viruses, respectively (**Figure 4.11 A, B**). The NP gene-segment of IAV-pH1N1 is closely related to IAV-PR8 (**Figure 4.11 A**), while the PB2 gene-segment is more closely related to avian viruses (**Figure 4.11 B**). Isogenic H7N3 and pH1N1 viruses, bearing synonymous mutations in the $NP_{1410\text{-}1495}$ or $PB2_{1823\text{-}1944}$ regions predicted to disrupt local RNA structure, were evaluated for virus replication *in vitro*. Synonymous mutations that are predicted to impact RNA secondary structure in these regions of IAV-H7N3 and IAV-pH1N1 attenuated viral replication, as measured by focus area (**Figure 4.11**

C, D). These results indicate that the RNA features, identified by PAR-CLIP in WT-PR8, are applicable to diverse avian and human strains of IAV.

4.3 Materials and methods

Cells

Madin-Darby Canine Kidney (MDCK) cells were maintained in minimal essential medium (MEM) with 5% fetal bovine serum (FBS, Biowest), MEM-vitamins (Gibco), L-glutamine (Gibco), and penicillin-streptomycin (Gibco). Human embryonic kidney cells (293T) were maintained in Opti-MEM (Life Technologies) with 10% FBS, L-glutamine, and penicillin-streptomycin.

Viruses

Eight bidirectional pHW2000 plasmids containing cDNA for A/Puerto Rico/08/1934 (H1N1), A/California/04/2009 (H1N1), or A/shorebird/Delaware/22/2006 (H7N3) were obtained from St. Jude Children's Research Hospital[23]. Viruses were generated by transfection of all eight plasmids into co-cultures of 293T and MDCK cells (1 μg per plasmid) with polyethylenimine (8 μg total). Wild-type (herein IAV-PR8 (WT-PR8), IAV-pH1N1 (WT-pH1N1) or IAV-H7N3 (WT-H7N3) and mutant viruses were generated in the same manner with the exception of substituting individual plasmids harboring mutagenized DNA for the single indicated wild-type plasmid. The next day, the transfection mixture was removed and replaced with Opti-MEM containing MEM-vitamins, L-glutamine and penicillin-streptomycin. Forty-eight hours post-transfection, an additional 1 ml of the same media containing 1 μg ml^{-1} TPCK-Trypsin (Worthington) was added to the co-culture. Seventy-two hours after addition of TPCK-Trypsin, culture supernatant was harvested and clarified by centrifugation (5 min, 350 x g). Viral stocks were generated by infection of MDCK cells in a T75 flask. Cells were washed once with PBS, and 200 μL of transfection supernatant was mixed with 25 ml of infection media (M0.1B)

composed of MEM containing vitamins, L-glutamine, penicillin-streptomycin, 0.1% Bovine Serum Albumin (Gibco) and 1 μg ml^{-1} TPCK-Trypsin for 48 hours. Stocks were aliquoted and stored at -80°C until use, and all studies were conducted with passage 1 stocks following verification of mutant sequence identity. All viruses were generated at least twice independently. Sequences for primers utilized during mutagenesis of IAV plasmids are available upon request.

IAV Nucleoprotein PAR-CLIP and RNA-seq Library Generation

To identify interactions between viral RNA and IAV nucleoprotein, we adapted the protocol for PAR-CLIP[20] coupled to Next Generation Sequencing to discern nucleotide resolution maps of protein-RNA interaction across the IAV genome. Confluent 293T cells were infected at an MOI of 1 for 16 h in the presence of 100 μM 4-thiouridine (4-SU) and then cross-linked with ultraviolet light (310nM, 500,000 μJ total energy). PAR-CLIP was performed essentially as described before[20] with the exception of antibodies used to immunoprecipitate NP-RNA complexes: monoclonal antibodies HB65 (ATCC) or MAb8258 (EMD Millipore). Briefly, protein-RNA complexes were separated on 4-12% SDS-PAGE gels and transferred to nitrocellulose membranes. ^{32}P-labeled RNA was identified by autoradiography, excised, and extracted following proteinase K (New England Biolabs (NEB)) digestion. To account for potential bias introduced during adaptor ligation and sequencing, we also generated RNA-seq libraries using RNA extracted from uncross-linked cell lysates of influenza infected 293T cells from which RNA was isolated by TRIzol (Invitrogen) extraction, fragmented by $Mg2^{+}$ at 95°C for 12 min (NEB, Magnesium RNA Fragmentation Module). ^{32}P-labeling of this RNA was performed in solution with T4 PNK (NEB), and size-selection of 10-100 nt fragments by Urea-

Page (15%) was performed. In both cases, total RNA was precipitated with ethanol, and used to prepare Illumina sequencing libraries. The adaptor sequence contains NNN-degenerate nucleotides in addition to a sequence used for demultiplexing to facilitate collapsing of redundantly generated PCR products and ascertain the frequency of individual NP-RNA interaction events. Following isolation of adaptor ligated and radiolabeled RNA, all subsequent library generation steps were identical. Independent PAR-CLIP or RNA-seq libraries were pooled and cDNA synthesized using a primer complementary to the 3' adaptor (SuperScript III, Invitrogen). Libraries were amplified by PCR for 9-15 cycles using Phusion DNA polymerase (Phusion HF Mastermix, Thermo) and primers annealing to the 5' and 3' termini of the DNA that enable flow-cell binding. Size selection of libraries was conducted by extraction of amplicons from 6% urea gel electrophoresis, and then libraries were precipitated. Molarity of libraries was determined by qPCR (NEBNext, New England Biosciences) and size by Bioanalyzer (Agilent). Illumina sequencing on the HiSeq 2500 instrument (1x50bp reads) was performed by the Genome Technology Access Center at Washington University in St. Louis.

Next Generation Sequencing

All NGS data was analyzed on the Washington University in Saint Louis School of Medicine's McDonnell Genome Institute (MGI) cluster using publically available analysis programs (FastX toolkit, Bowtie, and SAMtools) and in-house scripts[20]. PAR-CLIP and RNA-seq data were generated from four independent experiments. For PAR-CLIP and RNA-seq libraries we used an analysis pipeline that collapsed unique barcoded reads, removed adaptors, and aligned them to the viral genome. Strandedness was determined post-alignment and all

subsequent analysis was performed on negative-sense viral genomic RNA (vRNA). For reads that mapped to the viral RNAs, we normalized the number of reads per nucleotide to the total number of reads per genome segment to yield a normalized coverage ratio for both PAR-CLIP and RNA-seq libraries. We then compared the normalized coverage for each preparation at nucleotide resolution using an unpaired t-test with False Discovery Rate (FDR) correction (Benjamini-Hochberg)[24]. vRNA nucleotides with a FDR minima $Q < 0.01$ and fold-change >3 were identified and these regions were extended to the final nucleotide of each region where $Q \leq 0.1$. These areas represent low-NP or high-NP binding regions of interest (ROI). ROI ≥ 18nt were subjected to RNA structure analysis with mFold, RNAfold, and vsFold5 to determine computationally the theoretical minimal free energy (ΔG) of each region as well as the potential secondary or tertiary RNA structure formation[25,26]. Nucleotide composition of ROIs was determined by calculating the percent of A, G, C, or U in each region and comparing them to the percent of each nucleotide in the IAV genome. The approximate periodicity of NP-vRNA interactions was determined by assessing the nucleotide distance between U-to-C transition sites in vRNA. UV-crosslinking of 4-SU to IAV NP introduces a U-to-C transition upon reverse transcription of vRNA due to the residual covalent protein adduct left on the thiouridine base after proteinase K digestion and RNA recovery. PAR-CLIP data sets were compared to RNA-seq data sets at each nucleotide position on a per-segment basis and positions using unpaired t-test with FDR correction as above with a threshold of $Q = 0.05$. Sites of statistically enriched transition in PAR-CLIP data sets were tabulated and the sequence distance between these positions was calculated. Transition events with a step-size of 1, i.e. sequential nucleotides both with high transition rate, were excluded from analysis because they likely represent single NP-

RNA interactions of adjacent 4-SU nucleotides in vRNA. Transition rate for four PAR-CLIP libraries were averaged prior to distance calculation.

Selection, Design, and Generation of IAV Containing Altered RNA structures

Mutant selection was performed after analysis of open reading frames to assess potential nucleotide degeneracy while retaining coding sequence. Subsequently, we selected five underrepresented ROIs in four segments (1, 2, 5, and 8) for functional assessment of manipulating RNA structure on viral replication. We also identified four regions in segment 5, amenable to extensive silent mutagenesis that were either highly bound or represented at the same frequency in PAR-CLIP and RNA-seq data sets. We introduced between 2 and 7 synonymous mutations simultaneously *in silico* to the identified regions and reassessed structural stability or pseudoknot formation using the same structure prediction programs. We selected variant codon combinations that would disrupt or maintain the predicted vRNA structure but not change the encoded amino acid, alter codon usage, or disrupt alternative reading frames or splicing events. Once predicted destabilizing mutations were identified, mutant viruses were generated. Mutations were introduced into pHW2000 bidirectional plasmids by inverse PCR with primers including selected mutations and unique ligation sites. PCR products were gel extracted and digested with either BsmbI (New England Biolabs, (NEB)) or AarI (Thermo Fisher Scientific) restriction enzymes and DpnI (NEB) to remove residual parent plasmid. Digested PCR products were PCR purified and ligated using Instant Sticky End Ligase (NEB). Ligation products were transformed into *E. coli*, plated on selective LB Agar. Colonies were selected and grown overnight in LB broth and selective antibiotics. All plasmids were grown in 200 ml LB Broth prior to preparation by Qiagen HiSpeed Endotoxin-Free MaxiPrep and verified by Sanger Sequencing (Genewiz).

Influenza A Virus Focus Forming Assay

Six-well plates of MDCK cells were plated and inoculated with 1 ml of 10-fold serially diluted virus stock for 1 h in M0.1B. After 1h, the inoculum was removed, and replaced with an overlay of 1% Low Melting Point Agarose in MEM supplemented with 1 μg ml^{-1} TPCK-Trypsin. Seventy-two hours post infection cells were fixed in 5% formaldehyde, permeabilized with 0.1% saponin in HBSS, and stained for NP protein with biotinylated anti-NP (mAB8258b, Millipore). Foci were visualized by addition of HRP-conjugated streptavidin and TruBlue streptavidin substrate and imaged on a Biospot reader (Cellular Technology Limited). Individual foci were counted and area was calculated using the Analyze Particles extension within Fiji[27]. The area of the foci formed by mutant viruses was normalized to the average of WT-PR8 foci per experiment. A minimum of 60 foci was analyzed per virus.

Confocal Microscopy

We infected 293T cells in the presence or absence (mock) of 4-SU with WT-PR8 (MOI of 1) for 16 h. Briefly, 293T cells were seeded on glass coverslips coated with poly-D lysine overnight. Cells were infected with WT-PR8 for 16 h then fixed with 4% methanol-free PFA in PBS (pH 7.4), washed with PBS, permeabilized with saponin, then NP staining was performed using anti-NP MAb HB65 and goat-anti-mouse Alexa-488 (Invitrogen). Nuclei were counterstained with DAPI. Coverslips were mounted to slides with ProLong® Diamond Antifade Mountant (Molecular Probes). To determine if viral protein production was impacted by synonymous mutations, we utilized confocal microscopy to visualize the distribution of NP during infection as a marker of viral protein production and a proxy for trafficking of vRNPs. MDCK cells were infected with WT-PR8 and indicated mutant viruses (MOI of 0.2) for 8 h then

fixed and permeabilized as before. NP staining was performed as above using anti IAV-NP MAb HB65 and Goat-anti-Mouse Alexa-488 (Invitrogen). Slides were imaged with Zeiss LSM 880 Confocal Laser Scanning Microscopy with Airyscan and analyzed with Zeiss Zen Black Software performed within the Molecular Microbiology Imaging Facility at Washington University in Saint Louis.

Influenza A virus Reporter Assay

To assess the impact of silent mutations on polymerase complex activity, pHW2000 plasmids encoding WT or mutant PB2, PB1, PA, and NP were utilized in a dual-luciferase reporter assay as previously described[23]. A vRNA-like firefly luciferase reporter plasmid and a Renilla luciferase expression plasmid also were included. Briefly, 293T cells were seeded into 24-well plates and transfected with equal amounts of all six plasmids (500ng DNA total) in Opti-MEM containing TransIT-LT1 (Mirus). Cells were maintained at 37°C for the duration of the experiment. Forty-eight hours later, cells were lysed for analysis of luciferase activities (Promega). Each combination of polymerase proteins (set of plasmids) was examined in duplicate. The relative light units (RLU) of firefly luciferase activity were normalized to the RLU for Renilla luciferase activity within the same sample to account for differences in transfection efficiency between wells and experiments. The activity of each plasmid set containing a mutant segment was normalized to the activity of WT-PR8.

Flow Cytometry

To determine the amount of NP generated at a fixed time during infection, we infected 2 x 10^5 MDCK cells (MOI of 0.2). Eight hours post-infection, cells were harvested and fixed. Intracellular staining of IAV antigen was performed as above using an anti-IAV NP primary

antibody (HB65) and an Alexa-488-conjugate goat-anti-mouse secondary antibody (Invitrogen). The mean fluorescence intensity (MFI) of NP^+ cells was plotted and calculated from two experiments performed in duplicate. Co-expression of NP and M in MDCK cells was determined 16 hpi with an MOI of 5 and 0.05. Infected cells were fixed and stained as above with the additional staining step of utilizing mouse-anti-M primary antibody (M2-1C6) and goat-anti-mouse Alexa-647 secondary. A second intracellular staining step utilizing biotinylated mouse-anti-NP (mAb8258b) and Alexa-488 conjugated streptavidin was then performed. The frequency of co-expression was calculated by determining the number of cells expressing NP or M as well as those expressing both. The percentage of infected cells co-expressing NP and M was calculated by dividing co-expressing cells by all cells expressing one or more viral protein as previously described[22]. Analysis of viral proteins was determined using a flow cytometer and FlowJo software (Tree Star).

Multi-step Replication of Influenza A virus

MDCK cells (2×10^5) were seeded in 24-well plates and inoculated the next day with 200 $TCID_{50}$ of IAV. Cells were washed once with PBS before addition of inoculum in M0.1B and incubated for 1 h at 37°C. Subsequently, the cells were washed twice with PBS and 1 ml of M0.1B supplemented with 1 μg ml^{-1} TPCK-Trypsin was added to each well. Culture supernatants were collected at 24 and 48 hpi and the amount of infectious virus was quantified by titration on MDCK cells.

Titration of Influenza A virus ($TCID_{50}$)

Confluent monolayers of MDCK cells were grown overnight in 96-well tissue culture plates. The next day the cells were washed with PBS and inoculated with ten-fold serial dilutions

of culture supernatant or lung homogenate for one hour in M0.1B at 37°C. After 1 h the inoculum was removed and replaced with M0.1B supplemented with 1μg ml^{-1} TPCK-Trypsin and incubated for 72 hours. Presence of virus was determined by hemagglutination assay using 0.5% turkey red blood cells. $TCID_{50}$ was determined by the Reed-Muench method[28].

Infection of mice with influenza viruses

Male C57BL/6J mice (5 to 6 weeks of age) were bred in-house in a barrier facility at Washington University School of Medicine, St. Louis, MO or purchased from Jackson Laboratories. Mice received food and water *ad libitum* and all experiments were conducted in accordance with rules of the Institutional Animal Care and Use Committee. Mice were anesthetized with isoflurane in an airflow chamber and then inoculated intranasally with 30 μL of sterile PBS containing 1,000 $TCID_{50}$ of WT-PR8 or the indicated mutant virus. Forty-eight hours post-infection, lungs were harvested and homogenized in 1ml M0.1B, cleared by centrifugation at 1,200 x g for 5 minutes, then stored in aliquots at -80°C. Viral titers from lung homogenates were determined by $TCID_{50}$ assay.

Segment Abundance RT-qPCR

MDCK cell derived stocks of WT-PR8 or mutant viruses were clarified by centrifugation at 1,200 x g for 5 min, passed through a 0.22 μM filter, and pelleted on 30% sucrose cushion by ultracentrifugation (Beckman SW32ti swinging bucket rotor, 27K RPM, 4°C for 90min). Pelleted virus particles were directly resuspended in 350 μL TRK lysis buffer, and RNA purified immediately and eluted in 30 μL DEPC H_2O (Total RNA Kit I, Omega). cDNA was synthesized from 5μL of RNA with SSIII Reverse transcriptase (Invitrogen) and a vRNA specific primer[21]. Total cDNA was diluted 1:10,000 and used to quantify each of the 8 genome segments by SYBR

Green qPCR (PowerUp™ SYBR® Green Master Mix) with primer pairs previously published. Relative abundance of each genome segment was calculated as before, except with normalization to Segment 1 or 7 depending on the virus assessed.

Statistical Analysis

All statistical analysis was performed using GraphPad Prism 7.0. For comparison of PAR-CLIP and RNA-seq data sets, we used multiple unpaired t-tests with the Benjamini-Hochberg correction to identify areas in which these sequencing preparations were statistically different from each other ($Q < 0.1$ and $Q < 0.01$). Transition-distance was determined using FDR ($Q < 0.05$). For analysis of the focus forming assay, luciferase assay, and lung viral titers we used one-way ANOVA with multiple comparison corrections (Kruskal-Wallis test). $TCID_{50}$ per HA-unit was analyzed using an unpaired t-test.

4.5 Discussion

We have identified the interaction landscape of IAV NP with viral RNA in the context of infected cells. Our findings indicate that binding of NP to viral RNA is restricted to an average of 12 nucleotides and the distance between two crosslinking sites is 25 nucleotides. These estimates agree with molecular models of NP-RNA interactions and the NP-binding footprint of a related *orthomyxovirus*[1,19,29,30]. Within this model, NP is excluded from consistent interaction with specific regions of vRNA and allows trans-interactions with either other genome segment vRNA or host and virus factors. The interaction between NP and the viral RNA was non-uniform and characterized by regions that were consistently low- or high in NP binding. More than half of the low-NP regions have potential to form secondary and tertiary RNA structures, based on computational analysis, and mutations that are predicted to alter the stability of these RNA structures resulted in attenuated virus infection (summarized in **Table 4.2**). The presence of RNA structures in RNP complexes may explain the pleomorphic nature of RNPs[31]. Several, mostly shorter low-NP binding regions were not predicted to form stable structures. These regions may represent a portion of a larger RNA structure or have a yet unknown function during IAV replication to the fact that they are consisted of shorter sequences, had no detectable impact. These regions may represent a portion of a larger but less stable structure or have a different yet unknown function during IAV replication. We examined NP-vRNA interactions at a late time point when a majority of viral RNA is distributed throughout the cytoplasm and thought to be within vRNP complexes. In support of this, the great majority of PAR-CLIP reads were obtained from negative sense vRNAs.

We found many low-NP binding RNA regions that overlap with previously predicted packaging and bundling signals near segment termini (NP_{22-68}, $NP_{1410-1495}$, and NS_{23-88})[5,32]. However, additional low-NP binding regions were identified throughout the segment body and outside of traditional packaging singals[10,21,33]. Our findings agree with studies assessing structure-guided mutagenesis of hypothetical vRNA structures (segment 7) or biochemical analyses of vRNA (segment 8) that largely overlap with low-NP binding regions assessed here[7,34]. Additionally, we identified specific nucleotides required for co-packaging of segment 2 (PB1) and segment 6 (NA), in a region recently implicated in directing reassortment outcomes and co-segregation of these segments *in* vitro[35,36]. Therefore, NP PAR-CLIP allowed us to directly identify potentially functional RNA regions in IAV genomes required for coordinated genome packaging.

As predicted RNA structures in the low-NP binding regions are important for genome packaging, it suggests a requirement for inter-segment RNA interactions. Specific RNA interactions between genome segments have been demonstrated *in vitro* and ablation of these interactions lead to aberrant genome packaging; an outcome reversed by introduction of compensatory mutations in the interacting segment[11]. Many of the mutations that alter genome assembly also lead to an increase in number of defective viral particles[32]. The formation and nature of vRNA structures is likely contingent on the sequence composition and physical position of the nucleotides within a RNP complex.

IAVs can reassort and generate novel and potential pandemic strains. We tested if the regions identified in IAV-PR8 were required for replication of a divergent H7N3 virus. Synonymous structural changes in the same vRNA regions of two separate segments in both IAV-H7N3 and IAV-pH1N1 were attenuating and suggest conservations of these structural

elements. Future examination of NP-binding of viral RNA in diverse strains of IAV is likely to identify overlapping and distinct high and low-NP binding regions. These RNA features may be required for all viruses as coordinating packaging elements, but additional regions act as strain or lineage specific packaging enhancers. Further, these strain or lineage specific RNA features, required for genome packaging, may act as potential determinants of reassortment outcomes. Prior experimental systems demonstrated that co-segregation of genome segments occurs during natural reassortment as well as lab-adaptation of virus strains to create new vaccines[23,36]. Finally, attenuation of genome packaging through silent, structural mutations has the potential to accelerate live attenuated vaccine production using native genome constellations to allow vaccination with all protein epitopes of a novel pandemic virus, without the need for master donor strains.

4.6 Acknowledgements

I would like to thank the following colleagues for comments on experiments comprising this manuscript: Michael Diamond, MD, David Wang, PhD, and Deborah Lenschow, MD and Anshu Gounder all of Washington University in Saint Louis. Additionally, Traci Bricker and Christopher Edwards provided technical support. Chris Brooke provided anti-M1 hybridoma, M2-1C6. Funding for G.D.W. was provided by NIGMS training grant 5T32GM007067 and NIAID ID Training Grant 2T32AI007172, and the Victoria Fraser Fellowship for Infectious Disease Research.

4.7 References

1. Arranz, R. *et al*. The Structure of Native Influenza Virion Ribonucleoproteins. *Science (80-.)*. **338,** 1634–1637 (2012).

2. Moeller, A., Kirchdoerfer, R. N., Potter, C. S., Carragher, B. & Wilson, I. a. Organization of the Influenza Virus Replication Machinery. *Science (80-.)*. **338,** 1631–1634 (2012).

3. Stockley, P. G. *et al*. Packaging signals in single-stranded RNA viruses: Nature's alternative to a purely electrostatic assembly mechanism. *J. Biol. Phys*. **39,** 277–287 (2013).

4. Nicholson, B. L. & White, K. A. Functional long-range RNA-RNA interactions in positive-strand RNA viruses. *Nat. Rev. Microbiol*. **12,** 493–504 (2014).

5. Gultyaev, A. P. *et al*. RNA structural constraints in the evolution of the influenza A virus genome NP segment. *RNA Biol* **11,** 942–952 (2014).

6. Moss, W. N., Priore, S. F. & Turner, D. H. Identification of potential conserved RNA secondary structure throughout influenza A coding regions. 991–1011 (2011). doi:10.1261/rna.2619511.(IRES)

7. Kobayashi, Y., Dadonaite, B., Doremalen, N. Van, Barclay, W. S. & Pybus, O. G. Computational and molecular analysis of conserved influenza A virus RNA secondary structures involved in infectious virion production. *RNA Biol*. **13,** 883–894 (2016).

8. York, A., Hengrung, N., Vreede, F. T., Huiskonen, J. T. & Fodor, E. Isolation and characterization of the positive-sense replicative intermediate of a negative-strand RNA

virus. *Proc. Natl. Acad. Sci. U. S. A.* **110,** E4238-45 (2013).

9. Goto, H., Muramoto, Y., Noda, T. & Kawaoka, Y. The genome-packaging signal of the influenza A virus genome comprises a genome incorporation signal and a genome-bundling signal. *J. Virol.* **87,** 11316–22 (2013).

10. Marsh, G. a, Rabadán, R., Levine, A. J. & Palese, P. Highly conserved regions of influenza a virus polymerase gene segments are critical for efficient viral RNA packaging. *J. Virol.* **82,** 2295–304 (2008).

11. Fournier, E. *et al*. Interaction network linking the human H3N2 influenza A virus genomic RNA segments. *Vaccine* **30,** 7359–67 (2012).

12. Essere, B. *et al*. Critical role of segment-specific packaging signals in genetic reassortment of influenza A viruses. *Proc. Natl. Acad. Sci. U. S. A.* **110,** E3840-8 (2013).

13. Fournier, E. *et al*. A supramolecular assembly formed by influenza A virus genomic RNA segments. *Nucleic Acids Res.* **40,** 2197–2209 (2011).

14. Ozawa, M. *et al*. Nucleotide sequence requirements at the 5' end of the influenza A virus M RNA segment for efficient virus replication. *J. Virol.* **83,** 3384–8 (2009).

15. Fournier, E. *et al*. A supramolecular assembly formed by influenza A virus genomic RNA segments. *Nucleic Acids Res.* **40,** 2197–209 (2012).

16. Noda, T. *et al*. Three-dimensional analysis of ribonucleoprotein complexes in influenza A virus. *Nat. Commun.* **3,** 639 (2012).

17. Nakatsu, S. Complete and Incomplete Genome Packaging of Influenza A and B Viruses.

MBio **7,** 1–7 (2016).

18. Lakdawala, S. S. *et al*. Influenza a virus assembly intermediates fuse in the cytoplasm. *PLoS Pathog*. **10,** e1003971 (2014).

19. Gerber, M., Isel, C., Moules, V. & Marquet, R. Selective packaging of the influenza A genome and consequences for genetic reassortment. *Trends Microbiol*. **22,** 446–455 (2014).

20. Kutluay, S. B. *et al*. Global Changes in the RNA Binding Specificity of HIV-1 Gag Regulate Virion Genesis. *Cell* **159,** 1096–1109 (2014).

21. Marsh, G. a, Hatami, R. & Palese, P. Specific residues of the influenza A virus hemagglutinin viral RNA are important for efficient packaging into budding virions. *J. Virol*. **81,** 9727–36 (2007).

22. Brooke, C. B. *et al*. Most Influenza A Virions Fail to Express At Least One Essential Viral Protein. *J. Virol*. (2013). doi:10.1128/JVI.02284-12

23. Williams, G. D., Pinto, A. K., Doll, B. & Boon, A. C. M. A North American H7N3 influenza virus supports reassortment with 2009 pandemic H1N1 and induces disease in mice without prior adaptation. *J. Virol*. **90,** JVI.02761-15 (2016).

24. Gokhale, N. S. *et al*. N6-Methyladenosine in Flaviviridae Viral RNA Genomes Regulates Infection. *Cell Host Microbe* **20,** 654–665 (2016).

25. Gruber, A. R., Lorenz, R., Bernhart, S. H., Neuböck, R. & Hofacker, I. L. The Vienna RNA websuite. *Nucleic Acids Res*. **36,** W70-4 (2008).

26. Dawson, W. K., Fujiwara, K. & Kawai, G. Prediction of RNA pseudoknots using heuristic modeling with mapping and sequential folding. *PLoS One* **2,** (2007).

27. Schindelin, J. *et al*. Fiji: an open-source platform for biological-image analysis. *Nat. Methods* **9,** 676–682 (2012).

28. Reed, L. J. & Muench, H. A Simple Method of Estimating Fifty Per Cent Endpoints. *Am. J. Hygeine* **27,** 285–297 (1938).

29. Liu, C. L. *et al*. Using mutagenesis to explore conserved residues in the RNA-binding groove of influenza A virus nucleoprotein for antiviral drug development. *Sci Rep* **6,** 21662 (2016).

30. Zheng, W., Olson, J., Vakharia, V. & Tao, Y. J. The Crystal Structure and RNA-Binding of an Orthomyxovirus Nucleoprotein. *PLoS Pathog*. **9,** (2013).

31. Gallagher, J. R., Torian, U., McCraw, D. M. & Harris, A. K. Structural studies of influenza virus RNPs by electron microscopy indicate molecular contortions within NP supra-structures. *J. Struct. Biol*. (2016). doi:10.1016/j.jsb.2016.12.007

32. Hutchinson, E. C., Wise, H. M., Kudryavtseva, K., Curran, M. D. & Digard, P. Characterisation of influenza A viruses with mutations in segment 5 packaging signals. **27,** 6270–6275 (2009).

33. Gao, Q. *et al*. The influenza A virus PB2, PA, NP, and M segments play a pivotal role during genome packaging. *J. Virol*. **86,** 7043–51 (2012).

34. Yamanaka, K., Ishihama, A. & Nagata, K. Reconstitution of influenza virus RNA-

nucleoprotein complexes structurally resembling native viral ribonucleoprotein cores. *J. Biol. Chem.* **265,** 11151–11155 (1990).

35. Gilbertson, B. *et al*. Influenza NA and PB1 gene segments interact during the formation of viral progeny: Localization of the binding region within the PB1 gene. *Viruses* **8,** 1–17 (2016).

36. Cobbin, J. C. a *et al*. Influenza Virus PB1 and NA Gene Segments can Co-segregate during Vaccine Reassortment Driven by Interactions in the PB1 Coding Region. *J. Virol.* (2014). doi:10.1128/JVI.01022-14

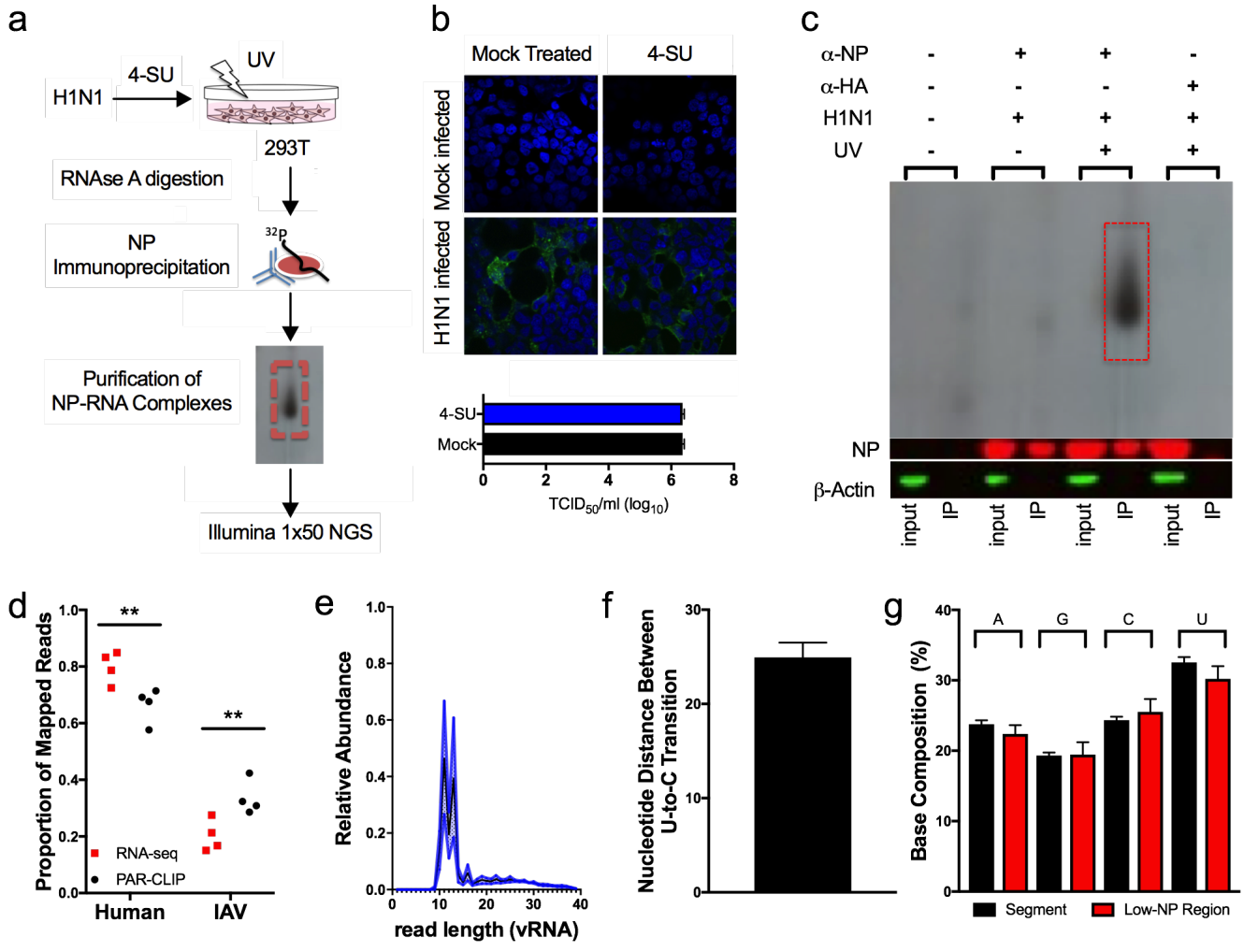

a
H1N1
4-SU
UV
293T
RNAse A digestion
NP
Immunoprecipitation
32P
Purification of
NP-RNA Complexes
Illumina 1x50 NGS
b
Mock Treated
4-SU
Mock infected
H1N1 infected
4-SU
Mock
0
2
4
6
8
TCID50/ml (log10)
c
α-NP
α-HA
H1N1
UV
NP
β-Actin
input
IP
d
Proportion of Mapped Reads
RNA-seq
PAR-CLIP
Human
IAV
e
Relative Abundance
read length (vRNA)
f
Nucleotide Distance Between
U-to-C Transition
g
Base Composition (%)
A
G
C
U
Segment
Low-NP Region

Figure 4.1. Development of PAR-CLIP for IAV NP

a, Schematic for IAV NP PAR-CLIP assay. **b,** Effects of 4-SU (100 μM) on attenuate IAV replication was assessed by $TCID_{50}$ assay (bottom) or confocal microscopy (top). Immunofluorescence staining for NP (green) was assessed in the presence or absence of 4-SU and counterstained with DAPI to identify cellular nuclei (blue). **c,** PAR-CLIP was conducted on 293T cells infected with WT-PR8 in the absence (lanes 1 and 2) or presence of a monoclonal antibody against to IAV NP (lanes 3-6) or viral hemagglutinin protein (HA) (lanes 7-8). The effect of UV cross-linking on binding of RNA to the viral NP is shown in lanes 3-4 (no UV) and lanes 5-6 (with UV). Radioactivity (^{32}P) is visualized by autoradiograph and the presence of absence of NP and cellular ß-actin was done by western blot. The input sample and eluate are loaded in the uneven and even lanes respectively. The results are representative of four independent experiments. **d,** Proportion of PAR-CLIP or RNA-seq derived reads mapping to human or IAV genomes (**, $P < 0.01$, n = 4). **e,** Length of negative-sense viral RNA (vRNA) aligning reads was determined using FastX Toolkit and the number of reads of a certain length is plotted as a proportion of total vRNA mapping reads. Mean (black line) +/- S.E.M. (blue shading) of 4 experiments. **f,** Nucleotide distance between crosslinking-induced A-to-G transitions in vRNA mapped PAR-CLIP data (mean + S.E.M. of 4 experiments). **g,** Nucleotide composition of low-NP binding regions and IAV genome (displayed as average base composition of all eight gene-segments) . No significant differences were detected between groups.

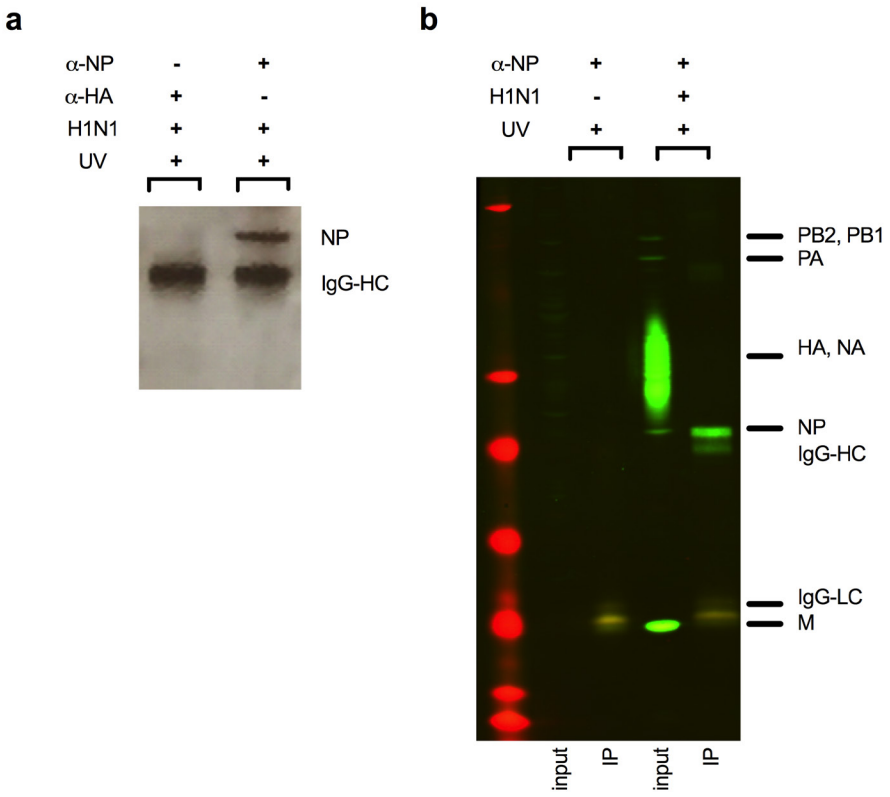


Figure 4.2. NP immunoprecipitation protocol results in pure NP elution. a, Silver stain of immunoprecipitation (IP) eluate for IAV infected cells with either anti-HA or anti-NP antibodies. The upper band represents NP and lower band represents IgG heavy chain. Gel image is representative of two independently performed experiments **b,** IP of NP from purified virus particles removes all other viral proteins. Western blot using polyclonal goat sera against H1N1 virus to detect all viral proteins in input and NP-immunoprecipitated eluate. Lane 1 is a marker, lanes 2 and 3 are control immunoprecipitation beads in which no virus lysate was added. Lanes 4 and 5 are input or IP eluate from virus lysate immunoprecipitated with anti-NP antibody HB65. The membrane was probed with goat-anti-IAV USSR (H1N1) (US Biological, I7650-78B) and IRDye 800CW donkey-anti-goat (Licor). Gel image is representative of two independently performed experiments.

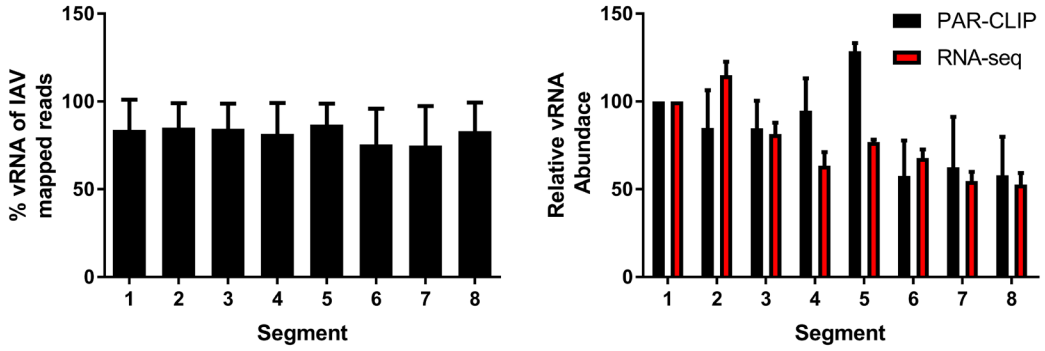


Figure 4.3. NP binds negative-strand IAV genome and immunoprecipitated RNAs are representative of total cellular IAV RNA levels. Percentage of mapped reads derived from negative-sense viral RNA (vRNA) for four PAR-CLIP experiments (mean + S.E.M., n = 4) and PAR-CLIP and RNA-seq sequence abundance per gene segment (vRNA only) of IAV. Calculated as the total number of base calls per segment and normalized to segment 1 (mean + S.E.M., n = 4 each).

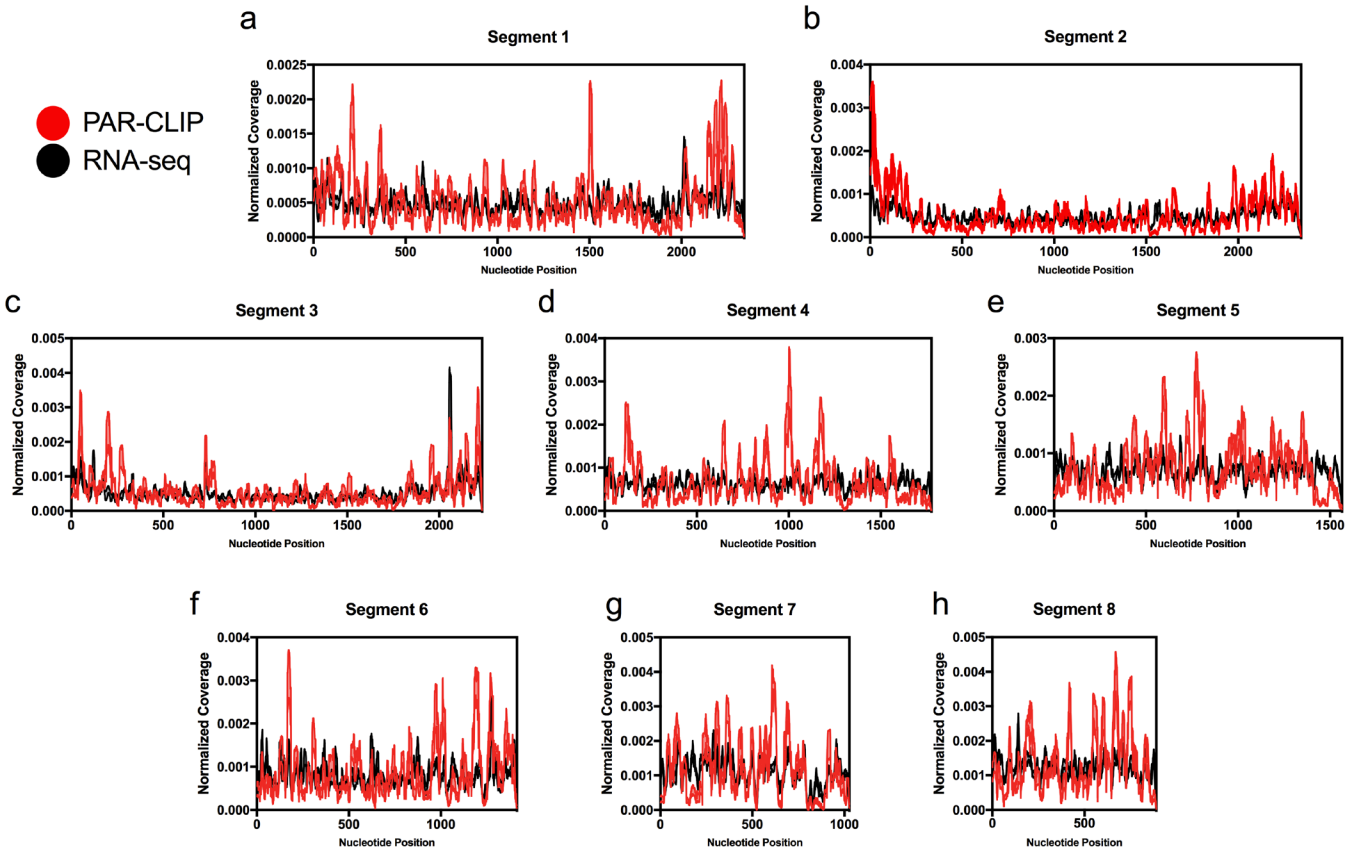


Figure 4.4. Nucleoprotein binding is non-uniform in all eight IAV genome segments. a-h, Normalized coverage for complete genome segments for PAR-CLIP (red) and RNA-seq (black). Lines represent mean +/- S.E.M from four experiments each.

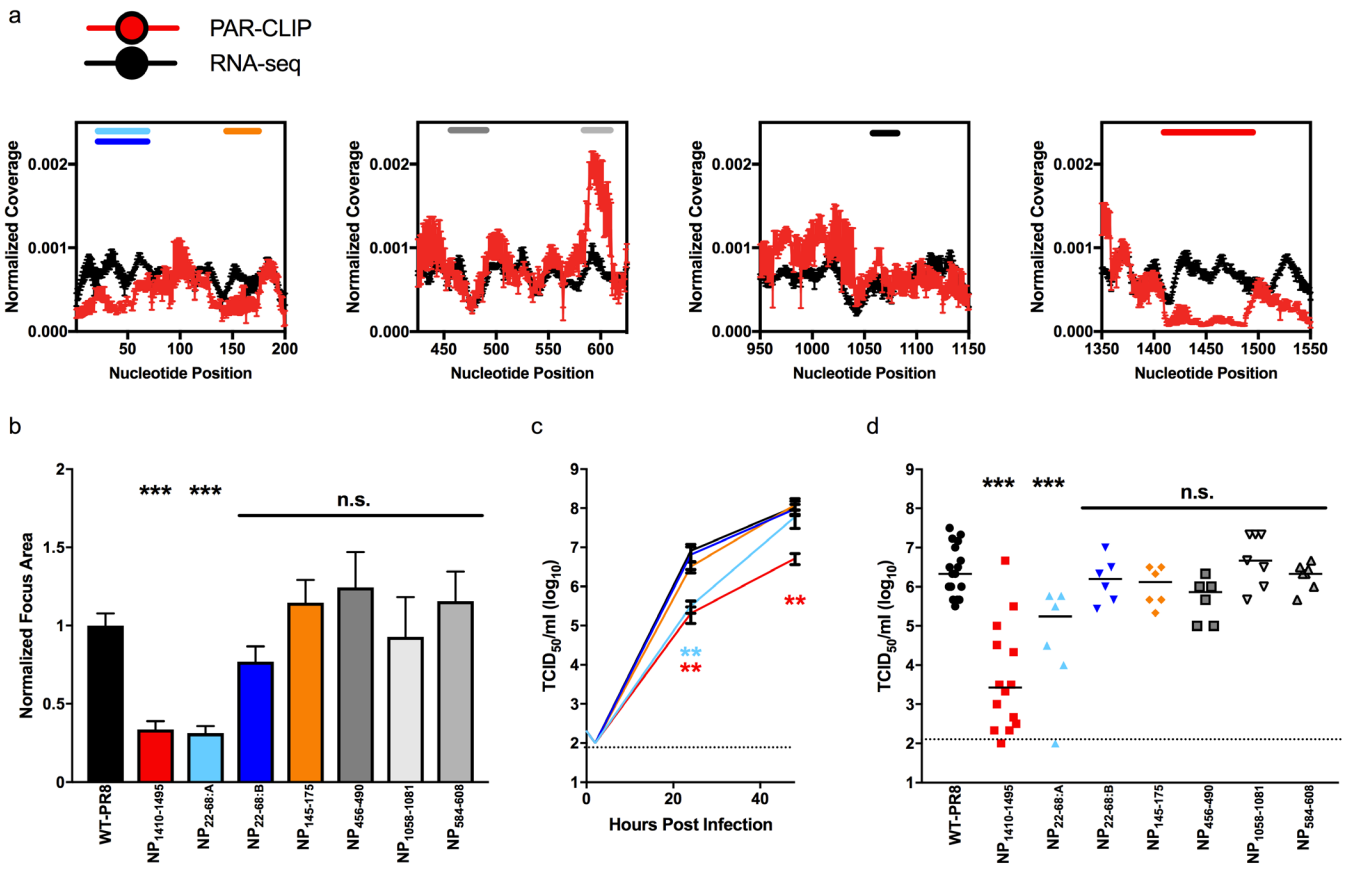

a
PAR-CLIP
RNA-seq
Normalized Coverage
Nucleotide Position
b
Normalized Focus Area
***
***
n.s.
WT-PR8
NP1410-1495
NP22-68:A
NP22-68:B
NP145-175
NP456-490
NP1058-1081
NP584-608
c
TCID50/ml (log10)
Hours Post Infection
**
**
**
d
TCID50/ml (log10)
***
***
n.s.

Figure 4.5. Manipulation of vRNA in low-NP binding regions that disrupts predicted RNA structure attenuates virus replication

a, Normalized coverage +/- S.E.M was determined for each nucleotide in the PAR-CLIP (red symbols) and RNA-seq (black symbols) libraries (n = 4 each). Six regions of interest (ROI) in segment 5 (NP) are highlighted including two low-NP binding regions ($NP_{1410-1495}$ and NP_{22-68}), one high-NP binding region ($NP_{584-608}$) and three intermediate NP-binding regions ($NP_{145-175}$, $NP_{456-490}$, and $NP_{1058-1081}$). Each ROI is indicated with a colored bar. **b,** Focus area of WT-PR8 and NP mutant IAV. MDCK cells were infected with serial dilutions of the indicated viruses and overlayed with infection media (M0.1B) containing 1% agarose and TPCK-trypsin. Seventy-two hours later cells were fixed, permeabilized, and stained for viral antigen (NP). Focus diameter was determined and normalized to WT-PR8 per experiment. Results are the average + S.E.M. of 3-5 experiments per virus (> 60 foci each). **c,** MDCK cells were infected with MOI=0.001 of the indicated viruses and culture supernatant was collected at indicated time points then titered on MDCK cells. Results are the average +/- S.E.M. $TCID_{50}$/ml of two experiments performed in duplicate. **d,** C57BL/6J mice were inoculated with 10^3 $TCID_{50}$ of the indicated viruses in 30μL. Lungs were collected, homogenized, and titered on MDCK cells. Each dot is a single mouse and the line is the geometric mean. Dotted line in **c** and **d** represents the limit of detection. (***, $P < 0.005$; **, $P < 0.01$; *, $P < 0.05$; n.s., not significant)

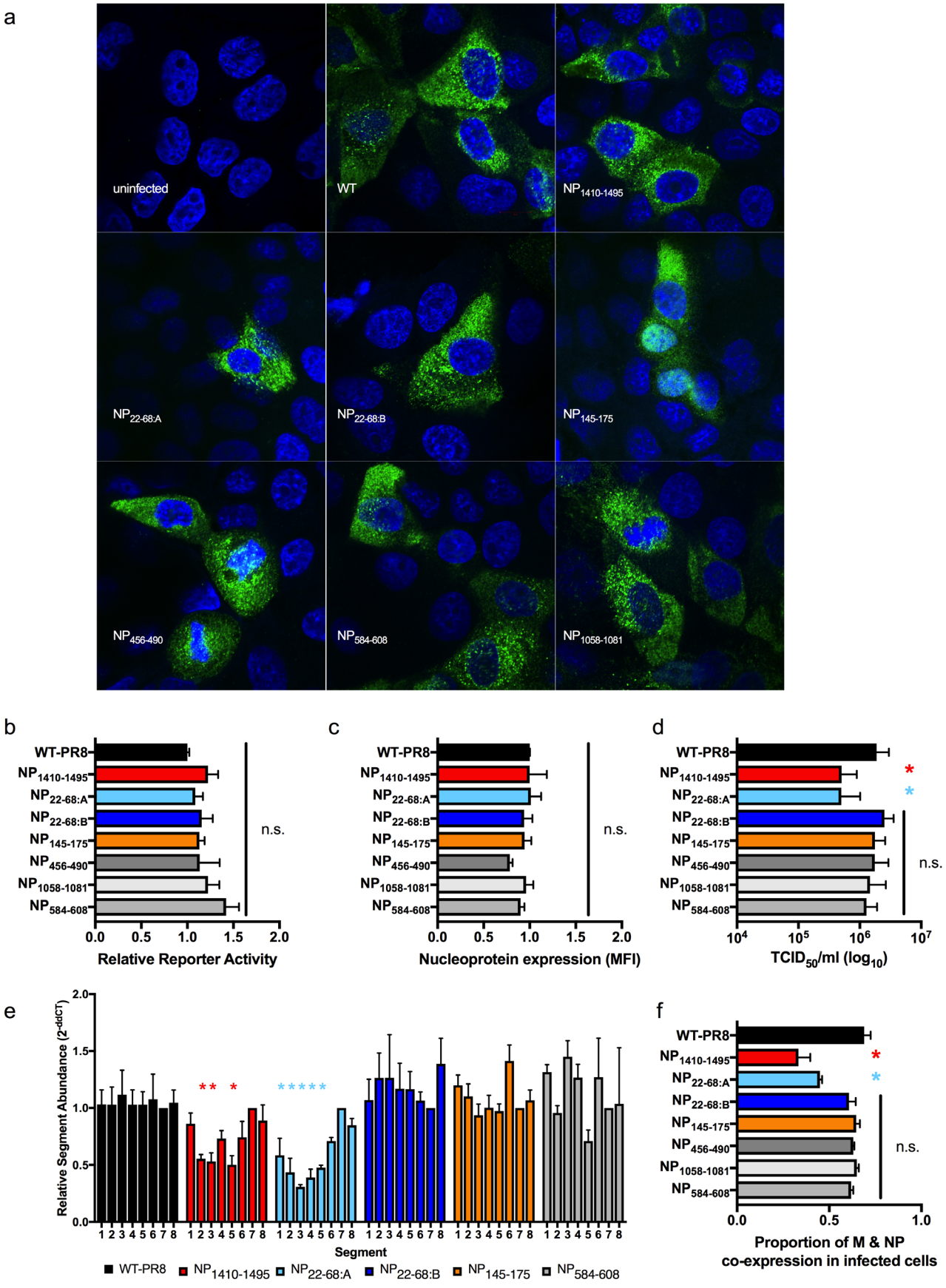

a
uninfected
WT
NP1410-1495
NP22-68:A
NP22-68:B
NP145-175
NP456-490
NP584-608
NP1058-1081
b
WT-PR8
NP1410-1495
NP22-68:A
NP22-68:B
NP145-175
NP456-490
NP1058-1081
NP584-608
n.s.
0.0 0.5 1.0 1.5 2.0
Relative Reporter Activity
c
WT-PR8
NP1410-1495
NP22-68:A
NP22-68:B
NP145-175
NP456-490
NP1058-1081
NP584-608
n.s.
0.0 0.5 1.0 1.5 2.0
Nucleoprotein expression (MFI)
d
WT-PR8
NP1410-1495
NP22-68:A
NP22-68:B
NP145-175
NP456-490
NP1058-1081
NP584-608
* *
n.s.
10^4 10^5 10^6 10^7
TCID50/ml (log10)
e
Relative Segment Abundance (2^-ddCT)
2.0 1.5 1.0 0.5 0.0
** * *****
1 2 3 4 5 6 7 8 1 2 3 4 5 6 7 8 1 2 3 4 5 6 7 8 1 2 3 4 5 6 7 8 1 2 3 4 5 6 7 8
Segment
WT-PR8 NP1410-1495 NP22-68:A NP22-68:B NP145-175 NP584-608
f
WT-PR8
NP1410-1495
NP22-68:A
NP22-68:B
NP145-175
NP456-490
NP1058-1081
NP584-608
* *
n.s.
0.0 0.5 1.0
Proportion of M & NP
co-expression in infected cells

Figure 4.6. Attenuating mutations in segment 5 impact coordinated genome packaging

a, Confocal microscopy images depicting the localization of nucleoprotein (NP) eight hours post infection (hpi). MDCK cells were infected with the indicated virus (white text), and NP was identified using the MAb HB65 and Alexa-488 conjugated goat-anti-mouse secondary antibody by immunofluorescence. Cell nuclei are counterstained with DAPI. Panels depict the merged image of DAPI and NP staining. Fields are representative of two independent experiments. **b,** Dual-luciferase reporter assay to assess viral transcription and genome replication, each combination of plasmids was assessed 3-5 times with corresponding the WT-PR8 combination. **c,** Mean fluorescence intensity (MFI) of NP in virally infected MDCK cells (MOI = 0.05, experiments were performed twice in duplicate) as revealed by flow cytometry. **d,** Viral titer ($TCID_{50}$/ml) of 4 HA-units of WT-PR8 and mutant viruses. Results are the average of three viral titrations and two HA assays per infectious virus titration ($TCID_{50}$ assay and HA titration experiments, average MFI + S.E.M). **e,** Proportion of infected cells co-expressing of matrix (M) and NP proteins in singly infected MDCK cells (MOI = 0.05) 16 hpi as revealed by flow cytometry. The average percentage of co-expression was calculated from two experiments performed in duplicate. **f,** Relative abundance of genome segments in purified WT or mutant viruses. All segments were compared to segment 7 (M) vRNA and normalized to the average of WT-PR8 values using the 2^{-ddCt} method[55]. Bars represent the mean of 3-6 independent virus preparations + S.E.M. (*, $P < 0.05$; n.s., not significant)

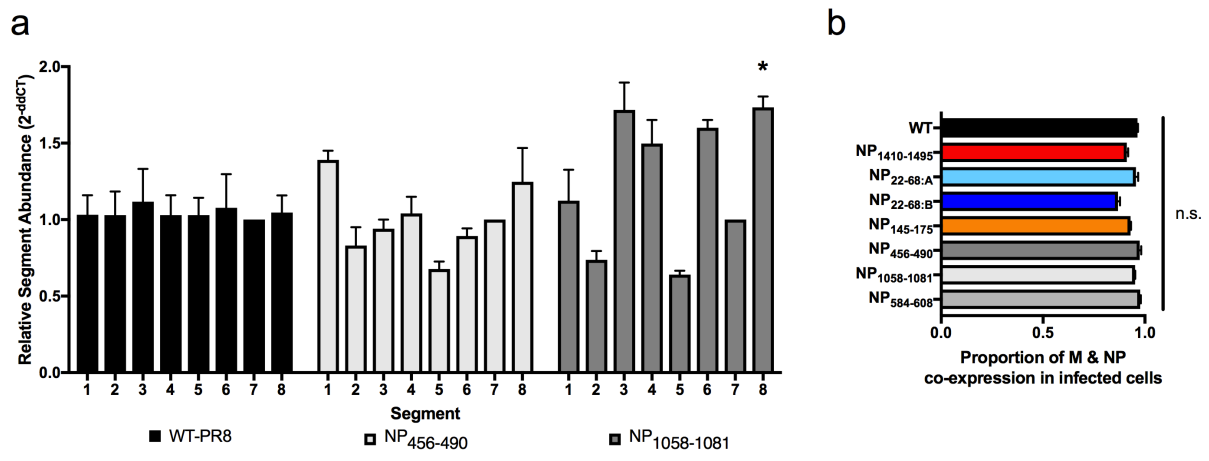


Figure 4.7. Additional mutations in NP-bound regions have minimal effect on coordinated genome packaging and decreases in viral protein co-expression may be overcome by high multiplicity of infection. a, Packaging qPCR of additional mutant viruses, as in Figure 3f (*, $P < 0.05$). The relative abundance of each genome segment in concentrated virus particles was assessed and normalized to Segment 7 and WT-PR8 virus by the 2^{-ddCT} method as stated in the materials and methods. **b,** co-expression of M and NP proteins in MDCK cells infected at high MOI (5) for 16 hours. High MOI infection of MDCK cells overcomes co-expression defects observed at low MOI (Figure 3e). Proportion of infected cells co-expressing of NP and M was determined by calculating the number of NP+M+ double positive cells and dividing by the total number of cells expressing NP or M when assessed by flow cytometry. Results are derived from two independent experiments performed in duplicate.

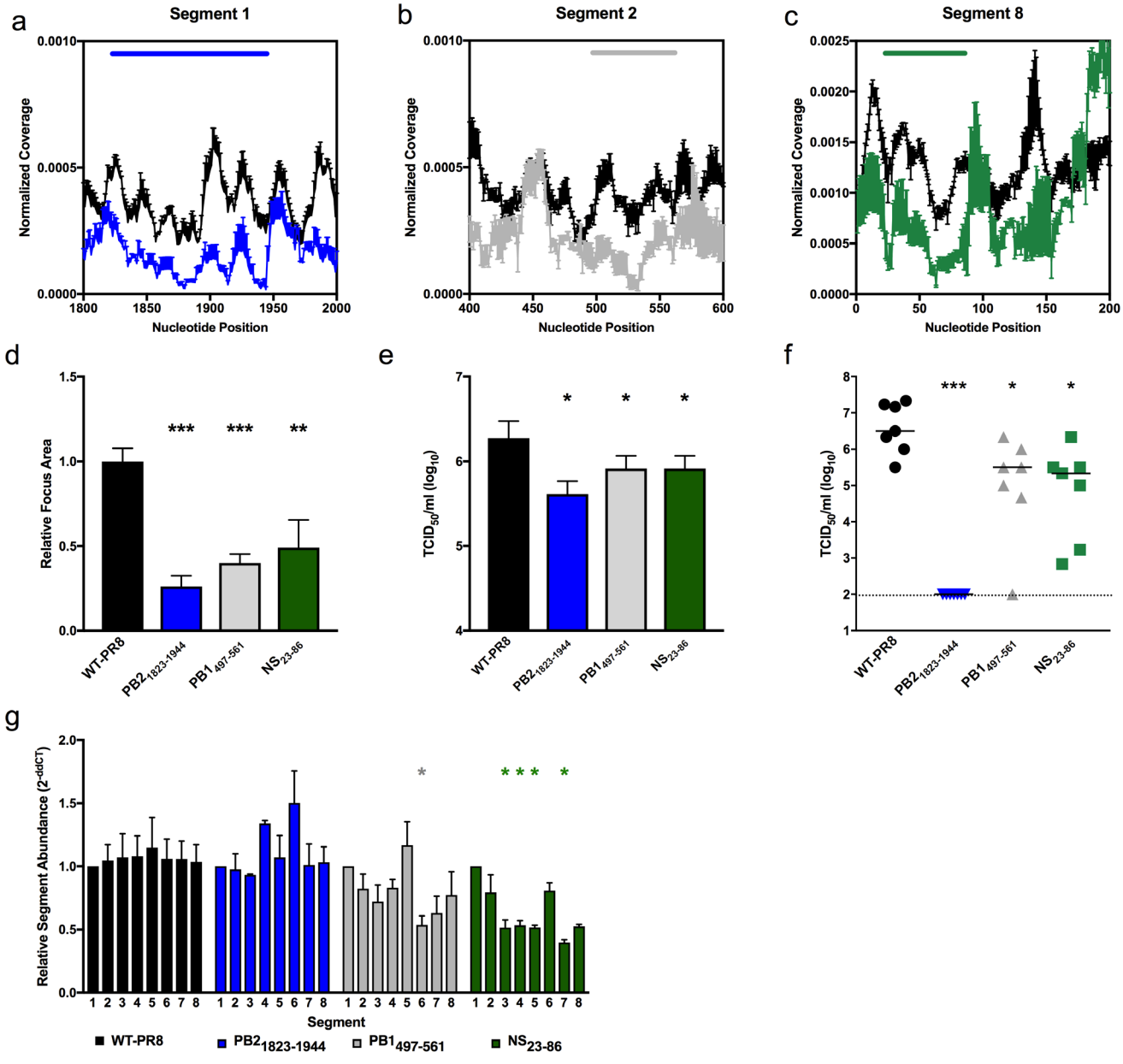

a
Segment 1
Normalized Coverage
Nucleotide Position
b
Segment 2
Normalized Coverage
Nucleotide Position
c
Segment 8
Normalized Coverage
Nucleotide Position
d
Relative Focus Area
WT-PR8
PB2 1823-1944
PB1 497-561
NS 23-86
e
TCID50/ml (log10)
f
TCID50/ml (log10)
g
Relative Segment Abundance (2-ddCT)
Segment
WT-PR8
PB2 1823-1944
PB1 497-561
NS 23-86

Figure 4.8. Synonymous structural mutations in low-NP binding regions in Segments 1, 2, and 8 attenuated virus replication and genome packaging.

a-c, Segment 1 (**a,** blue line), 2 (**b,** grey line), and 8 (**c,** green line) low-NP binding regions defined by PAR-CLIP analysis. Normalized coverage +/- S.E.M was determined for each nucleotide in the PAR-CLIP (blue, grey or green symbols) and RNA-seq (black symbols) libraries (n=4 each). **d,** Relative focus area of WT-PR8 and segment 1, 2, and 8 mutant viruses. Bars are the average + S.E.M. of 3 experiments per virus (> 60 foci each). **e,** Viral titer ($TCID_{50}$/ml) of 4 HA-units of WT-PR8 and mutant virus. Results are the average of three viral titrations and two HA assays per infectious virus titration ($TCID_{50}$ assay and HA titration experiments, average MFI + S.E.M). **f,** C57BL/6J mice were inoculated with 10^3 $TCID_{50}$ in 30μL. Lungs were collected 48 hpi, homogenized, and titered. Each dot is a single mouse and the line is the geometric mean. Dotted line in **f** represents the limit of detection. **e,** Relative abundance of genome segments in purified WT or mutant viruses. All segments were compared to segment 1 (PB2) vRNA and normalized to the average of WT-PR8 values using the 2^{-ddCt} method[55]. Bars represent the mean of 3-6 independent virus preparations + S.E.M. (***, $P < 0.005$; **, $P < 0.01$; *, $P < 0.05$; n.s., not significant).

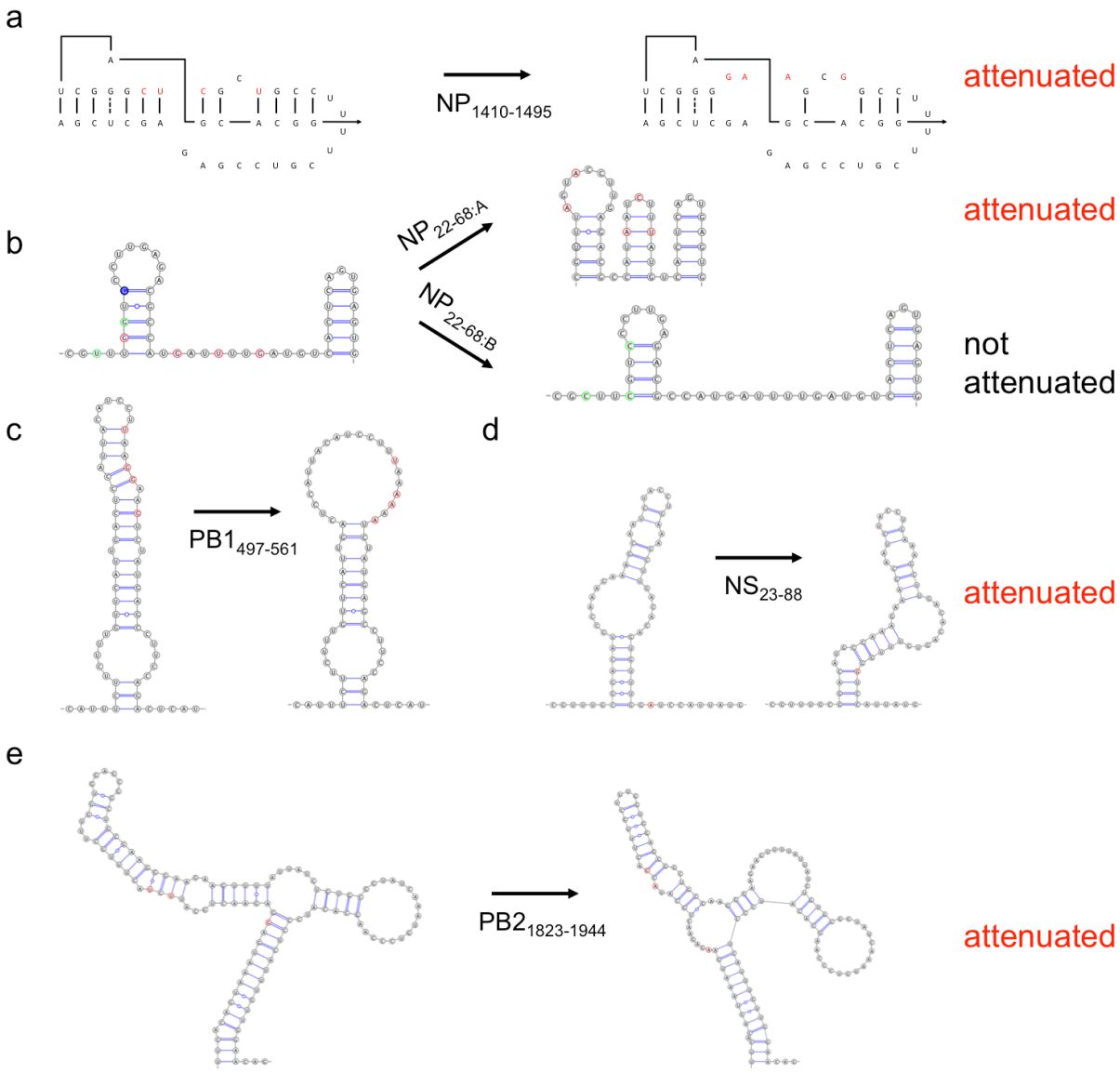


Figure 4.9. Predicted RNA structure of WT-PR8 and the indicated mutant viruses. In all cases the nucleotides substituted are highlighted and WT-PR8 structures are to the left of the arrow. **a,** alteration of a previously identified pseudoknot within the $NP_{1410-1495}$ vRNA region. **Alteration** of predicted secondary structures in the indicated vRNA regions for NP_{22-68} region (**b**), $PB1_{497-561}$ (**c**), NS_{23-88} (**d**). and $PB2_{1823-1944}$ (**e**).

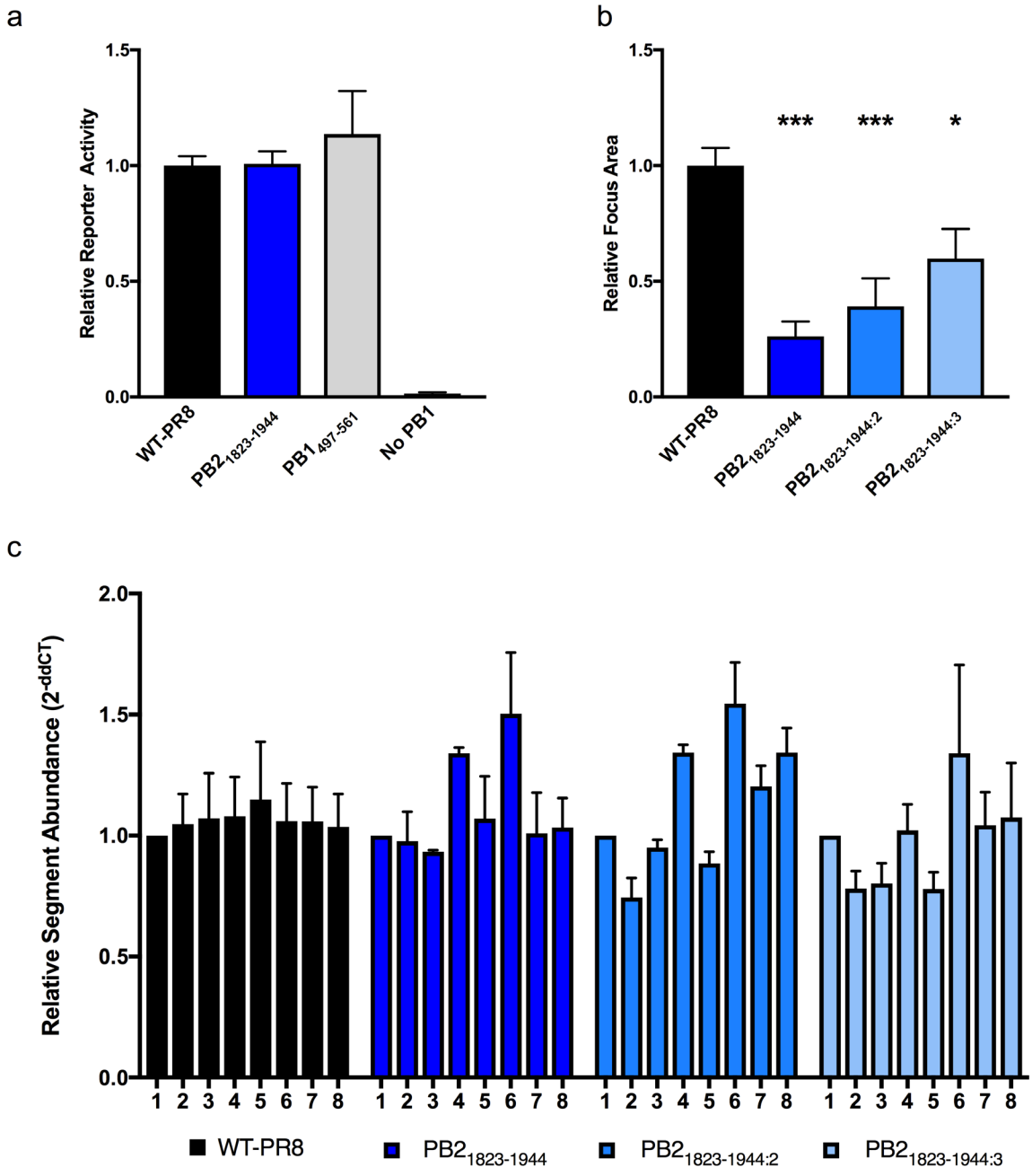
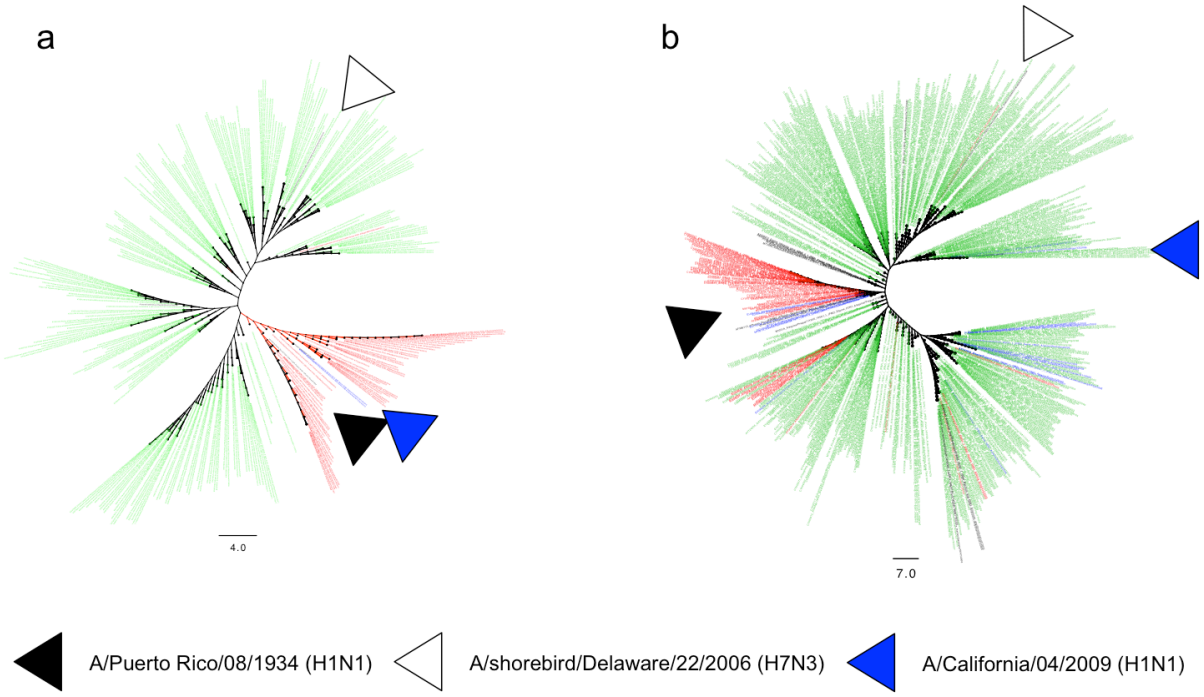

a
Relative Reporter Activity
1.5
1.0
0.5
0.0
WT-PR8
PB2 1823-1944
PB1 497-561
No PB1
b
Relative Focus Area
***
***
*
WT-PR8
PB2 1823-1944
PB2 1823-1944:2
PB2 1823-1944:3
c
Relative Segment Abundance (2-ddCT)
2.0
1.5
1.0
0.5
0.0
1 2 3 4 5 6 7 8
WT-PR8
PB2 1823-1944
PB2 1823-1944:2
PB2 1823-1944:3

Figure 4. 10. Multiple mutations in Segment 1 unbound region attenuate focus formation but do not lead to apparent segment-specific packaging defects. a, Dual-luciferase reporter assay to assess viral transcription and genome replication, no significant differences observed (average + S.E.M. of three experiments performed in duplicate). **b,** Relative focus area of WT-PR8 and segment 1 mutant viruses (n=3 experiments per virus, area for > 40 total foci per virus calculated). **c,** Relative abundance of genome segments in purified WT or mutant viruses. all segments were compared to segment 1 (PB2) vRNA and normalized to the average of WT-PR8 values using the 2^{-ddCt} method[55]. Bars represent the mean of 3-6 independent virus preparations + S.E.M. (***, $P < 0.005$; **, $P < 0.01$; *, $P < 0.05$; n.s., not significant).


a
b
4.0
7.0
A/Puerto Rico/08/1934 (H1N1)
A/shorebird/Delaware/22/2006 (H7N3)
A/California/04/2009 (H1N1)

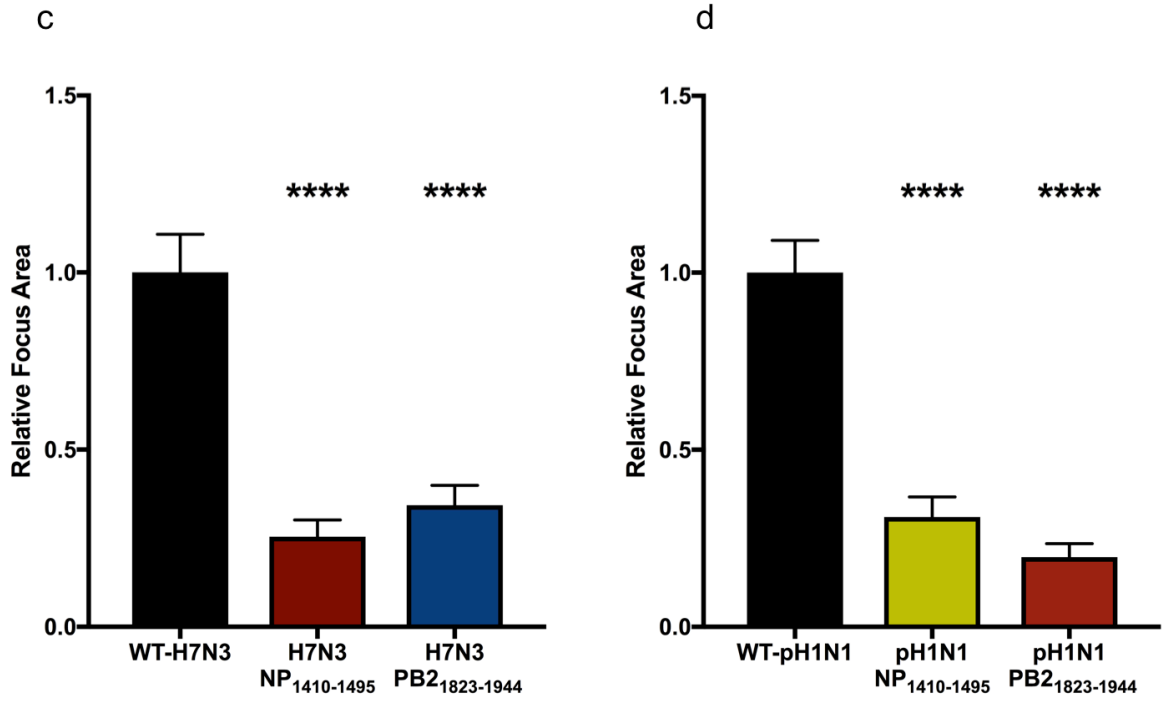

c
d
Relative Focus Area
1.5
1.0
0.5
0.0
****
****
WT-H7N3
H7N3 NP1410-1495
H7N3 PB21823-1944
WT-pH1N1
pH1N1 NP1410-1495
pH1N1 PB21823-1944

Figure 4.11. vRNA regions required for PR8 replication are required for replication of contemporary avian and human IAV.

a, Phylogenetic analysis of segment 5. WT-PR8, WT-H7N3, or WT-pH1N1 IAV are indicated by corresponding, labeled arrows. Phylogenies were created from randomly sampled full-length segment 5 sequences downloaded from NCBI IVR. Alignment performed in MEGA (version 7) using MUSCLE. Phylogenetic trees created using the Maximum Parsimony method included in MEGA. Phylogenies were visualized in FigTree and manually annotated. Green shading represents segment sequences derived from avian viruses; red represents human viruses; and blue represents swine viruses. **b,** Phylogenetic analysis of segment 2 performed as in (**a**). **c,** Relative focus area of WT-H7N3 and mutant viruses + S.E.M. d, Relative focus area of WT-pH1N1 and mutant viruses + S.E.M. ****, $P < 0.001$; 3 experiments, > 60 foci per virus with three independent experiments performed per virus.

Segment 1

Low (nt)	FDR min	Fold-change	Length	High (nt)	FDR min	Fold-change	Length
307-333*	0.01	>3	>18	117-150	0.01	<3	>18
408-431	0.05	>3	>18	2141-2159	0.05	>3	>18
1089-1121	0.05	>3	>18				
1823-1944*	0.01	>3	>18				
2321-2341	0.05	>3	>18				

Segment 2

Low (nt)	FDR min	Fold-change	Length	High (nt)	FDR min	Fold-change	Length
297-419*	0.01	>3	>18	39-56*	0.01	>3	>18
465-482	0.01	<3	>18	2178-2195	0.01	<3	>18
497-561*	0.01	>3	>18				
930-955*	0.01	>3	>18				
1294-1314	0.05	>3	>18				
1423-1440*	0.01	>3	>18				
1521-1576	0.05	>3	>18				
1746-1773*	0.01	>3	>18				
1790-1817*	0.01	>3	>18				

Segment 3

Low (nt)	FDR min	Fold-change	Length
619-640	0.05	>3	>18
792-817	0.05	>3	>18
1098-1113	0.01	>3	16

Segment 4

Low (nt)	FDR min	Fold-change	Length	Low (nt)	FDR min	Fold-change	Length
208-240*	0.01	>3	>18	98-151*	0.01	>3	>18
252-282	0.05	>3	>18	980-1009	0.05	>3	>18
376-430*	0.01	>3	>18	1446-1463	0.01	<3	>18
441-464	0.05	>3	>18				
491-528	0.05	>3	>18				
602-628	0.05	>3	>18				
1256-1340*	0.01	>3	>18				
1585-1661*	0.01	>3	>18				
1719-1770*	0.01	>3	>18				

Segment 5

Low (nt)	FDR min	Fold-change	Length	High (nt)	FDR min	Fold-change	Length
22-68*	0.01	>3	>18	634-661	0.01	<2	>18
261-329	0.01	>3	>18	766-787*	0.01	>3	>18
337-378	0.01	>3	>18				
1410-1495*	0.01	>3	>18				
1514-1560*	0.01	>3	>18				

Segment 6

Low (nt)	FDR min	Fold	Length	High (nt)	FDR min	Fold	Length
249-269*	0.01	>3	>18	1181-1199	0.05	>3	>18
849-866	0.05	>3	>18				
1038-1057*	0.01	>3	>18				

1080-1197	0.05	>3	>18

Segment 7

Low (nt)	FDR min	Fold	Length	High (nt)	FDR min	Fold	Length
1-16	0.01	>5	16	601-632*	0.01	>3	>18
142-201*	0.01	>3	>18				
640-660*	0.01	>3	>18				
839-866	0.05	>3	>18				

Segment 8

Low (nt)	FDR min	Fold	Length	High (nt)	FDR min	Fold	Length
23-86*	0.01	>3	>18	413-430	0.05	>3	>18
113-139	0.05	>3	>18				
867-885*	0.01	>3	>18				

Table 4.1. Low- and High-NP bound regions of vRNA.

Regions of vRNA greater than or equal to 18 nucleotides long and meeting criteria described in the methods. NP-low indicates the region was significantly lower in PAR-CLIP than RNA-seq and NP-high indicates the region was significantly higher in PAR-CLIP than RNA-seq. FDR Minima is the minimum False Discovery Rate for each region. Fold-change is absolute fold-difference between PAR-CLIP and RNA-seq data. The regions that meet all three criteria are denoted with asterisks, while those meeting two are not.

Virus	NP-bound	WT –ΔG (MFE)	Effect of Mutations	Effects on Virus				
				Focus Size	Lung Titer	HA:$TCID_{50}$ Ratio	Segment Packaging	H7N3 and pH1N1 Shared
NP 22-68:A	--	-8.7	↓↓–ΔG	↓↓	ê	↓↓	↓↓	N.D.
NP 22-68:B	--	-8.7	No effect	=	=	=	=	N.D.
NP 145-175	=	-8.0	↓–ΔG	=	=	=	=	N.D.
NP 456-490	=	-3.2	No effect	=	=	=	=	N.D.
NP 584-608	++	-0.0	No effect	=	=	=	=	N.D.
NP 1058-1081	=	-0.0	No effect	=	=	=	=	N.D.
NP 1410-1495	--	-19.8	Modified Pseudoknot	↓↓	↓↓	↓↓	↓↓	++
PB2 1823-1944	--	-33.0	Modified Pseudoknot	↓	↓↓	↓↓	↓↓	++
PB1 497-561	--	-12.9	↓ –ΔG	↓	↓	↓↓	↓↓	N.D.
NS 23-88	--	-10.0	Modified Pseudoknot	↓↓	↓	↓↓	↓↓	N.D.

Table 4.2. Disruption of potential RNA structural elements contributes to virus attenuation.

For NP-binding the following categories were considered: --, significantly lower that WT-PR8; =, equal to WT-PR8; ++, greater than WT-PR8. The regional stability was determined in WT-PR8 and mutant virus using Vienna RNAfold. MFE: minimum free energy. All calculations were performed using the default settings without imposing structural constraints. Pseudoknot formation potential was performed using vsFold5 and default settings. Results for focus area, lung titer, $TCID_{50}$/ml:HA ratio, relative segment packaging and conservation in A/shorebird/Delaware/22/2006 (H7N3) are summarized from previous figures (=, equivalent to WT-PR8; ↓, $P < 0.05$; ↓↓, $P < 0.01$).

Chapter 5:

Conclusions and Future Directions

5.1 Production and packaging of eight segments complicates IAV evolution

Influenza A viruses (IAV) possess eight genome segments that must be co-packaged into nascent virions to enable fully productive subsequent infection[1]. Throughout prior chapters, I have discussed the individual components of IAV genome packaging, reassortment, and how congruence between viral factors, RNA and protein, are required for generation of genetically diverse infectious virus, propagation, and pathogenesis. Herein, I will discuss the layers of genetic compatibility we found to be a pretext for the potential of reassortment, as well as unresolved and newly generated questions which future studies may be directed towards.

5.2 Environmental and cellular requirements for reassortment

The potential for pandemic formation of antigenically novel viruses necessitates simultaneous replication of parent viruses in a single host cell[2]. For contemporaneous cellular infection to occur a singular virus must have the ability to infect a host via either the respiratory (in mammals) or oral-fecal (in birds) route, establish replication, and spread within the target tissue. Additionally, once a singular virus has infected and established replication, the secondary virus must co-infect the host and disseminate to a shared site of replication. Recent studies demonstrate that tissue localized co-infection must be relatively contemporaneous, otherwise the host innate immune system likely suppresses the chances for super-infection[2,3]. The speed of co-infection on an organismal level has not been exhaustively studied, but the cellular time scale has been assessed *in vitro*[3]. In Chapters 2 and 4, I explored a method to bypass the required cellular co-infection by utilizing an *in vitro* co-transfection protocol[4]. By so doing, we minimized temporal and spatial issues related to co-infection. The results from those reassortment screens demonstrated that genetic as well as host factors are important determinants of what potential

new viruses may emerge. Subsequent sections of this chapter will focus on these findings and expound upon implications for viral evolution.

5.3 Reassortment may augment the fitness of viral quasispecies to overcome replication and transmission bottlenecks.

Segmentation, while complicating genome packaging, increases the potential for exploration of sequence space within and between multiple hosts following reassortment[5,6]. Influenza viruses have an error prone polymerase (PB1), generating $5x10^{-5}$ mutations per genome segment copied, therefore each new virus genome may contain 2 or 3 *de novo* mutations[7,8]. As such, the potential for acquisition of both beneficial and deleterious mutations at the RNA or protein level are possible. Recent studies on viral quasispecies demonstrate that variations in the mutation rate itself can be deleterious to viral fitness[7,9]. However, if a mutation is acquired in one segment, this allows transfer of evolved genetic information to a new genome constellation. In the case of beneficial mutations, such as those that overcome species-specific restriction factors, acquisition of a segment bearing this signature may enable host range diversification[5]. Conversely, if a deleterious mutation is acquired, it may be rapidly purged from the quasispecies with little consequence to the population at large. Additionally, the rapid within-host divergence likely alters the between host transmission potential, as this is both a physical and genetic bottleneck, especially when zoonosis occurs[10]. This has been demonstrated experimentally by sequencing viral populations at the site of infection in one host as well as the recipient in which a narrowing of genetic diversity has occurred within and between host quasispecies[11]. Co-infection presents many logistical issues during viral replication. Replication complexes comprised of heterogeneous polymerase subunits derived from either virus likely

form, potentially altering the copy number of certain segments. It is unknown if heterogeneous polymerase complexes faithfully replicate genome segments to the same extent as a homogenous complex; if not, the rate of spurious mutations may increase during the initial and subsequent rounds of infection following the original co-infection event and increasing again the diversity of a quasispecies. Similarly, when the mutation rate is enhanced, or a large proportion of genome segments are defective, reassortment between strains may be enhanced[12].

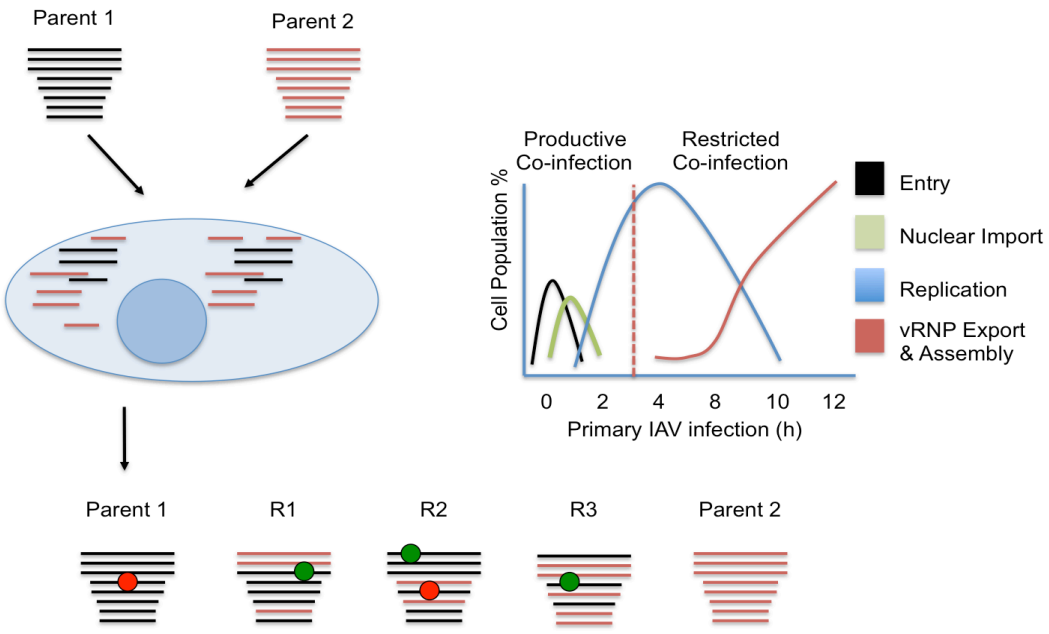


Figure 5.1. IAV co-infection may produce progeny viruses that possess mixed genomes through reassortment. When two parent viruses co-infect a cell (left), prior to the initiation of genome replication (right), generation of progeny IAV containing different combinations of genome segments (256 possible combinations) may occur. In addition to reassortment, or acquisiton of a new genome constellation component, the error prone polymerase of IAV may introduce mutations to genome segments during replication that alter the fitness of viruses in which they are packaged.

5.4 Decoding reassortment potential

Numerous studies demonstrate that protein compatibility is not the sole driver of reassortment potential despite critical interactions required for replication[13–16]. Additional studies indicate the potential for a segment-segment interaction network as a mechanism of genome packaging and therefore determinant of reassortment[17–21]. However, a theoretical problem in the field has persisted since a genome segment was first imaged in the 1970s[22]. The overwhelming majority of viral RNA should be bound tightly by nucleoprotein according to prevailing models[23].

This binding, if complete, would likely impede the formation of segment-segment interactions. However, others have shown the requirement for specific RNA features in IAV. This discord between the requirement for and ability to form RNA features drove our experimental design for the work described in Chapter 4. While we, and others, have reached similar conclusions pertaining to the interaction of NP and viral RNA, further investigation should be undertaken to complete the scaffolding our work has provided[24].

To determine how genetic interaction between diverse viruses occurs during co-infection, experiments to determine if vRNPs can be comprised of heterogeneous NP monomers from multiple viruses should be undertaken and if NPs from different strains interact with the same vRNA equivalently. Correspondingly, in the context of single virus and co-infection, the interaction network of vRNAs should be determined. Reassortment between closely related IAV occurs at high frequency and the rate of reassortment decreases, as segments are more divergent[3]. Certainly, a portion of reassortment rate is determined by protein-protein component interactions[5,25–27], which allows efficient replication and overcome host restriction[28,29], but the contribution of RNA-RNA interactions must be studied further. Here, I demonstrate that the

genetic and cellular context in which reassortment occurs couple to shape viral evolution. In chapter 2, I demonstrated that an avian H7N3 virus is capable of acquiring segments from the 2009 pandemic that enhance pathogenesis in mice. However, many highly functional polymerase complexes, *in vitro* at least, are grossly attenuated when used to infect mice[4]. Despite protein compatibility, some additional function of these multifunctional proteins or underlying RNA features rendered these viruses incapable of efficient replication. Chapter 3 addresses competitive incorporation of the same genome segments in a different genome constellation. We found stark differences in selection preferences of multiple segments, indicating that the genetic context in which reassortment occurs determines the output progeny that are able to cause subsequent infection. Surprisingly, manipulation of a single amino acid in one protein can alter the outcome of engineered reassortment experiments, as can alteration of the constituent genetic backbone in which these experiments take place. The constituent proteome of closely related host species, and even tissue and cell types within a specific organism are different. Differences in host proteins that may be utilized by the virus or more likely the presence of specific host restriction factors and strain-specific antagonists may drive virus evolution. Specifically, the identity of this PA amino acid (184) has previously been coupled to the ability of a different virus strain to efficiently initiation transcription and overcome a host-adaptation block in human cells via altered interaction with the host MCM complex[30,31]. Coupled with the data generated in Chapter 3, a multifactorial model of reassortment barriers, supported by extensive data, may be constructed in which many viral components must be congruent for efficient emergence of a new virus. A model for the road to reassortment and barriers that must be overcome is presented in figure 5.2. In total, the studies described herein demonstrate that both genomic RNA and encoded proteins simultaneously contribute to the potential fitness landscape of influenza A

viruses during co-infection and conditions amenable to reassortment.

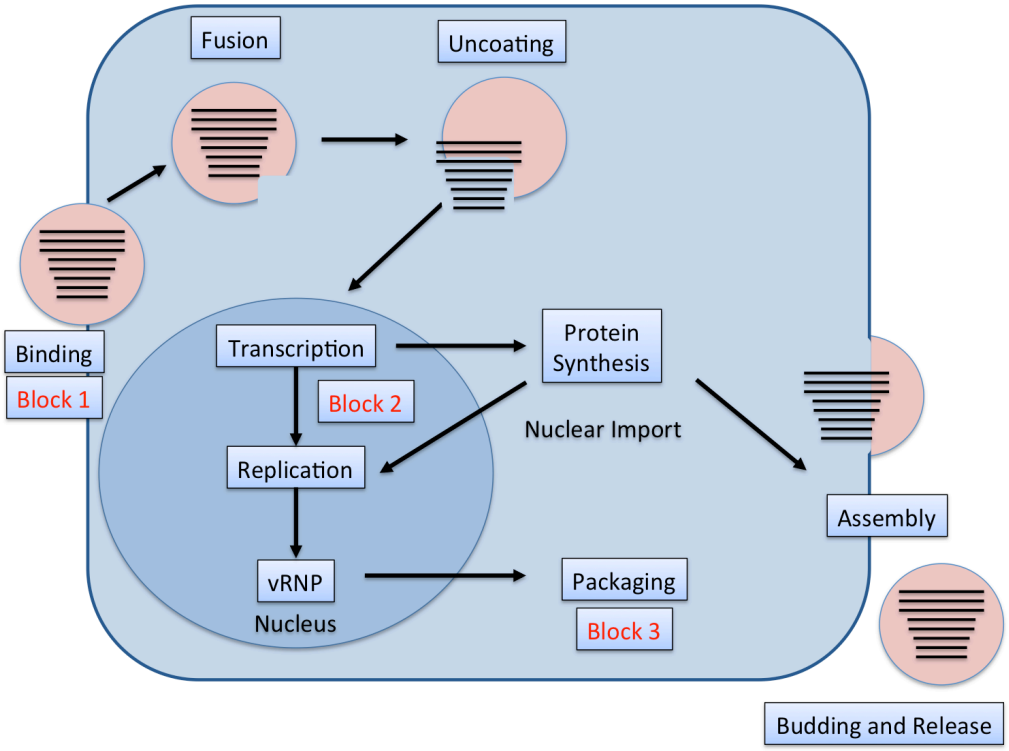


Figure 5.2. Host and Virus-intrinsic restriction of IAV replication and reassortment potential. Block 1: IAV must bind, enter, and fuse within a host cell. For this to occur, the HA molecule must recognize specific sialic-acid moieties, while evading neutralizing antibody, and fuse in the endosome at low-pH. Block 2: Transcription and replication, including nuclear import of newly synthesized viral proteins, are required for generation of progeny viruses. Host-specific factors are required for these processes and inadequate host-adaptation may preclude viral replication. Block 3: vRNP export and genome packaging are required for assembly of nascent viral particles. Efficient interaction between vRNAs are likely required for coordinated genome packaging and therefore may be a virus-intrinsic restriction of reassortment and evolution. (Figure adapted from Shi, et. al.[1]).

5.5 Recapitulating reassortment in a controlled manner

Prior studies of reassortment relied upon cellular co-infection and a model in which all genome segments from parent viruses were present, yielding 2^8 (256) possible progeny. To simplify this model and isolate the specific segment-based requirements and restrictions of reassortment potential, we generated a co-transfection model to mimic reassortment in cell culture (Figure 1.1). The resulting mixture of virus was then genotyped to determine the preferences of reassortant segment selection in the context of two distantly related viral backbones. We saw that for the H7N3 backbone, mostly avian origin segments were selected with the notable exception of three segments from the

2009 pandemic H1N1 virus. We then individually generated single gene reassortant viruses bearing seven segments of the avian virus and one segment from the pandemic virus. These viruses were characterized for their ability to replicate in tissue culture and in the lungs of mice. We found that substitution of three segments from the pandemic either did not alter the virulence of the H7N3 virus in mice. Interestingly, PB2 (segment 1) resulted in morbidity in mice surpassing that of the 2009 pandemic H1N1 itself. PB2 protein is known to be a determinant of host range and a number of protein coding changes present in the 2009 pandemic H1N1 PB2, but absent in the H7N3 PB2, allows increased replication in mammalian cells[4]. Additionally, the inclusion of NA (segment 6) or M (segment 7) segments from the 2009 pandemic H1N1 did not attenuate the replication and pathogenesis profiles of the H7N3 virus. Either together or individually, these segments from the 2009 pandemic H1N1 increased transmission of previously non-transmissible viruses[32,33]. Therefore, it may be important to know if North American avian H7N3 viruses that possess HA receptor binding signatures similar to

human viruses, and bearing one of these segments, are capable of animal-to-animal transmission[34]. Future studies to examine the potential for reassortment between close and distantly related human, swine, and canine segments may be insightful given the they are all derived from an avian origin, but replicate in distinct mammalian hosts. However, biosafety and ethical considerations must be carefully examined prior to experimentation as a result of the Pause on Gain-of-Function studies that occurred from 2014-2016. While, the basic knowledge of advantageous adaptations may enhance disease surveillance, vaccine design, or accelerate antiviral discovery, the potential for nefarious use of constructive scientific information should be weighed with future efforts to understand what genetic features enable better replication or transmission of IAV in humans.

5.6 Cartography of influenza A virus nucleoprotein-RNA interactions

How nucleoprotein interacts with viral RNA in cells prior to and during genome packaging has not been addressed comprehensively at nucleotide resolution. Our studies indicate that NP protects approximately 12 nucleotides of viral genomic RNA inside human cells, and the profile of NP-binding to viral RNA is largely non-uniform. These findings agree with prior estimates of the RNase protected footprint of NP and biochemical experiments demonstrating the minimal length of RNA required for NP oligomerization *in vitro*[35,36]. The interaction of NP and viral RNA is highly reproducible, indicating some means of phasing NP across vRNA molecules. We propose that RNA structural elements, as well as steric constraints of the vRNP itself, likely determine the interaction of NP and vRNA. Approximately 10% of the viral genome is unbound at a given time and has potential to form RNA elements, both secondary and tertiary, required for viral replication and genome packaging. NP, in either monomeric or dimeric form,

is added co-transcriptionally as PB1 synthesizes new viral RNA and coats a majority of the vRNA prior to nuclear export[37,38]. However, the addition of NP may be slower than the synthesis and processivity of PB1, allowing local RNA structure formation. In cases where a structure has a relatively stable fold, NP may be unable to deform the element and binds to the next adjacent available stretch of vRNA. The non-uniform landscape of NP-vRNA interactions suggests that these structural elements are important for the viral life cycle and corroborated by numerous mutagenesis and bioinformatics studies of sequence conservation in potentially structured regions[39–42]. An additional report of genome wide NP-vRNA association was published before the publication of our work and corroborates our findings in the context of two different virus strains[24]. A model to describe how favorable vRNA-vRNA interactions facilitate genome packaging, while sub-optimal interactions restrict genome packaging and reassortment is presented below (**Figure 5.3**) Taken together, these data suggest NP does not coat the entire IAV genome as previously hypothesized, and the unbound regions have required functions at a specific stage of infection.

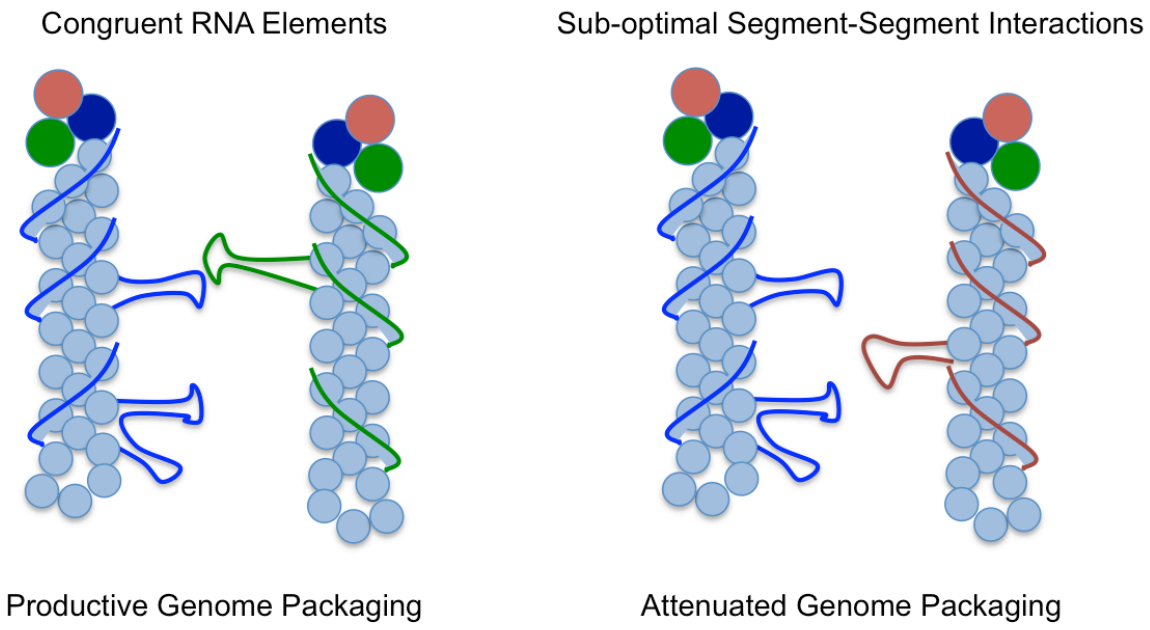


Figure 5.3. Model of vRNA-vRNA interactions required for IAV replication. IAV vRNA (blue) is bound by NP (light blue) as well as the polymerase complex of PB2 (red), PB1(navy), and PA (green). NP does not bind all vRNA, leaving a portion of the genome exposed and able to form RNA features required for formation of a trans-segment interaction network. Left: Presence of congruent RNA elements enable efficient genome packaging and potential genome reassortment. Right: Attenuation of segment-segment interactions between non-congruent RNA elements suppresses genome packaging and the potential for co-packaging and reassortment of these segments.

5.7 Probing RNA structure in the viral genome

Studies performed contemporaneously, by us and others, show the potential for NP to scaffold RNA structures, but there remains extremely limited biochemical and biophysical data for the authentic vRNA structure in the context of vRNPs. We hypothesize that certain RNA structures are present in NP unbound regions. However, biochemical evidence for any given native RNA structure to form in IAV is still lacking. Future investigations utilizing dimethyl sulfate reactivity mapping to identify what RNA bases are paired via high throughput sequencing of total viral RNA in various forms will be efficacious in determining the most populated structural elements[43]. Similarly, ongoing work in other laboratories utilizes SHAPE-seq to determine the reactivity and solvent accessibility of unpaired RNA bases[44,45]. Through personal communications, we discovered considerable correlation between our NP-unbound regions, and regions identified as unreactive to SHAPE reagents in the context of vRNPs, indicating NP does not bind highly structured elements. To elaborate on the function of these regions, we generated mutant viruses predicted to alter the fold of viral RNA in an isolated sequence. Probing if, and how, these changes in viral RNA alter RNA structure in isolation should also be undertaken. Utilization of methods including circular dichroism (CD) or small angle X-ray scattering (SAXS) may allow relative certainty that isolated NP-unbound regions do fold into specific structures and manipulation of the sequence in these regions augments the stability or structural envelope of the RNA. Together, these studies may show that NP provides an energetically and biochemically regulated platform for RNA structure formation and interaction between segments, thereby enabling complete multi-segment assembly and infectious particle production.

5.8 NP-unbound regions with no known structure

In addition to large, low-NP binding regions we identified a number of short areas in the viral genome with no known fold. Additional studies to examine the sequence requirements of these regions should be undertaken. While we looked for NP presence along the viral RNA, it is conceivable that the reproducible absence of NP is due to an additional host or viral factor that binds in a sequence specific manner at these sites. While these regions we identified appear to be of low structural complexity, a viral or host protein may compete off NP if it is present at great enough occupancy or has a significantly greater affinity for these specific sequences. Intriguingly, these short regions may also act as docking sites for larger structures from different segments through which a complex web of structured and unstructured elements interact to allow complete genome packaging.

5.9 Expanding the genomewide landscape and interaction networks for IAV vRNA

Evidence generated in our lab and others suggests vRNA features form a multi-segment interaction network[13,46,47]. Genome packaging is disrupted when mutations that are predicted to be structurally disruptive are introduced to virus, attenuating replication, and causing the loss of specific genome segments in total cell free virus, indicating interactions between may be required for genome assembly. We and others have described the interactions of NP and viral RNA in a limited set of IAV adapted to cell culture. Future studies must include a diverse array of IAV from human, swine, and avian sources in expanded studies on the NP-vRNA interaction. The nucleotide content of these viruses is unique and therefore each strain may have common as well as strain-specific RNA features protruding from the vRNP core. Analysis of additional

strain-specific elements may help elucidate how certain viruses tolerate slightly different levels of genome packaging efficacy as well as how likely it is for viruses containing congruous protein components to be generated following co-infection. In addition to mapping the scaffold of vRNA adorning the ribonucleoprotein complex, efforts should be made to biochemically assess the physical interaction network of viral RNAs with each other as well as with host RNA particularly given the association of multiple genome segments prior to packaging[20,48]. Recently, a number of proximity-dependent ligation and sequencing methods have been generated to allow access to nucleotide resolution maps of inter-RNA interaction networks[49,50]. Coupled with our knowledge of the basic NP-vRNA landscape – which enables determination of regions accessible for cross-segment interaction – specific mapping of the authentic network of sequence elements that contact one and other in cells and purified virus will yield the most complete understanding of how IAV packages a complete set of eight segments with high effectiveness. As with these studies, once a basic network map for one virus is established, it should be expanded to determine if the contact points for both close and distant relatives of the prototype virus are shared or dependent on genomic context.

5.10 Utilizing genome architecture data to monitor IAV evolution and guide treatment

Mapping the genetic structure and interaction network of IAV provides insight into the mechanisms of genome packaging and potential for reassortment. Computational models may be developed in the future to address the likelihood of productive interaction between vRNA that has been identified in nature and co-circulate geographically. In addition to the abundance of knowledge on protein features required for successful interaction of viral proteins, a predictive

model of reassortment success may be crafted to enable risk assessment of reassortant generation, if a large enough sample of virus genomes has been assessed by methods including PAR-CLIP and genome structure profiling. Additionally, therapeutics that target unstructured regions of vRNA with antisense oligos have been effectively utilized *in vitro* to target viral RNA for degradation, though the efficacy of these treatments *in vivo* has not been fully explored[24,51]. Finally, vaccination with live-attenuated viruses that retain all protein components, but reproducibly fail to replicate and package full genomes may enable generation of live-attenuated vaccines that evoke antibody as well as cell-mediated immunity[52]. Taken together, our understanding of how IAV RNA interacts between segments and if this interaction may be successfully translated to the clinic are exciting future directions.

5.11 Understanding the multifaceted viral compatibility and host restriction of IAV.

Herein, I have discussed three isolated but interconnected mechanisms that restrict the evolution and host adaptation of IAV. First, the genetic backgrounds of co-infecting viruses are a predominant driver of reassortment outcomes. Compatibility of viral proteins between strains, as well as RNA features, are capable of either enabling efficient reassortment or killing the possible formation of a new genome constellation. Second, the host environment in which co-infection occurs also dictates the potential for amplification of certain progeny genotypes. If a new reassortant combination is unable to subsequently replicate, or is attenuated relative to its parent, it will quickly disappear from the population at large. Conversely, the combination of multiple advantageous mutations, including escape from host immunity or enhanced utilization of a species-specific and required host factor, in a new constellation may enhance the ability of this

combination to spread throughout an organism, potentially leading to endemicity or pandemicity. Finally, congruence between RNA features, scaffolded by NP at the stage of genome packaging likely precludes the generation of multiple viruses despite compatible protein components. Together, these layers of complexity shape the evolutionary trajectories of IAV as well as the potential for emergence of novel genome constellations in animal and human populations. The research presented here adds to the growing understanding of how IAV successfully replicates in diverse hosts, with great and expanding genetic diversity, and elucidates some of the specific molecular mechanisms by which this pathogen continues to be an annual concern for animal and human health.

5.12 REFERENCES

1. Shi, Y., Wu, Y., Zhang, W., Qi, J. & Gao, G. F. Enabling the 'host jump': structural determinants of receptor-binding specificity in influenza A viruses. *Nat. Rev. Microbiol.* **12,** 822–31 (2014).

2. Dou, D. *et al.* Analysis of IAV Replication and Co-infection Dynamics by a Versatile RNA Viral Genome Labeling Method. *Cell Rep.* **20,** 251–263 (2017).

3. Marshall, N., Priyamvada, L., Ende, Z., Steel, J. & Lowen, A. C. Influenza virus reassortment occurs with high frequency in the absence of segment mismatch. *PLoS Pathog.* **9,** e1003421 (2013).

4. Williams, G. D., Pinto, A. K., Doll, B. & Boon, A. C. M. A North American H7N3 influenza virus supports reassortment with 2009 pandemic H1N1 and induces disease in mice without prior adaptation. *J. Virol.* **90,** JVI.02761-15 (2016).

5. Ince, W. L., Gueye-Mbaye, A., Bennink, J. R. & Yewdell, J. W. Reassortment complements spontaneous mutation in influenza A virus NP and M1 genes to accelerate adaptation to a new host. *J. Virol.* **87,** 4330–8 (2013).

6. Greenbaum, B. D., Li, O. T. W., Poon, L. L. M., Levine, A. J. & Rabadan, R. Viral reassortment as an information exchange between viral segments. *Proc. Natl. Acad. Sci. U. S. A.* **109,** 3341–6 (2012).

7. Moratorio, G. *et al.* Attenuation of RNA viruses by redirecting their evolution in sequence space. *Nat. Microbiol.* **2,** 17088 (2017).

8. Pauly, M. D., Procario, M. C. & Lauring, A. S. A novel twelve class fluctuation test reveals higher than expected mutation rates for influenza A viruses. *Elife* **6,** e26437 (2017).

9. Lauring, A. S. & Andino, R. Quasispecies theory and the behavior of RNA viruses. *PLoS Pathog.* **6,** 1–8 (2010).

10. Poon, L. L. M. *et al.* Quantifying influenza virus diversity and transmission in humans. *Nat. Genet.* **Advance On,** 8 (2016).

11. Mori, K., Murano, K., Ohniwa, R. L., Kawaguchi, A. & Nagata, K. Oseltamivir Expands Quasispecies of Influenza Virus through Cell-to-cell Transmission. *Sci. Rep.* **5,** 9163 (2015).

12. Fonville, J. M., Marshall, N., Tao, H., Steel, J. & Lowen, A. C. Influenza Virus Reassortment Is Enhanced by Semi-infectious Particles but Can Be Suppressed by Defective Interfering Particles. *PLoS Pathog.* **11,** 1–30 (2015).

13. Gavazzi, C. *et al.* An in vitro network of intermolecular interactions between viral RNA segments of an avian H5N2 influenza A virus: comparison with a human H3N2 virus. *Nucleic Acids Res.* **41,** 1241–54 (2013).

14. Gavazzi, C. *et al.* A functional sequence-specific interaction between influenza A virus genomic RNA segments. *Proc. Natl. Acad. Sci. U. S. A.* **110,** 16604–9 (2013).

15. Essere, B. *et al.* Critical role of segment-specific packaging signals in genetic reassortment of influenza A viruses. *Proc. Natl. Acad. Sci. U. S. A.* **110,** E3840-8 (2013).

16. Fournier, E. *et al.* Interaction network linking the human H3N2 influenza A virus genomic RNA segments. *Vaccine* **30,** 7359–67 (2012).

17. Noda, T. *et al.* Three-dimensional analysis of ribonucleoprotein complexes in influenza A virus. *Nat. Commun.* **3,** 639 (2012).

18. Calder, L. J., Wasilewski, S., Berriman, J. a & Rosenthal, P. B. Structural organization of a filamentous influenza A virus. *Proc. Natl. Acad. Sci. U. S. A.* **107,** 10685–90 (2010).

19. Gallagher, J. R., Torian, U., McCraw, D. M. & Harris, A. K. Structural studies of influenza virus RNPs by electron microscopy indicate molecular contortions within NP supra-structures. *J. Struct. Biol.* (2016). doi:10.1016/j.jsb.2016.12.007

20. Chou, Y. *et al.* Colocalization of different influenza viral RNA segments in the cytoplasm before viral budding as shown by single-molecule sensitivity FISH analysis. *PLoS Pathog.* **9,** e1003358 (2013).

21. Fujii, Y., Goto, H., Watanabe, T., Yoshida, T. & Kawaoka, Y. Selective incorporation of influenza virus RNA segments into virions. *Proc. Natl. Acad. Sci. U. S. A.* **100,** 2002–7 (2003).

22. Compans, R. W., Content, J. & Duesberg, P. H. Structure of the ribonucleoprotein of influenza virus. *J. Virol.* **10,** 795–800 (1972).

23. Baudin, F., Bach, C., Cusack, S. & Ruigrok, R. W. Structure of influenza virus RNP. I. Influenza virus nucleoprotein melts secondary structure in panhandle RNA and exposes the bases to the solvent. *EMBO J.* **13,** 3158–65 (1994).

24. Lee, N. *et al.* Genome-wide analysis of influenza viral RNA and nucleoprotein association. *Nucleic Acids Res.* **5,** e1000491 (2017).

25. Mehle, A. & Doudna, J. a. Adaptive strategies of the influenza virus polymerase for replication in humans. *Proc. Natl. Acad. Sci. U. S. A.* **106,** 21312–6 (2009).

26. Mehle, A., Dugan, V. G., Taubenberger, J. K. & Doudna, J. A. Reassortment and Mutation of the Avian Influenza Virus Polymerase PA Subunit Overcome Species Barriers. (2012). doi:10.1128/JVI.06203-11

27. Weber, M. *et al.* Influenza virus adaptation PB2-627K modulates nucleocapsid inhibition by the pathogen sensor RIG-I. *Cell Host Microbe* **17,** 309–19 (2015).

28. Naffakh, N., Tomoiu, A., Rameix-Welti, M.-A. & van der Werf, S. Host restriction of avian influenza viruses at the level of the ribonucleoproteins. *Annu. Rev. Microbiol.* **62,** 403–24 (2008).

29. Long, J. S. *et al.* Species difference in ANP32A underlies influenza A virus polymerase host restriction. *Nature* **529,** 101–104 (2016).

30. Kawaguchi, A. & Nagata, K. De novo replication of the influenza virus RNA genome is regulated by DNA replicative helicase, MCM. *Embo J* **26,** 4566–4575 (2007).

31. Hara, K., Nakazono, Y., Kashiwagi, T., Hamada, N. & Watanabe, H. Co-incorporation of the PB2 and PA polymerase subunits from human H3N2 influenza virus is a critical determinant of the replication of reassortant ribonucleoprotein complexes. *J. Gen. Virol.* **94,** 2406–2416 (2013).

32. Chou, Y. *et al.* The M segment of the 2009 new pandemic H1N1 influenza virus is critical for its high transmission efficiency in the guinea pig model. *J. Virol.* **85,** 11235–41 (2011).

33. Lakdawala, S. S. *et al.* Eurasian-origin gene segments contribute to the transmissibility, aerosol release, and morphology of the 2009 pandemic H1N1 influenza virus. *PLoS Pathog.* **7,** (2011).

34. Belser, J. a *et al.* Contemporary North American influenza H7 viruses possess human receptor specificity: Implications for virus transmissibility. *Proc. Natl. Acad. Sci. U. S. A.* **105,** 7558–63 (2008).

35. Resa-Infante, P., Jorba, N., Coloma, R. & Ortin, J. The influenza virus RNA synthesis machine: Advances in its structure and function. *RNA Biol.* **8,** 207–215 (2011).

36. Labaronne, A. *et al.* Binding of RNA by the nucleoproteins of influenza viruses A and B. *Viruses* **8,** 1–14 (2016).

37. Pflug, A., Guilligay, D., Reich, S. & Cusack, S. Structure of influenza A polymerase bound to the viral RNA promoter. *Nature* **516,** 355–60 (2014).

38. Reich, S., Guilligay, D. & Cusack, S. An in vitro fluorescence based study of initiation of RNA synthesis by influenza B polymerase. *Nucleic Acids Res.* **45,** 3353–3368 (2016).

39. Gultyaev, A. P. *et al.* RNA structural constraints in the evolution of the influenza A virus genome NP segment. *RNA Biol* **11,** 942–952 (2014).

40. Kobayashi, Y., Dadonaite, B., Doremalen, N. Van, Barclay, W. S. & Pybus, O. G.

Computational and molecular analysis of conserved influenza A virus RNA secondary structures involved in infectious virion production. *RNA Biol.* **13,** 883–894 (2016).

41. Moss, W. N., Priore, S. F. & Turner, D. H. Identification of potential conserved RNA secondary structure throughout influenza A coding regions. 991–1011 (2011). doi:10.1261/rna.2619511.(IRES)

42. Gog, J. R. *et al.* Codon conservation in the influenza A virus genome defines RNA packaging signals. **35,** 1897–1907 (2007).

43. Zubradt, M. *et al.* DMS-MaPseq for genome-wide or targeted RNA structure probing in vivo. *Nat. Methods* **14,** 75–82 (2016).

44. Loughrey, D., Watters, K. E., Settle, A. H. & Lucks, J. B. SHAPE-Seq 2.0: systematic optimization and extension of high-throughput chemical probing of RNA secondary structure with next generation sequencing. *Nucleic Acids Res.* **42,** 1–10 (2014).

45. Spitale, R. C. *et al.* RNA SHAPE analysis in living cells. *Nat. Chem. Biol.* **9,** 18–20 (2013).

46. Gerber, M., Isel, C., Moules, V. & Marquet, R. Selective packaging of the influenza A genome and consequences for genetic reassortment. *Trends Microbiol.* **22,** 446–455 (2014).

47. Gao, Q. *et al.* The influenza A virus PB2, PA, NP, and M segments play a pivotal role during genome packaging. *J. Virol.* **86,** 7043–51 (2012).

48. Lakdawala, S. S. *et al.* Influenza a virus assembly intermediates fuse in the cytoplasm.

PLoS Pathog. **10,** e1003971 (2014).

49. Sharma, E., Sterne-Weiler, T., O'Hanlon, D. & Blencowe, B. J. Global Mapping of Human RNA-RNA Interactions. *Mol. Cell* **62,** 618–626 (2016).

50. Lu, Z. *et al.* RNA Duplex Map in Living Cells Reveals Higher-Order Transcriptome Structure. *Cell* **165,** 1267–1279 (2016).

51. Giannecchini, S. *et al.* Packaging signals in the 5'-ends of influenza virus PA, PB1, and PB2 genes as potential targets to develop nucleic-acid based antiviral molecules. *Antiviral Res.* **92,** 64–72 (2011).

52. Nogales, A. *et al.* Influenza A Virus Attenuation by Codon Deoptimization of the NS Gene for Vaccine Development. *J. Virol.* **88,** 10525–10540 (2014).